

Learning Dynamics in Social Networks*

Simon Board[†] and Moritz Meyer-ter-Vehn[‡]

September 4, 2018

Abstract

This paper proposes a tractable model of Bayesian learning on social networks in which agents choose whether to adopt an innovation. We study the impact of network structure on learning dynamics and diffusion. In tree networks, we provide conditions under which all direct and indirect links contribute to an agent’s learning. Beyond trees, not all links are beneficial: An agent’s learning deteriorates when her neighbors are linked to each other, and when her neighbors learn from herself. These results imply that an agent’s favorite network is the directed star with herself at the center, and that learning is better in “decentralized” networks than “centralized” networks.

1 Introduction

How do groups of friends, organizations, or entire societies learn about innovations? Consider consumers learning about a new brand of electric car from their 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. Does an agent benefit from a more highly connected network? Do agents learn more in centralized networks or in decentralized networks? What network maximizes an agent’s information?

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 chooses whether to adopt 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, as the number of agents grows large. Understanding the full dynamics is important because empirical researchers

*We have received useful comments from Ben Golub, Alessandro Pavan, Klaus Schmidt, Bruno Strulovici, and Fernando Vega-Redondo. We also thank seminar audiences at Bar-Ilan, Bocconi, CEMFI, Chicago Theory Conference, Hebrew, HKU, Indiana, Michigan, Munich, Northwestern, NYU, Penn State, Pittsburgh, Princeton, Seoul National, Tel-Aviv, Toronto, UC3M, UCLA, UCSD, Utah. Keywords: networks, diffusion, social learning. JEL codes: D83, D85.

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

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

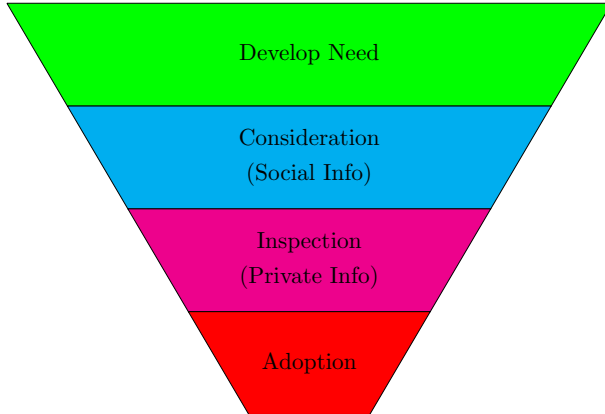


Figure 1: **The Social Purchasing Funnel.**

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. We thus recover the tractability of the reduced-form models of diffusion (e.g. Bass, 1969) in a model of Bayesian learning.

Our main results describe how learning dynamics depend on the network structure. Starting with a tree network, we show that an agent typically benefits when her neighbors have more links; but beyond tree networks, backward and correlating links can muddle her learning. This implies that an agent’s learning is maximized by a star network with herself at the center, and that “decentralized” networks are superior to “centralized” ones. These results help us understand how diffusion changes as social media raises interconnectedness, or when comparing the residents in a loosely connected city to those in a tightly connected village. They also inform the impact of interventions that form new social links to spread ideas (e.g. Cai and Szeidl, 2018).

In the model, we assume that agents consider a single innovation/product whose quality is high or low; inspection reveals this common quality and agents’ idiosyncratic preferences for the product. We additionally assume that an agent only adopts a product if it is high quality. Thus, when an agent sees a neighbor adopt, she knows that quality is high, and so inspects herself. Conversely, if an agent observes none of her neighbors adopt the product, she must infer whether (i) they have yet to develop a need for the product, (ii) they developed a need, but chose not to inspect, (iii) they found the quality to be high, but did not adopt for idiosyncratic reasons, or (iv) they found the quality to be low. The agent’s inspection decision is thus based on the hypothesized inspection decisions of her neighbors, which collectively generate her *social learning curve* (formally, the probability at least one of her neighbors adopts as a function of time). In turn, her adoption decision feeds into the social learning curves of her neighbors.

Our first main result characterizes 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. This motivates us to study directed tree networks, in which an agent’s neighbors receive independent information, as in a large random network with finite degree. In such trees, it is sufficient to keep track of individual adoption decisions, meaning the system reduces to I dimensions. Moreover, in a regular network, where everyone has the same number of neighbors, behavior is described by a one-dimensional differential equation.

We then use these differential equations to study adoption and learning rates across networks. First we show that, given a mild condition on the hazard rate of inspection costs, an agent’s adoption rate is increasing in her level of social information. In the context of tree networks, this implies inductively that an agent’s adoption rate increases if she adds an extra neighbor, if her neighbors’ add extra neighbors, and so on. Moreover, these direct and indirect links all raise the agent’s utility. We can even compare the value of direct and indirect links: for example, an agent prefers two direct neighbors to an infinite chain where everyone has one neighbor.

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 develops the need for the innovation; prior to this time, j could not have seen i adopt, and so the backward link makes j more pessimistic and lowers his adoption. All told, agent i ’s favorite network is the directed star with herself at the center. These results also imply that agents would prefer to have d links in a “decentralized” random graph than in a “centralized”, complete network of $d + 1$ agents.

1.1 Literature

The literature on observational learning originates with the classic papers of Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (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.

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 (2018) 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.^{1,2} A number of papers have analyzed related problems in complete networks. Guarino, Harmgart, and Huck (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, Sorensen, and Wiseman (2012) suppose an agent knows the order in which others move, but only sees

¹There is a wider literature on diffusion without learning. There are “awareness” models in which an agent becomes aware of the product when her neighbors adopt it. One can view Bass (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 (2018) puts these forces together, and studies diffusion of a network good where agents become aware from her neighbors.

²The inference problem is also related to a broader set of games in which agents move at privately known times (e.g. Kamada and Moroni, 2018).

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 3.2 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 varies with the form of the network, and characterize her preferences over the network structure.

2 Model

Network. A finite set of I agents is connected via a commonly known, exogenous, directed network G that represents which agents observe the actions of others.³ If i observes j we write $i \rightarrow j$ or $(i, j) \in G$, say i is *linked to* j and call j a *neighbor* of i . We denote the set of i ’s neighbors by N_i . Agent j is *successor* of i if there exists a path $i \rightarrow \dots \rightarrow j$.

States. The agents seek to learn about the quality of a single product of quality $\theta \in \{L, H\} = \{0, 1\}$. Time is continuous, $t \in [0, 1]$. At time $t = 0$, agents share a common prior $\Pr(\theta = H) = \pi_0 \in (0, 1)$.

Game. Agent i develops a need for the product, or *enters*, at privately observed time $t_i \sim U[0, 1]$.⁴ She observes which of her neighbors have adopted the product by time t_i and updates her belief about the quality of the product. The agent then chooses whether or not to inspect the product at cost c_i , with cdf F and bounded pdf f . Finally, if she inspects the product, she adopts it with probability $\alpha \in (0, 1]$ if $\theta = H$ and probability 0 if $\theta = L$. The parameter α captures taste heterogeneity across agents. Agents receive utility $1/\alpha$ from adopting a product, so taking expectations over idiosyncratic tastes, they have expected utility one from a high-quality product, and expected utility zero from a low-quality product. Entry times t_i , inspection costs c_i , and idiosyncratic tastes are independent within agents and iid across agents.

Remarks. The model makes several assumptions of note. First, 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

³When analyzing regular tree networks in Section 4, we also allow for a countably infinite number of agents.

⁴The uniform distribution is a normalization: t_i should not be interpreted as calendar time, but rather as time-quantile in the product life-cycle.

traditional observational learning models, and seems reasonable in several types of applications. An agent may not know the people she observes (e.g. she sees people on the street wear a new fashion item). The decisions may have small stakes (e.g. if she sees friends' social media posts about a new movie, she is unlikely to phone for further details). Or, there may be many possible alternatives (e.g. if a friend buys a new car, she is unlikely to enquire about all the other cars they chose not to buy).

Second, we assume that the agent only purchases the product if it is high quality. This simplifies the analysis since we need only keep track of whether an agent has seen at least one adoption. In ongoing work we consider a model variant in which agents adopt low-quality products with positive probability.

Third, 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 with Bayes' rule.

Finally, 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 in the spirit of traditional herding models rather than timing games such as Gul and Lundholm (1995).

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 .⁵ Since he enters uniformly over $t \in [0, 1]$, $\dot{x}_{j,t}$ equals the probability he adopts conditional on waking up at time t . This is given by

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

Given his prior π_0 , John's expected utility from inspecting the good is $\pi_0 - c$. He thus inspects with probability $F(\pi_0)$ and adopts with probability $\alpha F(\pi_0)$.

Now consider Iris. She learns by observing whether John has adopted. We thus interpret $x_{j,t}$ as Iris's social learning curve. Her adoption rate is given by

$$\begin{aligned} \dot{x}_{i,t} &= \Pr(i \text{ adopt}) = \alpha \Pr(i \text{ inspect}) = \alpha [1 - \Pr(i \text{ not inspect})] \\ &= \alpha [1 - \Pr(j \text{ not adopt}) \times \Pr(i \text{ not inspect} | j \text{ not adopt})] \\ &= \alpha \left[1 - (1 - x_{j,t})(1 - F(\pi_t^\theta)) \right], \end{aligned}$$

where Iris's posterior that the quality is high given John has not adopted is given by Bayes' rule,

$$\pi_t^\theta := \pi^\theta (1 - x_{j,t}) := \frac{(1 - x_{j,t})\pi_0}{(1 - x_{j,t})\pi_0 + (1 - \pi_0)}. \quad (2)$$

⁵Since no agent adopts when $\theta = L$, it suffices to keep track of the adoption probability conditional on $\theta = H$.

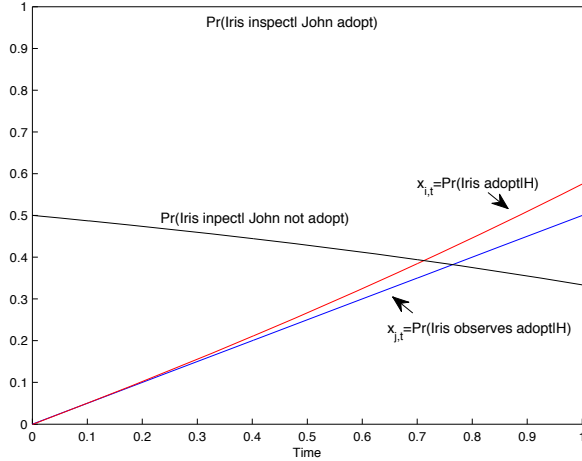


Figure 2: **Iris’s Social Learning and Adoption Curves from Example 1.** This figure assumes inspection costs are $c \sim U[0, 1]$, high-quality goods are adopted with probability $\alpha = 1$, and the prior is $\pi_0 = 1/2$.

Writing $\tilde{F}(1 - x_{j,t}) := F(\pi_t^\emptyset)$,⁶ Iris’s adoption curve becomes

$$\dot{x}_{i,t} = \alpha \left[1 - (1 - x_{j,t})(1 - \tilde{F}(1 - x_{j,t})) \right]. \quad (3)$$

This equation plays a central role throughout the paper. Figure 2 illustrates Iris and John’s adoption curves, as well as Iris’s choices. One sees that conditional on seeing John fail to adopt, Iris becomes more pessimistic over time, and her adoption probability falls. Overall, John provides useful information to Iris, and her adoption curve lies above his. \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_t . Analogous to Iris’s adoption in equation (3), this is governed by the ODE

$$\dot{x}_t = \alpha \left[1 - (1 - x_t)(1 - \tilde{F}(1 - x_t)) \right]. \quad (4)$$

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 only observe the previous agent (Celen and Kariy, 2004). One can show that the adoption curve is convex, lying above $x_{i,t}$ in Figure 2, meaning an agent’s adoption increases when her neighbor observes more information. We generalize this observation in Theorem 3. \triangle

3 General Networks

Consider an arbitrary, finite directed network G and suppose agent i enters at time t_i . Let $x_{N_i,t}^{-i}$ be the probability that one (or more) of i ’s neighbors adopts by time $t \leq t_i$. This acts as a summary

⁶We write the argument of \tilde{F} as $(1 - x)$ since the likelihood ratio given Iris sees “no adoption” is $(1 - x)/1$. We generalize this approach in Section 3.

statistic of other agents' adoption decisions, so we call it i 's *social learning curve*. As in equation (3), i 's adoption is governed by

$$\dot{x}_{i,t} = \alpha \left[1 - (1 - x_{N_i,t}^{-i})(1 - \tilde{F}(1 - x_{N_i,t}^{-i})) \right]. \quad (5)$$

Inspired by the examples in Section 2, one might hope to derive agent i 's social learning curve $x_{N_i,t}^{-i}$ from the adoption curves of her neighbors, $x_{j,t}$ for $j \in N_i$. Unfortunately, this is not possible in general.

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. We can determine k 's adoption curve, $x_{k,t}$, using equation (1) and j 's adoption curve, $x_{j,t}$, using equation (3). Agent i 's inspection decision in turn depends on the probability j or k adopts (or both). Since j observes k , their adoption decisions are correlated and it is not enough to keep track of the marginal adoption probabilities, $x_{j,t}$ and $x_{k,t}$; rather we must keep track of the joint distribution.

The second reason is the *self-reflection* problem. To illustrate, suppose i and j observe each other. 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 curve $x_{j,t}^{-i}$ on the event that i has not yet adopted, which differs from j 's actual, i.e. unconditional, adoption curve $x_{j,t}$. We return to these issues in Section 5.

Despite the problems of correlation and self-reflection, we can still show:

Theorem 1. *In any network G , there exists a unique equilibrium.*

Proof. We establish Theorem 1 by characterizing 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. In state λ , agent i 's information set is given by $\Lambda(i, \lambda) = \{\lambda' : \lambda'_i = \lambda_i, \lambda'_j = a \text{ iff } \lambda'_j = a \text{ for all } j \in N_i\}$; namely, i knows her own state and whether (or not) her neighbors have adopted. We can then describe the distribution over states at time t conditional on quality θ by $z = (z_{\lambda,t}^\theta)$, and the probability of sets of states Λ by $z_{\Lambda,t}^\theta := \sum_{\lambda \in \Lambda} z_{\lambda,t}^\theta$. Given state λ , let λ^{-i} denote the same state with $\lambda_i = \emptyset$.

Figure 3 illustrates the evolution of the state via a Markov chain. Suppose there are three agents and $\lambda = (\lambda_i, \lambda_j, \lambda_k) = (\emptyset, a, b)$. Then probability mass moves into state λ from states λ^{-j} as agent j enters and adopts, and from λ^{-k} as agent k enters and doesn't adopt. Similarly, probability mass moves out of state λ , and into states (a, a, b) and (b, a, b) , as agent i enters.

To quantify these effects, suppose quality is high. Agent i enters uniformly over time $[t, 1]$, meaning probability mass escapes at flow rate $z_t^H/(1-t)$. Similarly, in state λ^{-i} , agent i enters uniformly over time $[t, 1]$, compares the likelihood ratio of the quality to the cost, and thus adopts with probability $\alpha \tilde{F} \left(\frac{z_{\Lambda(i, \lambda^{-i}), t}^H}{z_{\Lambda(i, \lambda^{-i}), t}^L} \right)$; this inflow is then weighted by the mass $z_{\lambda^{-i}, t}^H$ in state λ^{-i} . The equilibrium distribution over the states λ thus evolves according to the ODE

$$\dot{z}_{\lambda,t}^H = -\frac{1}{1-t} \sum_{i:\lambda_i=\emptyset} z_{\lambda,t}^H + \frac{1}{1-t} \sum_{i:\lambda_i=a} z_{\lambda^{-i},t}^H \alpha \tilde{F} \left(\frac{z_{\Lambda(i, \lambda^{-i}), t}^H}{z_{\Lambda(i, \lambda^{-i}), t}^L} \right) + \frac{1}{1-t} \sum_{i:\lambda_i=b} z_{\lambda^{-i},t}^H \left[1 - \alpha \tilde{F} \left(\frac{z_{\Lambda(i, \lambda^{-i}), t}^H}{z_{\Lambda(i, \lambda^{-i}), t}^L} \right) \right] \quad (6)$$

with initial condition $z_{\lambda,0}^H = 1$ if $\lambda_i = \emptyset$ for all i , and 0 otherwise. When quality is low, the calculation

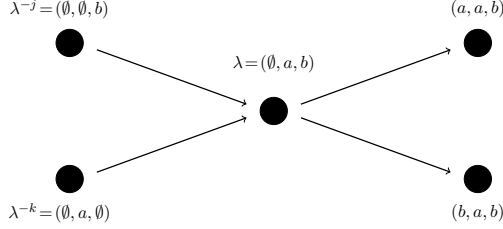


Figure 3: Illustrative Markov Transitions underlying Theorem 1.

is easier since no agents ever adopt, and the state distribution is determined solely by the exogenous entry process:

$$z_{\lambda,t}^L = (1-t)^{|\{i:\lambda_i=\emptyset\}|} t^{|\{i:\lambda_i=b\}|} 0^{|\{i:\lambda_i=a\}|}. \quad (7)$$

To establish existence of a unique equilibrium via the Picard-Lindelöf theorem, we need to argue that the RHS of (6) is Lipschitz continuous in z . In states λ where at least one of i 's neighbors has adopted, we have $z_{\Lambda(i,\lambda^{-i}),t}^L = 0$, meaning that the likelihood ratio is infinite and the inspection probability is $\tilde{F}\left(\frac{z_{\Lambda(i,\lambda^{-i}),t}^H}{z_{\Lambda(i,\lambda^{-i}),t}^L}\right) = 1$. In states λ where none of i 's neighbors has adopted, $\lambda_j \neq a$ for all $j \in N_i$, the denominator in (6) is constant, equal to $z_{\Lambda(i,\lambda^{-i}),t}^L = 1$, and the inspection probability simplifies to $\tilde{F}\left(z_{\Lambda(i,\lambda^{-i}),t}^H\right)$. Since the density $f(c)$ is bounded, one can verify that $\tilde{F}(z)$ is Lipschitz, as is the RHS of equation (6). \square

The system of ODEs (6) implies equilibrium existence and uniqueness, but is less useful as a tool to compute equilibrium numerically since the state is $3^{|I|}$ dimensional. For this reason we impose more structure on networks in the following sections, where we can provide simple formulas for diffusion that can be easily computed.

3.1 Social Learning and Adoption Curves

Having established existence and uniqueness in Theorem 1, we return to equation (5) to analyze how an agent's learning and adoption evolves over time. We will also use these tools to perform comparative statics on networks in later sections.

Recall that $x_{N_i,t}^{-i}$ is the probability that one (or more) of i 's neighbors adopt the high-quality product by time $t \leq t_i$. We can illustrate i 's information structure by the following table:

	≥ 1 adopt	0 adopt
$\theta = H$	$x_{N_i,t}^{-i}$	$1 - x_{N_i,t}^{-i}$
$\theta = L$	0	1

If $x_{N_i,t}^{-i} = 1$, then agent i has perfect information about the state, whereas if $x_{N_i,t}^{-i} = x$ it is as if the signal has been lost with probability $1 - x$. It follows that an increase in $x_{N_i,t}^{-i}$ Blackwell-improves agent i 's information and thereby increases her expected utility. This implies:

Lemma 1. *In any network, any agent's information Blackwell-improves over time.*

Proof. Adoption is irreversible, so the number of agents who have adopted the product increases over time. Hence, if we consider agent i , the probability that one (or more) of her neighbors adopt, $x_{N_i,t}^{-i}$, also increases over time, and her information Blackwell-improves. \square

Lemma 1 implies that agent i prefers to move later since the adoption of others (or lack thereof) provides useful information.

Next, we study the effect of information on adoption.

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

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

Lemma 2. *If F satisfies BHR then i 's adoption rate $\dot{x}_{i,t}$ increases in her information $x_{N_i,t}^{-i}$. Thus, if i 's information $x_{N_i,t}^{-i}$ increases for all t , then so does her adoption $x_{i,t}$.*

Under BHR, Lemmas 1 and 2 thus imply that i 's adoption rate $\dot{x}_{i,t}$ rises over time. That is, her adoption curve $x_{i,t}$ is convex in t , as shown in Figure 2. This result partly reflects our normalization that entry times are uniform on $[0, 1]$; if 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).

Proof. Rewriting the ODE (5) as $\dot{x}_{i,t} = \alpha\phi(x_{N_i,t}^{-i})$, we differentiate to obtain

$$\begin{aligned} \phi'(x) &= (1 - F(\pi^\emptyset(1 - x))) - (1 - x) \cdot \nabla\pi^\emptyset(1 - x) \cdot f(\pi^\emptyset(1 - x)) \\ &= (1 - F(\pi^\emptyset(1 - x))) - \pi^\emptyset(1 - x) \cdot (1 - \pi^\emptyset(1 - x)) \cdot f(\pi^\emptyset(1 - x)), \end{aligned}$$

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

$$(1 - x) \cdot \nabla\pi^\emptyset(1 - x) = (1 - x) \frac{\pi_0(1 - \pi_0)}{[1 - x\pi_0]^2} = \frac{(1 - x)\pi_0}{1 - x\pi_0} \frac{1 - \pi_0}{1 - x\pi_0} = \pi^\emptyset(1 - x) \cdot (1 - \pi^\emptyset(1 - x)).$$

Since $\pi^\emptyset(1 - x)$ is the belief after seeing ‘‘no adoption’’, it has range $[0, \pi_0]$. Thus, BHR implies that $\phi'(x) \geq 0$ for all $x \leq \pi_0$. Thus, better information $x_{N_i,t}^{-i}$ means higher slope $\dot{x}_{i,t}$ and level $x_{i,t}$. \square

For an intuition, recall that adoption probabilities x are conditional on high quality $\theta = H$. The expected value of the posterior $E[\pi_t|H]$ thus exceeds the prior π_0 and the adoption probability tends to increase in the amount of social information $x_{N_i,t}^{-i}$. But since the posterior after ‘‘no adoption’’ π_t^\emptyset falls in $x_{N_i,t}^{-i}$, this intuition does not apply to all cost distributions, and Lemma 2 requires assumption BHR. Formally, BHR is satisfied if f is weakly increasing;⁷ this includes $c \sim U[0, 1]$ as a special case. For other densities f , BHR is automatically satisfied when $c \approx 0$ since the RHS increases to infinity. For higher costs, BHR states that the density does not decrease too quickly. In particular, BHR holds with equality if $f(c) \propto 1/c^2$, meaning that it is satisfied if $d \log f(c)/dc \geq -2/c$.

Example 3 (Information can lower adoption). Assume the distribution $F(c)$ has support $[0, \pi_0]$; this violates BHR since the denominator of the RHS, $1 - F(\pi_0)$, vanishes. Without social information, agent i inspects with probability 1. With social information $x_{N_i,t}^{-i} \in (0, 1)$, agent i inspects with probability 1 if some neighbor adopts, and below 1 if no neighbor adopts, since $\pi_t^\emptyset < \pi_0$. After observing

⁷For then $f(c) \leq E[f(z)|z \geq c] = \frac{1 - F(c)}{1 - c} \leq \frac{1 - F(c)}{c(1 - c)}$.

the extra information, agent i 's inspection and adoption rate thus drop, contradicting Lemma 2. \triangle

3.2 Herding in Large Complete Networks

Much of the literature on social learning focuses on the asymptotic properties of learning, when agents observe many predecessors. When agents have imperfect information about the timing of other agents' moves, several papers have found that the asymmetry in the inference problem after observing adoption and non-adoption translates into an asymmetry of asymptotic learning.⁸ In particular, they find that bad products never succeed while good products sometimes fail. To connect our model to this strand of the literature, we show the analogous result holds in our model.

Let \underline{c} be the minimum cost in the support of F . Also, let p_I^θ be the ex-ante probability that an agent in a complete network of I agents will inspect a product of quality θ .

Theorem 2. *In large complete networks: Bad products always fail, $\lim_{I \rightarrow \infty} p_I^L = 0$; Good products always succeed, $\lim_{I \rightarrow \infty} p_I^H = 1$, iff $\underline{c} = 0$.*

Proof. See Appendix A.1. □

Intuitively, no-one ever adopts a low-quality product, so agents become pessimistic and stop inspecting. With a high quality-product, there exists a choke point \bar{x} such that, when the probability someone adopts hits \bar{x} , then seeing “no adoption” makes agents sufficiently pessimistic that they no longer inspect; that is, $\pi^\theta(1 - \bar{x}) = \underline{c}$. If $\underline{c} = 0$, then $\bar{x} = 1$ and some agents will keep inspecting until someone adopts the high-quality product. While, if $\underline{c} > 0$, the fate of the high-quality product in a large market is decided immediately after the product launch, $t = 0$. With probability \bar{x} , an early-mover inspects and adopts the product, and everybody else follows suit. With probability $1 - \bar{x}$, there is no such early-mover, and the posterior belief about product quality drops to \underline{c} before anybody adopts, and the product fails.

We now move on to the heart of our paper: studying the impact of network structure on social learning and diffusion.

4 Tree Networks

We have argued that the analysis is complicated by (i) self-reference and (ii) correlation. This motivates us to study learning in tree networks, where such problems do not arise. Tree networks are realistic in some applications such as hierarchical organizations where the information flow is unidirectional. They are also a good approximation of large random networks;⁹ these are relevant when a consumer learns about electric cars by observing the number on the road, or about movies through friends' posts on social media.

We say that network G is a *tree* if any two agents i and j are connected by at most one path $i \rightarrow \dots \rightarrow j$. Since i 's neighbors do not observe i , and since they have independent information, i 's

⁸See Guarino, Harmgart, and Huck (2011), Hendricks, Sorensen, and Wiseman (2012) and Herrera and Hörner (2013).

⁹Suppose there are I agents on a random directed graph G with out-degree distribution Φ ; for large I , such a network can be constructed, say, by the configuration model (Jackson, 2010, Section 4.1.4). If we look within n links of agent i , then i 's “ n -local” network is almost surely a tree as I grows large, and so i 's social learning curve in the random network G converges to the one on the tree with out-degree distribution Φ .

social learning curve depends only on the individual adoption probabilities of her neighbors $x_{N_i,t}^{-i} = x_{N_i,t} = 1 - \prod_{j \in N_i} (1 - x_{j,t})$. Thus, her adoption probability is

$$\dot{x}_{i,t} = \alpha \left[1 - \prod_{j \in N_i} (1 - x_{j,t}) \left(1 - \tilde{F} \left(\prod_{j \in N_i} (1 - x_{j,t}) \right) \right) \right]. \quad (8)$$

This forms an I -dimensional, time-invariant first-order ODE, which is easy to compute.

We say a tree G is *regular with degree d* if every node has d links (i.e. each agent observes d agents). Such a system has $|I| = \infty$ agents, so Theorem 1 does not apply as stated. Nevertheless, 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 a unique equilibrium exists. Agents are symmetric, so equilibrium adoption is the same for all agents, and we write it as x_t . The probability no neighbor adopts is $(1 - x_t)^d$, so agent i 's adoption curve is given by a one-dimensional ODE,

$$\dot{x}_t = \alpha \left[1 - (1 - x_t)^d (1 - \tilde{F}((1 - x_t)^d)) \right]. \quad (9)$$

This generalizes equation (4) by allowing for more than one neighbor per agent.

4.1 Comparative Statics

Clearly, an agent benefits when they add more neighbors since they can always ignore the extra links. But what if their neighbors add more links? Moreover, is it possible to compare learning in different tree networks? For example, would an agent learn more from an infinite chain of agents, or two independent, uninformed agents? In this section we answer these questions. This is in contrast to traditional herding models where behavior is too complicated for simple comparative statics.

Figure 4 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 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 $d = 1$ (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. We now show that, if the BHR assumption holds, an agent always benefits from both direct and indirect links.

Write adoption rates in tree networks G, \hat{G} as x, \hat{x} , and recall that agent i 's social learning in network \hat{G} is (Blackwell) superior to her social learning in network G if $\hat{x}_{N_i,t} \geq x_{N_i,t}$ for all t , implying that i prefers network \hat{G} over G .

Theorem 3. *Consider trees $G \subset \hat{G}$ and assume BHR holds. For any agent, social learning is superior in the larger tree,*

$$x_{N_i,t} \leq \hat{x}_{\hat{N}_i,t} \quad \text{for all } t. \quad (10)$$

Hence all agents prefer the larger tree \hat{G} .

Proof. First consider the leaves of G , who have no information in the small tree. By definition $x_{N_i,t} = 0 \leq \hat{x}_{\hat{N}_i,t}$. Now, consider some agent i with neighbors N_i in the small tree. By induction, assume that (10) holds for all $j \in N_i$. By Lemma 2, such neighbors adopt more in the larger network,

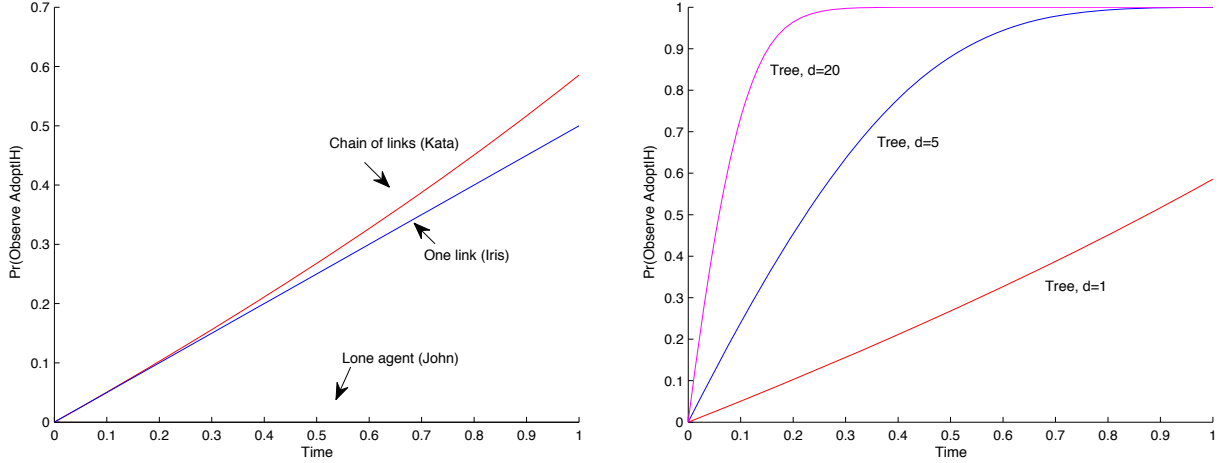


Figure 4: **Social Learning Curves in Tree Networks.** This figure assumes $c \sim U[0, 1]$, $\alpha = 1$, and $\pi_0 = 1/2$.

$x_{j,t} \leq \hat{x}_{j,t}$. Additionally, agent i has more neighbors in the large tree. Hence,

$$x_{N_i,t} = 1 - \prod_{j \in N_i} (1 - x_{j,t}) \leq 1 - \prod_{j \in \tilde{N}_i} (1 - \hat{x}_{j,t}) = \hat{x}_{\tilde{N}_i,t}$$

as required.¹⁰ □

Theorem 3 is proved by induction and hence applies to any finite tree. But the analogous result holds for (infinite) regular trees, as illustrated in Figure 4. That is, an increase in d raises the probability one neighbor adopts and, using BHR, raises the RHS of the law-of-motion (9). Thus, the adoption x_t also increases for all t .

Theorem 3 is silent about the quantitative impact of direct and indirect links. The next example shows that adding one direct link is more important than adding a whole chain of indirect 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 5. When agent i has two uninformed neighbors j, k , each neighbor has an adoption curve $x_{j,t} = x_{k,t} = \tilde{F}(1)\alpha t$. Hence the probability that at least one of them adopts is

$$x_{\{j,k\},t} = 1 - (1 - \tilde{F}(1)\alpha t)^2. \quad (11)$$

With an infinite chain, agent i 's social learning curve is given by

$$\dot{x}_t = \alpha \left[1 - (1 - x_t)(1 - \tilde{F}(1 - x_t)) \right] \leq \alpha \left[1 - (1 - x_t)(1 - \tilde{F}(1)) \right] = \alpha \left[(1 - \tilde{F}(1))x_t + \tilde{F}(1) \right].$$

Solving this ODE,

$$x_t \leq \frac{\tilde{F}(1)}{1 - \tilde{F}(1)} \left(\exp \left((1 - \tilde{F}(1))\alpha t \right) - 1 \right). \quad (12)$$

In Appendix A.2, we show that (11) exceeds (12) for any $\tilde{F}(1), \alpha, t \in [0, 1]$, as required.

¹⁰Without the BHR assumption, the result can break down. In Example 3, an agent observing agent i would prefer that she not see agent j 's action.

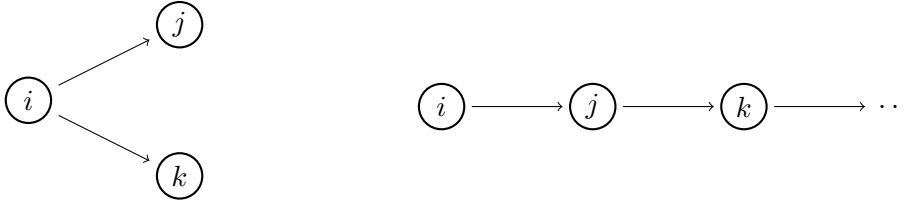


Figure 5: **Direct vs Indirect Links.** This figure illustrates the two networks compared in Example 4.

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^{th} removed neighbor in the chain is $(1/2)^n$, meaning that an infinite chain of signals is worth at most two direct signals. However, these indirect signals are intermediated (i.e. k 's signal must pass through j) which reduces their information value. Put differently, two direct links are better for $t \approx 0$ since k can communicate directly, rather than having to wait for j to enter. Over time, the information value of the chain increases and the slope of the social learning curve catches up, but the level never does. \triangle

5 Network Structure

In this section we move beyond trees, and study other aspects of network structure. In tree networks, Theorem 3 establishes that agents benefit from both direct and indirect links. Do they also benefit from backwards and correlating links? We first provide two examples and then present a more general theorem that combines these forces. See Figure 6 for illustrations.

Example 5 (Adding Correlating Links). First assume that agent i observes two uninformed agents, j and k . Given high quality, agent i sees no adoption with probability $\Pr(k \text{ not adopt}) \times \Pr(j \text{ not adopt})$ since j and k are independent. Now suppose we then add a link from j to k , correlating their adoption outcomes. The chance i sees no adoption rises to $\Pr(k \text{ not adopt}) \times \Pr(j \text{ not adopt} | k \text{ not adopt})$, thereby reducing i 's social learning.¹¹ 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. \triangle

Example 6 (Adding Backward Links). Recall Example 1 where agent i observes an uninformed agent j , and add a backwards link $j \rightarrow i$. 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. Let $x_{j,t}^{-i}$ be j 's adoption curve conditional on $t \leq t_i$. Equation (5) implies that i 's adoption curve is given by

$$\dot{x}_{i,t} = \alpha \left[1 - (1 - x_{j,t}^{-i})(1 - \tilde{F}(1 - x_{j,t}^{-i})) \right].$$

To solve for the adoption of the “ignorant j ” agent, $x_{j,t}^{-i}$, note that he in turn knows that i cannot

¹¹Formally, when agents j and k are independent, $\Pr(j \text{ not adopt}) = 1 - \alpha \tilde{F}(1)t$. However, when agent j observes k , $\Pr(j \text{ not adopt} | k \text{ not adopt}) = 1 - \int_0^t \alpha F(1 - x_{k,s}) ds$.



Figure 6: **Networks from Examples 5 and 6.** The **left** panel adds a correlating link. The **right** panel adds a backward link.

have seen him adopt at $t \leq t_j$, so thinks he is learning from an “ignorant i ” agent, $x_{i,t}^{-j}$. The ignorant agents i and j thus have adoption curves

$$\dot{x}_{j,t}^{-i} = \alpha \tilde{F}(1 - x_{i,t}^{-j}) \quad \text{and} \quad \dot{x}_{i,t}^{-j} = \alpha \tilde{F}(1 - x_{j,t}^{-i}). \quad (13)$$

It follows that $\dot{x}_{j,t}^{-i} \leq \alpha \tilde{F}(1)$, and so $x_{j,t}^{-i} \leq \alpha \tilde{F}(1)t$, which is j ’s adoption curve if he does not observe i . Thus, the link $j \rightarrow i$ lowers i ’s social learning curve, and her utility. Intuitively, when i makes her decision, j cannot have seen her adopt; the link $j \rightarrow i$ thus makes j more pessimistic in this event and reduces his adoption. \triangle

We generalize these examples in the following theorem. Define an i -tree G as a network where no $j \in N_i$ is its own successor, and no $j, j' \in N_i$ have a common successor. In particular, such a network has no *backward links* $B_i = \{j \rightarrow i\}$ for $j \in N_i$, and *correlating links* $C_i = \{j \rightarrow j'\}$ for $j, j' \in N_i$. We now consider adding such links, by considering networks $\hat{G} \subset G \cup B_i \cup C_i$ with $G \subsetneq \hat{G}$.

Theorem 4. *Adding backward and correlating links to an i -tree reduces i ’s social learning at all $t > 0$, $\hat{x}_{N_i,t}^{-i} < x_{N_i,t}^{-i}$, and her utility.*

Proof. See Appendix A.3. □

The intuition is the same as in the examples. Adding a backward link $j \rightarrow i$ makes j more pessimistic conditional on $t \leq t_i$, reducing j ’s adoption probability and thereby i ’s social learning. Adding a correlating link $j \rightarrow j'$ makes j less likely to adopt when j' does not adopt, which is exactly when j ’s information is valuable to i .

We have seen that i benefits from “independent” links (Theorem 3), but is harmed by backward and correlating links (Theorem 4). The effect of adding other links is less clear.

Example 7 (The “Niece” link). Take the four-person tree network with links $i \rightarrow j, k$ and $j \rightarrow l$ (so l is k ’s “niece”), and consider the effect of adding the link $k \rightarrow l$ on i ’s social learning. This new link has two effects: it provides extra information to k and can raise her adoption; it also correlates the decisions of j and k .

If costs have support $[0, \pi_0]$, as in Example 3, then both effects lower i ’s social learning and utility. However, if BHR holds, these forces counteract one another, and the niece link can benefit agent i . For an example, suppose costs have the following bi-modal distribution:

$$f(c) = \begin{cases} \underline{f} & \text{for } c \in [0, \pi'] \\ 0 & \text{for } c \in [\pi', \pi_0] \\ \bar{f} & \text{for } c \in [\pi_0, 1] \end{cases}$$

where $\pi' < \pi_0$ is small, and the constants \underline{f}, \bar{f} are such that BHR holds. Here, the link $k \rightarrow l$ always increases k 's probability of adopting: If l adopts, this induces k to adopt, which is useful if j has not yet adopted; if l does not adopt this makes k more pessimistic, but does not lower his inspection probability because his belief remains above π' (recall that π' is small). \triangle

5.1 Applications

We now address the implications of the economic forces we have identified. So far we have seen that an agent benefits from direct and indirect links (Theorem 3), prefers direct to indirect links (Example 4), but is harmed by self-reflecting and correlating links (Theorem 4). 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.

Theorem 5. *Among all networks with $|I|$ agents, the i -star maximizes agent i 's social learning, $x_{N_i, t}^{-i}$, and her utility.*

Proof. 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 , $t_j < t_i$, and has favorable idiosyncratic preferences, chooses not to inspect, $c_j > \pi_0$. 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 prove the result, 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 $c_\ell > \pi_0$ he chooses not to inspect, and thus does not adopt. Continuing by induction, agent ℓ also sees no adoption in G ; the lack of adoption is bad news, $\pi^\theta(1 - x_{N_\ell, t_\ell}^{-\ell}) \leq \pi_0 < c_\ell$, so he also does not inspect or adopt. Thus i 's social learning curve is higher in the i -star than in any other network, as is her utility. \square

Do agents learn more when learning from disparate sources, where each agent has many “weak ties”, or in a network where connections are clustered and lessons are reinforced? The famous “strength of weak ties” hypothesis (Granovetter, 1973) 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, consider two networks in which all agents have d links. In the *centralized* network there are $d + 1$ agents, all linked to one another. In the *decentralized* network there is an infinite number of agents connected via a random network.

Theorem 6. *Assume BHR holds and all agents have d neighbors. Social learning in the decentralized network is superior to the centralized network.*

Proof. Consider network G where agent i has d neighbors, who have no links themselves (i.e. an i -star). Since the decentralized network is a regular tree (see footnote 9), Theorem 3 implies that i 's

learning in the decentralized network is superior to G . In contrast, the centralized network is obtained from G by adding all self-reflecting and correlating links. By Theorem 4, i 's learning in G is superior to the centralized network. Finally, observe that both these networks treat all agent symmetrically, so all agents prefer the decentralized network to the centralized network. \square

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 show that an agent benefits from more direct and indirect links, but is harmed by correlating and backwards links. We also characterize an agent's optimal network, and show that agents prefer decentralized networks over centralized networks.

The paper has three 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. Third, it provides a base to understand policy experiments that affect network structure and the information of participants.

Our ongoing work pushes this paper into three different directions. Given the current model, we are interested in characterizing aggregate behavior (e.g. welfare, diffusion) across a variety of networks. We are also extending the model to allow for more general learning structures, whereby agents sometimes adopt low-quality products. Finally, we are studying diffusion when agents know their neighbors, but not the entire network structure.

Appendix

A Omitted Proofs

A.1 Proof of Theorem 2

When the probability of observing an adoption hits $\bar{x} := \sup\{x : \tilde{F}(1-x) > 0\}$, the absence of an adoption makes agents sufficiently pessimistic to shut down inspection completely, no matter the cost. Using the definition of \tilde{F} , this cutoff satisfies $\pi^0(1-\bar{x}) = \underline{c}$, so $\bar{x} = 1$ iff $\underline{c} = 0$. We also assume that $\pi_0 > \underline{c}$ so $\bar{x} > 0$; else, no agent ever inspects for any I , and the theorem is trivially true.

The key argument in the proof is that the adoption probability $x_{N_i,t}^{-i}$ converges to the choke-point \bar{x} instantaneously as the number of agents grows large. Formally, we claim that:

$$\lim_{I \rightarrow \infty} x_{N_i,t}^{-i} = \bar{x} \quad \text{for any } t > 0 \quad (14)$$

Intuitively, as long as $x_{N_i,t}^{-i} < \bar{x}$, some low cost types $c \in [\underline{c}, \pi(1-x_{N_i,t}^{-i})]$ are willing to inspect, pushing $x_{N_i,t}^{-i}$ up as I grows; as $x_{N_i,t}^{-i}$ approaches \bar{x} , the inspection probability vanishes, so clearly $x_{N_i,t}^{-i} \leq \bar{x}$. To formalize this argument, suppose agent i moves at time $t_i = t > 0$. Since the network is symmetric and adoption rises over time, the inspection probability of any agent j with $t_j < t$ (conditional on the most negative event that j observes no adoption) is bounded below by $\tilde{F}(1-x_{N_j,t_j}^{-j}) \geq \tilde{F}(1-x_{N_i,t}^{-i})$. Thus the ex-ante probability that agent i observes no adoption is bounded above by $1-x_{N_i,t}^{-i} \leq (1-\alpha t \tilde{F}(1-x_{N_i,t}^{-i}))^{I-1}$, where t is the probability an agent j moves before i and α is the probability he has a favorable idiosyncratic preference. Now, if for a subsequence I_ν we had $\lim_{\nu \rightarrow \infty} x_{N_i,t}^{-i} < \bar{x} \leq 1$, we run into the contradiction that

$$0 < \lim_{\nu \rightarrow \infty} (1-x_{N_i,t}^{-i}) \leq \lim_{\nu \rightarrow \infty} (1-\alpha t \tilde{F}(1-x_{N_i,t}^{-i}))^{I-1} = 0.$$

This establishes (14).

Now consider a low quality product, $\theta = L$. Since \tilde{F} is continuous, the inspection probability vanishes at any $t > 0$, $\lim_{I \rightarrow \infty} \tilde{F}(1-x_{N_i,t}^{-i}) = \tilde{F}(1-\bar{x}) = 0$, and thus its expected value $p_I^L = \int_0^1 \tilde{F}(1-x_{N_i,t}^{-i}) dt$ vanishes, too. For a high quality product, $\theta = H$, the inspection probability at any $t > 0$ converges to $\lim_{I \rightarrow \infty} (1-(1-x_{N_i,t}^{-i})(1-\tilde{F}(1-x_{N_i,t}^{-i}))) = \bar{x}$, which equals 1 iff $\underline{c} = 0$. Thus, good products succeed with probability \bar{x} ; this implies that good products succeed with certainty iff $\bar{x} = 1$, as required.

A.2 Example 4: Two Links vs Infinite Chain

Here, we provide the calculations for Example 4, arguing that $x(\alpha t) := x_{\{j,k\},t} = 1 - (1 - \tilde{F}(1)\alpha t)^2$ exceeds $y(\alpha t) := \frac{\tilde{F}(1)}{1-\tilde{F}(1)} \left(\exp\left((1-\tilde{F}(1))\alpha t\right) - 1 \right)$ which, in turn, exceeds the social learning curve in the infinite chain.

It suffices to show this inequality for $\alpha t = 1$. To see this, note that the difference $x - y$ is (i) concave in αt since x is concave and y is convex,¹² and (ii) increases at $\alpha t = 0$ since $x'(0) = 2\tilde{F}(1) >$

¹²Intuitively 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.

$\tilde{F}(1) = y'(0)$.¹³ Abbreviating $\delta := 1 - \tilde{F}(1)$, we thus need to show $x(1) = 1 - \delta^2 \geq \frac{1-\delta}{\delta}(e^\delta - 1) = y(1)$. Multiplying by $\delta/(1-\delta)$, this is equivalent to $1 + \delta(1 + \delta) \geq e^\delta = 1 + \delta + \delta^2/2 + \delta^3/6 + \dots$. Subtracting $(1 + \delta + \delta^2/2)$, and dividing by $\delta^2/2$, the LHS becomes 1, while the RHS equals $\frac{\delta}{3} + \frac{\delta^2}{12} + \dots \leq 2(\frac{1}{3!} + \frac{1}{4!} + \dots) = 2(e - 2\frac{1}{2}) < 0.5$.

A.3 Proof of Theorem 4

Consider the original network G and suppose that i sees no adoption of the high quality good. It must be that for every $j \in N_i$ with $t_j < t_i$ and favorable idiosyncratic preference: (i) agent j does not see an adoption, $\lambda_{k,t_j} \neq a$ for all $k \in N_j$, and (ii) agent j has relatively high costs, $c_j \geq \pi(1 - x_{N_j,t_j}^-)$. We now argue that for the same realization of costs, entry times and idiosyncratic preferences, i also sees no adoption in \hat{G} .

To prove the result, we consider the L neighbors of i 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 $\ell \in \{1 \dots L\}$. First, we establish that agent $\ell = 1$ does not observe an adoption in \hat{G} . By supposition (i) we know the ‘‘original neighbors’’ $k \in N_1$ do not adopt by t_1 ; since their inspection decision only depends on their successors, they also do not adopt in \hat{G} . Moreover, the ‘‘new neighbors’’ of $\ell = 1$, which may include both i ’s neighbors and i herself, have not yet entered and hence not adopted.

Second, we show $\ell = 1$ does not inspect in \hat{G} if none of his neighbors has adopted. By supposition (ii) we know that $\ell = 1$ does not inspect in G . The adoption probability of the original neighbors N_1 does not depend on the additional links in \hat{G} ; but the lack of adoption by the additional links provides additional bad news for $\ell = 1$, meaning he has less incentive to inspect in \hat{G} . Formally, $\pi^\theta(1 - \hat{x}_{N_1,t_1}^{-1}) \leq \pi^\theta(1 - x_{N_1,t_1}^{-1}) < c_1$, where the first inequality is strict if $\ell = 1$ has an extra neighbor in \hat{G} . All told, agent $\ell = 1$ does not adopt.

Now suppose by induction that agents $1, \dots, \ell - 1$ have not adopted and consider agent ℓ . First, he does not see an adoption in \hat{G} . This is because his ‘‘new neighbors’’ either move later, or move earlier and have not adopted, by induction. Second, the lack of adoption from the new neighbors makes him more pessimistic in \hat{G} and less likely to inspect conditional on seeing no adoption.

Summing up, we have argued that if i does not observe an adoption in network G , he does not observe an adoption in network \hat{G} either, and so $\hat{x}_{N_i,t}^{-i} \leq x_{N_i,t}^{-i}$. The strict inequality obtains since at least one of i ’s neighbors j has an additional link in \hat{G} , and so there are some realizations of entry times and idiosyncratic preferences such that for costs $c_j \in [x_{N_j,t_j}^{-j}, \hat{x}_{N_j,t_j}^{-j}]$, agent j adopts in G but not in \hat{G} .

¹³Intuitively, the twice removed link in the infinite chain is only half as useful as a direct link since agents need to enter in the right order, while further removed links are initially useless.

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.
- 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 Micro-finance,” *Science*, 341(6144), 1236498.
- BANERJEE, A. V. (1992): “A Simple Model of Herd Behavior,” *Quarterly Journal of Economics*, 107(3), 797–817.
- 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.
- 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.
- CONLEY, T. G., AND C. R. UDRY (2010): “Learning about a New Technology: Pineapple in Ghana,” *American Economic Review*, 100(1), 35–69.
- DEIMLING, K. (1977): *Ordinary Differential Equations in Banach Spaces*. Springer-Verlag, Berlin.
- 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.
- GUL, F., AND R. LUNDHOLM (1995): “Endogenous Timing and the Clustering of Agents’ Decisions,” *Journal of Political Economy*, 103(5), 1039–1066.
- 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.
- KAMADA, Y., AND S. MORONI (2018): “Games with Private Timing,” Working paper, University of Pittsburgh.
- LOMYS, N. (2018): “Sequential Collective Search in Networks,” Working paper, SSRN.
- 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. (2018): “Diffusion Games,” Working paper, Columbia University.

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.