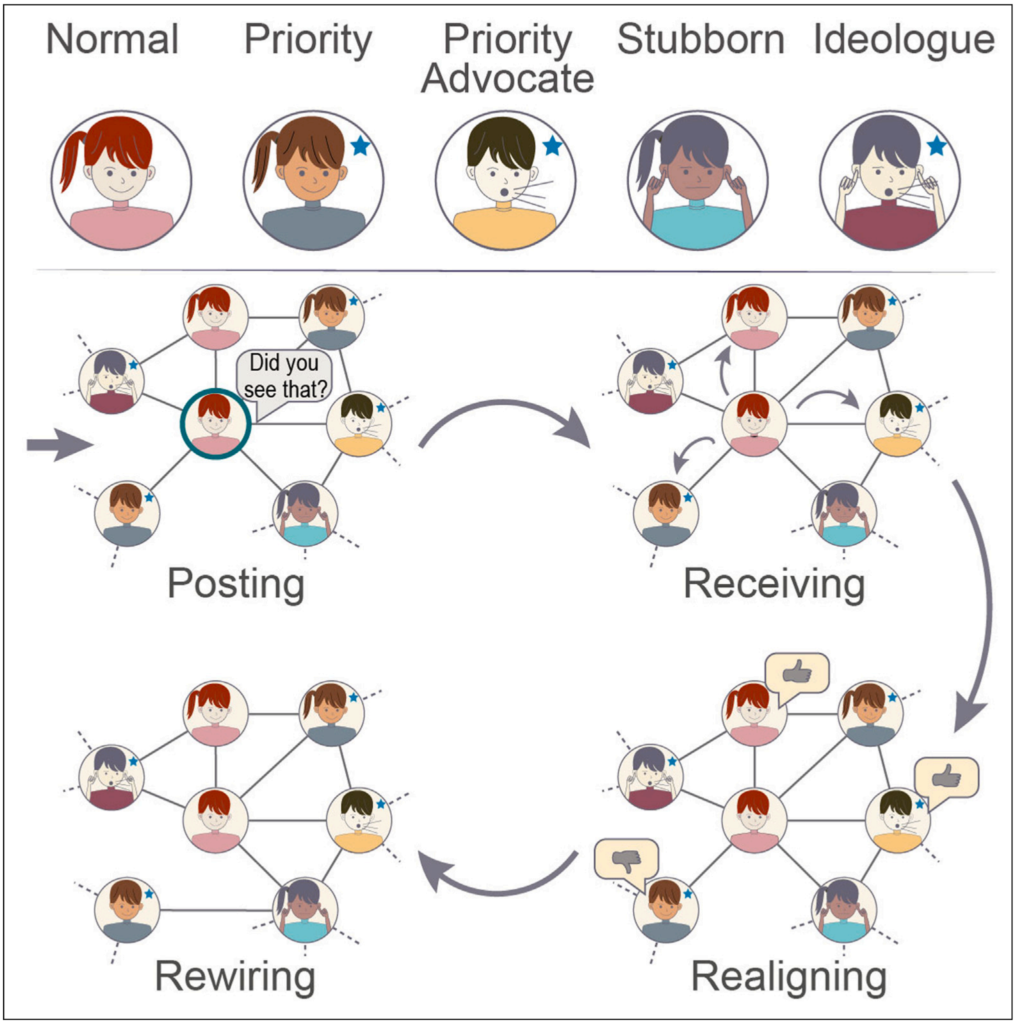
iScience

CellPress
OPEN ACCESS

Article

Echo chamber formation sharpened by priority users



Henrique Ferraz de Arruda, Kleber Andrade Oliveira, Yamir Moreno

h.f.arruda@gmail.com

Highlights

Investigates the impact of priority users on opinion dynamics in social media

Prioritization of users can mitigate polarization in certain contexts

Prioritizing users can change outcomes, exacerbating echo chamber formation

Ideologues can shift consensus to polarization more easily than stubborn users

Ferraz de Arruda et al., iScience 27, 111098
November 15, 2024 Published by Elsevier Inc.
https://doi.org/10.1016/j.isci.2024.111098

iScience

CellPress OPEN ACCESS

Article

Echo chamber formation sharpened by priority users

Henrique Ferraz de Arruda,[1,2,5,*] Kleber Andrade Oliveira,[1] and Yamir Moreno[1,3,4]

SUMMARY

On social media platforms, priority users (e.g., verified profiles on X) are users whose posts are promoted by recommendation algorithms. However, their influence on opinion dynamics, in particular polarization and echo chamber formation, is not well understood. Through computational modeling, we investigate this influence in a stylized setting. By introducing priority user accounts, we find that prioritization can mitigate polarization. However, by incorporating stubborn user behavior, we find that the results change and that priority accounts can exacerbate the formation of echo chambers. In other words, a minority of extremist ideologues can trigger a transition from consensus to polarization. Our study suggests careful monitoring of platform prioritization policies to prevent potential misuse of users with enhanced influence.

INTRODUCTION

After undergoing a recent shift in its business model, X (formerly called Twitter), a well-studied social media platform,[1] has introduced substantial revisions to its verification policy (from now on, we refer to Twitter as X). Users labeled as verified by the platform went from being "active, notable, and authentic"[2] to those who meet substantially more permissive criteria[3] and pay a monthly fee to a subscription plan. The benefits of the subscription involve forms of algorithmic boosting, such as prioritized ranking in replies and searches. However, inadvertently promoting these users may trigger unknown macroscopic effects in the system, especially regarding opinion dynamics.[4] Given this problem, our paper addresses the following research question: *what is the effect of inadvertent social media content promotion on opinion polarization and the formation of echo chambers?*

Social media platforms implemented verification processes to label some users differently than the rest. On Facebook[5] and Instagram,[6] it means the platform itself vouches for the authenticity of the associated user account. Concurrently, the role influential individuals play in the spread of information and opinion has been known from much before social media platforms.[7,8] Hence, many people in society who are famous or influential, regardless of social media, tend to rely on this feature so that their online influence is not misused or mischaracterized. In prominent cases, a verified account of an influential individual is an active public speech stage where they can perform and carry marketing or political campaigns.[9]

The rules for each verification procedure and benefits associated with verified accounts change according to the business model of the platform. As a consequence, there has been much debate[10–13] about how these verified users exert more influence compared to the other users. In the case of X, the change of verification policy has been a cause for concern for institutions that design policies involving situations affected by opinion polarization. For instance, the French Commission for Campaign Accounts and Political Financing (CNCCFP) decided not to allow candidates to subscribe to X's new verification system ahead of the 2024 French elections.[14] They understand the subscription constitutes an increase of visibility and reach of the subscribed account and, hence, a form of sponsored advertising.

It is not surprising to see policymakers acting preemptively toward perturbation on already polarized scenarios. Opinion polarization is not an easy phenomenon to intervene or control, and as such has been increasingly studied because it is critical to public debate and collective decisions.[15,16] Many researchers turned their attention to studying this phenomenon ever since measurement via social media data[17–19] become available, particularly in significant periods (such as upcoming elections or referenda) of democracies around the world.[20–23]

Computational models of opinion dynamics have been developed to explain the formative mechanisms behind polarization.[24] Agent-based models are a class of such models that computationally simulate the system through numerous agents whose automated behavior is ore-specified by a set of rules. The majority of the agent-based models encode opinion mechanisms formed from the interaction among agents.[25–29] As a result, network descriptions are largely used in the field. For instance, the persuasion mechanism, characteristic of social networks, has been modeled in the Sznajd model.[30,31] Adaptive networks have also been modeled and thus identified as a mechanism for forming echo chambers.[32–34] In addition, some models have incorporated ingredients of real social media, such as,[29] which took into account the mechanisms of posting, reposting, and three strategies for changing friendships.

[1]CENTAI Institute, Turin, Italy
[2]Geography and Geoinformation Science, College of Science, George Mason University, Fairfax, VA, USA
[3]Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza, Spain
[4]Department of Theoretical Physics, Faculty of Sciences, University of Zaragoza, Zaragoza, Spain
[5]Lead contact
*Correspondence: h.f.arruda@gmail.com

https://doi.org/10.1016/j.isci.2024.111098

In tandem with the change in the verification policy, X also started to limit its data access to researchers. To explore the implications of the platform alterations, we approach our research question computationally by incorporating heterogeneous content promotion into an agent-based simulation of opinion dynamics in a social media setting through different types of users. We modify a previously developed opinion model[35] on adaptive networks to incorporate heterogeneous user behavior. This model was capable of reproducing many opinion configurations (from consensus to echo chambers) even with its previous setup (homogeneous user behavior). This model was also found to reproduce real data on X in terms of polarization and echo chamber formation. Valensise et al.[36] conducted tests on similar mechanisms to verify which has better data support from different social media platforms. They found that the opinion model by Arruda et al., which we developed from, was better suited to reproduce polarization measurements from X than other social media platforms, adding to the results of the original paper.[35]

In this study, we investigate the effects of priority accounts on social media. We start by implementing priority users in the agent-based simulation and testing how they influence macroscopic measures of polarization. After this, we incorporate the behavior of stubborn users[37–39]. Here, we use the terms stubborn and zealot interchangeably. We also study whether the effect of stubborn users is stronger when they have priority accounts. All stubborn users with a priority account are defined here as ideologues. By varying the users and their priority, we study how far the gain of influence from prioritization can be pushed toward perturbing the dynamics. To do so, we investigate content promotion of extremist or centrist users via the priority policy.

To adequately measure echo chamber formation, we propose a novel measure that incorporates the relationship between users' opinions and the average opinions of their neighbors into a single unidimensional distribution. Our results reveal that the presence of extremist stubbornness, even in small proportions, can significantly drive opinion polarization and strengthen echo chambers. Moreover, we observe that stubborn centrists can reinforce polarization in the network. These findings highlight the potential risks of policy changes in social media and emphasize the complex interplay between network structure, user behavior, and opinion dynamics in social media environments. Overall, our study elucidates which mechanisms are enough to drive macroscopic changes to the dynamics and, by doing so, provides insights into understanding the role of content promotion policies in shaping public discourse and opinion polarization in online media. Finally, we emphasize that the change in verification policy also raises research questions about the dynamics of misinformation. However, in this paper, we limit the scope of our investigation to mechanisms of opinion polarization because empirical data are now limited after the change in the data access policy.

RESULTS

Background

In this section, we set clear working definitions to study how opinion polarization phenomena emerge in networked social systems. Many empirical characteristics of these phenomena need to be simplified so that explanatory computational models[40] are developed. In particular, we are interested in modeling ideological polarization.[41] In this sense, we say a system composed of many different opinions is polarized when its opinions are concentrated in extremes, as opposed to what is commonly perceived as center/moderate.

Throughout this work, we consider each opinion as a unidimensional and bounded continuous variable. Then, we opt to use numbers lying in the interval between -1 and 1 such that the "physical interpretation" of the extreme values (either 1 or -1) may be seen as a person who is completely for or against a particular matter, while those with an opinion 0 behave moderately. With this representation, systems can be characterized as polarized if the distribution of their opinions is densely accumulated away from the center (0) and closer to the extremes (-1 and 1). This is equivalent to observing bimodality around the center in the opinion distribution. Complementarily, when the opinions are densely accumulated close to a single value, we refer to the system configuration as a consensus.

While the distribution of opinions adequately describes the system's polarization, it may happen through different network configurations. Polarized opinions may be associated with a highly divided network, where intra-group opinion is aligned but inter-group is opposite, but they can also be found in a more homogeneously connected network, where many edges tie together people with opposite opinions.

The phenomenon of online echo chambers has been increasingly studied, even though it lacks a consensus as to how it is quantified and measured.[42] It may be defined as a specific media space in which information aligned with its constituents is amplified, while contradictory information is hindered.[43] Here, we focus on the opinion dynamics regarding how people post and receive messages on social media platforms and treat echo chambers as a specialized state of a polarized system. Thus, we work with a definition of "echo chambers" as a polarized system, which also happens to be divided in terms of how individuals are networked, such that users are predominantly connected with others who share similar opinions.

Several previous studies proposed approaches to measure how divided and polarized the networks are.[44–47] For example, Hohmann et al.[44] uses the concept of distances on the network, and Arruda et al.[47] measures how opinions are associated with network communities. Here, we rely on a method based on plotting a density map to quantify the presence of separate groups with user opinions in a social network, proposed by Cota et al.[45] In this approach, the distribution of user's opinions b is plotted against the distribution of the average opinion of each user's outgoing neighbors, denoted as b_{NN}. Correspondingly to how we define polarization with b, we refer to bimodality in b_{NN} as a "divided" state. The density map then captures the relationship between a user's opinion and the opinions of the users they follow in a bounded two-coordinate plane. When the network is divided into internally aligned groups that oppose each other externally, there is a higher density in the first and third quadrants of the associated map. This observation makes it possible to quantify the echo chamber effect within the network. This density map approach has been widely used in various studies.[34–36,44,45,48]

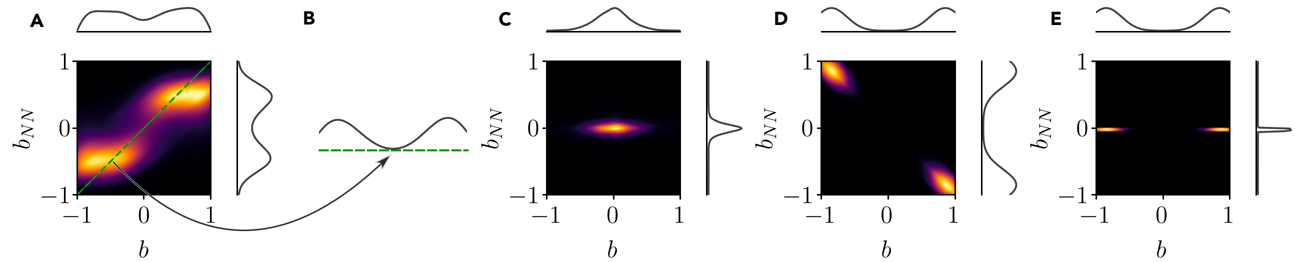


Figure 1. Measuring echo chambers with BC_{hom}

(A) shows the density map of b against b_{NN} and (B) shows the projected diagonal of the density map from (A). For this example, BC(b) = 0.583 and $BC_{hom}(b, b_{NN}) = 0.667$, which shows how the measure BC_{hom} enhances the echo chamber detection by encoding it into the bimodality of the rotated distribution. A consensus scenario (C) has unimodal distributions in both axes, which should not be significantly changed by the rotation (BC(b) = 0.237 and $BC_{hom} = 0.226$).(D) represents a scenario of conflict, which is polarized and divided, but its groups do not have aligned opinions; hence our measure penalizes it (BC(b) = 0.961 and $BC_{hom} = 0.415$). In the last case, in (E), we show a polarized scenario with no clear division in the network. In this case, BC(b) = 0.961 and $BC_{hom} = 0.926$. This is a case in which BC_{hom} does not inform better whether echo chambers formed in the system.

Homophilic bimodality coefficient

To measure polarization, Arruda et al. employed the bimodality coefficient BC[49], used to assess the degree of bimodality in the distribution of opinions. Given a sample distribution of opinions b of size n (the number of users), the bimodality coefficient BC(b)[35] is defined as

$$BC(b) = \frac{g^2+1}{k+\frac{3(n-1)^2}{(n-2)(n-3)}}, \qquad \text{(Equation 1)}$$

where g and k are the *sample skewness*[50] and *excess kurtosis*[50] of the distribution b, respectively. Empirically, if $BC(b) > 5/9$, the distribution is considered bimodal.[49]

In another study, Martin-Gutierrez et al.[46] projected the opinions through a principal component analysis[51,52] to find the best multipolar projection. However, this study was proposed to aim at measuring echo chambers of multi-dimensional opinions, and here, we are considering a single opinion.

Inspired by these studies, we propose a novel measure to enhance the neighborhood aspect and capture echo chamber formation as the bimodality in the combined dimensions of b and b_{NN}. First, we rotate the $b \times b_{NN}$ map by 45° using a rotation matrix R. By applying this rotation, we align the axes along the major and minor axes of the distribution, effectively combining the two dimensions. Subsequently, we calculate BC using only the first dimension of this rotated map. This novel measure is called the homophilic bimodality coefficient BC_{hom}.

Following the previous notation, a distribution $b^\dagger$ from the projected diagonal is obtained from

$$b^\dagger = [[b, b_{NN}]R]_{i1} \text{ for } 1 \leq i \leq n, \qquad \text{(Equation 2)}$$

where $[b, b_{NN}]$ is of order $n \times 2$ and R is the 2×2 rotation matrix of 45°. Thus, the homophilic bimodality coefficient $BC_{hom}(b, b_{NN})$ of the original distribution b associated with a b_{NN} is

$$BC_{hom}(b, b_{NN}) = BC(b^\dagger). \qquad \text{(Equation 3)}$$

In other words, the transformation combines both the overall opinion distribution in the system and the neighborhood alignment into a single distribution. This allows us to clarify whether the opinion poles also coincide with aligned networked clusters through the bimodality of the transformed distribution. Our measure represents the opinion densities accumulated in the first and third quadrants through a single value bounded between 0 and 1.

We use a simulated polarized scenario with echo chambers to illustrate the effectiveness of this measure. The details on how we obtained it are shown in Section base case. Figure 1A shows the density map of b against b_{NN} obtained through this base polarized scenario, and Figure 1B exhibits the corresponding transformed distribution $b^\dagger$ described previously. The other panels explain how the measure reacts to other scenarios, such as consensus (Figure 1C) and other types of polarization.

We notice that when BC_{hom} exceeds the 5/9 threshold, it does not mean the system is organized into exactly two groups, one with positive opinions and the other with negative ones. As pointed out by Hohnmann et al.,[44] echo chambers may happen in a network divided into several groups. It may well be the case that a high BC_{hom} points to the presence of several network clusters following the pattern of aligned intra-group opinion and opposing inter-group opinion.

To better understand BC_{hom}, we compare it with the measure proposed in[44] referred to as the generalized Euclidean distance $\delta_{G,b}$. This measure assesses ideological polarization by estimating how long it takes to navigate the network between nodes of opposite opinions. To compare them, we compute BC_{hom} and $\delta_{G,b}$ for a set of networks, varying connectivity between two communities. We used a stochastic block model[53] with two communities, each comprising 500 nodes. We also vary the degree of polarization, replicating tests from.[44] We focus on two communities because the rewiring process of our model often divides the networks into two distinct groups when echo chambers form. To assess whether our measure captures similar information to $\delta_{G,b}$, we compare the two measures using datasets with a fixed number of nodes and average degree. Specifically, we compute the Pearson correlation between $\delta_{G,b}$ and BC_{hom} for each of the tested datasets.

Following the pipeline used by Hohmann et al.,[44] we generate a series of networks varying the connectivity between two communities using a stochastic block model,[53] implemented using NetworkX.[54] Here, we consider only two communities with 500 nodes each. The two parameters tested count the edge density intra-communities and inter-communities, respectively; they are $\{(8.5\times10^{-3}, 8.5\times10^{-3}), (3.9\times10^{-2}, 4.2\times10^{-3}), (5.4\times10^{-2}, 2.4\times10^{-3}), (6.2\times10^{-2}, 1.2\times10^{-3}), (6.4\times10^{-2}, 6.0\times10^{-4}), (6.7\times10^{-2}, 3.0\times10^{-4})\}$.

Another approach is used to control how extreme the opinions are in the distribution. The bimodal distribution is given by a combination of two Gaussian distributions, i.e., half of the values are generated with the average μ, and the other half is copied with the negative sign. The values considered for the mean are $\mu \in \{0.0, 0.2, 0.4, 0.6, 0.8\}$. For all tests, the standard deviation is set to $\sigma = 0.2$. Note that for $\mu = 0$, the distribution is unimodal. Figure S1 shows the comparison between BC_{hom} and $\delta_{G,b}$, where each color represents a fixed pair of in- and out-community connection probabilities, which result in fixed expected average degrees. The lowest correlation value found is 0.97, indicating that both measures are linearly correlated, implying that BC_{hom} and $\delta_{G,b}$ convey similar information within this experimental setup.

Given the compatibility of both metrics in a fixed system state setup, we argue that BC_{hom} is better suited to capture echo chambers in this study. Since we work with adaptive networks, distances between nodes change in a non-trivial manner, rendering them much harder to interpret. In addition, BC_{hom} can be computed for networks formed by disconnected components, a capability lacking in $\delta_{G,b}$ but necessary when edges are often rewired. The Euclidean-based $\delta_{G,b}$ is unsuitable for comparing networks of different sizes and average degrees as opposed to our metric, which facilitates a straightforward interpretation of the results due to being normalized. In cases where the network is polarized but not with echo chambers (see example in Figure 1E), both measures are not suitable for our analysis. In this case, $\delta_{G,b}$ yields higher values when the distribution of b is bimodal than unimodal, as BC_{hom} do.

Model

The proposed opinion model is based on a modification of the one introduced by de Arruda et al.[35] The initial setup is a directed network $G = (V, E)$, where V is a set of nodes and E is a set of ordered pairs of nodes $\{v_i, v_j\}$, indicating the edge direction from i to j. In this representation, nodes are social media users, each with a continuous opinion bounded between -1 and 1. A user points to another if they receive content from them (followership relation); that is, the content is spread following the opposite direction of edges. Furthermore, the content has its own opinion value θ, which is always fixed and uniformly random.

Each iteration is summarized with the following steps.

(1) ***Activating:*** a node i is uniformly randomly picked to become active. Technically, we sort a random number in the interval $[1, |V|]$. This number represents the selected node i;
(2) ***Posting:*** The active node i posts an opinion with content θ according to a posting rule with probability denoted by P_p. This rule compares the opinion b_i of node i with θ, where θ is a uniformly random number between -1 and 1. In more detail, a uniformly random number $R^{posting}$ is generated in the interval [0, 1) and if $R^{posting} < P_p(b_i, \theta)$, the algorithm goes to the next step, otherwise the iteration finishes;
(3) ***Receiving:*** Each follower j of i receives the post content according to the receiving probability P_r, also a function of their opinion b_j and b_i. We say there is a recommendation algorithm controlling this particular rule. This step is implemented as follows. For each follower j of i, a uniformly random number $R_j^{receiving}$ in the interval $[0, 1)$ is generated, and if $R_j^{receiving} < P_r(b_i, b_j)$, the agent receives θ. Only agents that receive θ proceed to the next step;
(4) ***Realigning:*** Those who received the content may either move their opinion toward or away from θ by a constant $\Delta = 0.1$. They are repulsed away from the post with probability $|\theta - b_j|/2$ and attracted otherwise. This step is based on the fact that people tend to update their opinions after interacting in a discussion[55,56]; In terms of the implementation, for each node, j, another uniformly random number $R_j^{repulsing}$ in the interval $[0, 1)$ is generated, and for all cases where $R_j^{repulsing} < |\theta - b_j|/2$, b_j moves Δ away from θ, otherwise b_j moves Δ toward θ. If the resulting $|b_j| > 1$, it is set to $|b_j| = 1$. The iteration ends for all agents that are attracted by the post;
(5) ***Rewiring:*** For each follower j that is repulsed by the post, a stochastic bounded-confidence edge condition is checked. This condition is based on a probabilistic rule P_{rewire}, where the greater the difference between b_j and b_i, the more likely it is that the edge from j points to another random user instead of i. In other words, for each follower j that is repulsed by the post, a uniformly random number $R_j^{rewiring}$ in the interval $[0, 1)$ is generated. If $R_j^{rewiring} < P_{rewire}$ then the edge $\{v_i, v_j\}$ is rewired to $\{v_k, v_j\}$, where k, $k \neq i$ and $k \neq j$, is a random number in the range $[1, |V|]$.

Here, we modify the original model by introducing two additional types of users, each with their own behavior. A *priority user* posts content prioritized by the recommendation algorithm, which is always received by their followers. A *stubborn user* never changes their opinion. Notice that these two types may overlap; when they do, we call such a user an *ideologue*. Users not belonging to those types are called *normal users*. All other steps of the dynamics are implemented the same from the base model,[35] with the difference that now we consider directed edges (out-neighbors do not receive content).

For the sake of disambiguation, we note that naming users with the two simultaneous behaviors (priority and stubbornness) as ideologues does not directly imply they possess any topological advantage in terms of the network structure (e.g., increased number of followers). We only postulate these users have priority in terms of the recommendation algorithm and that they are committed to a point of view in the form of stubborn behavior. Although ideologues may not always be considered priority users in practical scenarios, their role in our model highlights their influence and shows how they can benefit from being prioritized.

In our extension, we choose two probabilistic rules for the posting step. The first one is calculated as

$$P_p^{con}(\theta, b_i) = \cos^2\left(\frac{\pi}{2}|\theta - b_i|\right). \quad \text{(Equation 4)}$$

Users who follow this rule do *conflicting posting*, since they may send content both aligned and opposed with their own opinion. This feature simulates two key mechanisms observed in real social networks. The first reflects confirmation bias, a well-researched phenomenon documented in previous studies,[57,58] which can be understood as the tendency to favor information that confirms their preexisting beliefs. Confirmation bias is also commonly represented in other computational models of social networks.[59,60] The second mechanism involves users' reactions to a piece of content; they actively post information that they disagree with, in some cases attached with small notes criticizing the content (e.g., a "quote retweet" in the X platform). Notice that this behavior is influenced by the difference between the user's own opinion b_i and the post θ. Consequently, users are more likely to exhibit confirmation bias behavior. Only users with extreme opinions (either 1 or -1) are equally likely to post both aligned and conflicting content. Normal users implement this posting behavior within our model dynamics.

We limit the behavior of stubborn users to only follow the confirmation bias. To do this, we use an alternative rule we call *aligned posting*, which is defined by

$$P_p^{ali}(\theta, b_i) = \begin{cases} \cos^2\left(\frac{\pi}{2}|\theta - b_i|\right), & \text{if } |\theta - b_i| \le 1 \\ 0, & \text{otherwise.} \end{cases} \quad \text{(Equation 5)}$$

This rule means they send content close to their opinion so as to represent a stylized zealot behavior. More specifically, P_p^{ali} represents users who are worried about presenting themselves as ideologically coherent to their followers so as to keep their loyalty.

In order to understand to what extent the effects we observe from priority users are generated by the prioritization algorithm or the user posting filter, we created another type of special user, the *priority advocate*. This user has priority in terms of the receiving filter but behaves according to a P_p^{ali} posting filter.

Both P_p^{con} and P_p^{ali} have been studied before in[35] (mapped to P_t^{pol} from Equation 2 and P_t^{sim} from Equation 3, respectively), but in a homogeneous scenario where all users follow a single rule. In this paper, we study various scenarios where users behave heterogeneously regarding which rule they use to post.

When the user i is not a priority user, the social network recommendation algorithm modulates how each follower j receives content from i through the probabilistic rule

$$P_r(b_i, b_j, \phi) = \cos^2\left(\frac{\pi}{2}|b_i - b_j| + \phi\right), \quad \text{(Equation 6)}$$

where the parameter ϕ controls the starting point of the cosine-squared function, effectively calibrating the recommendation algorithm. This is the same as the function P_d^l in Equation 5 of Ferraz de Arruda et al.[35]

The intuition behind the parameter ϕ is that when $\phi = 0$, for the majority of the cases, the recommendation algorithm is less permissive when the difference of opinions between the active user and their followers is around 1. That is, users are likely to receive posts from others who are very close to their opinion or very far away (in terms of our notation, $|b_i - b_j| \approx 0$ or $|b_i - b_j| \approx 2$, respectively).

On the other hand, for $\phi = \pi/2$, the algorithm operates in the opposite way; it is more permissive to posts when there is a medium distance between the active user's opinion and the followers. Conversely, it is likely to filter out posts when they strongly agree or strongly disagree. Further details of the behavior of this probabilistic filter are studied in the study by Ferraz de Arruda et al.[35] This mechanism turns out to be more permissive, allowing for broader communication between users with moderately different opinions. By interpreting this parameter, the algorithm considers that if two users have close opinions, the post should not be delivered because they already have similar opinions. This filter also avoids discussions between users with extreme opposite opinions.

In this paper, we introduce a novel rule for the recommendation algorithm. When the active user i is a priority user, the recommendation algorithm overrides the receiving step, and the content is automatically received by all followers. We highlight the post opinion θ is not considered in the receiving step, regardless of the rule adopted.

Finally, we define the rewiring rule as based on the difference between the opinions of the nodes i and j as

$$P_{rewire}(b_i, b_j) = \begin{cases} \cos^2\left(|b_i - b_j|\frac{\pi}{2}\right), & \text{if } |b_i - b_j| > 1 \\ 0, & \text{otherwise.} \end{cases} \quad \text{(Equation 7)}$$

Once again, this corresponds to P_{rewire} from Equation 7 in the study by Ferraz de Arruda et al.[35]

Notice the connections are typically rewired when individuals strongly disagree, but the new connection is chosen uniformly at random, which could mean a new disagreeing neighbor. That is, our model does not promote homophily based rewiring. It has been studied previously that homophily based connections contribute to the formation of echo chambers within social networks even when there are ideal conditions for opinion heterogeneity.[61] Given this observation, the random-based rewiring mechanism reveals how echo-chamber formation occurs dynamically and regardless of the opinion held by the new target node.

Algorithm 1. Opinion dynamics algorithm.

```
Input: Network G = (V,E), array of initial opinions b, posting filter P_p, receiving filter P_r(φ), number of iterations n_iter, attraction constant Δ, rewire flag r, rewire filter P_rewire, array indicating stubborn users stubborn, and array of priority users priority;
Output: Resulting network and opinion array.
for iteration ← 1 to n_iter do
  i ← random number in the range [1,|V|];                          ▷ Activating
  θ ← random number in the interval [−1,1);
  posting ← random number in the interval [0,1);                   ▷ Posting
  if posting < P_p(b[i],θ) then
    neighbors ← array of v_i neighbors;
    for j in neighbors do                                          ▷ Receiving
      receiving ← random number in the interval [0,1);
      if receiving < P_r(b[i],b[j]) or priority[i] == true then    ▷ Realigning
        repulsing ← random number in the interval [0,1);           ▷ Repulsion
        if repulsing < |θ − b[j]|/2 then
          if stubborn [j] == false then
            b[j] moves Δ away from θ;
            if b_j > 1 then
              b[j] ← 1;
            end if
            if b[j] < −1 then
              b[j] ← −1;
            end if
          end if
          rewire ← random number in the interval [0,1);            ▷ Rewiring
          if r and rewire < P_rewire(b[i],b[j]) then
            The edge {v_i,v_j} is rewired to {v_k,v_j}, where k, k≠i
            and k≠j, is a random number in the range [1,|V|];
          end if
        else                                                       ▷ Attraction
          if stubborn [j] == false then
            b[j] moves Δ toward θ;                                 ▷ If |b[j] − θ| < Δ, then b[j] ← θ
          end if
        end if
      end if
    end for
  end if
end for
```

When a user followed by stubborn users becomes active, the stubborn followers are still checked for rewiring even after skipping the re-aligning step because the active user may have had their opinion changed from other iterations. More details regarding the implementation can be seen in Algorithm 1.

The differences among the user types and their respective posting and receiving filters are summarized in Table 1. For convenience, in addition to the user type configurations, we also refer to a user with an opinion of either 1 or −1 as *extremist*, and another with an opinion of 0 as *centrist*.

In our experiments, we set up a fraction of users with one of the three special behaviors (described in Table 1) and then measure opinion polarization in the remaining users, which are normal users. Therefore, for convenience, we define $\widehat{b}$ as the distribution opinion of normal users only. Analogously, the corresponding $\widehat{b}_{NN}$ is the distribution of average neighbor opinions of normal users only, but considering all neighbors.

Our model could be tested with any other combination of the posting, receiving, and rewiring functions. Furthermore, other combinations of functions may also effectively represent the desired setup. However, here we chose this configuration set because it was previously studied in the study by Ferraz de Arruda et al.[35] and was found to be effective in modeling X. Valensise et al.[36] also studied the mechanisms of de Arruda et al.[35] in comparison with data from actual social networks. They found that Arruda et al. performs best when representing X data.

Table 1. Summary of the characteristics of each type of user

Behavior	Posting filter	Receiving filter	Can realign
Normal	P_p^{con} (Equation 4)	P_r (Equation 6)	Yes
Priority	P_p^{con} (Equation 4)	No filter	Yes
Priority Advocate	P_p^{ali} (Equation 5)	No filter	Yes
Stubborn	P_p^{ali} (Equation 5)	P_r (Equation 6)	No
Ideologue	P_p^{ali} (Equation 5)	No filter	No

The proposed user behaviors are defined by the posting and receiving filters and the ability to change opinions. The “Can realign” column refers to users who can change opinions. Normal users are also called non-priority users.

Base case

Before testing the modifications of the opinion model, we specify a base case to compare the changes in dynamics. This base case consists of a synthetic model configuration that approximates the bimodality coefficient measured on real social networks (see in the study by Ferraz de Arruda et al. [35]). To do this, we set all agents as normal users and run the dynamics for a range of ϕ values. We used an Erdős-Rényi[62] (ER) model of size $n = 10^4$ and average in-degree $z = 8$ (more details are shown in Section STAR Methods). The opinions of the agents are initialized with randomly uniform numbers between -1 and 1.

Figure 2 illustrates the bimodalities BC and BC_{hom} as a function of ϕ. The results regarding BC are different from those found in the previous study,[35] indicating t,hat the directionality of the network plays an important role in the behavior of the dynamics.

The highest value of $BC(b)$ was found for $\phi = 0.785$, where $\langle BC(b) \rangle = 0.58$. Moreover, this BC value (0.58) aligns closely with the values observed in empirical cases related to U.S. politics, specifically datasets on abortion, gun control, and Obamacare analyzed in the study by Ferraz de Arruda et al.[35] These datasets reported BC values ranging from 0.60 to 0.67. Therefore, we defined $\phi = 0.785$ as our base case. This base case represents a simplified polarized scenario in which the homogeneous behavior of users acts as a control for our experiments when we introduce heterogeneous behavior in fractions of users. A visual example of the initial network configuration is shown in Figure 3A. Furthermore, the base case is shown in Figure 3B.

Effect of priority users

We now systematically test what happens to the system as the dynamics are initialized with an increasing fraction of priority users. The initial configuration is the base case. We analyze the outcomes yielded by two posting functions concerning priority behavior, denoted as P_p^{con} and P_p^{ali}. It is important to clarify that the agents identified by P_p^{con} represent priority users and agents identified by P_p^{ali} represent priority advocate users (defined in Table 1). We also test different network structures (e.g., networks with communities and scale-free degree distributions), but the results are similar to those presented here. Thus, these results are not shown.

Figure 4 shows BC_{hom} for both tested cases. In both, increasing the number of priority users decreases BC_{hom}. Furthermore, this effect is observed to be stronger when priority users do align posting (priority advocate users), suggesting that echo chamber formation can be further reduced if priority users avoid conflicting content.

From the intuition gained from Figure 4 that the system moves away from echo chambers as the number of priority users doing aligned posting increases, we now build another step into our model. Up to now, we have not considered stubborn users. In order to understand how a fraction of priority users perturb the system with their influence, we observe what happens when these users are stubborn.

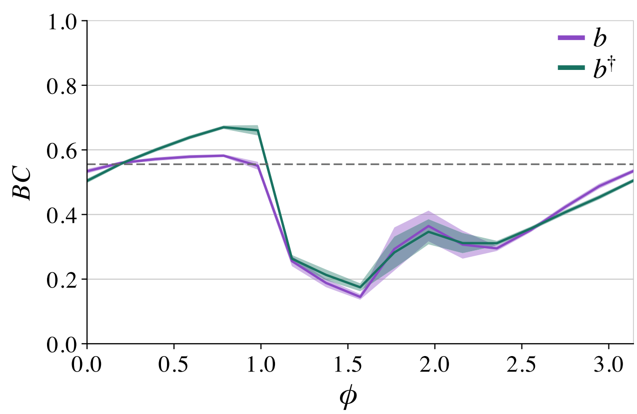


Figure 2. Echo chamber and polarization analysis by varying ϕ

BC and BC_{hom} by varying ϕ. The solid lines represent an average of 100 executions of the dynamics, and the shaded regions are the corresponding standard deviations. Distributions with the bimodality index above the dashed line BC = 5/9 can be considered bimodal.

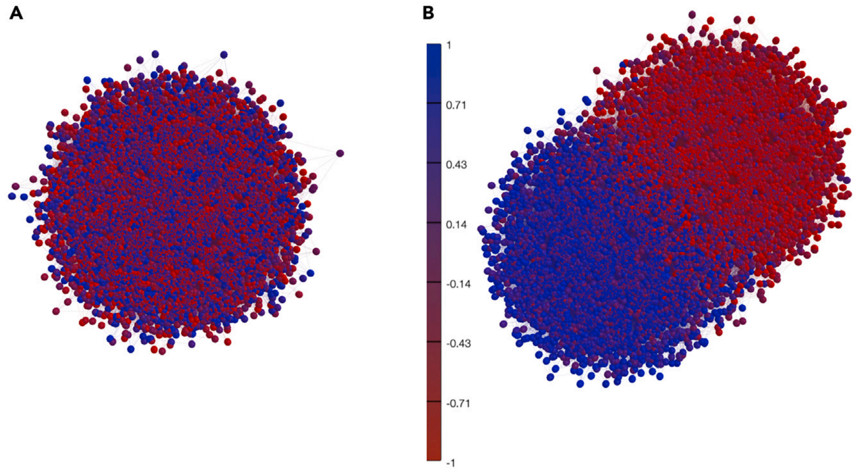


Figure 3. Visualization of the network used

(A) shows a random network (ER), with n = 10^4 nodes and average in-degree z = 8. (B) illustrates the resulting network after running the dynamics. The colors represent the user's opinions. The results shown in this figure represent the base case. These networks were plotted using the software implemented in Silva et al.[63]

How extremists can take advantage of being prioritized

Here, we investigate how stubbornness and extreme opinions affect the system when set on priority users (i.e., extreme ideologues) with two experiments. In the first, we set up a fraction of agents as ideologues with extreme opinions in the base case initialization to test the hypothesis that this type of user can benefit from being prioritized, in the sense of sharpening the measure of echo chamber formation. Then, we compare this setup with a similar one, but considering also stubborn users with extreme opinions instead of ideologues. Contrary to the results shown in the previous section, in both cases, increasing the percentage of ideologue or stubborn users strengthens their echo chambers. Furthermore, the ideologue case turns out to form stronger echo chambers for all values tested, showing that when ideologues are extremists, polarization, and division tend to increase faster. To understand whether the results are more related to the posting filter or the verified and stubborn users, we test an alternative posting rule (P_p^{con}) for ideological and stubborn users. The results are similar to those described here. Therefore, these results are not shown.

To have a more realistic comparison, we measure BC_{hom} against the fractions of priority and ideologue users. In this plot, the fraction of ideologues is always less than or equal to the percentage of priority users because all ideologues are also priority users. Figure 5 shows a contour map of this comparison. In Figure 5A, we keep the example of the homogeneous random network used for the base case. As expected, as the number of priority users increases, BC_{hom} tends to decrease. But with some stubbornness, BC_{hom} increases, showing that a few percent of stubbornness can make a difference.

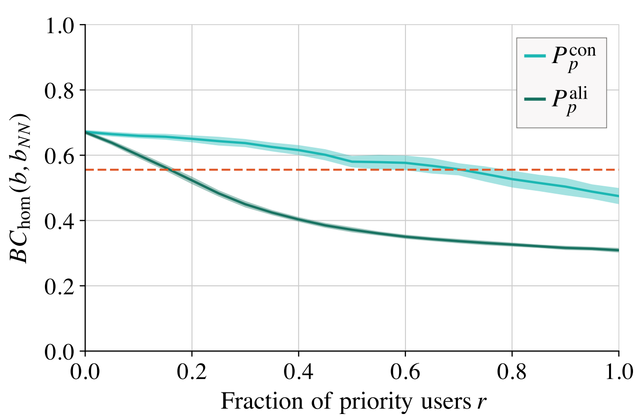


Figure 4. Effects of priority users with different posting behaviors

The measure BC_{hom} is used to compare two cases of posting functions of the priority users. When there is no priority user, the parametrization recovers the case shown in Figure 1A, meaning the system is parameterized at echo chamber formation. Priority users doing aligned posting drive the opinion distribution quickly out of polarization as they increase, but not when they do conflicting posting for most of the range shown. The solid lines represent an average of over 100 executions of the dynamics, and the shaded regions are the corresponding standard deviations.

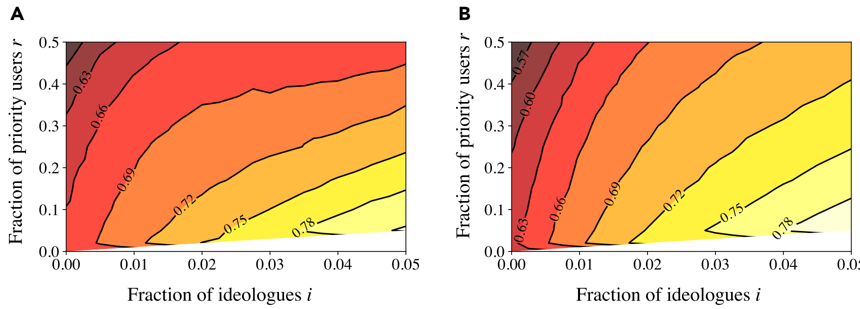


Figure 5. Effect of priority and ideologue users on echo chamber straightening

Contour maps of $BC_{hom}(\hat{b}, \hat{b}_{NN})$ by varying the percentages of ideologues (both stubborn and priority) and priority users. While both axes represent fractions of the total number of users N, the ideologues are randomly selected among the priority users to become ideologues (hence the white space for when there are more stubborn than priority users).(A) and (B) represent the dynamics executed on random homogeneous and heterogeneous networks, respectively. The plots represent an average of 100 runs of the dynamics, and for all tests performed to plot these numbers, the standard deviation was less than 0.02.

We also test another network structure, which consists of a network with a heterogeneous, scale-free (SF) degree distribution. The result obtained with this structure is shown in Figure 5B. For SF, the effect of the ideologues is stronger for higher percentages of priority users than for the homogeneous network. Without ideologues, the echo chambers are less well-defined. However, a smaller number of stubborn users can change the outcome of the dynamics and divide the network.

It is worth noting that our rewiring mechanism implies that, over time, the in-degree distribution will gradually become closer to a binomial (Poisson) distribution. Hence, using a synthetic network that starts with power-law distributed in-degree and Poisson-distributed out-degree produces similar results to the one with Poisson distributions for both degrees, which is confirmed by further testing (see Figure S2).

In the second experiment, we depart from the base case and investigate what happens when the system is parameterized to start at consensus. This is achieved by replicating all conditions of the base case but with $\phi = 1.473$. This value is shown to be close to the minimum bimodality coefficient in Figure 2. While the minimum coefficient was identified at $\pi/2$, that point is also near a ϕ range of increased dispersion of BC, which we opted to avoid. Subsequent tests with values close to our chosen one yield similar results.

This time, we gradually increase the fraction of stubborn extremist users while monitoring the BC_{hom} as an order parameter, which is shown in Figure 6A. A small minority of such users are able to drive the system from consensus to polarization with echo chambers, similar to a social norm change via critical mass.[64] We then proceed with this analysis but consider ideologues instead of stubborn users (see Figure 6C). In this case, the transition happens even faster (at a fraction about 20% smaller). While the transition looks smoother in the case these users are only stubborn (Figure 6B) for our choice of the order parameter (BC_{hom}), it has a richer behavior when they are ideologues as observed in the mass emerging right before the second peak of the distributions (Figure 6D).

This result is also related to the effect caused by the *committed minorities*,[65,66] which are groups of individuals strongly committed to a particular opinion or behavior. Here, the group of stubborn users can be understood as a committed minority. In summary, we find that when prioritization is available to every user, the committed minority can take advantage of it and change the system with an even smaller fraction of users. In other words, when stubborn users become ideologues, they are more likely to divide the opinions in the network. Finally, it is worth noting that at variance with some models involving committed minorities, our "committed minority" remains as such for the whole evolution of the dynamics.

Opinion polarization can still happen with centrist users

Here, we analyze the scenario in which the users with special behavior are centrists. To do so, we set their opinions to zero. In particular, we start again from the base case, whose ϕ is parameterized to reach polarization, and set out to find whether the centrist users are able to tip the system away from polarization.

In this case, we opt to measure polarization with BC instead of BC_{hom}. It should be noticed that the measure BC_{hom} was designed to accentuate the bimodality present in both the overall opinion distribution b and the opinion alignment b_{NN} by combining them into a single distribution. In the case either of these is not bimodal prior to the rotation, BC_{hom} does not reliably capture the simultaneous aspects of a polarized and divided system (as discussed in background). On the other hand, the angle of 45° is chosen to punish cases in which the density map is populated in the second and fourth quadrants (see Figure 1D). In this way, the resulting $b^{\dagger}$ from the rotation will likely return a value higher than the threshold 5/9 (see, for instance, Figure 1E). Since b_{NN} is not bimodal for all results shown in this section, we measure the following results with BC.

To understand the effect of priority on ideologue centrists, we compare BC between two scenarios: (i) the dynamics are initialized with an increasing fraction of stubborn users, and (ii) the dynamics are initialized with an increasing fraction of ideologue users. This result is shown in Figure 7). It shows that centrist stubborn users without priority are mostly not able to overcome the polarized setting. On the other hand, when they are set with priority (hence becoming ideologues), the system rapidly shifts toward consensus.

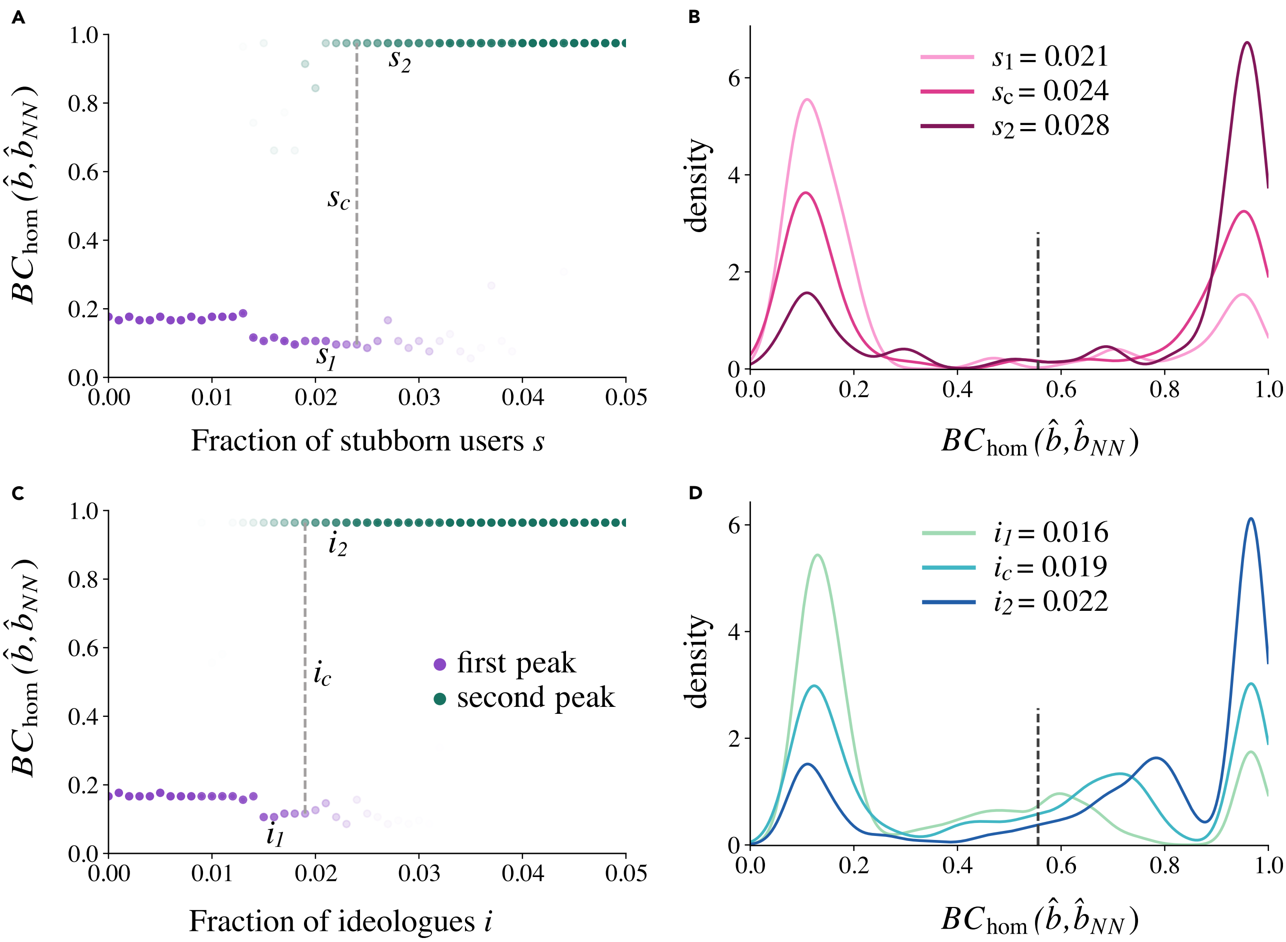


Figure 6. Echo chamber formation accelerated by extremist ideologues

The system undergoes a transition from consensus to polarization with echo chambers captured by the measure BC_{hom} as the fraction of extremist users increases. In (A), the extremist users are stubborn, and in (C), they are ideologues. In those panels, values of s and i are associated with the two peaks of the distribution of $BC_{hom}(\widehat{b}, \widehat{b}_{NN})$ values obtained from 500 simulations. These peaks are obtained by dividing the 500 values around the threshold 5/ 9. The transparency of points corresponds to how many simulations fall into each side. The purple dashed line is drawn for $s_c = 0.024$ in (A) and $i_c = 0.019$ in (C), where the threshold of bimodality divides the distribution of BC_{hom} into two sides with approximately equivalent mass. That is, at i_c and s_c, echo chambers are likely to be formed or not with similar odds. In (B) and (D), we show samples of those distributions for three values of s and three values of i, respectively, including s_c and i_c. The dark dashed vertical line marks the 5/9 threshold. For more information about the transition, see Figure S4.

When the fraction of stubborn users is close to 0.008, the opinion distribution $\widehat{b}$ tends to be unimodal, but for slightly higher values, it changes to be bimodal (see Figure 7B). Furthermore, for higher values of the fraction of stubborn users, $\widehat{b}$ is significantly bimodal (see Figure 7C). In the case of ideologues, as the fraction of stubborn users increases, $BC(\widehat{b})$ decreases until the scenario of consensus, but maintains a diverse range of opinions. In other words, the opinion distributions shown in Figures 7C and 7D have an average close to zero but with a considerably high dispersion.

In all cases, there are no echo chambers because of a mechanistic implication of the rewiring rule in our model. That is, centrist stubborn users have a probability 0 of losing followers (see Equation 7) and may only gain them over time. After a sufficient number of iterations, most of the normal users are not connected to each other but only to the centrist stubborn users, which is akin to a hub-to-spoke core-periphery structure.[67] This effect is visible in panels (b) to (e) of Figure 7, as the fraction s raises from 0.02 to 0.20.

In all scenarios examined, whether stubborn users are centrists or extremists, and whether they are ideologues or not, they tend to be more connected to others in the network. Seemingly, the stubborn users form a core structure that accumulates links within or around, but not outside of it, in a "hub-and-spoke" fashion.[67] Figure S3 shows that as the number of stubborn users increases, there is a tendency for users to be more connected to them. Figure S3A shows the case where the stubborn users are extremists. With and without being ideologues, the fraction of connections to stubborn users tends to be the same. For 50% of stubborn users, about one out of five connections are in the periphery, not involving stubborn users. In the case of centrists, Figure S3B, this effect is even clearer. At around 20% of stubborn users, the proportion of edges in the periphery vanishes, meaning they either connect normal users to stubborn ones or two stubborn users. However, for

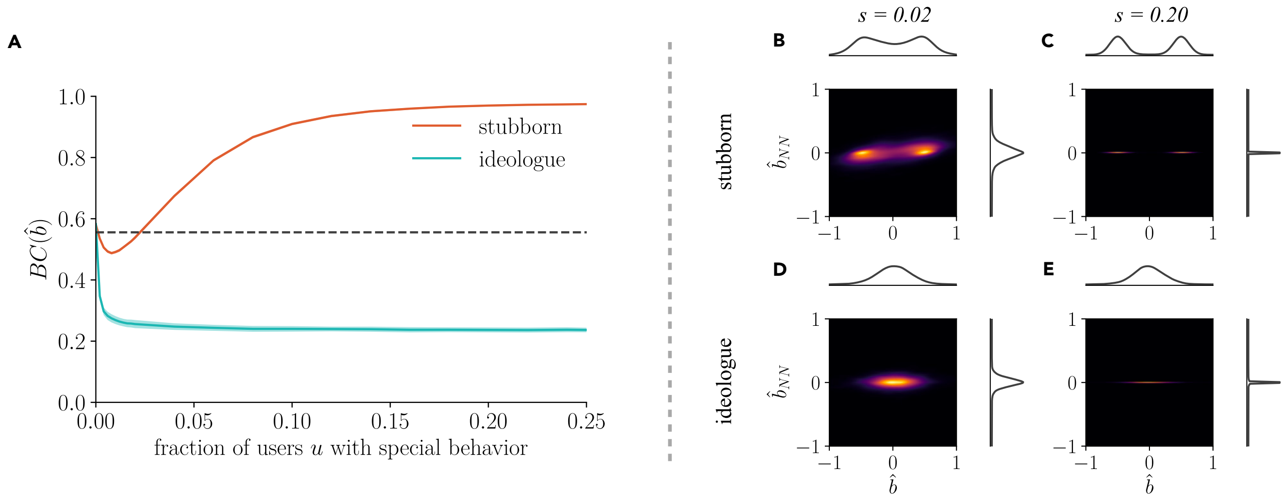


Figure 7. Impact of centrists on network polarization

(A) shows the opinion polarization measured by $BC(\hat{b})$ as the fraction of users u with special behavior (either stubborn or ideologue) increases. The solid lines represent an average of over 100 executions of the dynamics, and the shaded regions are the corresponding standard deviations. All the users in the fraction u of the total are centrists. (B–E) show the $\hat{b} \times \hat{b}_{NN}$ for a single run of the dynamics, for different values of u.(B) and (C) display examples obtained from normal users, and (D) and (E) for priority users. Even though both configurations start from a polarized point ($BC(\hat{b}) > 5/9$), stubborn centrists are not able to sustain a depolarization of the system, as opposed to centrist ideologues.

fractions of stubborn users lower than 0.15, the scenario with non-priority tends to retain stubborn users more connected than in the case with priority users. In both cases, the regular and stubborn users tend to be connected to the stubborn, which leads to the formation of a more connected core, creating the core-periphery structure in the network. We run these tests using ER networks with the same parameters as the previous tests.

DISCUSSION

In this paper, we have conducted a comprehensive mechanistic investigation of the impact of a policy change in social media on opinion dynamics. Motivated by X, which recently changed its policy regarding verified users, we proposed a model to simulate the impact of priority users (i.e., users whose content reaches more followers). As a complementary contribution, we also proposed a unified measure for polarization and echo chamber formation, the homophilic bimodality coefficient. Using this framework, we were able to test several scenarios, including cases where stubborn users take advantage of priority. Our model does not account for homophily based rewiring because homophily based links are known to contribute to echo chamber formation. Given this observation, by using a random-based rewiring mechanism (previously investigated in the study by Ferraz de Arruda et al[35]), we show the formation of echo chambers can also be explained via mechanisms that rely on platform activity rather than a static user-to-user comparison.

In the first experiment, we tested how the inclusion of priority users could change polarization and echo chamber formation. We showed that when priority users with normal behavior post conflicting content, the system remains polarized with echo chambers for much longer than when the same proportion of priority users post aligned content. In both cases, however, prioritizing users helps to depolarize and reduce echo chambers. This result illustrates that priority users can help reduce echo chambers, but their behavior also plays an important role. Moreover, in both cases, the proportion of priority users that could mitigate the echo chambers is too high for a real social network. In the best case, almost 20% of users should have priority to vanish the echo chambers.

Next, we set out to understand what happens when the dynamics admit ideologue and stubborn extremists. Starting from a polarized setting, we saw that for a fixed fraction of priority users, increasing the number of those who are ideologues amplifies the measurement of echo chambers. Then, we verified that even when the system is not supposed to polarize, a small fraction of stubborn extremists are able to induce polarization and cause echo chambers to form. Importantly, such a transition is accelerated when the extremists become ideologues. This is strikingly similar to the change of social norms through reaching a critical mass threshold found in the literature[64] and constitutes a cautionary tale for unintended emergent effects of policy changes in complex systems with intricate dynamics.

We analyzed the influence of centrist users and found that echo chambers disappear for a small percentage of centrist ideologues. However, when a similar situation involves a fraction of stubborn users instead of ideologues, the remaining users become highly polarized. In both cases, the changes in opinion polarization do not imply the formation of echo chambers.

In addition to the effects of priority users on polarization and echo chamber formation, our model also sheds light on the role of ideologues in shaping network dynamics. Regardless of whether users exhibit centrist or extremist behavior, we observe a remarkable phenomenon: when the number of ideologues in the network becomes sufficiently large, the majority of nodes tend to be connected to these ideologues.

In other words, a significant portion of the messages exchanged in the network are either sent to or received from these influential users. The latter feature has already been reported for collective human dynamics spurred by X, as in the case of social movements such as the 15M in Spain.[68] Our findings confirm that ideologues and priority behavior play a crucial role in shaping the flow of information and opinions within the social network. Their ability to reach a wide audience allows them to have a significant impact on the formation and reinforcement of echo chambers. When social network algorithms prioritize visibility over content control, the users may be able to reach others to reinforce their opinions in groups, potentially entrenching echo chamber structures.

Furthermore, our investigation suggests that the influence of these users is not limited to scenarios in which they are committed to extreme opinions. Even centrist ideologues, who may appear as a moderating force on the surface, can have a significant impact on the opinion dynamics when in enough numbers. This implies that addressing the issue of echo chambers and polarization requires a nuanced understanding of the influence exerted by all types of users, regardless of their position on the ideological spectrum.

When put in perspective, the two user behavior mechanisms we investigated, priority users and stubborn users, can be associated with a myriad of diverse behaviors encountered in empirical settings. Prioritizing a user amplifies their influence by means of amplifying their reach. In turn, the platform's minority of stubborn users can disturb the system in macroscopic terms. While the stubbornness characteristic can happen in principle by a cause unknown to the platform (for instance, from user self-organization or an externality), prioritization is something we modeled and worked through as a platform's policy. This implies that any policy that boosts user influence should be monitored closely by platforms to ensure the gain of influence is not used maliciously.

In conclusion, our model shows that both priority users and ideologues play critical roles in shaping the dynamics of polarization and echo chamber formation in social media. We acknowledge that all models simplify the explanation of how social networks and opinions behave. However, our findings highlight the need for careful consideration of platform policies that affect the visibility of certain users and the extent of their influence. Therefore, all stakeholders should carefully analyze the possible consequences before changing the policies of social media. Because small changes to the platforms can result in unexpected results, stakeholders should also monitor them on an ongoing basis to understand whether extremists are taking advantage of them and change their policies as needed. Regulatory bodies should examine whether to step in when mass media communication changes rapidly, possibly leading to unintended consequences that could weaken democracies and society. By understanding and quantifying the impact of these factors, effective strategies can be developed to mitigate division and promote healthier and more balanced information consumption.

In this study, we considered a mechanistic model to represent the changes in the social network. However, due to the lack of data from X, we were not able to validate our model by comparing it with real data. Similar to previous studies[69–71], in future work, we intend to consider data to parameterize our model and test it to simulate different scenarios and social media platforms.

Limitations of the study

The main limitation of this type of study lies in the inherent simplifications of the model, which does not fully capture the complexity of the real world. In addition, the lack of access to empirical data on the priority accounts limits the validation of the model results against real-world observations. We, however, believe that the reliability of the model is supported by its consistency with the results of previous studies. Besides, a potential methodological limitation concerns the interpretation of BC_{hom}. In cases where the social network is polarized but lacks clear structural divisions, BC_{hom} may yield high values that do not reflect true echo chambers. Consequently, BC_{hom} may not provide meaningful results in such a scenario.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Henrique Ferraz de Arruda (h.f.arruda@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- No data were generated as part of this study.
- All original code has been deposited at https://github.com/hfarruda/Echo-chamber-priority-users and is publicly available as of the date of publication.
- Any additional information required to carry out the simulations reported in this paper is available on request from the lead contact.

ACKNOWLEDGMENTS

The authors would like to thank Carolina Mattsson for fruitful discussions. Y.M. was partially supported by the Government of Aragon, Spain and "ERDF A way of making Europe" through grant E36-23R (FENOL), and by Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación (MCIN/AEI/10.13039/501100011033) Grant No. PID2020-115800GB-I00. We acknowledge the use of the computational resources of COSNET Lab at Institute BIFI, which was funded by Banco Santander (grant Santander-UZ 2020/0274) and the Government of Aragon (grant UZ-164255). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization: H.F.A., K.A.O., and Y.M.; methodology: H.F.A., K.A.O., and Y.M.; investigation: H.F.A., K.A.O., and Y.M.; writing—original draft: H.F.A., K.A.O., and Y.M.; writing—review an editing: H.F.A., K.A.O., and Y.M.; validation, formal analysis: H.F.A. and K.A.O.; software: H.F.A. and K.A.O. All authors have read, edited, and approved the final manuscript.

DECLARATION OF INTERESTS

There are no conflicts of interest to disclose.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111098.

Received: June 2, 2024
Revised: July 25, 2024
Accepted: September 30, 2024
Published: October 22, 2024

REFERENCES

1. Karami, A., Lundy, M., Webb, F., and Dwivedi, Y.K. (2020). Twitter and Research: A Systematic Literature Review Through Text Mining. IEEE Access *8*, 67698–67717. https://doi.org/10.1109/ACCESS.2020.2983656.
2. Legacy Verification Policy (2023. https://help.twitter.com/en/managing-your-account/legacy-verification-policy.
3. How to Get the Blue Checkmark on X (2023. https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts.
4. Sîrbu, A., Pedreschi, D., Giannotti, F., and Kertész, J. (2019). Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. PLoS One *14*, 1–20. https://doi.org/10.1371/journal.pone.0213246.
5. Facebook (2023). Meta Verified on Facebook. https://www.facebook.com/help/565322618581808.
6. Instagram (2023). Verified Badges. https://help.instagram.com/854227311295302.
7. Watts, D.J., and Dodds, P.S. (2007). Influentials, Networks, and Public Opinion Formation. J. Consum. Res. *34*, 441–458. https://doi.org/10.1086/518527. https://academic.oup.com/jcr/article-pdf/34/4/441/17929486/34-4-441.pdf.
8. Centola, D., Pescosolido, B., and Smith, E.B. (2021). Influencers, Backfire Effects, and the Power of the Periphery. In Personal Networks: Classic Readings and New Directions in Egocentric Analysis, M.L. Small and B.L. Perry, eds. (Cambridge University Press), pp. 73–86. Structural Analysis in the Social Sciences. https://doi.org/10.1017/9781108878296.005.
9. Ott, B.L. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. Crit. Stud. Mass Commun. *34*, 59–68. https://doi.org/10.1080/15295036.2016.1266686.
10. Morris, M.R., Counts, S., Roseway, A., Hoff, A., and Schwarz, J. (2012). Tweeting Is Believing? Understanding Microblog Credibility Perceptions. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. CSCW '12 (Seattle, Washington, USA: Association for Computing Machinery), pp. 441–450. https://doi.org/10.1145/2145204.2145274.
11. Paul, I., Khattar, A., Kumaraguru, P., Gupta, M., and Chopra, S. (2019). Elites Tweet? Characterizing the Twitter Verified User Network. In 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW), pp. 278–285. https://doi.org/10.1109/ICDEW.2019.00006.
12. Vaidya, T., Votipka, D., Mazurek, M.L., and Sherr, M. (2019). Does Being Verified Make You More Credible? Account Verification's Effect on Tweet Credibility Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19 (Glasgow, Scotland Uk: Association for Computing Machinery)9781450359702, pp. 1–13. https://doi.org/10.1145/3290605.3300755.
13. Alkhamees, M., Alsaleem, S., Al-Qurishi, M., Al-Rubaian, M., and Hussain, A. (2021). User trustworthiness in online social networks: A systematic review. Appl. Soft Comput. *103*, 107159. https://doi.org/10.1016/j.asoc.2021.107159.
14. Le, F. (2023). En France, les candidats aux élections ne pourront pas être abonnés à twitter blue. https://www.lefigaro.fr/secteur/high-tech/les-candidats-aux-elections-ne-pourront-pas-etre-abonnes-a-twitter-blue-20230713.
15. Bak-Coleman, J.B., Alfano, M., Barfuss, W., Bergstrom, C.T., Centeno, M.A., Couzin, I.D., Donges, J.F., Galesic, M., Gersick, A.S., Jacquet, J., et al. (2021). Stewardship of global collective behavior. Proc. Natl. Acad. Sci. USA *118*, e2025764118. https://doi.org/10.1073/pnas.2025764118.
16. Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C.R., and Hertwig, R. (2020). How behavioural sciences can promote truth, autonomy and democratic discourse online. Nat. Hum. Behav. *4*, 1102–1109. https://doi.org/10.1038/s41562-020-0889-7.
17. Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., and Quattrociocchi, W. (2016). The spreading of misinformation online. Proc. Natl. Acad. Sci. USA *113*, 554–559. https://doi.org/10.1073/pnas.1517441113.
18. Brady, W.J., Wills, J.A., Jost, J.T., Tucker, J.A., and Van Bavel, J.J. (2017). Emotion shapes the diffusion of moralized content in social networks. Proc. Natl. Acad. Sci. USA *114*, 7313–7318. https://doi.org/10.1073/pnas.1618923114.
19. Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., and Flammini, A. (2021). Political polarization on twitter. Proc. Intern. AAAI Conf. Web Soc. Media *5*, 89–96.
20. Narayanan, V., Kollanyi, B., Hajela, R., Barthwal, A., Marchal, N., and Howard, P.N. (2019). News and Information over Facebook and WhatsApp during the Indian Election Campaign. Data Memo 2. https://demtech.oii.ox.ac.uk/research/posts/news-and-information-over-facebook-and-whatsapp-during-the-indian-election-campaign/.
21. Baker, S.R., Baksy, A., Bloom, N., Davis, S.J., and Rodden, J.A. (2020). Elections, Political Polarization, and Economic Uncertainty (National Bureau of Economic Research), pp. 1–26. Working Paper. https://doi.org/10.3386/w27961.
22. Machado, C., Kira, B., Hirsch, G., Marchal, N., Kollanyi, B., Howard, P.N., Lederer, T., and Barash, V. (2018). News and Political Information Consumption in Brazil: Mapping the First Round of the 2018 Brazilian Presidential Election on Twitter. Data Memo 4. https://demtech.oii.ox.ac.uk/research/posts/news-and-political-information-consumption-in-brazil-mapping-the-2018-brazilian-presidential-election-on-twitter/.
23. Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A., and Quattrociocchi, W. (2017a). Mapping social dynamics on Facebook: The Brexit

debate. Soc. Network. *50*, 6–16. https://doi.org/10.1016/j.socnet.2017.02.002.
24. Noorazar, H., Vixie, K.R., Talebanpour, A., and Hu, Y. (2020). From classical to modern opinion dynamics. Int. J. Mod. Phys. C *31*, 2050101. https://doi.org/10.1142/S0129183120501016.
25. Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. (2000). Mixing beliefs among interacting agents. Adv. Complex Syst. *03*, 87–98.
26. DeGroot, M.H. (1974). Reaching a consensus. J. Am. Stat. Assoc. *69*, 118–121.
27. Rainer, H., and Krause, U. (2002). Opinion dynamics and bounded confidence: models, analysis and simulation. J. Artifical Soc. Soc. Simul. *5*, 1.
28. Galam, S. (2002). Minority opinion spreading in random geometry. Eur. Phys. J. B *25*, 403–406.
29. Sasahara, K., Chen, W., Peng, H., Ciampaglia, G.L., Flammini, A., and Menczer, F. (2021). Social influence and unfollowing accelerate the emergence of echo chambers. J. Comput. Soc. Sci. *4*, 381–402.
30. He, M., Li, B., and Luo, L. (2004). Sznajd model with “social temperature” and defender on small-world networks. Int. J. Mod. Phys. C *15*, 997–1003.
31. Sznajd-Weron, K., Sznajd, J., and Weron, T. (2021). A review on the sznajd model—20 years after. Phys. Stat. Mech. Appl. *565*, 125537.
32. Holme, P., and Newman, M.E.J. (2006). Nonequilibrium phase transition in the coevolution of networks and opinions. Phys. Rev. E *74*, 056108.
33. Benatti, A., de Arruda, H.F., Silva, F.N., Comin, C.H., and da Fontoura Costa, L. (2020). Opinion diversity and social bubbles in adaptive Sznajd networks. J. Stat. Mech. *2020*, 023407.
34. Baumann, F., Lorenz-Spreen, P., Sokolov, I.M., and Starnini, M. (2020). Modeling echo chambers and polarization dynamics in social networks. Phys. Rev. Lett. *124*, 048301.
35. Ferraz de Arruda, H., Maciel Cardoso, F., Ferraz de Arruda, G., R Hernández, A., da Fontoura Costa, L., and Moreno, Y. (2022). Modelling how social network algorithms can influence opinion polarization. Inf. Sci. *588*, 265–278.
36. Valensise, C.M., Cinelli, M., and Quattrociocchi, W. (2023). The drivers of online polarization: Fitting models to data. Inf. Sci. *642*, 119152. https://doi.org/10.1016/j.ins.2023.119152.
37. Mobilia, M., Petersen, A., and Redner, S. (2007). On the role of zealotry in the voter model. J. Stat. Mech. *2007*, P08029.
38. Mobilia, M. (2003). Does a single zealot affect an infinite group of voters? Phys. Rev. Lett. *91*, 028701.
39. Sîrbu, A., Loreto, V., Servedio, V.D., and Tria, F. (2017). Opinion Dynamics: Models, Extensions and External Effects. Participatory Sensing, Opinions and Collective Awareness, pp. 363–401.
40. Namatame, A., and Chen, S.-H. (2016). Agent-Based Modeling and Network Dynamics (Oxford University Press), pp. 1–341. https://doi.org/10.1093/acprof:oso/9780198708285.001.0001.
41. Kubin, E., and von Sikorski, C. (2021). The role of (social) media in political polarization: a systematic review. Ann. Intern. Comm. Assoc. *45*, 188–206. https://doi.org/10.1080/23808985.2021.1976070.
42. Terren, L., and Borge-Bravo, R. (2021). Echo chambers on social media: A systematic review of the literature. Rev. Commun. Res. *9*, 99–118.
43. Ross Arguedas, A., Robertson, C., Fletcher, R., and Nielsen, R. (2022). Echo Chambers, Filter Bubbles, and Polarisation: A Literature Review. Tech. Rep..
44. Hohmann, M., Devriendt, K., and Coscia, M. (2023). Quantifying ideological polarization on a network using generalized Euclidean distance. Sci. Adv. *9*, eabq2044. https://doi.org/10.1126/sciadv.abq2044.
45. Cota, W., Ferreira, S.C., Pastor-Satorras, R., and Starnini, M. (2019). Quantifying echo chamber effects in information spreading over political communication networks. EPJ Data Sci. *8*, 35.
46. Martin-Gutierrez, S., Losada, J.C., and Benito, R.M. (2023). Multipolar social systems: Measuring polarization beyond dichotomous contexts. Chaos. Solit. Fractals *169*, 113244.
47. Ferraz de Arruda, H., Benatti, A., Nascimento Silva, F., Comin, C.H., and da Fontoura Costa, L. (2021). Contrarian effects and echo chamber formation in opinion dynamics. J. Phys. Complex. *2*, 025010.
48. Cinelli, M., Morales, G.D.F., Galeazzi, A., Quattrociocchi, W., and Starnini, M. (2021). The echo chamber effect on social media. Proc. Nat. Acad. Sci. USA *118*, e2023301118.
49. Pfister, R., Schwarz, K.A., Janczyk, M., Dale, R., and Freeman, J.B. (2013). Good things peak in pairs: a note on the bimodality coefficient. Front. Psychol. *4*, 700.
50. Kokoska, S., and Zwillinger, D. (2000). CRC Standard Probability and Statistics Tables and Formulae (CRC Press).
51. Jolliffe, I.T. (2003). Principal component analysis. Technometrics *45*, 276.
52. Gewers, F.L., Ferreira, G.R., Arruda, H.F.d., Silva, F.N., Comin, C.H., Amancio, D.R., and Costa, L.d.F. (2021). Principal component analysis: A natural approach to data exploration. ACM Comput. Surv. *54*, 1–34.
53. Holland, P.W., Laskey, K.B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. Soc. Network. *5*, 109–137.
54. Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring Network Structure, Dynamics, and Function Using NetworkX (Los Alamos, NM: Los Alamos National Lab (LANL)). Tech. rep.
55. Moscovici, S., and Zavalloni, M. (1969). The group as a polarizer of attitudes. J. Person. Soc. Psychol. *12*, 125–135.
56. Isenberg, D.J. (1986). Group polarization: A critical review and meta-analysis. J. Person. Soc. Psychol. *50*, 1141–1151.
57. Molden, D.C., and Higgins, E.T. (2008). How preferences for eager versus vigilant judgment strategies affect self-serving conclusions. J. Exp. Soc. Psychol. *44*, 1219–1228.
58. Taber, C.S., and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. Am. J. Pol. Sci. *50*, 755–769.
59. Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H.E., and Quattrociocchi, W. (2017b). Modeling confirmation bias and polarization. Sci. Rep. *7*, 40391.
60. Alvim, M.S., Amorim, B., Knight, S., Quintero, S., and Valencia, F. (2021). A Multi-Agent Model for Polarization under Confirmation Bias in Social Networks. In International Conference on Formal Techniques for Distributed Objects, Components, and Systems (Springer), pp. 22–41.
61. Blex, C., and Yasseri, T. (2022). Positive algorithmic bias cannot stop fragmentation in homophilic networks. J. Math. Sociol. *46*, 80–97.
62. Erdős, P., and Rényi, A. (1960). On the evolution of random graphs. Pub. Math. Institute Hungarian Acad. Sci. *5*, 17–61.
63. Silva, F.N., Amancio, D.R., Bardosova, M., Costa, L.d.F., and Oliveira, O.N., Jr. (2016). Using network science and text analytics to produce surveys in a scientific topic. J. Informetrics *10*, 487–502. https://doi.org/10.1016/j.joi.2016.03.008.
64. Centola, D., Becker, J., Brackbill, D., and Baronchelli, A. (2018). Experimental evidence for tipping points in social convention. Science *360*, 1116–1119. https://doi.org/10.1126/science.aas8827.
65. Galam, S., and Jacobs, F. (2007). The role of inflexible minorities in the breaking of democratic opinion dynamics. Phys. Stat. Mech. Appl. *381*, 366–376.
66. Xie, J., Sreenivasan, S., Korniss, G., Zhang, W., Lim, C., and Szymanski, B.K. (2011). Social consensus through the influence of committed minorities. Physical Review E—Statistical, Nonlinear, and. Soft Matter Phys. *84*, 011130.
67. Gallagher, R.J., Young, J.-G., and Welles, B.F. (2021). A clarified typology of core-periphery structure in networks. Sci. Adv. *7*, eabc9800. https://doi.org/10.1126/sciadv.abc9800.
68. Borge-Holthoefer, J., Rivero, A., García, I., Cauhé, E., Ferrer, A., Ferrer, D., Francos, D., Iñiguez, D., Pérez, M.P., Ruiz, G., et al. (2011). Structural and Dynamical Patterns on Online Social Networks: The Spanish May 15th Movement as a Case Study. PLoS One *6*, 1–8. https://doi.org/10.1371/journal.pone.0023883.
69. Jones, M.I., Sirianni, A.D., and Fu, F. (2022). Polarization, abstention, and the median voter theorem. Humanit. Soc. Sci. Commun. *9*, 43.
70. Wang, X., Sirianni, A.D., Tang, S., Zheng, Z., and Fu, F. (2020). Public discourse and social network echo chambers driven by socio-cognitive biases. Phys. Rev. X *10*, 041042.
71. Evans, T., and Fu, F. (2018). Opinion formation on dynamic networks: identifying conditions for the emergence of partisan echo chambers. R. Soc. Open Sci. *5*, 181122.
72. Newman, M. (2018). Networks (Oxford University Press), pp. 1–1042. https://doi.org/10.1093/oso/9780198805090.001.0001.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Codes for reproducing all the results	The authors	https://github.com/hfarruda/Echo-chamber-priority-users

METHOD DETAILS

In this section, we present more details about the configuration of the opinion model used and how we analyze the results obtained. We started by setting up the base case (shown in Figure 1A). We set all the social network users with conflicting post contents (P_p^{con} from Equation 4) and varied ϕ of P_r (Equation 6) between 0 and π (shown in Figure 2). The base case was determined by the highest values of BC (Equation 1) and BC_{hom} (Equation 3), with $\phi = 0.785$, which was chosen as it represents the scenario of maximum polarization and echo chamber formation.

The parameter ϕ in P_r controls the starting point of the cosine-squared function, allowing us to explore the model at different levels of polarization and consensus. Each realization of the model has 10^8 iterations to ensure the system reaches a nearly stable state, as were all of the following simulations.

The base case test was run on an Erdős-Rényi[62] network with $n = 10^4$ nodes and an edge inclusion probability of $p = 1.6 \times 10^{-3}$, resulting in an undirected network with an average degree of 16. The choice of p was based on the desire to create a network with a moderate average degree, which would provide a balance between network connectivity and computational efficiency. Next, to convert this network to directed, we randomly chose the edge direction with equal probability for each possibility. In this way, the resulting network is directed, with an average in-degree and out-degree close to 8. The use of directed networks is crucial for our opinion dynamics model, as it allows for the implementation of the information flow from one user to another through the followership relation.

Furthermore, to create the networks of Figure 5B, we use a configuration model[72] whose degree sequence is inverse-sampled from a power law exponent of $\lambda = 2.43$ and a minimum degree of three, meaning the network has an average degree of about 8. In this case, both the degree distributions of the in- and out-edges follow this power law distribution.

Departing from the base case, we tested the other hypotheses described in this paper. This allows for a comprehensive comparison between each tested configuration and the results of the base case. In order to guarantee that the results obtained are not due to statistical fluctuations, we ran the dynamics 100 times for all results, except for the case of the results presented in Figure 7, which were run 500 times. This extensive simulation approach ensures the statistical robustness and reliability of our findings, enabling us to draw meaningful conclusions from the data.

As a complementary result, shown in Figure 7, we tested a different scenario. Specifically, we considered $\phi = 1.473$, a parameter value that leads to a scenario characterized by consensus and no echo chamber formation. In this parametrization, both BC and BC_{hom} approach their minimum values with low standard deviations, indicating convergence to a stable state. This scenario played a crucial role in our investigation as it allowed us to examine the proportion of extremist ideologues required to induce polarization and form echo chambers. By exploring the impact of different proportions of extremist ideologues in a low-polarization setting, we can understand the conditions under which our model can lead to significant shifts in the overall opinion distribution and the formation of echo chambers.

QUANTIFICATION AND STATISTICAL ANALYSIS

There are no quantification or statistical analyses to include in this study.