

Modified Axelrod Model Showing Opinion Convergence And Polarization In Scale-Free Networks

X. Zou^{1,2,*} and H. F. Chau^{1,†}

¹*Department of Physics, The University of Hong Kong, Pokfulam Road, Hong Kong*

²*Department of Physics, University of Toronto, Toronto, ON, Canada,*

(Dated: February 13, 2026)

Axelrod model is an opinion dynamics model such that each agent on a square lattice has a finite number of possible nominal opinions on a finite number of issues that are usually called features in the field. Moreover, its dynamics between two agents is assimilative in the sense that the number of agreeing features between them never decreases upon interaction. Here we modify this model to study opinion convergence, polarization and more importantly to find ways to reduce opinion polarization in an already polarized population. We do so by changing or adding several elements from complex network and continuous opinion dynamics research. First, we put agents in a scale-free network. Second, we adopt the bounded confidence model by representing our agent's opinions by numbers in $[-1, 1]$ those distances follow the standard Euclidean metric. Third, our rules allow both convergence and divergence of their resultant opinions after a pair of agents interacts. As a result, our modified model offers a more comprehensive exploration of opinion dynamics. Computer simulation results of our model show scaling behavior and a notable trend in opinion polarization on all features in the majority of reasonable simulation parameters. To mitigate this polarization, we introduce empathetic agents that work actively to reduce opinion differences. However, our findings indicate limited success in the approach for the most effective way is to change the behavior of a significant portion of highly connected agents. This research contributes to the understanding of opinion dynamics within society and highlights the nuanced complexities that arise when considering factors such as network structure and continuous opinion values. Our results prompt further exploration and open avenues for future investigations into effective methods of reducing opinion polarization.

I. INTRODUCTION

In socio-physics, opinion dynamics of individuals or groups is modeled via interactions among them. By investigating the emergence and convergence of individual opinions into collective opinions, the socio-physical approach aims to shed light on the underlying mechanisms. In particular, sociophysicists wanted to understand the cause as well as the evolution of opinion formation and dissemination through social influence and social networks. They also investigate and analyze the conditions that lead to consensus or polarization, and the factors governing the resulting distributions of opinions [1].

Numerous models of opinion dynamics have been proposed throughout the years. They include the Sznaid model [2, 3], Hegselmann-Krause (H-K) model [4], Galam models [5], Axelrod model [6], Deffuant-Weisbuch model [7, 8] and their variants [9–13]. Our research here focuses on the Axelrod model. In this model, individuals that we call agents from now on are located on a square lattice. The opinion of each agent, which takes on a discrete set of nominal values, evolves by interacting with neighboring agents according to the set of rules to be described in Sec. II below. In particular, Axelrod found that after a sufficiently long time, all agents on the lattice tend to share the same traits for all features provided that the number of features is sufficiently large or

the set of all possible opinions is sufficiently small. Otherwise, multiple regions of different homogeneous opinions on most features is observed in the long run [6]. Castellano *et al.* found that by varying the size of the discrete possible opinion set, Axelrod model shows the classical order-disorder phase transition on culture fragmentation when each agent has more than two features [14]. Besides, Klemm *et al.* showed that in the thermodynamic limit, adding an arbitrarily small noise rate to the Axelrod model leads to opinion fragmentation [15]. Opinions in this model are nominal numbers and its evolution rules on a pair of agents are assimilative in the sense that the number of differing opinions between them does not increase immediately after they interact. Thus, this model fails to capture the increasing divided and polarized political as well as ideological stances in numerous countries and societies in our contemporary world though modeling realistic opinions was not the original intention of the Axelrod model.

Various groups of social scientists have proposed reasons for this fragmentation and polarization in the real world [16–20]. For example, Ashraf and Galor argued that genetic diversity, shaped by the migration of humans out of Africa, is a fundamental determinant of cultural heterogeneity [16]. Dandekar *et al.* tried to explain the polarization by opinion formation [17].

Socio-physical models could offer a complementary approach to understanding the dynamics of opinion fragmentation. Different scholars have tried different simulation models attempting to capture the fragmentation behaviors in society. See, for example, the review by

* xiang.zou@mail.utoronto.ca

† Corresponding author, hfchau@hku.hk

Sîrbu *et al.* [9]. Could the Axelrod model be modified to give fragmented and polarized opinions or to incorporate cultural diversity in the long time limit? This is the first question we are going to address. Surely, various authors have studied this issue over the years with partial success. For example, MacCarron *et al.* considered a more complicated interaction rule and analyzed the clustering behaviors. They showed that by decreasing a certain tunable parameter of the model known as the agreement threshold, agents on the lattice will no longer be homogeneous in the sense that they have more than one set of features in the long run [21]. Radilo-Díaz *et al.* found that for an initial state with small cultural variability, introducing a small amount of rejection dynamics to the Axelrod model is already sufficient to drive the system eventually to a heterogeneous state [22]. Along another line, Battiston *et al.* proposed that the interactions in real society are carried out in multiple layers, which correspond to different interests or topics. By considering this multiplicity in their model, they qualitatively presented new dynamical behavior that leads to the emergence and stability of cultural diversity final state [23]. Törnberg proposed a variation of the Axelrod model in which agents have an additional static opinion denoting, say, their partisan or cultural affiliation. Moreover, they can interact nonlocally. As expected, dynamic opinions of agents with the same additional static opinion converge. Interestingly, dynamic opinions of agents with different additional static opinion can be very different. This so-called sorting behavior appears in spite of the absence of repulsive dynamics [24]. Worth noticing that MacCarron *et al.* [21], Battiston *et al.* [23] as well as Törnberg [24] considered discrete feature traits. In contrast, Amorim [25] modified the Axelrod model by aligning it with Bourdieu's theory on the relation between an individual's taste and the social environment.

This paper introduces a socio-physical model of opinion dynamics that builds upon the Axelrod model, incorporating continuous attitude change among agents within a scale-free network using idea borrowed from the bounded confidence model. First, agents are placed in a more realistic scale-free network of social interaction known as the Barabási-Albert (B-A) network [26]. Second, instead of using a discrete opinion set, we adopt the bounded confidence model [4, 7, 8, 27] by allowing agent's opinions to take on (continuous) values in the closed interval $[-1, 1]$ under the standard Euclidean distance metric. Third, interaction between a pair of agent may lead to either opinion convergence or divergence. As far as we know, this is the first opinion evolution model that incorporates both assimilative and repulsive dynamics into the Axelrod model with a bounded confidence favor.

Certainly, opinion dynamics studies on scale-free network are not new. For example, Bartolozzi *et al.* discovered intermittence in a probabilistic two-state model of opinion evolution on the B-A network [28]. Jacobmeier studied evolution of discrete cardinal opinion evolution for agents in the B-A network using assimila-

tive dynamics [29]. More importantly, Dinkelberg *et al.* found that the long term behavior of a modified Axelrod model with discrete cardinal opinions is insensitive to network topology [30]. Similarly, various groups have studied opinion dynamics using the bounded confidence model [4, 7, 8, 27]. (See, for example, the reviews in Refs. [31] and [32] for more details.) In particular, Defuant *et al.* put forward several agent-based models with continuous opinion change such that the involved agents have limited ranges of acceptance and non-commitment, where only the impact of assimilation is taken into account [7, 8]. Fortunato *et al.* showed that the eventual states of the combination of the H-K model and the bounded confidence model with assimilative interaction only may consist of multiple homogeneous opinion clustering regions [33]. Lorenz investigated the opinion dynamics of agents placed on d -dimensional cubes or simplexes using bounded confidence model with assimilative interaction [34]. Along a different line, Huet and Defuant proposed a bounded confidence model with major and minor feature traits whose interaction can be assimilative or repulsive [35]. Moreover, Chau *et al.* [13] proposed a model for opinion formation which is an extension of the Jager-Amblard (J-A) model [27] under the framework of bounded confidence. Chau *et al.*'s model incorporates assimilation and contrast, which influence agents' opinion acceptance or rejection. More importantly, it shows opinion clustering around extremes and/or moderate. Furthermore, with slow variation of model parameters, Chau *et al.* observed first-order phase transition in the distribution of opinion [13]. Recently, Liu *et al.* researched on the complicated dynamics of the bounded confidence opinion evolution with assimilative and repulsive interactions for agents arranged on a ring [36]. Last but not least, we remark that the works of Jacobmeier [29], Lorenz [34] and Dinkelberg *et al.* [30] mentioned earlier mixed the Axelrod model with the bounded confidence model. But unlike ours, they did not include repulsive dynamics.

Through computer simulation, we show that agents in our model gradually develop extreme opinions. These findings are consistent with observed social behavior and hence give novel insights into the factors influencing collective human opinions. Scaling is also observed in the opinion cluster size distribution whose precise definition will be given later in the text.

Most importantly, we also attempted to address our second question, namely, the polarization problem that is so urgently needed in our currently highly divided and politicized society by modifying our model in this paper. The intuition is that higher openness to ideas from the opposite camp may lead to mutually understanding and increase their willingness to seek common grounds. Therefore, we set a small portion of agents that are empathetic by actively trying to reduce opinion differences. (Note that introducing heterogeneous agents has been done before [8], but setting empathetic agents in this way is new.) Nonetheless, this approach can only

economically reduce opinion polarization of an already polarized population by making a significant portion of those highly connected agents empathetic. This work is an extension of a short project done by XZ [37] under the supervision of HFC, the teacher of the course “Data Analysis And Modeling In Physics”.

II. THE AXELROD MODEL

One of the most frequently used ways of measuring individual opinions in a society is to collect data from attitudinal surveys over a long period of time. It is instructive to model the attitude evolution of individuals and interest groups. One such model was proposed by Axelrod [6]. In the Axelrod model, agents are placed on a $N \times N$ square lattice with closed boundary conditions. An agent is conveniently labeled by its position (i, j) on the lattice. Each agent is characterized by the same fixed set of F features $\mathcal{F} = \{f_1, f_2, \dots, f_F\}$. Each feature takes only values from a fixed set of traits. For simplicity, each $f_i \in \mathcal{F}$ is taken from the same set of q traits labeled $\{0, 1, \dots, q-1\}$. As a result, each agent has F^q possible internal states. For example, one of the features could be the primary language spoken by the agent, and the corresponding set of traits is the set of all contemporary languages. Another possible feature is one’s religious belief. Each feature of an agent is randomly assigned in the beginning. Furthermore, the system evolves for several time steps as follows.

1. In each time step, an agent (i, j) is randomly picked.
2. One of the nearest neighbors of (i, j) is randomly chosen.
3. Denote the number of features that this pair of agents have in common by F_{comm} . And denote the features that the two agents are different by $\mathcal{F}_{\text{diff}}$.
4. With probability $P(\text{interaction}) = F_{\text{comm}}/F$, we randomly pick a feature in $\mathcal{F}_{\text{diff}}$ and the agent (i, j) then copies the trait of that feature from its chosen neighbor. In what follows, we call F_{comm}/F the cultural similarity between two agents.

Under the Axelrod model, agents’ opinions converge provided that F is sufficiently small or q is sufficiently large.

III. DEFICIENCIES OF THE AXELROD MODEL AND OUR REMEDY

The Axelrod model is simple and useful in opinion dynamics. However, it has several deficiencies even though being realistic was not the original goal of the Axelrod model. We now point them out and report ways to modify the model so as to produce an increasingly polarized

opinion distribution that is robust under small changes in tunable parameters.

A. Network Topology

In the Axelrod model, agents form a two-dimensional square lattice. This network topology is artificial. More importantly, it is not compatible with the fact that many social structures can be described by scale-free networks [38]. In simple terms, all but those agents in the lattice boundary in the Axelrod model have equal strength in influencing others or being influenced by others. In contrast, scale-free network topology can be used to model both social and parasocial interactions in the real world, including interactions between social media influencers and their followers. Thus, we propose putting agents in an (undirected) scale-free network in which only neighboring agents are allowed to interact directly. For simplicity, we generate such a network using the B-A model [26] as it is one of the most studied and used scale-free networks. Besides, it has the additional advantage that it can be efficiently generated. In this study, we simplify matters by considering non-evolving networks. It is instructive to consider evolving networks in future studies.

B. Feature Traits

In the Axelrod model, every agent is described by a set of features, and each feature can have several possible discrete nominal traits [6]. While there are features that are naturally described by discrete nominal traits, the Axelrod model is incomplete as traits of a feature need not be discrete or nominal. For example, one of the features could be the agent’s fluency in a particular spoken language. Here, the possible traits are more naturally described by a continuous real number than discrete nominal states. More importantly, using continuous traits is a natural platform to model the possibly increasingly polarized opinion distribution. We, therefore, modify the model by replacing the discrete set of traits with a real number representing the strength of the agent’s feature. We also call this type of continuous feature traits opinions of the agent.

How could we take care of the increasingly polarized opinion distribution? A simple and effective way to address this issue is to borrow the idea of the bounded confidence model [4, 7, 8, 27] by representing opinions using a real number in a bounded interval under the Euclidean metric. For simplicity, we set the possible opinions to be a real number in the interval $[-1, 1]$. Besides, we denote the distance between two opinions values p and q by

$$d(p, q) = |p - q|. \quad (1)$$

As for the initial values of each trait, instead of drawing from a uniform distribution in $[-1, 1]$, we use a more

realistic truncated normal distribution with respect to the Euclidean distance metric. Its detailed construction can be found in upcoming subsection.

C. Interaction Rules Among Agents

Recall that agents change their traits according to rule 4 in the Axelrod model. This rule is assimilative and hence can only increase the cultural similarity between the chosen pair of agents. In other words, rule 4 is too cohesive. Besides, it is not consistent with the observation that cultural diversity among groups of people appears even in apparently culturally homogeneous societies [16–19].

We may also need to take repulsive dynamics into consideration. The existence of repulsive dynamics and its psychological origin (if it exists) are current research topics. One possibility is that the effect is real. Supporters of this camp believe that social interaction where attempts to persuade or restrict someone’s freedom of opinion could result in the adoption of an opposing view. This effect occurs when persuasive communication happens in an unintended position, which often strengthens the original attitude an individual holds that it sought to alter. Hovland *et al.* first identified and named it the boomerang effect [39]. The boomerang effect has been observed in various contexts [40, 41], including consumer behavior, psychological interventions, health policies, and, relevant to our discussion here, political communication. In political contexts, attempts to correct misinformation through mock news articles sometimes lead to increased belief in inaccurate information among certain groups [42]. Nevertheless, we need to be cautious here as the evidence on the existence of the boomerang effect is controversial and mixed. Furthermore, this effect may not be systematic. For example, Schultz *et al.* observed it in a field experiment in promoting household energy conservation through the use of different normative messages. Moreover, boomerang effect was eliminated by adding an injunctive message [43]. Similar conclusions were drawn in the follow up study by Verkooijen *et al.* [44]. In addition, Liu *et al.* found that overinvestment in corporate social responsibility had a boomerang effect on the shareholder value of a company after announcing a recall. The effect is more evident when institutional ownership is low or when customer awareness is high [45]. On the other hand, Takács *et al.* failed to observe it in their recent study using the University of Groningen students as participants [46]. Bail *et al.* discovered that exposing Democrats (Republicans) with conservative (liberal) social media messages make their views more liberal (conservative). However, the change is statistically insignificant [47]. Another possibility is that boomerang effect is neither real nor systematic. At least, agents tend to be less affected by ideas from the opposite camp [24]. If it is so, what is the underlying mechanism for opinion polarization? Unfortunately, no

definitive answer has been identified.

Here we move on by adopting an empirical approach. That is to say, we model the repulsive dynamics as if it were coming from the boomerang effect without considering its psychological origin. Our goal is to reproduce the polarization seen in the real world and study if there is anything we can do to reduce it. Specifically, to study and model the positive persuasion as well as the repulsion effects in social interactions, we introduce a parameter $a \geq 0$ known as the agreement threshold. It is used to determine whether a social interaction results in opinion convergence or divergence. In each time step, we randomly pick an agent from the scale-free network plus one of its neighbors. We count the number of features in which the Euclidean distance between the opinions of the two agents is smaller than the agreement threshold a . If this number is greater than or equal to, say, half of the total number of features, the two agents are said to be “similar”. Otherwise, they are “dissimilar”. We hypothesize that upon interaction, opinions of similar agents are likely to converge whereas those of dissimilar ones are likely to diverge. Besides, an agent’s persuasion power is proportional to the number of neighbors it has. This models the asymmetric influence between agents of different connectivities. This hypothesis is consistent with a number of findings in the literature on the persuasion power of social media influencers. They include the meta-analysis by Pan *et al.* in which they found that the characteristics of a social media influencer, including the number of followers, is positively correlated with the purchasing behavior of his/her followers [48]. The report by Conde and Casais showed the positive correlation of the influencing power with the perceived popularity and opinion leadership from a sample of 140 Portuguese social media influencers [49]. The work by Peter and Muth who discovered that social media influencers were increasingly affecting German youth on political issues. The effects ranged from amplifying the effects of existing opinions to opinion formation as well as to change voting intentions [50]. Last but not least, the work by Liu and Zheng who argued that the influence between social media influencers and their followers is mutual when evaluating brand credibility and purchase intention. They further found that the influence could be repulsive sometimes [51]. Nevertheless, we have to be cautious that little is known on how the followers directly affecting the positions of social media influencers. It is instructive and timely to investigate the validity of the hypothesis that an agent’s persuasion power is proportional to its number of neighbors in the field experiments.

After the above preparation and discussion, we now state the details of our model below.

1. [Scale-free network generation] A B-A network with N vertices is generated with each newly added vertex connecting to m existing nodes using their standard preferential attachment mechanism [26]. Each vertex of the network represents an agent and only

those agents joined by an edge in the network can communicate directly with each other.

2. [Opinion initialization] Each agent is characterized by the same finite set of F features $\mathcal{F} = \{f_1, \dots, f_F\}$. The opinion of agent X on feature $f \in \mathcal{F}$, denoted by X_f is randomly and independently drawn from $\mathcal{N}_t(\bar{x}, \sigma) = \mathcal{N}_t(0, 0.25)$, namely, the truncated normal distribution with mean 0 and standard deviation 0.25 in the Euclidean distance metric. More precisely, if r is the value of the random variable drawn from $\mathcal{N}_t(0, 0.25)$, then the opinion is set to r if $|r| \leq 1$. And the opinion is set to 1 (-1) if $r > 1$ ($r < -1$).
3. [Opinion evolution] An agent X is randomly picked. Then, a neighbor of X , say Y , is randomly selected. A feature f is randomly selected from \mathcal{F} .
 - (a) [The case of similar agents] If the two agents are similar, namely, the Euclidean distances between at least $\kappa = 1/2$ of their opinions are less than the agreement threshold a , then their opinions of feature f move closer to each other in the sense that the Euclidean distance of the new opinions of the selected feature trait between the two agents decreases upon interaction. Specifically, X_f , the opinion of feature f for agent X , changes to \tilde{X}_f where

$$\tilde{X}_f = X_f - \frac{\mu c_Y |X_f - Y_f|}{c_X + c_Y}. \quad (2a)$$

Here, c_X and c_Y are the degrees of nodes X and Y , respectively. In addition, $\mu > 0$ is a parameter modeling the effectiveness of persuasion. Note that we borrow the preferential attachment idea in the B-A network [26] in this rule so as to model the effective persuasion power of social media influencers. Likewise, Y_f changes to \tilde{Y}_f where

$$\tilde{Y}_f = Y_f - \frac{\mu c_X |Y_f - X_f|}{c_X + c_Y} \quad (2b)$$

- (b) [The case of dissimilar agents] If the two agents are dissimilar (that is, they are not similar), then we randomly pick a feature f among those with $d(X_f, Y_f) \geq a$. The values of X_f and Y_f change to \tilde{X}_f and \tilde{Y}_f according to

$$\tilde{X}_f = \text{Trunc} \left[X_f + \frac{\mu c_Y |X_f - Y_f|}{c_X + c_Y} \right] \quad (3a)$$

and

$$\tilde{Y}_f = \text{Trunc} \left[Y_f + \frac{\mu c_X |Y_f - X_f|}{c_X + c_Y} \right], \quad (3b)$$

respectively. Here

$$\text{Trunc}(p) = \begin{cases} p & \text{if } |p| \leq 1, \\ 1 & \text{if } p > 1, \\ -1 & \text{if } p < -1. \end{cases} \quad (4)$$

Clearly, the Euclidean distance of the new opinions of the selected feature trait between the two agents increases or remains unchanged upon interaction.

The opinion evolution in Step 3 is repeated $Nt/2$ times, where t is the average number of interactions an agent has with its neighbors.

Note that we use the probability distribution $\mathcal{N}_t(0, 0.5)$ to initialize opinions to avoid opinion bias as well as to give an initial opinion diversity in our initial ensemble. In addition, our evolution rules do not change the nature of the agent pairs picked in Step 3 in the sense that they are similar (dissimilar) before and after the interaction. Nevertheless, the nature of each of the picked agents with its other neighbors may change after the interaction. Last but not least, we stress that for all the subsequent simulations that study the eventual system behavior, the average number of interactions an agent has with its neighbors t is chosen to be large enough for the system to approach equilibration.

IV. SIMULATION RESULTS AND EFFECTS OF DIFFERENT MODEL PARAMETERS

A. Methodology

We study the dynamics of our modified Axelrod model using Python [52]. The scale-free network is created using the Networkx package and most of the analytical work is done using the NumPy package. Figures are produced by the Matplotlib package.

We investigate the effect of each model parameter on the development of the system by varying one of them and fixing all the rest to their corresponding default values as tabulated in Table I. Each of the simulation results presented below is taken from a single run randomly selected over a sample of 10 runs. And a new scale-free network is used for each run. In fact, all runs give very similar distributions. We report our simulation results in histograms on the opinion distribution of feature traits. Since our initialization and evolution rules apply to different feature traits in the same way, opinion distributions of different feature traits averaged over a sufficiently large ensemble of independent runs should be the same. This is indeed what we have observed in our simulation. Consequently, to save running time, we lump the distributions of all F feature traits of every agent in plotting the histogram. In each histogram, we divide the opinion into 50 evenly spaced bins (in the usual Euclidean distance) each with a bin width of 0.04.

Parameter	Description	Default value
N	Number of agents	1000
m	Number of edges attached from new node to existing nodes while constructing the initial B-A network	4
F	Number of features	11
t	Mean number of interactions per agent	200
$\mathcal{N}_t(\bar{x}, \sigma)$	Initial probability distribution of opinions	$\mathcal{N}_t(0, 0.25)$
a	Agreement threshold that determines the similarity of agents	0.25
κ	Fraction of feature traits used to determine the similarity of agents	0.5
μ	Persuasion effectiveness parameter	0.25

TABLE I. Default values of different parameters used in our simulation.

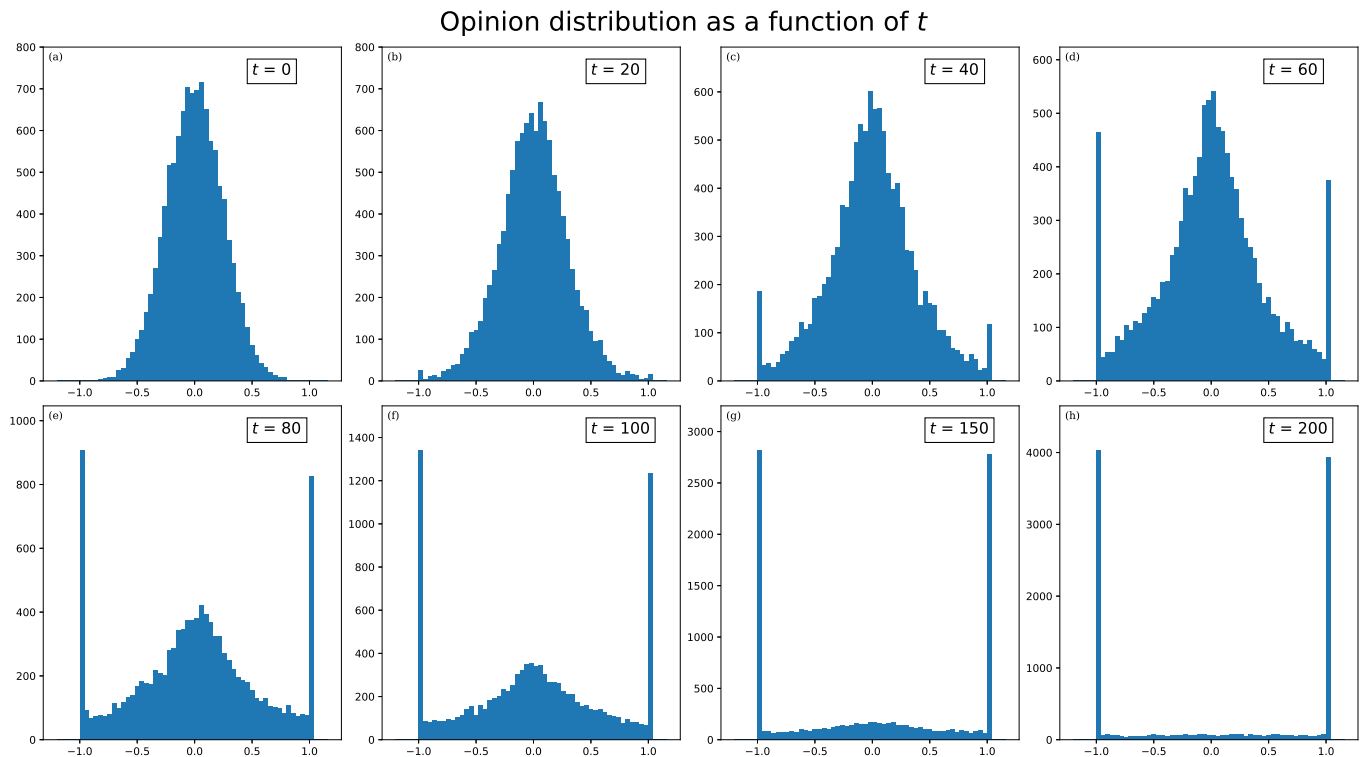


FIG. 1. Histograms showing the time evolution of opinion distribution with all but the parameter t taken from Table I. The x -axes of all subplots are the opinions in $[-1, 1]$ and the y -axes are the occurrence frequencies of the opinion. Although the initial opinions concentrate around 0, the system evolves to a state of extreme opinion polarization in the long time limit in which nearly all agents have extreme opinions on all features. Note that the y -axis ranges of all histogram subplots in this paper may be different.

B. Feature Evolution

We begin by investigating the time evolution of opinions by fixing all other model parameters to the ones listed in Table I. Fig. 1 shows the time evolution of agents' opinions. Clearly, opinion spreads as time progresses. When the mean number of interactions per agent t is less than ≈ 20 , the opinion distribution roughly follows the Gaussian distribution whose mean is zero and the standard deviation increases gradually. When $t \gtrsim 20$, extreme opinions start to emerge. In addition, the frac-

tional increase of extreme opinions accelerates. Eventually, extreme opinions dominate, and nearly all agents have extreme opinions on all features. That is to say, opinions become extremely polarized.

Our simulation result can be qualitatively understood as follows. Since the agreement threshold a is less than the standard deviation of the initial opinion σ , a sizable fraction of neighboring agents is dissimilar. Their interactions gradually lead to opinion polarization. Once a non-negligible portion of the opinion is polarized, our evolution rules accelerate further polarization for it is harder and harder to find a similar agent pair which

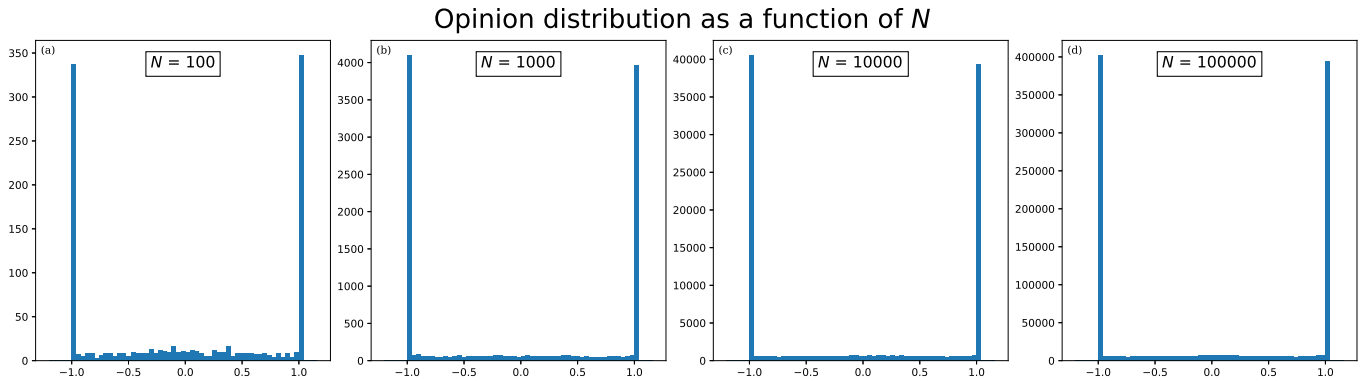


FIG. 2. Opinion distributions after $Nt/2$ time steps with $t = 200$ for different values of N . These subplots show no significant difference in opinion distribution for different N .

consists of an agent with moderate opinion and another agent with an extreme opinion on the same feature. As a consequence, it is no longer possible to pull enough extreme opinion agents back to moderate ones. Thus, an extremely polarized society is formed in the end. Readers will find in our subsequent analysis that the formation of extremely polarized opinion is a genuine and common feature of our model using reasonable parameters rather than the result of a cherry-picked set of parameters. (Surely, opinions are never polarized if $\sigma \approx 0$ or $\kappa \approx 0$. However, these sets of parameters are atypical and unrealistic.)

Let us compare our opinion distribution results with several models on market. The three closest models on market are the ones reported by Chau *et al.* [13], Jager and Amblard [27] as well as Liu *et al.* [36]. They used Euclidean metric bounded confidence with assimilation as well as contrast dynamics. Agents are placed on a complete graph in Refs. [13, 27], regular 2-d lattice in Ref. [27] and a ring in Ref. [36]. Using suitable parameters, the eventual agent opinion distributions in these two models exhibit bipolar extreme polarization very similar to ours.

C. Effect Of The Number Of Agents

We now study the effect of the number of agents N while fixing other parameters to their default values listed in Table I. As shown in Fig. 2, there is no significant difference in the distribution between different histograms even if N varies over four orders of magnitude from 10^2 to 10^5 . This result is not surprising because, upon coarse-graining, the resultant network is still scale-free. And since the formation of extreme opinion polarization is the generic behavior of the system, the effective parameters of the renormalized system likely fall within this generic region. Finally, we find that the convergence rate of the opinion polarization is almost independent of N . That is to say, for a given mean number of interactions per agent t , the distribution between different histograms is

statistically the same regardless of N .

D. Distributions Of Opinion Clusters And Their Sizes

We turn to study the opinion cluster distribution and its relation with N where an opinion cluster is defined to be a maximal (in the sense of set inclusion) path-connected set of agents such that any pair of adjacent agents (with respect to the underlying network topology) in the set is similar. Fig. 3(a) shows the distribution of opinion clusters for a typical run using parameters in Table I at $t = 200$ with N replaced by 400 to avoid over-crowding the graph. We find that in about 2/3 of the runs, one dominant opinion cluster consisting of about 50% to 60% of the population shows up. (One example is Fig. 3(a) whose dominant opinion cluster is made up of the pale yellow dots in the central region of the figure.) Whereas for the remaining 1/3 of the runs, one large and one much smaller dominant opinion clusters appear along side with many small clusters. Besides, Fig. 4(a) tells us that as N increases, both the chance of having more than one dominant opinion cluster and the standard deviation of the size of the dominant opinion cluster decrease. For instance, when $N = 10^5$, the distribution of the size of the dominant opinion cluster sharply peaks in the interval $[0.54N, 0.56N]$ up to 3σ level. Furthermore, for all the number of agents N used in our simulation, every remaining cluster contains much fewer number of agents. In other words, there is no intermediate sized opinion cluster in the system. To summarize, a highly fragmented society with a significant number of highly connected agents holding similar extreme opinions is formed. More importantly, numerous small clusters many of them holding opposite extreme opinions are also present. This polarization pattern resembles those observed in the real world such as the great opinion differences between those living in cities and in rural areas as well as the emergence of political extremism due to distrust of the elites. It is instructive to explore the po-

Opinion clusters in the network

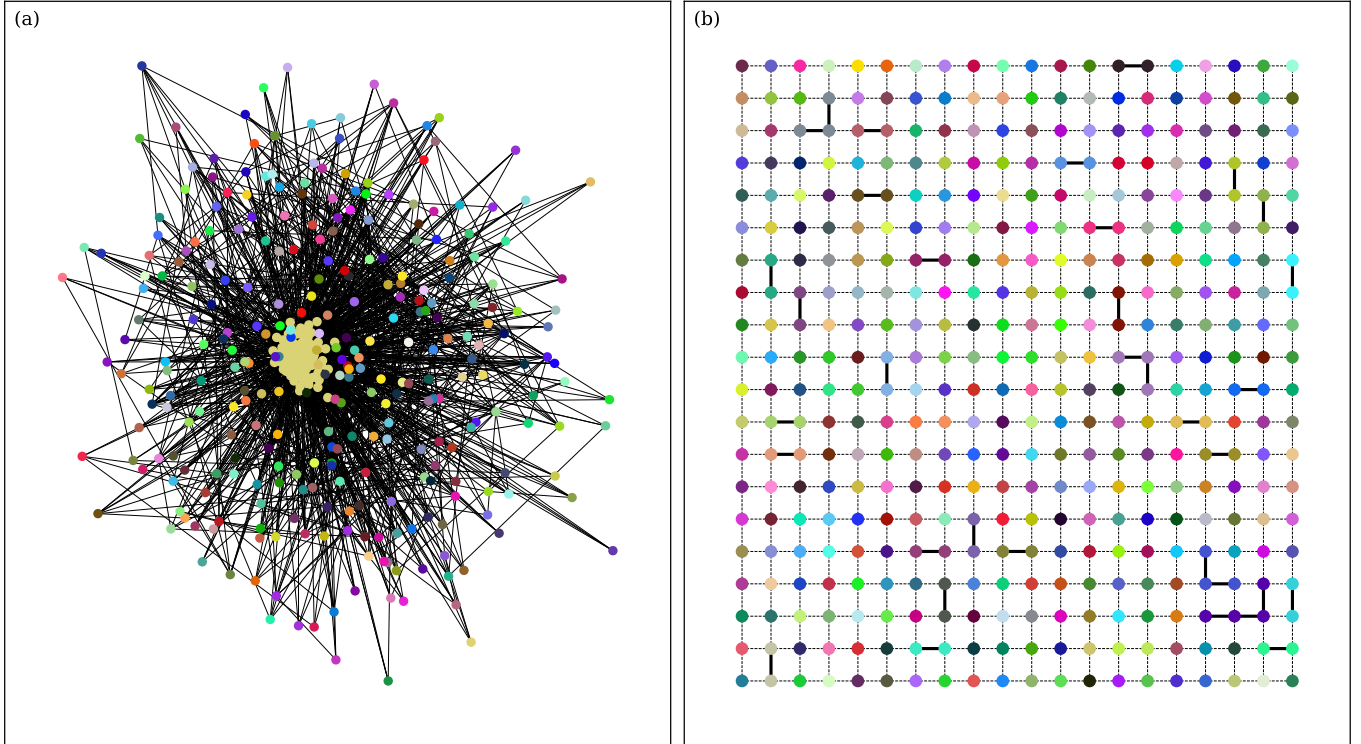


FIG. 3. Opinion cluster distributions of a typical run of our model on (a) a B-A network and (b) 20×20 lattice using parameters in Table I except that $N = 400$. Agents in the same opinion cluster are represented by dots of the same color; and no adjacent cluster is of the same color. Each adjacent agent pair in (a) is linked by a solid black line. In (b), a thick solid black (thin dashed black) line is drawn between a pair of adjacent similar (dissimilar) agents.

tential of applying our model to study this type of society as the general trend reported above is generic over typical sets of parameters. In contrast, Fig. 3(b) shows the opinion clusters of a typical run of our model when agents are placed in a 20×20 lattice using the same parameters as in Fig. 3(a). Unlike the case of using B-A network, here we see a much greater number of similarly sized opinion clusters. Besides, no dominant cluster is present. (For instance, the largest opinion cluster in Fig. 3(b) consists of 4 agents represented by purple dots near the bottom right corner of the figure.)

We further investigate the opinion cluster size distribution. Let us denote the average number of opinion cluster of size s in a system of N agents by $n(s, N)$. Figs. 4(a) and (b) show the log-log and semi-log plots of $n(s, N)/N$. The data points are obtained for $N = 10^3, 10^4$ and 10^5 each averaged over at least $S = 1024$ runs using parameters listed in Table I except that N can vary. Scaling is observed in $n(s, N)/N$ as curves for different values of N collapse. Moreover,

$$\frac{n(s, N)}{N} \approx \begin{cases} As^{-\tau} & \text{for } s \lesssim 10, \\ Be^{-\lambda s} & \text{for } 15 \lesssim s \lesssim 35, \end{cases} \quad (5)$$

with fitting parameters A, τ, B and λ tabulated in Table II. Figs. 4(a) and (b) depict that opinion cluster

Parameter	B-A network	square lattice
$\log_{10} A$	-0.51 ± 0.02	-0.02 ± 0.06
τ	3.2 ± 0.2	5.9 ± 0.4
$\log_{10} B$	-2.4 ± 0.1	-2.9 ± 0.1
λ	0.16 ± 0.01	0.19 ± 0.01

TABLE II. Best fit parameters for the subplots in Fig. 4.

distribution follows a power law over roughly one decade for $s \lesssim 10$ and an exponential decay law over roughly two decades for $15 \lesssim s \lesssim 35$. The power law distribution for small cluster size can be understood as follows. Since $n(s, N)$ obeys the sum rule $\sum_s s n(s, N) = N$ and agents are placed on a scale-free network, we expect that $n(s, N)/N \sim s^{-\tau}$ for sufficiently small s .

Next, we explore the effect of the underlying network topology on opinion cluster size distribution. Figs. 4(c) and (d) show the log-log and semi-log plots of n/N as a function of s for different N when agents are placed on a square lattice. Just like the B-A network case, scaling is observed as the curves collapse for any N . But unlike the B-A network situation, power law is observed in a much smaller region of $s \lesssim 4$ — a region too small to be of statistical significance. Furthermore, exponential

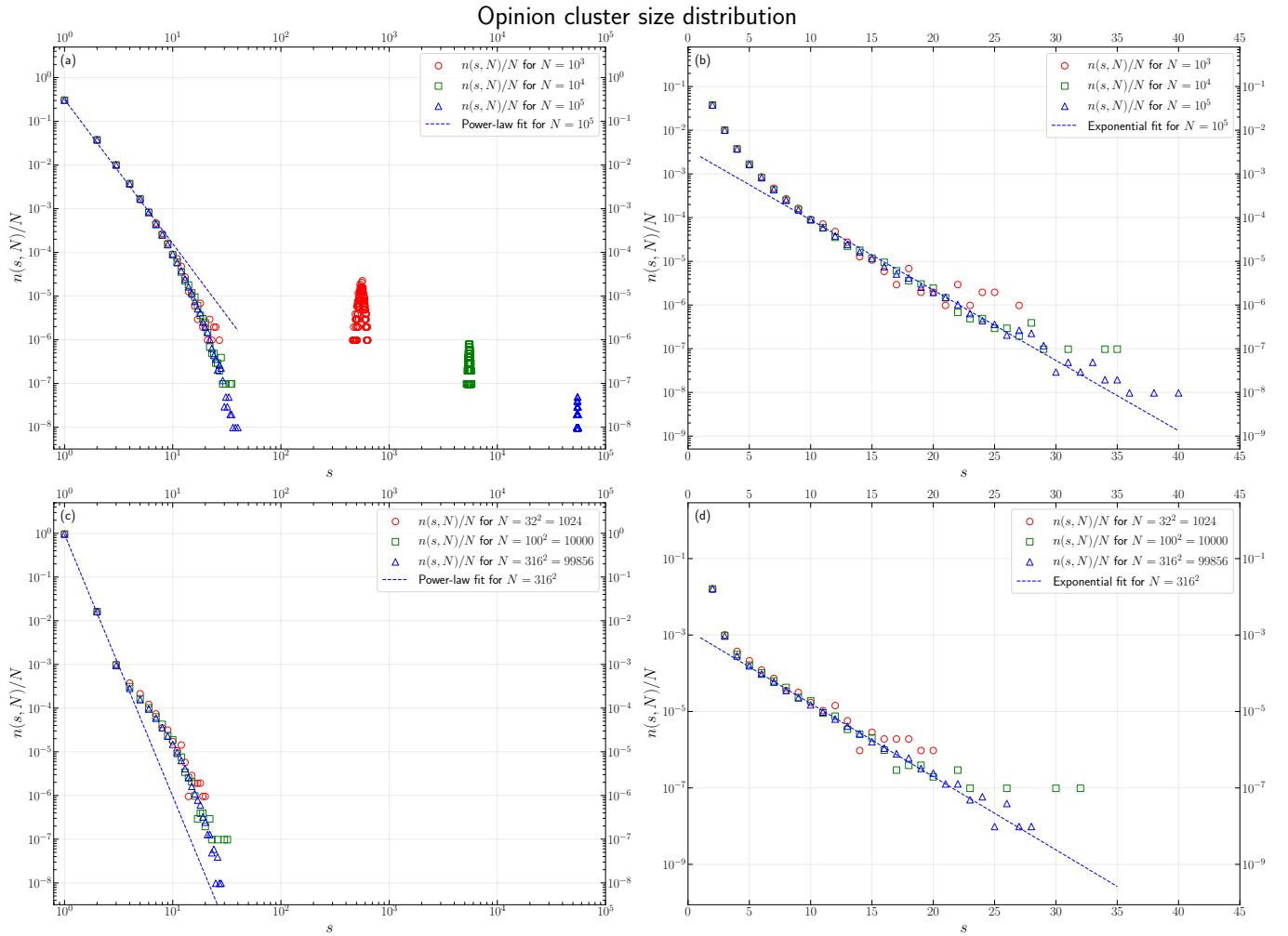


FIG. 4. The average number of opinion cluster divided by the number of agents n/N as a function of cluster size s and number of agents N on a B-A network in (a) log-log and (b) semi-log plots as well as on a square lattice in (c) log-log and (d) semi-log plots using parameters in Table I except that N can vary. The red dots, green squares and blue triangles are for $N = 10^3$, 10^4 and 10^5 , respectively for the B-A network and the corresponding closest square number for the square lattice. Each set of data is obtained by averaging over 1024 independent runs. The uncertainty of $\log_{10}[n(s, N)/N]$ of each data point is of order of $(NS)^{-1/2}/n$ and the corresponding error bar is omitted in the subfigures for clarity. The dashed line in each subfigure is the best fit curve whose parameters are shown in Table II.

decay is found for $5 \lesssim s \lesssim 25$. The corresponding best fit parameters are listed in Table II. Therefore, it appears that the use of scale-free network enlarges the power law regime of opinion cluster distribution.

We now compare our opinion cluster size distribution results with a few existing models in the literature. Opinion cluster size distribution of our model in B-A network is similar to that by Jacobmeier [29] that also used B-A network but only with assimilative dynamics. Both shows power law distribution but with very different exponents over very different regions — their exponent is close to 0 and is valid for intermediate values of s over a range of about 1.5 decades while our exponent is about 3.4 and is valid for $s \lesssim 10$. In contrast, their model shows clustering in opinions. Furthermore, our model using B-A network shares similar opinion cluster size distribution

with the model by Deffuant *et al.* [7] that used assimilative dynamics only in a regular 2-d lattice. However, their distribution exhibits power law but with no obvious scaling. Our 2-d square lattice result is also similar to the stable configuration of the polarization phase of the J-A model [27]. Besides, ours exhibits scaling but without any power law. In addition, the behavior of our model is different from that in Dinkelberg *et al.* [30] as the cluster size distribution of the later is weakly dependent on the underlying network used.

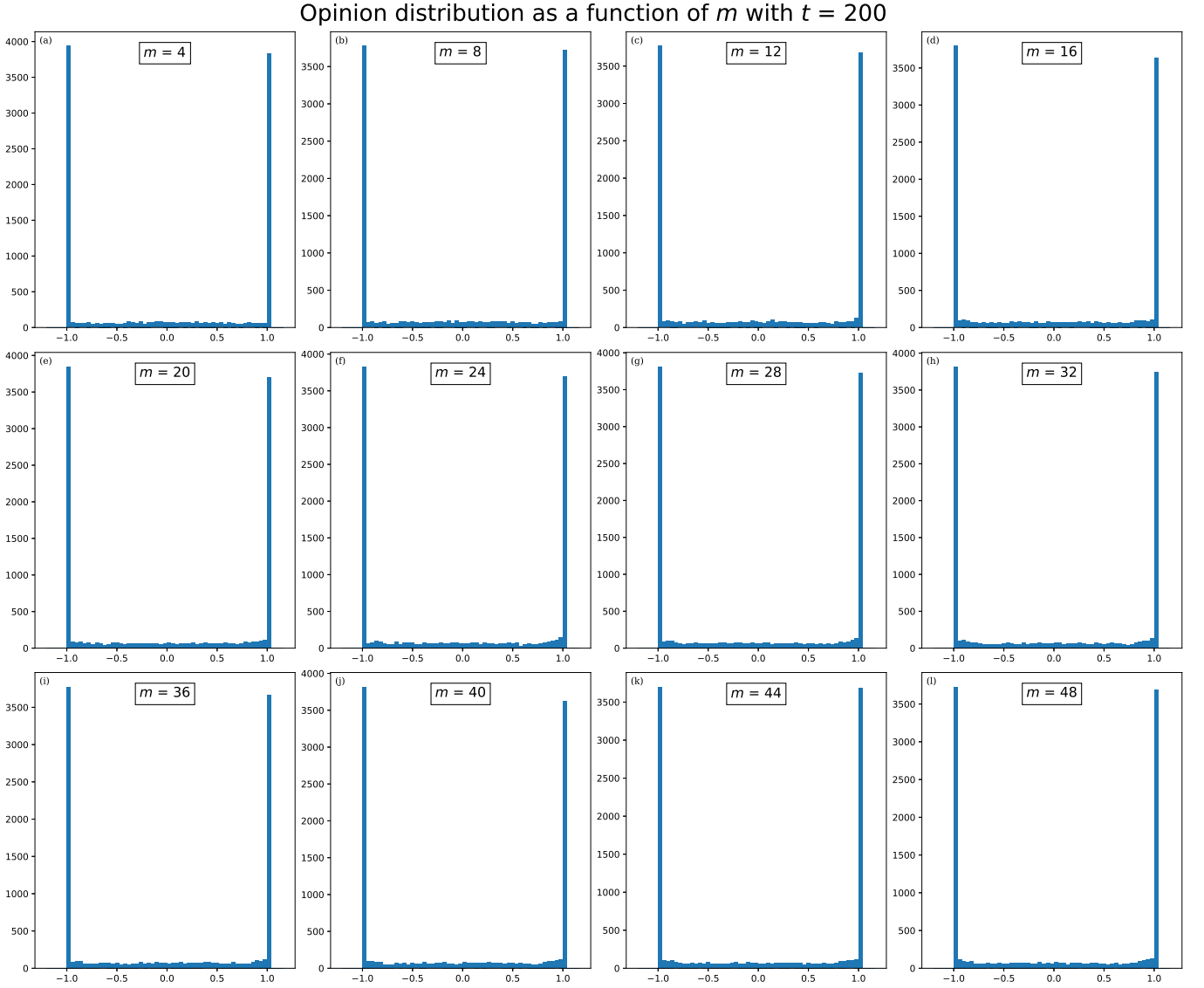


FIG. 5. Opinion distribution after $Nt/2$ time steps with $N = 1000$ and $t = 200$ for different values of m . These subplots show no significant difference in opinion distribution for different m .

E. Effect Of The Average Degrees Of Agents In The Scale-free Network

Here we investigate the effect of the average degree of the scale-free network m on the system evolution by varying it from 4 to 48. As shown in Fig. 5, there is no significant difference in the final steady state of the opinion distribution among all agents. They all evolve to a highly polarized opinion state.

To study the rate of convergence to this highly polarized state, we reduce the average number of interactions per agent t from 200 to 100. Fig. 6 depicts that the equilibration time to the extremely depolarized state increases as m increases. This result can be understood as follows. An agent has more neighbors on average as m increases. Consequently, the opinion of an agent is influenced by

more neighbors each has statistically independent opinions on each feature trait initially. Thus, it takes a longer time for opinions of the system to polarize.

F. Effect Of The Agreement Threshold

We now study the effect of agreement threshold a on the system. When $a \lesssim \sigma$, as shown in subplots (a)–(e) in Fig. 7, the system equilibrates to two dominating peaks at a polarized opinion as in the generic case. This is because the initial opinion distribution is sufficiently diverse to allow a sizable portion of neighboring dissimilar agents. Their interactions seed more extreme opinions. In contrast, when $a \gtrsim \sigma$, the system evolves to a single peak of moderate opinion and the extreme opinions

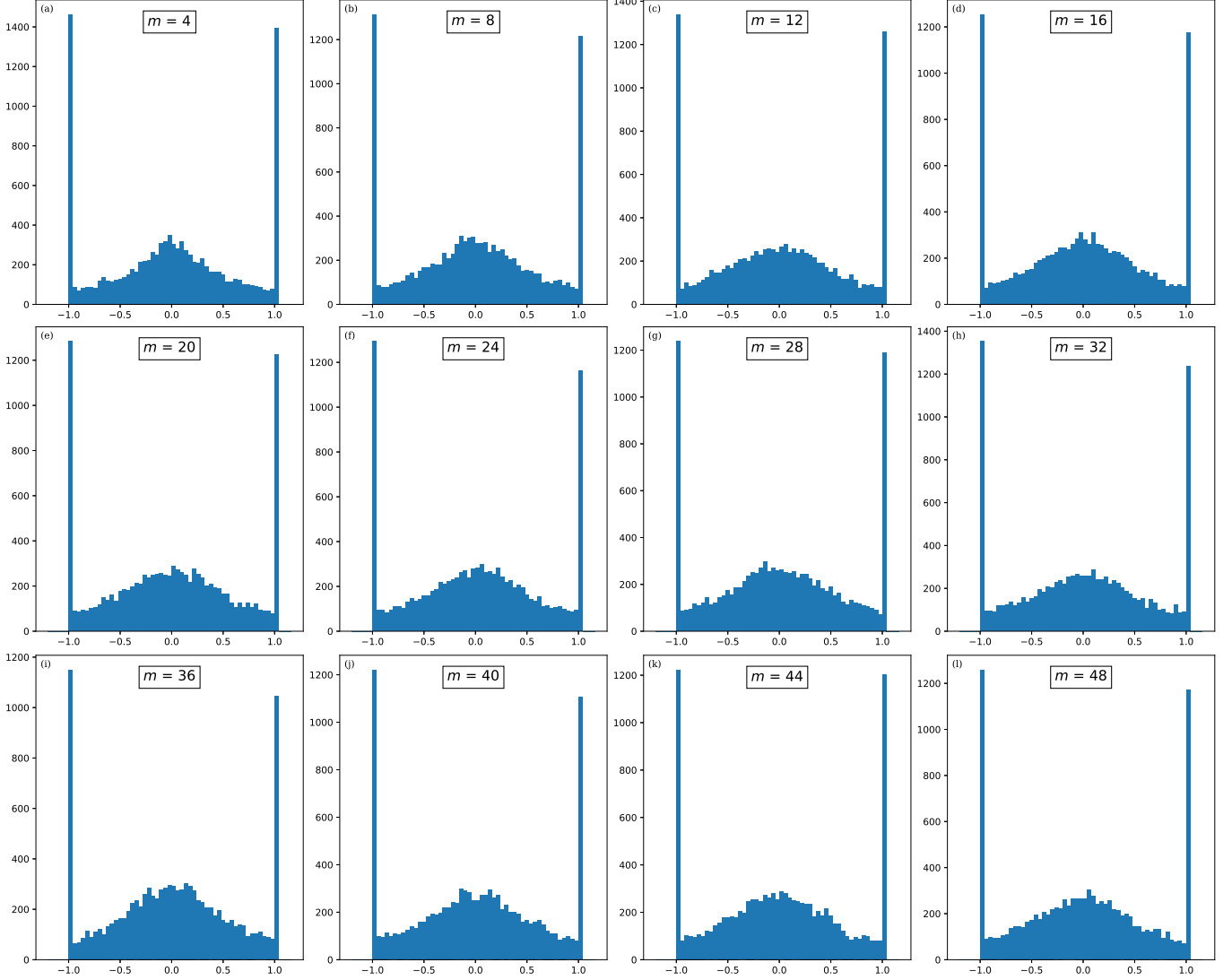
Opinion distribution as a function of m with $t = 100$ 

FIG. 6. Opinion distribution after $Nt/2$ time steps with $N = 1000$ and $t = 100$ for different values of m . Compared with Fig. 5, these subplots show that longer equilibration times are needed to reach the extremely opinion depolarization state as m increases.

population is very small. (See subplots (h)–(l) in Fig. 7.) The reason is that with a large value of a , almost all agent pairs are similar. This leads to opinion convergence for at least a majority portion of the population. For $a \approx \sigma$, as depicted in subplots (f) and (g) in Fig. 7, the coexistence of a non-negligible number of extreme and moderate opinions is observed.

In summary, we find that the value of a/σ is the most important factor determining the asymptotic behavior of the system. It is natural to expect that $a < \sigma$ in human societies for otherwise the repulsive dynamics can hardly be observed. The implication of our modified Axelrod model is that societies will evolve toward extreme opinion polarization — a phenomenon in our contemporary society that leads to all kinds of troubles making most people unhappy and some to go violence.

G. Effects Of Other Parameters

Last but not least, we consider the effects of varying each of the following parameters — F , κ and μ . We find that varying κ alters the final opinion distribution while varying the others do not. This is consistent with the analysis above. The reason is that by varying κ , we increase the fraction of similar agents and hence slow down the emergence of polarized opinion. Therefore, the simulation result is very similar to the result shown in Fig. 7. We do not show the corresponding histograms here to save space.

Surely, we may further consider the effects of \bar{x} and σ . However, we are not going to do so here. Setting a non-zero \bar{x} introduces initial bias on opinion distribution

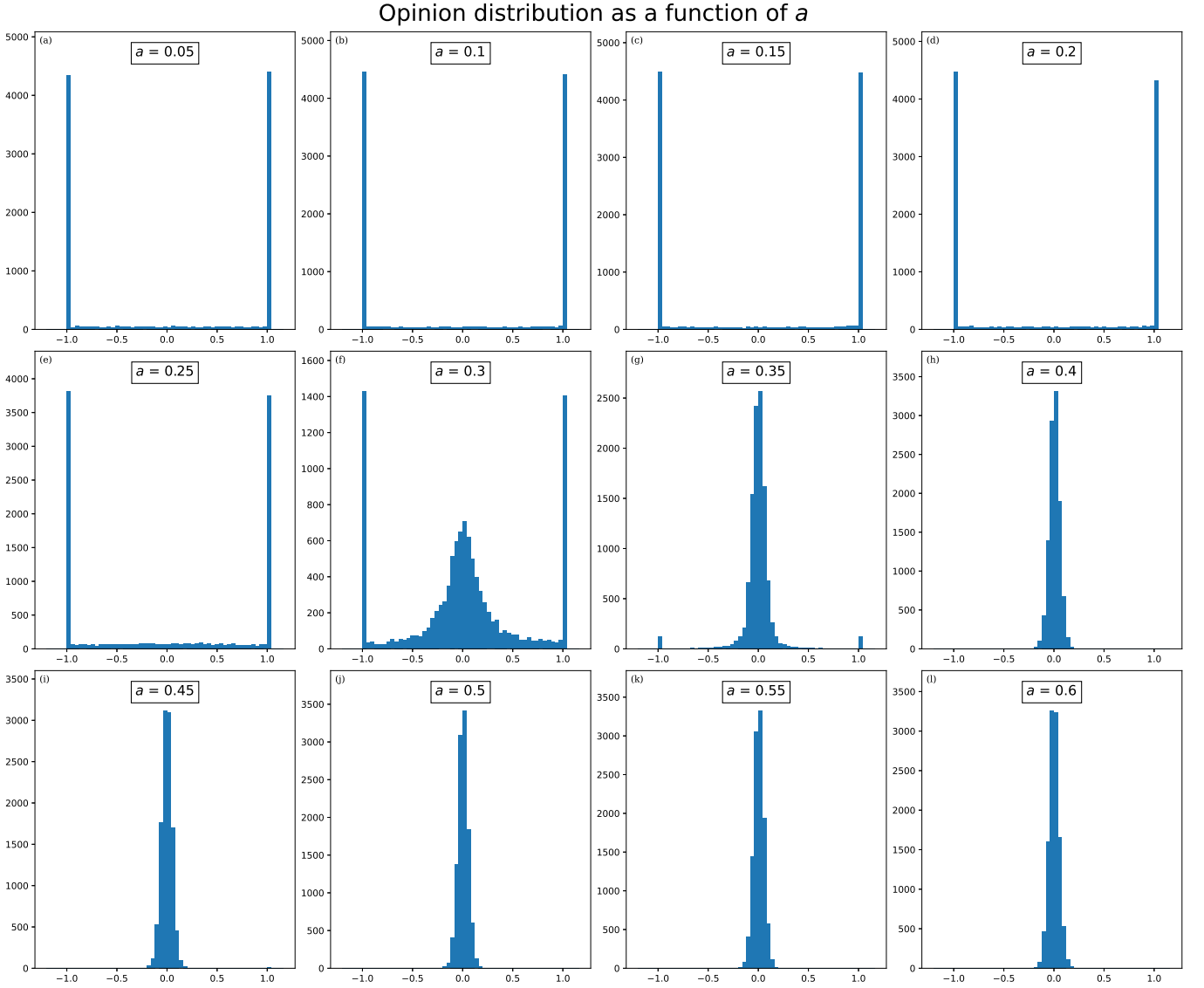


FIG. 7. Opinion distribution after $Nt/2$ time steps for $N = 1000$ and $t = 200$ for different values of a . It shows that opinion polarization occurs only when $a \lesssim \sigma$.

that may not be easy to justify. Whereas if $\sigma \ll 1$, by rescaling the X_f , it is clear that the ratio a/σ is a determining factor of the system dynamics. As we have already investigated the effect of a on the dynamics, this makes the study of σ redundant.

V. REDUCING OPINION POLARIZATION BY INTRODUCING EMPATHETIC AGENTS

Could we make the opinion distribution of the eventual states less polarized within the framework of our modified Axelrod model? We would like to investigate possible interventions that could alleviate extreme opinion polarization. As a general principle, we would like to make our interventions as small as possible, hopefully in

the form of making small changes in our rules. This is because changing human psychological behavior *en masse* is not easy, at least in the short term. Only these kinds of small changes have a real chance to occur in practice. Here we propose to alter the behavior of a small fraction of agents. Specifically, we fix a small portion p_s of the agents randomly chosen among those whose number of neighbors in the B-A network rank top $100\rho_s\%$ to be empathetic. A empathetic agent always interacts with its neighbors using the rules of a similar agent. One way to think of it is that each agent has a different agreement threshold, and that of a empathetic agent is arbitrarily large. Their readiness to be influenced and persuaded by all kinds of opinions is the key to reaching a consensus in the population.

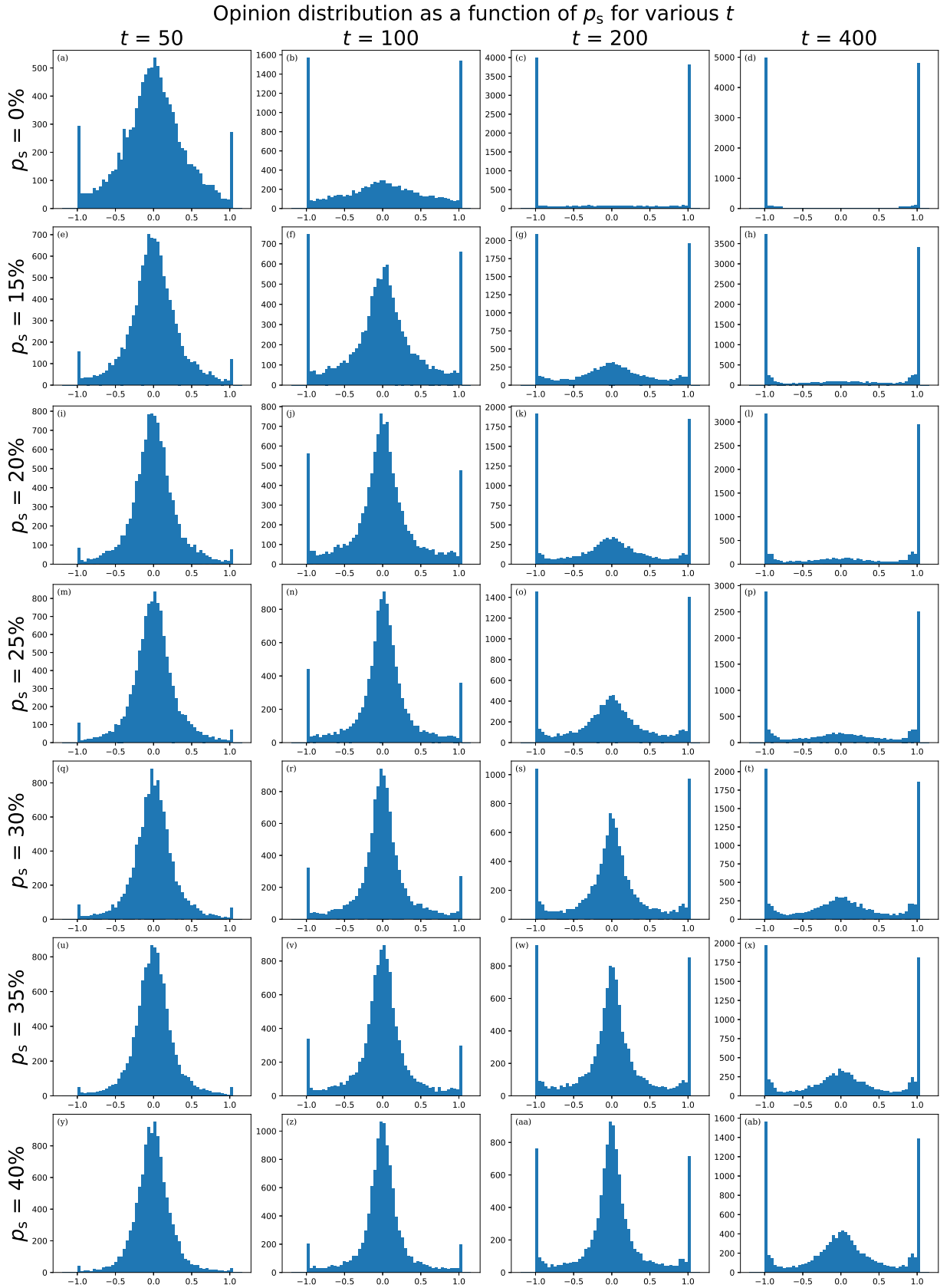


FIG. 8. Evolution of opinion distribution for different portion of empathetic agents p_s and mean number of interactions per agent t when $\rho_s = 1$. All other parameters are set to default values listed in Table I.

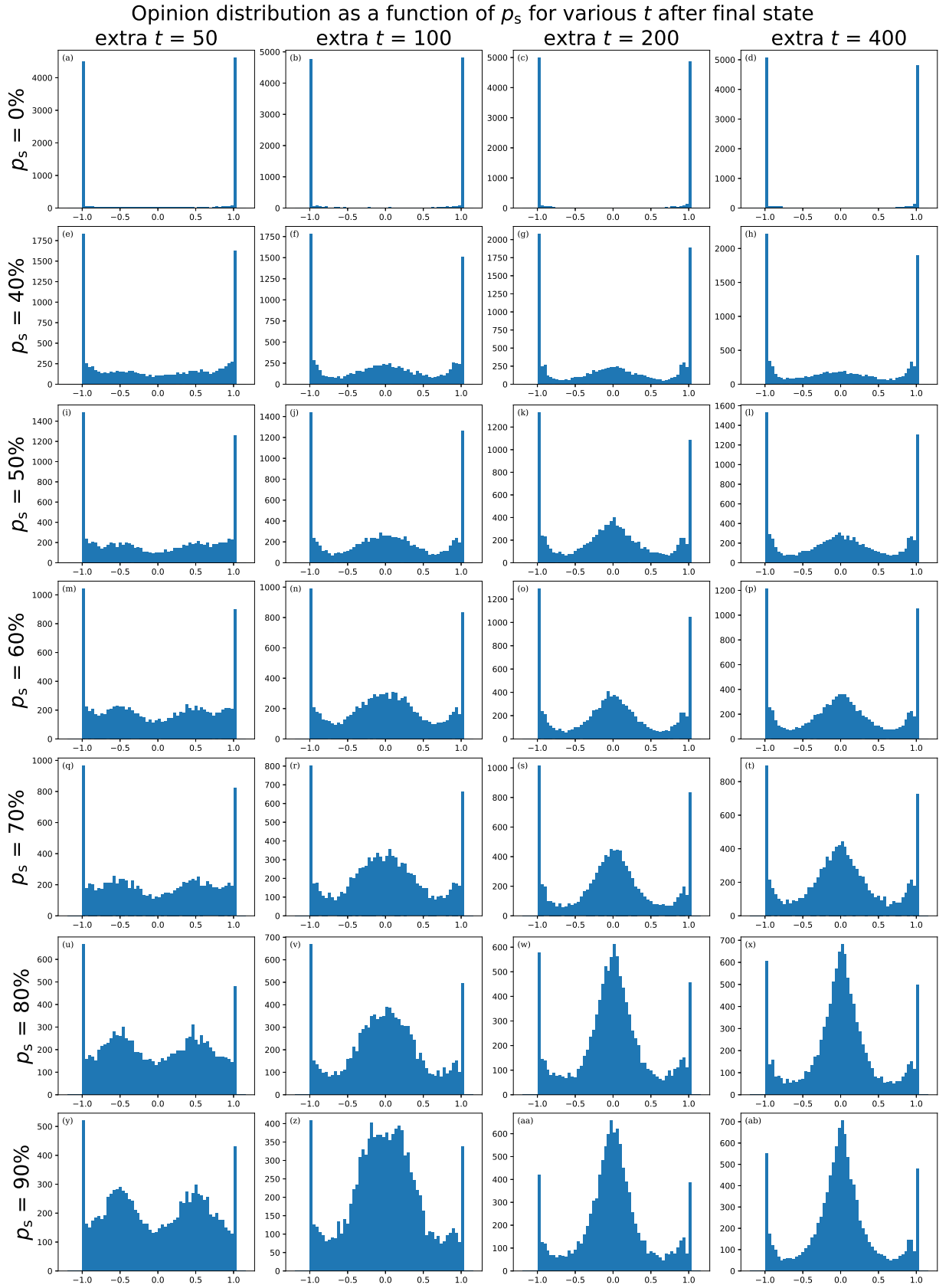


FIG. 9. Evolution of opinion distribution for different values of p_s after the system has almost reached a steady final state. All parameters taken are the same as those in Fig. 8.

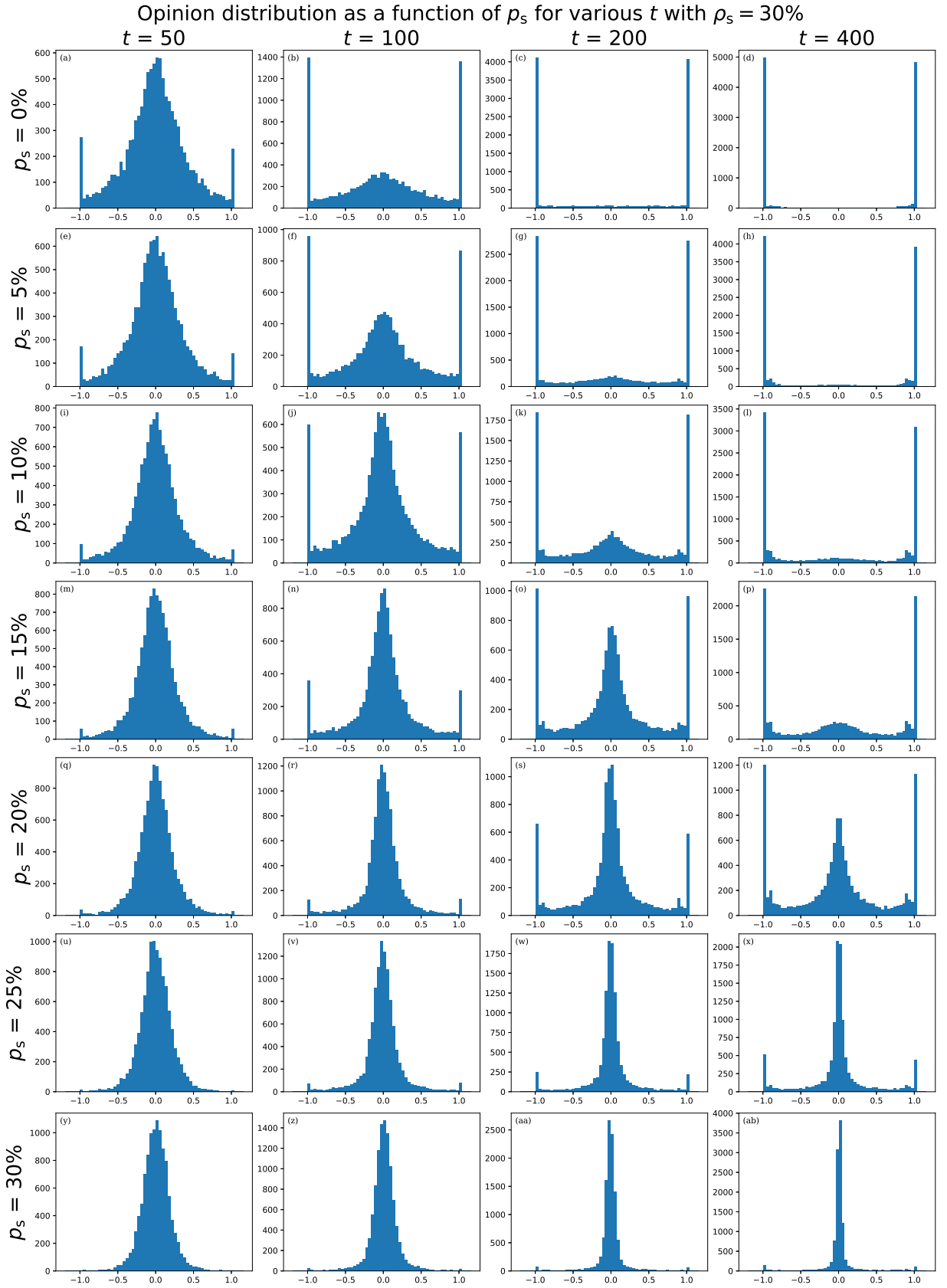


FIG. 10. Evolution of opinion distribution for different portion of empathetic agents p_s and mean number of interactions per agent t when $\rho_s = 0.3$. All other parameters are set to default values listed in Table I.

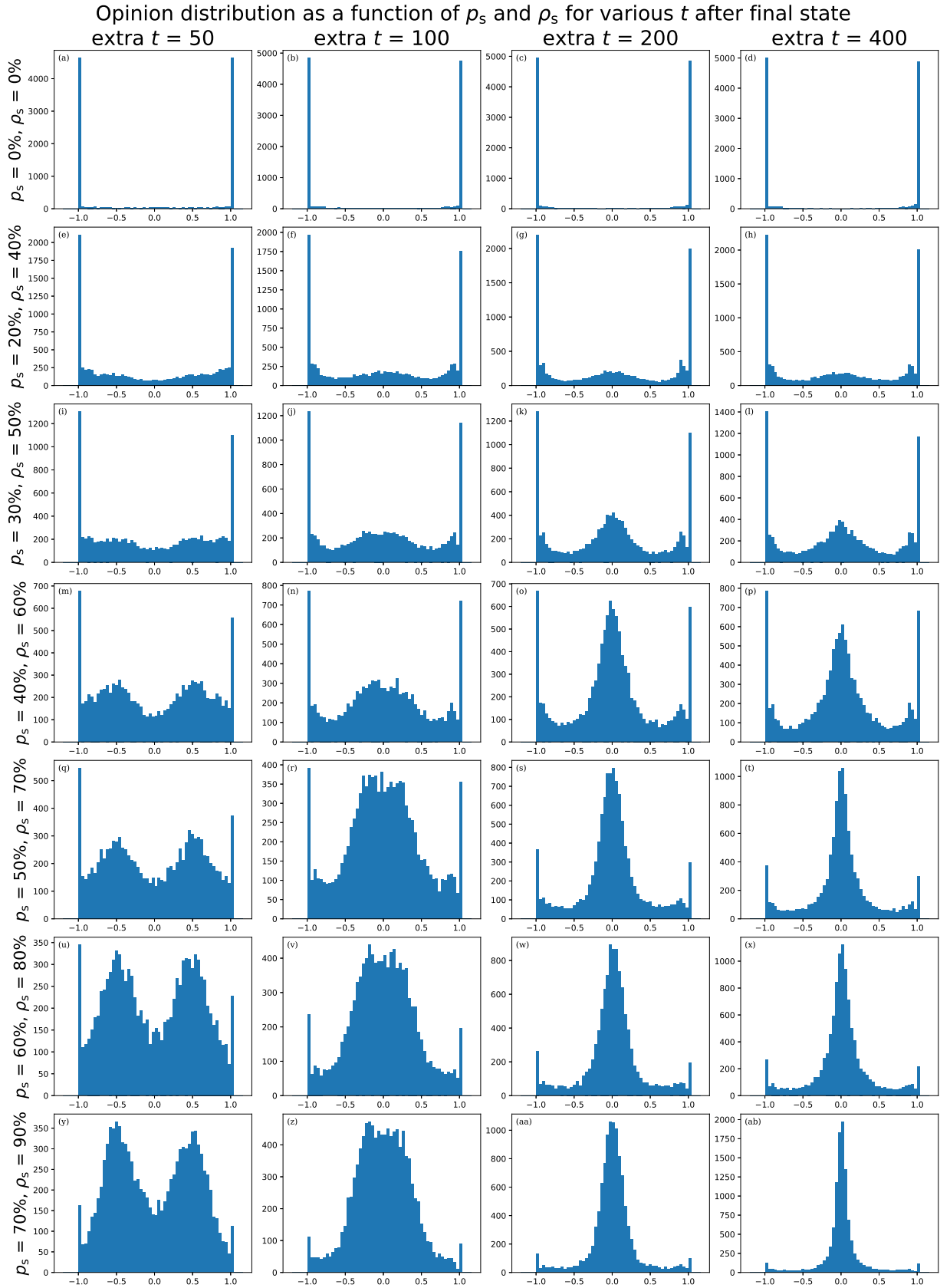


FIG. 11. Evolution of opinion distribution for different values of p_s after the system has almost reached a steady final state. All parameters taken are the same as those in Fig. 10.

A. The Case Of Unbiased Empathetic Agent Selection

We first consider the case when $\rho_s = 1$ so that each agent has equally likely chance of chosen to be empathetic. Here, we investigate the opinion evolution of systems with these empathetic agents in two ways. Fig. 8 depicts the opinion distribution evolution when empathetic agents are assigned in the beginning. As expected, the higher the value of p_s , the smaller the deviation in the attitudes of all agents. Although a sizable number of extremists is still present, a significant fraction of moderates remain. Nevertheless, this effect is significant only when $p_s \gtrsim 35\%$. Such a high percentage of empathetic agents is very hard if not impossible to reach in reality.

A more pressing issue is how to reduce society polarization in an already polarized society. Therefore, we study what happens if we change a fixed portion of agents from “normal” to empathetic *after* the society has evolved almost to its final highly polarized state. More precisely, we first let the system evolve with $t = 200$ using parameters listed in Table I so that it has almost equilibrated. Then, we will randomly pick and fix p_s of the agents and make them empathetic. Finally, we further evolve the system for an additional t up to 400. From Fig. 9, we observe that one can significantly reduce opinion polarization only when $p_s \gtrsim 40\%$, a goal that is even harder to attain.

B. The Case When Empathetic Agents Are Picked From The High Degree Nodes

In view of the ineffectiveness of picking empathetic agents randomly from the population, we consider the case of choosing them only from highly connected agents. In the real world, this corresponds to the case that a sizable number of social media influencers are empathetic. First we redo our simulations for the case of assigning empathetic agents right from the start by fixing $\rho_s = 0.3$ as an illustration. As shown in Fig. 10, a significant fraction of moderates remains when $p_s \gtrsim 20\%$ (which is much less than the critical value of $p_s \approx 35\%$ for the case when $\rho_s = 1$). The much more interesting case is when selected “normal” agents turn empathetic after the system has evolved with $t = 200$. Fig. 11 depicts that system becomes much less polarized if $\rho_s = 50\%$ and $p_s \gtrsim 30\%$. This corresponds to changing the behavior of

about 60% of those agents whose connectivity is above average. Furthermore, if picking $\rho_s = 40\%$ and $p_s = 60\%$, moderate opinion agents domain the system in the end. Compared to the case of randomly assigning empathetic agents, these strategies are much more viable although it is still highly challenging to implement.

VI. CONCLUSIONS

In this paper, we have identified deficiencies in the Axelrod model of opinion dynamics and proposed a modified one as an attempt to build a more sophisticated model to describe opinion dynamics in human society. Major modifications include using scale-free networks replacing the static two-dimensional lattice grid together with a new set of interaction rules for similar and dissimilar agents. Our simulation results show that over essentially all reasonable parameters agents’ opinions on all features eventually become more and more polarized. Besides, one or two dominating opinion clusters are formed together with numerous number of small opinion clusters. This is not unlike the big opinion differences between those living in cities and in rural areas as well as the emergence of political extremism due to distrust of the elites in the real world. We also discovered scaling in the distribution of the relative opinion cluster size $n(s, N)/N$. For B-A network, it follows a power law when $s \lesssim 10$ and switches to exponential decay when $15 \lesssim s \lesssim 35$. Finally, we tried to reduce this extreme opinion polarization by introducing empathetic agents that are open to talking and actively interacting with their neighbors. Their presence reduces opinion polarization but only when they appear *en masse*. Our study suggests that one effective and relatively economical but highly challenging way to lessen the difference in an already highly polarized opinion population in the real world is to change the behavior of a significant portion of social media influencers making them more empathetic.

Our model is flexible and can be modified to consider more realistic cases, even including evolution on the network. Our research is a small first step toward finding viable methods to understand opinion dynamics and reduce opinion polarization in our contemporary society.

ACKNOWLEDGMENTS

XZ was partially supported by The Innovation and Technology Scholarship of the Hong Kong Federation of Youth Groups.

[1] M. Jusup, P. Holme, K. Kanazawa, M. Takayasu, I. Romić, Z. Wang, S. Geček, T. Lipić, B. Podobnik, L. Wang, W. Luo, T. Klanjšček, J. Fan, S. Boccaletti, and M. Perc, *Phys. Rep.* **948**, 1 (2022).

[2] K. Sznajd-Weron and J. Sznajd, *Int. J. Mod. Phys. C* **11**, 1157 (2000).

[3] K. Sznajd-Weron, *Acta Phys. Pol. B* **36**, 2537 (2005).

[4] R. Hegselmann and U. Krause, *J. Artificial Societies and Social Simulation* **5**, 2 (2002),

- <https://www.jasss.org/5/3/2.html>.
- [5] S. Galam, *Int. J. Mod. Phys. C* **19**, 409 (2008).
 - [6] R. Axelrod, *J. Conflict Resolut.* **41**, 203 (1997).
 - [7] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, *Adv. Complex Sys.* **3**, 87 (2000).
 - [8] G. Weisbuch, G. Deffuant, F. Amblard, and J.-P. Nadal, *Complexity* **7**, 55 (2002).
 - [9] A. Sîrbu, V. Loreto, V. D. P. Servedio, and F. Tria, in *Participatory Sensing, Opinions and Collective Awareness*, edited by V. Loreto, M. Haklay, A. Hotho, V. D. Servedio, G. Stumme, J. Theunis, and F. Tria (Springer International Publishing, Cham, 2016) pp. 363–401.
 - [10] A. F. Peralta, J. Kertész, and G. Iñiguez, Opinion dynamics in social networks: From models to data, in *Handbook Of Computational Social Science*, edited by T. Yasserli (Elgaronline, 2025) pp. 384–406.
 - [11] J. Liu, J. He, Z. Qiu, and S. He, *Front. Phys.* **10**, 10.3389/fphy.2022.1042900 (2022).
 - [12] L. Helfmann, N. Djurdjevac Conrad, P. Lorenz-Spreen, and C. Schütte, *Sci. Rep.* **13**, 19375 (2023).
 - [13] H. Chau, C. Wong, F. Chow, and C.-H. F. Fung, *Physica A* **415**, 133 (2014).
 - [14] C. Castellano, M. Marsili, and A. Vespignani, *Phys. Rev. Lett.* **85**, 3536 (2000).
 - [15] K. Klemm, V. M. Eguiluz, R. Toral, and M. S. Miguel, *Phys. Rev. E* **67**, 045101(R) (2003).
 - [16] Q. Ashraf and O. Galor, *Am. Econ. Rev.* **103**, 528 (2013).
 - [17] P. Dandekar, A. Goel, and D. T. Lee, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5791 (2013).
 - [18] N. Koudenburg and Y. Kashima, *Pers. Soc. Psychol. Bull.* **48**, 1068 (2022).
 - [19] M. Ramos, J. Shao, S. D. S. Reis, C. Anteneodo, J. S. Andrade, S. Havlin, and H. A. Makse, *Sci. Rep.* **5**, 10032 (2015).
 - [20] O. Devauchelle, P. Szymczak, and P. Nowakowski, *Phys. Rev. E* **109**, 044106 (2024).
 - [21] P. MacCarron, P. J. Maher, S. Fennell, K. Burke, J. P. Gleeson, K. Durrheim, and M. Quayle, *PLOS ONE* **15**, e0233995 (2020).
 - [22] A. Radilo-Díaz, L. A. Pérez, and M. del Castillo-Mussot, *Phys. Rev. E* **80**, 066107 (2009).
 - [23] F. Battiston, V. Nicosia, V. Latora, and M. S. Miguel, *Sci. Rep.* **7**, 1809 (2017).
 - [24] P. Törnberg, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2207159119 (2022).
 - [25] C. S. Amorim, Bourdieu dynamics of fields from a modified Axelrod model (2014), arXiv:1407.3479 [physics.soc-ph].
 - [26] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
 - [27] W. Jager and F. Amblard, *Comput. Math. Organ. Theo.* **10**, 295 (2005).
 - [28] M. Bartolozzi, D. B. Leinweber, and A. W. Thomas, *Phys. Rev. E* **72**, 046113 (2005).
 - [29] D. Jacobmeier, *Int. J. Mod. Phys. C* **16**, 633 (2005).
 - [30] A. Dinkelberg, P. MacCarron, P. J. Maher, and M. Quayle, *Physica A* **578**, 126086 (2021).
 - [31] J. Lorenz, *Int. J. Mod. Phys. C* **18**, 1819 (2007).
 - [32] C. Bernardo, C. Altafini, A. Proskurnikov, and F. Vasca, *Automatica* **159**, 111302 (2024).
 - [33] S. Fortunato, V. Latora, A. Pluchino, and A. Rapisarda, *Int. J. Mod. Phys. C* **16**, 1535 (2005).
 - [34] J. Lorenz, in *Managing Complexity: Insights, Concepts, Applications*, edited by D. Helbing (Springer, Berlin, 2008) pp. 321–334.
 - [35] S. Huet and G. Deffuant, *Adv. Complex Sys.* **13**, 405 (2010).
 - [36] S. Liu, M. Mäs, H. Xia, and A. Flache, *J. Artificial Societies and Social Simulation* **26**, 8 (2023).
 - [37] X. Zou, *Opinion Dynamics: Go Beyond Axelrod Model and the Application on Voting Dynamics*, Project report of the course “Data Analysis And Modeling In Physics” (Univ. of Hong Kong, 2023).
 - [38] A.-L. Barabási, *Science* **325**, 412 (2009).
 - [39] C. I. Hovland, I. L. Janis, and H. H. Kelley, *Communication And Persuasion; Psychological Studies Of Opinion Change* (Yale U Press, 1953).
 - [40] A. Levy and Y. Maaravi, *Soc. Influ.* **13**, 39 (2017).
 - [41] D. Ringold, *J. Consumer Policy* **25**, 27 (2002).
 - [42] U. K. H. Ecker, S. Lewandowsky, J. Cook, P. Schmid, L. K. Fazio, N. Brashier, P. Kendeou, E. K. Vraga, and M. A. Amazeen, *Nat. Rev. Psychol.* **1**, 13 (2022).
 - [43] P. W. Schultz, J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius, *Psychol. Sci.* **18**, 429 (2007).
 - [44] K. T. Verkooijen, F. M. Stok, and S. Moolen, *Eur. J. Soc. Psychol.* **45**, 417 (2015).
 - [45] A. Z. Liu, A. X. Liu, R. Wang, and S. X. Xu, *J. Manag. Stud.* **57**, 1437 (2020).
 - [46] K. Takács, A. Flache, and M. Mäs, *PLOS One* **11**, e0157948 (2016).
 - [47] C. A. Bail, L. P. Argyle, T. W. Brown, J. P. Bumpus, H. Chen, M. B. F. Hunzaker, J. Lee, M. Mann, F. Merhout, and A. Volfovsky, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9216 (2018).
 - [48] M. Pan, M. Blut, A. Ghiassaleh, and Z. W. Y. Lee, *J. Acad. Mark. Sci.* **53**, 52 (2025).
 - [49] R. Conde and B. Casais, *J. Bus. Res.* **158**, 113708 (2023).
 - [50] C. Peter and L. Muth, *Media Commun.* **11**, 164 (2023).
 - [51] X. Liu and X. Zheng, *Humanit. Soc. Sci.* **11**, 15 (2024).
 - [52] X. Zou, Python code used in our simulation, available in https://github.com/SimbaZouXiang/Sociophysics_Option_dynamics (2025).