Hidden Opinions

Shaden Shabayek*

January 19, 2022

Abstract

This paper widens the scope of analysis of opinion dynamic models by introducing a novel heuristic: individuals choose to express their opinion or hide it, as a function of their local popularity. Intuitively, individuals who hide their opinion could be interpreted as individuals who have a low popularity such that even if they speak-up (or *tweet*) they will not be heard. Local popularity captures the idea that immediacy causes higher influence. Locally popular individuals express their opinion and can interact with like-minded or ideologically-opposed peers, namely expression entails debates and discussions. In the presence of hidden opinions, I show that the interactions between locally popular individuals and the magnitude of their influence explains whether consensus or polarization prevails. The primary mechanism at play is that the influence structure allows for consensus of opinion locally but communication between ideologically opposed expressers lead to global disagreement. The main contribution of this paper is to provide a unifying theoretical framework to assess different long-run opinion patterns with a focus on the topology of the network. I provide a measure of polarization and I run simulations to show the extent to which the topology of the network affects long-run opinion patterns.

Keywords: Naive learning, repulsive influence, opinion polarization.

JEL Classification numbers: D83, D91, Z1.

1 Introduction

Are all people equal on social media or do popular voices dominate the conversation? My paper investigates this question by introducing a new heuristic to the study of opinion formation in social contexts. Namely I introduce an *expression heuristic*, which assumes that individuals choose whether to express their opinion or hide it based on how popular they are within their social network. Individuals who hide their opinion can be interpreted as individuals who have a low popularity such that even if they speak-up they will not be heard or considered. One can think of an individual who has very few followers on Twitter. Alternatively, a second interpretation can be that hiding one's opinion is less costly than expressing it. The cost of expression can be the time

*Université Paris1 Panthéon-Sorbonne, Paris School of Economics, 48 boulevard Jourdan 75014 Paris, France. Contact information: shaden.shabayek@sciencespo.fr.

spent arguing with more eloquent and persuasive peers or the cost of social isolation when one's opinion drifts from the average group viewpoint.

The *expression heuristic* departs from early models of opinion formation, such as French (1956) [18], Harary (1959) [13], DeGroot (1971) [8]. 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, I introduce two types of individuals with different opinion updating rules. In doing so, I provide a unifying framework to assess different long-run opinion patterns, such as consensus, polarization of opinion or total disagreement.

Formally, I develop a model where a set of individuals are connected through an undirected network and exchange opinions about a given issue over a large number of periods. Each individual is exogenously allocated an initial opinion in a bounded interval, which represents their initial stance or attitude concerning the issue to be discussed. The influence of each individual is summarized by a given real valued centrality measure.[1] As mentioned, there are two types of individuals: individuals who *hide* and individuals *express* their opinion. If the influence of an individual is below a given threshold (hereafter *expression-threshold*), then they *hide* their opinion and update their opinions à la DeGroot. Otherwise, individuals *express* (hereafter *expresser*) and update their opinions according to a law of motion, which could incorporate either assimilation of opinions or distancing. At a given period, two directly connected expressers can either be like-minded or ideologically-opposed depending on the difference of their respective viewpoints. In particular, expressers only interact with peers who also choose to express. To be more precise, individuals who express undergo an attractive effect (positive influence or assimilation) when interacting at a given period with like-minded peers who also choose to express. But when they interact with expressing peers that are ideologically-opposed, they undergo a repulsive effect (negative influence or distancing) and push each other to the upper and lower bound of the opinion interval. Intuitively, expression allows for a debate or a discussion to take place. This discussion could lead to agreement or can escalate to conflict.

I provide a novel unifying framework that explains how different opinion patterns prevail in the long-run. In the presence of hidden opinions, the study of the interactions between locally *popular*

[1] In section 5 I define a centrality measure, which I call *local popularity*, and motivate its use.

individuals who interact with like-minded or ideologically-opposed peers can explain whether consensus or polarization prevails. Since influence is stronger locally, clusters can form. But some members within a given cluster who are popular enough and interact with ideologically-opposed peers can fall into disagreement, which causes opinions to polarize across clusters. Expressers only pay attention to neighbors who also express, while consensual individuals who update their opinions à la DeGroot pay attention to all of their direct social contacts. This paper makes several contributions. First, I start by characterizing the opinions of expressers. Second I characterize the overall process of interpersonal influence with the two types of individuals and the two opinion updating rules. I show that opinions converge in the long-run. Third, I provide simulations to relate the specific topology of the network to the level of opinion polarization; defined as the variance in the long-run opinions.

When an expresser has no neighbors who is also an expresser, this individual remains stubborn and their long-run opinion corresponds to their initial opinion. Furthermore, expressers can have neighbors who are locally popular and choose to express. For instance, consider three expressers 1, 2 and 3. Individual 1 has a direct link to 2 and 2 has a direct link to 3, but 1 and 3 are not directly connected. Assume that all the remaining friends of those three individuals are not popular enough to express. The opinion update of individual 1 depends on their opinion difference with individual 2 and indirectly depends on the interaction between individual 2 and 3. Individuals 1, 2 and 3 form a set of connected expressers. Namely, this set is defined as follows: all the elements of the set are nodes (individuals) who choose to express and any two nodes are connected by a path of expressers within this same set.

With this definition in hand, proposition 1 considers a group of individuals who belong to the same set of connected expressers and shows that, if each member of the set has only like-minded neighbors, then the long-run opinions of each member is exactly the average of the initial opinions of the members of this set. Each expresser of the group undergoes only the attractive effect. In particular, the group can have members who are ideologically-opposed that are not neighbors. But since they are not directly linked, they do not repulse each other to the upper and lower of the opinion interval and become extreme. Hence the whole group can reach consensus by gathering different viewpoints. This result sheds light on the design of media tools or the formation of discussion groups for initiatives related to participatory democracy.[2]

Proposition 2 characterizes the long-run opinions within a connected set of expressers, when a pair of initially ideologically-opposed neighbors belong to the group. I show that, generically, the long-run opinions of the group of connected expressers reach the upper or lower bound of the opinion interval. That is, all the members of the connected group of expressers become extreme. One special case where moderate opinions of expressers can survive in the long-run occurs when an

[2]For example, think of the Citizens' convention for climate (*convention citoyenne pour le climat*) in France, where 150 citizens were randomly selected to work in smaller groups on different propositions.

expresser has like-minded neighbors. But they are indirectly connected to at least two ideologically-opposed expressing individuals. This case depicts a political left-right spectrum where parties on the far right and far left are ideologically-opposed and interact often together but between both parties, many moderate parties survive.

Unlike linear models with assimilative influence, my model is non-linear; in the sense that the influence structure or the *hearing* matrix[3] can vary across periods in the presence of expressers who repulse or attract each other. Two expressers can initially be like-minded and influence each other positively. But in subsequent periods they can become ideologically-opposed and influence each other negatively, if one of them is repulsed by another expressing neighbor. Intuitively, this situation occurs when two individuals are somehow like-minded initially. But one of them starts adopting an extreme point of view in an unreasonable fashion under the influence of a third expressing peer, so her friend starts bringing to the table arguments which support the opposing view. In other words, the weights in the hearing matrix can depend on the opinion itself, as in Hegselmann and Krause (2002) [14].

Lemma 1 shows that the time or number of periods that two directly linked expressers take to repulse each other to the upper and lower bound of the opinion interval when they are ideologically-opposed, is smaller than the time they take to reach an agreement when they are like-minded. This motivates the study of the process of interpersonal influence with both types of individuals, starting from the time period by which repulsion opportunities in the course of a discussion are exhausted. I build a hearing matrix which accounts for both updating rules and show that opinions converge in the long-run. In particular, I show that the opinions of consensual individuals (who choose to *hide*) vanish in the long-run and remain hidden forever. Their long-run opinions simply become convex combinations of opinions of expressers to whom they are connected. Long-run consensus of opinions corresponds to an opinion vector such that the difference between any two opinions is small. Long-run bi-polarization of opinions corresponds to an opinion vector such that opinions belong to two groups of equal size: within each group opinions are close enough (below a threshold) and across both groups the opinion difference is large.

Proposition 3 provides necessary and sufficient conditions for consensus to prevail in the long-run. First, long-run consensus of opinions is obtained when there is a unique expresser. Second, consensus prevails when there exists multiple sets of connected expressers such that: (i) within each set, all members have like-minded neighbors, (ii) the average of initial opinions across two sets of connected expressers is close enough. Long-run opinions of consensual individuals depend on the long-run opinions of expressers to whom they are connected. Since those expressers all have similar viewpoints, consensual individuals adopt similar viewpoints when they take the average of opinions

[3]This matrix provides information on who listens to whom or who pays attention to whom and the magnitude of this attention.

expressed within their social circle. The *type* of consensus obtained will depend first on how many expressers there are in the society. Their number is directly related to the structure of the network. Second, it depends on the expressed opinions within the neighborhood of each expresser.

Mapping back a long-run bi-polarized opinion vector to an exact set of network structures and initial opinion distributions is tedious[4]. Consensual individuals are influenced by expressers to whom they are linked directly and indirectly. Put simply, they receive influence of different magnitudes from many expressers depending on how far or close they are from those expressers in the network. That is, one needs to account for all the possible paths of different length connecting consensual individuals to expressers who may be ideologically-opposed.

I provide compelling necessary conditions for long-run bi-polarization of opinions in lemma 2. I show that if a long-run opinion vector is bi-polarized then necessarily, (i) individuals who remain moderate in the long-run do not exist, (ii) both extreme influence groups influence an equal share of the society. The first condition means that there does not exist individuals who receive an equal amount of influence from two ideologically-opposed extreme opinion groups. Those individuals occupy a very particular location in the network, because they are not locally popular enough to express and they are equally influenced by two ideologically-opposed groups of expressers. Those individuals could be interpreted as neutral TV hosts, or non-biased journalists or intermediaries in general.

Simulations. I explore the model through simulations.[5] The objective is to relate the topology of the network to the long-run opinion patterns. Polarization of long-run opinions is measured by taking the variance of final opinions. I generate initial opinions uniformly at random in a bounded interval and study the evolution of opinions of a large set of individuals in scale-free networks. The degree distribution within scale-free networks follows a power law. Due to this inequality in the degree distribution, expressers and consensual individuals co-exist. In particular, this exercise shows that average polarization level can be relatively low even if expressers are densely connected among each other. This happens precisely for network topologies where consensual individuals are connected to many influence sources.

Related literature. The study of opinion dynamics is a multi-disciplinary topic. Different fields such as economics (learning in networks, for surveys see Golub and Sadler (2017) [12] and Acemoglu and Ozdaglar (2011) [2]), sociology (the community cleavage problem, see Flache et al. (2017) [9] or Friedkin (2015) [11]), statistical physics and computer science (community detection, Malliaros and Vazirgiannis (2013) [21] survey the literature) have tackled this problem from different

[4]Hegselmann and Krause (2002) [14] have a non-linear model and they discuss this point extensively: *though elementary, the model is nonlinear in that the structure of the model changes with the states of the model given by the opinions of the agents (see Section 2). Not only that helpful mathematical tools like Markov chains are no longer applicable, it turns out, moreover, that rigorous analytical results are difficult to obtain.*.

[5]The Matlab code can be found in the following repository https://github.com/shadenshabayek/Hidden-Opinions.

angles.

This paper is closely related to a strand of the literature on Naive Learning which introduces *stubborn* agents or agents that remain attached to their initial opinion to a certain extent, in order to model disagreement (See Friedkin and Johnsen (1990) [10] and Friedkin(2015) [11], Acemoglu et al (2013) [1]). Two papers that are closest to mine are Yildiz et al. (2013) [32] and Sadler (2019) [25]. Both papers introduce *stubborn* agents in a voter model set-up where opinions are discrete and can take only two discrete real 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 a stubborn player can have on others' long-run opinions, to their popularity.

My paper also fits in the family of bounded confidence models (See Hegselmann and Krauss (2002) [14], Jager and Amblard (2005) [15]). The key ingredient of those models is to consider the difference between the opinions of individuals when opinion updating is taking place. In particular, Hegselmann and Krauss (2002) [14] consider a model where agents update their opinions by taking an average over the opinions of neighbors whose opinion difference falls within a confidence interval. When neighbors opinions fall outside the confidence interval they are ignored. I extend this literature by introducing an opinion updating rule which treats the opinions of neighbors differently, depending on whether the opinions of neighbors fall within or outside the confidence interval.

Furthermore my paper is related to an active line of research about disagreement in social contexts and oppositional identity. Both literatures use different methodologies (dyadic interaction between agents versus interaction with a proportion of agents) but address very similar questions. Melguizo (2018) [22] studies persistent disagreement. She allows interactions and attitudes to coevolve, hence departing from the time independent weights used in averaging neighbors' opinions in models à la DeGroot. The key ingredient used is to assume that each individual has several attributes. Relations with other individuals sharing similar attributes become more intense, while relationships with dissimilar others deteriorate. Furthermore, disagreement can modeled by introducing repulsive or negative influence when an individual interacts with dissimilar others or by modeling individuals who are similarity biased (See Flache et al. (2017) [9] for a survey). As for oppositional identity, Bisin et al. (2016) [5] provide a model which incorporates cultural conformity and cultural distinction, in the context of mariage choice within the same ethnic group or outside. Individuals select their optimal choice by considering the psychological cost of interacting with the proportion of dissimilar others.

Finally, explaining polarization has been tackled by a handful of recent papers, yet there is no consensus in the literature about its main drivers. Bolletta and Pin (2019) [6] introduce a network formation model and argue that under certain conditions when agents optimally choose their links,

the network can become disconnected and consensus of opinions cannot be reached. Banisch and Olbrich (2019) [3] explain the emergence of polarization by introducing reinforcement learning, where agents optimally adopt one viewpoint when they get positive feedback from peers. But their focus is not on the network structure itself and how it could be one of the drivers, in particular they fix a random geometric network to account for the structure of interactions.

Outline. The remainder of the paper is organized as follows. Section 2 reviews relevant literatures in social psychology that lay the ground for the main behavioral assumptions of my model. In section 3, I present the model and I characterize the overall process of interpersonal influence in section 4. In section 5, through simulations I relate the topology of networks to long-run opinion patters. Section 6 concludes.

2 Related work in social psychology: Hidden Profiles

Stasser and Titus (1985) [29] document how individuals in social contexts, do not always share their own opinion or 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, Taylor and Hanna (1989) [28] 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 of opinion. In fact, Larson et al. (1996) [17] 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.

Using the findings of the above literature, I introduce an expression heuristic to a dynamic

opinion formation model. An individual chooses to express her opinion or hide it based on a popularity measure that is meant to capture different hierarchical and expertise levels.

3 The model

3.1 Set-up

A group of individuals $N = \{1, \ldots, n\}$ is embedded in a connected and symmetric network G of interpersonal relationships, with typical entries $g_{ij} = g_{ji} \in \{0, 1\}$. Each node represents an individual. The set of friends of individual $i \in N$ is denoted by $N_i = \{j \in N : g_{ij} = 1\}$ and $d_i = |N_i|$ is the cardinality of N_i. For all $i \in N$, I assume that $g_{ii} = 1$. A chain of friends of friends of length l between two individuals $i \neq j \in N$, hereafter called *path*, is defined as follows: there exists a sequence of distinct individuals $i = k_0, k_1, \ldots, k_l = j \in N$ such that $g_{ik_1} \times g_{k_1k_2} \times \ldots \times g_{k_lj} > 0$. The influence of each individual in the network G, is measured by a given real valued centrality measure which depends on the network topology, denoted by $\delta_i(G)$ for individual $i \in N$.

3.2 Expression heuristic

Each individual $i \in N$ is endowed with an exogenous initial opinion $\alpha_{i,0} \in [-1, 1]$ which represents their stance about a given issue.[6] Individuals exchange opinions about the issue over $t \geq 0$ periods and they can be like-minded or ideologically-opposed.

Definition 1 *At period $t \geq 0$, for $\tau \in (0, 1)$, two individuals $i \neq j \in N$ are :*

(i) like-minded, whenever $|\alpha_{i,t} - \alpha_{j,t}| < \tau$,

(ii) ideologically-opposed, whenever $|\alpha_{i,t} - \alpha_{j,t}| \geq \tau$.

At period t, each individual $i \in N$ observes the opinions in their neighborhood, denoted by $\alpha_{j,t} \in [-1, 1]$ for all $j \in N_i$. At period $t + 1$, each individual updates their opinion according to a rule which depends on their centrality:

$$\begin{cases} \text{Hide (rule 1)} & \text{if } \delta_i(G) < \delta^*(G) \\ \text{Express (rule 2)} & \text{if } \delta_i(G) \geq \delta^*(G), \end{cases}$$

where δ^* denotes an exogenous threshold. In the remainder of the paper, the set of individuals who express will be labelled $E = \{i \in N, \text{ s.t. } \delta_i(G) \geq \delta^*(G)\}$.

[6] Can model the issue as $\theta \in [-1, 1]$ and make the network depend on it.

Hide (rule 1 or DeGroot): when an individual hides their opinion, they update their opinion by taking the average of opinions expressed within their social circle at the previous period:

$$\alpha_{i,t} = \overline{\alpha}_{i,t-1} = \frac{1}{d_i} \sum_{j \in N} g_{ji} \alpha_{j,t-1}. \tag{1}$$

Express (rule 2): when an individual expresses, the opinion updating process depends on: (i) the opinions of neighbors, who also express and (ii) whether two neighbors are like-minded or ideologically opposed (see Definition 1). More precisely, expression allows for a debate and the opinion updating process of expressers follows a law of motion which allows for agreement and disagreement. It incorporates an attractive (agreement) and a repulsive (disagreement) effect among direct neighbors who express their opinions. For all $i \in E$ and $\mu \in (0, 1/2)$:

$$\alpha_{i,t} = \alpha_{i,t-1} + \mu\Big(\sum_{j \in \underline{N}_{i,t-1}} (\alpha_{j,t-1} - \alpha_{i,t-1}) - \sum_{j \in \overline{N}_{i,t-1}} (\alpha_{j,t-1} - \alpha_{i,t-1}) \Big) \text{ s.t. } \alpha_{i,t} \in [-1, 1], \tag{2}$$

Where $\underline{N}_{i,t} = \{j \in N_i \cap E, |\alpha_{i,t} - \alpha_{j,t}| < \tau\}$ and $\overline{N}_{i,t} = \{j \in N_i \cap E, |\alpha_{i,t} - \alpha_{j,t}| \geq \tau\}$.

Notice that, expressing neighbors receive influence from their own expressing neighbors (if any). Hence, an initially like-minded expressing neighbor of individual i can become in subsequent periods ideologically-opposed, if their opinion difference with i becomes larger than τ.

Example 1 *Consider a network G with only two connected individuals 1 and 2 who choose to express. Suppose that $\tau = 0.5$ and initial opinions are $\alpha_{1,0} = -0.7$ and $\alpha_{2,0} = 0.7$. In period $t = 1$, $\alpha_{1,1} = \alpha_{1,0} + \mu(\alpha_{1,0} - \alpha_{2,0}) = -0.7(1 + \mu) - \mu 0.7 < \alpha_{1,0} = -0.7$ and $\alpha_{2,1} = \alpha_{2,0} + \mu(\alpha_{2,0} - \alpha_{1,0}) = 0.7(1 + \mu) + \mu 0.7 > \alpha_{2,0} = 0.7$. The updated opinion of individual 1 becomes more negative or pushed-down towards -1, while the updated opinion of individual 2 more positive or pushed-up towards 1. Individuals 1 and 2 repulse each other.*

4 Opinion Dynamics

Given the above model, I am interested in studying the long-run opinions, for a network structure G and a vector of initial opinions α_0. Then I characterize in section 4.2 the overall dynamics of opinions and show that long-run opinions always converge.

4.1 Long-run opinions of expressers

Individuals who choose to express update their opinions at each time period according to the law of motion (2). Consequently, their opinion update will directly depend on the opinions of neighbors

who choose to *express* (if any). The opinions of the latter will depend on the opinions of their own neighbors (if any) who choose to *express* and so on. Recall that E is the set of individuals who choose to express because their local popularity is higher than the expression threshold δ^*. In order to account for the indirect effect of the opinions of expressers on other individuals who also express, I give a formal definition of a connected set of expressers.

Definition 2 (Connected set of expressers) *Let G be a given network structure and let $\mathcal{E}$ be a set of individuals such that (i) $\forall i \in \mathcal{E}$, $\delta_i \geq \delta^*$, (ii) $\mathcal{E} \subseteq E$ and (iii) any pair of individuals $i \neq j \in \mathcal{E} \subseteq E$ are connected in network G by a path of length l_{ij} of expressers belonging to the set $\mathcal{E}$, that is $\exists g_{ik_1} \times g_{k_1k_2} \ \ldots \times g_{k_lj} > 0$ for $k_1, \ldots k_l \in \mathcal{E} \subseteq E$. When expressers $k_1, \ldots k_l \in \mathcal{E} \subseteq E$ all have like-minded neighbors, I say that individual i is linked to individual j through a path of like-minded expressers and denote it by l_{ij}^+.*

In a given network there could be multiple connected sets of expressers $\mathcal{E}_1, \ldots, \mathcal{E}_k$ such that $E = \bigcup_{i=1}^{k} \mathcal{E}_i$. Those sets could be singletons or they could contain more than one expresser. In the network G_1 in figure 16, each of both individuals 1 and 2 form a set of connected expresser(s) on their own: $\mathcal{E}_1 = \{1\} \subset E$ and $\mathcal{E}_2 = \{2\} \subset E$. Moreover, $E = \mathcal{E}_1 \cup \mathcal{E}_2 = \{1, 2\}$.

For a given set of connected expressers, if each pair of neighbors are like-minded, then their long-run opinions will be the average of their initial opinions. I formalize this idea in the following proposition.

Proposition 1 *Let G be a network of interpersonal relationships, α_0 an initial opinion vector and consider $\mathcal{E} \subseteq E$ a given set of connected expressers.*

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

$$\forall t \ \geq 1, \ \alpha_{i,t} = \alpha_{i,0}$$

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

$$\exists t^* \geq 1, \ \forall t \geq t^*, \ \alpha_{i,t} = \frac{\alpha_{i,0} + \alpha_{j_1,0} + \ldots + \alpha_{j_\kappa,0}}{\kappa}$$

Proof. See Appendix 7.1. ■

When a set of connected expressers contains at least one pair of ideologically-opposed neighbors, *generically* long-run opinions become extreme, that is they take the value 1 or -1. For very specific initial opinion distributions some expressers within the same set can hold moderate opinions (strictly

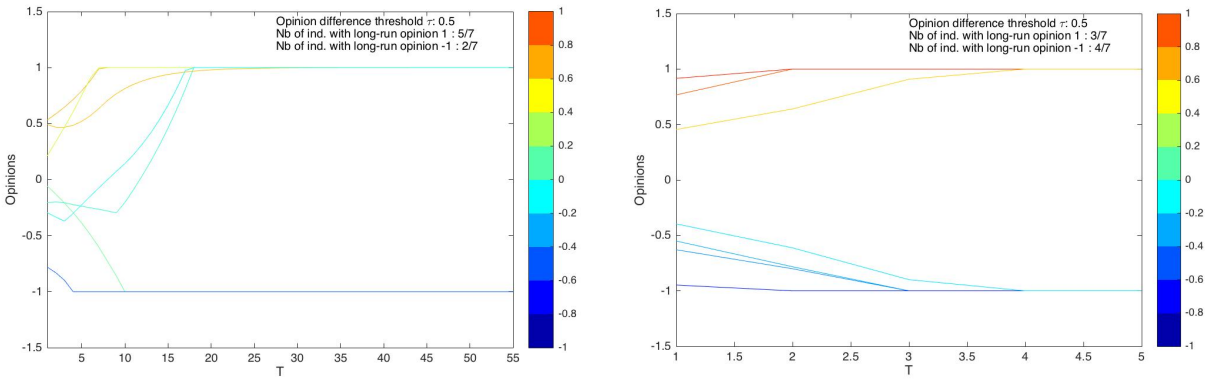


Figure 1: Non-monotonic and monotonic opinion updating in a circle

between -1 and 1) in the long-run. I give two examples to make clear: (i) why opinion updating within a set of connected expressers is non-monotonic, (ii) why moderate opinions can survive in the long-run. Finally, I summarize the discussion in proposition 2. In all the subsequent examples, I assume that the opinion difference of two directly linked individuals at a given time period is $\tau = 0.5$.

4.1.1 Non-monotonic opinion updating

The coexistence of pairs of directly linked individuals who are like-minded and pairs of directly linked individuals who are ideologically-opposed within the same connected set of expressers can lead to either monotonic or non-monotonic opinion updating. Non-monotonic opinion updating can occur when an individual is indirectly connected to an ideologically-opposed individual through a chain of like-minded expressers.

To see this, consider a wheel G of seven individuals $\{1, \ldots, 7\}$ such that $g_{12} = g_{23} = \ldots = g_{67} = g_{71} = 1$ and the remaining entries of G are zero. All individuals express because their local popularity is higher or equal to δ^* and together all seven form a unique connected set of expressers. Consider the following initial opinion vector:

$$\alpha_0 = (-0.21, -0.06, 0.53, 0.49, 0.21, -0.78, -0.29).$$

Individuals 5 and 6 are initially ideologically-opposed with initial opinions $\alpha_{5,0} = 0.21$ and $\alpha_{6,0} = -0.78$. The remaining individuals $\{1, 2, 3, 4, 7\}$ have initially like-minded neighbors. The evolution of opinions is plotted in the left panel in figure 1, with time periods on the x-axis and opinions on the y-axis. The colormap on the east side of the figure is associated with the opinion interval $[-1, 1]$ and the colors of the curves correspond to initial opinions.

Individuals 5 and 6 are pushed to the upper and lower bound of the opinion interval $[-1, 1]$ after few periods of interaction, as they are ideologically opposed. The evolution of their opinion is monotonic. The opinion of individual 5 becomes more and more positive, while the opinion of individual 6 becomes more and more negative. However, individuals 1 and 7 update their opinions non-monotonically. Individual 7 and 6 are initially like-minded and the attractive effect is at play in the first few periods of interaction. Hence the opinion of individual 7 starts becoming more negative, because it converges towards the opinion of individual 6. But after a few periods, individual 6 becomes extreme by reaching the lower bound -1. The opinion difference with their direct neighbor individual 7 becomes larger and larger, until a point where this difference becomes higher than the threshold τ and individual 7 starts getting repulsed by the extreme opinion of individual 6. In other words, the repulsive effect takes over. Intuitively, this situation occurs when a given individual i is having a discussion with an individual j who is initially more or less like-minded, but individual i is more neutral than j. As the discussion goes on, individual j becomes too extreme in an unreasonable fashion, such that individual i starts defending the opposite view.

Finally, in the right panel of figure 1, I provide an example where opinion updating is monotonic. Initial opinions are given by $\alpha_0 = (-0.55,\ 0.77,\ -0.63, -0.9478,\ 0.92, -0.4,\ 0.45)$. In other words, each individual has an ideologically-opposed neighbor and opinions converge *monotonically* to the upper or lower bound.

4.1.2 Moderate long-run opinions

There exist initial opinion distributions such that *moderate* opinions (i.e. with an opinion that is neither 1, nor -1) survive in the long-run within a set of connected expressers containing at least one ideologically-opposed pair. To see this, consider again a wheel G of 7 individuals such that $g_{12} = g_{23} = \ldots = g_{67} = g_{71} = 1$ and all the remaining entries of G are zeros. Suppose that initial opinions are : $\alpha_0 = (-0.9, -0.7, -0.4, -0.1, 0.2, 0.5, 0.8)$. Individuals 1 and 7 are initially ideologically-opposed. Individuals $\{2, \ldots, 6\}$ all have neighbors who are initially like-minded. The evolution of opinions is plotted in the left panel of figure 2. Individuals $i \in \{2, \ldots, 5\}$ remain moderate in the long-run and never adopt an extreme opinion of 1 or -1. To understand why moderate opinions can persist in the long-run, first notice than individual 7 is repulsed by her direct ideologically-opposed neighbor individual 1. Second, the opinion difference of individuals 6 and 7 remains smaller than τ even when 7 becomes extreme. Similarly, the opinion difference of individuals 1 and 2 remains smaller than τ even when 1 becomes extreme. Intuitively, extreme individuals 1 and 7 influence in opposite *directions* the chain of like-minded individuals that separate them so that in the long-run, each of these intermediate individuals have like-minded neighbors and remain moderate. This situation bears a resemblance to the left-right political spectrum in some countries, where the moderate parties survive in the long-run. To summarize, moderate opinions of expressers can persist in the long-run if such expressers only have like minded-neighbors and are linked to at

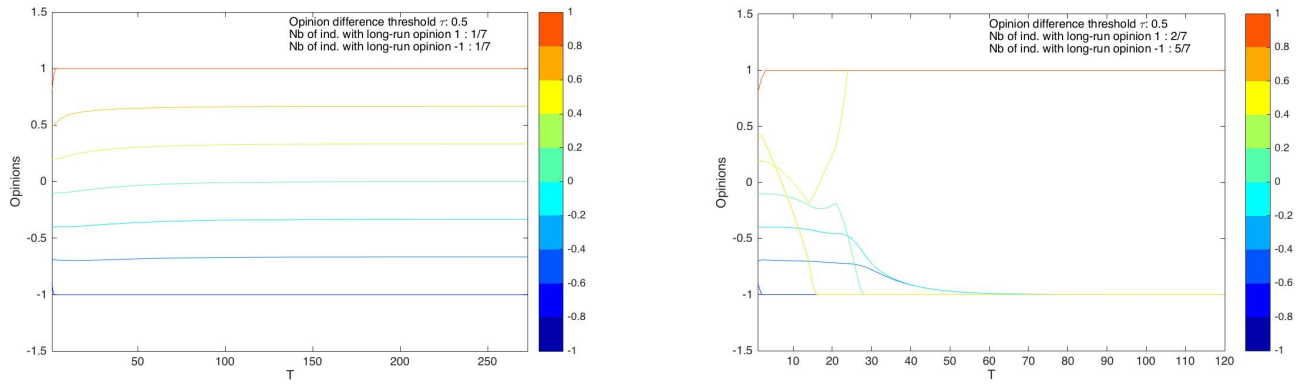


Figure 2: Moderate expressers

least two ideologically-opposed expressers by a path of like-minded neighbors.

In the right panel of figure 2, all individuals hold extreme opinions in the long-run. Similarly to the setting in the left panel of figure 2, individuals $i \in \{2, \ldots, 6\}$ have initially like-minded neighbors and are connected to individuals 1 and 7 directly or indirectly through a chain of like-minded neighbors. But the opinion difference of individuals 6 and 7 is initially larger than in the previous case (left panel of figure 2) and as individual 7 is pushed towards holding an extreme opinion, her opinion difference with individual 6 becomes larger and larger, up to a point where this difference becomes larger than τ.

Finally, consider a wheel G of three individuals such that $g_{12} = g_{23} = 0$ and all the remaining entries of the network G are zeros. All three individuals express and all three form together a unique connected set of expressers. Suppose that the initial opinion of individual 2 is $\alpha_{2,0} = 0$. Suppose that individuals 1 and 3 hold respectively the following initial opinions $\alpha_{1,0} = -1$ and $\alpha_{3,0} = 1$. In this case, the long-run opinion of individual i is $\alpha_{i,\infty} = 0$ because she is equally repulsed by both neighbors but in opposite directions.

To summarize, when a given connected set of expressers is formed of only initially like-minded expressers, influence is positive and opinions get *attracted* to the average opinion of the group. However, when the set contains at least one ideologically-opposed pair of neighbors, the repulsive and attractive effect can both be at play. I show in the following proposition, that when a set contains at least one ideologically-opposed pair of expressers, generically opinions get to the upper and lower bound of the opinion interval. Intuitively, some expressers are not updating their opinions by taking a convex combination of the opinions of their neighbors. Hence at each time step the length of the interval given by the opinion difference of an ideologically-opposed pair keeps growing until it

reaches the bounds. Such expressers also drive their like-minded neighbors to reach the upper and lower bound of the opinion difference. This happens, because when those like-minded neighbors are updating their opinion, they assign a positive weight to the opinions of their neighbors. Except that the opinion of the latter keeps getting pushed to either bounds under the effect of their own ideologically-opposed neighbor, rather than converging to the opinion of their like-minded neighbor. I given in points (i) and (ii) of the proposition, the two exceptions where some opinions remain moderate as explained in the previous examples.

Recall that $\underline{N}_{i,t}$ and $\overline{N}_{i,t}$ are the sets of neighbors of $i \in N$ who are respectively like-minded and ideologically-opposed at period $t \geq 0$. Denote the set of expressers that have at least one initially ideologically-opposed neighbor by $IO(\mathcal{E}) = \{i \in \mathcal{E} : \exists j \in \mathcal{E}, i \in \overline{N}_{j,0}, j \in \overline{N}_{i,0}\}$.

Proposition 2 *Let G be a network of interpersonal relationships, α_0 an initial opinion vector and consider $\mathcal{E} \subseteq E$ a set of connected expressers. Suppose that $IO(\mathcal{E}) \neq \emptyset$.*

(i) If there exists at each $t \geq 0$ at least two paths of expressers with only like-minded neighbors connecting $i \notin IO(\mathcal{E})$ to at least two elements of $\{i_1, i_2 \dots i_k\} \subset IO(\mathcal{E})$ then $\alpha_{i,\infty} \in conv(\alpha_{i_1,\infty}, \dots, \alpha_{i_k,\infty})$.

(ii) Let $i_1, i_2 \dots i_k \in IO(\mathcal{E})$. If there exists $i \in IO(\mathcal{E})$ such that (a) $\alpha_{i,0} = 0$, (b) $\forall t \geq 0$ $\underline{N}_{i,t} = \emptyset$, (c) and $\sum_{j \in \overline{N}_{i,t}} \alpha_{j,t} = 0$ then $\alpha_{i,\infty} = 0$.

(iii) Otherwise if (i) and (ii) don't hold and $IO(\mathcal{E}) \neq \emptyset$ then $\forall i \in \mathcal{E}$, $\alpha_{i,\infty} \in \{-1, 1\}$.

Proof. See Appendix 7.2. ■

Furthermore, when a pair of expressers are initially like-minded they take a longer *time* to reach consensus but when the pair is ideologically-opposed they disagree at a much faster rate. I formalize this idea in the following lemma.

Lemma 1 *Let $i \neq j \in \mathcal{E} \subset E$ such that $|\mathcal{E}| = 2$ and $t_{\alpha_\infty} = \min\{t : |\alpha_t - \alpha_\infty| < \epsilon\}$. If α^a_∞ is the long-run opinion vector when $|\alpha_{i,0} - \alpha_{j,0}| < \tau$ and α^r_∞ is the long-run opinion vector when $|\alpha_{i,0} - \alpha j, 0| \geq \tau$ then $t_{\alpha^r_\infty} < t_{\alpha^a_\infty}$.*

Proof. See Appendix 7.3. ■

4.2 The process of interpersonal influence

I build a hearing matrix which takes into account who listens to whom. This hearing matrix takes into account the two opinion updating rules depending on the type of individual: expresser (updates

according to the law of motion (2)) or consensual (updates à la DeGroot). In particular I study the long-run behavior starting at the time period where pairs of ideologically-opposed expressing neighbors have repulsed each other towards the most extreme opinion. As I have shown in the previous section, given the parameter μ in the law of motion (2), a pair of ideologically-opposed 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 neighbors who express, 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 $\tilde{g}_{ij}$.

Consensual individuals. For each individual $i \in N$ such that $\delta_i < \delta^*$, the entries in the hearing matrix become $\tilde{g}_{ij} = g_{ij}/d_i$, $\forall j \in N$.

Expressers. For individuals who choose to express there are four cases to consider.

(i) For all $i \in N$ in a connected set of expressers $\mathcal{E}$ 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 expressers $\mathcal{E}$ 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 expressers $\mathcal{E}$ such that $|\mathcal{E}| = \kappa > 1$ with ideologically-opposed neighbors, i.e. $\forall i \neq j \in \mathcal{E}$ and $j \in N_i \cap \mathcal{E}$, $|\alpha_{i,t^*} - \alpha_{j,t^*}| \geq \tau$, 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 $k \in N_j$.

Remark 1 *All the entries of the hearing matrix $\tilde{G}$ are positive and each row sums to one.*

Example 2 *Consider network G in figure 16. The hearing matrix $\tilde{G}$ has the following entries for expressers 1 and 2 who are both stubborn: $\tilde{g}_{11} = \tilde{g}_{22} = 1$ and $\tilde{g}_{1j} = \tilde{g}_{2j} = 0$ for all $j \in N_1 \cup N_2$. For the consensual individuals $i \in \{3, 4, 5\}$, $\tilde{g}_{ii} = \tilde{g}_{i2} = 1/2$. For $i \in \{6, 7, 8\}$, $\tilde{g}_{ii} = \tilde{g}_{i1} = 1/2$. Finally $\tilde{g}_{99} = \tilde{g}_{91} = \tilde{g}_{92} = 1/3$. All the remaining entries are zeros.*

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:

$$\alpha_{t+1} = \tilde{G}\alpha_t \tag{3}$$

By induction, the opinions at period $t \geq t^*$ are given by $\tilde{G}^t \alpha_{t^*}$ and the limit yields the long-run opinions. 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 $\lim_{t\to\infty} \tilde{G}^t$ exists. The entry on the row i and column j of the matrix $\lim_{t\to\infty} \tilde{G}^t$ 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 expressers 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) [26]) 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 expressers 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 expressers (see definition 2) that are like-minded. In other words, there exists a path connecting each pair in this connected set of expressers at each time period of interaction, but no paths from any of those expressers 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.

Theorem 2 *Given $\alpha_{t^*} \in [-1,1]^n$ a vector of opinion at period t^* and a hearing matrix $\tilde{G}$ associated with the network structure G, the long-run opinions are :*

$$\alpha_\infty = (\lim_{t\to\infty} \tilde{G}^t)\alpha_{t^*} = \mathcal{G}\mathbf{a}_{t^*} < \infty,$$

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. See Appendix 7.4 ■

The columns corresponding to consensual 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 expressers, they have at least one strictly positive entry. In particular, the long-run opinions of consensual individuals are exactly convex combinations of initial opinions of expressers. In other words, the long-run opinion of consensual individuals is affected by the long-run opinions of all the expressers to whom they are connected to through a path of other consensual 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 expresser has in the long-run opinions of other individuals. This motivates the introduction of the following statistic.

Definition 3 (Spectral influence) *Given a network structure G, a hearing matrix $\tilde{G}$ and its limit $\mathcal{G}$, the spectral influence of individual $i \in N = \{1, \ldots, n\}$ is:*

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

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

Example 3 *Consider network G in figure 16 and suppose that the initial opinions of expressers 1 and 2 are $\alpha_{0,1}$ and $\alpha_{0,2}$. Since both are not directly connected nor are they connected via a chain of expressers, each of them forms an essential class on their own. The spectral projector $\mathcal{G}$ associated to the eigenvalue 1 of the hearing matrix $\tilde{G}$ is a symmetric matrix of size 9 and is given by : $\mathcal{G}_{11} = \mathcal{G}_{i1} = 1$ for $i \in \{6, 7, 8\}$. That is the long-run opinion of individuals 6, 7 and 8 is fully determined by the initial opinion of individual 1. As for individual 2, $\mathcal{G}_{22} = \mathcal{G}_{j2} = 1$ for $j \in \{3, 4, 5\}$. The long-run opinion of individual 9 is equally determined by the initial opinions of expressers 1 and 2, that is $\mathcal{G}_{91} = \mathcal{G}_{92} = 1/2$. All the remaining entries of the matrix $\mathcal{G}$ are zero. Hence, long-run opinions are $\alpha_{\infty,i} = \alpha_{0,1}$ for $i \in \{1, 6, 7, 8\}$, $\alpha_{\infty,j} = \alpha_{0,2}$ for $j \in \{2, 3, 4, 5\}$ and $\alpha_{\infty,9} = \frac{1}{2}\alpha_{0,1} + \frac{1}{2}\alpha_{0,2}$. Expressers 1 and 2 have an identical spectral influence equal to $s_1 = s_2 = 1/2$.*

4.3 Patterns of long-run opinions: consensus and bi-polarization

In this section, I characterize patterns of long-run opinions. I focus on two patterns. First, I study consensus as a benchmark, where all the individuals in the long-run become like-minded.

Definition 4 (Consensus) *Long-run opinions form consensus if $\forall i \neq j \in N$, $|\alpha_{\infty,i} - \alpha_{\infty,j}| < \tau$.*

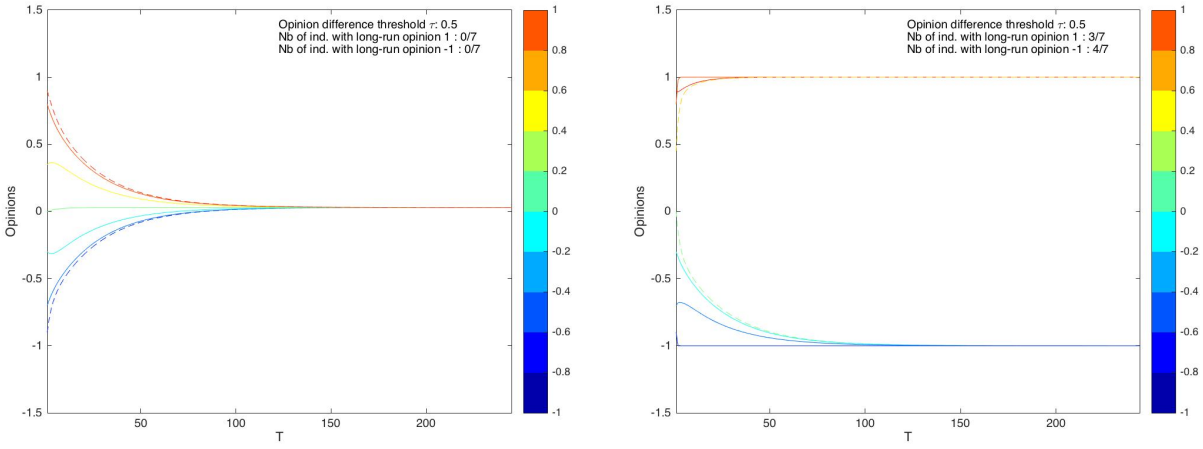


Figure 3: Consensus and bi-polarization

The above definition of consensus reflects the idea that the limiting opinions need not to be identical but the pairwise difference needs to be at most τ. In other words, the matrix $\mathcal{G}$ (see Theorem 2) in general will not have identical rows.[7] Second I consider bi-polarization[8] of long-run opinions, that is given a network structure and an initial opinion vector, the society of N individuals gets divided into two opinion groups of approximatively equal size with strong agreement within each group and disagreement across both groups.

Definition 5 (Bi-polarization) *Let $\epsilon > 0$ be a strictly positive real number. Long-run opinions are bi-polarized if (i) the society of N individuals is divided into two groups of size $|N_1|$ and $|N_2|$ such that $||N_1| - |N_2|| < \epsilon$, (ii) for $k \in \{1,2\}$, $\forall i \neq j \in N_k$, $|\alpha_{\infty,i} - \alpha_{\infty,j}| < \tau$, (iii) $|\overline{\alpha}_{\infty,N_1} - \overline{\alpha}_{\infty,N_2}| \geq \tau$ where $\overline{\alpha}_{\infty,N_i}$ is the average long-run opinion of group N_i for $i \in \{1,2\}$.*

In general, the higher the number of expressers induced by a given network structure G and the more the long-run opinion pattern depends on the distribution of initial opinions within each expresser's neighborhood. To see this, consider a situation where most individuals can *express*, as it is the case in regular network structures. When each individual has like-minded neighbors, then the group converges to consensus; even if there exists a pair of expressers that are not directly

[7]The spectral projector has identical rows if and only if the algebraic multiplicity of the unit eigenvalue is 1; meaning that computing the perron vector is sufficient to obtain the long-run opinions. But having more than two sets of connected expressers is translated into an algebraic multiplicity of the unit eigenvalue strictly higher than 1.

[8]I focus on opinion bi-polarization rather than the more general case of opinion polarization, where a polarized society is one that is divided into a small number (larger than two) of opposed groups. The special case of bi-polarization fits applications of the model where it would take a very large group - e.g. at least half of the population of interest - to over-turn a policy or to elect a president or to produce a divided public opinion. For example, one can think of Brexit, the election of Trump, the implementation of a Carbon tax in France or even at the beginning of 2020 divided views about risk reducing measures regarding the coronavirus.

connected and that hold ideologically-opposed views. For example, let G be a line network with 7 individuals such that $g_{12} = g_{23} = g_{34} = g_{45} = g_{56} = g_{67} = 1$ and the remaining entries are all zeros. Consider the following initial opinion vector $\alpha_0 = (-0.9, -0.7, -0.3, 0, 0.45, 0.8, 0.9)$. For $\tau = 0.5$, individuals reach consensus even though individuals 2 and 6 are ideologically-opposed, as shown in the left panel of figure 3. However, for the same network structure, if the initial opinion vector is $\alpha_0 = (0, -0.3, -0.7, -0.9, 0.8, 0.9, 0.45)$, then this group of individuals does not reach consensus, as shown in the right panel of figure 3. This is because individuals 3 and 4 repulse each other towards long-run opinions $a_{\infty,3} = -1$ and $a_{\infty,4} = 1$. In this case, long-run opinions become $\alpha_\infty = (-1, -1, -1, -1, 1, 1, 1)$ which corresponds to a bi-polarized group of individuals.

4.3.1 Consensus

Consensus of opinion depends on the number of expressers in the network, how they are connected (or not) to each other and the distribution of their initial opinions. Namely, consensus is reached whenever there is a unique expresser in the network or several expressers with like-minded neighbors who express, and for any two disjoint sets of connected expressers, individuals are in average like-minded across both sets. In the former case, the initial opinion of the unique expresser fully determines the long-run opinions of the remaining consensual individuals. While in the latter case, long-run opinions of individuals can be different but that difference is at most τ.

Proposition 3 *Given a network structure G, an initial opinion vector $\alpha_0 \in [-1, 1]^n$ and the set of expressers $E \subset N$, long-run opinions form a consensus if and only if exactly one of the following statements holds:*

(i) $|E| = 1$

(ii) There exists $\kappa \geq 1$ sets of connected expressers s.t. $E = \bigcup_{k=1}^{\kappa} \mathcal{E}_k$, $\forall k_i \neq k_j \in \{1, \ldots, \kappa\}$, $|\overline{\alpha}_{0,\mathcal{E}_{k_i}} - \overline{\alpha}_{0,\mathcal{E}_{k_j}}| < \tau$ and for each set $\mathcal{E}_{k_i} \subset E$, $\forall i \neq j \in \mathcal{E}_{k_i}$ and $j \in \mathcal{E}_{k_i} \cap N_i$, $|\alpha_{0,i} - \alpha_{0,j}| < \tau$.

Proof. See Appendix 7.5 ■

Corollary 1 (unique opinion leader) *Let $\alpha_0 \in [-1, 1]^n$ be a vector of initial opinions. If the network structure G is a star where the central node has a degree $k > 1/\delta^*$ then the long-run opinions a_∞ form a consensus.*

These necessary and sufficient conditions map long-run opinions to the network structure and the initial distribution of opinions. To see this, recall that choosing to express or hide one's opinion is determined by local popularity, which is a network statistic. Moreover, in the case where the structure of the network of interpersonal relationships allows the existence of more than one expresser, the distribution of initial opinions along with the structure of connections among expressers

become crucial in reaching consensus. In particular, consensus can never be reached if there exists a pair of ideologically-opposed expressers who are directly linked.

4.3.2 Bi-polarization

Unlike consensus, characterizing the initial opinion distribution of expressers within a network isn't sufficient to explain the emergence of a bi-modal long-run opinion distribution; that is, *a central interval that is sparsely populated and left and right intervals that are densely populated,* (Friedkin (2015) [11]). The characterization is challenging because the emergence of a bi-polarized society depends on where expressers are placed in the network, how they interact together and how many people they influence. If extreme opinion groups of expressers do form, the network structure together with the opinion difference threshold τ determine to which extreme opinion group, consensual individuals will belong.

Recall that the long-run opinion of each consensual individual is exactly a convex combination of the opinions of expressers to whom they are directly or indirectly linked by a path of other consensual individuals. For some network structures and expressers' opinion distributions, consensual individuals can receive influence from ideologically-opposed groups of expressers. In which case, consensual individuals remain moderate in the long-run. In figure 16, individual 11 holds a long-run opinion of *zero*, when individuals 1 and 2 hold respectively opinions 1 and -1. In other words, the existence of long-run moderate individuals can block the split of the population into two ideologically-opposed groups.

Intuitively, the network position of moderate consensual individuals allows them to receive influence from different opinion groups. They are initially linked to both influence groups and they don't express an opinion themselves. Such individuals could be interpreted as intermediaries or neutral TV hosts or moderators, who act as buffers against opinion polarization.

Nevertheless, the absence of moderate consensual individuals isn't sufficient for long-run opinions to be polarized. In particular, one needs to account for the size of both opinion groups. The split of expressers into two extreme opinion groups isn't sufficient to cause opinion bi-polarization. Both extreme opinion groups need to have influence over an equally large number of consensual individuals, so that the society becomes divided. In other words, for consensual individuals to belong to one of both extreme opinion groups, they need to receive *enough* influence from the members of one group so that they hold similar extreme opinions in the long-run. Finally, for two extreme opinion groups to form, at least one pair of ideologically-opposed expressers need to interact together or there must exist at least two individuals initially at each end of the opinion spectrum. Recall that the influence of each expresser $i \in E \subset N$ is summarized by their spectral influence s_i in definition 3. Define the set of expressers that hold long-run opinion 1 and -1 respectively by: $E^+ = \{i \in E : \alpha_{i,\infty} = 1\}$ and $E^- = \{i \in E : \alpha_{i,\infty} = -1\}$.

Lemma 2 *Suppose that* $E = E^+ \bigcup E^- \subset N$, $E^+ \neq \emptyset$ *and* $E^- \neq \emptyset$. *If long-run opinions are bi-polarized then*

(i) $|\sum_{i\in E^+} s_i - \sum_{i\in E^-} s_i| < \epsilon$,

(ii) $\nexists k \in N$ *s.t.* $1 - \tau/2 \leq \sum_{i\in E^+} \mathcal{G}_{ki} \leq \tau/2$.

Proof. See Appendix 7.6. ■

5 Network topology and opinion patterns

In this section I explore the model through simulations. The objective is to relate the topology of the network to long-run polarization of opinions. Polarization of opinions is measured simply by looking at the variance in final opinions. Clearly, the maximum level of polarization is 1 and the minimum level is 0. To focus on the network topology, I generate a large number of initial opinions vectors (1000) distributed uniformly at random over $[-1, 1]$ for the same network structure with n individuals, then I compute the average level of polarization of the (1000) final opinion vectors obtained.

Formally, I study the evolution of opinions of a large set of individuals in scale-free networks, as it is a general family of networks where the topology allows for the co-existence of expressers and consensual individuals. I provide a number of (inter-related) network statistics that measure inequality in the degree distribution such as assortativity, the Gini coefficient of the degree distribution and per capita average degree. I also look at network statistics that measure connectivity such as neighborhood connectivity and average path length. I also look at whether expressers are densely or sparsely connected[9], by identifying the number of connected sets of expressers in Definition 2. The description of the network statistics is in appendix 7.7.

The degree distribution within scale-free networks follows a power law. Due to this inequality in the degree distribution, expressers and consensual individuals co-exist. Barabási and Albert (1999) [4] explain the emergence of such networks by the following two mechanisms: the network expands because new individuals (vertices) keep getting added to an existing network and they attach preferentially to individuals who are already well connected. The mechanism leading to the formation of links is out of the scope of this paper. The objective of the simulations is to make obvious the impact of the interaction structure - location in the network and initial opinion - of expressers on the long-run opinion pattern of the whole group of individuals. To that end, I generate scale-free networks and work with them as a snapshot at a given point in time where no

[9]Density or sparsity of connections among expressers means whether they are isolated or interact with other expressers. In a graph with k expressers that have no expressing neighbors, the subgraph of G restricted to expressers will have exactly k components. One can think of some δ^*-core and count the number of components to account for sparsity or density.

new individuals are added and no new links are created. Each network topology is generated by providing an initial number of hubs h and a number of nodes n_c a newly added node connects to.

I start by providing two examples with high and low average polarization levels with 100 individuals, $h = 5$ initial hubs and $n_c = 1$. Then I give aggregate statistics by varying the number of hubs h and the number of connections n_c of each newly added nodes. All figures appear after the conclusion.

5.1 Local Popularity part in intro

To account for centrality, I define and use a measure that is: (i) increasing in the degree of a given individual (immediacy), (ii) decreasing in the average degree of direct friends of a given individual (strength). In other words, *local popularity* is defined as the degree of an individual divided by the average degree of friends. A locally popular individual has many friends, who themselves have very few influence sources. Naturally local and global centrality measures capture different aspects of influence. For example, when considering eigenvector centrality, the influence of an individual is high, when they are linked to other influential individuals. Global centrality measures require the knowledge of the whole network structure and are pertinent for many applications, e.g. the importance of website pages. Nevertheless, a local centrality measure may be better suited for a model of exchange of opinions or attitudes. In particular, well established strands of literature in social psychology, reviewed in section 2, argue that immediacy and strength of interaction are associated with higher social influence. Hence, an individual who may not be *globally* influential on the scale of the entire social network, can still have a high impact over their direct friends and local clustering of opinions can occur. It is the idea of having a devoted fan base.

5.2 Local Popularity

Finally, the social status or relative expertise is represented by a local centrality measure which I call *local popularity*.

Assumption 1 *Each individual $i \in N$ knows d_i and d_j for all $j \in N_i$.*

Definition 6 (Local popularity) *Let $i \in N$ be an individual and N_i the set of her/his direct friends. Local popularity is:*

$$\delta_i = \frac{d_i - 1}{\frac{1}{d_i - 1} \sum_{j \neq i \in N_i} (d_j - 1)} \tag{4}$$

Notice that the network G is connected and contains self-loops the above centrality measure is always defined. Alternatively, without making any assumptions, for any individual $i \in N$, if $d_i = 1$ then set $\delta_i \equiv 0$.

5.3 Dynamic social Impact theory

Latané's Dynamic Social Impact Theory (1981) [19] suggests that social influence has three determinants: strength, immediacy and the number of influence sources. Strength refers to social status, level of expertise or persuasiveness, while immediacy refers to closeness in space, time or the possibility of direct contact. The theory bridges the influence processes at an individual level using these three determinants, to outcomes at the level of a social system. The main statement of the theory is that *total impact of a group of people on an individual is a multiplicative function of their strength, immediacy, and number.* Latané, Nowak and Liu (1994) [20] use this theory to study through simulations the dynamics of attitude change in groups and societies. Rather than studying attitude distribution as the usual percentage frequencies of different attitude choices, they study the distribution of attitudes in *space*. Using *immediacy*, they are able to explain phenomena such as attitude clustering because individuals are more influenced by nearby individuals. In particular, *distance* between persons is used to compute *immediacy* and it is used along with *strength* to compute for each person the total persuasive impact and supportive impact. *A person will change his/her position if and only if the total persuasive impact (the pressure to change to a different position) outweighs the pressure to main- tain one's own position (the strength of the initial position plus any supportive impact).* Using computer simulations, they find that individuals cluster in the social space in terms of position similarity.

To incorporate *immediacy* in my model, I use a local popularity measure as opposed to global measures (e.g. eigenvector centrality). Individuals choose to express when their local popularity is above a given threshold. To incorporate *strength*, the local popularity of a given individual is inversely related to the average level of neighborhood connectivity. Meaning that an individual will have a higher impact over their neighbors if those same neighbors are not exposed to many other influence sources.

5.4 Example

I select two structures with the same number of hubs to illustrate the results of the model before moving to aggregate statistics. The first network structure G_1 contains 14 expressers, represented by squares in figure 6. Out of those 14 expressers, 11 are linked through a path of expressers, highlighted in red. In other words, the network topology allows for interaction between the majority of expressers. There are a number of empirical papers that support the idea that ideologically-opposed individuals interact together often (e.g. Conover et al. (2011) [7]). The first panel of figure 6 displays the initial opinions as colors given by the $[-1, 1]$ colormap on the east side of the figure. The second panel displays the long-run opinions. The third panel shows the evolution of opinions, where dotted lines correspond to the evolution of opinions of consensual individuals and the solid lines correspond to the evolution of opinions of expressers.

The second structure G_2 contains 11 expressers and only two pairs of expressers are linked as shown in figure 7. I select those two network structures by generating for the same initial opinions, a large number of scale-free networks with $n = 100$, $h = 5$ and $n_c = 1$. Then I pick out two network structures with a high and low average level polarization. The average level of polarization in network G_1 over the 1000 runs is 0.84, as opposed to only 0.43 for the network structure G_2. Figures 4 and 5 show the distribution over the 1000 runs of the size of the group of individuals holding extreme opinion -1, the group of individuals holding extreme opinion 1 and the group of individuals who hold a moderate opinion. For network G_1 the average size of extreme groups is almost 50%, while in network G_2 this same share is occupied by the group of moderate individuals. To plot the distribution of the size of the opinion groups, I compute for each of the 1000 runs, the number of individuals within each of the three groups. Then I use this output to plot the histogram and the kernel density.

5.5 Agregate statistics

In order to study the impact of a specific topology of scale-free networks on the mean level of polarization, I vary the number of initial hubs h and the number of nodes n_c to which a newly node is added. Doing so allows me to look at network topologies with different levels of connectivity and a wide range of degree distributions. With $n = 200$ individuals, $h \in \{10, 20, 30, 40, 50, 60, 70\}$ and $c_n \in \{1, 2, \ldots, 8\}$ I obtain 56 scale-free network topologies[10] and I compute for each the mean level of polarization over 1000 runs with the same 1000 initial opinion vectors. I do so to get information about the topology of the network, that are not related to a specific distribution of a given initial opinion vector. Then I compare across those networks the average level of polarization and provide several network statistics. The network statistics could be divided in two groups: (i) (inter-related) measures of degree inequality (ii) measures of connectivity. Furthermore, I provide the α parameter based on a maximum likelihood estimator, which is the exponent of the power-law that fits best the degree distribution of the considered scale-free network topology. In figures 8 to 12 the markers correspond to different sizes of initial hubs h and colors correspond to the number n_c of connections a newly added node will have. For all figures, the y-axis corresponds to the level of mean polarization while the x-axis gives the value of the different network statistic that will be considered.

Opinion polarization is expected to be observed for more skewed degree distribution where expressers have a very high degree. Following Newman (2002) [24], *a network is said to show assortative mixing if the nodes in the network that have many connections tend to be connected to other nodes with many connections.* Unsurprisingly, figure 8 shows that the mean level of polarization is

[10] Clearly with exactly the same number of hubs and the same n_c one obtains for different runs different network structures. But the variance in the estimated α parameter is very small, meaning that they display similar degree distributions and levels of connectivities.

positively correlated with the α parameter and negatively correlated with the assortativity coefficient of the degree distribution. In particular, in the right panel of figure 8 the two topologies with $c_n = 1$ and respectively with $h = 20$ and $h = 30$ show a high level of polarization and disassortative mixing. Furthermore, the right panel of figure 10 and left panel of figure 11 indicate high levels of polarization for high values of the Gini coefficient of the degree distribution and very low network sparsity computed as the per capita mean degree.

The more expressers interact together, the greater should be the possibilities of disagreement and consequently opinion polarization. Surprisingly, the left panel of figure 9 shows that the lowest levels of polarization are observed for topologies where the number of sets of connected expressers is smaller than 2. This means that many expressers are connected among each other, yet polarization is low. The right panel of both figures 9 and 10 bring an explanation. For those same network topologies, the average path length and the diameter are very small (respectively smaller than 3 and smaller than 5). Hence, those network topologies with a low level of polarization display high connectivity. Thereby, expressers interact among each other and can fall into disagreement, but consensual individuals are exposed to more influence sources. Intuitively, for those network topologies, expressers don't have a *fan base* which is solely devoted to each of them. This explanation is also supported by figure 11 and the right panel of figure 12. In the right panel of figure 12 low levels of polarization are related to higher neighborhood connectivity. Yet for these high levels of neighborhood connectivity (pink, purple and blue markers), the number of expressers is dispersed over the whole range as shown in figure 11.

6 Conclusion and the way forward

This paper introduces two novel ingredients to classic opinion formation models with assimilative influence. It relates the opinion updating rule to the network topology, by introducing two types of individuals who either update their opinions à la DeGroot, or update their opinions using a law of motion which incorporates assimilation or distancing. The chosen rule for opinion updating depends on whether the individual is locally popular or not. I show that opinion patterns can be explained by focusing on how influential individuals interact among each other and how they influence the (less popular) masses of users. In particular, individuals who are not popular enough to express but are connected to two extreme opinion groups can obstruct full opinion bi-polarization. Those individuals are neutral TV hosts or journalists. By means of simulations, I show that when popular individuals are densely connected among each other but the overall level of connectivity in the network is relatively low, then the average level of polarization is high. However when *all* individuals are densely connected, consensual individuals have access to more influence sources and the average polarization level is low. This present model can be extended in several ways. The choice of expression can depend on the opinion itself. Namely, even if an individual is popular

enough, they may not express when they hold an opinion that is too far from the average of opinions within their social circle (including both popular and unpopular individuals). Another promising direction for a follow up paper, is to endogenize the network structure and allow agents who *hide* to be able to *express* after a certain number of periods depending on the evolution of opinions in their neighborhood.

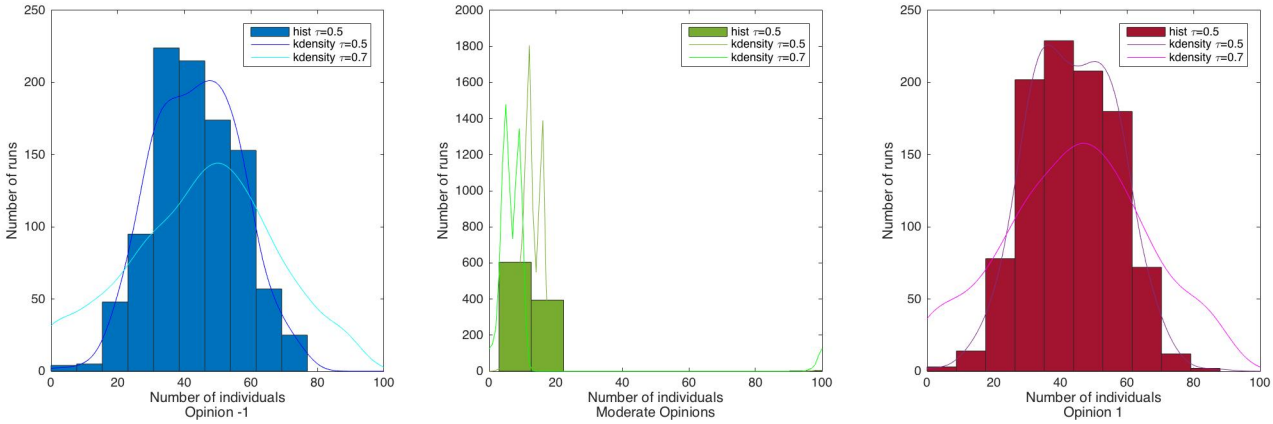


Figure 4: Kernel density over 1000 runs of groups of individuals with extreme opinion -1 (blue), 1 (red) and moderate (green) in scale-free network G_1

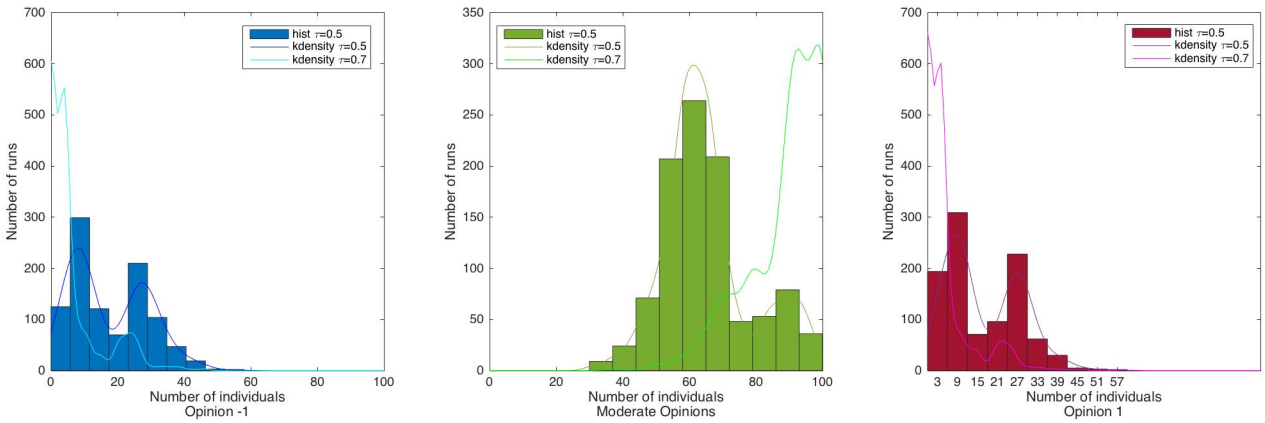


Figure 5: Kernel density over 1000 runs groups of individuals with extreme opinion -1 (blue), 1 (red) and moderate (green) in scale-free network G_2

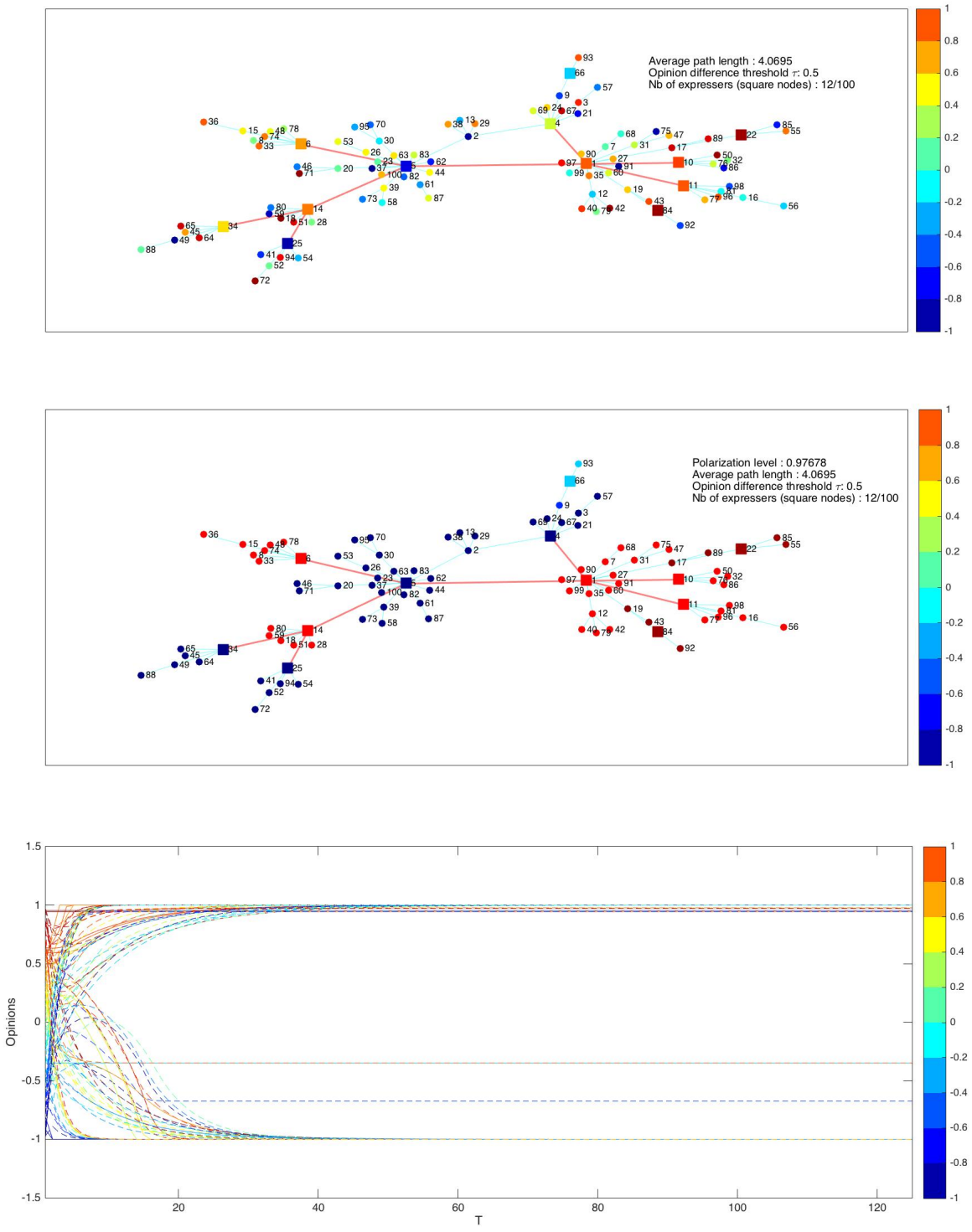


Figure 6: Initial opinions, final opinions and evolution of opinions in scale-free network G_1. Expressers are represented with a square marker. Links between expressers are highlighted in red.

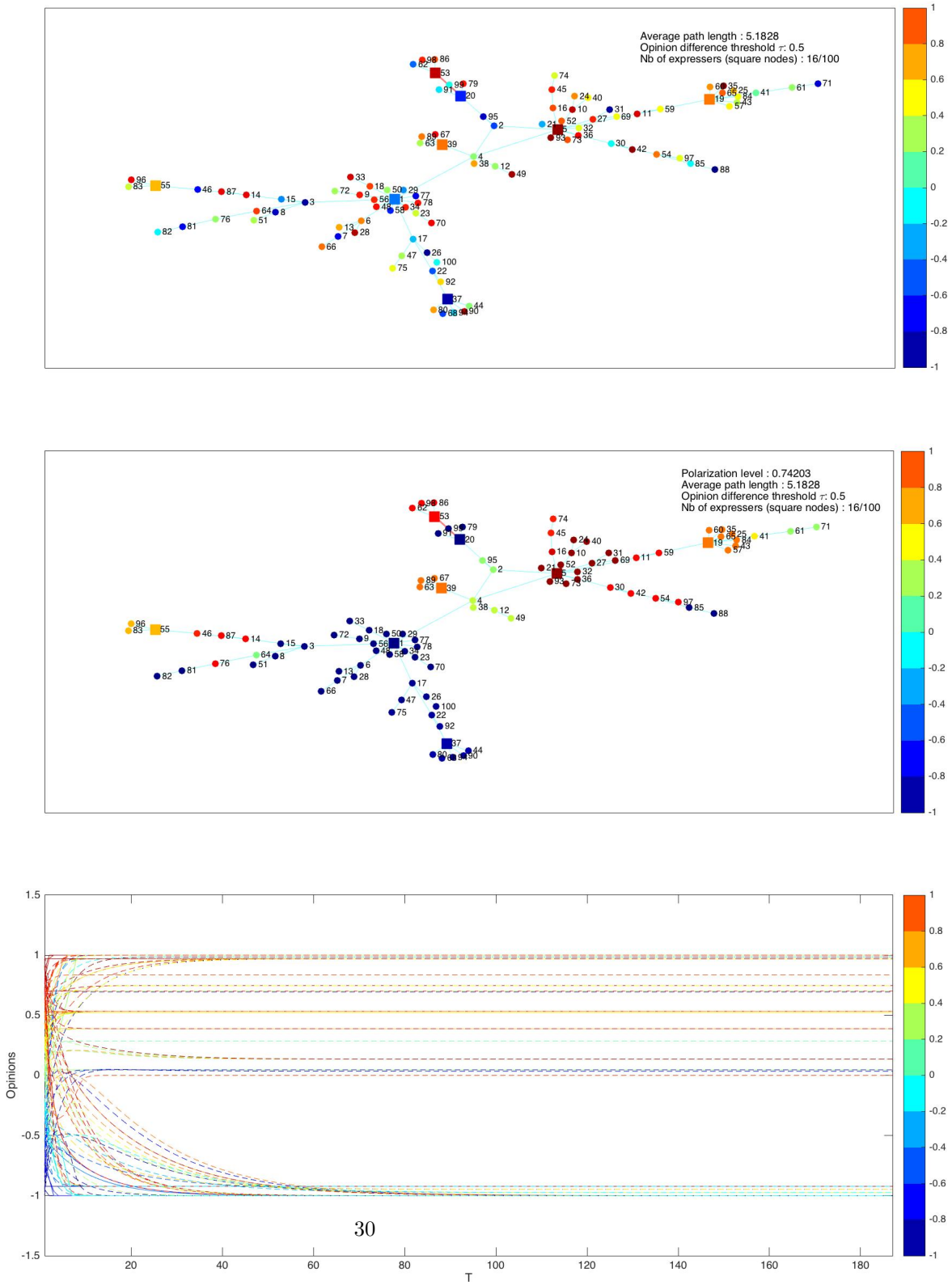
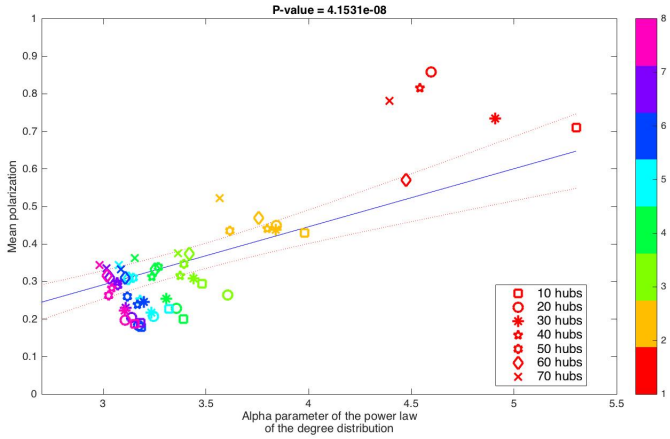


Figure 7: Initial and final opinions in the run with the highest level of polarization in scale-free network G_2. Links between expressers are highlighted in red.



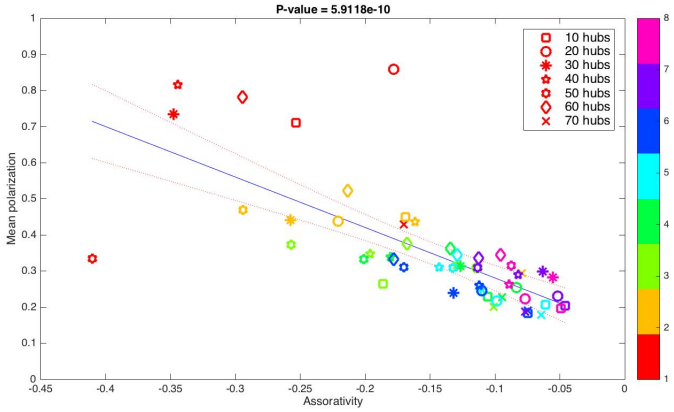
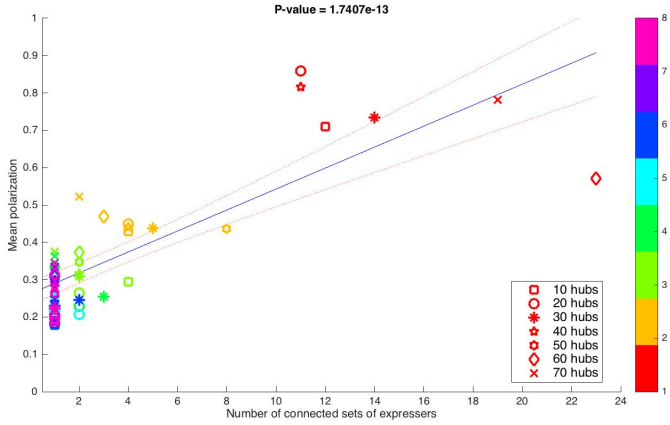


Figure 8: Markers correspond to different sizes of initial hubs h and colors correspond to the number n_c of connections a newly added node will have.



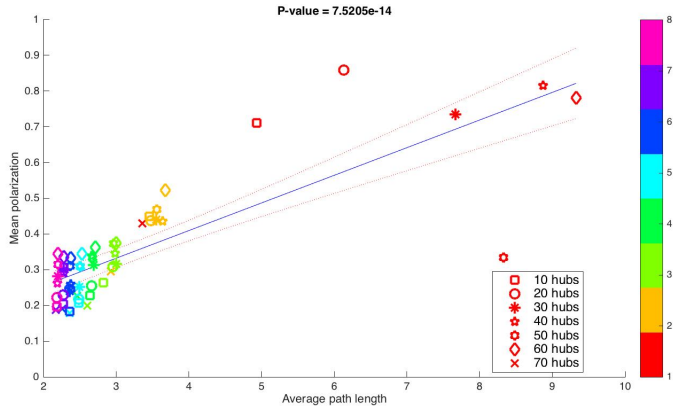
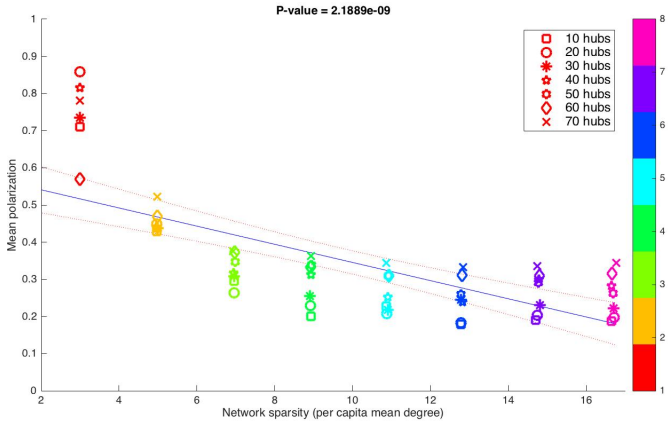


Figure 9



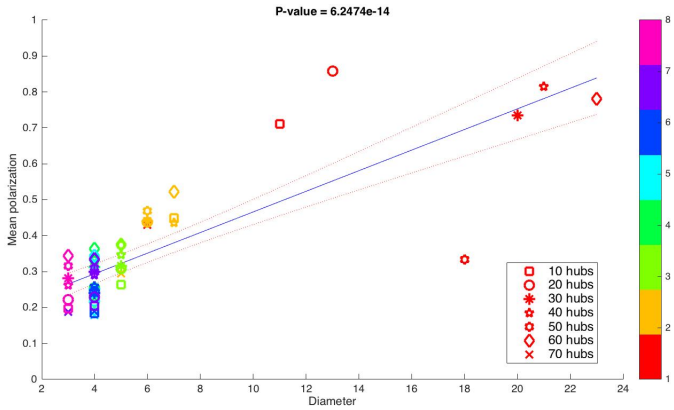
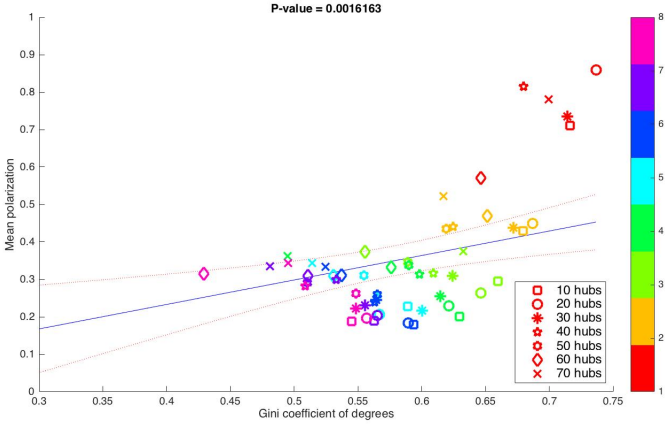


Figure 10



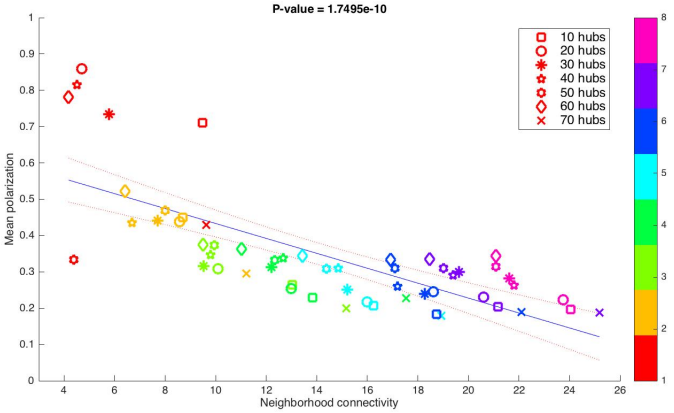


Figure 11

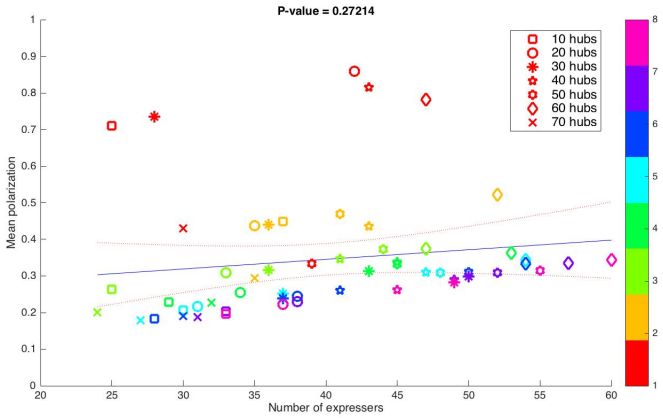

P-value = 0.27214
10 hubs
20 hubs
30 hubs
40 hubs
50 hubs
60 hubs
70 hubs
Mean polarization
Number of expressers
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
20
25
30
35
40
45
50
55
60
1
2
3
4
5
6
7
8

Figure 12

7 Appendix

7.1 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 $\alpha_{0,i}$.

(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 $\forall i \neq j \in \mathcal{E}$, $|\alpha_{0,i} - \alpha_{0,j}| < \tau$ then for $\mu \in (0, 1/\kappa)$ the opinions get updated in the following way:

$$\begin{cases} \alpha_{1,1} = \alpha_{0,1} + \mu \sum\limits_{j \neq 1 \in \mathcal{E}} g_{ij}(\alpha_{0,j} - \alpha_{0,1}), \\ \quad \vdots \\ \alpha_{1,\kappa} = \alpha_{0,\kappa} + \mu \sum\limits_{j \in \neq \kappa \in \mathcal{E}} g_{\kappa j}(\alpha_{0,j} - \alpha_{0,1}), \end{cases} \Leftrightarrow \begin{cases} \alpha_{1,1} = (1 - \mu d_i(\mathcal{E}))\alpha_{0,1} + \mu \sum\limits_{j \neq 1 \in \mathcal{E}} g_{1j}\alpha_{0,j}, \\ \quad \vdots \\ \alpha_{1,\kappa} = (1 - \mu d_\kappa(\mathcal{E}))\alpha_{0,\kappa} + \mu \sum\limits_{j \neq \kappa \in \mathcal{E}} g_{\kappa j}\alpha_{0,j}, \end{cases}$$

where $d_i(\mathcal{E}) = \sum_{j \in \mathcal{E}} g_{ij}$ corresponds to the number of expressers that are in the set of connected expressers $\mathcal{E}$ 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 :

$$\alpha_{t,\mathcal{E}} = M^t \alpha_{0,\mathcal{E}},$$

where $\alpha_{t,\mathcal{E}} = (\alpha_{t,1}, \ldots, \alpha_{t,\kappa})^T$, $\alpha_{0,\mathcal{E}} = (\alpha_{0,1}, \ldots, \alpha_{0,\kappa})^T$ 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 $\lim_{t\to\infty} \alpha_{t,\mathcal{E}} = \lim_{t\to\infty} M^t \alpha_{0,\mathcal{E}}$.

Claim 1 $\lim_{t\to\infty} M^t$ *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. In the case of the matrix M the Gershgorin disks[11] write for each $i \in \mathcal{E}$, $D_i = \{x \in \mathbb{R} : |x - m_{ii}| \leq \sum_{j \neq i} |m_{i,j}|\} = \{x \in \mathbb{R} : |x - (1 - d_i(\mathcal{E})\mu)| \leq d_i(\mathcal{E})\mu\}$. 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 $\lim_{t\to\infty} \alpha_{t,\mathcal{E}}$ is exactly the average of the initial opinions of individuals $1, \ldots, \kappa \in \mathcal{E}$.

[11] All the eigenvalues of M are real because M is a real symmetric matrix.

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

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) [31], which states that $\lim_{t\to\infty} M^t = \frac{1}{\kappa}\mathbf{1}_{\kappa,1}\mathbf{1}_{1,\kappa}$ 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 2. Hence the eigenvalue 1 is simple (Perron-Frobenius Theorem).

7.2 Proof of Proposition 2

Warm-up for the proof and notations. Recall that $\underline{N}_{i,t-1} = \{j \in N_i \cap E \text{ s.t. } |\alpha_{i,t-1} - \alpha_{j,t-1}| < \tau\}$ is the set of expressing neighbors of individual $i \in N$ such that their opinion difference is smaller than τ (like-minded). Similarly, recall that $\overline{N}_{i,t-1} = \{j \in N_i \cap E \text{ s.t. } |\alpha_{i,t-1} - \alpha_{j,t-1}| \geq \tau\}$. Clearly it follows that $\underline{N}_{i,t-1} \cup \overline{N}_{i,t-1} = N_i \cap E$. Without loss of generality, suppose that individual $i \in N$ chooses to express and belongs to the set of connected expressers $\mathcal{E} \subseteq E$ such that $|\mathcal{E}| > 1$ and there exists at least one pair of expressers in $\mathcal{E}$ who are neighbors and initially ideologically-opposed. Each individual $i \in \mathcal{E}$ updates their opinion at each time step according to the law of motion (2), given by :

$$\alpha_{i,t} = \alpha_{i,t-1} + \mu \sum_{j\in\underline{N}_{i,t-1}} (\alpha_{j,t-1} - \alpha_{i,t-1}) + \mu \sum_{j\in\overline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) \qquad \text{s.t } \alpha_{i,t} \in [-1,1] \tag{5}$$

$$\Leftrightarrow \alpha_{i,t} = \alpha_{i,t-1}(1 + \mu(|\overline{N}_{i,t-1}| - |\underline{N}_{i,t-1}|)) + \mu\Big(\sum_{j\in\underline{N}_{i,t-1}} \alpha_{j,t-1} - \sum_{j\in\overline{N}_{i,t-1}} \alpha_{j,t-1}\Big) \quad \text{s.t } \alpha_{i,t} \in [-1,1]. \tag{6}$$

The size of the sets $\overline{N}_{i,t-1}$ and $\underline{N}_{i,t-1}$ can vary between two periods because a like-minded neighbor at period $t-1$ can become ideologically-opposed at a subsequent period (see the examples in section 4.1). In other words, it's impossible to summarize the above system of equations by one time-invariant matrix because the entries or weights vary with time depending on the opinion differences between connected expressers and whether the opinion of a given individual has reached the upper

or lower bound. Hence we need to account for the *flow* of influence through *chains* of expressers within a given set $\mathcal{E}$. To do so, it's convenient to write the opinion of a given individual i at period t belonging to the connected set of expressers $\mathcal{E}$ as an **affine** combination of the opinions their neighbors j at period $t-1$:

$$\alpha_{i,t} = \sum_{j \in N_i \cap \mathcal{E}} a_{ij}(t-1)\alpha_{j,t-1}, \tag{7}$$

where

$$a_{ij}(t-1) = \begin{cases} \mu & \text{if } i \neq j, |\alpha_{i,t-1} - \alpha_{j,t-1}| < \tau \text{ and } \alpha_{i,t-1} \in (-1,1) \\ -\mu & \text{if } i \neq j, |\alpha_{i,t-1} - \alpha_{j,t-1}| \geq \tau \text{ and } \alpha_{i,t-1} \in (-1,1) \\ 1 + \mu(|\overline{N}_{i,t-1}| - |\underline{N}_{i,t-1}|) & \text{if } i = j \text{ and } \alpha_{i,t-1} \in (-1,1) \\ 0 & \text{if } i \neq j \text{ and } \alpha_{i,t-1} \in \{1+\epsilon, -1-\epsilon\} \\ 1 & \text{if } i = j \text{ and } \alpha_{i,t-1} \in \{1+\epsilon, -1-\epsilon\} \end{cases}$$

Let $A(t-1)$ be the matrix with entries $a_{ij}(t-1)$ for $i, j \in \mathcal{E}$. It follows that the opinions at period t of the expressers who belong to $\mathcal{E}$ can be written as:

$$\alpha_{\mathcal{E},t} = A(t-1)A(t-2)\dots A(0)\alpha_{\mathcal{E},0} = B(t-1,0)\alpha_{\mathcal{E},0},$$

where $B(t-1,0)$ is the matrix product of $A(t-1)A(t-2)\dots A(0)$. In other words it's a matrix that keeps track of the accumulated (positive and negative) weights between periods $t-1$ and 0. In particular, the entry $B_{ij}(t,t-1)$ reports the influence of j on i's opinion between periods t and $t-1$. Recall that the set of expressers that have at least one ideologically-opposed neighbor is given by:

$$IO(\mathcal{E}) = \{i \in \mathcal{E} : \exists j \in \mathcal{E}, i \in \overline{N}_{j,0}, j \in \overline{N}_{i,0}\}.$$

Part (i). Show that if i is connected indirectly to expressers with ideologically-opposed neighbors through a path of like-minded neighbors then $\alpha_{i,t}$ can be written as a convex combination of the opinions of neighbors for all $t \geq 1$. In this case individual i holds in the long-run a moderate opinion $\alpha_{i,\infty} \in (-1,1)$. Otherwise, their opinion keeps getting pushed to the upper or lower bound of the opinion interval and necessarily $\alpha_{i,\infty} \in \{-1,1\}$.

Claim 3 *Let $i_1, i_2 \dots i_k \in IO(\mathcal{E})$. If there exists at each $t \geq 0$ at least two paths of expressers with only like-minded neighbors connecting $i \notin IO(\mathcal{E})$ to at least two elements in $\{i_1, i_2 \dots i_k\}$ then $\alpha_{i,\infty} \in conv(\alpha_{i_1,\infty}, \dots, \alpha_{i_k,\infty})$.*

Proof. Suppose that $IO(\mathcal{E}) \neq \emptyset$, that is there exists at least one pair of neighbors in $\mathcal{E}$ that are initially ideologically-opposed. Let s be the time period such that for all $i_k \in IO(\mathcal{E})$, $a_{i_k j}(s) = 0$

for all $j \neq i_k$, $a_{i_k j}(s) = 1$ for $i_k = j$.

($\Rightarrow$) Suppose that there exists at each $t \geq 0$ paths of expressers with only like-minded neighbors connecting $i \notin IO(\mathcal{E})$ to at least two individuals $i_k \neq i_j \in IO(\mathcal{E})$. That is there exists $i, j_1, j_2 \ldots, j_k \notin IO(\mathcal{E})$ such that for all $t \geq s$, $B_{i,i_k}(t,s) > 0$ and $B_{i,i_j}(t,s) > 0$. Since,

1. the opinion of a given individual $j_k \in \{j_1, j_2, \ldots\} \cup \{i\} \notin IO(\mathcal{E})$ at period $t \geq s$ is a **convex** combination of the opinions of their neighbors at period $t-1$ (because they all have like-minded neighbors):
$$\alpha_{j_k,t} = \sum_{k \in N_i \cap \mathcal{E}} a_{j_k k}(t-1)\alpha_{k,t-1},$$
where $a_{j_k k}$ takes the value μ for any $k \neq j_k$ and $1 - |\underline{N}_{j_k,t}|\mu$ for $k = j_k$.
2. and at time period $t \geq s$, $\forall i_k \in IO(\mathcal{E})$, $\alpha_{i_k,t} = \alpha_{i_k,s} \in \{-1, 1\}$,
3. it follows that $\forall j_k \in \{j_1, j_2, \ldots\} \cup \{i\} \notin IO(\mathcal{E})$, $\alpha_{j_k,t} \in conv(\alpha_{i_1,s}, \ldots, \alpha_{i_k,s})$.
4. Moreover, since $IO(\mathcal{E}) \neq \emptyset$ and $\forall j_k \in \{j_1, j_2, \ldots\} \cup \{i\} \notin IO(\mathcal{E})$ are connected (indirectley) to at least two individuals in $IO(\mathcal{E})$ then $\alpha_{j_k,t} \in conv(\alpha_{i_1,s}, \ldots, \alpha_{i_k,s})$ and $\alpha_{i,t} \notin \{-1, 1\}$.
5. $\forall t \geq 0$, $\forall j_k \in \{j_1, j_2, \ldots\} \cup \{i\} \notin IO(\mathcal{E})$ have like-minded neighbors at each period, hence the argument extends to the limit.

■

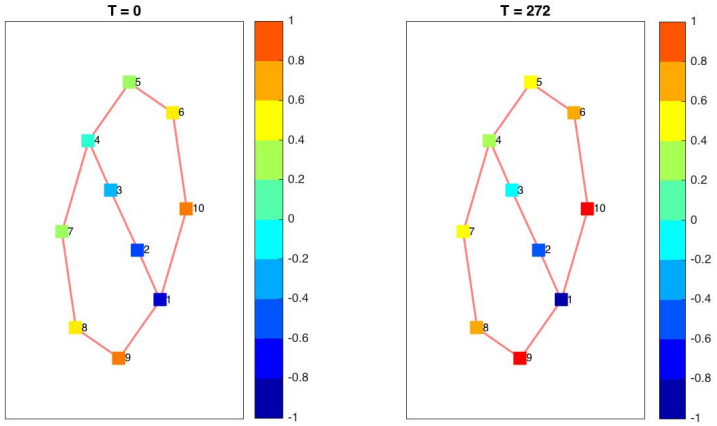


Figure 13

In figure 13, I provide an example that gives an intuition for the previous proof. I set δ^* such that all individuals can express for the sake of the example and expressers are hence represented by

a square. The colors of each node correspond to their opinion at period T indicated by the color map over $[-1, 1]$ on the east side of each figure. The set of connected expressers is $\mathcal{E} = E = N$ and $IO(\mathcal{E}) = \{1, 9, 10\}$. Individuals $\{2, \ldots, 8\}$ all have like-minded neighbors, as indicated by the colors in the left panel of figure 13 and they are all connected to the expressers in $IO(\mathcal{E})$ by three paths of like-minded expressers. The right panel indicates the long-run opinions and we see that individuals $\{2, \ldots, 8\}$ still have like-minded neighbors. In particular, their long-run opinion is a convex combination of the opinions of the individuals in $IO(\mathcal{E}) = \{1, 9, 10\}$. Individual 4 is exactly at distance 3 from each of the three individuals in $IO(\mathcal{E}) = \{1, 9, 10\}$ and their long-run opinion is $\frac{1}{3}\alpha_{1,\infty} + \frac{1}{3}\alpha_{9,\infty} + \frac{1}{3}\alpha_{10,\infty} = \frac{1}{3}(-1) + \frac{1}{3}(1) + \frac{1}{3}(1) = \frac{1}{3}$. Individuals in $IO(\mathcal{E}) = \{1, 9, 10\}$ reach the upper and lower bound of the opinion period just after one period of interaction.

Part (ii). Show that when a given individual i is initially neutral (opinion zero) with no like-minded friends, if she is repulsed by two neighbors who are themselves ideologically-opposed and the negative influence she receives at each period is exactly of the same magnitude, then she remains neutral. The simplest example is a circle with three individuals 1,2,3 where $\alpha_{1,0} = 0$, $\alpha_{2,0} = -0.8$ and $\alpha_{3,0} = 0.8$. Looking at expression (6), the opinions get updated in period $t = 1$ as follows:

$$\begin{aligned}
\alpha_{1,1} &= 0 + \mu(-(-0.8) - (0.8)) = 0 \\
\alpha_{2,1} &= -0.8 + \mu(-0 - (0.8) = -0.8 - \mu 0.8 \\
\alpha_{3,1} &= 0.8 + \mu(-0 - (-0.8) = 0.8 + \mu 0.8
\end{aligned}$$

In subsequent periods, the opinion of individual 1 remains zero because $\mu(-\alpha_{2,t} - \alpha_{3,t}) = 0$ for all $t \geq 0$. However, if $\alpha_{2,0} \neq -\alpha_{3,0}$ then the opinion of individual 1 will be pushed to the upper or lower bound after a certain number of periods depending on which of $|\alpha_{2,0}|$ and $|\alpha_{3,0}|$ is larger.

Claim 4 *Let $i_1, i_2 \ldots i_k \in IO(\mathcal{E})$. If there exists $i \in IO(\mathcal{E})$ such that (i) $\alpha_{i,0} = 0$, (ii) $\forall t \geq 0$ $\underline{N}_{i,t} = \emptyset$, (iii) and $\sum_{j \in \overline{N}_{i,t}} \alpha_{j,t} = 0$ then $\alpha_{i,\infty} = 0$.*

Proof. Let $i \in IO(\mathcal{E})$ such that $\alpha_{i,0} = 0$ and $\forall t \geq 0$, $\underline{N}_{i,t} = \emptyset$. Since individual i has no like-minded neighbors, $a_{i,k}(t) = -\mu$ for all $k \in \overline{N}_{i,0} = N_i$ and $\forall t \geq 0$. Their opinion gets updated as follows:

$$\alpha_{i,1} = 0 - \mu \sum_{j \in \overline{N}_{i,0}} \alpha_{j,0}$$

The opinion of i is zero at period 1 if and only if $\sum_{j \in \overline{N}_{i,0}} \alpha_{j,0} = 0$. Since, all the neighbors are in $\overline{N}_{i,0}$ and $\alpha_{i_0} = 0$ then even for an opinion threshold $\tau > \epsilon$, $\alpha_{j,0} \neq 0$. Hence, $\sum_{j \in \overline{N}_{i,0}} \alpha_{j,0} = 0$ if and only if the opinions $\alpha_{j,0}$ for all $j \in \overline{N}_{i,0}$ cancel out. By induction, the argument holds for all time periods.

Notice that if at any given period t this sum doesn't cancel out then the opinion of i will be different then zero in all subsequent period until it reaches the upper or lower bound.

Part (iii). Otherwise if (i) and (ii) do not hold and $IO(\mathcal{E}) \neq \emptyset$, then $\forall i \in N$, $\alpha_{i,\infty} \in \{-1, 1\}$. There are exactly three remaining cases to consider.

- If (i) does not hold then there are two cases to consider:

  1. There exists exactly 1 path of like-minded neighbors connecting $i \notin IO(\mathcal{E})$ to $i_k \in IO(\mathcal{E})$. In this case, by the same argument used to show part (ii), the long-run opinion of $i \notin IO(\mathcal{E})$ is $\alpha_{i,\infty} = \alpha_{i_k,\infty}$.
  2. There does not exist any path of like-minded neighbors connecting $i \notin IO(\mathcal{E})$ to $i_k \in IO(\mathcal{E})$. This is impossible when $|IO(\mathcal{E})| \geq 1$, because this would mean that i has herself an initially ideologically-opposed neighbor.

- If (i), (ii) do not hold and $|IO(\mathcal{E})| \geq 1$ then we need to consider the case where two individuals are directly linked and belong to $IO(\mathcal{E})$. Let $i \neq j \in IO(\mathcal{E})$ be two individuals such that $g_{ij} = 1$ and $|\alpha_{i,0} - \alpha_{j,0}| \geq \tau$ is the smallest element of the set $\{i \in N : \exists j \in N, |\alpha_{i,0} - \alpha_{j,0}| \geq \tau\}$. Without loss of generality suppose that $\alpha_{i,0} > \alpha_{j,0}$ so that $\alpha_{i,0} - \alpha_{j,0} \geq \tau$. I start by showing that :

  $$P_1 : \ \alpha_{i,1} - \alpha_{j,1} > \alpha_{i,0} - \alpha_{j,0} \geq \tau.$$

  Recall that: $\alpha_{i,1} = -\mu\alpha_{j,0} + \sum_{k \neq j \in N_i} a_{ik}(0)\alpha_{k,0}$ and $\alpha_{j,1} = -\mu\alpha_{i,0} + \sum_{k \neq i \in N_j} a_{jk}(0)\alpha_{k,0}$. It follows that their difference writes:

$$
\begin{aligned}
\alpha_{i,1} - \alpha_{j,1} &= -\mu\alpha_{j,0} + \sum_{k\neq j\in N_i} a_{ik}(0)\alpha_{k,0} + \mu\alpha_{i,0} - \sum_{k\neq i\in N_j} a_{jk}(0)\alpha_{k,0} \\
&= \mu(\alpha_{i,0} - \alpha_{j,0}) + \big(1 + \mu(|\overline{N}_{i,0}| - |\underline{N}_{i,0}|)\big)\alpha_{i,0} - \big(1 + \mu\big(|\overline{N}_{j,0}| - |\underline{N}_{j,0}|\big)\big)\alpha_{j,0} \\
&+ \sum_{\substack{k\neq j\in N_i \\ k\neq i}} a_{ik}(0)\alpha_{k,0} - \sum_{\substack{k\neq i\in N_j \\ k\neq j}} a_{jk}(0)\alpha_{k,0} \\
&= (\alpha_{i,0} - \alpha_{j,0}) + \mu(\alpha_{i,0} - \alpha_{j,0}) + \mu|\overline{N}_{i,0}|\alpha_{i,0} - \mu \sum_{\substack{k\neq j\in \overline{N}_{i,0} \\ k\neq i}} \alpha_{k,0} - \mu|\underline{N}_{i,0}|\alpha_{i,0} + \mu \sum_{\substack{k\neq j\in \underline{N}_{i,0} \\ k\neq i}} \alpha_{k,0} \\
&- \mu|\overline{N}_{j,0}|\alpha_{j,0} + \mu \sum_{\substack{k\neq i\in \overline{N}_{j,0} \\ k\neq j}} \alpha_{k,0} + \mu|\underline{N}_{j,0}|\alpha_{j,0} - \mu \sum_{\substack{k\neq i\in \underline{N}_{j,0} \\ k\neq j}} \alpha_{k,0} \\
&= (\alpha_{i,0} - \alpha_{j,0}) + \mu(\alpha_{i,0} - \alpha_{j,0}) + \mu \sum_{\substack{k\neq j\in \overline{N}_{i,0} \\ k\neq i}} (\alpha_{i,0} - \alpha_{k,0}) \\
&+ \mu \sum_{\substack{k\neq j\in \underline{N}_{i,0} \\ k\neq i}} (\alpha_{k,0} - \alpha_{i,0}) + \mu \sum_{\substack{k\neq i\in \overline{N}_{j,0} \\ k\neq j}} (\alpha_{k,0} - \alpha_{j,0}) + \mu \sum_{\substack{k\neq i\in \underline{N}_{j,0} \\ k\neq j}} (\alpha_{j,0} - \alpha_{k,0}) = (\alpha_{i,0} - \alpha_{j,0}) + R,
\end{aligned}
$$

where

$$
\begin{aligned}
R = \mu(\alpha_{i,0} - \alpha_{j,0}) + \mu \sum_{\substack{k\neq j\in \overline{N}_{i,0} \\ k\neq i}} (\alpha_{i,0} - \alpha_{k,0}) + \mu \sum_{\substack{k\neq j\in \underline{N}_{i,0} \\ k\neq i}} (\alpha_{k,0} - \alpha_{i,0}) + \\
\mu \sum_{\substack{k\neq i\in \overline{N}_{j,0} \\ k\neq j}} (\alpha_{k,0} - \alpha_{j,0}) + \mu \sum_{\substack{k\neq i\in \underline{N}_{j,0} \\ k\neq j}} (\alpha_{j,0} - \alpha_{k,0}).
\end{aligned}
$$

Statement P_1 holds if and only if $R > 0$. Assume by contradiction that $R < 0$. But, R is lower bounded by $\mu\tau + \mu\overline{N}_{i,0}\tau + \mu\epsilon\underline{N}_{i,0} + \mu\overline{N}_{j,0}\tau + \mu\epsilon\underline{N}_{j,0} > 0$.

General argument taking into account all cases:

Without loss of generality, let s be the time period such that, for all $t \geq s$, $\overline{N}_{i,t} = \overline{N}_{i,s}$ and $\underline{N}_{i,t} = \underline{N}_{i,s}$ for all $i \in N$. Now, consider the variation of opinion of individual $i \in N$ at period $t \geq s$:

$$
\begin{aligned}
\alpha_{t,i} &= \alpha_{t-1,i} + \mu \sum_{j\neq i\in \overline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) + \mu \sum_{j\neq i\in \underline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}), \\
\Leftrightarrow \alpha_{t,i} - \alpha_{t-1,i} &= \mu \sum_{j\neq i\in \overline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) + \mu \sum_{j\neq i\in \underline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}).
\end{aligned}
$$

Hence we can write the variation in the opinion of $i \in N$ as:

$$\frac{d\hat{\alpha}_i}{dt} = \mu \sum_{j \neq i \in \overline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) + \mu \sum_{j \neq i \in \underline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) = \nu_i.$$

For all $t \geq s$, $\frac{d\hat{\alpha}_i}{dt} = 0$ if and only if $\nu_i = 0$. Notice that it is impossible with $\alpha_{i,t-1} \neq 0$ that $\mu \sum_{j \neq i \in \overline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) = 0$ because $|\alpha_{i,t-1} - \alpha_{j,t-1}| \geq \tau$ for all $j \neq i \in \overline{N}_{i,t-1}$.

$$\nu_i = 0 \Leftrightarrow \mu \sum_{j \neq i \in \overline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) = -\mu \sum_{j \neq i \in \underline{N}_{i,t-1}} (\alpha_{i,t-1} - \alpha_{j,t-1}) \tag{8}$$

$$\Leftrightarrow g_{ij} = 0, \forall j \in \overline{N}_{i,t-1} \text{ and } \alpha_{i,t-1} = \alpha_{j,t-1}, \forall j \in \underline{N}_{i,t-1}, \text{ (consensus)} \tag{9}$$

$$\text{or } g_{ij} = 0, \forall j \in \overline{N}_{i,t-1} \text{ and } |\underline{N}_{i,t-1}|\alpha_{i,t-1} = \sum_{j \neq i \in \underline{N}_{i,t-1}} \alpha_{j,t-1}, \text{ (condition (i))} \tag{10}$$

$$\text{or } g_{ij} = 0, \forall j \in \underline{N}_{i,t-1} \text{ and } \alpha_{i,t-1} = 0 \text{ and } \sum_{j \neq i \in \overline{N}_{i,t-1}} \alpha_{j,t-1} = 0. \text{ (condition (ii))} \tag{11}$$

$$\text{or } \alpha_{i,t-1} = \frac{1}{|N_i|}\Big(\sum_{j \neq i \in \overline{N}_{i,t-1}} \alpha_{j,t-1} + \sum_{j \neq i \in \underline{N}_{i,t-1}} \alpha_{j,t-1}\Big), \text{ (condition (iii))} \tag{12}$$

Otherwise for all $t \geq s$, $|\frac{d\hat{\alpha}_i}{dt}| > \epsilon$ then the opinion of individual $i \in N$ keeps increasing or decreasing. But since opinions are bounded, her long-run opinion reaches 1 or -1. ■

7.3 Proof of Lemma 1

Case 1: $|\alpha_{i,0} - \alpha_{j,0}| < \tau$. The law of motion 2 rewrites:

$$\begin{cases} \alpha_{i,t} = \alpha_{i,t-1} + \mu(\alpha_{j,t-1} - \alpha_{i,t-1}) \\ \alpha_{j,t} = \alpha_{j,t-1} + \mu(\alpha_{i,t-1} - \alpha_{j,t-1}) \end{cases} \Leftrightarrow \begin{cases} \alpha_{i,t} = (1-\mu)\alpha_{i,t-1} + \mu\alpha_{j,t-1} \\ \alpha_{j,t} = (1-\mu)\alpha_{j,t-1} + \mu\alpha_{i,t-1} \end{cases}$$

We can write the above system in matrix notation:

$$\begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \begin{bmatrix} 1-\mu & \mu \\ \mu & 1-\mu \end{bmatrix} \begin{bmatrix} \alpha_{i,t-1} \\ \alpha_{j,t-1} \end{bmatrix} \Leftrightarrow \begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \overbrace{\begin{bmatrix} 1-\mu & \mu \\ \mu & 1-\mu \end{bmatrix}^t}^{=M^t} \begin{bmatrix} \alpha_{i,0} \\ \alpha_{j,0} \end{bmatrix} \text{ (by induction).}$$

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

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

For $\mu \in (0, 1/2)$, $\lim_{t\to\infty}(1-2\mu)^t = 0$. 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.[12] It follows that when the opinions of i and j are close enough then they converge exactly to their average:

$$\alpha_\infty^a = \lim_{t\to\infty} \begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} \alpha_{i,0} \\ \alpha_{j,0} \end{bmatrix} = \begin{bmatrix} \frac{\alpha_{i,0}+\alpha_{j,0}}{2} \\ \frac{\alpha_{i,0}+\alpha_{j,0}}{2} \end{bmatrix}.$$

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

Case 2: $|\alpha_{i,0} - \alpha_{j,0}| \geq \tau$. The law of motion (2) rewrites:

$$\begin{cases} \alpha_{i,t} = \alpha_{i,t-1} + \mu(\alpha_{i,t-1} - \alpha_{j,t-1}) \\ \alpha_{j,t} = \alpha_{j,t-1} + \mu(\alpha_{j,t-1} - \alpha_{i,t-1}) \end{cases} \Leftrightarrow \begin{cases} \alpha_{i,t} = (1+\mu)\alpha_{i,t-1} - \mu\alpha_{j,t-1} \\ \alpha_{j,t} = (1+\mu)\alpha_{j,t-1} - \mu\alpha_{i,t-1}. \end{cases}$$

We can write the above system in matrix notation:

$$\begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \begin{bmatrix} 1+\mu & -\mu \\ -\mu & 1+\mu \end{bmatrix} \begin{bmatrix} \alpha_{i,t-1} \\ \alpha_{j,t-1} \end{bmatrix} \Leftrightarrow \begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \overbrace{\begin{bmatrix} 1+\mu & -\mu \\ -\mu & 1+\mu \end{bmatrix}^t}^{=M^t} \begin{bmatrix} \alpha_{i,0} \\ \alpha_{j,0} \end{bmatrix} \text{ (by induction).}$$

Moreover, we can diagonalize the matrix M^t :

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

The limit opinions of i and j are:

$$\alpha_\infty^r \lim_{t\to\infty} \begin{bmatrix} \alpha_{i,t} \\ \alpha_{j,t} \end{bmatrix} = \lim_{t\to\infty} \frac{1}{2} \begin{bmatrix} 1+(1+2\mu)^t & 1-(1+2\mu)^t \\ 1-(1+2\mu)^t & 1+(1+2\mu)^t \end{bmatrix} \begin{bmatrix} \alpha_{i,0} \\ \alpha_{j,0} \end{bmatrix} = \lim_{t\to\infty} \frac{1}{2} \begin{bmatrix} \alpha_{i,0}+\alpha_{j,0}+(\alpha_{i,0}-\alpha_{j,0})(1+2\mu)^t \\ \alpha_{i,0}+\alpha_{j,0}+(\alpha_{j,0}-\alpha_{i,0})(1+2\mu)^t \end{bmatrix}.$$

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

[12] For more details on this topic in linear algebra See Silva, Silva and Fernandes (2016) [27].

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:

$$\begin{cases} \frac{1}{2}\big(\alpha_{i,0} + \alpha_{j,0} + (\alpha_{i,0} - \alpha_{j,0})(1+2\mu)^t\big) = 1 & \text{if } \alpha_{i,0} > \alpha_{j,0} \\ \frac{1}{2}\big(\alpha_{i,0} + \alpha_{j,0} + (\alpha_{i,0} - \alpha_{j,0})(1+2\mu)^t\big) = -1 & \text{if } \alpha_{i,0} < \alpha_{j,0}. \end{cases}$$

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

$$t_r = \begin{cases} \frac{\log\left(\frac{2-\alpha_{i,0}-\alpha_{j,0}}{\alpha_{i,0}-\alpha_{j,0}}\right)}{\log(1+2\mu)} & \text{if } 1 \geq \alpha_{i,0} > \alpha_{j,0} \geq -1 \\ \frac{\log\left(\frac{-2-\alpha_{i,0}-\alpha_{j,0}}{\alpha_{i,0}-\alpha_{j,0}}\right)}{\log(1+2\mu)} & \text{if } -1 \leq \alpha_{i,0} < \alpha_{j,0} \leq 1. \end{cases}$$

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 an -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 $\alpha_{i,0} > \alpha_{j,0}$ (similarly for the other case) and t_a at its lower bound:

$$\frac{\log(\epsilon)}{\log(1-2\mu)} > \frac{\log\left(\frac{2-\alpha_{i,0}-\alpha_{j,0}}{\alpha_{i,0}-\alpha_{j,0}}\right)}{\log(1+2\mu)} \Leftrightarrow \epsilon < \exp\left(\frac{\log\left(\frac{2-\alpha_{i,0}-\alpha_{j,0}}{\alpha_{i,0}-\alpha_{j,0}}\right)\log(1-2\mu)}{\log(1+2\mu)}\right).$$

7.4 Proof of Theorem 2

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) [23]). For $\tilde{G} \in \mathbb{R}^{n\times n}$, $\lim_{t\to\infty} \tilde{G}^t$ 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) [23]). Moreover, for every stochastic matrix, the spectral radius is 1 and it is semi-simple (p.696, Chapter 8 in Meyer (2000) [23] or see Corollary 2, page 2214, in Ding and Rhee (2011) [16]). 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 $\lim_{t\to\infty} \tilde{G}^t$ exists, it is equal to the spectral projector associated with eigenvalue 1 (again see p.630, chapter 7, in Meyer (2000) [23]).

Reminder from p.629 Meyer (2000) [23]. Recall that a row stochastix 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, $\lim_{t\to\infty} \tilde{G}_\theta^t = P_1Q_1$ 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) [26]: *we say that i leads to j and write $i \to j$ 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 $i \to j$ and $j \to i$ and write in this case $i \leftrightarrow j$. The index i is called essential when : $i \to j$ implies $i \leftrightarrow j$ and there is at least one j such that $i \to j$. 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 expressers 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. Form 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 $(nx1)$ vector, is a stationary distribution of the*

chain.

7.5 Proof of Proposition 3

Case (i)**:** $|E| = 1$**.** Suppose that $E = \{i\}$ and that the initial opinion of this individual is $\alpha_{0,i}$. Then using proposition 1, this expresser remains stubborn forever because she does not have neighbors who also express. Hence they never updates their opinion and $\alpha_{\infty,i} = \alpha_{0,i}$. Moreover from theorem 2 the long-run opinion of consensual individuals is a convex combination of opinions of expressers. Since there is only one expresser then all consensual individuals have long-run opinion of $\alpha_{0,i}$ and consensus prevails.

Case (ii)**:** $|E| > 1$**.** Long-run opinions form consensus when $\forall i \neq j \in N$, $|\alpha_{\infty,i} - \alpha_{\infty,j}| < \tau$.

1. Let $E = \bigcup_{k=1}^{\kappa} \mathcal{E}$ be the set of expressers such that $\kappa \geq 1$.

2. For $\kappa = 1$ there is a unique set of connected expressers $E = \mathcal{E}_1$ and long-run opinions form a consensus if and only if $\forall i \neq j \in \mathcal{E}_1$, such that $g_{ij} = 1$, $|\alpha_{0,i} - \alpha_{0,j}| < \tau$. Since each member within $\mathcal{E}_1$ has like-minded neighbors, the members of $\mathcal{E}_1$ converge to the average of their initial opinions as shown in proposition 1. Moreover, using theorem 2, the opinions of consensual individuals are convex combinations of the opinions of expressers. Since here there is only one set of connected expressers, the opinion of each consensual individual is exactly the average of opinions of the members of the set of connected expressers $\mathcal{E}_1$.

3. For $\kappa > 1$, without loss of generality suppose that the union of the two connected set of expressers $\mathcal{E}_1$ and $\mathcal{E}_2$ is equal to E. Long-run opinions form consensus if and only if (i) $\forall i \neq j \in \mathcal{E}_k$ for $k \in \{1, 2\}$, such that $g_{ij} = 1$, $|\alpha_{0,i} - \alpha_{0,j}| < \tau$ (ii) $|\overline{\alpha}_{0,\mathcal{E}_1} - \overline{\alpha}_{0,\mathcal{E}_2}| < \tau$ where $\overline{\alpha}_{0,\mathcal{E}_k}$ is the average of initial opinions within the set $\mathcal{E}_k$ for $k \in \{1, 2\}$. Since within each of both sets all members have like-minded neighbors within $\mathcal{E}_1$ and $\mathcal{E}_2$ opinions of expressers converge respectively to $\overline{\alpha}_{0,\mathcal{E}_1}$ and $\overline{\alpha}_{0,\mathcal{E}_1}$. Moreover, consensus can prevail if and only if $|\overline{\alpha}_{0,\mathcal{E}_1} - \overline{\alpha}_{0,\mathcal{E}_2}| < \tau$ because then any long-run opinion of a consensual individual $i \in N \setminus E$ is at a distance of at most τ from any other long-run opinion of other individuals in the network.

Formally,

$$\begin{aligned}\alpha_{\infty_i} = \sum_{j\in\mathcal{E}_1\cup\mathcal{E}_2} \mathcal{G}_{ij}\alpha_{j,\infty} &= \sum_{j_1\in\mathcal{E}_1} \mathcal{G}_{ij_1}\overline{\alpha}_{0,\mathcal{E}_1} + \sum_{j_2\in\mathcal{E}_2} \mathcal{G}_{ij_2}\overline{\alpha}_{0,\mathcal{E}_2} \\ &= \overline{\alpha}_{0,\mathcal{E}_1} \sum_{j_1\in\mathcal{E}_1} \mathcal{G}_{ij_1} + \overline{\alpha}_{0,\mathcal{E}_2} \sum_{j_2\in\mathcal{E}_2} \mathcal{G}_{ij_2} \\ &= \overline{\alpha}_{0,\mathcal{E}_1} \sum_{j_1\in\mathcal{E}_1} \mathcal{G}_{ij_1} + \overline{\alpha}_{0,\mathcal{E}_2}(1 - \sum_{j_1\in\mathcal{E}_1} \mathcal{G}_{ij_1}) \\ &= \overline{\alpha}_{0,\mathcal{E}_2} + (\overline{\alpha}_{0,\mathcal{E}_1} - \overline{\alpha}_{0,\mathcal{E}_2}) \sum_{j_1\in\mathcal{E}_1} \mathcal{G}_{ij_1}\end{aligned}$$

Hence for any expresser in $\mathcal{E}_2$ the opinion difference with a given consensual individual $i \in N \setminus E$ is at most τ (similarly for any expresser in $\mathcal{E}_1$) :

$$|\alpha_{\infty_i} - \overline{\alpha}_{0,\mathcal{E}_2}| = |(\overline{\alpha}_{0,\mathcal{E}_1} - \overline{\alpha}_{0,\mathcal{E}_2}) \sum_{j_1\in\mathcal{E}_1} \mathcal{G}_{ij_1}| \leq |(\overline{\alpha}_{0,\mathcal{E}_1} - \overline{\alpha}_{0,\mathcal{E}_2})| < \tau$$

Furthermore, for two consensual individuals $i \neq j \in N \setminus E$ their long-run is at most τ because:

$$\begin{aligned}|\alpha_{\infty_i} - \alpha_{\infty_j}| = |(\overline{\alpha}_{0,\mathcal{E}_1} - \overline{\alpha}_{0,\mathcal{E}_2})(\sum_{j_1\in\mathcal{E}_1} \mathcal{G}_{ij_1} - \sum_{j_1\in\mathcal{E}_1} \mathcal{G}_{jj_1})| &\leq |(\overline{\alpha}_{0,\mathcal{E}_1} - \overline{\alpha}_{0,\mathcal{E}_2})(1 - \sum_{j_1\in\mathcal{E}_1} \mathcal{G}_{jj_1})| \\ &\leq |\overline{\alpha}_{0,\mathcal{E}_1} - \overline{\alpha}_{0,\mathcal{E}_2}| \\ &< \tau\end{aligned}$$

The arguments easily extend for more than 2 sets of connected expressers. To see this, think of three sets of connected expressers $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}_3$ where the long-run opinions within each set are respectively $\overline{\alpha}_{0,\mathcal{E}_1}$, $\overline{\alpha}_{0,\mathcal{E}_2}$ and $\overline{\alpha}_{0,\mathcal{E}_3}$. For all $i \in N \setminus E$, $\alpha_{i,\infty} \in conv(\overline{\alpha}_{0,\mathcal{E}_1}, \overline{\alpha}_{0,\mathcal{E}_2}, \overline{\alpha}_{0,\mathcal{E}_3})$

and $|\overline{\alpha}_{0,\mathcal{E}_2} - \overline{\alpha}_{0,\mathcal{E}_1}| < \tau$, $|\overline{\alpha}_{0,\mathcal{E}_3} - \overline{\alpha}_{0,\mathcal{E}_1}| < \tau$, $|\overline{\alpha}_{0,\mathcal{E}_2} - \overline{\alpha}_{0,\mathcal{E}_3}| < \tau$, it follows that :

$$\begin{aligned}
|\alpha_{i,\infty} - \overline{\alpha}_{0,\mathcal{E}_1}| &= |\sum_{j_1\in\mathcal{E}_1} \mathcal{G}_{ij_1}\overline{\alpha}_{0,\mathcal{E}_1} + \sum_{j_2\in\mathcal{E}_2} \mathcal{G}_{ij_2}\overline{\alpha}_{0,\mathcal{E}_2} + \sum_{j_3\in\mathcal{E}_3} \mathcal{G}_{ij_3}\overline{\alpha}_{0,\mathcal{E}_3} \\
&\quad - (\overbrace{\sum_{j_1\in\mathcal{E}_1} \mathcal{G}_{ij_1} + \sum_{j_2\in\mathcal{E}_2} \mathcal{G}_{ij_2} + \sum_{j_3\in\mathcal{E}_3} \mathcal{G}_{ij_3}}^{=1})\overline{\alpha}_{0,\mathcal{E}_1}| \\
&= |\sum_{j_2\in\mathcal{E}_2} \mathcal{G}_{ij_2}\overline{\alpha}_{0,\mathcal{E}_2} + \sum_{j_3\in\mathcal{E}_3} \mathcal{G}_{ij_3}\overline{\alpha}_{0,\mathcal{E}_3} - \sum_{j_2\in\mathcal{E}_2} \mathcal{G}_{ij_2}\overline{\alpha}_{0,\mathcal{E}_1} - \sum_{j_3\in\mathcal{E}_3} \mathcal{G}_{ij_3}\overline{\alpha}_{0,\mathcal{E}_1}| \\
&= |\sum_{j_2\in\mathcal{E}_2} \mathcal{G}_{ij_2}(\overline{\alpha}_{0,\mathcal{E}_2} - \overline{\alpha}_{0,\mathcal{E}_1}) + \sum_{j_3\in\mathcal{E}_3} \mathcal{G}_{ij_3}(\overline{\alpha}_{0,\mathcal{E}_3} - \overline{\alpha}_{0,\mathcal{E}_1})| \\
&< \sum_{j_2\in\mathcal{E}_2} \mathcal{G}_{ij_2}\tau + \sum_{j_3\in\mathcal{E}_3} \mathcal{G}_{ij_3}\tau \\
&\leq \tau
\end{aligned}$$

7.6 Proof of Lemma 2

Let $k \in N$ be a consensual individual. It follows from theorem 2 that:

$$\alpha_{k,\infty} = \sum_{i\in E} \mathcal{G}_{k,i}\alpha_{i,\infty} = \sum_{i\in E^+} \mathcal{G}_{k,i} - \sum_{i\in E^-} \mathcal{G}_{k,i}$$

Moreover, recall that for all $k \in N$, $\sum_{i\in E} \mathcal{G}_{k,i} = \sum_{i\in E^+} \mathcal{G}_{k,i} + \sum_{i\in E^-} \mathcal{G}_{k,i} = 1$. Individual k doesn't hold the extreme opinion of the members of E^+, if and only if, for $i \in E^+$:

$$\begin{aligned}
|\alpha_{k,\infty} - \alpha_{i,\infty}| \geq \tau &\Leftrightarrow |(2\sum_{i\in E^+} \mathcal{G}_{ki} - 1) - 1| \geq \tau \\
&\Leftrightarrow 1 - \sum_{i\in E^+} \mathcal{G}_{ki} \geq \frac{\tau}{2} \\
&\Leftrightarrow 1 - \frac{\tau}{2} \geq \sum_{i\in E^+} \mathcal{G}_{ki}
\end{aligned}$$

Similarly, individual k doesn't hold the extreme opinion of the members of E^-, if and only if, for $i \in E^-$:

$$\begin{aligned}
|\alpha_{k,\infty} - \alpha_{i,\infty}| \geq \tau &\Leftrightarrow |(2\sum_{i\in E^+} \mathcal{G}_{ki} - 1) - (-1)| \geq \tau \\
&\Leftrightarrow \sum_{i\in E^+} \mathcal{G}_{ki} \geq \frac{\tau}{2}
\end{aligned}$$

Hence, if if long-run opinions are bi-polarized then necessarily all individuals belong to either of

both extreme opinion groups and there does not exist an individual $k \in N$ such that:

$$1 - \tau/2 \leq \sum_{i \in E^+} \mathcal{G}_{ki} \leq \tau/2$$

Now recall that from theorem 2 that the limit of $\tilde{G}^t$ when t is large, exists and is given by $\mathcal{G}$. Since $\tilde{G}$ is row stochastic, we can show by induction that $\tilde{G}^t$ is also row stochastic. The row stochasticity of $\tilde{G}^t$ is a linear condition, hence continuous so it is preserved by the limits. It follows that :

$$\sum_{j \in N} \sum_{i \in N} \mathcal{G}_{ij} = |N| \tag{13}$$

Let $\mathbf{a}_\infty$ be a bi-polarized long-run opinion vector (see definition 5) and suppose by contradiction that

$$|\sum_{i \in E^+} (\mathcal{G}^{'}\mathbf{1})_i - \sum_{i \in E^-} (\mathcal{G}^{'}\mathbf{1})_i| > \epsilon \tag{14}$$

Equation (14) can be rewritten as:

$$\sum_{i \in E^+} (\mathcal{G}^{'}\mathbf{1})_i - \sum_{i \in E^-} (\mathcal{G}^{'}\mathbf{1})_i = \nu \text{ for } |\nu| > 1$$

Moreover, $\sum_{i \in E^+} (\mathcal{G}^{'}\mathbf{1})_i = \sum_{j \in N} \sum_{i \in E^+} \mathcal{G}_{ji}$ and for all $j \in N$, $\sum_{i \in E^+} \mathcal{G}_{ji} + \sum_{i \in E^-} \mathcal{G}_{ji} = 1$. Hence:

$$\begin{aligned}
\sum_{i \in E^+} (\mathcal{G}^{'}\mathbf{1})_i - \sum_{i \in E^-} (\mathcal{G}^{'}\mathbf{1})_i = \nu &\Leftrightarrow \sum_{j \in N} \sum_{i \in E^+} \mathcal{G}_{ji} - \sum_{j \in N} \sum_{i \in E^-} \mathcal{G}_{ji} = \nu \\
&\Leftrightarrow \sum_{j \in N} (\sum_{i \in E^+} \mathcal{G}_{ji} - \sum_{i \in E^-} \mathcal{G}_{ji}) = \nu \\
&\Leftrightarrow \sum_{j \in N} (1 - 2 \sum_{i \in E^-} \mathcal{G}_{ji}) = \nu \\
&\Leftrightarrow \frac{|N| - \nu}{2} = \sum_{j \in N} \sum_{i \in E^-} \mathcal{G}_{ji} = \sum_{j \in N} \sum_{i \in N^-} \mathcal{G}_{ji}
\end{aligned}$$

Since $|\nu| > 1$ and there does not exist moderate individuals, the size of the group N^- is strictly smaller than $|N|/2$ which contradicts the definition of a bi-polarized long-run opinion vector.

7.7 Network Statistics

Consider a graph G, with N vertices and M edges.

- Assortativity or assortative mixing in Newman (2002) [24], assortativity of an observed net-

work is given by :

$$r = \frac{M^{-1}\sum_i j_i k_i - [M^{-1}\sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1}\sum_i \frac{1}{2}(j_i + k_i)]^2},$$

where j_i and k_i are the degrees of the vertices at the ends of the ith edge with $i = 1, \ldots, M$.

- Neighborhood connectivity: average degree in the neighborhood of a given node $i \in N$

$$\frac{1}{d_i}\sum_{j \in N_i} d_j,$$

where d_i is the degree of node i and N_i is the neighborhood.

- Per capita average degree:

$$\frac{\sum_{i \in N} d_i}{N},$$

where d_i is the degree of node i.

- Average path length :

$$l = \frac{1}{N(N-1)}\sum_{i \neq j} d(v_i, v_j),$$

where $d(v_i, v_j)$ is the shortest path between nodes v_1 and v_2 (computed here with Dijkstra).

7.8 Eigenvector centrality

For the simulation I have used the local popularity measure defined as the number of friends over the average number of friends of friends. I have argued in the introduction that a local measure is suitable for an opinion formation model, as an individual can have locally high impact over her own friends, even though she may not be globally influential in the network. Nevertheless, the convergence results go through for any other centrality measure. Figures 14 and 15 show the final opinions for the same initial opinion and network structure but two different centrality measures, respectively eigenvector centrality and local popularity. We have used as an expression threshold $\delta^* = 1/n$ when taking the eigenvector centrality measure so that in any regular network, individuals with no different levels of expertise express. We see that when using the eigenvector centrality measure, locally popular individuals do not get to express. Furthermore there is a lot of interaction and disagreement within the hub containing the agent with the highest eigenvector centrality and her friends.

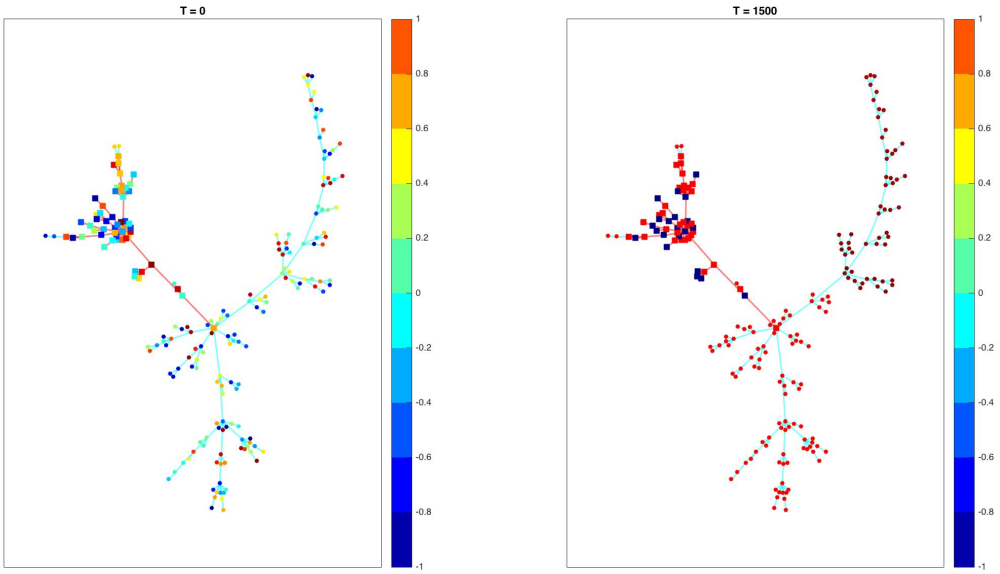


Figure 14: Eigenvector centrality. Squares are expressers, mean polarization 0.47

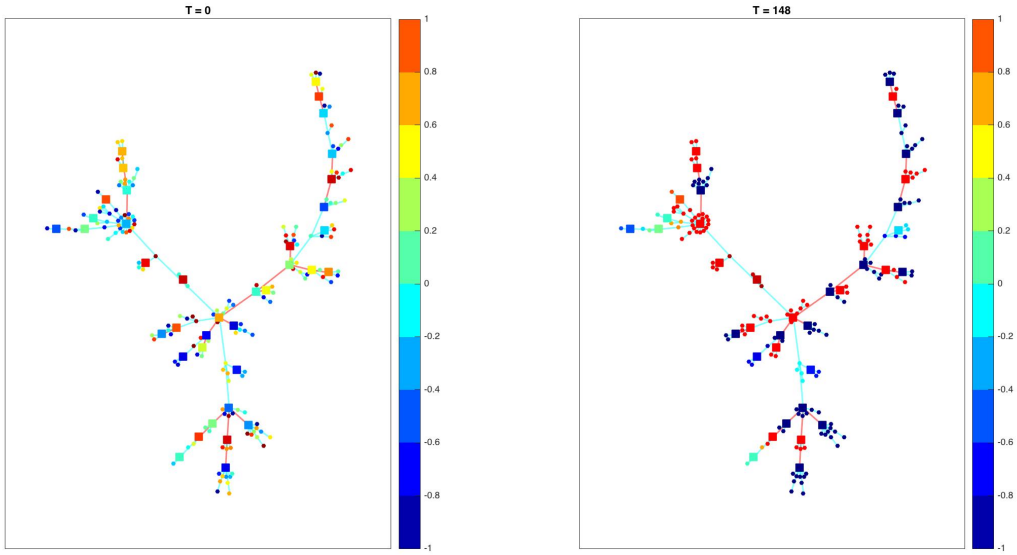


Figure 15: Local popularity. Squares are expressers, mean polarization 0.78

7.9 Micro-foundation

The *expression heuristic* can be micro-founded by a simple game where individuals face a social cost of expression. Intuitively this cost is a psychological cost of disagreement with peers or, the cost of a long lasting debate because of strong opposing views. At each time period, individuals observe the actions (opinions) of neighbors at the previous period and myopically best-respond at the current period. The payoff of individual $i \in N$ given action $a_{i,t} \in \{\text{express}, \text{hide}\} = \{\alpha_{i,t}, \overline{\alpha}_{i,t-1}\}$ is the following:

$$\pi_i(a_{i,t}, a_{-i,t-1}) = -f(\delta_i)(\alpha_{i,t} - a_{i,t})^2 - (1 - f(\delta_i))(\overline{\alpha}_{i,t-1} - a_{i,t})^2. \tag{15}$$

The function f has support $(0, 1)$ and $f' > 0$. The first term of the payoff function (15) is the cost borne by each player when they does not express their opinion, in other words it is the cost of being consensual. The second term is the cost of expression, which is a cost borne when a player expresses their own opinion rather than their neighborhood average opinion. Both terms are weighted by an increasing function of the local popularity parameter δ_i. The higher δ_i is, the higher the influence of player i within their own neighborhood and the lower the cost of expression or social disagreement. Given the payoff function (15), any $i \in N$ such that $\delta_i \geq \delta^* = f^{-1}(\frac{1}{2})$ best-responds with *express*, while any $i \in N$ such that $\delta_i < \delta^* = f^{-1}(\frac{1}{2})$ best-responds with *hide*. Hence the local centrality parameter of each individual is a sufficient statistic for the choice of the optimal binary action at each stage-game and individuals can be classified into two groups: $E = \{i \in N, \delta_i \geq \delta^*\}$ and $C = \{i \in N, \delta_i < \delta^*\}$.

Assumption 2 *For the remainder of the paper, the expression threshold is normalized to* $\delta^* = 1$.

Given the local popularity measure δ_i of individual i, the threshold $\delta^* = 1$ allows all individuals to play *express* when they are all equally popular. Namely, in any d-regular network where each individual has degree d, any player $i \in N$ expresses because $\delta_i = \frac{d}{\frac{1}{d} d \times d} = \delta^*$. Intuitively, d-regular networks have perfect assortativity meaning that there is no difference in the level of popularity, expertise or leadership.

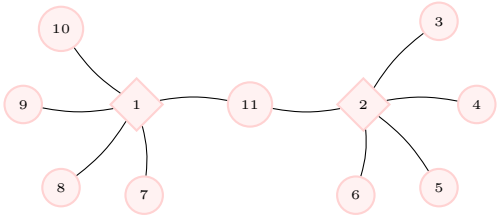


Figure 16: Network G_1, diamond shaped nodes correspond to individuals who choose to *express*.

Example 4 *Consider the network G_1 in Figure 16. Individuals* 1 *and* 2 *both have local popularity* $\delta_1 = \delta_2 = 5/\big((1/5)(1+1+1+1+2)\big) = 25/6 > \delta^*$*, therefore they play express and* $E = \{1, 2\}$*. While individuals* $j \in \{3, \ldots, 9, 10\}$ *have a local popularity* $\delta_j = 1/5 < \delta^*$ *and individual* 11 *has* $\delta_{11} = 4/10 < \delta^*$ *, hence they play hide and* $C = \{3, \ldots, 10, 11\}$*.*

References

[1] Daron Acemoglu, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar. Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research*, 38(1), 2013.

[2] Daron Acemoglu and Asuman Ozdaglar. Opinion dynamics and learning in social networks. *Dyn Games Appl*, 1:3–49, 2011.

[3] S. Banisch and E. Olbrich. Opinion polarization by learning from social feedback. *The Journal of Mathematical Sociology*, 43(2):76–103, 2019.

[4] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439), 1999.

[5] Alberto Bisin, Eleonora Patacchini, Thierry Verdier, and Yves Zenou. Bend it like beckham: Ethnic identity and integration. *European Economic Review*, 90:146–164, 2016.

[6] Ugo Bolletta and Paolo Pin. Polarization when people choose their peers. *mimeo*, 2019.

[7] M.D. Conover, J. Ratkiewicz, M. Francisco, B.Gonçalves, A. Flammini, and F. Menczer. Political polarization on twitter. *Proceedings of the Fifth International AAAI conference on weblogs and social media*, 2011.

[8] Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[9] 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.

[10] Noah Friedkin and Eugene Johnsen. Social influence and opinions. *Journal of mathematical sociology*, 15(3-4):193–205, 1990.

[11] 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.

[12] Benjamin Golub and Evan Sadler. *Learning in Social Networks*, chapter 19. The Oxford Handbook of the Economics of Networks, 2016.

[13] F. Harary. A criterion for unanimity in french's theory of social power. *Studies in Social Power*, 1959.

[14] Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence models, analysis, and simulationb. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.

[15] 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.

[16] J.Ding and Noah H. Rhee. On the equality of algebraic and geometric multiplicities of matrix eigenvalues. *Applied Mathematics letters*, 24, 2011.

[17] 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.

[18] J.R. P. French Jr. A formal theory of social power. *Psychol. Rev*, 63, 1956.

[19] Bibb Latané. The psychology of social impact. *American Psychologist*, 36(4):343–356, 1981.

[20] Bibb Latané, Andrej Nowak, and James H. Liu. Measuring emergent social phenomena: dynamism, polarization, and clustering as order parameters of social systems. *Behavioral Science*, 39, 1994.

[21] Fragkiskis Malliaros and Michalis Vazirgiannis. Clustering and community detection in directed networks: A survey. *Physics Reports*, 2013.

[22] Isabel Melguizo. Persistence of disagreement. *miemo*, 2018.

[23] Carl D. Meyer. *Matrix Analysis and Applied linear algebra*. SIAM, 2000.

[24] M.E.J. Newman. Assortative mixing in networks. *Phys. Rev. Lett*, 89, 2002.

[25] Evan Sadler. Influence campains. *mimeo*, 2019.

[26] E. Seneta. *Non-negative matrices and markov chains*. Springer Science and Business Media, 2006 edition, 1981.

[27] 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.

[28] 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.

[29] 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.

[30] Garold Stasser and William Titus. Hidden profiles: a brief history. *Psychological Inquiry*, 14(3):304–313, 2003.

[31] Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems and Control Letters*, 53, 2004.

[32] E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione. Binary opinion dynamics with stubborn agents. *ACM Trans. Econ. Comp.*, 2013.