# Bias in Social Interactions and Emergence of Extremism in Complex Social Networks

Vu X. Nguyen,[1] Gaoxi Xiao,[1,2,a)] Jie Zhou,[3] Guoqi Li,[4] and Beibei Li[5]
[1)] *School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore*
[2)] *Complexity Institute, Nanyang Technological University - 18 Nanyang Drive, Singapore 637723, Singapore*
[3)] *School of Physics and Materials Science, East China Normal University, Shanghai, China 200241, China*
[4)] *Center for Brain Inspired Computing Research, Department of Precision Instrument, Tsinghua University, Beijing, China 100084, China*
[5)] *College of Cybersecurity, Sichuan University, Chengdu, China 610065, China*

(Dated: 16 September 2020)

Emergence of extremism in social networks is among the most appealing topics of opinion dynamics in computational sociophysics in recent decades. Most of the existing studies presume that the initial existence of certain groups of opinion extremities and the intrinsic stubbornness in individuals' characteristics are the key factors allowing the tenacity or even prevalence of such extreme opinions. We propose a modification to the consensus making in bounded confidence models where two interacting individuals holding not so different opinions tend to reach a consensus by adopting an intermediate opinion of their previous ones. We show that if individuals make biased compromises, extremism may still arise without a need of an explicit classification of extremists and their associated characteristics. With such biased consensus making, several clusters of diversified opinions are gradually formed up in a general trend of shifting towards the extreme opinions close to the two ends of opinion range, which may allow extremism communities to emerge and moderate views to be dwindled. Further, we assume stronger compromise bias near opinion extremes. It is found that such a case allows moderate opinions a greater chance to survive compared to that of the case where the bias extent is universal across the opinion space. As to the extreme opinion holders' lower tolerances towards different opinions, which arguably may exist in many real-life social systems, they significantly decrease the size of extreme opinion communities rather than helping them to prevail. Brief discussions are presented on the significance and implications of these observations in real-life social systems.

**Extremism can be observed in many social systems. Understanding the mechanisms leading to the emergence and persistence of extreme opinions in social systems is of both research interest and importance in real life. While a well-known argument is that the persistence of extremism is largely due to the intrinsic characteristics and personalities of extremism-oriented individuals, such as possessing a low level of susceptibility to persuasion and a high commitment to inherent ideology, it remains as an important research topic on what else could be the driving forces pushing extremism to existence. We show in this paper that the proliferation of extremism may also be rooted from consensus making interactions where two interacting individuals tend to reach an agreement if they hold not so different views. Specifically, by quantifying the compromise bias in interpersonal interactions by a bias parameter on top of a well-known consensus making model, we show that bias in pairwise contacts may lead to opinion polarization on the population level, and in many cases, the emergence or even the prevalence of extreme entities. We also evaluate the effects that extreme opinion holders tend to have a stronger bias in consensus making while having a lower tolerance of different opinions. It is found that the stronger bias that extremists tend to have may actually give moderate opinion clusters a better chance to grow and sustain, while the lower tolerance of the extremist clusters significantly decreases their own growth. Such observations may help explain why extremism widely exists in various social systems yet does not enjoy a high chance to prevail, as well as shedding light on how we may counter extremism in social systems.**

## I. INTRODUCTION

Opinion dynamics is one of the most attractive topics in the past decades. A part of the studies on opinion dynamics aim to build mathematical models describing human interactions based on certain existing theories in sociology, social psychology, and complex sciences, such as social comparison theory[1], social power[2], cognitive dissonance theory[3], balance theory[4,5], and others[6]. Motivated by the developments of complex network theo-

[a)] Electronic mail: egxxiao@ntu.edu.sg

ries, opinion dynamics has been extensively studied based on social network models, in which individuals are typically represented as network vertices (nodes) and interactions between them as network edges (links)[7–13]. The opinion of each individual is typically represented as a value assigned to the corresponding vertex, while the strength of interactions may be reflected as edge weights when needed. While such a simplified approach may unavoidably miss revealing some crucial ingredients in complex interpersonal interactions, it nevertheless helps understand and evaluate the roles that certain key factors are playing in social actions and interactions as well as achieving valuable insights into complex social behaviors[14].

Many network-based opinion dynamics models have been proposed and extensively studied. One of the first works is the voter model[15], where each voter is endowed with a binary opinion, e.g., $-1$ or $+1$, and iteratively at each time step, s/he imposes her/his choice on one of her/his neighbors who is chosen at random. This model together with others including Galam[16], Sznajd[17], etc., are among the most studied on dynamics of discrete opinions, in general, and those of two competing opinions, in particular[10,18,19].

Another class of studies on opinion formation focuses on the evolution of the public opinion within a full spectrum ranging from totally agreeing to strongly disagreeing, rather than on evolution of for/against opinions on a certain issue. For such studies, continuous-valued opinion models are typically adopted. One of the well-known studies was undertaken by DeGroot[20], whose model was later extended in[21–23], etc. In the model, each agent holds a continuous-valued opinion and evolves by frequently adopting the weighted average of his/her neighbors' opinions. Another class of models which have probably even more extensively studied are bounded confidence models, which typically assume that a person only interacts with those whose opinions lie within a certain tolerance range (equivalently bound of confidence) of his/her opinion[24–27]. Among the family of bounded confidence models, the model proposed by Deffuant et al.[24] is widely recognized as a typical one. In the model, after an interpersonal interaction, two social entities either maintain their current views or compromise with each other to some extent depending on how much their prior opinions differ. Despite its algorithmic simplicity, the Deffuant model exhibits rich dynamics, grabbing much attention with many extended works, e.g., see[27–29].

Emergence of extremism, in which a part of population gradually forms up and persists on an extreme opinion towards a subject, is one of the most appealing properties of social systems. The coincidence, to a certain extent, between existing mathematical model-based observations and social phenomena triggers more studies to uncover the key factors leading to the formation, the persistence, and even the proliferation of extreme attitudes on social networks. In processes leading to such extremism, the notion of *extreme agents* who intrinsically maintain a low susceptibility to persuasion and high persistence of opinion is typically presented. Such agents are also labeled as zealots, extremists, or inflexibles interchangeably. A few existing studies demonstrate that the presence of extremists persisting on long-lasting attitudes may swap a large part or even a whole of the population, causing a social network, with an open conflict, to end up being separated into different or even opposing opinion communities[30–35]. Other studies devote to finding other possible explanations to the emergence of extremism, e.g., by introducing tolerance threshold as a function of opinion, such as[36,37], or basing mathematical models on social psychology theories such as biased assimilation (see[38]) in studies[39,40] or group polarization (see[41]) in[42,43].

Motivated by observations that (i) assuming the preexistence of extreme opinions does not explain how they emerge at the first place and (ii) polarization and extremism may appear in societies as a consequence of regular social interactions over a sufficiently long period of time, we propose to study on a simple approach that allows extremism to emerge from consensus/compromise making interactions where two interacting individuals holding not so different views tend to adopt a consensus opinion bounded in between their previous opinions. Specifically, we study on a simple model which may appear to be quite similar to consensus making in the original bounded-tolerance models with the only difference that the consensus making could be slightly or significantly biased towards the two ends of the opinion spectrum. That is, instead of always agreeing on a central value between two opinions in consensus making as that in the original Deffuant model, we let the consensus be biased to a certain extent depending on what the central value is. Specifically, if the central value is in the left/right half of the opinion distribution range, we let the consensus be biased towards left/right, respectively. Bias in consensus making would not happen if the central value happens to be at the middle of the opinion distribution range (e.g., at exactly 0.5 if we denote the opinion distribution range as [0,1]). It is worth noting that this is in line with the group polarization phenomena found in social psychology experiments, where a group, after social interactions, tends to make decisions that are more extreme than the initial inclination of its members[41,44]. It is shown that under such case, the consensus making, with a certain level of *bias* which may arguably be a part of human nature, may allow the emergence or even prevalence of extremism. This may help explain why extreme ideas can be observed in almost any human societies.

Note that the "bias in social interactions" is not a new term appearing in literature. In different references, however, it has different definitions and interpretations. Specifically, some have considered the bias in partner selection for social interactions where individuals, in a biased manner, tend to interact more frequently with those holding a similar opinion as themselves[12,29,45]. Studies on such a kind of bias may find their origin in confirmation bias phenomenon in social psychology[46] (also known as selective exposure[47–49] or congeniality bias[49,50]) that refers to individuals' tendency to expose themselves to favorable information which reinforces their pre-existing opinions while avoiding contradictory views. Another type of bias was called assimilation bias (also known as interpretation bias)[38,51,52], which refers to individuals' tendency to process new information with a bias towards their existing beliefs by accepting information that favors their beliefs while evaluating unfavorable ones critically. The effects of such type of bias have been examined in the DeGroot model and its variants[39,53]. A comprehensive discussion on biases in opinion dynamics can be found in[54]. The bias studied in our paper is different: as aforementioned, it is in line with group polarization theory where relatively more extreme decisions are made in social interactions. To the best of our knowledge, we are the first one to incorporate such bias factor into the Deffuant rule-grounded consensus formation model.

Our studies then further consider the effects of two factors which arguably may be observed in many, though definitely not all, social systems: (i) communities holding opinions closer to the two ends (far-left and far-right) may tend to be more biased towards their respective ends (extremist communities tend to be self-stimulated to be more extreme[38,39,55,56]); and (ii) extremists tend to be less tolerant of different opinions[57,58]. By slightly extending the basic biased consensus making model to take these two factors into account, we show that the former factor, in fact, offers moderate opinions a better chance to survive compared to that of the case where bias extent is universal across the opinion space. As to the latter factor, generally speaking, it contributes to significantly decreasing the size of the extremist communities rather than helping them to prevail. Such observations, to a certain extent, may help explain why, though extremism widely exists in most social systems, extreme opinions seldom prevail to become the mainstream opinions of human societies. Brief discussions are also given on why the latter factor, though it hinders the growth of extremist communities, can nevertheless be observed in so many social systems.

The rest of this report is organized as follows. First, the biased consensus making model is proposed, followed by some brief theoretical analyses. Then we present simulation results and discussions to demonstrate how the simple model may allow extremism to easily emerge. Also, we will evaluate and have some brief discussions on the effects of the relatively stronger bias and lower tolerance of extreme opinion holders. Finally, we shall conclude the report and briefly discuss some future research directions.

## II. BIASED CONSENSUS MAKING MODEL

### A. Model Definition

Let $G = (V, E)$ denote a connected undirected graph (network) composed of a set of vertices (nodes) $V$ and a set of edges (links) $E \subseteq V \times V$ on which opinions evolve. The sets $V$ and $E$ represent individuals and the social connections between them, respectively. Let $N = |V|$ and $\langle k \rangle = \frac{2|E|}{|V|}$ denote the network size and the average nodal degree, respectively. At time step $t = 0$, each node $u \in V$ holds a continuous-valued opinion drawn from a uniform distribution on the region [0,1], denoted by $o_u(t)$, where values 0 and 1 correspond to leftmost and rightmost extreme opinions, respectively, and values in between represent other relatively moderate, including neutral (with a value of 0.5), opinions. At each time step $t$, a node randomly chosen from $V$, say $u$, interacts with one of its neighbors, say $v$. They communicate and as a result, may change their current opinions following the updating rule: if their opinions differ an amount larger than a certain tolerance range $d \in [0, 1]$, i.e., $|o_u(t) - o_v(t)| > d$, they maintain their current opinions; otherwise, they adopt a new opinion as follows:

$$\begin{aligned} o_u(t+1) &= o_v(t+1) \\ &= \bar{o}_{u,v}(t) + \delta \frac{|o_u(t) - o_v(t)|}{2} \mathrm{sgn}\left(\bar{o}_{u,v}(t) - 0.5\right), \end{aligned} \tag{1}$$

where $\bar{o}_{u,v}(t) = \frac{o_u(t)+o_v(t)}{2}$, $\delta$ is a function quantifying the relative extent of bias in local consensus making, termed as *bias parameter*. A larger value of $\delta$ implies the existence of a stronger bias. In the present study, we investigate two cases of $\delta$:

- $\delta = \alpha$ is a scalar, $1 \geq \alpha \geq 0$, referred to as *opinion-independent bias* case, where there exists an identical extent of bias over the entire opinion range.
- $\delta = \delta(o_u(t), o_v(t)) = 2\beta\,|\bar{o}_{u,v}(t) - 0.5|$, $1 \geq \beta \geq 0$, referred to as *opinion-dependent bias* case, where consensus making interactions on regions closer to the two ends of the opinion spectrum are more biased, and vice versa.

The opinion evolution stops once the network reaches a steady state at which any two connected nodes are holding opinions with a difference either being larger than $d$ or being less than a small amount (set to be $10^{-3}$ in our simulations).

Note that other models certainly can also be proposed to reflect the non-uniform distribution of bias at different opinion ranges. As the main focus of this report is to evaluate the effects of bias on consensus making and how such bias may allow extremism to easily emerge and persist, rather than proposing a general framework, studies on more general bias models and their corresponding effects are largely out of the scope of this report.

It is also worth noting that there may be different ways to represent bias depending on particular models. In Eq. (1), we define bias in consensus/compromise making interactions where consensus opinion remains within the range bounded by the two prior opinions. In other models such as bounded confidence models with repulsion (or rejection) mechanism[59–62], where two dissenters tend to widen their difference in opinions after discussion, the bias towards respective extremes would likely be amplified. Such cases, with their very interesting dynamics, are out of the scope of this study.

### B. Dynamics equation of opinion evolution

Denote the probability density of opinion $o$ at time step $t$ as $P_o(t)$. For any $o_1$ and $o_2$ satisfying $|o_1 - o_2| \leq d$, denote $o_c = \frac{o_1+o_2}{2} + \delta\frac{|o_1-o_2|}{2}\mathrm{sgn}(\frac{o_1+o_2}{2} - 0.5)$, which is the opinion value of the consensus to be made as shown in Eq. (1). At time $t$, an interaction between two agents holding opinions $o_1$ and $o_2$ leads to consensus making. Mathematically, this corresponds to an increase in $P_{o_c}(t)$ and a decrease in $P_{o_1}$ and $P_{o_2}$. The probability that such an interaction occurs at $t$ is proportional to $P_{o_1}(t)P_{o_2}(t)$. Adopting the mean-field method, we derive the dynamics equation of $P_o(t)$ as follows:

$$\begin{aligned}\frac{\mathrm{d}P_o(t)}{\mathrm{d}t} =& 2\iint\limits_{|o_1-o_2|\leq d} P_{o_1}(t)P_{o_2}(t)\sigma(o-o_c)\mathrm{d}o_1\mathrm{d}o_2 \\ &- \iint\limits_{|o_1-o_2|\leq d} P_{o_1}(t)P_{o_2}(t)\sigma(o-o_1)\mathrm{d}o_1\mathrm{d}o_2 \\ &- \iint\limits_{|o_1-o_2|\leq d} P_{o_1}(t)P_{o_2}(t)\sigma(o-o_2)\mathrm{d}o_1\mathrm{d}o_2,\end{aligned} \tag{2}$$

where $\sigma(x) = \begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{otherwise} \end{cases}$. As $o_1$ and $o_2$ are symmetric variables and $P_o(t)$ is subject to normalization, Eq. (2) can be simplified to:

$$\begin{aligned}\frac{\mathrm{d}P_o(t)}{\mathrm{d}t} =& \iint\limits_{|o_1-o_2|\leq d} P_{o_1}(t)P_{o_2}(t)\sigma(o-o_c)\mathrm{d}o_1\mathrm{d}o_2 \\ &- \iint\limits_{|o_1-o_2|\leq d} P_{o_1}(t)P_{o_2}(t)\sigma(o-o_1)\mathrm{d}o_1\mathrm{d}o_2.\end{aligned} \tag{3}$$

Equation (3) is not amenable to closed-form solution. However, by numerical computations, the temporal evolution of opinions can be approximately produced when $t$ is given a sufficiently large value. We will show a good match between simulation and theoretical results in the next section.

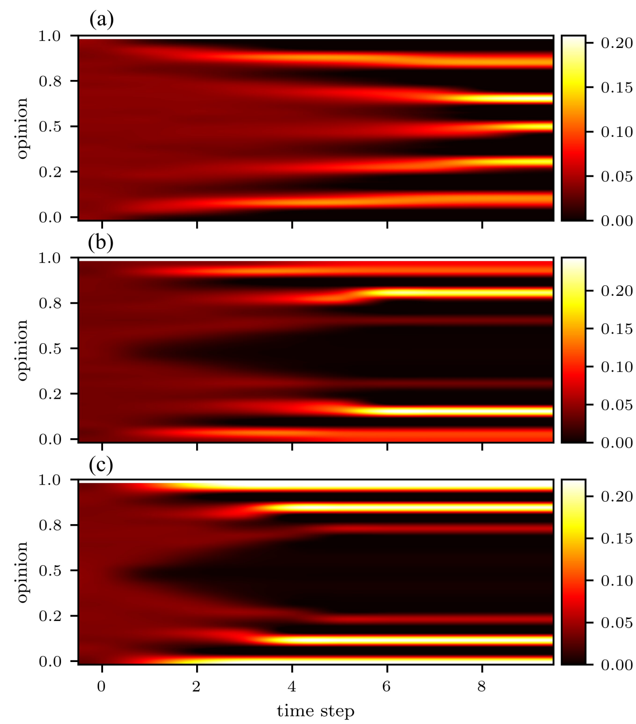


FIG. 1. Illustration of temporal distributions of opinions of 20000 agents connected into an ER network with an average nodal degree of 20. $d = 0.1$, $\alpha = 0$ (a), 0.5 (b), and 0.8 (c). Each time step at which the temporary opinion distribution is recorded corresponds to 20000 pair-wise interactions.

## III. RESULTS AND DISCUSSIONS

### A. Effects of bias on opinion cluster formation

We start by showing the temporal evolution of opinion distributions under different values of $\alpha$. As our extensive observations have confirmed that the system performances are similar on Erdös-Rényi (ER)[63] and scale-free (SF)[64] networks, unless otherwise specified, hereafter we shall mainly present simulation results on ER networks $G(N, p)$ with $N = 20000$ and $p = 10^{-3}$ and SF networks with a characteristic degree exponent of 2, a minimum degree of 7, and a degree cutoff of 140. With these settings, the average nodal degrees of the ER and SF networks are approximately 20 and 21, respectively. All synthetic networks are generated using the configuration method[65].

Figure 1 shows the results for a few different cases with different values of $\alpha$ while the tolerance range remains at $d = 0.1$. Note that the case where $\alpha = 0$ is equivalent to the original Deffuant model with a convergence rate – the rate of two agents moving towards each other in the opinion space in a local consensus making[24] – of 0.5. Over time, all nodes converge to 5 major opinion clusters (roughly equaling $\frac{1}{2d}$[24,29]) interpreted as communities of agents holding similar opinions.

We can observe that with an increasing value of $\alpha$, two peaks close to the two ends emerge more quickly in opinion evolution, and they get closer to their respective end values. We hereafter refer to those nodes bearing opinions right at, or relatively nearby, the two ends of the opinion spectrum as extremists. The bias in local interactions leads to the emergence and persistent existence of these extremist clusters, and concurrently the decline of clusters holding moderate opinions.

Figure 2 depicts, for the case with opinion-independent bias, the opinion distributions at equilibrium where the opinion evolution has come to an end. As can be observed, the opinions in the steady state diverge into several discontinuous regions holding different opinions. These regions are generally pushed towards the two ends of the opinion axis due to the effects of the bias. Figure 2 – *middle row* shows the co-existence of the two highest peaks corresponding to the co-existence of two major communities acquiring the majority of the population. Both of these communities hold moderate opinions until the bias in local interactions causes the polarized peaks to shift to the two ends. It is of interest to observe that such microscopic/local bias may eventually cause the extremism in a macro/global scope. Noticeably, even when $d = 0.3$, which is sufficiently large for the majority of the population to converge to the single cluster of neutral opinion where there is no bias in opinion formation[29], the existence of bias, when it is strong enough, may separate the peak into two shifting to the two ends, as can be observed in Fig. 2 – *bottom row*. Besides the general trend of agents shifting to the extremities under relatively high extents of bias, another common observation for the three cases of different $d$ values is the gradual vanishment of intermediate opinions, most noticeably the disappearance of the neutral opinion around 0.5. Also note that Fig. 2 shows the results for both the ER and the SF networks. In our extensive simulations, all main conclusions remain the same in the two networks.

It may be noted that while the middle and right panels of Fig. 2 steadily show similar shapes, the heights of the peaks are usually different. This is especially noticeable in the bottom row. Such differences come from the fact that there exist many very low peaks across the opinion axis that can only be observed when being zoomed in. These low peaks may significantly affect the heights of the high peaks.

We shall then study the effects of the function $\delta$ on opinion formation with opinion-dependent bias, where different system dynamics emerge. Specifically, it is found that while the unbalance in local consensus making still pushes the clusters towards the two ends, having an opinion-dependent bias may lower the chance that extremist clusters can dominate the society. Observing Fig. 2 and Fig. 3, we could see that, when extremists are keeping their bias level largely unchanged, lowering the bias level among the rest of the population gives moderate opinion clusters a much higher chance to survive or even prevail in the final steady state.

This can be explained. For the case where extremists have a stronger bias in the opinion-dependent bias model, the term $|\bar{o} - 0.5|$ lets opinion formation closer to the two ends of the opinion axis have a faster speed shifting towards their respective ends than the other opinion formations in between. This difference in shifting speed may quickly enlarge the distance, measured on the opinion axis, between extremists and other moderate opinion holders. Once this distance is larger than the tolerance range, extremists can no longer drag more moderate opinion holders to change their opinions. Such speed asynchronization and the consequent quick isolation of extremist clusters give moderate opinion clusters a better chance to sustain, or even prevail in some cases.

An interesting observation is since the quick isolation of extremist clusters may become a weaker effect when the tolerance range $d$ is relatively large, in a population that is relatively more *open-minded*, if the so-called open-mindedness is revealed as having a larger tolerance range rather than a weaker bias, extremist clusters may actually have a better chance to prevail when the bias extent is sufficiently large. Figure 3 – *bottom row* illustrates such a phenomenon where two polarized peaks dominate a system with a large tolerance range of $d = 0.3$. Note that theoretical analysis results are also presented in Fig. 3, which have a reasonably good match with simulation results.

Figure 3 also reveals that the local bias makes it more difficult for a social network to reach a complete consensus. To better demonstrate this, in the steady state, we removed those links connecting nodes with an opinion difference larger than $d$. The resulting graph is composed of disconnected subgraphs, each representing an opinion cluster. It is shown in Fig. 4 that the local bias prolongs the global polarization state (with respect to $d$) in which the two largest clusters holding roughly symmetric opposing and favoring opinions coexist in the steady state. As can be seen, the polarization-to-consensus transition point shifts to the right in respect of $d$-axis with an increasing value of $\beta$, meaning that in a system with a stronger bias, it requires a larger tolerance range to achieve system-wide complete consensus.

## B. Heterogeneous tolerance range

Intuitively, due to the intrinsic characteristic of the attitude resistance of extremists, it is relatively more difficult for them to have *effective* contacts that may lead to a change in their current attitude. Inferentially, a person holding a stronger stance tends to have a lower tolerance level. Such observations have been supported by quite a few studies on social psychology, e.g.,[57,58,66]. It is of interest to investigate the effects of the extremists' lower tolerance on the growth of their clusters. We hereby adopt an idea proposed by[37] to quantify the tolerance range of an individual according to his/her current state.

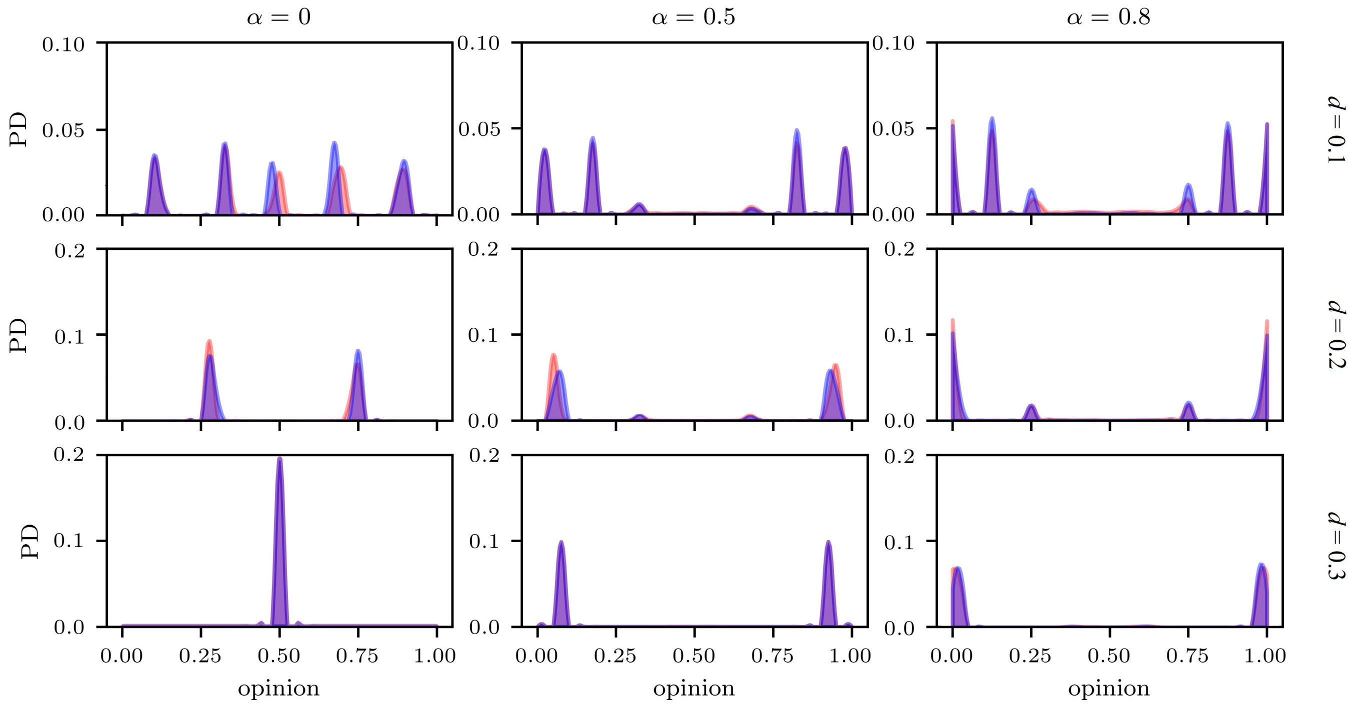


FIG. 2. Steady-state probability density (PD) of opinions for $\alpha = 0, 0.5$ and 0.8 (left, middle, and right column panels, respectively) and $d = 0.1, 0.2$, and 0.3 (top, middle, and bottom row panels, respectively) in ER (blue) and SF (red) networks.

Specifically, we let an agent holding a more extreme opinion have a narrower range of tolerance, and vice versa. As such, two agents would undergo an effective interaction only if the difference between their current opinions is within each other's tolerance ranges, or in other words, the difference is not bigger than the narrower one of the two agents' tolerance ranges. A simple formula is introduced as shown below, where $d$ is a symmetric function of $o$ with respect to $o = 0.5$ with the minimum and maximum values of $d$ being $d_{\min}$ and $d_{\max}$, respectively:

$$d(o) = 2(d_{\min} - d_{\max})|o - 0.5| + d_{\max}. \quad (4)$$

Clearly, $d(o)$ grows from $d_{\min}$ to $d_{\max}$ linearly as $o$ increases from 0 to 0.5 and then linearly returns to $d_{\min}$ as $o$ further increases to 1.

Figure 5 shows that, with or without bias (corresponding to different values of $\beta$), extremists with a narrower tolerance range would typically end up with having significantly smaller clusters than those of the moderate-opinion holders. This reveals that the extremists' lower tolerance towards different opinions may be one of the main reasons why, though we frequently witness the existence of extremist groups, we seldom see them prevail.

It is interesting to observe that having a lower tolerance level decreases the size of extremist clusters so significantly yet most, if not all, extremist groups do have relatively small tolerance ranges in opinion formation. Why don't these groups adopt a different strategy to become more tolerant, in which case they may have a much better chance to grow bigger and stronger? The answer may be that though having a larger tolerance range definitely helps cluster growth in opinion formation by consensus making, a more tolerant cluster may however unavoidably also be more tolerant towards other types of opinion changes (also known as mutation)[12,67,68]. As people may tend to change their minds when the cluster they belong to requests extensive devotion from its members emotionally, physically and/or financially, tolerating such opinion changes may allow the cluster to be quickly diminished under certain circumstances[69]. The best strategy for an extremist cluster to sustain, therefore, may be maintaining a low tolerance against any "unfavorable" opinion changes, even imposing penalties and punishments on such changes when necessary. Further studies are certainly needed to verify the above hypothesis.

Note that we have tested synthetic networks with different parameter values and observed that the main conclusions we presented in Subsec. III A and Subsec. III B basically always hold as long as the networks are not overly sparse or overly dense.

### C. Numerical simulations on real-life networks

In this section, we examine the opinion dynamics on networks retrieved from the real world. Ego networks are

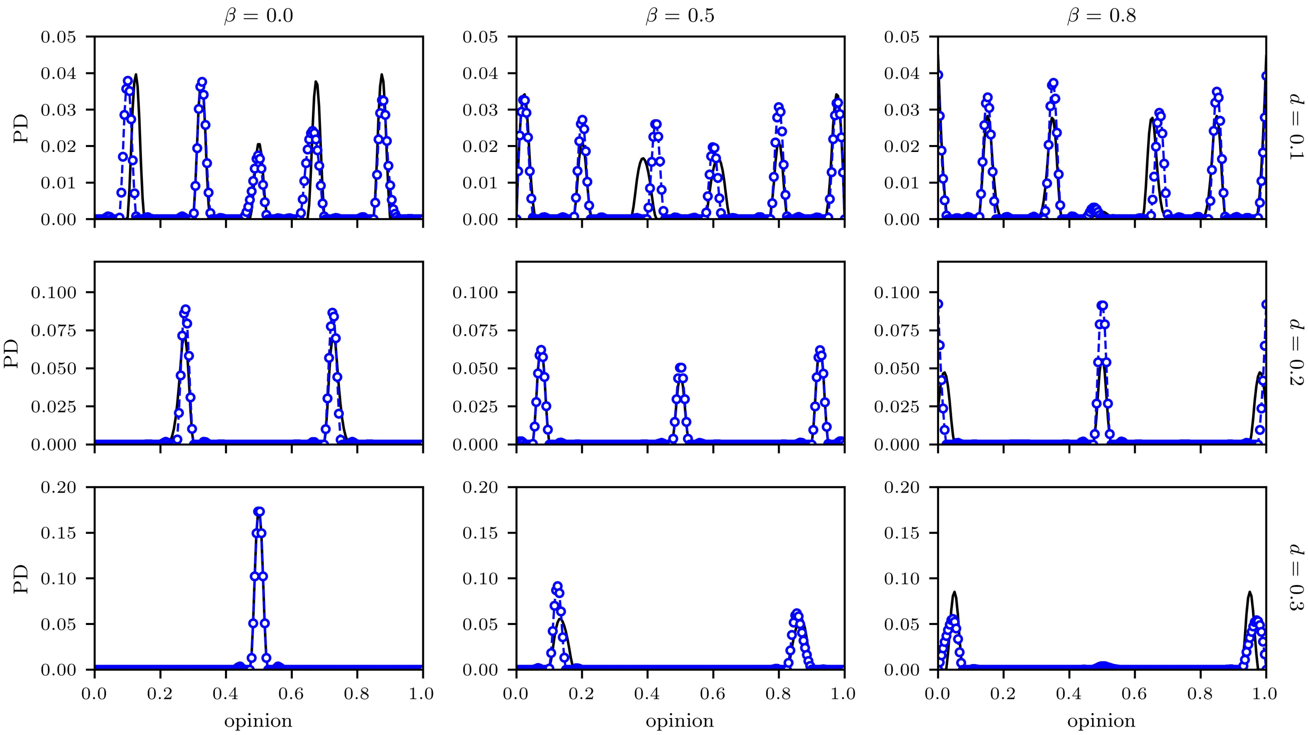


FIG. 3. Steady-state probability density (PD) of opinions for $\beta = 0, 0.5$, and 0.8 (left, middle, and right column panels, respectively) and ranges $d = 0.1, 0.2$, and 0.3 (top row, middle row, and bottom row panels, respectively) in ER networks. Both simulation (symbols) and theoretical (solid lines) results are presented. Each symbol corresponds to the fraction of nodes holding opinions in the respective range with a width of 0.005.

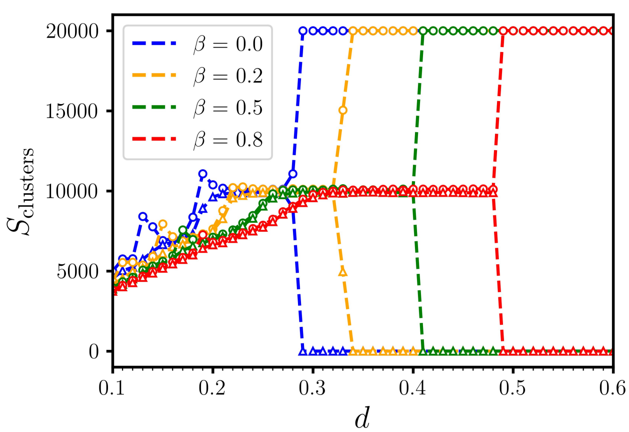


FIG. 4. Sizes of largest (circles) and second-largest (triangles) steady-state clusters in the ER networks.

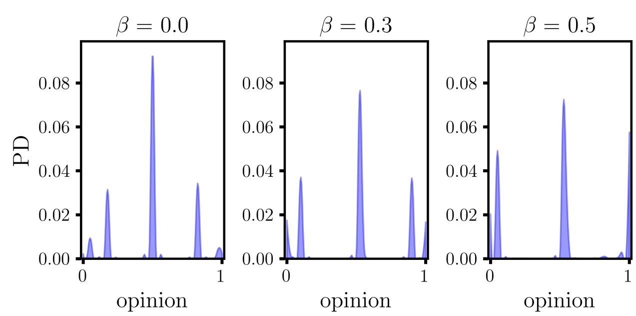


FIG. 5. Steady-state probability density (PD) of opinions versus $\beta$ in an ER network. The tolerance range is defined as in Eq. (4) with $d_{\min}, d_{\max} = 0.03, 0.3$, respectively.

of a typical topology in real-life social networks[70,71], each consisting of a focal node (ego) connected to its neighbors (alters) and other links among those alters. We carry out simulations on a network extracted from Facebook consisting of 10 connected component ego networks[72]. The network has 4039 nodes and 88234 links in total.

Figure 6 shows the temporal evolution of opinions in the network under the influences of opinion-dependent bias and heterogeneous tolerance. As that in the previous section, we set $d_{\min} = 0.03$ and $d_{\max} = 0.3$, respectively. Two different cases with different $\beta$ values reveal similar observations: the peaks closer to the two ends of the opinion axis quickly form up, largely due to their smaller tolerance ranges, stop adopting newcomers, and maintain their sizes while centrist opinions keep evolving.

Most of these moderates finally end up gathering together in the relatively large cluster holding opinions around 0.5. As $d_{\max} = 0.3$ is a relatively large value, only a single big cluster emerges in the moderate opinion

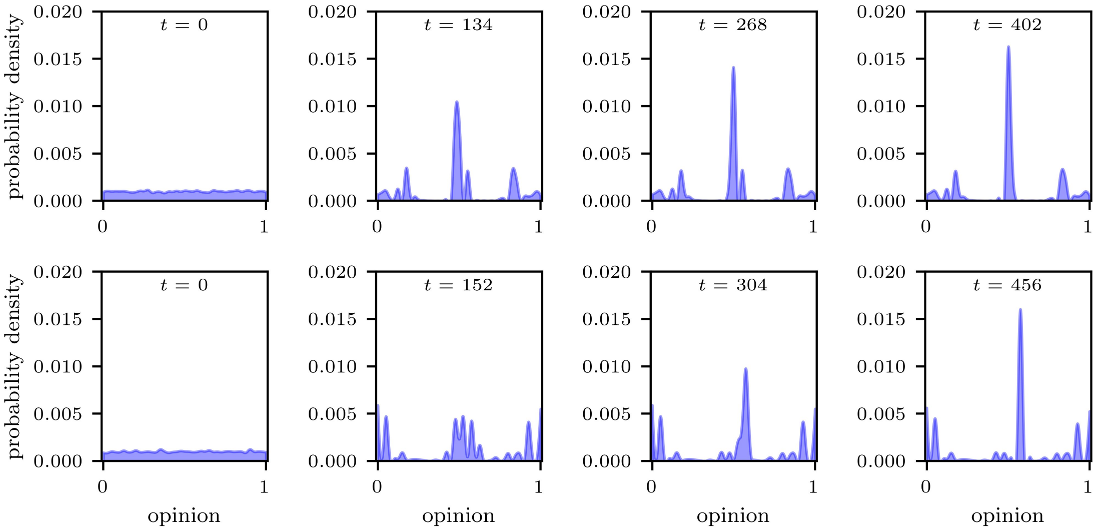


FIG. 6. Time evolution of opinions in the real-life network extracted from Facebook for $\beta = 0$ (top row panels) and $\beta = 0.5$ (bottom row panels). The tolerance range is defined as in Eq. (4) with $d_{\min}, d_{\max} = 0.03, 0.3$, respectively. Each time step at which the opinion distribution is recorded corresponds to 4039 pair-wise interactions.

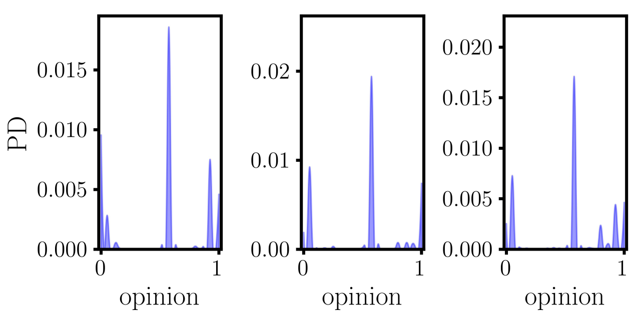


FIG. 7. Steady-state probability density (PD) of opinions in the three largest component ego networks for $\beta = 0.5$.

region. With a smaller $d_{\max}$ value, multiple moderate clusters would be formed up. More noticeably, the presence of the bias causes a significant increase in the heights of the peaks of extreme opinions compared to those in the original case ($\beta = 0$). Specifically, we may observe that extreme opinion peaks in the bottom right panel are roughly twice as high as those in the top right panel, and they are closer to the two ends as well. Such observations reveal that the conclusions holding for synthetic networks may remain valid in some real-life networks as well.

We note that the network used in this experiment has a strong community (modularity) structure, where each component ego network has a relatively denser internal connectivity than that among different component networks. This causes a higher frequency of intra-ego network interactions, resulting in an emergence of extremism in each component ego network. We demonstrate such an observation in Fig. 7.

It is worth noting that different network architectures and parameters, including community structures, clustering coefficient and so on, may contribute to making some differences in opinion formation. Opinion formation may also have different features in very sparse and overly dense networks, with certain initial opinion distributions. Discussions on these interesting issues, however, have to be left out to a separate report of our future studies.

## IV. CONCLUSIONS

In this paper, we studied the effects of bias on interpersonal consensus making. It is found that, when the bias extent is an identical scalar across the opinion spectrum, a general tendency of agents moving towards the two opinion extremities at a relatively similar pace can be observed. It is of interest that the microscopic bias in local pair-wise contacts could facilitate the public opinion polarization, allowing extremist clusters to emerge or even prevail. When there exist stronger biases in opinion formation closer to the two ends of opinion axis, which arguably may be the case in many social systems, extremist communities may quickly form up and get stabilized. Meanwhile, however, since they quickly isolate themselves from the rest of the population, moderate opinion clusters could have a better chance to grow up, flourishing a sustained diversity of opinions in the social systems. Lastly, we evaluated the effects where extreme opinion holders have a narrower tolerance range. It is found that such effects significantly decrease the size of the extremist clusters.

It is worth noting that, though having a narrower tolerance range decreases the size of extremist clusters, having a wider tolerance range may not necessarily help counter extremism. Our study suggests a hypothesis that it may be a more effective strategy to suppress bias, especially among moderate opinion holders, rather than encouraging everyone to become more tolerant of a wider range of different opinions. Further studies are needed along this direction.

As aforementioned, arguably we do not easily observe the phenomena of extremism domination in real life. For example, groups maintaining extreme attitudes on topics such as gun control, LBGT rights or immigration, exist concomitantly with individuals holding a wide spectrum of moderate views. While our work to a certain extent helps explain why extremism easily exists yet seldom prevails, further studies are needed at least on a few topics as follows: (i) how biases emerge and evolve in real-life social systems; (ii) whether social systems have a certain mechanism to make necessary self-correction to biases, and if they do, how the self-correction mechanism works (or fails to work); and (iii) when other factors, e.g., opinion mutation, are introduced into the model, whether and how any main conclusions would be affected or even totally changed. These topics, in addition to the one we briefly discussed in the last paragraph, shall be of our future research interest.

## ACKNOWLEDGMENTS

The research is partially supported by Ministry of Education, Singapore, under contract MOE2016-T2-1-119, Tsinghua University Initiative Scientific Research Program, and Tsinghua-Foshan Innovation Special Fund (TFISF). The work is also partially supported by the Fundamental Research Funds for the Central Universities, Sichuan University (Grant No. YJ201933), the China Postdoctoral Science Foundation (Grant No. 2019TQ0217), and the National Natural Science Foundation of China (Grant No. U1736212 and Grant No. 61572334). This work is also partially supported by the Future Resilient Systems Project - Stage II at the Singapore-ETH Centre (SEC), which is funded by the National Research Foundation of Singapore (NRF).

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

[1] L. Festinger, A theory of social comparison processes, Human Relations 7 (2) (1954) 117–140.
[2] J. R. French Jr, A formal theory of social power., Psychological Review 63 (3) (1956) 181.
[3] L. Festinger, A theory of cognitive dissonance, Vol. 2, Stanford University Press, 1957.
[4] D. Cartwright, F. Harary, Structural balance: a generalization of heider's theory., Psychological Review 63 (5) (1956) 277.
[5] F. Heider, Attitudes and cognitive organization, The Journal of Psychology 21 (1) (1946) 107–112.
[6] S. Galam, Sociophysics: a physicist's modeling of psycho-political phenomena, New York, NY: Springer, 2012.
[7] P. L. Krapivsky, S. Redner, Dynamics of majority rule in two-state interacting spin systems, Physical Review Letters 90 (23) (2003) 238701.
[8] M. Moussaïd, J. E. Kämmer, P. P. Analytis, H. Neth, Social influence and the collective dynamics of opinion formation, PloS One 8 (11) (2013) e78433.
[9] S. R. Etesami, T. Başar, Game-theoretic analysis of the hegselmann-krause model for opinion dynamics in finite dimensions, IEEE Transactions on Automatic Control 60 (7) (2015) 1886–1897.
[10] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, A. Scaglione, Binary opinion dynamics with stubborn agents, ACM Transactions on Economics and Computation (TEAC) 1 (4) (2013) 19.
[11] Y. Dong, Z. Ding, L. Martínez, F. Herrera, Managing consensus based on leadership in opinion dynamics, Information Sciences 397 (2017) 187–205.
[12] Y. Yu, G. Xiao, G. Li, W. P. Tay, H. F. Teoh, Opinion diversity and community formation in adaptive networks, Chaos: An Interdisciplinary Journal of Nonlinear Science 27 (10) (2017) 103115.
[13] V. X. Nguyen, G. Xiao, X.-J. Xu, G. Li, Z. Wang, Opinion formation on multiplex scale-free networks, EPL (Europhysics Letters) 121 (2) (2018) 26002.
[14] J. J. Schneider, C. Hirtreiter, The impact of election results on the member numbers of the large parties in bavaria and germany, International Journal of Modern Physics C 16 (08) (2005) 1165–1215.
[15] P. Clifford, A. Sudbury, A model for spatial conflict, Biometrika 60 (3) (1973) 581–588.
[16] S. Galam, Real space renormalization group and totalitarian paradox of majority rule voting, Physica A: Statistical Mechanics and its Applications 285 (1-2) (2000) 66–76.
[17] K. Sznajd-Weron, J. Sznajd, Opinion evolution in closed community, International Journal of Modern Physics C 11 (06) (2000) 1157–1165.
[18] S. Biswas, P. Sen, Model of binary opinion dynamics: Coarsening and effect of disorder, Physical Review E 80 (2) (2009) 027101.
[19] Y. Wang, G. Xiao, J. Liu, Dynamics of competing ideas in complex social systems, New Journal of Physics 14 (1) (2012) 013015.
[20] M. H. DeGroot, Reaching a consensus, Journal of the American Statistical Association 69 (345) (1974) 118–121.
[21] N. Friedkin, E. C. Johnsen, Social influence networks and opinion change, Advances in Group Processes (1999).
[22] Z. Xu, J. Liu, T. Başar, On a modified Degroot-Friedkin model of opinion dynamics, in: 2015 American Control Conference (ACC), IEEE, 2015, pp. 1047–1052.
[23] V. Amelkin, F. Bullo, A. K. Singh, Polar opinion dynamics in social networks, IEEE Transactions on Automatic Control 62 (11) (2017) 5650–5665.
[24] G. Deffuant, D. Neau, F. Amblard, G. Weisbuch, Mixing beliefs among interacting agents, Advances in Complex Systems 3 (01n04) (2000) 87–98.
[25] R. Hegselmann, U. Krause, et al., Opinion dynamics and bounded confidence models, analysis, and simulation, Journal of Artificial Societies and Social Simulation 5 (3) (2002).
[26] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, Reviews of Modern Physics 81 (2) (2009) 591.
[27] X. F. Meng, R. A. Van Gorder, M. A. Porter, Opinion formation and distribution in a bounded-confidence model on various networks, Physical Review E 97 (2) (2018) 022312.
[28] S. Fortunato, Universality of the threshold for complete consensus for the opinion dynamics of Deffuant et al., International Journal of Modern Physics C 15 (09) (2004) 1301–1307.
[29] B. Kozma, A. Barrat, Consensus formation on adaptive networks, Physical Review E 77 (1) (2008) 016102.
[30] S. Galam, F. Jacobs, The role of inflexible minorities in the breaking of democratic opinion dynamics, Physica A: Statistical Mechanics and its Applications 381 (2007) 366–376.

[31] G. Weisbuch, G. Deffuant, F. Amblard, Persuasion dynamics, Physica A: Statistical Mechanics and its Applications 353 (2005) 555–575.

[32] S. Chen, D. H. Glass, M. McCartney, Characteristics of successful opinion leaders in a bounded confidence model, Physica A: Statistical Mechanics and its Applications 449 (2016) 426–436.

[33] K. Tucci, J. González-Avella, M. Cosenza, Rise of an alternative majority against opinion leaders, Physica A: Statistical Mechanics and its Applications 446 (2016) 75–81.

[34] C. Jin, Y. Li, X. Jin, Political opinion formation: Initial opinion distribution and individual heterogeneity of tolerance, Physica A: Statistical Mechanics and its Applications 467 (2017) 257–266.

[35] A. C. Martins, S. Galam, Building up of individual inflexibility in opinion dynamics, Physical Review E 87 (4) (2013) 042807.

[36] G. Deffuant, F. Amblard, G. Weisbuch, Modelling group opinion shift to extreme: the smooth bounded confidence model, 2nd European Social Simulation Association (ESSA) Conference. Valladolid, Spain (2004) (2004).

[37] P. Sobkowicz, Extremism without extremists: Deffuant model with emotions, Frontiers in Physics 3 (2015) 17.

[38] C. G. Lord, L. Ross, M. R. Lepper, Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence., Journal of Personality and Social Psychology 37 (11) (1979) 2098.

[39] P. Dandekar, A. Goel, D. T. Lee, Biased assimilation, homophily, and the dynamics of polarization, Proceedings of the National Academy of Sciences 110 (15) (2013) 5791–5796.

[40] B. O. Baumgaertner, R. C. Tyson, S. M. Krone, Opinion strength influences the spatial dynamics of opinion formation, The Journal of Mathematical Sociology 40 (4) (2016) 207–218.

[41] C. R. Sunstein, The law of group polarization, Journal of Political Philosophy 10 (2) (2002) 175–195.

[42] S. Galam, M. A. Javarone, Modeling radicalization phenomena in heterogeneous populations, PloS One 11 (5) (2016).

[43] M. A. Javarone, S. Galam, Emergence of extreme opinions in social networks, in: International Conference on Social Informatics, Springer, 2014, pp. 112–117.

[44] D. G. Myers, H. Lamm, The group polarization phenomenon., Psychological Bulletin 83 (4) (1976) 602.

[45] A. Sîrbu, D. Pedreschi, F. Giannotti, J. Kertész, Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model, PloS One 14 (3) (2019) e0213246.

[46] E. Jonas, S. Schulz-Hardt, D. Frey, N. Thelen, Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information., Journal of Personality and Social Psychology 80 (4) (2001) 557.

[47] D. Frey, Recent research on selective exposure to information, in: Advances in Experimental Social Psychology, Vol. 19, Elsevier, 1986, pp. 41–80.

[48] N. Stroud, Revisiting the concept of selective exposure, in: Annual Meeting of the International Communication Association, San Francisco, CA, 2007.

[49] W. Hart, D. Albarracín, A. H. Eagly, I. Brechan, M. J. Lindberg, L. Merrill, Feeling validated versus being correct: a meta-analysis of selective exposure to information., Psychological Bulletin 135 (4) (2009) 555.

[50] A. H. Eagly, S. Chaiken, The psychology of attitudes., Harcourt Brace Jovanovich College Publishers, 1993.

[51] N. Schwarz, H. Bles, Scandals and the public's trust in politicians: Assimilation and contrast effects, Personality and Social Psychology Bulletin 18 (5) (1992) 574–579.

[52] J. Risen, T. Gilovich, R. Sternberg, D. Halpern, H. Roediger, Informal logical fallacies, Critical Thinking in Psychology 110 (2007).

[53] W. Xia, M. Ye, J. Liu, M. Cao, X.-M. Sun, Analysis of a nonlinear opinion dynamics model with biased assimilation, Automatica 120 (2020) 109113.

[54] D. Urbig, et al., Attitude dynamics with limited verbalisation capabilities, Journal of Artificial Societies and Social Simulation 6 (1) (2003) 2.

[55] H. Brandstätter, Social emotions in discussion groups (1978).

[56] C. R. Sunstein, Why they hate us: The role of social dynamics, Harv. JL & Pub. Pol'y 25 (2001) 429.

[57] J. O. Whittaker, Perception and judgment in the political extremist 1, Journal of Communication 17 (2) (1967) 136–141.

[58] J.-W. van Prooijen, A. P. Krouwel, Extreme political beliefs predict dogmatic intolerance, Social Psychological and Personality Science 8 (3) (2017) 292–300.

[59] C. Crawford, L. Brooks, S. Sen, Opposites repel: the effect of incorporating repulsion on opinion dynamics in the bounded confidence model, in: Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems, International Foundation for Autonomous Agents and Multiagent Systems, 2013, pp. 1225–1226.

[60] W. Jager, F. Amblard, Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change, Computational & Mathematical Organization Theory 10 (4) (2005) 295–303.

[61] E. Cornacchia, N. Singer, E. Abbe, Polarization in attraction-repulsion models, arXiv preprint arXiv:2006.05251 (2020).

[62] S. Huet, G. Deffuant, W. Jager, A rejection mechanism in 2D bounded confidence provides more conformity, Advances in Complex Systems 11 (04) (2008) 529–549.

[63] P. Erdös, A. Rényi, On random graphs I, Publicationes Mathematicae (Debrecen) 6 (1959) 290–297.

[64] R. Albert, H. Jeong, A.-L. Barabási, Internet: Diameter of the world-wide web, Nature 401 (6749) (1999) 130.

[65] M. Catanzaro, M. Boguná, R. Pastor-Satorras, Generation of uncorrelated random scale-free networks, Physical Review E 71 (2) (2005) 027103.

[66] J.-W. van Prooijen, A. P. Krouwel, Psychological features of extreme political ideologies, Current Directions in Psychological Science 28 (2) (2019) 159–163.

[67] M. Pineda, R. Toral, E. Hernandez-Garcia, Noisy continuous-opinion dynamics, Journal of Statistical Mechanics: Theory and Experiment 2009 (08) (2009) P08001.

[68] A. Carro, R. Toral, M. San Miguel, The role of noise and initial conditions in the asymptotic solution of a bounded confidence, continuous-opinion model, Journal of Statistical Physics 151 (1-2) (2013) 131–149.

[69] Y. Yu, G. Xiao, Influence of random opinion change in complex networks, in: 2015 IEEE International Conference on Digital Signal Processing (DSP), IEEE, 2015, pp. 750–754.

[70] S. P. Borgatti, P. C. Foster, The network paradigm in organizational research: A review and typology, Journal of Management 29 (6) (2003) 991–1013.

[71] M. Everett, S. P. Borgatti, Ego network betweenness, Social Networks 27 (1) (2005) 31–38.

[72] J. Leskovec, J. J. Mcauley, Learning to discover social circles in ego networks, in: Advances in Neural Information Processing Systems, 2012, pp. 539–547.