



HAL
open science

A model of Elite Interactions and Hidden Opinions

Shaden Shabayek

► To cite this version:

| Shaden Shabayek. A model of Elite Interactions and Hidden Opinions. 2024. hal-04649036

HAL Id: hal-04649036

<https://hal.science/hal-04649036>

Preprint submitted on 15 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A model of Elite Interactions and Hidden Opinions

Shaden Shabayek*

July 14, 2024

Abstract

I tackle an old question, which has received a revived interest with the rise of social media: what drives opinion polarisation? There is no consensus in both the empirical and the theoretical literature about how to model, measure and investigate Opinion heterogeneity. In particular, one drawback in the literature is that models rarely produce or explain the broad range of opinion distributions (consensus, weak consensus, dissensus, clustering, polarisation, strong polarisation) that one might observe in real life. In this research paper, I introduce essential elements into a theoretical framework which are directly inspired from the observation of online social networks today. First, I introduce the inequality of attention that individuals receive online, e.g. when public figures express themselves online they receive attention from thousands or millions of followers and can have an impact on a given discourse, as opposed to normal individuals with at most few hundred followers. Second, since only a happy few receive attention within the digital public sphere, I focus in my model on the interactions between Elite individuals (based on their centrality as a measure of influence) who can hold concordant or discordant opinions. It is documented in the literature that ideologically opposed individuals tend to interact together and push each other towards even more extreme opinions. I run simulations and show that depending on the fraction of Elite individuals and the level of tolerance in discussions, opinions within society can move from consensus to polarisation.

Keywords: opinion formation, opinion polarisation, Social network analysis, DeGroot Model.

1 Introduction

Opinion polarisation can be loosely defined as having two groups of individuals with discordant opinions across groups and concordant opinions within group. Over the past decades, a large scholarship has studied many definitional variants of this concept, such as group, political, attitudinal, ideological or affective polarisation, cultural differentiation, community cleavage, opinion disagreement, anti-conformism, oppositional identity, where the concept of opinion itself can be fused, depending on the disciplinary focus, with one or multiple components such as values, beliefs, attitudes or ideology. Before the rise of the internet,

*Associate researcher at Medialab Sciences Po, 27 rue Saint Guillaume 75337 Paris Cedex 07, France. Email of the corresponding author: shaden.shabayek@sciencespo.fr.

the web and social networking platforms, these variants were mainly circumscribed within the fields of social psychology (e.g. Sherif and Hovland [1961], Myers [1982], Huguet and Latané [1996]), political science (e.g. Axelrod [1977]), sociology (e.g. DiMaggio et al. [1996]) and applied mathematics (e.g. DeGroot [1974]).

However, a shift in the literature studying opinions, has occurred along two dimensions. First, a shift of attention has occurred, from studying dynamics of opinions within *small* or *big* groups from the perspective of their behaviour towards, the study of the impact of the medium itself used by these individuals to communicate, get informed, access others' stance about social issues, etc. In other words a lot of attention has been diverged towards the study of social networking platforms, along with the popularisation of personal computers and the normalisation of social media. Second, a shift in the methodology has occurred, from using experimental setups and surveys towards computational models, simulations and large amounts of collected data. This is because the development of personal computers and cheaper access to computing power, has made computational models, the collection and the processing of *big* data, within reach to many researchers.

Namely, the study of *opinion* dynamics within a society, has been adopted by scholars from, a priori distant fields, such as computer science (Jager and Amblard [2004], Malliaros and Vazirgiannis [2013]) and physics (e.g. Castellano et al. [2009]). As for scholars from historically related fields such as media studies, communication studies, political science, sociology, psychology, they have integrated new methods to account for the digital transformation. In particular, concepts like recommendation algorithms, filter bubbles, echo chambers, misinformation (e.g. Pariser [2011], Allcott and Gentzkow [2017], Sikder et al. [2020]), and in general the role of social media, as a technology driving society towards *more* polarised opinions, has been a popular, yet a criticised (Barberá [2020]), framing for the study of the dynamics of opinions of large groups immersed in daily digital spaces.

The present research seeks to take a step back in order to revisit simultaneously the two previously mentioned dimensions. To that end, I model opinion heterogeneity within a society, by (1) studying at the same time individual behaviour and the specificity of the technological component of our communication mediums today, and (2) by running simulations to explore opinion dynamics with the newly added components. More specifically, I formalise an opinion formation model which integrates two heuristics that reflect the functioning of social networks today, while at the same time modelling individual behaviour with an opinion updating rule. First, I account for differential visibility online. Eventhough, we could all express ourselves online, we do not all get the same attention. Only the content created by a happy few, gets to be seen, read and shared by active users (e.g. Wu et al. [2011], Dagoula [2019], Hughes et al. [2019], Wojcik and Hughes [2019]). I model the difference in visibility online, by introducing two types of individuals: *normal* individuals and *Elite*. Second, I account for communication practices in the digital public sphere, by allowing for two types of behaviours, correlated with the two types of individuals, where only Elite individuals get to interact with other Elite individuals. More precisely, I integrate an attractive (repulsive) effect, that occurs when two Elite individuals who are ideologically aligned (opposed) interact, they (attract) repulse each other by pulling (pushing) their opinions closer (further) to each other. This repulsive effect is well documented in the literature in the case of political discussions in the digital public sphere (e.g; Conover et al. [2011], ?), ?).

These two heuristics depart from early models of opinion formation, such as French [1956], Harary [1959], DeGroot [1974]. In these seminal models, individuals pool their opinions by taking the average of opinions expressed by all of their direct contacts at every period of interaction, irrespective of differences in

levels of expertise or influence. When referring to this classic opinion updating rule, I say that individuals update their opinions à la DeGroot. In that framework, regardless of the specific topology of the network, as long as it is (strongly) connected and aperiodic, individuals reach consensus of opinion in the long-run. The present paper refines these contributions by relating the opinion updating rule to the topology of the network. Namely, the two types of individuals (Elite and Normal) have different opinion updating rules, which depend on their network centrality. Furthermore, I run simulations where I vary the share of Elite individuals in society and the threshold of tolerance in discussions, which is a threshold that accounts for concordant or discordant opinions. I then study the patterns of opinion obtained in the long run by following the methodology suggested in Devia and Giordano [2022]. The latter provide a method which classifies long-run opinion distributions into multiple classes such as perfect consensus, consensus, polarisation, clustering, or dissensus. In doing so, I provide a unifying framework to assess different long-run opinion patterns, in relation to the existence of highly visible individuals who attract a lot of attention within the online public discourses.

The rest of the paper is organised as follows. First in section 2, I provide a review of the literature exploring disagreement in social contexts and briefly present the literature in social psychology about hidden profiles. In section 3, I formally present my heuristic model of opinion formation with two types of individuals and prove long-run convergence of opinions. In section 5, I run simulations of my model based on network structures randomly generated following the Barabási and Albert [1999] (BA) model of scale-free networks using a preferential attachment mechanism. Finally I provide in the last section a discussion and avenues of improvement.

2 Literature review

The workhorse models of French [1956], Harary [1959], DeGroot [1974] have been established the conditions under which consensus within a society of interconnected individuals. However, explaining disagreement, diversity in opinions (Flache et al. [2017]) or the *community cleavage problem* (Friedkin [2015]) has proven to be a challenging task and is still an active area of research.

In what follows I introduce the main strands of the literature which tackle disagreement in social contexts and relate them to my work. In particular, the role played by the specific topology of the network is not yet at the center stage of this literature and this is precisely one of the main contributions of my model. For example many papers consider dyadic interactions based on randomly paired individuals (e.g. Deffuant et al. [2000], Grow et al. [2017], Krueger et al. [2017], Axelrod [1977]). Finally, I introduce the papers in social psychology that formalise the idea of Hidden Profiles. This literatures provides further behavioural foundation to my expression heuristic, where people with a low level of popularity *hide* their opinions (update à la Degroot) while Elite individuals express their opinions and interact together.

2.1 Disagreement in social contexts

Stubborn agents

One main hypothesis introduced in order to model diversity of opinions or disagreement has been the introduction of stubborn agents. Individuals remain attached to their opinion in spite of repeated interactions with others (Friedkin and Johnsen [1990, 1997], Friedkin [2015]). In a close set-up, Yildiz et al. [2013] and

Sadler [2019] introduce stubborn agents in a model where opinions can take only two discrete values either a or b . Players can be either stubborn, that is they never update their opinion, or they can update à la DeGroot. Nevertheless, the stubbornness of a player is independent of their network position. Hence a stubborn player who is nor locally nor globally central can have a great impact on the long-run opinions of all individuals in the network. With that respect, I extend this approach by relating the impact that a stubborn player can have on others' long-run opinions, to their popularity or position in the network.

Similarity Bias

Modelling disagreement can also be achieved by assuming that only individuals that are similar *enough* can influence each other. Hence consensus cannot be achieved since dissimilar individuals even if they are connected within a network they do not influence each others' opinions. The similarity bias hypothesis has been formalised following different modelling assumptions. Deffuant et al. [2000] specify an opinion updating rule, where pairs of agents are randomly picked and influence each other only if their opinions are initially close enough; that is the distance between two real valued opinions is below a given threshold. In his seminal paper, Axelrod [1977] assumes that the chance of interaction of any pair of individuals is proportional to their (cultural) similarity; which is measured by assuming that culture can be summarised into a number of finite traits and then by counting the number of common cultural traits between two individuals. Hegselmann and Krause [2002] introduce the bounded confidence model, where the key ingredient is to consider the difference between the opinions of individuals when opinion updating is taking place. In other words, agents update their opinions by taking an average over the opinions of neighbours whose opinion difference falls within a confidence interval. When neighbours opinions fall outside the confidence interval they are ignored. Following the family of bounded confidence models, Grabisch et al. [2022] model diversity in opinion by allowing new links to get created when the difference of opinions between two individuals is close enough (below a threshold) and otherwise agents sever existing links with individuals who have distant opinions. The similarity bias is also revisited through the study of a confirmation bias on online social networks, where individuals are supposedly exposed to and process content produced by like-minded individuals, while content which is opposed to their beliefs is rejected. For example, Sikder et al. [2020] and Del Vicario et al. [2017] explore the confirmation bias hypothesis, in a model with linear opinion updating, in order to rationalise polarisation within society. Homophily has also been introduced as means of exploring the similarity bias hypothesis, both in a theoretical and empirical set-up with social networks data. Dandekar et al. [2013] show that homophily or biased assimilation can be a driver of persistent polarisation in a DeGroot set-up. Bakshy et al. [2015] provide empirical support for homophily by studying the political colouring of friends on Facebook and show that the composition of the network of friends is the main factor which limits exposure to opinions of dissimilar others.

I specifically extend this literature by introducing an opinion updating rule which accounts for the opinions of neighbours in different ways, depending on whether their opinions fall within or outside the confidence interval.

Repulsive and attractive influence

Disagreement can be further modelled by including repulsive or negative influence when an individual interacts with dissimilar others. Jager and Amblard [2004] introduce a model where pairs of agents are

randomly selected. For each pair, if the difference of their opinions is above a given threshold then their updated opinions drift further apart, otherwise if the difference is smaller than a given threshold then their opinions converge towards each other. Axelrod et al. [2021] formalise an attraction-repulsion model, where beyond a tolerance threshold, agents are too dissimilar and their difference increases. Melguizo [2018] studies persistent disagreement, she allows interactions and attitudes to co-evolve. Hence she departs from the time independent weights used in averaging neighbours' opinions in models à la DeGroot, where the network and the links between individuals are static through out all the interaction periods. The key ingredient used is to assume that each individual has several attributes. Relationships with other individuals sharing similar attributes become more intense, while relationships with dissimilar others deteriorate. On the empirical side of the literature, Conover et al. [2011] show that ideologically opposed individuals do interact together on Twitter without necessarily broadcasting content that is not aligned with their views, hence giving support to a theory of social influence which includes a repulsive effect.

2.2 Hidden profiles

The idea of introducing in an opinion formation model, two types of individuals depending on their influence or popularity in society, is also grounded in the *Hidden Profiles* literature in social psychology. Stasser and Titus [1985] document how individuals in social contexts, do not always share the information they hold. The starting point of their research is to challenge the common belief that a group of individuals should be able to take a better decision than each individual on their own by pooling the members' knowledge and expertise. Namely, group discussion or communication is believed to have a corrective function because members can each have incomplete information but together they can gather the different pieces of the puzzle. The authors ran an experiment in which they simulate a political set-up where a group has to elect one of three candidates: *Best*, *Okay* and *Ohum*. In a first protocol, they distributed a different subset of desirable traits of *Best* and a different subset of *Okay*'s undesirable traits over the members of the group, such that from each one's individual perspective *Okay* appeared more positive than *Best*. Before discussion *Best* received 25% of endorsement. Since the whole group had complete (but dispersed) information about *Best* they could exchange it and come to the conclusion that *Best* was actually the best candidate. Yet after group discussion, surprisingly the percentage of endorsement for *Best* remained at 24%. This finding suggests that unique information held by some members of the group about candidates were not being shared. In a later study, Stasser et al. [1989] showed that unique pieces of information are less likely to be mentioned during group discussion. One explanation is that social status, expertise or popularity can be a driver for expression. In fact, Jr. et al. [1996] suggest that repeating a unique piece of information, leading to the formation of group opinion during a discussion, is more likely by higher status members (experts, leaders, etc.) rather than lower status members. They ran an experiment with residents, interns and 3rd-year medical students and they show that residents were more likely to repeat (unique) information when compared to interns and students.

Hence, in my model I borrow the motivation behind the Hidden profiles literature to model the difference in opinion updating as a function of one's popularity. Clearly in my model there is no information to aggregate nor a "*best*" final decision to take. However studying opinions build up within a society can be extended to include a final phase where individuals have to take a decision, like for instance voting for political candidates.

3 The Model

3.1 A Heuristic model of Elite interactions

I consider a model where a set of individuals are embedded in a social network represented by a connected and undirected graph $\mathcal{G} = (V, E)$. The nodes V represent individuals where $|V| = n$ and the edges E represent a relationship or a conversational channel between two individuals. I denote by N_i the set of neighbours of individual i , where $N_i = \{j \in V : (i, j) \in E\}$. Let G be the adjacency matrix associated to the graph \mathcal{G} with entries $g_{ij} = 1$ whenever $(i, j) \in E$ and I assume that $g_{ii} = 1$ (or equivalently $(i, i) \in E$). Individuals hold opinions $\mathbf{x}(t) \in [-1, 1]^n$ which represent their stance about a given issue at time period $t \in \mathbb{N}$. The extreme points -1 and 1 represent opposing viewpoints. An opinion formation process describes how individuals update their opinions $\mathbf{x}(t)$ at time period t as a function of $\mathbf{x}(t-1)$. In the seminal DeGroot [1974] model, the updated opinion of an individual is the weighted average of their neighbours' opinions at the previous time period, including their own opinion, that is:

$$x_i(t) = \frac{1}{|N_i|} \sum_{j \in N_i} x_j(t-1). \quad (1)$$

If the social network structure is (strongly) connected and aperiodic, and individuals update their opinions according to (1), then they reach consensus of opinions in the long-run. In my model, I modify this rule by linking the opinion updating process to both the network topology and the distribution of initial opinions. Let $\mathbf{C}_i(\mathcal{G})$ be a real valued centrality measure, which represents the influence of individual i in network \mathcal{G} . Let $\mathbf{C}^*(\mathcal{G})$ be a threshold, which represents a cut-off value (e.g. 95th percentile) above which individuals belong to an influential *Elite*. The opinions of individuals with a centrality strictly below the *Elite* threshold $\mathbf{C}^*(\mathcal{G})$, get updated according to the DeGroot opinion formation process described in (1). Individuals with a centrality above the *Elite* threshold update their opinion according to a process which (i) includes an attractive and repulsive effect, reflecting respectively the opinions of two linked individuals which are concordant (discordant) and get further attracted (repulsed) to (from) each other, (ii) only takes into account the opinions of individuals with a centrality above the *Elite* threshold. More specifically, two individuals i and j have concordant opinions (like-minded) at period $t \geq 0$, whenever $|x_i(t) - x_j(t)| < \tau$, for $\tau \in (0, 2)$ the tolerance threshold, otherwise their opinions are discordant (ideologically-opposed). Formally, *Elite* individuals update their opinions according to the following process:

$$\begin{cases} x_i(t) = 1, & \text{if } x_i(t-1) \geq 1 \\ x_i(t) = -1, & \text{if } x_i(t-1) \leq -1 \\ x_i(t) = x_i(t-1) + \Delta(t-1), & \text{if } x_i(t-1) \in (-1, 1) \end{cases} \quad (2)$$

where

$$\begin{aligned} \Delta(t-1) = & \mu \left(\sum_{\substack{j \in N_i \\ \mathbf{C}_j(\mathcal{G}) \geq \mathbf{C}^*(\mathcal{G})}} \delta_{ij}(t-1)[x_j(t-1) - x_i(t-1)] \right. \\ & \left. - \sum_{\substack{j \in N_i \\ \mathbf{C}_j(\mathcal{G}) \geq \mathbf{C}^*(\mathcal{G})}} [1 - \delta_{ij}(t-1)][x_j(t-1) - x_i(t-1)] \right) \end{aligned} \quad (3)$$

and $\mu \in (0, \mu_i)$ and $\delta_{ij}(t-1) = \mathbb{1}_{\{|x_i(t-1) - x_j(t-1)| < \tau\}}$ is an indicator function that takes the value 1 when the opinions of individual i and their neighbour j are concordant at the previous period. The parameter μ , which controls the speed of the process, is in the interval $(0, \mu_i)$ to respect a Boundary condition explained in Appendix A.2. The first and second term in $\Delta(t-1)$ (3) reflect respectively the attractive and repulsive effect in the updating process of individual i , when interacting with an individual j whose opinion is concordant or discordant.

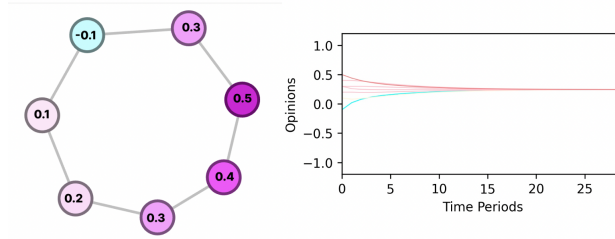


Figure 1: Left panel: wheel graph with Elite individuals who hold initial concordant opinions $\mathbf{x}_0 = [0.3, -0.1, 0.1, 0.2, 0.3, 0.4, 0.5]$. Right panel: the evolution of opinions over time, with the parameter $\mu = 0.1$.

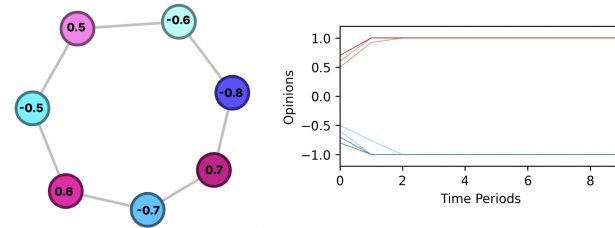


Figure 2: Left panel: wheel graph with Elite individuals who hold initial concordant opinions $\mathbf{x}_0 = [-0.5, 0.5, -0.6, 0.6, -0.7, 0.7, -0.8]$. Right panel: the evolution of opinions over time, with the parameter $\mu = 0.1$.

Figures 1 and 2 provide a simple example of the opinion formation process (2). I take 7 individuals connected in a wheel graph. The associated adjacency matrix G is such that $g_{12} = g_{23} = g_{34} = g_{45} = g_{56} = g_{67} = g_{71} = 1$. I further assume that the tolerance threshold, at which opinions are concordant or discordant, is $\tau = 0.5$. Figure 1 illustrates long-run consensus. Unlike the DeGroot [1974] model, consensus is reached here, only because the initial opinions of each pair of neighbors are concordant, that is their difference falls below the threshold $\tau = 0.5$. With the original DeGroot [1974] model, even if individuals have opinions which are very distant, they end up converging in this network. In Figure 2, each individual has at least one neighbour with a discordant opinion and in the long-run opinions become bipolarised, meaning that opinions are concentrated at the two extreme values 1 and -1 . Other initial opinion distributions, that are different than the ones selected for the previous toy examples, produce final opinion distributions that can fall between the extreme cases of consensus and bipolarisation.

4 Analytical results about opinions in the long-run

In this section, I provide a simple convergence result of opinions in the long-run. To reach that result, I use standard linear algebra theorems. In proposition 1, I characterise long-run opinions for two extreme cases. First, when an Elite individual is not connected to any other Elite individuals; then their long-run opinion is exactly their initial opinion, that is such individuals can be characterised as stubborn. Second, whenever an Elite individual is connected to other Elite individuals who are like minded, in the long-run they hold an opinion which is the average opinion of this group of individuals. Furthermore, I show in Lemma 1, that when a pair of Elite individuals are initially like-minded they take a longer time to reach consensus. But when the pair of Elite individuals have discordant opinions, they disagree at a much faster rate. Finally, I characterise the long-run opinions of all individuals by defining a hearing matrix, which summarises who influences whom by distinguishing Elite individuals and normal individuals.

Definition 1 (Connected set of Elite individuals \mathcal{E}) *Let \mathcal{G} be a given network structure. A connected set of Elite individuals $\mathcal{E} \subset V$ is a set of individuals such that:*

- (i) $\forall i \in \mathcal{E}, \mathbf{C}_j(\mathcal{G}) \geq \mathbf{C}^*(\mathcal{G})$
- (ii) $\forall i \neq j \in \mathcal{E}, \exists g_{ik_1} \times g_{k_1 k_2} \dots \times g_{k_l j} > 0$, for individuals $k_1, \dots, k_l \in \mathcal{E}$.

Proposition 1 *Let \mathcal{G} be a network of interpersonal relationships, $\mathbf{x}(0)$ an initial opinion vector and consider $\mathcal{E} \subseteq V$ a given connected set of Elite individuals.*

- (i) (Stubborn) *If $|\mathcal{E}| = 1$ and $\mathcal{E} = \{i\}$ then*

$$\forall t \geq 1, x_i(t) = x_i(0),$$

- (ii) (Like-minded) *If $|\mathcal{E}| = \kappa > 1$ and $\forall i \neq j \in \mathcal{E}, |x_i(0) - x_j(0)| < \tau$, then for $\mu \in (0, 1/\kappa)$ and $j_1, \dots, j_n \in \mathcal{E} \cap N_i$,*

$$\exists t^* \geq 1, \forall t \geq t^*, x_i(t) = \frac{x_i(0) + x_j(0) + \dots + x_{j_n}(0)}{|N_i|}.$$

Proof 1 *See Appendix A.2.*

Lemma 1 *Let $i \neq j \in \mathcal{E} \subset E$ such that $|\mathcal{E}| = 2$ and $t_{\mathbf{x}(\infty)} = \min\{t : |\mathbf{x}(t) - \mathbf{x}(\infty)| < \epsilon\}$. If \mathbf{x}_∞^a is the long-run opinion vector when $|\mathbf{x}_i(0) - \mathbf{x}_j(0)| < \tau$ and \mathbf{x}_∞^r is the long-run opinion vector when $|\mathbf{x}_i(0) - \mathbf{x}_j(0)| \geq \tau$ then $t_{\mathbf{x}_\infty^r} < t_{\mathbf{x}_\infty^a}$.*

Proof 2 *See Appendix A.3.*

Now, I build a hearing matrix which takes into account who listens to whom, depending on the type of individual. I study the long-run behaviour starting at the time period where pairs of ideologically-opposed connected Elite individuals have repulsed each other towards the most extreme opinion. As I have shown in Lemma 1, given the parameter μ in the law of motion (3), a pair of ideologically-opposed connected

Elite individuals repulse each other at a faster rate than a pair of like-minded individuals who debate to reach a consensus. Formally let $t^* \geq t$ be the time period by which the least ideologically-opposed pair of directly connected expressers, in the group of individuals N , have repulsed each other to reach opinions at the upper and lower bound of the opinion interval. That is, for any period t beyond time period t^* , ideologically-opposed neighbours, are no longer updating their opinions and have long-run opinions that are either 1 or -1 . Given a network G representing interpersonal relationships, denote by \tilde{G} the hearing matrix with typical entries \tilde{g}_{ij} , described below:

Normal individuals. For each individual $i \in N$ such that $\mathbf{C}_i(\mathcal{G}) < \mathbf{C}^*(\mathcal{G})$, the entries in the hearing matrix become $\tilde{g}_{ij} = g_{ij}/d_i, \forall j \in N$.

Elite individuals. For each individual $i \in N$ such that $\mathbf{C}_i(\mathcal{G}) \geq \mathbf{C}^*(\mathcal{G})$, there are three cases to consider.

- (i) For all $i \in N$ in a connected set of Elite individuals \mathcal{E} such that $|\mathcal{E}| = 1$ (stubborn), the entries of the hearing matrix \tilde{G} are: $\tilde{g}_{ii} = 1$ and $\tilde{g}_{ij} = 0$ for all $j \in N_i$.
- (ii) For all $i \in N$ in a connected set of Elite individuals \mathcal{E} such that $|\mathcal{E}| = \kappa > 1$ with like-minded neighbors at period t^* , the entries of the hearing matrix \tilde{G} are: $\tilde{g}_{ii} = 1 - |N_i \cap \mathcal{E}| \mu$, $\tilde{g}_{ij} = \mu$ for $j \in N_i \cap \mathcal{E}$ and $\tilde{g}_{ij} = 0, \forall j \notin N_i \cap \mathcal{E}$.
- (iii) For all $i \in N$ in a connected set of Elite individuals \mathcal{E} such that $|\mathcal{E}| = \kappa > 1$ with ideologically-opposed neighbors, i.e. $\forall i \neq j \in \mathcal{E}$ and $j \in N_i \cap \mathcal{E}, |\mathbf{x}_i(t^*) - \mathbf{x}_j(t^*)| \geq \tau$, the entries in the hearing matrix \tilde{G} are: $\tilde{g}_{ii} = \tilde{g}_{jj} = 1, \tilde{g}_{ik} = 0$ for all $k \in N_i$ and $\tilde{g}_{jk} = 0$ for all $k \in N_j$.

Notice that all the entries of the hearing matrix \tilde{G} are positive and each row sums up to one. For a given network structure G , the process of interpersonal influence describing the evolution of opinions at period $t \geq t^*$ is given by the following equation:

$$\mathbf{x}(t+1) = \tilde{G}\mathbf{x}(t) \quad (4)$$

By induction, the opinions at period $t \geq t^*$ are given by $\tilde{G}^t \mathbf{x}(t^*)$ and the limit yields the long-run opinions.

Theorem 1 Given $\mathbf{x}(t^*) \in [-1, 1]^n$ a vector of opinions at period t^* and a hearing matrix \tilde{G} associated with the network structure G , the long-run opinions are :

$$\mathbf{x}_\infty = \left(\lim_{t \rightarrow \infty} \tilde{G}^t \right) \mathbf{x}(t^*) = \mathcal{G}\mathbf{x}(t^*) < \infty,$$

where \mathcal{G} is the spectral projector associated with the eigenvalue 1. Moreover, the algebraic multiplicity of the eigenvalue 1 is equal to the number of essential classes of the hearing matrix \tilde{G} .

Proof 3 See Appendix A.4

The columns corresponding to Normal individuals in the matrix \mathcal{G} are all zero, meaning that in the long-run the initial opinions of such individuals vanish. Their opinions remain hidden through out the periods of interaction. As for the columns corresponding to Elite individuals, they have at least one strictly positive

entry. In particular, the long-run opinions of Normal individuals are exactly convex combinations of initial opinions of Elite individuals. In other words, the long-run opinion of Normal individuals is affected by the long-run opinions of all the Elite individuals to whom they are connected, through a path of other Normal individuals. Hence, the total impact of the initial opinion of a given expresser $i \in N$ over long-run opinions can be assessed by considering the total weight an Elite individual has in the long-run opinions of other individuals. This motivates the introduction of the following statistic, to assess the total influence of each Elite individual in the network.

Definition 2 (Spectral influence) *Given a network structure G , a hearing matrix \tilde{G} and its limit \mathcal{G} , the spectral influence of individual $i \in N = \{1, \dots, n\}$ is:*

$$s_i = \frac{1}{n}(\mathcal{G}' \mathbf{1}_n)_i,$$

where $\mathbf{1}_n$ is a column vector of ones.

5 Simulations

In this section I explore the model by means of simulations. I utilise the framework provided by Devia and Giordano [2022]. Their framework consists in classifying long-run opinion distributions into multiple qualitative categories, such as consensus, polarisation, clustering or dissensus.

5.1 Steps of the simulations

To run the simulations, I generate scale-free networks following the Barabási–Albert (BA) algorithm with n nodes and m hubs. I choose this specific algorithm, rather than small world networks or random networks, in order to be able to capture differences in nodes' (individuals) centrality in the generated graph. The BA algorithm starts with m nodes, which are the number of hubs. Then at each step creates a new node with m edges which get attached to m other nodes in the graph. Pre-existing nodes receive new links with a probability proportional to their degree, hence the algorithm exhibits preferential attachment. For my simulations, I set n the number of nodes to 2000 individuals and I run the simulations with 1 hub then 2 hubs, in order to capture increasing connectivity in the network.

Second, I generate initial opinions by drawing each individual's initial opinion from a uniform distribution over $[-1, 1]$. I use the same initial vector of opinions for multiple simulations in order to avoid any effect that would come from the uniform distribution.

Third, to identify Elite individuals as opposed to Normal individuals, I compute each individual's degree centrality. I select degree centrality for out of simplicity but more sophisticated centrality measures, such as Eigenvector centrality or more local centrality measures which would make more sense depending on the community of individuals that is being studied, can be easily used.

Fourth, to run the simulations with the previous characteristics, I test (i) different thresholds for the degree centrality that yield, for example a share of only 10% individuals that are Elite and, (ii) I vary the parameter $\tau \in [0, 2]$, which is a tolerance threshold such that if difference of opinions of two connected Elite individuals fall below this threshold then they are like-minded and attract each other, otherwise they repulse each other. When $\tau = 0$ it means that there is very little tolerance in discussions and any small difference in

opinions make individuals repulse each other, while $\tau = 2$ means that individuals are very patient and can interact with individuals with opinions that are very distant from theirs. Finally, I run robustness checks by allowing the selection of Elite individuals to be done at random, as opposed to individuals who have a degree centrality above a specific threshold, for the exact same generated network structure and the initial opinion vector.

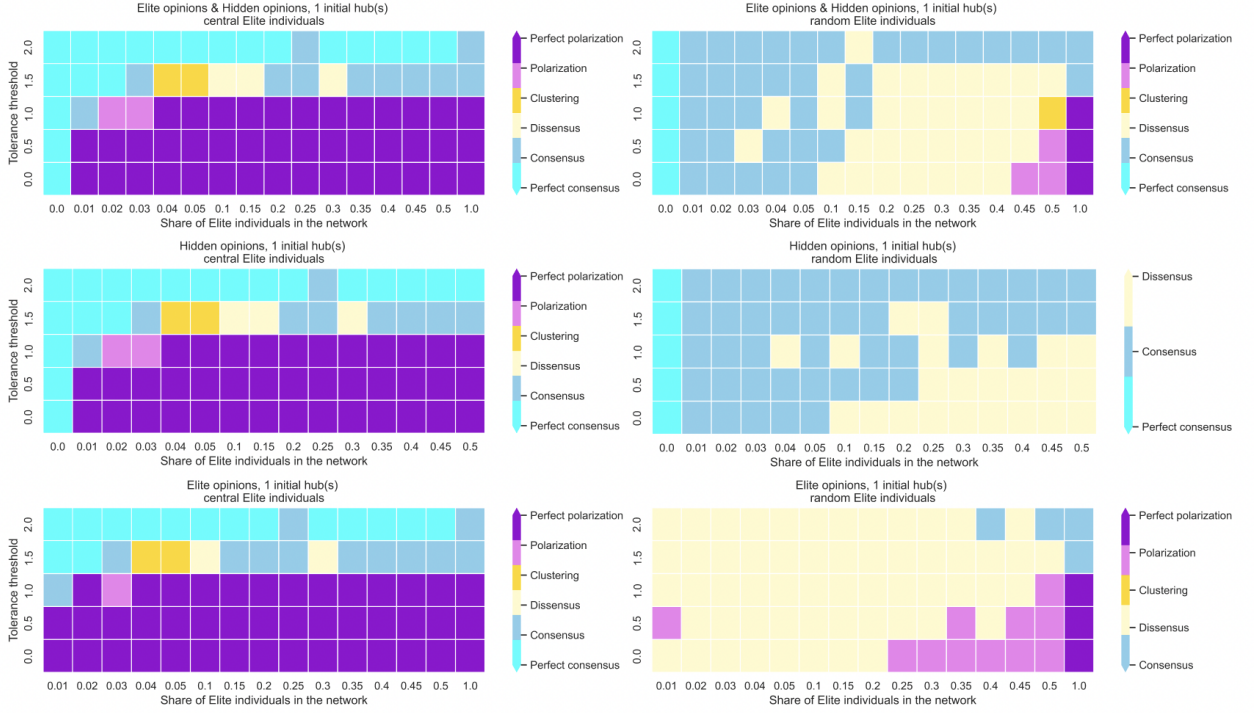


Figure 3: Simulations are run with 1 initial hub. The x-axis corresponds to the share of Elite individuals in society and the y-axis corresponds to the agreement threshold, which determines whether two Elite individuals have concordant or discordant opinions. Left panels: Elite individuals are selected such that they are the most central in terms of degree centrality. Right panels: Elite individuals are selected at random among the total population.

5.2 Results

To run the simulations I generate one single network structure and one single uniform opinion vector, then I use them for two sets of simulations: (i) first the case where a given share of Elite individuals are individuals who have the highest degree centrality in the network, (ii) second the case where Elite individuals are chosen at random in order to test the effect of my model. I use the same generated network and the same initial opinion vector in order to avoid confounding effects. I adopt the framework of Devia and Giordano [2022] for my model in order to classify long-run opinions distribution into multiple qualitative categories below,

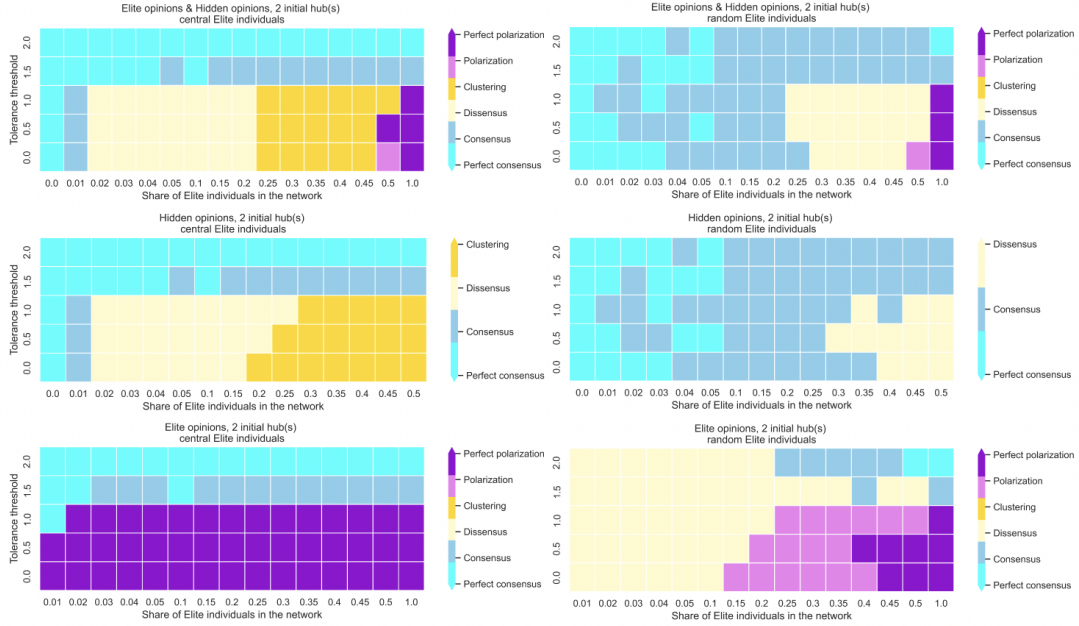


Figure 4: Simulations are run with 2 initial hubs. The x-axis corresponds to the share of Elite individuals in society and the y-axis corresponds to the agreement threshold, which determines whether two Elite individuals have concordant or discordant opinions. Left panels: Elite individuals are selected such that they are the most central in terms of degree centrality. Right panels: Elite individuals are selected at random among the total population.

to which I add perfect polarisation where the whole society is split between two distant groups:¹

- Perfect consensus the absolute majority chooses the very same opinion;
- Consensus the absolute majority chooses approximately the same opinion;
- Polarization the absolute majority is split between two ‘distant’ opinions;
- Clustering the absolute majority is split between two or more groups;
- Dissensus the majority of the opinions are uniformly distributed.

The top left panel of figure 3 shows the final opinions for the whole society when the network is generated with 1 initial hub, where an increasing share of Elite individuals leads to increasing polarisation and a higher agreement threshold leads to more consensus. Notice that since Elite individuals are selected such that they are the most central (influential) individuals within the network, only a share of 1% of Elite individuals can lead to polarisation when the agreement threshold is at $\tau = 0.5$. Furthermore for a very high share of Elite individuals in society (more than 35%) consensus can be reached provided that the agreement threshold is above $\tau = 1$. That is with more tolerance in discussions, ideologically opposed Elite individuals can discuss long enough to reach consensus. For an agreement threshold at $\tau = 1.5$, and a relatively small share of Elite individuals (4% or 5%) clustering of final opinions is observed.

¹The exact encoding of long-run opinions can be found in appendix A.1

When we compare the top left panel of figure 3 and 4 where the network was generated with 2 hubs, we obtain different results. With a share of Elite individuals between 2% and 20%, provided that the agreement threshold is strictly below $\tau = 1.5$, the society is in a situation of opinion dissensus and opinion clustering. Hence increasing the number of hubs leads to a higher density of the network, which translates into *Normal* individuals being exposed to more *Elite* opinions. With only 1 hub the diameter of the network, which is the longest shortest path between two nodes, is 21 while with 2 hubs the diameter drops to 8. Furthermore, with 2 hubs *Elite* opinions are similar to *Elite* Opinions with 1 hub, where perfect polarisation is observed for a share of Elite that ranges from 1% to 100% in society, with a tolerance threshold strictly smaller than $\tau = 1.5$. This is because the repulsive or attractive effect operate for pairs of interacting Elite individuals. Hence, even if the network is not very dense and an *Elite* individual is directly connected to one ideologically opposed neighbour, this link is sufficient to make them repulse each other to extreme opinions and since networks are generated with the BA algorithm *Elite* individuals are more likely to be connected to each other. However, the significant difference between the situation with 1 and 2 hubs is observed for the opinions of *Normal* individuals. Namely, with 2 hubs, the opinions of *Normal* individuals are clustered or form dissensus because they have a larger number of links to multiple *Elite* individuals, as opposed to the situation with only 1 hub where basically the opinions of *Normal* individuals can be captured by one specific *Elite* individual to whom they are connected.

Finally, both in figures 3 and 4, the right panels show the final opinions when *Elite* individuals are picked out at random in the population rather than assigning the most central individuals as *Elite* individuals. We see that the distribution of final opinions is significantly different, for all the considered cases of the simulations thus demonstrating the effect of the centrality of *Elite* individuals on final opinions within society. I provide in the appendix in figures 5, 6, 7 and 8 the exact final opinion distributions without breaking them down into the opinion categories of Devia and Giordano [2022].

Conclusion

The present research aimed to formalise an opinion formation model, which produces a broad range of opinion patterns that result from interactions between interconnected individuals, while accounting for modern means of communication and information dissemination. To do so, I have introduced a new heuristic in a DeGroot like opinion formation model, namely I distinguish two types of individuals, Elite individuals who receive a lot of attention and Normal individuals who represent the rest of the society. This heuristic is justified by the modern structure of online social networks where a happy few receive most of the attention online, such as politicians, TV anchors, journalists, musicians, actors, etc. With simulations I show that by varying the share of Elite individuals in society, the density of relationships in terms of exposure to a multiplicity of opinions of Elite individuals and the level of Tolerance in discussions, we can reproduce a broad range of opinion distributions. In particular, the opinions adopted by Normal individuals depend on which Elite individuals they are connected too and how many. Even though the opinions of Normal individuals are not influential, they contribute into making one or many opinions of Elite individuals become more popular.

More broadly, these modelling heuristics, aim at shedding light on our perception of opinion polarisation within society. Namely, I highlight the idea that Elite individuals do and can hold very polarised opinions, while normal individuals can have much more nuanced opinions since they can be exposed to many Elite individuals. In other words, our perception of what *Public Opinion* might be about a given topic, can be

simply biased by how we perceive the opinions of the most vocal and visible individuals.

For future research, I would like to extend the results by considering different initial opinion distributions and real life social network structures. Namely, it would be interesting to investigate which long-run opinions we get when normal or Elite individuals have skewed initial opinions rather than uniform.

References

- Muzafer Sherif and Carl Hovland. *Social Judgement: Assimilation and contrast effects in communication and attitude change*. Yale University Press, 1961.
- David G Myers. Polarizing effects of social interaction. *Group decision making*, 125:137–138, 1982.
- Pascal Huguet and Bibb Latané. Social representations as dynamic social impact. *Journal of communication*, 46(4), 1996.
- Robert Axelrod. The dissemination of culture. a model with local convergence and global polarization. *The journal of conflict resolution*, 1977.
- Paul DiMaggio, John Evans, and Bethany Bryson. Have american’s social attitudes become more polarized? *American journal of Sociology*, 102(3):690–755, 1996.
- Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.
- Wander Jager and Frédéric Amblard. Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. *Computational and Mathematical Organization Theory*, 10:295–303, 2004.
- Fragkiskis Malliaros and Michalis Vazirgiannis. Clustering and community detection in directed networks: A survey. *Physics Reports*, 2013.
- Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591, 2009.
- Eli Pariser. *The filter bubble: What the Internet is hiding from you*. penguin UK, 2011.
- Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–236, 2017.
- Orowa Sikder, Robert E Smith, Pierpaolo Vivo, and Giacomo Livan. A minimalistic model of bias, polarization and misinformation in social networks. *Scientific reports*, 10(1):1–11, 2020.
- Pablo Barberá. Social media, echo chambers, and political polarization. *Social media and democracy: The state of the field, prospects for reform*, 34, 2020.
- Shaomei Wu, Jake M Hofman, Winter A Mason, and Duncan J Watts. Who says what to whom on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 705–714, 2011.

- Chrysi Dagoula. Mapping political discussions on twitter: Where the elites remain elites. *Media and Communication*, 7(1):225–234, 2019.
- Adam Hughes, Alec Tyson, Emma Remy, Aron Smith, Patrick van Kessel, Regina Widjaya, and Stefan Wojcik. National politics on twitter: Small share of u.s. adults produce majority of tweets. *Pew Research Center*, 2019.
- Stefan Wojcik and Adam Hughes. Sizing up twitter users. *Pew Research Center*, 2019.
- French. A formal theory of social power. *Psychol. Rev.*, 63, 1956.
- F. Harary. A criterion for unanimity in french’s theory of social power. *Studies in Social Power*, 1959.
- Carlos Andres Devia and Giulia Giordano. A framework to analyze opinion formation models. *Scientific Reports*, 12(13441), 2022.
- Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439), 1999.
- Andreas Flache, Michael Mas, Thomas Feliciani, Edmund Chattoe-Brown, Guillaume Deffuant, Sylvie Huet, and Jan Lorenz. Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 2017.
- Noah E. Friedkin. The problem of social control and coordination of complex systems in sociology: a look at the community cleavage problem. *IEEE control systems magazine*, June, 2015.
- Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in complex systems*, 3:87–98, 2000.
- A. Grow, A. Flache, and R. Wittek. Global diversity and local consensus in status beliefs: The role of network clustering and resistance to belief change. *Sociological Science*, 4:611–640, 2017.
- Tyll Krueger, Janusz Szwabiński, and Tomasz Weron. Conformity, anticonformity and polarization of opinions: Insights from a mathematical model of opinion dynamics. *mimeo*, 2017.
- Noah Friedkin and Eugene Johnsen. Social influence and opinions. *Journal of mathematical sociology*, 15 (3-4):193–205, 1990.
- Noah E Friedkin and Eugene C Johnsen. Social positions in influence networks. *Social networks*, 19(3): 209–222, 1997.
- E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione. Binary opinion dynamics with stubborn agents. *ACM Trans. Econ. Comp.*, 2013.
- Evan Sadler. Influence campaigns. *mimeo*, 2019.
- Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.

- Michel Grabisch, Antoine Mandel, and Agnieszka Rusinowska. On the design of public debate in social networks michel grabisch (a)(b), antoine mandel (a)(b)*, agnieszka rusinowska (a)(c). *Operations Research*, 2022.
- Michela Del Vicario, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. Modeling confirmation bias and polarization. *Scientific reports*, 7(1):1–9, 2017.
- Pranav Dandekar, Ashish Goel, and David T Lee. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796, 2013.
- Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.
- Robert Axelrod, Joshua J Daymude, and Stephanie Forrest. Preventing extreme polarization of political attitudes. *Proceedings of the National Academy of Sciences*, 118(50):e2102139118, 2021.
- Isabel Melguizo. Persistence of disagreement. *miemo*, 2018.
- Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Goncalves, Filippo Menczer, and Alessandro Flammini. Political polarization on twitter. *Proceedings of the international aaai conference on web and social media*, 5(1):89–96, 2011.
- Garold Stasser and William Titus. Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and social psychology*, 48(6):1467–1478, 1985.
- G. Stasser, L. Taylor, and C. Hanna. Information sampling in structured and unstructured discussions of three and six person groups. *Journal of Personality and social psychology*, 57:67–78, 1989.
- James R. Larson Jr., Caryn Christensen, Ann S. Abbott, and Timothy M. Franz. Diagnosing groups: Charting the flow of information in medical decision-making teams. *Journal of Personality and social psychology*, 71(2):315–330, 1996.
- Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems and Control Letters*, 53, 2004.
- Teresa M. Silva, Luís Silva, and Sara Fernandes. Convergence time to equilibrium distributions of autonomous and periodic non-autonomous graphs. *Linear Algebra and its Applications*, 488:199–215, 2016.
- E. Seneta. *Non-negative matrices and markov chains*. Springer Science and Business Media, 2006 edition, 1981.
- Carl D. Meyer. *Matrix Analysis and Applied linear algebra*. SIAM, 2000.
- J.Ding and Noah H. Rhee. On the equality of algebraic and geometric multiplicities of matrix eigenvalues. *Applied Mathematics letters*, 24, 2011.

A Appendix

A.1 Coding of final opinions

Below I provide the exact shares and parameters I use for the classification of final opinions, as I adapt the framework of Devia and Giordano [2022].

- Perfect polarisation: the whole society is split between two *distant* groups where each group represents at least 40% of the population, and the opinions distance is at least $\tau = 1$;
- Polarisation: at least 75% is split between two *distant* groups where each group size is between 25% and 40% of the population, and the opinions distance is at least $\tau = 1$;
- Clustering: the absolute majority is split between two or more groups, whose opinions distance is strictly smaller than $\tau = 1$;
- Dissensus: the majority of the opinions are uniformly distributed;
- Consensus: at least the absolute majority of the society holds approximately the same opinion;
- Perfect consensus: at least 75% holds the exact same opinion.

A.2 Proof of Proposition 1

(i) When $|\mathcal{E}| = 1$ it means that individual $i \in \mathcal{E}$ has no direct neighbors who choose to express, hence individual i never updates their initial opinion and their long-run opinion is exactly their initial opinion $\mathbf{x}_i(0)$.

(ii) Suppose without loss of generality that $|\mathcal{E}| = \kappa$. If $|\mathcal{E}| = \kappa > 1$ and all individuals $i \in \mathcal{E}$ are like-minded, that is $\forall i \neq j \in \mathcal{E}, |\mathbf{x}_i(0) - \mathbf{x}_j(0)| < \tau$ then for $\mu \in (0, 1/\kappa)$ the opinions get updated in the following way:

$$\begin{cases} \mathbf{x}_1(1) = \mathbf{x}_1(0) + \mu \sum_{j \neq 1 \in \mathcal{E}} g_{ij}(\mathbf{x}_j(0) - \mathbf{x}_1(0)), \\ \dots \\ \mathbf{x}_\kappa(1) = \mathbf{x}_\kappa(0) + \mu \sum_{\substack{j \in \mathcal{E} \\ j \neq \kappa}} g_{\kappa j}(\mathbf{x}_j(0) - \mathbf{x}_\kappa(0)), \end{cases}$$

$$\Leftrightarrow \begin{cases} \mathbf{x}_1(1) = (1 - \mu d_i(\mathcal{E}))\mathbf{x}_1(0) + \mu \sum_{j \neq 1 \in \mathcal{E}} g_{1j}\mathbf{x}_j(0), \\ \dots \\ \mathbf{x}_\kappa(1) = (1 - \mu d_\kappa(\mathcal{E}))\mathbf{x}_\kappa(0) + \mu \sum_{\substack{j \in \mathcal{E} \\ j \neq \kappa}} g_{\kappa j}\mathbf{x}_j(0), \end{cases}$$

where $d_i(\mathcal{E}) = \sum_{j \in \mathcal{E}} g_{ij}$ corresponds to the number of Elite individuals that are in the set of connected Elite individuals \mathcal{E} and are also direct neighbors of individual $i \in \mathcal{E}$. Writing the above system in matrix notation and using induction we get the following relation :

$$\mathbf{x}_{\mathcal{E}}(t) = M^t \mathbf{x}_{\mathcal{E}}(0),$$

where $\mathbf{x}_{\mathcal{E}}(t) = (\mathbf{x}_1(t), \dots, \mathbf{x}_{\kappa}(t))^T$, $\mathbf{x}_{\mathcal{E}}(0) = (\mathbf{x}_1(0), \dots, \mathbf{x}_{\kappa}(0))^T$ and M an $\kappa \times \kappa$ symmetric matrix with diagonal entries $m_{ii} = 1 - d_i(\mathcal{E})\mu$ and off diagonal entries $m_{ij} = \mu g_{ij}$, for $j \neq i \in \mathcal{E}$. Hence, M is a symmetric matrix, with non-negative entries and whose columns and rows sum to one. In order to get the long-run opinions we need to compute $\lim_{t \rightarrow \infty} \mathbf{x}_{\mathcal{E}}(t) = \lim_{t \rightarrow \infty} M^t \mathbf{x}_{\mathcal{E}}(0)$.

Claim 1 $\lim_{t \rightarrow \infty} M^t$ exists.

This limit exists because all the eigenvalues of the matrix M are smaller or equal to 1. To see this, simply recall that by the Gershgorin Circle Theorem (1931), the eigenvalues of the square matrix M belong to the union of its Gershgorin disks.² In the case of the matrix M the Gershgorin disks² write for each $i \in \mathcal{E}$, $D_i = \{x \in \mathbb{R} : |x - m_{ii}| \leq \sum_{j \neq i} |m_{ij}|\} = \{x \in \mathbb{R} : |x - (1 - d_i(\mathcal{E})\mu)| \leq d_i(\mathcal{E})\mu\}$. Hence, the upper bound of the eigenvalues of M is given exactly by $\max_{i \in \mathcal{E}} (1 - d_i(\mathcal{E})\mu) + d_i(\mathcal{E})\mu = 1$. Now I will show that $\lim_{t \rightarrow \infty} \mathbf{x}_{\mathcal{E}}(t)$ is exactly the average of the initial opinions of individuals $1, \dots, \kappa \in \mathcal{E}$.

Claim 2 Let $\mathbf{1}_{p,q}$ be a matrix of ones of size $p \times q$. $\lim_{t \rightarrow \infty} M^t = \frac{1}{\kappa} \mathbf{1}_{\kappa,1} \mathbf{1}_{1,\kappa}$.

Intuitively, since at each time period every updated opinion of an expresser is a convex combination of the opinions of like-minded neighbors who also express, the long-run opinions converge to the average of initial opinions of the members of the connected set of expressers. Formally, I use theorem 1 in Xiao and Boyd (2004) Xiao and Boyd [2004], which states that $\lim_{t \rightarrow \infty} M^t = \frac{1}{\kappa} \mathbf{1}_{\kappa,1} \mathbf{1}_{1,\kappa}$ if and only if (i) the vector $\mathbf{1}$ is a left eigenvector of M associated with the eigenvalue one, (ii) the vector $\mathbf{1}$ is a right eigenvector of M associated with the eigenvalue one, (iii) one is a simple eigenvalue of M . Conditions (i) and (ii) hold for the matrix M because it is symmetric and row stochastic. To see this, one can simply sum the entries over a given row $i \in \mathcal{E}$: $m_{ii} + \sum_{j \neq i \in \mathcal{E}} m_{ij} = 1 - d_i(\mathcal{E})\mu + \sum_{j \neq i \in \mathcal{E}} g_{ij}\mu = 1 - d_i(\mathcal{E})\mu + d_i(\mathcal{E})\mu = 1$. Since the matrix M is symmetric, it is also column stochastic and the vector one is a left and right eigenvector of the matrix M associated with the eigenvalue one. Finally, condition (iii) holds because the matrix M is irreducible with non-negative entries; because the set of individuals in \mathcal{E} is connected and they are all like-minded, in the sense of definition 1. Hence the eigenvalue 1 is simple (Perron-Frobenius Theorem).

A.3 Proof of Lemma 1

Case 1: $|\mathbf{x}_i(0) - \mathbf{x}_j(0)| < \tau$. The law of motion 3 rewrites:

$$\begin{cases} \mathbf{x}_i(t) &= \mathbf{x}_i(t-1) + \mu(\mathbf{x}_j(t-1) - \mathbf{x}_i(t-1)) \\ \mathbf{x}_j(t) &= \mathbf{x}_j(t-1) + \mu(\mathbf{x}_i(t-1) - \mathbf{x}_j(t-1)) \end{cases}$$

$$\Leftrightarrow \begin{cases} \mathbf{x}_i(t) &= (1 - \mu)\mathbf{x}_i(t-1) + \mu\mathbf{x}_j(t-1) \\ \mathbf{x}_j(t) &= (1 - \mu)\mathbf{x}_j(t-1) + \mu\mathbf{x}_i(t-1) \end{cases}$$

We can write the above system in matrix notation:

$$\begin{bmatrix} \mathbf{x}_i(t) \\ \mathbf{x}_j(t) \end{bmatrix} = \begin{bmatrix} 1 - \mu & \mu \\ \mu & 1 - \mu \end{bmatrix} \begin{bmatrix} \mathbf{x}_i(t-1) \\ \mathbf{x}_j(t-1) \end{bmatrix}$$

²All the eigenvalues of M are real because M is a real symmetric matrix.

$$\Leftrightarrow \begin{bmatrix} \mathbf{x}_i(t) \\ \mathbf{x}_j(t) \end{bmatrix} = \overbrace{\begin{bmatrix} 1-\mu & \mu \\ \mu & 1-\mu \end{bmatrix}}^{=M^t} \begin{bmatrix} \mathbf{x}_i(0) \\ \mathbf{x}_j(0) \end{bmatrix} \quad (\text{by induction}).$$

Moreover, we can diagonalize the matrix M^t so that we can compute the limit easily:

$$\begin{aligned} M^t &= \begin{bmatrix} 1-\mu & \mu \\ \mu & 1-\mu \end{bmatrix}^t \\ &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1^t & 0 \\ 0 & (1-2\mu)^t \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}. \end{aligned}$$

For $\mu \in (0, 1/2)$, $\lim_{t \rightarrow \infty} (1-2\mu)^t = 0$. Notice that this is equivalent to upper bounding the distance between opinions at a given period t and the limiting opinions by the second highest eigenvalue.³ It follows that when the opinions of i and j are close enough then they converge exactly to their average:

$$\mathbf{x}_\infty^a = \lim_{t \rightarrow \infty} \begin{bmatrix} \mathbf{x}_i(t) \\ \mathbf{x}_j(t) \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_i(0) \\ \mathbf{x}_j(0) \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{x}_i(0) + \mathbf{x}_j(0)}{2} \\ \frac{\mathbf{x}_i(0) + \mathbf{x}_j(0)}{2} \end{bmatrix}.$$

For $\epsilon > 0$, the time t_a it takes to convergence is: $t_a \geq \frac{\log(\epsilon)}{\log(1-2\mu)}$.

Case 2: $|\mathbf{x}_i(0) - \mathbf{x}_j(0)| \geq \tau$. The law of motion (3) rewrites:

$$\begin{cases} \mathbf{x}_i(t) = \mathbf{x}_i(t-1) + \mu(\mathbf{x}_i(t-1) - \mathbf{x}_j(t-1)) \\ \mathbf{x}_j(t) = \mathbf{x}_j(t-1) + \mu(\mathbf{x}_j(t-1) - \mathbf{x}_i(t-1)) \end{cases}$$

$$\Leftrightarrow \begin{cases} \mathbf{x}_i(t) = (1+\mu)\mathbf{x}_i(t-1) - \mu\mathbf{x}_j(t-1) \\ \mathbf{x}_j(t) = (1+\mu)\mathbf{x}_j(t-1) - \mu\mathbf{x}_i(t-1). \end{cases}$$

We can write the above system in matrix notation:

$$\begin{bmatrix} \mathbf{x}_i(t) \\ \mathbf{x}_j(t) \end{bmatrix} = \begin{bmatrix} 1+\mu & -\mu \\ -\mu & 1+\mu \end{bmatrix} \begin{bmatrix} \mathbf{x}_i(t-1) \\ \mathbf{x}_j(t-1) \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{x}_i(t) \\ \mathbf{x}_j(t) \end{bmatrix} = \overbrace{\begin{bmatrix} 1+\mu & -\mu \\ -\mu & 1+\mu \end{bmatrix}}^{=M^t} \begin{bmatrix} \mathbf{x}_i(0) \\ \mathbf{x}_j(0) \end{bmatrix} \quad (\text{by induction}).$$

Moreover, we can diagonalize the matrix M^t :

$$\begin{aligned} M^t &= \begin{bmatrix} 1+\mu & -\mu \\ -\mu & 1+\mu \end{bmatrix}^t \\ &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1^t & 0 \\ 0 & (1+2\mu)^t \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}. \end{aligned}$$

The limit opinions of i and j are:

³For more details on this topic in linear algebra See Silva, Silva and Fernandes (2016)Silva et al. [2016].

$$\begin{aligned}
\mathbf{x}_\infty^r &= \lim_{t \rightarrow \infty} \begin{bmatrix} \mathbf{x}_i(t) \\ \mathbf{x}_j(t) \end{bmatrix} \\
&= \lim_{t \rightarrow \infty} \frac{1}{2} \begin{bmatrix} 1 + (1 + 2\mu)^t & 1 - (1 + 2\mu)^t \\ 1 - (1 + 2\mu)^t & 1 + (1 + 2\mu)^t \end{bmatrix} \begin{bmatrix} \mathbf{x}_i(0) \\ \mathbf{x}_j(0) \end{bmatrix} \\
&= \lim_{t \rightarrow \infty} \frac{1}{2} \begin{bmatrix} \mathbf{x}_i(0) + \mathbf{x}_j(0) + (\mathbf{x}_i(0) - \mathbf{x}_j(0))(1 + 2\mu)^t \\ \mathbf{x}_i(0) + \mathbf{x}_j(0) + (\mathbf{x}_j(0) - \mathbf{x}_i(0))(1 + 2\mu)^t \end{bmatrix}.
\end{aligned}$$

For any positive μ this limit explodes. However, recall that opinions have an upper 1 and lower bound -1 . It follows that when the opinions of i and j are faraway they diverge until they reach the upper and lower limit of opinions. Moreover, there exists a time t for a given $\mu > 0$ such that that we remain within the permitted bounds. To find this time t given μ , we must solve:

$$\begin{cases} \frac{1}{2}(\mathbf{x}_i(0) + \mathbf{x}_j(0) + (\mathbf{x}_i(0) - \mathbf{x}_j(0))(1 + 2\mu)^t) = 1, & \text{if } \mathbf{x}_i(0) > \mathbf{x}_j(0) \\ \frac{1}{2}(\mathbf{x}_i(0) + \mathbf{x}_j(0) + (\mathbf{x}_i(0) - \mathbf{x}_j(0))(1 + 2\mu)^t) = -1, & \text{if } \mathbf{x}_i(0) < \mathbf{x}_j(0). \end{cases}$$

Given μ , we get the following t_r (for integer values take the floor function):

$$t_r = \begin{cases} \frac{\log\left(\frac{2 - \mathbf{x}_i(0) - \mathbf{x}_j(0)}{\mathbf{x}_i(0) - \mathbf{x}_j(0)}\right)}{\log(1 + 2\mu)} & \text{if } 1 \geq \mathbf{x}_i(0) > \mathbf{x}_j(0) \geq -1 \\ \frac{\log\left(\frac{-2 - \mathbf{x}_i(0) - \mathbf{x}_j(0)}{\mathbf{x}_i(0) - \mathbf{x}_j(0)}\right)}{\log(1 + 2\mu)} & \text{if } -1 \leq \mathbf{x}_i(0) < \mathbf{x}_j(0) \leq 1. \end{cases}$$

For very small ϵ and $\mu \in (0, 1/2)$, it takes a very large number of periods to reach consensus while to reach the bounds 1 and -1 the individuals take a finite number of time periods. In other words, $t_r < t_a$ because we can always find a small enough ϵ such that the inequality holds. Formally, we solve the inequality $t_a > t_r$ for $\epsilon > 0$, for the case where $\mathbf{x}_i(0) > \mathbf{x}_j(0)$ (similarly for the other case) and t_a at its lower bound:

$$\begin{aligned}
\frac{\log(\epsilon)}{\log(1 - 2\mu)} &> \frac{\log\left(\frac{2 - \mathbf{x}_i(0) - \mathbf{x}_j(0)}{\mathbf{x}_i(0) - \mathbf{x}_j(0)}\right)}{\log(1 + 2\mu)} \\
\Leftrightarrow \epsilon &< \exp\left(\frac{\log\left(\frac{2 - \mathbf{x}_i(0) - \mathbf{x}_j(0)}{\mathbf{x}_i(0) - \mathbf{x}_j(0)}\right) \log(1 - 2\mu)}{\log(1 + 2\mu)}\right).
\end{aligned}$$

A.4 Proof of Theorem 1

A.4.1 Comments to explain the theorem

A few comments are in order.

First, the entries of the hearing matrix \tilde{G} are all non-negative and all the diagonal entries are strictly positive. Moreover it has rows and columns that sum to one. Hence, the eigenvalues of \tilde{G} are all lower or equal to 1 and $\lim_{t \rightarrow \infty} \tilde{G}^t$ exists. The entry on the row i and column j of the matrix $\lim_{t \rightarrow \infty} \tilde{G}^t$ is the weight (between 0 and 1) that the opinion of individual i at period t^* has in the final opinion of individual j .

Second, the hearing matrix \tilde{G} is a reducible. To see this, recall that consensual individuals account for the opinions of all their neighbors, while Elite individuals only account for the opinions of neighbors who also express (when such neighbors exist). Hence, there always exists at least one path starting at a node that represents a consensual individual and that ends at a node representing an expresser. However, there does not exist any paths that start at a node representing an expresser and that end at a node representing a consensual player. In particular, a set of individuals $\mathcal{C} \subset N$ is called an *essential class* (Seneta (1981) Seneta [1981]) if there does not exist a path starting at an individual $i \in \mathcal{C}$ and ending at an individual $j \in N \setminus \mathcal{C}$.

Third, the multiplicity of the eigenvalue 1 is equal to the number of essential classes in the hearing matrix \tilde{G} . To see this simply, consider a circle as a network structure with exactly k individuals, where each individual has two neighbors and where initial opinions are such that each individual has at least one neighbor who is ideologically-opposed. For this network structure, given the expression threshold $\delta^* = 1$, all individuals choose to express. Since each individual has at least one ideologically-opposed neighbor, each individual reaches an extreme opinion of 1 or -1 after few periods of interaction. In this setting, individuals no longer take into account the opinions of other Elite individuals in the long-run and each individual forms an essential class on their own. Hence, the hearing matrix \tilde{G} is simply the identity matrix of size k and the multiplicity of the eigenvalue 1 is exactly k . Beyond this example, the only case where an essential class is not a singleton is the case where there is a group of individuals that form a connected set of Elite individuals (see definition 1) that are like-minded. In other words, there exists a path connecting each pair in this connected set of Elite individuals at each time period of interaction, but no paths from any of those Elite individuals to an individual outside this set. I summarize the above discussion in the following theorem and provide a proof which makes use of standard linear algebra results.

A.4.2 Proof

Part (i) convergence: let λ be an eigenvalue of the matrix \tilde{G} . Recall that the algebraic multiplicity of λ is the number of times it is repeated as a root of the characteristic polynomial and the geometric multiplicity of λ is the maximum number of linearly independent eigenvectors associated with λ . An eigenvalue is semi-simple if its algebraic multiplicity is equal to its geometric multiplicity (definitions p.510, chapter 7, Meyer (2000) Meyer [2000]). For $\tilde{G} \in \mathbb{R}^{n \times n}$, $\lim_{t \rightarrow \infty} \tilde{G}^t$ exists if and only if $\rho(\tilde{G}) < 1$ (the spectral radius) or else $\rho(\tilde{G}) = 1$ where $\lambda = 1$ is the only eigenvalue on the unit circle and $\lambda = 1$ is semi-simple (see *Limits of Powers* page 630, chapter 7, in Meyer (2000) Meyer [2000]). Moreover, for every stochastic matrix, the spectral radius is 1 and it is semi-simple (p.696, Chapter 8 in Meyer (2000) Meyer [2000] or see Corollary 2, page 2214, in Ding and Rhee (2011) J.Ding and Rhee [2011]). Since, matrix \tilde{G} is a stochastic matrix, it has a spectral radius of 1 and it is semi-simple. Therefore, \tilde{G} is a convergent matrix.

Part (ii) spectral projector: when $\lim_{t \rightarrow \infty} \tilde{G}^t$ exists, it is equal to the spectral projector associated with eigenvalue 1 (again see p.630, chapter 7, in Meyer (2000) Meyer [2000]).

Reminder from p.629 Meyer (2000) Meyer [2000]. Recall that a row stochastic matrix A can be decomposed using its Jordan form J :

$$J = \begin{bmatrix} I_{p \times p} & \mathbf{0} \\ \mathbf{0} & \mathbf{K} \end{bmatrix}$$

where $I_{p \times p}$ is the identity matrix of size p , with p the algebraic multiplicity of the eigenvalue 1 and \mathbf{K} a diagonal matrix with entries corresponding to remaining eigenvalues which are strictly smaller than 1. Hence, $\tilde{G}_\theta^t = PJ^tP^{-1}$. Now write $P = (P_1, P_2)$ where P_1 are the columns that correspond to the eigenvectors associated with the eigenvalues 1 and P_2 are the columns that correspond the eigenvectors associated with the remaining eigenvalues which are strictly smaller than 1. Similarly $P^{-1} = Q = (Q_1; Q_2)$ with Q_1 the lines associated with the eigenvalues 1. Since K^t vanishes when t is large because all the diagonal entries are strictly smaller than one, $\lim_{t \rightarrow \infty} \tilde{G}_\theta^t = P_1Q_1$ which is the spectral projector of the eigenvalue 1.

Part (iii). The multiplicity of the eigenvalue 1 is equal to the number of essential classes. Recall that from Seneta (1981) Seneta [1981]: *we say that i leads to j and write $i \rightarrow j$ if there exists an integer $m \geq 1$ such that $t_{ij}^m > 0$ (chain between i and j). We say that i and j communicate if $i \rightarrow j$ and $j \rightarrow i$ and write in this case $i \leftrightarrow j$. The index i is called essential when : $i \rightarrow j$ implies $i \leftrightarrow j$ and there is at least one j such that $i \rightarrow j$. It is therefore clear that all essential indices (if any) can be subdivided into essential classes in such a way that all the indices belonging to one class communicate, but cannot lead to an index outside the class.*

The matrix \tilde{G} can contain several essential classes that are either: (i) singletons, when an expresser has reached the upper or lower bound of the opinion interval and is no longer updating their opinion (one self-loop), or (ii) contain more than one expresser, this occurs when individuals within a connected set of Elite individuals are like-minded and keep updating their opinions until they reach consensus. Each sub-matrix of \tilde{G} corresponding to an essential class is row stochastic, because (a) there are no outgoing edges from the members of the essential class to members outside the class by definition and (b) the matrix \tilde{G} is row stochastic. Furthermore, a sub-matrix corresponding to a single self communicating class is irreducible. Hence, each sub-matrix corresponding to an essential class is an irreducible aperiodic (because of self-loops) stochastic sub-matrix and by the Perron-Frobenius theorem of non-negative matrices, each such sub-matrix has an associated eigenvalue 1 that is simple.

Finally, the matrix \tilde{G} can be interpreted as an n -state Markov chain. From Seneta (1981) we further know that *if an n -state MC contains at least two essential classes of states, then any weighted linear combination of the stationary distribution vectors corresponding to each such class, each appropriately augmented by zeros to give an $(n \times 1)$ vector, is a stationary distribution of the chain.*

A.5 Supplementary figures for the simulations

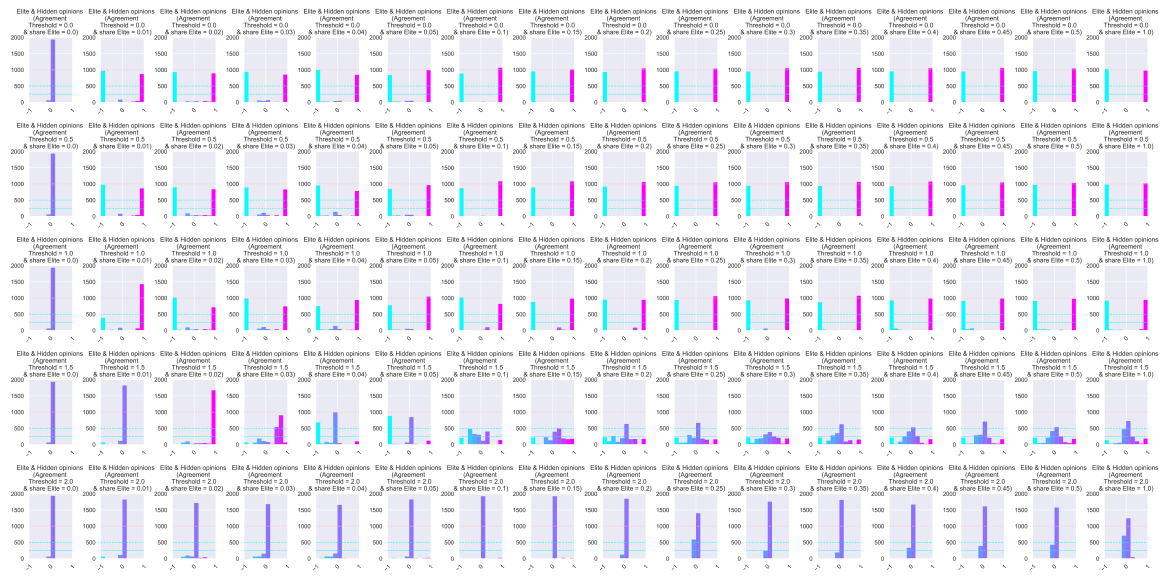


Figure 5: Distribution of final opinions for all combinations of share of Elite individuals and agreement thresholds, when Elite individuals are the most central and the network is generated with 1 hub.



Figure 6: Distribution of final opinions for all combinations of share of Elite individuals and agreement thresholds, when Elite individuals are the most central and the network is generated with 2 hubs.



Figure 7: Distribution of final opinions for all combinations of share of Elite individuals and agreement thresholds, when Elite individuals are chosen at random and the network is generated with 1 hub.



Figure 8: Distribution of final opinions for all combinations of share of Elite individuals and agreement thresholds, when Elite individuals are chosen at random and the network is generated with 2 hubs.