Social Media Networks, Fake News, and Polarization

Marina Azzimonti
Department of Economics
Stony Brook University
marina.azzimonti@gmail.com

Marcos Fernandes
Department of Economics
Stony Brook University
marcos.fernandes@stonybrook.edu

Abstract

We study how the structure of social media networks affects the degree of political polarization in society. We analyze a dynamic model of opinion formation in which individuals have imperfect information about the true state of the world and suffer from bounded rationality. Key to the analysis is the presence of partisan agents that communicate extremely biased opinions, interpreted as *fake news*. We characterize how the flow opinions evolves over time and evaluate the determinants of long-run disagreement among individuals in the network. To that end, we simulate a large set of random networks with different characteristics and quantify how the degrees of centrality, connectedness, and influence from partisan agents affect polarization in the long-run.

Keywords: Learning, Polarization, Social Networks, Social Media, Fake News

JEL Classification: C63, D83, D85

1 Introduction

The United States has experienced an unprecedented surge in political polarization over the last two decades. A recent survey conducted by The Pew Research Center indicates that Republicans and Democrats are further apart ideologically than at any point since 1994 (see Figure 1).

What could be causing this increase in polarization? Traditional theories in economics and political science point to the recent rise in income inequality, the influence of PACs through campaign financing, party sorting among voters, re-districting (Gerrymanding), and changes in

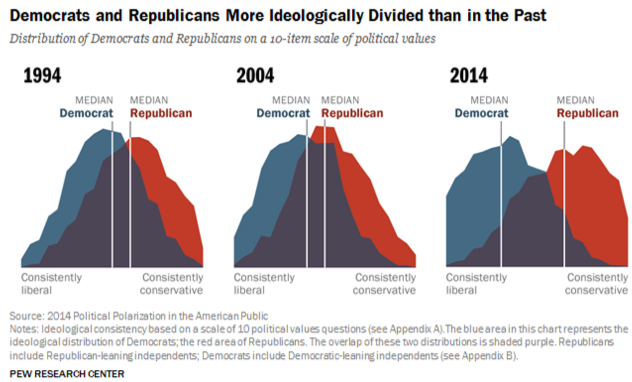


Figure 1: Political Polarization in the American Public (2014, Pew Research Center)

the media environment as potential determinants (see Barber and McCarthy, 2015 for an excellent discussion). More recently, attention has focused to the internet as an alternative candidate explanation. Cass Sunstein (2002) argues that the internet creates 'echo chambers' where individuals find their own biases and opinions endlessly reinforced, and writes that 'people restrict themselves to their own points of view—liberals watching and reading mostly or only liberals; moderates, moderates; conservatives, conservatives; Neo-Nazis, Neo-Nazis' (p. 5-6). This reduces the 'unplanned, unanticipated encounters central to democracy itself' and significantly increases polarization (p. 9). With the dispersion of news through social media, and more generally the internet, and given that a growing proportion of individuals, politicians, and media outlets are relying more intensively on this networked environment to get information and to spread their world-views, it is natural to ask whether and to what extent political polarization might be exacerbated by social media communication.

In this paper, we analyze how the structure of social media communication affects the degree of polarization in society. To that end, we study a dynamic model of opinion formation in the spirit of Jadbabaie, Molavi, Sandroni, and Tahbaz-Salehi (2012, JMST henceforth) in which individuals who are connected in a social network have imperfect information about the true state of the world. The true state of the world can be interpreted as the relative quality of two candidates competing for office, the optimal size of the government, the degree of government intervention (through the provision of public goods such as healthcare), etc. Individuals can obtain information (e.g. signals) about the true state of the world from unbiased sources (reports from non-partisan research institutions, mainstream media, etc.), but are unable process all the available information. They can also obtain information from their social neighbours (e.g. individuals connected to them through the network) who are potentially exposed to other sources.

Due to limited observability about the structure of the network and the probability distribution of signals observed by others, individuals would need to update opinions on the state of the world as well as on the topology of the network. This makes Bayesian updating complex and impractical. We assume instead that individuals suffer from bounded rationality, and update their opinions partly based on information obtained from their social network in an inhomogeneous stochastic gossip model of communication based on JMST(2012) and Acemoglu, Como, Fagnati, and Ozdaglar (2013, ACFO henceforth).

There are two types of individuals in this economy: *regular* agents and *partisan* agents. Regular agents receive signals from unbiased sources and are also influenced by the opinion of their social neighbours. Partisan agents, on the other hand, restrict attention to biased media (e.g. news sources supporting their preferred policy or candidate) and ignore the opinion of others.[1] Because their world-views are necessarily biased, they introduce mis-information into the network in the form of *fake news*. The opinions generated from the exchange of information forms a Markov process which never leads to consensus among regular agents. In such environment, it can be shown that society's beliefs fail to converge almost surely. Moreover, under some conditions, the belief profile fluctuates in an ergodic fashion leading to polarization cycles.

The structure of the graph representing the social media network and the degree of influence of partisan agents in it shape the dynamics of opinion and the degree of polarization in the long-run. More specifically, long-run polarization depends on three factors: behavioral assumptions (e.g. the updating rule), communication technology (e.g. the speed at which information flows), and the network topology (e.g. existence of echo chambers, the centrality of partisan agents, etc.). Because a theoretical characterization of the relationship between the topology of the network and the degree of polarization is unfeasible, we simulate a large set of random networks with different characteristics. We then quantify how the degrees of centrality, connectedness, and influence affect long-run polarization, defined as in Esteban and Ray (1994) and Esteban (2007).

The arrival of the internet allowed individuals to access an extremely rich set of sources of information. On the flip side, the abundance of signals made it more difficult for individuals to process all the available information. We find that to the extent that agents rely more heavily on the opinion of others (and less on incorporating signals from unbiased sources), polarization rises. This happens because they are more likely to be influenced by fake news. The speed of communication, on the other hand, *reduces* polarization. We interpret the rapid growth in social media outlets as a technological change that allowed agents to share information at a faster rate.

[1] In that sense, they have similar characteristics to the 'stubborn' agents in ACFO (2013).

On the one hand, this could exacerbate polarization, as fake news spread faster. On the other hand, it is possible to aggregate information coming from unbiased signals more efficiently. To the extent that the partisan agents constitute a small share of the population, the second force dominates. This is consistent with findings by Barbera (2015) who documents that the expansion of Twitter resulted in lower political polarization. In terms of the effects of network structure, we find that a higher degree of centrality of partisan agents exacerbates polarization. This effect is mitigated by the degree of influence of their audience. The reason being that if a partisan agent manages to obtain the exclusive attention of very influential nodes, it can influence the beliefs of a large set of agents in the network. This reduces the influence of partisan agents at the other end of the spectrum, hence reducing polarization. Even though long-run disagreement goes down, the degree of misinformation (e.g. the distance between the society beliefs and the true state of the world) can be significant. Finally, we find that networks with a high degree of clustering tend to exhibit larger polarization in the long run, as echo chambers are more likely observed.

Related Literature Our paper is related to a growing number of articles studying information transmission in networks under both, bounded and fully rational agents.

The strand of the literature assuming that agents are fully rational typically considers a dynamic game where individuals interact sequentially and exchange opinions only once. Examples are Banerjee (1992), Smith and Sorensen (2000), Banerjee and Fundenberg (2004), and Acemoglu et. al (2011). Because the theoretical characterization of equilibria is complex, these papers restrict attention to very stylized networks. Moreover, they typically study environments in which society eventually reaches consensus, implying that polarization arises only in the short-run.

The strand of literature focusing on bounded rational agents (also referred to as 'De-Grootian') assumes that individuals follow simple heuristic rules to update beliefs. Examples are Ellison and Fundenberg (1993, 1995), Bala and Goyal (1998,2001), De Marzo et al (2003), Golub and Jackson (2010), and ACFO (2013). In these environments, long-run polarization arises in equilibrium because individuals receive information only once—at the outset of the initial period. There is no sense in which new information (such as news) may arrive and modify regular agents' opinions. JMST (2012) show that when this assumption is relaxed, that is, when individuals receive a constant flow of information, polarization eventually disappears. This occurs even though individuals are not fully Bayesian, but requires the network to be strongly connected (i.e. no partisan agents are present).

In this paper, we consider simultaneously the possibility of learning from the news and being exposed to partisan agents. As a result, our environment encompasses ACFO (2013) and JMST (2012) as special cases. We first show that their results can be replicated by an appropriate choice of parameters. That is, we can show that by shutting down the degree of influence of partisan agents, individuals eventually learn the truth. But if partisan agents are influential, social media communication is not effective in aggregating information, and polarization persists in the long run. Our main contribution relative to the existing literature is that we simulate a large set of complex social networks and quantify the relative importance of behavioral assumptions, technological characteristics, and network topology on long-run polarization.

There is a growing empirical literature analyzing the effects of social media in opinion formation and voting behavior (Halberstam and Knight, 2016). Because individual opinions are unobservable from real network data, these papers typically use indirect measures of ideology to back-out characteristics of the network structure (such as homophyly) potentially biasing their impact. By creating a large number artificial networks, we can directly measure how homophily and other network characteristics affect opinion. Finally, our paper complements the literature on the role of biased media (Campante and Hojman, 2013, Gentzkow and Shapiro, 2011 and Flaxman et al. 2013) and social media (Weber at al 2013 and Barbera, 2016) on political polarization.

2 Baseline Model

Agents and Information Structure The economy is composed by a finite number of agents $i \in N = \{1, 2, \ldots, n\}$ interacting through a social network. Individuals have imperfect information about the true state of the world θ belonging to a parameter space $\Theta = [0, 1]$. This parameter is interpreted as the relative quality of two candidates, L and R, competing for office. A value of $\theta = 0$ implies that candidate L is better suited for office, whereas $\theta = 1$ implies that R is more qualified.

Each agent starts with a prior belief $\theta_{i,0}$ assumed to follow a Beta distribution,

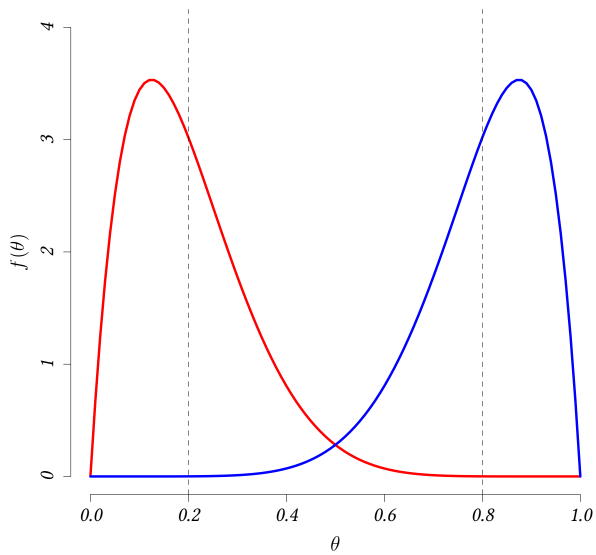
$$\theta_{i,0} \sim \mathcal{B}e\Big(\alpha_{i,0}, \beta_{i,0}\Big).$$

This distribution or *world-view* is characterized by initial parameters $\alpha_{i,0} > 0$ and $\beta_{i,0} > 0$. Note that individuals agree upon the parameter space Θ and the functional form of the probability distribution, but have different world-views as they disagree on $\alpha_{i,0}$ and $\beta_{i,0}$. Given their prior beliefs, we define her initial *opinion* $y_{i,0}$ about the relative quality of the candidates as her best guess of θ given the available information,[2]

$$y_{i,0} = \mathbb{E}[\theta|\Sigma_0] = \frac{\alpha_{i,0}}{\alpha_{i,0} + \beta_{i,0}}$$

where $\Sigma_0 = \{\alpha_{i,0}, \beta_{i,0}\}$ denotes the information set available at time 0.

Example 1. *In the Figure below, we depict the world-views of two individuals (distributions) and their associated opinions (vertical lines). The world-view that is skewed to the right is represented by the distribution $\mathcal{B}e(\alpha = 2, \beta = 8)$. The one skewed to the left is represented by the distribution $\mathcal{B}e(\alpha = 8, \beta = 2)$. The opinions are, respectively, 0.2 and 0.8.*



At each point in time $t \geq 1$ regular agent i reads a series of mainstream newspapers and reports from unbiased sources that are jointly informative about the relative quality of these candidates. We formalize the information obtained from mainstream media as a draw $s_{i,t}$ from a Bernoulli distribution with parameter θ. This implies that the signal is unbiased, as signals are on average equal to the true state. Partisan agents, on the other hand, disregard mainstream media and focus on information provided by sources aligned with their political color. We formalize their signals $s_{i,t}^p$ as draws from a Bernoulli distribution with parameter θ^p for $p \in \{L, R\}$, with $\theta^L = 0$ and $\theta^R = 1$. Note that these signals are always biased, this could either be due to slant in certain media sources or the production of fake news.

[2] Note that $\mathbb{E}[\theta/\Sigma_0]$ is the Bayesian estimator of θ that minimizes the mean squared error given a Beta distribution.

Social Network We assume that regular agents update their world-views and opinions based not only on private signals $s_{i,t}$, but also through the influence of individuals connected to them in a social network.

The connectivity among agents in the network at each point in time is described by a directed graph $\mathbf{G}^t = (N, \mathbf{g}^t)$, where $\mathbf{g}^t$ is a real-valued $n \times n$ adjacency matrix. Each regular element g_{ij}^t in the directed-graph represents the connection between agents i and j at time t. More precisely, $g_{ij}^t = 1$ if i is paying attention to (e.g. receiving information from) j, and 0 otherwise. Since the graph is directed, it is possible that some agents pay attention to (e.g. receive information from) others who are not necessarily paying attention to (e.g. obtaining information from) them, i.e. $g_{ij}^t \neq g_{ji}^t$. The out-neighborhood of any agent i at any time t represents the set of agents that i is receiving information from (e.g. i's references), and is denoted by $N_i^{out}(\mathbf{g}^t) = \{j | g_{ij}^t = 1\}$. Similarly, the in-neighborhood of any agent i at any time t, denoted by $N_i^{in}(g^t)$, represents the set of agents that are receiving information from i (e.g. i's audience), $N_i^{in}(\mathbf{g}^t) = \{j | g_{ji}^t = 1\}$. We define a directed path in $\mathbf{G}^t$ from agent i to agent j as a sequence of agents starting with i and ending with j such that each agent is a neighbour of the next agent in the sequence. We say that a social network is *strongly connected* if there exists a directed path from each agent to any other agent.

In the spirit of Acemoglu, Ozdaglar, and ParandehGhebi (2010) and ACFO (2012), we allow the connectivity of this graph g_{ij}^t to change over time stochastically. This structure captures rational inattention, incapacity of processing all information, or impossibility to pay attention to all individuals in the agent's social clique. More specifically, for all $t \geq 1$, we associate a *clock* to every directed link of the form (i,j) in the initial adjacency matrix g^0 to determine whether the link is activated or not at time t. The ticking of all clocks at any time is then dictated by i.i.d. samples from a Bernoulli Distribution with fixed and common parameter $\rho \in [0, 1]$, meaning that if the (i,j)-clock ticks at time t (realization 1 in the Bernoulli draw), then agent i receives information from agent j. The Bernoulli draws are represented by the $n \times n$ matrix $\mathbf{c}^t$, with regular element $c_{ij}^t \in \{0, 1\}$. Thus, the adjacency matrix of the network evolves stochastically across time according to the following equation[3]:

$$\mathbf{g}^t = \mathbf{g}^0 \circ \mathbf{c}^t, \tag{1}$$

where the initial structure of the network, represented by the initial adjacency matrix $\mathbf{g}^0$, remains

[3]The notation $\circ$ denotes the Hadamard Product, or equivalently, the element-wise multiplication of the matrices.

unchanged.

Example 2 (Bernoulli Clock)**.** *In this example we intend to illustrate the network dynamics. The figure in Panel 2a represents the original network and its adjacency matrix, whereas the figure in Panel 2b depicts a realization such that agent 1 does not pay attention to agents 2 and 4 in period* 1*. Agents 2 and 3, on the other hand, pay attention to agent 1 in both periods.*

$$\mathbf{g}^0 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{g}^1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

(a) Original Network at $t = 0$ (b) Potential Network at $t = 1$

Figure 2: Bernoulli Clock and Network Dynamics

Evolution of Beliefs Before the beginning of each period, agent i meets individuals in her out-neighbourhood $N_i^{out}(\mathbf{g}^t)$, a set determined by the realization of clock. These neighbors share their world-views, summarized by $\alpha_{j,t}$ and $\beta_{j,t}$ for all $j \in N_i^{out}(\mathbf{g}^t)$. At the beginning of period t, a signal profile is realized and the signal $s_{i,t}$ is privately observed by agent i.

Regular agents

After observing this signal, regular agent i computes her Bayesian posterior conditional on $s_{i,t}$. We assume that parameters $\alpha_{i,t+1}$ and $\beta_{i,t+1}$ are convex combinations between her Bayesian posterior and the weighted average of the information obtained from her neighbors.

$$\alpha_{i,t+1} = b_{i,t}[\alpha_{i,t} + s_{i,t}] + (1 - b_{i,t}) \sum_{j \in N_i^{out}(\mathbf{g}^t)} \hat{g}_{ij}^t \alpha_{j,t} \tag{2}$$

$$\beta_{i,t+1} = b_{i,t}[\beta_{i,t} + 1 - s_{i,t}] + (1 - b_{i,t}) \sum_{j \in N_i^{out}(\mathbf{g}^t)} \hat{g}_{ij}^t \beta_{j,t}, \tag{3}$$

where

$$b_{i,t} = \mathbb{1}_{\left\{\sum_j \hat{g}_{ij}^t = 0\right\}} 1 + \left(1 - \mathbb{1}_{\left\{\sum_j \hat{g}_{ij}^t = 0\right\}}\right) b \tag{4}$$

denotes the reliance weight given to mainstream media and $1 - b_{i,t}$ captures the influence of friends through social media. The parameter $b_{i,t} \in [0, 1]$ captures the attention span: a regular agent's full attention span is split between processing information from unbiased sources (e.g. reading the news from mainstream media, technical and scientific reports, etc.) and that provided by their friends in the network (e.g. reading a Facebook or Twitter feed). If no friends are found in the neighborhood of agent i, then this agent attaches weight 1 to the signal received. Conversely, if at least one friend is found, this agent uses a common weight $b \in [0, 1]$. The term $\hat{g}_{i,j}^t = \frac{g_{i,j}^t}{|N_i^{out}(\mathbf{g}^t)|}$ represents the weight given to the information received from her out-neighbor j. When $b_{i,t} = 1$ for some t agent i fully relies on her private signal behaving like a standard Bayesian agent. As $b_{i,t}$ approaches zero, she is more influenceable by social media, as more weight is given to her friends' opinions.

Finally, note that this updating rule implies that the posterior distribution determining world-views of agent i will also be a Beta distribution with parameters $\alpha_{i,t+1}$ and $\beta_{i,t+1}$. Hence, an agent's opinion regarding the true state of the world at t can be computed as

$$y_{i,t} = \frac{\alpha_{i,t}}{\alpha_{i,t} + \beta_{i,t}}.$$

Our heuristic rule resembles the one in JMST (2012), but there are two important distinctions. First, their adjacency matrix is fixed over time, whereas ours is stochastic (an element we borrowed from ACFO, 2012). Second, we restrict attention to a specific family of distributions (e.g. Beta) and assume that individuals exchange parameters that characterize this distribution (e.g. $\alpha_{i,t}$ and $\beta_{i,t}$). So the heuristic rule involves updating these parameters, whereas JMST (2012)'s heuristic rule involves a convex combination of the whole distribution. The latter implies that posterior distributions may not belong to the same family as the prior distribution.

Partisan agents

We assume that there are two types of partisan agents, L-wing partisan and R-wing partisan, with extreme views. They disregard information from other agents in the network, implying $b_{i,t} = 1$ for all t and only based their beliefs on the information obtained from the biased signal

$s_{i,t}^p$, for $p \in \{L, R\}$

$$\alpha_{i,t+1}^p = \alpha_{i,t}^p + s_{i,t}^p$$
$$\beta_{i,t+1}^p = \beta_{i,t}^p + 1 - s_{i,t}^p,$$

where $s_{i,t}^L$ is drawn from a Bernoulli with parameter $\theta^L = 0$ and $s_{i,t}^R$ is drawn from a Bernoulli with parameter $\theta^R = 1$. To the extent that they are influential (e.g. that their in-neighborhood is large), their presence in the network will be key for the persistence of polarization over time. This is due to the fact that they will consistently communicate biased signals (i.e., fake news) to other agents in the network.

3 Polarization and network structure

We base our notion of polarization on the seminal work by Esteban and Ray (1994), adapted to the context of this environment. At each point in time, we partition the $[0, 1]$ interval into $K \leq n$ segments. Each segment represents significantly-sized groups of individuals with similar opinions. We let the share of agents in each group $k \in \{1, ..., K\}$ be denoted by $\pi_{k,t}$, with $\sum_k \pi_{k,t} = 1$.

Esteban and Ray (1994)'s polarization measure aggregates both 'identification' and 'alienation' across agents in the network. Identification between agents captures a sense of political alignment: an individual feels a greater sense of identification if a large number of agents in society shares his or her opinion about the relative quality of the candidate. In this sense, identification of a citizen at any point in time is an increasing function of the share of individuals with a similar opinion. The concept of identification captures the fact that *intra*-group opinion homogeneity accentuates polarization. On the other hand, an individual feels alienated from other citizens if their opinions diverge. The concept of alienation captures the fact that *inter*-group opinion heterogeneity amplifies polarization. Mathematically, we have the following representation.

Definition 1 (Polarization). *Polarization P_t aggregates the degrees of 'identification' and 'alienation' across groups at each point in time.*

$$P_t = \sum_{k=1}^{K} \sum_{l=1}^{K} \pi_{k,t}^{1+a} \, \pi_{l,t} \, |\tilde{y}_{k,t} - \tilde{y}_{l,t}| \tag{5}$$

where $a \in [1, 1.6]$ and $\tilde{y}_{k,t}$ is the average opinion of agents in group k and $\pi_{k,t}$ is the share of agents in group k at time t.

We are interested in understanding how the existence of partisan agents and the structure of the network affect the evolution of polarization.

Polarization without partisan agents The following two results show conditions under which polarization vanishes in the limit. The first one is analogous to Sandroni et al (2012), whereas the second one extends it to a network with dynamic link formation as in Acemoglu et al (2010).

Proposition 1. *If the network $\boldsymbol{G}^0 = (N, \mathbf{g}^0)$ is strongly connected and if the directed links are activated every period (e.g., $\mathbf{g}^t = \mathbf{g}^0$), all agents eventually learn the true θ*

$$\max_i |\operatorname*{plim}_{t\to\infty} y_{i,t} - \theta| < \epsilon$$

As a consequence, polarization converges to zero,

$$\operatorname*{plim}_{t\to\infty} P_t = 0.$$

Proof. See Appendix A. □

When the network is strongly connected all opinions and signals eventually travel through the network allowing agents to perfectly aggregate information. Note that strong connectedness precludes the existence of partisan agents, as these agents do not internalize other people's opinions. The proposition shows that the society reaches consensus (e.g. there is no polarization) and uncovers the true relative quality of the political candidates, θ. We refer to this as a 'wise' society, as defined below.

Definition 2 (Wise Society). *We say that a society is wise if*

$$\max_i |\operatorname*{plim}_{t\to\infty} y_{i,t} - \theta| < \epsilon.$$

The result in Proposition 1 is in line with the findings in JMST (2012) despite the difference in heuristic rules being used. Proposition 2 shows that the assumption of a fixed listening matrix can be relaxed. In other words, even when $\mathbf{g}^t$ is not constant, polarization vanishes in strongly connected networks.

Proposition 2. *If the network $\boldsymbol{G}^0 = (N, \mathbf{g}^0)$ is strongly connected, even when the edges are not activated every period, polarization still converges to zero,* $\text{plim}_{t\to\infty} P_t^a = 0$.

Proof. See Appendix B. □

Polarization with partisan agents The presence of partisan agents breaks the strong connectivity in the network, but this does not necessarily imply that the society will exhibit polarization. The following example depicts two networks, with three regular agents (2, 3, and 4) and one partisan agent—left-winged in panel (a) and right-winged in the panel (b)—.

(a) Society with L-partisan (b) Society with R-partisan

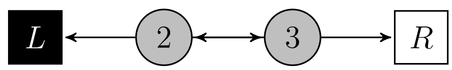
Figure 3: Two societies with partisan agents

Polarization in both societies converges to zero in the long-run. However, neither society is wise. This illustrates that the influence of partisan agents may generate mis-information in the long run, preventing agents from uncovering θ, but does not necessarily create polarization. This insight is formalized in Proposition 3

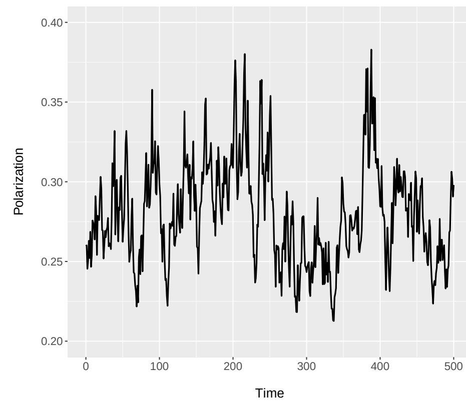
Proposition 3. *A wise society experiences null social polarization. However, not all societies that experience null social polarization are wise.*

Proof. If perfect information aggregation is reached at any particular time $\bar{t}$, then we know that $y_{i,\bar{t}} = \theta$ for all $i \in G$, thus all alienation terms in the polarization function are zero because $|y_{i,\bar{t}} - y_{j,\bar{t}}| = |\theta - \theta| = 0$, for all i and j in N. Therefore, Polarization $P_{\bar{t}}$ is zero for any particular choice of parameter a. Conversely, if polarization at time $\bar{t}$ is zero, then all alienation terms are necessarily zero, since the measure of groups is non-negative. This means that $|y_{i,\bar{t}} - y_{j,\bar{t}}| = 0$ implies $y_{i,\bar{t}} = y_{j,\bar{t}}$ and, therefore, any opinion consensus of the form $y_{i,\bar{t}} = y_{j,\bar{t}} = \tilde{\theta}$, such that $\tilde{\theta} \in \Theta = [0, 1]$ and $\tilde{\theta} \neq \theta$, meets this requirement. □

In other words, it is possible for a society to reach consensus (i.e. experience no polarization of opinions) to a value of θ that is incorrect. In order for a society to be polarized, individuals need to be exposed to partisan agents with opposing views.



(a) Society with both L and R-wing partisans



(b) Cycles

Figure 4: Two societies with partisan agents

Consider the social network depicted in Figure 4a, in which both L-partisan and R-partisan agents are present. Even though agents 2 and 3 receive unbiased signals and communicate with each other (e.g. update their beliefs according to eqs. 2 and 3), this society exhibits polarization in the long run. This happens because partisan agents subject to different biases (e.g. left-wing and right-wing) are influential.

Another noticeable characteristic of the evolution of P_t over time is that rather than settling at a constant positive value, it fluctuates in the interval [0.2,0.4]. The example illustrates that polarization cycles are possible in this environment. In the simulation analysis of Section 4 we will discuss what determines the presence or absence of such cycles.

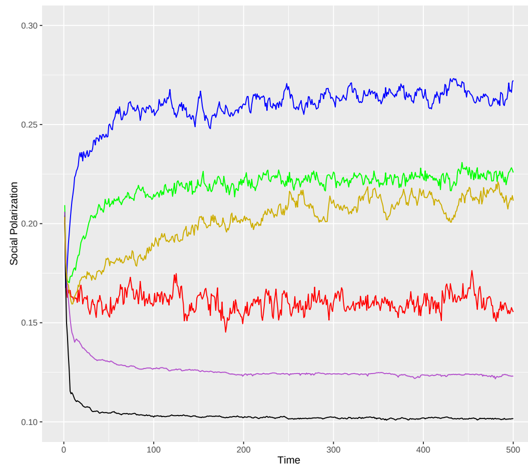


Figure 5: Different polarization levels

Finally, we want to point out that whether polarization increases, decreases, or fluctuates over time depends importantly on the topology of the network, the number and degree of influence of partisan agents, the frequency of meetings between individuals (e.g. the clock) and the degree of rationality of agents. Figure 5 depicts the behavior of P_t over time for a series of larger random networks (e.g. there are 100 nodes, an arbitrary number of partisan agents, and different rationality levels). The next section is devoted to uncovering what drives these different dynamics.

4 Simulation and Regression Analysis

One of the biggest challenges when using network analysis is to ascertain analytical closed forms and tractability. The combinatorial nature of social networks that exhibit a high degree of heterogeneity makes them very complex objects, imposing a natural challenge for theoretical analysis. In our work, limiting properties can be characterized only when we assume strong connectivity (absence of partisan agents). As we drop this assumption, we observe that different networks might experience different limiting polarization levels, even if the initial characteristics are relatively similar.

We resort to computer simulations where a large number of random networks are generated according to classical random network models. Besides emulating real-world networks characteristics, these models allow us to create a variety of initial networks with different characteristics (e.g. degree of centrality, presence of partisan agents, partisan agents' influence, homophily, etc) and learning standards (i.e., exposition to signals from mainstream and/or social media). The simulation exercise helps us to better understand the relative importance of the network topology and other social characteristics in driving polarization by producing enough variability in a controlled environment.

4.1 Random networks models

We will run the simulation for random networks generated from three basic models: the Erdos-Renyi model, the Barabasi-Albert model, and a Combo model which includes characteristics of the other two.

Erdos-Renyi model. Erdos and Renyi (1960) is a seminal study exploring purely random networks formation. In their canonical model, they fix a set of nodes n and assume that each link between two agents (possibly directed or undirected) is formed with a given probability. The link formation is independent across links, implying that the probability p that a node has k edges follows a Poisson distribution. A particular feature of this model is that the probability of finding a highly connected node (a large k) decreases exponentially with k. Thus, nodes with large connectivity are practically absent. An example is presented in Figure 6, in which we observe an absence of influential agents and a relatively high degree of connectivity among individuals.

Even after the upsurge of topological information of real-world networks, this model remains very popular and useful to understand properties of connected and unconnected networks due to its simplicity. Moreover, since this is a very mature model, many useful mathematical regularities are well established, providing a strong basis for basic understanding, comparisons, and extensions.

In contrast to this random network model, most real-world networks exhibit preferential connectivity, a common characteristic of business and social networks. This is a feature produced by the following random network model proposed by Barabasi and Albert (1999).

Barabasi-Albert model. Barabasi and Albert were mainly motivated by the emergence of the World Wide Web and the evolution of popularity of some web pages. They noted that popular web pages would show a tendency to get more popular over time. The popularity of web pages in this context refers to the number of other web pages pointing a direct link to them. This characteristic means that new entrant nodes (web pages) tend to link themselves to already existent nodes that are very well connected (popular web pages), indicating that the probability with which a new node connects to the existing nodes is not uniform. Contrarily, there is a higher probability that it will be linked to a node that already has a large number of connections. An implication of this characteristic is that a few nodes in the network are very well connected while most of the other nodes are not as well connected and “hubs” will be formed.

In this context, Barabasi and Albert (1999) developed an algorithm to generate random networks with such characteristics using a process called *preferential attachment*. In this process, starting with a small number n_0 of nodes, at every time step a new node with $m(\leq n_0)$ edges is added to the network. Thus, the new node links to m different nodes already present in the system. To incorporate preferential attachment, they assume that the probability Π that a new node will be connected to node i depends on the connectivity k_i (in-degree, or the number of nodes pointing to them) of that node, so that $\Pi = \frac{k_i}{\sum_j k_j}$. After t periods, this protocol leads to a random network with $t + n_0$ nodes and mt edges. Figure 7 illustrates a random network generated following this procedure. In it, there is a small subset of agents in the center of the network with a relatively large audience. In our context, each node represents an agent. Individuals with a larger audience are more influential.

While this model allow us to introduce influential agents, it rules out reciprocity. That is, if agent i pays attention to agent j, agent j never pays attention to agent i. In other words, agents do not exchange information as in the Erdos and Renyi model. Because social media networks exhibit both characteristics, namely influential individuals and exchange of information, we will consider a *combo* model which combines both properties.

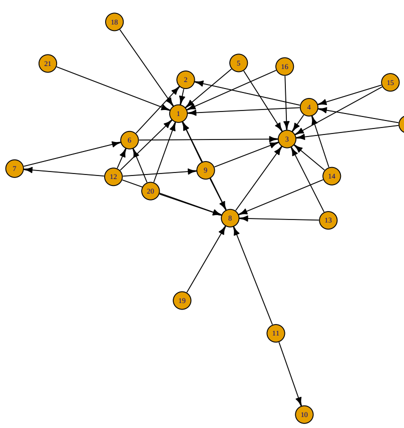


Figure 6: Barabasi-Albert
$n = 28$, Power$= 1.5$,
Out Dist = $\mathcal{G}(3, 1)$

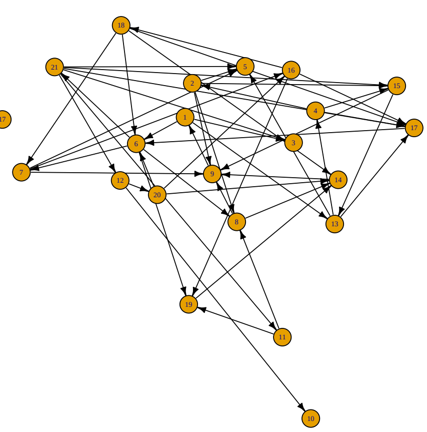


Figure 7: Erdos-Renyi
$n = 28, p = \frac{\log(n)}{n}$

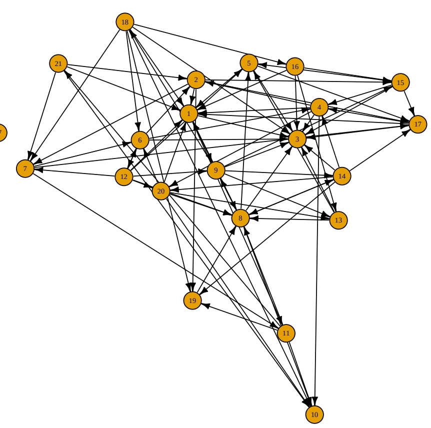


Figure 8: **Combo**
$n = 28$, **Power**$= 1.5$
Out Dist = $\mathcal{G}(3, 1)$

Combo model. We implement a small variant in the mechanism to produce Barabasi-Albert random network in order to produce heterogeneity of connections. In each step, instead of allowing only a fixed number $m(\leq n_0)$ of links to be formed, the number of edges of an entrant node at any time t will be the realization of a draw from a pre-specified distribution defined over the potential number of edges to add in each time step. If this rule is not implemented, we would implicitly assume that every regular agent would be paying attention to exactly $m(\leq n_0)$ agents. Figure 8 illustrates the network generated from the Combo model. In it, we can see that information flows in both directions and that some agents are more influential than others. This specification constitutes the benchmark case of our study.

4.2 Simulation: generating the dataset

Our artificial dataset is created as follows. We fix the true state of the world, $\theta = 0.5$, the number of agents (or nodes) $n = 35$, the number of partisan agents, and the average degree of polarization P^0 across simulations. We restrict attention to networks with a symmetric number of L-wing and R-wing partisan agents. Given these parameters, we draw $M = 575$ initial random networks $\mathbf{G}^0$ following the Combo model.[4] We then assign initial conditions and other characteristics that vary across simulations. More specifically, we vary the speed of communication through the clock parameter ρ, the weight given to unbiased sources b, the location of partisan agents in the network, the degree of initial homophily, clustering, and reciprocity. This produces the basic structure for social communication and determines the initial dispersion of beliefs about θ. Finally, we simulate social media communication for $t \in T = \{1, \ldots, 2000\}$ periods given the network structure, and use the resulting opinions to compute the evolution of polarization. We only consider the last 1000 periods to compute average polarization.

Network heterogeneity: For each network m, we fix the initial distribution of opinions so that the same mass of the total population lies in the middle point of each one of the 7 groups. This rule basically distributes our agents evenly over the political spectrum $[0, 1]$ such that each of the 7 groups contains exactly $\frac{1}{7}$ of the total mass of agents, as shown in Figure 9.

[4] The results for the simulations from the Erdos Reni and Barabasi-Albert model are presented in the appendix.

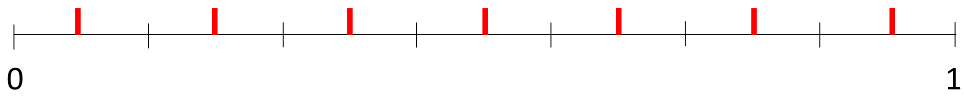


Figure 9: Initial distribution of opinions in all simulations

Moreover, we set the same variance for each agent world-view to be $\sigma^2 = 0.03$. With both opinion and variance, we are able to compute the initial parameter vector (α_0 , β_0) [5]. Among the agents populating our network, a predetermined number of agents is chosen, uniformly at random among those nodes with in-degree of at least one, to be partisan in each simulation. In the benchmark estimation, we consider only one partisan agent of each type, but we then analyze the robustness of our findings to a larger share of partisan agents (2, 3, and 4 of each type). [6]

The location and degree of influence of partisan agents, which are key to determining polarization, vary randomly across simulations. The only restriction imposed is that partisan agents must have at least one individual in their in-neighborhood at $t = 0$.

We also allow the degree of rationality b and the parameter ρ of the Bernoulli distribution determining the ticking of the clock (e.g. the persistence of connections) to vary across networks. We draw b_m and ρ_m from discrete Uniform distributions defined over the interval $[0, 1]$ (with band-with 0.05) for each network $m \leq M$.

This approach is able to emulate many initial configurations, each one representing a potential society with significant variability in terms of connections, beliefs, exposure to social media, homophily, etc.

Simulation: For each network m, we draw a signal $s_{i,t}^m$ for individual $i \in N$ at time $t \in T$ from a Bernoulli distribution with parameter $\theta = 0.5$ for regular agents, or parameters $\theta^p \in \{0, 1\}$ for partisan agents. We also draw the $n \times n$ matrix $\mathbf{c}^t$ at each period t from a Bernoulli distribution with parameter ρ_m, which determines the evolution of the network structure according to eq. (**??**). Together, the signals and the clock determine the evolution of world-views according to eqs. (2) and (3). Polarization $P_{m,t}$ is computed according to eq. (5) at each point in time. Our variable of interest is the level of polarization in the long-run, $\bar{P}_m$. This is defined as the average

[5] In this case, we only need to use the relationships $\mu = \frac{\alpha}{\alpha+\beta}$ and $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ to fully determine α and β. Algebraic manipulation yields $\alpha = -\frac{\mu(\sigma^2+\mu^2-\mu)}{\sigma^2}$ and $\beta = \frac{(\sigma^2+\mu^2-\mu)(\mu-1)}{\sigma^2}$.

[6] The initial beliefs (α, β) for both L-winged and R-winged partisan agents are given by the rule described above. These imply that the initial opinions about the relative quality of candidates, θ, are not extreme in the very beginning of the interactions, however, it becomes relatively extreme after a few rounds once they update their parameters using the biased update rule.

value of $P_{m,t}$ for t larger than a threshold $\bar{t}$, which in our simulations is set to be $\bar{t} = 1000$. We chose this threshold because simulations converge after about 1000 periods to an ergodic set. Mathematically,

$$\bar{P}_m \equiv \sum_{t>\bar{t}} \frac{1}{T - \bar{t} - 1} P_{m,t}$$

The advantage of this approach with respect to others that use real-world networks is that we can combine any network structure with any initial belief structure. Moreover, we can guarantee that all variability is statistically independent. On the other hand, as with any form of analysis, this approach has its disadvantages as well. There are important considerations in terms of how sensitive our conclusions are to the number of simulations performed M, the length of each simulation in T, the size of the society n, and the values that other variables capturing heterogeneity can take. We will perform some robustness to these results. Finally, the curse of dimensionality is a challenging aspect of this approach due to its effects on computation time.

4.3 Regression analysis

We are interested in estimating the effect of network characteristics on long-run polarization. To assess the quantitative importance of each explanatory variable, we estimate the coefficients of a simple ordinary least squares linear model,

$$\bar{P}_m = \mathbf{X}_m \beta + \epsilon_m.$$

where the $m \times 1$ vector $\bar{P}_m$ denotes long-run polarization obtained from simulation $m \in \{1 \ldots M\}$, $\mathbf{X}_m$ denotes $m \times k$ matrix of network characteristics (where k is the number of explanatory variables described ahead), and the $m \times 1$ vector ϵ_m is the error term.

The set of explanatory variables can be split in three categories: *behavioral, technological*, and *topological*. On the *behavioral* dimension, we consider the effect of changing the extent that agents rely more heavily on the opinion of others (and less on incorporating signals from unbiased sources) by varying the parameter b on the updating rule. A higher value of b gives more weight to the Bayesian posterior from unbiased signals. The initial conjecture is that the higher the value of this parameter is, the lower the polarization in any society is: agents "mute" the network channel through which fake news permeate and restrict attention to the unbiased signal, facilitating information aggregation and reducing polarization.

The main parameter capturing communication *technology* is ρ, which controls the speed at which links are activated. A higher value of ρ could be interpreted as higher speed of information flow, i.e. the easiness to access all friends opinions. The overall effect of a higher ρ is a priori ambiguous: on the one hand, it is more likely hat a regular agent will (indirectly) incorporate fake news from those paying attention to the extreme views of partisan agents; on the other hand, a faster flow of information makes it more likely to form consensus among regular agents.

In terms of the network *topology* we consider i) the degree of popularity of partisan agents, capturing the number of 'followers' in real-life social media (such as Facebook or Twitter) and proxied by their in-degree centrality; ii) the degree of influence of the partisan agents' followers, proxied by the Google PageRank centrality, iii) average clusteringing, i.e. the presence of "triangles" in the social network, measuring the tendency that friends of my friends are also my friends), iv) reciprocity, measured as the percentage of the relationships that are mutual, and v) initial homophily, measured by the degree to which initial nodes with similar beliefs are mutually tied[7]. The last three characteristics jointly capture echo chambers effects, interpreted as the tendency that people find their biases confirmed and augmented in social networks.

Our initial conjectures regarding the effect of the topology over polarization are that i) larger number of partisan followers increase polarization, since they will spread the extreme fake news among agents; ii) partisan followers centrality is most likely negative since the more central these agents are, the easier it is to nudge other agents to the their political position, reducing alienation and increasing identification; iii) average clustering is a priori ambiguous: on one hand, it might reinforce biases and on the other hand it increases the degree of connectivity of the society; iv) reciprocity is likely to reduce polarization since it facilitates consensus between any two agents communication; and finally v) initial homophily in opinions is expected to increase polarization or be neutral, if its effect vanishes in the long run.

The estimated parameters using a standard linear model are summarized in Table 1[8]. The first column shows the result of a linear regression using data of societies with one partisan agent of each type, whereas the second to fourth columns show results for networks including 2, 3, and 4 partisan agents of each type, respectively.

[7] In political science and economic networks literatures, homophily is a characteristic that drives link formation. In our case, initial homophily is simply a statistic of assortativity computed over opinions after the initial random network is fully characterized and populated with different agents and beliefs. The degree of homophily in the long-run is endogenously determined. In an environment with no partisan agents, for example, all agents converge to the same opinion.

[8] Since polarization level is a variable bounded between zero and one, the effect of any explanatory variables tends to be non-linear and regular linear models are limited to account for this feature. In this situation, it is appropriate to use a either a logit/probit (link) of a binomial family or a generalized beta regression models.

Table 1: Linear regression results

	Dependent Variable: Average Polarization [1000,2000]			
	2 partisans (1:1)	4 partisans (2:2)	6 partisans (3:3)	8 partisans (4:4)
	(1)	(2)	(3)	(4)
Bayes λ	−0.019***	−0.021***	−0.037***	−0.029***
	(0.005)	(0.004)	(0.004)	(0.004)
Clock p	−0.047***	−0.042***	−0.052***	−0.047***
	(0.005)	(0.004)	(0.004)	(0.004)
Audience of Both (% of nodes)	0.133***	0.207***	0.294***	0.382***
	(0.041)	(0.028)	(0.026)	(0.023)
Audience of Both × PageRank	−0.103**	−0.246***	−0.244***	−0.255***
	(0.049)	(0.033)	(0.030)	(0.028)
Audience of L (% of nodes)	0.144***	0.246***	0.378***	0.427***
	(0.028)	(0.025)	(0.025)	(0.025)
Audience of L × PageRank	−0.142***	−0.184***	−0.202***	−0.157***
	(0.029)	(0.025)	(0.025)	(0.026)
Audience of R (% of nodes)	0.115***	0.232***	0.393***	0.467***
	(0.028)	(0.023)	(0.026)	(0.025)
Audience of R × PageRank	−0.110***	−0.168***	−0.220***	−0.199***
	(0.029)	(0.023)	(0.025)	(0.026)
Initial Homophily	−0.013	−0.029	0.013	−0.015
	(0.029)	(0.023)	(0.020)	(0.017)
Average Cluster	0.203***	0.206***	0.156***	0.080***
	(0.011)	(0.011)	(0.013)	(0.012)
Reciprocity	−0.147***	−0.058*	−0.023	−0.039
	(0.035)	(0.031)	(0.033)	(0.034)
Observations	575	850	868	824
Adjusted R^2	0.817	0.890	0.928	0.957
Residual Std. Error	0.037 (df = 564)	0.036 (df = 839)	0.034 (df = 857)	0.030 (df = 813)
F Statistic	233.702*** (df = 11; 564)	625.679*** (df = 11; 839)	1,016.950*** (df = 11; 857)	1,661.630*** (df = 11; 813)

Note: *p<0.1; **p<0.05; ***p<0.01

From the first column, we can see that if agents place more weight on the unbiased signal (and less on the social media friends' opinions), polarization falls. On average, a one percentage point (p.p.) increase in this parameter leads to a 1.9 p.p. reduction in polarization. The speed of communication, as measured by the ticking persistence of our clock, reduces polarization by 4.7 p.p. for every 1 p.p. increase. This result suggests that the effect of internalizing a larger number of opinions outweighs the effect of higher fake news exposure. Moreover, the marginal effect of this parameter seems to be very stable across all 4 cases.

As the popularity of partisan agents rises — measured by the number of their followers — polarization is exacerbated. The marginal effect of an additional follower on polarization depends on the popularity of the follower itself. To capture this non-trivial effect, we introduce an interaction term between followers (quantity centrality) and influence (quality centrality, as measured by Google PageRank centrality). The coefficient on the interaction term is negative, confirming the conjecture that if the partisan agent's audience is sufficiently influential, the average opinion in society may be nudged towards the partisan's political point. This happens through a reduction in alienation and an increase in identification (see Definition 1 for a description of these terms). The net effect of increasing the number of partisan's followers on polarization can be computing by evaluating PageRank centrality at its sample mean. The resulting coefficients are presented in the first column of Table 2 in Appendix C. We find that on average, a higher number of followers rises polarization.

When comparing these parameters across experiments we find that the net marginal effect increases as the percentage of partisan agents increases: in the case with 2 fanatics, if the audience of both partisan agents increase by 1 p.p., polarization increases 5.9 p.p. However, when the percentage of partisan agents present in the network increases, this marginal effect becomes 18.4 p.p. (last column in Table 2). A similar effect is observed when the number of followers of just one partisan agent increases. Finally, we find that an increase in the number of followers of just one partisan agent rises polarization more that an increase in the number of followers paying attention to both types of partisan agents (see Table 2).

Finally, the effects of initial homophily on polarization vanish over time, as seen by the fact that the coefficient is statistically insignificant. Average clustering and reciprocity play opposite roles. Average clustering increases polarization, which suggests that the higher connectivity reached with higher clustering is not sufficiently strong to countervail the bias reinforcement associated to it. When there are 2 partisan agents present, the effect of increasing average clustering by 1 p.p. leads to an increase of 20.3 p.p. in polarization. The marginal effect of this parameter, however, decreases significantly as the percentage of partisan agents increases (e.g.

it is just 8.0 p.p. when there are 8 partisan agents present). Reciprocity reduces polarization in networks populated by a small number of partisan agents (by 14.7 p.p. and 5.8 p.p. for every increase of 1 p.p., when there are 2 and 4 partisans), but loses significance after some threshold of partisans present in the network is reached. These last three parameters increase polarization on average. This evidence is consistent with Sunstein's theory (2002, 2009) that echo chambers increase polarization.

5 Conclusions

We simulated a large number of social media networks by varying their characteristics in order to understand what the most important drivers of polarization are. A premise in all of them is the presence of partisan agents with opposite extreme views who purposely spread fake news (i.e. biased signals) aligned with their political color. To the extent that regular agents can be partially influenced by these signals—directly by 'following' the partisan agent, or indirectly by following friends who are themselves influenced by fake news—, this generates polarization in the long run. In other words, fake news prevent information aggregation and consensus in the population. An important assumption is that the links in the network evolve stochastically. It would be interesting to extend the model to consider a case in which links are endogenously determined. That is, in which agents could choose to 'unfollow' friends who provide mostly biased information.

Our preliminary results suggest that the recent increase in polarization could be associated with the radical shift in communication technology experienced in the last twenty years. Even though the internet expanded the access to raw information and allowed individuals to share it at a faster speed, it also provided a channel for partisan agents with extreme views to manipulate information via fake news, factually inaccurate facts, and/or slanted and misleading rhetoric. The popularity of social media networks such as Facebook and Twitter has exacerbated this problem even further, as individuals may be exposed to fake news through their friends opinions. This results from the inability of regular agents to filter out the sources of information that shape their friends opinions. Two other characteristic of social media networks may have contributed to increased polarization. First, they tend to be highly clustered. Second, they make it costless for partisan agents to strengthen their influence by simply increasing their number of followers. Social media networks allow partisan agents to by-pass mainstream media, and communicate their unfiltered message directly to followers.

Having identified the main determinants of polarization, it would be interesting to parameterize a real-life social media network (e.g. calibrate it) in order to quantify what percentage of the increase in polarization can be explained by these channels. In addition, it is possible to carry forward a key-player analysis to better understand what is the most efficient way to reduce polarization. Finally, as can be noticed from the random networks simulation, polarization cycles are a strong regularity. Analyzing their determinants could be a fruitful avenue for future research.

A Proof of Proposition 1

Lemma 1. *The matrix* $W_t = B_t + (\mathbb{I}_n - B_t)\,\hat{g}_t$ *is row-stochastic in any period* t*, where* $B_t =$ *diag*$(b_{1,t}, b_{2,t}, \ldots, b_{n,t})$.

Proof. It is sufficient to show that $W_t\mathbf{1} = B_t\mathbf{1} + (\mathbb{I}_n - B_t)\,\hat{g}_t\mathbf{1} = \mathbf{1}$. For that we can show that the vector $W_t\mathbf{1}$, for every t, has all entries equal to

$$b_{i,t} + (1 - b_{i,t})\hat{g}^t_{i,*}\mathbf{1} = \begin{cases} b_{i,t} = \mathbb{1}_{\{\hat{g}^t_{i,*}\mathbf{1}=0\}}1 + \left(1 - \mathbb{1}_{\{\hat{g}^t_{i,*}\mathbf{1}=0\}}\right) b = 1 & \text{, if } \hat{g}^t_{i,*}\mathbf{1} = 0 \\ 1 & \text{, if } \hat{g}^t_{i,*}\mathbf{1} = 1 \end{cases}$$

, where $\hat{g}^t_{i,*}$ is the i-th row of matrix $\hat{g}^t$. □

Lemma 2. *The iteration of the row-stochastic matrix* W *is convergent and therefore there exists a threshold* $\bar{\tau} \in \mathbb{N}$ *such that* $|W^{\tau+1}_{ij} - W^{\tau}_{ij}| < \epsilon$ *for any* $\tau \geq \bar{\tau}$ *and* $\epsilon > 0$

Proof. In order to see how W^τ behaves as τ grows large, it is convenient to rewrite W using its diagonal decomposition. In particular, let v be the squared matrix of left-hand eigenvectors of W and $\boldsymbol{D} = (d_1, d_2, \ldots, d_n)'$ the eigenvector of size n associated to the unity eigenvalue $\lambda_1 = 1$[9]. Without loss of generality, we assume the following normalization $\mathbf{1'}\boldsymbol{D} = 1$. Therefore, $W = v^{-1}\Lambda v$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is the squared matrix with eigenvalues on its diagonal, ranked in terms of absolute values. More genreally, for any time τ we write

$$W^\tau = v^{-1}\Lambda^\tau v.$$

Noting that v^{-1} has ones in all entries of its first column, it follows that

$$[W^\tau]_{ij} = d_j + \sum_r \lambda^\tau_r v^{-1}_{ir} v_{rj},$$

for each r, where λ_r is the r-th largest eigenvalue of W. Therefore, $\lim_{\tau\to\infty} [W^\tau]_{ij} = \boldsymbol{D}\mathbf{1'}$, i.e. each row of W^τ for all $\tau \geq \bar{\tau}$ converge to $\boldsymbol{D}$, which coincides with the stationary distribution. Moreover, if the eigenvalues are ordered the way we have assumed, then $\|W^\tau - \boldsymbol{D}\mathbf{1'}\| = o(|\lambda_2|^\tau)$, i.e. the convergence rate will be dictated by the second largest eigenvalue, as the others converge to zero more quickly as τ grows. □

[9] This is a feature shared by all stochastic matrices because having row sums equal to 1 means that $\|W\|_\infty = 1$ or, equivalently, $W\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the unity n-vector.

With these two auxiliary lemmas, we start by considering the parameter update process described in the Section (XX). Since the network's edges are activated every single period, $\hat{g}_t = \hat{g}$ and $B_t = B_{n\times n} = B = \text{diag}(b, b, \ldots, b)$, where $b \in [0, 1]$, since $\sum_j g_{ij}^t \neq 0$ for any i and t. Thus, the update process for the parameter-vector α of size n is

$$\begin{aligned}\alpha_{t+1} &= B(\alpha_t + s_{t+1}) + (\mathbb{I}_n - B)\hat{g}\alpha_t \\ &= [B + (\mathbb{I}_n - B)\hat{g}]\,\alpha_t + Bs_{t+1}.\end{aligned}$$

We define the matrix inside the squared bracket as W for any t. We re-write the update process above as follows

$$\alpha_{t+1} = W\alpha_t + Bs_{t+1}$$

When $t = 0$,

$$\alpha_1 = W\alpha_0 + Bs_1$$

When $t = 1$,

$$\begin{aligned}\alpha_2 &= W\alpha_1 + Bs_2 \\ &= W\,(W\alpha_0 + Bs_1) + Bs_2 \\ &= W^2\alpha_0 + WBs_1 + Bs_2\end{aligned}$$

When $t = 3$,

$$\begin{aligned}\alpha_3 &= W\alpha_2 + Bs_3 \\ &= W\left(W^2\alpha_0 + WBs_1 + Bs_2\right) + Bs_3 \\ &= W^3\alpha_0 + W^2Bs_1 + WBs_2 + Bs_3\end{aligned}$$

So on and so forth, resulting in the following expression for any particular period τ

$$\alpha_\tau = W^\tau\alpha_0 + \sum_{t=0}^{\tau-1} W^t Bs_{\tau-t} \tag{6}$$

Similarly for the parameter β, we have

$$\beta_\tau = W^\tau\beta_0 + \sum_{t=0}^{\tau-1} W^t B(\mathbf{1} - s_{\tau-t}). \tag{7}$$

where $\mathbf{1}$ is the vector of ones of size n. From Equations (6) and (7), the sum of this two parameter-vectors is given by the following expression

$$\begin{aligned}\alpha_\tau + \beta_\tau &= W^\tau(\alpha_0 + \beta_0) + \sum_{t=0}^{\tau-1} W^t B\mathbf{1} \\ &= W^\tau(\alpha_0 + \beta_0) + \sum_{t=0}^{\tau-1} W^t \mathbf{b} \\ &= W^\tau(\alpha_0 + \beta_0) + \tau\mathbf{b}. \end{aligned} \tag{8}$$

Therefore, at any point in time τ, the opinion of any agent i is given by $y_{i,\tau} = \dfrac{\alpha_{i,\tau}}{\alpha_{i,\tau} + \beta_{i,\tau}}$. From equation (6), we write

$$\begin{aligned}\alpha_{i,\tau} &= W_{i*}^\tau \alpha_0 + \sum_{t=0}^{\tau-1} W_{i*}^t b s_{\tau-t} \\ &= W_{i*}^\tau \alpha_0 + \tau b \frac{1}{\tau} \sum_{t=0}^{\tau-1} W_{i*}^t s_{\tau-t} \\ &= W_{i*}^\tau \alpha_0 + \tau b \tilde{\theta}_i(\tau), \end{aligned} \tag{9}$$

where the symbol W_{i*}^τ is used to denote the i-th row of matrix W^τ and $W^0 = \mathbb{I}_n$. From equations (9) and (8), we write $y_{i,\tau}$ as

$$\begin{aligned}y_{i,\tau} &= \frac{W_{i*}^\tau \alpha_0 + \tau b \tilde{\theta}_i(\tau)}{W_{i*}^\tau(\alpha_0 + \beta_0) + \tau b} \\ &= \frac{\tau}{\tau}\left(\frac{\frac{1}{\tau} W_{i*}^\tau \alpha_0 + b\tilde{\theta}_i(\tau)}{\frac{1}{\tau} W_{i*}^\tau(\alpha_0 + \beta_0) + b}\right), \end{aligned} \tag{10}$$

From Equation (10), we have that the limiting opinion (in probability) of any agent i, at any point in time τ, is described as

$$\begin{aligned}\operatorname*{plim}_{\tau\to\infty} y_{i,\tau} &= \operatorname*{plim}_{\tau\to\infty} \tilde{\theta}_i(\tau) \\ &= \operatorname*{plim}_{\tau\to\infty} \frac{1}{\tau} \sum_{t=0}^{\tau-1} W_{i*}^t s_{\tau-t} \\ &= \operatorname*{plim}_{\tau\to\infty} \frac{1}{\tau} \sum_{t=0}^{\bar{\tau}} W_{i*}^t s_{\tau-t} + \operatorname*{plim}_{\tau\to\infty} \frac{1}{\tau} \sum_{t=\bar{\tau}+1}^{\tau} W_{i*}^t s_{\tau-t}. \end{aligned} \tag{11}$$

From Lemma 2, we can split the series in Equation (12) into two parts. The first term describes a series of $\bar{\tau}$ terms that represent the "most recent" signals coming in to the network. Notice that every weight-matrix W^t in the interval from $t = 0$ to $t = \bar{\tau}$ is different from one another, since the matrix W^t does not converge to a row-stochastic matrix with unity rank for low t. It is straight-forward to see that this term converges to zero as $\tau \to \infty$. The second term represents describes a series of $\tau - \bar{\tau}$ terms that represent the"older signals" that entered in the network and fully reached all agents. As $\tau \to \infty$, this term becomes a series with infinite terms. From the i.i.d. property of the Bernoulli signals, we can conclude that

$$\begin{aligned}
\operatorname*{plim}_{\tau\to\infty} y_{i,\tau} &= \operatorname*{plim}_{\tau\to\infty} \frac{1}{\tau} \sum_{t=\bar{\tau}+1}^{\tau} W_{i*}^{t} s_{\tau-t} \\
&= \operatorname*{plim}_{\tau\to\infty} \frac{1}{\tau} \sum_{t=\bar{\tau}+1}^{\tau} \boldsymbol{W}_{i*} s_{\tau-t} \quad \text{(by Lemma 2)} \\
&= \operatorname*{plim}_{\tau\to\infty} \boldsymbol{W}_{i*} \frac{1}{\tau} \sum_{t=\bar{\tau}+1}^{\tau} s_{\tau-t} \\
&\overset{\text{asy}}{=} \operatorname*{plim}_{\tau\to\infty} \boldsymbol{W}_{i*} \frac{1}{\tau - \bar{\tau}} \sum_{t=\bar{\tau}+1}^{\tau} s_{\tau-t} \\
&\overset{\text{asy}}{=} \boldsymbol{W}_{i*} \boldsymbol{\theta}^* = \theta^*, \qquad \text{(i.i.d. Bernoulli signals)} \qquad (12)
\end{aligned}$$

where $\boldsymbol{W} = \boldsymbol{D}\mathbf{1'}$. From equation (12), we conclude that society is wise and because of that, $\operatorname{plim}_{t\to\infty} |\tilde{y}_{k,t} - \tilde{y}_{l,t}| = 0$, i.e. the K groups reach consensus, impliying $\operatorname{plim}_{t\to\infty} P_t = |\theta^* - \theta^*| = 0$. **(Q.E.D.)**

B Proof of Proposition 2

Consider again the update process described in the Section (XX)

$$\begin{aligned}\alpha_{t+1} &= B_t(\alpha_t + s_{t+1}) + (\mathbb{I}_n - B_t)\hat{g}_t\alpha_t \\ &= \left[B_t + (\mathbb{I}_n - B_t)\hat{g}_t\right]\alpha_t + B_t s_{t+1}.\end{aligned}$$

Notice that B_t is not fixed over time now. We re-write the stochastic matrix (see lemma 1) inside the squared bracket as

$$\alpha_{t+1} = W_t\alpha_t + B_t s_{t+1}.$$

When $t = 0$,

$$\alpha_1 = W_0\alpha_0 + B_0 s_1.$$

When $t = 1$,

$$\begin{aligned}\alpha_2 &= W_1\alpha_1 + B_1 s_2 \\ &= W_1\left(W_0\alpha_0 + B_0 s_1\right) + B_1 s_2 \\ &= W_1 W_0\alpha_0 + W_1 B_0 s_1 + B_1 s_2.\end{aligned}$$

When $t = 2$,

$$\begin{aligned}\alpha_3 &= W_2\alpha_2 + B_2 s_3 \\ &= W_2(W_1 W_0\alpha_0 + W_1 B_0 s_1 + B_1 s_2) + B_2 s_3 \\ &= W_2 W_1 W_0\alpha_0 + W_2 W_1 B_0 s_1 + W_2 B_1 s_2 + B_2 s_3.\end{aligned}$$

So on and so forth and similarly for the parameter vector β.

Following Chaterjee and Seneta (1977), Seneta (2006) and Tahbaz-Salehi and Jadbabaie (2008), we let $\{W_k\}$, for $k \geq 0$, be a fixed sequence of stochastic matrices (see lemma 1), and let $U_{r,k}$ be the stochastic matrix defined by the following *backward product*

$$U_{r,k} = W_{r+k} \cdot W_{r+(k-1)} \ldots W_{r+2} W_{r+1} W_r, \tag{13}$$

where $W_k = \{w_{ij}(k)\}$, $U_{r,k} = \{u_{ij}^{(r,k)}\}$[10].

[10]Our backward product has last term equals to W_r, rather than W_{r+1}. This is because our *first* period is 0, rather than 1. This notation comes without costs or loss of generality.

Then, the update process of both parameters can be represented in the following form for any period τ

$$\alpha_\tau = U_{0,\tau-1}\alpha_0 + \left(\sum_{r=1}^{\tau-1} U_{r,\tau-1-r}B_{r-1}s_r\right) + B_{\tau-1}s_\tau \tag{14}$$

$$\beta_\tau = U_{0,\tau-1}\beta_0 + \left(\sum_{r=1}^{\tau-1} U_{r,\tau-1-r}B_{r-1}\left(\mathbf{1}-s_r\right)\right) + B_{\tau-1}\left(\mathbf{1}-s_\tau\right) \tag{15}$$

From equation (14), we write its entries as

$$\begin{aligned}
\alpha_{i,\tau} &= \sum_j u_{ij}^{(0,\tau-1)}\alpha_{j,0} + \left(\sum_j \sum_{r=1}^{\tau-1} u_{ij}^{(r,(\tau-1-r))} b_{j,r-1}s_{j,r}\right) + b_{i,\tau-1}s_{i,\tau} \\
&= \sum_j u_{ij}^{(0,\tau-1)}\alpha_{j,0} + \tau\frac{1}{\tau}\left[\left(\sum_j \sum_{r=1}^{\tau-1} u_{ij}^{(r,(\tau-1-r))} b_{j,r-1}s_{j,r}\right) + b_{i,\tau-1}s_{i,\tau}\right] \\
&= \sum_j u_{ij}^{(0,\tau-1)}\alpha_{j,0} + \tau\tilde{\theta}_{i,1}(\tau)
\end{aligned} \tag{16}$$

Each entry of the parameter vector β is written in a similar way

$$\beta_{i,\tau} = \sum_j u_{ij}^{(0,\tau-1)}\beta_{j,0} + \left(\sum_j \sum_{r=1}^{\tau-1} u_{ij}^{(r,(\tau-1-r))} b_{j,r-1}(1-s_{j,r})\right) + b_{i,\tau-1}(1-s_{i,\tau}).$$

The sum of both parameters $\alpha_{i,\tau}$ and $\beta_{i,\tau}$ yields

$$\begin{aligned}
\alpha_{i,\tau}+\beta_{i,\tau} &= \sum_j u_{ij}^{(0,\tau-1)}\left(\alpha_{j,0}+\beta_{j,0}\right) + \left(\sum_j \sum_{r=1}^{\tau-1} u_{ij}^{(r,(\tau-1-r))} b_{j,r-1}\right) + b_{i,\tau-1} \\
&= \sum_j u_{ij}^{(0,\tau-1)}\left(\alpha_{j,0}+\beta_{j,0}\right) + \tau\frac{1}{\tau}\left[\left(\sum_j \sum_{r=1}^{\tau-1} u_{ij}^{(r,(\tau-1-r))} b_{j,r-1}\right) + b_{i,\tau-1}\right] \\
&= \sum_j u_{ij}^{(0,\tau-1)}\left(\alpha_{j,0}+\beta_{j,0}\right) + \tau\tilde{\theta}_{i,2}(\tau)
\end{aligned} \tag{17}$$

In which $\sum_j u_{ij}^{(r,(\tau-1))} = 1$, for all $r \geq 0$ since $U_{r,k}$ is a stochastic matrix. Therefore, the opinion of each agent i in this society, at some particular time τ, is $y_{i,\tau} = \frac{\alpha_{i,\tau}}{\alpha_{i,\tau}+\beta_{i,\tau}}$, where each entry of the parameter vectors can be written as follows:

$$y_{i,\tau} = \frac{\sum_j u_{ij}^{(0,\tau-1)}\alpha_{j,0} + \tau\tilde{\theta}_{i,1}(\tau)}{\sum_j u_{ij}^{(0,\tau-1)}\left(\alpha_{j,0}+\beta_{j,0}\right) + \tau\tilde{\theta}_{i,2}(\tau)}$$

Asymptotically we have:

$$\begin{aligned}
\operatorname*{plim}_{\tau\to\infty} y_{i,\tau} &= \operatorname*{plim}_{\tau\to\infty}\left(\frac{\sum_j u_{ij}^{(0,\tau-1)}\alpha_{j,0}+\tau\tilde{\theta}_{i,1}(\tau)}{\sum_j u_{ij}^{(0,\tau-1)}(\alpha_{j,0}+\beta_{j,0})+\tau\tilde{\theta}_{i,2}(\tau)}\right)\\
&= \operatorname*{plim}_{\tau\to\infty}\frac{\tau}{\tau}\left(\frac{\frac{\sum_j u_{ij}^{(0,\tau-1)}\alpha_{j,0}}{\tau}+\tilde{\theta}_{i,1}(\tau)}{\frac{\sum_j u_{ij}^{(0,\tau-1)}(\alpha_{j,0}+\beta_{j,0})}{\tau}+\tilde{\theta}_{i,2}(\tau)}\right)\\
&= \operatorname*{plim}_{\tau\to\infty}\frac{\tilde{\theta}_{i,1}(\tau)}{\tilde{\theta}_{i,2}(\tau)} \qquad (18)
\end{aligned}$$

Our main concern in order to prove that equation (18) converges in probability to θ^* is the behavior of $U_{r,k}$ when $k \to \infty$ for each $r \geq 0$. For that, we need to define two concepts of ergodicity. The sequence $\{W_k\}$ is said to be *weakly ergodic*, as $k \to \infty$, if for all $i, j, s = 1, 2, \dots, n$ and $r \geq 0$

$$|u_{i,s}^{(r,k)} - u_{j,s}^{(r,k)}| \to 0$$

On the other hand, we say that this very same sequence is *strongly ergodic* for all $r \geq 0$, and elementwise, if:

$$\lim_{k\to\infty} U_{r,k} = \mathbf{1}D_r'$$

Where $\mathbf{1}$ is a size n vector of ones and D_r is a probability vector in which $D_r \geq 0$ and $D_r'\mathbf{1} = \mathbf{1}$.

Both weak and strong ergodicity describe a tendency to consensus. In the strong ergodicity case, all rows of the stochastic matrix $U_{r,k}$ are becoming the same as k grows large and reaching a stable limiting vector, whereas in the weak ergodicity case, every row is converging to the same vector, but each entry not necessarily converges to a limit.

The three following lemmas are auxiliary helps to conclude the proof. Lemma 1 do xxx, Lemma 2 do yyy, whereas Lemma 3 do zzz.

Lemma 3. *For the backward product* (13)*, weak and strong ergodicity are equivalent.*

Proof. Following Seneta (1977)'s Theorem 1, we only need to prove that weak ergodicity implies strong ergodicity. Fix $r \geq 0$ and $\epsilon > 0$. Then, by *weak* ergodicity, we have

$$-\epsilon \leq u_{i,s}^{(r,k)} - u_{j,s}^{(r,k)} \leq \epsilon \Longleftrightarrow u_{i,s}^{(r,k)} - \epsilon \leq u_{j,s}^{(r,k)} \leq u_{i,s}^{(r,k)} + \epsilon$$

for $k \geq \bar{\tau}$ for all $i, h, s = 1, \dots, n$. Since $U_{r,k+1} = W_{r+k+1}U_{r,k}$,

$$\sum_{j=1}^{n} w_{hj}(r+k+1)(u_{i,s}^{(r,k)} - \epsilon) \leq \sum_{j=1}^{n} w_{hj}(r+k+1)u_{j,s}^{(r,k)} \leq \sum_{j=1}^{n} w_{hj}(r+k+1)(u_{i,s}^{(r,k)} + \epsilon).$$

The inequality above shows that for any h and $k \geq \bar{\tau}$

$$u_{i,s}^{(r,k)} - \epsilon \leq u_{h,s}^{(r,k)} \leq u_{i,s}^{(r,k)} + \epsilon.$$

Thus, by induction, for any $i, h, s = 1, 2, \ldots, n$, for any $k \geq \bar{\tau}$ and for any integer $q \geq 1$

$$|u_{h,s}^{(r,k+q)} - u_{i,s}^{(r,k)}| \leq \epsilon.$$

By setting $i = h$, it is clear that $u_{i,s}^{k,r}$ is a Cauchy sequence that aproaches a limit as $k \to \infty$. □

Definition 3. *The scalar function $\mu(\cdot)$ continuous on the set of $n \times n$ stochastic matrices W and satisfying $0 \leq \mu(W) \leq 1$ is called a coefficient of ergodicity. It is said to be proper if $\mu(W) = 1 \Leftrightarrow W = \mathbf{1}\boldsymbol{v}'$, where $\boldsymbol{v}'$ is any probability vector (i.e. whenever W is a row-stochastic matrix with unity rank).*

In particular, we will focus on the *proper* coefficient of ergodicity $\mu(W) = 1 - a(W)$, where

$$a(W) = \frac{1}{2} \max_{i,j} \sum_{s=1}^{n} |w_{is} - w_{js}|.$$

Therefore, weak ergodicity is then equivalent to $\mu(U_{r,k}) \to 1$ as $k \to \infty$ and $r \geq 0$.

Lemma 4. *Suppose that $1 - a(\cdot)$ and $\mu(\cdot)$ are both proper coefficients of ergodicity. Then $\{W_k\}$, $k \geq 0$, is ergodic if and only if there exists a strictly increasing subsequence $\{i_j\}$, $j = 1, 2, \ldots$ of the positive integers such that*

$$\sum_{j=1}^{\infty} \mu\left(U_{i_j, i_{j+1} - i_j}\right) = \infty$$

Proof. Soon □

Lemma 5. *The weak ergodicity of the sequence $\{W_k\}$, $k \geq 0$ is a trivial event when $\mathbf{g}^t$ follows equation* (1).

Proof. Soon □

With the results of these three lemmas, we can proceed with

$$
\begin{aligned}
\operatorname*{plim}_{\tau\to\infty} \frac{\tilde{\theta}_{i,1}(\tau)}{\tilde{\theta}_{i,2}(\tau)} &= \operatorname*{plim}_{\tau\to\infty} \frac{\frac{1}{\tau}\sum_j \sum_{r=1}^{\tau-\bar{\tau}} u_{ij}^{(r,(\tau-1-r))} b_{j,r-1} s_{j,r}}{\frac{1}{\tau}\sum_j \sum_{r=1}^{\tau-\bar{\tau}} u_{ij}^{(r,(\tau-1-r))} b_{j,r-1}} \qquad \text{(by lemma 5)} \\
&= \operatorname*{plim}_{\tau\to\infty} \frac{\sum_j \bar{u}_{ij} \frac{1}{\tau} \sum_{r=1}^{\tau-\bar{\tau}} b_{j,r-1} s_{j,r}}{\sum_j \bar{u}_{ij} \frac{1}{\tau} \sum_{r=1}^{\tau-\bar{\tau}} b_{j,r-1}} \qquad \text{(by lemma 4)} \\
&\overset{\text{asy}}{\equiv} \operatorname*{plim}_{\tau\to\infty} \frac{\sum_j \bar{u}_{ij} \frac{1}{\tau-\bar{\tau}} \sum_{r=1}^{\tau-\bar{\tau}} b_{j,r-1} s_{j,r}}{\sum_j \bar{u}_{ij} \frac{1}{\tau-\bar{\tau}} \sum_{r=1}^{\tau-\bar{\tau}} b_{j,r-1}} \\
&= \frac{\sum_j \bar{u}_{ij} \mathbb{E}(b_j s_j)}{\sum_j \bar{u}_{ij} \mathbb{E}(b_j)} \qquad \text{(by weak law of large numbers)} \\
&= \frac{\sum_j \bar{u}_{ij} \mathbb{E}(b_j) \mathbb{E}(s_j)}{\sum_j \bar{u}_{ij} \mathbb{E}(b_j)} \qquad \text{(by independence of } b_j \text{ and } s_j\text{)} \\
&= \frac{\theta^* \sum_j \bar{u}_{ij} \mathbb{E}(b_j)}{\sum_j \bar{u}_{ij} \mathbb{E}(b_j)} = \theta^* \qquad \text{(since } \mathbb{E}(s_j) = \theta^*, \forall j \text{)}
\end{aligned} \tag{19}
$$

C Regression analysis: alternative specifications

Table 2: Linear regression results: Interaction terms in Table 1 evaluated at the average level of PageRank

	Dependent Variable: Average Polarization [1000,2000]			
	2 partisans (1:1)	4 partisans (2:2)	6 partisans (3:3)	8 partisans (4:4)
	(1)	(2)	(3)	(4)
Bayes λ	-0.019^{***}	-0.021^{***}	-0.037^{***}	-0.029^{***}
	(0.005)	(0.004)	(0.004)	(0.004)
Clock p	-0.047^{***}	-0.042^{***}	-0.052^{***}	-0.047^{***}
	(0.005)	(0.004)	(0.004)	(0.004)
Audience of Both (% of nodes)	0.059^{***}	0.031^{***}	0.110^{***}	0.184^{***}
	(0.016)	(0.010)	(0.009)	(0.009)
Audience of L (% of nodes)	0.024^{**}	0.084^{***}	0.192^{***}	0.272^{***}
	(0.010)	(0.011)	(0.011)	(0.012)
Audience of R (% of nodes)	0.026^{**}	0.083^{***}	0.192^{***}	0.272^{***}
	(0.011)	(0.010)	(0.011)	(0.012)
Initial Homophily	-0.013	-0.029	0.013	-0.015
	(0.029)	(0.023)	(0.020)	(0.017)
Average Cluster	0.203^{***}	0.206^{***}	0.156^{***}	0.080^{***}
	(0.011)	(0.011)	(0.013)	(0.012)
Reciprocity	-0.147^{***}	-0.058^{*}	-0.023	-0.039
	(0.035)	(0.031)	(0.033)	(0.034)
Observations	575	850	868	824
Adjusted R^2	0.817	0.890	0.928	0.957
Residual Std. Error	0.037 (df = 564)	0.036 (df = 839)	0.034 (df = 857)	0.030 (df = 813)
F Statistic	233.702^{***} (df = 11; 564)	625.679^{***} (df = 11; 839)	$1{,}016.950^{***}$ (df = 11; 857)	$1{,}661.630^{***}$ (df = 11; 813)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$