Rumors and Social Networks

Francis Bloch, Gabrielle Demange, Rachel Kranton

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2014 – 15

Rumors and Social Networks

Francis Bloch
Gabrielle Demange
Rachel Kranton

JEL Codes: C72, D83

Keywords: Bayesian updating, rumors, misinformation, social networks

PARIS-JOURDAN SCIENCES ECONOMIQUES
48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES
ÉCOLE DES PONTS PARISTECH – ECOLE NORMALE SUPÉRIEURE – INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQU

Rumors and Social Networks

Francis Bloch, Gabrielle Demange, Rachel Kranton *

March 25, 2014

Abstract: Why do people spread rumors? This paper studies the transmission of possibly false information—by rational agents who seek the truth. Unbiased agents earn payoffs when a collective decision is correct in that it matches the true state of the world, which is initially unknown. One agent learns the underlying state and chooses whether to send a true or false message to her friends and neighbors who then decide whether or not to transmit it further. The papers hows how a social network can serve as a filter. Agents block messages from parts of the network that contain many biased agents; the messages that circulate may be incorrect but sufficiently informative as to the correct decision.

Keywords: Bayesian updating, rumors, misinformation, social networks.

JEL Classification: C72, D83.

*Francis Bloch: Paris School of Economics-Paris I, Gabrielle Demange: Paris School of Economics-EHESS, Rachel Kranton: Duke.University. We thank Aaron Kolb and Margaux Luflade for invaluable research assistance, and seminar participants at various universities and conferences for comments. Gabrielle Demange is supported by the grant NET from ANR. Rachel Kranton thanks Chaire Blaise Pascal/Paris School of Economics and the National Science Foundation for support.

I Introduction

Why do people spread rumors? Rumors are opinions spread from person to person with no discernible source.[1] In a recent book Cass Sunstein (2009) documents the pervasiveness of rumors, their public benefits, and their perils. Rumors abound concerning the efficacy of vaccines, the birthplace of presidential candidates, the propriety of politicians, the fabrication of data in academic research, and the impact of fracking on the water table. This paper studies why rumors are spread—by rational agents who seek the truth.

In a simple model people communicate to neighbors and friends. Agents' individual payoffs depend on a collective decision, such as election of a candidate or authorizing the use of new technology. Collective-decision making is modeled as a stylized "vote" that reflects each agent's expected utility from the decision. Some agents are unbiased and prefer that the decision correctly matches the true state of the world. Other agents are biased and prefer a particular decision regardless of the true state. (Such agents might personally benefit, say, from the decision.) Agents have prior beliefs as to the true state. One agent, selected random, receives precise information about the true state. This agent, whose identity is not known, can create a message, a rumor, to send to her friends about the state; the message may or may not convey the true state, and biased agents have the incentive to create a false message. Agents who receive a message decide whether or not to pass it along. Agents strategically spread the message, in order to influence how others will vote on the collective outcome.

The paper derives network conditions for a *full communication equilibrium*, where all unbiased agents transmit messages and, therefore, spread possibly false rumors. They do so because there is a sufficiently large probability the rumor is true. The equilibrium conditions rely on the distribution of biased and unbiased agents in the network. For any agent, the set of possible senders of a message must contain sufficiently few biased agents.

When this condition fails so that full communication is not possible, there is an equilibrium in which communication is maximized. We construct an algorithm (which runs in finite time) that precisely identifies subgraphs of the network where communication takes place. A main feature of this equilibrium is that information can flow from one part of the network to another but not in the reverse direction. Unbiased agents maintain the credibility of messages by blocking those

[1] Webster's English dictionary definition.

that come from a part of the network that contains too many biased agents. This same agent, however, will transmit messages coming from another direction. These *maximal equilibria* yield the highest expected payoffs of all perfect Bayesian equilibria of the game.

We have two main economic insights. First, networks can serve as a filter and aid communication. We contrast the network outcomes to a situation where agents can communicate to everyone simultaneously. In this *public broadcast model*, there are only two equilibrium outcomes: one with full communication and one with no communication. Full communication arises if and only if there are sufficiently few biased agents in the population. The network can replicate the full communication outcome when biased agents are evenly distributed in the network. While all biased agents send only messages that match their bias, there are enough unbiased agents sending truthful messages that agents are willing to transmit to their neighbors. The network, however, can allow partial communication when no communication is the only outcome in the public broadcast model. In a network, agents can block messages that originate in parts of the network that contain many biased agents. The messages that do circulate contain sufficient information for agents to take them into account when voting on the collective decision.

Second, biased agents wishing to influence a population could be better off limiting their numbers. As unbiased agents are strategic, they block the transmission of opinions that originate in a part of the network that contains many biased agents. Hence, it can serve biased agents to limit their numbers and to spread themselves throughout the network, so as to maximize the transmission of messages between agents.

Relative to previous literature, the innovation of this paper is to study the strategic decision to create and transmit rumors in order to influence general opinions. In a large literature, agents somewhat mechanically adopt the opinions of their neighbors and eventually the population converges on a set of beliefs, which could be unduly influenced by a set of well-located biased agents. In some models, opinions spread like diseases; i.e., individuals become infected (adopt an opinion) by contact with another agent with that disease (see e.g. Chapter 7 of Jackson (2008)). Such diffusion processes are being studied also in computer science, physics, and sociology. For a review article in physics, see for example, Castellano, Fortunato & Loreto (2009) For complex contagion where agents need multiple exposure to become infected see Centola & Macy (2007) and Romero, Meeder & Kleinberg (2011). In such models, biased agents are always better off when

there are more biased agents, in contrast with the present paper. Another strand of literature of opinion formation in social networks builds on DeGroot (1974) model of beliefs' exchange. Agents, with possibly different initial priors, repeatedly exchange their beliefs with their neighbors and adopt some statistic (the weighted average, say) of their neighbors' opinions. Such agents fail to take into account the repetition of information that can propagate through a network, leading to a persuasion bias as referred to by DeMarzo, Vayanos, & Zweibel (2003). Golub & Jackson (2010) find sufficient network conditions under which such a naive rule leads to convergence to the truth—there can be no prominent groups, for example, that have disproportionate influence. Research on Bayesian learning in networks (e.g. Gale & Kariv (2003), Bala & Goyal (1998), Acemoglu, Dahleh, Lobel & Ozdaglar (2011)) characterizes convergence or not to common opinions for different network architectures. In our model, there is a single unknown source of information and agents are bayesian, but due to differences in their preferences and the possibility of falsification and blocking, they may end up with different beliefs and choose different actions.

A large economic literature also studies the transmission and communication of information through the observation of other agents' actions. Observation helps them to discern the true state of the world Knowledge or information costlessly spreads (Banerjee (1992), Bikhchandani, Hirshleifer & Welch (1992), or spills over, to others, as occurs when people observe others' use of a new technology (e.g., Foster & Rosensweig (1995), Conley & Udry (2010)). In these models, though individuals influence others through their actions, they derive no benefit in influencing them and, contrary to this paper, the decision to communicate is not strategic.

A new literature studies the incentive to communicate private information to others. In a recent advance, Niehaus (2011) adds a cost to sharing information; an agent will weigh the benefits to her friends and neighbors against the personal cost. Other papers analyze influence in networks where agents all have private information and have an since, for example, agents derive a benefit from adopting the same action as others (Calvó-Armengol, de Martí & Prat (2011), Hagenbach & Koessler (2011), Galeotti, Ghiglino, and Squintani (2013)). In contrast to this work, the present paper features a situation in which information is not disseminated and strategic agents may possibly falsify information with the desire to influence public opinion.

In its foundation, the model in this paper combines two classic elements of information games: "cheap talk" (Crawford & Sobel (1982)) in the decision of the initial receiver of the signal as

to whether or not to create a truthful message, and "persuasion" (Milgrom (1981), Milgrom & Roberts (1986)) in the decision of agents who subsequently choose whether to transmit the message, which they cannot transform. We draw on insights from both in the analysis. On one hand, it is well known that cheap talk games have multiple equilibria (e.g., babbling, fully revealing, and mixed). On the other hand, in persuasion games, agents send truthful (verifiable) information to individuals with similar preferences. In our model, there are multiple equilibria, along the lines of cheap talk games. However, as in persuasion games, at the transmission stage in the present model, agents have an incentive to pass on credible information to other agents. Our analysis features these information game elements in a network setting, and the network plays a primary role in the outcomes. The analysis focuses on the network conditions that allow fully revealing strategies by unbiased agents and identifies the paths in a network along which agents are willing to listen to messages and persuade others.

The rest of the paper is organized as follows. Section II specifies the two benchmark models of communication: public broadcast and network. Section III studies full communication equilibria in both settings, where all unbiased agents create truthful messages. Section IV studies maximal communication equilibria in networks, building the algorithm that yields the maximal paths along which unbiased agents are willing to transmit messages. Section V studies, from the point of view of biased agents, the tradeoffs between more or less biased agents in the population. Section VI considers extensions to the basic network. Section VII concludes.

II Benchmark Models of (Possibly Biased) Communication

A Utility and Agents' Types

There is a population of $|N| = n$ agents, and two possible states of nature, $\theta \in \{0, 1\}$. Individual agents earn payoffs from a collective decision, or outcome, which can be understood, for example, as a public policy, a verdict, or election of a particular candidate. Let $x \in \{0, 1\}$ denote the outcome. There are two types of agents, with different preferences. *Unbiased agents*, set $\mathcal{U}$, prefer the outcome to match the state of nature and have utility

$$w(x, \theta) = -(x - \theta)^2.$$

Biased agents, set $\mathcal{B}$, prefer outcome $x = 1$ to be implemented, regardless of the state of nature. The utility for a biased agent is

$$v(x, \theta) = -(x - 1)^2.$$

The number of biased and unbiased agents in the population is common knowledge. For any subset of agents S, b_S denotes the fraction of biased agents in S and u_S the fraction of unbiased agents, where necessarily $b_S + u_S = 1$. For any unbiased individual, let $b \equiv \frac{|\mathcal{B}|}{|N|-1}$denote the fraction of biased agents in the remainder of the population.

B Prior Beliefs, Signals, and Communication

Agents have a common prior belief that $\theta = 1$ with probability π. This common prior is common knowledge. We assume $\pi < 1/2$ so that agents initially believe the true state is 0 with higher probability. With probability $p < 1$, one agent is randomly selected and receives a perfect signal $s \in \{0, 1\}$ of the state of nature. This agent – and this agent only – has the opportunity to create a message $m \in \{0, 1\}$.

We consider two benchmark models of communication.

The *public broadcast model* represents an environment where agents are anonymous and, while the number of biased and unbiased agents are known, individual agents' types are private information. The agent who receives the signal can send a message to the public at large; i.e., the message simultaneously and anonymously reaches all other agents. Formally, the agent who receives the signal chooses an action $M(s) \in \{\varnothing, 0, 1\}$, where $M(s) = \varnothing$ denotes that the agent chooses not to create any message.

The *network model* represents an environment where agents communicate with friends, family, colleagues, etc. When a pair of agents i and j have a link, denoted ij, they can communicate, and we say agent i is agent j's *neighbor* and vice versa. To distinguish the direction of communication (i, j) denotes the directed link from i to j, and (j, i) the directed link from j to i, and G denotes the set of all directed links. We assume G and individual agents' types are common knowledge.[2]

Communication in a network proceeds as follows: The agent who (possibly) receives the signal s chooses a message $M(s) \in \{\varnothing, 0, 1\}$. Subsequently, agents who receive a message m

[2]We discuss extensions of the model where agents have incomplete information about the network in Section VII.

cannot transform it but they can choose whether or not to transmit the message to all their neighbors,[3] i.e., agent i who receives a message m from neighbor j, denoted $m(j)$, chooses an action $t_i(m(j)) \in \{\varnothing, m(j)\}$. Notice that for any strategies, the event $\varnothing$ occurs with positive probability because every agent could receive no message since no signal is received with probability $1-p > 0$.

We suppose throughout the paper that agents are connected in such a way that a message can reach any individual through only one route. That is, the network is a *tree*, where there is a unique path from any agent a to any agent b. With a tree, we can neatly parse the network and study agents' posterior beliefs as to the veracity of a received message. Section VII discusses general networks.

C Collective Outcome

We abstract from time and use a reduced-form decision-making process to allow us to focus on agents' incentives to create and transmit possibly false messages. Suppose after all possible communication is exhausted, agents each "vote" for an outcome, and the more agents who vote for an outcome, the more likely it is to be implemented. When z agents vote for outcome 1, let $f(z)$ be the probability that outcome 1 is implemented, with $1 - f(z)$ the probability that outcome 0 is implemented. We will assume here probabilistic voting: $f(z) = z/n$, to simplify the analysis as it precludes strategic voting (Lemma 1).

Agents vote for the outcome that maximizes their expected utility. Biased agents always vote for $x = 1$, but unbiased agents vote given their posterior beliefs about the true state of nature. Let ρ_i denote agent i's posterior belief that $\theta = 1$. Given z other agents vote for $x = 1$, an unbiased agent's expected utility from voting for $x = 1$ is

$$Ew(x, \theta) = -\rho_i \left(1 - f(z+1)\right) - \left(1 - \rho_i\right) f(z+1).$$

The expected utility from voting for $x = 0$ is

$$Ew(x, \theta) = -\rho_i \left(1 - f(z)\right) - \left(1 - \rho_i\right) f(z).$$

[3] All messages and transmission are assumed to be multi-cast; agents send/transmit messages to either none or all of their neighbors. As we will see, this assumption is made without loss of generality in our baseline model where the underlying network is a tree.

Lemma 1 *With probabilistic voting, $f(z) = \frac{z}{n}$, it is optimal for unbiased agents to vote according to their beliefs: An unbiased agent votes for outcome $x = 1$ if $\rho_i > 1/2$, and votes for outcome 0 if $\rho_i < \frac{1}{2}$. If $\rho_i = \frac{1}{2}$, we assume agent i votes for 0 and 1 with equal probability*

Similarly, if an unbiased agent can influence the beliefs of other unbiased agents in order to make them vote according to her beliefs by creating or transmitting a message, she has an incentive to do so.

The same behavior holds under more general increasing functions f if one assumes agents to be 'naive', meaning that they do not account for the possible correlation between others' vote and their information on the state: for $\rho_i > 1/2$ ($\rho_i < 1/2$) agent i's utility is larger when he votes for 1 instead of 0 (0 instead of 1) for a fixed z, hence also for any distribution of z provided the correlation between this distribution and the true state is neglected.[4]

D Equilibrium Concept and Maximal Equilibria in Networks

We consider pure-strategy perfect Bayesian equilibria (henceforth simply equilibria) in each benchmark model.

Public broadcast model: An equilibrium consists of message creation strategies M_i and posterior beliefs ρ_i for each agent i such that each agent's strategy is sequentially rational given the beliefs and strategies of others, and beliefs are formed using Bayes rule from the strategies whenever possible.

Network model: A network equilibrium consists of message creation strategies, transmission strategies, and beliefs (M_i, t_i, ρ_i) for each agent i such that each agent's strategy is sequentially rational given the beliefs and strategies of others, and beliefs are formed using Bayes rule from the strategies whenever possible. In the analysis below, we make precise the strategies for every possible history of play and beliefs at every information set. In a network model, let η denote a collection of strategies and beliefs that constitute an equilibrium.

In both communication games, there are possibly many equilibria where agents do not communicate or communication does not contain any information. As in cheap talk games, there exist babbling equilibria, where messages do not contain any information about the true state

[4]Such correlation matter in situations such as common values. For example, this correlation is the basis of the winner's curse in auctions or of the strategic behavior of a pivotal voter in the Condorcet jury.

and thus no agents update their priors. Furthermore, even when unbiased agents choose revealing strategies when they create messages, there exist equilibria where communication fails at the transmission stage.[5] Because of the presence of biased agents, there do not exist equilibria where all messages reflect the true state of nature.[6]

Our main interest is equilibria in which unbiased agents create truthful messages, and, in the network case, transmission of messages is the highest possible. Since biased agents always create message $m = 1$, we are interested in equilibria in which unbiased agents are always willing to transmit such messages. To compare communication in networks with public broadcasting, We first study *full communication*. All unbiased agents create truthful messages, and, in a network, all unbiased agents transmit all messages from all their neighbors. When full communication is not possible we can characterize an (essentially unique) equilibrium where communication is maximal. Section IV defines and constructs this equilibrium. We also show that this equilibrium Pareto dominates all other equilibria for unbiased agents.

III Full Communication

A Full Communication: Public Broadcast

In the public broadcast game, consider the following strategies and beliefs. Strategies: any unbiased agent who receives the signal sends the message $m = s$. Any biased agent who receives the signal sends the message $m = 1$. Beliefs: Upon receiving a message $m = 0$, each unbiased agent i has posterior belief $\rho_i = 0$, since $m = 0$ can only originate from an unbiased agent sending a truthful message. Upon receiving a message $m = 1$, following Bayes' rule each unbiased agent has posterior belief

$$\rho_i = \frac{\pi}{b + (1-b)\pi}.$$

Upon receiving no message, each unbiased agent maintains her prior belief, $\rho_i = \pi$ (This event occurs in equilibrium when no signal is received which occurs with strictly positive probability). Following Lemma 1 when $\rho_i > 1/2$ ($\rho_i < 1/2$), agent i votes for outcome 1 (0).

It is easy to see that these strategies and beliefs constitute an equilibrium in this communica-

[5] See Section V for a discussion of multiplicity of equilibria in our game.

[6] In a truthful equilibrium, unbiased agents always believe that the state is 1 when they receive message 1. These beliefs give biased agents an incentive to create message 1 irrespective of the signal they receive.

tion structure. No unbiased agent that receives the signal has an incentive to deviate and choose $M(s) \neq s$, since, given agents' posterior beliefs, this will decrease the number of agents that vote for the outcome corresponding to the true state. No biased agent has an incentive to deviate and and choose $m = \oslash$ or $m = 0$, since these actions will decrease the number of agents that vote for outcome 1.

It is also easy to see that no other equilibrium can yield higher expected utility for unbiased agents and that no partial communication equilibria exist. Consider the possibility that in equilibrium a subset of unbiased agents send truthful messages but others do not. One of the latter unbiased agents would have an incentive to deviate and send a truthful message, since it will make the true outcome more likely to be implemented.

These arguments give us our first result concerning communication:

Proposition 1 *In a public broadcast game, an equilibrium exists where all unbiased agents broadcast truthful messages if and only if*

$$\frac{\pi}{1-\pi} \geq b. \tag{1}$$

This equilibrium maximizes unbiased agents' expected payoffs. It is an equilibrium for no unbiased agents to broadcast messages, but there is no equilibrium where a strict subset of unbiased agents broadcast truthful messages.

Proof. Proofs of all results are provided in the Appendix.

B Full Communication: Networks

We now consider full communication in a network. Consider the following strategies and beliefs:

Strategies: Upon receipt of the signal, biased agents create a message that matches their bias, i.e., $M(s) = 1$. Biased agents only transmit messages that match their bias, i.e., $t(0) = \varnothing$, $t(1) = 1$. Unbiased agents create true messages upon receiving a signal; i.e., $M(s) = s$, and transmit any message they receive, i.e., $t(m) = m$.

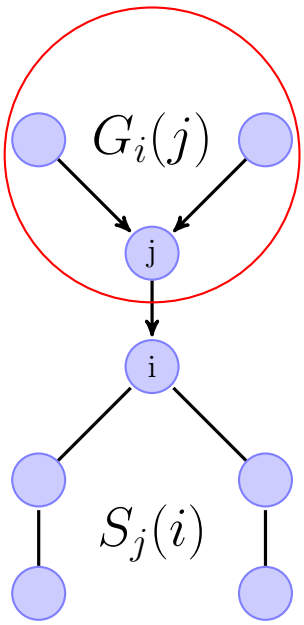


Figure 1: Decomposition of the tree

Beliefs: Along the equilibrium path, beliefs follow Bayes' rule. Consider an agent i who has received a message from a neighbor j. Let an agent i's belief that $\theta = 1$ be $\widetilde{\rho}_i(m(j))$. To construct these beliefs, consider the directed edge (j, i). Since the network is a tree, agents in the network can be divided into two disjoint subsets, with one subset on either side of the edge. Let $S_i(j)$ be the set of agents whose messages can reach i by going through j (this set includes j). The set $S_i(j)$ corresponds to the nodes in the oriented subgraph of G flowing toward i and ending with the directed edge (j, i); we denote this oriented subgraph $G_i(j)$. The other set $S_j(i)$ is the set of agents whose messages can reach j by going through i (this set includes i). $G_j(i)$ is defined similarly. Figure 1 illustrates $G_i(j)$.

Beliefs are as follows. Consider first information sets which can be reached using these strategies: (1) For an agent i who has received a message $m = 0$ from an unbiased neighbor j, $\widetilde{\rho}_i(0(j)) = 0$, since only unbiased agents create and transmit message 0, and they create truthful messages. (2) Messages $m = 1$, on the other hand, are created by both biased and unbiased agents. Following Bayes' rule, and our discussion above concerning the partition of the graph into disjoint subsets $S_i(j)$ and $S_j(i)$, an agent i who has received message $m = 1$ from j has the beliefs

$$\rho_i = \widetilde{\rho}_i(1(j)) = \frac{\pi}{b_{S_i(j)} + u_{S_i(j)}\pi}, \tag{2}$$

where recall $b_{S_i(j)}$ is the proportion of biased agents in $S_i(j)$ and $u_{S_i(j)}$ is the proportion of unbiased agents. (3) For an agent i who receives no message, her beliefs take into account the probability that no signal has been sent and the fact that biased agents block messages 0. Hence the posterior beliefs are surely smaller than π.

The only events with zero probability for which we need to specify beliefs are when an agent i receives a message zero from a biased agent.We suppose i's posterior belief is equal to his prior; $\rho_i(0(j)) = \pi$ for all $j \in \mathcal{B}$.

These strategies and beliefs constitute an equilibrium of the network game, depending on the location of biased agents in the network. In particular, an unbiased agent will only pass on a message $m(j) = 1$ when $\widetilde{\rho}_i(1(j)) \geq 1/2$; that is, the message could induce the agent to vote for outcome 1. This condition will be true for all unbiased agents i with neighbors j only when there are sufficiently few biased agents in all subgraphs of the network $S_i(j)$. We have the following result which is illustrated in Example 1.

Theorem 1 *In the network model, a full communication equilibrium (FCE) exists if and only if for each unbiased agent i and each of his neighbors j:*

$$b_{S_i(j)} \leq \frac{\pi}{1-\pi}. \tag{3}$$

Example 1 *Consider 8 agents in a line, as shown in Figure 2, with 7 unbiased agents and 1 biased agent, and the biased agent is 5th from the left. The equilibrium condition is then tightest for agent 4, since the subset $S_4(5)$ has the highest proportion of biased agents of all such subsets. In order for agent 4 to transmit messages to agent 3, π must satisfy $\frac{\pi}{1-\pi} \geq \frac{1}{4}$, which implies a bound of $\pi \geq \frac{1}{5}$. Thus, for $\frac{1}{2} > \pi \geq \frac{1}{5}$, there exists a FCE in this network.*

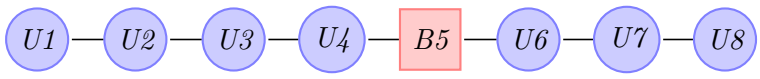


Figure 2: Eight Agent Line with One Biased Agent

C Full Communication: Public Broadcast vs. Network

Comparing public broadcast to a network, we see that full communication is possible in both structures, depending on the number and distribution of biased agents in the network. In public

broadcast, full communication exists for $\frac{\pi}{1-\pi} \geq b$. In the network, in contrast, biased agents must be dispersed so that no subset $S_i(j)$ violates the condition $\frac{\pi}{1-\pi} \geq b_{S_i(j)}$ for any unbiased agent i with neighbor j. Let $\max_{(j,i)} b_{S_i(j)}$ be the subgraph with the highest proportion of biased agents. Necessarily, $\max_{(j,i)} b_{S_i(j)} \geq b$. We then have the following result which is illustrated in Example 2.

Proposition 2

- *If* $\frac{\pi}{1-\pi} \geq \max_{(j,i)} b_{S_i(j)}$ *full communication is an equilibrium in both the public broadcast and the network models.*
- *If* $\max_{(j,i)} b_{S_i(j)} \geq \frac{\pi}{1-\pi} \geq b$*, full communication is an equilibrium in the public broadcast model, but not in a network.*
- *If* $b \geq \frac{\pi}{1-\pi}$*, no communication occurs in equilibrium in the public broadcast model.*

Example 2 *Consider a population of 8 agents with one biased agent. A FCE exists in the public broadcast model if and only if* $\pi \geq \frac{1}{8}$*. Consider again the network of 8 agents in Figure 2. For* $\frac{1}{5} \geq \pi \geq \frac{1}{8}$*, full communication is possible in the public broadcast setting but not in this network.*

The next section shows that communication is possible in a network when it is not possible in the public broadcast setting. In a network, agents can block messages that originate in parts of the network with higher concentrations of biased agents. The messages that do circulate, then, are sufficiently credible.

IV Maximal Communication Equilibria in Networks

In this section we construct strategies allowing for maximal communication among unbiased agents and prove that they form an equilibrium. Of course, these strategies coincide with those of the full communication equilibrium when it exists.

Here, we parse the network and construct an algorithm to find the subgraphs within which communication can occur. These subgraphs are directed and represent paths along which agents are willing to transmit messages, since the agents believe the message with sufficiently high probability. The algorithm eliminates directed edges from G, and we denote the remaining set of directed edges G^*. In the strategies constructed below, unbiased agents transmit messages from a neighbor j if and only if the directed link (j, i) is contained in G^*. We show that these

strategies maximize the possible communication in the graph at an equilibrium and yield the highest possible expected utility for unbiased agents.

A Algorithm to Identify Subgraphs of Transmission

When a full communication equilibrium does not exist, $b_{S_i(j)} > \frac{\pi}{1-\pi}$ for at least one unbiased agent i and directed edge (j,i) (Theorem 1). Consider the following algorithm in this case. Let V be the (non-empty) set of all directed edges in G that violate the condition $b_{S_i(j)} \leq \frac{\pi}{1-\pi}$. The algorithm will eliminate such *violating edges*. In the process, some violating edges may become non-violating and some edges in V will not be eliminated; on the other hand all non-violating edges will remain non-violating, so that V is the maximal set of edges that can be eliminated.

A directed edge $(j,i) \in V$ is said to be of *level 1* in G if there is no directed edge (k,l) such that $(k,l) \neq (j,i)$ in $V \cap G_i(j)$.[7] A directed edge $(j,i) \in V$ is a *level* ℓ edge in G if all violating directed edges in $G_i(j) \cap V$, distinct from (j,i) are of level less than ℓ.

Pick one level 1 edge $(j,i) \in V$. Remove (j,i) from G and let

$$\begin{aligned} G^1 &= G \setminus (j,i), \\ \Gamma^1 &= G \setminus G_i(j). \end{aligned}$$

For each unbiased agent l and directed edge (k,l) in Γ^1, let $S_l^1(k)$ be the set of agents whose message 1 can reach l through k in Γ^1. Compute the proportion $b_{S_l^1(k)}$ of biased agents in that set, $b_{S_l^1(k)} = \frac{|B \cap S_l^1(k)|}{|S_l^1(k)|}$, and define V^1 to be the set of directed edges in Γ^1 such that $b_{S_l^1(k)} \leq \frac{\pi}{1-\pi}$. If the set V^1 is empty, the algorithm stops.

Otherwise, pick a directed edge (k,l) of V^1 which is of level 1 in the graph Γ^1.[8] Remove (k,l) from G^1 to obtain G^2 and let $\Gamma^2 = \Gamma^1 \setminus G_l(k)$, and search for violating edges (if any) in Γ^2.

In the general step $t+1$ of the algorithm, given directed graphs G^t, Γ^t and a non-empty set V^t of violating edges in Γ^t, pick a level 1 violating edge (a,b). Eliminate this edge from G^t to obtain G^{t+1}. Define $\Gamma^{t+1} = \Gamma^t \setminus G_b(a)$ and accordingly the proportions $b_{S_b^{t+1}(a)}$ of biased agents in Γ^{t+1}. Let V^{t+1} be the set of edges (j,i) in G^{t+1} such that $b_{S_i^{t+1}(j)} > \frac{\pi}{1-\pi}$.

If the set V^{t+1} is empty, the algorithm stops and define $G^* = G^{t+1}$. Let W be the set of

[7] There is always at least one level 1 edge because the graph G is finite.

[8] Notice that the level of an edge in Γ^1 may differ from the level in G.

directed edges that have been eliminated by the algorithm, $G^* = G \setminus W$.

Lemmas in the Appendix show (i) that the sequence $\{V^t\}$ is decreasing, (ii) the set W does not depend on the order links are chosen to be eliminated and (iii) for any $(j, i) \in W$, j is biased (and i is unbiased by definition of a violating edge). That is, the algorithm parses the network into directed links of biased and unbiased agents. We illustrate the output of the algorithm in the following example:

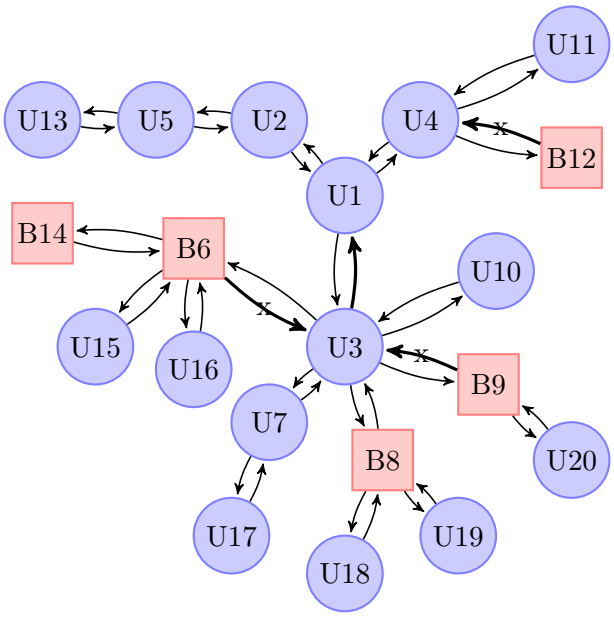


Figure 3: Complex Network and Algorithm

Example 3 *Subgraphs of Communication. The network in Figure 3 was generated by a random process for 20 agents, with an overall target fraction of 0.3 biased agents. Let the initial belief be* $\pi = \frac{2}{7}$. *Given* π, *the threshold for the proportion of biased agents is* $b = \frac{\pi}{1-\pi} = \frac{2}{5}$. *The edges in bold are the violating edges in* G. *They are all of level 1 except (U3-U1) which is of level 2. The edges which are crossed out are the edges in* W *which are eliminated by the algorithm.* G^* *does not contain the edges (B9-U3) (B12-U4) and (B6-U3), since messages flowing from these biased agents would not be believed. The edge (U3-U1) is violating in the original graph but not in the final graph because the proportion of biased agents whose messages would flow through U3 to the rest of the graph decreases from* $\frac{4}{9}$ *to* $\frac{2}{7}$.

B Maximal Communication Equilibrium Strategies and Beliefs

We next construct the strategies in which communication flows along all edges except those in W; i.e. along edges in G^*.

Consider the following strategies and beliefs.

Strategies: Biased agents, upon receipt of the signal, create a message that matches their bias, i.e., $M(s) = 1$. Biased agents only transmit messages that match their bias, i.e., $t(0) = \oslash$, $t(1) = 1$. Unbiased agents, upon receipt of a signal, create true messages; i.e., $M(s) = s$. All unbiased agents i transmit message 1 received from agent j if $(j, i) \in G^*$, otherwise agent i does not transmit the message. All unbiased agents transmit messages $m = 0$ received from any agent.

Beliefs: For information sets which are reached with positive probability, beliefs follow Bayes rule consistent with the above strategies: (1) For an agent i who has received a message $m = 0$ from an unbiased neighbor j, $\widetilde{\rho}_i(0(j)) = 0$, since only unbiased agents create and transmit $m = 0$, and they create truthful messages. (2) For an agent i who has received a message $m = 1$ from a neighbor j, her beliefs reflect the strategies of agents to only submit messages along edges in G^* and to not transmit messages otherwise. Posteriors are then given by $\widetilde{\rho}_i(1(j)) \geq 1/2$ for $(j, i) \in G^*$ and $\widetilde{\rho}_i(1(j)) < 1/2$ for $(j, i) \notin G^*$. (3) For an agent i who receives no message, her posterior beliefs take into account the probability that no signal has been received and the fact that biased and unbiased agents block messages that originate in particular parts of the network. These posteriors are surely less than π.[9]

The only event for which beliefs need to be specified is when an agent receives a message zero from a biased agent. As previously, we suppose i's posterior belief is equal to his prior in this case; i.e., $\rho_i(0(j)) = \pi$, for all $j \in \mathcal{B}$.

These strategies constitute an equilibrium of the network game. Furthermore, communication is maximal among all equilibria. The intuition as follows. If (i, j) is a violating edge of level 1, j is surely a biased agent and agent i never believes message 1 received from j[10] hence in no equilibrium communication flows from j to i. Inspection of the algorithm shows that all violating

[9] To see this, consider first the impact of an agent i who does not transmit a message 1 from j for $(j, i) \notin G^*$. Recall that j is biased. So, not only does i not transmit $m = 1$ received from j, but i never transmits $m = 0$ from j because j, being biased, does not create or transmit 0: all the signals received by an agent in $S_i(j)$, be them 0 or 1, are lost for the other agents, those in $N - S_i(j)$. As for biased agents, they block $m = 0$, which, by the strategies, is circulated only when the true state is 0. This implies that the posterior can only be lower than π: $\rho_i(\varnothing) < \frac{1}{2}$.

[10] Whatever behavior of the unbiased agents in $S_i(j)$, the posterior belief is not larger than $\frac{\pi}{b_{S_i(j)} + u_{S_i(j)}}$ which is the posterior belief when all unbiased agents transmit the message as seen from (2).

edges of level 1 in G are eliminated. Recursively, the edges in W—eliminated by the algorithm—are edges along which communication surely never flows in any equilibrium.

Theorem 2 *The above strategies and beliefs form an equilibrium of the network game. We call this equilibrium the "maximal communication equilibrium"(MCE) as communication is maximal among all equilibria in the following sense: in any equilibrium, if* $(j,i) \notin G^*$ *(equivalently* $(j,i) \in W$*), then* j *is biased and* i *does not transmit* $m = 1$ *received from* j*.*

C The network as a filter: public broadcasting vs. network communication

The above analysis shows that when full communication is not possible in the public broadcast model, communication is still possible in a network. In the network unbiased agents block messages from certain parts of the network, limiting the influence of localized biased agents. The network serves as a filter, allowing for credible communication. In particular, two unbiased agents always communicate between each other in a MCE since the corresponding edges are not in W. We have the following result.

Proposition 3 *If* $b > \frac{\pi}{1-\pi}$*, no communication is possible in the public broadcast game whereas partial communication exists in equilibrium in the network game. For any* $\pi > 0$*, as long as at least one unbiased agent is linked to another unbiased agent, there is an equilibrium with partial communication.*

V Multiplicity of equilibria and optimality property of an MCE

This section discusses other equilibria in our network model, relates the analysis to cheap talk and persuasion games, shows that the MCE is Pareto optimal for the unbiased agents and provides a refinement criterium, referred to as *activity* that distinguishes these equilibria.

First, as in cheap talk games, there are babbling equilibria in which no valuable information is created. Suppose each unbiased agent who has not received the signal takes the same action independent of any message received, and votes for 0 according to his prior. In this case, all unbiased agents are indifferent between all actions: creating, or not, true or false messages and transmitting, or not, messages. A simple equilibrium then consists of the following strategies:

Unbiased agents never create or transmit any messages, and biased agents always create $m = 1$ upon receipt of the signal, and transmit any $m = 1$, but no other message.[11] The only messages that are generated are those from the biased agents, and hence they are not informative. These strategies form an equilibrium supported by (consistent) posterior beliefs equal to the prior, except for the agent who has received the signal.

Second, there are equilibria that satisfy sequential rationality where unbiased agents create truthful messages but do not transmit all messages. These equilibria involve a coordination failure. The standard perfection argument which generates active transmission in a persuasion game does not hold in our model. Specifically, assuming that unbiased agents create truthful messages, one needs to consider only their behavior at a transmission stage. This stage is like a persuasion game since agents cannot falsify the message. However, because of the presence of biased agents, messages are not perfectly informative and it may be rational not to transmit message 1. This is illustrated by the following example.

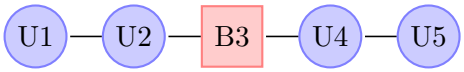


Figure 4: Five Agent Line with One Biased Agent

Example 4 *Consider 5 agents in a line, as shown in Figure 4 above with agents 1 and 2 unbiased, 3 biased, and 4 and 5 unbiased. Consider* $\pi \geq 1/4$*, in which case there is a full communication equilibrium. (The largest proportion of biased to unbiased agents in any subgraph* $S_i(j)$ *is* $1/3$*.) Change the strategies of the FCE as follows:* $U2$ *does not transmit message* $m = 1$ *received from* $U1$ *to* $B3$*;* $U4$ *does not transmit any message from* $B3$*. Note that all unbiased agents still create truthful messages. It is easy to check that these strategies form an equilibrium for* $\pi \leq 1/3$*: When* $U4$ *receives* $m = 1$ *from* $B3$*, the proportion of biased agents among the initiators is* $1/2$ *(instead of* $1/3$ *in the FCE), so the posterior on the true state being state 1 is lower than 1/2.* $U2$ *has no incentive to transmit* $m = 1$ *received from* $U1$ *since it will not influence the vote of* $U4$*, who maintains his prior upon receipt of any message from* $B3$*. Since for* $U4$ *receiving* $m = 1$ *from* $B3$ *is on the equilibrium path, a perturbation argument does not destabilize this equilibrium.*

[11] More formally consider the following strategies. For message creation, biased agents adopt the strategy $M(s) = 1$ and unbiased agents adopt the strategy $M(s) = \varnothing$. For transmission, biased agents adopt the strategy $t(m) = \varnothing$ for $m = 0$ and $t(m) = 1$ for $m = 1$. Unbiased agents have the strategy $t(m) = \varnothing$ for all m.

This equilibrium exhibits a coordination failure, and the equilibrium payoffs of unbiased agents are lower than the payoffs in an FCE. In the FCE, if a signal that the state is 1 is received by any agent, all agents receive a message $m = 1$ and all agents vote for 1. If a signal that the state is 0 is received by any unbiased agent, an unbiased agent either receives the signal, receives a message $m = 0$, or receives no message (as it is blocked by B3). Hence all unbiased agents vote for 0 and the biased agent votes for 1. The only mismatch between the circulated message and the state occurs when the state is 0 and a signal is received by B3; all agents vote for 1 in this case. The expected loss (for $p \to 1$) for unbiased agents is therefore $\frac{4}{5}(1-\pi)\frac{1}{5} + \frac{1}{5}(1-\pi)\frac{5}{5}$. In the above equilibrium with coordination failure, unbiased agents $U4$ and $U5$ do not update their priors in the events that a signal $s = 1$ is received by agents U1, U2, or B3. They vote for 0 in these events, and their votes do not match the state. On the other hand, $U4$ and $U5$ also do not change their prior in the event $s = 0$ is received by B3 (who then sends message $m = 1$). The expected loss (for $p \to 1$) for unbiased agents is therefore $\frac{3}{5}\pi\frac{2}{5} + \frac{4}{5}(1-\pi)\frac{1}{5} + \frac{1}{5}(1-\pi)\frac{3}{5}$. As $\pi \geq \frac{1}{4}$, the expected payoff of all unbiased agents is higher in the FCE than in the alternative equilibrium.

To refine the equilibrium set, consider restricting attention to the following simple strategies. A biased agent is *active* if and only if she creates message $M(s) = 1$ and only transmits message 1. An unbiased agent is *active* if and only if she creates a message that matches the signal and transmits message m if she thinks the probability that the true state is m is higher than $\frac{1}{2}$. This refinement allows us to single out the MCE.

Proposition 4 *The MCE is the only equilibrium where all agents are active.*

In an equilibrium where all agents are active, coordination failures are ruled out both at the message creation and transmission stages. This results in the highest expected payoff for the unbiased agents.

Theorem 3 *Among all equilibria, the MCE yields the highest expected payoffs for unbiased agents.*

It is straightforward to rank the utility of unbiased agents in the equilibria for each communication structure. By the arguments of Theorem 3, the expected utility of unbiased agents is higher in any full communication equilibrium than in any partial communication equilibrium and

higher in any partial communication equilibrium than in an equilibrium without communication. Furthermore, biased agents rank the three types of equilibria in the same way. Biased agents prefer equilibria with communication. Their messages are transmitted and more unbiased agents are likely to vote for outcome 1. Thus, both biased and unbiased agents would prefer network communication for lower values of π.

VI Application: number and placement of biased agents.

In this section we apply our analysis to two questions. First, what is the effect of the replacement of an unbiased agent by a biased agent on the welfare of individuals? Second, from the point of view of biased agents, what is the optimal number of biased agents in the population and where should they be placed?

The replacement of an unbiased agent by a biased agent j in the network has three effects: a direct effect on the number of votes for collective action 1, a direct effect on information transmission because a message $m = 1$ is always created when the signal is received by agent j, and an indirect effect on information transmission as messages $m = 1$ are more likely to be blocked by unbiased agents, since the message is less likely to be credible. For unbiased agents who receive the same message as the unbiased agent whose status has switched, all effects concur to reduce expected utility.

Proposition 5 *Consider two assignments of biased and unbiased agents in the network, σ and σ' such that one unbiased agent under σ is replaced by a biased agent in σ'. Then the expected utility of any unbiased agent at the MCE under σ' is lower than at the MCE under σ.*

For biased agents, there is a tradeoff. Both direct effects result in an increase in expected utility, but the indirect effect may induce a decrease in the number of unbiased agents who receive and believe message $m = 1$. As the following example shows, the indirect effect may dominate the two direct effects so that the replacement of an unbiased agent by a biased agent may reduce the utility of biased agents.

Example 5 *Placement of Biased Agents on a Line. As in Figure 2, consider eight agents be arranged on a line. Under the assignment σ, agents $1, 2, 3, 4, 6, 7, 8$ are unbiased and agent 5 is*

biased. Under the assignment σ', agents $1,2,3$ and $6,7,8$ are unbiased and agents $4,5$ are biased. Suppose that the prior π satisfies $\frac{1}{5} \leq \pi < \frac{2}{7}$. Then, in the MCE under σ, message $m=1$ is believed and transmitted by all unbiased agents – the MCE is a FCE, whereas in the MCE under σ', message $m=1$ received from agent 4 is not believed by agent 3 and message $m=1$ received from agent 5 is not believed by agent 6. A simple computation shows that the expected loss of a biased agent under σ (as $p \to 1$) is

$$\mathcal{L} = \frac{7}{8}\frac{7(1-\pi)}{8} = \frac{49(1-\pi)}{64},$$

whereas the expected loss of a biased agent under σ' is

$$\mathcal{L}' = \frac{3}{8}\frac{6\pi}{8} + \frac{6}{8}\frac{2+6(1-\pi)}{8} = \frac{48-18\pi}{64}.$$

For values $\pi \in [\frac{1}{5}, \frac{2}{7})$, $\mathcal{L}' > \mathcal{L}$.

The negative effect of adding biased agents stands in sharp contrast to models of rumors and opinion formation based on fixed laws of diffusion or adoption. In such models, it is always beneficial for biased agents to increase their numbers. Here, where agents strategically transmit messages from others, the introduction of a biased agents can reduce their expected utility, depending on where the agent is located in the network.

This observation in turn raises the following question: How can a biased operator select k nodes in the network to implant biased agents in order to maximize the expected probability that collective action 1 is taken? We analyze this problem in the simple case where n agents are located along a line.

Proposition 6 *Consider n agents on a line. Let $k^* = \frac{n\pi}{1-\pi} + 2\pi - 1$. In the MCE, unbiased agents will transmit a message from a subset of agents that contains at most k^* biased agents. If $k \leq k^*$, there is a full communication equilibrium when biased agents are spaced evenly at locations: $\{\llcorner \frac{n-k}{k+1} + 1 \lrcorner, \llcorner \frac{2(n-k)}{k+1} + 2 \lrcorner, ..., \llcorner \frac{k(n-k)}{k+1} + k \lrcorner\}$. If $k > k^*$, there is a maximal equilibrium with partial communication when $k - k^*$ biased agents are located at the end of the line, and the remaining k^* biased agents are spaced evenly along the line at locations $\{k - k^* + \llcorner \frac{n-k}{k+1} + 1 \lrcorner, k - k^* + \llcorner \frac{2(n-k)}{k+1} + 2 \lrcorner, k - k^* + \llcorner \frac{k(n-k)}{k+1} + k \lrcorner\}$.*

Proposition 6 provides an upper bound on the number of biased agents on a line for full communication, and characterizes uniform spacing as an optimal way for an operator to implant biased agents in the network. The uniform spacing may not be the only optimal location strategy of the operator. For example, if $k = 1$ and π is close to $\frac{1}{2}$, all unbiased agents will transmit a message that could have originated from the biased agent wherever she is located, except at the end of the line. But, as π decreases, unbiased agents are less likely to transmit a message that could have been created by a biased agent, and in the end, the only way to guarantee that the biased agent's message is transmitted is to locate the biased agent exactly in the middle of the line.

VII Extensions

A Two Types of Biased Agents

This section considers a tree and two types of biased agents: 0-biased and 1-biased. The network and all agents' types are common knowledge. We adapt the strategies for the base case and find conditions for existence of a full communication equilibrium. The conditions mirror those for the 1-type case; unbiased agents must be sufficiently confident in the content of a message in order to transmit it to their neighbors. Here this confidence depends on the message and the proportions of agents of both biases, as they are distributed in the network.

Consider the following strategies and beliefs.

Strategies: Biased agents create messages that match their bias, and only transmit messages that match their bias. That is, each β-biased agent has the strategy $M(s) = \beta$ and $t(m) = m$ only if $m = \beta$, otherwise $t(m) = \varnothing$. Unbiased agents create true messages upon receiving a signal; i.e., $M(s) = s$ transmit any message they receive, i.e., $t(m) = m$ and vote for m..

Beliefs: Agents base their beliefs on the strategies and on their knowledge of the paths through the network. They know that message 1 will never be transmitted by 0-biased agents and message 0 by 1 biased agents. This leads us to construct two subgraphs from the original network , one subgraph is free of 1-biased agents, the other is free of 0-biased agents. Formally, consider the network G and remove all the 1-biased agents together with their links. This defines the subgraph G^{-1}, which contains all the 0-biased, all the unbiased agents, and contains no 1-biased agents.

Note that G^{-1} is typically formed of several components. For any directed edge (j,i) in G^{-1}, let $S_i^{-1}(j)$ be the set of agents whose path to i goes through j in G^{-1} that is $S_i^{-1}(j)$ is the set of agents who reach i through j and the path does not contain any 1-biased agent. Define similarly G^{-0} and $S_i^{-0}(j)$. If there is no 0-biased agent, G^{-1} s a collection of components only containing unbiased agents and G^{-0} is the entire graph G.

We now define the beliefs. Consider first information sets which are reached with positive probability. Let agent i receive message 0 from a neighbor j in G^{-1}. By Bayes rule her belief that the true state is 0 is

$$\frac{(1-\pi)}{b_{S_i^{-1}(j)} + u_{S_i^{-1}(j)}(1-\pi)}$$

where $b_{S_i^{-1}(j)}$ and $u_{S_i^m(j)}$ denote the proportions of 0-biased and unbiased agents in $S_i^{-1}(j)$, since, according to the strategies, the message has traveled along a path of non 1-biased agents who access i through j. Similarly, if agent i receives message 1 from neighbor j in G^{-0} her posterior belief that the state is 1 is

$$\frac{\pi}{b_{S_i^{-0}(j)} + u_{S_i^{-0}(j)}\pi}.$$

If an unbiased agent receives no message, either no signal was received or a message m was blocked by a non-m biased agent. The exact computation of the posterior belief depends in a nontrivial way of the location of 0 and 1 biased agents in the network.

The only events with zero probability are event where agent i receives message m from a non-m biased agent. We make the same assumption on beliefs as in the one bias case. When an unbiased agent receives $m = 1$ from a 0-biased agent or $m = 0$ from a 1-biased agent, she keeps her prior belief π.

The equilibrium conditions, then, involve the proportions of biased agents in each subgraph G^{-0} and G^{-1}. We adapt the arguments used with one type of biased agent to prove that these strategies form a full communication equilibrium under simple conditions on the proportions of β-biased agents.

Theorem 4 *There exists a full communication equilibrium when there are two types of biased agent if for each unbiased player i and directed edge (j,i) in G^{-1}*

$$b_{S_i^{-1}(j)} \leq \frac{1-\pi}{\pi} \tag{4}$$

and for any directed edge (j,i) *in* G^{-0}

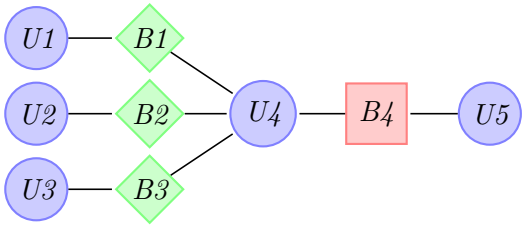
$$b_{S_i^{-0}(j)} \leq \frac{\pi}{1-\pi}. \tag{5}$$

If one inequality in (4) or (5) does not hold, the strategies and beliefs do not constitute an equilibrium.

A difference between the one and two biased agents cases is the interpretation of the absence of a message. With one type of biased agents, the absence of a message signals that message 0 was blocked by a 1-biased agent, and hence the posterior belief that the state is 1 is lower than the prior belief. Unbiased agents always vote for zero when they don't receive any message. When there are two types of biased agents, in the absence of message, the posterior belief that the state is 1 may increase or decrease. In addition, this depends on the location of the unbiased agent in the network. When they do not receive any message, different unbiased agents may update their beliefs in different directions and vote for different outcomes.

Example 6 *Consider the graph depicted in the following figure with five unbiased agents, 3 0-biased agents (green diamonds) and one 1-biased agent (pink square). We first check the conditions under which an FCE exists. Agents* $U1$*,* $U2$ *and* $U3$ *can only receive message* 0 *from their neighbor, and the fraction of 0-biased agents in* $S_i^{-1}(j)$ *is* $\frac{3}{6} = \frac{1}{2}$*. Agent* $U4$ *can receive message 1 from his 1-biased neighbor (the proportion of biased agents in* $S_i^{-0}(j)$ *is then* $\frac{1}{2}$*) and can receive message 0 from his 0-biased neighbors (the proportion of biased agents in* $S_i^{-1}(j)$ *is then* $\frac{1}{2}$*). Agent* $U5$ *can only receive message 1 from his 1-biased neighbor* $B4$ *and the fraction of biased agents in* $S_i^{-0}(j)$ *is* $\frac{1}{2}$*. Hence an FCE exists if and only if* $\frac{\pi}{1-\pi} \geq \frac{1}{2}$ *and* $\frac{1-\pi}{\pi} \geq \frac{1}{2}$*, that is* $\frac{1}{3} \leq \pi \leq \frac{2}{3}$*.*

In an FCE, if the unbiased agents do not receive any msessage, they update their prior beliefs using Bayes rule. Agents $U1, U2$ *and* $U3$ *form a posterior that the state is 1 (for* $p \to 1$*) equal to* $\frac{5\pi}{2+3\pi} > \pi$*. Agent* $U4$ *forms a posterior equal to* $\frac{3\pi}{2\pi+1} > \pi$ *and agent* $U5$ *a posterior* $\frac{6\pi}{7-pi} < \pi$*.*



If either condition (4) or condition (5) fails, we cannot easily extend the algorithm to compute the maximal communication equilibrium. In fact, it may be in the interest of a β- biased agent to transmit message $m \neq \beta$, as the posterior probability of an unbiased agent that s occurs may be higher when no message is received than when message s is sent.

B General Networks

While the base case of a tree allows precise characterization of network subgraphs, a similar analysis for the full communication equilibrium would apply to more general situations. When a network contains cycles, an agent may receive a message several times from the same neighbor or simultaneously from different neighbors, though it would be necessarily be the same message. The existence, uniqueness and optimality of a full communication equilibrium remains true under some conditions on how agents process information.

For example, suppose that agents only transmit a message once. That is, suppose agents react at the first receipt, either by transmitting or not and the second time they receive it, they refrain from transmitting it, nor make any additional inference. To specify behavior, we would just modify the sets $S_i(j)$ to take into account any additional information obtained about the probability the message is generated by a biased agent. An agent could, for example, consider the length of paths. If an agent i receives a message from only one neighbor j, let $T_i(j)$ be the set of agents k for which the shortest path between k and i goes through j. If i receives the message from several neighbors, a set J, say, then $T_i(J)$ denotes the set of agents k such that J is the set of neighbors j of i such that all j lie on the shortest path between k and i. The analysis of the baseline case can be repeated simply by replacing the sets $S_i(j)$ by the sets $T_i(J)$.

VIII Conclusion

This paper studies why agents purposefully and rationally spread rumors. Biased agents desire a particular outcome to occur; unbiased agents want the outcome to match the true state of nature. Both types of agents create and transmit messages in order to influence the common outcome. The analysis compares two benchmark models of communication. In a public broadcast setting, agents anonymously send a message about the state of nature to all other agents. In a network setting, agents send a message to their friends and neighbors, who can then transmit the message to their friends and neighbors.

The analysis shows when each setting has an advantage in generating truthful communication. When agents are less sure about the true state, information is valuable and in both structures full communication is possible. When agents have more confidence in the true state, however, they are less willing to believe messages that could come from a biased source. In the public broadcast model, no communication becomes the only outcome, since agents cannot discern at all the source of the message. In a network, however, agents can discriminate among messages received from different neighbors. They can choose to not transmit messages that originate in parts of the network that are heavily populated with biased agents. We construct an algorithm to identify, in any network, the paths along which messages can flow in an equilibrium.

A feature that emerges in the network maximal equilibria is one-way flow of information. A message can flow from a part of the network to another, but not in the opposite direction, since the proportion of biased agents on either side of the link determine the credibility of the message. Thus, studies of the spread of information and rumors in networks should consider that links are not always used and not always used in both directions.

We also find that in order to influence outcomes, biased agents might prefer to limit their numbers and to spread themselves within the population. If there are too many biased agents, unbiased agents do not believe any messages and do not send them along to their neighbors. With fewer biased agents located sporadically in the network, unbiased agents transmit all messages, since the likelihood of the messages being false is sufficiently small. Hence biased agents are better off since, if they have the opportunity, they can create a false message which then spreads.

Future research would consider situations where agents have some information about the network but not complete information. For example, agents may know the number and biases of

their neighbors, but more distantly in the network they know only know the degree distribution and proportions of biased and unbiased agents. In order to make a judgement about a received message, agents would need to make inferences about the the fraction of biased agents in different directions of the nework. Another example would be networks with homophily—biased agents are more likely to have biased neighbors and unbiased agents are more likely to have unbiased neighbors. Following the insights of the current paper, agents would use this information to make inferences about the credibility of a message.

Appendix

Proof of Lemma 1. Let us consider unbiased agent i at the end of the transmission stage and $\mathcal{I}$ her information: $\mathcal{I}$ includes others' strategies as well as the message/signal i may have received and from whom, or the absence of message. Others' votes depend on their own information but not on i's vote. Let us denote their number by $\tilde{z}$ and by $Proba(\theta, z|\mathcal{I})$ their joint probability with the state as is perceived by i. Thus i's expected utility from voting for a, $a = 0, 1$, is

$$E[w(\tilde{x}, \tilde{\theta})|a, \mathcal{I}] = \sum_{z,\theta}[w(1,\theta)f(z+a) + w(0,\theta)(1 - f(z+a))]Proba(\theta, z|\mathcal{I}) \tag{6}$$

The incentive to vote for 1 instead of 0 are thus determined by the sign of

$$\sum_{z,\theta}[w(1,\theta) - w(0,\theta)](f(z+1) - f(z))Proba(\theta, z|\mathcal{I})$$

For $f(z) = z/n$, this expression writes

$$\frac{1}{n}\sum_{z,\theta}[w(1,\theta) - w(0,\theta)]Proba(\theta, z|\mathcal{I}).$$

Since $[w(1,\theta) - w(0,\theta)]$ is equal to 1 for $\theta = 1$ and to -1 for $\theta = 0$ the above expression is equal to

$$\frac{1}{n}[2\sum_{z} Proba(\theta = 1, z|\mathcal{I}) - 1]$$

As $\sum_z Proba(\theta = 1, z|\mathcal{I}) = Proba(\theta = 1|\mathcal{I})$, we finally obtain that the incentives to vote for 1 or 0 only depends on the sign of $2Proba(\theta = 1|\mathcal{I}) - 1$, i.e. on $2\rho_i - 1$ where ρ_i is the posterior on the state being 1 at the time of the vote. ■

Proof of Proposition 1. The proof is provided in the text. ■

Proof of Theorem 1.

Sufficiency. Given $b_{S_i(j)} \leq \frac{\pi}{1-\pi}$, we consider possible deviations from the specified strategies.

Biased agents. For message strategies, a biased agent's expected payoff cannot increase by adopting the strategy $M(s) = 0$ or $M(s) = \varnothing$. Given the strategies of other agents and beliefs consistent with these strategies, creating message $m = 0$ rather than $m = 1$ would decrease the probability that unbiased agents receive a message $m = 1$ and thus decrease the number of agents that vote for outcome $x = 1$. For transmission strategies, the same argument applies.

Unbiased agents. For message strategies, an unbiased agent who receives a signal s believes with probability 1 that the true state is s. Given other agents' strategies, creating any message other than s then lowers the expected number of agents who will vote for outcome $x = s$. For transmission strategies, all unbiased agents believe with greater than probability $1/2$ that the true state is $1(0)$ upon receiving a message $m = 1(m = 0)$. Since any unbiased agent's expected utility is increasing in the number of agents who share his beliefs, an agent cannot benefit by not transmitting a message or blocking a message, given other agents' transmission strategies and beliefs.

Necessity. Suppose $b_{S_i(j)} > \frac{\pi}{1-\pi}$ for some unbiased agent i and one of his neighbors j. This agent would have an incentive to deviate from the specified transmission strategy and adopt the strategy to block a message $m = 1$ received from neighbor j. In this case, agent i holds that state 1 is less likely than state 0, despite having received the message $m = 1$. Given other agents' strategies and beliefs, agent i can improve his expected payoffs by not transmitting the message.■

Proof of Proposition 2. The argument is provided in the text.■

Lemma 2 *A If $V^t \neq \varnothing$, then $V^{t+1} \subset V^t$. Hence, there exists a step T such that $V^T = \varnothing$ and $V^{T-1} \neq \varnothing$.*

Proof of Lemma 2.

Consider V^t and let (j,i) be the level 1 directed edge in V^t that is eliminated in this step. Consider a directed edge $(k,l) \in G^{t+1}$ and suppose that (k,l) is not in V^t. We show that it is not in V^{t+1}. As $(k,l) \in G^{t+1}$, (k,l) cannot belong to $G_i(j)$. Now, consider two possibilities: either $(j,i) \in G_l(k)$ or not.

If $(j,i) \notin G_l(k)$ when k receives message $m = 1$ from l, the message cannot have traveled through (j,i): the elimination of (i,j) does not affect $G_l^t(k)$ nor $S_l^t(k)$. Hence,

$$b_{S_l^{t+1}(k)} = b_{S_l^t(k)} \leq \frac{\pi}{1-\pi},$$

so that (k,l) does not belong to V^{t+1}.

If $(j,i) \in G_l(k)$, then $G_i(j) \subset G_l(k)$. Hence, $S_i(j)^t \subset S_l(k)^t$. Following the elimination of (j,i), $S_i(j)^t$ has been withdrawn from $S_l(k)^t$, hence $S_l(k)^{t+1} = S_l(k)^t - S_i(j)^t$ and

$$b_{S_k^{t+1}(l)} = \frac{|B \cap S_l(k)^{t+1}|}{|S_l(k)^{t+1}|} = \frac{b^t_{S_l(k)}|S_l(k)^t| - b^t_{S_i(j)}|S_i(j)^t|}{|S_l(k)^t| - |S_i(j)^t|}. \tag{7}$$

As $b_{S_k^t(l)} \leq \frac{\pi}{1-\pi} < b_{S_i^t(j)}$,

$$\frac{b^t_{S_l(k)}|S_l(k)^t| - b^t_{S_i(j)}|S_i(j)^t|}{|S_l(k)^t| - |S_i(j)^t|} < \frac{b^t_{S_l(k)}|S_l(k)^t|}{|S_l(k)^t|},$$

so that

$$b_{S_k^{t+1}(l)} < b_{S_k^t(l)} \leq \frac{\pi}{1-\pi},$$

(k,l) is not in V^{t+1}, concluding the proof of the lemma.■

Lemma 3 *A The set W of eliminated edges is independent of the order in which directed edges are chosen at each step of the algorithm.*

Proof of Lemma 3.

The proof is by induction on the *initial* levels of edges in the set V, that is their levels in G^0.

All level 1 directed edges $(j,i) \in V$ are always eliminated in the algorithm, irrespective of the order in which edges are chosen: as the subgraph $G_i(j)$ contains no edge in V, the proportion $b^t_{S_i(j)}$ stays constant and any level 1 edge $(j,i) \in V$ remains level 1.

Suppose the induction assumption holds for all directed edges (k,l) of initial levels smaller than ℓ: either (k,l) is eliminated by the algorithm in all possible orders or never. Consider a directed edge (j,i) of level ℓ. Since $G_i(j)$ only contains edges of initial levels smaller than ℓ, the final graph $G_i^T(j)$ obtained when the algorithm stops is independent of the order. As a result, the proportion of biased agents in $G_i^T(j)$ is independent of the order in which edges are chosen, and we can unambiguously determine whether (j,i) is eliminated or not, proving the induction step. ■

Lemma 4 *A For any $(j,i) \in W$, j is biased and i is unbiased.*

Proof of Lemma 4.

Recall that i is unbiased by definition of directed edges in V. As for j, j must be biased for any level 1 edge in V (otherwise there would be a violating edge in $G_i(j)$. The same argument holds for any edge in W because each edge in W is a level 1 edge of V^t in G^t at the step t it is eliminated. ■

Proof of Theorem 2.

Consider first the behavior of a biased agent. A biased agent does not have an incentive to deviate and either create $m=0$ upon receipt of a signal, transmit $m=0$, or not transmit a $m=1$. Given agents' beliefs, any of these action would (weakly) increase the probability that an agent votes for outcome 0 instead of outcome 1.

Consider next unbiased agents. An agent i who receives $s=0$ or $m=0$ has the belief that the true state is 0. She then does not have an incentive to deviate and create or transmit $m=1$, since this action will (weakly) increase the probability that more agents vote for outcome 1. An agent i who receives $s=1$ knows that the true state is 1. She does not have an incentive to deviate and create $m=0$ since, given the beliefs, this will (weakly) increase the number of agents who vote for outcome 0. For transmission of $m=1$, an unbiased agent i who receives $m=1$ and places sufficiently high probability that the true state is 1, cannot gain by blocking the message. For $(j,i) \in G^*$, then, an agent i who receives $m=1$ from j cannot gain by blocking the message. If on the other hand, she receives $m=1$ from a neighbor j where $(j,i) \notin G^*$, she cannot gain by transmitting the message: her beliefs are $\widetilde{\rho}_i(1(j)) < \frac{1}{2}$, and given the strategies of others, more agents will then vote for 1 and lower her expected utility.

We now show that there cannot be an equilibrium where $m=1$ is transmitted along a directed edge (j,i) not in G^*. (j,i) is in W hence in V. The proof is by induction on the level of (j,i) in the initial graph G^0.

First suppose that (j,i) is a level 1 edge. In the specified strategies, for any edge (k,l) with l unbiased in $G_i(j)$ l transmits $m=1$ when he receives it from k (this is also the case for l biased by assumption). Consider an alternative equilibrium.

If for any edge (k,l) in $G_i(j)$, with l unbiased, l behaves as in the original equilibrium, then i's posterior $\widetilde{\rho}_i(1(j))$ is the same as in the original equilibrium, hence is lower than 1/2: i must block the message.

Otherwise, there are edges (k,l) in $G_i(j)$, with l unbiased, for which l does not transmit $m=1$ received from k. Call such an edge deviating and denote D the set of deviating edges. The subgraph $G_i'(j)$ along which $m=1$ can reach i through j in the alternative equilibrium is made of all the paths to i in $G_i(j)$ that contains no edge in D. We show that the proportion of biased agents in $S_i'(j)$ (with obvious notation) is larger than $\frac{\pi}{1-\pi}$.

Let (k,l) be in D such that $G_l(k)$ contains no edge in D (such a (k,l) surely exists). Since (k,l) is not in V (because (j,i) is of level 1), the proportion of biased agents in $S_l(k)$ is not larger than $\frac{\pi}{1-\pi}$. Since (i,j) is in V the proportion of biased agents in $S_l(k)$ is larger than $\frac{\pi}{1-\pi}$. Hence the proportion of biased agents in $S_i(j) - S_l(k)$ is strictly larger than $\frac{\pi}{1-\pi}$.[12]

Consider the directed tree $G_i(j) - G_k(l)$. If it contains no element in D, $G'_i(j) = G_i(j) - G_k(l)$. The set of nodes of $G_i(j) - G_k(l)$ is $S_i(j) - S_l(k)$, which has a proportion of biased agents strictly larger than $\frac{\pi}{1-\pi}$: i's posterior $\widetilde{\rho'}_i(1(j))$ is lower than 1/2: i must block the message.

If the directed tree $G_i(j) - G_k(l)$ contains an element in D we can use the previous argument to that tree (i.e. replacing $G_i(j)$ by $G_i(j) - G_k(l)$) and obtain a sub-tree by deleting an element of D. Continuing this way, we obtain a decreasing sequence of subgraphs whose nodes have a proportion of biased agents larger than $\frac{\pi}{1-\pi}$ by deleting at each step a subgraph containing an edge in D. The process stops when all deviating elements have been eliminated and the tree $G'_i(j)$ is reached; this proves that $S'_i(j)$ has surely a proportion of biased agents larger than $\frac{\pi}{1-\pi}$ so that i's posterior $\widetilde{\rho}_i(1(j))$ is lower than 1/2: i must block message 1 from j at the alternative equilibrium.

Next, at the induction step, suppose that all unbiased agents l with $(k,l) \notin G^*$ of level smaller than ℓ block message 1. Consider a level ℓ edge $(i,j) \notin G^*$. Consider the directed subtree $G^*_i(j)$ of $G_i(j)$. $G^*_i(j)$ is the subgraph along which $m = 1$ travels i through j in the original equilibrium and contains violating edges of level smaller than ℓ. Therefore, by the induction assumption, at an alternative equilibrium, $m = 1$ can reach i through j only along a path included in $G^*_i(j)$. We can therefore apply exactly the same argument as above, replacing the tree $G_i(j)$ by $G^*_i(j)$.■

Proof of Proposition 3. The argument is provided in the text.■

Proof of Proposition 4. As the behavior of biased agents and of unbiased agents at the initial stage are fixed, we only need to consider the transmission of unbiased agents at edges in G^*. If the message is 0, unbiased agents surely want to transmit the message because they know it is thruthful. Suppose that the message is 1. We show that the activity rule shows that the message will also be believed and transmitted.

We implement the following coloring of directed edges in G^*. Start by coloring edge (j,i) in green if i is biased and in white if i is unbiased. Each white edge will be colored in blue if i believes the message with probability grater than $\frac{1}{2}$. By the activity rule, this implies that i transmits the message as in the MCE. Uniqueness is proved by coloring all white edges in blue.

Consider all white dangling edge (j,i) with j as a leaf. As (j,i) is in G^*, i believes with probability greater than $\frac{1}{2}$ that the message is truthful (in fact the posterior probability is 1). Color the directed edge (j,i) in blue. At the end of the first step, all dangling edges are green or blue. Furthermore, there are no other blue edges in the graph. This initial step shows that whenever the diameter of $G_i(j)$ is equal to one, all edges are colored in blue or green.

[12] By a computation similar to (7)

$$b_{S_i(j)-S_l(k)} = \frac{b_{S_i(j)}|S_i(j)| - b_{S_l(k)}|S_l(k)|}{|S_i(j)| - |S_l(k)|}. \tag{8}$$

As $b_{S_l(k)} \leq \frac{\pi}{1-\pi}$ and $b_{S_i(j)} > \frac{\pi}{1-\pi}$, we obtain $b_{S_i(j)-S_l(k)} > \frac{\pi}{1-\pi}$.

Suppose now that in all graphs $G_i(j)$ with diameter smaller than d, all directed edges are colored in blue or green. Pick a white edge (j,i) such that the diameter of $G_i(j)$ is equal to d. As all paths in strict subgraphs of $G_i(j)$ are either blue or green, agents in $S_i(j)$ behave as in the MCE, and the posterior belief of agent i receiving message 1 from j is the same as in the MCE. As this posterior is larger than $\frac{1}{2}$, by the activity rule he transmits the message. Hence, the edge (j,i) is colored in blue.■

Proof of Theorem 3. We provide the proof in the case where a biased agent always creates message $m = 1$ when he receives the signal. The proof in the general case is more involved, and is available on request.

We compare the expected utility, or, its opposite, the expected loss of an unbiased agent in the MCE and in an alternative equilibrium, denoted $\prime$.

A strategy profile determines the number of votes $\tilde{z}$ in each state given who receives the signal or whether no signal is sent. As each (ex ante) individual loss depends only on these votes $\tilde{z}$, all unbiased agents derive the same ex ante loss (even if they do not always vote identically as they may not have the same information). For $f(z) = z/n$, this common expected loss is directly related to the total number of votes not matching the state; up to the factor $\frac{1}{n}$ it is equal to

$$\pi \times [\text{number of votes for } 0 \ |\theta = 1] + (1-\pi) \times [\text{number of votes for } 1 \ |\theta = 0]$$

Since biased agents always vote for 1, this expression is equal to the sum of the constant $(1-\pi)b_N$ and

$$\pi \times [\text{number of U-votes for } 0 \ |\theta = 1] + (1-\pi) \times [\text{number of U-votes for } 1 \ |\theta = 0].$$

which writes by disaggregating over all U-agents

$$\sum_{i \in U} \pi[\text{number of } i\text{-votes for } 0 \ |\theta = 1] + (1-\pi)[\text{number of } i\text{-votes for } 1 \ |\theta = 0].$$

i's term inside the square brackets is the probability she casts a vote that does not match the state. In what follows, we show that this probability is minimized at the MCE relative to other equilibria. This will prove the theorem.

To simplify the presentation we take $p = 1$. Consider an unbiased agent i. Given a strategy profile, i's vote is a function of who receives the signal and the state, i.e. the value of the signal (this does not assume that i has this information). Let $v(j,\theta) \in \{0,1\}$ denote i's vote when the signal lands on agent j in N and the state is θ, i.e. j has received signal θ, where we omit index i for simplification. The probability that i's vote does not match the state is

$$\mathcal{L}^i = \frac{1}{n} \sum_{j \in N \setminus i} [\pi \mathbf{1}_{v(j,1)=0} + (1-\pi)\mathbf{1}_{v(j,0)=1}].$$

where $\mathbf{1}$ is the indicator function.

We will need the following lemma, which follows from the construction and properties of the MCE.

Lemma 5 *There is a collection of disjoint sets $S_k(\ell)$ where (ℓ,k) is in W such that if the signal lands on*

an agent in one of these sets, i receives no message at any equilibrium. Let us denote by N_{-i} the union of these sets.

Proof of Lemma 5. Let (ℓ, k) be a directed edge in W where k is between ℓ and i. We know from Theorem 2 that k does not transmit $m = 1$ from ℓ neither at the MCE nor at any equilibrium. Furthermore ℓ is biased and k is unbiased, so k never receives $m = 0$ from ℓ. We thus obtain that no message goes through the directed edge (ℓ, k) in W. This proves that i receives no message at any equilibrium when the signal lands on an agent in one of these sets. As the sets are either disjoint or included into one another, we may pick up the maximal sets and obtain a partition of N_{-i}. (Each maximal set is such that the path from ℓ to i contains no edge in W.) ■

In the MCE, if unbiased agents receive message m they vote for m ; if they receive message $\emptyset$ they vote for 0. Hence, there are two sources of incorrect votes:

1. Agent i does not receive any message and $\theta = 1$
2. Agent i receives message $m = 1$, the signal is received by $\ell \in B$ and $\theta = 0$.

Consider another equilibrium denoted by $\prime$. We show that if i votes for the correct outcome when i does not in a MCE, i must vote for the wrong outcome in a sufficiently large number of situations where he is correct at the MCE so that the incurred loss overweighs the benefit. We deal with each source of incorrect votes in turn.

1. Agent i does not receive any message and $\theta = 1$.

At the MCE, everyone creates $m = 1$ upon the receipt of signal $\theta = 1$. Hence the signal has been received by some j who has sent message $m = 1$ but the message has not reached i.[13] Thus there must exist a directed edge (ℓ, k) in W on the path from j to i: j belongs to N_{-i}.

At the $\prime$ equilibrium, no message can go through (ℓ, k) (lemma 5), and i votes for the same outcome whenever $j \in S_k(\ell)$ (even if i does not know that $j \in S_k(\ell)$). In the MCE, i votes for 0. In the $\prime$ equilibrium, i might vote for 1. As (ℓ, k) is in W, $b_{S_k(\ell)} > \frac{\pi}{1-\pi}$, so that voting for 0 matches the state more often than voting for 1 when $j \in S_k(\ell)$.

As the argument works for any set $S_k(\ell)$ in the partition of N_{-i}, the probability that i's vote matches the state when the signal is received by $j \in N_{-i}$ is at least as high in the MCE as in any other equilibrium.

2. The signal is received by ℓ in B and $v(\ell, 0) = 1$ in the MCE equilibrium.

ℓ is surely in $N_i = N - N_{-i}$. As biased ℓ creates message $m = 1$ when he receives the signal 0, either (a) $m = 1$ does not reach i and i votes for 0 when $m = \varnothing$ or (b) $m = 1$ reaches i through an agent j; she votes for 0 because the posterior is less than 1/2.

Assume that in the alternative $\prime$ equilibrium, $v'(\ell, 0) = 0$ so that the expected loss when ℓ receives the signal is less at this $\prime$ equilibrium than at the MCE. Consider the edge (ℓ, k) on the path from ℓ to i ($k = i$ is possible). Let $S_k^*(\ell)$ denote the set of agents from which i can receive a message transiting through (ℓ, k) in the MCE.[14] In the $\prime$ equilibrium, i possibly receives a message transiting through (ℓ, k) from these agents only. We show that the probability that i's vote matches the state when the signal is received by $j \in S_k^*(\ell)$ is at least as high in the MCE than in any other equilibrium.

[13] Or the signal has not been received, which is impossible for $p = 1$.

[14] N_i is the component to which i belongs in the graph where all the elements W have been dropped, i.e. it is the component of the graph Γ obtained at the end of the algorithm. $S_k^*(\ell)$ is the set corresponding to (ℓ, k) in Γ.

Assume first that k is unbiased. As $m = 0$ does not travel through ℓ, i can receive only $m = 1$ or $m = \varnothing$. If i receives $m = 1$ it is through j (case (b) above) and this triggers vote 0. If $m = \varnothing$, there are two possible cases :

1- i votes for 0 when $\varnothing$. Then, whatever the signal received in $S_k^*(\ell)$, i votes for 0 at the $\prime$ equilibrium. At the MCE, k is better off following the messages from strategy than voting constantly for 0.

2- i votes for 1 when $\varnothing$. Then, i votes for the wrong outcome when the signal 0 is received by an unbiased agent in $S_k^*(\ell)$. Furthermore when a biased agent receives the signal, the vote is constant and the minimal probability of a wrong outcome is π. Hence, denoting $b = b_{S_k^*(\ell)}$, the probability of an incorrect vote in the $\prime$ equilibrium when the signal lands on $S_k^*(\ell)$ is at least $(1-b)(1-\pi) + b\pi$. This has to be compared with $b(1-\pi)$ which is the expected loss at the MCE. By construction of the MCE, we have $b < \pi/(1-\pi)$; it is easy to check that this implies $(1-b)(1-\pi) + b\pi > b(1-\pi)$.[15] Hence, the probability that i's vote matches the state when the signal is received by $j \in S_k^*(\ell)$ is at least as large in the MCE than in any other equilibrium.

Assume now k to be biased. Consider the first unbiased agent k' on the directed path from ℓ to i ($k' = i$ is possible): there is a biased agent ℓ' on this path linked to k' and all agents between ℓ and ℓ' are biased. Thus when ℓ sends 1, all these biased agents transmit it. It implies that when ℓ' sends $m = 1$ (either transmits or creates), agent i is in the same situation as when ℓ sends $m = 1$: either (a) message 1 does not reach i or (b) $m = 1$ reaches i through j. As i cannot distinguish whether ℓ or ℓ' have sent the message, he votes 0. We thus have a pair (k', ℓ') where k' is unbiased, ℓ' is biased and i votes for 0 when ℓ' receives the signal and we can apply the previous result.

To conclude, in each situation in which agent j receives the signal and i's vote does not match the state in the MCE, there is a set of agents that contains j such that when the signal lands on this set, the expected number of i's incorrect votes in the $\prime$ equilibrium is at least as high as in the MCE, and furthermore the sets are disjoint.■

Proof of Proposition 6. We first show that, in a maximal communication equilibrium, no unbiased agent can believe message $m = 1$ received from more than k^* biased agents. Suppose by contradiction that there exists an unbiased agent and $k > k^*$ biased agents such that, whenever message $m = 1$ is originated by one of the k biased agents, the unbiased agent switches his prior to $\rho_i > \frac{1}{2}$. First note that, as $k \geq k^* + 1$,

$$k \geq \frac{n\pi}{1-\pi} + 2\pi - 1 + 1 = \frac{n\pi}{1-\pi} + 2\pi.$$

Fix the unbiased agent i and consider two sets $S_i(j)$ and $S_i(j')$. Let $0 \leq l \leq k$ be the number of biased agents in $S_i(j)$ and $k - l$ the number of biased agents in $S_i(j')$. By Theorem 1,

$$\begin{aligned} l(1-\pi) &\leq |S_i(j)|\pi, \\ (k-l)(1-\pi) &\leq |S_i(j')|\pi. \end{aligned}$$

[15] $(1-b)(1-\pi) + b\pi > b(1-\pi)$ is equivalent to $b \leq (1-\pi)/(2-3\pi)$. Now $\pi/(1-\pi) \leq (1-\pi)/(2-3\pi)$ (as this is equivalent to $(1-2\pi)^2 \geq 0$, hence $b \leq \pi/(1-\pi)$, implies $b \leq (1-\pi)/(2-3\pi)$.

Summing up,

$$k \leq \frac{(|S_i(j)| + |S_i(j')|)\pi}{1-\pi} < \frac{n\pi}{1-\pi},$$

yielding a contradiction.

We now show that, in the optimal location described in the Proposition, all unbiased agents believe message $m = 1$ received from $\hat{k} = \min\{k, k^*\}$ biased agents. Pick an unbiased agent i and consider the set $S_i(j)$. By construction, if the set $S_i(j)$ contains l biased agents,

$$|S_i(j)| \geq l + l\frac{n-\hat{k}}{\hat{k}+1}.$$

so that the fraction of biased agents in $S_i(j)$ satisfies:

$$b_{S_i(j)} = \frac{l}{|S_i(j)|} \leq \frac{\hat{k}+1}{n+1} \leq \frac{k^*+1}{n+1} \leq \frac{\pi}{1-\pi}.$$

By Theorem 1, this implies that a full communication equilibrium among $\hat{k}$ biased agents and the unbiased agents exists. ■

Proof of Theorem 4 :

Sufficiency:

We consider possible deviations from the specified full communication strategies. Observe first that the votes for outcome x are (weakly) maximized if message $m = x$ is transmitted. To show this, let m be created, and contemplate changing m into m'. This change can switch the vote of an unbiased agent i under two cases. First, i receives now m' hence votes for m' whereas she was voting for $x = m$ (because either she received m or she received nothing and votes for $x = m$ in that case). Second, i receives nothing and vote for $x = m'$ whereas she voted for $x = m$ because she received m. In both cases, the change is in favor of a vote for $x = m'$. Similarly let m be created, and contemplate not creating any message. The previous argument applies (only the second case can happen) hence the same result holds. The same argument also applies at a transmission stage: not transmitting message m can only decrease the number of votes for $x = m$. To summarize, creating m or transmitting maximizes the number of votes for outcome $x = m$.

Suppose conditions (4) and (5) are satisfied. We show that no possible deviation is beneficial.

Biased agents. For creation strategies, a $\beta-$biased agent creates $m = \beta$. By the monotonicity of the votes, he cannot improve the number of votes for his bias by creating a message that does not match his type or by not creating any message. For transmission strategies, he only transmits a message that does not match his type; the same argument applies.

Unbiased agents. For creation strategies, an unbiased agent who receives a signal s believes with probability 1 that the true state is s. Creating any message other than s (s' or $\emptyset$) can only lower the number of agents who will vote for the true state $x = s$, which is harmful for an unbiased agent. As for the vote and transmission strategies, under the conditions (4) and (5), all unbiased agents believe with probability greater than 1/2 that the true state is 1 (resp. 0) upon receiving a message $m = 1$ (resp. $m = 0$) from any unbiased or 1-biased (resp. 0-biased) neighbor. When the probability is strictly greater than 1/2, the unbiased agent's expected utility is increasing in the number of agents who vote for 1 (resp.0)

so that she cannot benefit by not voting for 1 or by not transmitting the message. When the probability is just equal to 1/2, she is indifferent hence the specified strategy is optimal as well.

To check perfectness, consider an unbiased agent who received a message m from a non-m-biased agent. We specified that her posterior that the state is m is larger than 1/2, so the specified strategies are optimal by the same argument.

Necessity. If (4) does not hold, then agent i believes that the state is 1 with probability strictly greater than $\frac{1}{2}$, and can improve her expected payoffs by not transmitting the message $m = 0$. And similarly, if (5) does not hold not transmitting the message $m = 1$ from j is optimal. ■

REFERENCES

Acemoglu, Daron, Munther Dahleh, Ilan Lobel, and Asuman Ozdaglar. 2011. "Bayesian Learning in Social Networks," *Review of Economic Studies*, 78, pp. 1201-1236.

Acemoglu, Daron and Asuman Ozdaglar. 2011. "Opinion Dynamics and Learning in Social Networks," *Dynamic Games and Applications* 1(1), pp. 3-49.

Banerjee, Abhijit. 1992. "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107(3), August 1992, pp. 797-817.

Bikhchandani, Sushil, David Hirshleifer and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," Journal of Political Economy," 100,pp. 992-1026.

Calvó-Armengol, Joan de Martí and Andrea Prat, "Communication and Influence" working paper 2011.

Castellano, Claudio, Santo Fortunato, and Vittorio Loreto. 2009. "Statistical Physics of Socal Dynamics," *Review of Modern Physics* 81(2), pp. 591-646.

Centola, Damon and Michael Macy. 2007. "Complex Contagion and the Weakness of Long Ties in Social Networks," *American Journal of Sociology* 113(3), November, pp. 702-34.

Conley, T. and C. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, 100, pp. 35-69.

Crawford, Vincent P. and Joel Sobel. 1982. "Strategic Information Transmission". *Econometrica* 50 (6), pp. 1431–1451

DeMarzo, Peter M. Dimitri Vayanos, and Jeffrey Zweibel. 2003. "Persuasion bias, social influence, and uni-dimensional opinions," *Quarterly Journal of Economics* 113(3), pp. 909-968.

DeGroot, Morris H. 1974. "Reaching a Consensus," *Journal of the American Statistical Association*, 69 (345) March, pp. 118-121.

Foster, Andrew and Mark Rosenzweig. 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," Journal of Political Econ-

omy 103 (December 1995),pp. 1176-1209.

Gale, Douglas and Shachar Kariv. 2003. "Bayesian Learning in Social Networks," *Games and Economic Behavior*, November 2003, 45(2), pp. 329-346.

Galeotti, Andrea, C. Ghiglino and F. Squintani. 2013. "Strategic Information in Networks," forthcoming *Journal of Economic Theory.*

Golub, Benjamin and Matthew Jackson. 2010. "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2, pp. 112-149.

Hagenbach, Jeanne and Frédéric Koessler. 2010. "Strategic Communication in Networks," *Review of Economic Studies* 77(3), pp. 1072-1099.

Jackson, Matthew. 2008. *Social and Economic Networks.* Princeton: Princeton University Press.

Milgrom, Paul R. 1981. "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12(2), pp. 380–391.

Milgrom, Paul and John Roberts. 1986. "Relying on the Information of Interested Parties." *Rand Journal of Economics,* 17, pp. 18-32.

Niehaus, Paul. 2011. "Filtered Social Learning," *Journal of Political Economy* 119(4), pp. 686-720.

Romero, Daneil, Brendan Meeder, and Jon Kleinberg. 2011. "Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter." *Proc. 20th International World Wide Web Conference*, 2011.

Sunstein, Cass R. 2009. *On Rumors: How Falsehoods Spread.* New York: Farrar, Straus and Giroux.