Homophily is (not) Bad for Learning

Abstract

The consensus in the social learning literature is that homophily –defined as the tendency of people to associate more with those who are similar to themselves– is not conducive to learning. This is because increasing homophily leads to a more segregated society and insular communities that cannot efficiently aggregate information. In this paper, we show that this common view of homophily is inaccurate when some of the information that agents use to learn the state of the world is provided by a strategic actor who may be interested in distorting their learning process for his own gain. In this setting, increased homophily has an ambiguous effect on learning, and can even protect agents from misinformation and help them learn the state of the world. This finding stands in stark contrast to almost all of the social learning literature on homophily, with potential policy implications. For example, we show that a policy maker may not want to encourage the integration of highly-segregated communities if this integration will only slightly reduce homophily, as this might actually facilitate the spread of misinformation. We use our model to propose optimal policies that try to stop this spread of misinformation, and show that these policies might require creating inequality –defined as unequal access to network resources– in the population.

1 Introduction

The central question of the social learning literature is whether a society or a group of people will manage to learn an unknown state of the world.[1] Agents update their beliefs about this state based on their own information (that they obtain from news sources) as well as from communicating with their friends, and various results exist that detail when learning is or is not possible, with a specific emphasis on how homophily –the tendency of people to associate with those who are similar to themselves– impacts the ability of agents to learn. The consensus is that homophily is disadvantageous to learning (e.g. Golub and Jackson (2012)), since it results in agents being less exposed to different opinions and beliefs, which slows down the learning process and leads to echo chamber effects.

The prevailing assumption in all of the previous literature is that the news that agents obtain are organic: it may or may not be accurate, but it is provided by sources that have no stake in

[1] For example, whether temperatures on earth are increasing over time, as stated in Golub and Jackson (2012).

what the agents' beliefs are. In reality, such sources are rarely neutral, and may have an interest in shaping these beliefs in order to direct agents towards taking certain actions. The interference of Cambridge Analytica in the 2016 US presidential election is one example of such belief manipulation. Similarly, recent outbreaks of measles in Eastern Europe and parts of the US have been linked to Russian interference and propaganda whose goal is to convince people that vaccines are harmful in order to make them opt against vaccinating themselves and their children.[2]

This paper combines the two ingredients from the previous paragraphs to answer the following question: if a strategic actor – a *principal* – tampers with the learning process in order to get agents to mislearn the state of the world, and if this happens in the presence of a phenomenon (homophily) that is believed to be bad for learning the state of the world, then does that make things doubly worse for the agents? As we show, the somewhat unexpected answer is no. Homophily, long regarded as a negative and undesirable force in social learning, has an ambiguous effect when information is provided to agents by strategic sources, and can some times even shield agents from misinformation. This novel finding is particularly compelling because of its potential consequences to social and economic policy: in addition to its effects on learning, homophily is also connected to segregation in society (Currarini et al. (2009)), which leads to well-documented economic inequality.[3] This makes it natural to suggest policies that alleviate the negative effects of homophily, but as we show, efforts to integrate communities can end up making society as a whole more prone to misinformation.

The model we develop in this paper builds upon the work of Chandrasekhar et al. (2019) and Mostagir et al. (2019). The first paper shows, experimentally, that agents update their beliefs in ways that are consistent with Bayesian or DeGroot updating,[4] and suggests that the proportion of agent types in society may be driven by contextual factors (e.g. level of education). The latter paper takes this finding and builds a theoretical model where a principal tries to manipulate a society with mixed learning types, and shows that Bayesian agents are never manipulated and that, once they figure out the true state of the world, can help spread this knowledge to other agents. Whether these other agents are affected by misinformation is a function of the (deterministic and known) social network structure.

In this paper, we assume no knowledge of the network structure; instead, we assume that the network is generated randomly from a distribution and exploit the connection between ran-

[2] https://www.newsweek.com/russian-trolls-promoted-anti-vaccination-propaganda-measles-outbreak-1332016

[3] Wage inequality and differences in labor market participation across different groups are documented in Card and Krueger (1992); Chandra (2000), and Heckman et al. (2000). Calvo-Armengol and Jackson (2004) and Calvó-Armengol and Jackson (2007) show that these inequalities can be explained through network models of homophily.

[4] A Bayesian agent integrates the opinions of her friends using Bayes' law, while a less-sophisticated DeGroot agent takes the opinions of her friends on face value and incorporates them into her own beliefs through a weighted average.

dom graphs and homophily to examine the effects of the latter on the spread of misinformation. The discussion in the previous paragraph suggests that, by protecting other agents from manipulation, Bayesian agents serve as a resource for their communities. This gives rise to a natural definition of inequality that is rooted in homophily: "affluent" communities with easy access to Bayesian agents are more likely to be protected against misinformation, but if homophily is too high and communities become more insular, then vulnerable communities with few or no Bayesians will have little access to the more affluent communities and their resources. How does this affect the spread of misinformation in the individual communities and society as a whole?

The above setup naturally gives rise to important questions about resource allocation and policy interventions. Should a social planner try to curb the spread of misinformation by encouraging the integration of different communities (i.e. by *decreasing* homophily)? Or by endowing some agents in the network with "Bayesian abilities" (for example through education or increasing awareness), and how should these agents be selected? As we discuss next, while these questions are simple to state, their corresponding answers are not straightforward.

Contribution and Overview of Results. In this paper, we consider the manipulation and learning questions presented above in the presence of homophily, arguably a more realistic representation of real-world networks (see Marsden (1987); McPherson et al. (2001)). In the process, we develop new technical results and generate novel insights that invite a rethinking of the role that homophily plays in social learning.

In particular, the paper makes three types of contributions:

Methodological: Our methodological contribution answers the following question: Given a random network model, what is the likelihood that any realized network from this model will be susceptible to misinformation? Theorem 1 is a technical result that shows that under certain regularity conditions, studying manipulation in these random networks can be reduced to applying a centrality measure to the *average* network, which greatly simplifies their analyses.

This result is quite useful because, unlike prior work (e.g. Mostagir et al. (2019)), it assumes no direct knowledge of the network and is robust to changes in the network structure, which makes our model well-suited to the fact that social networks are rarely static configurations; people break ties and form new relationships over time, and the network structure itself is constantly evolving. Our work provides a technical reduction that allows us to study these more realistic models of network formation.

Conceptual: With Theorem 1 in hand, we turn our attention to studying the effects of homophily on the spread of misinformation. We start by giving a definition of inequality as a measure of how communities differ in their access to Bayesian agents, which is shaped by how those agents are distributed in the different communities and by the level of homophily in society. Our main conceptual finding in Theorem 2 is that the effect of inequality is not monotone: manipulation is quite difficult when there is little inequality or when there is extreme inequality, but agents can be manipulated as the level of inequality falls in some intermediate range. This non-monotonicity is shaped by a combination of network connectivity and the principal's costs of sending misinformation to the agents. When there is little inequality, manipulation is quite difficult because information flows freely, and Bayesian agents anywhere in the network can exert the influence necessary for other agents to learn the true state of the world and overcome any misinformation injected by the principal. At the other extreme, when inequality is very high and agents form insular communities, manipulation can still be difficult for the principal because he cannot use the network to his advantage when spreading propaganda. Instead, he has to target communities individually and, as we show, might end up having no profitable strategy for manipulating *any* one in the population. When inequality is in an intermediate range is where things align for the principal's objective. Connectivity is not too high, which prevents influence from the Bayesians scattered through the network, but at the same time the network is connected enough that the principal gets more "bang for his buck", i.e. he only has to target a small subset of agents with misinformation, but can then rely on the homophily structure and the resulting connectivity to spread this misinformation throughout the network.

We then show in Theorem 3 that the narrative of homophily being bad for learning can be recovered in our model, albeit under special conditions. When these conditions do not hold, homophily can again lead to an improvement in learning. This non-monotonicity and the generally ambiguous role that homophily plays in learning and the spread of misinformation has a non-trivial impact on possible network interventions, as we discuss next.

Policy Implications: The findings from our model can be used as input to policy makers trying to curb manipulation. By understanding the homophily structure of society, policy makers can best direct their resources towards interventions that will protect this society from misinformation. We consider two natural interventions: first, increasing connectivity across communities in order to make the network less segregated and provide more access to Bayesian agents, and second, educating specific agents in the network who will then serve as Bayesian agents to protect their (and potentially other) communities.

As we show in Section 5, some of the optimal policy recommendations that arise from these interventions are initially not intuitive. For example, a policy maker may not want to encourage the integration of highly-segregated communities if this integration will only slightly reduce homophily. This is because, as mentioned earlier, moving society from a highly-homophilous, highly-segregated regime to an intermediate homophily regime may make society as a whole more susceptible to misinformation. Thus, while the desire to reduce homophily is reasonable both from a normative perspective and based on the literature cited earlier, our model suggests that it should only be acted upon if the degree of homophily can be drastically reduced, otherwise the intervention will result in the opposite effect for which it was designed.

As another example, we show in Section 5.1 that it may sometimes be better for a policy maker to willfully create inequality in how she spends her budget and how the Bayesian agents are allocated across communities. By unevenly splitting the budget across communities, society as a whole becomes more impervious to manipulation compared to the case where the budget is evenly split. This and other examples demonstrate how subtle deviations from our main theorems can lead to unexpected policy recommendations.

Related Literature The seminal paper of Golub and Jackson (2012) shows the negative effects that homophily has on a society that learns in a DeGroot fashion. Instead of learning rates, our focus is on understanding the role of homophily in whether a network is impervious to manipulation *in the limit*, and when information is potentially provided by a strategic source. Lobel and Sadler (2015) show how the role of homophily in a sequential learning model depends on the density of the network. Homophily in that paper is used to describe alignment of preferences over the agents' decision problem, e.g. an agent who is deciding on a restaurant weighs the opinion of her friend differently if she and her friend prefer the same type of food, whereas homophily in our model captures similarities along dimensions (race, age, profession, income, etc.) that can be orthogonal to whatever state the agents are trying to learn.

Bayesian agents in our model are stubborn agents who know the truth. Opinion dynamics with stubborn agents have been studied in Acemoglu et al. (2013) and Yildiz et al. (2013) among others. The recent work of Sadler (2019) extends Yildiz et al. (2013) to random graphs. What differentiates our paper from this literature is the presence of a strategic principal, which gives rise to completely different learning dynamics and implications.

There is recent work on fake news and manipulation. In Candogan and Drakopoulos (2017) and Papanastasiou (2020), there is no strategic news provider; fake news already exists in the system and the focus is on how it can be identified and controlled. Keppo et al. (2019) pay less attention to the social and network aspect and focus instead on how how Bayesian agents can be manipulated through selective information dissemination. The manipulation problem as presented in our paper was introduced in Mostagir et al. (2019). In that paper, agents interact over a fixed network topology and conditions on that topology are provided to determine whether the agents will be affected by misinformation. Our paper embeds this model in a random network structure in order to study the role of homophily in the spread of misinformation. In addition, our paper provides a prescriptive component to evaluate which policies may be effective in stopping the spread of misinformation as a function of inequality in society. As mentioned earlier, these policies also speak to issues of community integration and resource allocation.

Our paper assumes no knowledge of the network structure, but of the random process from which the network is generated. There is recent literature that tries to recover the network structure from relational data, e.g. Alidaee et al. (2020); Ata et al. (2018) consider a seller who does not know the network structure but, in the presence of externalities, estimates it from transaction data. Other recent work, e.g. Auerbach (2019), tests whether a network was generated from an inhomogeneous random graph model (which includes the class of stochastic block matrices commonly used to model homophily). These methods can be used to estimate the structure of homophily in society and applied as input to our model.

Finally, our paper is also related to diffusion and seeding in random networks, as exemplified by the recent work of Manshadi et al. (2018), Akbarpour et al. (2018), and Sadler (2019). These papers consider the classic problem of which agents to select in order to spread information throughout the network. The primary difference with our work is that we consider an adversarial, strategic principal who is trying to spread his own influence in the network, and our goal is to identify conditions and policies under which we can stop this principal from spreading misinformation, with a specific emphasis on the role of homophily and the social structure of the network in propagating such information.

Organization As mentioned, our paper builds on the model of Mostagir et al. (2019), so we present a summary of that model and the necessary definitions and results in Section 2. Section 3 introduces the random network formation process and provides regularity conditions under which the study of random networks can be reduced to the study of the average network. We use this reduction to study the effects of different models of homophily on learning in Section 4. We then discuss possible interventions to prevent the spread of misinformation in Section 5 and conclude the paper in Section 6.

2 Deterministic Networks: Reduced-Form Model and Solution

We build a framework that embeds the model of Mostagir et al. (2019) into a random network formation model. This random network model is introduced in Section 3, but to make the paper self-contained, we present here a high-level summary of the primitives and results from Mosta-

		Agent	
		R	**S**
State y	**R**	$1, 1+b$	$0, 0$
	S	$1, b$	$0, 1$

Table 1. Terminal Payoffs. The parameter b is in $[-1, 1]$.

gir et al. (2019) that are releveant for our setup.[5] The preliminaries presented below assume connectivity in a deterministic network of known topology, and we provide conditions in Section 3 that guarantee that the realized random network is connected with high probability.

We consider a social network with n agents trying to learn a binary state of the world $y \in \{S, R\}$ over time. Time is continuous and agents learn over a finite horizon, $t \in [0, T)$. At time $t = 0$, the underlying state $y \in \{S, R\}$ is drawn, with $\mathbb{P}(y = S) = q \in (0, 1)$. Agents try to learn the state of the world in order to take an action at time T, with the goal of taking the action that matches the true state, i.e. take action S if the state is S or action R if the state is R.[6] The principal is interested in agents taking action R regardless of what the state of the world is. These payoffs are represented in Table 1, where the two numbers in each cell represent the payoffs to the principal and to the agent, respectively, for the state of nature and agent action combination corresponding to that cell. From the table, we can see that an agent would take action R if her belief that the true state is R is at least equal to $\frac{1-b}{2}$. Equivalently, she would take action S if her belief that the true state is S is at least equal to $\frac{1+b}{2}$.

News Agents receive news over time in the form of signals, where $s_t^i \in \{S, R\}$ is the signal that agent i receives at time t. News is either organic or strategic. Organic news is informative and is generated from a process that is correlated with the state of the world, and we assume that the probability that s_t^i correctly represents that state is strictly greater than $1/2$. However, a principal may choose specific agents in the network and jam their news process by periodically sending them message $\hat{y} \in \{S, R\}$ that corresponds to the state that he would like them to believe. Importantly, agents who are targeted by the principal do not know that they are targeted and cannot tell whether a signal they are receiving is organic or strategic. This is akin to agents scrolling through their news feed and seeing both organic and strategic stories without knowing which stories are which.

[5]For completeness, we include the full technical details in Appendix A.1. These details (e.g. arrival rates of news, explicit form of DeGroot updating, etc.) are not pertinent for our goals but we include them in the appendix for the interested reader.

[6]For example, as in Mostagir et al. (2019), they want to learn whether a particular vaccine is safe (state of the world $y = S$) so that they vaccinate (take action S) or whether a vaccine is risky (state of the world $y = R$) so that they would choose to avoid vaccination (take action R).

Manipulation The principal picks an influence strategy $x_i \in \{0, 1\}$ for each agent i in the network, where $x_i = 1$ indicates that agent i is targeted by the principal, so that some of the news she receives is strategic. The principal may play any influence strategy $\mathbf{x} \equiv \{x_i\}_{i=1}^n$ over the network, and incurs an upfront investment cost $\varepsilon > 0$ for each agent with $x_i = 1$, thus the utility of the principal is the number of agents taking action R less total investment cost.

Agent i is manipulated if she would figure out the correct state in the absence of interference from the principal (i.e. when $\mathbf{x} = \mathbf{0}$), but would mislearn the state when the principal has a profitable network strategy $\mathbf{x} \neq \mathbf{0}$. Notice that the agent does not have to be directly targeted by the principal to be manipulated.

A network is impervious to manipulation if no agents can be manipulated, i.e. if there is no profitable strategy for the principal that results in any agent mislearning the true state.

Learning Society consists of two types of agents. Bayesian agents always learn the state of the world correctly and become stubborn agents as in Theorem 1 of Mostagir et al. (2019). We provide conditions in Appendix A.3 that guarantee that this result extends to the random network domain.

DeGroot agents follow the model of Jadbabaie et al. (2012), with two components that go into their learning and belief update process. The first component comes from the DeGroot agent's own personal experience, which is a Bayesian update on the true state *taking the information received (personally) at face value*. Given a vector of signals $\mathbf{s}_{i,t}$, we denote this personal experience by $\text{BU}(\mathbf{s}_{i,t})$. The second comes from a linear aggregation of the opinions from friends. These components are detailed in Appendix A.1. In general, for periods of new information $\tau = 1, 2, \ldots$, we assume DeGroot agent i updates her beliefs $\pi_{i,\tau+1}$ at time $\tau + 1$ as follows:

$$\pi_{i,\tau+1} = \theta_i \cdot \text{BU}(\mathbf{s}_{i,\tau}) + \sum_{i=1}^{n} \alpha_{ij} \pi_{j,\tau}$$

where θ_i is the weight that agent i places on her own Bayesian update and $\theta_i + \sum_{j=1}^{n} \alpha_{ij} = 1$.

DeGroot Centrality Determining whether an agent is manipulated is equivalent to computing her limit belief of the incorrect state and checking whether the belief is higher than the cutoff obtained from the payoff table. One key contribution in Mostagir et al. (2019) is *DeGroot Centrality* (DC), which is a generalization of Katz-Bonacich centrality and can be computed in the same way using weighted walks (see Jackson (2010), Figures 1 and 2 in this paper, and the expression for limit beliefs in Appendix A.1). DeGroot Centrality captures how the network structure propagates the principal's injected signals to any specific agent; it corresponds to how much influence other DeGroots (who receive these signals) have on agent i's own belief, and under appropriate normalization, DC is exactly equal to that agent's belief of the incorrect state. This means that agents with high DC are those most vulnerable to manipulation, while the closer a DeGroot agent is to a Bayesian agent, the lower their DC is and therefore the more resistant to manipulation they become.

Based on the previous paragraph, characterizing manipulation in deterministic networks becomes an exercise in computing the DC of agents. With the network forming randomly however, there is no easy way to determine which agents will have high DC and which agents to target for intervention. The following section shows that under some regularity conditions, we can overcome this difficulty by reducing the study of the random network to studying the average network instead.

3 Random Networks: Model and Theory

Models of homophily are traditionally built on top of random network formation models (see, for e.g. Golub and Jackson (2012). In this section, we introduce the theoretical foundations for our random network model and give our main technical result in Theorem 1, which allows us to study the role of homophily in learning and manipulation. We start by defining some notation for a random network model on n agents. This random network is specified by a matrix $\boldsymbol{\rho}$ of link probabilities, where

$$\boldsymbol{\rho} = \begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{pmatrix}$$

with the probability that agent i listens to agent j equal to ρ_{ij}. We let $a_{ij} \in \{0, 1\}$ denote whether a link from i to j is realized. Take χ to be any sequence of realized links $(a_{i_1 j_1}, a_{i_2 j_2}, \cdots, a_{i_k j_k})$, and moreover let $\boldsymbol{\chi}_{-ij}$ denote the set of χ where both a_{ij} and a_{ji} do not appear. For additional simplification, we impose:

Assumption 1. For every i, j, the conditional probability of link formation satisfies $\mathbb{P}[a_{ij} = 1|\chi] = \rho_{ij}$ and $\mathbb{P}[a_{ij} = 1|a_{ji} = 1] = 1$, for all $\chi \in \boldsymbol{\chi}_{-ij}$.

In other words, (i) the probability of a link $i_1 \rightleftarrows j_1$ forming does not depend on whether $i_2 \rightleftarrows j_2$ forms, unless $i_1 = i_2$ *and* $j_1 = j_2$, and (ii) the link $i \to j$ exists if and only if $j \to i$ also exists (and hence we write $i \rightleftarrows j$). While the former assumption is made for simplification, we point out that it captures a broad range of networks seen in practice. For instance, it does not rule out highly-clustered networks because some subset of agents may be much more likely to connect to each other than any other agents. On the other hand, the latter assumption avoids the possibility of "informational sinks," whereby an agent is influenced by her environment but has no influence herself, and therefore social learning has no impact on the evolution of her beliefs.

We consider two network objects: (i) the realized network $\tilde{\mathbf{G}}$, and (ii) the "average" network $\bar{\mathbf{G}}$. In the realized network $\tilde{\mathbf{G}}$, we assume DeGroot agents are influenced equally by all neighbors, so the influence weights satisfy (recall that θ_i is the weight that agent i places on her own Bayesian update):

$$\tilde{\alpha}_{ij} = \begin{cases} (1-\theta_i)/d_i, & \text{if } a_{ij} = 1 \\ 0, & \text{otherwise} \end{cases}$$

where $d_i = \sum_{j=1}^{n} a_{ij}$ is the realized degree of agent i. If $d_i = 0$, then we set $\tilde{\alpha}_{ii} = 1 - \theta_i$ and $\tilde{\alpha}_{ij} = 0$ for all $i \neq j$. As mentioned in Section 2, Bayesian agents eventually are uninfluenced by additional social interaction, and instead can be treated as "stubborn agents" under some mild technical assumptions detailed in Appendix A.3.

On the other hand, in the expected network $\bar{\mathbf{G}}$, expected DeGroot influence weights are given by:

$$\bar{\alpha}_{ij} = (1-\theta_i)\rho_{ij}/\bar{d}_i$$

where $\bar{d}_i = \sum_{j=1}^{n} \rho_{ij}$, which is *expected degree* of agent i.[7] As before, if $\bar{d}_i = 0$, then we set $\bar{\alpha}_{ii} = 1 - \theta_i$ and $\bar{\alpha}_{ij} = 0$ for all $i \neq j$. In both the realized and expected networks, Bayesian agents can be modeled as agents with $\theta = 1$ and whose signals cannot be jammed by the principal.

3.1 Deterministic Reduction

Consider a sequence of growing societies $\mathcal{S}_n$ each with n agents. The main connection we develop in this section is between a growing sequence of random networks $\tilde{\mathbf{G}}_n$ and the corresponding "average" networks given by $\bar{\mathbf{G}}_n$. Agent n is born in society $\mathcal{S}_n$ with persistent sophistication type $\tau_n \in \{B, D\}$.[8] Moreover, agent i in society $\mathcal{S}_n$ has personal-experience weight $\theta_i^{(n)} \in (0, 1)$, and for simplicity, we write $\boldsymbol{\theta}^{(n)}$ as the vector of θ's when the population is of size n. Finally, we make the following assumption:

[7]Note that $\bar{\alpha}_{ij}$ is not technically the expectation of α_{ij} for finite n, but these two expressions are shown to be consistent as $n \to \infty$. Moreover, it is more natural to think of $\bar{\alpha}_{ij}$ as the expected influence than a literal expectation over α_{ij}.

[8]All of our results hold regardless of what the sequence τ_n actually is. The types B and D stand for Bayesian and DeGroot, respectively.

Assumption 2. Consider the sequence of vectors $\boldsymbol{\theta} \equiv \{\boldsymbol{\theta}^{(n)}\}_{n=1}^{\infty}$ along with the expected degree matrix $\bar{\mathbf{D}}_n$ and the link probability matrix $\boldsymbol{\rho}_n$. These satisfy the following two conditions:

(a) The *Laplacian matrix*, $\bar{\mathbf{D}}_n - \boldsymbol{\rho}_n$ has its second-smallest eigenvalue bounded away from 0;

(b) $\boldsymbol{\theta}^{(n)}$ is uniformly bounded away from $\mathbf{1}$: there exists $\eta < 1$ and N such that for all $n \geq N$, $\boldsymbol{\theta}^{(n)} \leq \eta\mathbf{1}$ component-wise.

The first condition requires the Laplacian matrix of the realized network not have multiple zero eigenvalues, as this would imply the network is not connected. The second condition rules out situations where agents become increasingly more reliant on their own news, to the point of almost entirely ignoring social influence. Together, these conditions ensure the social learning dynamics are ergodic and non-vanishing.

Each society $\mathcal{S}_n$ comes with its own random network generation process given by:

$$\boldsymbol{\rho}_n = \begin{pmatrix} \rho_{11}(n) & \rho_{12}(n) & \dots & \rho_{1n}(n) \\ \rho_{21}(n) & \rho_{22}(n) & \dots & \rho_{2n}(n) \\ \dots & \dots & \dots & \dots \\ \rho_{n1}(n) & \rho_{n2}(n) & \dots & \rho_{nn}(n) \end{pmatrix}$$

We denote the adjacency matrix of the (realized) random network as $\tilde{\mathcal{A}}_n = \{a_{ij}\}_{i,j}^{n}$ in society $\mathcal{S}_n$. It is easy to see that the expected adjacency matrix is given simply by the matrix of link probabilities in society $\mathcal{S}_n$, i.e. $\bar{\mathcal{A}}_n = \boldsymbol{\rho}_n$. Note by Assumption 1, both $\tilde{\mathcal{A}}_n$ and $\bar{\mathcal{A}}_n$ are symmetric. There are two regularity assumptions we impose on $\boldsymbol{\rho}_n$. These are closely related to the conditions needed in Dasaratha (2019) that we adapt to out setting to account for the two facts that (i) DeGroot influences are formed using the *normalized* adjacency matrix instead of the adjacency matrix itself and, (ii) the mixed-learning environment with both Bayesian and DeGroot agents complicates the belief dynamics.

Our first condition requires that agents' degrees grow at a sufficiently fast rate. This is important because in very sparse networks, individual realizations of links have a significant effect on centrality (see Example 4 in Appendix B). For example, when the expected degree is bounded above as $n \to \infty$, then any individual link represents a non-vanishing contribution toward an agent's DeGroot centrality. Therefore, the realization of this link will be of crucial importance in determining the agent's susceptibility to manipulation.

Definition 1 (Expected Degrees)**.** We say that expected adjacency matrix $\bar{\mathcal{A}}_n$ satisfies the *expected-degrees* condition if $\lim_{n\to\infty} \min_{i\in\mathcal{S}_n} \bar{d}_i^{(n)} / \log n = \infty$.

In other words, the expected degrees condition requires that as the society $\mathcal{S}_n$ grows, *all* agents in the society have expected degrees which are uniformly growing with $\log n$, regardless of the time of their birth.[9] Note that this is a stronger condition than just requiring every agent in the network to have an expected degree that grows strictly faster than $\log n$. The difference between the two conditions may be subtle, but without the stronger condition, there is a significant chance that the realized network is susceptible to manipulation, even though the average network $\bar{\mathbf{G}}_n$ itself never is. For the interested reader, this difference is illustrated in Example 4 given in Appendix B.

Secondly, the society needs to be somewhat homogeneous in the sense that DeGroot agents weigh their own experiences roughly in the same proportion to other DeGroots:

Definition 2 (Normal Society)**.** We say that $\boldsymbol{\theta}$ satisfies the *normal society* condition if there exists a constant $\nu < \infty$ such that for all DeGroot agents i, j:

$$\limsup_{n\to\infty} \frac{\theta_i^{(n)}}{\theta_j^{(n)}} \leq \nu$$

As a special case, the "normal society" condition is always satisfied when all DeGroots have the exact same θ. The analysis of manipulation is most interesting in this case because a non-normal society will typically admit some degree of manipulation. In particular, agents whose $\theta_i^{(n)}$ decay more quickly compared to other agents will be highly influenced by the experiences of their peers, and hence more susceptible to indirect influence by the principal. Again, interested readers can refer to Example 5 in Appendix B to see how, in the absence of the normal society condition, two networks that occur with equal probabilities can have different properties as $n \to 0$.

Under the above regularity conditions, we get the following reduction to deterministic network analysis:

Theorem 1. *Suppose Assumptions 1 and 2 hold, the sequence of* $\boldsymbol{\rho}_n$ *satisfies the expected-degree condition, and* $\boldsymbol{\theta}$ *is a normal society. Then, with high probability, the random network* $\tilde{\mathbf{G}}_n$ *is impervious (resp. susceptible) if and only if* $\bar{\mathbf{G}}_n$ *is impervious (resp. susceptible).*

Theorem 1 states that under the conditions above, it suffices to consider $\bar{\mathbf{G}}_n$ instead of $\tilde{\mathbf{G}}_n$ when looking at manipulation in the large network limit. This theorem offers a major technical simplification: we can apply the DeGroot centrality analysis from Mostagir et al. (2019)

[9]Formally, Definition 1 can be replaced with the "uniformity requirement" that $\exists \beta_n$ such that $d_i^{(n)} \geq \beta_n \log n$ for all $i \leq n$ with $\lim_{n\to\infty} \beta_n = \infty$.

as if the expected networks are (with high probability) the randomly-generated networks themselves. It also proves that network-imperviousness is a 0-1 property: as the network grows larger, the probability a random network is impervious converges to either zero or one (e.g., unlike in Example 4 in Appendix B, where it alternates between zero and one as n increases).

4 Homophily, Inequality, and the Spread of Misinformation

We now apply Theorem 1 to analyze a class of networks known as *stochastic block* networks. This is a class of networks that was introduced in Holland et al. (1983) and is the focus of the study of homophily in Golub and Jackson (2012). In stochastic block networks, agents interact in well-connected communities, with few links between communities. We differentiate between two cases of interest. In *weakly assortative* networks, agents are more likely to be linked to agents within their community, but when they reach out to agents outside that community, they are equally likely to connect to agents in any other community. In that sense, homophily has a flat hierarchy. This model turns out to be rather accurate in describing friendship and communication patterns, as documented in, e.g. Marsden (1987). We show in Theorem 2 that learning is sensitive to the amount of homophily in society and that this effect is not monotone. We then identify special circumstances in Theorem 3 and show that under these circumstances, the role of homophily resembles the results in the literature, i.e. more homophily impedes learning, but that this is generally not true when the conditions for the special case do not hold.

Finally, we introduce *strongly assortative* networks, which are communities ordered by similarity, with agents in neighboring communities more likely to be linked than agents in communities that are farther apart. This model captures the more hierarchical structure that is sometimes observed in society. While this is a natural homophily model, we are not aware of any literature that studies it compared to the much stronger focus on weakly-assortative networks. We compare the effects of weak and strong homophily on the spread of misinformation in Proposition 2.

4.1 Weak Homophily

Consider k "islands" (an equivalence class on all agents) where each island has a proportionate amount of the population $\{s_1, \ldots, s_k\}$ with $s_1 + \ldots + s_k = 1$. The vector $\mathbf{m} = (m_1, ..., m_k)$ denotes the number of Bayesian agents on each island.[10] We assume that if agents i and j are on the same island, then $p_{ij} = p_s$, and if i and j are on different islands, then $p_{ij} = p_d$, with $p_s > p_d$. This is the *weakly assortative* homophily model referred to earlier. The key feature is that

[10]We use the words "island" and "community" interchangeably.

within each island, agents are more likely to be connected, but there is no differentiation when communicating across islands. For simplicity, we will assume throughout that $\theta_i^{(n)} = \mathbf{1}/n$ for all DeGroot agents i, so as to isolate the effect of homophily on manipulation.[11] The corresponding mean adjacency matrix for the weak homophily model is given by:

$$\bar{\mathcal{A}}_n^{WH} = \begin{pmatrix} p_s \mathbf{1}_{ns_1 \times ns_1} & p_d \mathbf{1}_{ns_1 \times ns_2} & \cdots & p_d \mathbf{1}_{ns_1 \times ns_k} \\ p_d \mathbf{1}_{ns_2 \times ns_1} & p_s \mathbf{1}_{ns_2 \times ns_2} & \cdots & p_d \mathbf{1}_{ns_2 \times ns_k} \\ \cdots & \cdots & \cdots & \cdots \\ p_d \mathbf{1}_{ns_k \times ns_1} & p_d \mathbf{1}_{ns_k \times ns_2} & \cdots & p_s \mathbf{1}_{ns_k \times ns_k} \end{pmatrix}$$

4.1.1 General Methods

One can easily check that the weak homophily model satisfies all the regularity conditions in Section 3, and so we can use Theorem 1 and work with deterministic systems. Here, we will introduce the general methodology for determining whether a population (or community) whose structure is randomly drawn from the weak homophily model is susceptible to manipulation. We provide an upper bound on the extent of incorrect beliefs (recall that these beliefs are equivalent to DeGroot centralities of the agents) in the population by first assuming the principal targets *all* DeGroot agents, which can be generalized to arbitrary principal strategies, albeit with more complicated formulations. DeGroot centralities can be computed in a similar way to other centrality measures (e.g., Bonacich centrality), by counting the sum of weighted walks to Bayesian agents. This is highlighted in Figure 1 and an analytical procedure is illustrated in Figure 2.

Computing DeGroot centralities is easiest by considering the linear recursive formulation of weighted walks, as in Figure 2 and noting that, by symmetry, the walks on each island are identical. Suppose m_ℓ is the number of Bayesian agents on island ℓ and $m_{-\ell}$ is the number of Bayesian agents on all other islands besides ℓ. Denote by w_ℓ the sum of weighted walks to a Bayesian for an agent i living on island ℓ. Then,

$$\frac{w_\ell}{1-\theta_i^{(n)}} = \frac{p_s m_\ell + p_d m_{-\ell}}{p_s s_\ell n + p_d (1-s_\ell) n} + \frac{p_s (s_\ell n - m_\ell)}{p_s s_\ell n + p_d (1-s_\ell) n} w_\ell + \sum_{\ell' \neq \ell} \frac{p_d (s_{\ell'} n - m_{\ell'})}{p_s s_\ell n + p_d (1-s_\ell) n} w_{\ell'}$$

[11]If we let $\theta_i^{(n)}$ vary with say, the size of agent i's neighborhood, then we are simultaneously measuring the effects from changing the homophily structure and the experience weights. An increase in p_s, while increasing homophily, also decreases θ and may decrease manipulation simply by virtue of allowing the Bayesians to communicate more effectively.

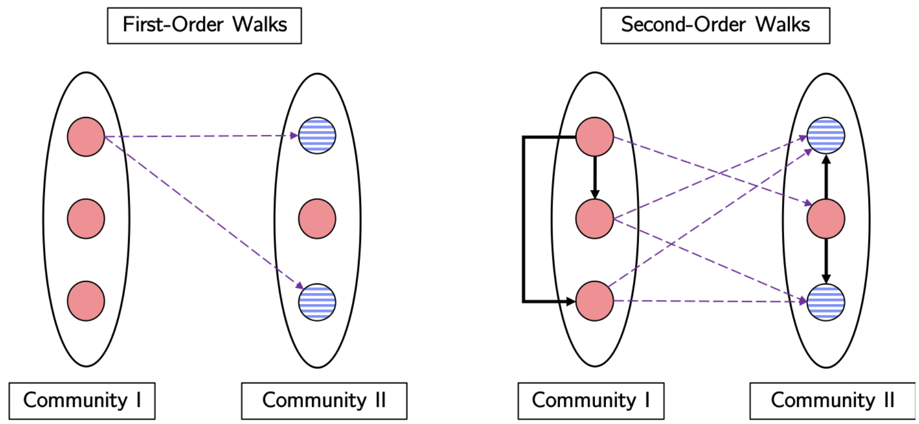


Figure 1. An illustration of computing weighted walks to Bayesians agents. Solid circles are DeGroot agents and shaded circles are Bayesians. Solid lines represent higher weights "within-community links" than dotted lines. Consider the top-left agent, and for each walk, multiply the weights of the links along the walk. The figure on the left shows a first-order walk, i.e. a walk of length 1, which consists of the link directly connecting that agent to a Bayesian agent. The second-order walk displayed on the right consists of walks of length 2, so that there is a link to another DeGroot agent who is linked to a Bayesian agent, and the weight of that walk is the product of the two link weights and so on. Total weighted walks is the sum over all orders (i.e., walk lengths) of walks $1, 2, \ldots$.

In particular, this admits a linear matrix equation with a closed-form solution:

$$\mathbf{I}_\theta^{(n)}\mathbf{w} = \boldsymbol{a} + \boldsymbol{B}\mathbf{w} \Longrightarrow \mathbf{w} = (\mathbf{I}_\theta^{(n)} - \boldsymbol{B})^{-1}\boldsymbol{a} \tag{1}$$

where

$$\mathbf{I}_\theta^{(n)} = \begin{pmatrix} \frac{1}{1-\theta_1^{(n)}} & 0 & \ldots & 0 \\ 0 & \frac{1}{1-\theta_2^{(n)}} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \frac{1}{1-\theta_k^{(n)}} \end{pmatrix}$$

which implies that $1 - w_\ell$ for agent i living on island ℓ is equal to her belief of the false state when the principal sends fake signals to all agents in the network. Using this calculation, it is possible to provide sufficiency conditions on the exact parameters (i.e., $p_s, p_d, \{m_\ell\}_{\ell=1}^k, \{s_\ell\}_{\ell=1}^k$) so that the network is impervious (by basically showing that the DC of all agents is less than the threshold $\frac{1-b}{2}$). In general, for tight (i.e. necessary and sufficient) conditions, one must consider more general principal strategies, as we will demonstrate in later sections.[12] The next example illustrates the methodology described above and serves as an entry point for discussing inequal-

[12]As an aside: to prove the network is susceptible, it is always sufficient to prove that some strategy obtains positive profit.

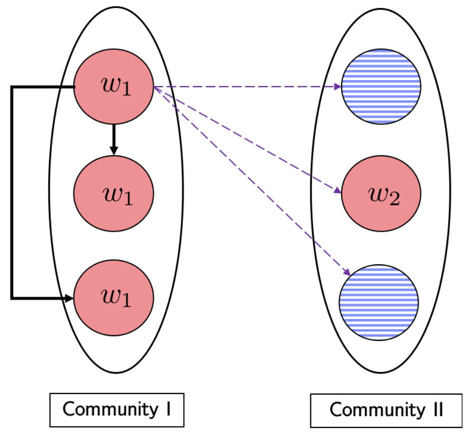


Figure 2. An analytic approach to computing Bayesian walks for the top-left agent (which equals her belief of the true state). Each agent's sum of walks equals a weighted-average of her neighbors' sums of walks.

ity.

Example 1 (2-Island Model)**.** We consider two islands of equal size and explore the degree of manipulation under different homophily structures, similar to Figures 1 and 2. In this example, we measure inequality by the distribution of Bayesian agents across communities. These inequalities may exist, for instance, if sophisticated agents are more concentrated in one island and mostly talk amongst each other.

Suppose that each of the two islands has an equal share of the population, i.e. $s_1 = s_2 = 1/2$. Let there be $n = 100$ agents, of which 10% are Bayesian and let $p_s = 0.9$ and $p_d = 0.1$, so there a significant degree of homophily. We consider two distinct cases: (i) The distributions of Bayesians $\mathbf{m}$ is given by $\mathbf{m} = (5, 5)$, i.e. the Bayesians are split evenly between the two islands, and (ii) the Bayesians all reside on the first island, i.e. $\mathbf{m} = (10, 0)$. Finally, we consider a modification of case (ii) where the Bayesians all reside on one island but the homophily structure is $p_s = 0.5$ and $p_d = 0.5$.

We use Equation (1) to compute the weighted walks. For case (i), we get:

$$\frac{100}{99}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} 0.81 & 0.09 \\ 0.09 & 0.81 \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \Longrightarrow \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0.9083 \\ 0.9083 \end{pmatrix}$$

Therefore, for every DeGroot i, we have that her belief of the false state is only around 0.0917, and so, for $b \lesssim 0.816$, there is no manipulation for any value of ε (recall that the cutoff belief from Table 1 is given by $(1-b)/2 = \frac{1-0.816}{2} = 0.092$, so that an agent is manipulated only if her belief in the false state falls above this cutoff).

Now for case (ii), assume island 1 has all of the Bayesians, then:

$$\frac{100}{99}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0.18 \\ 0.02 \end{pmatrix} + \begin{pmatrix} 0.72 & 0.1 \\ 0.08 & 0.9 \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \Longrightarrow \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0.9114 \\ 0.8438 \end{pmatrix}$$

Notice that the belief (of the false state) of the agents on the first island has decreased to 0.0886, which implies they cannot be manipulated for this same range of b. On the other hand, for $0.68 < b < 0.82$, agents on the second island can now be manipulated (whenever $\varepsilon < 1/2$). That is, inequality in Bayesian distribution coupled with the homophily structure leads one community to hold beliefs that are farther from the truth because this community has less access to Bayesian agents.

Let us continue with the case of unequal distribution of Bayesians but assume that the homophily structure comes from an Erdos-Renyi model with $p_s = p_d = 0.5$, then:

$$\frac{100}{99}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} 0.4 & 0.5 \\ 0.4 & 0.5 \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \Longrightarrow \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0.9083 \\ 0.9083 \end{pmatrix}$$

which gives the same result as case (i), i.e. society as a whole is impervious because there is no longer inequality in terms of access to Bayesian agents. Therefore, extreme degrees of homophily can be seen to increase the inequality of DeGroot centralities (and thus beliefs of the incorrect state), which in turn decreases the imperviousness of the network.

In summary, when ε is small, homophily exhibits a phase transition whereby increasing homophily eventually transitions the network from impervious to susceptible. This is largely consistent with the social learning literature that studies homophily and its (negative) impacts on information aggregation, but as we discuss shortly, this structure is only applicable under special circumstances. □

The aforementioned example highlights how increasing homophily can generate informational inequalities between groups with different compositions of agent sophistication types. Importantly, inequality arises as a combination of uneven distribution of Bayesian agents between communities and a homophily structure that impedes one community from being able to draw some of their information from those Bayesian agents. The example above can be strengthened by performing basic comparative statics on homophily. Towards this end, we parametrize a random network $\tilde{\mathbf{A}}_n$ by the homophily model $(p_s, p_d, \mathbf{m})$ from which it is drawn and operationalize inequality in the following definition.

Definition 3. (Inequality) For two random networks $\tilde{\mathbf{A}}_n$ and $\tilde{\mathbf{A}}'_n$, we say $\tilde{\mathbf{A}}_n$ exhibits *less inequal-*

ity than $\tilde{\mathbf{A}}'_n$ if the following conditions hold:

(a) There is more communication across islands; namely, the between-group link probabilities satisfy $p_d \geq p'_d$;

(b) There is less communication within islands; namely, the within-group link probabilities satisfy $p_s \leq p'_s$;

(c) The distribution of Bayesian agents across groups is more "equally distributed" for all populations n; formally, $\mathbf{m}'/\mathbf{s}$ is a majorization[13] of $\mathbf{m}/\mathbf{s}$, for all n.[14]

with at least one condition strict.

If network A has less inequality than network A', this suggests two features. First, any agent in network A is more likely to talk to agents outside her own island, relative to network A'. Second, there is less inequality in terms of direct connections to Bayesian agents: any two agents in network A are more likely to have a similar number of weighted connections to Bayesians as compared to network A'. The most equitable distribution of Bayesians occurs when Bayesians are the same constant fraction of the population in every island.

4.1.2 Inequality and Network Homophily

Based on Definition 3, we can decompose inequality into the inequality resulting from *network homophily* and inequality resulting from *Bayesian placement*. To most transparently demonstrate the effects from homophily, we begin by assuming there is one island consisting of only Bayesians, whereas all other islands consist of only DeGroots (i.e., $\mathbf{m} = (ns_1, \mathbf{0})$). This way inequality between two networks can be measured in terms of (un)equal access to Bayesians as a result of different homophily structures. With this understanding, we present our main result, which establishes that homophily has an ambiguous effect on learning:

Theorem 2. *Assume there exists $p_d < p_s$ where the network is impervious. Then there exist cutoffs $0 < \underline{p}_d < p'_d < \bar{p}_d < 1$ such that:*

(i) Whenever $p_d > \bar{p}_d$, the network is impervious;

(ii) If there is manipulation when $p_d < \underline{p}_d$, there is manipulation when $\underline{p}_d < p_d < p'_d$;

[13] A majorization $\mathbf{x}'$ of $\mathbf{x}$ satisfies (i) $\sum_{\ell=1}^{k} x_\ell = \sum_{\ell=1}^{k} x'_\ell$ and (ii) $\sum_{\ell=1}^{\ell^*} x_\ell \geq \sum_{\ell=1}^{\ell^*} x'_\ell$ for all $\ell^* \in \{1, \ldots, k\}$, where the components of $\mathbf{x}$ and $\mathbf{x}'$ are sorted in ascending order (see Marshall et al. (2011)).

[14] An equivalent condition is whether one can transform $\mathbf{m}'$ into $\mathbf{m}$ via a sequence of "Robin Hood" operations: one can recover $\mathbf{m}$ from $\mathbf{m}'$ via a sequence of transferring Bayesians from islands with more Bayesians to those islands with fewer (see Arnold (1987)).

(iii) There exists some (ε, b) such that the network is impervious when $p_d < \underline{p}_d$, but susceptible when $\underline{p}_d < p_d < p'_d$.

This result establishes the *weak homophily* effect. Loosely, Theorem 2 states that an "intermediate" amount of homophily is worse than an extreme amount of homophily, which in turn is worse than no homophily at all. While removing all inequality improves learning, simply reducing inequality in an extremely homophilous society can actually lead to worse learning and manipulation. Theorem 2 can be easily generalized to consider other measures of inequality (i.e., p_s and $\mathbf{m}$) as given in Definition 3.

Underlying the previous result is the fact that social connections have both positive and negative externalities. On one hand, they serve as a transmission mechanism for spreading the (correct) beliefs of the Bayesian agents. However, they also allow the principal to spread misinformation in a less costly way, by only targeting a subset of the agents, and using social forces to manipulate other agents as well. When homophily is strong, the principal cannot use one community to influence another, which can make manipulation costly, and ultimately unprofitable. When homophily is quite weak, all Bayesians work together to spread correct information, despite the presence of any possible misinformation from the principal. Intermediate homophily often acts as the perfect breeding ground for manipulating beliefs.

This result provides a sleek connection to models of contagion in financial networks (see Acemoglu et al. (2015), Babus (2016), Kanak (2017)). Similar to the degree of homophily in our setting, in these models, connections both serve to reduce and exacerbate the propagation of negative forces. On one hand, when a bank's related institutions are in distress, the bank finds itself less well-capitalized and more likely to default. However, when a bank faces an idiosyncratic or temporary problem, it can rely on neighboring institutions to protect it from insolvency. Hence, the stability of a financial network can be subtle, and the effect of increased interconnectivity is typically ambiguous, just as with social learning in the presence of homophily.

4.1.3 Inequality and Bayesian Placement

Theorem 2 shows that homophily typically has a non-monotone effect on learning. This is generally inconsistent with the social learning models where news is not provided by a motivated principal. We next show that the general result from these models (that increased homophily is bad for learning) can be recovered in a special case of our setup where a) the cost of sending jammed signals is low and b) the populations of different islands are of the same size (this is the equal-size islands model from Golub and Jackson (2012)). We then provide two examples that violate each of these conditions to show that they are necessary to generate the special case result, and that in general, the effect of homophily on learning is ambiguous and increased homophily can still help learning. The following comparative static applies to all forms of inequality (network homophily *and* Bayesian placement):

Theorem 3. *Suppose there are k islands of equal size ($s_1 = s_2 = \ldots = s_k = 1/k$). There exists $\bar{\varepsilon} > 0$ such that for all $\varepsilon < \bar{\varepsilon}$, if network $\tilde{\mathbf{A}}$ is susceptible to manipulation, any network $\tilde{\mathbf{A}}'$ with more inequality is also susceptible to manipulation, with high probability.*

If agents increase the likelihood of forming connections outside their own community, this decreases homophily and reduces the level of manipulation possible, even in communities which have few Bayesian agents. When investment costs are low, the principal finds it profitable to jam the signals of all agents if this means he can manipulate even just one. Therefore, it suffices to consider the DeGroot centrality when the principal targets all agents, and in particular look at the island with the highest DeGroot centrality, as their beliefs are "farthest" from the truth. Because decreasing homophily equalizes DeGroot centralities across communities of the same size, this is necessarily beneficial to the community with the most incorrect beliefs.

As we mentioned, there are important caveats to the effect described above:

Sizable costs: when the costs of influence ε are intermediate or large, sizable investments that lead to only a small fraction of manipulated individuals are unprofitable, but by decreasing homophily, these same investments can lead to a much larger manipulated population, thereby making them profitable (similar to Theorem 2).

Different population sizes: when communities consist of different size populations, beliefs are sculpted in more complex ways, and it is possible that an increase in homophily can move all agents' beliefs closer to the truth.

We demonstrate these two effects via Examples 2 and 3 below, which offer contrary evidence to the generally-accepted idea that homophily always hurts learning. Lastly, note that while Theorem 3 states that increased homophily cannot make the network impervious, it does not rule out that such an increase can still *decrease* the number of manipulated agents.

Example 2 (Sizable Costs)**.** Consider the following simulated example. Suppose there are two islands with equal population ($n = 100$, 50 agents on each island), with two Bayesians on one island and none on the other. Suppose ε is slightly less than one and $b = 1/3$ and for simplicity that $p_s = p_d$ so there is plenty of communication and the beliefs on both islands are identical. If the principal exerts effort into sending signals to all agents, their beliefs in the correct state fall below the cutoff $(1 + b)/2 = 2/3$ and so all agents are manipulated. This gives the principal a higher payoff than sending no signals whatsoever, so the network is susceptible. This is shown

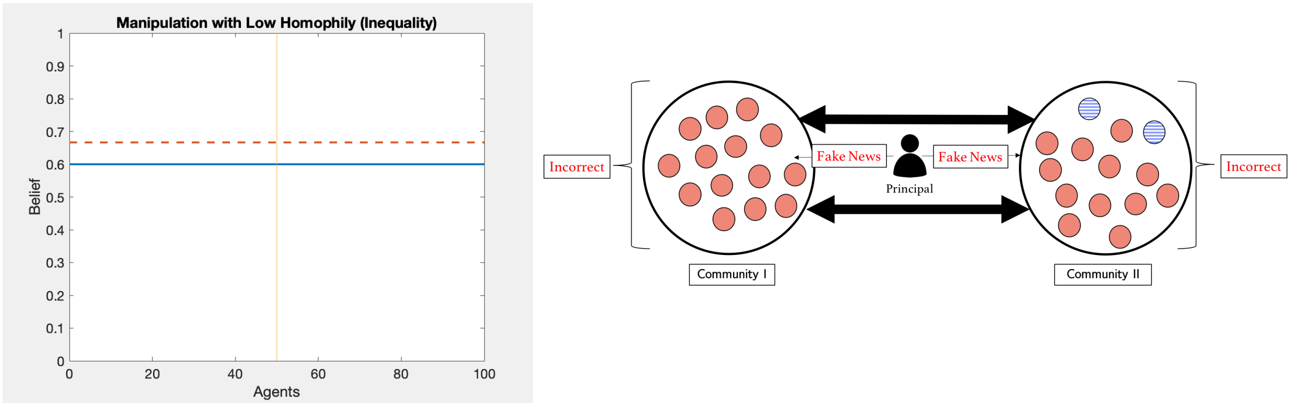


Figure 3. **Low Homophily** - Agent's Beliefs and Principal's Strategy (heavy lines between islands indicate strong connectivity).

in Figure 3, where the principal sends signals to all agents and beliefs in both populations fall below the $2/3$ threshold (all agents are manipulated).

Now suppose homophily increases, so that $p_s = 0.8$ and $p_d = 0.2$. Once again, consider the case where the principal sends jammed signals to all of the agents in the network. After the increase in homophily, the island with two Bayesians moves closer to the truth whereas the island with no Bayesians moves farther away. All agents in the first island now hold beliefs (of the true state) which exceed the cutoff of $(1 + b)/2$, and so are not manipulated (see Figure 4). Given that $\varepsilon > 1/2$ and only half the population is manipulated, this strategy is not profitable for the principal and he would prefer to talk no action rather than send jammed signals to the entire population.

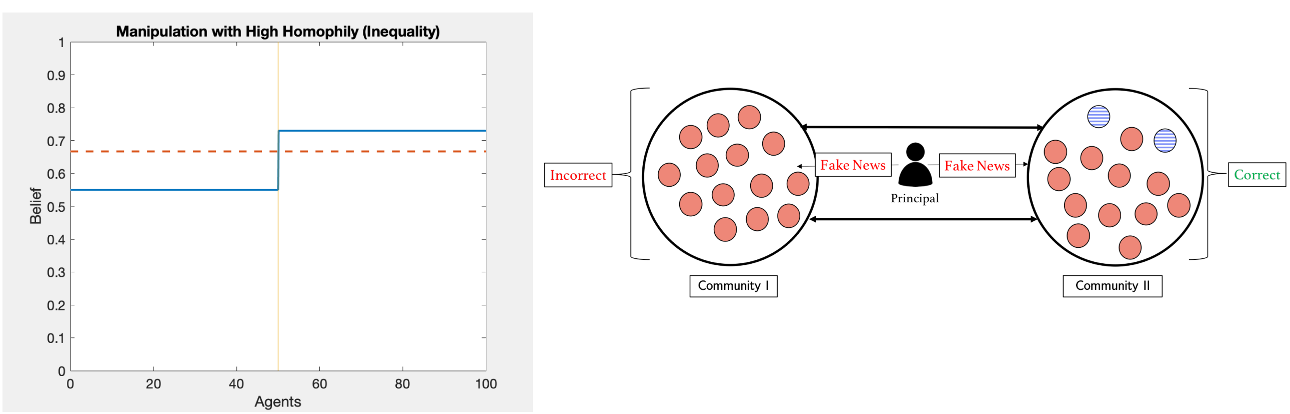


Figure 4. **High Homophily** - Agents' beliefs and principal's strategy. Agents in Community I are manipulated by the strategy is still not profitable for the principal.

It then remains to check that the principal cannot manipulate the island with no Bayesians ("Community I" in the figure) while sending fewer signals to the second island ("Community II"),

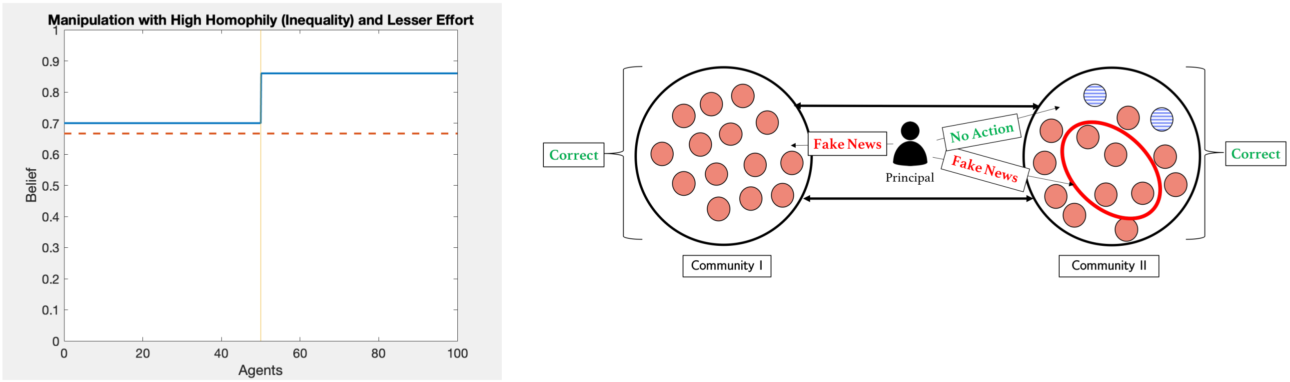


Figure 5. **High Homophily** - Agents' beliefs and principal's strategy. There is no profitable strategy that targets only a subset of agents.

as shown in the right part of Figure 5.[15] To be cost effective, the principal must almost entirely stop sending signals to the second island. However, when doing this, the beliefs of the agents in that island become even closer to the truth, as can be seen in the plot on the left in Figure 5, and these beliefs are strong enough such that they seep through to the first island despite the weak ties, making the principal unable to manipulate agents on that island as well and leading to the entire society becoming impervious. $\square$

Example 3 (Different Population Sizes)**.** We consider one large island (more sophisticated) and one small island (less sophisticated), as pictured in Figure 6. We illustrate how it is possible for decreases in homophily to make the network more susceptible, even when ε is small.

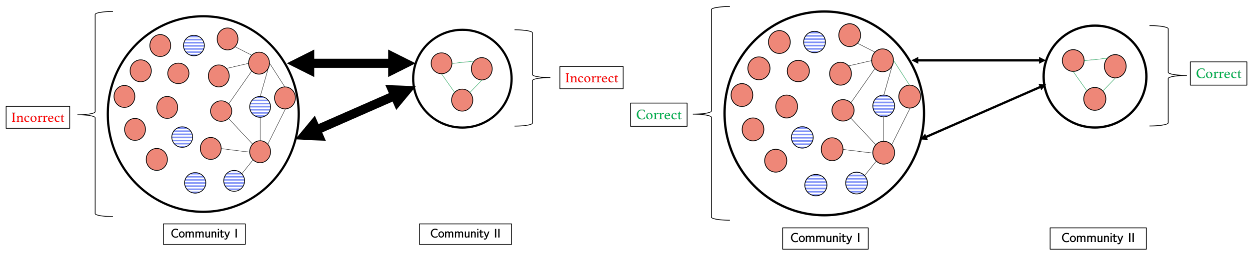
Consider one large island with 80 agents (five of whom are Bayesians) and a small island with 20 agents (only one of whom is Bayesian). Notice that the fraction of Bayesians on the first island is 6.25% whereas it is only 5% on the smaller island. Suppose we originally start with the case of no homophily; that is, let $p_s = p_d = 0.5$. Then we can see that:

$$\frac{100}{99}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0.06 \\ 0.06 \end{pmatrix} + \begin{pmatrix} 0.75 & 0.19 \\ 0.75 & 0.19 \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \Longrightarrow \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0.8559 \\ 0.8559 \end{pmatrix}$$

As expected, the beliefs on the two islands are the same. Now let us increase homophily by setting $p_s = 0.8$ and $p_d = 0.2$, in which case:

$$\frac{100}{99}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0.0618 \\ 0.0562 \end{pmatrix} + \begin{pmatrix} 0.8824 & 0.0559 \\ 0.4687 & 0.4750 \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \Longrightarrow \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0.8584 \\ 0.8571 \end{pmatrix}$$

[15] It can be shown with equal island sizes that the principal does not prefer to stop sending some signals to the second island.



(a) Low Homophily. (b) High Homophily.

Figure 6. Example 3: Homophily can be beneficial when island sizes are unequal.

which shows that the beliefs of *both islands* moved closer to the truth after introducing homophily.[16]

If there is a concentration of Bayesian agents on the large island, decreasing homophily leads to increasing within-group links. This allows the DeGroot agents on the large island to communicate more with their fellow Bayesians, which strengthens the beliefs of agents on that island about the true state. However, because of the size imbalance between the islands, the small island still has plenty of connections to agents in the large island. In the matrix above B_{21}, the amount island 2 draws from island 1's belief, is still substantial as compared to B_{22}, the amount island 2 draws from its own island (whereas on island 1 B_{11} is much larger than B_{12}). Therefore, the second island still draws much of its opinions from those agents on island 1, which in turn moves the beliefs in its community closer to the truth. In this example, *increased homophily leads all islands to move their beliefs closer to truth.* One can show that the exact opposite happens when most Bayesian agents reside on the small island: increasing homophily leads both communities (including the small one) to move further away from the truth (since agents on that island still absorb the misinformation persistent on the larger island), thereby making manipulation easier in both communities. $\square$

As a final interesting point, we note that the beliefs of an agent on island ℓ do not necessarily move closer to the truth when a Bayesian agent is moved from some other island to ℓ. Thus, decreasing inequality can lead to a *Pareto-improvement* in the (correct) beliefs of all agents in society. This occurs because of the *decreasing marginal benefit* from adding Bayesians to a given island. This is because of the following opposing forces: moving a Bayesian to an island (call it island 1) provides a direct influence towards truth on that island. At the same time, the island from which that Bayesian was removed (island 2) is now more prone to misinformation, and this misinformation propagates back and impacts the beliefs of agents in island 1. For instance,

[16]Note that while this difference is small, for different values of θ, the effect can be much more exaggerated. We emphasize the directional change as opposed to its magnitude for this reason.

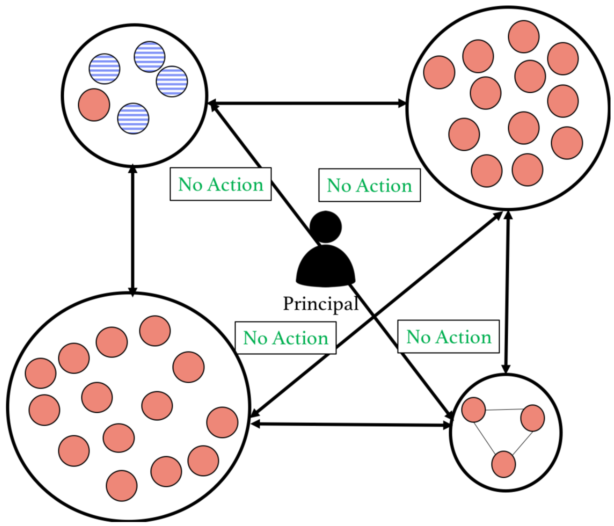


Figure 7. An illustration of Proposition 1: under weak homophily, a constant number of Bayesian agents is enough to prevent manipulation anywhere in the network.

with two communities of equal size, if there are 10 Bayesian agents, the beliefs of agents in *both communities* are closer to the truth when there are 8 Bayesians on island 1 and 2 Bayesians on island 2 as compared to 9 Bayesians on island 1 and a single Bayesian on island 2. So even though agents in island 1 talk directly to more Bayesians when there are 9 of them, the more misinformed beliefs of island 2 make the entire population worse off.

We conclude this section by giving a lower bound on the number of Bayesian agents needed to achieve imperviousness, regardless of the position of these agents or the homophily structure of the network.

Proposition 1. *Fix* (p_s, p_d) *and* (ε, b)*. There exists* $c > 1$ *such that as* $n \to \infty$*, if there are* $m = \Omega((c - p_s + p_d)^{-1})$ *Bayesian agents anywhere, then any weak homophily network* $\tilde{\mathbf{A}}_n$ *with communities* $\{s_\ell\}_{\ell=1}^{k}$ *is impervious with high probability. In particular, if the number of Bayesian agents grows unboundedly with* n*, any weak homophily network is impervious with high probability.*

In summary, in the weakly-assortative model, agents are more likely to form connections within their own community, but all connections outside of the community are equally likely. For general weak homophily networks, the bound on the number of Bayesians needed for imperviousness does best when there is *little* homophily. This reinforces the result of Theorem 3 and applies even if the communities are of unequal sizes.

4.2 Strong Homophily

Suppose now we introduce a different model of homophily: the *strongly-assortative* model. Most studies of homophily in the literature is of the stochastic block weak homophily type; for example, the equal-size islands model in Golub and Jackson (2012). Strong homophily is a novel twist on stochastic block networks that provide an apt representation of a society where connections are formed in a hierarchical fashion. The effects of strongly-assortative homophily drastically contrast with the weakly-assortative case studied in the previous section. We assume that each island j has a vector of qualities, $\Lambda_j \in \mathbb{R}^L$. Qualities can capture different variables like education, profession, income, etc. Islands can be sorted according to their similarity to other islands, with the distance metric between islands i and j given by $d(i,j) = ||\Lambda_i - \Lambda_j||_2$. For simplicity, we assume that $L = 1$ (the quality vector is one-dimensional) and link probabilities are given as follows: within-community links are still formed according to probability p_s, similar to the weak homophily model. On the other hand, communities are (strongly) ordered by their Λ_j on a *line topology*. In our model, agents in community i are linked to agents in community $i+1$ or $i-1$ with probability p_d, whereas agents in "farther" communities are linked with probability 0.[17] The corresponding mean adjacency matrix for the strong homophily model is given by:

$$\bar{\mathcal{A}}_n^{SH} = \begin{pmatrix} p_s \mathbf{1}_{ns_1 \times ns_1} & p_d \mathbf{1}_{ns_1 \times ns_2} & \mathbf{0}_{ns_1 \times ns_3} & \cdots & \mathbf{0}_{ns_1 \times ns_k} \\ p_d \mathbf{1}_{ns_2 \times ns_1} & p_s \mathbf{1}_{ns_2 \times ns_2} & p_d \mathbf{1}_{ns_2 \times ns_3} & \cdots & \mathbf{0}_{ns_2 \times ns_k} \\ \mathbf{0}_{ns_3 \times ns_1} & p_d \mathbf{1}_{ns_3 \times ns_2} & p_s \mathbf{1}_{ns_3 \times ns_3} & \cdots & \mathbf{0}_{ns_3 \times ns_k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0}_{ns_k \times ns_1} & \mathbf{0}_{ns_k \times ns_2} & \mathbf{0}_{ns_k \times ns_3} & \cdots & p_s \mathbf{1}_{ns_k \times ns_k} \end{pmatrix}$$

For illustration, we assume that one community has m_n Bayesian agents and all other communities consist of DeGroot agents. We assume that m_n is any superconstant for the following result:

Proposition 2. *Fix (p_s, p_d). There exists (ε, b) such that for every large n, there is a strongly assortative homophily network $\tilde{\mathbf{A}}_n$ with communities $\{s_\ell\}_{\ell=1}^k$ susceptible to manipulation with high probability (as $n \to \infty$). On the other hand, the weak homophily network with the same communities $\{s_\ell\}_{\ell=1}^k$ is impervious with high probability.*

Proposition 2 shows the stark difference between weak and strong homophily. Furthermore, we note that a significant proportion of the agents are manipulated in the presence of strongly-

[17] We can equivalently assume that these link probabilities are positive but decay sufficiently quickly, such as on the order of $\exp(-||\Lambda_i - \Lambda_j||_2)$. For simplicity of exposition and illustration of the effects of our strong assortative property, we simply set the link probabilities to 0.

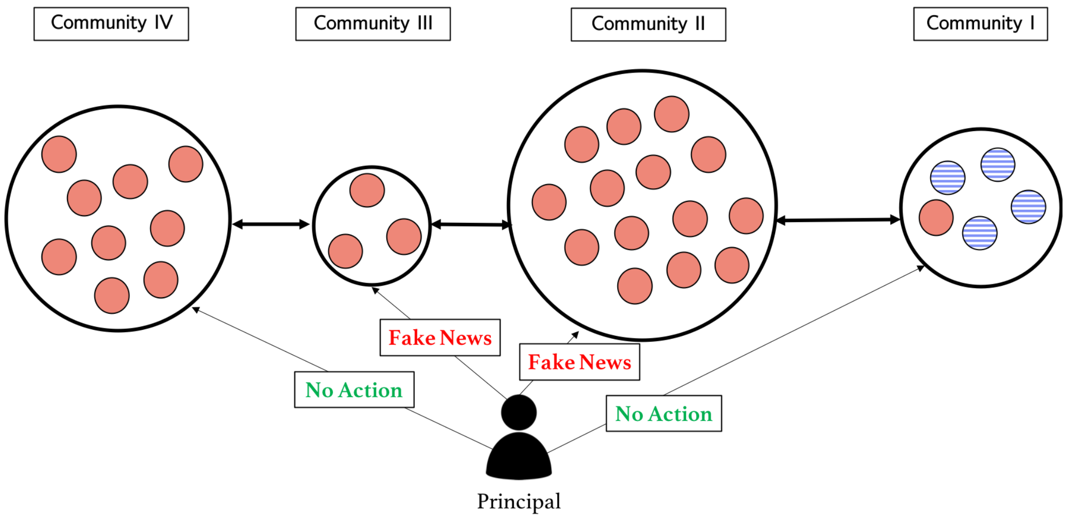


Figure 8. An illustration of Proposition 2: even with many Bayesian agents, strong homophily allows the principal to manipulate many agents in the network. In the figure above, Community IV believes the false state without even being directly exposed to fake news from the principal.

assortative links. The proof of Proposition 2 shows how the principal has a strategy that can manipulate all but $O(\sqrt{n})$ agents under this line topology (see Appendix C). Moreover, we note that increasing strong homophily (by either increasing p_s or decreasing p_d) exacerbates the extent of manipulation. We provide an intuition in what follows.

Because links are undirected in the strong homophily network, agents receiving fake news communicate their experiences both forwards and backwards, which leads to an overall more effective propagation of incorrect beliefs. When communication goes both ways, agents believe the state that is more "local" to their position in the line. This is largely due to an *echo chamber effect*, where the influence from misinformation, as reflected in the agents' beliefs, gets inflated because they fail to recognize their own influence on their neighbors' beliefs. While this echo chamber aids the principal in manipulating more agents, it also makes it more difficult to manipulate agents at the end of the line. In particular, the principal must continue to send fake signals for all but the last $O(\sqrt{n})$ agents. This again is for the same reason as before: if the principal fails to send fake signals to too many communities at the end of the line, their beliefs will reflect mostly the organic news they receive.

Nonetheless, the deep separation between the Bayesians and some DeGroots in the strongly-assortative model creates pervasive manipulation not present in weak homophily, where strong connections within communities and weaker ones between communities do not necessarily lead to manipulation in the large limit. On the other hand, when agents within a certain community are more likely to interact with agents in other communities of a particular kind, then some manipulation will generally be unavoidable.

5 Optimal Interventions

We now discuss the role that a social planner has in shaping the beliefs of society in the presence of the principal. We consider two possible interventions: Bayesian interventions and homophily interventions. In the former, we assume the planner may improve the sophistication type of a subset of agents, perhaps through targeted education. In the latter, the planner may decrease the extent of homophily through efforts to integrate communities (i.e. by increasing p_d). We focus on two types of policies:

(a) *Preventative policy*: Ideally, a planner would like to make the network impervious, so that there is no manipulation at all. A policy is *preventative* if it accomplishes such a goal. Having as many Bayesian agents as given in the bound of Proposition 1 would yield such a policy. Sometimes however, preventative policies may not exist because of budget constraints,[18] which leads to the second type of policy we consider.

(b) *Protective policy*: This policy aims to protect as many agents as possible by minimizing the number manipulated. Recall from Equation (1) that we can write the beliefs of the agents, $\boldsymbol{\pi}$, as:

$$\boldsymbol{\pi}(\mathbf{m}) = (\mathbf{I}_\theta^{(n)} - \mathbf{B}(\mathbf{m}))^{-1}\mathbf{a}$$

where $\mathbf{B}(\mathbf{m})$ is a linear function of the distribution $\mathbf{m}$ of Bayesians across islands. Then the planner solves $\min \ \sum_{i=1}^{n} \mathbb{1}_{\boldsymbol{\pi}_i<(1+b)/2}$, i.e. minimize the number of agents whose beliefs fall below the manipulation cutoff.

If the optimal solution to the minimization problem above is zero, then the protective and preventative policies coincide. In general however, a preventative policy –if it exists– is the unique Pareto-optimal outcome for society, whereas optimal protective policies may be forced to benefit some agents over others. We derive the optimal policies for some special cases, but show through simulated examples that inequality can often be beneficial to society.

5.1 Bayesian Interventions

While up until now we have considered an agent's sophistication type as immutable, we consider the possibility of endowing some agents in the population with Bayesian abilities. Equivalently,

[18] It is possible that there is no preventative policy even with unlimited budget, as we discuss in Section 5.2

since Bayesian agents are stubborn agents who hold correct beliefs, one can think of providing an agent with verifiable information about the state (e.g. through teaching a class or program). We assume that this process is costly and that the planner's budget constraint is of the form $\sum_{i=1}^{n} \mathbb{1}_{\tau_i = B} \leq M$ for some integer number of Bayesians M. Based on our observations in Theorem 3 we note:

Proposition 3. *The preventative policy when ε is sufficiently small is to set $m_\ell = M \cdot s_\ell$ for all islands ℓ (i.e., distribute the Bayesians equally).*

In other words, when sending jammed signals is cheap for the principal, the planner is best-off distributing the M Bayesian agents proportionally to the population on each island. If such placement does not lead to imperviousness, then no preventative policy exists, and the planner should focus on protective policies instead. As a special case, when all islands are exactly the same size and a preventative policy exists, then that policy places the same number of Bayesian agents on each island.

As the cost of sending jammed signals increases, i.e. $\varepsilon \gg 0$, the preventative policy might require an unequal placement of Bayesians across islands – similar to what we have observed in Example 2. Therefore, while Proposition 3 states that equal placement of Bayesian agents is optimal when ϵ is small, for general ε, inequality can be first-best for *all agents* in society, as we show next.

To demonstrate Proposition 3 and the above discussion, consider Figure 9, where ε ranges from small to large on two islands of size 50 each and a budget that is large enough for four Bayesians (in this example, the homophily structure is $p_s = 0.8$ and $p_d = 0.2$). As ε ranges from 0 to 0.5 (small ε region), there is no distribution of Bayesians that admits imperviousness. Proposition 3 allows us to quickly check this by distributing the Bayesians evenly across the two islands and checking if manipulation exists, which it does since the principal has a profitable strategy that sends jammed signals to all agents. As ε becomes slightly higher than 0.5, the most inequitable distribution (4 Bayesians on one island and 0 on the other) or the most equitable distribution $(m_1, m_2) = (2, 2)$ lead to manipulation, even when only some of the agents are targeted by the principal. On the other hand, an unequal distribution of $(m_1, m_2) = (3, 1)$ or $(m_1, m_2) = (1, 3)$ leads to imperviousness. When ε continues to increase, *only the even distribution* $(m_1, m_2) = (2, 2)$ *makes society susceptible*, i.e., $(0, 4)$ and $(4, 0)$ are also impervious, and splitting the Bayesians equally across both islands is the worst distribution for society. Of course, eventually, all distributions are impervious when the cost becomes too prohibitive for the principal to have a profitable strategy. In summary, for the planner, the most equitable distribution $(m_1, m_2) = (2, 2)$ is weakly dominated by $(m_1, m_2) = (3, 1)$ over the *entire* cost range of ε.

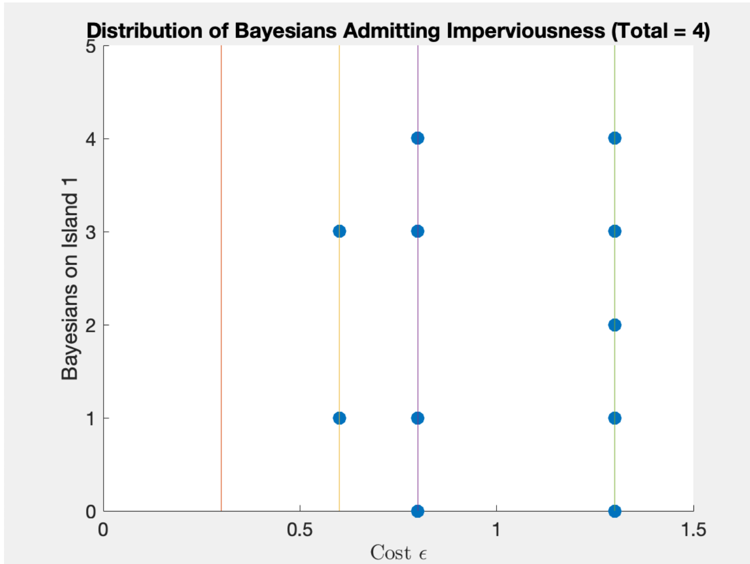


Figure 9. Preventative Policies for Bayesian Placement.

It is also straightforward to see that optimal protective policies might necessitate an uneven number of Bayesians across the islands. If there are two islands, enforcing an equitable distribution might lead to both islands being manipulated, whereas placing all of the Bayesians on one island would protect at least all the agents on this island.

5.2 Homophily Interventions

We now consider a fixed homophily model with parameters (p_s, p_d^o) and Bayesian distribution $\mathbf{m}$. We assume the social planner pays a positive, convex cost $\phi(p_d - p_d^o)$ with $\phi(0) = 0$ to increase (or decrease) connections between islands. As before, we assume the planner has a budget to spend; that is, the planner must satisfy $\phi(p_d - p_d^o) \leq Budget$. Based on our observations in Theorem 2 and Theorem 3 we have:

Proposition 4. *If Bayesians are distributed equally, then $p_d = p_d^o$ is the optimal protective policy. On the other hand, if the budget is large and either: (i) each island is all Bayesian or all DeGroot or (ii) all islands are equal sizes and ε is small, then $p_d = p_s$ is the optimal protective policy.*

When Bayesians are distributed equally (i.e., proportionate to the population), the beliefs of all agents in society are the same regardless of the homophily parameters. Therefore, no intervention is necessary. When all of the Bayesians and DeGroots are segregated (or the islands are equally-sized), but the planner's budget is limited, implementing $p_d = p_s$ may not be feasible. But most importantly, when there is *some inequality* in the distribution of Bayesians, some degree of homophily may actually be desirable, even with a large budget. In Example 2, homophily

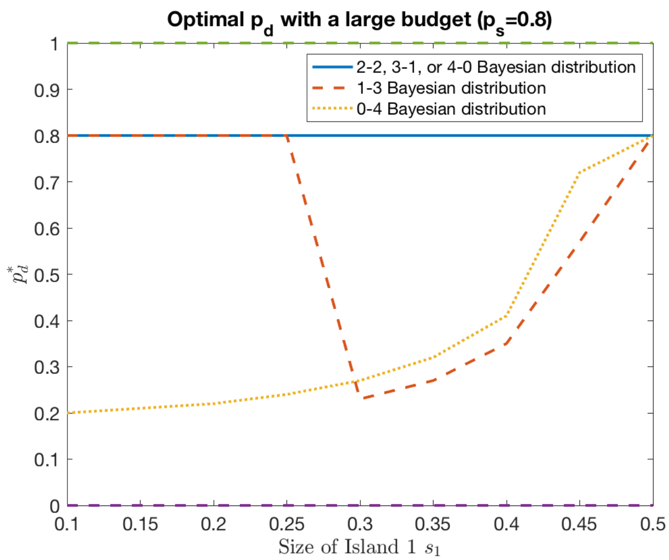


Figure 10. Optimal (Simulated) p_d with $\varepsilon = 0$.

pushes the community with more Bayesians closer to the truth and makes manipulation unprofitable; in Example 3, homophily can move *all islands* closer to the truth. In other words, a homophilous network can balance out the inequality in the Bayesian distribution.

Unlike with Bayesian interventions, the planner may not want to use the entire budget because p_d may have an interior global optimum. In fact, it is possible the planner may even want to use the budget to *decrease* p_d, i.e increase homophily. Setting $p_d^o = p_s$ and given a big enough budget, we see this to be true in Figure 10. Moreover, the optimal p_d is not always p_s when the island sizes are different and possess an inequitable distribution of Bayesian agents. When ε is large, as with Bayesian interventions, homophily can be beneficial even when the island sizes are identical.

To summarize, while sometimes the optimal policies that prevent or minimize the spread of misinformation take the familiar and expected form of equally distributing resources in proportion to community sizes and of encouraging the integration of communities, it is also possible that these policies may sometimes necessitate an unequal distribution of resources/Bayesian agents or require weakening the ties between individual communities.

6 Conclusion

This paper analyzes the role of homophily in social learning when information may be provided to the agents by a strategic actor. This setup resembles many realistic scenarios where an information provider may have their own agenda and exerts effort to influence agents to follow that agenda. We show that in these environments increased homophily can sometimes be beneficial to society in terms of learning the correct state of the world, which is a novel finding that stands in contrast to most of the literature. We also show how a special case of our model recovers the standard results about the negative effects of homophily on learning.

Our main technical result, which reduces the study of stochastic networks into the study of an average 'deterministic' network allows us to capture the best of both worlds: on one hand we have a model rooted in random network theory and stochastic block matrices, which offers a more realistic description of real-world networks and allows us to study a phenomenon like homophily, but on the other hand we can develop insights and policy recommendations that apply to all networks from this class through studying a much simpler topology.

The finding that homophily can help learning in the presence of strategic injection of news into society manifests in ways that are not immediately obvious. Because Bayesian agents know the truth and can help spread it around to their neighbors, then one would imagine that easier access to these agents would always help society as a whole, but as demonstrated in Theorem 2, Examples 2 and 3, and Section 5, this is not always the case. By integrating communities, it is possible that the incorrect beliefs of one community dilute the influence of Bayesian agents within their own community and lead to an overall failure of learning for *all* communities. On the other hand, less integration can result in an increase in costs for injecting misinformation that results in the principal having no profitable manipulation strategy. Another contribution of our paper is the introduction of the strong homophily model, which captures another realistic aspect of social connections –the hierarchical structure that is sometimes present in society– and contrasting the effect of this model with the standard weak homophily model.

Finally, and as we mention in the introduction, homophily is associated with different negative social, educational, and economic outcomes, and hence policies that aim to integrate communities in order to alleviate these outcomes are usually desired. The fact that homophily has an ambiguous effect on learning highlights that one should be careful when designing these policies, and one takeaway from our paper is that reducing homophily should be approached as a multi-criteria optimization problem, because of the simple fact that less homophily and more integration do not necessarily imply an automatic improvement along all metrics of interest, as we demonstrated in this paper in the case of social learning.

Appendix

A Technical Conditions and Model Details

Appendix A.1 provides more technical details about the deterministic model, while Appendices A.2 and A.3 give conditions under which the Bayesian learning results hold in the random network generation model.

A.1 News Generation and Belief Evolution

The following model details are from Mostagir et al. (2019) and are presented here for contextualization of Section 2.

(a) **Organic News**: We assume agents receive organic information about the state y over time. News is generated according to a Poisson process with unknown parameter $\lambda_i > 0$ for each agent i; for simplicity, assume λ_i has atomless support over $(\underline{\lambda}_i, \infty)$ and $\underline{\lambda} > 0$. Let us denote by $(t_1^(i), t_2^{(i)}, \ldots)$ the times at which news occurs for agent i. For all $\tau \in \{1, 2, \ldots\}$, the organic news for agent i generates a signal $s_{t_\tau^{(i)}} \in \{S, R\}$ according to the distribution:

$$\mathbb{P}\left(s_{t_\tau^{(i)}} = S \middle| y = S\right) = \mathbb{P}\left(s_{t_\tau^{(i)}} = R \middle| y = R\right) = p_i \in [1/2, 1)$$

i.e., the signal is correlated with the underlying truth.

(b) **News from Principal**: In addition to the organic news process, there is a principal who may also generate news of his own. At $t = 0$, the principal picks an influence state $\hat{y} \in \{R, S\}$. The principal then picks an influence strategy $x_i \in \{0, 1\}$ for each agent i in the network. If the principal chooses $x_i = 1$, for any agent i, then he (the principal) generates news according to an independent Poisson process with (possibly strategically chosen) intensity λ_i^* which is received by all agents where $x_i = 1$. We assume the principal commits to sending signals at this intensity, which may not exceed some threshold $\bar{\lambda}$.

(c) **News Observations**: Agents are unable to distinguish news sent by the principal or that organically generated. We denote by $\hat{t}_1^{(i)}, \hat{t}_2^{(i)}, \ldots$ the arrival of *all* news, either from organic sources or from the principal, for agent i. At each time $\hat{t}_\tau^{(i)}$, if the news is organic, the agent gets a signal according to the above distribution, whereas if the news is sent from the principal, she gets a signal of $\hat{y}$.

(d) **DeGroot Update**: DeGroots use a simple learning heuristic to update beliefs about the underlying state from other agents. We assume every DeGroot agent believes signals arrive according to a Poisson process and all signals are independent over time with $\mathbb{P}\left(s_{i,\hat{t}_\tau} = y\right) = p_i$ (i.e., takes the news at face value). DeGroot agents form their opinions about the state both through their own experience (i.e., the signals they receive). Given history $h_{i,t} = (s_{i,\hat{t}_1^{(i)}}, s_{i,\hat{t}_2^{(i)}}, \ldots, s_{i,\hat{t}_{\tau_i}^{(i)}})$ up until time t with $\tau_i = \max\{\tau : \hat{t}_\tau^{(i)} \leq t\}$, each agent forms a personal belief about the state according to Bayes' rule. Let $z_{i,t}^S$ and $z_{i,t}^R$ denote the number of S and R signals, respectively, that agent i received by time t; then the DeGroot agent has a direct "personal experience" (note that the $g_{i,t}$ function is the mathematical representation of $\text{BU}(\mathbf{s}_{i,t})$ from Section 2):

$$g_{i,t}(S|h_{i,t}) = \frac{p_i^{z_{i,t}^S}(1-p_i)^{z_{i,t}^R}q}{p_i^{z_{i,t}^S}(1-p_i)^{z_{i,t}^R}q + p_i^{z_{i,t}^R}(1-p_i)^{z_{i,t}^S}(1-q)}$$

As mentioned in Section 2, each DeGroot i then updates her belief for all $k\Delta < t \leq (k+1)\Delta$ according to:

$$\pi_{i,t} = \theta_i g_{i,t}(h_{i,t}) + \sum_{i=1}^{n} \alpha_{ij}\pi_{j,k\Delta}$$

for some weights θ_i, α_{ij} with $\theta_i + \sum_{j=1}^{n} \alpha_{ij} = 1$, and Δ is a time period of short length.

(e) **Bayesian Update**: Bayesian agents know the network $\mathbf{G}$ and the signal structures $\{p_i\}_{i=1}^{n}$. Each Bayesian observes the history of beliefs in her neighborhood $\mathcal{N}_i$ for all time $t' \leq t$. Moreover, the Bayesian is aware the principal may be strategic and has accurate conjectures about the equilibrium (influence) strategy of the principal. At time $t+dt$, the Bayesian agents makes a Bayesian update about the state given her private history of signals and her history of observed neighbor beliefs, forming π_{t+dt}.

(f) **DeGroot Centrality Vector**: To figure out the limit beliefs of the DeGroot agents when the principal targets everyone who is not a Bayesian, denote by $\boldsymbol{\gamma}$ the vector in $\{0,1\}^n$ that designates which agents are targeted by the principal and let $\gamma_i = x_i = 1$ wherever agent i is DeGroot and $\gamma_i = x_i = 0$ everywhere else. DeGroot Centrality, which is equivalent to the belief in the incorrect state in the limit is then given by $\mathcal{D}(\boldsymbol{\gamma}) = (I-A)^{-1}\boldsymbol{\gamma} = \sum_{k=0}^{\infty} A^k \boldsymbol{\gamma}$, where I is the identity and A is the adjacency matrix. DeGroot agent i is manipulated if her belief in the false state is below the cutoff, i.e. if $\mathcal{D}_i(\boldsymbol{\gamma}) < (1-b)/2$.

A.2 Conditions for Bayesian Learning

We assume throughout this paper that Bayesian agents are stubborn agents who hold correct beliefs about the truth (which is Theorem 1 from Mostagir et al. (2019)). In this section, we provide formal conditions for this to follow when Bayesians in fact must learn the state.

Let B be the set of Bayesian agents and D the set of DeGroot agents. Every network consists of (i) a neighborhood for each Bayesian agent i, $\mathcal{N}_i \subset \{1,\ldots,n\}$, and (ii) an influence vector for each DeGroot agent i, $[\mathbf{A}]_i = (\alpha_{i1}, \cdots \alpha_{in})$. In the realized network $\tilde{\mathbf{G}}$, the neighborhood of each Bayesian agent is given simply by her realized connections, $\tilde{\mathcal{N}_i} = \{j | a_{ij} = 1\}$; we assume DeGroot agents are influenced equally by all neighbors, so the influence weights satisfy:

$$\tilde{\alpha}_{ij} = \begin{cases} (1-\theta_i)/d_i, & \text{if } a_{ij} = 1 \\ 0, & \text{otherwise} \end{cases}$$

where $d_i = \sum_{j=1}^{n} a_{ij}$ is the degree of agent i. If $d_i = 0$, then we set $\tilde{\alpha}_{ii} = 1 - \theta_i$ (as before).

Each expected network, $\bar{\mathbf{G}}$, is parametrized by a set of neighborhoods for each Bayesian agent $\{\bar{\mathcal{N}_i}\}_{i\in B}$, which denote the "expected" neighborhoods of the Bayesians, and are taken as given. On the other hand, DeGroot influence weights are given by:

$$\bar{\alpha}_{ij} = (1-\theta_i)\rho_{ij}/\bar{d}_i$$

where $\bar{d}_i = \sum_{j=1}^{n} \rho_{ij}$, which is *expected degree* of agent i, the same as before.

A.3 Reduction to Stubborn Agents

Consider a sequence of growing societies $\mathcal{S}_n$ each with n agents. Agent n is born in society $\mathcal{S}_n$ with sophistication type $\tau_n \in \{B, D\}$, signal intensity $\lambda_n \in \mathbb{R}_+$, and signal strength $p_n \in (1/2, 1)$, which never changes. We make the following assumption:

Assumption 3. Consider the vectors $\boldsymbol{\lambda}_n \equiv \{\lambda_i\}_{i=1}^n, \mathbf{p}_n \equiv \{p_i\}_{i=1}^n$, and the sequence of vectors $\boldsymbol{\theta} \equiv \{\boldsymbol{\theta}^{(n)}\}_{n=1}^\infty$. Then, $\mathbf{p}_n, \boldsymbol{\lambda}_n$ satisfy Assumption 2 from Mostagir et al. (2019) for all n.

Recall that our average network consists of a set of "average" neighborhoods $\{\bar{\mathcal{N}_i}\}_{i\in B}$ for the Bayesians. For example, in the context of Example 4(a), agent 1 forms a link with agent n with probability $1/2$. But in the average network, should we expect agent n to be included in the neighborhood of Bayesian agent 1? Because this question is difficult to answer, we instead take the approach that, for the most part, *the answer does not matter*. For this, we make the additional assumption:

Definition 4. The average Bayesian network $\bar{\mathbf{G}}_n^B$ consists of all agents j who are connected to some Bayesian agent i, and all links between these agents that exist in $\bar{\mathbf{G}}$.

Assumption 4. The directed network of Bayesian agents $\bar{\mathbf{G}}_n^B$ is connected and contains at least one DeGroot agent. Moreover, the weights α_{ij} between any two agents i, j in the realized network is perturbed by some random ϵ_{ij} from a continuous distribution over finite support $F(\cdot)$, with $\lim_{n\to\infty} \epsilon_{ij} \overset{a.s.}{\to} 0$ for all agents i, j (i.e., $F(\cdot)$ converges to $\delta(0)$.

The first condition requires the "average" neighborhoods force the Bayesian agents to be connected to the society of DeGroots. This is a fairly tame condition; for instance, as long as every Bayesian observes one DeGroot the condition is satisfied. The second condition, while more stringent, considers small perturbations to the network weights to guarantee genericity. Without it, there are some special cases where a Bayesian agent might have an identification problem when observing the beliefs of her neighbors. Under both of these assumptions, Bayesian agents may be treated as stubborn agents as the learning horizon $T \to \infty$.

B Random Network Theory Failure Examples

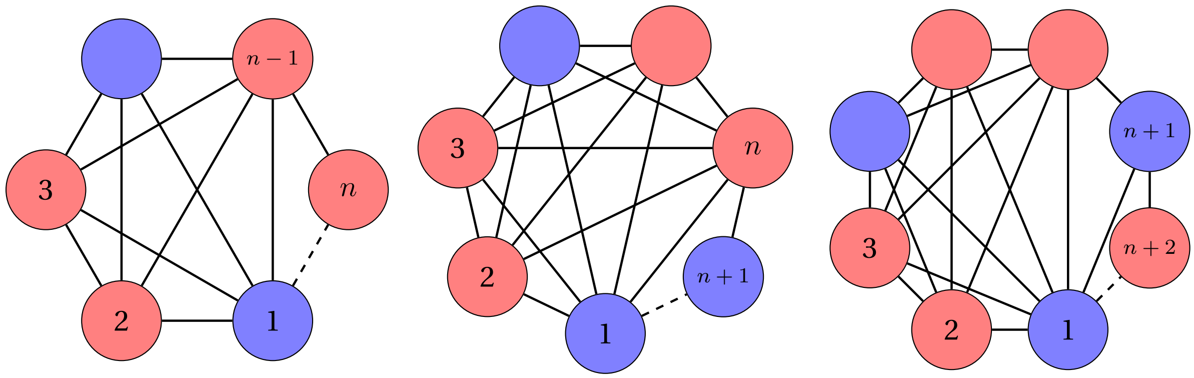
In this section, we provide examples to show that the (Expected Degrees and Normal Society) conditions required for Theorem 1 cannot be dispensed with.

B.1 Non-Uniform (Slow) Degree Growth

Example 4. Suppose the society $\mathcal{S}_n$ has the link probability matrix:

$$\boldsymbol{\rho}_n = \begin{pmatrix} 0 & 1 & \cdots & 1 & 1/2 \\ 1 & 0 & \cdots & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 0 & 1 \\ 1/2 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

with $\theta_i^{(n)} = 1/2$, so each agent weighs her own experience equally with that of all other neighbors. We assume every third agent is Bayesian. The expected network grows as society grows according to Figure 11.

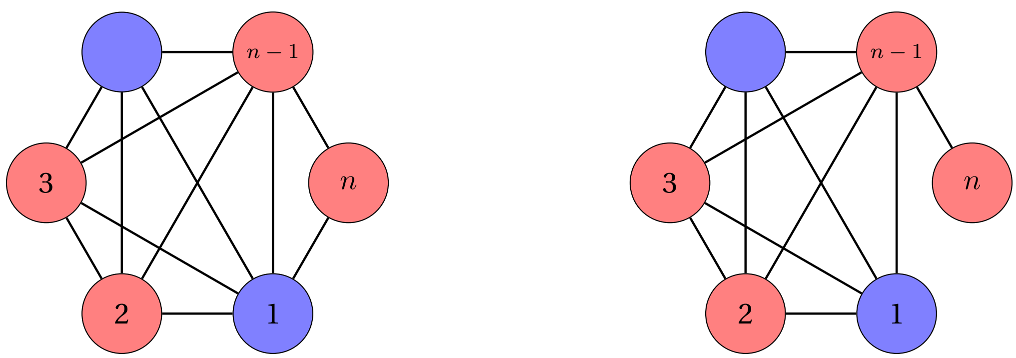


(a) n agents (b) $n+1$ agents (c) $n+2$ agents

Figure 11. The expected network (solid links = weight 1, dashed lines = weight $1/2$).

It can be shown via a walk-counting argument that the DeGroot centrality of all DeGroot agents $2 \leq i \leq (n-1)$ is at most approximately $2/3$ for large n. Similarly, for DeGroot agent n (in average network (a)), her centrality is equal to at most (approximately) $13/18 > 2/3$. Therefore, the average network is impervious to manipulation for all $b < -4/9$ for any ε.

On the other hand, consider the realized network for society $\mathcal{S}_n$. In this case, $\tilde{\mathbf{G}}_n$ looks like one of the two networks in Figure 12, each with equal probability. For large n the DeGroot centralities of all agents $2 \leq i \leq (n-1)$ are approximately unchanged from the average network.



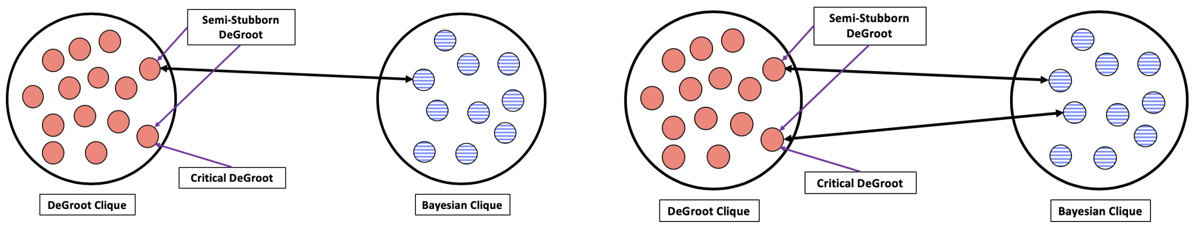
(a) Impervious (b) Susceptible

Figure 12. The realized network (all links are the same).

In network (a), the maximal DeGroot centrality of agent n is strictly less than in the average network; in particular, it is equal to $2/3$, the same as all other DeGroots in the network. Thus, network (a) is impervious to manipulation for all $b < -1/3$, which implies that if $b < -4/9$, then realized network (a), $\tilde{\mathbf{G}}_n$, is impervious as in the average network $\bar{\mathbf{G}}_n$. In network (b), the maximal DeGroot centrality of agent n is strictly more than in the average network; in particular, it is equal to $5/6$. For $b = -5/9 < -4/9$ and sufficiently small ε, agent n will be manipulated under realized network (b). Therefore, $\tilde{\mathbf{G}}_n$ in network (b) is susceptible to manipulation even though $\bar{\mathbf{G}}_n$ is not.

Of course, for any agent i, as $n \to \infty$, the expected degree of agent i grows faster than $\log n$ (it grows linearly!), but every third society $\mathcal{S}_n$ is susceptible to manipulation with probability $1/2$, even though the average network $\bar{\mathbf{G}}_n$ never is. $\square$

B.2 Non-Normal Society



(a) Susceptible. (b) Impervious.

Figure 13. Example that non-normal society may violate the random network theory: two networks which occur with equal probabilities, but possess different properties as $n \to \infty$.

Example 5. For ease of notation, let us define $n_* = n - \log^2(n)$. Suppose the first $\log^2(n)$ agents are Bayesian, the next two agents are DeGroot with $\theta = 1/n_*$ (who we will call *semi-stubborn*), and the remaining agents are DeGroot with $\theta = 1/\exp(n_*)$.[19] All of the DeGroots are pairwise adjacent with probability 1 in a clique of size $|D|$; all of the Bayesians are pairwise adjacent with probability 1 in a clique of size $|B|$. Suppose the first semi-stubborn DeGroot agent is adjacent to the first Bayesian with probability 1 (so the network is always connected), whereas the second semi-stubborn agent is adjacent to the first Bayesian with only probability $1/2$ (called the *critical DeGroot*). All other (non semi-stubborn) DeGroot agents are never adjacent to a Bayesian.

When critical DeGroot is adjacent to the Bayesian, the walks to the Bayesian are given by:

$$w = \frac{2}{n_*} w_{ss} + \frac{n_* - 2}{n_*} w$$
$$w_{ss} = \frac{n_* - 1}{n_*^2} + \frac{(n_* - 2)(n_* - 1)}{n_*^2} w$$

which as $n \to \infty$ (so $n_* \to \infty$) satisfies $w = w_{ss} = 1/3$. Now, consider the case where the DeGroot is *not* adjacent to the Bayesian; the walks to the Bayesians are given by:

$$w = \frac{1}{n_*} w_c + \frac{1}{n_*} w_{ss} + \frac{n_* - 2}{n_*} w$$
$$w_{ss} = \frac{n_* - 1}{n_*^2} + \frac{(n_* - 1)(n_* - 2)}{n_*^2} w + \frac{n_* - 1}{n_*^2} w_c$$
$$w_c = \frac{n_* - 2}{n_*} w + \frac{1}{n_*} w_{ss}$$

which as $n \to \infty$ satisfies $w = w_{ss} = w_c = 0$. When $(1+b)/2 = 1/3$, $b = -1/3$, whereas when $(1+b)/2 = 0$, $b = -1$. Thus, for $b \in (-1, -1/3)$ (and $\varepsilon \approx 0$), the network is impervious

[19] Note, that technically this violates the principle that $\tau : \mathcal{N} \to \{B, D\}$ is a fixed map. However, it is a basic fact that there exists a fixed map τ such that for any n, the number of Bayesians is approximately $\log^2(n)$; for simplicity, we omit this detail.

with probability $1/2$ (if the critical agent is adjacent to a Bayesian) and susceptible with probability $1/2$ (if the critical agent is not adjacent to the Bayesian). But clearly the normal society condition is violated since the semi-stubborn agents hold $\theta = 1/n_*$ and other DeGroots hold $\theta = 1/\exp(n_*)$, and $\lim_{n\to\infty}(1/n_*)/(1/\exp(n_*)) = \infty$. The average network $\bar{\mathbf{G}}$ is impervious if and only if $b \in (-1, -2/3)$.

Note that if $\theta = 1/n_*$ for all DeGroot agents, then $w = w_{ss} = w_c = 0$ regardless of the realization of this link; on the other hand, if $\theta = 1/\exp(n_*)$, then $w = w_{ss} = w_c = 1$ regardless of the realization of this link.

C Proofs

Preliminaries The following notation is used throughout the proofs. The vector $\boldsymbol{\gamma} \in \{0,1\}^n$ denotes which agents are targeted by the principal, and the DeGroot Centrality vector resulting from this targeting is denoted by $\mathcal{D}(\boldsymbol{\gamma})$. DeGroot agent i is manipulated if $\mathcal{D}_i(\boldsymbol{\gamma}) < (1-b)/2$.

Proof of Theorem 1. Let us denote by $\tilde{\mathbf{E}}_n$ the "expected" (normalized) adjacency matrix, $\tilde{\mathbf{E}}_n = \mathbb{E}[\tilde{\boldsymbol{\rho}}_n]\mathbb{E}[\tilde{\mathbf{D}}_n]^{-1}$, and the "mean" influence network, $\bar{\mathbf{A}}_n = \mathbf{O}_\theta^{(n)}\tilde{\mathbf{E}}_n$, where:

$$\mathbf{O}_\theta^{(n)} = \begin{pmatrix} (1-\theta_1^{(n)}) & 0 & \dots & 0 \\ 0 & (1-\theta_2^{(n)}) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & (1-\theta_n^{(n)}) \end{pmatrix}$$

The first step of the proof establishes that the difference between the "mean" (normalized) adjacency matrix and the realized (normalized) adjacency matrix, $||\tilde{\boldsymbol{\rho}}_n\tilde{\mathbf{D}}_n^{-1} - \bar{\mathbf{E}}_n||_2$, is small with high probability. In the second step, we prove that for any sequence of slack parameters $\boldsymbol{\gamma}_n$, the difference between the expected and realized DeGroot centrality vector, $||\mathcal{D}(\boldsymbol{\gamma}_n) - \mathbb{E}[\mathcal{D}(\boldsymbol{\gamma}_n)]||_2$, is also small with high probability.[20] Finally, we argue that this proves that $\mathbf{A}_n$ is with high probability impervious (resp. susceptible) if and only if $\bar{\mathbf{A}}_n$ is impervious (resp. susceptible).

Step 1: Call $\tilde{\mathbf{E}}_n \equiv \tilde{\boldsymbol{\rho}}_n\tilde{\mathbf{D}}_n^{-1}$. Let $\psi > 0$. Let $\underline{d}^{(n)} = \min_i d_i^{(n)}$; that is, $\underline{d}^{(n)}$ is the expected minimum degree. We first show that the Laplacian matrices $\tilde{\mathcal{L}}_n = \mathbf{I} - \tilde{\mathbf{D}}_n^{-1/2}\tilde{\boldsymbol{\rho}}_n\tilde{\mathbf{D}}_n^{-1/2}$ and $\bar{\mathcal{L}}_n = \mathbf{I} - \bar{\mathbf{D}}_n^{-1/2}\bar{\boldsymbol{\rho}}_n\bar{\mathbf{D}}_n^{-1/2}$ satisfy $\lim_{n\to\infty}\mathbb{P}[||\tilde{\mathcal{L}}_n - \bar{\mathcal{L}}_n||_2 \geq \psi] = 0$ (i.e., they are equal with high probability). It follows from Theorem 2 in Chung and Radcliffe (2011) that with probability at least $1-\psi$:

$$||\tilde{\mathcal{L}}_n - \bar{\mathcal{L}}_n||_2 \leq 2\sqrt{\frac{3\log(4n/\psi)}{\underline{d}^{(n)}}}$$

By the expected-degrees condition, we know that $\lim_{n\to\infty} \underline{d}^{(n)}/\log n \to \infty$, which implies that:

$$\limsup_{n\to\infty} ||\tilde{\mathcal{L}}_n - \bar{\mathcal{L}}_n||_2 \leq \lim_{n\to\infty} 2\sqrt{\frac{3\log(4n/\psi)}{\underline{d}^{(n)}}} = 0$$

[20]From Appendix A.1, the DeGroot Centrality Vector is given by $\mathcal{D}(\boldsymbol{\gamma}) = (I-A)^{-1}\boldsymbol{\gamma} = \sum_{k=0}^{\infty} A^k\boldsymbol{\gamma}$, where $\boldsymbol{\gamma}$ is the vector in $\{0,1\}^n$ with $\gamma_i = 1$ if Agent i is targeted by the principal, I is the identity and A is the adjacency matrix.

establishing the desired result. It is clear that the same implication is true of the matrices:

$$\tilde{\mathbf{N}} \equiv \tilde{\mathbf{D}}_n^{-1/2}\tilde{\boldsymbol{\rho}}_n\tilde{\mathbf{D}}_n^{-1/2} = \tilde{\mathbf{D}}_n^{-1/2}\tilde{\mathbf{E}}_n\tilde{\mathbf{D}}^{1/2}$$
$$\bar{\mathbf{N}} \equiv \bar{\mathbf{D}}_n^{-1/2}\bar{\boldsymbol{\rho}}_n\bar{\mathbf{D}}_n^{-1/2} = \bar{\mathbf{D}}^{-1/2}\bar{\mathbf{E}}_n\bar{\mathbf{D}}^{1/2}$$

Thus, let us write:

$$\begin{aligned}
\limsup_{n\to\infty} ||\tilde{\mathbf{E}}_n - \bar{\mathbf{E}}_n||_2 &\leq \limsup_{n\to\infty} ||\tilde{\mathbf{D}}^{1/2}\tilde{\mathbf{N}}\tilde{\mathbf{D}}^{-1/2} - \bar{\mathbf{D}}^{1/2}\bar{\mathbf{N}}\bar{\mathbf{D}}^{-1/2}||_2 \\
&\leq \limsup_{n\to\infty} \left(\max\{||\tilde{\mathbf{D}}^{1/2}||_2, ||\bar{\mathbf{D}}^{1/2}||_2\}\right) ||\tilde{\mathbf{N}} - \bar{\mathbf{N}}||_2 \left(\max\{||\tilde{\mathbf{D}}^{-1/2}||_2, ||\bar{\mathbf{D}}^{-1/2}||_2\}\right) \\
&\leq \limsup_{n\to\infty} \psi \cdot \left(\max\{||\tilde{\mathbf{D}}^{1/2}||_2, ||\bar{\mathbf{D}}^{1/2}||_2\}\right) \cdot \left(\max\{||\tilde{\mathbf{D}}^{-1/2}||_2, ||\bar{\mathbf{D}}^{-1/2}||_2\}\right)
\end{aligned}$$

Let L_{ij} be the binary random variable for if there exists a link $j \to i$, note that:

$$||\tilde{\mathbf{D}}_n^{1/2}||_2||\bar{\mathbf{D}}_n^{-1/2}||_2 = \sqrt{\frac{\max_i \sum_{j=1}^n L_{ij}}{\max_i \sum_{j=1}^n p_{ij}}} \leq \max_i \sqrt{\frac{\sum_{j=1}^n L_{ij}}{\sum_{j=1}^n p_{ij}}}$$
$$||\tilde{\mathbf{D}}_n^{1/2}||_2||\bar{\mathbf{D}}_n^{-1/2}||_2 = \sqrt{\frac{\max_i \sum_{j=1}^n p_{ij}}{\max_i \sum_{j=1}^n L_{ij}}} \leq \max_i \sqrt{\frac{\sum_{j=1}^n p_{ij}}{\sum_{j=1}^n L_{ij}}}$$

which are both bounded above almost surely. To see this, note that for n large, we can apply the Lyapunov Central Limit Theorem (see Billingsley (1995)):

$$\begin{aligned}
\lim_{n\to\infty} \frac{\sum_{j=1} L_{ij}}{\sum_{j=1}^n p_{ij}} - 1 &\sim \frac{1}{\sum_{j=1}^n p_{ij}} N\left(0, \sum_{j=1}^n p_{ij}(1-p_{ij})\right) \\
&\sim \frac{1}{\sqrt{\log n}} N\left(0, \Omega_n\right)
\end{aligned}$$

where $\Omega_n \to 0$. If $z_1, \ldots, z_n$ are normally distributed with variance σ^2, then by the Fisher-Tippet-Gnedenko theorem (see Charras-Garrido and Lezaud (2013) and Taylor (2011)), we see that:

$$\mathbb{E}\left[\max_i z_i\right] \in O(\sigma\sqrt{\log n})$$

Therefore, we have by Jensen's inequality:

$$\begin{aligned}
\mathbb{E}\left[\max_i \sqrt{\frac{\sum_{j=1}^n L_{ij}}{\sum_{j=1}^n p_{ij}}}\right] &= \mathbb{E}\left[\sqrt{\max_i \frac{\sum_{j=1}^n L_{ij}}{\sum_{j=1}^n p_{ij}}}\right] \\
&\leq \sqrt{\mathbb{E}\left[\max_i \frac{\sum_{j=1}^n L_{ij}}{\sum_{j=1}^n p_{ij}}\right]} \in 1 + O(\sqrt{\Omega_n}) \overset{n\to\infty}{\to} 1
\end{aligned}$$

which by Markov's inequality suggests for any $\kappa > 0$:

$$\lim_{n\to\infty} \mathbb{P}\left[\max_i \sqrt{\frac{\sum_{j=1}^n L_{ij}}{\sum_{j=1}^n p_{ij}}} \leq (1+\kappa)\right] = 1$$

Similar reasoning proves the second case where $\max_i \tilde{\mathbf{D}}_n < \max_i \bar{\mathbf{D}}_n$. This establishes that $||\tilde{\mathbf{E}}_n - \bar{\mathbf{E}}_n||_2$ is small with high probability.

Step 2: In this step, we model Bayesian agents i by taking $\theta_i^{(n)} = 1$ for all n. We note then that $\tilde{\mathbf{A}}_n = \mathbf{O}_\theta^{(n)} \tilde{\mathbf{E}}_n$ and $\bar{\mathbf{A}}_n = \mathbf{O}_\theta^{(n)} \bar{\mathbf{E}}_n$. Fix a sequence of influence vectors, $\boldsymbol{\gamma}_n$. For every $\psi > 0$, we can write for large enough n:

$$
\begin{aligned}
\left|\left|(\mathbf{I} - \tilde{\mathbf{A}}_n)^{-1} - (\mathbf{I} - \bar{\mathbf{A}}_n)^{-1}\right|\right|_2 &= \left|\left|\sum_{k=0}^{\infty} \left(\tilde{\mathbf{A}}_n^k - \bar{\mathbf{A}}_n^k\right)\right|\right|_2 \\
&= \left|\left|\sum_{k=0}^{\infty} \mathbf{O}_n^k (\tilde{\mathbf{E}}_n^k - \bar{\mathbf{E}}_n^k)\right|\right|_2 \\
&\leq \sum_{k=0}^{\infty} \left|\left|\mathbf{O}_n^k\right|\right|_2 \left|\left|\tilde{\mathbf{E}}_n^k - \bar{\mathbf{E}}_n^k\right|\right|_2 \\
&\leq \sum_{k=0}^{\infty} \left(1 - \inf_i \theta_i^{(n)}\right)^k \psi \\
&\leq \frac{\psi}{\inf_i \theta_i^{(n)}}
\end{aligned}
$$

Note that this implies for large n any $\boldsymbol{\gamma}_n$:

$$
\begin{aligned}
\left((\mathbf{I} - \tilde{\mathbf{A}}_n)^{-1} - (\mathbf{I} - \bar{\mathbf{A}}_n)^{-1}\right)_{DD} \left(\boldsymbol{\gamma}_n \otimes \boldsymbol{\theta}^{(n)}\right) &\leq \frac{\psi}{\inf_i \theta_i^{(n)}} \mathbf{1} \otimes \boldsymbol{\theta}^{(n)} \\
&\leq \frac{\psi \sup_i \theta_i^{(n)}}{\inf_i \theta_i^{(n)}} \mathbf{1} \leq \psi \nu \mathbf{1}
\end{aligned}
$$

by the normal society condition on $\boldsymbol{\theta}^{(n)}$. Therefore, we can bound this difference in influence from above by any constant $C > 0$ we choose; in particular, for any $C > 0$:

$$
\lim_{n \to \infty} \mathbb{P}\left[\left|\left|\left((\mathbf{I} - \tilde{\mathbf{A}}_n)^{-1} - (\mathbf{I} - \bar{\mathbf{A}}_n)^{-1}\right)_{DD} \left(\boldsymbol{\gamma} \otimes \boldsymbol{\theta}^{(n)}\right)\right|\right|_\infty \geq C\right] = 0
$$

Lastly, we know that,

$$
\left((\mathbf{I} - \tilde{\mathbf{A}}_n)^{-1} - (\mathbf{I} - \bar{\mathbf{A}}_n)^{-1}\right)_{DB} \mathbf{0} = \mathbf{0}
$$

Thus, as $n \to \infty$, with high probability we have for every $\mu > 0$ and $\boldsymbol{\gamma}_n$:

$$
\lim_{n \to \infty} \mathbb{P}\left[\left|\left|\tilde{\mathcal{D}}(\boldsymbol{\gamma}_n) - \mathbb{E}[\mathcal{D}(\boldsymbol{\gamma}_n)]\right|\right|_\infty \geq \mu\right] = 0
$$

as desired.

Step 3: Under Assumption 2(a), $\bar{\boldsymbol{\rho}}_{DD,n} \bar{\mathbf{D}}_{DD,n}^{-1}$ has a non-vanishing spectral gap, then for sufficiently large n, we know any two DeGroot agents i and j in the realized network $\tilde{\mathbf{A}}_n$ are connected with high probability. To see this, we first construct a directed network T as follows:

$$t_{ij} = \begin{cases} [\bar{\boldsymbol{\rho}}_n \bar{\mathbf{D}}_n^{-1}]_{ij}, \text{ if } i, j \in D \\ 0, \text{ if } i \in D, j \in B \text{ or } i \in B \\ 1, \text{ if } i = j \in B \end{cases}$$

The (normalized) Laplacian matrix of the directed network T is given by:

$$\mathcal{L} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \bar{\mathbf{D}}_{DD,n}^{-1/2} \bar{\boldsymbol{\rho}}_{DD,n} \bar{\mathbf{D}}_{DD,n}^{-1/2} \end{pmatrix}$$

which has the same $|D|$ largest eigenvalues as $\mathbf{I} - \bar{\boldsymbol{\rho}}_{DD,n} \bar{\mathbf{D}}_{DD,n}^{-1}$. We use the result of Chung (2005) which provides a Cheeger inequality for directed networks: namely, that the conductance of T, $\phi(T)$, is bounded below by $\lambda_2^T/2$. Note that $\phi(T) = \phi(\bar{\mathbf{E}}_{DD,n})$, and therefore we know that as $n \to \infty$, $\phi(\bar{\mathbf{E}}_{DD,n}) \geq (1-\eta)/2 \equiv \kappa > 0$. Since $\sup_{i,n\in\mathbb{N}} \theta_i^{(n)} \ll 1$, we know that $\bar{\mathbf{A}}_{DD,n}$ is connected if and only if $\bar{\mathbf{E}}_{DD,n}$ is connected; similarly, $\tilde{\mathbf{A}}_{DD,n}$ is connected w.h.p. if and only if $\bar{\mathbf{E}}_{DD,n}$ is.

Consider the network $\bar{\mathbf{E}}^*_{DD,n} = \bar{\mathbf{E}}_{DD,n} \bar{\mathbf{D}}_n \left(\bar{\mathbf{D}}_n^*\right)^{-1}$, where:

$$[\bar{\mathbf{D}}_n^*]_{ij} = \begin{cases} 0, \text{ if } i \neq j \\ \sum_{k \in D} p_{ik}(n), \text{ if } i = j \end{cases}$$

Then $\bar{\mathbf{E}}^*_{DD,n}$ is symmetric, so by Chung and Radcliffe (2011) and the same reasoning as in Step 1 (along with the fact that $\bar{\mathbf{E}}^*_{DD,n}$ and $\mathbf{I} - \bar{\mathcal{L}}^*_{DD,n}$ have the same eigenvalues), we know that:

$$||\tilde{\lambda}_\mu^{*DD} - \bar{\lambda}_\mu^{*DD}||_2 \leq 2\sqrt{\frac{3\log(4n/\psi)}{\underline{d}^{DD,(n)}}}$$

for $\mu = 1, 2$, where $\underline{d}^{DD,(n)}$ denotes the minimum degree of a DeGroot to other DeGroot agents. Note that $\lim_{n\to\infty} \underline{d}^{DD,(n)}/\log n = \infty$ since $\bar{\mathbf{E}}_n$ has a non-vanishing spectral gap (and the previous conductance argument), and $\lim_{n\to\infty} \underline{d}^{(n)}/\log n = \infty$ by the expected-degrees condition. This implies that with high probability, $\tilde{\mathbf{E}}^*_{DD,n}$ has no vanishing spectral gap, and so using the standard Cheeger inequality proves that it is connected w.h.p. Since $\tilde{\mathbf{E}}^*_{DD,n}$ is connected if and only if $\tilde{\mathbf{E}}_{DD,n}$ is, we see that $\tilde{\mathbf{E}}_{DD,n}$ is connected w.h.p. and therefore, the network of DeGroot agents is connected.

Finally, consider the problem given in Theorem 4 of Mostagir et al. (2019):

$$\boldsymbol{\Gamma}^* = \arg \max_{\boldsymbol{\gamma} \in \{0,1\}^{n-m}} \sum_{i=m+1}^{n} z_i - \varepsilon \gamma_i$$

subject to the condition that $z_i \leq \mathcal{D}_i(\boldsymbol{\gamma}) + (1+b)/2$ and $z_i, \gamma_i \in \{0,1\}$ for all DeGroots i. We can rewrite z_i as the indicator function $1_{\mathcal{D}_i(\boldsymbol{\gamma}) > (1-b)/2}$. The difference in the objective function between the “expected” network and the realized network is:

$$\left|\sum_{i=m+1}^{n}\left(1_{\tilde{\mathcal{D}}_i(\boldsymbol{\gamma}_n)>(1-b)/2}-1_{\bar{\mathcal{D}}_i(\boldsymbol{\gamma}_n)>(1-b)/2}\right)\right|$$
$$\leq \sum_{i=m+1}^{n} 1_{|\tilde{\mathcal{D}}_i(\boldsymbol{\gamma}_n)-\bar{\mathcal{D}}_i(\boldsymbol{\gamma}_n)|>\beta}1_{|\bar{\mathcal{D}}_i(\boldsymbol{\gamma}_n)-(1-b)/2|<\beta}$$

for any choice of $\beta > 0$. By step 2, this implies that as $n \to \infty$, and then $\beta \to 0$, the difference in objective functions is upper bounded by $\sum_{i=m+1}^{n} 1_{\bar{\mathcal{D}}_i(\boldsymbol{\gamma}_n)} = (1-b)/2$ which holds for at most countably many b as $n \to \infty$, and therefore the difference in objective functions is zero for each γ_n. Because the principal knows only the random network generation process but not the realized network, there are no differences in the objective function for any given $\boldsymbol{\gamma}_n$. This implies that the set $\boldsymbol{\Gamma}^*$ for the realized influence network $\tilde{\mathbf{A}}_n$, $\tilde{\boldsymbol{\Gamma}}^*$, and the “expected” influence network $\bar{\mathbf{A}}_n$, $\bar{\boldsymbol{\Gamma}}^*$, are equivalent sets: $\tilde{\boldsymbol{\Gamma}}^* = \bar{\boldsymbol{\Gamma}}^*$. Therefore, $\mathbf{0} \in \tilde{\boldsymbol{\Gamma}}^*$ if and only if $\mathbf{0} \in \bar{\boldsymbol{\Gamma}}^*$. $\square$

Proof of Theorem 2. For part (a), it is enough to show that for a fixed $\boldsymbol{\gamma}_n$, when $p_d = p_s$ the minimum belief of any DeGroot agent is maximized. First, note that when $p_d = p_s$, the beliefs of all DeGroot agents are the same, say π^*. For any other p_d, the weighted walks to the Bayesian agents decreases for all DeGroots. Moreover, there must exist some DeGroot agent j whose weighted walks to agents in $\{i : \gamma_i = 1\}$ has increased. Therefore, agent j’s belief of the state under $p_d < p_s$ is strictly less than π^*, which establishes (a).

For parts (b) and (c), note that when $p_d \to 0$, each island is isolated. The belief of the agents on island ℓ are given by $\frac{1}{ns_\ell}\sum_{i\in[\ell]}\gamma_i$, where $[\ell]$ is the set of agents on island ℓ. Therefore, the principal can manipulate on island ℓ if and only if $\varepsilon < ns_\ell \cdot \lceil \frac{2}{ns_\ell(1-b)} \rceil \equiv \varepsilon^*_\ell$. Once again, by continuity this holds for sufficiently small p_d as well i.e., on some interval $(0, \underline{p}_d)$. It just remains to find p'_d such that there is manipulation when $\varepsilon < \min_\ell \varepsilon^*_\ell$, but for some b there is also manipulation for some $\varepsilon' > \min_\ell \varepsilon^*_\ell$. Consider $\underline{p}_d$ which is the supremum of all p_d where there is manipulation if and only if $\varepsilon < \min_\ell \varepsilon^*_\ell$ (such a $\underline{p}_d$ exists because $p_d = p_s$ is impervious); call this property **Island Independence**. It is clear that island independence is satisfied for all $(0, \underline{p}_d)$. Also notice for large enough n, under island independence either all (DeGroot) agents are manipulated or none are.

Because the influence of the Bayesians is bounded above by a constant fraction strictly less than 1 (this is maximized when $p_s = p_d$), there exists $p'_d > \underline{p}_d$ (i.e., p'_d which violates island independence) and b such that the principal can manipulate all DeGroots by targeting fewer than $\sum_{\ell=1}^{k}\lceil \frac{n(1-b)s_\ell}{2} \rceil$. Thus, there exists an appropriate $\varepsilon > \min_\ell \varepsilon^*_\ell$ where the network with $p_d \in (0, \underline{p}_d)$ is impervious but $p_d \in (\underline{p}_d, p'_d)$ is not. $\square$

Proof of Theorem 3. We prove the case for $k = 2$ islands; the more general case can be proved analogously via induction.

Notice the network is susceptible to manipulation (with high probability) if there exists an agent j with $\mathcal{D}_j(\mathbf{1}) > (1-b)/2$; similarly, the network is impervious to manipulation (with high probability) if for all agents j, $\mathcal{D}_j(\mathbf{1}) < (1-b)/2$. We show that increasing homophily leads to an increase in the inequality of DeGroot centralities (i.e., $\mathcal{D}(\mathbf{1})$). We have the system of equations:

$$\frac{n}{n-1}\mathbf{w} = \frac{2}{n(p_s+p_d)}\begin{pmatrix} p_s m_1 + p_d m_2 \\ p_d m_1 + p_s m_2 \end{pmatrix} + \frac{2}{n(p_s+p_d)}\begin{pmatrix} p_s(n/2-m_1) & p_d(n/2-m_2) \\ p_d(n/2-m_1) & p_s(n/2-m_2) \end{pmatrix}\mathbf{w}$$

which is equivalent to:

$$n^2\frac{p_s+p_d}{2n-2}\mathbf{w} = \begin{pmatrix} p_s m_1 + p_d m_2 \\ p_d m_1 + p_s m_2 \end{pmatrix} + \begin{pmatrix} p_s(n/2-m_1) & p_d(n/2-m_2) \\ p_d(n/2-m_1) & p_s(n/2-m_2) \end{pmatrix}\mathbf{w}$$

Without loss of generality suppose that island 1 has more Bayesian agents (i.e., $m_1 > m_2$). Consider the map T given by:

$$T : \mathbf{w} \mapsto \begin{pmatrix} p_s m_1 + p_d m_2 \\ p_d m_1 + p_s m_2 \end{pmatrix} + \begin{pmatrix} p_s(n/2-m_1) & p_d(n/2-m_2) \\ p_d(n/2-m_1) & p_s(n/2-m_2) \end{pmatrix}\mathbf{w}$$

We claim that T has the property that $(w_1 \geq w_2) \implies T(w_1) \geq T(w_2)$. Suppose that $w_1 \geq w_2$, then:

$$\begin{aligned} m_1 + w_1(n/2-m_1) &\geq m_2 + w_1(n/2-m_2) \\ &\geq m_2 + w_2(n/2-m_2) \end{aligned}$$

which moreover implies that

$$p_s(m_1+w_1(n/2-m_1)) + p_d(m_2+w_2(n/2-m_2)) \geq p_d(m_1+w_1(n/2-m_1)) + p_s(m_2+w_2(n/2-m_2))$$

because $p_s > p_d$. Because p_s, p_d, n are fixed, this suggests that $(2n-2)/n^2(p_s+p_d)\cdot T$ also has this property, so any fixed point of T must have $w_1 \geq w_2$ by Brouwer's fixed point theorem. Since the system is linear and non-singular, there is a unique fixed-point with $w_1 \geq w_2$. This implies the DeGroot centrality of the island with more Bayesians is no larger than the DeGroot centrality of the other island.

For the remainder of this part of the proof, we define a new operator T which maps $\mathbf{w}$, parametrized by p_s, p_d, and (m_1, m_2), respectively. We show the following: (i) $T|p_s$ is increasing in p_s for w_1 but decreasing for w_2, (ii) $T|p_d$ is decreasing in p_d for w_1 but increasing for w_2, and (iii) $T|(m_1, m_2)$ subject to $m_1 + m_2 = m$ and $m_1 \geq m_2$ is minimized at $m_1 = m/2$. This result suffices in order to show that there exists a fixed-point of T which obeys the desired properties of Theorem 3 (by Brouwer[21]), and by linearity, this fixed-point is unique.

1. Increasing p_s: Let us define T as:

$$T|p_s : \mathbf{w} \mapsto \frac{1}{p_s+p_d}\left[\begin{pmatrix} p_s m_1 + p_d m_2 \\ p_d m_1 + p_s m_2 \end{pmatrix} + \begin{pmatrix} p_s(n/2-m_1) & p_d(n/2-m_2) \\ p_d(n/2-m_1) & p_s(n/2-m_2) \end{pmatrix}\mathbf{w}\right]$$

   Computing directly:

$$\begin{aligned} \frac{\partial T(w_1|p_s)}{\partial p_s} &= p_d\frac{(m_1+(n/2-m_1)w_1)-(m_2+(n/2-m_2)w_2)}{(p_s+p_d)^2} > 0 \\ \frac{\partial T(w_2|p_s)}{\partial p_s} &= p_d\frac{(m_2+(n/2-m_2)w_2)-(m_1+(n/2-m_1)w_1)}{(p_s+p_d)^2} < 0 \end{aligned}$$

   where the inequalities follow from the analysis above.

[21] In particular, let (w_1, w_2) be the old fixed-point and (w_1', w_2') the new fixed point. We illustrate for the case of increasing p_s: all other cases are similar. By increasing p_s, we know that T maps all w_1 larger and all w_2 smaller. Therefore, the convex compact set $[w_1, 1] \times [0, w_2]$ maps into itself, which implies the new fixed-point (w_1', w_2') lies in this set.

2. Increasing p_d: Let us define T in the same way as in (1), except parametrized by p_d. Then in exactly the same way:

$$\frac{\partial T(w_1|p_d)}{\partial p_d} = \frac{p_s}{p_d}\frac{\partial T(w_2|p_s)}{\partial p_s} < 0$$
$$\frac{\partial T(w_2|p_d)}{\partial p_d} = \frac{p_s}{p_d}\frac{\partial T(w_1|p_s)}{\partial p_s} > 0$$

which is the desired result.

3. Setting $m_1 = m_2 = m/2$: Let us define T as:

$$T|(m_1,m_2) : \mathbf{w} \mapsto \left[\begin{pmatrix} p_s m_1 + p_d m_2 \\ p_d m_1 + p_s m_2 \end{pmatrix} + \begin{pmatrix} p_s(n/2-m_1) & p_d(n/2-m_2) \\ p_d(n/2-m_1) & p_s(n/2-m_2) \end{pmatrix}\mathbf{w}\right]$$

Let us suppose $(m_1,m_2) = (m/2, m/2)$. Computing the directional derivative along the gradient $\mathbf{u} = (1,-1)$:

$$\nabla_{\mathbf{u}} T(w_1)|(m_1,m_2) = p_s(1-w_1) - p_d(1-w_2) = (p_s - p_d)(1-w_1) > 0$$

because $w_1 = w_2$ when m_1, m_2 are equal. This implies that $(m_1,m_2) = (m/2,m/2)$ is a local maximum for the DeGroot centralities of agents on island 1 subject to the constraint $w_1 \geq w_2$. To show that it is a global maximum, we prove there are no other local maxima. First, notice that

$$\nabla_{\mathbf{u}} T(w_1)|(m_1,m_2) = p_s(1-w_1) - p_d(1-w_2)$$
$$\nabla_{\mathbf{u}} T(w_2)|(m_1,m_2) = p_d(1-w_1) - p_s(1-w_2)$$

which implies that $p_s(1-w_1) - p_d(1-w_2)$ decreases as m_1 increases (and m_2 increases) along the direction $\mathbf{u}$:

$$\begin{aligned} -p_s \nabla_{\mathbf{u}} T(w_1)|(m_1,m_2) + p_d \nabla_{\mathbf{u}} T(w_2)|(m_1,m_2) &= -p_s^2(1-w_1) + p_d^2(1-w_1) \\ &= -(p_s^2 - p_d^2)(1-w_1) < 0 \end{aligned}$$

Therefore, it is impossible for (m_1^*, m_2^*) and (m_1^{**}, m_2^{**}) to exist such that

$$p_s(1-w_1^*) + p_d(1-w_2^*) < 0$$
$$p_s(1-w_1^{**}) + p_d(1-w_2^{**}) > 0$$

with $m_1^{**} > m_1^*$ and $m_2^{**} < m_2^*$ such that $m_1^* + m_2^* = m_1^{**} + m_2^{**} = m$. There are no other local maxima, so it just remains to check that $(m_1,m_2) = (m,0)$ admits a lower DeGroot centrality for island 1 (higher w_1) than when $(m_1,m_2) = (m/2,m/2)$. In these cases, we get,

$$\begin{aligned} T(w_1)|(m,0) &= p_s m + p_s(n/2-m)w_1 + p_d n w_2/2 \\ T(w_1)|(m/2,m/2) &= p_s m/2 + p_d m/2 + p_s(n-m)w_1/2 + p_d(n-m)w_2/2 \\ &= p_s m/2 + p_d m/2 + p_s n w_1/2 + p_d n w_2/2 - (p_s+p_d)m w_1/2 \end{aligned}$$

Observe then that:

$$\begin{aligned} T(w_1)|(m,0) - T(w_1)|(m/2,m/2) &= m(p_s - p_d)/2 - m(p_s - p_d)w_1/2 \\ &= m(p_s - p_d)(1-w_1)/2 > 0 \end{aligned}$$

Thus, $(m_1, m_2) = (m/2, m/2)$ is the global maximum for island 1's DeGroot centrality. Because there are also no other local maxima, this claim can be generalized to any majorization of $\mathbf{m}$ (where the island with fewer Bayesians receives Bayesians from the island with more).

Lastly, we need to argue that in the case of islands of equal size, increased DeGroot centrality inequality (i.e., the island with larger centrality increases its centrality while the other's centrality decreases) cannot make the network go from susceptible to impervious. To check if the network is impervious, all that needs to be checked is $\max_i \mathcal{D}_i(\mathbf{1}) > (1-b)/2$. When inequality of the DeGroot centrality increases, then $\max_i \mathcal{D}_i(\mathbf{1})$ increases, and so the above inequality is more likely to be satisfied when inequality is increased. Therefore, the network can go from impervious to susceptible, but not the other direction. $\square$

Proof of Proposition 1. Suppose there are n agents in the network, and there are m Bayesians. We denote by w_r^ℓ the weighted walks from an agent on island ℓ to any Bayesian on island r. For $\ell \neq r$, we can write:

$$\begin{aligned} w_r^\ell &\geq \frac{(n-1)p_d m_r}{n^2 p_s s_\ell + n^2 p_d(1-s_\ell)} + \left(\frac{n-1}{n}\right)\left(\frac{np_s s_\ell w_r^\ell + p_d(ns_r - m_r)w_r^r + np_d \sum_{\tau\neq r,\ell} s_\tau w_r^\tau}{np_s s_\ell + np_d(1-s_\ell)}\right) \\ &\geq \frac{(n-1)p_d m_r}{n^2 p_s s_\ell + n^2 p_d(1-s_\ell)} + \left(\frac{n-1}{n}\right)\left(\frac{np_s s_\ell \underline{w}_r + p_d(ns_r - m_r)(\underline{w}_r + w_r^r - \underline{w}_r) + np_d(1-s_\ell - s_r)\underline{w}_r}{np_s s_\ell + np_d(1-s_\ell)}\right) \\ &= \frac{(n-1)p_d m_r}{n^2 p_s s_\ell + n^2 p_d(1-s_\ell)} + \left(\frac{n-1}{n}\right)\left(\underline{w}_r + \frac{np_d s_r(w_r^r - \underline{w}_r) - p_d m_r w_r^r}{np_s s_\ell + np_d(1-s_\ell)}\right) \\ &= \frac{(n-1)p_d m_r}{n^2 p_s s_\ell + n^2 p_d(1-s_\ell)} + \left(\frac{n-1}{n^2}\right)\left(\frac{n(p_s s_\ell + p_d(1-s_\ell - s_r))}{p_s s_\ell + p_d(1-s_\ell)}\underline{w}_r + \frac{p_d(ns_r - m_r)}{p_s s_\ell + p_d(1-s_\ell)}w_r^r\right) \end{aligned}$$

where $\underline{w}_r = \min_{\tau\neq r} w_r^\tau$. This implies that:

$$\underline{w}_r \geq \frac{(n-1)p_d m_r + (n-1)p_d(ns_r - m_r)w_r^r}{n^2 p_d s_r + n(p_s s_\ell + p_d(1-s_\ell - s_r))}$$

Otherwise,

$$\begin{aligned} w_r^r &= \frac{(n-1)p_s m_r}{n^2 p_s s_r + n^2 p_d(1-s_r)} + \left(\frac{n-1}{n}\right)\frac{p_s(ns_r - m_r)w_r^r + p_d\sum_{\tau\neq r} s_\tau w_r^\tau}{np_s s_r + np_d(1-s_r)} \\ &\geq \frac{(n-1)p_s m_r}{n^2 p_s s_r + n^2 p_d(1-s_r)} + \left(\frac{n-1}{n}\right)\left[\underline{w}_r + \frac{np_s s_r(w_r^r - \underline{w}_r) - p_s m_r w_r^r}{np_s s_r + np_d(1-s_r)}\right] \\ &= \frac{(n-1)p_s m_r}{n^2 p_s s_r + n^2 p_d(1-s_r)} + \left(\frac{n-1}{n^2}\right)\left[\frac{np_d(1-s_r)}{p_s s_r + p_d(1-s_r)}\underline{w}_r + \frac{p_s(ns_r - m_r)}{p_s s_r + p_d(1-s_r)}w_r^r\right] \end{aligned}$$

which moreover implies that

$$w_r^r \geq \frac{(n-1)p_s m_r + (n-1)np_d(1-s_r)\underline{w}_r}{n^2 p_d(1-s_r) + p_s[n(m_r + s_r) - m_r]}$$

Combining these two results we get:

$$\begin{aligned}
&\underline{w}_r \geq \frac{(n-1)p_d m_r + (n-1)p_d(ns_r - m_r)\frac{(n-1)p_s m_r + (n-1)np_d(1-s_r)\underline{w}_r}{n^2 p_d(1-s_r)+p_s[n(m_r+s_r)-m_r]}}{n^2 p_d s_r + n(p_s s_\ell + p_d(1-s_\ell - s_r))} \\
\implies &\left(n^2 p_d s_r + n(p_s s_\ell + p_d(1-s_\ell-s_r))\right)\left(n^2 p_d(1-s_r) + p_s[n(m_r+s_r)-m_r]\right)\underline{w}_r \\
&\geq (n-1)p_d m_r\left(n^2 p_d(1-s_r) + p_s[n(m_r+s_r)-m_r]\right) \\
&+ (n-1)p_d(ns_r - m_r)\left((n-1)p_s m_r + (n-1)np_d(1-s_r)\underline{w}_r\right)
\end{aligned}$$

Note that:

$$n^4 p_d^2 s_r(1-s_r) - n(n-1)^2 p_d^2(ns_r - m_r)(1-s_r) = mp_d^2(1-s_r)(m_r(n-1)^2 + n(2n-1)s_r)$$

Therefore, we can write $\underline{w}_r \geq N(n)/D(n)$, where:

$$\begin{aligned}
N(n) \equiv& (n-1)p_d m_r(n^2 p_d(1-s_r) + p_s[n(m_r+s_r)-m_r]) + (n-1)^2 p_s p_d(ns_r - m_r)m_r \\
D(n) \equiv& p_d^2(1-s_r)(m_r(n-1)^2 + n(2n-1)s_r) + n^2 p_d p_s s_r[n(m_r+s_r)-m_r] \\
&+ n(p_s s_\ell + p_d(1-s_\ell-s_r))(n^2 p_d(1-s_r) + p_s[n(m_r+s_r)-m_r])
\end{aligned}$$

But, as $n \to \infty$, we can write:

$$\underline{w}_r \geq \frac{(p_d(1-s_r)+p_s s_r)m_r}{(p_d(1-s_r)+p_s s_r)m_r + p_s s_r^2 + (1-s_r)(p_s\bar{s} + p_d(1-\bar{s}-s_r))} = \frac{A_r m_r}{A_r m_r + C_r}$$

where $\bar{s} = \max_{\tau\in\{1,\ldots,k\}} s_\tau$ and,

$$\begin{aligned}
A_r &\equiv (p_d(1-s_r) + p_s s_r) \\
C_r &\equiv p_s s_r^2 + (1-s_r)(p_s\bar{s} + p_d(1-\bar{s}-s_r))
\end{aligned}$$

From the preliminaries at the beginning of this section and Theorem 1, the network is impervious as long as $\underline{w}_r > (1+b)/2$. Therefore, the network is impervious if $m_r \geq \frac{(1+b)C_r}{(1-b)A_r}$ for any island r. Taking $r^* \equiv \arg\max_{\ell\in\{1,\ldots,k\}} \frac{(1+b)C_r}{(1-b)A_r}$, we know that if $m \geq \frac{(1+b)C_{r^*}}{(1-b)A_{r^*}}$, the network is impervious to manipulation. $\square$

Proof of Proposition 2. Without loss of generality we assume that the first community in the network consists entirely of Bayesian agents. We will construct, for each n, a strong homophily model that is susceptible to manipulation. There will be $k = \lfloor n/\log^2 n\rfloor$ islands each of size $\log^2 n$; this satisfies the expected-degrees condition needed to apply Theorem 1. We assume that the principal attempts to manipulate the last r communities along the line:

$$[\boldsymbol{\gamma}]_i \equiv \begin{cases} 1, \text{ if } S(i) \leq k-r \\ 0, \text{ if } S(i) > k-r \end{cases}$$

where $S(i)$ denotes the island of agent i. We first compute $\mathcal{D}_\ell(\boldsymbol{\gamma})$ for every island $2 \leq \ell(k-r)$ by counting Bayesian walks *plus* walks to DeGroot agents with jammed signals (i.e., $\gamma_i = 1$), weighted by their $\theta_{i,n} = 1/n$; we denote these walks by w_ℓ which is equivalent to $1 - \mathcal{D}_\ell(\boldsymbol{\gamma})$. For

$2 \leq \ell \leq (k-r)$, we have:

$$\begin{aligned} w_\ell &= \frac{n-1}{n} \frac{p_d(w_{\ell-1} + w_{\ell+1}) + p_s w_\ell}{2p_d + p_s} \\ \implies w_\ell &= \frac{(n-1)p_d(w_{\ell-1} + w_{\ell+1})}{2np_d + p_s} \end{aligned}$$

with $w_1 = 1$ because island 1 consists entirely of Bayesian agents and are absorbing states of the walk. Letting $\Delta w_\ell = w_\ell - w_{\ell-1}$, we re-arrange and note that for $2 \leq \ell \leq (k-r)$:

$$\Delta w_{\ell+1} = \Delta w_\ell + \frac{p_s + 2p_d}{(n-1)p_d} w_\ell$$

Solving for Δw_ℓ explicitly when $2 \leq \ell \leq (k-r)$:

$$\Delta w_\ell = \frac{p_s + 2p_d}{(n-1)p_d} \sum_{\tau=2}^{\ell-1} w_\tau + (w_2 - w_1)$$

We also know that:

$$\begin{aligned} w_\ell &= w_1 + \sum_{\kappa=2}^{\ell} \Delta w_\kappa \\ &= w_1 + \sum_{\kappa=2}^{\ell} \left[\frac{p_s + 2p_d}{(n-1)p_d} \sum_{\tau=2}^{\kappa-1} w_\tau + (w_2 - w_1) \right] \\ &= w_1 + \frac{p_s + 2p_d}{(n-1)p_d} \sum_{\kappa=3}^{\ell} \sum_{\tau=2}^{\kappa-1} w_\tau + \sum_{\kappa=2}^{\ell} (w_2 - w_1) \\ &= w_1 + \frac{p_s + 2p_d}{(n-1)p_d} \sum_{\tau=2}^{\ell-1} (\ell - \tau) w_\tau + (\ell - 1)(w_2 - w_1) \end{aligned}$$

We prove by induction that $w_\tau \sim \exp(-\alpha\tau/\sqrt{n})$ for all $\tau \leq r$, where $\alpha = \sqrt{(p_s + 2p_d)/p_d}$. As $n \to \infty$, we have:

$$\begin{aligned} w_\ell \sim& 1 + \frac{p_s + 2p_d}{(n-1)p_d} \sum_{\tau=2}^{\ell-1} \left(\ell \exp(-\alpha\tau/\sqrt{n}) - \tau \exp(-\alpha\tau/\sqrt{n}) \right) + (\ell - 1) \left(\exp(-\alpha/\sqrt{n}) - 1 \right) \\ \sim& 1 + \frac{p_s + 2p_d}{(n-1)p_d} \int_2^{\ell-1} \left[\ell \exp(-\alpha\tau/\sqrt{n}) - \tau \exp(-\alpha\tau/\sqrt{n}) \right] d\tau + (\ell - 1) \left(\exp(-\alpha/\sqrt{n}) - 1 \right) \\ \sim& 1 + \frac{p_s + 2p_d}{(n-1)p_d} \left(\frac{\sqrt{n} e^{-\alpha(1+\ell)/\sqrt{n}} \left(e^{\alpha(\ell-1)/\sqrt{n}} (\alpha(\ell-2) - \sqrt{n}) + e^{2\alpha/\sqrt{n}} (\sqrt{n} - \alpha) \right)}{\alpha^2} \right) \\ &+ (\ell - 1) \left(\exp(-\alpha/\sqrt{n}) - 1 \right) \\ \sim& (1 - \alpha) + \left(\frac{p_s + 2p_d}{p_d} \right) \left(\frac{\alpha - 1 + e^{-\alpha\ell/\sqrt{n}}}{\alpha^2} \right) \\ \sim& e^{-\alpha\ell/\sqrt{n}} \end{aligned}$$

which confirms our inductive hypothesis.

Now, suppose the principal attempts to manipulate $r \sim \beta\sqrt{n}$ communities at the end of the line. By our previous result, we know that $w_{k-r} \to 0$ as $n \to \infty$. Now, let us compute $\mathcal{D}_i(\boldsymbol{\gamma})$ directly for communities $(k-r+1) \leq \ell \leq (k-1)$; we know that $\mathcal{D}_i(\boldsymbol{\gamma}) = 1$ for all agents i on island $k-r$. Furthermore, we can write for all islands $(k-r+1) \leq \ell \leq (k-1)$ the recursion:

$$\mathcal{D}_\ell = \frac{n-1}{n} \frac{p_d(\mathcal{D}_{\ell-1} + \mathcal{D}_{\ell+1}) + p_s \mathcal{D}_\ell}{2p_d + p_s}$$

which is the exact same recursion we saw before with walks to Bayesian agents and DeGroots with $\gamma_i = 1$. This implies that $\mathcal{D}_\ell \sim \exp(-\alpha(\ell - (k-r))/\sqrt{n})$, and in particular at the end of the line:

$$\mathcal{D}_{k-1} \sim \exp(-\alpha(r-1)/\sqrt{n}) \sim \exp(-\alpha\beta) > 0$$

Finally, we can write:

$$\mathcal{D}_k = \frac{n-1}{n} \frac{p_d \mathcal{D}_{k-1} + p_s \mathcal{D}_k}{p_d + p_s}$$
$$\implies \mathcal{D}_k = \frac{(n-1)p_d \mathcal{D}_{k-1}}{np_d + p_s} \sim \exp(-\alpha\beta)$$

For $(k-r+1) \leq \ell \leq k$, the principal will successfully manipulate $\beta\sqrt{n}$ communities without even jamming their signals provided that:

$$\mathcal{D}_\ell(\boldsymbol{\chi}^*) \geq \exp(-\alpha\beta) > \frac{\epsilon}{1+\epsilon}$$
$$\implies \beta < \frac{1}{\alpha} \ln\left(1 + \epsilon^{-1}\right)$$

On the other hand, the communities $2 \leq \ell \leq \lambda\sqrt{n}$ will not be manipulated even though the principal jams their signals, where λ satisfies:

$$1 - \mathcal{D}_\ell(\boldsymbol{\chi}^*) \geq \exp(-\alpha\lambda) > \frac{1}{1+\epsilon}$$
$$\implies \lambda < \frac{1}{\alpha} \ln(1+\epsilon)$$

Thus, the principal's limit payoff (as compared to the zero payoff of $\mathbf{x} = \mathbf{0}$) by manipulating all but the first $\lambda\sqrt{n}$ communities is given by:

$$n \cdot (1 - \lambda\sqrt{n}/n) - \varepsilon(1 - \beta\sqrt{n}/n)$$

Thus, for some ε bounded away from zero, the principal finds manipulation profitable. Note that β is increasing in p_s and decreasing in p_d; therefore, the benefits from manipulation increase as the degree of first-order homophily in the strong homophily model increases. □

Proof of Proposition 3. When ε is small, all that needs to be checked for imperviousness is $\max_i \mathcal{D}_i(\mathbf{1})$. The optimal Bayesian placement minimizes $\max_i \mathcal{D}_i(\mathbf{1})$ (or equivalently, maximizes

$\min_\ell w_\ell$). Consider the map T:

$$T \,:\, \mathbf{w} \mapsto \left[\begin{pmatrix} \frac{p_s m_1 + \sum_{\ell \neq 1} p_d m_\ell}{np_s s_1 + np_d(1-s_1)} \\ \cdots \\ \frac{p_s m_k + \sum_{\ell \neq k} p_d m_\ell}{np_s s_k + np_d(1-s_k)} \end{pmatrix} + \begin{pmatrix} \frac{p_s(ns_1 - m_1)}{np_s s_1 + np_d(1-s_1)} & \frac{p_d(ns_2 - m_2)}{np_s s_1 + np_d(1-s_1)} & \cdots & \frac{p_d(ns_k - m_k)}{np_s s_1 + np_d(1-s_1)} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{p_d(ns_1 - m_1)}{np_s s_k + np_d(1-s_k)} & \frac{p_d(ns_2 - m_2)}{np_s s_k + np_d(1-s_k)} & \cdots & \frac{p_s(ns_k - m_k)}{np_s s_k + np_d(1-s_k)} \end{pmatrix} \mathbf{w} \right]$$

Consider the directional derivative along $\mathbf{u}$:

$$\nabla_{\mathbf{u}} T(w_\ell) = \frac{u_1 p_s(1-w_1) + \sum_{\ell' \neq \ell} u_{\ell'} p_d (1 - w_\ell)}{np_s s_\ell + np_d(1-s_\ell)}$$

Consider $w_{\ell^*} \in \min_\ell w_\ell$; by the above formula (with directional derivative $u_{\ell^*} = 1$, $u_{\ell^{**}} = -1$ for some ℓ^{**} and everywhere else zero) adding a Bayesian to this island would increase w_ℓ. Therefore, to maximize the smallest w_ℓ, one should figure out the Bayesian placement as to equalize all w_ℓ.

We prove the optimal policy which is to set the Bayesians proportional to population, i.e., $m_\ell = M \cdot s_\ell$ by showing this equalizes DeGroot centrality across all of the islands. To see this, we simply plug in $\mathbf{w} = w^* \mathbf{1}$ to T and show it is a fixed point. For island ℓ:

$$\begin{aligned} w_\ell &= \frac{n-1}{n} \cdot \frac{p_s m_\ell + \sum_{\ell' \neq \ell} p_d m_{\ell'} + p_s(ns_\ell - m_\ell)w_\ell + \sum_{\ell' \neq \ell} p_d(ns_{\ell'} - m_{\ell'})w_{\ell'}}{np_s s_\ell + np_d(1-s_\ell)} \\ &= \frac{n-1}{n} \cdot \frac{p_s s_\ell M + \sum_{\ell' \neq \ell} p_d s_{\ell'} M + p_s(ns_\ell - s_\ell M)w^* + \sum_{\ell' \neq \ell} p_d(ns_{\ell'} - s_{\ell'} M)w^*}{np_s s_\ell + np_d(1-s_\ell)} \\ &= \frac{n-1}{n} \cdot \frac{M(p_s s_\ell + p_d(1-s_\ell)) + w^*(np_s s_\ell + np_d(1-s_\ell) - M(p_s s_\ell + p_d(1-s_\ell)))}{np_s s_\ell + np_d(1-s_\ell)} \\ &= \frac{n-1}{n} \cdot \frac{M + w^*(n-M)}{n} \end{aligned}$$

The above expression has no dependence on ℓ. Letting $w^* = \frac{(n-1)M}{(n-1)M+n}$, we see then that $w_\ell = w^*$, which completes the proof. $\square$

Proof of Proposition 4. When Bayesians are distributed equally, we saw in Proposition 3 that homophily has no effect (beliefs are the same on every island), so setting $p_d = p_d^o$ is optimal, and always feasible because it costs nothing. By Theorem 2, if each island is all Bayesian or DeGroot, then the network is impervious for some p_d if and only if it is impervious for $p_d = p_s$. By Theorem 3, when ε is small and the islands are equal sizes, decreasing homophily may lead to imperviousness but not vice-versa, so setting $p_d = p_s$ is optimal. $\square$

References

Acemoglu, Daron, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar (2013), "Opinion fluctuations and disagreement in social networks." *Mathematics of Operations Research*, 38, 1–27.

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi (2015), "Systemic Risk and Stability in Financial Networks." *American Economic Review*, 105, 564–608, URL https://www.aeaweb.org/articles?id=10.1257/aer.20130456.

Akbarpour, Mohammad, Suraj Malladi, and Amin Saberi (2018), "Diffusion, seeding, and the value of network information." In *Proceedings of the 2018 ACM Conference on Economics and Computation*, 641–641, ACM.

Alidaee, Hossein, Eric Auerbach, and Michael P Leung (2020), "Recovering network structure from aggregated relational data using penalized regression." *arXiv preprint arXiv:2001.06052*.

Arnold, Barry C. (1987), *Majorization and the Lorenz Order: A Brief Introduction*. Lecture Notes in Statistics, Springer-Verlag, New York, URL https://www.springer.com/gp/book/9780387965925.

Ata, Baris, Alexandre Belloni, and Ozan Candogan (2018), "Latent agents in networks: Estimation and pricing." *arXiv preprint arXiv:1808.04878*.

Auerbach, Eric (2019), "Measuring differences in stochastic network structure." *arXiv preprint arXiv:1903.11117*.

Babus, Ana (2016), "The formation of financial networks." *The RAND Journal of Economics*, 47, 239–272, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1756-2171.12126.

Billingsley, Patrick (1995), *Probability and Measure*, third edition. John Wiley and Sons.

Calvo-Armengol, Antoni and Matthew O Jackson (2004), "The effects of social networks on employment and inequality." *American economic review*, 94, 426–454.

Calvó-Armengol, Antoni and Matthew O Jackson (2007), "Networks in labor markets: Wage and employment dynamics and inequality." *Journal of economic theory*, 132, 27–46.

Candogan, Ozan and Kimon Drakopoulos (2017), "Optimal signaling of content accuracy: Engagement vs. misinformation." *Operations Research, forthcoming*.

Card, David and Alan B Krueger (1992), "School quality and black-white relative earnings: A direct assessment." *The Quarterly Journal of Economics*, 107, 151–200.

Chandra, Amitabh (2000), "Labor-market dropouts and the racial wage gap: 1940-1990." *American Economic Review*, 90, 333–338.

Chandrasekhar, Arun G, Horacio Larreguy, and Juan Pablo Xandri (2019), "Testing models of social learning on networks: Evidence from two experiments." *Econometrica*.

Charras-Garrido, Myriam and Pascal Lezaud (2013), "Extreme Value Analysis : an Introduction." *Journal de la Societe Française de Statistique*, 154, pp 66–97, URL https://hal-enac.archives-ouvertes.fr/hal-00917995.

Chung, Fan (2005), "Laplacians and the cheeger inequality for directed graphs." *Annals of Combinatorics*, 9, 1–19.

Chung, Fan and Mary Radcliffe (2011), "On the spectra of general random graphs." *Electronic Journal of Combinatorics*, 18.

Currarini, Sergio, Matthew O Jackson, and Paolo Pin (2009), "An economic model of friendship: Homophily, minorities, and segregation." *Econometrica*, 77, 1003–1045.

Dasaratha, Krishna (2019), "Distributions of centrality on networks." Forthcoming in Games and Economic Behavior.

Golub, Benjamin and Matthew O Jackson (2012), "How homophily affects the speed of learning and best-response dynamics." *The Quarterly Journal of Economics*, 127, 1287–1338.

Heckman, James J, Thomas M Lyons, and Petra E Todd (2000), "Understanding black-white wage differentials, 1960-1990." *American Economic Review*, 90, 344–349.

Holland, Paul W, Kathryn Blackmond Laskey, and Samuel Leinhardt (1983), "Stochastic blockmodels: First steps." *Social networks*, 5, 109–137.

Jackson, Matthew O (2010), *Social and economic networks*. Princeton university press.

Jadbabaie, Ali, Pooya Molavi, Alvaro Sandroni, and Alireza Tahbaz-Salehi (2012), "Non-bayesian social learning." *Games and Economic Behavior*, 76, 210–225.

Kanak, Zafer (2017), "Rescuing the Financial System: Capabilities, Incentives, and Optimal Interbank Networks." Technical Report 17-17, NET Institute, URL https://ideas.repec.org/p/net/wpaper/1717.html.

Keppo, Jussi, Michael Jong Kim, and Xinyuan Zhang (2019), "Learning manipulation through information dissemination." *Available at SSRN 3465030*.

Lobel, Ilan and Evan Sadler (2015), "Preferences, homophily, and social learning." *Operations Research*, 64, 564–584.

Manshadi, Vahideh, Sidhant Misra, and Scott Rodilitz (2018), "Diffusion in random networks: Impact of degree distribution." *Operations Research, forthcoming*.

Marsden, Peter V (1987), "Core discussion networks of americans." *American sociological review*, 122–131.

Marshall, Albert W., Ingram Olkin, and Barry C. Arnold (2011), *Inequalities: Theory of Majorization and Its Applications*, 2 edition. Springer Series in Statistics, Springer-Verlag, New York, URL https://www.springer.com/gp/book/9780387400877.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001), "Birds of a feather: Homophily in social networks." *Annual review of sociology*, 27, 415–444.

Mostagir, Mohamed, Asu Ozdaglar, and James Siderius (2019), "When is society susceptible to manipulation?" *Working paper*.

Papanastasiou, Yiangos (2020), "Fake news propagation and detection: A sequential model." *Management Science*.

Sadler, Evan (2019), “Influence campaigns.” *Available at SSRN 3371835*.

Taylor, Chris (2011), “Expectation of the maximum of gaussian random variables.” Mathematics Stack Exchange, URL https://math.stackexchange.com/q/89030. URL:https://math.stackexchange.com/q/89030 (version: 2011-12-06).

Yildiz, Ercan, Asuman Ozdaglar, Daron Acemoglu, Amin Saberi, and Anna Scaglione (2013), “Binary opinion dynamics with stubborn agents.” *ACM Transactions on Economics and Computation (TEAC)*, 1, 19.