

# Partisan Confidence Model for Group Polarization

Armineh Rahmanian Kooshkaki, Sadegh Bolouki, S. Rasoul Etesami, and Abolfazl Mohebbi

**Abstract**—Models of opinion dynamics play a major role in various disciplines, including economics, political science, psychology, and social science, as they provide a framework for analysis and intervention. In spite of the numerous mathematical models of social learning proposed in the literature, only a few models have focused on or allow for the possibility of popular extreme beliefs' formation in a population. This paper closes this gap by introducing the Partisan Confidence (PC) model inspired by the foundations of the well-established socio-psychological theory of groupthink. The model hints at the existence of a tipping point, passing which the opinions of the individuals within a so-called "social bubble" are exaggerated towards an extreme position, no matter how the general population is united or divided. The results are also justified through numerical experiments, which provide new insights into the evolution of opinions and the groupthink phenomenon.

**Index Terms**—Opinion dynamics, groupthink, group polarization, partisan confidence

## 1 INTRODUCTION

OPINION dynamics is an important area of research with a wide range of applications in political campaigning, marketing, transportation management, public opinion management [1], and group recommender systems [2]. In particular, due to the rapid growth of online social networks and unprecedented ease of opinion exchange on these platforms, there has been growing interest in how individuals' opinions are formed and perceived within a population [3]. An interesting phenomenon frequently observed on these platforms, yet largely rejected by the existing opinion dynamics models, is that extremist positions can emerge and become mainstream [4]–[7].

The vast majority of previous models have focused on the notion of *conformity* [8] and references therein. The upshot is that conformist individuals avoid extreme beliefs and occupy an intermediate and middle-ground position as time flows. To better capture real-world situations where a compromised position is not always achievable, classic conformity-based models, such as DeGroot's [9] or Abelson's models [10], have gone through several modifications [11]–[15]. A noteworthy advancement in this area is the development of several models with attention to cognitive biases, particularly confirmation bias, such as the bounded confidence model [16], [17] and its variants [18]–[25], each focuses on different behaviors, including agreement, disagreement, fragmentation, and polarization.

Some attempts have been made to modify classical conformity-based models to ones where opinion polarization is not beyond

the realms of possibility. A notable example is the Altafini model [26]–[28], which incorporates the notion of antagonism among individuals. Remarkable variations have also included *bounded confidence* and *biased assimilation*, each of which can be viewed as a type of confirmation bias, as an individual's characteristics in the dynamics [29]–[32]. While antagonistic interactions/relationships can contribute to and justify the tendency toward more extreme beliefs in a divided population, such as that of the United States with respect to political ideology (see Fig. 1, extracted from [33]), extreme beliefs have also been observed to form in fully collaborative environments [34]. Furthermore, these models often need the pre-existence of extreme tendencies at the onset of the opinion evolution to explain how such tendencies become mainstream.

A handful of frameworks have been developed to address the emergence of popular extreme beliefs in fully collaborative environments where antagonistic relationships are absent. For instance, a model based on Persuasive Argument Theory and homophily has been proposed in [35], [36], where individuals who share similar viewpoints are more likely to engage in conversation with and influence each other. As another example, in [37], a model has been introduced to explain the prevalence of extremism if stubborn extremists exist at the onset of opinion evolution. Other remarkable examples are [38]–[40], where biased assimilation is incorporated into Degroot's opinion averaging model [9] to create the possibility of extreme beliefs emerging and becoming popular [38]–[40]. We refer the reader to [8], [41] for other recent developments in the field of opinion dynamics.

Our main objective in this work is to propose, conceptualize, and investigate a mathematical model inspired by the socio-psychological analysis of the groupthink phenomenon to study opinion formation. More specifically, in contrast to other models in the literature, where interaction among like-minded agents leads to an intermediate and middle-ground position, we aim to propose a model of opinion formation where interaction among like-minded agents may lead to resonance and extremism. For this purpose, we leverage the causes of a socio-psychological phenomenon called groupthink. Groupthink corresponds to a situation where the tendency of agents to seek consensus becomes dominant in a cohesive in-group, which results in the members irrationally disregarding or

- A. Rahmanian Kooshkaki is with the Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran, 14155. E-mail: armineh.rahmanian@modares.ac.ir
- S. Bolouki is with Department of Mechanical Engineering, Polytechnique Montréal, Montreal, QC, Canada, H3T 1J4. E-mail: sadegh.bolouki@polymtl.ca
- S.R. Etesami is with the Department of Industrial & Enterprise Systems Engineering and Coordinated Science Lab, University of Illinois at Urbana-Champaign, Urbana, IL, USA, 61801. E-mail: etesami1@illinois.edu
- A. Mohebbi is with Department of Mechanical Engineering, Polytechnique Montréal, Montreal, QC, Canada, H3C3A7. E-mail: abolfazl.mohebbi@polymtl.ca
- This work is partially supported by the NSF CAREER Award under Grant No. EPCN-1944403.

discounting unpopular realistic views [42]–[44]. Notable antecedent conditions for groupthink include high cohesiveness within the group, isolation of the members from contrary views, and a lack of impartial leadership. Furthermore, according to the *ubiquity model* of groupthink [45], social identification, salient norms, and low self-efficacy are replaced by previous ones. The antecedent conditions cause individuals to suppress dissenting views despite their realistic spirit, where inevitable defective consequences and poorly made decisions are often observed [46]–[49].

*Group polarization*, as a direct, immediate consequence of the groupthink phenomenon, is a group's tendency to make decisions that are more extreme than the initial positions of its members [43]. The term group polarization, with which the current work is mainly concerned, should not be confused with opinion polarization, even though the two phenomena are correlated and may coexist within a population. Opinion polarization captures how far from consensus the opinions are distributed in a population.

We also note that groupthink differs from the *wisdom of crowds*, the theory that crowds make better and more accurate decisions than individuals. The primary point is removing social influence due to conformity behavior [50]. A group can make better decisions if this static one-step process is transformed into an iterative one with equal social influence [51]. In contrast, groupthink is an iterative process where individuals make extreme and defective decisions under certain conditions.

The core aim of this paper is to offer, develop, and analyze new models of opinion formation, namely *Partisan Confidence-lite (PC-lite)* and *Partisan Confidence (PC)*, that can justify the emergence and prevalence of popular extreme beliefs established on the rich theory of groupthink. The key contributions are as follows.

- 1) Decomposition of opinions based on Partisan Confidence: The models are built upon a unique understanding of opinions and social learning, which is called partisan confidence herein. Opinions are decomposed based on the concept of partisan confidence, where confidence refers to the intensity (absolute value), and partisan is the direction (sign) of an individual's opinion.
- 2) Social influence characterization: By deploying this partisan confidence viewpoint: (i) Interaction among like-minded individuals lead to resonance, (ii) The impact of opposing opinions have diminished, (iii) More confidence yields more impact on others.
- 3) Introducing the PC and PC-lite models: Furthermore, two models for social learning, PC-lite and PC, are established upon the characterization of social influence described above. While the PC model is more compelling since it also accounts for the confirmation bias (item (ii) above), the PC-lite model is of great importance as it demonstrates that group polarization may occur even without the contribution of confirmation bias.
- 4) Deriving conditions for group polarization: We derive conditions for group polarization through rigorous analysis of the PC and PC-lite models. For the PC-lite model, these conditions are primarily established on isolation and cohesiveness as groupthink's antecedent conditions. However, in the PC model, incorporating confirmation bias allows for lower degrees of isolation for the emergence of extremism, aligning with the recent antecedent conditions of groupthink.

There are several differences between this work and works in [24], [35], [38]. The partisan confidence decomposition of opinions is different from viewpoints in [24], [38]. The work in [24] considered publicly known discrete choices when the exact opinions

are not known. The model in [38] assumes the distance between an individual's opinion and each endpoint of the opinion spectrum represents her support for that marginal belief. In addition, works in [35], [38] demonstrate group polarization due to confirmation bias; however, the PC-lite model in this work is capable of presenting such behavior without the presence of confirmation bias. Furthermore, in contrast to [35], [38], both PC-lite and PC models consider time-varying ties and influences among individuals. Moreover, in classic diffusion models such as Independent Cascade and Linear Threshold models [52], [53], active and inactive states may resemble the partisan confidence (partisan part) viewpoint deployed in this work. However, these states are binary and cannot demonstrate the confidence of an opinion and its social influence. In addition, confirmation bias, which is the core idea of the PC model, is not captured by mentioned models.

The remainder of this paper is organized as follows. In Section 2, after setting out some notations, we introduce and investigate a basic model, the so-called *PC-lite* model, the generalization of which leads to the PC model in Section 3. We discuss the inherent properties of the proposed models in their corresponding sections. In Section 4, we provide simulations to demonstrate the properties of the proposed models. We conclude the paper with a discussion on stubbornness and also identify some future directions of research in Section 5. We relegate all the proofs and lemmas to Section 6.

## 2 PARTISAN CONFIDENCE-LITE MODEL AND ITS PROPERTIES

Before proceeding to model definition, we introduce some preliminaries and then the model.

### 2.1 Preliminaries And Notation

A weighted, time-varying digraph  $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t), W(t))$  is assumed to represent the topology of the network of agents over time, where  $\mathcal{V} = \{1, \dots, n\}$  is the set of nodes,  $\mathcal{E}(t) \subseteq \{(i, j) | i, j \in \mathcal{V}, i \neq j\}$  is the set of edges at time  $t$  indicating the interactions among the agents, and an element  $w_{ij}(t)$  of the weight matrix  $W(t)$  indicates the weight of influence of agent  $j$  on agent  $i$  at time  $t$ . It is assumed that  $w_{ij}(t) > 0$  if  $(i, j) \in \mathcal{E}(t)$ , and  $w_{ij}(t) = 0$  otherwise. The non-negative matrix  $W(t)$  is called *row-stochastic* if  $\sum_{j \in \mathcal{V}} w_{ij}(t) = 1$  for all  $i \in \mathcal{V}$ , and it is called *row-substochastic* if  $\sum_{j \in \mathcal{V}} w_{ij}(t) \leq 1$  for all  $i \in \mathcal{V}$ , and there exists some  $k \in \mathcal{V}$  such that  $\sum_{j \in \mathcal{V}} w_{kj}(t) < 1$  [54]. Throughout the paper, and in the proofs in particular, the argument  $t$  of time-varying functions is dropped for the benefit of notational convenience. For instance,  $x_i$  and  $z$ , and  $w_{ij}$  often replace  $x_i(t)$ ,  $z(t)$ , and  $w_{ij}(t)$ , respectively. Furthermore, an agent's update value,  $x_i(t+1) - x_i(t)$ , will be denoted by  $\Delta x_i$ , that itself is short for  $\Delta x_i(t)$ .

### 2.2 Partisan Confidence-lite Model

In this section, we introduce, justify, and investigate the Partisan Confidence-lite (PC-lite) model for the evolution of opinions in a social network. Let  $x_i(t) \in [-1, 1]$  denote the opinion of agent  $i \in \mathcal{V}$  at discrete time  $t \geq 0$ , where  $\mathcal{V} = \{1, \dots, n\}$  is the set of all agents. In the PC-lite model, the opinion of every agent  $i$  evolves according to the following discrete-time dynamics:

**(PC-lite dynamics)**

$$\Delta x_i = \sum_{j \neq i} [w_{ij} |x_j| (\text{sgn}(x_j) - x_i)]. \quad (1)$$

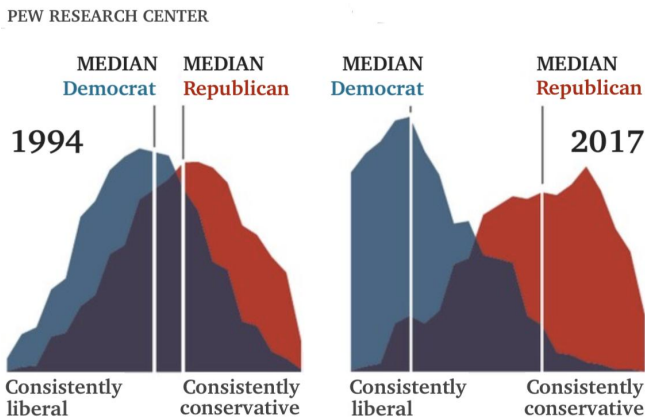


Fig. 1. Growing ideological divide in the United States (Source: Pew Research Center [33]).

To be clear, as described in Subsection 2.1, the dynamics (1) should be read as

$$x_i(t+1) - x_i(t) = \sum_{j \neq i} [w_{ij}(t)|x_j(t)|(\text{sgn}(x_j(t)) - x_i(t))]. \quad (2)$$

According to the PC-lite dynamics (1), a self-weight  $w_{ii}(t)$  is not present and does not contribute to the opinion change of agent  $i$  at time  $t$ . Thus, we can assume  $w_{ii}(t) = 0, \forall i \in \mathcal{V}, t \geq 0$ . In addition, we assume that the influence matrix  $W(t) = [w_{ij}(t)]$  is either row-stochastic or row-substochastic in the above dynamics.

### 2.3 Justification of the PC-lite Model

One notices that the PC-lite dynamics (1) follows, in principle, the same rule of social influence as the time-varying version of the DeGroot model [9],

**(DeGroot dynamics)**

$$\Delta x_i = \sum_{j \neq i} [w_{ij}(x_j - x_i)]. \quad (3)$$

However, (1) is set up on a fundamentally distinctive interpretation of opinion perception, that is, agent  $i$  perceives the opinion  $x_j$  of agent  $j$  as an approval of  $\text{sgn}(x_j)$  with confidence level  $|x_j|$ . Subsequently, the weight of influence of agent  $j$  on agent  $i$  is discounted by the factor  $|x_j|$ , while the opinion  $x_j$  of agent  $j$  is replaced by  $\text{sgn}(x_j)$ . Therefore,  $x_j$  is decomposed into two parts: a direction part  $\text{sgn}(x_j)$ , which can be viewed as party affiliation in political terms, and an intensity or confidence part  $|x_j|$ . It is this decomposition of agents' opinions that justifies the appellation *Partisan Confidence*. Indeed,  $\text{sgn}(x_j)$  is often concerned with a much more specific issue than one's political party. For instance, on the issue of abortion rights, it addresses whether a person generally supports or is against abortion rights.

The consideration that the influence weights  $w_{ij}(t)$  are time-varying adds to the practicality of the PC-lite model since (i) no agent has to interact with the same set of agents at all times, i.e., there are asynchronous interactions, and (ii) the dynamics (1) with fixed weights  $w_{ij}$  cannot accurately model human thinking, i.e., there is model uncertainty.

We may use  $w_{ij}(t)|x_j(t)|$  to refer to the overall influence of agent  $j$  on agent  $i$  at time  $t$ . The amount of this overall influence

is determined by the intensity of an opinion  $|x_j(t)|$ . On the other hand, the direction  $\text{sgn}(x_j(t))$  determines whether the overall influence is in favor of or against an opinion. In fact, one can think of  $|x_j(t)|$  as the intensity of the emotion being transferred to another agent that either advocates or disapproves of some position or idea, and this intensity governs the overall influence. Thus, the more extreme the emotion, the greater the overall influence would be, and vice versa. When the opinion of agent  $j$  at time  $t$  is zero, we assume that she is completely neutral; thus, her opinion will not drag the opinion of agent  $i$  at time  $t$  in either directions on the opinion spectrum. In other words, a neutral opinion does not contribute to the opinion change of an agent.

### 2.4 Properties of the PC-lite Model

We now investigate the PC-lite dynamics (1) in detail and discuss why it can explain groupthink behavior. A key antecedent condition for groupthink is the isolation of the group members from the outside population. We start off with a simple but important result that highlights the unique capability of the PC-lite model (and also the PC model discussed in the next section) in explaining group polarization and the emergence of popular, extreme beliefs in a network of agents.

**Definition 1 (Connectedness).** Given the model (1), a subset  $\mathcal{B} \subseteq \mathcal{V}, |\mathcal{B}| > 1$ , of agents is said to be connected if

$$\sum_{j \in \mathcal{B}} \sum_{t=0}^{\infty} w_{ij}(t) = \infty, \forall i \in \mathcal{B}. \quad (4)$$

**Proposition 1.** Given the model (1), let a connected subset  $\mathcal{B} \subseteq \mathcal{V}$  of agents be isolated from outside, i.e., for any  $i \in \mathcal{B}$ , assume that

$$\sum_{j \in \mathcal{V} \setminus \mathcal{B}} w_{ij}(t) = 0. \quad (5)$$

If for some  $t_0, x_i(t_0) > 0, \forall i \in \mathcal{B}$ , then we have  $\lim_{t \rightarrow \infty} x_i(t) = 1, \forall i \in \mathcal{B}$ .

*Proof.* Proposition 1 is a special case of Proposition 2, stated later on in this section. See Subsection 6.3 for more details.  $\square$

To give an intuition about the validity of Proposition 1, without considering the Proposition 2, which will be stated later in the upcoming sections, one can simply take summation in (1) over  $j \in \mathcal{B}$  and substitute the sign by  $+1$ . Since the bubble  $\mathcal{B}$  is connected and for all  $j \in \mathcal{B}$  we have  $\text{sgn}(x_j) = +1$ , the only equilibrium point is  $x_i = +1$ . Proposition 1 addresses an exaggerated but important situation in which a set of connected agents is completely isolated from the rest of the population. It states that if the agents in this set are initially in agreement, no matter how weak this agreement is, they reach an extremely strong agreement as time passes. In other words, in the absence of opposing views, given consistent interactions among the agents, group polarization is inevitable.

**Definition 2 (Bubble number).** Given the model (1) and a subset  $\mathcal{B} \subseteq \mathcal{V}, |\mathcal{B}| > 1$ , of agents, the bubble number of  $\mathcal{B}$ , denoted by  $\gamma_{\mathcal{B}}$ , is defined as the largest non-negative constant  $\gamma$  that satisfies the following equation for any  $i \in \mathcal{B}$ :

$$\sum_{j \in \mathcal{B}} w_{ij}(t) \geq \gamma_{\mathcal{B}} \sum_{j \in \mathcal{V} \setminus \mathcal{B}} w_{ij}(t). \quad (6)$$

The bubble number is well-defined for any  $\mathcal{B}$  since (6) is satisfied by an upper-bounded, closed interval in  $\mathbb{R}$  containing 0.

Equation (6) states that the sum of the weights inside  $\mathcal{B}$  is at least  $\gamma_{\mathcal{B}}$  times larger than the sum of the weights to agents outside that bubble. Thus,  $\gamma_{\mathcal{B}}$  indicates the isolation level of  $\mathcal{B}$  from outside in the sense of opinion influence. The greater the bubble number  $\gamma_{\mathcal{B}}$ , the greater the isolation of the members in the subset. It is worth noting that the bubble number is closely related to the so-called *cut ratio* of a weighted graph [55]. More precisely, if we sum (6) over all  $i \in \mathcal{B}$ , we obtain

$$\frac{1}{\gamma_{\mathcal{B}}} \geq \frac{\sum_{i \in \mathcal{B}, j \in \mathcal{V} \setminus \mathcal{B}} w_{ij}(t)}{\sum_{i, j \in \mathcal{B}} w_{ij}(t)}, \quad (7)$$

where the expression on the right side is the ratio of the sum of the edge weights crossing the cut  $\mathcal{B}$  over the sum of the edge weights inside that cut. It is known that the minimum cut ratio over all the cuts can be bounded by the algebraic connectivity of the graph [55]. Therefore, one can bound the bubble number in terms of the eigenvalues of the adjacency matrix  $W(t)$ .

**Definition 3 (Social bubble).** *Given the model (1), a subset  $\mathcal{B} \subseteq \mathcal{V}$ ,  $|\mathcal{B}| > 1$ , of agents is loosely called a social bubble, or simply a bubble, if it has a large bubble number, meaning that it is, to a great extent, isolated from outside influence.*

To clarify the concept of a social bubble, we shall give an example in the following. See the digraph shown in Fig. 2, where  $\mathcal{V} = \{1, \dots, 10\}$  and  $\mathcal{B}_1 = \{1, 2, 3, 4\}$  and  $\mathcal{B}_2 = \{5, 6, 7\}$ . It is easy to check that the bubble property exists both in  $\mathcal{B}_1$  and  $\mathcal{B}_2$  with  $\gamma_{\mathcal{B}_1} = 4$  and  $\gamma_{\mathcal{B}_2} = 3$ . Suppose that for bubble  $\mathcal{B}$ ,  $w_{in}(i) = \sum_{j \in \mathcal{B}} w_{ij}$  and  $w_{out}(i) = \sum_{j \in \mathcal{V} \setminus \mathcal{B}} w_{ij}$ . For bubble  $\mathcal{B}_1$ , we have  $[w_{in}(k)]/[w_{out}(k)] = 4$ ,  $k \in \{1, 3\}$ ,  $w_{out}(2) = 0$ , and  $[w_{in}(4)]/[w_{out}(4)] = 5$ . For bubble  $\mathcal{B}_2$ , we have  $w_{out}(5) = w_{out}(7) = 0$ , and  $[w_{in}(6)]/[w_{out}(6)] = 3$ .

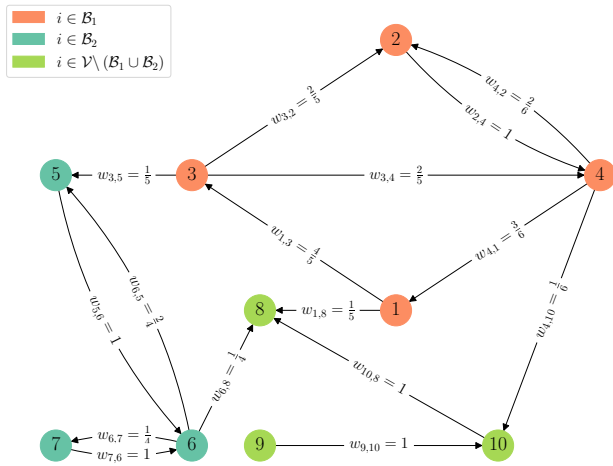


Fig. 2. An example of a social bubble with bubble number  $\gamma_{\mathcal{B}_1} = 4$  and  $\gamma_{\mathcal{B}_2} = 3$  where  $\mathcal{B}_1 = \{1, 2, 3, 4\}$  and  $\mathcal{B}_2 = \{5, 6, 7\}$ , at a fixed time instance  $t$ .

From the theory of groupthink, it is expected that agents in a connected bubble will intensify cohesiveness (if it exists) in a discussion, seeking stronger agreement within the bubble. According to the following proposition, this phenomenon is very well captured by the PC-lite model for social learning.

**Proposition 2.** *Given the PC-lite dynamics (1), let a connected*

*subset  $\mathcal{B} \subseteq \mathcal{V}$  of agents have the bubble number*

$$\gamma_{\mathcal{B}} > 3 + 2\sqrt{2}, \quad (8)$$

*and assume that  $\alpha_1$  and  $\alpha_2$ , where  $\alpha_1 < \alpha_2$ , are the two positive solutions of the equation*

$$\frac{1 + \alpha}{\alpha(1 - \alpha)} = \gamma_{\mathcal{B}}. \quad (9)$$

*If, for some  $t_0$ , it happens that*

$$x_i(t_0) > \alpha_1, \quad \forall i \in \mathcal{B}, \quad (10)$$

*then we have*

$$\liminf_{t \rightarrow \infty} x_i(t) \geq \alpha_2, \quad \forall i \in \mathcal{B}. \quad (11)$$

*Proof.* The proof can be found in Section 6.  $\square$

The threshold  $3 + 2\sqrt{2}$  in Proposition 2 marks the smallest possible  $\gamma_{\mathcal{B}}$  for which (9) has two positive real solutions for  $\alpha$ . It can also be viewed as the threshold that makes the loosely defined notion of a “social bubble” in Definition 3 precise. Therefore, Proposition 2 implies that if the agents in a bubble reach a certain degree of cohesiveness, that is, are at least  $\alpha_1$ -confident in advocating in favor of a common position, then their confidence tends to grow higher, beyond degree  $\alpha_2$ . As the bubble number  $\gamma_{\mathcal{B}}$  increases,  $\alpha_1$  and  $\alpha_2$  will decrease and increase, respectively, as demonstrated in Fig. 3. In limit, as  $\gamma_{\mathcal{B}}$  goes to infinity,  $\alpha_1$  reaches 0 while  $\alpha_2$  reaches 1, making the case for Proposition 1. For  $\gamma_{\mathcal{B}} \approx 7.83$ ,  $\alpha_1$  and  $\alpha_2$  are 1/2 apart. It should also be noted that the same can be said about a bubble in which the agents disapprove of a position. Hence, in summary, Proposition 2 shows group polarization occurring within a connected bubble if the opinions of the agents in the bubble have reached a certain degree of support/disapproval of any given position at some time  $t_0$ . Furthermore, we shall note that the spirit of the Proposition 2 and the concept of a social bubble applies to small or intermediate groups. More specifically, although a very large and connected group that is loosely connected to the outside can be a social bubble, the emergence of a weak agreement among members of a huge group does not seem probable to take place in reality (equation (10)).

We note that the term social bubble is broader than what is loosely known as an *echo chamber* or a *filter bubble* in that a social bubble, unlike an echo chamber or a filter bubble, may include individuals with diverse or even opposite beliefs. An echo chamber refers to a setting where a group of individuals only receive information consistent with their viewpoint, causing them to strengthen their beliefs. In addition, a filter bubble is a term used for an echo chamber built as a result of recommender systems and search engine algorithms designed to maximize user entertainment and consumption. There has been a great deal of controversy on whether these algorithms are reinforcing political segregation and causing the users to drown in their filter bubbles and get disconnected from opposing views. However, evidence from both sides of the argument suggests that both statements are true to a certain degree [56]. Proposition 2 demonstrates that once a social bubble turns into an “echo chamber”, i.e., once the condition (10) is satisfied, one should expect exaggeration of those beliefs as time grows (equation (11)).

*Remark 1.* Suppose that  $\mathcal{V}$  contains at least  $m$  connected, pairwise disjoint social bubbles  $\mathcal{B}_1, \dots, \mathcal{B}_m$ . Now, depending on whether, for each bubble  $\mathcal{B}_k, k = 1, \dots, m$ , we have  $x_i(t_{0_k}) > \alpha_1, \forall i \in \mathcal{B}_k$  or  $x_i(t_{0_k}) < -\alpha_1, \forall i \in \mathcal{B}_k$ , where

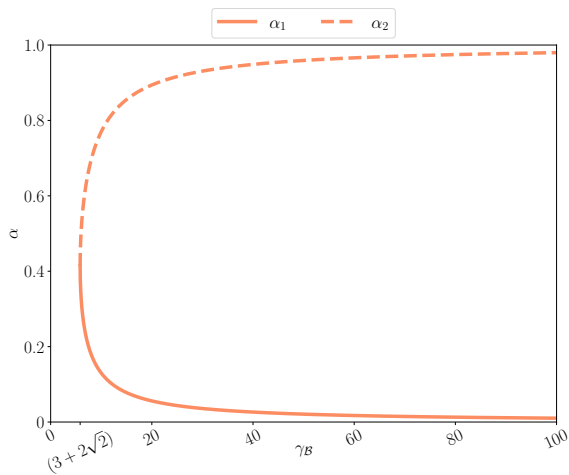


Fig. 3. Changes of  $\alpha_1$  and  $\alpha_2$  with respect to  $\gamma_B$

$t_{0_k} \geq 0$ , we have

$$\liminf_{t \rightarrow \infty} x_i(t) \geq \alpha_2, \quad \forall i \in \mathcal{B}_k, \quad (12)$$

or

$$\liminf_{t \rightarrow \infty} x_i(t) \leq -\alpha_2, \quad \forall i \in \mathcal{B}_k, \quad (13)$$

respectively. In particular, the asymptotic structure of the bubbles can be represented via one of the  $2^m$  vectors  $s \in \{-1, 1\}^m$  such that  $s_k = +1$  if (12) holds, and  $s_k = -1$  if (13) holds. Thus, the network can exhibit  $2^m$  substantially different limiting behaviors.

### 3 PARTISAN CONFIDENCE MODEL AND ITS PROPERTIES

Acting toward opinions with a bias has been well documented in confirmation bias theory [57]. Agents tend to respond with a bias toward information inconsistent with their own information, beliefs, and old experiences. Also, agents tend to willingly ignore some nonconforming information and opinions only to fit into their social groups [58]. In summary, this bias can be due to receiving information that is inconsistent or in conflict with one's social norm or identity.

In this section, we introduce Partisan Confidence (PC) model, which is a generalization of the PC-lite model (1) that accounts for the agents' confirmation bias. As we discussed earlier, the PC-lite model (1) can describe the group polarization caused by the groupthink behavior described in Irving L. Janis's seminal work [47]. To fit that model into Robert S. Baron's more advanced model of groupthink [45], we assume that each agent  $i$  discounts the influence of contrary views received from any other agent and propose the following opinion dynamics model:

#### (PC dynamics)

$$\Delta x_i = \sum_{j \neq i} [d_i(x_i, x_j) w_{ij} |x_j| (\text{sgn}(x_j) - x_i)], \quad (14)$$

where  $d_i : [-1, 1]^2 \rightarrow [0, 1]$  is a discounting function elaborated in the following subsection, before investigating the properties of the PC model. The inclusion of the discounting function in the PC model (14), that is, discounting of opposing views, in a sense amplifies the isolation degree of a cohesive bubble. Hence, in view of Proposition 2, one expects that the bubble number threshold for group polarization should now be lower, as it is made concrete later.

### 3.1 Discounting Function

As implied from its title, the discounting function is assumed to always return a number within  $[0, 1]$ ; a trivial assumption which will not be repeated but made throughout. Furthermore, in view of the confirmation bias, agent  $i$  discounts the influence of agent  $j$  with a general belief opposite to hers. No discount is expected otherwise, meaning that

$$d_i(x_i, x_j) = 1 \text{ if } \text{sgn}(x_i) = \text{sgn}(x_j). \quad (15)$$

We also assume that the discount value for opposing general beliefs is at all times upper bounded as

$$d_i(x_i, x_j) \leq \hat{d}_i(|x_i|) \text{ if } \text{sgn}(x_i) \neq \text{sgn}(x_j), \quad (16)$$

where  $\hat{d}_i : [0, 1] \rightarrow [0, 1]$  is an arbitrary non-increasing function. It should be noted that the non-increasing assumption on  $\hat{d}_i$  is reasonable, as it implies that confirmation bias increases with confidence. While our analysis shall remain valid for any discounting function satisfying (15) and (16) for a non-increasing  $\hat{d}_i$ , to shed some light on the PC model, we consider the following candidate for  $\hat{d}_i$ :

$$\hat{d}_i(|x_i|) = 1 - (1 - d)|x_i|^\beta \quad (17)$$

where  $d$  and  $\beta$  are constants satisfying  $0 \leq d \leq 1$  and  $\beta > 0$ . This means that the condition (16) now translates to

$$d_i(x_i, x_j) \leq 1 - (1 - d)|x_i|^\beta \text{ if } \text{sgn}(x_i) \neq \text{sgn}(x_j). \quad (18)$$

In what follows, the interpretation of the parameters  $d$  and  $\beta$ , along with the justification of the upper bound assumption, are given. The case for a general discounting function satisfying (15) and (16) will be discussed at the very end of the section.

Let us start with the reason why (18) only imposes an upper bound on the discounting function instead of assuming an exact formulation. We believe that any exact formulation is too restrictive and unrealistic in a social network setting. An upper bound, with two degrees of freedom in  $d$  and  $\beta$ , allows for a great deal of uncertainty and agents' variability in the model, which means the results derived based upon the PC dynamics remain credible in a practical setting. It also addresses the case where the discounting function also varies over time, that is if  $d_i$  is a function of  $t$  besides  $x_i$  and  $x_j$ .

With the exact discounting function approach ruled out, one wonders why upper-bounding is selected for approximating the discounting function among various possible non-exact formulations. The answer to that lies in the fact that the issue at hand is group polarization, which is reasonably expected to strengthen with the strength of the discount of opposing beliefs. Thus, if some group polarization result is valid for a given discounting function, a group polarization result at least as strong should hold for discounting functions with smaller values.

We now discuss the properties of the upper bound function in (18), that is  $1 - (1 - d)|x_i|^\beta$ . First of all, for neutral agents, i.e., when  $x_i \rightarrow 0$ , it returns 1, which allows for the continuity with respect to  $x_i$  of the broader  $d_i$  characterized via (15) and (18). This is a very important property to satisfy if  $d_i$  is to be realistic in any shape or form. Then, we focus on how  $d$  and  $\beta$ , earlier branded as the degrees of freedom in the upper bound function, are interpreted. We first notice that  $1 - (1 - d)|x_i|^\beta$  is non-decreasing in both  $d$  and  $\beta$ . The parameter  $d$  can be viewed as a uniform discount factor when  $\beta$  is small. It also serves as an upper bound for the discount value employed by the extremely confident individuals,

i.e., those with opinions close to 1 in absolute value. If  $d = 1$ , the PC model converts to the PC-lite model. The parameter  $\beta$  can be viewed as the discount's decay rate with respect to  $|x_i|$ . In other words, it captures the contribution of an individual's confidence to her discount value of opposing beliefs. The case where  $\beta \rightarrow \infty$  turns the PC dynamics to its PC-lite counterpart.

### 3.2 Properties of the PC Model

We now aim to investigate the behavior of the PC dynamics (14), with the discounting function  $d_i$  characterized through (15) and (18). Special cases of (18), corresponding to marginal values of  $d$  and  $\beta$ , are of particular interest to better understand the behavior of a general discounting function under conditions (15) and (18), and later a more general discounting function only restricted by (15). As discussed earlier, either case of  $d = 1$  and  $\beta \rightarrow \infty$  eliminates the confirmation bias and consequently simplifies the PC dynamics to the PC-lite dynamics, which was thoroughly investigated in Subsection 2.4. The marginal case  $d = 0$  will be later in Remark 2 argued to transpire the "ultimate" group polarization, where the opinions within a social bubble reach one of the very most extreme values  $\pm 1$ . The last marginal case, which will prove to be both interesting and informative, is that of  $\beta \rightarrow 0$ , that allows for an infinitely fast decay in the discount value with respect to  $|x_i|$  and in limit amounts to a uniform upper bound on the discount value of opposing beliefs, i.e.,

$$d_i(x_i, x_j) \leq d \text{ if } \text{sgn}(x_i) \neq \text{sgn}(x_j). \quad (19)$$

In this case, Proposition 2 can be generalized as follows.

**Proposition 3.** *Given the PC dynamics (14), with the discounting function  $d_i$  satisfying (15) and (19), let a connected subset  $\mathcal{B} \subseteq \mathcal{V}$  of agents have the bubble number*

$$\gamma_{\mathcal{B}} > (3 + 2\sqrt{2})d, \quad (20)$$

and assume that  $\alpha_1$  and  $\alpha_2$ , where  $\alpha_1 < \alpha_2$ , are the two positive solutions of the equation

$$\frac{1 + \alpha}{\alpha(1 - \alpha)} = \frac{\gamma_{\mathcal{B}}}{d}. \quad (21)$$

If, for some  $t_0$ , it happens that

$$x_i(t_0) > \alpha_1, \quad \forall i \in \mathcal{B}, \quad (22)$$

then we have

$$\liminf_{t \rightarrow \infty} x_i(t) \geq \alpha_2, \quad \forall i \in \mathcal{B}. \quad (23)$$

*Proof.* Proposition 3 will turn out to be a special case of Proposition 4, stated later on in this section (see Subsection 6.3).  $\square$

In Proposition 3, it is assumed that all agents share the same upper-bound function, i.e., for all  $i \in \mathcal{V}$  the upper-bound is considered as  $\hat{d}_i(|x_i|) = d$  with uniform  $d \in [0, 1]$ . Just like Proposition 2, Proposition 3 also implies that if all agents within a social bubble reach a certain level of advocacy  $+\alpha_1$  or disapproval  $-\alpha_1$  of a position, as the interactions continue, agents in that relaxed bubble will reach a more extreme level of advocacy  $+\alpha_2$  or disapproval  $-\alpha_2$  of that position. Therefore, opinions are intensified and become more extreme or, equivalently, the opinions become polarized within that group. Consequently, a group polarization will occur in the direction of support/disapproval of a specific position. One also notices that as  $d$  decreases,  $\alpha_1$  and  $\alpha_2$  will decrease and increase, respectively, as can be seen in

Fig. 4. Therefore, if the conditions of Proposition 3 are satisfied in a social bubble, the result will be that agents with relatively low initial levels of advocacy/disapproval on a position will later have relatively extreme levels of advocacy/disapproval on that position.

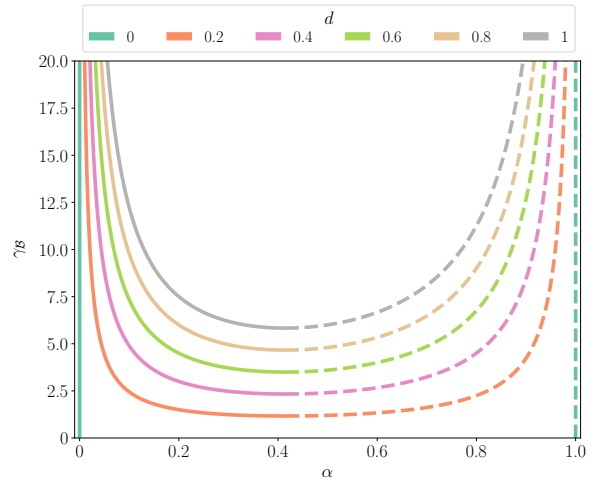


Fig. 4. Changes of  $\alpha_1$  and  $\alpha_2$  with respect to  $\gamma_{\mathcal{B}}$  and  $d$ . Solid and dashed lines represent  $\alpha_1$  and  $\alpha_2$  respectively. Each curve corresponds to a different value for  $d = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ .

For the purpose of completeness, we generalize Proposition 3 to the following proposition, which addresses group polarization under the PC dynamics for general  $d$  and  $\beta$ .

**Proposition 4.** *Given the PC dynamics (14), with the discounting function  $d_i$  satisfying (15) and (18), let a connected subset  $\mathcal{B} \subseteq \mathcal{V}$  of agents have the bubble number  $\gamma_{\mathcal{B}}$ . Assume that equation*

$$\frac{1 + \alpha}{\alpha(1 - \alpha)} = \frac{\gamma_{\mathcal{B}}}{1 - (1 - d)\alpha^\beta} \quad (24)$$

has two positive solutions for  $\alpha \in (0, 1)$ , namely  $\alpha_1$  and  $\alpha_2$ , where  $\alpha_1 \leq \alpha_2$ . If, for some  $t_0$ , it happens that

$$x_i(t_0) > \alpha_1, \quad \forall i \in \mathcal{B}, \quad (25)$$

then we have

$$\liminf_{t \rightarrow \infty} x_i(t) \geq \alpha_2, \quad \forall i \in \mathcal{B}. \quad (26)$$

*Proof.* Proposition 4 is a special case of Theorem 1, stated later on in this section, which is proved in Section 6.  $\square$

In Proposition 4, it is assumed that all agents share the same upper-bound function, i.e., for all  $i \in \mathcal{V}$  the upper-bound is considered as  $\hat{d}_i(|x_i|) = 1 - (1 - d)|x_i|^\beta$  with uniform  $d \in [0, 1]$  and  $\beta$ . The statement of Proposition 4 is different from those of Propositions 2 and 3 in that a succinct condition, such as (8) and (20), under which (24) is guaranteed to have solutions has not been provided. However, given any  $d$  and  $\beta$ , it is straightforward to verify whether such solutions exist. We should also point out that a larger  $\gamma_{\mathcal{B}}$ , smaller  $d$ , and smaller  $\beta$ , all work in favor of (24) having solutions for  $\alpha$ . The assumption of having two positive solutions for equation (24) for big enough  $\gamma_{\mathcal{B}}$  is reasonable. Since the left-hand side is a concave-parabolic-like function, confined in  $[0, 1]$  and has its minimum inside  $(0, 1)$ . While the right-hand side is a non-decreasing function within  $(0, 1)$ . Therefore, inside the interval  $[0, 1]$ , for big enough  $\gamma_{\mathcal{B}}$ , these two curves will coincide at two points. In addition, a lower bound of the bubble number in equation (24) can be obtained by finding the minimum of  $[(1 - (1 - d)\alpha^\beta)(1 + \alpha)] / [\alpha(1 - \alpha)]$  over  $\alpha \in (0, 1)$ .

*Remark 2.* In view of Proposition 4, the marginal case  $d = 0$  can be interpreted to represent the “ultimate” group polarization. More precisely, given  $d = 0$ , (24) will have two positive solutions,  $\alpha_1 < 1$  and  $\alpha_2 = 1$ , with the latter solution indicating the convergence of the opinions within the bubble to one of the very most extreme values  $+1$  or  $-1$ .

The two degrees of freedom incorporated in the discounting function (17) can indicate sensitivity to an issue being discussed among the individuals. The higher the sensitivity to the issue for a specific population, the more intense the population acts in a biased manner towards it (smaller  $d$  or  $\beta$ ), the more probable/intense the polarization of opinions on that issue.

Finally, Proposition 4 can be extended as follows to any discounting function restricted to conditions (15) and (16) for a non-increasing function  $\hat{d}_i$ .

**Theorem 1.** *Given the PC dynamics (14), with the discounting function  $d_i$  satisfying (15) as well as (16) for a non-increasing  $\hat{d}_i : [0, 1] \rightarrow [0, 1]$ , let a connected subset  $\mathcal{B} \subseteq \mathcal{V}$  of agents have the bubble number  $\gamma_{\mathcal{B}}$ . Assume that inequality*

$$\frac{1 + \alpha}{\alpha(1 - \alpha)} < \frac{\gamma_{\mathcal{B}}}{\hat{d}_i(\alpha)} \quad (27)$$

is satisfied for any  $i \in \mathcal{B}$  and  $\alpha \in (\alpha_1, \alpha_2) \subseteq (0, 1)$ . If, for some  $t_0$ , it happens that

$$x_i(t_0) > \alpha_1, \quad \forall i \in \mathcal{B}, \quad (28)$$

then we have

$$\liminf_{t \rightarrow \infty} x_i(t) \geq \alpha_2, \quad \forall i \in \mathcal{B}. \quad (29)$$

*Proof.* The proof of Theorem 1 is given in Subsection 6.2.  $\square$

In Theorem 1, the discounting function  $d_i(x_i, x_j)$  for each agent  $i$  is assumed to be arbitrary. However, a corresponding upper-bound  $\hat{d}_i(|x_i|)$  is considered for each  $d_i(x_i, x_j)$ . Let us denote the solution to inequality (27) with upper-bound  $\hat{d}_i(|x_i|)$  by the interval  $I_i = (a_i, b_i)$ , and let  $\mathcal{I} = \cap_{i \in \mathcal{B}} I_i = (\alpha_1, \alpha_2)$  be the intersection among the  $|\mathcal{B}|$  solutions. Equations (28) and (29) are stated according to  $\mathcal{I} = (\alpha_1, \alpha_2)$ . Furthermore, if there are separate multiple intervals as the solutions for inequality (27), the results are valid only within those intervals separately and not in between them. In addition, as the discounting functions become more intense, smaller values of agreement within a bubble will lead to higher levels of confidence and group polarization. In other words, for discounting functions  $\hat{d} \geq \hat{d}'$ , it is easy to see that for the same value of  $\gamma_{\mathcal{B}}$ , the results are such that  $(\alpha_1, \alpha_2) \subseteq (\alpha'_1, \alpha'_2)$ , where the first and second interval corresponds to the solutions obtained from discounting functions  $\hat{d}$  and  $\hat{d}'$ , respectively.

Furthermore, one may be interested in finding the lower bound for the bubble number in (27) for having positive solutions. We point out that once the discounting function is given, it is straightforward to find the minimum of  $\gamma_{\mathcal{B}}$  in which (27) can have solutions. More precisely, a lower bound for  $\gamma_{\mathcal{B}}$  can be obtained by finding the minimum of  $[\hat{d}_i(\alpha)(1 + \alpha)] / [\alpha(1 - \alpha)]$  over  $\alpha \in (0, 1)$ .

## 4 NUMERICAL EXPERIMENTS

In this section, we illustrate the behaviors of PC-lite dynamics (1) and PC dynamics (14) through numerical examples. In all examples, a fixed Erdős–Rényi random graph [59] embodies the underlying graph of the network. It consists of  $|\mathcal{V}| = 500$  nodes and, for each pair of nodes  $i, j \in \mathcal{V}$ , an edge  $e_{ij}$  exists with the uniform probability  $p_G = 0.06$ , independently of other edges. Each edge is then

independently activated at any time step with the uniform probability  $p_L = 0.8$ , which results in asynchronous interactions in the network. Numerical examples of the PC-lite model and the PC model are provided in Subsections 4.1 and 4.2 below, respectively. For subsections, the initial opinions of the agents within each bubble are selected according to a normal probability distribution function with near-zero mean and low variance, with  $\mu$  and  $\sigma^2$  representing the corresponding mean and variance, truncated to the range  $[-1, 1]$ . Moreover, the median of each bubble is denoted by  $Med(\cdot)$ .

### 4.1 Numerical Examples for the PC-lite Model

For the numerical examples of PC-lite dynamics (1), we consider two cases, (i) a case where there are three bubbles within the population, while some agents do not belong to any of these bubbles, and (ii) a case where the entire population is divided into two bubbles. Other parameters used in the simulations of these two cases, including the size of the bubbles, their bubble numbers, and their respective  $\alpha_1$  and  $\alpha_2$  values, are given in Tables 1 and 2. The initial opinions of the agents outside the bubbles in the first case are selected according to a normal probability distribution function with zero mean and variance equal to 0.11. Finally, we note that the weight values at any time step are generated randomly but scaled in such a way not to violate the bubble numbers listed in Tables 1 and 2. More precisely, given an agent inside a bubble, the outside weights influencing over her are uniformly scaled down.

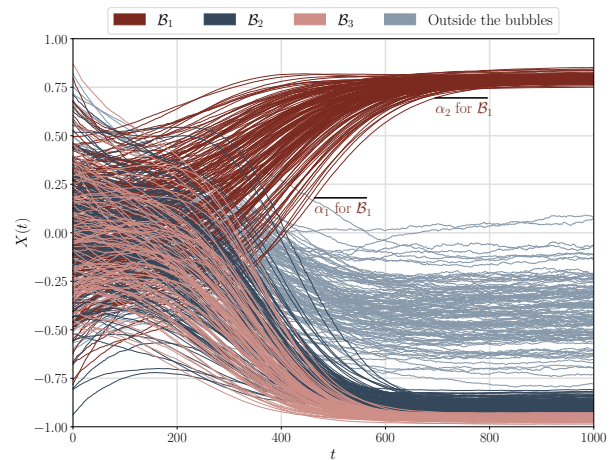


Fig. 5. Opinion evolution according to PC-lite dynamics (1) with parameters specified in Table 1.

Fig. 5 demonstrates the evolution of opinions under the PC-lite dynamics for the first case. It confirms the statement of Proposition 2 that if the opinions of the agents in Bubble 1 become greater than  $\alpha_1$  in finite time, they will in the long run become more extreme than  $\alpha_2$ . The same conclusion can be drawn for Bubble 2 and Bubble 3. Furthermore, Bubble 3 appears to show a stronger group polarization than Bubble 1 and Bubble 2, which is consistent with it having a larger  $\alpha_2$  than the other bubbles, that in view of Fig. 3 is a result of its relatively large bubble number.

Fig. 6 shows the evolution of opinions under the PC-lite dynamics for the second case. While it confirms the statement of Proposition 2 like the previous simulation, it is designed to resemble the opinion polarization of the US population depicted in Fig. 1. More specifically, it shows how the medians of the opinions

TABLE 1  
Parameters of the PC-lite model simulation (Fig. 5)

Parameter	Bubble 1	Bubble 2	Bubble 3
$ \mathcal{B} $	159	127	90
$\gamma_{\mathcal{B}}$	8	6	12
$\alpha_1$	0.18	0.333	0.102
$\alpha_2$	0.695	0.5	0.814
$\mu$	0.04	-0.02	-0.01
$\sigma^2$	0.1	0.09	0.08

in the two bubbles diverge over time. The distribution of opinions at time steps 0 and 200 are separately drawn in Fig. 7, bearing a resemblance to Fig. 1.

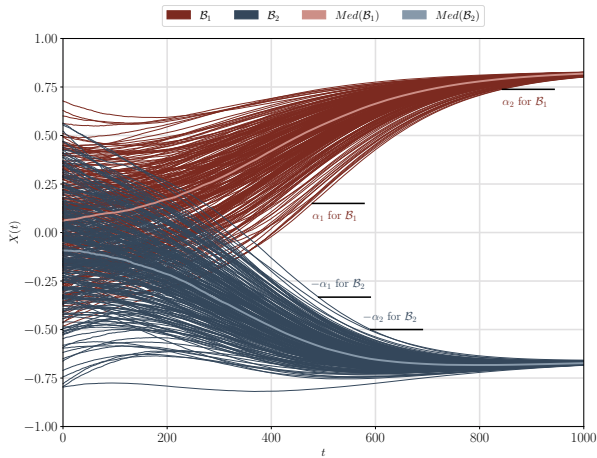


Fig. 6. Opinion evolution according to PC-lite dynamics (1) with parameters specified in Table 2.

TABLE 2  
Parameters of the PC-lite model simulation (Fig. 6)

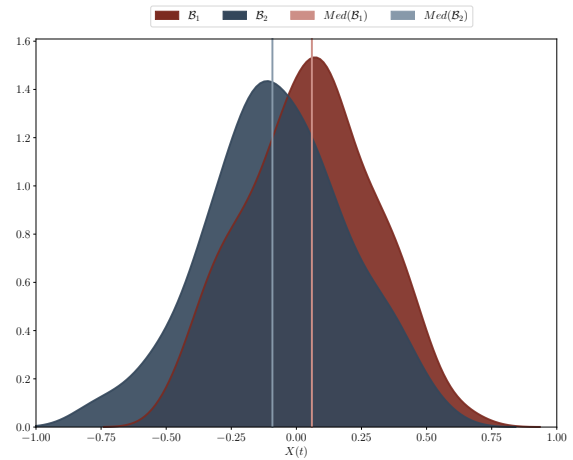
Parameters	Bubble 1	Bubble 2
$ \mathcal{B} $	252	248
$\gamma_{\mathcal{B}}$	9	6
$\alpha_1$	0.15	0.333
$\alpha_2$	0.738	0.5
$\mu$	0.06	-0.08
$\sigma^2$	0.25	0.3

## 4.2 Numerical Examples for the PC Model

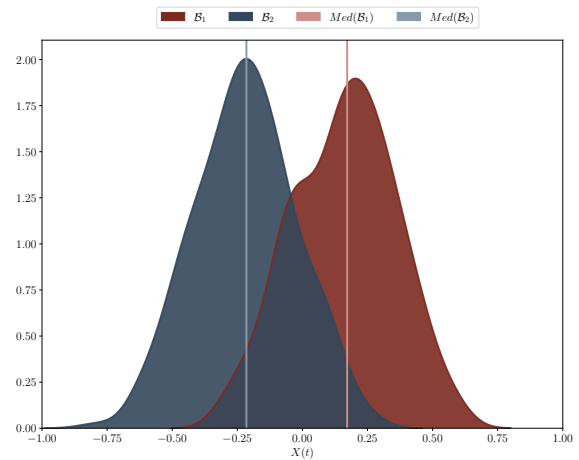
For the numerical examples of PC dynamics (14), we consider a case where the network is divided into two bubbles, the parameters of which are given in Table 3. Fig. 8 demonstrates the evolution of opinions under the PC dynamics (14) where

$$d_i(x_i, x_j) = \begin{cases} 1 & \text{if } \text{sgn}(x_i) = \text{sgn}(x_j) \\ 1 - (1 - d)|x_i|^\beta & \text{if } \text{sgn}(x_i) \neq \text{sgn}(x_j) \end{cases} \quad (30)$$

with  $d = 0.4$  and  $\beta = 0.4$ . One can observe in Fig. 8 that all the opinions of agents in the bubbles will asymptotically exceed their respective  $\alpha_2$ 's in magnitude, confirming Proposition 3.



(a)



(b)

Fig. 7. Opinion distribution in each bubble appearing in Fig. 6 at (a)  $t = 0$  and (b)  $t = 200$ .

TABLE 3  
Parameters of the PC model simulation (Fig. 8)

Parameters	Bubble 1	Bubble 2
$ \mathcal{B} $	247	253
$\gamma_{\mathcal{B}}$	3.5	4
$\alpha_1$	0.3721	0.2869
$\alpha_2$	0.6287	0.7149
$\mu$	0.09	-0.08
$\sigma^2$	0.1	0.11

## 5 DISCUSSION AND CONCLUSION

In the following, we first discuss the issue of stubborn agents, compare this concept with the viewpoint of partisan confidence and end it with concluding remarks and some directions for future research.

The inclusion of stubborn agents and their role has been the issue for many pieces of research [60]. There is no consensus on the terminology of a stubborn agent within the literature, i.e., other terms frequently used are opinion leader, media source, inflexible or close-minded agent. Nevertheless, all refer to an agent



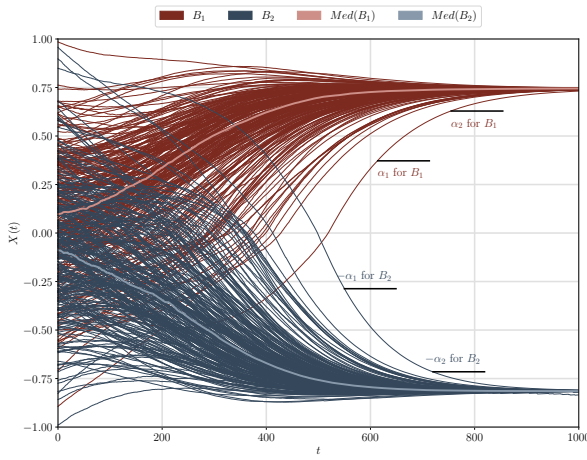


Fig. 8. Opinion evolution according to PC dynamics (14) with parameters specified in Table 3.

who does not change her position due to peers' influence.

There are multiple realizations for stubbornness. The simplest one is removing all influences from the outside, i.e.  $w_{ij} = 0, \forall j \neq i$  for a stubborn agent  $i$ , which is widely used in DeGrootian models [61], [62]. A remarkable realization is the concept of adherence to an agent's initial position used in Friedkin-Johnsen's (FJ) model [11], and its variations [15], [63]–[65]. Specifically, a complete stubborn agent  $\lambda = 1$  never forgets her initial position, whereas an oblivious agent  $\lambda = 0$  tends to conform and completely forget her initially held position.

Another realization is the use of confidence levels in bounded-confidence models [66], [67], in which the zero confidence level  $\epsilon = 0$  implies a completely closed-minded agent. In addition, the concept of zero confidence, in fact, also implies the first type of stubbornness, which is stated as  $w_{ij} = 0, \forall j \neq i$ .

The rule of thumb is that stubborn agents can guide and control the opinion of others if they have a great deal of influence on non-stubborn agents. Various outcomes, such as consensus, disagreement, and polarization, might happen depending on the number of stubborn agents and their initial voice and the topology of the network [61], [62], [65], [68]. For instance, consensus and polarization might occur if there is only one stubborn agent (or several stubborn agents with the same voice) in the network, the influence of stubborn(s) agents is significant, and the network of non-stubborn agents is connected. Interestingly, by observing the asymptotic behavior of individuals and having knowledge of stubborn agents, the system (influence matrix) is tractable [69]. Disagreement may occur if there are several stubborn agents with a connected network of non-stubborn agents [70].

Getting back to our proposed model, we note that the notion of partisan confidence is different from the stubbornness concept. More specifically, the opinions of individuals are not fixed during the evolution of opinions, and actually change with respect to time. Moreover, agents tend to increase their confidence in either political affiliations if the conditions are prepared. Note that complete stubbornness does not fit into the definition of a social bubble due to connectivity. However, regardless of the definition, the increase in confidence in either direction is not guaranteed if the confidence of stubborn agents within a bubble (not all members) is not greater than  $\pm\alpha_2$ . In addition, the concept of

discounting function can be interpreted as a flexible degree of close-mindedness in an individual's political affiliation.

In this paper, we proposed the PC-lite and PC models of opinion dynamics based on an approach that views an opinion via its intensity and direction. We established a result on the occurrence of opinion polarization in a social bubble, referring to a group of individuals who are highly cohesive and isolated from outside influence. Both of the models developed are inspired by the notion of groupthink, widely studied in the socio-psychological literature. We also justified our results using numerical simulations.

The ultimate goal of the proposed models is to analyze, predict, and possibly intervene in the process of group polarization and opinion polarization in a population. While the analysis and prediction goals were discussed in this work, the intervention techniques will remain as part of future work. As another future research direction, it will be interesting to study the multidimensional version of the proposed models that captures the simultaneous evolution of opinions on a multitude of correlated topics.

## 6 PROOFS

This section is composed of a subsection containing some preliminary definitions and lemmas that are integral in the proof of Theorem 1, an entire subsection detailing the proof of Theorem 1, and another subsection on the derivations of Propositions 1, 2, 3, and 4 from Theorem 1.

### 6.1 Preliminaries to the Proof of Theorem 1

We recall that the discounting function in Theorem 1 is assumed to satisfy (15) and (16) for non-increasing functions  $\hat{d}_i$ . Since the marginal case of (16) will prove to be of great importance, we define an auxiliary discounting function  $d' : [-1, 1]^2 \rightarrow [0, 1]$  by

$$d'_i(x_i, x_j) = 1 \text{ if } \text{sgn}(x_i) = \text{sgn}(x_j), \quad (31)$$

$$d'_i(x_i, x_j) = \hat{d}_i(|x_i|) \text{ if } \text{sgn}(x_i) \neq \text{sgn}(x_j), \quad (32)$$

where (31) is identical to (15), while (32) is the marginal case of (16). For any  $y \in [-1, 1]^n$ , it should be clear that

$$d_i(y_i, y_j) \leq d'_i(y_i, y_j), \forall i, j. \quad (33)$$

We also define a function  $f : [-1, 1]^n \rightarrow [-1, 1]^n$  with its  $i$ th coordinate formulated as

$$f_i(y) \triangleq y_i + \sum_{j \neq i} \left[ d'_i(y_i, y_j) w_{ij} |y_j| (\text{sgn}(y_j) - y_i) \right], \quad (34)$$

There is a slight abuse of notation in (34), in that  $w_{ij}$  is in general a function of time, while  $f$  does not seem to be treated as one. This will not cause a problem since the time index will be fixed whenever  $f_i$  will show up in future arguments. In particular,  $f_i(x(t))$  can be expressed as

$$f_i(x(t)) = x_i(t) + \sum_{j \neq i} \left[ d'_i(x_i(t), x_j(t)) w_{ij}(t) |x_j(t)| (\text{sgn}(x_j(t)) - x_i(t)) \right]. \quad (35)$$

In contrast, according to (14),

$$x_i(t+1) = x_i(t) + \sum_{j \neq i} \left[ d_i(x_i(t), x_j(t)) w_{ij}(t) |x_j(t)| (\text{sgn}(x_j(t)) - x_i(t)) \right]. \quad (36)$$

The following lemmas will be used in the proof of Theorem 1.

**Lemma 1.** For an arbitrary agent  $i$ , if  $x_i(t) \geq 0$ , then  $x_i(t+1) \geq f_i(x(t))$ .

*Proof.* From (35) and (36),

$$\begin{aligned} x_i(t+1) - f_i(x(t)) &= \\ &= \sum_{j \neq i} \left[ (d_i(x_i(t), x_j(t)) - d'_i(x_i(t), x_j(t))) \times \right. \\ &\quad \left. w_{ij}(t) |x_j(t)| (\text{sgn}(x_j(t)) - x_i(t)) \right]. \end{aligned} \quad (37)$$

To complete the proof, it is sufficient to show that each summand (term appearing in a summation) in (37) is non-negative, which can be done simply by considering the two cases for  $\text{sgn}(x_j(t))$ . If  $\text{sgn}(x_j(t)) = +1$ , given the assumption  $x_i(t) > 0$ , the summand becomes zero since both  $d_i$  and  $d'_i$  would equal 1. If  $\text{sgn}(x_j(t)) = -1$ , then from (33), the summand is non-negative.  $\square$

**Lemma 2.** The function  $f$  is non-decreasing, i.e., for any pair of vectors  $y^1, y^2 \in [-1, 1]^n$ ,

$$y^1 \leq y^2 \Rightarrow f(y^1) \leq f(y^2), \quad (38)$$

where the inequalities in (38) are to be understood element-wise.

*Proof.* Without loss of generality, we can assume that  $y^1$  and  $y^2$  differ only in one coordinate, say the  $k$ th coordinate. For notational convenience, we may write  $y_i$  for both  $y_i^1$  and  $y_i^2$ , when  $i \neq k$ . We show that  $f_i(y^1) \leq f_i(y^2)$  for each  $i$  by considering the following cases:

**Case 1:** ( $i \neq k$ ) In this case,

$$\begin{aligned} &f_i(y^2) - f_i(y^1) \\ &= \left( y_i + \sum_{j \neq i} \left[ d'_i(y_i, y_j^2) w_{ij} |y_j^2| (\text{sgn}(y_j^2) - y_i) \right] \right) \\ &\quad - \left( y_i + \sum_{j \neq i} \left[ d'_i(y_i, y_j^1) w_{ij} |y_j^1| (\text{sgn}(y_j^1) - y_i) \right] \right) \\ &= w_{ik} d'_i(y_i, y_k^2) |y_k^2| (\text{sgn}(y_k^2) - y_i) \\ &\quad - w_{ik} d'_i(y_i, y_k^1) |y_k^1| (\text{sgn}(y_k^1) - y_i) \end{aligned} \quad (39)$$

Now, if  $\text{sgn}(y_k^1) = \text{sgn}(y_k^2)$ , then from (15), (31), and (32) we have  $d'_i(y_i, y_k^1) = d'_i(y_i, y_k^2)$ , which combined with (39) leads to

$$\begin{aligned} &f_i(y^2) - f_i(y^1) \\ &= w_{ik} d'_i(y_i, y_k^1) \left( |y_k^2| (\text{sgn}(y_k^2) - y_i) - |y_k^1| (\text{sgn}(y_k^1) - y_i) \right) \\ &= w_{ik} d'_i(y_i, y_k^1) \left( (y_k^2 - y_k^1) - y_i (|y_k^2| - |y_k^1|) \right) \\ &\geq w_{ik} d'_i(y_i, y_k^1) (1 - |y_i|) (y_k^2 - y_k^1) \\ &\geq 0, \end{aligned} \quad (40)$$

where in the first inequality of (40), we used the inequality  $||y_k^2| - |y_k^1|| \leq |y_k^2 - y_k^1| = y_k^2 - y_k^1$ . On the other hand, if  $\text{sgn}(y_k^1) \neq \text{sgn}(y_k^2)$ , given  $y_k^2 \geq y_k^1$ , we have  $\text{sgn}(y_k^1) = -1$  and  $\text{sgn}(y_k^2) = 1$ , which considered together with (39), immediately results in  $f_i(y^2) - f_i(y^1) \geq 0$ .

**Case 2:** ( $i = k$ ) In this case,

$$\begin{aligned} &f_i(y^2) - f_i(y^1) \\ &= \left( y_k^2 + \sum_{j \neq k} \left[ d'_k(y_k^2, y_j) w_{kj} |y_j| (\text{sgn}(y_j) - y_k^2) \right] \right) \end{aligned}$$

$$\begin{aligned} &- \left( y_k^1 + \sum_{j \neq k} \left[ d'_k(y_k^1, y_j) w_{kj} |y_j| (\text{sgn}(y_j) - y_k^1) \right] \right) \\ &= (y_k^2 - y_k^1) + \sum_{j \neq k} \left[ w_{kj} |y_j| d'_k(y_k^2, y_j) (\text{sgn}(y_j) - y_k^2) \right] \\ &\quad - \sum_{j \neq k} \left[ w_{kj} |y_j| d'_k(y_k^1, y_j) (\text{sgn}(y_j) - y_k^1) \right] \\ &= (y_k^2 - y_k^1) + \sum_{j \neq k} \left[ w_{kj} |y_j| d'_k(y_k^1, y_j) (y_k^1 - y_k^2) \right] \\ &\quad + \sum_{j \neq k} \left[ w_{kj} |y_j| (d'_k(y_k^2, y_j) - d'_k(y_k^1, y_j)) (\text{sgn}(y_j) - y_k^2) \right]. \end{aligned} \quad (41)$$

Considering the two cases  $\pm 1$  for  $\text{sgn}(y_j)$ , and remembering  $y_k^2 \geq y_k^1$  as well as (15), (31), and (32), the term

$$(d'_k(y_k^2, y_j) - d'_k(y_k^1, y_j)) (\text{sgn}(y_j) - y_k^2) \quad (42)$$

which appears in the last line of (41), can be easily shown to be zero or positive. Hence, (41) results in

$$\begin{aligned} &f_i(y^2) - f_i(y^1) \geq (y_k^2 - y_k^1) \\ &\quad + \sum_{j \neq k} \left[ w_{kj} |y_j| \left( d'_k(y_k^1, y_j) (y_k^1 - y_k^2) \right) \right] \\ &= (y_k^2 - y_k^1) \left( 1 - \sum_{j \neq k} \left[ w_{kj} |y_j| d'_k(y_k^1, y_j) \right] \right) \\ &\geq (y_k^2 - y_k^1) \left( 1 - \sum_{j \neq k} w_{kj} \right) \\ &\geq 0. \end{aligned} \quad (43)$$

$\square$

## 6.2 Proof of Theorem 1

Defining  $z(t) = \min_{i \in \mathcal{B}} x_i(t)$ , the inequality (29) holds for each  $i \in \mathcal{B}$  if and only if

$$\liminf_{t \rightarrow \infty} z(t) \geq \alpha_2. \quad (44)$$

To show (44), we first note that, by the assumption of the Theorem 1, that is (28), we have

$$z(t_0) > \alpha_1. \quad (45)$$

Given an arbitrary but fixed  $t$ , we then take the following six steps to fully examine  $\liminf_{t \rightarrow \infty} z(t)$  and show (44).

**Step 1:** We show that the following statement is true:

$$(\alpha_1 < z(t) < \alpha_2) \implies (z(t+1) \geq z(t)). \quad (46)$$

To this aim, construct a vector  $y$  from  $x(t)$  as

$$y_i = \begin{cases} z(t) & \text{if } i \in \mathcal{B} \\ -1 & \text{if } i \notin \mathcal{B}. \end{cases} \quad (47)$$

It should be clear that  $x(t) \geq y$ . Thus, from Lemma 2, we must have  $f(x(t)) \geq f(y)$ , and in particular,  $f_i(x(t)) \geq f_i(y), \forall i \in \mathcal{B}$ . On the other hand, since  $x_i(t) \geq 0$  for  $i \in \mathcal{B}$ , from Lemma 1, we conclude that  $x_i(t+1) \geq f_i(x(t))$ . Hence,  $x_i(t+1) \geq f_i(y)$ . Therefore, for each  $i \in \mathcal{B}$ ,

$$x_i(t+1) \geq f_i(y)$$

$$\begin{aligned}
&= y_i + \sum_{j \in \mathcal{V}} d'_i(y_i, y_j) w_{ij}(t) |y_j| (\text{sgn}(y_j) - y_i) \\
&= y_i + \sum_{j \in \mathcal{B}} d'_i(y_i, y_j) w_{ij}(t) |y_j| (\text{sgn}(y_j) - y_i) \\
&+ \sum_{j \in \mathcal{V} \setminus \mathcal{B}} d'_i(y_i, y_j) w_{ij}(t) |y_j| (\text{sgn}(y_j) - y_i) \\
&= z(t) + \sum_{j \in \mathcal{B}} w_{ij}(t) z(t) (1 - z(t)) \\
&+ \sum_{j \in \mathcal{V} \setminus \mathcal{B}} \hat{d}_i(z(t)) w_{ij}(t) (-1 - z(t)) \\
&= z(t) + z(t)(1 - z(t)) \times \\
&\left[ \sum_{j \in \mathcal{B}} w_{ij}(t) - \frac{\hat{d}_i(z(t))(1 + z(t))}{z(t)(1 - z(t))} \sum_{j \in \mathcal{V} \setminus \mathcal{B}} w_{ij}(t) \right]. \tag{48}
\end{aligned}$$

Furthermore, since  $\alpha_1 < z(t) < \alpha_2$ , from the assumption (27) we have

$$\frac{\hat{d}_i(z(t))(1 + z(t))}{z(t)(1 - z(t))} < \gamma_{\mathcal{B}}. \tag{49}$$

Combining (48) and (49) implies that

$$\begin{aligned}
&x_i(t+1) \geq z(t) \\
&+ z(t)(1 - z(t)) \left[ \sum_{j \in \mathcal{B}} w_{ij}(t) - \gamma_{\mathcal{B}} \sum_{j \in \mathcal{V} \setminus \mathcal{B}} w_{ij}(t) \right] \\
&\geq z(t). \tag{50}
\end{aligned}$$

Consequently,  $z(t+1) \geq z(t)$ , which completes the proof of statement (46).

**Step 2:** We show that

$$(z(t) \geq \alpha_2) \Rightarrow (z(t+1) \geq \alpha_2). \tag{51}$$

To prove statement (51), we first point out that since (27) is assumed to hold for any  $i$  and  $\alpha \in (\alpha_1, \alpha_2)$ , one can write

$$\begin{aligned}
\hat{d}_i(\alpha_2) &\leq \lim_{\alpha \rightarrow \alpha_2} \hat{d}_i(\alpha) \leq \lim_{\alpha \rightarrow \alpha_2} \frac{\alpha(1 - \alpha)\gamma_{\mathcal{B}}}{1 + \alpha} \\
&= \frac{\alpha_2(1 - \alpha_2)\gamma_{\mathcal{B}}}{1 + \alpha_2}, \tag{52}
\end{aligned}$$

where the first inequality of (52), as well as the existence of  $\lim_{\alpha \rightarrow \alpha_2} \hat{d}_i(\alpha)$ , are both immediate results of  $\hat{d}_i$  being non-increasing, while the second inequality of (52) is a direct consequence of (27). Rewriting (52), we have

$$\frac{1 + \alpha_2}{\alpha_2(1 - \alpha_2)} \leq \frac{\gamma_{\mathcal{B}}}{\hat{d}_i(\alpha_2)} \tag{53}$$

Now, we follow the same line of arguments as in Step 1, only here we construct the vector  $y$  as

$$y_i = \begin{cases} \alpha_2 & \text{if } i \in \mathcal{B} \\ -1 & \text{if } i \notin \mathcal{B}, \end{cases} \tag{54}$$

and replace  $z(t)$  in (48) by  $\alpha_2$ . Then, we write (49) for  $\alpha_2$  in place of  $z(t)$  by employing (53) in place of (27). These modifications of (48) and (49) together imply  $x_i(t+1) \geq \alpha_2$ ,  $\forall i \in \mathcal{B}$ , which proves statement (51).

**Step 3:** Combining Steps 1 and 2, from (45), (46) and (51), we conclude that (44) holds unless  $z(t)$  is non-decreasing for every  $t \geq t_0$  and  $\lim_{t \rightarrow \infty} z(t)$  exists and lies in the interval  $(\alpha_1, \alpha_2)$ . Thus, assume on the contrary that  $z(t)$  is non-decreasing for  $t \geq t_0$  and

$$\lim_{t \rightarrow \infty} z(t) = z^* \in (\alpha_1, \alpha_2). \tag{55}$$

Let  $\epsilon > 0$  be sufficiently small that it satisfies

$$\gamma_{\mathcal{B}} > \frac{\hat{d}_i(z^* - \epsilon)(1 + z^* + \epsilon)}{(z^* - \epsilon)(1 - (z^* + \epsilon))}, \tag{56}$$

for any  $i \in \mathcal{V}$ . One notices that (56) holds for any sufficiently small  $\epsilon$  since, as  $\epsilon$  vanishes, the right-hand expression of (56) converges to  $\hat{d}_i(z^*)(1 + z^*)/[z^*(1 - z^*)]$ , which is less than  $\gamma_{\mathcal{B}}$ . According to (55), there exists  $T > t_0$  such that

$$z(t) > z^* - \epsilon, \forall t > T. \tag{57}$$

**Step 4:** Let  $i \in \mathcal{B}$  be arbitrary. We show that there is a time instant  $t_i > T$  such that

$$x_i(t_i) > z^* + \epsilon. \tag{58}$$

Assume to the contrary that  $x_i(t)$  never exceeds  $z^* + \epsilon$  after time  $T$ , that is  $x_i(t) \leq z^* + \epsilon$  for any  $t > T$ . Now, for any  $t > T$ , given the PC dynamics we have

$$\begin{aligned}
&x_i(t+1) - x_i(t) \\
&= \sum_{j \in \mathcal{V}} d_i(x_i(t), x_j(t)) w_{ij}(t) |x_j(t)| (\text{sgn}(x_j(t)) - x_i(t)) \\
&\geq \left( \sum_{j \in \mathcal{B}} w_{ij}(t) \right) (z^* - \epsilon)(1 - (z^* + \epsilon)) \\
&+ \hat{d}_i(z^* - \epsilon) \left( \sum_{j \in \mathcal{V} \setminus \mathcal{B}} w_{ij}(t) \right) (-1 - (z^* + \epsilon)) \\
&\geq \left( \sum_{j \in \mathcal{B}} w_{ij}(t) \right) \times \\
&\left[ (z^* - \epsilon)(1 - (z^* + \epsilon)) + \frac{\hat{d}_i(z^* - \epsilon)}{\gamma_{\mathcal{B}}} (-1 - (z^* + \epsilon)) \right]. \tag{59}
\end{aligned}$$

Summing up (59) over consecutive time instants, we conclude that

$$\begin{aligned}
&x_i(t') - x_i(t) \geq \left( \sum_{\tau=t}^{t'-1} \sum_{j \in \mathcal{B}} w_{ij}(\tau) \right) \\
&\left[ (z^* - \epsilon)(1 - (z^* + \epsilon)) + \frac{\hat{d}_i(z^* - \epsilon)}{\gamma_{\mathcal{B}}} (-1 - (z^* + \epsilon)) \right]. \tag{60}
\end{aligned}$$

The right-hand expression in (60) explodes as  $t'$  grows since

$$\left[ (z^* - \epsilon)(1 - (z^* + \epsilon)) + \frac{\hat{d}_i(z^* - \epsilon)}{\gamma_{\mathcal{B}}} (-1 - (z^* + \epsilon)) \right] \tag{61}$$

is lower-bounded by a positive number according to (56) and

$$\sum_{\tau=t}^{t'-1} \sum_{j \in \mathcal{B}} w_{ij}(\tau) \tag{62}$$

grows unbounded as  $t' \rightarrow \infty$  since  $\mathcal{B}$  is connected. This is a contradiction, meaning that there is  $t_i > T$  for which (58) holds.

**Step 5:** Let  $i \in \mathcal{B}$  be arbitrary. We show that if  $t > T$ ,

$$(x_i(t) \geq z^* + \epsilon) \implies (x_i(t+1) \geq z^* + \epsilon). \quad (63)$$

According to Lemma 1,  $x_i(t+1) \geq f_i(x(t))$ . For the arbitrary but fixed  $i$ , we construct the vector  $y$  as

$$y_j = \begin{cases} z^* + \epsilon & \text{if } j = i \\ z^* - \epsilon & \text{if } j \in \mathcal{B}, j \neq i \\ -1 & \text{if } i \notin \mathcal{B}. \end{cases} \quad (64)$$

Since  $x(t) \geq y$  and  $f$  is non-decreasing according to Lemma 2,  $f_i(x(t)) \geq f_i(y)$ , and consequently,  $x_i(t+1) \geq f_i(y)$ . Thus, it is sufficient to show that  $f_i(y) \geq z^* + \epsilon$ . Hence, we write

$$\begin{aligned} f_i(y) &= y_i + \sum_{j \in \mathcal{V}} d'_i(y_i, y_j) w_{ij}(t) |y_j| (\text{sgn}(y_j) - y_i) \\ &= z^* + \epsilon + \left( \sum_{j \in \mathcal{B}} w_{ij}(t) \right) (z^* - \epsilon)(1 - (z^* + \epsilon)) \\ &\quad + \hat{d}_i(z^* + \epsilon) \left( \sum_{j \in \mathcal{V} \setminus \mathcal{B}} w_{ij}(t) \right) (-1 - (z^* + \epsilon)) \\ &\geq z^* + \epsilon + \left( \sum_{j \in \mathcal{B}} w_{ij}(t) \right) \times \\ &\quad \left[ (z^* - \epsilon)(1 - (z^* + \epsilon)) + \frac{\hat{d}_i(z^* + \epsilon)}{\gamma_{\mathcal{B}}} (-1 - (z^* + \epsilon)) \right] \\ &\geq z^* + \epsilon, \end{aligned} \quad (65)$$

where the last inequality in (65) is a result of

$$\left[ (z^* - \epsilon)(1 - (z^* + \epsilon)) + \frac{\hat{d}_i(z^* + \epsilon)}{\gamma_{\mathcal{B}}} (-1 - (z^* + \epsilon)) \right] > 0, \quad (66)$$

which itself is implied from (56) considering  $\hat{d}_i(z^* + \epsilon) \leq \hat{d}_i(z^* - \epsilon)$  according to Lemma 2.

**Step 6:** Combining Steps 4 and 5, we conclude that for each  $i \in \mathcal{B}$ , there is a time  $t_i$  such that  $x_i(t) \geq z^* + \epsilon$  for any  $t \geq t_i$ . Hence,  $z(t) \geq z^* + \epsilon$  for any  $t \geq \max(t_1, \dots, t_n)$ , which contradicts the assumption  $\lim_{t \rightarrow \infty} z(t) = z^*$  made in Step 3, completing the proof.  $\square$

### 6.3 Derivations of Propositions 1, 2, 3, and 4

In this subsection, starting from Proposition 1, we demonstrate that each proposition stated in this paper can be derived from the one coming next, while the last proposition, that is Proposition 4, is a result of Theorem 1 proved previously.

Condition (5) in Proposition 1 indicates that  $\mathcal{B}$  has an infinite bubble number. Thus, assuming that Proposition 2 is true, in view of (9), we obtain  $\alpha_1 = 0$  and  $\alpha_2 = 1$ , immediately resulting in Proposition 1. Setting  $d = 1$  in Proposition 3 simply converts it into Proposition 2. Proposition 3 is a special case of Proposition 4 where  $\beta \rightarrow 0$ . Thus, it only remains to derive Proposition 4 from Theorem 1.

The assumption in Proposition 4 that (24) has two positive solutions  $\alpha_1$  and  $\alpha_2$  in  $(0, 1)$  means that for any  $\alpha \in (\alpha_1, \alpha_2)$  we have

$$\frac{1 + \alpha}{\alpha(1 - \alpha)} < \frac{\gamma_{\mathcal{B}}}{1 - (1 - d)\alpha^\beta}. \quad (67)$$

Thus, setting  $\hat{d}_i(\alpha) = 1 - (1 - d)\alpha^\beta$ , which is a non-increasing function in  $(0, 1)$ , in Theorem 1 immediately leads to the statement of Proposition 4.

## REFERENCES

- [1] Y. Dong, M. Zhan, G. Kou, Z. Ding, and H. Liang, "A survey on the fusion process in opinion dynamics," *Information Fusion*, vol. 43, pp. 57–65, 2018.
- [2] J. Castro, J. Lu, G. Zhang, Y. Dong, and L. Martínez, "Opinion dynamics-based group recommender systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2394–2406, 2018.
- [3] P. V. L. Mastroeni and M. Naldi, "Agent-based models for opinion formation: A bibliographic survey," *IEEE Access*, vol. 7, pp. 58 836–58 848, 29 April 2019.
- [4] A. Gustafson, S. A. Rosenthal, M. T. Ballew, M. H. Goldberg, P. Bergquist, J. E. Kotcher, E. W. Maibach, and A. Leiserowitz, "The development of partisan polarization over the green new deal," *Nature Climate Change*, vol. 9, no. 12, pp. 940–944, 2019.
- [5] J. Green, J. Edgerton, D. Naftel, K. Shoub, and S. J. Cranmer, "Elusive consensus: Polarization in elite communication on the covid-19 pandemic," *Science Advances*, vol. 6, no. 28, p. eabc2717, 2020.
- [6] J. Lang, W. W. Erickson, and Z. Jing-Schmidt, "# maskon!# maskoff! digital polarization of mask-wearing in the united states during covid-19," *PloS one*, vol. 16, no. 4, pp. 1–25, 2021.
- [7] M. Bastos, D. Mercea, and A. Baronchelli, "The geographic embedding of online echo chambers: Evidence from the brexit campaign," *PloS one*, vol. 13, no. 11, pp. 1–16, 2018.
- [8] H. Noorazar, K. R. Vixie, A. Talebanpour, and Y. Hu, "From classical to modern opinion dynamics," *International Journal of Modern Physics C*, vol. 31, no. 07, p. 2050101, 2020.
- [9] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.
- [10] R. P. Abelson, "Mathematical models of the distribution of attitudes under controversy," *Contributions to mathematical psychology*, vol. 14, pp. 1–160, 1964.
- [11] N. E. Friedkin and E. C. Johnsen, "Social influence and opinions," *Journal of Mathematical Sociology*, vol. 15, no. 3-4, pp. 193–206, 1990.
- [12] F. B. V. Amelkin and A. K. Singh, "Polar opinion dynamics in social networks," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5650–5665, 2017.
- [13] S. M. Nematollahzadeh, S. Ozgoli, A. Jolfaei, and M. S. Haghghi, "Modeling of human cognition in consensus agreement on social media and its implications for smarter manufacturing," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2902–2909, 2021.
- [14] C. G. J. Semonsen, A. Squicciarini, and S. Rajtmajer, "Opinion dynamics in the presence of increasing agreement pressure," *IEEE Transactions on Cybernetics*, vol. 49, no. 4, pp. 1270–1278, 2019.
- [15] M. Ye, Y. Qin, A. Govaert, B. D. Anderson, and M. Cao, "An influence network model to study discrepancies in expressed and private opinions," *Automatica*, vol. 107, pp. 371–381, 2019.
- [16] R. Hegselmann, U. Krause *et al.*, "Opinion dynamics and bounded confidence models, analysis, and simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, no. 3, pp. 1–33, 2002.
- [17] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, "Mixing beliefs among interacting agents," *Advances in Complex Systems*, no. 3, p. 11, 2001.
- [18] X. Chen, X. Zhang, Y. Xie, and W. Li, "Opinion dynamics of social-similarity-based hegselmann–krause model," *Complexity*, vol. 2017, 2017.
- [19] A. Sirbu, D. Pedreschi, F. Giannotti, and J. Kertész, "Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model," *PloS one*, vol. 14, no. 3, p. e0213246, 2019.
- [20] H. P. Maia, S. C. Ferreira, and M. L. Martins, "Adaptive network approach for emergence of societal bubbles," *Physica A: Statistical Mechanics and its Applications*, vol. 572, p. 125588, 2021.
- [21] C. Cheng and C. Yu, "Opinion dynamics with bounded confidence and group pressure," *Physica A: Statistical Mechanics and Its Applications*, vol. 532, p. 121900, 2019.

- [22] C. Altafini and F. Ceragioli, "Signed bounded confidence models for opinion dynamics," *Automatica*, vol. 93, pp. 114–125, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0005109818301596>
- [23] G. He, J. Liu, H. Hu, and Jian An Fang, "Discrete-time signed bounded confidence model for opinion dynamics," *Neurocomputing*, vol. 425, pp. 53–61, 2021.
- [24] V. X. Nguyen, G. Xiao, J. Zhou, G. Li, and B. Li, "Bias in social interactions and emergence of extremism in complex social networks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 10, p. 103110, 2020.
- [25] Y. Mao, S. Bolouki, and E. Akyol, "Spread of information with confirmation bias in cyber-social networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 688–700, 2020.
- [26] C. Altafini, "Dynamics of opinion forming in structurally balanced social networks," in *In 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 5876–5881.
- [27] —, "Consensus problems on networks with antagonistic interactions," *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 935–946, 2012.
- [28] A. V. Proskurnikov, A. S. Matveev, and M. Cao, "Opinion dynamics in social networks with hostile camps: Consensus vs. polarization," *IEEE Transactions on Automatic Control*, vol. 61, no. 6, pp. 1524–1536, 2016.
- [29] K. Fan and W. Pedrycz, "Emergence and spread of extremist opinions," *Physica A: Statistical Mechanics and its Applications*, vol. 436, pp. 87–97, 2015.
- [30] P. Cisneros-Velarde, K. S. Chan, and F. Bullo, "Polarization and fluctuations in signed social networks," *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3789–3793, 2021.
- [31] S. P. Hassan Dehghani Aghbolagh, Mohsen Zamani and Z. Chen, "Balance seeking opinion dynamics model based on social judgment theory," *Physica D: Nonlinear Phenomena*, vol. 403, p. 132336, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167278919302684>
- [32] L. Wang, Y. Hong, G. Shi, and C. Altafini, "A biased assimilation model on signed graphs," in *2020 59th IEEE Conference on Decision and Control (CDC)*, 2020, pp. 494–499.
- [33] Pew Research Center, *Partisan Divides over Political Values Widen*, October 5, 2017 (accessed October 24, 2020). [Online]. Available: <https://www.pewresearch.org/politics/2017/10/05/1-partisan-divides-over-political-values-widen>
- [34] G. L. Cohen, J. Aronson, and C. M. Steele, "When beliefs yield to evidence: Reducing biased evaluation by affirming the self," *Personality and social psychology bulletin*, vol. 26, no. 9, pp. 1151–1164, 2000.
- [35] M. Mäs and A. Flache, "Differentiation without distancing. explaining bi-polarization of opinions without negative influence," *PLoS one*, vol. 8, no. 11, p. e74516, 2013.
- [36] G. Fu and W. Zhang, "Opinion formation and bi-polarization with biased assimilation and homophily," *Physica A: Statistical Mechanics and its Applications*, vol. 444, pp. 700–712, 2016.
- [37] M. Ramos, J. Shao, S. D. Reis, C. Anteneodo, J. S. Andrade, S. Havlin, and H. A. Makse, "How does public opinion become extreme?" *Scientific reports*, vol. 5, no. 1, pp. 1–14, 2015.
- [38] P. Dandekar, A. Goel, and D. T. Lee, "Biased assimilation, homophily, and the dynamics of polarization," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5791–5796, 2013. [Online]. Available: <https://www.pnas.org/content/110/15/5791>
- [39] Z. Chen, J. Qin, B. Li, H. Qi, P. Buchhorn, and G. Shi, "Dynamics of opinions with social biases," *Automatica*, vol. 106, pp. 374–383, 2019.
- [40] W. Xia, M. Ye, J. Liu, M. Cao, and X.-M. Sun, "Analysis of a nonlinear opinion dynamics model with biased assimilation," *Automatica*, vol. 120, p. 109113, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0005109820303113>
- [41] B. D. Anderson and M. Ye, "Recent advances in the modelling and analysis of opinion dynamics on influence networks," *International Journal of Automation and Computing*, vol. 16, no. 2, pp. 129–149, 2019.
- [42] I. L. Janis, "Groupthink," *Psychology Today*, vol. 5, no. 6, pp. 43–46, 1971.
- [43] E. Aronson, T. D. Wilson, R. M. Akert, and S. R. Sommers, *Social Psychology*, 9th ed. Pearson, 2013, ch. 9, pp. 285–289.
- [44] S. P. Robbins and T. A. Judge, *Essentials of Organizational Behavior*, 14th ed. Prentice Hall Upper Saddle River, NJ, 2018, ch. 10, pp. 187–197.
- [45] R. S. Baron, "So right it's wrong: Groupthink and the ubiquitous nature of polarized group decision making," *Advances in Experimental Social Psychology*, vol. 37, no. 2, pp. 219–253, 2005.
- [46] I. Janis, *Victims of Groupthink: A Psychological Study of Foreign-policy Decisions and Fiascoes*. Houghton, Mifflin, 1972.
- [47] I. L. Janis, "Groupthink," *IEEE Engineering Management Review*, vol. 36, no. 1, pp. 235–246, 2008.
- [48] A. Hermann and H. G. Rammal, "The grounding of the "flying bank"," *Management Decision*, vol. 48, no. 7, pp. 1048–1062, 2010.
- [49] C. Lees, "Brexit, the failure of the British political class, and the case for greater diversity in UK political recruitment," *British Politics*, vol. 16, no. 1, pp. 1–22, 2020.
- [50] J. Lorenz, H. Rauhut, F. Schweitzer, and D. Helbing, "How social influence can undermine the wisdom of crowd effect," *Proceedings of the National Academy of Sciences*, vol. 108, no. 22, pp. 9020–9025, 2011.
- [51] J. Becker, D. Brackbill, and D. Centola, "Network dynamics of social influence in the wisdom of crowds," *Proceedings of the National Academy of Sciences*, vol. 114, no. 26, pp. E5070–E5076, 2017. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.1615978114>
- [52] P. Shakerian, A. Bhatnagar, A. Aleali, E. Shaabani, and R. Guo, *Diffusion in Social Networks*. Cham: Springer International Publishing, 2015, ch. Independent Cascade and Linear Threshold Models.
- [53] S. S. Singh, K. Singh, A. Kumar, H. K. Shakya, and B. Biswas, "A survey on information diffusion models in social networks," in *International Conference on Advanced Informatics for Computing Research*. Springer, 2018, pp. 426–439.
- [54] F. Bullo, *Lectures on Network Systems*, 1st ed. Kindle Direct Publishing, 2020, ch. 5, with contributions by J. Cortes, F. Dorfler, and S. Martinez. [Online]. Available: <http://motion.me.ucsb.edu/book-Ins>
- [55] C. Godsil and G. F. Royle, *Algebraic Graph Theory*. Springer Science & Business Media, 2013, vol. 207.
- [56] S. Flaxman, S. Goel, and J. M. Rao, "Filter Bubbles, Echo Chambers, and Online News Consumption," *Public Opinion Quarterly*, vol. 80, no. S1, pp. 298–320, Mar. 2016, \_eprint: <https://academic.oup.com/poq/article-pdf/80/S1/298/17120810/nfw006.pdf>. [Online]. Available: <https://doi.org/10.1093/poq/nfw006>
- [57] L. Breuning, "The neurochemistry of science bias," in *Groupthink in Science*. Springer, 2020, pp. 3–14.
- [58] D. M. Allen, "The mental and interpersonal mechanisms of groupthink maintenance," in *Groupthink in Science*. Springer, 2020, pp. 27–35.
- [59] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, no. 1, pp. 17–60, 1960.
- [60] H. Noorazar, "Recent advances in opinion propagation dynamics: a 2020 survey," *The European Physical Journal Plus*, vol. 135, no. 6, pp. 1–20, 2020. [Online]. Available: <https://doi.org/10.1140/epjp/s13360-020-00541-2>
- [61] O. Abrahamsson, D. Danev, and E. G. Larsson, "Opinion dynamics with random actions and a stubborn agent," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2019, pp. 1486–1490.
- [62] Q. Zhou, Z. Wu, A. H. Altalhi, and F. Herrera, "A two-step communication opinion dynamics model with self-persistence and influence index for social networks based on the degroot model," *Information Sciences*, vol. 519, pp. 363–381, 2020.
- [63] S. Huang, B. Xiu, and Y. Feng, "Modeling and simulation research on propagation of public opinion," in *2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, 2016, pp. 380–384.
- [64] G. Fu, W. Zhang, and Z. Li, "Opinion dynamics of modified hegselmann–krause model in a group-based population with heterogeneous bounded confidence," *Physica A: Statistical Mechanics and its Applications*, vol. 419, pp. 558–565, 2015.
- [65] S. E. Parsegov, A. V. Proskurnikov, R. Tempo, and N. E. Friedkin, "Novel multidimensional models of opinion dynamics in social networks," *IEEE Transactions on Automatic Control*, vol. 62, no. 5, pp. 2270–2285, 2017.
- [66] R. Hegselmann and U. Krause, "Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model," *Networks & Heterogeneous Media*, vol. 10, no. 3, p. 477, 2015.
- [67] H. Z. Brooks and M. A. Porter, "A model for the influence of media on the ideology of content in online social networks," *Physical Review Research*, vol. 2, no. 2, p. 023041, 2020.
- [68] Y. Yi and S. Patterson, "Disagreement and polarization in two-party social networks," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 2568–2575, 2020.
- [69] H.-T. Wai, A. Scaglione, and A. Leshem, "Active sensing of social networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 3, pp. 406–419, 2016.
- [70] D. Acemoğlu, G. Como, F. Fagnani, and A. Ozdaglar, "Opinion fluctuations and disagreement in social networks," *Mathematics of Operations Research*, vol. 38, no. 1, p. 1–27, Feb 2013. [Online]. Available: <http://dx.doi.org/10.1287/moor.1120.0570>