

# Learning Dynamics in Social Networks\*

Simon Board<sup>†</sup> and Moritz Meyer-ter-Vehn<sup>‡</sup>

July 3, 2020

## Abstract

This paper proposes a tractable model of Bayesian learning on social networks in which agents choose whether to adopt an innovation. We consider both deterministic and random networks, and study the impact of the network structure on learning dynamics and diffusion. In directed “tree like” networks (e.g. stochastic block networks), all direct and indirect links contribute to an agent’s learning. In comparison, learning and welfare are lower in undirected networks and networks with clusters. In a broad set of networks, behavior can be captured by a small number of differential equations, making the model appropriate for empirical work.

---

\*We have received useful comments from Ben Golub, Niccolo Lomys, Alessandro Pavan, Andy Postlewaite, Klaus Schmidt, Bruno Strulovici, Fernando Vega-Redondo, Chen Zhao, and especially Evan Sadler. Hersh Chopra provided excellent research assistance. We also thank seminar audiences at Bar-Ilan, Bocconi, CEMFI, Chapman, Chicago Theory Conference, Columbia Theory Conference, Conference on Network Science and Economics, Hebrew, HKU, Indiana, Michigan, Munich, NASM, Northwestern, NYU, Oxford, Penn State, Pittsburgh, Princeton, Purdue, Seoul National, Tel-Aviv, Toronto, Toulouse, UC3M, UCLA, UCSD, Utah, Warwick, Yale. Keywords: networks, diffusion, social learning. JEL codes: D83, D85.

<sup>†</sup>UCLA, <http://www.econ.ucla.edu/sboard/>

<sup>‡</sup>UCLA, <http://www.econ.ucla.edu/mtv/>

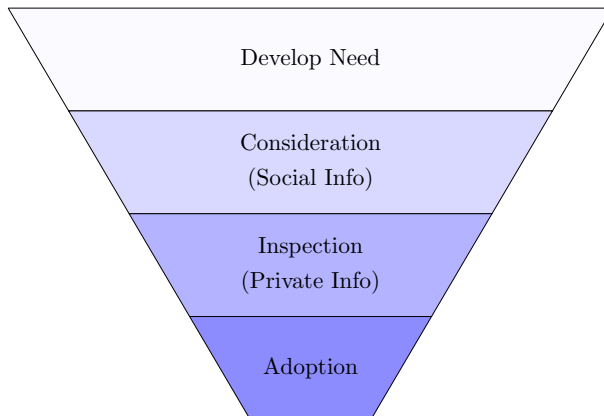


Figure 1: **The Social Purchasing Funnel.**

## 1 Introduction

How do communities, organizations, or entire societies learn about innovations? Consider consumers learning about a new brand of electric car from friends, farmers learning about a novel crop from neighbors, or entrepreneurs learning about a source of finance from nearby businesses. In all these instances agents learn from others' choices, so the diffusion of the innovation depends on the social network. One would like to know: Do agents learn more in a highly connected network? Do products diffuse faster in a more centralized network?

This project proposes a tractable, Bayesian model to answer these questions. The model separates the role of social and private information as illustrated in the “social purchasing funnel” in Figure 1. First, at an exogenous time, an agent develops a need for an innovation. For example, a person’s car breaks down, and she contemplates buying a new brand of electric car. Second, at the consideration stage, she observes how many of her friends drive the car and makes an inference about its quality. Third, if the social information is sufficiently positive, she inspects the car by taking it for a test drive. Finally, she decides whether to buy the car, which in turn provides information for her friends.

We characterize the diffusion of innovation in the social network via a system of differential equations. In contrast to most papers in the literature (e.g. Acemoglu et al., 2011), our results speak to learning dynamics at each point in time, rather than

focusing on long-run behavior. We thus recover the tractability of the reduced-form models of diffusion (e.g. Bass, 1969) in a model of Bayesian learning. Understanding the full dynamics is important because empirical researchers must identify economic models from finite data, and because in practice, governments and firms care about *when* innovations take off, not just *if* they take off.

Our paper has two contributions. First, we describe how learning dynamics depend on the network structure. For directed “tree-like” networks, we show that an agent typically benefits when her neighbors have more links. Beyond tree-like networks, additional links that correlate the information of an agent’s neighbors or create feedback loops can muddle her learning, meaning that welfare is higher in “decentralized” networks than in “centralized” ones. These results can help us understand how the diffusion of products and ideas changes with the introduction of social media, differs between cities and villages, and is affected by government programs that form new social links (e.g. Cai and Szeidl, 2018).

Our second contribution is methodological. The complexity of Bayesian updating means that applied and empirical papers typically study heuristic updating rules on the exact network (e.g. Golub and Jackson (2012), Banerjee et al. (2013)). In comparison, we take a “macroeconomic approach” by studying exact Bayesian updating on the approximate network. Figure 2 illustrates three random networks exhibiting clusters and homophily that we can analyze with low-dimensional ODEs. Given a “real life” network one can then study diffusion and learning on an approximate network which shares the same network statistics (e.g. agents’ types, degree distributions, cluster coefficients).

In the model, agents are connected via a directed network. They may know the entire network (our “deterministic” networks) or only their local neighborhood (our “random” networks). An agent “enters” at a random time and considers a product (or innovation) whose quality is high or low. The agent observes which of her neighbors has adopted the product when she enters and chooses whether to inspect the product at a cost. Inspection reveals the common quality and the agent’s idiosyncratic preferences for the product. The agent then adopts the product if it is high quality and it fits her personal preference.

The agent learns directly from her neighbors via their adoption decisions; she also learns indirectly from further removed agents as their adoption decisions influence her neighbors' inspection (and adoption) decisions. The agent's own inspection decision is thus based on the hypothesized inspection decisions of her neighbors, which collectively generate her *social learning curve* (formally, the probability one of her neighbors adopts a high-quality product as a function of time). In turn, her adoption decision feeds into the social learning curves of her neighbors.

In Section 3 we characterize the joint adoption decisions of all agents via a system of ordinary differential equations. For a general network, the dimension of this system is exponential in the number of agents,  $I$ , since one must keep track of correlations between individual adoption rates; e.g. if two agents have a neighbor in common, their adoption decisions are correlated. To reduce the dimensionality of this system and derive substantive insights about social learning, we consider two ways of imposing structure on the network.

In Section 4 we consider simple deterministic networks, so agents know how everyone in the economy is connected. These “elemental” networks serve as building blocks for large, random networks, and also for our economic intuition. We start with a directed tree, where an agent's neighbors receive independent information. In such trees, it suffices to keep track of individual adoption decisions, meaning the system of ODEs reduces to  $I$  dimensions (or fewer). Given a mild condition on the hazard rate of inspection costs, an agent's adoption rate rises in her level of social information. This means that more neighbors lead to more adoption, which leads to more information, which leads to more adoption all the way down the tree. Thus, an agent benefits from both direct and indirect links.

Beyond tree networks, we show that agent  $i$  need not benefit from additional links of her neighbors. First, adding a *correlating* link between two of  $i$ 's neighbors harms  $i$ 's learning because the correlation raises the probability that neither of them adopts the product. Second, when we add a *backward* link, from  $i$ 's neighbor  $j$  to  $i$ , this lowers  $j$ 's adoption rate and thereby  $i$ 's information and utility. Intuitively,  $i$  cares about  $j$ 's adoption when she enters; prior to this time,  $j$  could not have seen  $i$  adopt, and so the backward link makes  $j$  more pessimistic and lowers his adoption. Using these insights, we characterize adoption in a cluster (i.e. a complete network) via a

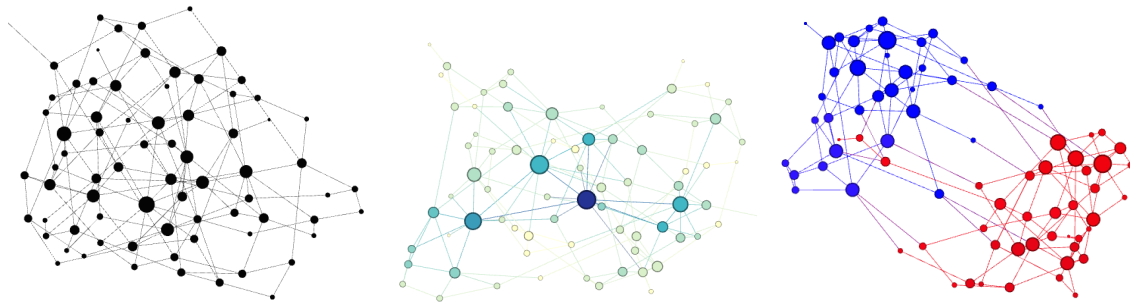


Figure 2: **Illustrative Networks.** The **left** panel shows an Erdős-Rényi random network, where social learning is described by a one-dimensional ODE, (21). The **middle** panel shows a random network with triangles, where social learning is described by a two-dimensional ODE, (25-26). The **right** panel shows a random network with homophily, where social learning is described by a four-dimensional ODE, (36).

one-dimensional ODE, and use it to show that for a given number of neighbors agents prefer a “decentralized” tree network to a “centralized” complete network.

In Section 5 we turn to large random networks, where agents know their neighbors, but not their neighbors’ neighbors. Such incomplete information is both realistic and can simplify the analysis: the less agents know, the less they can condition on, and the simpler their behavior. Formally, we model network formation via the configuration model: Agents are endowed with link-stubs that we randomly connect in pairs or triples. In the limit economy with infinitely many agents, we characterize adoption behavior both for directed networks with multiple types of agents (e.g. Twitter), and undirected networks with clusters (e.g. Facebook) in terms of a low-dimensional system of ODEs.<sup>1</sup> We validate our analysis, by showing that equilibrium behavior in large finite networks converges to the solution of these ODEs.

The ODEs allow for sharp comparative statics of social learning as a function of the network structure. Large networks with bilateral links locally look like trees, so we confirm the result that learning and welfare improve in the number of links. We also show that learning is superior in directed networks than in undirected networks and clustered networks.

Finally, we connect our theory to prominent themes in the literature on learning in

---

<sup>1</sup>The directed network with multiple types has one ODE per type. The undirected network with clusters has one ODE per type of link (e.g. bilateral links, triangles,...).

networks. First, we extend the model to allow for correlation neglect (e.g. Eyster and Rabin (2014)), and show that it reduces learning and welfare. Intuitively, agent  $i$ 's mis-specification causes her to over-estimate the chance of observing an adoption, and means she grows overly pessimistic when none of her neighbors adopt; this reduces  $i$ 's adoption and other agents' social information. Second, we reconsider the classic question of information aggregation, (e.g. Smith and Sørensen (1996), Acemoglu et al. (2011)) by letting the network and average degree grow large. When the network remains sparse, agents aggregate information perfectly; yet when it becomes clustered, information aggregation may fail. Thus, adding links may lower social welfare.

## 1.1 Literature

The literature on observational learning originates with the classic papers of Banerjee (1992) and Bikhchandani et al. (1992). In these models, agents observe both a private signal and the actions of all prior agents before making their decision. Smith and Sørensen (2000) show that “asymptotic learning” arises if the likelihood ratios of signals are unbounded. Smith and Sørensen (1996) and Acemoglu et al. (2011) dispense with the assumption that an agent can observe all prior agents' actions, and interpret the resulting observation structure as a social network. The latter paper generalizes Smith and Sorensen's asymptotic learning result to the case where agents are (indirectly) connected to an unbounded number of other agents.<sup>2</sup>

Our model departs from these papers in two ways. First, the “inspection” aspect of our model separates the role of social and private information, endogenizing the latter. A few recent papers have considered models with this flavor. Assuming agents observe all predecessors, Mueller-Frank and Pai (2016) and Ali (2018) show asymptotic learning is perfect if experimentation costs are unbounded below. In a network setting, Lomys (2019) reaches the same conclusion if, in addition, the network is sufficiently connected.

Second, the “adoption” aspect of our model complicates agents' inference problem when observing no adoption.<sup>3</sup> A number of papers have analyzed related problems in

---

<sup>2</sup>For further limit results see Monzón and Rapp (2014) and Lobel and Sadler (2015).

<sup>3</sup>There is a wider literature on product adoption without learning. There are “awareness” models in which an agent becomes aware of the product when her neighbors adopt it. One can view Bass

complete networks. Guarino et al. (2011) suppose an agent sees how many others have adopted the product, but not the timing of others’ actions or even her own action. Herrera and Hörner (2013) suppose an agent observes who adopted and when they did so, but not who refrained from adopting. Hendricks et al. (2012) suppose an agent knows the order in which others move, but only sees the total number of adoptions; as in our model, the agent then uses this public information to acquire information before making her purchasing decision. These papers characterize asymptotic behavior, and find an asymmetry in social learning: good products may fail but bad products cannot succeed. In Section 5.6 we show a similar result applies to our setting.

Our key contribution over this literature lies in the questions we ask. Traditionally, herding papers ask whether society correctly aggregates information as the number of agents grows. In their survey of observational learning models, Golub and Sadler (2016) write:

“A significant gap in our knowledge concerns short-run dynamics and rates of learning in these models. [...] The complexity of Bayesian updating in a network makes this difficult, but even limited results would offer a valuable contribution to the literature.”

In this paper we characterize such “short-run” learning dynamics in social networks. We then study how an agent’s information and welfare varies with the network structure.<sup>4</sup>

## 2 Model

**The Network.** A finite set of  $I$  agents is connected via an exogenous, directed network  $G \subseteq I \times I$  that represents which agents observe the actions of others. If  $i$

---

(1969) as such a model with random matching; Campbell (2013) studies diffusion on a fixed network. There are also models of “local network goods” where an agent wants to adopt the product if enough of her neighbors also adopt. Morris (2000) characterizes stable points in such a game. Sadler (2020) puts these forces together, and studies diffusion of a network good where agents become aware from her neighbors. Banerjee (1993) and McAdams and Song (2020) integrate awareness and social learning, allowing people to infer the quality from the time at which they become aware of a good.

<sup>4</sup>A different approach is to look at the rate at which agents’ beliefs converge. For example, Hann-Caruthers, Martynov, and Tamuz (2018) compare the cases of “observable signals” and “observable actions” in the classic herding model of Bikhchandani et al. (1992).

(Iris) observes  $j$  (John) we write  $i \rightarrow j$  or  $(i, j) \in G$ , say  $i$  is *linked to*  $j$  and call  $j$  a *neighbor* of  $i$ . We write the set of  $i$ 's neighbors as  $N_i(G)$ . Agents may have incomplete information about the network. We capture such information via finite *signals*  $\xi_i \in \Xi_i$  and a joint prior distribution over networks and signal profiles  $\mu(G, \xi)$ . A random network is given by  $\mathcal{G} = (I, \Xi, \mu)$ .

To be more concrete, consider several special cases. We describe the configuration models in detail in Section 5.

- **Deterministic network.** Signal spaces are degenerate,  $|\Xi_i| = 1$ , and the prior  $\mu$  assigns probability one to one network,  $G$ . While complete information might appear to be a simplifying assumption, in fact equilibrium analysis becomes very complicated once we move beyond the simplest networks  $G$ ; this motivates us to study random networks with incomplete information.
- **Directed configuration model with finite types**  $\theta \in \Theta$ .<sup>5</sup> Each agent has a type  $\theta$  and draws random stubs for each type  $\theta'$ . We then randomly connect the type  $\theta'$  stubs to type  $\theta'$  agents. Agents know how many outlinks of each type they have. For example, Twitter users know what kind of other users they follow.
- **Undirected configuration model with binary links and triangles.**<sup>6</sup> Each agent draws  $\bar{d}$  binary stubs and  $\hat{d}$  pairs of triangle stubs. We then randomly connect binary stubs in pairs, and pairs of triangle stubs in triples. Agents know how many binary and triangle links they have. For example, consider groups of friends linked on Facebook.

**The Game.** The agents seek to learn about the quality of a single product of quality  $q \in \{L, H\} = \{0, 1\}$ . Time is continuous,  $t \in [0, 1]$ . At time  $t = 0$ , agents share a common prior  $\Pr(q = H) = \pi_0 \in (0, 1)$ , independent of network  $G$  and signals  $\xi$ .

Agent  $i$  develops a need for the product, or *enters*, at time  $t_i \sim U[0, 1]$ .<sup>7</sup> She observes which of her neighbors have adopted the product by time  $t_i$  and updates

---

<sup>5</sup>This generalizes the directed Erdős-Rényi network and the directed stochastic block network to allow for arbitrary distributions of degrees.

<sup>6</sup>This generalizes the undirected Erdős-Rényi network to allow for arbitrary distributions of degrees and clusters.

<sup>7</sup>The uniform distribution is a normalization:  $t_i$  should not be interpreted as calendar time, but rather as time-quantile in the product life-cycle.



her belief about the quality of the product. The agent then chooses whether or not to inspect the product at cost  $\kappa_i \sim F[\underline{\kappa}, \bar{\kappa}]$ , with bounded pdf  $f$ . The agent inspects when indifferent. If she inspects the product, she adopts it with probability  $\alpha \in (0, 1]$  if  $q = H$  and with probability 0 if  $q = L$ ; this random adoption is independent across agents, reflecting idiosyncratic taste or fit. Entry times  $t_i$ , inspection costs  $\kappa_i$ , and idiosyncratic tastes are private information, independent within agents, and iid across agents. All of these are independent of product quality  $q$ , the network  $G$ , and agents' signals  $\xi$ .

Agents receive utility  $1/\alpha$  from adopting a product; in expectation over idiosyncratic tastes, they thus receive utility one from a high-quality product, and zero from a low-quality product. For simplicity, we assume that low-cost agents inspect given the prior information,  $\underline{\kappa} < \pi_0$ , and all cost types inspect when they observe an adoption,  $\bar{\kappa} \leq 1$ .

**Remarks.** The model makes several assumptions of note, which we discuss in the order of their timing in the social purchasing funnel in Figure 1. First, we assume the agent must act at her exogenous entry time  $t_i$  and cannot delay her decision. For example, when the consumer's current car breaks down she needs to buy a new one. Methodologically, this means our model is in the spirit of traditional herding models rather than timing games such as Chamley and Gale (1994).

Second, we assume that an agent only observes the adoption decisions of her neighbors, but not their entry times or inspection decisions. Learning is thus asymmetric: If agent  $i$  sees that  $j$  has adopted, she knows that the product is high quality. Conversely, if she sees that  $j$  has not adopted, she must infer whether (i) he has yet to develop a need for the product, (ii) he developed a need, but chose not to inspect, (iii) he inspected and found the quality to be high, but did not adopt for idiosyncratic reasons, or (iv) he inspected and found the quality to be low. The assumption that only product adoption is observable is consistent with traditional observational learning models, and seems reasonable in several types of applications. For example, consider an agent learning about the quality of a new movie from social media. She infers that people recommending the movie watched it and liked it, but she would typically not ask all the people who did not post about the movie why they did not

do so.<sup>8</sup>

Third, we assume that an agent cannot adopt a good without inspecting it. This is optimal if the agents’ cost from adopting a good with a poor idiosyncratic fit is sufficiently high (e.g. she needs to inspect an electric car to ensure it can be charged at home). It also simplifies the analysis since the agent only has one decision: whether or not to inspect.

Fourth, we assume that the agent only adopts the product if it is high quality. Adoption is thus proof of high quality and an agent who observes an adoption always chooses to inspect. This simplifies the analysis since we need only keep track of whether an agent has seen at least one adoption. In Appendix C we extend the analysis to the case where agents adopt low-quality products with positive probability.

Finally, we assume that the agent learns product quality perfectly by inspecting the good. Thus, social information determines the inspection decision, but is rendered obsolete in the adoption decision. This makes the model more tractable than traditional herding models, where social and private information need to be aggregated via Bayes’ rule.

## 2.1 Examples

The next two examples illustrate agents’ inference problem.

**Example 1 (Directed pair  $i \rightarrow j$ ).** Suppose there are two agents, Iris and John. John has no social information, while Iris observes John. Let  $x_{j,t}$  be the probability that John adopts product  $H$  by time  $t$ .<sup>9</sup> Since he enters uniformly over  $t \in [0, 1]$ , the time-derivative  $\dot{x}_{j,t}$  equals the probability he adopts conditional on entering at time  $t$ . Dropping the time-subscript, we have

$$\dot{x}_j = \Pr(j \text{ adopt}) = \alpha \Pr(j \text{ inspect}) = \alpha F(\pi_0). \quad (1)$$

---

<sup>8</sup>In this application, “inspection” corresponds to watching the movie and “adoption” corresponds to recommending it on social media. This contrasts with applications to more expensive products such as electrical cars, where “inspection” is interpreted literally while “adoption” corresponds to the purchase decision.

<sup>9</sup>Since no agent adopts when  $\theta = L$ , it suffices to keep track of the adoption probability conditional on  $\theta = H$ .

Given his prior  $\pi_0$ , John's expected utility from inspecting the good is  $\pi_0 - \kappa_j$ . He thus inspects with probability  $F(\pi_0)$  and adopts with probability  $\alpha F(\pi_0)$ .

Now consider Iris, and suppose she enters at time  $t$ . She learns by observing whether John has adopted. We thus interpret John's adoption curve  $x_j$  as Iris's social learning curve. If John has adopted, then Iris infers that quality is high and chooses to inspect. Conversely, if John has not adopted, then Iris's posterior that the quality is high at time  $t$  is given by Bayes' rule,

$$\pi(1 - x_j) := \frac{(1 - x_j)\pi_0}{(1 - x_j)\pi_0 + (1 - \pi_0)}, \quad (2)$$

and Iris inspects if  $\kappa_i \leq \pi(1 - x_j)$ . All told, Iris's adoption rate is given by

$$\begin{aligned} \dot{x}_i &= \Pr(i \text{ adopt}) = \alpha \Pr(i \text{ inspect}) = \alpha [1 - \Pr(i \text{ not inspect})] \\ &= \alpha [1 - \Pr(j \text{ not adopt}) \times \Pr(i \text{ not inspect} | j \text{ not adopt})] \\ &= \alpha [1 - (1 - x_j)(1 - F(\pi(1 - x_j)))] \end{aligned} \quad (3)$$

For shorthand, we write

$$\dot{x}_i = \Phi(1 - x_j) \quad (4)$$

where  $\Phi(y) := \alpha(1 - y(1 - F(\pi(y))))$ , and  $y$  measures the probability that  $j$  has not adopted. Equation (4) plays a central role throughout the paper.

Figure 3 illustrates the equilibrium from Iris's perspective. When Iris sees John adopt, she inspects with certainty; the probability of this event rises linearly over time. When Iris sees John fail to adopt, she becomes pessimistic; initially, this event carries little information but it becomes worse news over time, lowering her adoption probability. On net, the former effect dominates: Observing John raises Iris's inspection and adoption probability at all times.  $\triangle$

**Example 2 (Chain).** Suppose there is an infinite chain of agents, so Kata observes Lili, who observes Moritz, and so on ad infinitum. All agents are in identical positions, so it is natural to consider equilibria where they have the same adoption curve  $x$  (indeed, this is the unique equilibrium). Analogous to equation (4), adoption is

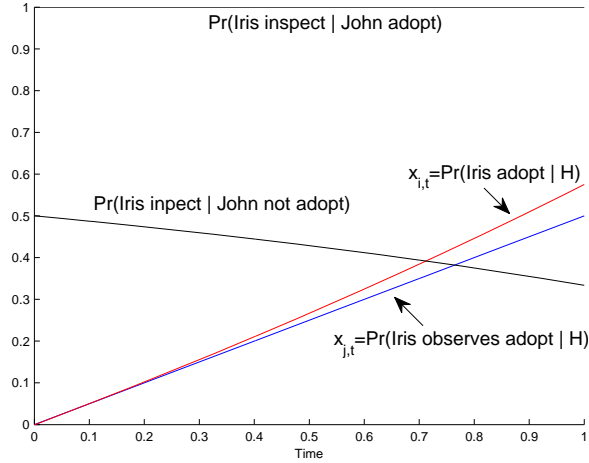


Figure 3: **Iris’s Social Learning and Adoption Curves from Example 1.** This figure assumes inspection costs are  $\kappa \sim U[0, 1]$ , high-quality goods are adopted with probability  $\alpha = 1$ , and the prior is  $\pi_0 = 1/2$ .

governed by the ODE

$$\dot{x} = \Phi(1 - x). \quad (5)$$

This captures the idea that Kata’s decision takes into account Lili’s decision, which takes into account Moritz’s decision, and so on. The simplicity of the adoption curve is in stark contrast to the cyclical behavior seen in traditional herding models when agents observe only the previous agent (Celen and Kariv, 2004). As we will see, the adoption curve is convex, lying above  $x_i$  in Figure 3, meaning an agent’s adoption increases when her neighbor obtains more information. We generalize this observation in Theorem 1.  $\triangle$

### 3 General Networks

In this section, we extend our analysis to arbitrary random networks  $\mathcal{G} = (I, \Xi, \mu)$ . Section 3.1 studies the connection between an agent’s adoption probability and her social learning curve, and shows that adoption rises with social information under a “bounded hazard rate” assumption. Section 3.2 closes the model and establishes that it admits a unique equilibrium.

### 3.1 Social Learning and Adoption Curves

We start with some definitions. As in the examples of Section 2.1, we generally denote agent  $i$ 's probability of adopting the high-quality product by  $x_i$ . Agent  $i$  may not know the network, and so we need to keep track of agents' adoption across realizations of  $G$  and signals  $\xi$ . Let  $x_{i,G,\xi}$  be agent  $i$ 's *realized adoption curve*, given  $(G, \xi)$  after taking expectations over entry times  $(t_i)$ , cost draws  $(\kappa_i)$ , and idiosyncratic tastes. Taking expectation over  $(G, \xi_{-i})$ , let  $x_{i,\xi_i} := \sum_{G, \xi_{-i}} \mu(G, \xi_{-i} | \xi_i) x_{i,G,\xi}$  be  $i$ 's *interim adoption curve* given her signal  $\xi_i$ . Throughout, we drop the time-subscript  $t$  for these and all other curves.

Bayesian agents form beliefs over their neighbors' adoption decisions. Since a single adoption by one of  $i$ 's neighbors perfectly reveals high quality, she only keeps track of this event. Specifically, let  $y_{i,G,\xi}$  be the probability that *none* of  $i$ 's neighbors adopt product  $H$  by time  $t \leq t_i$  in network  $G$  given signals  $\xi$ , and  $y_{i,\xi_i} := \sum_{G, \xi_{-i}} \mu(G, \xi_{-i} | \xi_i) y_{i,G,\xi}$  the expectation conditional on  $\xi_i$ .

To solve for  $i$ 's realized adoption curve  $x_{i,G,\xi}$ , suppose agent  $i$  enters at time  $t$ . If she sees one of her neighbors adopt, she always inspects. Conversely, if she sees no adoption, she inspects if her inspection cost is below a cutoff,  $\kappa_i \leq c_{i,\xi_i} := \pi(y_{i,\xi_i})$ . Analogous to equation (3), we get

$$\dot{x}_{i,G,\xi} = \alpha [1 - y_{i,G,\xi}(1 - F(\pi(y_{i,\xi_i})))] =: \phi(y_{i,G,\xi}, y_{i,\xi_i}). \quad (6)$$

Note that equation (6) depends on both the realized and the interim adoption-probability of  $i$ 's neighbors,  $y_{i,G,\xi}$  and  $y_{i,\xi_i}$ . The former determines whether  $i$  actually observes an adoption, given  $(G, \xi)$ ; the latter determines the probability that  $i$  inspects, which depends only on  $i$ 's coarser information  $\xi_i$ . Taking expectations over  $(G, \xi_{-i})$  given  $\xi_i$ , agent  $i$ 's interim adoption curve is then

$$\dot{x}_{i,\xi_i} = \alpha [1 - y_{i,\xi_i}(1 - F(\pi(y_{i,\xi_i})))] = \phi(y_{i,\xi_i}, y_{i,\xi_i}) = \Phi(y_{i,\xi_i}). \quad (7)$$

Equation (7) captures the positive implications of our theory for the diffusion of new products.

Our primary results concern normative implications, quantifying the value of social learning, as captured by  $i$ 's *social learning curve*  $1 - y_{i,\xi_i}$ . To see that this curve indeed

measures  $i$ 's learning, observe that the probability that  $i$  sees an adoption is given by,

|         |                   |               |
|---------|-------------------|---------------|
|         | $\geq 1$ adopt    | 0 adopt       |
| $q = H$ | $1 - y_{i,\xi_i}$ | $y_{i,\xi_i}$ |
| $q = L$ | 0                 | 1             |

If  $y_{i,\xi_i} = 0$  then agent  $i$  has perfect information about the state; otherwise, she has effectively lost the signal with probability  $y_{i,\xi_i}$ . It follows that a decrease in  $y_{i,\xi_i}$  Blackwell-improves agent  $i$ 's information and thereby increases her expected utility.

We thus refer to an increase in  $1 - y_{i,\xi_i}$  as a rise in agent  $i$ 's social information. Clearly,  $i$ 's social information improves over time: Since adoption is irreversible,  $y_{i,G,\xi}$  decreases in  $t$  for every  $G$ , and hence also in expectation. Much of our paper compares social learning curves across networks. For networks  $\tilde{\mathcal{G}}$  and  $\mathcal{G}$  (with overlapping agents  $i$  and types  $\xi_i$ ) we write  $1 - \tilde{y}_{i,\xi_i} \geq 1 - y_{i,\xi_i}$  if social information is greater in  $\tilde{\mathcal{G}}$  for all  $t$ , and  $1 - \tilde{y}_{i,\xi_i} > 1 - y_{i,\xi_i}$  if it is strictly greater for all  $t > 0$ .

Social learning and adoption are intimately linked by (7). One would think that as  $i$  collects more information then her adoption of product  $H$  increases. Indeed, with perfect information,  $y_{i,\xi_i} = 0$  she always inspects the product. More generally, monotonicity requires an assumption.

**Assumption:** The distribution of costs has a *bounded hazard rate* (BHR) if

$$\frac{f(\kappa)}{1 - F(\kappa)} \leq \frac{1}{\kappa(1 - \kappa)} \quad \text{for } \kappa \in [0, \pi_0]. \quad (8)$$

**Lemma 1.** *If  $F$  satisfies BHR then  $i$ 's interim adoption rate  $\dot{x}_{i,\xi_i}$  increases in her information  $1 - y_{i,\xi_i}$ . Thus, if  $i$ 's information  $1 - y_{i,\xi_i}$  increases, then so does her interim adoption probability  $x_{i,\xi_i}$ .*

*Proof.* Recalling (7), we differentiate

$$\begin{aligned} \Phi'(y) &= \partial_1 \phi(y, y) + \partial_2 \phi(y, y) = -\alpha(1 - F(\pi(y))) + \alpha y \cdot \pi'(y) \cdot f(\pi(y)) \\ &= -\alpha(1 - F(\pi(y))) + \alpha \pi(y) \cdot (1 - \pi(y)) \cdot f(\pi(y)), \end{aligned} \quad (9)$$

where the second equality uses Bayes' rule (2) to show,

$$y \cdot \pi'(y) = y \frac{\pi_0(1 - \pi_0)}{[1 - (1 - y)\pi_0]^2} = \frac{y\pi_0}{1 - (1 - y)\pi_0} \frac{1 - \pi_0}{1 - (1 - y)\pi_0} = \pi(y) \cdot (1 - \pi(y)).$$

Equation (9) captures two countervailing effects: The first term in (9) is negative because  $i$  is certain to inspect if she observes an adoption. The second term is positive because a rise in  $y$  makes  $i$  is less pessimistic when she sees no adoption. The aggregate effect is negative if and only if BHR holds. Thus if BHR holds for all  $\kappa \in [0, \pi_0]$  then  $\Phi'(y) \leq 0$  for all  $y \in [0, 1]$ , and better information  $1 - y_{i,\xi_i}$  means higher slope  $\dot{x}_{i,\xi_i}$  and level  $x_{i,\xi_i}$ .  $\square$

For an intuition, recall that adoption probabilities  $x$  are conditional on high quality  $q = H$ . The agent's expected posterior belief that quality is high, conditional on high quality, thus exceeds the prior  $\pi_0$  and rises with her information. This suggests that the agent's adoption should also increase in her information. However, her adoption decision also depends on the curvature of the cost distribution  $F$ . In particular, BHR requires that the increasing adoption from seeing others adopt outweighs the decreasing adoption from the event when she sees no adoption. Formally, BHR is satisfied if  $f$  is increasing on  $[0, 1]$ , which includes  $\kappa \sim U[0, 1]$  as a special case.<sup>10</sup>

**Example 3 (Information can lower adoption).** Assume the distribution  $F(\kappa)$  has support  $[0, \pi_0]$ ; this violates BHR since the denominator in the LHS of (8),  $1 - F(\pi_0)$ , vanishes. Without social information, agent  $i$  inspects with probability 1. With social information, agent  $i$  inspects with probability 1 if some neighbor adopts, and below 1 if no neighbor adopts, since  $\pi(y_{i,\xi_i}) < \pi_0$ . In expectation, social information lowers agent  $i$ 's inspection and adoption rate, contradicting Lemma 1.  $\triangle$

When we compare networks, we state the results in terms of agent  $i$ 's social learning curve and thus her welfare (e.g. Theorems 1-4). Lemma 1 means that, if we assume BHR, then we can also compare adoption rates across networks.

---

<sup>10</sup>An increasing density guarantees  $f(\kappa) \leq \frac{\int_{\kappa}^1 f(z)dz}{1-\kappa} \leq \frac{1-F(\kappa)}{\kappa(1-\kappa)}$ . For other densities  $f$ , BHR is automatically satisfied when  $\kappa \approx 0$  since the RHS increases to infinity. For higher costs, BHR states that the density does not decrease too quickly,  $d \log f(\kappa)/d\kappa \geq -2/\kappa$ . In particular, BHR holds with equality at all  $\kappa$  if  $f(\kappa) \propto 1/\kappa^2$ .

As a corollary, BHR implies that adoption curves are convex in time. Indeed, recall that the agent’s social information  $1 - y_{i,\xi_i}$  increases over time; thus, BHR implies that  $i$ ’s interim adoption rate  $\hat{x}_{i,\xi_i}$  rises over time, as shown in Figure 3. This result partly reflects our normalization that entry times are uniform on  $[0, 1]$ ; if instead, agents enter according to a Poisson process, adoption curves would look more like the familiar “S-shape” seen in diffusion models (e.g. Bass (1969)).

## 3.2 Equilibrium

We now turn to the problem of characterizing equilibrium by closing equation (7). Inspired by the examples in Section 2.1, one might hope to derive agent  $i$ ’s social learning curve from the adoption curves of her neighbors. Unfortunately, it is *not* generally the case that  $y_{i,G,\xi} = \prod_{j \in N_i(G)} (1 - x_{j,G,\xi})$ , even if the network is commonly known. There are two reasons.

The first reason is the *correlation* problem. To illustrate, suppose  $i$  observes both  $j$  and  $k$ ,  $j$  observes only  $k$ , while  $k$  has no information (see left panel of Figure 7). Agent  $k$ ’s adoption curve  $x_k$  follows (1), while  $x_j$  follows (4). Agent  $i$ ’s inspection decision in turn depends on the probability that neither  $j$  nor  $k$  adopts. Since  $j$  observes  $k$ , their adoption decisions are correlated and it is not enough to keep track of the marginal probabilities,  $x_j$  and  $x_k$ ; rather we must keep track of the joint distribution. We return to this issue in Section 4.2.

The second reason is the *self-reflection* problem. To illustrate, suppose  $i$  and  $j$  observe each other (see right panel of Figure 7). When making her inspection decision at time  $t_i$ , agent  $i$  must infer whether or not  $j$  has already inspected. However, since she just entered, agent  $i$  knows that agent  $j$  cannot have seen  $i$  adopt. Thus,  $i$  conditions  $j$ ’s adoption probability on the event that  $i$  has not yet adopted, which differs from  $j$ ’s actual (i.e. unconditional) adoption curve  $x_j$ . We return to this issue in Section 4.3.

Despite these problems, we can show:

**Proposition 1.** *In any random network  $\mathcal{G} = (I, \Xi, \mu)$ , there exists a unique equilibrium.*



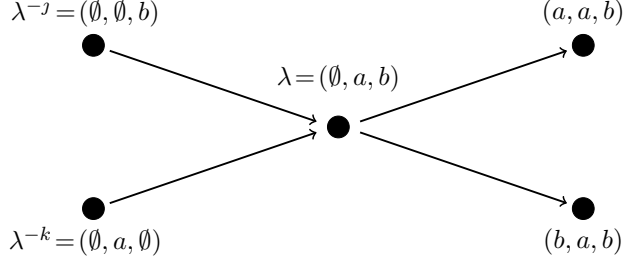


Figure 4: **Illustrative Markov Transitions underlying Proposition 1.**

*Proof.* We will characterize equilibrium adoption via a system of ODEs, albeit in a large state space. Denote the *state* of the network by  $\lambda = \{\lambda_i\}_{i \in I}$ , where  $\lambda_i \in \{\emptyset, a, b\}$ . Let  $\lambda_i = \emptyset$  if  $i$  has yet to enter,  $t \leq t_i$ ;  $\lambda_i = a$  if  $i$  has entered and adopted; and  $\lambda_i = b$  if  $i$  has entered and not adopted. Given state  $\lambda$ , let  $\lambda^{-i}$  denote the same state with  $\lambda_i = \emptyset$ .

Fix a network  $G$  and agents' signals  $\xi$ , and condition on a high-quality product,  $q = H$ . We can then describe the distribution over states by  $z_{\lambda, G, \xi}$  (as always omitting the dependence on time  $t$ ). Figure 4 illustrates the evolution of the state via a Markov chain in a three-agent example. Probability mass flows into state  $\lambda = (\lambda_i, \lambda_j, \lambda_k) = (\emptyset, a, b)$  from state  $\lambda^{-j}$  as agent  $j$  enters and adopts, and from  $\lambda^{-k}$  as agent  $k$  enters and doesn't adopt. Similarly, probability flows out of state  $\lambda$ , and into states  $(a, a, b)$  and  $(b, a, b)$ , as agent  $i$  enters.

Quantifying these transition rates, we now argue that the equilibrium distribution over the states  $\lambda$  evolves according to the following ODE

$$\begin{aligned}
(1-t)\dot{z}_{\lambda, G, \xi} = & - \sum_{i: \lambda_i = \emptyset} z_{\lambda, G, \xi} + \sum_{i: \lambda_i = a, \exists j \in N_i(G): \lambda_j = a} z_{\lambda^{-i}, G, \xi} \alpha + \sum_{i: \lambda_i = a, \forall j \in N_i(G): \lambda_j \neq a} z_{\lambda^{-i}, G, \xi} \alpha F(\pi(y_{i, \xi_i})) \\
& + \sum_{i: \lambda_i = b, \exists j \in N_i(G): \lambda_j = a} z_{\lambda^{-i}, G, \xi} (1 - \alpha) + \sum_{i: \lambda_i = b, \forall j \in N_i(G): \lambda_j \neq a} z_{\lambda^{-i}, G, \xi} (1 - \alpha F(\pi(y_{i, \xi_i}))).
\end{aligned} \tag{10}$$

To understand (10), first observe that the probability that  $i$  observes no adoption at time  $t$ , given  $G, \xi$  and conditional on  $t_i = t$ , is given by  $y_{i, G, \xi} = \frac{1}{1-t} \sum z_{\lambda, G, \xi}$  where the sum is over all  $\lambda$  with  $\lambda_i = \emptyset$  and  $\lambda_j \neq a$  for all  $j \in N_i(G)$ .

Next fix a state  $\lambda$ . Agents  $i$  that have not yet entered,  $\lambda_i = \emptyset$ , enter uniformly

over time  $[t, 1]$ , and so probability escapes at rate  $z_{\lambda, G, \xi}/(1 - t)$  for each such agent. This out-flow is the first term in (10). Turning to in-flows, if  $\lambda_i = a$  then in state  $\lambda^{-i}$  agent  $i$  enters uniformly over time  $[t, 1]$  and adopts with probability  $\alpha$  if one of her neighbors  $j \in N_i(G)$  has adopted (the second term in (10)), and with probability  $\alpha F(\pi(y_{i, \xi_i}))$  if none of her neighbors has adopted (the third term in (10)). If  $\lambda_i = b$ , inflows from  $\lambda^{-i}$  are similarly captured by the fourth and fifth term in (10).

The existence of a unique equilibrium now follows from the Picard-Lindelöf theorem since the boundedness of  $f$  implies the system is Lipschitz.<sup>11</sup>  $\square$

The system of ODEs (10) implies equilibrium existence and uniqueness. But it is less useful as a tool to compute equilibrium numerically since there are  $3^I \times 2^{I \times I} \times |\Xi|$  triples  $(\lambda, G, \xi)$ , making it impossible to evaluate the ODE numerically. Even if the network  $G$  is common knowledge, there are still  $3^I$  states  $\lambda$ . For this reason we impose more structure on networks in the following sections, where we can provide simple formulas for diffusion curves.

## 4 Elemental Deterministic Networks

In this section we consider simple deterministic networks including trees, undirected pairs, clusters and stars. Collectively, these provide the building blocks for richer, random networks in Section 5. The equilibrium characterizations (in both sections) allow us to study how social information is shaped by the network. In particular, we will see that in a tree all links are beneficial but, beyond trees, backward and correlating links harms agent  $i$ 's learning. Since the network  $G$  is fixed we omit it from the notation, e.g. writing  $i$ 's neighbors as  $N_i$  instead of  $N_i(G)$ .

### 4.1 Trees

Network  $G$  is a *tree* if for any pair of agents  $i, j$  there is at most one path  $i \rightarrow \dots \rightarrow j$ . Trees avoid the correlation and self-reflection problems. Such networks are realistic in some applications such as hierarchical organizations where the information flow is

---

<sup>11</sup>In fact, if we assume that time  $t$  is discrete rather than continuous, this proof shows more strongly that our game - much like most herding models - is dominance-solvable.

uni-directional. As we will see in Section 5, they are also a good approximation of large random networks. In this section, we characterize adoption decisions, show that agents benefit from both direct and indirect links, and compare the value of direct and indirect links.

Since the adoption decisions of  $i$ 's neighbors in a tree are mutually independent, and independent of her entry time  $t_i$ , her social learning curve depends only on the individual adoption probabilities of her neighbors,  $y_i = \prod_{j \in N_i} (1 - x_j)$ . Agent  $i$ 's adoption curve (7) becomes

$$\dot{x}_i = \Phi \left( \prod_{j \in N_i} (1 - x_j) \right). \quad (11)$$

Equilibrium is thus characterized by this  $I$ -dimensional ODE. One can further reduce the state space by grouping all nodes  $i$  with same “descendant trees” (e.g. all leaves) into the same “type”. In the extreme, we say a tree  $G$  is *regular with degree  $d$*  if every agent has  $d$  links.<sup>12</sup> Agents are symmetric, so equilibrium adoption is the same for all agents, and we write it as  $x$ . The probability no neighbor adopts is  $(1 - x)^d$ , so agent  $i$ 's adoption curve is given by a one-dimensional ODE,

$$\dot{x} = \Phi((1 - x)^d). \quad (12)$$

This generalizes equation (5) by allowing for more than one neighbor per agent.

We now argue that, within trees, both direct and indirect links are beneficial. While not surprising, such simple comparative static results are beyond the scope of traditional herding models. Figure 5 illustrates the social learning curves as we add links to the network. The left panel compares a lone agent (John in Example 1), an agent with one link (Iris in Example 1) and an infinite chain (Kata, Lili, and Moritz in Example 2). The social learning curves shift up as neighbors add more links, and so the Blackwell-ranking implies that Kata is better off than Iris, who is better off than John. The right panel shows the social learning curves in regular networks with

---

<sup>12</sup>Such a system has  $|I| = \infty$  agents, so Proposition 1 does not apply as stated. But, Proposition 2 shows that equilibrium adoption probabilities in large random networks converge to the solution of (11). Alternatively, joint adoption probabilities  $(x_i) \in [0, 1]^\infty$  equipped with the sup-norm define a Banach-space, so an infinite-dimensional version Picard-Lindelöf theorem (e.g. Deimling, 1977, Section 1.1) implies that the equilibrium described by (12) is unique.

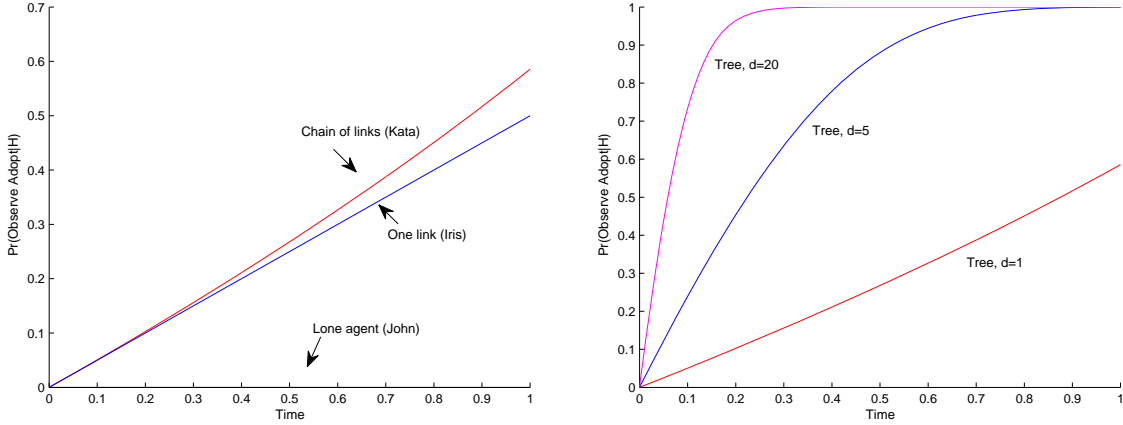


Figure 5: **Social Learning Curves in Tree Networks.** The **left** panel illustrates Examples 1 and 2. The **right** panel shows regular trees with degree  $d$ . This figure assumes  $\kappa \sim U[0, 1]$ ,  $\alpha = 1$ , and  $\pi_0 = 1/2$ .

$d = 1$  (i.e. an infinite chain),  $d = 5$  and  $d = 20$ . Again these social learning curves shift up, so agents benefit from making the tree denser.

Write the adoption curve of agent  $i$  in tree networks  $G, \tilde{G}$  as  $x_i, \tilde{x}_i$ , and the social learning curves as  $y_i, \tilde{y}_i$ . We say that  $G$  is a subtree of  $\tilde{G}$  and write  $G \subseteq \tilde{G}$  if  $\tilde{G}$  includes all agents and links of  $G$ .

**Theorem 1.** *Consider trees  $G \subseteq \tilde{G}$  and assume BHR holds. For any agent  $i$  in  $G$ , social learning is superior in the larger tree,*

$$1 - y_i \leq 1 - \tilde{y}_i \quad \text{for all } t. \quad (13)$$

*Proof.* First consider the leaves of  $G$ , who have no information in the small tree. That is, for an agent  $i$  who is a leaf in  $G$ ,  $y_i = 1 \geq \tilde{y}_i$ . Continuing by induction down the tree, consider some agent  $i$  with neighbors  $N_i$  in the small tree and  $\tilde{N}_i$  in the large tree, and assume that (13) holds for all  $j \in N_i$ . By Lemma 1, such neighbors adopt more in the larger tree,  $x_j \leq \tilde{x}_j$ . Additionally, agent  $i$  has more neighbors in the large tree. Hence

$$y_i = \prod_{j \in N_i} (1 - x_j) \geq \prod_{j \in \tilde{N}_i} (1 - \tilde{x}_j) = \tilde{y}_i,$$

as required.<sup>13</sup> □

<sup>13</sup>Without the BHR assumption, the result can break down. In Example 3, an agent observing

Theorem 1 is proved by induction and hence applies to any finite tree. But the analogous result holds for (infinite) regular trees, as illustrated in Figure 5. That is, an increase in  $d$  raises the probability one neighbor adopts and, using BHR, raises the RHS of the law-of-motion (12). Thus, the adoption  $x_t$  also increases for all  $t$ .

Theorem 1 has practical implications. The fact that denser networks raise social learning and adoption confirms the intuition that social information is more valuable for “visible” products (e.g. laptops) that are represented by a dense network, than for “hidden” products (e.g. PCs). The proof of Theorem 1 can be adapted to show that a rise in the adoption probability  $\alpha$  also raises social learning.<sup>14</sup> Thus, social information is more important for “broad” products that most people will buy if high quality (e.g. Apple Watch) than for “niche” products that appeal to more idiosyncratic tastes (e.g. Google Glass).

Theorem 1 is silent about the quantitative impact of direct and indirect links. The next example emphasizes the importance of direct links.

**Example 4 (Two Links vs Infinite Chain).** Compare an agent with two uninformed neighbors, and an agent in an infinite chain where everyone has one neighbor, as shown in in Figure 6. When agent  $i$  has two uninformed neighbors  $j, k$ , each neighbor’s adoption curve equals  $x_j = x_k = \alpha F(\pi_0)t$ . Hence the probability that neither adopts is

$$y_i = (1 - \alpha F(\pi_0)t)^2. \tag{14}$$

With an infinite chain, agent  $i$ ’s social learning curve is given by  $\dot{y} = -\Phi(y) \geq -\alpha[1 - y(1 - F(\pi_0))]$ . Solving this ODE,

$$y \geq \frac{1 - F(\pi_0)e^{\alpha(1-F(\pi_0))t}}{1 - F(\pi_0)}. \tag{15}$$

In Appendix A.1, we show that the RHS of (15) exceeds (14), and hence  $i$  gains more social information from the pair. Intuitively, if  $i \rightarrow j \rightarrow k$ , then agent  $k$  only affects  $i$ ’s action if  $k$  enters first, then  $j$  enters, and then  $i$  enters. Thus, the chance of learning information from the  $n^{\text{th}}$  removed neighbor in the chain is  $\frac{1}{n!}$ , meaning

---

agent  $i$  would prefer that  $i$  not have links.

<sup>14</sup>Indeed, a rise in  $\alpha$  scales up adoption in the leaf  $j$  of  $G$ , which raises the social learning curves of agents  $i$  who observe  $j$ , which raises agent  $i$ ’s adoption (by BHR), and so on down the tree.

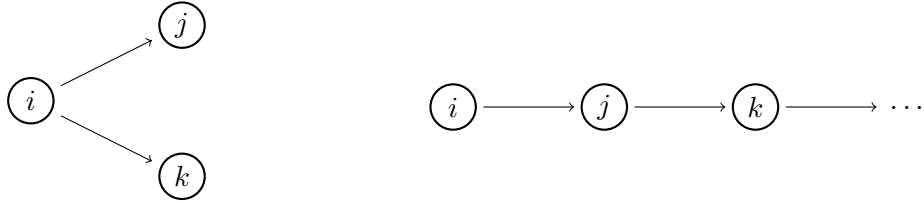


Figure 6: **Direct vs Indirect Links.** This figure illustrates the two networks compared in Example 4.

that an infinite chain of signals is worth less than two direct signals. Moreover, these indirect signals are intermediated (i.e.  $k$ 's signal must pass through  $j$ ) which reduces their information value.  $\triangle$

## 4.2 Correlating Links

Within the class of tree networks, agents benefit from all direct and indirect links. In contrast, here we show by example that agents can be made worse off by indirect, *correlating* links. This analysis also illustrates that correlation requires us to study joint adoption probabilities, rather than just marginals.

Assume agent  $i$  initially observes two uninformed agents  $j$  and  $k$ , as in the left panel of Figure 7. As in Example 4, the probability that neither adopts is  $(1 - \alpha F(\pi_0)t)^2$ . Now, suppose we add a link from  $j$  to  $k$ , correlating their adoption outcomes. Agent  $k$ 's behavior is unchanged, but the probability that agent  $i$  sees no adoption increases. This is because the probability  $x_{j|-k}$  that  $j$  adopts conditional on  $k$  not adopting follows

$$\hat{x}_{j|-k} = \alpha F(\pi(1 - x_k)) < \alpha F(\pi_0).$$

Intuitively, agent  $i$  just needs one of her neighbors to adopt. Adding the link  $j \rightarrow k$  makes  $j$  more pessimistic and lowers his adoption probability exactly in the event when his adoption would be informative for  $i$ , namely when  $k$  has not adopted. Thus, the correlating link makes agent  $i$  worse off. Of course, the link also makes  $j$  better



Figure 7: **Networks from Sections 4.2 and 4.3.** The **left** panel adds a correlating link. The **right** panel adds a backward link.

off. We return to this issue and discuss the overall welfare effects in Section 5.3.<sup>15</sup>

### 4.3 Undirected Pairs

We now study the simple undirected pair by adding a backward link,  $j \rightarrow i$ , to the directed pair in Example 1. This is illustrated in the right panel of Figure 7. We first derive the agents' adoption probabilities, and then show that the backward link harms agent  $i$ .

Define  $\bar{x}_j$  to be  $i$ 's expectation of  $j$ 's adoption curve at  $t \leq t_i$ . To solve for  $\bar{x}_j$  we must distinguish between two probability assessments of the event that  $j$  does not observe  $i$  adopt. From  $i$ 's "objective" perspective, this probability equals one since  $i$  knows she has not entered at  $t \leq t_i$ . From  $j$ 's "subjective" perspective, the probability equals  $1 - \bar{x}_i$ , since  $j$  thinks he is learning from agent  $i$  given  $t \leq t_j$ . This objective and subjective probabilities correspond to the "realized" and "interim" probabilities in equation (6). Thus,

$$\dot{\bar{x}}_j = \phi(1, 1 - \bar{x}_i) = \alpha F(\pi(1 - \bar{x}_i)). \quad (16)$$

By symmetry,  $\bar{x}_i = \bar{x}_j =: \bar{x}$ . Thus, both agents' social information satisfies  $y = 1 - \bar{x}$ , and their actual (unconditional) adoption probability follows  $\dot{x} = \Phi(1 - \bar{x})$ .

We can now study the effect of the backward link  $j \rightarrow i$  on  $i$ 's social information. Equation (16) implies that  $\dot{\bar{x}}_j \leq \alpha F(\pi_0)$ , and so  $\bar{x}_j \leq \alpha F(\pi_0)t$ , which is  $j$ 's adoption

<sup>15</sup>This argument leverages our assumption that agents  $j, k$  never adopt low-quality products, and hence  $i$  only cares about whether at least one of them has adopted the product. In Appendix C we reconsider this example in the context of a more general imperfect learning model where  $j, k$  sometimes adopt low-quality products, and where  $i$  benefits from seeing both  $j$  and  $k$  adopt. We show by example that the correlating link  $j \rightarrow k$  may then improve  $i$ 's social information.

curve if he does not observe  $i$ . Thus, the link  $j \rightarrow i$  lowers  $i$ 's social information and her utility. Intuitively, when agent  $i$  enters the market at  $t_i$ , she knows that  $j$  cannot have seen her adopt; however,  $j$  does not know the reason for  $i$ 's failure to adopt. This makes  $j$  more pessimistic, reduces his adoption probability, and lowers  $i$ 's social learning curve. Of course, the backward link also makes  $j$  better off. We return to this issue and discuss the overall welfare effects in Section 5.2.

## 4.4 Clusters

Our final building block is the complete network of  $I + 1$  agents. When  $I = 1$ , this is equivalent to the undirected pair. With more agents, agent  $j$ 's adoption before  $i$  enters depends on agent  $k$ 's adoption before both  $i$  and  $j$  enter. One might worry about higher order beliefs as  $I$  gets large. Fortunately, we can side-step this complication by thinking about the game from the first mover's perspective, before anyone else has adopted.

To be specific, let the *first-adopter probability*  $\hat{x}$  be the probability that an agent adopts given that no one else has already adopted. Since everyone is symmetric, intuition suggests that the first adopter attaches subjective probability  $(1 - \hat{x})^I$  to the event that none of the other potential first-adopters has adopted. A first adopter then inspects with probability  $F(\pi((1 - \hat{x})^I))$ , generalizing equation (16). But objectively, the first adopter is certain to observe no adoption, and so we define  $\hat{x}$  to be the solution of

$$\hat{x} = \phi(1, (1 - \hat{x})^I) = \alpha F(\pi((1 - \hat{x})^I)) \quad (17)$$

and show

**Lemma 2.** *In the complete network with  $I + 1$  agents, any agent's social learning curve is  $1 - \hat{y} = 1 - (1 - \hat{x})^I$ ; actual (unconditional) adoption probabilities follow  $\hat{x} = \Phi((1 - \hat{x})^I)$ .*

*Proof.* See Appendix A.2. □

Our characterization of the cluster allows us to answer a classic question: Do agents prefer to be in “centralized” networks or “decentralized” networks? Granovet-



ter’s (1973) “strength of weak ties” hypothesis argues that social behavior spreads more quickly in loosely connected networks (as in a big city), whereas Centola’s (2010) experiment suggests that clusters may be important for learning and diffusion (as in a tight-knit village).

To address this question, compare a cluster with  $d + 1$  agents to a regular directed tree of degree  $d$ . In both these cases each agent has  $d$  neighbors, yet the former network is more “centralized”. Fix any agent  $i$ . In the cluster, her social information is  $1 - (1 - \hat{x})^d$ , where  $\hat{x}$  is given by (17). In the tree, her social information is  $1 - (1 - x)^d$ , where  $x$  is given by (12). Since  $\phi$  is decreasing in its first argument, the Single-Crossing Lemma (see Appendix A.3) implies that  $x \geq \hat{x}$ , meaning the agents’ have higher social information and higher utility in the “decentralized” network than the “centralized” network.

## 4.5 Stars

So far we have seen that an agent benefits from independent direct and indirect links (Section 4.1), prefers direct to indirect links (Section 4.1), but is harmed by correlating links (Section 4.2) and self-reflecting links (Section 4.3). This suggests that agent  $i$ ’s optimal network is the  $i$ -star, in which agent  $i$  observes all other agents, and other agents observe nobody.<sup>16</sup>

We conclude our analysis of deterministic networks by arguing that, indeed, among all networks with  $|I|$  agents the  $i$ -star maximizes agent  $i$ ’s social learning (i.e. minimizes  $y_i$ ) and thus maximizes her utility. To see this, first consider the  $i$ -star, and suppose  $i$  sees no adoption of the high-quality good. It must be the case that any agent  $j$  who enters before  $i$  and has favorable idiosyncratic preferences, chooses not to inspect,  $\kappa_j > \pi_0$ . To complete the argument, we now argue that for the same realizations of costs, entry times, and idiosyncratic preferences, agent  $i$  observes no adoption in any other network  $G$ . To do so, we consider the  $L$  agents who move before  $i$  and have favorable idiosyncratic preferences, relabel them by their entry times  $t_1 < t_2 < \dots < t_L$ , and argue by induction over  $\ell \in \{1 \dots L\}$ . Agent  $\ell = 1$  moves first and thus sees no adoption in network  $G$ ; since  $\kappa_\ell > \pi_0$  she chooses not to inspect, and

---

<sup>16</sup>Relatedly, in the classic herding model, SgROI (2002) asks what is the optimal number of guinea pigs who move before other agents.

thus does not adopt. Continuing by induction, agent  $\ell + 1$  also sees no adoption in  $G$ ; the lack of adoption is bad news,  $\pi(y_\ell) \leq \pi_0 < \kappa_\ell$ , so she also does not inspect or adopt. Thus  $i$ 's social learning curve is higher in the  $i$ -star than in any other network.

## 5 Large Random Networks

Our analysis of deterministic networks demonstrates the major economic forces, but quickly becomes intractable. To move beyond the examples of Section 4 and characterize adoption in a rich class of networks that may resemble reality, we return to the general random network model of Sections 2-3 and specialize them by encoding local network information into agents' types  $\xi_i$ . This allows us to expand our welfare analysis, and develop a tractable model of diffusion that can be used for empirical applications.

Formally, we introduce two configuration models, one for directed networks with multiple types (Section 5.1) and one for undirected networks with clusters (Sections 5.2-5.4), and study their limit equilibria as the networks grow large. This suggests a “macroeconomic” approach to studying diffusion empirically: First, calibrate the random network to the real-life network by matching the pertinent network parameters (agents' types, degree distributions, cluster coefficients). Second, solve for equilibrium behavior on this approximate network. We round the paper off by examining the model's implications for correlation neglect (Section 5.5) and information aggregation (Section 5.6).

### 5.1 Directed Networks with Multiple Types

We first consider diffusion in large directed networks with different types of agents. For example, think of Twitter users as celebrities, posters and lurkers: Celebrities only follow celebrities, while posters and lurkers follow posters and celebrities. Agents know who they follow and know the distribution over networks, but do not know exactly who others follow.

To formalize this idea, we generate the random network  $\mathcal{G}_I$  via the configuration model (e.g. Jackson (2010), Sadler (2020)). For any agent  $i$ , independently draw a

finite type  $\theta \in \Theta$  according to some distribution with full support. For any agent with type  $\theta$ , independently draw a vector of labeled outlinks  $d = (d_{\theta'})_{\theta'} \in \mathbb{N}^{\Theta}$ ; these are realizations of a random vector  $D_{\theta} = (D_{\theta,\theta'})_{\theta'}$  with finite expectations  $E[D_{\theta,\theta'}]$ . We call  $D = (D_{\theta,\theta'})_{\theta,\theta'}$  the *degree distribution*. To generate the network, connect type- $\theta'$  outlinks to type- $\theta'$  agents independently across outlinks. Finally, prune self-links from  $i$  to  $i$ , multi-links from  $i$  to  $j$ , and – in the unlikely case that no type  $\theta'$  agent was realized – all of the unconnectable type- $\theta'$  outlinks. Agent  $i$ 's signal  $\xi_i$  consists of her degree  $d \in \mathbb{N}^{\Theta}$  after the pruning.<sup>17</sup>

Since  $\mathcal{G}_I$  is symmetric across agents, we drop their identities  $i$  from the notation of Section 3, and write adoption probabilities, learning curves, and cost thresholds of a degree- $d$  agent as  $x_d$ ,  $1 - y_d$ , and  $c_d = \pi(y_d)$ . Taking expectation over the degree of a type- $\theta$  agent, we write  $x_{\theta} = E[x_{D_{\theta}}]$ .

In principle, one can use Theorem 1 to characterize equilibrium in this network by a  $3^I \times 2^{I \times I} \times |\Theta|$  dimensional ODE. But that characterization is complicated because it keeps track of all possible network realizations, and the joint distribution of adoption decisions. Fortunately things are much simpler here since, for  $I$  large, the random network locally resembles a tree where every type  $\theta$  agent has  $D_{\theta,\theta'}$  links to independent type- $\theta'$  agents. So motivated, in analogy to the equilibrium adoption probabilities in finite trees (11), define  $(x_{\theta}^*)$  to be the solution of

$$\dot{x}_{\theta} = E \left[ \Phi \left( \prod_{\theta'} (1 - x_{\theta'})^{D_{\theta,\theta'}} \right) \right]. \quad (18)$$

This is a  $\Theta$ -dimensional ODE, which is easy to compute. For example, in the Twitter example, we get one ODE each for celebrities, posters and lurkers, even though the number of possible degrees  $d$  for each type of agent is infinite. We write

$$c_d^* := \pi \left( \prod_{\theta'} (1 - x_{\theta'}^*)^{d_{\theta'}} \right) \quad (19)$$

---

<sup>17</sup>This definition differs from the literature on configuration models, e.g. Sadler (2020), in at least three ways. (a) Sadler considers undirected networks, to which we turn in Section 5.2. (b) We model agent  $i$ 's degree as a random variable  $D_{\theta}$ , while Sadler fixes the realized degrees  $d$  and imposes conditions on the empirical distribution of degrees as  $I$  grows large. (c) When a realized network  $G$  has self-links or multi-links, we prune these links from  $G$ , while Sadler discards  $G$  by conditioning the random network on realizing no such links. We view (b) and (c) as technicalities, and deviate from the literature because doing so simplifies our analysis.

for the cutoff-costs associated with  $(x_\theta^*)$ .

We now show that this simple,  $\Theta$ -dimensional ODE is a reasonable description of adoption behavior for large  $I$ . Formally, say that a vector of cutoff costs  $(c_d)$  is a *limit equilibrium* of the large directed random network with degree distribution  $D$  if there exist  $(\epsilon_I)$  with  $\lim_{I \rightarrow \infty} \epsilon_I = 0$  such that  $(c_d)$  is a  $\epsilon_I$ -equilibrium in  $\mathcal{G}_I$ .

**Proposition 2.**  $(c_d^*)$  is the unique limit equilibrium of the large directed random network with degree distribution  $D$ .

*Proof.* The key idea is that social learning in our model only depends on local properties of the network, and our large random networks locally approximate trees. We formalize this intuition in Appendix B.1.  $\square$

The notion of a limit equilibrium extends our “macroeconomic perspective” from the modeler to the agents. While the real network is finite, agents treat it as infinite; in large networks, the resulting behavior is approximately optimal. For completeness, Appendix B.2 provides a “microeconomic perspective” by showing that the equilibria of the finite models  $\mathcal{G}_I$  indeed converge to  $(c_d^*)$ .

We now leverage Proposition 2 to double-down on our finding that both direct and indirect links improve agents’ social learning (Theorem 1). Thus lurkers are better off if celebrities or posters increase their number of links. Formally, consider two degree distributions such that  $\tilde{D} \succeq_{FOSD} D$  in the usual multi-variate stochastic order (see Shaked and Shanthikumar, 2007, Chapter 6). Let  $x_\theta^*, \tilde{x}_\theta^*$  be the associated adoption probabilities derived from (18), and  $y_d^*$  and  $\tilde{y}_d^*$  be the corresponding social learning curves.

**Theorem 1’.** *Assume BHR. In large random networks, social learning and welfare improves with links: If  $\tilde{D} \succeq_{FOSD} D$ ,*

- (a) *For any degree  $d$ ,  $1 - \tilde{y}_d^* \geq 1 - y_d^*$ ,*
- (b) *For any type  $\theta$ ,  $E[1 - \tilde{y}_{D_\theta}^*] \geq E[1 - y_{D_\theta}^*]$ .*

*Proof.* BHR means that  $\Phi' \leq 0$ . Thus, the RHS of (18) FOSD-rises in both the (exogenous) degree  $D$  and the (endogenous) adoption probabilities  $x_\theta$ . Thus, the

solution  $(x_\theta^*)$  rises in  $D$ ,  $\tilde{x}_\theta^* \geq x_\theta^*$ , and so  $1 - \tilde{y}_d^* \geq 1 - y_d^*$ . Taking expectations over  $d$  then yields part (b).<sup>18</sup>  $\square$

Part (a) says that social information rises if we fix the degree  $d$ , and thus only speaks to the value of additional indirect links. Obviously the additional direct links also help, as shown in part (b).

## 5.2 Undirected Networks

We now consider undirected random networks, representing friends on Facebook, or drivers who learn about cars from observing other drivers' choices. To formalize this, we use a different variant of the configuration model. For simplicity we here abstract from types, but re-introduce them in Appendix B.3. We illustrate a single-type (multi-type) network in the left (right) panel of Figure 2. Each agent independently draws  $d \in \mathbb{N}$  *link-stubs* generated by a random variable  $D$ . We then independently connect these stubs in pairs, and prune self-links, multi-links, and residual unconnected stubs (if the total number of stubs is odd).

An important feature of random undirected networks is the *friendship paradox*. Namely,  $i$ 's neighbors typically have more neighbors than  $i$  herself. Formally, we define the neighbor's degree distribution  $D'$  by

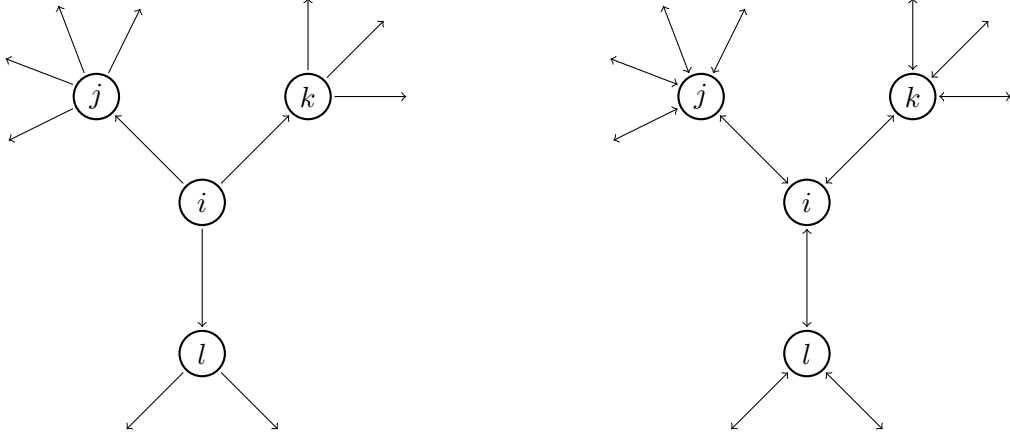
$$\Pr(D' = d) := \frac{d}{E[D]} \Pr(D = d). \quad (20)$$

For example, in an Erdős-Rényi network,  $D$  is Poisson and  $D' = D + 1$ , whereas in a regular deterministic network  $D' = D \equiv d$ .

We now study the behavior of the limit economy as the number of agents grows large. This allows us to treat neighbors' adoption probabilities as independent; Proposition 2' in Section 5.4 justifies this approach. The simplest such network, corresponding to  $D \equiv 1$ , is the dyad from Section 4.3. There we argued that, due to the backward link, agent  $i$  knows that  $j$  cannot have seen her adopt before  $t_i$  and wrote  $\bar{x}$  for  $i$ 's learning curve conditional on this event. Since  $j$  expects to see no adoption with

---

<sup>18</sup>If we additionally assume that the inequality  $\tilde{D} \succeq_{FOSD} D$  is strict and that  $\tilde{D}$  is irreducible, we more strongly get the strict inequality  $1 - \tilde{y}_d^* > 1 - y_d^*$ .



**Figure 8: Comparing Directed and Undirected Networks.** This figure illustrates two Erdős-Rényi networks, from agent  $i$ 's perspective, showing  $i$ 's neighbors and their outlinks. For simplicity, the directed picture does not show inlinks to  $i$  or her neighbors (in a large network, these inlinks do not affect  $i$ 's learning). Observe that  $i$ 's neighbors have one more outlink in the undirected network, namely the link to  $i$ ; this reflects the friendship paradox.

subjective probability  $1 - \bar{x}$  but objectively never sees an adoption at  $t \leq t_i$  we concluded that  $\dot{\bar{x}} = \phi(1, 1 - \bar{x})$ . In our random network,  $i$ 's neighbor  $j$  additionally learns from another  $D' - 1$  independent links, from which he observes no adoption with probability  $(1 - \bar{x})^{2(D'-1)}$ . All told, define  $\bar{x}^*$  as the solution of

$$\dot{\bar{x}} = E \left[ \phi((1 - \bar{x})^{D'-1}, (1 - \bar{x})^{D'}) \right]. \quad (21)$$

Agent  $i$ 's actual, unconditional adoption rate then equals  $E[\Phi(1 - \bar{x}^*)^D]$ . Equation (21) implies that, as in Theorems 1 and 1', social learning increases with the number of links  $D$  (see Appendix B.4).

We now evaluate the impact of backward links. In Section 4.3 we showed that adding a  $j \rightarrow i$  link to the dyad  $i \rightarrow j$  makes  $i$  worse off and  $j$  better off. Here we evaluate the overall effect by comparing a network where agents have  $D$  directed links to one with  $D$  undirected links, as illustrated in Figure 8. Recalling neighbors' limit adoption probabilities in directed and undirected networks  $x^*, \bar{x}^*$  from equations (18) and (21), we write  $y_d^* = (1 - x^*)^d$  and  $\bar{y}_d = (1 - \bar{x}^*)^d$  for the respective social learning curves.

**Theorem 2.** *Assume  $D$  is Poisson or deterministic. Social learning in large random*

networks is higher when the network is directed rather than undirected: For any degree  $d$ ,  $1 - y_d^* > 1 - \bar{y}_d^*$ .

*Proof.* Rewriting (18) for a single-type, and taking expectations over  $D$ , the adoption probability of any given neighbor in the directed network follows

$$\dot{x} = E\left[\phi\left((1-x)^{D'}, (1-x)^D\right)\right]. \quad (22)$$

In the undirected network, the conditional adoption probability of any given neighbor follows equation (21). For Poisson links  $D' = D + 1$ , and so the RHS of (22) exceeds the RHS of (21) when  $x = \bar{x}$ , because  $\phi$  rises in its second argument. For deterministic links  $D' = D$ , and so the RHS of (22) exceeds the RHS of (21) when  $x = \bar{x}$ , because  $\phi$  falls in its first argument. In either case, the Single-Crossing Lemma implies that the solution of (22) exceeds the solution of (21),  $x^* > \bar{x}^*$ , and so  $1 - y_d^* > 1 - \bar{y}_d^*$ .  $\square$

Theorem 2 says that, fixing the degree distribution, directed networks generate better information than undirected networks. Intuitively, in an undirected network an agent's neighbors cannot have seen her adopt when she enters; this makes them more pessimistic and reduces social learning. The result assumes that  $D$  is Poisson or deterministic, which means  $i$ 's neighbors have, on average, at most one neighbor more than her, meaning we are comparing "like with like".<sup>19</sup>

When  $D$  is deterministic, Theorem 2 tells us that backward links are bad when we fix the total number of links. But what if we increase the number of links?

**Example 5 (Directed vs Undirected Chain).** With the infinite directed line ( $i \rightarrow j \rightarrow k \rightarrow \dots$ ), Example 2 shows that the social learning curve is given by the solution  $y^d$  of

$$\dot{y} = -\alpha [1 - y[1 - F(\pi(y))]] = -\Phi(y).$$

With the infinite undirected line ( $\dots \leftrightarrow i \leftrightarrow j \leftrightarrow k \leftrightarrow \dots$ ) the adoption of each of  $i$ 's neighbors,  $\bar{x}$ , is given by (21) for  $D' = D \equiv 2$ , and the probability that neither of

---

<sup>19</sup>To see how the friendship paradox can overturn the result, suppose that agents are equally likely informed,  $d = 100$ , or uninformed,  $d = 1$ . In the directed network, agents are equally likely to be looking at an informed or uninformed neighbor. In contrast, in the undirected network, agents are far more likely to be looking at informed neighbors.

them adopts is  $y = (1 - \bar{x})^2$ . Differentiating, the social learning curve is given by the solution  $y^u$  of

$$\dot{y} = -\alpha [2\sqrt{y} - 2y[1 - F(\pi(y))]] =: -\Psi(y).$$

Initially, learning in the undirected network is twice as fast as that in the directed network,  $\Psi(1) = 2\Phi(1)$ . If  $\alpha$  is sufficiently small, one can prove that  $y$  stays relatively high, so that  $\Psi(y) \geq \Phi(y)$  for the relevant range of  $y$ , and learning is better in the undirected network,  $1 - y^u > 1 - y^d$ .<sup>20</sup> Our numerical simulations suggest a more general result: For all parameter values  $\alpha, \pi_0, F$  we have tried, the undirected network performs better than the directed one. Intuitively, the value of having an additional direct link outweighs the problems with the backward link.  $\triangle$

### 5.3 Clustering

One prominent feature of real social networks is clustering, whereby  $i$ 's neighbors  $j, k$  are also linked to each other. For example, consider an agent who gets information from her family, her geographic neighbors and her colleagues; we think of information as independent across groups, but correlated within.

The configuration models in the previous subsections do not give rise to such clustering since the chance that any two neighbors of  $i$  are linked vanishes for large  $I$ . To capture clustering, we consider the following variant of the configuration model. Each agent independently draws  $D$  pairs of link-stubs, which are then randomly connected to two other pairs of link-stubs to form a triangle. As in Section 5.2 we then prune self-links and multi-links, as well as left-over pairs if the total number of pairs is not divisible by three. Also, recall the weighted distribution  $D'$  from (20) that captures the number of link-pairs of a typical neighbor by accounting for the friendship paradox.

We now study the behavior of the limit economy, as the number of agents grows large; Proposition 2' in Section 5.4 justifies this approach. Adoption is independent across neighboring triangles but correlated within. The simplest such network, corre-

<sup>20</sup>For example, if  $\alpha \leq 1 - (1 - F(\pi_0)/2)^2$  then the fact that  $x \leq \alpha$  implies that  $y \geq (1 - F(\pi_0)/2)^2$ . If BHR also holds then  $\Psi(y) - \Phi(y) = \alpha [2\sqrt{y} - 1 - y[1 - F(\pi(y))]] \geq \alpha [2\sqrt{y} - 2 + F(\pi_0)] \geq 0$  where the first inequality comes from BHR and the second from the lower bound on  $y$ .



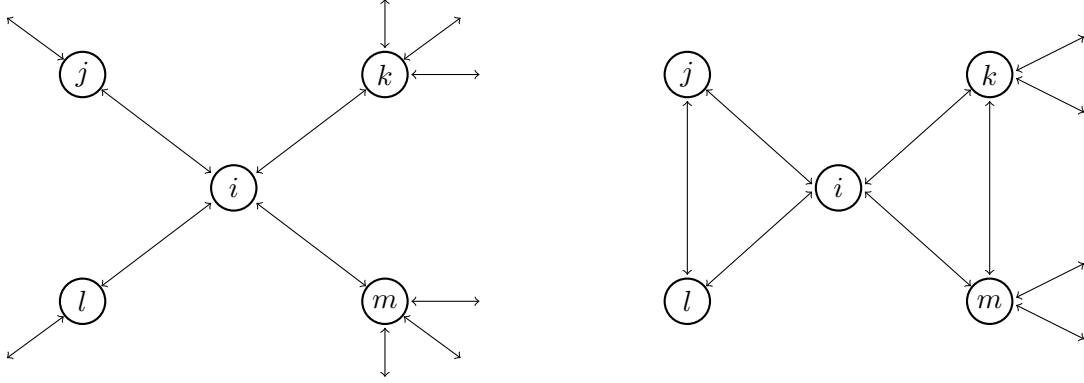


Figure 9: **Bilateral Links vs. Triangles.** This figure illustrates two networks from agent  $i$ 's perspective. In the left network, everyone has  $2D$  bilateral links, where  $D = 2$  for  $\{i, k, m\}$  and  $D = 1$  for  $\{j, l\}$ . In the right network, everyone is part of  $D$  triangles.

sponding to  $D \equiv 1$ , is the triangle from Section 4.4 with  $I = 2$ . There we argued that the learning curve is determined by the adoption probability  $\hat{x}$  of the first adopter. Since the first adopter expects to see no adoption with subjective probability  $(1 - \hat{x})^2$  but objectively never observes an adoption, we concluded that  $\dot{\hat{x}} = \phi(1, (1 - \hat{x})^2)$ . In our random network, agent  $i$ 's neighbors additionally learn from another  $D' - 1$  independent triangles, from which they observe no adoption with probability  $(1 - \hat{x})^{2(D'-1)}$ . All told, define  $\hat{x}^*$  as the solution of

$$\dot{\hat{x}} = E \left[ \phi \left( (1 - \hat{x})^{2(D'-1)}, (1 - \hat{x})^{2D'} \right) \right]. \quad (23)$$

Agent  $i$ 's actual, unconditional adoption rate then equals  $E[\Phi(1 - \hat{x})^{2D}]$ .

We now assess the effect of clustering on social information and welfare. In Section 4.2 we showed that if agent  $i$  observes two agents  $j, k$  then adding a correlating link  $j \rightarrow k$  makes  $i$  worse off and  $j$  better off. Here we evaluate the overall effect by comparing an undirected network with  $D$  pairs of link-stubs to one with  $2D$  bilateral link-stubs, as illustrated in Figure 9. The social learning curve equals  $\hat{y}_{2d}^* = (1 - \hat{x}^*)^{2d}$  in the former network, and  $\bar{y}_{2d}^* = (1 - \bar{x}^*)^{2d}$  in the latter, where  $\hat{x}^*$  solves (23) and  $\bar{x}^*$  solves (21).

**Theorem 3.** *Clusters reduce learning and welfare in large random networks: For each  $d$ ,  $1 - \hat{y}_{2d}^* < 1 - \bar{y}_{2d}^*$ .*

*Proof.* Equation (20) implies that with  $2D$  bilateral links, the link distribution of a neighbor equals  $(2D)' = 2D'$ .<sup>21</sup> Thus, the conditional adoption of one's neighbor follows

$$\dot{\hat{x}} = E\left[\phi\left((1 - \bar{x})^{2D'-1}, (1 - \bar{x})^{2D'}\right)\right]. \quad (24)$$

Since  $\phi$  decreases in its first argument, the RHS of (24) exceeds the RHS of (23) when  $\bar{x} = \hat{x}$ . Thus, the Single-Crossing Lemma implies  $\hat{x}^* < \bar{x}^*$  and so  $1 - \hat{y}_{2d}^* < 1 - \bar{y}_{2d}^*$ .  $\square$

Agents learn slower in clusters than with an equivalent number of independent links. Intuitively, agent  $i$  needs one of her neighbors to be sufficiently optimistic that they are willing to experiment. Clusters correlate the decisions of  $i$ 's neighbors, making them more pessimistic in exactly the event when  $i$  most wants them to experiment.

## 5.4 General Model: A Framework for Empirical Analysis

We now define random networks that encompass the undirected links and clusters from the last two sections; the middle panel of Figure 2 illustrates such networks. Proposition 2' shows that limit equilibria of this model are described by a simple, two-dimensional ODE (25-26).

To define these networks  $\hat{\mathcal{G}}_I$ , every agent independently draws  $\bar{D}$  bilateral stubs and  $\hat{D}$  pairs of triangle stubs. Both distributions can be arbitrary, but have a finite expectation. We connect pairs of bilateral stubs and triples of triangular stubs at random, and then prune self-links, multi-links, and left-over stubs (if  $\sum \bar{d}_i$  is odd or  $\sum \hat{d}_i$  not divisible by three). Agents know their number of bilateral links and triangle links after the pruning.

Assume that  $I$  is large, and define neighbors' link distributions  $\bar{D}'$  and  $\hat{D}'$  as in (20). Since  $\bar{D}$  and  $\hat{D}$  are independent, a neighbor on a bilateral link has  $\bar{D}'$  bilateral links and  $\hat{D}$  triangle link pairs, whereas a neighbor on a triangular link has  $\bar{D}$  bilateral links and  $\hat{D}'$  triangle link pairs. As in Sections 5.2 and 5.3, agents condition on the fact that their neighbors cannot have seen them adopt. So motivated, define  $(\bar{x}^*, \hat{x}^*)$

---

<sup>21</sup>Indeed,  $\Pr((2D)' = 2d) = \Pr(2D = 2d) \frac{2d}{E[2D]} = \Pr(D = d) \frac{d}{E[D]} = \Pr(D' = d) = \Pr(2D' = 2d)$ .

as the solution to the two-dimensional ODE

$$\dot{\bar{x}} = E \left[ \phi \left( (1 - \bar{x})^{\bar{D}'-1} (1 - \hat{x})^{2\hat{D}}, (1 - \bar{x})^{\bar{D}'} (1 - \hat{x})^{2\hat{D}} \right) \right] \quad (25)$$

$$\dot{\hat{x}} = E \left[ \phi \left( (1 - \bar{x})^{\bar{D}} (1 - \hat{x})^{2(\hat{D}'-1)}, (1 - \bar{x})^{\bar{D}} (1 - \hat{x})^{2\hat{D}'} \right) \right]. \quad (26)$$

As in Section 5.2,  $\bar{x}$  is the probability  $i$ 's bilateral neighbor  $j$  adopts before  $t_i$ . Agent  $j$ 's "subjective" probability of observing no adoption conditions on  $\bar{D}'$  bilateral links and  $\hat{D}$  triangle link pairs; but from  $i$ 's "objective" perspective, the number of bilateral links on which  $j$  could observe an adoption drops to  $\bar{D}' - 1$ . Similarly, as in Section 5.3,  $\hat{x}$  is the probability that the first adopter  $j$  in one of  $i$ 's triangles adopts before  $t_i$ . Agent  $j$ 's "subjective" probability of observing no adoption conditions on  $\bar{D}'$  bilateral links and  $\hat{D}$  triangle link pairs; but from  $i$ 's "objective" perspective, the number of triangle link pairs on which  $j$  could observe an adoption drops to  $\hat{D}' - 1$ .

Given beliefs  $(\bar{x}^*, \hat{x}^*)$ , an agent with  $\bar{d}$  bilateral neighbors and  $\hat{d}$  pairs of triangular neighbors adopts cutoffs  $c_{\bar{d}, \hat{d}}^* = \pi((1 - \bar{x}^*)^{\bar{d}} (1 - \hat{x}^*)^{2\hat{d}})$ , and her unconditional adoption probability follows  $\dot{x}_{\bar{d}, \hat{d}} = \Phi((1 - \bar{x})^{\bar{d}} (1 - \hat{x})^{2\hat{d}})$ . We can now extend the limit result, Proposition 2, to undirected networks with clusters.

**Proposition 2'.**  $(c_{\bar{d}, \hat{d}}^*)$  is the unique limit equilibrium of  $\hat{\mathcal{G}}_I$ .

*Proof.* See Appendix B.5. □

Social learning in this complex network is thus characterized by a simple, two-dimensional ODE, (25-26). The model in this section is already rich enough to match important network statistics, like the degree distribution or the clustering coefficient. But the logic behind equations (25-26) and Proposition 2' easily accommodates additional features and alternative modeling assumptions. Allowing for larger  $(d + 1)$ -clusters amounts to replacing the "2" in the exponents of (25-26) by " $d$ ". Allowing for multiple types  $\theta$  of agents is slightly more complicated since it requires keeping track of the conditional adoption probability of type  $\theta$ 's neighbor  $\theta'$  for all pairs  $(\theta, \theta')$ ; we spell this out in Appendix B.3. Our analysis also extends to correlation of  $\bar{D}$  and  $\hat{D}$ , and to the alternative assumption that agents only know their total number of links, but cannot distinguish bilateral from triangle links.

## 5.5 Correlation Neglect

Bayesian updating on networks can be very complex as agents try to decompose new and old information. For example, Eyster and Rabin’s (2014) “shield” example shows that a Bayesian agent  $i$  should counter-intuitively “anti-imitate” a neighbor  $j$  if  $j$ ’s action is also encoded in the actions of  $i$ ’s other neighbors. Instead of anti-imitating, agents may adopt heuristics, such as ignoring the correlations between neighbors’ actions (e.g. Eyster et al. (2018), Enke and Zimmermann (2019), Chandrasekhar et al. (2020)). Our model can be adapted to capture such mis-specifications, and predicts that correlation neglect reduces social learning.<sup>22</sup>

Consider a configuration model with  $D$  pairs of undirected triangular stubs. We model correlation neglect by assuming that all agents believe all their information is independent. That is,  $i$  believes that her neighbors are not connected, believes her neighbors think their neighbors are not connected, and so on. All told, agents think that the network is as in Section 5.2, while in reality it is as in Section 5.3.<sup>23</sup>

Consider the limit as  $I$  grows large. Since agent  $i$  believes that links are generated bilaterally, her “subjective” probability assessment that any of her neighbors has adopted,  $\bar{x}^*$ , solves (24). An agent with  $2d$  links thus uses cutoff  $\pi((1 - \bar{x}^*)^{2d})$  when choosing whether to inspect. In reality, agent  $i$ ’s neighbors form triangles  $(i, j, k)$ , and so the “objective” probability  $\check{x}^*$  that the first-adopter in a triangle adopts follows a variant of the usual first-adopter triangle formula (23),

$$\check{x} = E\left[\phi\left((1 - \check{x})^{2(D'-1)}, (1 - \bar{x}^*)^{2D'}\right)\right]. \quad (27)$$

Intuitively, the first adopter in  $(i, j, k)$  expects to see no adoption with probability  $(1 - \bar{x}^*)^{2D'}$  but only actually sees adoption at a rate of  $(1 - \check{x}^*)^{2(D'-1)}$ . Formally, (27) is an instance of equation (38) in Appendix B.5.

Turning to the effect of correlation neglect on social learning, suppose agent  $i$  has  $2d$  links. Her social information in a network of triangles is  $\hat{y}_{2d}^* = (1 - \hat{x}^*)^{2d}$ , while her social information on the same network with correlation neglect is  $\check{y}_{2d}^* = (1 - \check{x}^*)^{2d}$ .

<sup>22</sup>There is a growing literature studying herding models with mis-specified beliefs. Eyster and Rabin (2010) and Bohren and Hauser (2019) study complete networks, while Eyster and Rabin (2014) and Dasaratha and He (2020) consider the role of network structure.

<sup>23</sup>To formally capture mis-specification in the general model of Section 2, we drop the assumption that agents’ beliefs  $\mu(G, \xi_{-i}|\xi_i)$  are deduced from a common prior  $\mu(G, \xi)$ .

**Theorem 4.** *Correlation neglect reduces social learning: For any  $d$ ,  $1 - \check{y}_{2d}^* < 1 - \hat{y}_{2d}^*$ .*

*Proof.* The proof of Theorem 2 established that clustering decreases neighbors' adoption rates,  $\bar{x}^* > \hat{x}^*$ . Since  $\phi$  increases in its second argument, the RHS of (27) is smaller than the RHS of (23) when  $\check{x} = \hat{x}$ . Then the Single-Crossing Lemma implies that  $\check{x}^* < \hat{x}^*$  and so  $1 - \check{y}_{2d}^* < 1 - \hat{y}_{2d}^*$ .  $\square$

Intuitively, when agent  $j$  believes all his sources of information are independent, he over-estimates the chance of observing at least one adoption, and grows overly pessimistic when he observes none. This reduces  $j$ 's adoption probability and reduces agent  $i$ 's social information.

As a corollary, correlation neglect reduces welfare: By Theorem 4 it reduces  $i$ 's social information and additionally causes her to react suboptimally to the information she does have. Ironically, while correlation neglect lowers  $i$ 's objective expected utility, it raises her subjective expected utility: Formally,  $1 - \bar{y}_{2d}^* > 1 - \hat{y}_{2d}^*$ , so her subjective social information is higher than in the actual network. Intuitively,  $i$  thinks she is learning from lots of independent signals, rather than correlated ones, which makes her too optimistic. It is precisely this over-optimism about observing an adoption that reduces actual adoption probabilities by flipping into over-pessimism when the adoptions fail to materialize.

## 5.6 Information Aggregation and the Value of Links

We end this section on one of the most central issues in social learning: Does society correctly aggregate dispersed information? The configuration model provides a novel perspective on this question. Consider a regular network as both the degree  $d$  and the total number of agents  $I$  grow large. We claim that if  $I$  grows sufficiently faster than  $d$ , then agents have access to a large number of almost independent signals and society correctly aggregate information. However, if  $d$  grows too quickly, the network becomes “clustered” and information aggregation can fail. As a corollary, in a large society with a fixed number of agents, more links can introduce excessive correlation and lower everyone's utility.

To see the problem with clustering, recall the first-adopter probability  $\hat{x}_I$  in the complete network with  $I + 1$  agents from Section 4.4. Also recall the lowest cost type

$\underline{\kappa} < \pi_0$  and let  $\underline{y} = \pi^{-1}(\underline{\kappa})$  be the belief below which even type  $\underline{\kappa}$  stops inspecting. As  $I \rightarrow \infty$  the social information  $y_I = (1 - \hat{x}_I)^I$  immediately drops to the “choke point”  $\underline{y}$  and stays there for all  $t > 0$ .<sup>24</sup> Intuitively, observing no adoption from an exploding number of agents  $I$  makes agents grow pessimistic so fast that they are only willing to inspect at the very first instant. Learning is perfect when  $\underline{y} = 0$ , which is the case if and only if  $\underline{\kappa} = 0$ . Thus, unboundedly low costs substitute for unboundedly strong signals as necessary and sufficient condition for information aggregation, as in Ali (2018) and Mueller-Frank and Pai (2016). In particular, when  $\underline{\kappa} > 0$ , high quality products fail to take off with probability  $\underline{y} > 0$ , in which case they fizzle out immediately.<sup>25</sup>

This failure of information aggregation does not arise when the network remains sparse as the degree grows. Specifically consider the limit  $d \rightarrow \infty$  of the limit equilibria, where adoption and social information are given by  $\hat{x}_d^* = \Phi(y_d^*)$  and  $y_d^* = (1 - \hat{x}_d^*)^d$ , recalling (12). In other words, this is the double-limit of a regular random network where first  $I \rightarrow \infty$  and then  $d \rightarrow \infty$ . Since  $\Phi(y)$  is bounded away from zero,<sup>26</sup> we have  $y_d^* \rightarrow 0$  for all  $t > 0$  as  $d \rightarrow \infty$ , meaning that  $i$ 's information becomes perfect irrespective of the cost distribution. Intuitively, agents' signals on a sparse network are independent and their joint adoption decisions become perfectly informative as the degree grows. This contrasts with Acemoglu et al. (2011) where “for many common deterministic and stochastic networks, bounded private beliefs are incompatible with asymptotic learning” and positive results (e.g. their Theorem 4) rely on a slowly exploding number of “guinea pig” agents, who have little social information, and whose actions are thus very informative about their private signals. In our model, the guinea pigs arise naturally, by virtue of being the first to enter the market.<sup>27</sup>

<sup>24</sup>*Proof:* To see  $\lim_{I \rightarrow \infty} (1 - \hat{x}_I)^I = \underline{y}$  for all  $t > 0$ , first note that the first-adopter stops experimenting,  $\hat{x}_I = 0$ , when  $(1 - \hat{x}_I)^I$  drops below  $\underline{y}$ , so  $(1 - \hat{x}_I)^I \geq \underline{y}$  for any  $I, t$ . For the opposite inequality, if  $\limsup_{I \rightarrow \infty} (1 - \hat{x}_I)^I > \underline{y}$  for some  $t > 0$ , then  $\hat{x}_I = \alpha F(\pi((1 - \hat{x}_I)^I))$  is bounded away from zero on  $[0, t]$ , hence  $\hat{x}_I$  is bounded away from 0 at  $t$ . Then  $\limsup_{I \rightarrow \infty} (1 - \hat{x}_I)^I \rightarrow 0$  at  $t$ , contradicting the initial assumption that  $\limsup_{I \rightarrow \infty} (1 - \hat{x}_I)^I > \underline{y}$ .

<sup>25</sup>As  $I \rightarrow \infty$ , agents always stop inspecting low-quality products. The asymmetry between good and bad products is seen elsewhere in the literature (e.g. Guarino et al. (2011), Hendricks et al. (2012) and Herrera and Hörner (2013)).

<sup>26</sup>*Proof:* Since  $F(\pi_0) < 1$ , there exists  $\epsilon > 0$  with  $\alpha F(\pi(1 - \epsilon/\alpha)) > \epsilon$ . Then  $\Phi(y) = \alpha[1 - y(1 - F(\pi(y)))]$  exceeds  $\alpha(1 - y) \geq \epsilon$  for  $y \leq 1 - \epsilon/\alpha$ , and  $\alpha F(\pi(y)) \geq \epsilon$  for  $y > 1 - \epsilon/\alpha$ . All told,  $\Phi(y) \geq \epsilon > 0$  for all  $y$ .

<sup>27</sup>The distinguishing feature between our model and Acemoglu et al. (2011) is that each agent's

## 6 Conclusion

Social learning plays a crucial role in the diffusion of new products (e.g. Moretti, 2011), financial innovations (e.g. Banerjee et al., 2013), and new production techniques (e.g. Conley and Udry, 2010). This paper proposes a tractable model of social learning on networks, describes behavior via a system of differential equations, and studies the effect of network structure on learning dynamics. We characterize learning and adoption on small deterministic networks, and use this as the building blocks to study large random networks. We show that an agent benefits from more direct and indirect links in “tree like” networks, but is harmed by correlating and backward links. We started this paper by asking about the effect of connectedness and centralization on learning and adoption. The connectedness of a network is a two-sided sword, improving social information as long as the network remains sparse, but eventually harming it as the network becomes clustered. For the same reason, centralization - as measured by the amount of clustering - unambiguously slows learning.

The paper has two broad contributions. First, it develops intuition for how network structure affects learning. Second, it can be used to structurally estimate diffusion in real-world networks while maintaining Bayesian rationality. To focus on the effect of the network on social learning, the model is stylized in some aspects. For example, we assume that a single adoption is proof of high quality. In Appendix C we extend our analysis to more general, imperfect information structures where agents sometimes adopt low-quality products, and show that most of our analysis extends to this richer model.

Moving forwards, one can take the model in a number of different directions. One could study the effect of policies, like changing the price of the product (e.g. Campbell (2013)) or seeding the network (e.g. Akbarpour, Malladi, and Saberi (2018)). While we studied correlation neglect, one could allow for other mis-specifications (e.g. Bohren and Hauser (2019)). Finally, one could allow agents to choose when they move, allowing skeptical agents to delay (e.g. Chamley and Gale (1994)).

---

neighborhood naturally resembles a tree in configuration models  $\mathcal{G}_I$  for exploding  $I$ ; this feature does not arise naturally in a model of one infinite sequence of agents. Other differences between the models, such as the fact that their agents enter the game with private information whereas ours choose to acquire private information, are not important for this contrast.

## References

- ACEMOGLU, D., M. A. DAHLEH, I. LOBEL, AND A. OZDAGLAR (2011): “Bayesian Learning in Social Networks,” *Review of Economic Studies*, 78(4), 1201–1236.
- AKBARPOUR, M., S. MALLADI, AND A. SABERI (2018): “Just a Few Seeds More: Value of Network Information for Diffusion,” Working paper, Stanford University.
- ALI, S. N. (2018): “Herding with Costly Information,” *Journal of Economic Theory*, 175, 713–729.
- BANERJEE, A., A. G. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2013): “The Diffusion of Microfinance,” *Science*, 341(6144), 1236498.
- BANERJEE, A. V. (1992): “A Simple Model of Herd Behavior,” *Quarterly Journal of Economics*, 107(3), 797–817.
- (1993): “The Economics of Rumours,” *Review of Economic Studies*, 107(2), 309–327.
- BASS, F. M. (1969): “A New Product Growth for Model Consumer Durables,” *Management Science*, 15(5), 215–227.
- BIKHCHANDANI, S., D. HIRSHLEIFER, AND I. WELCH (1992): “A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades,” *Journal of Political Economy*, 100(5), 992–1026.
- BOHREN, J. A., AND D. N. HAUSER (2019): “Social Learning with Model Misspecification: A Framework and a Characterization,” Working paper, University of Pennsylvania.
- BÖRGERS, T. (2009): “Notes on Blackwell’s Comparison of Experiments,” Working paper, University of Michigan.
- CAI, J., AND A. SZEIDL (2018): “Interfirm Relationships and Business Performance,” *Quarterly Journal of Economics*, 3(133), 1229–1282.



- CAMPBELL, A. (2013): “Word-of-Mouth Communication and Percolation in Social Networks,” *American Economic Review*, 103(6), 2466–98.
- CELEN, B., AND S. KARIV (2004): “Observational Learning under Imperfect Information,” *Games and Economic Behavior*, 47(1), 72–86.
- CENTOLA, D. (2010): “The Spread of Behavior in an Online Social Network Experiment,” *Science*, 329(5996), 1194–1197.
- CHAMLEY, C., AND D. GALE (1994): “Information Revelation and Strategic Delay in a Model of Investment,” *Econometrica*, pp. 1065–1085.
- CHANDRASEKHAR, A. G., H. LARREGUY, AND J. P. XANDRI (2020): “Testing Models of Social Learning on Networks: Evidence from Two Experiments,” *Econometrica*, 88(1), 1–32.
- CONLEY, T. G., AND C. R. UDRY (2010): “Learning about a New Technology: Pineapple in Ghana,” *American Economic Review*, 100(1), 35–69.
- DASARATHA, K., AND K. HE (2020): “Network Structure and Naive Sequential Learning,” *Theoretical Economics*, 15(2), 415–444.
- DEIMLING, K. (1977): *Ordinary Differential Equations in Banach Spaces*. Springer-Verlag, Berlin.
- ENKE, B., AND F. ZIMMERMANN (2019): “Correlation Neglect in Belief Formation,” *Review of Economic Studies*, 86(1), 313–332.
- EYSTER, E., AND M. RABIN (2010): “Naive Herding in Rich-Information Settings,” *American Economic Journal: Microeconomics*, 2(4), 221–43.
- (2014): “Extensive Imitation is Irrational and Harmful,” *Quarterly Journal of Economics*, 129(4), 1861–1898.
- EYSTER, E., M. RABIN, AND G. WEIZSÄCKER (2018): “An Experiment on Social Mislearning,” Working paper, Harvard University.

- GOLUB, B., AND M. O. JACKSON (2012): “How Homophily affects the Speed of Learning and Best-Response Dynamics,” *Quarterly Journal of Economics*, 127(3), 1287–1338.
- GOLUB, B., AND E. SADLER (2016): “Learning in Social Networks,” in *The Oxford Handbook of the Economics of Networks*, ed. by Y. Bramoullé, A. Galeotti, and B. Rogers, chap. 19, pp. 504–542. Oxford University Press, Oxford.
- GRANOVETTER, M. S. (1973): “The Strength of Weak Ties,” *American Journal of Sociology*, 78(6), 1360–1380.
- GUARINO, A., H. HARMGART, AND S. HUCK (2011): “Aggregate Information Cascades,” *Games and Economic Behavior*, 73(1), 167–185.
- HANN-CARUTHERS, W., V. V. MARTYNOV, AND O. TAMUZ (2018): “The Speed of Sequential Asymptotic Learning,” *Journal of Economic Theory*, 173, 383–409.
- HENDRICKS, K., A. SORENSEN, AND T. WISEMAN (2012): “Observational Learning and Demand for Search Goods,” *American Economic Journal: Microeconomics*, 4(1), 1–31.
- HERRERA, H., AND J. HÖRNER (2013): “Biased Social Learning,” *Games and Economic Behavior*, 80, 131–146.
- JACKSON, M. O. (2010): *Social and Economic Networks*. Princeton University Press.
- LOBEL, I., AND E. SADLER (2015): “Information Diffusion in Networks through Social Learning,” *Theoretical Economics*, 10(3), 807–851.
- LOMYS, N. (2019): “Sequential Collective Search in Networks,” Working paper, SSRN.
- MCADAMS, D., AND Y. SONG (2020): “Viral Social Learning,” Working paper, SSRN.
- MILGROM, P., AND R. WEBER (1982): “A theory of auctions and competitive bidding,” *Econometrica*, 50, 1089–1122.

- MONZÓN, I., AND M. RAPP (2014): “Observational Learning with Position Uncertainty,” *Journal of Economic Theory*, 154, 375–402.
- MORETTI, E. (2011): “Social Learning and Peer Effects in Consumption: Evidence from Movie Sales,” *Review of Economic Studies*, 78(1), 356–393.
- MORRIS, S. (2000): “Contagion,” *Review of Economic Studies*, 67(1), 57–78.
- MUELLER-FRANK, M., AND M. M. PAI (2016): “Social Learning with Costly Search,” *American Economic Journal: Microeconomics*, 8(1), 83–109.
- SADLER, E. (2020): “Diffusion Games,” *American Economic Review*, 110(1), 225–270.
- SGROI, D. (2002): “Optimizing Information in the Herd: Guinea Pigs, Profits, and Welfare,” *Games and Economic Behavior*, 39(1), 137–166.
- SHAKED, M., AND J. G. SHANTHIKUMAR (2007): *Stochastic Orders*. Springer Science & Business Media.
- SMITH, L., AND P. SØRENSEN (1996): “Rational Social Learning with Random Sampling,” Working paper, MIT.
- (2000): “Pathological Outcomes of Observational Learning,” *Econometrica*, 68(2), 371–398.

# Online Appendix

## A Proofs from Section 4

### A.1 Example 4: Two Links vs Infinite Chain

When agent  $i$  has two uninformed neighbors, the probability neither of them adopts is  $y(\alpha t) := (1 - F(\pi_0)\alpha t)^2$ . Compare this to the lower bound on the non-adoption probability in the chain from (15),

$$\zeta(\alpha t) := \frac{1 - F(\pi_0)e^{(1-F(\pi_0))\alpha t}}{1 - F(\pi_0)}.$$

We wish to show that  $\zeta(\alpha t) - y(\alpha t) \geq 0$  for all  $\alpha t$ . It suffices to show this inequality for  $\alpha t = 1$ . This follows since (i)  $y(\alpha t)$  is convex while  $\zeta(\alpha t)$  is concave, and (ii)  $\zeta(\alpha t) - y(\alpha t)$  is zero and increasing at  $\alpha t = 0$ .<sup>28</sup>

Setting  $\alpha t = 1$ , abbreviating  $\delta := 1 - F(\pi_0) \in (0, 1]$ , and multiplying by  $\delta$ , we thus wish to show that  $\xi(\delta) := 1 - (1 - \delta)e^\delta - \delta^3 \geq 0$ . Differentiating, one can see that  $\xi$  has a unique local extremum  $\delta^*$  on  $[0, 1]$  and that  $\xi''(\delta^*) \leq 0$ . Thus, it is quasi-concave with  $\xi(0) = \xi(1) = 0$ , and hence is positive everywhere.

### A.2 Proof of Lemma 2

We wish to study the adoption decision of an agent given that no other agent in the cluster has yet adopted. Formally, let  $y$  be the probability that no other agent has yet adopted when  $i$  enters (or “wakes up”) at time  $t$ . That is,

$$y = \Pr(\lambda_j \neq a \text{ for all } j \neq i | \lambda_i = \emptyset).$$

When agent  $i$  wakes up and sees that no other agent has yet adopted, she thus adopts with probability  $\alpha F(\pi(y)) = \phi(1, \pi)$ . Agents who have not adopted may either be

---

<sup>28</sup>Intuitively, part (i) follows since the marginal benefit from two links eventually falls because of double-counting, while the marginal benefit from the infinite line increases with the increased likelihood of benefiting from the indirect links. For part (ii), we have  $y'(0) = -2F(\pi_0)$  and  $\zeta'(0) = -F(\pi_0)$ . Intuitively, the indirect links in the chain are initially useless, so all that matters is the one direct link.

asleep ( $\lambda_j = \emptyset$ ) or decided not to adopt ( $\lambda_j = b$ ). Let  $y_\nu$  be the probability that  $\nu$  agents are asleep,

$$y_\nu := \Pr(\lambda_j = \emptyset \text{ for } \nu \text{ others } j \neq i, \text{ and } \lambda_j = b \text{ for the other } I - \nu | \lambda_i = \emptyset).$$

Clearly,  $y = \sum_{\nu=0}^I y_\nu$ .

We now characterize  $\{y_\nu\}$  recursively. For  $\nu = I$ , the probability that everyone is asleep is  $y_I = (1-t)^I$ . For states  $\nu < I$ , probability flows in from state  $\nu + 1$  as one of these agents wakes up and chooses not to adopt given that no-one else had adopted; this rate is given by  $\frac{\nu+1}{1-t} y_{\nu+1} (1 - \phi(1, y))$ . There is also outflow at rate  $\frac{\nu}{1-t} y_\nu$  as one of the  $\nu$  agents wakes up; this scales down the time- $s$  inflow by a factor  $(\frac{1-t}{1-s})^\nu$  at time- $t$ . All told,

$$y_{\nu,t} = \int_0^t \left( \frac{1-t}{1-s} \right)^\nu y_{\nu+1,s} \frac{\nu+1}{1-s} (1 - \phi(1, y_s)) ds = (\nu+1)(1-t)^\nu \int_0^t y_{\nu+1,s} \frac{1 - \phi(1, y_s)}{(1-s)^{\nu+1}} ds \quad (28)$$

We claim that this can be written as

$$y_{\nu,t} = \frac{I!}{\nu!} (1-t)^\nu \underbrace{\int_0^t \int_0^s \dots \int_0^r}_{I-\nu \text{ integrals}} [(1 - \phi(1, y_q)) \dots (1 - \phi(1, y_s))] dq \dots dr ds. \quad (29)$$

which reduces to  $y_{I,t} = (1-t)^I$  for  $\nu = I$ . To verify the claim, assume (29) holds for  $\nu + 1$  and substitute into (28). This becomes,

$$y_{\nu,t} = (\nu+1)(1-t)^\nu \int_0^t \left[ \frac{I!}{(\nu+1)!} (1-s)^{\nu+1} \underbrace{\int_0^s \dots \int_0^r}_{I-\nu-1 \text{ integrals}} [(1 - \phi(1, y_q)) \dots (1 - \phi(1, y_r))] dq \dots dr \right] \frac{1 - \phi(1, y_s)}{(1-s)^{\nu+1}} ds$$

which collapses to (29).

To further simplify (29), note that the integration domain consists of all  $(I - \nu)$ -tuples  $(q, \dots, r, s)$  with  $0 \leq q \leq \dots \leq r \leq s \leq t$  and the integrand is symmetric in  $(q, \dots, r, s)$ . Since there are  $(I - \nu)!$  permutations of the integration variables, (29)

equals

$$\begin{aligned}
y_{\nu,t} &= \frac{I!}{\nu!(I-\nu)!} (1-t)^\nu \underbrace{\int_0^t \int_0^t \dots \int_0^t}_{I-\nu \text{ integrals}} [(1-\phi(1,y_q)) \dots (1-\phi(1,y_s))] dq dr \dots ds \\
&= \binom{I}{\nu} (1-t)^\nu \left( \int_0^t (1-\phi(1,y_s)) ds \right)^{I-\nu}.
\end{aligned}$$

Summing over  $\nu$  and using the binomial formula,

$$y_t = \sum_{\nu=0}^I y_{\nu,t} = \left( (1-t) + \int_0^t (1-\phi(1,y_s)) ds \right)^I = \left( 1 - \int_0^t \phi(1,y_s) ds \right)^I$$

as required.

### A.3 The Single-Crossing Lemma

**Lemma 3.** *Let  $(x_t), (\tilde{x}_t)$  solve  $\dot{x} = \psi(x)$  and  $\dot{\tilde{x}} = \tilde{\psi}(\tilde{x})$  with  $x_0 = \tilde{x}_0 = 0$ , where  $\psi(x) > \tilde{\psi}(x) > 0$  for all  $x \in (0, 1]$  and  $\psi(0) = \tilde{\psi}(0) > 0$ . Then  $x_t > \tilde{x}_t$  for all  $t > 0$ .*

The reason that we formalize this seemingly obvious fact instead of invoking, say, Milgrom and Weber (1982, Lemma 2), is that the trajectory  $(x_t)$  is steeper than  $(\tilde{x}_t)$  only when they coincide, and the inequality is weak at  $x_0 = \tilde{x}_0 = 0$ .

*Proof.* For any  $\epsilon > 0$ , define  $x^\epsilon : [\epsilon, 1] \rightarrow \mathbb{R}$  as the solution of  $\dot{x}^\epsilon = \psi(x^\epsilon)$  with initial condition  $x^\epsilon_\epsilon = \tilde{x}_\epsilon$ . Then  $x^\epsilon_t > \tilde{x}_t$  for all  $t > \epsilon$ : The set  $\{t \in [\epsilon, 1] : x^\epsilon_t > \tilde{x}_t\}$  is non-empty and open, and if there was a smallest  $t > \epsilon$  with  $x^\epsilon_t = \tilde{x}_t$ , we'd run into the contradiction that  $\dot{x}^\epsilon_t > \dot{\tilde{x}}_t$  yet  $x^\epsilon_s > \tilde{x}_s$  for all  $s \in (\epsilon, t)$ .

Since the solution of a differential equation is continuous in its initial conditions, we have  $\lim_{\epsilon \rightarrow 0} x^\epsilon_t = x_t$  and so  $x_t \geq \tilde{x}_t$  for all  $t > 0$ . But  $x_t \geq \tilde{x}_t > 0$  implies  $x_{t'} > \tilde{x}_{t'}$  for all  $t' > t$ , and so we get  $x_t > \tilde{x}_t$  for all  $t > 0$ .  $\square$

## B Proofs from Section 5

### B.1 Proof of Proposition 2

We break the proof into four steps:

- (1) Define the branching process.
- (2) Characterize “limit adoption”  $X_\theta(c)$  and “equilibrium limit adoption”  $x_\theta^*$  associated with the branching process.
- (3) Show that  $(c_d^*)$  from (19) is a limit equilibrium by showing that learning curves  $y_{I,d}^*$  in  $\mathcal{G}_I$  converge to those associated with the branching process.
- (4) Show that the limit equilibrium is unique.

While steps 1 and 2 may seem clear enough for the directed networks studied here, we spell them out in formal detail to prepare the ground for the more involved case of undirected networks with clusters in Proposition 2’.

### B.1.1 Branching process

Here we formalize the idea that the random network  $\mathcal{G}_I$  locally approximates a tree as  $I$  grows large. Following Sadler (2020, Section 8.2), for any degree  $d \in \mathbb{N}^\Theta$  we consider a multi-type branching process  $\mathcal{T}_d$  where offspring is given by  $d$  in the first step, and distributed according to  $D$  in all subsequent steps.<sup>29</sup> For any radius  $r \in \mathbb{N}$ , let  $\mathcal{T}_{d,r}$  be the random rooted graph generated by the branching process  $\mathcal{T}_d$ , truncated after step  $r$ .

Turning to our finite networks, for agent  $i$  with degree  $d$  in network  $G$ , define  $i$ ’s  $r$ -neighborhood  $G_{i,r}$  as the subgraph consisting of all nodes and edges in  $G$  that can be reached from  $i$  via paths of length at most  $r$ ; e.g. for  $r = 1$ ,  $G_{i,r}$  consists of  $i$ , her neighbors and her outlinks. Let  $\mathcal{G}_{I,d,r}$  be the random rooted graph generated by realizing a network  $G$  from  $\mathcal{G}_I$ , choosing an agent  $i$  with degree  $d$  at random, and truncating  $G$  to  $i$ ’s  $r$ -neighborhood  $G_{i,r}$ .

We can now state our formal result, which mirrors Sadler’s Lemma 1.

**Lemma 4.**  $\mathcal{G}_{I,d,r}$  can be coupled to  $\mathcal{T}_{d,r}$  with probability  $\tilde{\rho}_{I,d,r}$ , where  $\lim_{I \rightarrow \infty} \tilde{\rho}_{I,d,r} = 1$  for all  $d, r$ .

*Proof.* We uncover the rooted graph  $G_{i,r}$  following a breadth first search procedure: Start by connecting the  $d$  outlinks of root  $i$  to randomly chosen nodes of the appropriate types; then connect the outlinks of these neighbors, and so on until  $G_{i,r}$  is realized.

---

<sup>29</sup>In contrast, Sadler (2020) uses the forward distribution  $D' - 1$  to account for the friendship paradox in his undirected networks. We follow that approach in Appendix B.5.1.

The coupling with the truncated branching process  $T_{d,r}$  succeeds if at every point in this process, the respective type- $\theta$  outlink connects to a previously unvisited type- $\theta$  node; this could fail (i) if type- $\theta$  outlinks cannot be connected because no node of type  $\theta$  was realized, or (ii) by realizing a self-link or multi-link (which we then prune from the network), or (iii) by realizing a link to a node that has already been visited in the process (then  $G_{i,r}$  is not a tree). Since the expected number of nodes in  $G_{i,r}$  is finite, the chance of either of these three causes of failure (aggregated over all  $|G_{i,r}|$  nodes) vanishes for large  $I$ .  $\square$

### B.1.2 Limit adoption

Here we compute limit adoption probabilities of agents on an infinite random tree, generated by the branching process. Specifically, for any cost-cutoffs  $c = (c_d)$  define the limit adoption probabilities as the solution  $X_\theta(c)$  of the ODE

$$\dot{x}_\theta = E \left[ \phi \left( \prod_{\theta'} (1 - x_{\theta'})^{D_{\theta,\theta'}}, \pi^{-1}(c_{D_\theta}) \right) \right]. \quad (30)$$

That is, when all agents in the branching process employ cost-cutoffs  $c$ , an agent with degree  $d = (d_{\theta'})$  sees no adoption with probability  $\prod_{\theta'} (1 - x_{\theta'})^{d_{\theta'}}$ , in which case she adopts if  $\kappa \leq c_d$ . Taking expectations over  $D_\theta$  yields (30). This nests as a special case the solution  $x_\theta^* = X_\theta(c^*)$  of (18) for cost cutoffs  $c_d^* := \pi \left( \prod_{\theta'} (1 - x_{\theta'}^*)^{d_{\theta'}} \right)$ .

### B.1.3 Limit equilibrium

We now turn to the proof of Proposition 2 proper, and show that  $c^* = (c_d^*)$  is a limit equilibrium. In analogy to the limit probabilities  $X_\theta(c)$  and  $x_\theta^*$ , we write  $Y_{I,d}(c)$  for the social learning curve in  $\mathcal{G}_I$  when agents use cutoffs  $c = (c_d)$ , and  $y_{I,d}^* = Y_{I,d}(c^*)$ . Then,  $c^*$  is a limit equilibrium iff  $\lim_{I \rightarrow \infty} \pi(y_{I,d}^*) = c_d^*$  or equivalently, iff

$$\lim_{I \rightarrow \infty} y_{I,d}^* = \prod_{\theta} (1 - x_\theta^*)^{d_\theta}. \quad (31)$$

Given a finite number of agents  $I$ , equation (31) fails for the usual two reasons: correlating and backward links. We now show that these concerns vanish as  $I$  grows large. Lemma 4 showed that any agent  $i$ 's neighborhood in  $G$  resembles a tree. We



complement this argument by showing that  $i$ 's social learning in our model only depends on  $i$ 's neighborhood in  $G$ .

To formalize this statement, say that a path  $i = i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_r$  of length  $r$  is a *learning path* of agent  $i$  if  $t_{i_{\nu-1}} > t_{i_\nu}$  for all  $\nu = 1, \dots, r$ ; the chance of this is  $1/(r+1)!$ . Let  $p_{\theta,r}$  be the probability that a type- $\theta$  node has no learning path of length  $r$  in the infinite random tree generated by the branching process. The expected number of length- $r$  paths is bounded above by  $(\max_\theta \sum_{\theta'} E[D_{\theta,\theta'}])^r$ . Thus,  $p_{\theta,r} \geq (\max_\theta \sum_{\theta'} E[D_{\theta,\theta'}])^r / (r+1)!$  and so  $\lim_{r \rightarrow \infty} p_{\theta,r} = 1$ . For an agent  $i$  with degree  $d = (d_\theta)$ , the probability of the event  $\mathcal{E}$  that ‘‘all of  $i$ 's neighbors have no learning path of length  $r-1$ ’’ equals  $p_{d,r} := \prod_\theta p_{\theta,r-1}^{d_\theta}$ , which also converges to 1 as  $r \rightarrow \infty$ .

Turning from the branching process to the random network  $\mathcal{G}_I$ , note that the probability of event  $\mathcal{E}$  depends on the network  $G$  only via  $i$ 's  $r$ -neighborhood  $G_{i,r}$ . Thus, conditional on the coupling of  $\mathcal{G}_{I,d,r}$  and  $\mathcal{T}_{d,r}$  in Lemma 4,  $p_{d,r}$  also equals the probability of  $\mathcal{E}$  in  $\mathcal{G}_I$ . All told, write  $\rho_{I,d,r} = \tilde{\rho}_{I,d,r} p_{d,r}$  for the joint probability that the coupling succeeds and of event  $\mathcal{E}$ . Then

$$\lim_{r \rightarrow \infty} \lim_{I \rightarrow \infty} \rho_{I,d,r} = \lim_{r \rightarrow \infty} p_{d,r} \lim_{I \rightarrow \infty} \tilde{\rho}_{I,d,r} = 1.$$

We can now study adoption probabilities on  $i$ 's neighborhood. Write  $y_{I,d,r}^*$  for  $i$ 's probability of not observing an adoption, conditional on the intersection of three events:  $i$ 's  $r$ -neighborhood being coupled to the branching process,  $i$  having  $d$  neighbors, and none of these neighbors having a learning path of length  $r$ . Similarly, write  $x_{\theta,r}^*$  for the adoption probability of a type- $\theta$  agent in the branching process, conditional on her not having a learning path of length  $r$ . By construction  $y_{I,d,r}^* = \prod_\theta (1 - x_{\theta,r}^*)^{d_\theta}$ .

We now return to equation (31), which states that the social learning curve on  $\mathcal{G}_I$  converges to the learning curve on the branching process. The triangle inequality implies

$$\begin{aligned} \left| y_{I,d}^* - \prod_\theta (1 - x_\theta^*)^{d_\theta} \right| &\leq \left| y_{I,d}^* - y_{I,d,r}^* \right| + \left| y_{I,d,r}^* - \prod_\theta (1 - x_{\theta,r}^*)^{d_\theta} \right| + \left| \prod_\theta (1 - x_{\theta,r}^*)^{d_\theta} - \prod_\theta (1 - x_\theta^*)^{d_\theta} \right| \\ &\leq (1 - \rho_{I,d,r}) + 0 + \sum_\theta d_\theta (1 - p_{\theta,r}) \end{aligned} \quad (32)$$

for any  $r$ . Since the LHS does not depend on  $r$ , we get

$$\limsup_{I \rightarrow \infty} \left| y_{I,d}^* - \prod_{\theta} (1 - x_{\theta}^*)^{d_{\theta}} \right| \leq \limsup_{r \rightarrow \infty} \limsup_{I \rightarrow \infty} (1 - \rho_{I,d,r}) + \sum_{\theta} d_{\theta} (1 - p_{\theta,r}) = 0.$$

This implies (31) and thereby establishes that  $(c_d^*)$  is indeed a limit equilibrium.

### B.1.4 Uniqueness

Uniqueness of the limit equilibrium follows immediately: Since the asymptotic independence of adoptions (31) does not rely on the optimality of the cutoffs  $(c_d^*)$ , the same argument implies

$$\lim_{I \rightarrow \infty} Y_{I,d}(c) = \prod_{\theta} (1 - X_{\theta}(c))^{d_{\theta}}$$

for any other cutoffs  $c = (c_d) \neq c^*$ . But as the solution to (18) is unique, we have  $\pi(\prod_{\theta} (1 - X_{\theta}(c))^{d_{\theta}}) \neq c_d$ . Thus,  $\lim_{I \rightarrow \infty} \pi(Y_{I,d}(c)) \neq c_d$ , and so  $c$  is not a limit equilibrium.

## B.2 Limit of Equilibria

Proposition 2 shows that  $(c_d^*)$  are the unique cost-cutoffs that are approximately optimal in  $\mathcal{G}_I$  for all large  $I$ . This is the appropriate equilibrium notion for our macro-economic perspective, whereby the finite agents simplify their problem by treating the economy as infinite, and are vindicated by the fact that their solution to the simplified problem is indeed approximately optimal for large  $I$ .

An alternative “micro-economic” solution concept might assume that agents can somehow overcome the complexity of the finite models  $\mathcal{G}_I$  and play the exact equilibria  $(c_{I,d})$ . The uniqueness of the limit equilibrium suggests that  $(c_{I,d})$  converge to  $(c_d^*)$ .<sup>30</sup> Here, we confirm this conjecture. For notational simplicity we state the proof for a single type  $\theta$ , so the number of outlinks (or degree)  $d$  is an integer rather than a

---

<sup>30</sup>As always, the cost cutoffs also depend on  $t \in [0, 1]$ , which we omit to simplify notation; when talking about convergence, we refer to the topology of uniform convergence in  $t$ .

vector. All told, we need to prove that for all  $d$

$$\lim_{I \rightarrow \infty} c_{I,d} = c_d^*. \quad (33)$$

As a preliminary step, note that in the equilibrium  $c_I = (c_{I,d})$  of  $\mathcal{G}_I$ , social information  $y_{I,d}$  is Lipschitz as a function of  $t$  and so, too, are the cutoffs  $c_{I,d} = \pi(y_{I,d})$ . By the Arzela-Ascoli theorem, the sequence of cutoff vectors  $c_I = (c_{I,d})$  has a subsequence which converges to some  $c_\infty = (c_{\infty,d})$  (pointwise for all  $d$ ). We write  $x_\infty := X(c_\infty)$  for the adoption probabilities associated with this strategy in the branching process, as defined in (30).

Equation (33) now follows from the claim (proved below) that the limit behavior  $c_\infty$  is indeed optimal, given the induced adoption probabilities  $x_\infty$ , i.e.

$$c_{\infty,d} = \pi((1 - x_\infty)^d). \quad (34)$$

Indeed, given (34) we substitute into (30) (for a single type  $\theta$ ) to get

$$\dot{x}_\infty = E [\phi((1 - x_\infty)^D, \pi^{-1}(c_{\infty,D}))] = E [\Phi((1 - x_\infty)^D)].$$

That is,  $x_\infty$  solves (18) and so  $x_\infty = x^*$ . Thus, the limit of the (subsequence of) equilibria in  $\mathcal{G}_I$  is an equilibrium in the branching process, i.e.  $(c_{\infty,d}) = (c_d^*)$ . Since the solution to (18) and the associated cost cutoffs are unique, the entire sequence  $c_I$  (rather than just a subsequence) must converge to  $c_\infty$ , finishing the proof.

*Proof of (34).* By the triangle inequality,

$$|c_{\infty,d} - \pi((1 - x_\infty)^d)| \leq |c_{\infty,d} - c_{I,d}| + |c_{I,d} - \pi((1 - X(c_I))^d)| + |\pi((1 - X(c_I))^d) - \pi((1 - X(c_\infty))^d)|$$

Along the subsequence of  $I$  as  $c_I$  converges to  $c_\infty$ , the first term on the RHS vanishes and so, too, does the third term since the operator  $X$  and the function  $\pi$  are continuous. Turning to the second term, note that the proof of (31) and in particular the upper bound in (32) do not depend on the strategy  $c^*$ , and so implies more strongly that

$$\lim_{I \rightarrow \infty} \sup_c |Y_{I,d}(c) - (1 - X(c))^d| = 0. \quad (35)$$

The equilibrium cost-cutoffs  $c_I = (c_{I,d})$  of  $\mathcal{G}_I$  additionally satisfy  $\pi(Y_{I,d}(c_I)) = c_{I,d}$ , and so

$$\lim_{I \rightarrow \infty} |c_{I,d} - \pi((1 - X(c_I))^d)| = 0.$$

□

### B.3 Undirected, Multi-type Networks

Here we introduce heterogeneous types into the undirected networks of Section 5.2. As in Section 5.1, every agent independently draws a finite type  $\theta$  and then every agent with type  $\theta$  independently draws a vector of link-stubs  $(D_{\theta,\theta'})_{\theta'}$  to agents of type  $\theta'$ . We additionally impose the accounting identity  $\Pr(\theta)E[D_{\theta,\theta'}] = \Pr(\theta')E[D_{\theta',\theta}]$  and an additional independence assumption on  $(D_{\theta,\theta'})_{\theta'}$  across  $\theta'$ . Next, we connect matching link-stubs (i.e. type  $(\theta, \theta')$ -stubs with  $(\theta', \theta)$ -stubs) at random, and finally discard self-links, multi-links, and left-over link-stubs; the accounting identity guarantees that a vanishing proportion of link-stubs are discarded as  $I \rightarrow \infty$ . The additional independence assumption in turn implies that the typical type- $\theta'$  neighbor of a type- $\theta$  agent  $i$  has  $D'_{\theta',\theta}$  links to type- $\theta$  agents (including  $i$ ), and  $D_{\theta',\theta''}$  links to type- $\theta''$  agents for all  $\theta'' \neq \theta$ . That is, the friendship paradox only applies to agent  $i$ 's own type  $\theta$ .

When agent  $i$  enters, write  $\bar{x}_{\theta,\theta'}$  for the probability that her neighbor  $\theta'$  has adopted (conditional on not having observed  $i$  adopt earlier). By the same logic as in the body of the paper, adoption probabilities in the branching process follow

$$\dot{\bar{x}}_{\theta,\theta'} = E \left[ \phi \left( (1 - \bar{x}_{\theta',\theta})^{D'_{\theta',\theta} - 1} \prod_{\theta'' \neq \theta} (1 - \bar{x}_{\theta',\theta''})^{D_{\theta',\theta''}}, (1 - \bar{x}_{\theta',\theta})^{D'_{\theta',\theta}} \prod_{\theta'' \neq \theta} (1 - \bar{x}_{\theta',\theta''})^{D_{\theta',\theta''}} \right) \right] \quad (36)$$

### B.4 Adding Links in Undirected Networks

Here we prove the claim from Section 5.2 that additional links contribute to social learning in undirected networks. As in Theorem 1', given link distributions  $D, \tilde{D}$ , write  $y_d^* = (1 - \bar{x}^*)^d$  and  $\tilde{y}_d^*$  as the corresponding social learning curves. Letting  $\succeq_{LR}$  represent the likelihood ratio order, we then have

**Theorem 1''.** *Assume BHR. Learning and welfare increase with links: If  $\tilde{D} \succeq_{LR} D$*

(a) For any degree  $d$ ,  $1 - \tilde{y}_d^* \geq 1 - y_d^*$ ,

(b) In expectation over the degree,  $E[1 - \tilde{y}_D^*] \geq E[1 - y_D^*]$ .

*Proof.* First observe that  $\tilde{D} \succeq_{LR} D$  implies  $\tilde{D}' \succeq_{LR} D'$  since

$$\frac{\Pr(\tilde{D}' = \bar{d})}{\Pr(\tilde{D}' = \underline{d})} = \frac{\bar{d}}{\underline{d}} \cdot \frac{\Pr(\tilde{D} = \bar{d})}{\Pr(\tilde{D} = \underline{d})} \geq \frac{\bar{d}}{\underline{d}} \cdot \frac{\Pr(D = \bar{d})}{\Pr(D = \underline{d})} = \frac{\Pr(D' = \bar{d})}{\Pr(D' = \underline{d})}.$$

Hence  $\tilde{D}' \succeq_{FOSD} D'$ . Given BHR,  $\phi((1-x)^{d-1}, (1-x)^d) = \alpha(1 - \frac{1}{1-x}[(1-x)^d(1 - F((1-x)^d))])$  rises in  $d$  since the term in square brackets increases in  $(1-x)^d$ . Thus the RHS of (21) FOSD-increases in  $D'$ , and so too does its solution  $\bar{x}^*$  by the Single Crossing Lemma. This implies  $1 - \tilde{y}_d^* \geq 1 - y_d^*$ . Part (b) then follows from the fact that  $E[1 - y_D] = E[1 - (1 - \bar{x})^D]$  increases in  $D$ .  $\square$

## B.5 Proof of Proposition 2'

The proof of Proposition 2' mirrors the proof Proposition 2. Instead of repeating all of the arguments, we only discuss where they need to be adapted.

### B.5.1 Branching Process

Define a two-type branching process with bilateral and triangle types. In the first step, the number of offspring are given by some fixed degree  $(\bar{d}, 2\hat{d})$ . In every subsequent step, the (forward) degree is drawn from  $(\bar{D}' - 1, 2\hat{D})$  for bilateral offspring, and from  $(\bar{D}, 2(\hat{D}' - 1))$  for triangle offspring; all draws are independent and generate distinct nodes, including the draws from two triangle offspring on the same triangle. The resulting (undirected) network consists of the tree generated by the branching process and the links connecting neighbors on any given triangle.

Next, we couple the  $r$ -neighborhoods of agent  $i$  with degree  $(\bar{d}, 2\hat{d})$  in the finite network<sup>31</sup> and the branching process,  $\mathcal{G}_{I,(\bar{d},2\hat{d}),r}$  and  $\mathcal{T}_{(\bar{d},2\hat{d}),r}$ . This is where we have to account for the friendship paradox: When uncovering  $i$ 's neighbor  $j$  on a bilateral link, the probability distribution of  $j$ 's bilateral degree  $\Pr(\bar{D}_j = k)$  must be re-weighted by  $k/E[\bar{D}]$ ; that is, it is drawn from  $D'$ , defined in (20). Finally, one of  $j$ 's  $D'$

<sup>31</sup>In undirected networks, we define an agent's  $r$ -neighborhood of agent  $i$  as all agents and all *undirected* edges that can be reached from agent  $i$  via paths of length at most  $r$ .

bilateral links goes back to  $i$ , and so only  $\bar{D} - 1$  go forward to additional nodes. Since  $j$ 's bilateral and triangle links  $\bar{D}_j, \hat{D}_j$  are independent,  $\hat{D}_j$  simply follows  $\hat{D}$ . All told, conditional on a successful coupling, the ‘‘forward-degree’’ of a bilateral neighbor follows  $(\bar{D} - 1, 2\hat{D})$ . The argument that the degree of a triangle neighbor follows  $(\bar{D}, 2(\hat{D}' - 1))$  is analogous.

### B.5.2 Limit Adoption

Following Section B.1.2, we now characterize neighbors' adoption probabilities in the infinite network generated by the branching process for arbitrary strategies  $c = (c_{\bar{d}, \hat{d}})$ . Indeed, let  $\bar{X}(c), \hat{X}(c)$  be the solution of

$$\dot{\bar{x}} = E \left[ \phi \left( (1 - \bar{x})^{\bar{D}' - 1} (1 - \hat{x})^{2\hat{D}}, \pi^{-1}(c_{\bar{D}', \hat{D}}) \right) \right] \quad (37)$$

$$\dot{\hat{x}} = E \left[ \phi \left( (1 - \bar{x})^{\bar{D}} (1 - \hat{x})^{2(\hat{D}' - 1)}, \pi^{-1}(c_{\bar{D}, \hat{D}'}) \right) \right]. \quad (38)$$

We claim (i) that  $\bar{X}(c)$  is the adoption probability of  $i$ 's bilateral neighbor  $j$  at  $t_i$ , and (ii) that  $\hat{X}(c)$  is the probability that the first-adopter in a triangle  $(i, j, k)$  has adopted at  $t_i$ ; more precisely,  $(1 - \hat{X}(c))^2$  is the probability that neither  $j$  nor  $k$  has adopted at  $t_i$ .

Claim (i) follows by the standard argument that  $j$  has  $D'$  neighbors, but that includes  $i$ , who has not yet adopted at  $t < t_i$ . Claim (ii) is more subtle and follows by the proof of Lemma 2. As in that proof define  $y_\nu$  as the probability that  $\nu \in \{0, 1, 2\}$  of  $i$ 's neighbors  $j, k$  in a given triangle have not yet entered, while the other  $2 - \nu$  have entered but chose not to adopt. In the triangle, when one of the remaining  $\nu$  neighbors enters she observes no adoption and hence adopts herself with probability  $\phi(1, y)$ , where  $y := \sum_{\nu=0}^2 y_\nu$ . Here a triangular neighbor  $j$  has additional  $\bar{D}$  bilateral links and  $\hat{D}' - 1$  additional triangular link pairs, so observes an adoption with probability  $(1 - \bar{x})^{\bar{D}} (1 - \hat{x})^{2(\hat{D}' - 1)}$ , and is assumed to employ the exogenous cutoffs  $c_{\bar{D}, \hat{D}'}$  upon seeing no adoption. Thus, from  $i$ 's perspective,  $j$ 's adoption rate is given by the RHS of (38). Subject to substituting this term for  $\phi(1, y)$ , the proof of Lemma 2 applies as stated, yielding (38).

The solution  $\bar{X}(c), \hat{X}(c)$  of (37-38) nests as a special case the solution  $\bar{x}^* = \bar{X}(c^*), \hat{x}^* = \hat{X}(c^*)$  of (25-26) for cost cutoffs  $c_{\bar{d}, \hat{d}}^* = \pi((1 - \bar{x}^*)^{\bar{d}} (1 - \hat{x}^*)^{2\hat{d}})$ .

### B.5.3 Limit equilibrium, Uniqueness, and Limit of equilibria

The arguments in Sections B.1.3, B.1.4, and B.2 only require adapting the notation. Indeed, write  $Y_{I,\bar{d},\hat{d}}(c)$  for the social learning curve in  $\hat{\mathcal{G}}_I$  when agents use cutoffs  $c = (c_{\bar{d},\hat{d}})$ , and  $y_{I,\bar{d},\hat{d}}^* = Y_{I,\bar{d},\hat{d}}(c^*)$ . Then, at the most general level, (35) generalizes to

$$\lim_{I \rightarrow \infty} \sup_c |Y_{I,\bar{d},\hat{d}}(c) - (1 - \bar{X}(c))^{\bar{d}}(1 - \hat{X}(c))^{2\hat{d}}| = 0. \quad (39)$$

The fact that  $c^*$  is a limit equilibrium then follows by substituting  $c^*$  into (39) and recalling that  $c_{\bar{d},\hat{d}}^* = \pi((1 - \bar{x}^*)^{\bar{d}}(1 - \hat{x}^*)^{2\hat{d}})$ .

Uniqueness follows by substituting any other cutoffs  $c \neq c^*$  into (39) and noting that  $c_{\bar{d},\hat{d}} \neq \pi((1 - \bar{X}(c))^{\bar{d}}(1 - \hat{X}(c))^{2\hat{d}})$ .

Finally, the exact equilibria  $c_I = (c_{I,\bar{d},\hat{d}})$  of  $\hat{\mathcal{G}}_I$  converge to the limit equilibrium  $c_{\bar{d},\hat{d}}^*$  by the same reasoning as in Appendix B.2, invoking Arzela-Ascoli to obtain a convergent subsequence of  $c_I$  and then leveraging (39) to show that its limit must equal  $c^*$ .

## C Imperfect News

Throughout this paper we have assumed that agents never adopt the low-quality product,  $q = L$ . Observing an adoption is thus “perfect good news”, causing the agent’s belief  $\pi = \Pr(q = H)$  to jump to 1. This allows us to summarize  $i$ ’s social information (at time- $t$ ) by a single number, namely the probability that none of her neighbors has adopted the product  $y_i$ .

In this appendix we consider “imperfect news”. Agents must then keep track of how many, and which of their neighbors have adopted. Much of the “perfect news” analysis in the body of the paper carries over. In particular, social information improves with additional independent links (Theorem 1) but deteriorates when adding a backward link (Section 4.3). But we qualify the finding of Section 4.2 with an example where a correlating link improves social learning, in line with Centola (2010)’s experimental findings.

Formally, the structure of the network and timing of the game is the same as in Section 2. The only difference is that, upon inspection, agents adopt the quality- $q$

product with probability  $\alpha^q$  where  $0 < \alpha^L < \alpha^H \leq 1$ . For simplicity suppose agents know the network  $G$ , as in Section 4. To simplify further, our results will assume an even prior,  $\pi_0 = 1/2$ .

## C.1 Social Information and Adoption in General Networks

With imperfect news, the main complicating factor is that we must keep track of the adoption of all of  $i$ 's neighbors, and not simply whether or not there has been at least one adoption. Let  $x_{A,i}^q$  be the probability that exactly  $A \subseteq N_i$  of  $i$ 's neighbors have adopted a quality- $q$  product before  $t_i$ . The posterior probability of high quality in this event equals

$$\pi_{A,i} = \frac{x_{A,i}^H \pi_0}{x_{A,i}^H \pi_0 + x_{A,i}^L (1 - \pi_0)} = \pi(x_{A,i}^H / x_{A,i}^L). \quad (40)$$

Thus  $i$ 's adoption probability follows

$$\dot{x}_i^q = \alpha^q \sum_{A \subseteq N_i} x_{A,i}^q F(\pi_{A,i}). \quad (41)$$

Unpacking (41),  $x_{A,i}^q$  is the probability that neighbors  $A$  adopt,  $F(\pi_{A,i})$  is the probability that  $i$  inspects given that  $A$  adopt, and  $\alpha^q$  is the probability that  $i$  adopts given inspection.<sup>32</sup>

With perfect good news,  $i$ 's learning curve  $1 - y_i$  is Blackwell-sufficient for her social information. Here we capture  $i$ 's information by a more general “experiment”  $X_i = (x_{A,i}^q)_{q,A}$  and denote  $i$ 's random posterior by  $\Pi_i$ : Specifically, the outcome of  $i$ 's experiment is the set of adopting neighbors  $A$ , which induces posterior  $\pi_{A,i}$  and occurs with unconditional probability  $x_{A,i} := \pi_0(x_{A,i}^H + x_{A,i}^L)$ . Blackwell-sufficiency of experiments  $X_i, \succeq_{BW}$ , is characterized by a mean-preserving spread of  $\Pi_i$  (e.g. Börgers (2009)).<sup>33</sup>

As in Section 4.1, the key argument is that an increase in agent  $i$ 's social infor-

<sup>32</sup>With perfect good news,  $N_i = \{j\}$  and  $A = \emptyset$ , we have  $x_{A,i}^H = 1 - x_j$  and  $x_{A,i}^L = 1$ . Thus, (40) reduces to (2), while (41) reduces to (3) for  $q = H$ .

<sup>33</sup>With perfect good news we have  $x_{\emptyset,i}^H = y_i$  and  $x_{\emptyset,i}^L = 1$ . Thus, the random posterior  $\Pi_i$  equals  $\pi_{A,i} = 1$  for  $A \neq \emptyset$ , which happens with probability  $\pi_0(1 - y_i)$ , and  $\pi_{\emptyset,i} = \pi(y_i)$  with the residual probability. Since  $\pi(\cdot)$  is an increasing function,  $\Pi_i$  is decreasing in  $y_i$  in the order  $\succeq_{BW}$ .



mation renders her own adoption more informative. To establish this, we strengthen the BHR condition and assume that

$$\kappa F(\kappa) \text{ is convex and } (1 - \kappa)F(\kappa) \text{ is concave.} \quad (42)$$

which is satisfied if  $\kappa \sim U[0, 1]$ .

**Lemma 1'.** *Assume  $\pi_0 = 1/2$  and (42) holds. If  $i$ 's social information  $X_i$  Blackwell-improves, her adoption rate  $\dot{x}_i^q$  rises for  $q = H$  and falls for  $q = L$ .*

*Proof.* Given  $\pi_0 = 1/2$ , equation (40) implies that  $x_{A,i}^H = 2\pi_{A,i}x_{A,i}$  and  $x_{A,i}^L = 2(1 - \pi_{A,i})x_{A,i}$  with the unconditional probability  $x_{A,i} = (x_{A,i}^H + x_{A,i}^L)/2$ . Thus, we can rewrite (41) as

$$\dot{x}_i^H = 2\alpha^H \sum_{A \subseteq N_i} x_{A,i} [\pi_{A,i} F(\pi_{A,i})] = 2\alpha^H E_i[\Pi_i F(\Pi_i)]. \quad (43)$$

$$\dot{x}_i^L = 2\alpha^L \sum_{A \subseteq N_i} x_{A,i} [(1 - \pi_{A,i}) F(\pi_{A,i})] = 2\alpha^L E_i[(1 - \Pi_i) F(\Pi_i)]. \quad (44)$$

By (42), a mean-preserving spread of  $\Pi_i$  raises the RHS of (43) and lowers the RHS of (44).  $\square$

## C.2 Trees

On tree networks, all of  $i$ 's neighbors act independently, so we can close the analysis of  $i$ 's adoption (41) by computing the probability  $A$  adopt as

$$x_{A,i}^q = \prod_{j \in A} x_j^q \prod_{j \in N_i \setminus A} (1 - x_j^q).$$

We can now compare a large tree  $\tilde{G}$  to a small tree  $G$ , and show that all agents in  $G$  receive Blackwell-superior information in  $\tilde{G}$ .

**Theorem 1'''.** *Consider trees  $\tilde{G} \supseteq G$  and assume  $\pi_0 = 1/2$  and (42) holds. For any  $i$  in  $G$ , social learning is Blackwell-superior in  $\tilde{G}$ , i.e.  $\tilde{X}_i \succeq_{BW} X_i$ .*

*Proof.* As in the proof of Theorem 1 we proceed by ‘‘induction down the tree’’, starting at the leaves. If  $i$  is a leaf of  $G$  she has no information in  $G$ , and so trivially  $\tilde{X}_i \succeq_{BW} X_i$

$X_i$ . Continuing by induction, consider some agent  $i$  with neighbors  $\tilde{N}_i$  in the large tree and  $N_i$  in the small tree, and assume  $\tilde{X}_j \succeq_{BW} X_j$  for all  $j \in N_i$ . Using Lemma 1' these neighbors adopt more in the high state and less in the low state,

$$\tilde{x}_j^H \geq x_j^H \text{ and } \tilde{x}_j^L \leq x_j^L. \quad (45)$$

We claim that  $(\tilde{x}_j^H, \tilde{x}_j^L)$  is Blackwell-superior to  $(x_j^H, x_j^L)$ . With a single neighbor,  $N_i = \{j\}$ , the relevant subsets are  $A = \emptyset$  or  $A = \{j\}$ , so we need to show  $\tilde{\pi}_{\{j\},i} \geq \pi_{\{j\},i}$  and  $\tilde{\pi}_{\emptyset,i} \leq \pi_{\emptyset,i}$ . Using (40), we can rewrite this in terms of likelihood ratios,

$$\frac{\tilde{x}_j^H}{\tilde{x}_j^L} \geq \frac{x_j^H}{x_j^L} \quad \text{and} \quad \frac{1 - \tilde{x}_j^H}{1 - \tilde{x}_j^L} \leq \frac{1 - x_j^H}{1 - x_j^L}. \quad (46)$$

Equation (45) implies (46), so adoption by each neighbor  $j$  is more informative in the larger network. Since neighbors are independent and  $N_i \subseteq \tilde{N}_i$ , adoption  $(\tilde{x}_j^H, \tilde{x}_j^L)_{j \in \tilde{N}_i}$  is Blackwell-superior to  $(x_j^H, x_j^L)_{j \in N_i}$ , completing the induction step.  $\square$

In Section 4.1 we argued that social information increases with the adoption probability  $\alpha$ . Here we generalize and extend this finding.

**Claim.** *Assume  $\pi_0 = 1/2$  and (42) holds. Agent  $i$ 's social information rises if*

- (a)  $\alpha^H$  rises and  $\alpha^L$  falls, or
- (b) both  $(\alpha^H, \alpha^L)$  rise and  $\alpha^H/\alpha^L$  stays constant.

Part (a) reflects the idea that the more people adopt for idiosyncratic reasons, the harder it is to learn about common quality. Part (b) reflects the idea that lowering  $(\alpha^H, \alpha^L)$  while fixing  $\alpha^H/\alpha^L$  amounts to losing the signal that  $i$  adopted with some probability, leading to a Blackwell-decrease in information.<sup>34</sup>

*Proof.* As in the proof of Theorem 1''', we argue by induction down the tree. The claim trivially holds true for leaves  $i$ , which have no social information. For the induction step, it suffices to argue that a change of  $\alpha^q$  as in (a) or (b) together with a rise in agent  $i$ 's social information increases the informativeness of  $i$ 's own adoption, i.e. it raises  $x_i^H/x_i^L$  and lowers  $(1 - x_i^H)/(1 - x_i^L)$ , cf. (46).

---

<sup>34</sup>In the baseline model, increasing  $\alpha = \alpha^H$  while fixing  $\alpha^L = 0$  is a special case of both part (a) and part (b).

Case (a) is immediate: The RHS of (43) rises (because  $\alpha^H$  rises and  $i$ 's information rises) and the RHS of (44) falls, raising  $x_i^H$  and lowering  $x_i^L$ . Case (b) follows because the ratio of the RHS of (43) and (44) rises (since  $\alpha^H/\alpha^L$  is constant and  $i$ 's information improves) and so  $x_i^H/x_i^L$  rises and observing an adoption is as informative as before. Since  $x_i^H$  rises, too,  $(1 - x_i^H)/(1 - x_i^L) = (1/x_i^H - 1)/(1/x_i^H - x_i^L/x_i^H)$  falls because it is a decreasing function in  $x_i^H$  and  $x_i^H/x_i^L$ ; thus, observing no adoption becomes more informative.  $\square$

### C.3 Backward Links

As in the baseline model, backward links continue to lower  $i$ 's learning. To see this, compare the directed and undirected pair in Section 4.3. With  $i \rightarrow j$ , agent  $j$  has no information and adoption curve

$$\dot{x}_j^q = \alpha^q F(\pi_0). \quad (47)$$

With  $i \leftrightarrow j$ , agent  $j$  cannot have seen  $i$  adopt at time  $t \leq t_i$ , making  $j$  more pessimistic and reducing his inspection probability. To see this, define  $\bar{x}^q$  to be  $j$ 's adoption probability given quality  $q$  and conditional on  $t \leq t_i$ . Analogous to (21),

$$\dot{\bar{x}}^q = \alpha^q F\left(\pi\left(\frac{1 - \bar{x}^H}{1 - \bar{x}^L}\right)\right). \quad (48)$$

We claim that agent  $i$  has Blackwell-more information in the directed pair than in the undirected pair. Indeed,  $\bar{x}^q = \zeta x_j^q$  for  $\zeta := F\left(\pi\left(\frac{1 - \bar{x}^H}{1 - \bar{x}^L}\right)\right) / F(\pi_0) < 1$ . That is, observing  $j$  in the undirected pair is like observing him in the directed pair losing the signal with probability  $1 - \zeta$ , and hence Blackwell-inferior.

### C.4 Correlating Links

In the baseline model, correlating links are detrimental to social information. Intuitively, if  $i$ 's neighbors observe each other, then any given neighbor is less likely to adopt when the others fail to adopt. Thus, correlating links lower the probability that at least one of  $i$ 's neighbors adopts, and thereby lowers  $i$ 's social information. Central to this argument is the notion that one adoption is enough. In contrast, with

imperfect information one might imagine that correlating links reinforce a message and provide useful information to a skeptical agent. For example, Centola's (2010) experiment shows the importance of strong links in social learning, arguing that it takes multiple positive signals to sway an agent's decision.

We now construct an example in which correlating links improve welfare with imperfect news. As in Section 4.2, suppose agent  $i$  initially observes agents  $j, k$ , and consider the effect of an additional  $j \rightarrow k$ . Suppose the agent's cost is equally likely to be low or high,  $\kappa \in \{0, \bar{\kappa}\}$ , with  $\bar{\kappa} = \pi(\alpha^H/\alpha^L)$ , and recall that the agent inspects if indifferent. The low-cost agent always inspects, and information has no value to her. The high-cost agent only benefits from social information if it pushes her posterior strictly above  $\bar{\kappa}$ , which only ever happens in either network when  $i$  observes both  $j$  and  $k$  adopt. Thus, we need to compare the (unconditional) probability of this event  $x_{\{j,k\},i} = \pi_0 x_{\{j,k\},i}^H + (1 - \pi_0) x_{\{j,k\},i}^L$  and the induced posterior belief  $\pi_{\{j,k\},i}$  across the two networks. Given  $\pi_0 = 1/2$ , the value of information then equals  $\frac{1}{2} x_{\{j,k\},i} (\pi_{\{j,k\},i} - \bar{\kappa})$ .

In either network,  $k$  has no information and only her low-cost type inspects; the probability that she adopts good  $q$  by time  $t$  thus equals  $\frac{1}{2} \alpha^q t$ . Without the correlating link, by symmetry and independence of  $j$  and  $k$ , the probability they both adopt product  $q$  equals  $x_{\{j,k\},i}^q = (\frac{1}{2} \alpha^q t)^2$  and the posterior belief equals  $\pi_{\{j,k\},i} = \pi((\alpha^H/\alpha^L)^2)$ .

The correlating link raises  $j$ 's adoption probability conditional on  $k$  having adopted: If  $j$  enters first, only low-cost  $j$  adopts and the probability is unchanged; but if  $k$  enters first,  $j$  inspects with certainty. Thus, the joint probability that both  $j$  and  $k$  adopt rises to  $\hat{x}_{\{j,k\},i}^q = \frac{1}{2} \alpha^q t \cdot \frac{3}{4} \alpha^q t$ , while the posterior belief is unchanged, equal to  $\hat{\pi}_{\{j,k\},i} = \pi((\alpha^H/\alpha^L)^2)$ . All told, the correlating link  $j \rightarrow k$  is valuable to  $i$  because it increases the probability of observing  $j$  adopt in the event that  $k$  also adopts, which is precisely when this information is most valuable to  $i$ . Note the contrast to Section 4.2, where  $j$ 's adoption was valuable to  $i$  in the complementary event, when  $k$  had not adopted.