**ᵒ Computational Social Networks**

# A model of opinion and propagation structure polarization in social media

Hafizh A. Prasetya* iD and Tsuyoshi Murata

*Correspondence:
hafizh@net.c.titech.ac.id
Murata Laboratory, Tokyo
Institute of Technology,
W8-59 2-12-1, Ookayama,
Meguro, Tokyo 152-8552,
Japan

## Abstract

The issue of polarization in online social media has been gaining attention in recent years amid the changing political landscapes of many parts of the world. Several studies empirically observed the existence of echo chambers in online social media, stimulating a slew of works that tries to model the phenomenon via opinion modeling. Here, we propose a model of opinion dynamics centered around the notion that opinion changes are invoked by news exposure. Our model comes with parameters for opinions and connection strength which are updated through news propagation. We simulate the propagation of multiple news under the model in synthetic networks and observe the evolution of the model's parameters and the propagation structure induced. Unlike previous models, our model successfully exhibited not only polarization of opinion, but also segregated propagation structure. By analyzing the results of our simulations, we found that the formation probability of echo chambers is primarily connected to the news polarization. However, it is also affected by intolerance to dissimilar opinions and how quickly individuals update their opinions. Through simulations on Twitter networks, we found that the behavior of the model is reproducible across different network structure and sizes.

**Keywords:** Echo chambers, Polarization, Opinion modeling, News propagation

## Background

In recent years, multiple empirical studies have revealed the existence of segregated communities in online social media interaction [4, 7, 9, 14, 18, 39, 73], where like-minded people interact with little exposure to different viewpoints. This phenomenon is commonly referred to as the echo chamber effect. More concretely, the term echo chambers refers to a state where online interactions are conducted in a polarized pattern; within an online social network, there exist groups that correspond to a certain belief (e.g., political left or right), and most interactions happen between users from the same group. A study into the nature of these polarized communities is consequential; echo chambers has been increasingly studied as of late due to its association with significant political events such as Brexit and the surprise win of Donald Trump in the 2016 US election. It is also known to cause significant harms to the process of discussion and democracy [68, 69, 72]. At its worst, echo chambers can limit one's viewpoints, reinforce personal biases, and foster environments where hoaxes and misinformation thrives.

The ubiquity of online social media adds to the urgency of this line of work. A recent survey from the Pew Research Institute [64] has found that at least 70% of adults in the United States currently use social media sites; the majority of them admitted to visiting the sites daily. Other researches [53, 61, 75] also indicate that social media sites have been also increasingly used as news sources, with around two-thirds of adults in the United States admitting to being exposed to news from social media sites. At least 30% of internet users worldwide used social media sites as news sources in the past week. In the context of politics, online social media can also be seen as an important political tool [2].

One popular approach to investigating echo chambers and opinion polarization is through modeling [15, 24, 30, 49, 52, 55, 63, 67, 71, 74]. This approach borrows ideas from the study of opinion dynamics which has been established since as early as the 1970s. Early models of opinion focus on the establishment of consensus [16, 48, 70]. In time, the study diversified into models that enable the coexistence of multiple opinions, with the existence of two conflicting opinions (i.e., polarization) as a special case. These new models are either based on an established consensus model (e.g., [24, 49, 55, 67]) or new mechanisms drawn from sociological and psychological theories [6, 50]. Latest polarization models introduce mechanisms based on the idiosyncrasies of online social media; e.g., the rewiring of social connection [59] and algorithmic bias [63].

As previously mentioned, we can intuitively split polarization study into empirical and modeling works. However, we observe that previous works can also be divided into those that focus on the segregation of the propagation structure and those that focuses on the polarization of opinion (typically represented as some kind of score/value). Together, these two phenomena constitute the echo chamber effect.

In this work, we concern ourselves with both of the aforementioned focuses and adopt a different approach to modeling polarization. We propose a news propagation-centered model where opinions change as a consequence of news propagation. By adopting this point of view, it becomes possible to capture both opinion polarization and propagation structure segregation. We consider it very likely that this assumed scenario is happening in real social network for two reasons. As expounded earlier, the first one is how recent studies had found that online social media sites are also widely used as news sources. The second reason is related to the effect partisan news (i.e., news that are framed to promote a specific viewpoint rather than remaining neutral) have been found to have on individual opinions. Feldman [17, Jones 40] found apparent shifts of opinion/stance in people who are exposed to partisan news. Together, these findings indicate that not only people are actively searching for news through social media, but exposure to such news can lead to a change in opinion. Propagation/diffusion is an established topic in the field of network science and our model is inspired by the well-studied Independent Cascade (IC) model [42].

This article extends on the initial modeling work done in [56]. In addition to the contents in [56], we explain in more detail the background works and include detailed results of the ablation study on the model's mechanism. Focusing on the effect of network topology which was previously unexplored, results from additional simulations on real Twitter networks are presented. The rest of the article is structured as follows. "Related works" section establishes the foundation of the proposed model by going over the related works. "Model description" section then describes the model in detail. The setup of the simulation, including the description of the network data, is given in

"Experimental setup" section. "Simulation result" section provides the result of our simulation, which is then discussed in the following "Discussion and analysis" section. Finally, we conclude the work in "Conclusions" section.

## Related works

### Modeling opinion dynamics

Early works of opinion dynamics emerged from the field of sociophysics, pioneered by works drawing from the magnetic spin model. This means that the opinion of the population is limited to a binary range: 0 and 1 (or − 1 and + 1). A classic example of this would be the voter model [38], in which an individual tries to adopt the majority of its neighbors' opinions in every time step. A more sophisticated concept of opinion adoption is formulated in the Sznajd model [70], in which interaction between two picked individuals will result in opinion change of their neighbors. This change is determined by the opinion agreement of the two original individuals.

While a binary, discrete opinion range might be deemed sufficient in initial investigations of opinion consensus, it could not portray nuanced opinions. In time, opinion models which deal with a continuous opinion spectrum were developed. Two popular examples are the averaging model [16] and the bounded confidence model (BCM) [48]. These models produce a more realistic depiction of opinion dynamics in exchange for mathematical tractability. The averaging model is a discrete averaging process wherein each time step, a node in a network updates its own opinion according to the weighted average of its neighbors' opinions until convergence. In the BCM, interactions only happen among those whose difference in opinion is less than a given threshold. As such it can be seen as modeling selective exposure.

It should be noted that the aforementioned works do not deal with the phenomenon of polarization. Subsequent works that tried to model polarization did so by implementing various mechanisms capable of inducing polarization. These models either extend the previous four models (e.g., [15, 24, 30, 49, 52, 55, 63, 67, 71, 74]) or take inspiration from other branches of science to introduce new mechanisms [6, 50, 57]. The dynamics and the characteristics of the models belonging to the former group are roughly similar to their parent models; on the other hand, members of the latter group generally introduce new point-of-views. Mäs and Flache [50] proposed a model where the process of opinion influence is modeled as an evaluation and exchange of arguments. The distribution of opinions in the model is then a function of both the internal opinions of the system and the external opinions of the argument. Banisch and Olbrich [6] provide another explanation of polarization by proposing an opinion model based on social feedback; the opinion of an individual is based on his evaluation of his neighbors' feedbacks. The result shows that existing community structures of the underlying network confine the opinion of individuals inside them.

In terms of introducing polarization, the most common method involves assuming either the existence of a persistent agent with a fixed opinion (e.g., [30, 52, 71]) or negative influence where the opinion of an individual will move away from any opinion he/she disagrees with (e.g., [24, 60]). More recent polarization models introduce mechanisms based on the idiosyncrasies of online social media; e.g., the rewiring of social connection [59, 74], and algorithmic bias [63]. Recent polarization modeling

works generally strive to contextually capture the process of polarization by including rules that directly mimic online social media. These efforts align with a recent call for more complex and realistic opinion model previously voiced in [66] which our model also tries to answer.

In comparison, our proposed model centers on the idea that opinion dynamics are affected by news propagation. It draws from two different research areas (opinion dynamics and network propagation) and allow us to also observe the segregation of propagation structure, similar to what has been observed in empirical studies [9, 14, 18, 39]. As seen in Table 1, our model mainly differentiates itself from the rest in terms of interaction mode. The issue with opinion exchanges in models where a random pair of nodes interact every time step is that they become a series of disconnected episodes. This is not the case in online social media where opinions are broadcast and propagated throughout the network via news; a successful propagation can lead to another one. This process is exactly described by our model. Another difference is the fact that we also model the connection strength between individuals, an aspect that has yet to be fully explored in previous works. While our model involves multiple parameters and exhibits more complexity than the majority of previous works, it reflects the real interaction system in a particular way that has never been done before.

### Polarization drivers

While there are multiple factors that have been attributed as the driving force behind echo chambers and polarization, these factors can be roughly categorized into two: micro-level behavior of individuals collectively causing a macro-level phenomenon (polarization) and external factors originating from the system that governs the interaction between individuals.

Factors belonging to the former category can be seen as originating from the phenomenon of cognitive dissonance. Cognitive dissonance is an extensively studied phenomenon within the field of psychology, and refers to the discomfort experienced by people when presented by information that challenges their beliefs or decisions [22, 23]. This discomfort in turn encourages certain behavior that causes polarization.

Among these behaviors, one that is attributed the most as the cause of polarization is selective exposure [25, 43] which is the tendency for individuals to expose themselves only to information that aligns with their preexisting beliefs. Research on selective exposure predates the mass adoption of internet; the behavior has been observed in control studies where participants were asked to examine pieces of information after forming an initial decision/opinion [41]. More recently within the context of online social networks, selective exposure has also been observed empirically on Facebook [4].

Another behavior is called confirmation bias, or biased assimilation [54]. This concept is related the assimilation or interpretation of information, where regardless of what the actual information is (contradictory or supportive), individuals have a higher tendency to interpret it as supportive, reinforcing their beliefs. Similar to selective exposure, research on biased assimilation have existed prior to the internet in the form of controlled experiments [47]. Other than selective exposure and confirmation bias, other behaviors have also been proposed as the driver of polarization such as free choice [11]

**Table 1  Comparison of the proposed model to popular opinion model families and recent polarization models**

| Model name | Opinion spectrum | Stochastic process | Interaction mode | Model parameters | Modeled aspect |
|---|---|---|---|---|---|
| Families of opinion models | | | | | |
| Voter [13] | Binary | Node selection | Singular | Opinion | Majority influence |
| Sznajd [70] | Binary | Pair selection | Pair | Opinion | Social validation |
| Averaging [16] | Continuous | None | Global | Opinion | Opinion averaging, positive influence |
| Bounded confidence [48] | Continuous | Pair selection | Pair | Opinion, deviation threshold, confidence value | Positive influence, selective exposure, confidence |
| Recent polarization models | | | | | |
| Argument exchange [50] | Continuous | Pair selection, argument selection | Pair | Opinion, arguments | Argument exchange, homophily |
| Social feedback [6] | Binary | Pair selection | Global | Opinion, opinion perception, learning rate, deviation probability | Social validation |
| Algorithmic bias [63] | Continuous | Pair selection | Pair | Opinion, deviation threshold, bias strength | Positive influence, selective exposure, algorithmic bias |
| Bounded confidence w/ propaganda [71] | Continuous | Pair selection | Pair | Opinion, deviation threshold, confidence value, propaganda threshold | Positive influence, selective exposure, confidence, propaganda influence |
| Bounded confidence w/ emotion [67] | Continuous | Pair selection | Pair | Opinion, deviation threshold, confidence value | Positive influence, selective exposure, confidence, emotion |
| Algorithmic bias [74] | Continuous | Pair selection | Pair | Opinion | Positive influence, selective exposure, algorithmic bias |
| Confidence-tolerance [59] | Continuous | Pair selection | Pair | Opinion, deviation threshold, confidence, rewiring probability | Positive influence, selective exposure, confidence dynamic, connection rewiring |
| News percolation [72] | Continuous | Seed selection | Propagating broadcast | Opinion, news fitness, sharing threshold | News influence, homogeneity |
| Proposed model | Continuous | Seed selection, propagation success | Propagating broadcast | Individual opinion, connection strength, news opinion, update rate, tolerance | News influence, selective exposure, connection strength |

and induced compliance [21]. Interested readers can consult [29] for more information on these behavior.

Moving on to the second category, one factor that has been connected to polarization is the algorithmic personalization employed by many social media sites [10]. As widely known, contents which appear in the home page of a Facebook or Twitter users are curated by an algorithm which aims order to maximize engagement or relevance. The general idea is that these relevant contents are usually contents which align with a user's predisposition, creating a string of posts which can reinforce personal opinions. Another factor is the existence of 'stubborn' agents whose opinions are static and extreme (e.g., political twitter bots [19, 76], partisan news sources). The idea is through exchanging opinions with these stubborn agents, opinions will inevitably shift to the extreme ends of the spectrum since the opinion of these agents are unchangeable.

The model proposed in this work incorporates concepts from both categories: selective exposure and the existence of partisan news items. This is not to say that other factors have no contribution in inducing polarization. As described above, each factor plausibly contributes to the complex social system in some way. Our model simply provides an abstraction and focus on the concepts we chose to include in our mechanism.

### Opinion drivers

Most models of polarization were designed with the underlying assumption that the polarization of a network is reflected by the polarization of individuals that construct it. In these models, opinions of individuals are commonly driven by another individual(s)' opinion; interaction between individuals will result in some sort of change to their opinion. There are two ways opinion influence could happen between individuals in these models: positive and negative. Positive influence means when individuals interact, their opinion move in the positive direction to each other in order to reach consensus. While the exact reason behind this kind of conformist behavior can yet be conclusively pointed out, several explanations have been offered. For example, classic social comparison theory states that people use similar other as guides when evaluating their own opinions [20]. Another explanation is that this positive influence is done in order to gain acceptance from others [12]. Positive influence is the normal kind of opinion influence commonly assumed in opinion modeling [13, 16, 48].

The other kind of opinion influence is termed negative influence. This refers to the condition when two individuals with differing opinion interact and their opinion moves away from each other, amplifying their difference as a result. The existence of negative influence can be understood as implied by the same explanation for positive influence. For example, people may use dissimilar others when evaluating their opinions, guiding their opinion to the negative directions. However, this kind of negative influence is a stronger assumption placed only on a few models of polarization [24, 60] due to weak empirical support. As such, only the former positive influence is assumed in the model we propose in this work. A more comprehensive review and discussion on positive and negative can be found in [37].

Instead of being directly driven by the opinion of others, this work proposes an indirect opinion driver in the form of information news. As previously explained, the scenario of news and information spread in networks extends intuitively to the phenomenon of polarization. This approach is motivated by works documenting the power partisan news have over the opinions of readers [17, 40]. Given that partisan news have been found to be able to shift opinion, we argue that the opinion polarization—and in turn, echo chambers—observed in online social medias happens through opinion shifts caused by news sharing. The idea of a news-driven opinion model has been proposed before [49, 72] but never in tandem with the exploration of the news spreading pattern itself, which is something that we particularly focus on.
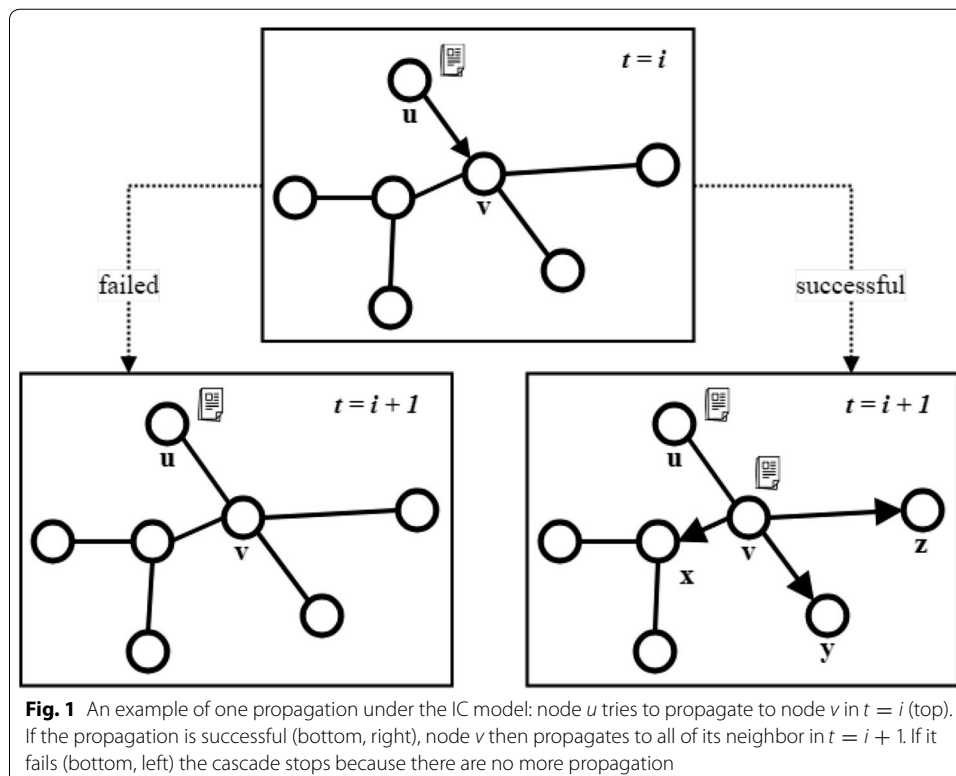
### Propagation in complex network

Propagation or diffusion in complex networks is a well-studied topic within the field of network science and encompasses many different kinds of spreading process, e.g., information, ideas, and diseases. Our model is influenced in particular by the IC model,

originally proposed by Kempe [42] to formulate the problem of *Influence Maximization*: given a group of people connected in a certain network topology and assuming that information propagates through two connected individuals, which $k$ seed individuals will lead to the biggest spread of information? An example of one propagation under this model can be seen in Fig. 1. The propagation-focused scenario of the influence maximization problem is very similar to the one studied within this work, making it natural to extend the IC model to the problem of polarization modeling.

Several previous works pertaining to online polarization have also considered polarization through the lens of propagation in complex networks. Garimella et al. formulated the problem of minimizing polarization in news propagation [26] in a scenario where two pieces of opposing information are propagating simultaneously in a network. Polarization is then defined as an imbalance in the cascade size of the two information and thus minimizing it is equal to finding two sets of seed nodes that produced cascades with similar sizes. Another work formulated the problem as a maximization problem of propagation diversity [3]. Given a set of individuals with their own leanings and a set of information also equipped with their own leanings, the problem is to find the individual-information assignment that maximizes the diversity of the resulting cascade.

Both of the works defined above uses the IC model as their propagation model. Our work is more closely related to the latter in the way that they both consider the leanings of not only individuals but the spreading information. While the view that online polarization can be mainly seen as a process of propagation in complex networks is relatively new, the aforementioned works serve as precedents and motivation of this work.



**Fig. 1** An example of one propagation under the IC model: node *u* tries to propagate to node *v* in $t = i$ (top). If the propagation is successful (bottom, right), node *v* then propagates to all of its neighbor in $t = i + 1$. If it fails (bottom, left) the cascade stops because there are no more propagation

### Social learning

Previously, "Modeling opinion dynamics" section presented a brief digest of opinion modeling works. It should be noted that these models are operating within the assumption that users/agents continuously update their opinions over multiple cascades. There exists a line of work that models sequential decision-making processes of agents over a social network. In this setting commonly referred to as bayesian social learning, agents in a given social network are required to make a single optimized decision based on two sources of information: its personal knowledge on the given subject and decisions of previous agents (hence the bayesian namesake). This process, initially investigated by [5, 8, 45], can be thought of as a single cascade of opinion adoption. Interested readers can refer to [1, 58] for a comprehensive review of the field and exposition on how they relate with our study.

### Model description

The proposed model, called Polarizing Independent Cascade (PIC), adopts the news propagation scenario of the IC model with three key differences: the existence of model parameters, a new formula for propagation probability, and an update process for the

---

**Algorithm 1:** PIC

**Given:**
- $G = (V, E)$: An undirected graph representing a set of connected users
- $Q = \{q_u \in [-1, 1], u \in V\}$: The set of every node's opinion
- $C = \{c_{uv}, (u, v) \in E\}$: The set of every edge's connection strength
- $I = \{i_i \in [-1, 1], 1 < i < T\}$: The set of every simulated news items' opinion score
- $S = \{s_i, 1 < i < T\}$: The set of initial spreader nodes for every news item
- $\eta$: The update rate for opinion and connection strength
- $B, \mu$: Parameters for the logistic function (optional)

**Procedure:**

```
for i = 1 to T do  // For every news item

    /* Initialize variables:
        - t: current time step, set to 1 at the beginning of every news item's spread
        - A_t^i: the set of nodes that can still spread news item i at t
        - Inf^i: the set of nodes that have been infected by news item i
                                                                                */
    t ← 1 ;
    A_t^i ← S_i ;
    Inf^i ← Inf^i ∪ S_i ;
    while A_t^i ≠ ∅ do
        for u ∈ A_t^i do // While an active node exists

            for v : e_uv ∈ E, v ∉ Inf^i do // Spread to uninfected neighbors

                if Random() < P_uv(q_v, c_uv, i_i) then
                    // Neighbor become infected and active
                    Inf^i ← Inf^i ∪ {v} ;
                    A_{t+1}^i ← A_{t+1}^i ∪ {v} ;
                    // Opinion shifts and connection strengthens
                    q_v ← U_q(u_v, i_i, η) = q_v + (η × sign(i_i − q_v)) ;
                    c_uv ← U_c(q_uv) = c_uv + η ;
                else
                    // Connection weakens
                    c_uv ← U_c(q_uv) = c_uv − η ;
                end
            end
        end
        // Go to next time step
        t ← t + 1;
    end
end
```

---

parameters. A pseudocode of the model's spreading procedure is given in Algorithm 1.

This model is based on the idea that the expression of opinion in online social media happens in the form of news sharing, i.e., a change in opinion is a result of news
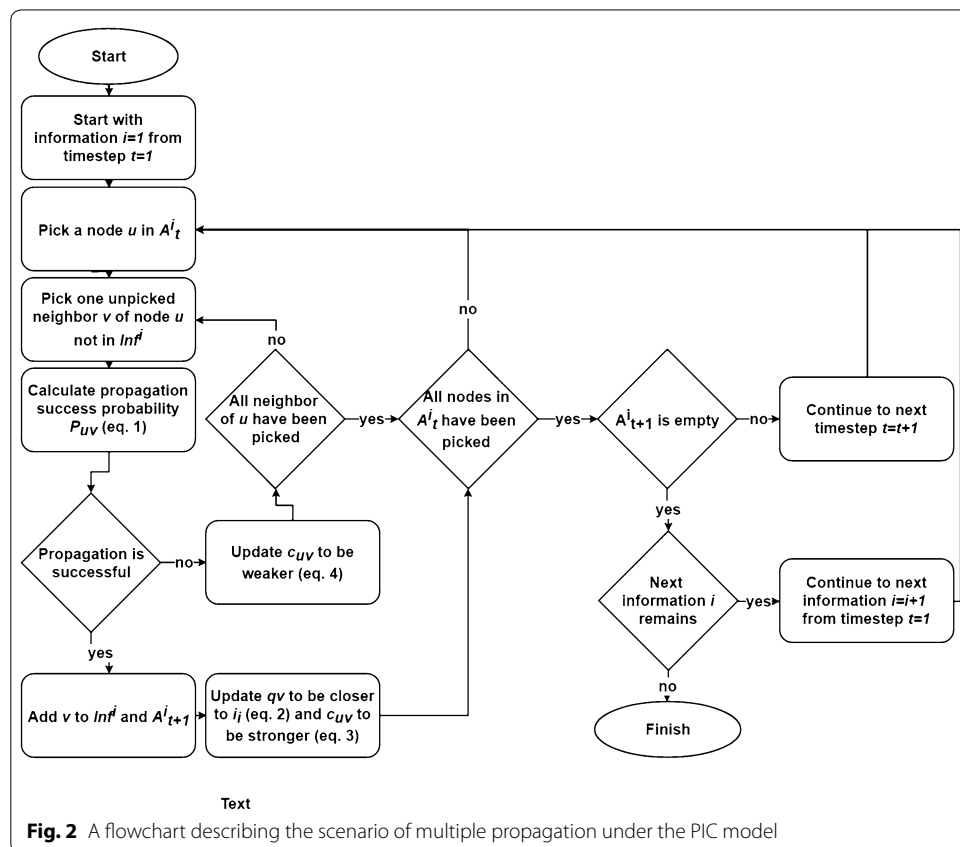
propagating through users in the network. As listed in the beginning of the pseudocode, the PIC model has 7 parameters: $Q, C, I, S, \eta, B$, and $\mu$. Later, we will investigate how some of these parameters affect the polarization dynamic through sets of simulations. Note also that throughout the procedure, some values are stored inside the variables $t, A_t^i$, and $Inf^i$. These variables should not be mistaken as parameters of the model since they are temporary in nature. Figure 2 explains how the PIC model works in the scenario of multiple propagation.

One fundamental change to the original IC model that we propose pertains to the propagation probability. In the original model, it is directly reflected by the connection strength and is assumed to be static. This is not the case in our scenario where the opinion of an individual factors in his/her decision to propagate information. To mimic this behavior, we formulate $P_{uv}$, the probability of a successful propagation, as follows:

$$P_{uv} = (c_{uv} \times c_{uv}) + ((1 - c_{uv}) \times (1 - f(|i_i - q_v|/2))), \tag{1}$$

where $f(x)$ is a function that models the growth of opinion similarity's influence. Here, we test the effect of using three functions: linear, logistic, and step function.

The behavior of $P_{uv}$ can be simply described as follows: the probability that a neighbor $v$ of user $u$ will accept a piece of news with opinion $i_i$ is a combination of the strength of his connection with $u$ and the similarity between his opinion and $i$. The proportion of these two factors depends on how strongly connected $u$ and $v$ are. If $u$ and $v$ are strongly connected, then the probability will depend more on the strength



**Fig. 2** A flowchart describing the scenario of multiple propagation under the PIC model

of this connection. Otherwise, the probability will depend more on how similar the opinion of $v$ is to the opinion of the news.

The first half of the equation directly takes from the original IC model, which assumes that propagation success is determined by a probability value associated with each edge. This probability illustrates the strength of influence between the connected nodes. Here, we refer to this strength as *connection strength* and denote it as $c_{uv}$ Meanwhile, the second half of the equation simulates the concept of selective exposure observed in [4]. The similarity function $f(x)$ in which the similarity value is feed to enables the tweaking of the strength of selective exposure. Putting $c_{uv}$ as the variable that controls the proportion of the two factors creates a duality of behavior between weak and strong connections. This design is inspired by the work of Granovetter on weak and strong ties [33]. Weakly connected users can be thought of as acquaintances (weak ties) and strongly connected users as close friends (strong ties).

Updates to the model parameters happen as a consequence of propagation; depending on whether the information is accepted, the update functions behave differently. If $v$ decides to accept the information and spread it, it updates its own opinion $q_v$ to be closer to $i_i$ and increases its connection strength with $v$ $c_{uv}$ according to the equations:

$$q'_v = U_q(u_v, i_i, \eta) = q_v + (\eta \times \text{sign}(i_i - q_v)) \tag{2}$$

$$c_{uv} = U_c(q_{uv}) = c_{uv} + \eta \tag{3}$$

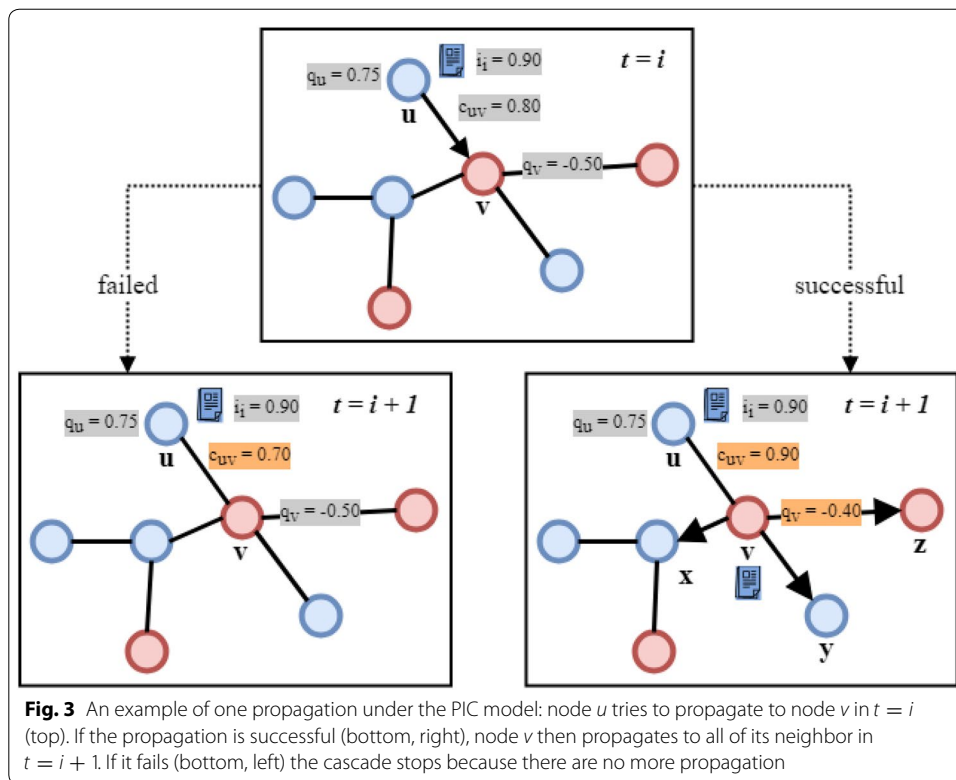Should $v$ decide to reject the information, $c_{uv}$ will instead decrease according to the equation:

$$c_{uv} = U_c(q_{uv}) = c_{uv} - \eta \tag{4}$$

$U_q$ is motivated by the effect partisan news has been found to cause on people's opinion [17, 40]. Essentially, the equation stipulates that any accepted news will directly persuade users to shift their opinions towards itself, following the results found in [17]. On the other hand, $U_c$ basically assumes that the more information successfully propagates between two users, the stronger the bond between them will become. Conversely, the more information failed to propagate, the weaker the bond will become. Finally, it should be noted that the proposed model can be perceived as a dynamic extension of the IC model. It reduces to the normal IC model when $\eta = 0$ and the ratio is 1 instead of $c_{uv}$. An example of one propagation can be seen in Fig. 3.

## Experimental setup
### General procedure
The simulation scenario performed on the model consists of sequentially propagating $T$ items that represent news on the social network and observing the changes that happen to the parameters, $Q$ and $C$. However, if all $T$ news have the same opinion score $i$, it is obvious that opinions will eventually converge to $i$. Therefore, the particular scenario of interest in this work is when items whose opinion scores are sampled from a certain distribution are propagated. The distribution of $I$ then will reflect the polarization of news, making it possible to see how the polarization of news translates to the polarization of

**Fig. 3** An example of one propagation under the PIC model: node $u$ tries to propagate to node $v$ in $t = i$ (top). If the propagation is successful (bottom, right), node $v$ then propagates to all of its neighbor in $t = i + 1$. If it fails (bottom, left) the cascade stops because there are no more propagation

opinion and the segregation of propagation structure. All simulations are performed with $T = 10,000$. This number is obtained from initial experiments and has shown to always be sufficient for the model to enter a state in which the distribution of opinions either converges or diverges completely, making any further change in modality unlikely. $I$ is sampled from a symmetrical beta distribution:

$$f(x, \Theta) = \frac{\gamma(2\Theta)((x+1)/2)^{\Theta-1}((1-x)/2)^{\Theta-1}}{2\gamma(a)\gamma(b)}, \tag{5}$$

where $\gamma(z)$ stands for the gamma function. The simulation parameter $\Theta$ then signifies the degree of polarization of the news items. A $\Theta$ value of 1 yields a uniform distribution and anything below yields a u-shaped polarized distribution. Every news item starts propagating from one random seed node.

Initial values for $q$ and $c$ are randomly sampled from a given range. For $q$, this is the full opinion range, $[-1, 1]$. However, this is not the case for $c$, which are sampled from the low range of $[0, 0.5]$. This is due to the fact that in preliminary simulations, it was found that pre-established strong connections cause news to always propagate between nodes with different opinions, often leading to an opinion convergence.

Every value presented in "Simulation result" section is an average over multiple simulations. The number of repetition for simulations performed on synthetic networks is 50 while for simulations performed on real networks it is 20 due to constraints on time and computational resource.

### Network dataset

For this work, we utilize two network types for different purposes: synthetic and real. First, synthetic networks are used to understand the model through ablation studies and extensive parameter exploration. Afterward, the model is simulated over real Twitter follow networks in order to ascertain the consistency of its behavior on real networks.

#### Synthetic networks

All of the synthetic networks used in our simulations are generated using the Lanchichinetti Fortunato Radicchi (LFR) algorithm [44]. This algorithm works by first creating a network with power-law degree distribution before assigning them to communities and rewiring the edges according to the given parameter. The power-law degree distribution of the generated network has also been found in multiple social networks. In addition, the algorithm also has the ability to control community structures, making it suitable for the task of model exploration. Unless specified otherwise, the synthetic networks comprise 250 nodes with approximately around 2000 edges. This size was decided through several experiments with consideration to the running time of the simulation. With the exception of the set of simulation that involves different community number, the synthetic networks are divided into two communities with roughly equal sizes. The mixing value $\mu$ between these communities is set to 0.1.

#### Twitter networks

For validation, simulations are performed over real network structure taken from Twitter's follow relationships between users that discuss certain topics. In these networks, nodes signify users while an edge signifies a follow relationship. While Twitter's Follow mechanism is one-directional, the resulting graph is considered as undirected. The Twitter follow data set used in the research was first constructed and used by [28] to develop a measure of polarization over a given network. The data set consists of multiple networks each constructed from users that participated in a Twitter of a certain topic through a given hashtag. A brief overview of the Twitter networks is presented along with the simulation results in "Simulation result" section.

### Polarization measures

Here we explain two measures that are used in our analysis of the simulation results. Each of these measures pertains to polarization in different ways. The first measure, called $m$-value, concerns itself with the polarization in the opinion distribution; the second one, called edge homogeneity, is more closely related to the segregation of propagation structure and the relationship between connection strengths and opinions in the social network.

#### m-value

$M$-value is a simple measure of modality introduced by [34]. In this work, it is used to measure the modality of the users' final opinion distribution. Given a histogram $x$ with $n$ bins (including zero bin terminators) and maximum bin value $M$, $m$-value is defined as:

$$m\text{-value}(x) = \frac{1}{M} \sum_{i=2}^{n} |x_i - x_{i-1}|. \tag{6}$$

A perfectly bimodal distribution (two peaks of the same height) yields $m$-value $= 4$ and a perfectly unimodal distribution yields $m$-value $= 2$. Any number higher than 4 means the distribution is spread across multiple peaks and is not considered to be polarized. As proposed by [34], we use the value of 2.4 as a rough threshold to determine whether a distribution can be considered as bimodal.

It should be noted $m$-value does not take into account the position of the peaks on the polarization spectrum. This makes it by itself insufficient for determining polarization; after all, a distribution with two separated peaks on $-0.1$ and $0.1$ will have the same $m$-value with a distribution two peaks at the opposite ends of the opinion spectrum, $-1$ and 1. Therefore, we also measure the mean absolute opinion score of the distribution, $\overline{q}_{\text{abs}}$, and visually inspect individual histograms. We use all three measures in the interpretation of our results. Together, these measures can give us an idea of the number of peaks and their position in the spectrum of the final opinion distribution.

### Edge homogeneity

Contrary to models which only updates opinion, our proposed model also updates the connection strength between users. This adds an extra dimension to the notion of polarization; we can now observe whether users with a similar opinion are strongly connected to each other and weakly connected to everyone else. To do this we use *edge homogeneity*, a measure of opinion similarity between two connected nodes [72] defined as $\text{hom}(u, v) = q_u \times q_v$. We modify this measure to include the strength of the edge and averages it for every edge in a given graph to obtain *average weighted mean edge homogeneity*:

$$\overline{\text{hom}}_w = \frac{\sum_{e(u,v)} c_{uv} q_u q_v}{\sum_{e(u,v)} c_{uv}}. \tag{7}$$

High $\overline{\text{hom}}_w$ indicates a network structure where nodes of similar, strong opinions are connected by strong edges and nodes with different or weak opinions are connected by weak edges. By observing how $\overline{\text{hom}}_w$ changes through our simulation, we can see how the propagation structure evolves w.r.t. to the evolution of opinions and connection strengths.

## Simulation result

In this section, we present the result of our simulations. The sets of simulations performed in this study are divided into three main parts, each performed with different aims in mind: ablation study, parameter exploration, and simulations on Twitter networks.

### Ablation study

The PIC model comes with multiple mechanisms and rules. Therefore, it is important to ascertain that each mechanism is necessary to induce polarization. The model should strive to be as simple as possible while still being able to produce the phenomenon of interest (i.e., polarization). While the majority of the mechanisms are backed by empirical findings (selective exposure [4], opinion update [17, 40]), several remaining ones are adaptation of

concepts in sociology and psychology (connection duality, connection strength update). Acknowledging this fact, a preliminary ablation analysis is performed on each of the latter mechanisms.

Every experiment performed for this ablation study uses the synthetic LFR network and assumes a polarized news distribution with $\Theta = 0.5$.

### Connection duality

First, we investigate whether the usage of $c_{uv}$ as the variable ratio in Eq. 1 is indeed necessary to induce polarization. We do this by modifying the equation to:

$$P_{uv} = (c_{uv} \times \alpha) + ((1 - \alpha) \times (1 - f(|i_i - q_v|/2))), \tag{8}$$

where $\alpha$ is a static variable ratio replacing $c_{uv}$. Three values of $\alpha$ are chosen to simulate, representing 3 different scenarios: $\alpha = 1$ (propagation only depends on connection strength), $\alpha = 0.5$, and $\alpha = 0$ (propagation only depends on opinion similarity). The $m$-value and $\overline{q}_{abs}$ produced by these simulations, plus another from a simulation using the original equation, is shown in Table 2.

As can be seen from the table, no value of static $\alpha$ is able to induce polarization on the model (characterized by both $m$-value in the range of $[2.4 - 4]$ and a high $\overline{q}_{abs}$). The usage of $c_{uv}$ on the other hand, leads to the fulfillment of both characteristics. Figure 4 shows the final distributions of $Q$ and $C$ taken from one instance of simulation for all three values of $\alpha$ and $c_{uv}$.

**Table 2 Statistics of the final opinion distribution for simulations on varying static propagation success ratio $\alpha$ and the original ratio $c_{uv}$**

| Ratio ($\alpha$) | $m$-value | $\overline{q}_{abs}$ |
|---|---|---|
| 1.0 | 5.814 | 0.232 |
| 0.5 | 2.174 | 0.282 |
| 0 | 2.255 | 0.685 |
| *$c_{uv}$* | *3.021* | *0.712* |

Italic values indicate a polarized distribution



**Fig. 4** Final distributions of *Q* (top) and *C* (bottom) for varying propagation success ratio. From left to right: { 1.0, 0.5, 0, $c_{uv}$ }

### Connection strength update

We next evaluate the update to connection strength, which is underpinned by the assumption that propagation success affects the strength of connection ($C$). It is possible to assess the importance of this mechanism by performing simulations where the mechanism is omitted and analyzing the result. The question of interest is whether the polarization of opinions happens even when connection strength is not updated. If so, we can conclude that the update is not a necessary mechanism to induce polarization.

Similar to the previous part of the ablation study, a simulation set is performed on a modified version of the model where the updates using Eqs. 2 and 4 are skipped. The average $m$-value for 50 iterations of this simulation is 2.027 and the average $\overline{q}_{abs}$ is 0.615; we see that polarization is not induced without connection strength update.

## Effect of model parameters

Next, we look into how the parameters of the model affect the polarization process happening inside the social network. We perform one set of simulation for each parameter; for each set, we vary the values only for the parameter we are interested in and keep the values of the other constant. This procedure of isolating one parameter at a time prevents unwanted noises from the interaction of multiple parameters, making it easier to interpret our results. This also reduces the number of simulations that needs to be done considering the large amount of combination possible for all parameter values.

As with the ablation study, every experiment performed here uses the synthetic LFR network.

### News polarization $\Theta$

First, we investigate how polarization in the propagating news items translates to the polarization of the population's opinion. For this we set $B = 10$, $\mu = 0.15$, and $\eta = 0.01$ .13 simulations were performed for $\Theta = 0.3$ to 0.9, with the increment of 0.05. As every simulated value of $\Theta$ is less than 1, news distributions for this set of simulations are always polarized. However, this range of $\Theta$ is sufficient to explore the behavior of the model considering its tendency to converge. The relationship between $\Theta$ and both measurements of distribution polarization is plotted in Fig. 5.

From the top curve for $m$-value, it can be seen that for $\Theta < 0.5$ opinions are separated to a bimodal final distribution ($m$-value $\approx 4$). Furthermore, this separation is accompanied by a $\overline{q}_{abs}$, indicating that the majority of opinions lies on the extreme ends of the opinion spectrum. Contrarily for $\Theta > 0.5$, $m$-value falls and stabilizes at around 2, signaling opinion convergence to a unimodal final distribution. As can be seen from the accompanying $\overline{q}_{abs}$, this convergence happens towards the middle of the opinion spectrum ($-0.5$ to $0.5$).

### Update rate $\eta$

Next we investigate the effect of $\eta$. We increment $\eta$ by half an order of magnitude in the simulation set; 5 simulations were performed with $\eta = \{0.001, 0.005, 0.01, 0.05, 0.1\}$.

$\Theta = 0.5$ is used for all of these simulations. $m$-value and $\overline{q}_{abs}$ of the final distribution of $Q$ is calculated and plotted in Fig. 6.

It is immediately apparent that only $\eta$ values of 0.05 and 0.01 produce a polarized final opinion distribution ($m$-value $\approx 4$ and large $\overline{q}_{abs}$).

Sample final opinion histograms taken from one instance of simulation for different $\eta$ are shown in Fig. 7. It can be seen that the large $m$-values for low $\eta$ are caused by the existence of multiple peaks. Despite that, this low value of $\eta$ seems to still induce a tendency for opinion to polarize. On the contrary, larger values of $\eta$ tend to induce convergence.

### Similarity function

We explore two alternative similarity functions $f(x)$: a generalized logistic function and a step function. A generalized logistic function is given by the following equation:

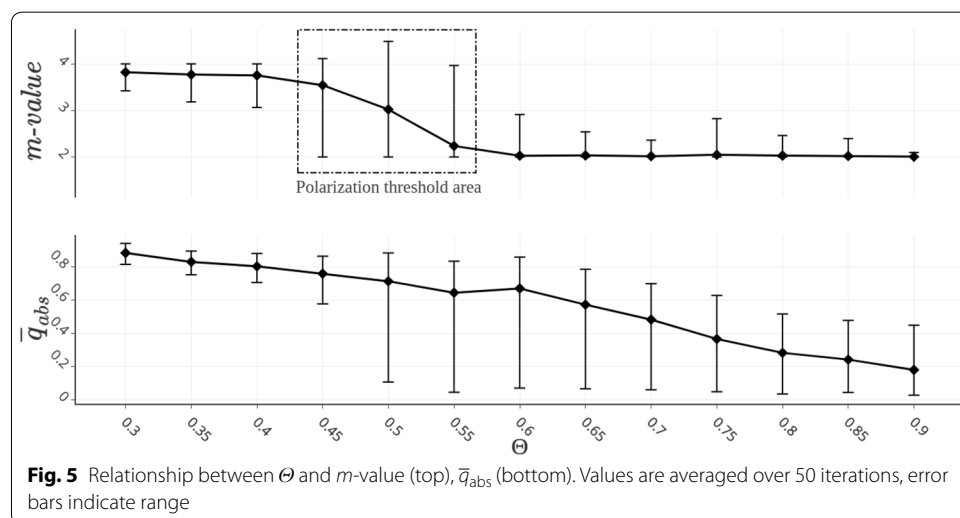$$f(x) = \frac{1}{1 + e^{-B(x-\mu)}} \tag{9}$$

with two parameters, $B$ and $\mu$. Meanwhile, a step function is given by the following equation:

$$f(x) = \begin{cases} 1 & \text{if } x \geq \mu \\ 0 & \text{if } x < \mu \end{cases} \tag{10}$$

with only $\mu$ as the parameter.

Changing the value of $\mu$ translates the function along the $x$-axis, changing the value of $x$ where these functions start outputting 1. The second parameter $B$ for the logistic function controls the uncertainty of the logistic function by changing the steepness of the logistic slope.

Simulations for the two functions were performed for $B = \{10, 20, 50\}$ (for the logistic function) and $\mu = \{0.1, 0.2, 0.3, 0.4, 0.5\}$. For all of the simulations, $\Theta$ was set to the value of 0.5. The results from the two simulation sets can be seen in Fig. 8.



**Fig. 5** Relationship between $\Theta$ and $m$-value (top), $\overline{q}_{abs}$ (bottom). Values are averaged over 50 iterations, error bars indicate range

Results for both similarity functions indicate that opinion polarization ($m$-value $\approx 4$ and large $\overline{q}_{abs}$) is only induced on $\mu$ lower than 0.3. Any increase in $\mu$ leads to convergence. Figure 9 shows several examples of the final opinion distribution from the simulations, which are very reminiscent of the results from BCM [48].
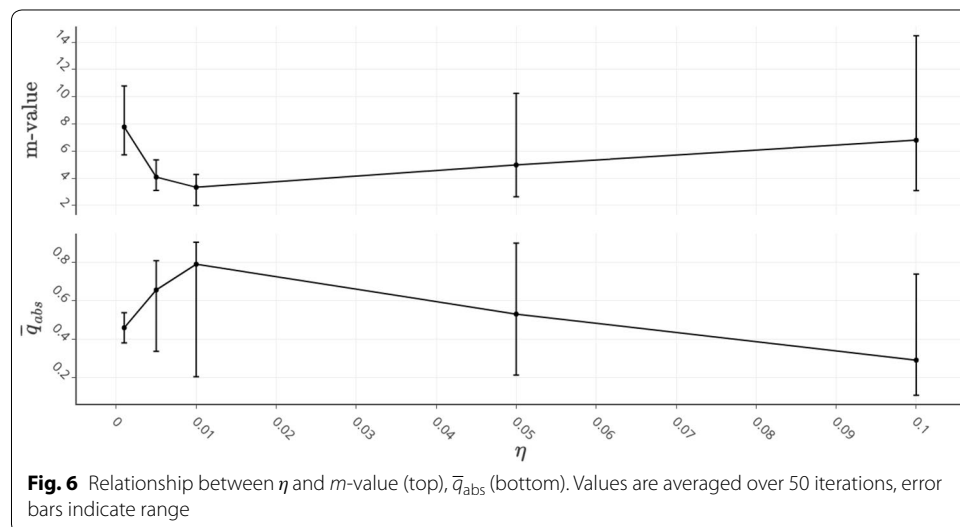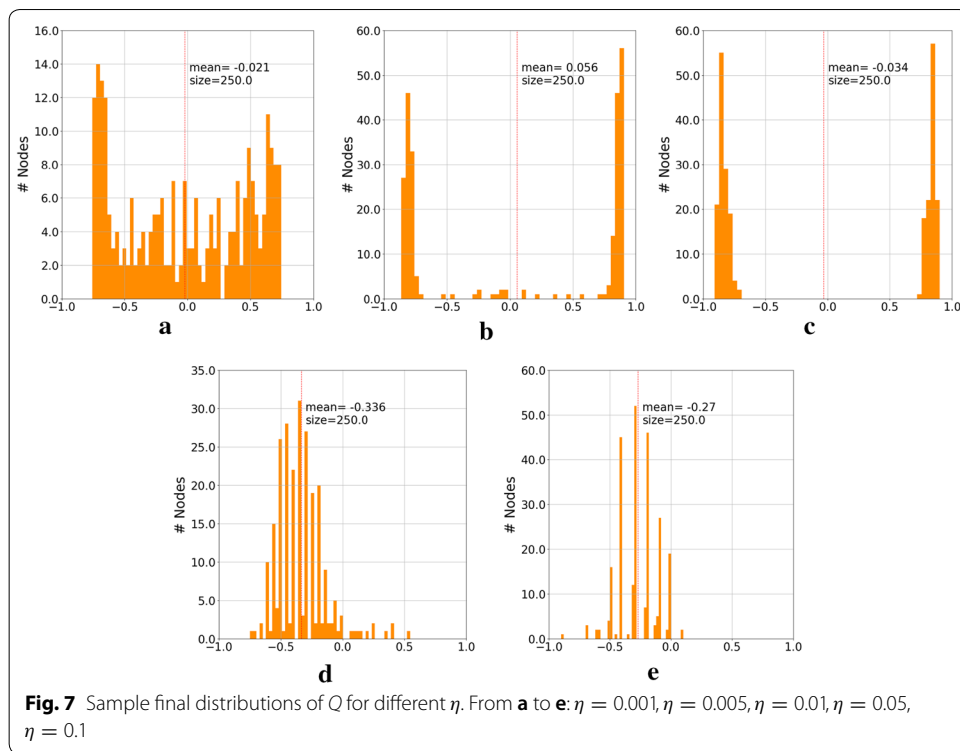
### Simulations on Twitter networks

One way to validate our model is to make sure that the observed behavior persists when the simulation is done on real networks. To this end, a set of simulations is performed over 18 Twitter follow networks. For all of these simulations, $\Theta$ is set to 0.50, $\eta - 0.01$, and the logistic similarity function is used with $\mu = 0.2$, $B = 10$. Table 3 shows the final $m$-value and $\overline{q}_{abs}$ for all of the simulations along with several measures for the networks.

As can be seen in the table, there exist Twitter networks which polarizes on average and networks that does not. The numbers in the table suggest that the combination of a high clustering coefficient and a low average shortest path length produce more susceptibility to polarization. Nevertheless, a more extensive investigation is needed to draw a more definite conclusion; note that the result presented in Table 3 is an average of only 20 iterations.

All defining characteristics of a polarized network found in the simulations over synthetic networks are also produced in Twitter networks. Figure 10 shows the final distribution of opinions $Q$ and connection strength $C$ for two instances of simulations on the Twitter follow networks; one converging and one polarized. The bimodality of the final opinion distribution and the clear separation for the weak edges can be observed for the polarizing simulation. Conversely, the converging simulation exhibits neither of those characteristics.

Upon plotting the $\overline{\text{hom}}_w$ growth of the cascades for both instances in Fig. 11, it is apparent that the $\overline{\text{hom}}_w$ growth for the converging instance's cascades are on a downward trend with low value. The cascades for the polarizing instance exhibit the exact opposite. Finally, Fig. 12 visualizes the strong graph of both of these instances after the final cascade.
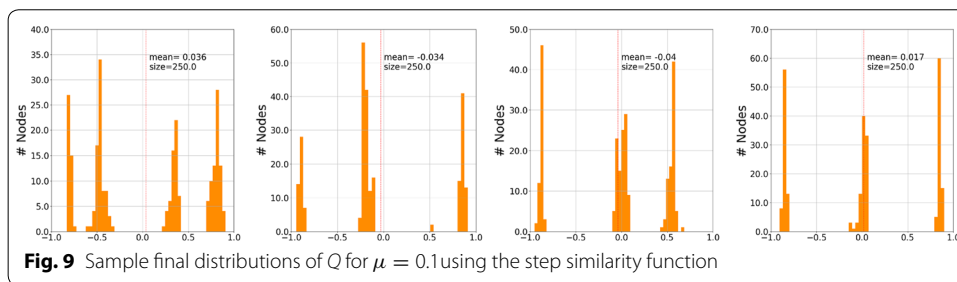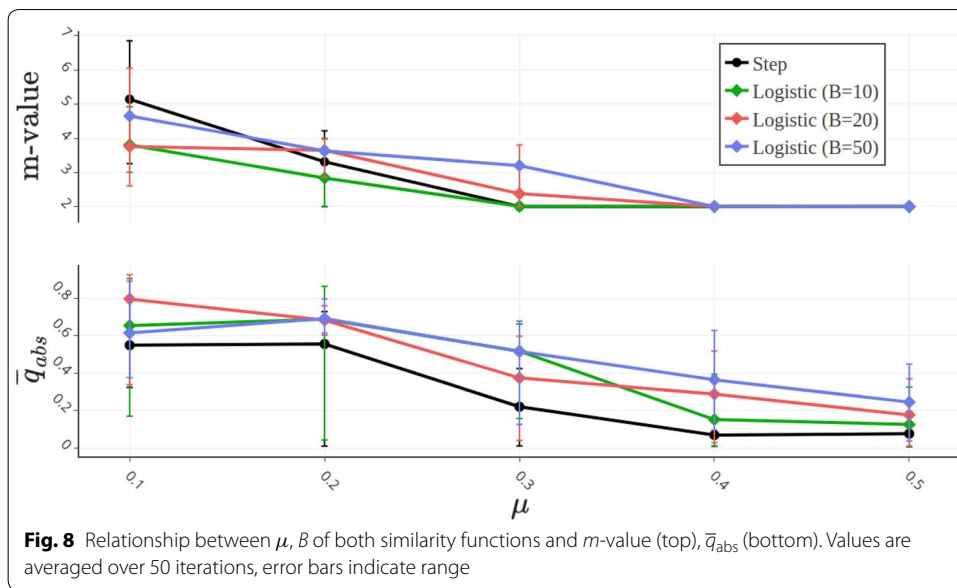


**Fig. 6** Relationship between $\eta$ and $m$-value (top), $\overline{q}_{abs}$ (bottom). Values are averaged over 50 iterations, error bars indicate range

**Fig. 7** Sample final distributions of $Q$ for different $\eta$. From **a** to **e**: $\eta = 0.001, \eta = 0.005, \eta = 0.01, \eta = 0.05, \eta = 0.1$

## Discussion and analysis

### Model design and validation

As shown by the ablation study, both of our proposed mechanisms (connection strength duality and update) are indeed necessary in order to induce polarization of opinion. To understand why this is the case, it is helpful to see how each of the chosen values of static $\alpha$ failed in inducing polarization. Interestingly, failures happened in different ways for each value of $\alpha$, as suggested by the different combination of $m$-values and $\overline{q}_{\mathrm{abs}}$ in Table 2. This difference can be explained by closely inspecting the distribution of connection strength in each case of $\alpha$ in Fig. 4.

First, when propagation success depended entirely on connection strength $c_{uv}$ ($\alpha = 1.0$), the behavior of the model became predictable considering how $c_{uv}$ is updated. Edges with weak initial strengths (i.e., $c_{uv} < 0.5$) most likely failed in propagating news, decreasing their connection strengths even further until eventually they hit 0. Conversely, initially strong edges eventually hit $c_{uv} = 1$. Propagation became essentially arbitrary, subject to only the initial values of connection strength. Since the initial value of $C$ was sampled on the range [0, 0.5] for all simulations, every edge eventually became very weak. As can be seen in the leftmost pair of histograms of Fig. 4, this resulted in a very slow convergence of opinion.

For the opposite scenario ($\alpha = 0.5$), it can be seen that opinions convincingly converged (Fig. 4, second from right). This behavior exhibits the model's innate converging tendency, which can be intuitively understood. Consider any point in the opinion spectrum $[-1, 1]$ which is not the center, 0. At any point, the range of opinion that will pull a user's opinion to the middle will always be larger than the ones attracting

**Fig. 8** Relationship between $\mu$, $B$ of both similarity functions and $m$-value (top), $\overline{q}_{abs}$ (bottom). Values are averaged over 50 iterations, error bars indicate range



**Fig. 9** Sample final distributions of $Q$ for $\mu = 0.1$ using the step similarity function

one's opinion to the extreme. Given that the distribution of $I$ is symmetrical, there will always be more news that attracts users opinion to converge.
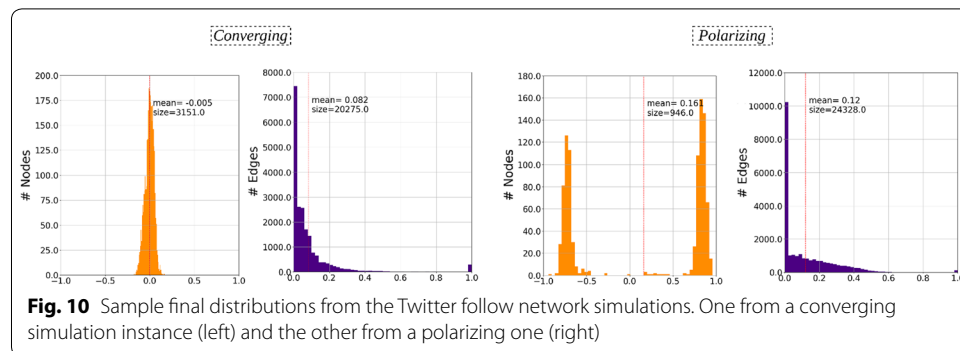
Finally, consider the middling scenario where both factors contributed equally to propagation success ($\alpha = 0.5$). In this case, to successfully propagate news a user needs both strong connection and opinion similarity to his/her target. This led to a general decrease in the number of successful propagation. As a result, connection strengths $c_{uv}$ of the entire graph once again plummeted to near 0 as a consequence of the connection strength update (Fig. 4; second from left, bottom). When this happened, the behavior of the model was again reduced to the second case explained above.

The usage of dynamic propagation success ratio is necessary to resolve the issues explained above. As exhibited in the ablation study, it led to a clear separation of edges in terms of their connection strength and a polarized opinion (Fig. 4, rightmost). This was caused exactly by the duality of behavior in determining propagation success: weak edges that connected nodes with a similar opinion were able to consolidate when they became strong enough, which then proceeded to amplify the differences in opinion to the point of extreme polarization. This duality of behavior can then be deemed as necessary to produce polarization.
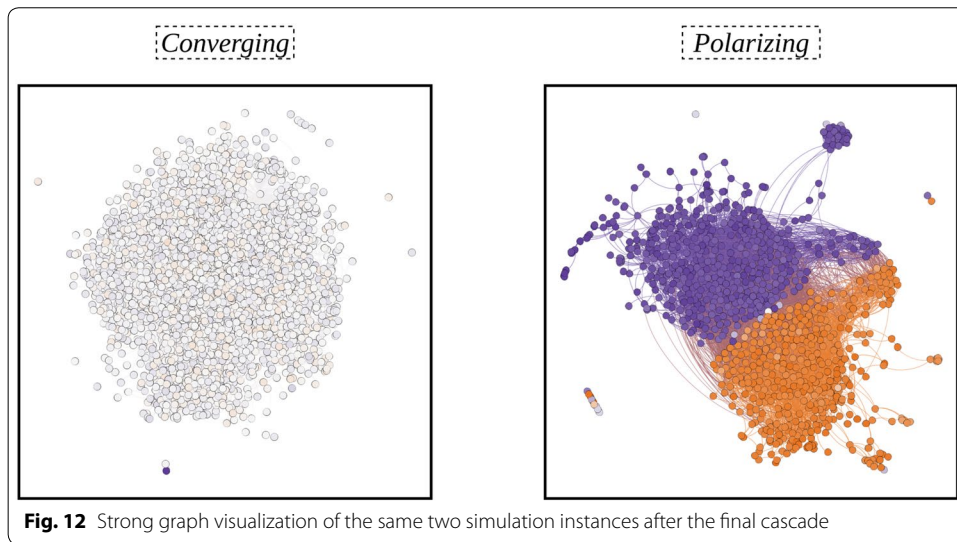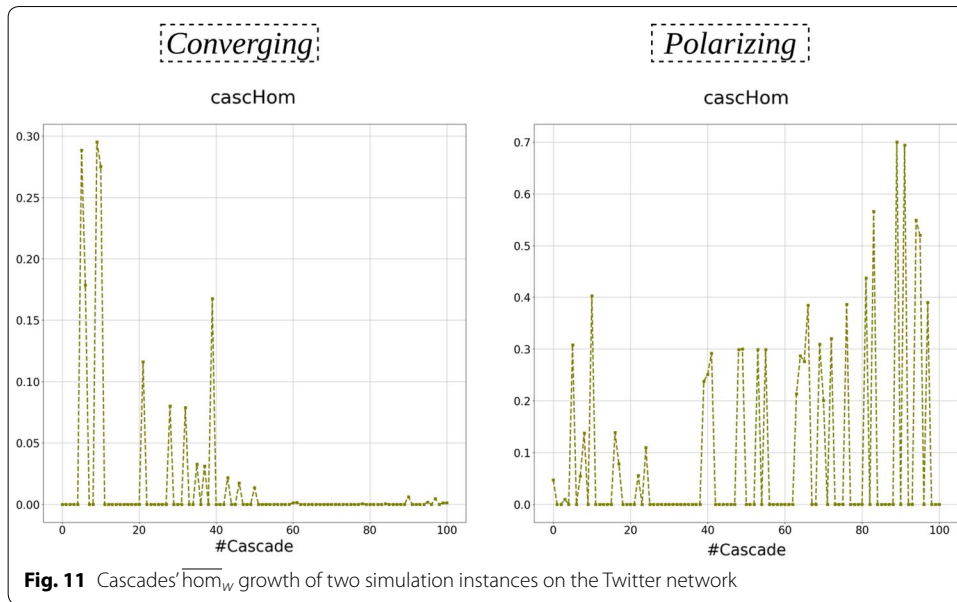
**Table 3 Statistics and final *m*-value of the Twitter follow dataset**

| Hashtag | Nodes | Edges | Density | Average shortest path | Average clustering coefficient | Assortativity | Transitivity | *m*-value | $\overline{q}_{abs}$ |
|---|---|---|---|---|---|---|---|---|---|
| nepal | 4242 | 42,833 | 0.0048 | 4.038 | 0.265 | − 0.191 | 0.125 | 2.019 | 0.411 |
| german-wings | 2111 | 7329 | 0.0033 | 4.267 | 0.133 | − 0.127 | 0.111 | 2.031 | 0.278 |
| indias-daugh-ter | 1542 | 9480 | 0.0079 | 3.614 | 0.184 | − 0.078 | 0.174 | 2.042 | 0.446 |
| mothers-day | 2245 | 14,160 | 0.0057 | 4.364 | 0.314 | − 0.016 | 0.364 | 2.057 | 0.445 |
| ff | 3899 | 63,672 | 0.0084 | 3.696 | 0.328 | 0.072 | 0.277 | 2.076 | 0.432 |
| onedirec-tion | 3151 | 20,275 | 0.0041 | 3.226 | 0.103 | − 0.087 | 0.040 | 2.096 | 0.516 |
| jurrasic-world | 4395 | 31,802 | 0.0033 | 3.998 | 0.219 | − 0.143 | 0.142 | 2.215 | 0.408 |
| beefban | 799 | 6026 | 0.0189 | 3.338 | 0.235 | − 0.063 | 0.215 | 2.244 | 0.422 |
| ukraine | 3382 | 84,035 | 0.0146 | 3.104 | 0.302 | − 0.89 | 0.256 | 2.300 | 0.511 |
| baltimore | 1441 | 28,291 | 0.0272 | 2.861 | 0.227 | − 0.241 | 0.192 | 2.344 | 0.523 |
| indiana | 946 | 24,328 | 0.544 | 2.530 | 0.351 | − 0.202 | 0.286 | 2.399 | 0.473 |
| *ultralive* | *2113* | *16,070* | *0.0072* | *2.464* | *0.359* | *− 0.250* | *0.031* | *2.504* | *0.699* |
| *leadersde-bate* | *9566* | *344,088* | *0.0075* | *2.547* | *0.310* | *− 0.231* | *0.146* | *2.643* | *0.521* |
| *nemtsov* | *2156* | *46,529* | *0.0200* | *2.692* | *0.322* | *− 0.126* | *0.236* | *2.660* | *0.693* |
| *russia-march* | *1189* | *16,471* | *0.0233* | *2.798* | *0.285* | *− 0.222* | *0.215* | *2.723* | *0.613* |
| *sxsw* | *4558* | *91,356* | *0.0088* | *2.638* | *0.257* | *− 0.135* | *0.086* | *2.824* | *0.619* |
| *netanyahu* | *4292* | *297,136* | *0.0322* | *2.341* | *0.329* | *− 0.274* | *0.183* | *2.896* | *0.563* |
| *gunsense* | *1821* | *103,840* | *0.0626* | *2.342* | *0.434* | *− 0.159* | *0.312* | *3.288* | *0.720* |

Italic fonts indicate polarization



**Fig. 10** Sample final distributions from the Twitter follow network simulations. One from a converging simulation instance (left) and the other from a polarizing one (right)

The same thing can be said for connection strength update. In the ablation study, removing updates to $c_{u,v}$ also resulted in no polarization. This can be easily understood by following the line of thoughts above; since $C$ was initialized in the range [0, 0.5] and remains static throughout the simulation, the model operated almost completely by selective exposure. As seen previously, the PIC model has an innate tendency to converge when it relies completely on selective exposure. It is then clear that the update to connection strength is also imperative in inducing polarization in opinion.

**Fig. 11** Cascades' $\overline{hom}_w$ growth of two simulation instances on the Twitter network



**Fig. 12** Strong graph visualization of the same two simulation instances after the final cascade

While we have established the importance the model's mechanisms, it is also important to confirm that the behavior of the PIC model can persists over different network structure. It is quite possible that the polarization showed in the ablation study will not be reproduced when the model is operating inside a larger and more complex network structure. However, the results on the Twitter networks simulation show that this is not the case.

The results for simulations done in Twitter networks show that the PIC model was still able to induce polarization. While this does not validate the model in the sense that it confirms that the dynamics of the real social system is 100% as predicted by the model, it does testify to the generalizability of the model by showing that its behavior are reproducible across different networks. Furthermore, all defining characteristics of a polarized

network found in the simulations over synthetic networks were also produced in Twitter networks. The strong graph for the polarizing instance exhibited a clear modular structure that corresponds to node opinion, signaling the existence of echo chambers. These signature characteristics were not observed when opinion converges.

In sum, all of these results prove that the PIC model is able to retain its behavior across networks of different sizes and structure.

### Parameter initialization

One vital aspect of the experiment design is the initialization of the model parameters, namely $Q$, and $C$. The question of finding proper values for these parameters is not something that can be answered easily. One obvious reason is the difficulty in quantifying these parameters. It goes without saying that the ideal values to initialize these parameters are ones that matches exactly what is observed empirically. However, assigning exact numerical values to user opinions and connection strengths given only traces of interaction/propagation is a difficult question in itself. Earlier works that attempted to do this eventually rely on the notion of interaction frequencies [4, 7, 73]: more interaction translates to stronger connection/tendency for a side in the opinion spectrum.

The second problem presents itself when we consider that we are looking for initial values of these parameters before any interaction occurs. Suppose we chose to initialize our opinions and connection strengths based on empirical findings that measure opinion/connection strengths based on frequency, such initialization would make little sense, because these values can only be calculated after interactions happened.

As explained in "Experimental setup" section, in the end we opt to simply do a random initialization for $Q$ and $C$ within a certain range. This make sense for us since we are looking to study how polarization was formed; there is no reason to initialize $Q$ and $C$ from any non-uniform distribution with any degree of polarization. We initialize $Q$ in the full opinion range $[-1, 1]$. However, initializing $C$ in the full range makes less sense because in combination with the initialization for $Q$, this leads to unlikely situations such as two users with opposite opinions but very strong connection. In turn, these situations produces an arbitrary behavior for the PIC model (since high connection strength translates to very high acceptance) which almost always lead to convergence. Given that it's impossible to study the behavior of the model this way, we settle with initializing $Q$ in a lower range of $[0 - 0.5]$, i.e., everyone in the social network starts off as weakly connected.

While the aforementioned scenario of initialization in the full connection strength range may be the case in the real world (e.g., family member with differing views), a low-ranged initial connections strength produces an organic connection formed solely through selective exposure (i.e., strangers with the same views bonding together). This kind of scenario is likely to happen in online social media where people from different places and background can meet. In fact, the evolution of connection strengths from a lower range of value through the mechanism described by the PIC model can be seen as a process of trust establishment in online social media as reviewed by [32]. The assignment of initial values of $C$ could be considered as the initial phase of trust building where users' trust are based on on rapid, cognitive cues or first impressions, as opposed to personal interactions [51]. Considering that high values of $C$ in the PIC model signifies

unconditional trust (as news acceptance and propagation is almost certain), the $C$ initialization on the low range of values is a likely scenario. Early interactions, governed by opinion similarity can also be seen as representing *calculus-based trust* based on calculation of cost and reward of interaction.

After a number of interactions, judgments of users become the function of the interaction itself. This represents the second stage of trust developments which is governed by established knowledge and understanding between users *knowledge- or experience-based trust* [46]. In the PIC model, we update $C$ based on interaction results; if we consider $C$ as trust, then the process described in PIC is exactly this second stage of trust development. Within the broader context of trust, this process can be considered as the *relational* facet of trust. This particular facet of trust acts as a basis for many trust model in computer science, as expounded in [62].

Design choices for the entire simulation process, from initialization to the growth of the parameters, were made with the considerations explained above. While online trust was not the primary focus of this work, the particular scenario we chose for our simulations can be sufficiently motivated seeing how the dynamics of connection strength in the PIC model can also be viewed as a trust development in online social networks. Readers interested in the topic of online trust can consult either [32] or [62].

### Effect of model parameters

Results from the set of simulations involving the model parameters indicate that each parameter has a significant amount of contribution in inducing polarization. Without a correct combination of parameters, opinion polarization will likely not form.

One of the more intuitive result came from the simulation set for the effect of news polarization degree $\Theta$. As seen in Table 5, both $m$-value and $\overline{q}_{abs}$ increased along with $\Theta$. This shows that news polarization translates to the polarization of opinion; this behavior is both intuitive and observable in the real world. The result also confirms the model's tendency to converge. This tendency can be assumed as a semblance of realism: the more extreme one's opinion is, the more argument is available to make it less extreme.

Note that the relation between news polarization $\Theta$ and the produced modality of the final distribution is nonlinear. On individual examination of each simulation instance, we found that opinion distribution either ended in two peaks or one peak but never in-between. This is also apparent from the wide $m$-value range on the critical area, where the final opinion distribution alternated between convergence and polarization. This suggests the existence of a certain threshold condition in our model that determines whether the network will converge or polarize. This threshold is at least determined by the news polarization $\Theta$ and for the parameter values used in this set of experiment, the threshold lies around $\Theta \approx 0.50$. Given the state transition-like nature exhibited by the model, the average $m$-value can then be interpreted as the tendency or probability for the distribution to polarize.

On the other hand, $\overline{q}_{abs}$ is linearly correlated with $\Theta$; the more polarized the propagating news is, the more extreme the average opinion of the population. This means that

while a certain polarizing threshold exists, increasing or reducing news polarization beyond the threshold will always have an impact on the overall distribution of opinions.

Results from the set of simulations involving the update rate $\eta$ can also be easily understood; small values of $\eta$ induced polarization and large values did not. The obvious takeaway is that opinions must update slowly for polarization to happen. However, the large updates of opinions seem to produce big and unusual $m$-values. To comprehend this, we examine individual histograms for each simulation shown in Fig. 7.

It turns out that large $m$-values were caused by opinions taking only a set of values, resulting in small gaps throughout the histogram. This is easily understandable. Suppose updates happen in increments of 0.1; if any opinion hits the maximum/minimum bound of the $[-1, 1]$ range, its value will always be a multiple of 0.1 from that point on. Given enough time, eventually, every opinion will only be a multiple of 0.1. Despite this, large values generally led to a converging opinion distribution shape. On the other hand, the reason why $\eta = 0.001$ seemed to not result in a polarized state (indicated by high $m$-value) is that the update to opinion was too slow. When we examine the histogram, it can be seen that the distribution was actually leaning towards a bimodal shape.

In the final set of simulation on the parameters of the similarity function, $\mu$ can be interpreted as the tolerance of individuals to opinion difference. Smaller $\mu$ means even for small differences, individuals tend to reject propagating news. Conversely, large $\mu$ means individuals are tolerant. In a way, the step function could be viewed as an extreme case of the logistic function. $B$ on the other hand can be interpreted as the how consistent users are with their decision to accept news items. With this in mind, we can try to understand the simulation results contained in Fig. 8.

We see that polarization of opinion happened when the tolerance of individual is low enough. For the simulated parameter values, the limit lies in $\mu \approx 0.3$, which translates to around 0.6 point differences on the $[-1, 1]$ opinion spectrum. This behavior, in fact, aligns with the findings of previous models, particularly BCM [48] which is exactly designed to model tolerance.

Increasing the steepness of the logistic function through the parameter $B$ seemed to increase the probability of polarization happening. While the general behavior of the two alternative functions is similar, a particular behavior was observed when the steepness of the function becomes maximal in the form of a step function. Consider a pair of connected nodes $u$ and $v$ in a network where the overall connection strength is low. For the step function, if the opinion difference between $u$ and $v$ is at least four times $\mu$, it will be almost impossible for news to propagate between the two. Neither of these nodes will accept and propagate news with an opinion the other can accept. The result is a more deterministic process (relative to when using the logistic function) which will lead to the existence of multiple isolated groups based on their opinions. This is what happened when we set $\mu = 0.1$, as indicated by the large $m$-value.

Overall, the results from the simulations involving the model parameters indicates that each parameters needs to be in a specific range of values so that polarization can happen. In the case that even one of these parameters is out of range, opinions will most likely converge. In our sets of simulations, requirements for these parameters was $\Theta < 0.5$, $\eta \approx 0.1$, and $\mu \approx 0.3$.
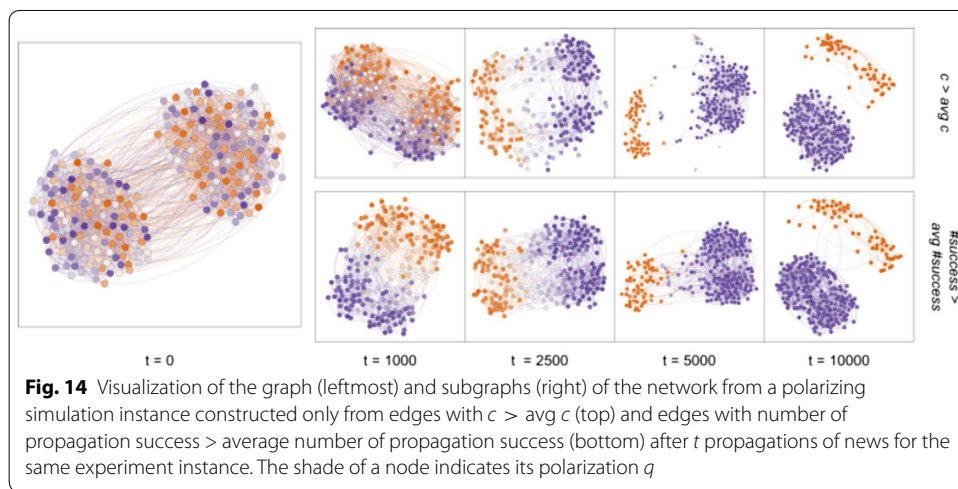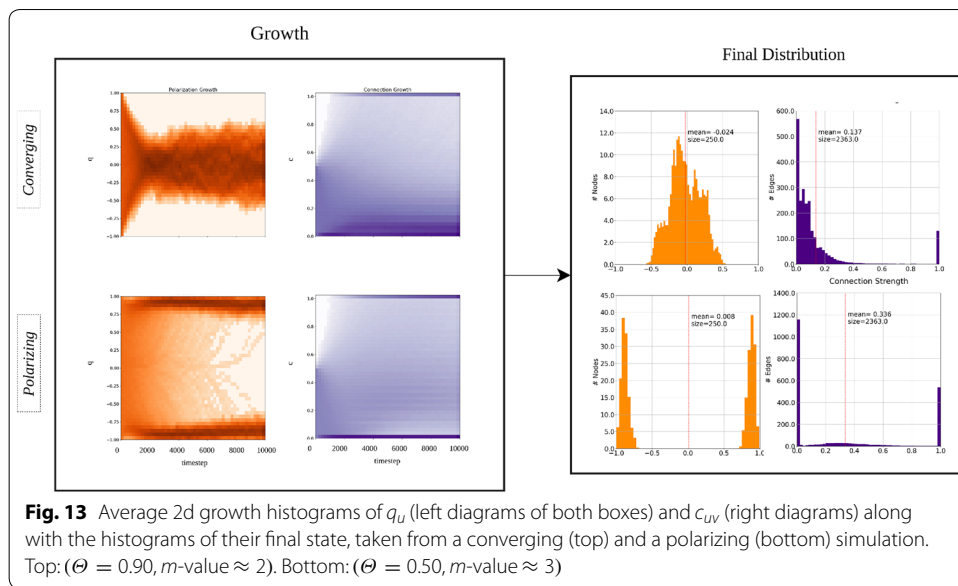
**Segregated propagation structure**

To study the propagation structure of the network, we turn to the distribution of the connection strength parameter, $c_{uv}$ and its relation to the opinion parameter, $q_u$. Figure 13 shows the growth and the final distribution of $Q$ and $C$ taken from two simulation instances from the simulation described in "Effect of model parameters" section. One of these simulations converged (top) and the other polarized (bottom). The final distribution of $c$ shown in the bottom part of the figure shows a clear separation between weak edges ($c_{uv} = 0$) and the rest. While this separation happened in both simulation instances, the shape of the separation is different. Strong edges are clearly separated in both case, but weak edges are not distinctly separated in the converging instance as indicated by the smooth decrease along the beginning of the $x$-axis. In addition, the number of edges with $c_{uv} = 0$ is less ($\sim$500) compared to the polarizing instance ($\sim$1000).

This means that when opinions converge, the propagation structure is not segregated. What is even more important than the simple existence of this edge separation is whether they function differently throughout the simulation. Even if the edges on the converging simulation can be considered to be separated, all of them is still identical in the sense that they connect nodes with similar opinions. For the polarized simulation, the distinction is functional: weak edges are seldom used in news propagation between nodes with opposing opinions and strong edges are responsible for the propagation of news in the network between nodes with similar opinions. The latter can be thought of as representing the true propagation structure of the network.

The subgraph of the network induced by the strong edges is visualized in Fig. 14. Initially, two communities with diverse opinion existed. After $t = 1000$, opinion started to polarize. While the initial community structure still persisted, each community was then divided with regard to opinion. At $t = 2500$, the initial structure faded and the opinion of the network was split into three: the extreme left, the moderates, and the extreme right. The propagation structure also clearly highlights this separation. It can be seen that nodes of moderate opinion acted as intermediaries between the two extremes. The behavior shown here is similar to the notion of polarization proposed in [35]. Subsequent propagation further exacerbated both structural segregation and opinion polarization until finally the two opinion groups were fully separated. Note how the similarity between the subgraphs based on $c_{uv}$ and number of success increased until becoming almost identical at the end.

The main takeaway from the figure is the fact that polarization of opinion (signified by the darkening hue of the nodes and the disappearance of the intermediary white nodes) happened at the same time with the segregation in propagation structure (signified by the increasingly modular layout of the network). As a result of these two processes, echo chambers are formed in the network. This can also be observed earlier in the 2d growth histogram of Fig. 13 which also suggests that the opinion polarization and propagation structure segregation happens concurrently. One does not cause the other, but both of them happens as a consequence of the behavior of the PIC model. The same conclusion is also drawn upon individual examinations of the final opinion and connection strength histograms of every simulation instances. The polarization of opinion and connection strength always happened hand-in-hand; in no simulation instance either of them happened exclusively.
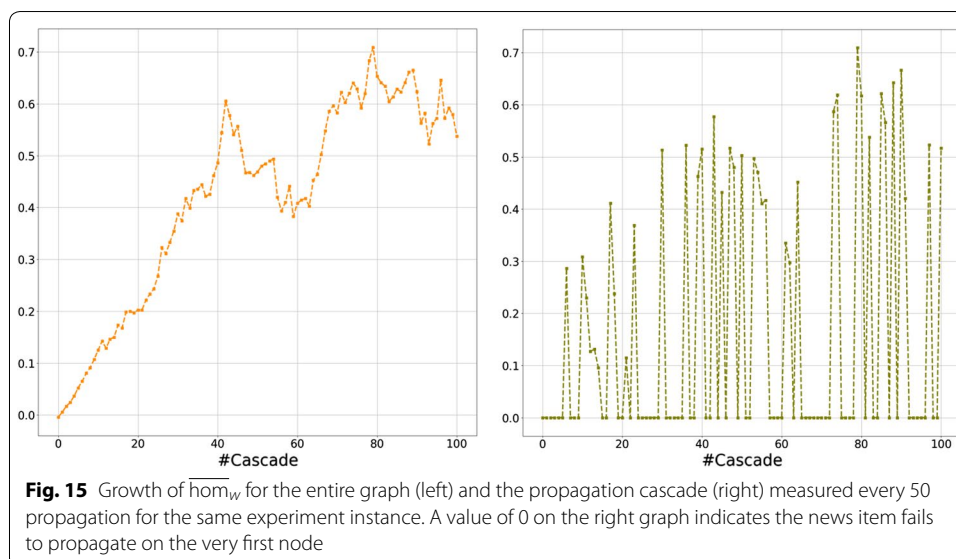
**Fig. 13** Average 2d growth histograms of $q_u$ (left diagrams of both boxes) and $c_{uv}$ (right diagrams) along with the histograms of their final state, taken from a converging (top) and a polarizing (bottom) simulation. Top: ($\Theta = 0.90$, $m$-value $\approx 2$). Bottom: ($\Theta = 0.50$, $m$-value $\approx 3$)



**Fig. 14** Visualization of the graph (leftmost) and subgraphs (right) of the network from a polarizing simulation instance constructed only from edges with $c >$ avg $c$ (top) and edges with number of propagation success > average number of propagation success (bottom) after $t$ propagations of news for the same experiment instance. The shade of a node indicates its polarization $q$

To further confirm this idea, we examine the growth of $\overline{\text{hom}}_w$ for the polarizing simulation instance. As shown in Fig. 15, $\overline{\text{hom}}_w$ increased overtime in both the entire network and individual news propagation cascades. This means as opinion became more polarized (Fig. 13, top), so did both the structure of the entire network and the individual cascades. In the beginning, news propagated through nodes with different opinions, reflecting an inter-belief discussion in social media. However, this kind of discussion diminished as the network became more polarized. We also found that in later stages of polarization, the opinion of the news circulating inside a community was similar to the opinion of the member of the community, which is exactly what the echo chamber effect is.

## Other resulting network structures

We established that the PIC model is capable of producing a polarized network structure (i.e., echo chambers) as observed in many empirical researches. Looking forward, one interesting question to be explored is what other network structure could it produce/replicate. We know that the PIC model itself is conceptually very particular in its design; each mechanism was chosen with the aim of producing polarization. However, there are specific parameter configurations that end up producing propagation structures that can not exactly be categorized as converging or polarizing. Given this ability, it would be insightful to check whether the PIC model can reproduce common archetypes of network structures in online social media.

While there are many past works that identified various network structures in online social medias, [65] in particular observed 6 typical Twitter network structures, one of which is exactly the segregated propagation structure produced by the PIC model. Here we briefly discuss the possibility of producing the other 5 with the model described in this work, and the modification it will take to do so should any of it is impossible for produce. This discussion can serve as pointers in future works to generalize the PIC model to encompass more network structures and phenomenon.

[65] identified 6 network archetypes, which are divided further into 3 pairs of related two. The first of these three pairs are *polarized crowd* and *tight crowd*. The former is characterized by the existence of two big communities with very few connections among them. This is exactly a segregated propagation structure explained in "Segregated propagation structure" section. The PIC model is able to reliably produce this kind of network. The required conditions as identified in our simulations sets are propagation of polarized news items, low tolerance, and slow updates to opinion. The findings of [65] regarding the *polarized crowd* network archetype agrees with the first requirement, as most web URLs which circulates in it was found to be highly political and partisan.



**Fig. 15** Growth of $\overline{\text{hom}}_w$ for the entire graph (left) and the propagation cascade (right) measured every 50 propagation for the same experiment instance. A value of 0 on the right graph indicates the news item fails to propagate on the very first node
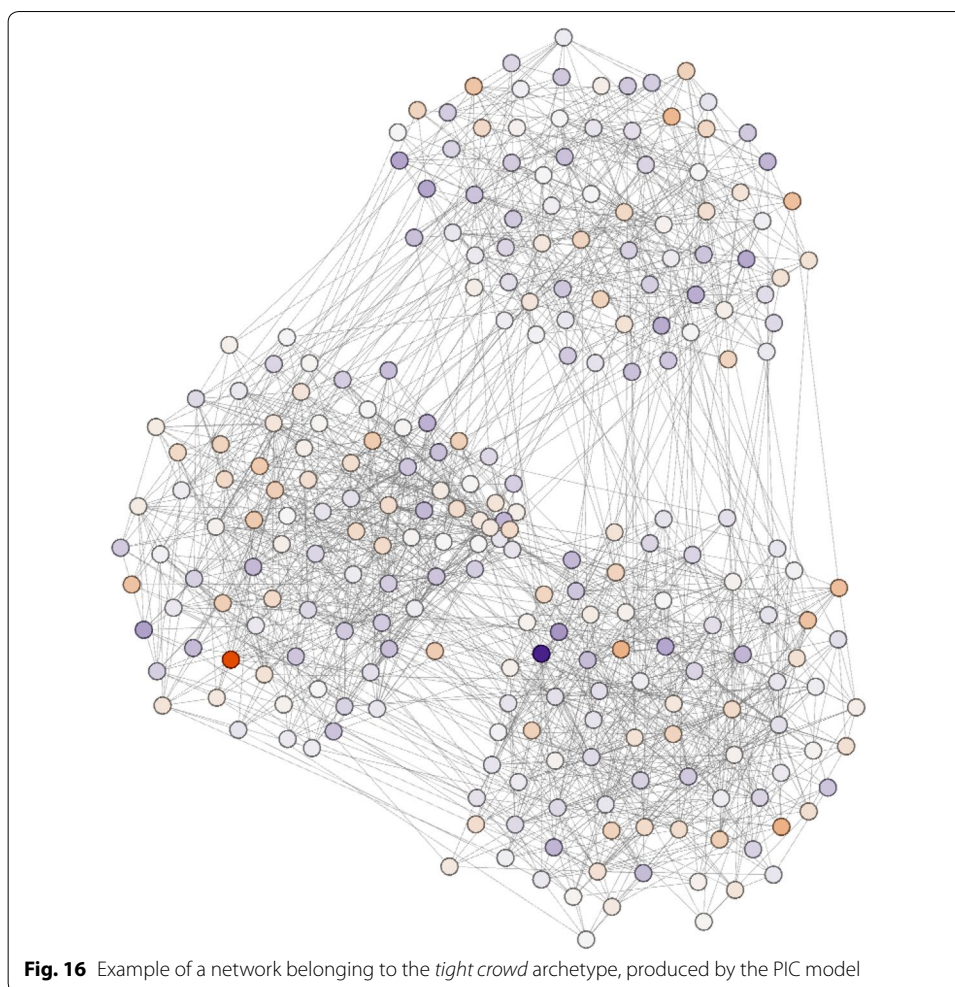
On the other hand, the *tight crowd* network archetype features multiple large communities which are highly connected. Interaction networks with more a large number of interacting communities were not produced in any simulation sets we performed. However, converging simulation instances matches other characteristics of the *tight crowd* network types: high participation among all members of the network and the absence of disconnected members. We believe that the absence of multiple communities in the network structure produced by the PIC model is simply a byproduct of simulating in LFR networks with two communities, i.e., the PIC behave is indifferent to initial network structure. In fact, looking at Fig. 14, we see that polarization happened almost independently w.r.t. community structure. Each of the two initial communities splits into two. The same goes happens when opinions converge. Network with two initial communities will end up having the same two communities, only with converging opinions.

Figure 16 shows a sample of the resulting strong graph, representing the propagation structure for a simulation done in an LFR network with 3 communities. As can be seen, in the end the network still comprised exactly the 3 initial communities. We believe this is also the case with the *tight crowd* networks found in Twitter; their community structure is carried over from the actual communities of the Twitter users (e.g., group of friends, acquaintances, or families). Other than the community structure, the non-divided trait of this *tight crowd* is identical to converging simulation instances of the PIC model. The required condition for this type of network is non-polarized news (high $\Theta$).

The second network archetype pair found by [65] are called *brand clusters* and *community clusters*. Both are characterized with a large proportion of lone nodes, representing those who are not engaged in any interactions with other members of the node. Another trait of these two network archetypes are multiple small/medium-sized communities that are isolated from one another. Within the context of our opinion model, multiple communities inside the interaction network should be signified by multiple peaks in the opinion distribution. Each peak will represent one community with similar opinion values. This matches well with one explanation from [65] about these clusters of communities; that they may be formed around different influencers, news sources, and audience. In our case, one community is centered around audiences with similar opinions.

One example of propagation structure matching the characteristics of *brand clusters* and *community clusters* produced by the PIC model can be seen in Fig. 17. This propagation structure is produced using the logistic similarity function, with $\mu = 0.025$ (low tolerance) and $B = 1000$ (high certainty in news acceptance). In this simulation instance, $\Theta$ was set to 0.9, and $\eta$ was set to 0.01 for connection strength update and 0.005 for opinion update. We found that in order for multiple peaks of opinion to form, the most important condition to fulfill is low $\mu$. Within the context of polarization study, this means that people are intolerant to different opinion. In a broader sense, the value of $\mu$ could also just be viewed as the amount of interest in different discussion groups. Low $\mu$ means members of the network in general have little interest in interacting with different sets of people, preferring to stick within a small community.
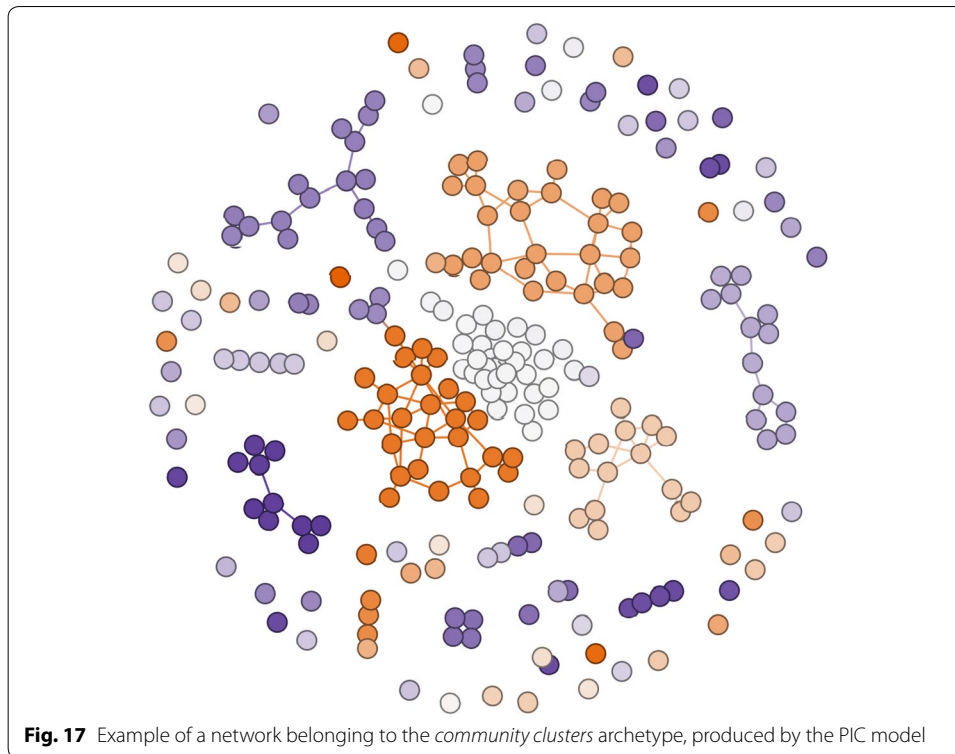
Having said that, we are unable to find parameters that exactly control the size of these small communities and connection between them so that we are able to confidently

**Fig. 16** Example of a network belonging to the *tight crowd* archetype, produced by the PIC model

differentiate between producing interaction networks of the *brand clusters* type and the *community clusters* type. Subsequent simulations with the previously mentioned parameters produced different sizes of communities and isolated nodes. Thus, the distinction between the two is determined by a factor not captured by the PIC model.

Among the 3 pairs of network archetypes explained in [65], we found that the only pair we are unable to reproduce is the last. The final two network archetypes, *broadcast network* and *support network*, are characterized with the existence of very large hub nodes which are connected to the rest of the network. The former network type happens when these hubs broadcast news to the network, and the latter happens if the opposite happens. In addition, non-hub nodes are also not connected to each other.

There are several obvious reasons why our PIC model is unable to reproduce this kind of network. First, the relation between hub nodes and common nodes in these networks are asymmetric. As of current, the PIC model operates only on undirected networks, which means every connection is symmetric. This also makes it almost impossible to naturally develop any node in the network to become a hub node. In Twitter, many accounts are natural hub (e.g., celebrity, news outlet). This is not the case in simulations with the PIC model. To reproduce these two types of networks, the model needs to be

**Fig. 17** Example of a network belonging to the *community clusters* archetype, produced by the PIC model

modified to work on directed network. In addition, the underlying network also needs to have designated hub. In short, formation of the final two network types is most likely unrelated to the PIC model.

## Conclusions

### Concluding remarks

The phenomenon of polarization and echo chambers has been empirically observed in multiple corners of the online community. High degree of polarization has been connected to unfavorable consequences such as extremism. Thus, the investigation of its behavior and evolution is consequential especially in the increasing prevalence of social media.

The PIC model we proposed is rooted in the idea that news affects the opinion of the masses. It incorporated a dynamic propagation probability rule based on selective exposure and connection strength. The simulations we performed exhibited how news propagates inside a social network and affect opinion of the nodes.This process will in turn govern future propagation pattern. The broadcast nature of information propagation and opinion updates is a natural translation of news sharing in social media, making the PIC model a more realistic depiction of opinion polarization than most existing models. In addition, the temporally dependent and consecutive nature of news propagation allowed us to observe the structure and growth of news propagation and opinion change.

We found that opinion polarizes under a reasonable degree of news polarization ($\Theta \approx 0.50$) with a state transition-like behavior. In relation to this, one important point

of discussion then is whether this degree of news polarization observed in the real world or not. While it is impossible to answer this question directly since news polarization $I$ is immeasurable in real life, the common public perception of news media is that they are partisan [31]. In addition, empirical findings indicate that major online news outlets cater to and are mainly consumed by individuals from identical political leaning [53]. These suggest that the notion of polarized online news is not unfounded in real social media, especially with the increasing popularity of partisan news [17, 40].

The state transition-like behavior of polarization may seem to suggest that that the any effort to reduce partisan news will have little impact in eliminating polarization up until a certain threshold point. However, this is not true when we consider that news polarization still retains a linear relationship with the average extremity of people's opinion. What this means for policy-makers or stakeholders in the social media industry is that disregarding whether it would eradicate polarization or not (by crossing the polarization threshold), efforts in reducing news polarization will always be a worthwhile; even if people are still divided, their opinions will not be as extreme.

Other than the polarization of the cascading news itself, we found that important contributors to polarization are slow opinion update and low tolerance for opinion difference. All of these findings can be understood intuitively and matches what has been observed in previous models without news propagation [48]. The polarization of opinion happens concurrently with the segregation of the news propagation structure, reflected by strong edges of the network, resulting in the formation of echo chambers. This matches what has been empirically observed in [7, 9, 14, 18, 39, 73] and connects these aforementioned researches to the modeling of opinion dynamics. Structure segregation and opinion polarization share no apparent causal relationship.

We note that the design choices of our model are supported by the results of our ablation study. In addition, the model is also able to reproduce its behavior on large, complex network. This means, similar to the IC model [42] it is based on, our model possesses a certain degree of validity for it to be used to explore problems related to news propagation and opinion. Problems such as finding parts of the social network which contribute the most to the flow of information and opinion change can be an extremely valuable asset for companies who are interested in generating favor for a new product or a politician on a campaign, for example.

In summary, the contributions of our work is as follows: it proposes a novel opinion model which describes how the propagation of news drive opinion dynamics; it studies the condition needed for polarization to happen inside a network of interconnected individuals; and it uncovers the relationship between the polarization of opinion and the segregation of propagation structure, which was never explained by previous opinion models.

### Limitations and future works

The work described in this article is limited in more ways than one. First of all, the model is not truly validated in an empirical manner in the sense that there are no real data pulled from social networks that can exactly confirm that it was formed from the mechanisms described by the model. This kind of empirical validation is a non-trivial task

plagued with many difficult questions such as how to ethically collect such empirical data or how to truly quantify the opinion of a social network user at any given time. Considering the absence of such validation, it should be understood that causation between the mechanism explained by the PIC model and online polarization is not explicitly guaranteed. Instead, this work can be viewed as one of many mechanistic explanations [36] of online polarization. Nevertheless, we believe that our proposed mechanism will be invaluable in guiding the effort of identifying the actual mechanism behind echo chambers.

The second limitation is that behavior comparison with other recent models of polarization has not yet been conducted. As such, we urge readers to refrain from referring to the PIC model as *state-of-the-art.* This would imply not only the existence of an absolute measure for goodness in predicting polarization, but also that other models are obsolete. In the first place, the aim of this study was never to offer one true model that exhaustively and most accurately illustrate echo chambers, but to give an abstraction consisting of mechanisms we believe are of importance. Nevertheless, we reaffirm that the mechanism proposed through the PIC model is one that has not yet been explored thoroughly by other past models.

Finally, the effect of network topology on the behavior of the model has yet to be investigated. In the results for the Twitter networks simulation we found that different networks display different degree of polarization tendency. We have yet to found a network measure that could match exactly this tendency.

The obvious future direction to this line of research is to overcome these limitations. Empirical validation or a systematic comparison with other models of polarization to evaluate is one important direction to take if one is aiming to identify the actual mechanism of online polarization. It would also be particularly useful to find out exactly how network topologies affect polarization, as implied by the results of the Twitter networks simulation. This can be done by further scrutinizing the model through more extensive sets of simulation or even a rigorous mathematical analysis similar to what have been done to other families of opinion models. After all, the illustrative power of the model comes with additional parameters and complexity allowing rooms for more explorations such as investigating the effect of other news polarization distribution shapes.

Another interesting aspect to study in particular is the polarization threshold and the state transition-like behavior of the model. As is the case for the epidemic threshold in epidemiology study, finding the polarization threshold would immensely assist policy-makers to prevent polarization in the most cost-efficient way possible. Finally, the model can be extended or modified to fit different scenarios such as the inclusion of biased assimilation, negative influence, or a stubborn agent. A more algorithmic take on the research could be done in the future, similar to [26, 27] in which the model is used to formulate a polarization minimization problem.

**References**
1. Acemoglu D, Dahleh MA, Lobel I, Ozdaglar A. Bayesian learning in social networks. Rev Econ Stud. 2011;78(4):1201–36.
2. Anderson M, Toor S, Rainie L, Smith A. Activism in the social media age. Pew internet & American Life project; 2018. http://www.pewinternet.org/2018/07/11/activism-in-the-social-media-age/. Accessed 10 Feb 2019.
3. Aslay C, Matakos A, Galbrun E, Gionis A. Maximizing the diversity of exposure in a social network. arXiv preprint; 2018. arxiv:1809.04393. Accessed 10 Feb 2019.
4. Bakshy E, Messing S, Adamic LA. Exposure to ideologically diverse news and opinion on Facebook. Science. 2015;348(6239):1130–2.
5. Banerjee AV. A simple model of herd behaviour. Quart J Econ. 1992;107(3):797–817.
6. Banisch S, Olbrich E. Opinion polarization by learning from social feedback. J Math Sociol. 2018;43(2):76–103. https://doi.org/10.1080/0022250X.2018.1517761.
7. Bessi A, Zollo F, Del Vicario M, Puliga M, Scala A, Caldarelli G, Uzzi B, Quattrociochi W. Users polarization on Facebook and Youtube. PLoS ONE. 2016;. https://doi.org/10.1371/journal.pone.0159641.
8. Bikhchandani S, Hirshleifer D, Welch I. A theory of fads, fashion, custom, and cultural change as informational cascades. J Polit Econ. 1992;100(5):992–1026.
9. Boutet A, Kim H, Yoneki E. Whats in Twitter, i know what parties are popular and who you are supporting now!. Soc Netw Anal Min. 2013;3(4):1379–91.
10. Bozdag E. Bias in algorithmic filtering and personalization. Ethics Inf Technol. 2013;15(3):209–27.
11. Brehm JW. Postdecision changes in the desirability of alternatives. J Abnorm Soc Psychol. 1956;52(3):384.
12. Campbell JD, Fairey PJ. Informational and normative routes to conformity: the effect of faction size as a function of norm extremity and attention to the stimulus. J Pers Soc Psychol. 1989;57(3):457.
13. Clifford P, Sudbury A. A model for spatial conflict. Biometrika. 1973;60(3):581–8. https://doi.org/10.2307/2335008.
14. Conover M, Ratkiewicz J, Francisco MR, Gonçalves B, Menczer F, Flammini A. Political polarization on twitter. In: Proceedings of the 5th international AAAI conference on weblogs and social media, Barcelona; 2011.
15. Dandekar P, Goel A, Lee DT. Biased assimilation, homophily, and the dynamics of polarization. Proc Natl Acad Sci. 2013;110(15):5791–6. https://doi.org/10.1073/pnas.1217220110.
16. DeGroot MH. Reaching a consensus. J Am Stat Assoc. 1974;69(345):118–21.
17. Feldman L. The opinion factor: the effects of opinionated news on information processing and attitude change. Polit Commun. 2011;28(2):163–81. https://doi.org/10.1080/10584609.2011.565014.
18. Feller A, Kuhnert M, Sprenger TO, Welpe IM. Divided they tweet: the network structure of political microbloggers and discussion topics. In: Proceedings of the 5th international AAAI conference on weblogs and social media, Barcelona; 2011.
19. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Commun ACM. 2016;59(7):96–104.
20. Festinger L. A theory of social comparison processes. Hum Relat. 1954;7(2):117–40.
21. Festinger L, Carlsmith JM. Cognitive consequences of forced compliance. J Abnorm Soc Psychol. 1959;58(2):203.
22. Festinger L. A theory of cognitive dissonance, vol. 2. Palo Alto: Stanford university press; 1962.
23. Fischer P, Frey D, Peus C, Kastenmuller A. The theory of cognitive dissonance: state of the science and directions for future research. Clashes of Knowledge. Dordrecht: Springer; 2008. p. 189–98.
24. Flache A, Macy MW. Small worlds and cultural polarization. J Math Sociol. 2011;35(1–3):146–76. https://doi.org/10.1080/0022250X.2010.532261.
25. Frey D. Recent research on selective exposure to information. Adv Exp Soc Psychol. 1986;19:41–80. https://doi.org/10.1016/S0065-2601(08)60212-9.
26. Garimella K, Gionis A, Parotsidis N, Tatti N. Balancing information exposure in social networks. Adv Neural Inf Process Syst. 2017;2017:4663–71.
27. Garimella K, De Francisci Morales G, Gionis A, Mathioudakis M. Reducing controversy by connecting opposing views. In: Proceedings of the tenth ACM international conference on web search and data mining, Cambridge; 2017.
28. Garimella K, Morales GDF, Gionis A, Mathioudakis M. Quantifying controversy on social media. ACM Trans Soc Comput. 2018;1(1):3.

29. Garimella K, De Francisci Morales G, Gionis A, Mathioudakis M. Polarization on social media (Tutorial); 2018. https://gvrkiran.github.io/polarization/. Accessed 15 Aug 2019.

30. Ghaderi J, Srikant R. Opinion dynamics in social networks with stubborn agents: equilibrium and convergence rate. Automatica. 2014;50(12):3209–15. https://doi.org/10.1016/j.automatica.2014.10.034.

31. Gottfried J, Stocking G, Grieco L. Partisans remain sharply divided in their attitudes about the news media. Pew Research Center; 2018. https://pewrsr.ch/2pqyCOQ. Accessed 10 Feb 2019.

32. Grabner-Krauter S, Bitter S. Trust in online social networks: a multifaceted perspective. Forum Soc Econ. 2015;44(1):48–68.

33. Granovetter MS. The strength of weak ties. Am J Sociol. 1973;78(6):1360–80.

34. Gregg BD. Frequency trails: modes and modality. http://www.brendangregg.com/FrequencyTrails/modes.html. Accessed 10 Feb 2019.

35. Guerra PHC, Meira Jr W, Cardie C, Kleinberg R. A measure of polarization on social media networks based on community boundaries. In: Proceedings of the 7th international AAAI conference on weblogs and social media, Cambridge; 2013.

36. Hedström P, Ylikoski P. Causal mechanisms in the social sciences. Annu Rev Sociol. 2010;36:49–67.

37. Hilmert CJ, Kulik JA, Christenfeld NJ. Positive and negative opinion modeling: the influence of another's similarity and dissimilarity. J Pers Soc Psychol. 2006;90(3):440.

38. Holley RA, Liggett TM. Ergodic theorems for weakly interacting infinite systems and the voter model. Ann Prob. 1975;3(4):643–63.

39. Himelboim I, McCreery S, Smith M. Birds of a feather tweet together: integrating network and content analyses to examine cross-ideology exposure on twitter. J Comput Mediat Commun. 2013;18(2):154–74. https://doi.org/10.1111/jcc4.12001.

40. Jones DA. The polarizing effect of new media messages. Int J Public Opin Res. 2002;14(2):158–74. https://doi.org/10.1093/ijpor/14.2.158.

41. Jonas E, Schulz-Hardt S, Frey D, Thelen N. Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. J Pers Soc Psychol. 2001;80(4):557–71.

42. Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. In: Proceedings of the nACM SIGKDD international conference on knowledge discovery and data mining, Washington; 2003.

43. Klapper JT. The effects of mass communication. New York: Free Press; 1960.

44. Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms. Phys Rev E. 2008;. https://doi.org/10.1103/PhysRevE.78.046110.

45. Lee IH. On the convergence of informational cascades. J Econ Theory. 1993;61(2):395–411.

46. Lewicki RJ, Bunker BB. Developing and maintaining trust in work relationships. Trust Org Front Theory Res. 1996;114:139.

47. Lord CG, Ross L, Lepper MR. Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. J Pers Soc Psychol. 1979;37(11):2098–109. https://doi.org/10.1037/0022-3514.37.11.2098.

48. Lorenz J. Continuous opinion dynamics under bounded confidence: a survey. Int J Mod Phys C. 2007;18(12):1819–38. https://doi.org/10.1142/S0129183107011789.

49. Malarz K, Gronek P, Kulakowski K. Zaller-Deffuant model of mass opinion. J Artif Soc Soc Simul. 2011;. https://doi.org/10.18564/jasss.1719.

50. Mäs M, Flache A. Differentiation without distancing. Explaining bi-polarization of opinions without negative influence. PLoS ONE. 2013;8(11):e74516. https://doi.org/10.1371/journal.pone.0074516.

51. McKnight DH, Cummings LL, Chervany NL. Initial trust formation in new organizational relationships. Acad Manag Rev. 1998;23(3):473–90.

52. Mobilia M. Does a single zealot affect an infinite group of voters? Phys Rev Lett. 2003;91(2):028701. https://doi.org/10.1103/PhysRevLett.91.028701.

53. Newman N, Fletcher R, Kalogeropoulos A, Levy DA, Nielsen RK. Reuters institute digital news report. Oxford: University of Oxford; 2018.

54. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. Rev Gen Psychol. 1998;2(2):175. https://doi.org/10.1037/1089-2680.2.2.175.

55. Nyczka P, Sznajd-Weron K. Anticonformity or independence?—insights from statistical physics. J Stat Phys. 2013;151(1–2):174–202.

56. Prasetya HA, Murata T. Modeling the co-evolving polarization of opinion and news propagation structure in social media. In: International conference on complex networks and their applications, Cambridge; 2018.

57. Rajpal H, Rosas F, Jensen H. Tangled worldview model of opinion dynamics. arXiv preprint; 2019. arxiv:1901.06372. Accessed 7 Aug 2019.

58. Rosas F, Hsiao JH, Chen KC. A technological perspective on information cascades via social learning. IEEE Access. 2017;5:22605–33.

59. Rychwalska A, Roszczyńska-Kurasińska M. Polarization on social media: when group dynamics leads to societal divides. In: Hawaii international conference on system sciences 2018, Hawaii; 2018.

60. Salzarulo L. A continuous opinion dynamics model based on the principle of meta-contrast. J Artif Soc Soc Simul 9(1):1–13.

61. Shearer E, Gottfried J. News use across social media platforms 2017. Pew Research Center; 2017. http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/. Accessed 10 Feb 2019.

62. Sherchan W, Nepal S, Paris C. A survey of trust in social networks. ACM Comput Surv (CSUR). 2013;45(4):47.

63. Sîrbu A, Pedreschi D, Gianotti F, Kertesz J, Algorithmic Bias Amplifies Opinion Polarization: A Bounded Confidence Model. arXiv preprint. https://arxiv.org/abs/1803.02111 2018. Accessed 10 February 2019.

64. Smith A, Anderson M. Social media use in 2018. Pew internet & American life project; 2018. http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/. Accessed 10 Feb 2019.

65.  Smith MA, Rainie L, Shneiderman B, Himelboim I. Mapping Twitter topic networks: from polarized crowds to community clusters. Pew Res Center. 2014;20:1–56.

66.  Sobkowicz P. Modelling opinion formation with physics tools: call for closer link with reality. J Artif Soc Soc Simul. 2009;12(1):11.

67.  Sobkowicz P. Extremism without extremists: Deffuant model with emotions. Front Phys. 2015;3:17. https://doi.org/10.3389/fphy.2015.00017.

68.  Sunstein CR. The law of group polarization. J Polit Philos. 2002;10(2):175–95. https://doi.org/10.1111/1467-9760.00148.

69.  Sunstein CR. Republic.com 2.0. Princeton: Princeton University Press; 2007.

70.  Sznajd-Weron K, Sznajd J. Opinion evolution in closed community. Int J Mod Phys C. 2000;11(06):1157–65. https://doi.org/10.1142/S0129183100000936.

71.  Timothy JJ. How does propaganda influence the opinion dynamics of a population?; 2017. arXiv preprint. arxiv:1703.10138. Accessed 10 Feb 2019.

72.  Vicario MD, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Stanley HE, Quattrociocchi W. The spreading of misinformation online. Proc Natl Acad Sci. 2016;113(3):554–9. https://doi.org/10.1073/pnas.1517441113.

73.  Del Vicario M, Zollo F, Caldarelli G, Scala A, Stanley HE. Quattrociocchi: mapping social dynamics on facebook: the brexit debate. Soc Netw. 2017;50:6–16. https://doi.org/10.1016/j.socnet.2017.02.002.

74.  Vicario MD, Scala A, Caldarelli G, Stanley HE, Quattrociocchi W. Modeling confirmation bias and polarization. Scientific Rep. 2017;7:40391. https://doi.org/10.1038/srep40391.

75.  Villi M, Matikainen J, Khaldarova I. Recommend, Tweet, share: user-distributed content (UDC) and the convergence of news media and social networks. Media convergence handbook, vol. 1. Berlin: Springer; 2016.

76.  Woolley SC. Automating power: social bot interference in global politics. First Monday. 2016;. https://doi.org/10.5210/fm.v21i4.6161.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.