Theoretical Economics 10 (2015), 807–851 1555-7561/20150807

Information diffusion in networks through social learning

Ilan Lobel
IOMS Department, Stern School of Business, New York University

Evan Sadler
IOMS Department, Stern School of Business, New York University

We study perfect Bayesian equilibria of a sequential social learning model in which agents in a network learn about an underlying state by observing neighbors' choices. In contrast with prior work, we do not assume that the agents' sets of neighbors are mutually independent. We introduce a new metric of information diffusion in social learning that is weaker than the traditional aggregation metric. We show that if a minimal connectivity condition holds and neighborhoods are independent, information always diffuses. Diffusion can fail in a well connected network if neighborhoods are correlated. We show that information diffuses if neighborhood realizations convey little information about the network, as measured by network distortion, or if information asymmetries are captured through beliefs over the state of a finite Markov chain.

Keywords. Social networks, Bayesian learning, information aggregation, herding.

JEL classification. C72, D83.

1. Introduction

Social networks play an increasingly important role in our lives, particularly in the formation and spread of beliefs and opinions. Many of our decisions are inherently social in nature, and we frequently rely on information gleaned from observing the actions of others. Our choices of what college to attend, what job offer to take, what smart phone to use, or what politician to support are often influenced by the decisions of our friends, colleagues, and neighbors.

In this paper, we study how the structure of the social network affects the equilibrium outcome in a sequential model of social learning. We consider a countably infinite set of agents, each endowed with a private signal about an underlying state of the world, taking actions according to an exogenous ordering. Following Acemoglu et al. (2011), we assume each agent observes a stochastically generated subset of predecessors' actions, which we call the agent's neighborhood. There is a common prior on the joint

Ilan Lobel: ilobel@stern.nyu.edu
Evan Sadler: esadler@stern.nyu.edu
We thank Nicola Persico and three anonymous referees for their thoughtful comments and suggestions. We thank the National Science Foundation (Grant CCF-1216004) and IBM (OCR Grant) for their financial support.

DOI: 10.3982/TE1549

distribution of all agents' neighborhoods, and we call this joint distribution the network topology. Unlike Acemoglu et al. (2011), we do not assume neighborhoods are generated independently, allowing arbitrary correlations between agents' neighborhoods. Consequently, when an agent observes her neighborhood, she forms updated beliefs about the structure of the rest of the network. This allows us to capture asymmetric information regarding the network topology that could influence how observed choices are interpreted.

Studying learning with correlated neighborhoods takes us a step closer to understanding the efficiency of learning in realistic settings. A classic example in the literature involves people arriving sequentially and choosing between two restaurants (Banerjee 1992). Suppose you observe that one restaurant has a long line while the other is nearly empty, and you consider two plausible scenarios to account for this observation. In one scenario, the individuals in line arrive sequentially and choose a restaurant using their own private information as well as their inferences based on the growing line. Alternatively, suppose a tour bus pulls up and the entire line of customers follows the tour leader to this particular restaurant. In the first scenario, we might infer that some individuals had strong signals indicating that this restaurant is of higher quality. In the second scenario, we might infer that everyone in line had a weak signal and simply copied the rest of the group. Our inference about the restaurant's quality will depend on how likely we think the first scenario is relative to the second—it will depend on our beliefs about the observation network, and those beliefs are far more complex if observations are correlated.

In a more contemporary example, suppose we observe via social media two friends sequentially purchase a coupon from a company like Groupon or Living Social. If these two friends know each other well, we should realize the decisions might provide partially redundant information because one made the purchase decision after seeing the other do so. If we believe the two are unlikely to know one another, we can make a stronger inference about the value of the coupon since the decisions were made independently. Two people observing this same pair of decisions may reach different conclusions when given different knowledge of the social network: information about the network structure can impact how observations are interpreted. Differences in information about the network, even when agents make similar observations, could alter the outcome of social learning.

According to the last two decades of economics scholarship, herding is the key inefficiency that arises in social learning models. Herding occurs when rational agents choose to disregard their own private signals, preferring to copy their neighbors' actions instead. When agents herd, they prevent others from obtaining any information about their private signals; this limits the amount of new information that becomes available to the broader society. While our model still captures herd behavior, we find that correlated neighborhoods may lead to more severe modes of failure.

To distinguish this new kind of failure, we introduce a new metric of social learning. The traditional metric for a successful outcome is one of aggregation: whether as society grows large, later agents approach certainty about the underlying state of the world. This is a stringent criterion, yet society can sometimes achieve complete aggregation because there are many independent signals dispersed among different people. However, full aggregation is unlikely when agents only observe the discrete actions taken by their neighbors, rather than actual signals or beliefs. Complete aggregation requires either strong assumptions on the network topology (Acemoglu et al. 2011) or that private signals are of unbounded strength (Smith and Sørensen 2000).

Our new metric of social learning highlights that individuals can benefit substantially from observing the actions of their peers even if full information aggregation does not occur. Suppose you have a friend who always carefully researches every purchase she makes; when you decide to copy one of her purchases, you know there is a chance she erred, but copying likely leads to a better decision than you would make independently. Formally, consider an expert as an agent outside the network who draws a signal from a different distribution than the other agents. This signal takes one of two possible values, leading respectively to the strongest private belief that an agent in the network could have in favor of either state. Experts may not be infallible, but if all members of society can achieve the same ex ante probability of making a good decision as an expert, we say that information *diffuses*. Information diffusion captures the idea that the strongest available signals are transmitted throughout the network.

Both aggregation and diffusion have a role to play in the study of social learning in networks. When everyone in society already has access to the strongest signals or the strongest signals provide little information, aggregation is a far more interesting metric: diffusion is achieved trivially in these contexts. However, if strong signals are rare but informative, the question of diffusion becomes paramount. The two metrics also provide complementary insights. Previous work studying aggregation calls attention to how the signal structure impacts learning. As long as the network is sufficiently connected, the presence of bounded or unbounded private beliefs usually determines whether information aggregates. Our new metric of diffusion provides results that are independent of the private signals, shifting the focus onto the role of network structure and beliefs about that structure.

One reason prior literature has focused on aggregation rather than diffusion is that diffusion trivially obtains when agents have identical information about the broader network. As a special case of Theorem 1, we show that the complete network studied in early papers (e.g., Banerjee 1992, Bikhchandani et al. 1992, and Smith and Sørensen 2000) always diffuses information. Our theorem also captures results on deterministic network topologies (Çelen and Kariv 2004) as well as stochastic topologies in which agents have independently drawn neighborhoods (Banerjee and Fudenberg 2004, Smith and Sørensen 2013, Acemoglu et al. 2011). As long as agents' neighborhoods are independent, a mild connectivity condition referred to as expanding observations is both necessary and sufficient for the network to diffuse information.

As we emphasize in Proposition 1, this implies that information cascades—when all agents after some time disregard their private signals and copy their neighbors—only occur after information has diffused through the network. In the complete network, a cascade always occurs in a limiting sense, suggesting an equivalence between diffusion and cascades; however, we show through an example that the two are distinct. In more general networks, some realizations may lead agents to continue relying on their private signals, even if on average the available social information overwhelms these signals.

Once we allow neighborhood correlations, neighborhood realizations convey information about the network, and expanding observations is no longer a sufficient condition for information diffusion. Theorem 2 generalizes the basic intuition that a successful network must be sufficiently connected. More surprisingly, we find that information diffusion can fail even in strongly connected networks. A series of examples demonstrates how failures of diffusion can result from agents having substantially different information about the network. In one case, diffusion fails in a well connected network because agents cannot identify an information path even though it exists. A more complex example shows how rational behavior can mimic overconfidence, leading agents to copy the actions of their neighbors too frequently for information to diffuse. Finally, we demonstrate that the actions of a large collection of neighbors could be rendered uninformative in some realizations of a network. That even the weaker metric of information diffusion can fail in a well connected network is an important finding of our work.

These examples raise serious doubts about the robustness of positive learning results in networks. Individuals often do have different information about the networks in which they are embedded, and popular generative network models—for instance, those based on preferential attachment mechanisms—induce neighborhood correlations. The existing social learning literature is silent on outcomes in such contexts. We provide two positive results in an effort to fill this gap, suggesting that information diffusion is generally robust.

First, Theorem 3 demonstrates that information diffuses in a well connected network as long as the network satisfies a *low distortion* property. We interpret distortion as a focused measure of the information that neighborhood realizations provide to the agents. If the realizations provide little information relative to this measure, then diffusion occurs through the same mechanisms as in simpler network models. Interestingly, low distortion is a far weaker condition than neighborhood independence assumed in prior work. Using this result, we show that any network with long deterministic information paths diffuses information regardless of how informative neighborhood realizations are about the rest of the network structure. An additional example shows that preferential attachment networks exhibit zero distortion despite strong neighborhood correlations. These findings highlight that the failures of information diffusion in our earlier examples are not the result of neighborhood correlations per se; rather, it is the information neighborhood correlations provide about the broader network that can create problems.

A second result concerns neighborhood correlations that arise through aggregate shocks to the network. We can imagine situations in which correlations between the observations different agents make are the result of some exogenous event, such as an Internet outage, that similarly affects all of the agents. If the neighborhoods of all agents are mutually independent conditional on the realization of such a shock, Theorem 4 shows that a connectivity condition still characterizes networks that diffuse information. In fact, our theorem establishes a stronger result, showing that as long as neighborhood correlations are fully captured via an underlying Markov process with finitely many states, connectivity is sufficient for diffusion.

Our paper makes two key contributions to the study of social learning. First, we explore the impact of correlated neighborhoods on learning, demonstrating potentially severe failures. Our findings illustrate previously unstudied outcomes of social learning, and we shed light on the robustness of earlier results. Second, we introduce and study a new metric of information diffusion. This metric helps unify the findings of prior work, particularly the disconnect between learning results obtained with bounded and unbounded private signals, and provides some measure of how significant herding inefficiencies can be. Moreover, our metric provides a lens to focus attention on the impact of the network structure on the learning process, in contradistinction to the impact of the signal structure.

Our work also suggests a few empirical implications. We highlight that network transparency may play a role that facilitates information diffusion independently of the actual structure of the network. To the extent that the visibility of connections in online social networks reduces information asymmetry about the social network structure, we would predict that these communities encourage the diffusion of information over and above what we would expect based purely on the increased visibility of individuals' actions. Researchers have motivated the study of information cascades to explain widespread conformity in the adoption of certain behaviors and products (Bikhchandani et al. 1992). If the greater diffusion of information facilitated through network transparency leads to more information cascades, we should expect more conformity in individual behavior following the growth of social media.

1.1 *Related literature*

The social learning literature has its origins in the seminal papers of Banerjee (1992) and Bikhchandani et al. (1992). These papers demonstrated that fully rational agents who learn by observing the actions of their peers are susceptible to herding, an inefficient equilibrium outcome in which agents ignore their own private signals and copy the actions of their neighbors instead. This outcome of social learning is often cited as an explanation for widespread behavioral conformity in a variety of contexts. Smith and Sørensen (2000) showed that in a complete network, inefficient herding outcomes occur as long as private signals are of bounded strength. Çelen and Kariv (2004) extended the analysis to a line topology, in which agents observe only the most recent action. Banerjee and Fudenberg (2004) and Smith and Sørensen (2013) incorporated the idea of neighbor sampling, with the former considering a model with a continuum of agents and the latter studying a generalization of their own earlier model. Our work builds most closely on that of Acemoglu et al. (2011), in which agents are situated in a social network and can only observe the actions of their neighbors. These earlier analyses suggest that Bayesian agents will generally settle on the right action as long as there exist arbitrarily strong signals and the network satisfies some minimal connectivity condition.

While the papers above study perfect Bayesian equilibria of social learning games, a parallel literature has emerged on non-Bayesian or partially Bayesian network learning models. Some early papers in this stream of work include Ellison and Fudenberg (1993, 1995), Bala and Goyal (1998), DeMarzo et al. (2003), and Gale and Kariv (2003).

More recent papers include Golub and Jackson (2010, 2012), Acemoglu et al. (2010), and Guarino and Ianni (2010). Non-Bayesian models are often motivated by the complexity of equilibrium behavior in dynamic network games with incomplete information. Our model provides an alternative approach to modeling learning in complex networks. The model we construct, while still imposing constraints on the problem (such as the sequential nature of the game), allows us to consider the perfect Bayesian equilibria of learning models in networks that are both complex and realistic, such as the preferential attachment network of Barabási and Albert (1999).

An important question that remains largely unanswered is the extent to which the Bayesian and the non-Bayesian network learning models offer similar predictions. If the dynamics of the easier-to-analyze non-Bayesian models are similar to the equilibria of Bayesian models, then one could argue in favor of considering the simpler models rather than the more complex ones. Jadbabaie et al. (2012) demonstrate that the predictions of both kinds of models often converge asymptotically. However, we argue that the key driver of inefficiency in learning is not the network structure itself, but persistent differences in beliefs about the network structure, leading to different interpretations of the same information. Deriving simple non-Bayesian models that accurately capture the learning difficulties created by this asymmetry is a substantive challenge.

The social learning literature is, of course, larger than the set of papers we discuss here. It also includes models where agents communicate signals or beliefs, such as Eyster and Rabin (2011), Acemoglu et al. (2014), and Fan et al. (2012), as well as models in which agents have collective preferences, such as Ali and Kartik (2012). Jackson (2007) and Acemoglu and Ozdaglar (2011) provide excellent surveys of the field.

1.2 *Organization*

The rest of the paper is structured as follows. Section 2 presents our model. Section 3 discusses measures used to assess learning outcomes, arguing that the measures used in prior work are too demanding to study settings with correlated neighborhoods. Section 4 explores how social learning can fail to diffuse information in complex networks. Section 5 presents our positive results, establishing conditions under which diffusion succeeds. Section 6 concludes. All proofs can be found in the Appendix.

2. Model

A set of agents, indexed by $n \in \mathbb{N}$, sequentially choose between two mutually exclusive actions labeled 0 and 1. Denote the action of agent n by $x_n \in \{0, 1\}$. The payoff from this choice depends on an unknown underlying state of the world θ. There are two possible underlying states, again labeled 0 and 1, and we specify the utility of agent n derived from act x_n:

$$u_n(x_n, \theta) = \begin{cases} 1 & \text{if } x_n = \theta \\ 0 & \text{if } x_n \neq \theta. \end{cases}$$

That is, all agents derive strictly higher utility from choosing the action with the same label as the underlying state of the world. To simplify notation, we further specify that a priori the probability of each state is $\mathbb{P}(\theta=0)=\mathbb{P}(\theta=1)=\frac{1}{2}$. We discuss how to generalize our results to other priors and utility functions in Section 3.

To make an optimal decision, agents seek to learn the true state of the world. Agents have access to two types of relevant information: private information and social information. An agent's private information consists of a private signal observable only to the agent. Prior to making a decision, agent n learns the value of a random variable s_n taking values in some metric space $\mathcal{S}$. Conditional on the state θ, each agent's signal is independently drawn from the distribution $\mathbb{F}_\theta$, and the pair of probability measures $(\mathbb{F}_0, \mathbb{F}_1)$ constitutes the signal structure of the model. The only assumption we impose on the signal structure is that the two measures are not almost everywhere equal; thus, the probability of receiving an informative signal is positive.

The probability $p_n=\mathbb{P}(\theta=1 \mid s_n)$ constitutes agent n's private belief, and the support of the private beliefs is the region $[\underline{\beta}, \overline{\beta}]$, where

$$\underline{\beta}=\inf\{r\in[0,1] \mid \mathbb{P}(p_1\leq r)>0\}, \quad \text{and} \quad \overline{\beta}=\sup\{r\in[0,1] \mid \mathbb{P}(p_1\leq r)<1\}.$$

The distinction between bounded ($\underline{\beta}>0$ and $\overline{\beta}<1$) and unbounded ($\underline{\beta}=0$ and $\overline{\beta}=1$) private beliefs has proved important in prior work on social learning,[1] but plays a minor role in this paper due to our focus on a different learning metric.

An agent's social information is derived from observing some subset of the past actions of other agents. The set of agents whose actions are observed by agent n, denoted by $B(n)\subseteq\{1,2,\ldots,n-1\}$, is called n's neighborhood. The set $B(n)$ itself is a random variable, and the sequence of neighborhood realizations describes a social network of connections between the agents. We assume that the neighborhood realizations $B(n)$ are independent of the state θ. The probability $q_n=\mathbb{P}(\theta=1 \mid B(n), x_m, m\in B(n))$ is called agent n's social belief.

We describe the structure of the social network via a probability measure $\mathbb{Q}$, giving a distribution over all possible sequences of neighborhoods. Formally, $\mathbb{Q}$ is a probability measure on the product space $\mathbb{B}=\prod_{i=1}^{\infty} 2^{\mathbb{N}_i}$, where $\mathbb{N}_i=\{x : x\in\mathbb{N}, x<i\}$. A particular measure $\mathbb{Q}$ on the space $\mathbb{B}$ is called a network topology. The topology is deterministic if $\mathbb{Q}$ is a Dirac distribution concentrated on a single element $\mathbf{B}\in\mathbb{B}$; otherwise the topology is stochastic.

We assume that the signal structure $(\mathbb{F}_0, \mathbb{F}_1)$ and the network topology $\mathbb{Q}$ are common knowledge to all agents. Before making a decision, agent n observes the values of s_n, $B(n)$, and x_m for each $m\in B(n)$. Let $\mathcal{I}_n$ denote the set of all possible values of agent n's information set. Agent n's strategy is defined by a function $\sigma_n : \mathcal{I}_n\to\{0,1\}$ that maps a realization of this information set to a decision. A strategy profile, denoted by σ, is a sequence of strategies for each agent. For notational convenience, let $\sigma_{-n}=\{\sigma_1,\ldots,\sigma_{n-1},\sigma_{n+1},\ldots\}$ denote the set of all strategies other than agent n's, allowing us to represent the strategy profile as $\sigma=(\sigma_n,\sigma_{-n})$. Given a particular strategy profile, the sequence of actions $\{x_n\}_{n\in\mathbb{N}}$ is a stochastic process with a measure $\mathbb{P}_\sigma$ generated by the signal structure and the network topology.

[1]See Smith and Sørensen (2000), Smith and Sørensen (2013), Acemoglu et al. (2011), and Mossel et al. (2012) among others.

The solution concept we consider is the set of perfect Bayesian equilibria of this social learning game, hereafter referred to simply as the equilibria of the game. A strategy profile is an equilibrium if for every n, the strategy σ_n maximizes agent n's expected utility given the strategies of the other agents, σ_{-n}. For a particular information set $I_n \in \mathcal{I}_n$, agent n's expected utility from action y is simply $\mathbb{P}_{(y,\sigma_{-n})}(y = \theta \mid I_n)$. Therefore, agent n's decision in equilibrium is

$$x_n = \sigma_n(I_n) \in \underset{y \in \{0,1\}}{\arg\max}\, \mathbb{P}_{(y,\sigma_{-n})}(y = \theta \mid I_n).$$

Given any set of strategies for the agents acting prior to n and any realization of n's information set $I_n \in \mathcal{I}_n$, this maximization problem has a well defined solution. Thus, an inductive argument shows that the set Σ of perfect Bayesian equilibria of this game is nonempty. The set Σ will contain multiple equilibria if some agents are indifferent between their two choices.

The model studied by Acemoglu et al. (2011) is a special case of the one considered here. Acemoglu et al. (2011) assume all neighborhoods are generated independently from one another, but in our paper, arbitrary topologies are allowed. This difference between the two models might appear small at first glance, but it significantly changes our conclusions about social learning outcomes. Correlations between agents' neighborhoods imply that agents sometimes disagree over the likely composition of others' neighborhoods, potentially leading to a lack of consensus over who is well connected or well informed.

3. The metrics of social learning

The classical metric of learning is whether decisions converge on the fully informed optimal action. As society grows, the number of independent private signals grows. In an ideal world, social learning would aggregate all of these signals, allowing later agents to approach certainty by drawing wisdom from an ever expanding group of peers.[2] This represents the best possible asymptotic result of the social learning process—the same limiting outcome that would occur if each agent directly observed the private signals of all prior agents.[3] Given a network topology, a signal structure, and an equilibrium strategy profile, we say that *information aggregates* if

$$\lim_{n\to\infty} \mathbb{P}_\sigma(x_n = \theta) = 1.$$

While this presents a high bar for social learning to meet, aggregation does indeed occur in many models. In prior papers, successful aggregation often turns on whether the signal structure features unbounded or bounded private beliefs. A consensus has

[2]Acemoglu et al. (2011) capture this idea in their definition of asymptotic learning. Smith and Sørensen (2000) refer to it as complete learning.

[3]The notion of perfect learning studied by Lee (1993) and Arieli and Mueller-Frank (2012) is an even more stringent, nonasymptotic criterion, requiring individuals to act as though they have observed all previous agents' private signals.

emerged that unbounded beliefs robustly lead to aggregation in many models and contexts. Under varying assumptions, Smith and Sørensen (2000), Smith and Sørensen (2013), Acemoglu et al. (2011), and Mossel et al. (2012) each establish positive learning results when private beliefs are unbounded. In contrast, the seminal herding results of Banerjee (1992) and Bikhchandani et al. (1992), expanded by Smith and Sørensen (2000) and Acemoglu et al. (2011), demonstrate that aggregation often fails when private beliefs are bounded. One way to achieve aggregation with bounded private beliefs is to allow a rich enough action space that the content of an individual's private signal is largely revealed through her choice. With finite action spaces, there are a few examples in the literature of bounded beliefs leading to successful aggregation—notably, Theorem 4 of Acemoglu et al. (2011)—but aggregation only occurs in networks that are precisely constructed to achieve this outcome. For those rare networks in which social learning aggregates information in any equilibrium regardless of the signal structure,[4] we say the *network aggregates information.*

We argue two reasons why new metrics deserve attention. First, results based on the aggregation metric rely heavily on the distinction between bounded and unbounded private beliefs. While we can measure such details of the signal structure in some contexts, in others we may need to base predictions on other salient model parameters. In particular, a metric leading to results based on properties of the network structure could provide a useful alternative.[5]

Second, aggregation is often too stringent a criterion for judging the efficiency of equilibrium outcomes. Individuals may derive a substantial benefit from their social information without obtaining full knowledge of the underlying state of the world. Consider a technology adoption problem in which most consumers are completely uninformed about the merits of a particular product, having only an even chance of making a good adoption decision on their own. However, a few individuals have strong signals; they understand the technology and are 95% confident in their assessment of its value. If this information diffuses so that all agents have a 95% chance of making the correct choice, the traditional metric would say that learning has failed. While social learning has not fully aggregated information, agents all perform (weakly) better than any individual would by herself. When this occurs, learning has reached a significant threshold, and a theory of social learning ought to separate this outcome from more significant failures. To differentiate such outcomes, we motivate a less exacting benchmark of information diffusion.

Consider a hypothetical expert situated outside the social network. We suppose this expert has access to the most informative binary signal $\tilde{s}$ supported on $[\underline{\beta}, \overline{\beta}]$. That is, the expert always receives one of the strongest signals an agent in the network could

[4]Recall that the model assumes that an agent's probability of receiving an informative signal is positive; this assumption is essential for the definition to be meaningful.

[5]Mueller-Frank and Pai (forhcoming) take an another approach to this issue; their work makes private information endogenous by modeling it as the result of a costly search process. A major finding is that aggregation occurs in the complete network if and only if some agents have arbitrarily low search costs, effectively trading an assumption of strong signals for one of low search costs.

observe, leading to a private belief of either $\underline{\beta}$ or $\overline{\beta}$. In the absence of any social information, a straightforward calculation shows that the expert chooses the optimal action with probability

$$\mathbb{P}(x^* = \theta) \equiv \beta^* = \frac{(1-\underline{\beta})(2\overline{\beta}-1) + \overline{\beta}(1-2\underline{\beta})}{2(\overline{\beta}-\underline{\beta})}.$$

Given a signal inducing belief $\overline{\beta}$, the expert chooses action 1 and is correct with probability $\overline{\beta}$. Similarly, given a signal inducing belief $\underline{\beta}$, the expert chooses action 0 and is correct with probability $1-\underline{\beta}$. The probability β^* is a weighted average of these two outcomes. When the support of private beliefs is symmetric ($\underline{\beta} = 1-\overline{\beta}$), the expert is equally likely to receive each signal, and this probability reduces to an intuitive expression $\beta^* = \overline{\beta} = 1-\underline{\beta}$. When the support of private beliefs is asymmetric, the expert is relatively more likely to receive the less informative of the two signals. When the support of private beliefs is extremely asymmetric (for instance, $\underline{\beta} = 0.49$, $\overline{\beta} = 0.99$), the expert does only slightly better than flipping a coin; given even prior probabilities on the two states, the strongly informative signal must occur infrequently relative to the weakly informative one.

Given a network topology, a signal structure, and an equilibrium strategy profile, we say that *information diffuses* if

$$\liminf_{n\to\infty} \mathbb{P}_\sigma(x_n = \theta) \geq \beta^*.$$

As we did for aggregation, we say a *network diffuses information* if information diffuses for any signal structure and any equilibrium.

This notion of diffusion does not require society to fully incorporate infinitely many signals, and it does not require agents to make near-perfect decisions as society grows large. The aim is more modest, requiring only that the strongest signals spread through the network, and agents perform as well as our fictitious expert in the limit as the network grows large. The following definition summarizes our learning metrics.

Definition 1. Given a network topology, a signal structure, and an equilibrium strategy profile, *information aggregates* if $\lim_{n\to\infty} \mathbb{P}_\sigma(x_n = \theta) = 1$, and *information diffuses* if $\liminf_{n\to\infty} \mathbb{P}_\sigma(x_n = \theta) \geq \beta^*$. If information aggregates (diffuses) in any equilibrium for any signal structure, we say *the network aggregates (diffuses) information*.

We do not argue that diffusion, rather than aggregation is the single correct metric for social learning. The two metrics capture different aspects of learning. If the signal structure is binary, then the agents' signal distribution is the same as an expert's, and our notion of learning is trivial. When all signals give at most 51% confidence about the state of the world, successful diffusion is not a particularly striking outcome either. Diffusion can be a very weak metric in some cases. However, most realistic scenarios will include a mix of agents, some with strong signals and some with weak signals. In these settings, finding that all agents perform as well as an expert in the limit is far from a trivial statement.

Note that although this metric is generally weaker than aggregation, the two metrics coincide when private beliefs are unbounded: unbounded private beliefs imply that the fictitious expert has perfect information. Since much of the past literature required unbounded private beliefs to obtain positive results, one could interpret such findings as characterizations of information diffusion, rather than information aggregation.

The literature on social learning has focused on the more strict criterion of aggregation for a clear reason: failure of information diffusion, as we have defined it, is atypical when neighborhoods are independently generated. For example, the complete network in which $B(n) = \{1, \dots, n-1\}$ for all n diffuses information. More generally, any network with a minimal level of connectivity diffuses information. To make this precise, recall the definition of expanding observations from Acemoglu et al. (2011).

Definition 2. A network topology $\mathbb{Q}$ features *expanding observations* if for all positive integers K,

$$\limsup_{n\to\infty} \mathbb{Q}\Big(\max_{b\in B(n)} b < K\Big) = 0.$$

As argued by Acemoglu et al. (2011), this is a mild condition guaranteeing that agents at least have indirect access to enough information for learning to occur. This restriction is necessary to rule out trivial failures—for instance, the one that occurs if $B(n) = \varnothing$ for every agent n.

Theorem 1. *Suppose neighborhoods are mutually independent in the network topology $\mathbb{Q}$. Then $\mathbb{Q}$ diffuses information if and only if $\mathbb{Q}$ features expanding observations.*

Barring neighborhood correlations and barring networks without expanding observations, information diffusion always occurs. The prior literature has concerned itself solely with aggregation because diffusion is almost trivially attained in these models without neighborhood correlations. The inefficient herding results of Banerjee (1992) and Bikhchandani et al. (1992) are failures of aggregation but not of diffusion. Although agents may herd on the incorrect action, they are far more likely to herd on the correct one. To clarify the role of herding in social learning models, we extend the definition of information cascades from Bikhchandani et al. (1992).

Definition 3. An *information cascade* occurs if there exists an integer valued random variable N such that all agents $n \geq N$ ignore their private signals, relying entirely on their social information.

Proposition 1. *If an information cascade occurs, information diffuses.*

When an information cascade occurs, agents act according to their social information regardless of their private signals. Even agents with the strongest private information ignore their signals in favor of social information. This can only occur if the information obtained from the social network is at least as informative as the strongest signals, which by definition implies that information diffuses. This is not to say that we consider herding a positive outcome, but that the diffusion metric provides a bound on how inefficient a herd can be.

The converse is false. Diffusion of information does not imply that we have an information cascade, even in the weak sense of a limit cascade as defined by Smith and Sørensen (2000). Consider a deterministic network with three distinct groups of agents. Agents in the set $S_1 \equiv \{n : n = 3k - 2, k \in \mathbb{N}\}$ observe all previous agents in S_1 and only those agents. Likewise, agents in the set $S_2 \equiv \{n : n = 3k - 1, k \in \mathbb{N}\}$ observe all previous agents in S_2 and only those agents. The remaining agents in $S_3 \equiv \{n : n = 3k, k \in \mathbb{N}\}$ have the neighborhoods $B(n) = \{n - 1, n - 2\}$.

An agent in S_3 observes the most recent member of each of the other two groups. If signals are symmetric and the two neighbors choose opposite actions, the two observations balance exactly, generating a social belief of $\frac{1}{2}$. Hence, the agent in S_3 relies on her own signal as though there were no neighbors. If the signals induce bounded private beliefs, then with positive probability, agents in S_1 herd on one action while agents in S_2 herd on the other. With positive probability, all later agents in S_3 have social belief equal to $\frac{1}{2}$. There is no sense in which the social belief of agents in S_3 converges to a set of cascading beliefs, but our theorem implies we still get diffusion. This shows an important conceptual distinction between cascades and diffusion: the former represents a notion of belief convergence, while the latter is fundamentally about utility levels.

Nevertheless, the notion of a cascade and our metric of diffusion share a close relationship. Define the *cascade set* for social beliefs as the union $[0, 1 - \overline{\beta}] \cup [1 - \underline{\beta}, 1]$. If q_n is in the cascade set, then no realization of the private signal will compel agent n to change her behavior; at most she is indifferent to changing her a priori preferred action. Let $\tilde{h}$ denote a binary signal inducing private beliefs supported on $1 - \overline{\beta}$ and $1 - \underline{\beta}$. This is the weakest signal that guarantees a private belief in the cascade set. Call this the *minimal cascade signal*, and call an agent outside the network with access to this signal the *cascade expert*. We could just as easily have defined the diffusion metric in terms of the cascade expert as the cascade expert is exactly as likely to match the state as the expert.

This relationship between the expert signal and the minimal cascade signal holds more generally for an arbitrary prior over the two states and an arbitrary utility function $u(x, \theta)$ satisfying $u(1, 1) > u(0, 1)$ and $u(0, 0) > u(1, 0)$. We can define the expert signal $\tilde{s}$ as before, and the minimal cascade signal $\tilde{h}$ is characterized by the property that an agent receiving both signals is indifferent between the two actions whenever the signals conflict. Consequently, always following the expert signal is optimal, and always following the minimal cascade signal is optimal, so the expected utility derived from having either signal (or even both signals) is the same. If we then define the utility level of the expert u^*, an improvement principle still holds, and our later positive results have a clear analog in the general case.

We can interpret Theorem 1 in terms of the underlying mechanisms that lead to inefficient outcomes. In the absence of neighborhood correlations, there are two potential sources of inefficiency: limited observations and information externalities. The most severe failures are caused by the limited observations a disconnected network offers. We can attribute the distance between diffusion and aggregation to an information externality. To the extent that individual decisions rationally rely on social as opposed to private information, this private information is left unrevealed. The asymptotic impact of this externality is limited as a consequence of the "overturning principle" that Smith and Sørensen (2000) describe. Societal learning continues as long as it remains possible for a single agent to receive a signal overturning the sum of the available social information. Hence, the support of private beliefs provides one measure of the severity of this externality.

We see in the next section that our more general setting with correlated neighborhoods introduces asymmetric information about the network as a new source of inefficiency in the learning process. In a deterministic network, there is no uncertainty about the network structure, so there can be no information asymmetry regarding the network. When neighborhoods are stochastic, but independently generated, there are information asymmetries; however, the asymmetries are localized. The realization of an agent's neighborhood conveys no information about the rest of the network. These local asymmetries are minor and disappear in the limit. One can intuitively think of this as a "law of large numbers" type of result. With neighborhood correlations, asymmetries in information about the network are no longer localized, and they can persist, leading to new modes of failure. A major finding is that even the weaker standard of diffusion can fail when there are significant asymmetries.

4. Failure to diffuse information

This section examines several of the interesting and surprising ways that information may fail not just to aggregate, but even to diffuse across a network when there are significant correlations between the neighborhoods of different agents. We first revisit networks in which agents lack the connections needed for information to diffuse. In simple network topologies, nonexpanding observations sharply characterizes this type of failure, but an example shows that our model requires a more robust condition to capture the same intuition.

Example 1 (Failure with expanding observations). Let E_n denote the set of agents $m < n$ whose neighborhoods are empty, and define the network topology $\mathbb{Q}$ such that for each n, $B(n)$ is empty with probability $1/2^{|E_n|}$ and $B(n) = \{m \mid m = \max_{i \in E_n} i\}$ otherwise. $\Diamond$

This network topology plainly features expanding observations, yet all agents in the network either have an empty neighborhood or observe one agent whose neighborhood is empty in turn. The network will fail to diffuse information for exactly the same reason that a network with nonexpanding observations fails. To more precisely characterize failures caused by limited observation, we define an agent's *personal subnetwork* and the *expanding subnetworks* property.

Definition 4. An agent m is a member of agent n's *personal subnetwork* if there exists a sequence of agents, starting with m and terminating on n, such that each member of the sequence is contained in the neighborhood of the next. The personal subnetwork of agent n is denoted by $\hat{B}(n)$.

A network topology $\mathbb{Q}$ features *expanding subnetworks* if

$$\limsup_{n\to\infty} \mathbb{Q}\big(|\hat{B}(n)| < K\big) = 0$$

for all positive integers K.

Agent n's personal subnetwork represents the set of all agents in the network connected to n, either directly or indirectly, as of the time n must make a decision. If n's personal subnetwork is bounded in size, then the information available to n is fundamentally limited. Without expanding subnetworks, a network topology generates an infinite subsequence of agents whose social information is subject to a fixed limit on the number of private signals from which to draw. We intuitively expect information diffusion to fail in these networks, and our next theorem obliges.

Theorem 2. *If the network topology $\mathbb{Q}$ does not feature expanding subnetworks, then $\mathbb{Q}$ fails to diffuse information.*

When neighborhood realizations are mutually independent, expanding subnetworks and expanding observations are, in fact, equivalent conditions. This result is in the same spirit as Theorem 1 from Acemoglu et al. (2011); our refined notion of expanding subnetworks more directly captures the idea of drawing information from a growing number of signals and rightly excludes networks like that described in Example 1.

When neighborhoods are independent, Theorem 2 describes the only way failure occurs and the only mechanism through which information diffusion fails that is studied in prior work. This mode of failure differs fundamentally from those we study in the rest of this section. Failure in a poorly connected network has nothing to do with information asymmetries or externalities: even a social planner with full knowledge of the network realization is unable to dictate a sequence of decision rules that diffuses the information. When we introduce neighborhood correlations, the situation changes. Asymmetric information about the network structure can lead to similarly severe failures, but in all of the examples to follow, a social planner with full knowledge of the network could ameliorate the problem.

There are at least two analytically distinct ways that a network topology can feature expanding subnetworks. One is to ensure the existence of arbitrarily long "chains" of agents where each agent in the chain observes the previous one. Another occurs if agents directly observe unboundedly many others. In neither case is the network guaranteed to diffuse information. We begin by examining the first type of network, providing the following definitions to facilitate our discussion.

Definition 5. An *information path* for agent n is a set $\{a_1, a_2, \dots, a_k\}$ of agents such that $a_k = n$ and $a_i \in B(a_{i+1})$ for each $i < k$. Consider the set $\mathbf{S}_n$ comprised of all information paths for agent n. The *depth* of an agent n is defined as $d(n) = \max_{s\in\mathbf{S}_n} |s|$. A network topology is *deep* if for any positive integer K,

$$\limsup_{n\to\infty} \mathbb{Q}(d(n) < K) = 0.$$

Deep network topologies are precisely those that always develop long paths. We analyze two examples of how information can fail to diffuse in a deep network topology—described in turn as "failure due to unidentifiable paths" and "failure due to overconfidence." Failure due to unidentifiable paths can occur when an agent observes several neighbors and knows one of them has a long information path. A problem arises if the agent cannot identify this neighbor. We use the notation $S_i = \{2^i, 2^i + 1, \ldots, 2^{i+1} - 1\}$ for any $i \in \mathbb{N}$ to facilitate the exposition of examples.

Example 2 (Failure due to unidentifiable paths). Consider the following network topology:

- Exactly one of $B(2)$ and $B(3)$ is empty, and the other is equal to $\{1\}$: $B(2)$ and $B(3)$ are equally likely to be empty.
- For each S_i with $i > 1$, there is exactly one $n \in S_i$, chosen uniformly at random, such that $B(n)$ is empty. This agent is denoted by e_i.
- All other agents $n \in S_i$ have $B(n) = \{e_{i-1}, m_{i-1}\}$, where m_{i-1}, fixed for all $n \in S_i$, is a member of $S_{i-1} \setminus \{e_{i-1}\}$ chosen uniformly at random.

This network topology fails to diffuse information. ◊

In the network topology of Example 2 (shown in Figure 1), there are two types of agents. A small minority are unlucky and must make a decision without observing any neighbors, while the rest have two neighbors. The network features expanding subnetworks since for all $n \in S_i$, n lies at the end of an information path of length $i + 1$ with probability $1 - 2^{-i}$. However, whenever n has a long information path, n observes two neighbors: one always has an empty neighborhood, while the other is sure to have a long information path. Unfortunately, there is no way for agent n to tell her two neighbors apart. The two neighbors will occasionally choose different actions, and when this happens, agent n is forced to rely on her own signal.

We can avoid the challenge of identifiability of paths in deep networks if we limit ourselves to networks in which agents all have at most one neighbor. However, the next example shows that a well connected network topology can still fail to diffuse information. In the example, agents have no difficulty identifying a neighbor with a long information path, but agents disagree about the realized structure of the network. There are persistent information asymmetries regarding the network structure, leading agents to form different posterior beliefs about the network realization. As a result, key agents copy their neighbors too frequently. While we do not intend to imply that the agents suffer from some psychological bias—they are Bayesians after all—we refer to this as a failure due to overconfidence because the dynamics mimic what might happen if some agents were misguidedly optimistic about their neighbors' knowledge.

Example 3 (Failure due to overconfidence). Consider the following network topology: with probability $\frac{1}{2}$, $B(n) = \{n - 1\}$ for all $n > 1$; otherwise, the following statements hold.

- We have $B(2) = \{1\}$ and $B(3) = \{2\}$.

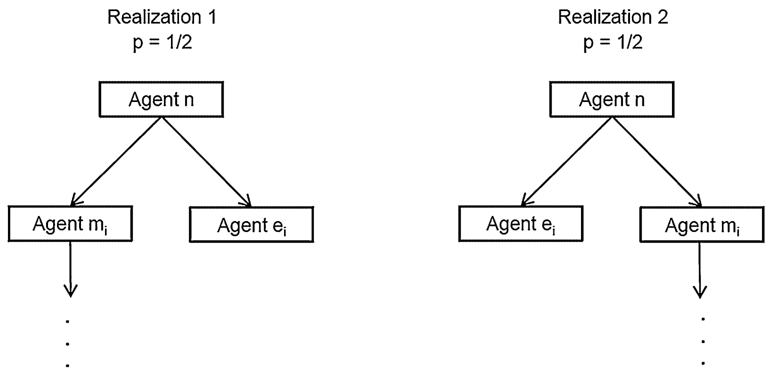

FIGURE 1. The network of Example 2. A typical agent n does not know which realization has occurred.

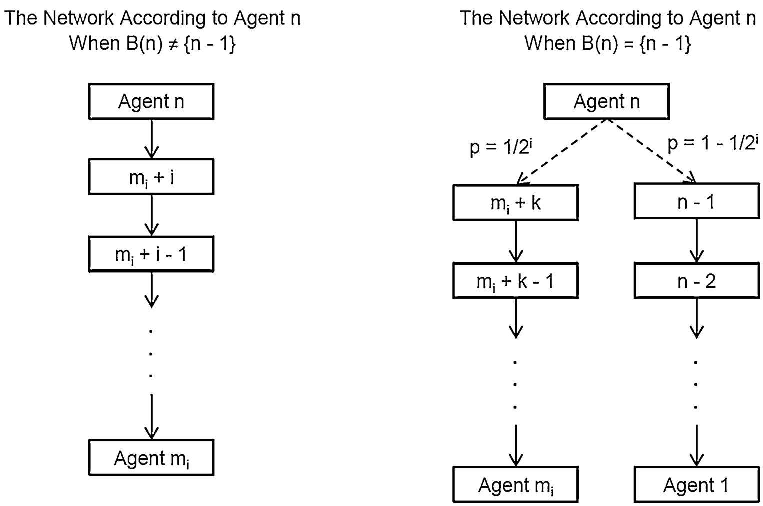

FIGURE 2. The network of Example 3. If agent n observes her immediate predecessor, she believes that most likely all agents have observed their immediate predecessors.

- For each $i > 1$, a set $S_i' = \{m_i, m_i + 1, \ldots, m_i + i\}$ of $i + 1$ consecutive agents contained entirely in S_i is chosen uniformly at random. We have $B(m_i) = \varnothing$ and $B(n) = \{n - 1\}$ for each $n \in S_i' \setminus \{m_i\}$.
- For each $n \in S_i$ with $n \notin S_i'$, $B(n) = \{m_{i-1} + i - 1\}$.

This network topology fails to diffuse information. Figure 2 illustrates this example. $\Diamond$

The realizations of this network topology fall into two distinct regimes. In the "good" regime, all agents observe their immediate predecessor and information diffuses. In the "bad" regime, the growing information paths are broken from time to time and must be rebuilt. When this happens, the agents in S_i' who rebuild the path are nearly certain the network has realized the good regime. The agents in $S_i' \setminus \{m_i\}$ each observe their immediate predecessor, and conditional on this fact, the probability that the bad regime has been realized is small. As a result, agents in S_i' are far too confident in their neighbors and copy too frequently, leading to the failure of information diffusion.

Failure due to overconfidence is related to information cascades, but this phenomenon is distinct in several important ways. When the bad regime is realized, agents in S_i' are far too likely to copy, but there is no information cascade: each agent still has a small chance of following her own signal. However, the agents in S_i' approach herd behavior asymptotically as i grows. The probability that an agent follows her own signal approaches zero faster than the chain grows; thus, the agents $m_i + i$ approach the performance of an agent with an empty neighborhood. Another distinction between the phenomenon in this network and an information cascade is that not all agents join the herd in the limit. Most agents in S_i are fully aware when the bad regime has been realized. The problem is that the agents they observe and the agents further down the information path do not share this awareness. The severe failure we see is driven by the interaction of these incongruent perspectives, rather than by simple herding.

In networks with independent neighborhoods, all topologies featuring expanding subnetworks are deep, but this is not true in general. Our framework allows the construction of network topologies with expanding subnetworks and bounded depth, and these networks exhibit unique learning dynamics. The final failure we catalogue—failure due to correlated actions—occurs in one such network: as the network grows, agents observe unboundedly many others. In these situations, we might expect both information diffusion and aggregation to easily succeed. However, despite the apparent abundance of social information, cases arise in which no number of neighbors yields any information relevant to the state θ.

Example 4 (Failure due to correlated actions). Let F_n denote the set $\{m < n \mid B(m) = \{1\}\}$. Consider a network topology in which $B(n) = \{1\}$ with probability $1/2^{|F_n|}$ and $B(n) = F_n$ otherwise for all $n > 1$. This network topology fails to diffuse information. This example is illustrated in Figure 3. ◊

Under some signal structures, if $\theta = 0$ and $x_1 = 1$, then the social information and private information of an agent $m \in F_n$ effectively cancel; m is equally likely to choose $x_m = 0$ and $x_m = 1$. The same outcome occurs when $\theta = 1$ and $x_1 = 0$. With one of these signal structures, if $x_1 \neq \theta$, agents who observe F_n cannot tell from their social information whether $\theta = 0$ and $x_1 = 1$ or $\theta = 1$ and $x_1 = 0$. They learn merely that the first agent erred.

Our examples reflect the rich variety of phenomena that may appear in general networks when agents have asymmetric network information. These asymmetries give rise to several ways for information diffusion to fail beyond those that prior researchers have studied. In the next section, we provide sufficient conditions for a network to successfully diffuse information despite the presence of neighborhood correlations that induce information asymmetries.

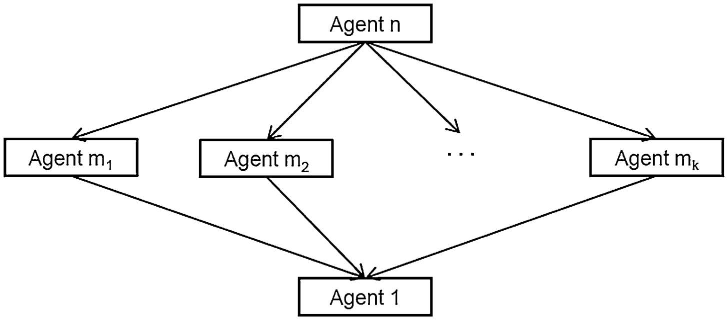

FIGURE 3. The network of Example 4. A typical agent n observes many agents, each of whom in turn observes the first agent.

5. Conditions for success

The positive results of this section establish sufficient conditions for successful information diffusion in the presence of asymmetric network information. Theorem 3 shows that learning succeeds in a sufficiently connected network if each agent's neighborhood realization provides little knowledge about agents along some information path. Theorem 4 demonstrates that if information asymmetries concern an aggregate shock to the entire network, information diffusion proceeds without any complications.

Both of our theorems are based on a generalization of the improvement principle introduced by Acemoglu et al. (2011). For an improvement principle to work, long information paths must exist and must be identifiable. Moreover, agents along the path need reasonably accurate information about the network realization. The essential proof technique is to benchmark the performance of fully Bayesian agents against the performance of a heuristic that is simpler to analyze. Imagine that agents behave in the following way: on seeing who their neighbors are, each selects one neighbor on whom to rely. An agent considers the decision of this chosen neighbor along with her own signal, and chooses an action without regard for what other neighbors have done. If agents could diffuse information by following the heuristic, then the Bayesian agents of our model must do at least as well.

More formally, we define the concepts of neighbor choice functions and chosen neighbor topologies. A neighbor choice function represents a particular agent's means of selecting a neighbor, and a chosen neighbor topology represents a network in which agents discard all observations of the unselected neighbors.

DEFINITION 6. A function $\gamma_n : 2^{\mathbb{N}_n} \to \mathbb{N}_n \cup \{0\}$ is a *neighbor choice function* for agent n if for all sets $B_n \in 2^{\mathbb{N}_n}$, we have either $\gamma_n(B_n) \in B_n$ or $\gamma_n(B_n) = 0$. Given a neighbor choice function γ_n, we say that $\Gamma^n_m = \{B_n \mid \gamma_n(B_n) = m\}$ is agent n's *neighbor-m choice set.*

That is, agent n's neighbor choice function either selects an agent m contained in her neighborhood $B(n)$ or it selects no one. Agent n's neighbor-m choice set is the set of realizations of $B(n)$ such that agent n selects agent m.

Definition 7. A *chosen neighbor topology*, denoted by $\mathbb{Q}_\gamma$, is derived from a network topology $\mathbb{Q}$ and a sequence of neighbor choice functions $\{\gamma_n\}_{n\in\mathbb{N}}$. It consists only of the links in $\mathbb{Q}$ selected by the sequence of neighbor choice functions $\{\gamma_n\}_{n\in\mathbb{N}}$.

Our first positive result depends on controlling what we call network distortion. Given some information about the realized network, the distortion of a particular agent captures how much the distribution of historical neighborhood realizations changes.

Definition 8. Let $\mathbf{E}$ denote a $\mathbb{Q}$-measurable event. Let $\mathbf{B}(n)$ be the n vector comprised of the neighborhoods of the first n agents, and let $\mathbf{B}_n$ denote a particular realization of $\mathbf{B}(n)$. The *network distortion* of agent n with respect to event $\mathbf{E}$ is

$$\delta_n(\mathbf{E}) = \sum_{\mathbf{B}_n} \left| \mathbb{Q}(\mathbf{B}(n) = \mathbf{B}_n \mid \mathbf{E}) - \mathbb{Q}(\mathbf{B}(n) = \mathbf{B}_n) \right|.$$

The network distortion of agent n is a bound on how much some piece of information about the network changes the likelihood of each possible realization of $\mathbf{B}(n)$. When agent n's network distortion with respect to event $\mathbf{E}$ is low, the probability of realizing a given sequence of n neighborhoods and the conditional probability of realizing this sequence given $\mathbf{E}$ are approximately the same.

Given a sequence of neighbor choice functions, for each agent $m < n$, we consider the distortion of m with respect to the event that m is n's chosen neighbor. Since the neighborhoods are drawn from a joint distribution, the unconditional distribution of agent m's neighborhood (and personal subnetwork) is generally different from the distribution of agent m's neighborhood (and personal subnetwork) conditional on agent n's neighborhood. The neighbor choice functions allow us to consider the change in this distribution conditional on agent n's chosen neighbor realization rather than the change in the distribution given agent n's neighborhood realization. If for all $\epsilon > 0$, the probability that agent n's chosen neighbor has distortion greater than ϵ approaches zero as n approaches infinity, then the chosen neighbor topology has low distortion.

Definition 9. The chosen neighbor topology derived from $\mathbb{Q}$ and $\{\gamma_n\}_{n\in\mathbb{N}}$ has *low network distortion* if for any $\epsilon > 0$, we have

$$\lim_{n\to\infty} \sum_{m:\delta_m(B(n)\in\Gamma_m^n)<\epsilon} \mathbb{Q}(B(n) \in \Gamma_m^n) = 1.$$

Low distortion and sufficient connectivity are jointly sufficient for successful information diffusion. For any network topology, if we can find a sequence of neighbor choice functions such that the derived chosen neighbor topology satisfies both the expanding subnetworks condition and the low distortion condition, then information diffuses.

Theorem 3. *Consider an arbitrary network topology $\mathbb{Q}$. If there exists a sequence of neighbor choice functions $\{\gamma_n\}_{n\in\mathbb{N}}$ such that the corresponding chosen neighbor topology features both expanding subnetworks and low network distortion, then $\mathbb{Q}$ successfully diffuses information.*

This theorem characterizes a large class of network topologies in which information diffuses. Use of the theorem centers on the selection of neighbor choice functions. For instance, if we take the agent with the largest index in $B(n)$ as the chosen neighbor of agent n, i.e., $\gamma_n(B(n)) = \max_{m\in B(n)} m$, we can easily prove Theorem 1. With mutually independent neighborhoods, there is no network distortion, and expanding subnetworks is equivalent to expanding observations. Thus, expanding observations is a sufficient condition for learning when neighborhoods are generated independently. Theorem 2 then shows that expanding observations is a necessary condition for information diffusion in this setting, completing the proof of Theorem 1.

Another immediate application of Theorem 3 comes in a setting that allows significant neighborhood correlations. If a network topology has long deterministic information paths, we can consider the chosen neighbor topology these paths generate to show that information diffusion is successful. Agents may have other neighbors with complex, correlated relationships in the network, but the unambiguous existence of one information path for each agent allows information to diffuse.

Definition 10. A sequence of agents, $\{a_1, a_2, \dots, a_k\}$, is a *deterministic information path* if $\mathbb{Q}(a_i \in B(a_{i+1})) = 1$ for all $i < k$.

Proposition 2. *Let $d^*(n)$ denote the length of the longest deterministic information path terminating on agent n, and suppose we have*

$$\lim_{n\to\infty} \mathbb{Q}(d^*(n) < K) = 0$$

for any positive integer K. Then $\mathbb{Q}$ successfully diffuses information.

Two additional examples illustrate the broad applicability of Theorem 3. The first example is based on the Barabási–Albert random graph model. Barabási and Albert (1999) introduced a model for generating random graphs that feature the "scale-free" property commonly associated with complex networks that emerge naturally in many fields. The key feature of this model is that new vertices in the graph attach preferentially to those that are already well connected—a property colloquially described by the phrase, "the rich get richer." We consider the equilibria of the social learning game in a network based on this generative mechanism. Although there are strong correlations in this model, there is no network distortion, and information diffuses.

Prior work has modeled preferential attachment networks in a setting without neighborhood correlations (see Acemoglu et al. (2011)). These networks can be described by designating predetermined sets of agents with varying probabilities of being observed: some groups are persistently more likely to be observed than others, and we can think of agents in the network attaching preferentially to these groups. However, the likelihood of a given agent being observed in the future cannot vary with the actual network realization in such models. Without correlated neighborhoods, the natural generative mechanism cannot be captured, and more importantly, the information conveyed by observing an agent's local neighborhood cannot be captured in the model.

Example 5 (Learning with preferential attachment). Define $\lambda_n(m)$ to be the number of agents acting prior to n who either observe agent m or are observed by agent m:

$$\lambda_n(m) = \left|\{i < n : m \in B(i) \text{ or } i \in B(m)\}\right|.$$

Consider the network topology

- $B(2) = \{1\}$
- $\mathbb{Q}(B(n) = \{m\} \mid \mathbf{B}(n-1) = \mathbf{B}_{n-1}) = (\lambda_n(m))/(2(n-2))$ for all $n > 2$ and $m < n$.

This network topology diffuses information. ◊

In this network, agent n observes exactly one agent chosen at random from those who have already acted: the probability of observing an agent $m < n$ is proportional to the number of links already connecting to m in the network when n's neighborhood is realized. Social learning succeeds because the distortion of m conditional on being observed by n is zero, despite the highly correlated neighborhood realizations. By observing m, agent n learns that her recent predecessors are likely to have observed m as well, but no information is conveyed about agents who acted prior to m.

We next present a network where learning occurs despite nonzero distortion. If agents' neighborhoods are not strongly correlated with neighborhoods realized in the distant past, then social learning can still diffuse information.

Example 6 (Learning with nonzero distortion). Consider a network topology $\mathbb{Q}$ where all agents observe exactly one neighbor and the probability of observing any particular agent m depends on the depth of m, $d(m)$. In particular, define the sets $\{T_k\}_{k\in\mathbb{N}}$ by $T_k = \{2k-1, 2k\}$ for each $k \in \mathbb{N}$. For agent $n \in T_i$ and a fixed realization of $\mathbf{B}(n-1)$,

$$\mathbb{Q}\big(B(n) = \{m\} \mid \mathbf{B}(n-1) = \mathbf{B}_{n-1}\big) = \begin{cases} \frac{1}{2(i-1)}\left(1 - \frac{1}{i-j}\right) & \text{if } d(m) > d(m') \\ \frac{1}{2(i-1)}\left(1 + \frac{1}{i-j}\right) & \text{if } d(m) < d(m') \\ \frac{1}{2(i-1)} & \text{if } d(m) = d(m') \end{cases}$$

for $m, m' \in T_j$, $m \neq m'$, and $j < i$. If $m \in T_i$, $\mathbb{Q}(B(n) = \{m\} \mid \mathbf{B}(n-1) = \mathbf{B}_{n-1}) = 0$. This network topology diffuses information. ◊

Each agent in this network observes one neighbor chosen randomly from the past, but past agents have different likelihoods of being chosen. Past agents are assigned weights such that agents with longer information paths are less likely to be chosen; however, differences in weightings are small for agents in the distant past. Therefore, if an agent's neighbor acted far enough in the past, the distortion of that neighbor is low, and this happens with increasing frequency as n grows large. Despite a bias toward observing agents with short information paths, distortion in this network is low enough that information diffuses.

An empirically minded researcher might wonder how to utilize this characterization and, in particular, how to determine whether a network has low or high distortion. We suggest a few approaches, noting that tests to establish low distortion are easier to design than tests to establish high distortion. Tests of low distortion can rely entirely on statistical features of observed networks. One approach is to show that the network is well described by a member of some family of networks with low distortion. For instance, through Example 5 we show that preferential attachment models have low distortion, so one might test whether a Barabási–Albert model is a good description of the network. After fitting the model parameters from data, we can judge how well the model fits by simulating many instances of the network and checking whether features like the degree distribution, clustering coefficient, and average path length match their empirical counterparts.

Similarly, one could test whether the network fits a configuration model (Bender and Canfield 1978, Bollobás 1980). Configuration models are a family of random networks parameterized by their degree distributions. These models lead to neighborhoods that are approximately independent, so large networks will exhibit low distortion. Given data on a network, one can measure the degree distribution and obtain its corresponding configuration model. One can measure structural properties of this model by generating samples via simulation. If the observed network has structural properties similar to the corresponding configuration model, we would conclude that the configuration model is a good description of the data and the network has low distortion.

Demonstrating high distortion is a more difficult task. This requires us to show that an agent's neighborhood conveys significant information about the broader network—that beliefs about the network conditional on the set of neighbors are substantially different than unconditional beliefs. In the absence of a well understood parameterized network model with high distortion, this is hard to test based on network data alone. It may require surveying individuals on their beliefs about the overall network topology. Supposing data on these subjective beliefs were available, the most direct way to test for high distortion would be to check whether individuals have similar beliefs about the broader network. If individuals have very dissimilar beliefs, this would be a strong indication of high distortion.

Our other positive finding concerns correlations that stem from an underlying Markov process. Let $\mathcal{C}$ be a finite space and for each $i \in \mathcal{C}$, let $\mathbb{Q}_n^{(i)}$ denote a probability measure on the collection of subsets of $\{1, 2, \dots, n-1\}$. We say that the network topology $\mathbb{Q}$ is Markovian if we can find a finite set $\mathcal{C}$, a Markov chain Z_n on $\mathcal{C}$, and a collection of measures $\{\mathbb{Q}_n^{(i)}\}_{n\in\mathbb{N}, i\in\mathcal{C}}$ such that $B(n)$ is drawn according to $\mathbb{Q}_n^{(Z_n)}$ independently from the rest of the network for all n. We call the chain Z_n and the collection $\{\mathbb{Q}_n^{(i)}\}$ the Markov decomposition of $\mathbb{Q}$. In a Markovian network topology, all dependencies between the neighborhoods of different agents are captured in the underlying process Z_n. Earlier positive learning results are robust to Markovian network topologies.

Definition 11. A Markovian network topology $\mathbb{Q}$ features *statewise expanding observations* if there exists a decomposition $\{Z_n, \{\mathbb{Q}_n^{(i)}\}\}$ such that for all $i \in \mathcal{C}$ and each positive integer K, we have

$$\limsup_{n\to\infty} \mathbb{Q}_n^{(i)}\Big(\max_{b\in B(n)} b < K\Big) = 0.$$

Statewise expanding observations is a slightly stronger condition than the expanding observations property that characterizes learning networks with mutually independent neighborhoods. The intuition is similar though. This represents a mild connectivity condition and is sufficient for information diffusion if the network is Markovian.

Theorem 4. *If $\mathbb{Q}$ is a Markovian network topology with statewise expanding observations, then information diffuses.*

A simple and interesting application of Theorem 4 is the case in which all states of the Markov chain are absorbing. In this scenario, some state of network is realized and neighborhoods will be generated independently conditional on that network state. This could represent an aggregate shock to the network, such as a mass Internet disruption, that affects all agents similarly. As long as the chain has finitely many states, asymmetric information about this shock will not impede information diffusion.

This result describes an overlapping, but distinct set of networks that diffuse information compared with those captured in Theorem 3. For instance, Theorem 3 fails to encompass a network in which, with equal probability, either all agents observe their immediate predecessor, or all agents observe a single neighbor drawn uniformly at random. Likewise, Theorem 4 fails to capture the preferential attachment example. The notion of a Markov decomposition offers some additional insight on how significant information asymmetries must be to disrupt information diffusion. Examples 2–4 in the previous section are all sufficiently connected networks, featuring expanding subnetworks, but the topologies cannot be represented through a Markov decomposition nor do they feature low network distortion for any sequence of neighbor choice functions. Some agents receive strongly conflicting information about the network structure inducing vastly different posterior beliefs about this structure.

6. Conclusions

Studying more complex network structures that feature correlated neighborhoods adds richness to our view of observational social learning. When we move beyond the confines of networks with independently realized neighborhoods, we find striking new phenomena that fail to appear in simpler models. Our weaker learning metric of information diffusion highlights the severity of learning failures that can appear in such models. Asymmetric information about the network structure creates inefficiencies that are more harmful than simple herding.

Distortion has emerged in our paper as an especially useful measure of the information neighborhood realizations provide to the agents. When the act of observing neighbors substantially alters our beliefs about their personal subnetworks, learning becomes markedly more difficult. Together with expanding subnetworks, low distortion is a broadly applicable sufficient condition for information diffusion. We also demonstrate that positive learning results are robust to aggregate network shocks and more generally to Markovian network structures.

Our results also contribute to the discussion on appropriate metrics for social learning. The distinction between diffusion and aggregation provides finer detail on what social learning can achieve in different settings and connects prior results in the literature. Networks that always lead to the strongest learning results are rare, but common generative models for social networks do reach our lower bar of diffusion. If no one is well informed to begin with, then social learning has difficulty making any headway, but when some individuals have strong signals, we should expect this information to spread effectively. Importantly, by rendering metrics independently of the signal structure of the model, we offer a clearer picture of the network topology's effect on the learning process.

Appendix

We begin with some preliminaries before providing proofs of the results in the paper in the order that they appear. Subsequently, we analyze each of the examples that were presented, both failures and successes. The first lemma characterizes an agent's decision in equilibrium.

Lemma 1. *Let σ be an equilibrium strategy profile and let $I_n \in \mathcal{I}_n$ be agent n's information set. Then the decision $x_n = \sigma_n(I_n)$ satisfies*

$$x_n = \begin{cases} 1 & \text{if } \mathbb{P}_\sigma(\theta = 1 \mid s_n) + \mathbb{P}_\sigma(\theta = 1 \mid B(n), x_k, k \in B(n)) > 1 \\ 0 & \text{if } \mathbb{P}_\sigma(\theta = 1 \mid s_n) + \mathbb{P}_\sigma(\theta = 1 \mid B(n), x_k, k \in B(n)) < 1, \end{cases}$$

and $x_n \in \{0, 1\}$ otherwise.

This lemma is Proposition 2 in Acemoglu et al. (2011).

We shall find ourselves working with the signal structure frequently. Generally, we find working with the distributions of the private beliefs more convenient than working directly with the signal structure. The private belief distributions are given by $\mathbb{G}_i(r) = \mathbb{P}(p_1 \le r \mid \theta = i)$ for each $i \in \{0, 1\}$. The private belief distributions and the signal structure provide equivalent representations of the agents' private information. The next lemma notes some important properties of the private belief distributions.

Lemma 2. *The private belief distributions, $\mathbb{G}_0$ and $\mathbb{G}_1$, satisfy the following properties:*

(a) For all $r \in (0, 1)$, $d\mathbb{G}_0(r)/d\mathbb{G}_1 = (1 - r)/r$.

(b) For all $0 < z < r < 1$, $\mathbb{G}_0(r) \ge ((1 - r)/r)\mathbb{G}_1(r) + ((r - z)/2)\mathbb{G}_1(z)$.

(c) For all $0 < r < w < 1$, $1 - \mathbb{G}_1(r) \ge (r/(1 - r))(1 - \mathbb{G}_0(r)) + ((w - r)/2)(1 - \mathbb{G}_0(w))$.

(d) The term $\mathbb{G}_0(r)/\mathbb{G}_1(r)$ is nonincreasing in r and is strictly larger than 1 for all $r \in (\underline{\beta}, \overline{\beta})$.

This is Lemma 1 in Acemoglu et al. (2011)

Proof of Theorem 1. This theorem is a direct consequence of Theorems 2 and 3. The proof of this result is presented in the paragraph immediately following Theorem 3. $\square$

Proof of Proposition 1. We need only consider the probability that agent N makes a correct decision. Since agent N has begun an information cascade, we have from Lemma 1 that the social belief q_N satisfies either $q_N > 1-\underline{\beta}$ or $q_N < 1-\overline{\beta}$. The probability that agent N is correct can be expressed as

$$\begin{aligned}\mathbb{P}_\sigma(x_N=\theta) &= \tfrac{1}{2}[\mathbb{P}_\sigma(\text{cascade on } 0 \mid \theta=0)+\mathbb{P}_\sigma(\text{cascade on } 1 \mid \theta=1)] \\ &= \tfrac{1}{2}[\mathbb{P}_\sigma(q_N<1-\overline{\beta} \mid \theta=0)+\mathbb{P}_\sigma(q_N>1-\underline{\beta} \mid \theta=1)].\end{aligned}$$

Since N has begun an information cascade, we have

$$\mathbb{P}_\sigma(q_N<1-\overline{\beta} \mid \theta=j)=1-\mathbb{P}_\sigma(q_N>1-\underline{\beta} \mid \theta=j) \tag{1}$$

for $j\in\{0,1\}$. Further, the definition of q_N and an application of Bayes' rule give

$$\mathbb{P}_\sigma(q_N<1-\overline{\beta} \mid \theta=0) \ge \frac{\overline{\beta}}{1-\overline{\beta}}\mathbb{P}_\sigma(q_N<1-\overline{\beta} \mid \theta=1) \tag{2}$$

$$\mathbb{P}_\sigma(q_N>1-\underline{\beta} \mid \theta=1) \ge \frac{1-\underline{\beta}}{\underline{\beta}}\mathbb{P}_\sigma(q_N<1-\overline{\beta} \mid \theta=0). \tag{3}$$

We now consider the problem of minimizing $\mathbb{P}_\sigma(x_N=\theta)$ subject to the constraints given by (1), (2), and (3). Let $\underline{x}_i$ denote $\mathbb{P}_\sigma(q_N>1-\underline{\beta} \mid \theta=i)$ and let $\overline{x}_i$ denote $\mathbb{P}_\sigma(q_N<1-\overline{\beta} \mid \theta=i)$. The problem can then be written as minimizing $\overline{x}_0+\underline{x}_1$ subject to $\overline{x}_i=1-\underline{x}_i$, $\overline{x}_0\ge(\overline{\beta}/(1-\overline{\beta}))\overline{x}_1$, and $\underline{x}_1\ge((1-\underline{\beta})/\underline{\beta})\underline{x}_0$. We can simplify the problem to two variables by substituting (1) into the inequality constraints, obtaining as new constraints $(\overline{\beta}/(1-\overline{\beta}))(1-\underline{x}_1)-\overline{x}_0\le 0$ and $((1-\underline{\beta})/\underline{\beta})(1-\overline{x}_0)-\underline{x}_1\le 0$. The Karush–Kuhn–Tucker (KKT) conditions imply that both constraints are binding in an optimal solution and, therefore, the solution is determined by a pair of linear equations. Solving the resulting equations gives the minimum value of $\mathbb{P}_\sigma(x_N=\theta)$ as β^*, thus proving that information diffuses. □

Proof of Theorem 2. Given an equilibrium σ, fix a realization of $\mathbf{B}(n)$, and define $\mathfrak{z}_n$ as a decision that maximizes the conditional probability of making a correct decision given $\mathbf{B}(n)=\mathbf{B}_n$ and given the private signals of agent n and each agent in $\hat{B}(n)=\hat{B}_n$:

$$\mathfrak{z}_n\in\arg\max_{y\in\{0,1\}}\mathbb{P}_{(y,\sigma_{-n})}(y=\theta \mid \mathbf{B}(n)=\mathbf{B}_n, s_i=\mathfrak{s}_i,\ \text{for } i\in\hat{B}_n\cup\{n\}).$$

We use $\mathfrak{s}_i$ to denote a particular realization of the private signal s_i. Observe that for any realization of $\hat{B}(n)$, we have $B(m)\subset\hat{B}(n)$ for all $m\in\hat{B}(n)$. Thus, given a realization of the signals $\{\mathfrak{s}_i\}_{i\in\hat{B}_n\cup\{n\}}$, the decisions x_i for $i\in\hat{B}_n\cup\{n\}$ are completely determined by the equilibrium condition. Therefore,

$$\begin{aligned}&\mathbb{P}_\sigma(x_n=\theta \mid \mathbf{B}(n)=\mathbf{B}_n, s_i=\mathfrak{s}_i,\ \text{for } i\in\hat{B}_n\cup\{n\}) \\ &\qquad\le\mathbb{P}_\sigma(\mathfrak{z}_n=\theta \mid \mathbf{B}(n)=\mathbf{B}_n, s_i=\mathfrak{s}_i,\ \text{for } i\in\hat{B}_n\cup\{n\}).\end{aligned}$$

Integrating over all possible sets of signals $\{\mathfrak{s}_i\}_{i\in\hat{B}(n)\cup\{n\}}$, we have

$$\mathbb{P}_\sigma(x_n=\theta \mid \mathbf{B}(n)=\mathbf{B}_n) \le \mathbb{P}_\sigma(\mathfrak{z}_n=\theta \mid \mathbf{B}(n)=\mathbf{B}_n)$$

for any $\mathbf{B}_n$. Since neighborhoods are realized independently from θ and the private signals, $\mathfrak{z}_n$ only depends on the signals of agents in $\hat{B}_n\cup\{n\}$:

$$\mathfrak{z}_n \in \operatorname*{arg\,max}_{y\in\{0,1\}} \mathbb{P}_\sigma(y=\theta \mid s_i=\mathfrak{s}_i,\ \text{for } i\in\hat{B}_n\cup\{n\}).$$

Since signals are identically distributed, this means $\mathbb{P}_\sigma(\mathfrak{z}_n=\theta \mid s_i=\mathfrak{s}_i,\ \text{for } i\in\hat{B}_n\cup\{n\})=\mathbb{P}_\sigma(z_{|\hat{B}_n|+1}=\theta)$, where

$$z_k \in \operatorname*{arg\,max}_{y\in\{0,1\}} \mathbb{P}_\sigma(y=\theta \mid s_i=\mathfrak{s}_i,\ \text{for } i\in\{1,2,\dots,k\}).$$

Therefore, $\mathbb{P}_\sigma(x_n=\theta \mid \mathbf{B}(n)=\mathbf{B}_n)\le \mathbb{P}_\sigma(z_{|\hat{B}_n|+1}=\theta)$.

We now examine the probability that z_k is incorrect. Consider a signal structure where the probability of a partially informative signal is positive such that

$$\mathbb{P}_\sigma(z_1=1 \mid \theta=1)=1-\mathbb{G}_1\Big(\frac{1}{2}\Big)<1 \quad \text{and} \quad \mathbb{P}_\sigma(z_1=0 \mid \theta=0)=\mathbb{G}_0\Big(\frac{1}{2}\Big)<1.$$

Let $\overline{S}_\sigma$ denote the set of signals such that in equilibrium σ, $x_1=0$ if $s_1\in\overline{S}_\sigma$. We have $\mathbb{P}_\sigma(s_1\in\overline{S}_\sigma \mid \theta=1)=\mathbb{G}_1(\frac{1}{2})>0$. Since private signals are independent conditional on θ, we have $\mathbb{P}_\sigma(s_i\in\overline{S}_\sigma \text{ for all } i\le k \mid \theta=1)=\mathbb{P}_\sigma(s_1\in\overline{S}_\sigma \mid \theta=1)^k>0$.

Now suppose that $s_i\in\overline{S}_\sigma$ for each $i\le k$. Applying Bayes' rule gives

$$\begin{aligned}\mathbb{P}_\sigma(\theta=0 \mid s_i\in\overline{S}_\sigma \text{ for all } i\le k) &= \left[1+\frac{\mathbb{P}_\sigma(s_i\in\overline{S}_\sigma \text{ for all } i\le k \mid \theta=1)}{\mathbb{P}_\sigma(s_i\in\overline{S}_\sigma \text{ for all } i\le k \mid \theta=0)}\right]^{-1}\\ &= \left[1+\left(\frac{\mathbb{P}_\sigma(s_1\in\overline{S}_\sigma \mid \theta=1)}{\mathbb{P}_\sigma(s_1\in\overline{S}_\sigma \mid \theta=0)}\right)^k\right].\end{aligned}$$

Another application of Bayes' rule gives

$$\begin{aligned}\mathbb{P}_\sigma(s_1\in\overline{S}_\sigma \mid \theta=1) &= \frac{\mathbb{P}_\sigma(\theta=1 \mid s_1\in\overline{S}_\sigma)\mathbb{P}_\sigma(s_1\in\overline{S}_\sigma)}{\mathbb{P}_\sigma(\theta=1)}\\ \mathbb{P}_\sigma(s_1\in\overline{S}_\sigma \mid \theta=0) &= \frac{\mathbb{P}_\sigma(\theta=0 \mid s_1\in\overline{S}_\sigma)\mathbb{P}_\sigma(s_1\in\overline{S}_\sigma)}{\mathbb{P}_\sigma(\theta=0)}.\end{aligned}$$

Since $\mathbb{P}(\theta=0)=\mathbb{P}(\theta=1)=\frac{1}{2}$,

$$\begin{aligned}\frac{\mathbb{P}_\sigma(s_1\in\overline{S}_\sigma \mid \theta=1)}{\mathbb{P}_\sigma(s_1\in\overline{S}_\sigma \mid \theta=0)} &= \frac{\mathbb{P}_\sigma(\theta=1 \mid s_1\in\overline{S}_\sigma)}{\mathbb{P}_\sigma(\theta=0 \mid s_1\in\overline{S}_\sigma)}\\ &= \frac{1}{\mathbb{P}_\sigma(\theta=0 \mid s_1\in\overline{S}_\sigma)}-1.\end{aligned}$$

For $s_1 \in \overline{S}_\sigma$, we have $x_1 = 0$, so $\mathbb{P}_\sigma(\theta = 0 \mid s_1 \in \overline{S}_\sigma) \geq \frac{1}{2}$. The function

$$\mathbb{P}_\sigma(\theta = 0 \mid s_i \in \overline{S}_\sigma \text{ for all } i \leq k) = \left[1 + \left(\frac{1}{\mathbb{P}_\sigma(\theta = 0 \mid s_1 \in \overline{S}_\sigma)} - 1\right)^k\right]^{-1}$$

is nondecreasing in $\mathbb{P}_\sigma(\theta = 0 \mid s_1 \in \overline{S}_\sigma)$ for any value in $[\frac{1}{2}, 1]$, so $\mathbb{P}_\sigma(\theta = 0 \mid s_i \in \overline{S}_\sigma \text{ for all } i \leq k) \geq \frac{1}{2}$. Thus, $z_k = 0$ whenever $s_i \in \overline{S}_\sigma$ for each $i \leq k$ (we may take $z_k = 0$ in cases of indifference). Similarly, $z_k = 1$ whenever $s_i \notin \overline{S}_\sigma$ for each $i \leq k$. Therefore,

$$\begin{aligned}
\mathbb{P}_\sigma(z_k \neq \theta) &\geq \mathbb{P}(\theta = 0)\mathbb{P}(s_i \notin \overline{S}_\sigma, i \leq k \mid \theta = 0) + \mathbb{P}(\theta = 1)\mathbb{P}(s_i \in \overline{S}_\sigma, i \leq k \mid \theta = 1) \\
&= \tfrac{1}{2}[\mathbb{P}(s_1 \notin \overline{S}_\sigma \mid \theta = 0)^k + \mathbb{P}(s_1 \in \overline{S}_\sigma \mid \theta = 1)^k] \\
&= \tfrac{1}{2}\left[\left(1 - \mathbb{G}_0\left(\tfrac{1}{2}\right)\right)^k + \mathbb{G}_1\left(\tfrac{1}{2}\right)^k\right].
\end{aligned}$$

Now, suppose the network topology does not feature expanding subnetworks. Then there exists some positive integer K, some $\epsilon > 0$, and an infinite sequence of agents n such that $\mathbb{Q}(|\hat{B}(n)| < K) \geq \epsilon$. For these agents we have

$$\begin{aligned}
\mathbb{P}_\sigma(x_n = \theta) &= \sum_{\mathbf{B}_n} \mathbb{P}_\sigma(x_n = \theta \mid \mathbf{B}(n) = \mathbf{B}_n)\mathbb{Q}(\mathbf{B}(n) = \mathbf{B}_n) \\
&\leq 1 - \epsilon + \sum_{\mathbf{B}_n \mid |\hat{B}_n| < K} \mathbb{P}_\sigma(x_n = \theta \mid \mathbf{B}(n) = \mathbf{B}_n)\mathbb{Q}(\mathbf{B}(n) = \mathbf{B}_n) \\
&\leq 1 - \epsilon + \epsilon\mathbb{P}_\sigma(z_K = \theta) \\
&\leq 1 - \epsilon + \epsilon\left[1 - \tfrac{1}{2}\left[\left(1 - \mathbb{G}_0\left(\tfrac{1}{2}\right)\right)^K + \mathbb{G}_1\left(\tfrac{1}{2}\right)^K\right]\right].
\end{aligned}$$

If this last expression is less than β^*, then information does not diffuse. We can easily find signal structures for which this is true; for instance, information will not diffuse if the signal structure exhibits unbounded private beliefs. Therefore, the network topology fails to diffuse information. □

The improvement principle

The proofs of Theorems 3 and 4 center on an improvement principle, which is built over several lemmas that collectively give us a lower bound on the improvement an agent makes over one of her neighbors. We begin by characterizing the optimal decision using information from only one neighbor. In the event that agent b is selected by agent n's neighbor choice function, let $\tilde{x}_n$ be a coarse version of agent n's decision:

$$\tilde{x}_n = \underset{y \in \{0,1\}}{\arg\max}\, \mathbb{P}_\sigma(y = \theta \mid s_n, B(n) \in \Gamma_b^n, x_b).$$

Note that $\mathbb{P}_\sigma(\tilde{x}_n = \theta \mid B(n) \in \Gamma_b^n)$ provides a lower bound on the probability that n makes a correct decision given $B(n) \in \Gamma_b^n$. Further note that if $|B(n)| = 1$ whenever agent b is selected, then $B(n) \in \Gamma_b^n$ is equivalent to $B(n) = \{b\}$, and $\tilde{x}_n = x_n$. Define the probabilities

$$Y_b^\sigma = \mathbb{P}_\sigma(x_b = 1 \mid B(n) \in \Gamma_b^n, \theta = 1), \qquad N_b^\sigma = \mathbb{P}_\sigma(x_b = 0 \mid B(n) \in \Gamma_b^n, \theta = 0).$$

Further, define the decision thresholds

$$L_b^\sigma = \frac{1 - N_b^\sigma}{1 - N_b^\sigma + Y_b^\sigma}, \qquad U_b^\sigma = \frac{N_b^\sigma}{N_b^\sigma + 1 - Y_b^\sigma}.$$

Observe that whenever $\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n) = \frac{1}{2}(Y_b^\sigma + N_b^\sigma) \geq \frac{1}{2}$, we have $L_b^\sigma \leq \frac{1}{2} \leq U_b^\sigma$. Our first lemma characterizes the decision $\tilde{x}_n$ when $\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n) \geq \frac{1}{2}$. Our second shows that networks with $\mathbb{Q}(|B(n)| \leq 1) = 1$ for all n always satisfy this condition.

Lemma 3. *Suppose $\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n) \geq \frac{1}{2}$. Then the decision $\tilde{x}_n$ satisfies*

$$\tilde{x}_n = \begin{cases} 0 & \text{if } p_n < L_b^\sigma \\ x_b & \text{if } p_n \in (L_b^\sigma, U_b^\sigma) \\ 1 & \text{if } p_n > U_b^\sigma. \end{cases}$$

Apply Bayes' rule to determine $\mathbb{P}_\sigma(\theta = 1 \mid x_b = j)$ for each $j \in \{0, 1\}$; the proof follows immediately from Lemma 1.

Lemma 4. *If $\mathbb{Q}(|B(n)| \leq 1) = 1$ for all $n \geq 1$, then $\mathbb{P}_\sigma(x_n = \theta \mid \mathbf{B}(n) = \mathbf{B}_n) \geq \frac{1}{2}$ for any realization $\mathbf{B}_n$ that occurs with positive probability. It follows that $\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n) \geq \frac{1}{2}$ for any b and n.*

Proof. Proceed by induction. The first agent has an empty neighborhood with probability 1, so the equilibrium condition guarantees $\mathbb{P}_\sigma(x_1 = \theta \mid \mathbf{B}(1) = \mathbf{B}_1) = \mathbb{P}_\sigma(x_1 = \theta) \geq \frac{1}{2}$. Suppose that $\mathbb{P}_\sigma(x_n = \theta \mid \mathbf{B}(n) = \mathbf{B}_n) \geq \frac{1}{2}$ for all $n \leq k$ and all $\mathbf{B}_n$. Given a realization $\mathbf{B}(k+1) = \mathbf{B}_{k+1}$, if $B(x_{k+1})$ is empty, then agent $k+1$ is just as likely to be correct as the first agent, and we immediately get the desired conclusion. If $B(k+1) = \{b\}$, take $\gamma_{k+1}(\{b\}) = b$. By Lemma 3,

$$\begin{aligned} \mathbb{P}_\sigma(x_{k+1} = \theta \mid \mathbf{B}(k+1) = \mathbf{B}_{k+1}) = {} & \mathbb{P}_\sigma(p_n < L_b^\sigma)\mathbb{P}_\sigma(\theta = 0 \mid p_n < L_b^\sigma) \\ & + \mathbb{P}_\sigma(p_n > U_b^\sigma)\mathbb{P}_\sigma(\theta = 1 \mid p_n > U_b^\sigma) \\ & + \mathbb{P}_\sigma(p_n \in [L_b^\sigma, U_b^\sigma])\mathbb{P}_\sigma(\theta = x_b \mid \mathbf{B}(k+1) = \mathbf{B}_{k+1}) \\ \geq {} & \tfrac{1}{2}\big[\mathbb{P}_\sigma(p_n < L_b^\sigma) + \mathbb{P}_\sigma(p_n > U_b^\sigma) + \mathbb{P}_\sigma(p_n \in [L_b^\sigma, U_b^\sigma])\big] \\ = {} & \tfrac{1}{2}, \end{aligned}$$

where we have assumed that agent $k+1$ copies in the event of indifference; this assumption does not affect our result. □

Using Lemma 3, a straightforward calculation shows that $\mathbb{P}_\sigma(\tilde{x}_n = \theta \mid B(n) \in \Gamma_b^n)$ is equal to

$$\tfrac{1}{2}\big[\mathbb{G}_0(L_b^\sigma) + (\mathbb{G}_0(U_b^\sigma) - \mathbb{G}_0(L_b^\sigma))N_b^\sigma + 1 - \mathbb{G}_1(U_b^\sigma) + (\mathbb{G}_1(U_b^\sigma) - \mathbb{G}_1(L_b^\sigma))Y_b^\sigma\big]. \quad (4)$$

Some additional effort yields a lower bound on the improvement over x_b.

Lemma 5. *Suppose* $\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n) \geq \frac{1}{2}$. *Then*

$$\mathbb{P}_\sigma(\tilde{x}_n = \theta \mid B(n) \in \Gamma_b^n) \geq \mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n) + \frac{(1-N_b^\sigma)(L_b^\sigma - \underline{\beta})}{8}\mathbb{G}_1\left(\frac{L_b^\sigma + \underline{\beta}}{2}\right) + \frac{(1-Y_b^\sigma)(\overline{\beta} - U_b^\sigma)}{8}\left[1 - \mathbb{G}_0\left(\frac{\overline{\beta} + U_b^\sigma}{2}\right)\right].$$

Proof. We use properties of the belief distributions given in Lemma 2. In Lemma 2(b), take $r = L_b^\sigma$ and $z = \min(L_b^\sigma, (L_b^\sigma + \underline{\beta})/2)$ to obtain

$$(1-N_b^\sigma)\mathbb{G}_0(L_b^\sigma) \geq Y_b^\sigma \mathbb{G}_1(L_b^\sigma) + \frac{(1-N_b^\sigma)(L_b^\sigma - \underline{\beta})}{4}\mathbb{G}_1\left(\frac{L_b^\sigma + \underline{\beta}}{2}\right).$$

Even if $L_b^\sigma < \underline{\beta}$, the second term on the right is zero, so the inequality still holds. Similarly, using Lemma 2(c) with $r = U_b^\sigma$ and $w = \max(U_b^\sigma, (U_b^\sigma + \overline{\beta})/2)$ yields

$$(1-Y_b^\sigma)(1-\mathbb{G}_1(U_b^\sigma)) \geq N_b^\sigma(1-\mathbb{G}_0(U_b^\sigma)) + \frac{(1-Y_b^\sigma)(\overline{\beta} - U_b^\sigma)}{4}\left[1 - \mathbb{G}_0\left(\frac{U_b^\sigma + \overline{\beta}}{2}\right)\right].$$

Substitute these two results into (4) and note that $Y_b^\sigma + N_b^\sigma = 2\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n)$ to complete the proof. □

The next lemma shows the improvement is uniformly bounded away from zero whenever $\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n) < \beta^*$. When $\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n) \geq \beta^*$, Lemma 5 still guarantees that the improvement is nonnegative.

Lemma 6. *Let* α *denote* $\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n)$. *Suppose that* $\frac{1}{2} \leq \alpha < \beta^*$ *and let* $\Delta = \beta^* - \alpha$. *Further, define* $\underline{c} = 1 - 2\underline{\beta}$, $\overline{c} = 2\overline{\beta} - 1$, *and* $c^* = 2\beta^* - 1$. *Then*

$$\mathbb{P}_\sigma(\tilde{x}_n = \theta \mid B(n) \in \Gamma_b^n) \geq \alpha + \frac{\Delta^2}{8c^*}\min\left\{\underline{c}^2\mathbb{G}_1\left(\underline{\beta} + \frac{\underline{c}^2\Delta}{2c^*}\right), \overline{c}^2\left[1 - \mathbb{G}_0\left(\overline{\beta} - \frac{\overline{c}^2\Delta}{2c^*}\right)\right]\right\}.$$

Proof. One of the following two inequalities must hold:

$$N_b^\sigma \leq \frac{1 - \underline{\beta} - 2\alpha\underline{\beta}}{1 - 2\underline{\beta}} - \Delta \quad \text{or} \quad Y_b^\sigma \leq \frac{\overline{\beta} - 2\alpha(1-\overline{\beta})}{2\overline{\beta} - 1} - \Delta. \tag{5}$$

Suppose not. We then have

$$\alpha = \frac{1}{2}(N_b^\sigma + Y_b^\sigma) > \frac{1}{2}\left(\frac{1 - \underline{\beta} - 2\alpha\underline{\beta}}{1 - 2\underline{\beta}} + \frac{\overline{\beta} - 2\alpha(1-\overline{\beta})}{2\overline{\beta} - 1}\right) - \Delta,$$

which is equivalent to

$$\beta^* > \frac{1}{2}\left(\frac{1 - \underline{\beta} - 2\alpha\underline{\beta}}{1 - 2\underline{\beta}} + \frac{\overline{\beta} - 2\alpha(1-\overline{\beta})}{2\overline{\beta} - 1}\right).$$

Rearranging and simplifying gives $\alpha > \beta^*$, contradicting our assumptions.

Suppose the first inequality in (5) holds. The definition of L_b^σ, along with the relationship $2\alpha = N_b^\sigma + Y_b^\sigma$, yields

$$L_b^\sigma = \frac{1 - N_b^\sigma}{1 - 2N_b^\sigma + 2\alpha}.$$

Taking the first derivative with respect to N_b^σ and recalling that $\alpha \geq \frac{1}{2}$, we see that L_b^σ is a nonincreasing function of N_b^σ. Therefore, the minimum occurs at the corner and

$$L_b^\sigma \geq \frac{2\alpha - 1}{2\beta^* - 1 - 4\underline{\beta}\Delta}\left(\underline{\beta} + \frac{(1 - 2\underline{\beta})\Delta}{2\alpha - 1}\right).$$

It follows that $1 - N_b^\sigma \geq \Delta$ and $L_b^\sigma - \underline{\beta} \geq (1 - 2\underline{\beta})^2\Delta/(2\beta^* - 1) = (\underline{c}^2\Delta)/c^*$. Substituting gives

$$\frac{(1 - N_b^\sigma)(L_b^\sigma - \underline{\beta})}{8}\mathbb{G}_1\left(\frac{L_b^\sigma + \underline{\beta}}{2}\right) \geq \frac{(\underline{c}\Delta)^2}{8c^*}\mathbb{G}_1\left(\underline{\beta} + \frac{\underline{c}^2\Delta}{2c^*}\right).$$

Similarly, if the second inequality in (5) holds, then $1 - Y_b^\sigma \geq \Delta$, $\overline{\beta} - U_b^\sigma \geq (\overline{c}^2\Delta)/c^*$, and

$$\frac{(1 - Y_b^\sigma)(\overline{\beta} - U_b^\sigma)}{8}\left[1 - \mathbb{G}_0\left(\frac{\overline{\beta} + U_b^\sigma}{2}\right)\right] \geq \frac{(\overline{c}\Delta)^2}{8c^*}\left[1 - \mathbb{G}_0\left(\overline{\beta} - \frac{\overline{c}^2\Delta}{2c^*}\right)\right].$$

Substituting these inequalities into Lemma 5 completes the proof. □

The last two lemmas describe the improvement that a single agent can make over her neighbor by employing a heuristic that discards the information from all other neighbors. To study the limiting behavior of these improvements, we define the function

$$\overline{\mathcal{Z}}(\alpha) = \alpha + \frac{\Delta^2}{8c^*}\min\left\{\underline{c}^2\mathbb{G}_1\left(\underline{\beta} + \frac{\underline{c}^2\Delta}{2c^*}\right), \overline{c}^2\left[1 - \mathbb{G}_0\left(\overline{\beta} - \frac{\overline{c}^2\Delta}{2c^*}\right)\right]\right\}$$

for $\alpha \in [1/2, \beta^*]$ and $\overline{\mathcal{Z}}(\alpha) = \alpha$ for $\alpha \in (\beta^*, 1]$. By Lemmas 5 and 6, we have

$$\mathbb{P}_\sigma(\tilde{x}_n = \theta \mid B(n) \in \Gamma_b^n) \geq \overline{\mathcal{Z}}\big(\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n)\big).$$

We note some important properties of the function $\overline{\mathcal{Z}}$ in the next lemma.

Lemma 7. *The function $\overline{\mathcal{Z}}$ has the following properties:*

(a) The function $\overline{\mathcal{Z}}$ is left-continuous and has no upward jumps:

$$\overline{\mathcal{Z}}(\alpha) = \lim_{r\uparrow\alpha}\overline{\mathcal{Z}}(r) \geq \lim_{r\downarrow\alpha}\overline{\mathcal{Z}}(r).$$

(b) For any $\alpha \in [1/2, 1]$, $\overline{\mathcal{Z}}(\alpha) \geq \alpha$.

(c) For any $\alpha \in [1/2, \beta^*)$, $\overline{\mathcal{Z}}(\alpha) > \alpha$.

PROOF. The $\mathbb{G}_0$ and $\mathbb{G}_1$ are cumulative distribution functions, so they are left-continuous with no downward jumps. Part (a) follows. For parts (b) and (c), observe that the improvement term in the definition of $\overline{\mathcal{Z}}(\alpha)$ is nonnegative and it is strictly positive whenever $\alpha < \beta^*$. □

We further define a related function $\mathcal{Z}(\alpha)$ that is continuous and monotonic while maintaining the same improvement properties as $\overline{\mathcal{Z}}(\alpha)$:

$$\mathcal{Z}(\alpha) = \frac{1}{2}\Big(\alpha + \sup_{r\in[1/2,\alpha]} \overline{\mathcal{Z}}(r)\Big).$$

LEMMA 8. *The function $\mathcal{Z}(\alpha)$ has the following properties:*

(a) For any $\alpha \in [1/2, 1]$, $\mathcal{Z}(\alpha) \geq \alpha$.

(b) For any $\alpha \in [1/2, \beta^*)$, $\mathcal{Z}(\alpha) > \alpha$.

(c) The function $\mathcal{Z}$ is increasing and continuous.

PROOF. Parts (a) and (b) are immediate consequences of Lemma 7. Note that $\sup_{r\in[1/2,\alpha]} \overline{\mathcal{Z}}(r)$ is nondecreasing and α is an increasing function. Thus, the average of these two is an increasing function, establishing the first part of (c).

We now show that $\mathcal{Z}$ is continuous on $[\frac{1}{2}, 1]$. For $\alpha \in (\beta^*, 1]$, this is obvious since $\mathcal{Z}(\alpha)$ is constant on that interval. For $\alpha \in [\frac{1}{2}, \beta^*)$, we argue by contradiction. Suppose $\mathcal{Z}(\alpha)$ is discontinuous at $\alpha' \in [\frac{1}{2}, \beta^*)$; this implies $\sup_{r\in[1/2,\alpha]} \overline{\mathcal{Z}}(r)$ is discontinuous at $\alpha = \alpha'$. Since this is a nondecreasing function, we have

$$\lim_{\alpha\downarrow\alpha'} \sup_{r\in[1/2,\alpha]} \overline{\mathcal{Z}}(r) > \sup_{r\in[1/2,\alpha']} \overline{\mathcal{Z}}(r).$$

However, this contradicts property (a) of Lemma 7: $\overline{\mathcal{Z}}$ has no upward jumps. It remains to show that $\mathcal{Z}$ is continuous at β^*. This follows from property (b) of Lemma 7 once we note that the improvement term in the definition of $\overline{\mathcal{Z}}$ is less than $\beta^* - \alpha$. □

LEMMA 9. *Let $\sigma \in \Sigma$ be an equilibrium, and suppose that $\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n) \geq \frac{1}{2}$. Then $\mathbb{P}_\sigma(\tilde{x}_n = \theta \mid B(n) \in \Gamma_b^n) \geq \mathcal{Z}(\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n))$.*

PROOF. Let α denote $\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n)$. If $\mathcal{Z}(\alpha) = \alpha$, the result follows from Lemma 5. If $\mathcal{Z}(\alpha) > \alpha$, then $\mathcal{Z}(\alpha) \leq \sup_{r\in[1/2,\alpha]} \overline{\mathcal{Z}}(r)$, and there exists $\overline{\alpha} \in [\frac{1}{2}, \alpha]$ such that $\overline{\mathcal{Z}}(\overline{\alpha}) \geq \mathcal{Z}(\alpha)$. We now show that $\mathbb{P}_\sigma(\tilde{x}_n = \theta \mid B(n) \in \Gamma_b^n) \geq \overline{\mathcal{Z}}(\overline{\alpha})$.

Agent n can render her decision even coarser by choosing not to observe the action of agent b with some probability. Suppose that instead of considering b's action directly, agent n bases her decision on the coarse observation $\tilde{x}_b$ generated as

$$\tilde{x}_b = \begin{cases} x_b & \text{with probability } (2\overline{\alpha} - 1)/(2\alpha - 1) \\ 0 & \text{with probability } (\alpha - \overline{\alpha})/(2\alpha - 1) \\ 1 & \text{with probability } (\alpha - \overline{\alpha})/(2\alpha - 1), \end{cases}$$

with the realizations of $\tilde{x}_b$ independent of the rest of n's information set. Note that $\mathbb{P}_\sigma(\tilde{x}_b = \theta \mid B(n) \in \Gamma_b^n) = \overline{\alpha}$. Lemma 6 implies that the decision based on this coarse observation of b is correct with probability at least $\overline{\mathcal{Z}}(\overline{\alpha})$, and the desired result follows. □

Proof of Theorem 3. We proceed by constructing sequences $\{\alpha_k\}$ and $\{\phi_k\}$ such that $\mathbb{P}_\sigma(x_n = \theta) \geq \phi_k$ for all $n \geq \alpha_k$. Define the sequence ϕ_k by $\phi_1 = \frac{1}{2}$ and $\phi_{k+1} = (\phi_k + \mathcal{Z}(\phi_k))/2$. By the hypothesis of the theorem, for any positive integer K and any $\epsilon > 0$, we can find a positive integer $N(K, \epsilon)$ and a sequence of neighbor choice functions $\{\gamma_k\}_{k\in\mathbb{N}}$ such that the following condition holds: For all $n \geq N(K, \epsilon)$, the probability that either $\gamma_n(B(n)) < K$ or $\gamma_n(B(n)) = b$ such that $\delta_b^\sigma(B(n) \in \Gamma_b^n) \geq \epsilon$ is less than ϵ. Define the sequence

$$\epsilon_k = \min\Big[\tfrac{1}{2}\big(1 + \mathcal{Z}(\phi_k) - \sqrt{1 + 2\phi_k + \mathcal{Z}(\phi_k)^2}\big), \mathbb{P}_\sigma(x_1 = \theta) - \tfrac{1}{2}\Big].$$

Under the assumption that $d\mathbb{G}_0/d\mathbb{G}_1$ is not uniformly equal to 1, some of the signals are informative and $\mathbb{P}_\sigma(x_1 = \theta) > \frac{1}{2}$, so $\epsilon_k > 0$ for all k. Now let $\alpha_1 = 1$ and define the rest of the sequence by $\alpha_{k+1} = N(\alpha_k, \epsilon_k)$. Proceed by induction on k. Since agents can always do as well as the first agent by ignoring their social information, we have $\mathbb{P}_\sigma(x_n = \theta) \geq \mathbb{P}_\sigma(x_1 = \theta) > \frac{1}{2}$ for all n, establishing the result for $k = 1$.

Assume the result for some arbitrary k and consider $n \geq \alpha_{k+1}$. First note that for any $\mathbb{Q}$-measurable event $\mathbf{E}$, we have the bound

$$\begin{aligned}
|\mathbb{P}_\sigma(x_n = \theta \mid \mathbf{E}) - \mathbb{P}_\sigma(x_n = \theta)| &= \Big|\sum_{\mathbf{B}_n} \mathbb{P}_\sigma(x_n = \theta \mid \mathbf{B}(n) = \mathbf{B}_n, \mathbf{E})\mathbb{Q}(\mathbf{B}(n) = \mathbf{B}_n \mid \mathbf{E}) \\
&\qquad - \mathbb{P}_\sigma(x_n = \theta \mid \mathbf{B}(n) = \mathbf{B}_n)\mathbb{Q}(\mathbf{B}(n) = \mathbf{B}_n)\Big| \\
&= \Big|\sum_{\mathbf{B}_n} \mathbb{P}_\sigma(x_n = \theta \mid \mathbf{B}(n) = \mathbf{B}_n) \\
&\qquad \cdot \big[\mathbb{Q}(\mathbf{B}(n) = \mathbf{B}_n \mid \mathbf{E}) - \mathbb{Q}(\mathbf{B}(n) = \mathbf{B}_n)\big]\Big| \\
&\leq \sum_{\mathbf{B}_n} \big|\mathbb{Q}(\mathbf{B}(n) = \mathbf{B}_n \mid \mathbf{E}) - \mathbb{Q}(\mathbf{B}(n) = \mathbf{B}_n)\big| = \delta_n^\sigma(\mathbf{E}),
\end{aligned}$$

where the second equality follows because, conditional on $\mathbf{B}(\mathbf{n}) = \mathbf{B}_n$, the probability that agent n makes a correct decision is independent of any later neighborhood realizations. Using Lemma 9, we have

$$\begin{aligned}
\mathbb{P}_\sigma(x_n = \theta) &= \sum_{b=0}^{n-1} \mathbb{P}_\sigma(x_n = \theta \mid B(n) \in \Gamma_b^n)\mathbb{Q}(B(n) \in \Gamma_b^n) \\
&\geq \sum_{b=0}^{n-1} \mathbb{P}_\sigma(\tilde{x}_n = \theta \mid B(n) \in \Gamma_b^n)\mathbb{Q}(B(n) \in \Gamma_b^n)
\end{aligned}$$

$$
\begin{aligned}
&\geq \sum_{b=0}^{n-1} \mathcal{Z}\big(\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n)\big)\mathbb{Q}(B(n) \in \Gamma_b^n) \\
&\geq \sum_{\substack{b: b \geq \alpha_k \\ \delta_b^\sigma(B(n)\in\Gamma_b^n)<\epsilon}} \mathcal{Z}\big(\mathbb{P}_\sigma(x_b = \theta \mid B(n) \in \Gamma_b^n)\big)\mathbb{Q}(B(n) \in \Gamma_b^n) \\
&\geq \sum_{\substack{b: b \geq \alpha_k \\ \delta_b^\sigma(B(n)\in\Gamma_b^n)<\epsilon}} \mathcal{Z}(\phi_k - \epsilon_k)\mathbb{Q}(B(n) \in \Gamma_b^n) \\
&\geq \mathcal{Z}(\phi_k - \epsilon_k)(1 - \epsilon_k) \geq \phi_{k+1}.
\end{aligned}
$$

From here it remains only to show that $\lim_{k\to\infty} \phi_k \geq \beta^*$. The term $\{\phi_k\}_{k\in\mathbb{N}}$ is a bounded, nondecreasing sequence, so the sequence converges to a limit, ϕ^*. Now note

$$
2\phi^* = 2 \lim_{k\to\infty} \phi_k = \lim_{k\to\infty} (\phi_k + \mathcal{Z}(\phi_k)) = \phi^* + \mathcal{Z}(\phi^*)
$$

since $\mathcal{Z}$ is continuous. Therefore, ϕ^* is a fixed point of the function $\mathcal{Z}$, i.e., $\phi^* = \mathcal{Z}(\phi^*)$. By Lemma 8(b), this means that $\phi^* \geq \beta^*$, completing the proof. □

Proof of Proposition 2. Suppose $\mathbb{Q}(m \in B(n)) = 1$ and define $\gamma_n(B(n)) = m$ for all realizations of $B(n)$. Then the distortion $\delta_m^\sigma(B(n) \in \Gamma_m^n)$ is zero. We can form a chosen neighbor topology with low distortion and expanding subnetworks as follows. For every n, set $\gamma_n(B(n))$ equal to the neighbor with the longest deterministic information path. □

Proof of Theorem 4. The proof of this theorem follows a similar inductive argument as in Theorem 3. Define neighbor choice functions that always select the neighbor of highest index. Since $\mathcal{C}$ is finite, by the hypothesis of the theorem, for any positive integer K and any $\epsilon > 0$, we can find $N(K, \epsilon)$ such that $\mathbb{Q}_n^{(i)}(\gamma_n(B(n)) < K) < \epsilon$ for all $i \in \mathcal{C}$ and all $n \geq N(K, \epsilon)$. Define the sequences $\{\alpha_k\}$ and $\{\phi_k\}$ recursively. Let $\alpha_1 = 1$ and $\phi_1 = \frac{1}{2}$; for $k \geq 1$, we have

$$
\alpha_{k+1} = N\left(\alpha_k, \frac{1}{2}\left[1 - \frac{\phi_k}{\mathcal{Z}(\phi_k)}\right]\right), \qquad \phi_{k+1} = \frac{\phi_k + \mathcal{Z}(\phi_k)}{2}.
$$

We apply the improvement principle to the minimal value of $\mathbb{P}_\sigma(x_n = \theta \mid Z_n = i)$ over all possible realizations of Z_n. Suppose $\min_{i\in\mathcal{C}} \mathbb{P}_\sigma(x_n = \theta \mid Z_n = i) \geq \phi_k$ for all $n \geq \alpha_k$. Since $\mathcal{Z}$ is increasing, we have the bound

$$
\begin{aligned}
\mathbb{P}_\sigma(x_n = \theta \mid Z_n = i) &= \sum_{b=0}^{n-1} \mathbb{P}_\sigma(x_n = \theta \mid Z_n = i, B(n) \in \Gamma_b^n)\mathbb{Q}_n^{(i)}(B(n) \in \Gamma_b^n) \\
&\geq \sum_{b=0}^{n-1} \mathbb{P}_\sigma(\tilde{x}_n = \theta \mid Z_n = i, B(n) \in \Gamma_b^n)\mathbb{Q}_n^{(i)}(B(n) \in \Gamma_b^n)
\end{aligned}
$$

$$\begin{aligned}
&\geq \sum_{b=0}^{n-1} \mathcal{Z}\Big(\min_{j\in\mathcal{C}} \mathbb{P}_\sigma(x_b=\theta \mid Z_b=j)\Big)\mathbb{Q}_n^{(i)}(B(n)\in\Gamma_b^n)\\
&\geq \sum_{b\geq\alpha_k} \mathcal{Z}\Big(\min_{j\in\mathcal{C}} \mathbb{P}_\sigma(x_b=\theta \mid Z_b=j)\Big)\mathbb{Q}_n^{(i)}(B(n)\in\Gamma_b^n)\\
&\geq \mathcal{Z}(\phi_k)\frac{1}{2}\bigg[1-\frac{\phi_k}{\mathcal{Z}(\phi_k)}\bigg] \geq \phi_{k+1}.
\end{aligned}$$

Since this is true for each $i\in\mathcal{C}$, we have $\min_{i\in\mathcal{C}}\mathbb{P}_\sigma(x_n=\theta \mid Z_n=i)\geq\phi_{k+1}$. Since $\liminf_{k\to\infty}\phi_k\geq\beta^*$, the proof is complete. □

We now turn to the analysis of the examples presented in the body of the paper.

Example 2 (Failure due to unidentifiable paths). An agent $n\in S_i$ with $i\geq 1$ has an empty neighborhood with probability $1/2^i$, and this probability approaches zero as n grows. For all other agents $n\in S_i$, there is an information path of length $i+1$ terminating on agent 1. These agents observe two neighbors: one with an empty neighborhood and one with an information path of length i. By construction, each neighbor is equally likely to have an empty neighborhood. If the two neighbors choose the same action, no trouble arises: n knows the decision of her well informed neighbor. If the two neighbors choose different actions, n has a problem.

Since m_{i-1} and e_{i-1} are indistinguishable, agent n's social belief given $x_{m_{i-1}}\neq x_{e_{i-1}}$ is the same regardless of who chooses 0 and who chooses 1:

$$q_n^*\equiv\mathbb{P}_\sigma(\theta=1 \mid B(n), x_{m_{i-1}}=0, x_{e_{i-1}}=1)=\mathbb{P}_\sigma(\theta=1 \mid B(n), x_{m_{i-1}}=1, x_{e_{i-1}}=0).$$

We can bound the probability of agent n making an error as

$$\begin{aligned}
\mathbb{P}_\sigma(x_n\neq\theta) &= \tfrac{1}{2}[\mathbb{P}_\sigma(x_n=1 \mid \theta=0)+\mathbb{P}_\sigma(x_n=0 \mid \theta=1)]\\
&\geq \tfrac{1}{2}[\mathbb{P}_\sigma(x_n=1 \mid x_{m_{i-1}}\neq x_{e_{i-1}}, \theta=0)\mathbb{P}_\sigma(x_{m_{i-1}}\neq x_{e_{i-1}} \mid \theta=0)\\
&\qquad + \mathbb{P}_\sigma(x_n=0 \mid x_{m_{i-1}}\neq x_{e_{i-1}}, \theta=1)\mathbb{P}_\sigma(x_{m_{i-1}}\neq x_{e_{i-1}} \mid \theta=1)]\\
&= \tfrac{1}{2}\big[(1-\mathbb{G}_0(q_n^*))\mathbb{P}_\sigma(x_{m_{i-1}}\neq x_{e_{i-1}} \mid \theta=0)+\mathbb{G}_1(q_n^*)\mathbb{P}_\sigma(x_{m_{i-1}}\neq x_{e_{i-1}} \mid \theta=1)\big].
\end{aligned}$$

Since neither m_{i-1} nor e_{i-1} is in the other's personal subnetwork, their actions are independent conditional on θ. To simplify notation, define $p(x_i)_{a,b}=\mathbb{P}(x_i=a \mid B(n), \theta=b)$. We have

$$\begin{aligned}
\mathbb{P}_\sigma(x_{m_{i-1}}\neq x_{e_{i-1}} \mid \theta=0) &= p(x_{m_{i-1}})_{0,0}\,p(x_{e_{i-1}})_{1,0}+p(x_{m_{i-1}})_{1,0}\,p(x_{e_{i-1}})_{0,0}\\
&= p(x_{m_{i-1}})_{0,0}\Big[1-\mathbb{G}_0\Big(\tfrac{1}{2}\Big)\Big]+p(x_{m_{i-1}})_{1,0}\mathbb{G}_0\Big(\tfrac{1}{2}\Big)\\
&\geq \min\Big[1-\mathbb{G}_0\Big(\tfrac{1}{2}\Big), \mathbb{G}_0\Big(\tfrac{1}{2}\Big)\Big]\equiv p_0^*,
\end{aligned}$$

where the inequality follows since $p(x_{m_{i-1}})_{0,0}+p(x_{m_{i-1}})_{1,0}=1$. A similar calculation gives

$$\mathbb{P}_\sigma(x_{m_{i-1}}\neq x_{e_{i-1}} \mid \theta=1)\geq\min\Big[\mathbb{G}_1\Big(\tfrac{1}{2}\Big), 1-\mathbb{G}_1\Big(\tfrac{1}{2}\Big)\Big]\equiv p_1^*.$$

Define $p^* = \min(p_0^*, p_1^*)$. Substituting into our previous bound for $\mathbb{P}_\sigma(x_n \neq \theta)$ gives

$$\mathbb{P}_\sigma(x_n \neq \theta) \geq \frac{p^*}{2}[1 - \mathbb{G}_0(q_n^*) + \mathbb{G}_1(q_n^*)].$$

Now define g by $g = \min_{q\in[0,1]}\{1 - \mathbb{G}_0(q) + \mathbb{G}_1(q)\}$. Clearly, $g > 0$ and $\mathbb{P}_\sigma(x_n \neq \theta) \geq (p^* g)/2$. The quantity on the right hand side is strictly positive and depends solely on the signal structure. If $(p^* g)/2 \geq 1 - \beta^*$, then information does not diffuse. Since this is true for any signal structure with unbounded private beliefs, the network topology does not diffuse information. ◊

Example 3 (Failure due to overconfidence). Our proof is divided into three parts. In the first part, we establish a variation on the improvement principle derived in the proof of Theorem 3. The second part uses the improvement principle to provide a lower bound on the level of confidence agent n has in agent $n-1$ whenever $B(n) = \{n-1\}$ and, hence, a lower bound on the probability that n copies $n-1$. The final part of our proof combines these results to show that the network topology fails to diffuse information.

Part 1: The improvement principle

We extend the result of Lemma 6 to give a lower bound on the probability that n makes a correct choice conditioned on any realization of a network with $\mathbb{Q}(|B(n)| \leq 1) = 1$ for all n. Let $\mathbf{B}_n$ denote a realization of the first n neighborhoods $\mathbf{B}(n)$ and suppose that $B(n) = \{b\}$ in this realization. Further, let α, $\underline{c}$, $\overline{c}$, c^*, and Δ be defined as in the statement of Lemma 6. We show that if $\frac{1}{2} \leq \alpha < \beta^*$ and $\alpha' = \mathbb{P}_\sigma(x_b = \theta \mid \mathbf{B}(n) = \mathbf{B}_n)$, then

$$\begin{aligned}\mathbb{P}_\sigma(x_n = \theta \mid \mathbf{B}(n) = \mathbf{B}_n) &\geq \min(\alpha, \alpha') \\ &\quad + \frac{\Delta^2}{8c^*}\min\left\{\underline{c}^2\mathbb{G}_1\left(\underline{\beta} + \frac{\underline{c}^2\Delta}{2c^*}\right), \overline{c}^2\left[1 - \mathbb{G}_0\left(\overline{\beta} - \frac{\overline{c}^2\Delta}{2c^*}\right)\right]\right\}.\end{aligned}$$

By Lemma 6, we have

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\}) \geq \alpha + \frac{\Delta^2}{8c^*}\min\left\{\underline{c}^2\mathbb{G}_1\left(\underline{\beta} + \frac{\underline{c}^2\Delta}{2c^*}\right), \overline{c}^2\left[1 - \mathbb{G}_0\left(\overline{\beta} - \frac{\overline{c}^2\Delta}{2c^*}\right)\right]\right\}.$$

The probability $\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\})$ can be split into two components:

$$\begin{aligned}&\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\}) \\ &\quad = \mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\}, p_n \in (L_b^\sigma, U_b^\sigma))\mathbb{P}_\sigma(p_n \in (L_b^\sigma, U_b^\sigma)) \\ &\qquad + \mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\}, p_n \notin (L_b^\sigma, U_b^\sigma))\mathbb{P}_\sigma(p_n \notin (L_b^\sigma, U_b^\sigma)) \\ &\quad = \alpha\mathbb{P}_\sigma(p_n \in (L_b^\sigma, U_b^\sigma)) \\ &\qquad + \mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\}, p_n \notin (L_b^\sigma, U_b^\sigma))\mathbb{P}_\sigma(p_n \notin (L_b^\sigma, U_b^\sigma)).\end{aligned}$$

Similarly,

$$\begin{aligned}&\mathbb{P}_\sigma(x_n = \theta \mid \mathbf{B}(n) = \mathbf{B}_n)) \\ &\quad = \mathbb{P}_\sigma(x_n = \theta \mid \mathbf{B}(n) = \mathbf{B}_n, p_n \in (L_b^\sigma, U_b^\sigma))\mathbb{P}_\sigma(p_n \in (L_b^\sigma, U_b^\sigma))\end{aligned}$$

$$+\mathbb{P}_\sigma(x_n=\theta \mid \mathbf{B}(n)=\mathbf{B}_n,\, p_n\notin(L_b^\sigma,U_b^\sigma))\mathbb{P}_\sigma(p_n\notin(L_b^\sigma,U_b^\sigma))$$

$$=\alpha'\mathbb{P}_\sigma(p_n\in(L_b^\sigma,U_b^\sigma))$$

$$+\mathbb{P}_\sigma(x_n=\theta \mid B(n)=\{b\},\, p_n\notin(L_b^\sigma,U_b^\sigma))\mathbb{P}_\sigma(p_n\notin(L_b^\sigma,U_b^\sigma)).$$

The second equality above follows because private signals are independent of the network realization, n relies on her private signal whenever $p_n\notin(L_b^\sigma,U_b^\sigma)$, and the interval (L_b^σ,U_b^σ) is deterministic given $B(n)=\{b\}$. Consider the difference between the last two expressions and note that $\mathbb{P}_\sigma(p_n\in(L_b^\sigma,U_b^\sigma))\le 1$; our claim follows immediately.

Part 2: Level of confidence

In this part of our proof, we focus on the network topology of the example and establish a lower bound on the probability, $\mathbb{P}_\sigma(x_n=\theta\mid B(n+1)=\{n\})$ for $n\in S_i$. Let $\mathbf{B}_n^*$ denote the event that $B(k+1)=\{k\}$ for $1\le k<n$. Observe

$$\begin{aligned}\mathbb{P}_\sigma(x_n=\theta\mid B(n+1)=\{n\}) &\ge \mathbb{P}_\sigma(x_n=\theta\mid \mathbf{B}_n^*)\mathbb{Q}(\mathbf{B}_n^*\mid B(n+1)=\{n\})\\ &\ge \mathbb{P}_\sigma(x_n=\theta\mid \mathbf{B}_n^*)(1-\epsilon_i)\\ &\ge \mathbb{P}_\sigma(x_n=\theta\mid \mathbf{B}_n^*)-\epsilon_i,\end{aligned}$$

where

$$\epsilon_i=1-\min_{n\in S_i}\mathbb{Q}(\mathbf{B}_n^*\mid B(n+1)=\{n\})=\frac{i}{2^i}$$

for $i\ge 1$. Let $\alpha_n=\mathbb{P}_\sigma(x_n=\theta\mid\mathbf{B}_n^*)$. The rest of this part centers on bounding α_n; $\mathbb{P}_\sigma(x_n=\theta\mid B(n+1)=\{n\})\ge\alpha_n-\epsilon_i$ then provides a bound on the quantity of interest. By Part 1 and a straightforward adaptation of the coarsening argument in Lemma 9,

$$\alpha_{n+1}\ge\overline{\mathcal{Z}}(\alpha_n-\epsilon_i). \tag{6}$$

From this point forward, we restrict our attention to a particular class of signal structures that will be shown to preclude information aggregation in this network. We assume the signal structure exhibits unbounded private beliefs and polynomial shape.

Definition 12. The private belief distributions $\mathbb{G}_0$ and $\mathbb{G}_1$ have *polynomial shape* of degree K if there exist constants $C'>0$ and $C''>0$ such that

$$\begin{aligned}\mathbb{G}_1(1-\alpha)&\ge C'(1-\alpha)^K, &\quad 1-\mathbb{G}_0(\alpha)&\ge C'(1-\alpha)^K\\ \mathbb{G}_1(\alpha)&\ge 1-C''(1-\alpha)^K, &\quad 1-\mathbb{G}_0(1-\alpha)&\ge 1-C''(1-\alpha)^K\end{aligned}$$

for all $\alpha\in[\frac{1}{2},1]$.

For a signal structure with unbounded beliefs and polynomial shape of degree K, the function $\overline{\mathcal{Z}}$ simplifies to

$$\overline{\mathcal{Z}}(\alpha)=\alpha+\frac{(1-\alpha)^2}{8}\min\left\{\mathbb{G}_1\left(\frac{1-\alpha}{2}\right),1-\mathbb{G}_0\left(\frac{1+\alpha}{2}\right)\right\}$$

and there exists a constant $C > 0$ such that $\overline{\mathcal{Z}}(\alpha) - \alpha \geq C(1-\alpha)^{K+2}$. From (6), we have

$$\alpha_{n+1} \geq \alpha_n - \epsilon_i + C(1 - (\alpha_n - \epsilon_i))^{K+2}. \tag{7}$$

Without loss of generality, we assume that (7) is increasing in α_n. If not, we can replace C by $C^* < C$, with $C^* < 2^{K+1}/(K+2)$, and the desired condition will hold. Now, define the sequence $\{\varphi_k\}_{k\in\mathbb{N}}$ by $\varphi_1 = \frac{1}{2}$ and

$$\varphi_{k+1} = \max\Bigl(\tfrac{1}{2}, \varphi_k - \epsilon_i + C(1-(\varphi_k - \epsilon_i))^{K+2}\Bigr)$$

for all $k \in S_i$, $i \geq 1$. We now establish a lower bound on α_n when $n \in S_i$, thereby providing a lower bound on $\mathbb{P}_\sigma(x_n = \theta \mid B(n+1) = \{n\})$. First, assuming a signal structure with polynomial shape, we show that $\alpha_n \geq \varphi_n$ for all n.

Proceed inductively. The case $n = 1$ is trivially satisfied since $\alpha_n \geq \frac{1}{2}$ for all n by Lemma 4. Suppose the inequality holds for an arbitrary n. Then

$$\begin{aligned}\alpha_{n+1} &\geq \alpha_n - \epsilon_i + C(1-(\alpha_n - \epsilon_i))^{K+2} \\ &\geq \varphi_n - \epsilon_i + C(1-(\varphi_n - \epsilon_i))^{K+2}.\end{aligned}$$

Therefore, $\alpha_{n+1} \geq \varphi_{n+1}$ as desired.

We further show that under the same assumptions, for all $i > 1$ and $n \in S_i$, we have

$$C(1-(\alpha_n - \epsilon_i))^{K+2} \leq 2\epsilon_{i-1}.$$

It follows that

$$\alpha_n \geq 1 - \left(\frac{2\epsilon_{i-1}}{C}\right)^{1/(K+2)}.$$

We prove the result for the sequence $\{\varphi_n\}$, and our knowledge that $\alpha_n \geq \varphi_n$ completes the proof. To simplify notation, let $f(\alpha)$ denote $C(1-\alpha)^{K+2}$. We begin by noting a few important properties of the function f:

(a) The term $f(\alpha)$ is differentiable and strictly decreasing for $\alpha \in [\frac{1}{2}, 1]$.

(b) The term $\alpha + f(\alpha)$ is increasing due to the constraint placed on C.

(b) As a result of (b), $f'(\alpha) > -1$.

Proceed by induction on i; the case $i = 2$ is trivially satisfied as $2\epsilon_1 = 1$. Assuming the result for an arbitrary i, we proceed in two steps to demonstrate the result for $i+1$:

Claim 1. *If $f(\varphi_{2^{i+1}} - \epsilon_{i+1}) \leq 2\epsilon_i$, then $f(\varphi_n - \epsilon_{i+1}) \leq 2\epsilon_i$ for all $n \in S_{i+1}$.*

Claim 2. *If $f(\varphi_{2^i} - \epsilon_i) \leq 2\epsilon_{i-1}$, then $f(\varphi_{2^{i+1}} - \epsilon_{i+1}) \leq 2\epsilon_i$, completing the inductive step.*

For Claim 1, take $n \in S_{i+1}$ and suppose $f(\varphi_n - \epsilon_{i+1}) \leq 2\epsilon_i$. We show that $f(\varphi_{n+1} - \epsilon_{i+1}) \leq 2\epsilon_i$. Consider two cases.

Case 1: $f(\varphi_n - \epsilon_{i+1}) \geq \epsilon_{i+1}$. In this case, $\varphi_{n+1} \geq \varphi_k$. $f(\alpha - \epsilon_{i+1})$ is decreasing in α, so the result follows.

Case 2: $f(\varphi_n - \epsilon_{i+1}) < \epsilon_{i+1}$. This quantity is never less than zero, so $\varphi_{n+1} \geq \varphi_n - \epsilon_{i+1}$. Using property (c),

$$f(\varphi_{n+1} - \epsilon_{i+1}) \leq f(\varphi_n - 2\epsilon_{i+1}) \leq f(\varphi_n - \epsilon_{i+1}) + \epsilon_{i+1} < 2\epsilon_{i+1} < 2\epsilon_i,$$

proving our first claim.

We now prove Claim 2. For any $n \in S_i$, the argument from Claim 1 above shows that if

$$f(\varphi_n - \epsilon_i) \leq 2\epsilon_i, \quad \text{then } f(\varphi_{n+1} - \epsilon_i) \leq 2\epsilon_i.$$

Therefore, if $f(\varphi_{2^{i+1}} - \epsilon_{i+1}) > 2\epsilon_i$, we have $f(\varphi_n - \epsilon_i) > 2\epsilon_i$ for all $n \in S_i$. However, this implies that $\varphi_{n+1} \geq \varphi_n + \epsilon_i$ for all $n \in S_i$. Since $\epsilon_i \geq 1/2^i$ and $\varphi_{2^i} \geq \frac{1}{2}$, this would imply $\varphi_{2^{i+1}} > 1$, which is not possible. Thus, $f(\varphi_{2^{i+1}} - \epsilon_i) \leq 2\epsilon_i$, completing the proof.

In light of this result, we now have the following bound for $n \in S_i$:

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n+1) = \{n\}) \geq 1 - \left(\frac{2\epsilon_{i-1}}{C}\right)^{1/(K+2)} - \epsilon_i. \tag{8}$$

Part 3: Failure to diffuse

Suppose $B(n+1) = \{n\}$ for some $n \in S_i$. By Lemma 3, agent $n+1$ chooses the same action as agent n if either $x_n = 0$ and $p_{n+1} < U_n^\sigma$ or $x_n = 1$ and $p_{n+1} > L_n^\sigma$. Therefore, given some realization $\mathbf{B}(n+1) = \mathbf{B}_{n+1}$, with $B(n+1) = \{n\}$, then

$$\begin{aligned}
&\mathbb{P}_\sigma(x_{n+1} \neq \theta \mid \mathbf{B}(n+1) = \mathbf{B}_{n+1}) \\
&\quad = \tfrac{1}{2}\big[\mathbb{P}_\sigma(x_{n+1} = 0 \mid \mathbf{B}(n+1) = \mathbf{B}_{n+1}, \theta = 1) \\
&\qquad\qquad + \mathbb{P}_\sigma(x_{n+1} = 1 \mid \mathbf{B}(n+1) = \mathbf{B}_{n+1}, \theta = 0)\big] \\
&\quad \geq \tfrac{1}{2}\big[\mathbb{P}_\sigma(x_n = 1 \mid \mathbf{B}(n+1) = \mathbf{B}_{n+1}, \theta = 0)(1 - \mathbb{G}_0(L_n^\sigma)) \\
&\qquad\qquad + \mathbb{P}_\sigma(x_n = 0 \mid \mathbf{B}(n+1) = \mathbf{B}_{n+1}, \theta = 1)\mathbb{G}_1(U_n^\sigma)\big] \\
&\quad \geq \mathbb{P}_\sigma(x_n \neq \theta \mid \mathbf{B}(n+1) = \mathbf{B}_{n+1}) \min\big[(1 - \mathbb{G}_0(L_n^\sigma)), \mathbb{G}_1(U_n^\sigma)\big].
\end{aligned}$$

Let $P_n^\sigma = \min[(1 - \mathbb{G}_0(L_n^\sigma)), \mathbb{G}_1(U_n^\sigma)]$. Further, let $\alpha = \mathbb{P}_\sigma(x_n = \theta \mid B(n+1) = \{n\})$, and rewrite L_n^σ and $1 - U_n^\sigma$ as

$$L_n^\sigma = \frac{1 - N_n^\sigma}{1 - 2N_n^\sigma + 2\alpha}, \qquad 1 - U_n^\sigma = \frac{1 - Y_n^\sigma}{2\alpha + 1 - 2Y_n^\sigma}.$$

Recall that these are decreasing in N_n^σ and Y_n^σ, respectively. Since Y_n^σ and N_n^σ are both between 0 and 1, we have $Y_n^\sigma, N_n^\sigma \geq 2\alpha - 1$, giving

$$L_n^\sigma \leq \frac{2 - 2\alpha}{3 - 2\alpha} \leq 2 - 2\alpha,$$

and similarly, $U_n^\sigma \geq 2\alpha - 1$. Thus, using the assumption that the signal structure has polynomial shape of degree K,

$$P_n^\sigma \geq \min[1 - \mathbb{G}_0(2 - 2\alpha), \mathbb{G}_1(2\alpha - 1)] \geq 1 - C'(1-\alpha)^K \tag{9}$$

for some constant C'.

Recall that with probability $\frac{1}{2}$, we have $B(k+1) = \{k\}$ for all $k \geq 1$. If this happens, the network has realized the "good" regime, and otherwise, the network has realized the "bad" regime. If the network does not realize the good regime, then in S_i, there exists an agent m_i, whose neighborhood is empty. Moreover, there is a chain of agents, starting on $m_i + 1$ and terminating on $m_i + i$, who each observe their immediate predecessor. Using Part 2, we have

$$\begin{aligned}\mathbb{P}_\sigma(x_{m_i+i} \neq \theta \mid \text{bad regime}) &\geq \mathbb{P}_\sigma(x_{m_i} \neq \theta \mid \text{bad regime}) \prod_{k=m_i}^{m_i+i-1} P_k^\sigma \\ &\geq \mathbb{P}_\sigma(x_1 \neq \theta)\left[1 - C'\left(\epsilon_i + \left(\frac{2\epsilon_{i-1}}{C}\right)^{1/(K+2)}\right)^K\right]^i,\end{aligned}$$

where the second inequality follows from (8) and (9). Since ϵ_i decays exponentially, the second term approaches 1 as n approaches infinity. For large i, the probability that agent $m_i + i$ makes a correct decision is bounded away from 1 and, in fact, is almost the same as the probability that the first agent makes a correct decision. Therefore, since the majority of agents in S_{i+1} observe only agent $m_i + i$, these agents are correct with probability bounded away from 1. For an unbounded signal structure with polynomial shape, we plainly see that information does not diffuse. We can easily construct such signal structures; for instance, the private belief distributions given by $\mathbb{G}_0(r) = 2r - r^2$ and $\mathbb{G}_1(r) = r^2$ satisfy these properties. ◊

Example 4 (Failure due to correlated actions). We say that a signal structure is symmetric if $\mathbb{G}_0(r) = 1 - \mathbb{G}_1(1-r)$ for all $r \in [0,1]$ and $\mathbb{G}_1(\mathbb{G}_0(\frac{1}{2})) = \mathbb{G}_0(\mathbb{G}_1(\frac{1}{2})) = \frac{1}{2}$. Suppose the signal structure is symmetric. Let $p(x_m)_{a,b}$ denote $\mathbb{P}_\sigma(x_m = a \mid B(m), \theta = b)$, and suppose that $\theta = 1$ and $x_1 = 0$. Then for $m \in F_n$ for any n, we have

$$\begin{aligned}p(x_m)_{1,1} &= \mathbb{P}_\sigma(p_m + q_m > 1 \mid \theta = 1) = \mathbb{P}_\sigma\left(p_m + \frac{p(x_1)_{0,1}}{p(x_1)_{0,1} + p(x_1)_{0,0}} > 1 \,\Big|\, \theta = 1\right) \\ &= \mathbb{P}_\sigma\left(p_m + \frac{\mathbb{G}_1(\frac{1}{2})}{\mathbb{G}_1(\frac{1}{2}) + \mathbb{G}_0(\frac{1}{2})} > 1 \,\Big|\, \theta = 1\right) \\ &= \mathbb{P}_\sigma\left(p_m > 1 - \mathbb{G}_1\left(\tfrac{1}{2}\right) \mid \theta = 1\right) = 1 - \mathbb{G}_1\left(1 - \mathbb{G}_1\left(\tfrac{1}{2}\right)\right) = \tfrac{1}{2},\end{aligned}$$

where the last three equalities follow from the symmetry of the signal structure. A similar calculation shows that if $\theta = 0$ and $x_1 = 1$, then $p(x_m)_{0,0} = p(x_m)_{1,0} = \frac{1}{2}$. Thus, if $x_1 \neq \theta$, all agents $m \in F_n$ are equally likely to choose $x_m = 0$ or $x_m = 1$, regardless of the true state. Furthermore, $\mathbb{P}_\sigma(x_1 = 0 \mid \theta = 1) = \mathbb{G}_1(\frac{1}{2}) = 1 - \mathbb{G}_0(\frac{1}{2}) = \mathbb{P}_\sigma(x_1 = 1 \mid \theta = 0) \equiv p$.

Now, since the agents in F_n all have exactly the same social information, they are interchangeable from the perspective of an agent observing F_n; conditional on θ and x_1, the decisions x_m for $m \in F_n$ are independent and identically distributed. Thus, the social belief of an agent n with $B(n) = F_n$ depends on the number of members of F_n who have chosen action 1. Let $P^k_{i,j}$ denote the probability that k members of F_n choose action 1 conditional on $\theta = i$ and $x_1 = j$, and let η denote the number of members in F_n who choose action 1. We calculate

$$q_n = \frac{P^\eta_{1,0} p + P^\eta_{1,1}(1-p)}{P^\eta_{1,0} p + P^\eta_{1,1}(1-p) + P^\eta_{0,1} p + P^\eta_{0,0}(1-p)}.$$

Since agents in F_n are equally likely to be right or wrong when $x_1 \neq \theta$, we have $P^\eta_{1,0} = P^\eta_{0,1}$ for any η. If $x_1 = \theta$, then each agent $m \in F_n$ makes a correct decision with probability greater than $\frac{1}{2}$. This, together with the symmetry of the signal structure, gives $P^\eta_{1,1} \geq P^\eta_{0,0}$ whenever $\eta \geq |F_n|/2$ and $P^\eta_{1,1} \leq P^\eta_{0,0}$ whenever $\eta \leq |F_n|/2$. Therefore, if $\eta \geq |F_n|/2$, then $q_n \geq \frac{1}{2}$, and if $\eta \leq |F_n|/2$, then $q_n \leq \frac{1}{2}$.

Suppose $\theta = 0$ and $x_1 = 1$. We have $\mathbb{P}_\sigma(\eta \geq |F_n|/2 \mid \theta = 0, x_1 = 1) \geq \frac{1}{2}$. For any n with $B(n) = F_n$,

$$\begin{aligned}\mathbb{P}_\sigma(x_n = 0 \mid \theta = 0, x_1 = 1, B(n) = F_n) &\leq 1 - \mathbb{P}_\sigma\left(\eta \geq \frac{|F_n|}{2}, p_n > \frac{1}{2} \;\middle|\; \theta = 0, x_1 = 1\right) \\ &\leq 1 - \tfrac{1}{2}\left(1 - \mathbb{G}_0\left(\tfrac{1}{2}\right)\right).\end{aligned}$$

Similarly, $\mathbb{P}_\sigma(x_n = 1 \mid \theta = 1, x_1 = 0, B(n) = F_n) \leq 1 - (\mathbb{G}_1(\frac{1}{2}))/2$. Combining these results with the probability that the first agent errs gives us

$$\mathbb{P}_\sigma(x_n \neq \theta \mid B(n) = F_n) \geq \tfrac{1}{4} p\left[1 - \mathbb{G}_0\left(\tfrac{1}{2}\right) + \mathbb{G}_1\left(\tfrac{1}{2}\right)\right].$$

As n approaches infinity, the probability that $B(n) = F_n$ approaches 1. All that remains to complete the proof is to establish the existence of symmetric signal structures for which $\mathbb{P}_\sigma(x_n \neq \theta \mid B(n) = F_n) > 1 - \beta^*$. Taking $\mathbb{G}_0(r) = 2r - r^2$ and $\mathbb{G}_1(r) = r^2$ provides one example. ◊

Example 5 (Learning with preferential attachment). We show that this network satisfies the hypotheses of Theorem 3. We first show that $\delta_m(B(n) = \{m\}) = 0$ for all $m < n$. Proving that this network features expanding subnetworks then establishes the result.

Suppose $\mathbf{B}_m$ is a realization of $\mathbf{B}(m)$. We proceed by induction to show that

$$\mathbb{Q}(B(n) = \{m\} \mid \mathbf{B}(m) = \mathbf{B}_m) = \mathbb{Q}(B(n) = \{m\})$$

for $n > m$. Once we establish this, a simple application of Bayes' theorem gives

$$\mathbb{Q}(\mathbf{B}(m) = \mathbf{B}_m) = \mathbb{Q}(\mathbf{B}(m) = \mathbf{B}_m \mid B(n) = \{m\}),$$

which implies $\hat{\delta}_m(B(n) = \{m\}) = 0$.

We will actually employ a slightly stronger inductive hypothesis: we show that for any $n > m$, $\mathbb{Q}(\lambda_n(m) = i \mid \mathbf{B}(m) = \mathbf{B}_m) = \mathbb{Q}(\lambda_n(m) = i)$. The desired result follows from the calculation

$$\mathbb{Q}(B(n) = \{m\} \mid \mathbf{B}(m) = \mathbf{B}_m) = \frac{1}{2(n-1)} \sum_{i=1}^{n-m} i\mathbb{Q}(\lambda_n(m) = i \mid \mathbf{B}(m) = \mathbf{B}_m).$$

For $n = m+1$, we trivially have $\lambda_n(m) = 1$ with probability 1. Now suppose the identity holds for $m < n \leq m+k-1$. For $n = m+k$, we can write $\mathbb{Q}(\lambda_n(m) = i \mid \mathbf{B}(m) = \mathbf{B}_m)$ as

$$\begin{aligned}
&\mathbb{Q}(\lambda_{n-1}(m) = i \mid \mathbf{B}(m) = \mathbf{B}_m)\mathbb{Q}(B(n-1) \neq \{m\} \mid \lambda_{n-1}(m) = i) \\
&\qquad + \mathbb{Q}(\lambda_{n-1}(m) = i-1 \mid \mathbf{B}(m) = \mathbf{B}_m)\mathbb{Q}(B(n-1) = \{m\} \mid \lambda_{n-1}(m) = i-1) \\
&\quad = \mathbb{Q}(\lambda_{n-1}(m) = i)\mathbb{Q}(B(n-1) \neq \{m\} \mid \lambda_{n-1}(m) = i) \\
&\qquad + \mathbb{Q}(\lambda_{n-1}(m) = i-1)\mathbb{Q}(B(n-1) = \{m\} \mid \lambda_{n-1}(m) = i-1) = \mathbb{Q}(\lambda_n(m) = i),
\end{aligned}$$

where the first equality follows from the inductive hypothesis. Thus, there is no distortion in this network.

We now show that the network has expanding subnetworks. We show inductively that $\limsup_{n\to\infty} \mathbb{Q}(d(n) < K) = 0$ for all positive integers K. We clearly have $d(n) \geq 2$ for all $n > 1$, establishing the base case. We now assume the result for some integer K and prove that it holds for $K+1$. Since there is no distortion, we have

$$\begin{aligned}
\mathbb{Q}(d(n) < K+1) &= \sum_{m<n} \mathbb{Q}(B(n) = \{m\})\mathbb{Q}(d(m) < K \mid B(n) = \{m\}) \\
&= \sum_{m<n} \mathbb{Q}(B(n) = \{m\})\mathbb{Q}(d(m) < K).
\end{aligned}$$

By our inductive hypothesis, given $\epsilon > 0$, there exists a positive integer $N(K, \epsilon)$ such that for all $n \geq N(K, \epsilon)$, $\mathbb{Q}(d(n) < K) < \epsilon/2$. By the above calculation, if we can find $N(K+1, \epsilon)$ such that for any $n \geq N(K+1, \epsilon)$,

$$\sum_{m<N(K,\epsilon)} \mathbb{Q}(B(n) = \{m\}) < \frac{\epsilon}{2},$$

then $\mathbb{Q}(d(n) < K+1) < \epsilon$ for any $n \geq N(K+1, \epsilon)$. We now show that

$$\limsup_{n\to\infty} \sum_{m<N} \mathbb{Q}(B(n) = \{m\}) = 0$$

for any positive integer N, completing the proof.

Define $\lambda_{n,N}$ to be the number of network connections to the first $N-1$ agents after the first $n-1$ neighborhood realizations. That is,

$$\lambda_{n,N} = \begin{cases} 2(n-2) & \text{if } n < N \\ 2(N-2) + |\{l : l \geq N, B(l) = \{m\}, m < N\}| & \text{otherwise.} \end{cases}$$

Given an agent $n > N$, let P_n denote the probability that $B(n) = \{m\}$ with $m < N$:

$$P_n = \sum_{i=2(N-2)+1}^{n+N-4} \mathbb{Q}(\lambda_{n,N} = i)\frac{i}{2(n-1)}.$$

Alternatively, we can express P_n in terms of P_{n-1} by noting that $\lambda_{n,N} = \lambda_{n-1,N}$ if agent $n-1$ observes an agent with index at least N, and $\lambda_{n,N} = \lambda_{n-1,N} + 1$ otherwise. We have

$$\begin{aligned} P_n &= \sum_{i=2(N-2)+1}^{n+N-5} \mathbb{Q}(\lambda_{n-1,N} = i)\left[P_{n-1}\frac{i+2}{2(n-2)} + (1-P_{n-1})\frac{i+1}{2(n-2)}\right] \\ &= \sum_{i=2(N-2)+1}^{n+N-5} \frac{\mathbb{Q}(\lambda_{n-1,N} = i)(i+1)}{2(n-2)} \\ &\qquad + \sum_{j=2(N-2)+1}^{n+N-5}\sum_{i=2(N-2)+1}^{n+N-5} \frac{j\mathbb{Q}(\lambda_{n-1,N} = j)}{2(n-3)}\frac{\mathbb{Q}(\lambda_{n-1,N} = i)}{2(n-2)} \\ &= \frac{1}{2(n-2)} + \frac{n-3}{n-2}P_{n-1} + \frac{1}{2(n-2)}P_{n-1} = \frac{1+(2n-5)P_{n-1}}{2(n-2)}. \end{aligned}$$

Note that $P_n > P_{n-1}$ as long as $P_{n-1} < 1$ and, furthermore, $P_n - P_{n-1} = (1-P_{n-1})/(2(n-2))$, which goes to zero as n goes to infinity. Therefore, $\{P_n\}$ is an increasing Cauchy sequence with a limit P^*. We finish our proof by showing that $P^* = 1$. Suppose $P^* < 1$. We have $P_n < P^*$ for all n, from which we obtain $P_n - P_{n-1} > (1-P^*)/(2(n-2))$ for all n. However, $\sum_{k=1}^{\infty} 1/(2(k-2)) = \infty$, implying that $P^* = \infty$, a contradiction. ◊

Example 6 (Learning with nonzero distortion). We proceed as in the last example to show that this network satisfies the hypotheses of Theorem 3. Since each agent observes only one neighbor, we may take $\gamma_n(B(n))$ to be the lone agent contained in $B(n)$. We first establish low distortion by bounding $\delta_m(B(n) = \{m\})$ for $m < n$.

Suppose $n \in T_i$ and $T_j = \{m, m'\}$ for some $j < i$. We compute the network distortion $\delta_m(B(n) = \{m\})$ from the definition:

$$\begin{aligned} \delta_m(B(n) = \{m\}) &= \sum_{\mathbf{B}_m}\left|\mathbb{Q}(\mathbf{B}(m) = \mathbf{B}_m \mid B(n) = \{m\}) - \mathbb{Q}(\mathbf{B}(m) = \mathbf{B}_m)\right| \\ &= \sum_{\mathbf{B}_m}\left|\frac{\mathbb{Q}(B(n) = \{m\} \mid \mathbf{B}(m) = \mathbf{B}_m)\mathbb{Q}(\mathbf{B}(m) = \mathbf{B}_m)}{\mathbb{Q}(B(n) = \{m\})} - \mathbb{Q}(\mathbf{B}(m) = \mathbf{B}_m)\right|. \end{aligned}$$

From the definition of the network topology, we have

$$\frac{i-j-1}{i-j+1} \le \frac{\mathbb{Q}(B(n) = \{m\} \mid \mathbf{B}(m) = \mathbf{B}_m)}{\mathbb{Q}(B(n) = \{m\})} \le \frac{i-j+1}{i-j-1},$$

and it follows that $\delta_m(B(n) = \{m\}) \le 1/(i-j)$.

To show that the chosen neighbor topology has low network distortion, we must show that for any $\epsilon > 0$, the probability that n's neighbor has network distortion at least ϵ goes to zero as n approaches infinity. Given the above bound on the network distortion, this is equivalent to showing that given any positive integer K, $\lim_{n\to\infty} \mathbb{Q}(B(n) = \{m\}, m > n - K) = 0$. The bound $\mathbb{Q}(B(n) = \{m\}, m > n - K) \leq (K + 1)/(n-2)$ establishes this result.

We now show that the chosen neighbor topology features expanding subnetworks by inductively showing that $\limsup_{n\to\infty} \mathbb{Q}(d(n) < K) = 0$ for all positive integers K. For all $n > 2$, $d(n) \geq 2$ by our definition of the network topology. If the result holds for some positive integer K, then given $\epsilon > 0$, we can find $N_1(K, \epsilon)$ such that $\mathbb{Q}(d(n) < K) \leq \epsilon/4$ whenever $n \geq N_1(K, \epsilon)$. We have $\mathbb{Q}(B(n) = \{m\}, m < N_1(K, \epsilon)) \leq (N_1(K, \epsilon))/(n-2)$, which approaches zero as n approaches infinity. Thus, we can find $N_2(K, \epsilon)$ such that for any $n \geq N_2(K, \epsilon)$, we have

$$\mathbb{Q}(B(n) = \{m\}, m < N_1(K, \epsilon)) \leq \frac{\epsilon}{4}.$$

Since the chosen neighbor topology has low network distortion, we can find $N(\epsilon)$ such that

$$\sum_{m:\hat{\delta}(B(n)=\{m\})\geq\epsilon/4} \mathbb{Q}(B(n) = \{m\}) < \frac{\epsilon}{4}$$

whenever $n \geq N(\epsilon)$. Thus, if $n \geq \max(N_2(K, \epsilon), N(\epsilon))$, then $\mathbb{Q}(d(n) < K + 1) < \epsilon$, completing our proof. $\Diamond$

References

Acemoglu, Daron, Kostas Bimpikis, and Asuman Ozdaglar (2014), "Dynamics of information exchange in endogenous social networks." *Theoretical Economics*, 9, 41–97. [812]

Acemoglu, Daron, Munther A. Dahleh, Ilan Lobel, and Asuman Ozdaglar (2011), "Bayesian learning in social networks." *Review of Economic Studies*, 78, 1201–1236. [807, 808, 809, 811, 813, 814, 815, 817, 820, 824, 826, 830]

Acemoglu, Daron, Asuman Ozdaglar, and Ali ParandehGheibi (2010), "Spread of misinformation in social networks." *Games and Economic Behavior*, 70, 194–227. [812]

Acemoglu, Daron and Asuman Ozdaglar (2011), "Opinion dynamics and learning in social networks." *Dynamic Games and Applications*, 1, 3–49. [812]

Ali, S. Nageeb and Navin Kartik (2012), "Herding with collective preferences." *Economic Theory*, 51, 601–626. [812]

Arieli, Itai and Manuel Mueller-Frank (2012), "Generic outcomes of observational learning." Unpublished paper. [814]

Bala, Venkatesh and Sanjeev Goyal (1998), "Learning from neighbours." *Review of Economic Studies*, 65, 595–621. [811]

Banerjee, Abhijit and Drew Fudenberg (2004), "Word-of-mouth learning." *Games and Economic Behavior*, 46, 1–22. [809, 811]

Banerjee, Abhijit V. (1992), "A simple model of herd behavior." *Quarterly Journal of Economics*, 107, 797–817. [808, 809, 811, 815, 817]

Barabási, Albert-László and Réka Albert (1999), "Emergence of scaling in random networks." *Science*, 286, 509–512. [812, 826]

Bender, Edward A. and E. Rodney Canfield (1978), "The asymptotic number of labeled graphs with given degree sequences." *Journal of Combinatorial Theory*, 24, 296–307. [828]

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992), "A theory of fads, fashion, custom, and cultural change as informational cascades." *Journal of Political Economy*, 100, 992–1026. [809, 811, 815, 817]

Bollobás, Béla (1980), "A probabilistic proof of an asymptotic formula for the number of labelled regular graphs." *European Journal of Combinatorics*, 1, 311–316. [828]

Çelen, Boğaçhan and Shachar Kariv (2004), "Observational learning under imperfect information." *Games and Economic Behavior*, 47, 72–86. [809, 811]

DeMarzo, Peter M., Dimitri Vayanos, and Jeffrey Zwiebel (2003), "Persuasion bias, social influence, and unidimensional opinions." *Quarterly Journal of Economics*, 118, 909–968. [811]

Ellison, Glenn and Drew Fudenberg (1993), "Rules of thumb for social learning." *Journal of Political Economy*, 101, 612–643. [811]

Ellison, Glenn and Drew Fudenberg (1995), "Word-of-mouth communication and social learning." *Quarterly Journal of Economics*, 110, 93–125. [811]

Eyster, Erik and Matthew Rabin (2011), "Rational observational learning." Unpublished paper.[812]

Fan, Jianqing, Xin Tong, and Yao Zeng (2012), "Communication learning in social networks: Finite population and the rates." Unpublished paper. [812]

Gale, Douglas M. and Shachar Kariv (2003), "Bayesian learning in social networks." *Games and Economic Behavior*, 45, 329–346. [811]

Golub, Benjamin and Matthew O. Jackson (2010), "Naïve learning in social networks and the wisdom of crowds." *American Economic Journal: Microeconomics*, 2, 112–149. [812]

Golub, Benjamin and Matthew O. Jackson (2012), "How homophily affects the speed of learning and best-response dynamics." *Quarterly Journal of Economics*, 127, 1287–1338. [812]

Guarino, Antonio and Antonella Ianni (2010), "Bayesian social learning with local interactions." *Games*, 1, 438–458. [812]

Jackson, Matthew O. (2007), "The study of social networks in economics." In *The Missing Links: Formation and Decay of Economic Networks* (James E. Rauch, ed.), Russell Sage Foundation, Thousand Oaks. [812]

Jadbabaie, Ali, Pooya Molavi, Alvaro Sandroni, and Alireza Tahbaz-Salehi (2012), "Non-Bayesian social learning." *Games and Economic Behavior*, 76, 210–225. [812]

Lee, In Ho (1993), "On the convergence of informational cascades." *Journal of Economic Theory*, 61, 395–411. [814]

Mossel, Elchanan, Allan Sly, and Omer Tamuz (2012), "On agreement and learning." Unpublished paper. [813, 815]

Mueller-Frank, Manuel and Mallesh M. Pai (forthcoming), "Social learning with costly search." *American Economic Journal: Microeconomics*. [815]

Smith, Lones and Peter Sørensen (2000), "Pathological outcomes of observational learning." *Econometrica*, 68, 371–398. [809, 811, 813, 814, 815, 818, 819]

Smith, Lones and Peter N. Sørensen (2013), "Rational social learning with random sampling." Unpublished paper.[809, 811, 813, 815]

Submitted 2013-5-28. Final version accepted 2014-8-18. Available online 2014-9-2.