



Link recommendation algorithms and dynamics of polarization in online social networks

Fernando P. Santos^{a,b,1} , Yphtach Lelkes^c , and Simon A. Levin^a

^aDepartment of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544; ^bInformatics Institute, University of Amsterdam, 1098XH Amsterdam, The Netherlands; and ^cAnnenberg School for Communication Research, University of Pennsylvania, Philadelphia, PA 19104

Edited by Charles Perrings, Arizona State University, Tempe, AZ, and accepted by Editorial Board Member Kenneth W. Wachter July 14, 2021 (received for review March 1, 2021)

The level of antagonism between political groups has risen in the past years. Supporters of a given party increasingly dislike members of the opposing group and avoid intergroup interactions, leading to homophilic social networks. While new connections offline are driven largely by human decisions, new connections on online social platforms are intermediated by link recommendation algorithms, e.g., “People you may know” or “Whom to follow” suggestions. The long-term impacts of link recommendation in polarization are unclear, particularly as exposure to opposing viewpoints has a dual effect: Connections with out-group members can lead to opinion convergence and prevent group polarization or further separate opinions. Here, we provide a complex adaptive-systems perspective on the effects of link recommendation algorithms. While several models justify polarization through rewiring based on opinion similarity, here we explain it through rewiring grounded in structural similarity—defined as similarity based on network properties. We observe that preferentially establishing links with structurally similar nodes (i.e., sharing many neighbors) results in network topologies that are amenable to opinion polarization. Hence, polarization occurs not because of a desire to shield oneself from disagreeable attitudes but, instead, due to the creation of inadvertent echo chambers. When networks are composed of nodes that react differently to out-group contacts, either converging or polarizing, we find that connecting structurally dissimilar nodes moderates opinions. Overall, our study sheds light on the impacts of social-network algorithms and unveils avenues to steer dynamics of radicalization and polarization in online social networks.

polarization | social networks | complex systems | link recommendation | opinion dynamics

Online social networks are increasingly used to access political information (1), engage with political elites, and discuss politics (2). These new communication platforms can benefit democratic processes in several ways: They reduce barriers to information and, subsequently, increase citizen engagement, allow individuals to voice their concerns, help debunk false information, and improve accountability and transparency in political decision-making (3). In principle, individuals can use social media to access ideologically diverse viewpoints and make better-informed decisions (4, 5).

At the same time, internet and online social networks reveal a dark side. There are mounting concerns over possible linkages between social media and affective polarization (6, 7). Other than healthy political deliberation, social networks can foster so-called “echo chambers” (8, 9) and “information cocoons” (3, 10) where individuals are only exposed to like-minded peers and homogeneous sources of information, which polarizes attitudes (for counterevidence, see ref. 5). As a result, social media can trigger political sectarianism (6, 7, 11–13) and fuel misinformation (14, 15). Averting the risks of online social networks for political institutions, and potentiating their advantages, requires multidisciplinary approaches and novel methods to understand long-term dynamics on social platforms.

That is not an easy task. As pointed out by Woolley and Howard, “to understand contemporary political communication we must now investigate the politics of algorithms and automation” (16). While traditional media outlets are curated by humans, online social media resorts to computer algorithms to personalize contents through automatic filtering. To understand information dynamics in online social networks, one needs to take into account the interrelated subtleties of human decision making [e.g., only share specific contents (17), actively engage with other users, follow or befriend particular individuals, interact offline] and the outcomes of automated decisions (e.g., news sorting and recommendation systems) (18, 19). In this regard, much attention has been placed on the role of news filters and sorting (1, 18, 19). Shmargad and Klar (20) provide evidence that algorithms sorting news impact the way users engage with and evaluate political news, likely exacerbating political polarization. Likewise, Levy (21) notes that social media algorithms can substantially affect users’ news consumption habits.

While past studies have examined how algorithms may affect which information appears on a person’s newsfeed, and subsequent polarization, social matching (22) or link recommendation (23) algorithms [also called user, contact, or people recommender systems (24, 25)] constitute another class of algorithms that can affect the way users engage in (and with) online social networks (examples of such systems in *SI Appendix*, Fig. S13).

Significance

Polarization is rising while political debates are moving to online social platforms. In such settings, algorithms are used to recommend new connections to users, through so-called link recommendation algorithms. Users are often recommended based on structural similarity (e.g., nodes sharing many neighbors are similar). We show that preferentially establishing links with structurally similar nodes potentiates opinion polarization by stimulating network topologies with well-defined communities (even in the absence of opinion-based rewiring). When networks are composed of nodes that react differently to out-group contacts—either converging or polarizing—connecting structurally dissimilar nodes enhances moderate opinions. Our study sheds light on the impacts of social-network algorithms in opinion dynamics and unveils avenues to steer polarization in online social networks.

Author contributions: F.P.S., Y.L., and S.A.L. designed research; F.P.S., Y.L., and S.A.L. performed research; F.P.S. analyzed data; and F.P.S., Y.L., and S.A.L. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. C.P. is a guest editor invited by the Editorial Board.

Published under the PNAS license.

¹To whom correspondence may be addressed. Email: f.p.santos@uva.nl.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2102141118/DCSupplemental>.

Published December 6, 2021.

These algorithms are implemented to recommend new online connections—“friends” or “followees”—to social network users, based on supposed offline familiarity, likelihood of establishing a future relation, similar interests, or the potential to serve as a source of useful information. Current data provide evidence that link recommendation algorithms impact network topologies and increase network clustering: Daly et al. (26) show that an algorithm recommending friends-of-friends, in an IBM internal social network platform, increases clustering and network modularity. Su et al. (27) analyzed the Twitter graph before and after this platform implemented link recommendation algorithms and show that the “Who To Follow” feature led to a sudden increase in edge growth and the network clustering coefficient. Similarly, Zignani et al. (28) show that, on a small sample of the Facebook graph, the introduction of the “People You May Know” (PYMK) feature led to a sudden increase in the number of links and triangles [i.e., motifs comprising three nodes (A, B, C) where the links AB, AC, and BC exist] in the network. The fact that PYMK is responsible for a significant fraction of link creations is alluded to in other works (29). Furthermore, recent work shows, through experiments with real social media users (30) and simulations (31), that link recommendation algorithms can effectively be used as an intervention mechanism to increase networks’ structural diversity (30, 31) and minimize disagreements (32). It is thereby relevant to understand, 1) How do algorithmic link recommendations interplay with opinion formation? and 2) What are the long-term impacts of such algorithms on opinion polarization?

Here, we tackle the previous questions from a complex adaptive-systems perspective (33), designing and analyzing a simple model where individuals interact in a dynamic social network. While several models explain the emergence of polarization through link formation based on opinion similarity (34–41) and information exchange (42), here we focus instead on rewiring based on “structural similarity,” which is defined as similarity based on common features that exclusively depend on the network structure (43). This contrasts with the broader concept of homophily, which typically refers to similarity based on common characteristics besides network properties (e.g., opinions, taste, age, background). Compared with rewiring based on homophily—which can also contribute to network fragmentation—rewiring based on structural similarity can be less restrictive in contexts where information about opinions and beliefs is not readily available to individuals before the connection is established. Furthermore, rewiring based on structural similarity is a backbone of link recommendation algorithms [e.g., “People you may know” or “Whom to follow” (25) suggestions], which rely on link prediction methods to suggest connections to users (43, 44). Importantly, our model combines three key ingredients: 1) Links are formed according to structural similarity, based on common neighbors, which is one of the simplest link prediction methods (43); this way, we do not assume a priori that individuals with similar opinions are likely to become connected [as recent works underline, sorting can be incidental to politics (45, 46)]. 2) Then, to examine opinion updating, we adapt a recent model that covers the interplay of social reinforcement and issue controversy to promote radicalization on social networks (39). 3) Last, we explicitly consider that nodes can react differently to out-group links, either converging in their opinions (10, 47) or polarizing further (48–50).

We find that establishing links based on structural similarity alone [a process that is likely to be reinforced by link recommendation algorithms—see *SI Appendix*, Fig. S10 and previous work pointing that such algorithms affect a social network topology and increase their clustering coefficient (26–28)] contributes to opinion polarization. While our model sheds light on the effect of link recommendation algorithms on opinion formation and polarization dynamics, we also offer a justification

for polarization to emerge through structural similarity-based rewiring, in the absence of explicit opinion-similarity rewiring (34, 36, 39, 51), confidence-bounds (37, 38, 40), or rewiring based on concordant messages (42).^{*} Second, we find that the effects of structural similarity-based rewiring are exacerbated if even moderate opinions have high social influence. Finally, we combine nodes that react differently to out-group contacts: “converging” nodes, which converge if exposed to different opinions (10, 21, 52), and “polarizing” nodes, which diverge when exposed to different viewpoints (48–50). We observe that the coexistence of both types of nodes can contribute to moderate opinions. Polarizing nodes develop radical opinions, and converging nodes, influenced by opposing viewpoints, yield more temperate ones. However, again, link recommendation algorithms impact this process: Given the existence of communities isolated to a greater degree through link recommendation, converging nodes may find it harder to access diverse viewpoints, which, in general, contributes to increasing the adoption of extreme opinions.

Results

As fully described in *Materials and Methods*, we simulate a networked population of N individuals that adapt their opinions through social influence. Each individual i is characterized by an opinion $x_i \in (-\infty, +\infty)$. The sign of x_i represents individuals’ stance toward a certain topic (e.g., pro or against gun control). The process of opinion updating that we consider is inspired by recent models developed to capture dynamics of polarization on social media platforms (39, 53) and by earlier models of collective decision-making (54, 55) and opinion formation (56, 57). We consider that the opinion of individual i at time $t+1$ is updated following $x_i(t+1) = \gamma x_i(t) + K \sum_j^N A_{ij} \tanh(\alpha x_j(t)) / k_i$. This reflects the assumption that in the absence of social reinforcement, opinions’ magnitudes decay with factor $0 < \gamma < 1$. This means that we do not include intrinsic preferences for a specific opinion—a condition that can be altered in future iterations of the model, by assuming that certain individuals have a baseline tendency to adopt more extreme opinions over time, $\gamma > 1$, or are strongly opinionated (58) by default. Furthermore, $\gamma < 1$ prevents the magnitude of opinions from growing without bounds. We assume that opinions are updated based on the average influence of neighbors in an unweighted network and hence the division by k_i , the degree of individual i . The hyperbolic tangent (\tanh) introduces nonlinearity on the social impact of opinions and imposes a bound on the level of influence that individuals have on each other $[-1 < \tanh(x) < 1]$. α controls how one’s opinion translates into social influence: If $\alpha = 0$, the opinions of others are non-influential, even if they are extreme. For high α , even moderate opinions (low $|x|$) have a strong social influence. If α is low, only extreme opinions (high $|x|$) have a meaningful impact in social dynamics. In refs. 39 and 53, a higher α is associated to a more controversial topic, which individuals are more passionate about, more attentive to and, as a result, more susceptible to be socially influenced.

Individuals are influenced by their neighbors, according to an adjacency matrix A . K represents the social interaction strength. At this point, we can formalize the concept of “polarization” and “radicalization” as we will use throughout the manuscript: We assume that the polarization level of a population (P) is given by the SD in opinions [$P = SD(x_0, \dots, x_N)$] and the radicalization level of individual i (R_i) is given by the absolute value of its

^{*}Alternative ways of forming links that do not rely directly on opinions include checking news sources and breaking ties with neighbors that likely fuel misinformation, as also studied in this Special Feature (52).

opinion ($R_i = |x_i|$)—the radicalization level of a population (R) is given by the mean of absolute opinions ($R = \sum_i^N R_i/N$).

The networks determining social contacts vary over time. Individuals can break and form new ties. While this process depends on individual decision-making, nowadays, as mentioned in the Introduction, online social networks use link recommendation algorithms to suggest new connections (22). The number of common neighbors is perhaps the simplest measure to predict a missing link between two nodes and recommend it (43, 44). Inspired by this quantity, we assume that i will form a new link with j with a probability that depends on structural similarity, $s_{ij} = |N_i \cap N_j|^\eta / \sum_k^N |N_i \cap N_k|^\eta$. The set of neighbors of i (and respectively, j and k) is represented by N_i (N_j and N_k) and the number of common neighbors between i and j is $|N_i \cap N_j|$. At each time step, links will be removed randomly, and new links will be created with a probability proportional to s_{ij} . The parameter η is central in our analysis. It determines how structural similarity impacts the formation of new links. If $\eta = 0$, links will be added regardless of structural similarity. If $\eta > 0$, the probability of forming a link between i and j will be an increasing function of the structural similarity between i and j . To allow links between independent modules to occasionally be formed, we add a noise term ϵ (detailed in *Materials and Methods*). While previous works show that interactions based on homophily (34, 35, 39) and confidence bounds (37, 38, 40) can trigger polarization, here we observe that forming links based on structural similarity provides an alternative route to the emergence of polarization. We report results for rewiring based on the absolute number of common neighbors, yet we observed similar results with rewiring based on the Adamic–Adar Index and the Resource Allocation Index (43), both penalizing the contribution of high-degree common neighbors for the similarly index.

We start by noting that, as observed in previous work (39), the impact of opinions on social influence (α) determines whether individuals reach a state of neutral consensus or radicalization. As Fig. 1A shows, if only extreme opinions exert social influence (low α), the population will converge to a state of neutral consensus, where every agent is characterized by an opinion close to $x_i = 0$. If moderate opinions exert strong social influence (high α), the absolute value of opinions will grow over time (i.e., opinions will radicalize). This is evident in Fig. 1B. If rewiring is not strongly impacted by structural similarity, the population will converge to the same opinion. If rewiring is based on structural similarity (high η), different independent modules emerge, and each can converge to a different opinion (Fig. 1C). Such a rewiring process—which to some extent aims at capturing the effect of link recommendation algorithms—generates a social network topology, highly clustered, that is amenable to sustaining polarization.

While in Fig. 1 we focus on the examples of particular simulations, in Fig. 2 we can observe that the previous observations extend to a wide range of (α, η) and remain if simulations are repeated for different initial conditions. We observe that, if both α and η are high, radicalization and polarization are likely to emerge.

Rewiring based on structural similarity leads to polarization, as well-defined communities are likely to emerge. This is evident in Fig. 3A, where we explicitly measure the average number of connected components visible in the last time step of our simulations. As Fig. 3B reveals, if η is low, a single connected component prevails. For intermediate values of η ($\eta = 2$) the number of connected components is maximized. We observe many small cliques (size 2) and a few large components. For $\eta > 2$, we observe fewer and larger connected components. Our conclusions extend to connected networks, where isolated (yet connected) modules are present (*SI Appendix*, Figs. S1, S2, S4, S7, S8, and

S9). In such cases, we never remove links that, if detached, would lead to disconnected components.

So far, we have been considering that links connecting individuals with opposing viewpoints lead to convergence in opinions (converging nodes). On the other hand, links connecting individuals with the same opinions lead to social reinforcement and opinion radicalization. This is in agreement with previous results on intergroup contact theory (52) and group polarization (47, 59). Recently, however, research points out that, on social media, exposure to opposing viewpoints can in fact lead opinions to diverge further (48, 49, 60, 61); along the same lines, earlier works show that encouraging individuals to take out-group perspectives can increase intolerance and dislike toward out-groups (50). This indicates that, when exposed to opposing opinions, individuals can either converge (which we will refer to as “converging nodes”) or diverge (“polarizing nodes”). To capture the dynamics resulting from both types of nodes coexisting in a population—and investigate the impact of link recommendation in such settings—we now consider that a node will be a “converging” node with probability ρ and will be a “polarizing” node with probability $1 - \rho$. The class of each node is attributed initially, and it is fixed over time. Polarizing nodes will update their

$$x_i(t+1) = \rho x_i(t) + K \sum_j^N A_{ij} \tanh(\alpha \sigma_i \sigma_j x_j(t)) / k_i,$$

where σ_i corresponds to the sign of individual i 's opinion. In Fig. 4, we can observe that, if all nodes are polarizing ($\rho = 0$), the population will both radicalize and polarize. For intermediate values of ρ (e.g., $\rho \approx 0.5$), rewiring links based on structural similarity is likely to exacerbate both polarization and radicalization. In these contexts, connecting dissimilar nodes can reduce polarization and radicalization.

This effect is further investigated in Fig. 5. If all nodes are converging nodes ($\rho = 1$) and η is low, the population will converge to radical opinions (Fig. 5A). If all nodes are polarizing ($\rho = 0$), the population will evolve to a polarization state, where positive and negative radical opinions will coexist in the population regardless η (Fig. 5B). More interesting scenarios are observed when converging and polarizing nodes coexist in the population ($\rho = 0.5$). Depending on initial opinions, some individuals will become polarized through outgroup contacts (48), and others will converge through outgroup contacts (52). Polarizing nodes will sustain the coexistence of different (radical) viewpoints in the network. This coexistence will drive converging nodes to moderate their opinions. In Fig. 5C, this is visible through a mass of individuals evolving to more radical opinions (negative and positive) and a mass of individuals adopting moderate opinions. Related with this observation, a paper in the present Special Feature (62) also notes that repulsive extremists can contribute for maintaining a moderate majority. For low η , this process occurs globally. For high η this process will occur locally, within “communities.” Isolated modules will likely be composed of a mixture of converging and polarizing nodes (represented through circles and squares, respectively, in the bottom panels of Fig. 5). Importantly, the moderate (converging) individuals may find it harder to access a diverse pool of radical opinions, when interacting within isolated groups. They will themselves become radical. As a result, high η will, on average, lead to more radical opinions when there is a mixture of polarizing and converging nodes. We observe that heterogeneity on how individuals react to opposing views (either converging or polarizing) provides an opportunity for link recommendation algorithms to attenuate both polarization and radicalization. A rewiring process that does not only rely on structural similarity (low η) can lead individuals to sustain temperate viewpoints and reduce polarization. We provide additional intuition about this process in *SI Appendix*, Fig. S3.

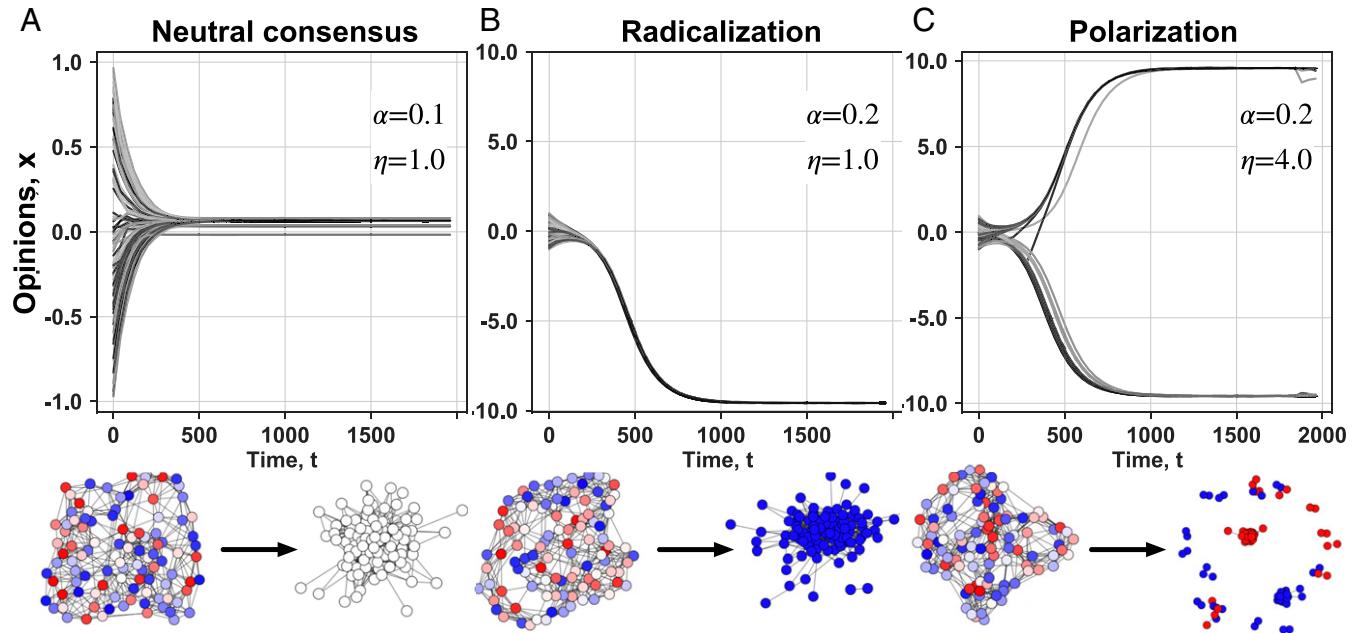


Fig. 1. The impact of social influence (α) and rewiring based on structural similarity (η) in opinion dynamics. On Top, time-series representing the evolving opinions of $N = 100$ individuals, each curve corresponding to one individual (with some overlap between lines, denoting individuals with the same opinion). Initial opinions are sampled with uniform probability from $[-1, 1]$. With low α , all opinions converge to $x \approx 0$ (A). With high α , opinions converge to extreme values, $|x| \rightarrow K/(1 - \gamma)$ (B). In a network where links are randomly rewired (low η), all individuals tend to converge to the same opinion. Increasing η leads to isolated modules that sustain diverging opinions (C). In the Bottom, we represent the initial [Watts-Strogatz network (91) with $p = 0.1$] and the final networks for the representative simulations. Node colors represent opinions: Red is a positive opinion $x > 0$ and blue is a negative opinion $x < 0$. Color intensity increases with $|x|$; white is a neutral opinion, $x = 0$. Other parameters (all detailed in *SI Appendix, Table S1*): $N = 100$, $\epsilon = 0.00$, $K = 0.1$, $\gamma = 0.99$, $\langle k \rangle = 10$, $\rho = 1.0$.

Discussion

Political opinions are increasingly shaped by interactions in online social networks. Importantly, to understand opinion dynamics in such settings, we must consider the combined effect of human social influence and algorithmic decisions. Link recommendation algorithms can possibly bias the creation of new ties

toward individuals that have a high number of common acquaintances, which constitutes perhaps the simplest form of link prediction (43, 44, 63). Here, we show that social media algorithms that recommend that users establish a link based on structural similarity (25–27, 64), potentiate opinion polarization. We observe that this effect is exacerbated if issues are such that

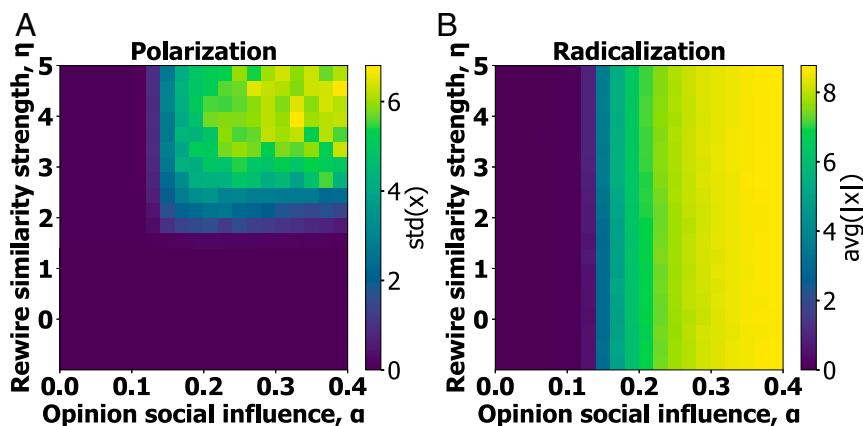


Fig. 2. Polarization emerges for high values of social influence (α) and high rewire based on structural similarity strength (η). As introduced in Fig. 1, high η leads to well-defined, isolated, modules to emerge in the network. Such modules can sustain opposing opinions. Here, we observe that this conclusion extends to a wide interval of α and η . We can identify a region of low polarization and low radicalization ($\alpha < 0.15$) where a neutral consensus is reached. If $\alpha > 0.15$ and $\eta < 2$ (approximately), we observe the low-polarization, high-radicalization regime represented in Fig. 1B. If $\alpha > 0.15$ and $\eta > 2$ (approximately), we observe the high-polarization, high-radicalization regime represented in Fig. 1C, where opinions in the population converge to a distribution with two well-defined peaks. (A) The average SD of opinions in the population. (B) The average absolute opinion values. Averages are taken over the last 20% of time steps and over 100 independent runs, starting in random initial conditions where the opinion of each individual is independently and uniformly sampled from the interval $[-1, 1]$. Other parameters (*SI Appendix, Table S1*): $N = 100$, $\epsilon = 0.01$, $K = 0.1$, $\gamma = 0.99$, $\langle k \rangle = 10$, $\rho = 1.0$.

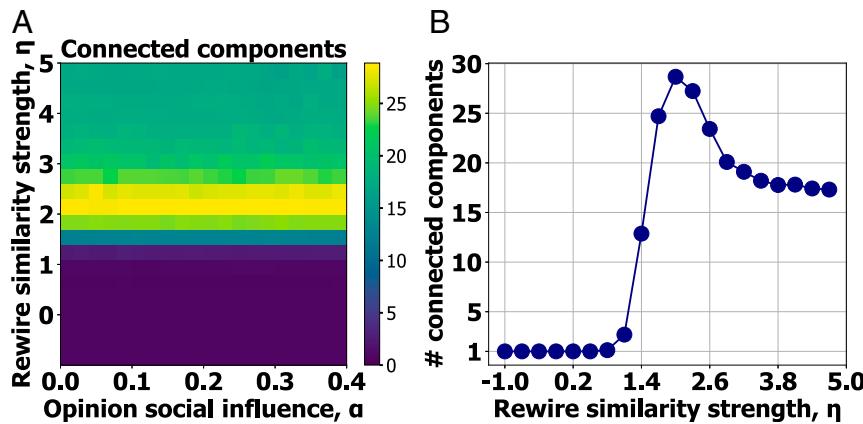


Fig. 3. Increasing η leads to independent modules in the social network. As individuals connect preferentially with those structurally similar to them—which here is measured as a function of common neighbors—disconnected components are likely to appear. Nodes in a given component do not share neighbors with nodes in other components, and links are only established across disconnected components with a small probability (proportional to ϵ). Such independent components drive the dynamics of polarization observed in Fig. 1C. In A, yellow/green represent a higher number of connected components. We observe that high η leads to a larger number of connected components; this effect is nonlinear, though. This can be better observed in B, where we fix $\alpha = 0.15$. The results reported refer to the average number of components in the final networks obtained over 100 independent runs, each starting with a randomly generated network (Watts–Strogatz small-world network with initial rewire probability 0.1). In *SI Appendix*, Fig. S5, we explore different network sizes and average degree; in *SI Appendix*, Fig. S7, we explore different initial network topologies. Other parameters (*SI Appendix*, Table S1): $N = 100$, $\epsilon = 0.01$, $K = 0.1$, $\gamma = 0.99$, $\langle k \rangle = 10$, $\rho = 1.0$.

even moderate opinions are likely to have high social influence (high α). The consequences of adding new ties in a social network depend on how individuals react to outgroup others. This said, we explicitly combine polarizing and converging nodes that, respectively, diverge or converge in opinions when exposed to different viewpoints. We show that the coexistence of both types of nodes contributes to moderate opinions. By stimulating the existence of isolated communities, rewiring based on structural similarity may prevent converging nodes from accessing diverse viewpoints, which, in general, contributes to increased levels of extreme views.

A natural question to ask at this point is whether link recommendation algorithms have a real impact in online social networks' topology. This naturally depends on the willingness of users to follow algorithmic recommendations and how that process compares with many other heuristics to form connections (e.g.,

linking based on personal interests or personal recommendations by offline friends). Future experiments in this context are in order. However, to better motivate the impactful role of link recommendation and the relevance of our analysis, in *SI Appendix*, Fig. S10 we provide empirical data [based on a dataset explored in a previous work (65)] revealing that the introduction of such algorithms changes the rewiring pattern of networks and favors the creating of links between structurally similar nodes. Other empirical studies of interest point out that social networks often reveal a fragmented shape with tight-knit communities (66–70), which constitutes a setting over which establishing links based on structural similarity is likely to be more impactful. Regarding the evolution of intra- versus intercommunity links, previous work provides data showing that the community structure of online social networks tends to be reinforced, becoming more fragmented over time (71). Furthermore, empirical data on

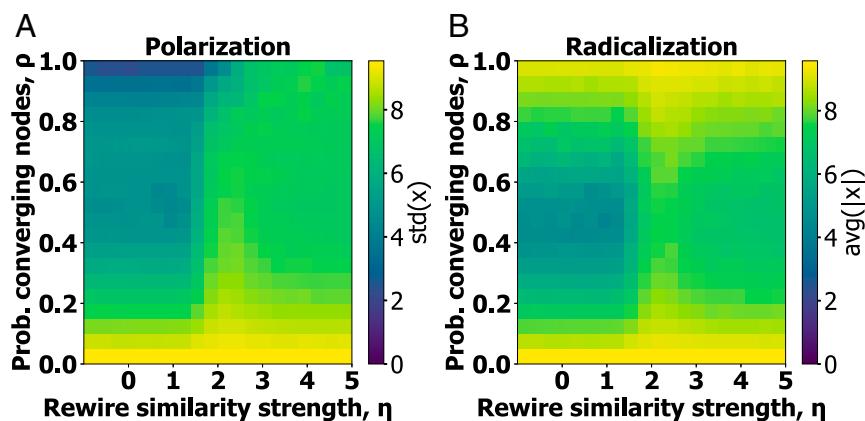


Fig. 4. Effect of network similarity-based rewiring when individuals can react differently to opposing views: polarizing or converging. We consider that nodes can belong to two different classes: i) “polarizing” when interacting with opposing opinions or ii) “converging” when interacting with opposing opinions. Nodes are “converging nodes” with a probability ρ and “polarizing nodes” with a probability $1 - \rho$. Here, ρ is represented in the vertical axis. The horizontal axis represents η , the strength of similarity-based rewiring. We can observe that, for intermediate values of ρ ($\rho=0.5$) increasing η increases both polarization (A) and radicalization (B). In *SI Appendix*, Fig. S4, we observe that this also holds if networks remain connected. The results reported refer to an average over the last 20% time steps of 100 independent runs, each starting from random initial conditions where the opinions are independently and uniformly sampled from the interval $[-1, 1]$. Other parameters (*SI Appendix*, Table S1): $N = 100$, $\epsilon = 0.01$, $K = 0.1$, $\gamma = 0.99$, $\langle k \rangle = 10$, $\alpha = 0.2$.

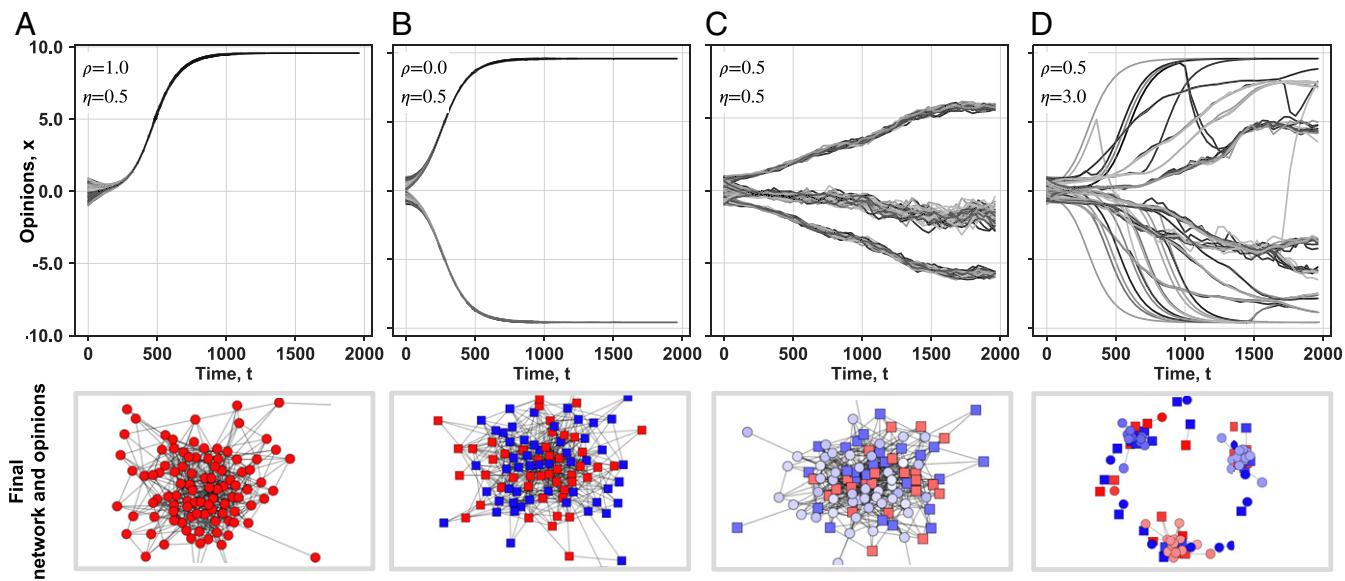


Fig. 5. Impact of network similarity-based rewiring in populations composed of converging and polarizing nodes. On *Top*, time series representing the evolving opinions of $N = 100$ individuals. On the *Bottom*, each graph represents the resulting social networks, opinions (following the color code detailed in Fig. 1), and class of each node: Circles represent “converging nodes,” and squares represent “polarizing nodes”. (A) If all nodes are converging ($\rho = 1.0$) and α is high, we typically obtain a radicalization dynamic as detailed in Fig. 2B. If all nodes are polarizing ($\rho = 0.0$), we observe polarization (B). If half of the nodes are polarizing and half are converging ($\rho = 0.5$), the ensuing dynamics depends strongly on η : Under low η (and with just one large, connected component), we can clearly distinguish polarizing nodes converging to more radical opinions and a mass of converging nodes adopting moderate opinions (C). For high η , the scenario is more complex: Isolated modules, possibly composed of polarizing and converging nodes, reveal distinguishing dynamics. In certain communities, converging nodes will not find a diversity of opinions and so they will get more extreme. As this depends strongly on each community composition, we observe the highly heterogeneous outcome represented in D. Other parameters (SI Appendix, Table S1): $N = 100$, $\epsilon = 0.01$, $K = 0.1$, $\gamma = 0.99$, $\langle k \rangle = 10$, $\alpha = 0.2$.

network dynamics and opinions during the Hong Kong occupy movement reveal that social network fragmentation precedes opinion polarization (72). On a final note concerning connections between our work and empirical data, while here we opted to confirm our main conclusions in a wide parameter space future research can be developed to estimate precisely the parameters to apply our model to specific contexts: Network parameters (N and $\langle k \rangle$) can be directly set to the number of nodes and links of a specific social network; individual behavior parameters (ρ, γ, α, K) require longitudinal data on opinion evolution to assess how fast—and in which direction—individuals alter their opinions when in contact with others’ viewpoints; alternatively, experiments (such as those reported in refs. 47–50, 59) can be conducted. Finally, network dynamics parameters (η, ϵ) can be estimated from longitudinal network data [or, once again, experiments (51)] and would require data on who initiated a connection in the first place. In *SI Appendix*, Fig. S10, we provide an example of how to visualize whether link-rewiring algorithms have a positive or negative impact on η (a parameter that correlates with the Jaccard similarity metric we explicitly compute).

Our simple model has limitations that hopefully can trigger future extensions. First, we focus on a one-dimensional opinion space. In reality, considering multiple dimensions can allow us to study phenomena of ideological alignment, dimensionality reduction in opinions, ideological consistency (53, 73), and the potential role of link recommendations in such processes. Furthermore, here we do not focus on behavioral changes accompanying opinion dynamics. Recent results indicate that opinion divergence can evolve concurrently with outgroup bias (12, 13), which suggests combining opinions with dynamics on strategies, norms, reputations, and out-group/ingroup cooperation (74–77)—in line with some works in this Special Feature on polarization, cooperation, and group identity (78–80). While here we determine exogenously which nodes are polarizing or converging,

future extensions can focus on explaining the mechanisms that determine why different individuals may react differently to outgroup opinions—an attitude that can depend on individuals’ experience in contacting with diverse viewpoints (49) or on information quality, abundance, and individuals’ assumptions about how informed their neighbors are (17). Also, we note that while here—and in related models (39, 53–57)—more extreme opinions are assumed to be more influential, alternative models assume that more extreme opinions can be less influential as moderate individuals may only listen to similar others and ignore distant opinions (35, 37, 38, 40, 81). In future work we will study the impacts of link recommendation in bounded confidence models of opinion dynamics and in combination with homophily-based rewiring. Finally, future extensions shall address 1) the impact of structural similarity rewiring in combination with complex contagion (82, 83), 2) the role of alternative link recommendation algorithms (25, 26, 29, 43), or 3) individuals’ heterogeneity regarding information sharing (17, 83), information levels (58), influence (84) and susceptibility to the perils of online information consumption (85).

We acknowledge that forming links based on common acquaintances is a natural social process, which precedes the existence of online social network algorithms—just as content curation by algorithms coexists with curation by own users (18, 19). In fact, some works underline that offline social networks can even be more homogeneous than their online counterparts (5, 19). Our argument is that, in online platforms, link recommendation algorithms can exacerbate the formation of isolated communities, building on top of the tendency for homophily and triadic closure that individuals already reveal (51, 86). Algorithms result from designers’ decisions. If properly tuned, link recommendation algorithms provide an opportunity to establish links across parties, and, in certain cases, provide chances to establish healthy political dialogues (87), balancing the natural tendency of individuals to establish links with members of the

same party (51). With this work, we hope to contribute to better understanding opinion dynamics in online social networks and shed light on avenues to curb polarization in social media (88).

Implications and Future Directions. Our study examines the impact of social media's link recommendation algorithms on opinion dynamics. First of all, we demonstrate that recommendations that incentivize a rewiring process based on nodes sharing a large number of friends—as likely to occur with current link recommendation algorithms—contributes to opinion polarization. Likewise, we show that such a rewiring process can disrupt the balance between converging and polarizing individuals that sustains moderate opinions: By introducing well-defined modules, link recommendation algorithms may prevent converging nodes from being exposed to a diverse pool of opinions and, as a result, contribute to more extreme viewpoints.

In addition to uncovering a potential mechanism underpinning opinions polarization and radicalization, our results have implications for the design of novel intervention mechanisms that can steer opinion dynamics. Our results suggest that link recommendation algorithms can also curb polarization, if structurally dissimilar nodes are sporadically recommended and followed (see also refs. 32, 60, 62, and 89, and *SI Appendix*, Fig. S9). These types of network interventions only require data about network properties and pose fewer issues vis-a-vis privacy and information availability than interventions that require background on individuals' personal characteristics, opinions, and/or political affiliation. The implementation of this type of interventions calls for future research, namely new experiments on incentives for individuals to follow recommendations to connect to dissimilar others, on the actual fraction of polarizing and converging individuals in different platforms—and the different contexts that may lead the same individual to polarize or converge after intergroup contacts—and, in general, on the ethical aspects of having link recommendation algorithms possibly isolating specific groups and communities (24). Finally, we note that recent work points out that an important distinction among social media platforms is providing the opportunity for users to tweak their newsfeed algorithm: Reddit offers that option and may be one reason to explain why this platform reveals less segregation than Facebook (90). Providing the opportunity for users to tune their link recommendation algorithm can inspire new intervention mechanisms.

Materials and Methods

Opinions. We consider a population of N individuals. Each individual is characterized by an opinion $x_i(t)$, at time t , in a one-dimensional opinion space, $x_i \in (-\infty, +\infty)$. Opinions can be positive or negative: The sign of x_i [denoted by $\sigma(x_i)$] represents the positioning of agent i regarding a certain issue (e.g., pro or against abortion); $x_i = 0$ represents a neutral opinion. The absolute value of x_i , $|x_i|$, captures how radical (extreme) individual i 's opinion is.

Opinion Dynamics. Let us assume that individuals' opinions change through social influence, by interactions with neighbors as defined by a (dynamic) network with adjacency matrix $A(t) \in \{0, 1\}^{N \times N}$. If $A(t)_{ij} = 1$, individual i is connected with individual j at time t and they can co-influence each other. We assume an undirected network, i.e., $A(t)_{ij} = A(t)_{ji}$. Inspired by previous works on opinion dynamics (39, 53, 56) and collective decision-making (54, 55), we assume that opinions change following:

$$x_i(t+1) = \gamma x_i(t) + K \sum_j^N A_{ij} \tanh(\alpha x_j(t)) / k_i. \quad [1]$$

Parameter K controls social influence, that is, the impact of neighbors' opinions on a focal agent opinion. The hyperbolic tangent introduces non-linearity on the social impact of opinions and imposes a bound on the level of influence that individuals have on each other (39, 53, 54, 56, 57); α controls how one's opinion translates into social influence (as discussed in the

Introduction). We assume that opinions are updated based on the average influence of neighbors in an unweighted network and hence the division by k_i . We assume that, in the absence of social interactions, individuals' opinions decay toward the neutral state $x_i = 0$, with a decay factor γ . k_i is the degree of individual i , $k_i = \sum_j^N A_{ij}$. As such, $K \sum_j^N A_{ij} \tanh(\alpha x_j(t)) / k_i$ is bounded within $[-K, K]$. Note that the maximum difference in opinions between two consecutive time steps is $\max|\Delta x_i| = \max|x_i(t+1) - x_i(t)| = |x_i(t)(\gamma - 1) + K|$. Thereby, as $x_i(t)$ approaches $K/(1-\gamma)$, the maximum difference in opinion approaches 0; γ and K determine the maximum magnitude of opinions, $x_i < |K/(1-\gamma)|$. In most of the results presented, we set $\gamma = 0.99$ and $K = 0.1$, which leads to $-10 < x_i < 10$. In *SI Appendix*, Figs. S5–S9, we set $K = 1$ and $\gamma = 0.9$, thus maintaining the same interval in opinions' domain, and reproducing the conclusions presented in the main text.

When individuals interact, they will influence each other: If their opinions have the same sign, $\sigma_i(t) = \sigma_j(t)$, they will reinforce their opinions—which is a phenomenon associated with group polarization (47). If they have different opinion signs, $\sigma_i(t) \neq \sigma_j(t)$, opinions will converge. It was also noted, however, that individuals can become more polarized when exposed to opposing viewpoints on social media—a phenomenon called “backfire effect” (48), also discussed in the context of misinformation and fact checkers (61). A related result was observed in an experiment showing that a talk show that encourages listeners to consider out-group perspectives caused listeners to be less tolerant toward disliked out-groups (50). As a result, we consider a modified version of Eq. 1:

$$x_i(t+1) = \gamma x_i(t) + K \sum_j^N A_{ij} \tanh\left(\alpha(\sigma_i(t)\sigma_j(t))^\lambda x_j(t)\right) / k_i. \quad [2]$$

We note that $\sigma_i(t)\sigma_j(t) = -1$ if individuals have opposing viewpoints. Parameter $\lambda \in \{0, 1\}$ controls the impact of intergroup links on opinions. If $\lambda = 0$, we recover Eq. 1: out-group contacts contribute for opinions to converge and only in-group contracts contribute for individuals to reinforce their opinions (group polarization; ref. 47). If $\lambda = 1$, having a different opinion than a given contact also contributes for opinions to be reinforced (backfire effect; ref. 48).

Social Networks. We consider a dynamic, time-evolving, social network. This is akin to previous approaches [e.g., the adaptive voters' model (34, 41) or recent models that capture online social network dynamics (40, 42)]. Typically, previous models assume that links are established (or broken) based on opinions' distance: Individuals tend to connect with those having similar opinions (34, 39, 40, 81). Here, we assume that individuals connect based on structural similarity,

$$s_{ij} = \frac{|N_i \cap N_j|^\eta}{\sum_k^N |N_i \cap N_k|^\eta}, \quad [3]$$

where N_i denotes the set of neighbors of i and $|N_i \cap N_j|$ the number of common neighbors between i and j . This allows us to introduce a process of link-formation that is likely to be biased by link recommendation algorithms. With effect, s_{ij} is inspired in the common-neighbors index, a local similarity index typically used in link prediction algorithms (23, 43, 44) to anticipate missing or future links—and recommend them. To allow links between independent modules to occasionally be formed, we add with a noise term ϵ ,

$$s_{ij} = \frac{\left(|N_i \cap N_j|(1-2\epsilon) + \epsilon\right)^\eta}{\sum_k^N (|N_i \cap N_k|(1-2\epsilon) + \epsilon)^\eta}. \quad [4]$$

New links will be established with a randomly selected individual, with a probability proportional to the similarity index (s_{ij}). Self-loops and multiple links between the same pair of nodes are not allowed—only simple graphs are considered. η and ϵ control the rewiring process dependence on the similarity index. $\eta = 0$ implies that new links are created with individuals sampled with uniform probability distribution. $\eta > 0$ implies that similar nodes are likely to become connected and very high values of η mean that, at each time step, only the most similar node j will become connected with i . In the simulations performed, we start from a Watts-Strogatz small-world network with probability of rewiring 0.1 (91), which corresponds to a configuration with high clustering coefficient and low average path length—likely to characterize social contacts.

Full Opinion and Network Dynamics. Overall, opinions and links evolve as following:

At each time step t (from $t = 0$ to $t = T$), we repeat:

1. Select, in random order, each individual i out of the N possible individuals.
2. Update i 's social network:
 - 2.a. Randomly select one neighbor of i (say, k) to break the link with. To prevent loners, k needs to have at least 2 neighbors; in *SI Appendix, Figs. S1, S2, S4, S7, S8, and S9*, we do not allow to break links that disconnect the network.
 - 2.b. Select a new neighbor j ($j \neq k$), out of all individuals that i is not yet connected with, with a probability proportional to s_{ij} .
 - 2.c. Set $A_{i,k} = A_{k,i} = 0$ and $A_{i,j} = A_{j,i} = 1$. A new link is only formed if the previous one was allowed to be broken; average degree is thereby constant.
3. Update x_i according with Eq. 2.

The results reported refer, in general, to the (time) average of the last 20% time steps ($T = 2,000$) and (ensemble) average over 100 independent runs starting from random initial conditions (the exception being *SI Appendix, Figs. S5 and S6* where, for $N = 1,000$, we only consider 50 independent runs). Regarding initial conditions, we assume that opinions are attributed to each individual independently and are sampled uniformly from the interval $[-1.0, 1.0]$; the initial topology considered consists in a Watts–Strogatz small-world network with probability of rewiring p (in general we use $p = 0.1$; in *SI Appendix, Fig. S7* we vary this parameter). Here we assume that each time a link is created another one is broken. In reality, link removal can occur at a different rate than link creation; unfriending someone can be rare, and the number of connections of a given user tends to, in general, increase over time. Real social networks are sparse while their size varies over time: Links and nodes can increase at a rate that prevents the network average degree from increasing. Here, we break a link any time a link is created to prevent the network average degree (i.e., average number of connections per node) from increasing, which would lead a network to become denser over time, ultimately turning into a complete graph. This allows us to focus our analysis on a sparse network and, by keeping the number of nodes and average degree of the network constant, we avoid introducing additional effects that could affect opinion dynamics (such as size and density). The networks we test vary, instead, in properties such as modularity, clustering, distances between nodes, and degree distribution. In this regard, we follow previous

1. E. Shearer, E. Grieco, Americans are wary of the role social media sites play in delivering the news. Pew Research Center (2 October 2019). https://www.journalism.org/wp-content/uploads/sites/8/2019/09/PJ_2019.09.25_Social-Media-and-News_FINAL.pdf (Accessed 11 December 2020).
2. R. M. Bond *et al.*, A 61-million-person experiment in social influence and political mobilization. *Nature* **489**, 295–298 (2012).
3. C. R. Sunstein, Is social media good or bad for democracy? *SUR-Int. J. Hum. Rights* **27**, 83 (2018).
4. P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, R. Bonneau, Tweeting from left to right: Is online political communication more than an echo chamber? *Psychol. Sci.* **26**, 1531–1542 (2015).
5. A. Guess, B. Lyons, B. Nyhan, J. Reifler, “Avoiding the echo chamber about echo chambers: Why selective exposure to like-minded political news is less prevalent than you think” (White Paper, Knight Foundation, 2018; https://www.researchgate.net/publication/330144926_Avoiding_the_echo_chamber_about_echo_chambers_Why_selective_exposure_to_like-minded_political_news_is_less_prevalent_than_you_think).
6. Y. Leikkes, G. Sood, S. Iyengar, The hostile audience: The effect of access to broadband internet on partisan affect. *Am. J. Pol. Sci.* **61**, 5–20 (2017).
7. H. Allcott, L. Braghieri, S. Eichmeyer, M. Gentzkow, The welfare effects of social media. *Am. Econ. Rev.* **110**, 629–676 (2020).
8. K. Garimella, G. De Francisci Morales, A. Gionis, M. Mathioudakis, “Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship” in *Proceedings of the 2018 World Wide Web Conference* (Association for Computing Machinery, 2018), pp. 913–922.
9. M. Del Vicario *et al.*, The spreading of misinformation online. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 554–559 (2016).
10. C. R. Sunstein, *Republic.com 2.0* (Princeton University Press, 2007).
11. E. J. Finkel *et al.*, Political sectarianism in America. *Science* **370**, 533–536 (2020).
12. S. Iyengar, Y. Leikkes, M. Levendusky, N. Malhotra, S. J. Westwood, The origins and consequences of affective polarization in the United States. *Annu. Rev. Polit. Sci.* **22**, 129–146 (2019).
13. S. Iyengar, G. Sood, Y. Leikkes, Affect, not ideology: Social identity perspective on polarization. *Public Opin. Q.* **76**, 405–431 (2012).
14. D. M. J. Lazer *et al.*, The science of fake news. *Science* **359**, 1094–1096 (2018).
15. S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
16. S. C. Woolley, P. N. Howard, Political communication, computational propaganda, and autonomous agents: Introduction. *Int. J. Commun.* **10**, 4882–4890 (2016).
17. R. Bowen, D. Dmitriev, S. Galperti, *Learning from Shared News: When Abundant Information Leads to Belief Polarization* (National Bureau of Economic Research, 2021).
18. D. Lazer, Social sciences. The rise of the social algorithm. *Science* **348**, 1090–1091 (2015).
19. E. Bakshy, S. Messing, L. A. Adamic, Political science. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
20. Y. Shmargad, S. Klar, Sorting the news: How ranking by popularity polarizes our politics. *Polit. Commun.* **37**, 423–446 (2020).
21. R. Levy, Social media, news consumption, and polarization: Evidence from a field experiment. *Am. Econ. Rev.* **111**, 831–870 (2021).
22. L. Terveen, D. W. McDonald, Social matching: A framework and research agenda. *ACM Trans. Comput. Hum. Interact.* **12**, 401–434 (2005).
23. Z. Li, X. Fang, O. R. L. Sheng, A survey of link recommendation for social networks: Methods, theoretical foundations, and future research directions. *ACM Trans. Manag. Inf. Syst.* **9**, 1–26 (2017).
24. F. Fabbri, F. Bonchi, L. Boratto, C. Castillo, “The effect of homophily on disparate visibility of minorities in people recommender systems” in *Proceedings of the International AAAI Conference on Web and Social Media* (Association for the Advancement of Artificial Intelligence, 2020), pp. 165–175.
25. P. Gupta *et al.*, “Wtf: The who to follow service at Twitter” in *Proceedings of the 22nd International Conference on World Wide Web* (Association for Computing Machinery, 2013), pp. 505–514.
26. E. M. Daly, W. Geyer, D. R. Millen, “The network effects of recommending social connections” in *Proceedings of the Fourth ACM Conference on Recommender Systems* (Association for Computing Machinery, 2010), pp. 301–304.
27. J. Su, A. Sharma, S. Goel, “The effect of recommendations on network structure” in *Proceedings of the 25th International Conference on World Wide Web* (Association for Computing Machinery, 2016), pp. 1157–1167.

models of opinion dynamics that keep the number of nodes and average degree constant (34, 40, 81).

While the model we analyze here assumes opinion dynamics on top of a dynamic network structure [and in that context resembling previous works (39, 54, 56, 57)], we shall stress that our methodological approach deviates from previous models by 1) considering that interaction probability and network rewiring depend exclusively on structural similarly; 2) considering the coexistence of converging and polarizing nodes in what concerns interactions with the out-group; and 3) assuming a network dynamics that can evolve in parallel with opinions. Besides numerical simulations, in *SI Appendix* we provide an analytically tractable model that allows studying the coupled dynamics of binary opinions and link-rewiring within two communities (*SI Appendix, Supplementary Text*).

Measuring Polarization and Radicalization. Defining a unified measure of polarization is challenging (92), and the magnitude of polarization can be captured along different dimensions (93). Options include measuring the spread of opinions (difference between maximum and minimum opinions), coverage (how much of the opinion spectrum is covered) or opinions' relative representativeness (92). Here, we consider that polarization is associated with the SD of opinions (94). This resembles the dispersion measure defined in refs. 92 and 93. We consider that opinions are more radicalized if the absolute value of $x_i(t)$ is higher.

Data Availability. The raw data from simulations (used to generate figures) have been deposited in <https://github.com/fp-santos/link-recommendation-polarization>. Previously published data (65) were used for this work.

ACKNOWLEDGMENTS. This work was supported by the James S. McDonnell Foundation 21st Century Science Initiative in Understanding Dynamic and Multi-scale Systems Postdoctoral Fellowship Award (Grant 200200555) and Collaborative Award (Grant 220020542), the National Science Foundation (Grant CCF1917819), the C3.ai Inc. and Microsoft Corporation (Award AWD1006615), and the Army Research Office (Grant W911NF-18-1-0325). We thank the College of Liberal Arts and Sciences at Arizona State University for providing the funding for the workshops that led to this paper. We thank the participants of the PNAS Political Polarization Conference, and especially Sara Constantino and Vitor Vasconcelos, for enriching comments. We also thank participants of the Theoretical Ecology Lab Tea (Department of Ecology and Evolutionary Biology, Princeton University) for useful suggestions. We thank Alan Mislove for providing us a dataset used in a previous paper (65).

28. M. Zignani *et al.*, "Link and triadic closure delay: Temporal metrics for social network dynamics" in *Proceedings of the International AAAI Conference on Web and Social Media* (Association for the Advancement of Artificial Intelligence, 2014).
29. L. Backstrom, J. Leskovec, "Supervised random walks: Predicting and recommending links in social networks" in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining* (Association for Computing Machinery, 2011), pp. 635–644.
30. J. Su, K. Kamath, A. Sharma, J. Ugander, S. Goel, "An experimental study of structural diversity in social networks" in *Proceedings of the International AAAI Conference on Web and Social Media* (Association for the Advancement of Artificial Intelligence, 2020), pp. 661–670.
31. J. Sanz-Cruzado, P. Castells, "Enhancing structural diversity in social networks by recommending weak ties" in *Proceedings of the 12th ACM Conference on Recommender Systems* (Association for Computing Machinery, 2018), pp. 233–241.
32. C. Musco, C. Musco, C. E. Tsourakakis, Minimizing Polarization and Disagreement in Social Networks. *Proceedings of the 2018 International World Wide Web Conference* (Association for Computing Machinery, 2018), 369–378 (2018).
33. S. A. Levin, Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* **1**, 431–436 (1998).
34. P. Holme, M. E. Newman, Nonequilibrium phase transition in the coevolution of networks and opinions. *Phys. Rev. E* **74**, 056108 (2006).
35. R. Axelrod, The dissemination of culture: A model with local convergence and global polarization. *J. Conflict Resolut.* **41**, 203–226 (1997).
36. M. W. Macy, J. A. Kitts, A. Flache, S. Benard, "Polarization in dynamic networks: A Hopfield model of emergent structure" in *Dynamis Social Network Modeling and Analysis*, R. Breiger, K. Carley, P. Pattison, Eds. (National Academy of Sciences, 2003), pp. 162–173.
37. R. Hegselmann, U. Krause, Opinion dynamics and bounded confidence models, analysis, and simulation. *J. Artif. Soc. Soc. Simul.* **5**, 3 (2002).
38. G. Deffuant, D. Neau, F. Amblard, G. Weisbuch, Mixing beliefs among interacting agents. *Adv. Complex Syst.* **3**, 87–98 (2000).
39. F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, M. Starnini, Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.* **124**, 048301 (2020).
40. X. Wang, A. D. Sirianni, S. Tang, Z. Zheng, F. Fu, Public discourse and social network echo chambers driven by socio-cognitive biases. *Phys. Rev. X* **10**, 041042 (2020).
41. O. J. Chu, J. F. Donges, G. B. Robertson, G. Pop-Eleches, The microdynamics of spatial polarization: A model and an application to survey data from Ukraine. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104194118 (2021).
42. K. Sasahara *et al.*, Social influence and unfollowing accelerate the emergence of echo chambers. *J. Comput. Soc. Sci.* **4**, 381–402 (2021).
43. L. Lü, T. Zhou, Link prediction in complex networks: A survey. *Physica A* **390**, 1150–1170 (2011).
44. D. Liben-Nowell, J. Kleinberg, The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.* **58**, 1019–1031 (2007).
45. G. J. Martin, S. W. Webster, Does residential sorting explain geographic polarization? *Political Sci. Res. Methods* **8**, 215–231 (2020).
46. J. Mummaloo, C. Nall, Why partisans do not sort: The constraints on political segregation. *J. Polit.* **79**, 45–59 (2017).
47. C. R. Sunstein, The law of group polarization. *J. Polit. Philos.* **10**, 175–195 (2002).
48. C. A. Bail *et al.*, Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9216–9221 (2018).
49. Y. Shmargad, S. Klar, How partisan online environments shape communication with political outgroups. *Int. J. Commun.* **13**, 27 (2019).
50. E. L. Paluck, Is it better not to talk? Group polarization, extended contact, and perspective taking in Eastern Democratic Republic of Congo. *Pers. Soc. Psychol. Bull.* **36**, 1170–1185 (2010).
51. M. Mosleh, C. Martel, D. Eckles, D. G. Rand, Shared partisanship dramatically increases social tie formation in a Twitter field experiment. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022761118 (2021).
52. T. F. Pettigrew, L. R. Tropp, A meta-analytic test of intergroup contact theory. *J. Pers. Soc. Psychol.* **90**, 751–783 (2006).
53. F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, M. Starnini, Emergence of polarized ideological opinions in multidimensional topic spaces. *Phys. Rev. X* **11**, 011012 (2021).
54. R. Gray, A. Franci, V. Srivastava, N. E. Leonard, Multiagent decision-making dynamics inspired by honeybees. *IEEE Trans. Control Netw. Syst.* **5**, 793–806 (2018).
55. A. Franci, V. Srivastava, N. E. Leonard, A realization theory for bio-inspired collective decision-making. arXiv [Preprint] (2015). <https://arxiv.org/abs/1503.08526> (Accessed 15 February 2021).
56. A. Bizyaeva, A. Franci, N. E. Leonard, A general model of opinion dynamics with tunable sensitivity. arXiv [Preprint] (2020). <https://arxiv.org/abs/2009.04332> (Accessed 15 February 2021).
57. N. E. Leonard, K. Lipsitz, A. Bizyaeva, A. Franci, Y. Lelkes, The nonlinear feedback dynamics of asymmetric political polarization. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102149118 (2021).
58. I. D. Couzin *et al.*, Uninformed individuals promote democratic consensus in animal groups. *Science* **334**, 1578–1580 (2011).
59. D. J. Isenberg, Group polarization: A critical review and meta-analysis. *J. Pers. Soc. Psychol.* **50**, 1141 (1986).
60. C. S. Taber, M. Lodge, Motivated skepticism in the evaluation of political beliefs. *Am. J. Pol. Sci.* **50**, 755–769 (2006).
61. B. Swire-Thompson, J. DeGutis, D. Lazer, Searching for the backfire effect: Measurement and design considerations. *J. Appl. Res. Mem. Cogn.* **9**, 286–299 (2020).
62. R. Axelrod, J. J. Daymude, S. Forrest, Preventing extreme polarization of political attitudes. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102139118 (2021).
63. A. Ghasemian, H. Hosseiniardi, A. Galstyan, E. M. Airolidi, A. Clauset, Stacking models for nearly optimal link prediction in complex networks. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23393–23400 (2020).
64. Facebook, People You May Know—Facebook Help Center. <https://www.facebook.com/help/336320879782850/people-you-may-know>. Accessed 20 February 2021.
65. B. Viswanath, A. Mislove, M. Cha, K. P. Gummadi, "On the evolution of user interaction in Facebook" in *Proceedings of the 2nd ACM Workshop on Online Social Networks* (Association for Computing Machinery, 2009), pp. 37–42.
66. L. A. Adamic, N. Glance, "The political blogosphere and the 2004 US election: Divided they blog" in *Proceedings of the 3rd International Workshop on Link Discovery* (Association for Computing Machinery, 2005), pp. 36–43.
67. A. Feller, M. Kuhnert, T. Sprenger, I. Welpe, "Divided they tweet: The network structure of political microbloggers and discussion topics" in *Proceedings of the International AAAI Conference on Web and Social Media* (Association for the Advancement of Artificial Intelligence, 2011).
68. M. Conover, *et al.*, "Political polarization on Twitter" in *Fifth International AAAI Conference on Weblogs and Social Media* (Association for the Advancement of Artificial Intelligence, 2011).
69. A. Gruzd, J. Roy, Investigating political polarization on Twitter: A Canadian perspective. *Policy Internet* **6**, 28–45 (2014).
70. P. Aragón, K. E. Kappler, A. Kaltenbrunner, D. Laniado, Y. Volkovich, Communication dynamics in Twitter during political campaigns: The case of the 2011 Spanish national election. *Policy Internet* **5**, 183–206 (2013).
71. S. Du, S. Gregory, *The Echo Chamber Effect in Twitter: Does Community Polarization Increase?* (Springer, 2016), pp. 373–378.
72. C. Chan, K. Fu, The relationship between cyberbalkanization and opinion polarization: Time-series analysis on Facebook pages and opinion polls during the Hong Kong Occupy Movement and the associated debate on political reform. *J. Comput. Mediat. Commun.* **22**, 266–283 (2017).
73. Y. Lelkes, Mass polarization: Manifestations and measurements. *Public Opin. Q.* **80**, 392–410 (2016).
74. F. Fu *et al.*, Evolution of in-group favoritism. *Sci. Rep.* **2**, 460 (2012).
75. F. P. Santos, F. C. Santos, J. M. Pacheco, Social norm complexity and past reputations in the evolution of cooperation. *Nature* **555**, 242–245 (2018).
76. A. J. Stewart, N. McCarty, J. J. Bryson, Polarization under rising inequality and economic decline. *Sci. Adv.* **6**, eabd4201 (2020).
77. F. P. Santos, F. C. Santos, J. M. Pacheco, S. A. Levin, "Social network interventions to prevent reciprocity-driven polarization" in *Proceedings of the 20th Conference on Autonomous Agents and MultiAgent Systems* (Association for Computing Machinery, 2021).
78. M. Kawakatsu, Y. Lelkes, S. A. Levin, C. E. Tarnita, Interindividual cooperation mediated by partisanship complicates Madison's cure for "mischiefs of faction." *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102148118 (2021).
79. V. V. Vasconcelos *et al.*, Segregation and clustering of preferences erode socially beneficial coordination. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102153118 (2021).
80. A. J. Stewart, J. B. Plotkin, N. McCarty, Inequality, identity, and partisanship: How redistribution can stem the tide of mass polarization. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102140118 (2021).
81. M. Del Vicario, A. Scala, G. Caldarelli, H. E. Stanley, W. Quattrociocchi, Modeling confirmation bias and polarization. *Sci. Rep.* **7**, 40391 (2017).
82. V. V. Vasconcelos, S. A. Levin, F. L. Pinheiro, Consensus and polarization in competing complex contagion processes. *J. R. Soc. Interface* **16**, 20190196 (2019).
83. C. K. Tokita, A. M. Guess, C. E. Tarnita, Polarized information ecosystems can reorganize social networks via information cascades. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e210247118 (2021).
84. J. Flamino *et al.*, Shifting polarization and Twitter news influencers between two U.S. presidential elections. arXiv [Preprint] (2021). <https://arxiv.org/pdf/2111.02505.pdf> (Accessed 24 November 2021).
85. A. M. Guess, B. Nyhan, J. Reifler, Exposure to untrustworthy websites in the 2016 US election. *Nat. Hum. Behav.* **4**, 472–480 (2020).
86. M. McPherson, L. Smith-Lovin, J. M. Cook, Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
87. D. C. Mutz, *Hearing the Other Side: Deliberative Versus Participatory Democracy* (Cambridge University Press, 2006).
88. M. Nelimarkka, S.-M. Laaksonen, B. Semaan, "Social media is polarized, social media is polarized: Towards a new design agenda for mitigating polarization" in *DIS '18: Proceedings of the 2018 Designing Interactive Systems Conference* (2018), pp. 957–970.
89. K. Garimella, G. De Francisci Morales, A. Gionis, M. Mathioudakis, "Reducing controversy by connecting opposing views" in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (Association for Computing Machinery, 2017), pp. 81–90.
90. M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, M. Starnini, The echo chamber effect on social media. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e202301118 (2021).
91. D. J. Watts, S. H. Strogatz, Collective dynamics of "small-world" networks. *Nature* **393**, 440–442 (1998).
92. A. Bramson *et al.*, Understanding polarization: Meanings, measures, and model evaluation. *Philos. Sci.* **84**, 115–159 (2017).
93. P. DiMaggio, J. Evans, B. Bryson, Have Americans social attitudes become more polarized? *Am. J. Sociol.* **102**, 690–755 (1996).
94. N. McCarty, *Polarization: What Everyone Needs to Know* (Oxford University Press, 2019).