THE SPEED OF INNOVATION DIFFUSION IN SOCIAL NETWORKS

ITAI ARIELI
Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology

YAKOV BABICHENKO
Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology

RON PERETZ
Department of Economics, Bar Ilan University

H. PEYTON YOUNG
Department of Mathematics, London School of Economics and Nuffield College, Oxford

New ways of doing things often get started through the actions of a few innovators, then diffuse rapidly as more and more people come into contact with prior adopters in their social network. Much of the literature focuses on the speed of diffusion as a function of the network topology. In practice, the topology may not be known with any precision, and it is constantly in flux as links are formed and severed. Here, we establish an upper bound on the expected waiting time until a given proportion of the population has adopted that holds independently of the network structure. Kreindler and Young (2014) demonstrated such a bound for regular networks when agents choose between two options: the innovation and the status quo. Our bound holds for directed and undirected networks of arbitrary size and degree distribution, and for multiple competing innovations with different payoffs.

KEYWORDS: Innovation diffusion, social networks, speed of equilibrium convergence.

1. INTRODUCTION

SOCIAL AND TECHNOLOGICAL ADVANCES are essential to economic development, but the mere existence of new and better ways of doing things does not guarantee that they will be widely used. The time it takes to dislodge inferior practices is another crucial factor in explaining how rapidly development can occur. This lag time depends on several crucial factors. One is lack of information: it may not be immediately evident that the innovation is in fact superior to the status quo. A second factor is network externalities: the desirability of an innovation depends not only on its inherent payoff but on how many others in one's social network have also adopted. Dislodging an inferior practice or technology requires a coordinated shift in expectations and behaviors among members of the group, which may take a long time even if it is already evident that everyone would be better off if they were to do so.

There is a substantial theoretical and empirical literature on these issues that we shall discuss below. In contrast to much of this literature, which is concerned with learning

Itai Arieli: iarieli@technion.ac.il

Yakov Babichenko: yakovbab@technion.ac.il

Ron Peretz: ron.peretz@biu.ac.il

H. Peyton Young: peyton.young@economics.ox.ac.uk

about payoffs from the actions of prior adopters, our focus here will be on the time it takes to dislodge an inferior practice when there are increasing returns from adoption and the payoffs are already known. This allows us to separate the effects of pure network externalities from the problem of learning about the payoffs by observing the behavior of others. Moreover, unlike much of the literature, we shall focus on the question of how long it takes to dislodge an inferior practice or technology when little or nothing is known about the topology of social interactions. Although this would appear to omit the main variable of interest, this is not the case. In particular, Kreindler and Young (2014) demonstrate that the expected waiting time to overturn an inferior equilibrium can be usefully bounded from above for all undirected regular networks.

The theoretical contributions of this paper are three-fold. First, we establish an upper bound on the expected waiting time that holds for networks of *any size and degree distribution, whether directed or undirected.* Second, we show how to extend the analysis to multiple competing innovations, instead of a single innovation versus the status quo, which is the usual assumption in the literature. Third, we show how the network topology affects the results by deriving waiting time bounds that hold for specific classes of networks, including regular networks and star networks. The mathematical techniques are novel and rely on results characterizing the distribution of arbitrarily large sums of independent random variables.

1.1. *Related Literature*

The importance of social interactions in spreading new ideas and practices has been documented in a wide variety of situations. Ryan and Gross (1943) demonstrated that farmers' decisions to adopt an agricultural innovation—hybrid corn—were strongly influenced by the adoption decisions of their neighbors. Subsequently, Griliches (1957) showed that the decision to adopt also depended on the expected gains in payoff from the innovation relative to the status quo. At about the same time, sociologists Coleman, Katz, and Menzel (1957) analyzed the role of social networks among doctors in the adoption of a new medical treatment (in this case tetracycline), and showed that adoption was driven to a significant extent by peer effects.[1]

Since then a substantial theoretical literature has developed on the rate of innovation diffusion as a function of the network topology. A recurrent theme is that interaction among small close-knit groups can speed up the adoption process. The logic is that the innovation can gain a local foothold relatively quickly, and from different local footholds it then spreads throughout the network (Ellison (1993), and Young (1998, 2011); Montanari and Saberi (2010)). Experimental studies of games played on networks are consistent with these predictions (Centola and Baronchelli (2015)).[2]

A related line of work is concerned with the time it takes for a new idea to spread when it is seeded at one or more locations. Here, the key features are the centrality of the nodes where the new idea is seeded, and the degree of connectivity of the network (Morris (2000); Banerjee, Chandrasekhar, Duflo, and Jackson (2013)). These are instances of "threshold" models, in which a given individual adopts once a sufficient number of his

---

[1]For subsequent empirical work on innovation diffusion in networks, see Valente (1995, 2005); Foster and Rosenzweig (1995, 2010), Kohler (1997), Kohler, Behrman, and Watkins (2001), Udry and Conley (2001), Rogers (2003), and Munshi (2006).

[2]There is also a recent literature on the speed of convergence in games where agents interact globally instead of with a fixed set of neighbors; see, in particular, Ellison, Fudenberg, and Imhof (2016) and Arieli and Young (2016).

neighbors have adopted (Watts (2002)). In this literature, adoption decisions are typically treated as irreversible, whereas in the present paper we treat adoption (and disadoption) as random variables that depend on the distribution of choices by one's neighbors and stochastic shocks to their perceived payoffs; they are not deterministic.

Yet another branch of the literature investigates how network structure affects the rate at which agents update their priors about the desirability of the innovation, based on observation of their neighbors' choices. issue here is the identification of conditions on the network topology and updating rules under which the process converges to correct beliefs and optimal actions.[3] The actual learning behavior of subjects who are embedded in different types of networks has been investigated experimentally by Gale and Kariv (2003) and Mäs and Nax (2016).

In this paper, we focus instead on the situation where the payoffs are known in advance or have already been learned. In particular, we assume that people know that the new practice or technology would be inherently better than current practice provided that a sufficient number of people in one's network adopted it. The source of such network externalities depends very much on the situation. A social website is more valuable the more people who use it. A market is more valuable the more traders it attracts. A similar logic holds for communication technologies, technological standards, and many other innovations with increasing returns.[4]

A different type of network externality arises when people are sanctioned for not conforming to a current norm or practice. Demographers have found, for example, that social norms are a significant factor in explaining the pace and pattern of contraceptive use in developing countries.[5] In the United States, norms of medical treatment for a given medical condition differ widely among states and even among counties within the same state. These differences appear to be the product of two types of network externality: peer effects and information sharing. Physicians tend to conform to the choices of local opinion leaders, and once a given practice becomes established its benefits are enhanced by local knowledge sharing within the group.[6]

These and other sources of network externalities can be modelled as a network game in which individuals repeatedly play a coordination game against their neighbors. Individuals periodically update their choices according to a random arrivals process. When individuals update they choose an optimal response given the inherent payoff from the choice as modified by the current choices of their neighbors (the coordination payoff) plus an idiosyncratic utility shock.[7]

The contribution of the current paper is to establish an upper bound on the expected waiting time until a given proportion of the population has adopted, where the bound is independent of the size and topology of the network itself. This result is similar in spirit to the framework of Kreindler and Young (2013, 2014), who establish an upper bound on

---

[3]See Banerjee (1992), Bikhchandani, Hirschleifer, and Welch (1992), Ellison and Fudenberg (1993), Bala and Goyal (1998), Jackson (2008), Solan, Rosenberg, and Vieille (2009), Golub and Jackson (2010, 2012), Acemoglu, Dahleh, Lobel, and Ozdaglar (2011), Mueller-Frank (2013), Mueller-Frank and Pai (2016), and Mossel, Sly, and Tamuz (2015).

[4]David (1985) and Arthur (1994).

[5]See Bongaarts and Watkins (1996), Montgomery and Casterline (1996), Kohler (1997), Kohler, Behrman, and Watkins (2001), and Munshi and Myaux (2006).

[6]See Wennberg and Gittelsohn (1973), Phelps and Mooney (1993), Chandra and Staiger (2007), and Burke, Fournier, and Prasad (2007, 2010).

[7]See among others Blume (1993), Jackson and Yariv (2007), Jackson (2008), Vega-Redondo (2007), and Golub and Jackson (2010, 2010).

the waiting time for all regular undirected networks and a logit model of errors. Here, we employ different mathematical methods to establish a more general bound on the waiting time that holds for *all directed and undirected networks of arbitrary size*, and for a very broad class of error distributions. These bounds are especially useful when the network is difficult to observe and is constantly changing as agents form and sever links with one another.[8] The mathematical techniques rely on recent advances in estimating the distribution of arbitrarily large sums of independent random variables (Feige (2006), Garnett (2020)).

A common technique for the analysis of innovation diffusion in large networks is the mean field approach (see, e.g., Young (2009)). We use this approach in Proposition 6.1 to derive a tight bound on the expected waiting time in large regular networks. The mean field approach, however, is not suitable for proving our main result (Theorem 3.1) which holds even when agents are arbitrarily heterogeneous with respect to their degree and the influence they exert on one another.

The plan of the paper is as follows. In the next section, we formulate the stochastic updating model. In Sections 3–4, we establish a general bound on the expected waiting time until a target proportion of the population has adopted, starting from the state where everyone is playing the status quo. The bound depends on the shape of the error distribution, on the payoff gap between the innovation and the status quo, and on the magnitude of the coordination payoffs relative to the inherent payoffs from different choices, but it does not depend on the topology of the network per se. Section 5 extends the analysis to multiple competing innovations. In Section 6, we show how the waiting time depends on the topology of the network by considering two contrasting cases: regular networks and star networks. In particular, we show that the bound established by Kreindler and Young (2014) for large regular networks can be substantially improved.

2. MODEL

A weighted directed network with m nodes can be represented by an $m \times m$ row-stochastic matrix $P(t) = \{P_{ij}(t)\}$. We interpret $P_{ij}(t)$ as the probability that agent i interacts with j during the current period.[9] Alternatively, we can view $P_{ij}(t)$ as the relative weight that i attaches to interactions with j in the current period. For expositional simplicity, we shall begin by considering the situation where each agent chooses one of two actions: the innovation (action 1) or the status quo (action 0). In Section 5, we shall extend the analysis to multiple innovations with different payoffs.

The payoff from choosing an action consists of two parts: (i) an inherent payoff that is independent of how many others choose it, and (ii) a coordination payoff that results from coordinating with others. For an interaction with a single neighbor, we represent the payoffs by the following 2×2 matrix:

	0	1
0	c	0
1	a	$a + c$

---

[8]There is a large literature on the dynamics of link formation in social networks. See, among others, Skyrms and Pemantle (2000), Jackson and Watts (2002), Goyal and Vega-Redondo (2007), and Jackson (2002).

[9]We allow for the possibility that $P_{ii}(t) > 0$, that is, i's own action in the previous period increases the probability that he chooses it again next period. This can be interpreted as a form of inertia.

Here, $a > 0$ is the payoff difference between the innovation 1 and the status quo 0, and $c > 0$ is the coordination payoff that results from making the same choice as someone with whom one interacts. Note that when $c < a$ the innovation is a dominant strategy and there is a unique equilibrium whereas if $c > a$ there are three equilibria: two pure and one mixed. The subsequent analysis holds in either case but we shall typically assume that there are multiple equilibria.

The *state* of the process at the end of each period t is an m-vector $s(t) \in \{0 \ 1\}^m$, where $s_i(t) = 1$ if agent i chooses the innovation at t, and $s_i(t) = 0$ otherwise. The updating process works as follows. Time is continuous and the initial state is $s(0) = (0 \quad 0)$. We suppose that every agent receives updating opportunities according to a Poisson arrival process with rate one per time period, where these processes are independent among the individuals. Suppose that i receives such an opportunity at time t. Given the current network structure $P = P(t)$ and the current state $s(t)$, let

$$x_i(t) = \sum_{j \in [m]} P_{ij} s_j(t).$$

Thus i's expected payoff from interacting with a randomly drawn neighbor (drawn according the distribution $P(t)$) is $u_i(1) = a + cx_i(t)$ if he chooses action 1, and $u_i(0) = c(1 - x_i(t))$ if he chooses action 0.

Let us assume that the difference between the payoff from 1 and the payoff from 0 is perturbed by a random payoff shock $_i(t)$ with c.d.f. $F( )$. For ease of interpretation, we shall assume that F has a density $f( )$ and that all shocks are i.i.d. among agents and among time periods. If the perturbed payoff difference is positive, agent i chooses action 1; otherwise he chooses action 0. This class of perturbed best reply dynamics is very general and includes such standard dynamics as the logit response, where the payoff shocks are extreme-valued distributed. Experimental evidence on learning in networks shows that subjects do deviate from best reply with a probability that is decreasing in the resulting payoff loss, which is consistent with this class of models (Maes and Nax (2016)).

Conditional on receiving an updating opportunity at time t, the probability that i chooses 1 is

$$\mathbf{P} \ s_i(t) = 1 | x_i(t) \ = \mathbf{P} \ a + \ _i(t) + cx_i(t) > c \ 1 - x_i(t)$$

$$= \mathbf{P} \ _i(t) > c - a - 2cx_i(t) \ = 1 - F \ c - a - 2cx_i(t) \tag{1}$$

We shall call this the *response function* and write

$$r \ x_i(t) \ = 1 - F \ c - a - 2cx_i(t) \tag{2}$$

Note that $r(\cdot)$ is monotone, nondecreasing, continuous, and $0 \le r(0) \le r(1) \le 1$. (Continuity follows from our assumption that F has a density.) In what follows, we shall assume that $r(0) > 0$, that is, $F(c - a) < 1$. This amounts to saying that even when none of i's neighbors has adopted the innovation, there is a positive probability that i will switch from 0 to 1 due to receiving a sufficiently large payoff shock.

Figure 1 shows an example of a response function based on the normal distribution.

Although we have derived the response function from a specific shock distribution F, in what follows we can view the response function as a primitive of the model: $r(x)$ specifies the probability that an agent chooses action 1 whenever a (weighted) proportion x of the agent's neighbors choose action 1. Note that the same response function holds for all agents and is independent of the particular network through which they interact.
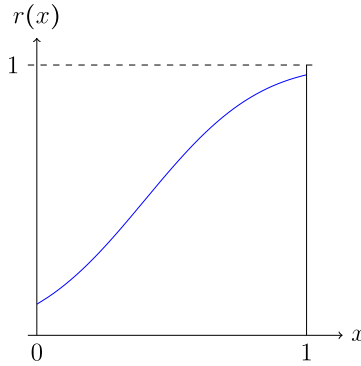
FIGURE 1.—A response function based on the normal distribution.

Given a response function $r(x)$, define *the supporting line* $L(x)$ to be the unique line that solves the following maximization problem:

$$\max L(0)$$

$$\text{subject to } L(1) = r(1) \text{ and } L(x) \leq r(x) \text{ for } x \in [0\ 1]$$

The *slope* of L will be denoted by $1 - \alpha := L(1) - L(0) < 1$, and its *fixed point* will be denoted by $p = \frac{L(0)}{\alpha}$, as shown in Figure 2.[10]

In what follows, we shall focus on the case $L(0) > 0$. The conditions under which this assumption holds are discussed in Section 3.1. In particular, it fails to hold if the response function is too "bowed," and crosses the 45° line at a low value of x as shown in Figure 3. In this case, the adoption process can become trapped in a low-level equilibrium for a very long time; indeed the expected waiting time until a low proportion q of the agents have adopted is unbounded in the network size for some classes of networks, as we show in Section 6.3.
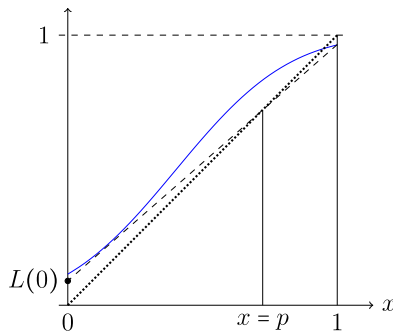

FIGURE 2.—Supporting line $L(x)$ (dashed) and response function $r(x)$ (solid).

---

[10]We note that the fixed point p of the supporting line is always smaller than the fixed point of the response function.
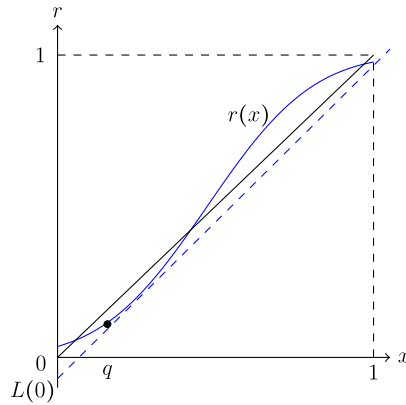
FIGURE 3.—Response function with $L(0) < 0$, and a low value q such that $r(q) < q$.

3. A GENERAL RESULT ON WAITING TIMES

The question we wish to address is the following: starting from the state $s(0) = (0 \quad 0)$ and given a target proportion $q \in [0 \ 1]$, how long does it take in expectation to reach a state in which at least qm agents have adopted the innovation? More precisely, we wish to find an upper bound on the expected waiting time defined as follows:

$$T_q \ F, P(t) \ = \mathbb{E} \ \min \ t : \frac{1}{m} \sum_{i \in [m]} s_i(t) \geq q, s(0) = (0 \quad 0)$$

We can now state our main result.

THEOREM 3.1: *Let F be an error distribution function with density f. Suppose that the supporting line $L(x)$ satisfies $L(0) > 0$, has slope $1 - \alpha$, fixed point p, and $q < p$. Then for every dynamic network $P(t)$ we have*

$$T_q \ F, P(t) \ \leq \frac{7 \ 2}{p\alpha} \ 1 + \ln \ \frac{p}{p - q} \ = \frac{7 \ 2}{L(0)} \ 1 + \ln \ \frac{p}{p - q} \tag{3}$$

The crucial point is that the right-hand side does not depend on the network structure or on how rapidly it is evolving.

As we mentioned earlier, the closest result in the literature is due to Kreindler and Young (2014). Using martingale methods, they derive the following upper bound on the expected waiting time until at least qm agents have adopted and $P(t)$ is any regular, undirected network, namely[11]

$$T_q \ F, P(t) \ \leq \frac{1}{\alpha(p - q)} \tag{4}$$

---

[11]To illustrate their approach, let G be a complete undirected network. Given $q < p$, let $\mu = L(q) - q$. Assuming that $L(0) > 0$, we have $L(x) - x > \mu \geq 0$, for all $x \in [0, q]$. Given a proportion $x(t) \leq q$, the expected increase next period is $E[x(t + 1) - x(t)|x(t)] \geq \mu > 0$. It follows that the expected waiting time until $x(t) \geq q$ is at most $\frac{1}{\mu}$. Since $L(x) = (1 - \alpha)x + \alpha p$, $\mu = \alpha(p - q)$, and hence $T \leq \frac{1}{\alpha(p-q)}$. In fact, the same bound holds for all regular networks; moreover, if $q = \frac{1}{2}$ the bound can be improved to $T \leq \frac{0 \ 42}{\alpha(p-q)}$ (Kreindler and Young (2014), Lemma 2).

The same bound holds for irregular networks provided that the target q is expressed as the degree-weighted proportion of adopting agents. This criterion is different from insisting that a given *proportion* of agents adopt. In a star network, for example, a degree-weighted majority ($q = \frac{1}{2}$) is achieved when just the central node has adopted. A comparison of (3) and (4) shows that the latter is a better (i.e., lower) bound unless q is very close to p. Indeed (4) is less than (3) provided that

$$\frac{p}{p-q} > 7 \, 2 \left(1 + \ln\left(\frac{p}{p-q}\right)\right)$$

which holds whenever $q < 0 \, 969 \, p$. However, our bound is much more general because it holds for directed and undirected networks of any size and degree distribution. In Section 6.1, we show that for large regular networks the waiting time is precisely $\frac{1}{\alpha} \ln(\frac{p}{p-q})$, which is considerably tighter than the Kreindler–Young bound. Note, however, that for a fixed p our general bound (3) is just a linear transformation of the special bound $\frac{1}{\alpha} \ln(\frac{p}{p-q})$ which is tight for regular networks. This result shows that our bound is of the right order with respect to q for a given p.

3.1. *Discussion of the Condition $L(0) > 0$*

A key assumption underlying Theorem 3.1 is that the tangent line $L(x)$ must have a strictly positive intercept $L(0)$. If this is not the case, the theorem does not apply. When will $L(0)$ be negative? The answer is when the payoff advantage a is small relative to the coordination payoff c, and/or the variance of the distribution F is small. Recall that $r(0) = 1 - F(c - a)$ and $r(1) = 1 - F(-a - c)$. For ease of exposition suppose also that f is symmetric about 0, so that the inflection point of F occurs at 0. Then $r(0) = F(a - c)$ and $r(1) = F(a + c)$. Increasing a to $a + b$ shifts the origin as shown in Figure 4. When b is large enough, the intercept of the supporting line become positive (see Figure 4, bottom panel).

The second case is illustrated in Figure 5. Rescaling the standard deviation by an amount σ is equivalent to rescaling the horizontal axis by $\frac{1}{\sigma}$, which leads to a positive intercept if σ is large enough, as shown in the bottom panel of Figure 5.

More generally, given a response function $r(x)$ we have

$$L(0) = \inf_{x \in [0 \ 1)} \frac{r(x) - xr(1)}{1 - x} \tag{5}$$

Therefore, $L(0) > 0$ iff $r(x) > xr(1)$ for every $x \in [0 \ 1)$. If r is induced by a c.d.f. F, then $r(x) = 1 - F(c - a - 2cx)$. In this case, the condition becomes $1 - F(c - a - 2cx) > x(1 - F(-a - c))$ for every $x \in [0 \ 1)$.[12]

3.2. *Proof Sketch of Theorem 3.1*

Conditional on receiving a revision opportunity, an agent chooses action 1 with probability $r(x)$, where x is the weighted average proportion of his neighbors who are currently

---

[12]In case $a \geq c$ and the density f is symmetric and unimodal, the response function is concave, and hence $L(0) > 0$.
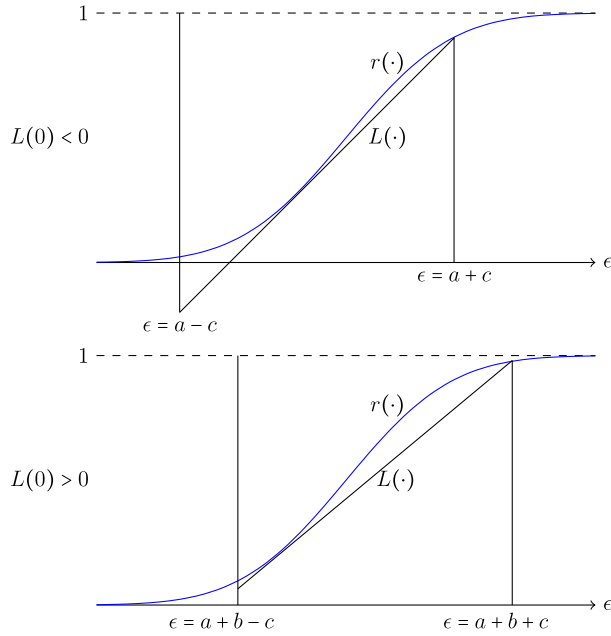
FIGURE 4.—For a given c.d.f. $F(\ )$, increasing the payoff advantage of the innovation by an amount b increases the intercept $L(0)$. Here, $F(a + b - c) = r(0)$ and $F(a + b + c) = r(1)$.
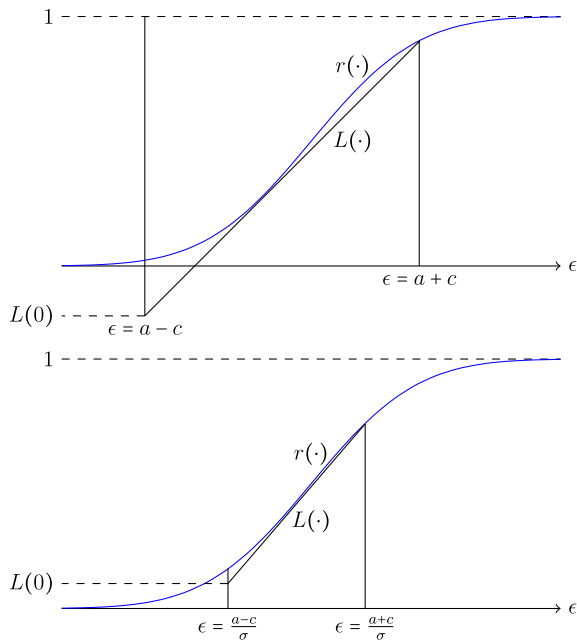

FIGURE 5.—Increasing the standard deviation of the error distribution by a factor σ increases the intercept $L(0)$. Here, $F(\frac{a-c}{\sigma}) = r(0)$ and $F(\frac{a+c}{\sigma}) = r(1)$.

choosing 1. The idea is to analyze the *linear dynamic* where an agent that faces the proportion x updates his action to 1 with probability $L(x)$ instead of $r(x)$. Since $L(x) \leq r(x)$, for every time t we have

$$r\ x_i(t) \geq p\alpha + (1-\alpha) \sum_j P_{ij}(t)s_j(t).$$

In particular, the expected time it takes for the linear dynamic to reach the threshold q is at least as long as the expected time under the actual dynamic.

An advantage of analyzing the linear dynamic is that it can be viewed as an *imitation dynamic*. Namely, at every revision opportunity, the revising agent i acts as follows: with probability α he chooses his action according to a coin toss with probability of success p. With probability $(1-\alpha)$, he imitates a neighbor that is drawn according to the distribution $P_i(t)$.[13] Thus the probability that i chooses 1 is

$$\alpha p + (1-\alpha)x_i(t) = L\ x_i(t) \tag{6}$$

In what follows, it will be notationally convenient to analyze the corresponding *discrete time* linear dynamic such that at each time t one agent is chosen uniformly at random to update his action. (This discrete time dynamic is m times slower than the corresponding continuous time dynamic, where m is the number of agents.)

A history of the imitation dynamic up to time T induces an *imitation forest*, which is defined as follows. Each vertex of the imitation forest is a pair $(i\ t) \in [m] \times [T]$. There is an edge from $(i\ t)$ to $(j\ t-1)$ if at time t agent i copied agent's j action at time $t-1$. In addition, there is an edge from $(i\ t)$ to $(i\ t-1)$ if i was not chosen to update at time t. This construction is illustrated in Figure 6.

The imitation forest admits two types of roots. A *coin-toss root* is a vertex $(i\ t)$ such that i tossed a coin at time t. An *origin root* is a vertex $(i\ 0)$ for $i \in [m]$. This induces a partition of the population at time t into two groups. A *coin-toss agent* is one that belongs to a tree with a coin-toss root. Similarly, an *origin agent* is one that belongs to a tree with an origin root.

The proof of Theorem 3.1 consists of two key steps.

Step 1. First, we show that the proportion of coin-toss agents increases at a rate that is bounded below by a simple formula that is independent of the network. This provides an explicit formula for the *expected* number of coin toss agents at each time t. We then apply Markov's inequality to show that *with high probability* the proportion of coin-toss agents is large at all sufficiently large times t.

Step 2. At each point in time, the proportion of agents that play 1 is a weighted average of a sequence of 0–1 random variables that correspond to prior coin tosses. Since we allow the network to evolve over time, analyzing these weights is extremely difficult. Nevertheless, we are able to overcome this hurdle by the following lemma, which follows readily from Feige's inequality (Feige (2006)) and its subsequent improvement by Garnett (2020). (See the Appendix for further details).[14]

---

[13]This set-up bears some resemblance to the "voter model" in which a randomly drawn agent imitates a randomly drawn neighbor Liggett (1999). In our model, by contrast, imitation only occurs with probability $1-\alpha$; otherwise the agent chooses action 1 with probability p. This leads to a fully ergodic process whereas the voter model eventually absorbs into the all-0 or all-1 state.

[14]Billingsley ((2008), Theorem 9.2) has a different bound that depends on higher moments of the Bernoulli random variable, but the bound is not as good for values of p close to 1. See Appendix A for further details.
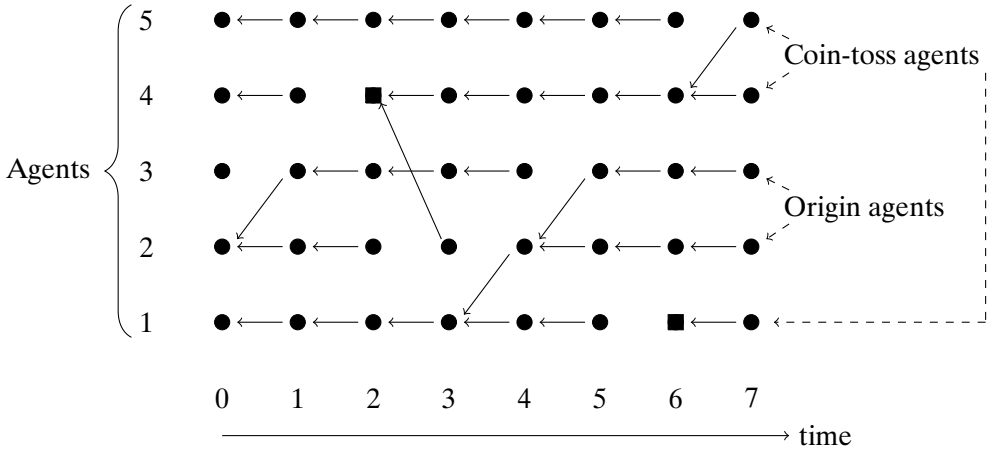
FIGURE 6.—The imitation forest of the following history. At time $t = 1$, agent 3 imitates agent 2. At time $t = 2$, agent 4 tosses a coin. At time $t = 3$, agent 2 imitates agent 4. At time $t = 4$, agent 2 imitates agent 1. At time $t = 5$, agent 3 imitates agent 2. At time $t = 6$, agent 1 tosses a coin. At time $t = 7$, agent 5 imitates agent 4. Square nodes are points where an agent tossed a coin.

LEMMA 1: *Let* $c_1, \ldots, c_k$ *be i.i.d.* Bernoulli(p) *random variables. For all k and every sequence of weights* $\beta_1, \ldots, \beta_k \geq 0$,

$$\mathbf{P}\left(\sum_{i=1}^{k} \beta_i c_i \geq \sum_i \beta_i p \right) \geq 0{\cdot}14 p.$$

This corollary allows us to place a nonzero lower bound on the probability that a given proportion of the agents are choosing action 1 by time t, which leads to an upper bound on the expected waiting time to reach a given target proportion q for any interaction structure.

4. PROOF OF THEOREM 3.1

As was mentioned in the proof sketch (Section 3.2), we shall analyze the discrete-time imitation dynamic in which one randomly drawn agent updates per period. Since each agent's Poisson clock has rate 1, the discrete process is m times slower then the actual continuous process. We shall prove that for the discrete dynamic the expected waiting time satisfies the bound

$$T_q(F, P(t)) \leq \frac{7 \cdot 2m}{p\alpha}\left(1 + \ln\left(\frac{p}{p-q}\right)\right)$$

First, we analyze the diffusion of coin-toss agents. Consider the following process: the initial state is $y(0) = (y_i(0))_{i \in [m]} = (0, \ldots, 0)$, and in each period one player $i \in [m]$ is drawn at random to update. With probability α, $y_i(t) = 1$, and with probability $1 - \alpha$, $y_i(t) = y_j(t-1)$, where j is drawn according to $P_i(t)$.

LEMMA 2: *The probability that an agent is coin-toss at time t is independent of the network structure and is equal to* $p(t) = \mathbf{P}(y_i(t) = 1) = 1 - (1 - \frac{\alpha}{m})^t$.

PROOF: This claim is established by induction on t. For $t = 0$, the claim is trivial. Assume that the equality holds for t. Than it holds for $t + 1$, because

$$\mathbf{P}\left[y_i(t+1) = 1\right] = \frac{1}{m}\left[\alpha + (1-\alpha)p(t)\right] + \left(1 - \frac{1}{m}\right)p(t)$$

$$= \frac{\alpha}{m} + \left(1 - \frac{\alpha}{m}\right)p(t)$$

$$= \frac{\alpha}{m} + \left(1 - \frac{\alpha}{m}\right)\left(1 - \left(1 - \frac{\alpha}{m}\right)^t\right)$$

$$= 1 - \left(1 - \frac{\alpha}{m}\right)^{t+1} = p(t+1).$$
$\hspace{2cm}$ Q.E.D.

PROOF OF THEOREM 3.1: Let $\tilde{T}_q$ be the first time at which at least qm agents have adopted. We seek to bound $E[\tilde{T}_q] = T_q(F[P(t)])$ from above independently of $P(t)$. Lemma 2 shows that the expected number of coin-toss agents at time t is given by $\mathbb{E}[\sum_i y_i(t)] = m - m(1 - \frac{\alpha}{m})^t$. Let us set a threshold of $\frac{q}{p}m$ for the number of coin-toss agents. We use Markov's inequality to bound the probability that the number of coin-toss agents is less than $\frac{q}{p}m$, namely,

$$\mathbf{P}\left[\sum_i y_i(t) < \frac{q}{p}m\right] = \mathbf{P}\left[\sum_i \left(1 - y_i(t)\right) > \left(1 - \frac{q}{p}\right)m\right]$$

$$\leq \frac{\mathbb{E}\left[\sum_i \left(1 - y_i(t)\right)\right]}{\left(1 - \frac{q}{p}\right)m} = \frac{\left(1 - \frac{\alpha}{m}\right)^t}{1 - \frac{q}{p}} \quad (7)$$

At each point in time t, let $R(t) = \{r_1(t), \ldots r_{k(t)}(t)\}$, be the relevant coin-toss roots at t, and denote by $c_1(t), \ldots c_{k(t)}(t) \in \{0, 1\}$ their realized actions. Note that for any fixed t, $c_1(t), \ldots c_{k(t)}(t)$ are i.i.d. Bernoulli(p) random variables. For notational convenience, we let $c_0 = 0$. Let $\varphi(i, t)$ be the function that associates each coin-toss agent with the corresponding root. Formally, if i is a coin-toss agent at time t and (i, t) belongs to a tree with root $r_j(t)$ then $\varphi(i, t) = j$. If i is an origin agent at time t, we let $\varphi(i, t) = 0$. Finally, for every $j \in [k(t)]$ we let $\beta_j(t) = |\{i : \varphi(i, t) = j\}|$ be the number of appearances of the coin-toss root j, with the corresponding action $c_j(t)$. Note that $\sum_{i \in [m]} y_i(t) = \sum_{j \in [k(t)]} \beta_j(t)$. The state of the dynamic $s(t)$ can be expressed as $s_i(t) = c_{\varphi(i,t)}$, hence $\sum_i s_i(t) = \sum_{j \in [k(t)]} \beta_j(t)c_j(t)$.

In order to bound $\tilde{T}_q$ from above, we define the stopping time

$$\tau := \min\left\{t : \sum_i y_i(t) > \frac{q}{p}m\right\}$$

and the event

$$\mathcal{E} := \left\{\omega : \sum_j \beta_j\left(\tau(\omega)\right)c_j\left(\tau(\omega)\right) \geq \sum_j \beta_j\left(\tau(\omega)\right)p\right\}$$

Note that $\omega \in \mathcal{E}$ implies that $\tilde{T}_q(\omega) \leq \tau(\omega)$. Since τ is measurable with respect to the imitation process and $\{c_1(\tau), \ldots c_{k(\tau)}(\tau)\}$ are i.i.d. Bernoulli(p) random variables conditioned on the imitation process, Lemma 1 implies that

$$\mathbf{P}(\mathcal{E}|\tau) \geq 0\,14p. \tag{8}$$

We next show that

$$\mathbb{E}[\tau] \leq \frac{m}{\alpha}\left(1 + \ln\left(\frac{p}{p-q}\right)\right) \tag{9}$$

Indeed,

$$\mathbb{E}[\tau] = \sum_{t=0}^{\infty} \mathbf{P}(\tau > t) \leq \sum_{t=0}^{\infty} \min\left\{1, \frac{1 - \left(\frac{\alpha}{m}\right)^t}{1 - \frac{q}{p}}\right\} \leq \frac{\ln\left(1 - \frac{q}{p}\right)}{\ln\left(1 - \frac{\alpha}{m}\right)} + \sum_{t=0}^{\infty}\left(1 - \frac{\alpha}{m}\right)^t$$

$$= -\frac{\ln\left(\frac{p}{p-q}\right)}{\ln\left(1 - \frac{\alpha}{m}\right)} + \frac{m}{\alpha} \leq \frac{m}{\alpha}\ln\left(\frac{p}{p-q}\right) + \frac{m}{\alpha}$$

The first inequality follows from (7), and the last inequality from the inequality $\ln(1+x) \leq x$.

By restarting a new imitation process after τ, we can define a sequence of stopping times $\tau = \tau_1 < \tau_2 < \cdots$, and corresponding sequence of events $\mathcal{E}_1, \mathcal{E}_2, \ldots$ in the above fashion. Namely, τ_{i+1} is the first time at which there are at least $\frac{q}{p}m$ coin-toss agents in the imitation process that starts at time τ_i, and $\mathcal{E}_{i+1}$ is the event that the corresponding weighted sum of coin-toss realizations reaches or exceeds its expectation. Letting i^* be the first success of one of the events $\mathcal{E}_i$, we have $\tilde{T}_q \leq \tau_{i^*}$. Since the law of $(\tau_{i+1} - \tau_i, \mathcal{E}_{i+1})$ (conditioned on the history up to time τ_i) obeys the same conditions as those of $(\tau, \mathcal{E})$, the uniform bounds (8) and (9) apply also to the pairs $(\tau_{i+1} - \tau_i, \mathcal{E}_{i+1})$, $i = 1, 2, \ldots$ conditioned on the history up to time τ_i. Therefore,

$$T_q(F, P(t)) = \mathbb{E}[\tilde{T}_q] \leq \mathbb{E}[\tau_{i^*}] \leq \frac{1}{0\,14p}\frac{m}{\alpha}\left(1 + \ln\left(\frac{p}{p-q}\right)\right) = \frac{7\,2m}{p\alpha}\left(1 + \ln\left(\frac{p}{p-q}\right)\right)$$

Recalling that the actual adoption dynamic with i.i.d. Poisson updating is m times faster than the discrete time dynamic, the statement of Theorem 3.1 follows. *Q.E.D.*

5. MULTIPLE TECHNOLOGIES

So far we have analyzed the case where there are two competing technologies: the status quo and the innovation. In practice, multiple innovations may be competing for acceptance at any given point in time. In this section, we explain how our results can be extended to this more general case.

Consider the case where the action set $A = \{1 \quad , k\}$ consists of k alternatives, including the status quo. We shall assume that the payoff matrix U has the following form:

$$
U = \quad
\begin{array}{c|c|c|c|c|}
 & 1 & 2 & \cdots & k \\
\hline
1 & a_1 + c & a_1 & \cdots & a_1 \\
\hline
2 & a_2 & a_2 + c & \cdots & a_2 \\
\hline
 & & & & \\
\hline
k & a_k & a_k & \cdots & a_k + c \\
\hline
\end{array}
$$

Here, a_i is the utility from choosing technology i and c is the utility from coordination, which for simplicity we assume is the same for all technologies. (The model can also accommodate heterogeneous coordination payoffs but the notation is more cumbersome.) We shall assume, without loss of generality, that $a_1 > a_2 > \cdots > a_k$. Let $_j$ be the idiosyncratic payoff shock from playing alternative j. We assume that the shocks $(\epsilon_j)_{j=1 \quad ,k}$ are nonatomic, i.i.d., and have mean zero.

Given a distribution $x = (x_1, ..., x_k) \in \Delta(A)$, let $r_i(x)$ be the probability that technology i is the best-reply alternative for an agent facing this distribution:

$$
r_i(x) = \mathbf{P} \ (Ux)_i \geq (Ux)_j \text{ for all } j \in A
$$

Let $r(x) = (r_1(x) \quad , r_k(x))$ denote the *multidimensional response function*.

For every value $x \in [0 \ 1]$, let $\rho(x)$ be the minimum probability that an agent adopts technology 1 when the proportion x of his neighbors have adopted technology 1:

$$
\rho(x) = \min_{x \in \Delta^k \text{ s.t. } x_1 = x} r_1(x). \tag{10}
$$

We replace the function r of the two technologies case with the function ρ. We now approximate ρ using a linear function L as described above. Note that for every vector of proportions x,

$$
r_1(x) \geq \rho(x_1) \geq L(x_1).
$$

We can therefore use Theorem 3.1 to approximate the expected waiting time until a proportion $q < p$ have adopted the superior technology 1.

Deriving the function ρ explicitly from the multidimensional response function r is not always straightforward. However, for a wide class of shock distributions used in empirical applications, ρ can easily be derived from r, as the following result shows.

LEMMA 3: *If the payoff shock distribution F is log concave, then for every $x \in [0 \ 1]$,*

$$
\rho(x) = r_1(x, 1 - x, 0 \qquad 0).
$$

In other words, if F is log concave, then the technology distribution that minimizes the switching probability to technology 1 across all distributions with $x_1 = x$ occurs when the proportion playing technology 2 equals $1 - x$.[15]

---

[15]For example, the extreme value distribution, which generates the logit response, is log concave, and so is the normal distribution, which generates the probit response. In contrast, the Cauchy and the Pareto distributions are not log concave.

Before providing the proof, note that

$$Ux = (a_1 + cx_1, a_2 + cx_2, ..., a_n + cx_n).$$

PROOF: Consider the case where the realized payoff shock for technology 1 equals ε_1. In this case the payoff from using technology 1 is $u_1 = a_1 + cx + \varepsilon_1$. It is sufficient to show that across all $x \in \Delta^k$ with $x_1 = x$, the vector $(x, 1 - x, 0 \quad 0)$ minimizes the probability

$$\mathbf{P}\,(Ux)_j + \varepsilon_j \le u_1 \text{ for } j = 2 \quad ,k \tag{11}$$

Note that the probability in (11) equals

$$F(u_1 - a_2 - cx_2) \cdots F(u_1 - a_k - cx_k). \tag{12}$$

We wish to find $x \in \Delta^k$ that minimizes (12) subject to $x_1 = x$. It follows from the log concavity of F that the product in (12) is also log concave. Since the domain, $\{x \in \Delta^k : x_1 = x\}$, is convex and compact it follows that the minimum is attained at an extreme point.

It remains to show that the extreme point at which the minimum is attained is $(x, 1 - x, 0 \quad 0)$. We contend that the log concavity of F implies that for every $j \in A \setminus \{1\ 2\}$ the function

$$g(z) = \ln\,F(u_1 - a_j + z) \;-\ln\,F(u_1 - a_2 + z)$$

is decreasing in z. To see this, note that $u_1 - a_j > u_1 - a_2$ for all $j \ge 3$. Since the function $h(z) = \ln(F(z))$ is concave, the slope

$$\frac{\ln\,F(u_1 - a_j + z) \;-\ln\,F(u_1 - a_2 + z)}{a_2 - a_j}$$

is decreasing in z. In particular, $g(0) \le g(-(1 - x)c)$. This implies that for every $j \in A \setminus \{1\ 2\}$,

$$F\,u_1 - a_2 - (1 - x)c\,F(u_1 - a_j) \le F(u_1 - a_2)F\,u_1 - a_j - (1 - x)c$$

Therefore, for every $j \in A \setminus \{1\ 2\}$

$$F\,u_1 - a_2 - (1 - x)c\,F(u_1 - a_3) \cdots F(u_1 - a_k)$$
$$\le F(u_1 - a_2) \cdots F\,u_1 - a_j - (1 - x)c\,\cdots F(u_1 - a_k).$$

This concludes the proof of the lemma. *Q.E.D.*

6. THE ROLE OF THE NETWORK

Our main result shows that, under certain conditions on the payoff shock distribution, the expected waiting time until the innovation is adopted by a large fraction of the population is uniformly bounded *for all networks of arbitrary size*. Moreover, the bound remains valid even when the network varies over time. In this section, we discuss the role that the network topology plays in the diffusion process when the network remains fixed. We focus on simple classes of networks such as regular networks and star networks and demonstrate *network-dependent* behavior in these cases.

We address three closely related questions. First, for regular networks and star networks, we derive bounds on the speed of convergence that improve on the bound in Theorem 3.1. Furthermore, we show that for star networks, rapid convergence holds even for values of q that are greater than the fixed point p.

Second, we study the question of persistence. Once the adoption process reaches a high proportion q of adopters, how likely is it that the proportion remains high for an extended period of time? As we shall see, the answer depends on both the size of the network and its topological structure. When the network is small, it only takes a handful of agents to revert to playing action 0 for the proportion to drop substantially below q. Indeed, reversion for any particular agent will occur with probability at least $1 - r(1)$, which will typically be positive. Therefore, for small networks we cannot expect that a high proportion of adopters will be maintained for very long. Even if the network is large, the degree of persistence is sensitive to the network topology. We shall show in particular that for large regular networks persistence is high, whereas for large star networks persistence is low (see Section 6.2).

Finally we ask, given a response function r, for which values of q is the expected waiting time unbounded? The answer depends on the topology of the network. In star networks, the expected waiting time to reach a target $q < r(1)$ is bounded above for all monotone increasing response functions (Proposition 6.2). In contrast, for regular networks, the expected waiting time to reach a target $q < r(q)$ increases exponentially in the degree of the network (Proposition 6.5).

6.1. *Convergence Rates for Specific Networks*

We begin by providing an exact formula for the convergence time of the linear dynamic in large regular networks. An undirected network G is *regular* if *all* agents have the same degree $d \geq 0$. As before, for every $q \in [0\ 1]$, we let $T_q(G)$ be the expected waiting time until at least the proportion q of the agents have adopted.

PROPOSITION 6.1: *Let* $r(x) = L(x) = (1 - \alpha)x + \alpha p$ *be a linear dynamic and let* $q \in [0, p)$. *For every* $m \geq 1$, *let* R_m *be an undirected, regular network with m nodes. Then* $\lim_{m \to \infty} T_q(R_m) = \frac{1}{\alpha} \ln(\frac{p}{p-q})$.

PROOF: We rely on stochastic approximation results due to Benaim (1998) and Roth and Sandholm (2013). Let $\{x_{m,t}\}$ be the discrete time process representing the proportion of adopters in R_m at time t. Let $\mathcal{F}_t$ be the sigma algebra that is generated by the sequence of actions of the agents in R_m up to and including time t. For any proportion $x \in [0\ 1]$, let $V(x) = \alpha(p - x)$. We contend that

$$E(x_{m,t+1} - x_{m,t} | \mathcal{F}_t) = \frac{V(x_{m,t})}{m}$$

In particular, the expected change in next period's proportion depends only on the current proportion and not on the network configuration.

To see this, consider the case where the current proportion $x_{m,t} = x$ and an agent is selected uniformly at random to revise his strategy choice. With probability α, the selected agent chooses action 1 with conditional probability p, and chooses action 0 with conditional probability $1 - p$. With probability $1 - \alpha$, the agent imitates the action of a randomly

chosen neighbor. Let $d(t)$ be the number of edges i, j where the actions of agents i and j are distinct. We can write

$$\mathbf{P}\left[x_{m,t+1} - x_{m,t} = \frac{1}{m} \middle| \mathcal{F}_t\right] = (1 - x_{m,t})\alpha p + (1 - \alpha)\frac{d(t)}{2|E|} \tag{13}$$

Similarly,

$$\mathbf{P}\left[x_{m,t+1} - x_{m,t} = -\frac{1}{m} \middle| \mathcal{F}_t\right] = \alpha x_{m,t}(1 - p) + (1 - \alpha)\frac{d(t)}{2|E|} \tag{14}$$

It follows from (13) and (14) that

$$E(x_{m,t+1} - x_{m,t}|\mathcal{F}_t) = \frac{\mathbf{P}\left[x_{m,t+1} - x_{m,t} = \frac{1}{m}\middle|\mathcal{F}_t\right] - \mathbf{P}\left[x_{m,t+1} - x_{m,t} = -\frac{1}{m}\middle|\mathcal{F}_t\right]}{m}$$

$$= \frac{V(x_{m,t})}{m} \tag{15}$$

Hence for any m and t we can write

$$x_{m,t+1} - x_{m,t} - \frac{U_{m,t+1}}{m} = \frac{V(x_{m,t})}{m} \tag{16}$$

where $U_{m,t+1} = m(x_{m,t+1} - x_{m,t}) - V(x_{m,t})$ is a bounded $\mathcal{F}_{t+1}$-measurable random variable. Furthermore, equations (15) and (16) imply that $E(U_{m,t+1}|\mathcal{F}_t) = 0$. Therefore, the process $\{(U_{m,t})_t\}_m$ satisfies the conditions of Proposition 2.3 in Roth and Sandholm (2013) and we can use Theorem 3.2 in Roth and Sandholm (2013) to approximate our process $\{(x_{m,t})_t\}_m$ by a differential equation as illustrated below.

Let $\{x_m(s)\}_{s\geq 0}$ be the continuous-time process defined by

$$x_m(s) = x_{m, \, ms}$$

For any $x \in [0 \; 1]$, let z_x be the solution of the following differential equation:

$$\dot{z} = L(z) - z = V(z) \qquad z(0) = x.$$

Theorem 3.2 in Roth and Sandholm (2013) implies that, for every $\delta > 0$, $x \in [0 \; 1]$ and $T > 0$,

$$\lim_{m\to\infty} \mathbf{P}\left[\sup_{s\in[0,T]} \left[x_m(s + t) - z_x(s)\right] \geq \delta \middle| x_m(t) = x\right] = 0$$

where the limit holds uniformly in x. Note that $z_0(s) = p(1 - \exp(-\alpha s))$ and for $s = \frac{1}{\alpha}\ln(\frac{p}{p-q})$ we have $z_0(s) = q$. In addition, for every $x \in [0 \; 1]$ and every $\epsilon > 0$ we have $z_x(s + \epsilon) > q$. Therefore, for every $\epsilon > 0$, $x \in [0 \; 1]$ and $t > 0$, $\lim_{m\to\infty} \mathbf{P}(x_m(s + t + \epsilon) < q|x_m(t) = x) = 0$, and hence

$$\lim_{m\to\infty} T_q(R_m) = \frac{1}{\alpha}\ln\left[\frac{p}{p - q}\right]$$

as was to be shown.   Q.E.D.

Next, we turn to estimating the expected convergence time for large star networks. Let S_m consist of m vertices with central vertex 1 and edges $\{1, j\}$ for all $j \in [m]$ such that $j > 1$. We shall show that for every monotone increasing response function, rapid convergence can hold for values $q > p$ beyond the fixed point.

Let $T_q(S_m)$ be the expected waiting time to reach at least qm adopters in the star network with m agents.

PROPOSITION 6.2: *Consider any star network S_m and an adoption dynamic that is governed by a monotone increasing response function $r : [0 \ 1] \to [0 \ 1]$. For every $q < r(1)$,*

$$\limsup_{m \to \infty} T_q(S_m) \leq \frac{r(1)}{r(1) - q} \left[\frac{1}{r(0)} + \ln \left(\frac{r(1)}{r(1) - q} \right) \right]$$

PROOF: Consider the following event: agent 1 updates his action to 1 and thereafter $\frac{q}{r(1)}m$ agents receive a revision opportunity. We claim that the expected waiting time for this event is bounded above by

$$\frac{1}{r(0)} + \ln \left(\frac{r(1)}{r(1) - q} \right)$$

Indeed, $\frac{1}{r(0)}$ bounds the expected waiting time for agent 1 to switch to action 1, and $\ln(\frac{r(1)}{r(1)-q})$ bounds the expected waiting time for at least $\frac{mq}{r(1)}$ agents to have a revision opportunity (this is an instance of the coupon collector problem; see, e.g., Section 3.6 in [Motwani and Raghavan (1995)](#)). Note that with probability $\frac{r(1)-q}{r(1)}$ agent 1 is not part of the fraction $\frac{q}{r(1)}$ of agents who receive a revision opportunity.

Now consider the following event: agent 1 updates his action to 1 and thereafter at least $\frac{qm}{r(1)}$ of the agents, not including 1, receive a revision opportunity. By the preceding, the expected waiting time for this event is at most

$$\frac{r(1)}{r(1) - q} \left[\frac{1}{r(0)} + \ln \left(\frac{r(1)}{r(1) - q} \right) \right]$$

Each of the $\frac{qm}{r(1)}$ agents who receives an updating opportunity chooses action 1 with probability $r(1)$ because during their revision time agent 1 is playing action 1. Therefore, conditional on this event, the probability of reaching a threshold of $q - \varepsilon$ adopters approaches one as m goes to infinity, for every $\epsilon > 0$. This establishes Proposition 6.2. *Q.E.D.*

We note, in particular, that for a linear dynamic on star networks, the determinant of the fast adoption regime is $L(1)$ and not the fixed point $p < L(1)$. Hence even when q is close to p, the expected waiting time to reach at least qm adopters is much faster in large star networks than in large regular networks.

6.2. *The Persistence of Innovation*

In this section, we show that innovation is persistent for large regular networks, whereas it is not persistent for star networks.

Recall that p denotes the fixed point of the linear dynamic. Fix $q', q \in (0, p)$ with $q' < q$ and consider the expected waiting time for the population to reach a state with at most

q m adopters from a state with at least qm adopters. Say that innovation is *persistent* for a class of networks if this expected waiting time grows to infinity with the population size m *for any* such pair q, q.

Formally given a pair $q < q < p$ and a network G, let $I_{q,q}(G)$ be the minimal expected waiting time to reach a state in which the proportion of adopters is q or less, where the minimum is taken across all initial conditions with proportion at least q. We next show that innovation is persistent for the class of regular networks.

PROPOSITION 6.3: *Let R_m be a regular network with m agents. For every $0 < q < q < p$, there exists a constant $c = c(q\ q) > 1$ (independent of m) such that $I_{q,q}(R_m,) \geq c^m$.*

It follows that the expected waiting time until the fraction of adopters is lower than q, starting from a state where the fraction is at least $q > q$, grows exponentially with the number of agents m. (The proof is contained in the Appendix). In contrast, for the class of star networks, persistence fails and the adoption dynamic escapes from any targeted proportion relatively quickly, as the following proposition shows.

PROPOSITION 6.4: *Let S_m be a star network with m agents and let $r(x)$ be a response function with $r(1) < 1$ and $r(0) < \frac{1}{2}$. There exists a constant $c > 0$ (independent of m) such that $I_{1\ \frac{1}{2}}(S_m) \leq c$.*

Thus, even if we start in the state where all agents have adopted, the expected waiting time to reach a state with a proportion of adopters below $\frac{1}{2}$ is bounded above.

PROOF: The proof is similar to the the proof of Proposition 6.2. Let $\gamma = \frac{3-2r(0)}{4(1-r(0))}$. Note that since $r(0) < 1/2$ we have $\frac{\gamma - 1/2}{\gamma} > r(0)$ and $\gamma < 1$. Consider the following event: agent 1 updates his action to 0 and thereafter a fraction γ of agents receive a revision opportunity. As above, the expected waiting time for this event is bounded above by

$$\frac{1}{1 - r(1)} + \ln\ \frac{1}{1 - \gamma}$$

Note that with probability $1 - \gamma$ agent 1 is not part of the fraction $\frac{q}{r(1)}$ of agents who receive a revision opportunity.

Therefore, the expected waiting time of the following event: agent 1 updates his action to 0 and thereafter a fraction γ of the agents, *which does not include agent* 1, receive a revision opportunity is

$$\frac{1}{(1 - \gamma)}\ \frac{1}{1 - r(1)} + \ln\ \frac{1}{1 - \gamma}$$

Each of the γm agents who receives an updating opportunity chooses action 1 with probability $r(0)$ because during their revision time agent 1 is playing action 0. By the choice of γ, there exists a positive probability $w > 0$ that at most $\frac{\gamma - 1/2}{\gamma} > r(0)$ of the γm agents have chosen action 1, independently of m. If this holds, the number of adopters is at most $\gamma \frac{\gamma - 1/2}{\gamma} + (1 - \gamma) \leq 1/2$.

This shows that

$$I_{1\ \frac{1}{2}} \leq \frac{1}{w(1 - \gamma)}\ \frac{1}{1 - r(1)} + \ln\ \frac{1}{1 - \gamma}$$

which establishes Proposition 6.4.   *Q.E.D.*

6.3. *Slow Convergence*

A crucial assumption for our main result (Theorem 3.1) is that the supporting line $L(x)$ of the response function $r(x)$ must intersect the y-axis at a positive value. If this is not the case, the corresponding imitation dynamic is not well-defined. In this section, we ask, given a monotone increasing response function $r(\cdot)$, when does fast convergence fail to hold for a given target q? A natural candidate would be any value of q such that $r(q) < q$, for then the adoption process gets caught in a bottleneck. In particular, this situation arises when r is convex-concave and is too sharply bowed in the convex part (see Figure 3 for an example). Proposition 6.2 shows, however, that this is not a sufficient condition: star networks exhibit fast convergence for any value $q < r(1)$. In this section, we show that the condition $r(q) < q$ does lead to slow convergence for large regular networks; in particular, the expected waiting time to reach such a target q grows exponentially with the degree of the network.

PROPOSITION 6.5: *Let R_m be a sequence of undirected d_m-regular networks, and let the response function be monotone increasing with $r(q) < q$. There exists a constant $c > 0$ (independent of m) such that $T_q(R_m) \geq \frac{e^{cd_m}}{m}$.*

PROOF IDEA: Instead of considering the first time τ when the fraction of adopting agents reaches a threshold q, we study the first time $\tilde{\tau}$ where there exists an agent in the network who has a fraction of at least q adopting neighbors. (The regularity of the network implies that $\tilde{\tau} \leq \tau$.)

We rely on the fact that prior to time $\tilde{\tau}$, all updating agents in all periods revise their action to 1 with probability at most $r(q)$. (This follows from the monotonicity of r.) From the perspective of a single player, his neighbors update their actions in an i.i.d. manner with probability of success $c \leq r(q)$. We can therefore deduce (using Lemma 1 in Benaim and Weibull (2003)) that the expected waiting time ro reach a fraction q of adopters in a d_m-regular network with m agents is bounded by $\frac{e^{cd_m}}{m}$ for some constant $c > 0$.   Q.E.D.

Proposition 6.5 does not hold for regular networks of *fixed* degree d. To see this, note that for every constant d, the union of $\frac{m}{d+1}$ cliques of size $d+1$ has the fast-convergence property. Since the mixing time T of a single clique of size $d+1$ is constant, after time T we will obtain a large number of cliques whose configurations are distributed according to the invariant distribution, and this fraction will be very close to the expected fraction under the invariant distribution (i.e., a high fraction of adopters).

7. CONCLUSION

In this paper, we have established an upper bound on the expected waiting time until an innovation is adopted by a large fraction of a given population. The formula for the bound is universal in the sense that it holds for all directed and undirected networks of arbitrary size and degree distribution. Moreover, the bound holds when the network itself is evolving over time. Previous results on this topic rely on the existence of a potential function and hold only for regular networks with symmetric interactions, as in Kreindler and Young (2014). We have also established waiting time bounds for the diffusion of multiple innovations instead of just a single innovation, which is the usual case treated in the literature.

Our bound holds for a wide variety of perturbed best response processes where agents choose optimal responses under random payoff shocks. Particular cases include errors

that are normally or extreme-value distributed, for example. These models are consistent with empirical evidence on subjects' learning behavior in network games (Mäs and Nax (2016)). The formula for the bound is expressed in terms of the slope and intercept of a suitably chosen 'linearization' of the perturbed response process. In particular, the expected waiting time to reach a given proportion of adopters is inversely proportional to the initial amount of 'noise' in the linear process when no one has yet adopted ($L(0)$ in expression (3)).

Given more information on the topological structure of the network, one can obtain significantly tighter bounds on the expected waiting time, as we showed in Section 6. The usefulness of our main result arises from the fact that in practice it is very difficult to observe the topology of interactions and the degree of influence that different actors exert on one another. By contrast, it may be possible to estimate the response probabilities of agents conditional on the choices of their neighbors using longitudinal data. From this, one can derive the linearized process and the upper bound on the expected waiting time to reach a given adoption threshold.

APPENDIX A: FEIGE'S INEQUALITY AND LEMMA 1

The following theorem is due to Feige (2006).

THEOREM A.1: *Let* $X_1,...,X_n$ *be nonnegative independent random variables, with* $E[X_i] \leq 1 \, \forall i$, *and let* $X = \sum_{i=1}^{n} X_i$. *Then for any* n,

$$\mathbf{P}\big(X < E[X]+1\big) \geq \gamma > 0$$

for some $\gamma \geq 1/13$.

Garnett (2020) improved upon the constant γ and show that $\gamma \geq 0.14$. We next prove Lemma 1 from the above theorem.[16]

PROOF OF LEMMA 1: Let $c_1,...,c_k$ be i.i.d. Bernoulli(p), and let $\beta_1,...,\beta_k$ be positive real numbers. We can assume without loss of generality that $\beta_i \leq \frac{1}{1-p}$ for all $i \in [k]$, and that $\beta_1 = 1/(1-p)$. Let $X_i = \beta_i(1-c_i)$, $m_i = \beta_i(1-p)$, $X = X_2 + \cdots + X_k$, and $m = m_2 + \cdots + m_k$. We must show that $\mathbf{P}(\sum_{i\in[k]} \beta_i c_i \geq \sum_{i\in[k]} \beta_i p) = \mathbf{P}(X_1 + X \leq m+1) \geq \gamma p$. Indeed, by the above theorem,

$$\mathbf{P}(X_1 + X \leq m+1) \geq \mathbf{P}(X_1 = 0)\mathbf{P}(X < m+1) \geq \gamma p \qquad Q.E.D.$$

A related result due to Billingsley ((2008), Theorem 9.2) says the following.

THEOREM: *If* $E(Z) = 0$, $E(Z^2) = s^2$, *and* $E(Z^4) = \xi^4$, *then* $\mathbf{P}(Z \geq 0) \geq \frac{s^4}{4\xi^4}$.

To apply this in our setting, let $c_1,...,c_k$ be i.i.d. Bernoulli(p) and let $\beta_1,...,\beta_k \geq 0$. Let $Z = \sum_{i=1}^{k} \beta_i(c_i - p)$. Note that $\mathbf{P}(Z \geq 0) = \mathbf{P}(\sum_{i=1}^{k} \beta_i c_i \geq (\sum_{i=1}^{k} \beta_i)p)$. The preceding

---

[16]The proof of Lemma 1 was first communicated to us by Terence Tao (see https://mathoverflow.net/questions/278117/convex-combination-iid-bernoulli-random-variables). We thank Yuval Peres for pointing out to us the result by Feige (2006).

theorem shows that

$$\mathbf{P}(Z \geq 0) \geq \frac{p(1-p)\left(\sum_{i=1}^{k}\beta_i^2\right)^2}{4\left[\left(1-3p+3p^2\right)\sum_{i=1}^{k}\beta_i^4 + 3p(1-p)\sum_{i,j:i\neq j}\beta_i^2\beta_j^2\right]}.$$

For values of p that are close to $\frac{1}{2}$, Billingsley's inequality provides a better bound than the one by Feige and Garnett. However, for p close to 1 the right-hand side is close to zero whereas the corollary to Feige's lemma shows that $\mathbf{P}(Z \geq 0) \geq 0.14p$. For certain values of p, the bound of Theorem 3.1 can be further improved. Concretely, a result by Garnett (2020, Theorem 1.2), can be applied to establish the bound $\min\{\frac{1}{6}, \frac{p(1-p)}{2(1-3p+3p^2)}\}$ for $p \geq \frac{1}{2}$. This bound is slightly better than 0.14 for $p < 0.835$. To see this let $X = \sum_i \beta_i(p - c_i)$. By Garnett's theorem $\mathbf{P}(X < 0) \geq \frac{1}{2c}$ provided $E(X^4) \leq c[E(X^2)]^2$ and $E(X^3) \geq 0$. The latter holds for $p \geq \frac{1}{2}$. Now $\frac{E(X^4)}{[E(X^2)]^2} = \lambda\frac{1-3p+3p^2}{p(1-p)} + (1-\lambda)3$ where $\lambda = \frac{\sum_i \beta_i^4}{(\sum_i \beta_i^2)^2}$. Hence we may take $c = \max\{3, \frac{(1-3p+3p^2)}{p(1-p)}\}$ from which the bound follows.[17] Since we prefer not to limit the range of p, however, the Feige–Garnett inequality is more useful for our purposes.

APPENDIX B: PROOFS OF PROPOSITIONS 6.3 AND 6.5

PROPOSITION 6.3: *Let R_m be a regular network with m agents. For every $0 < q' < q < p$, there exists a constant $c = c(q, q') > 1$ (independent of m) such that $I_{q,q'}(R_m,) \geq c^m$.*

PROOF OF PROPOSITION 6.3: As in the proof of Theorem 3.1, we analyze the discrete-time linear dynamic. Recall that $1 - \alpha = L(1) - L(0)$, and that $s(t)$ represents the imitation dynamic: with probability α the updating agent tosses a coin with success probability p, and with probability $1 - \alpha$ he imitates a random neighbor.

If we prove that the expected time to reach the fraction q under the discrete-time dynamic is at least $\hat{c}^m$ for some $\hat{c} > 1$, then the original dynamic satisfies the exponential bound $\hat{c}^m/m$. (The factor m is a result of changing from discrete to continuous time.) Note that for all sufficiently small $1 < c < \hat{c}$ we have $\hat{c}^m/m \geq c^m$, hence the exponential lower bound on the discrete-time dynamic provides an exponential lower bound on the original dynamic.

Given $s(t) \in \{0,1\}^m$, let $z(t) = \sum_i s_i(t) \in \mathbb{N}$ be the number of adopters. We denote by $d(t) = |\{[i,j] \in E : s_i(t) = 0, s_j(t) = 1\}|/|E| \in [0,1]$ the fraction of "disagreement" edges in the graph, where the agents are playing opposite actions.

The process $z(t)$ is a random walk with $z(t+1) - z(t) \in \{-1, 0, 1\}$. We shall provide an explicit formula for the probability that $z(t)$ moves to the left, to the right, or stays put. Given $s(t)$ one agent is drawn at random in period $t + 1$. The probability is α that this agent tosses a coin. Conditional on the coin toss, the agent switches from 0 to 1 with probability $(1 - z(t)/m)p$. (The first term is the probability that he was initially choosing 0, and the second term is the probability that he updates to 1). With probability $1 - \alpha$, the agent imitates. Let $d(t)$ be the number of edges $\{i, j\}$ where the actions of agents i and j

---

[17] We thank an anonymous referee for pointing this out.

are different. Then

$$\mathbf{P}\big(z(t+1)-z(t)=1|s(t)\big)=\alpha\big(1-z(t)/m\big)p+(1-\alpha)d(t)/2 \tag{17}$$

Similarly, we deduce that

$$\mathbf{P}\big(z(t+1)-z(t)=-1|s(t)\big)=\alpha\big(z(t)/m\big)(1-p)+(1-\alpha)d(t)/2 \tag{18}$$

For $z(t)\in[mq\,,mq]$, the ratio between moving left and right can be bounded by

$$\frac{\mathbf{P}\big(z(t+1)-z(t)=-1|s(t)\big)}{\mathbf{P}\big(z(t+1)-z(t)=1|s(t)\big)}$$

$$=\frac{\alpha\big(z(t)/m\big)(1-p)+(1-\alpha)d(t)/2}{\alpha\big(1-z(t)/m\big)p+(1-\alpha)d(t)/2}$$

$$\leq\frac{\alpha q(1-p)+(1-\alpha)d(t)/2}{\alpha(1-q)p+(1-\alpha)d(t)/2}\leq\frac{\alpha q+1}{\alpha p+1}$$

We denote this constant by $\beta=\frac{\alpha q+1}{\alpha p+1}<1$. Given $s(t)$ with $z(t)=mq-1$, we want to estimate the probability that the process will reach the low value $z=nq$ before it visits $z=nq$. This question has a precise answer for any biased random walk with bias $\beta<1$ (see Feller (2008) Chapter XIV.2). Namely, if we let $n=(q-q)m$, the probability of reaching $z=mq$ is $\frac{\beta^{n-1}-\beta^{n}}{1-\beta^{n}}$, which is bounded above by β^{n-1}. Therefore, the expected time until our process reaches $z(t)=q\,m$ is at least $(\frac{1}{\beta})^{n-1}\geq\hat{c}^{m}$ for some sufficiently small constant $\hat{c}>1$. $\qquad$ *Q.E.D.*

PROPOSITION 6.5: *Let R_m be a sequence of undirected d_m-regular networks, and let the response function be monotone increasing with $r(q)<q$. There exists a constant $c>0$ (independent of m) such that $T_q(R_m)\geq\frac{e^{cd_m}}{m}$.*

PROOF OF PROPOSITION 6.5: Consider a d_m-regular graph with m agents. Let $\tau:=\min\{\frac{1}{m}\sum_i s_i(t)\geq q\}$ be the first time such that the fraction of adopters is at least q. Let $\tilde{\tau}$ be the first time such that there exists an agent i with at least $d_m q$ neighbors who have adopted. It follows from the regularity of the network that $\tilde{\tau}\leq\tau$. We shall bound $E(\tau)$ from below by providing a lower bound for $E(\tilde{\tau})$.

It follows from the monotonicity of r that, prior to time $\tilde{\tau}$, any agent that updates his strategy chooses action 1 with probability at most $r(q)$. Thus in order to bound $E(\tilde{\tau})$ from below we can assume that, prior to $\tilde{\tau}$, all agents choose action 1 with probability $r(q)$. Assume that at time $t<\tilde{\tau}$ the fraction of i's neighbors that have adopted lies weakly below $\frac{r(q)+q}{2}$. Consider the following two events: (a) at time $t+1$ the fraction of i's neighbors who have adopted lies above $\frac{r(q)+q}{2}$; (b) there exists a time $s\in[t,t+1]$ where the fraction of i's neighbors who have adopted exceeds q. It follows from Lemma 1 in Benaim and Weibull (2003) that if all neighbors of i choose action 1 with probability $r(q)$ at every revision opportunity, then the probability that either (a) or (b) holds is at most e^{-cd_m} for some constant $c>0$.

Now assume that the fraction of adopting neighbors of *every* agent i at time $t<\tilde{\tau}$ lies below $\frac{r(q)+q}{2}$. Then the probability that there exists an agent i for which either (a) or (b)

holds is at most me^{-cd_m}. If neither (a) nor (b) holds for every agent at every time period $t = 0,...,n$, then $\tilde{\tau} > n + 1$. It follows that $E(\tilde{\tau}) \geq \frac{e^{cd_m}}{m}$. This concludes the proof of Proposition 6.5. *Q.E.D.*

REFERENCES

ACEMOGLU, D., M. Z. DAHLEH, I. LOBEL, AND A. OZDAGLAR (2011): "Bayesian Learning in Social Networks," *Review of Economic Studies*, 78, 1201–1236. [571]

ARIELI, I., AND H. P. YOUNG (2016): "Stochastic Learning Dynamics and Speed of Convergence in Population Games," *Econometrica*, 84, 627–676. [570]

BALA, V., AND S. GOYAL (1998): "Learning From Neighbours," *Review of Economic Studies*, 65 (3), 595–621. [571]

BANERJEE, A. (1992): "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107, 797–817. [571]

BANERJEE, A., A. G. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2013): "The Diffusion of Microfinance," *Science*, 341 (6144). [570]

BENAIM, M. (1998): "Recursive Algorithms, Urn Processes and Chaining Number of Chain Recurrent Sets," *Ergodic Theory and Dynamical Systems*, 18 (1), 53–87. [584]

BENAIM, M., AND J. W. WEIBULL (2003): "Deterministic Approximation of Stochastic Evolution in Games," *Econometrica*, 71 (3), 873–903. [588,591]

BIKHCHANDANI, J. S., D. HIRSCHLEIFER, AND I. WELCH (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades," *Journal of Political Economy*, 100, 992–1026. [571]

BILLINGSLEY, P. (2008): *Probability and Measure*. John Wiley & Sons. [578,589]

BLUME, L. E. (1993): "The Statistical-Mechanics of Strategic Interactions," *Games and Economic Behavior*, 5, 387–424. [571]

BONGAARTS, J., AND S. C. WATKINS (1996): "Social Interactions and Contemporary Fertility Transitions," *Population and Development Review*, 22 (4), 639–682. [571]

BRIAN, A. W. (1994): *Increasing Returns and Path Dependence in the Economy*. Ann Arbor, MI: Univ. of Michigan Press. [571]

BURKE, M. A., G. M. FOURNIER, AND K. PRASAD (2007): "The Diffusion of a Medical Innovation: Is Success in the Stars?" *Southern Economic Journal*, 73, 588–603. [571]

BURKE, M. A., G. M. FOURNIER, AND L. PRASAD (2010): "Geographic Variations in a Model of Physician Reatment Choice With Social Interactions," *Journal of Economic Behavior and Organization*, 73, 418–432. [571]

CENTOLA, D., AND A. BARONCHELLI (2015): "The Spontaneous Emergence of Conventions: An Experimental Study," *PNAS*, 112 (7), 1989–1994. [570]

CHANDRA, A., AND D. STAIGER (2007): "Productivity Spillovers in Health Care: Evidence From the Treatment of Heart Attacks," *Journal of Political Economy*, 115, 103–140. [571]

COLEMAN, J., E. KATZ, AND H. MENZEL (1957): "The Diffusion of an Innovation Among Physicians," *Sociometry*, 20, 253–270. [570]

DAVID, P. A. (1985): "Clio and the Economics of QWERTY," *American Economic Review Papers and Proceedings*, 75, 332–337. [571]

ELLISON, G. (1993): "Learning, Social Interaction, and Coordination," *Econometrica*, 61, 1047–1071. [570]

ELLISON, G., AND D. FUDENBERG (1993): "Rules of Thumb for Social Learning," *Journal of Political Economy*, 101, 612–643. [571]

ELLISON, G., D. FUDENBERG, AND L. IMHOF (2016): "Fast Convergence in Evolutionary Models: A Lyapunov Approach," *Journal of Economic Theorey*, 161, 1–36. [570]

FEIGE, U. (2006): "On Sums of Independent Random Variables With Unbounded Variance, and Estimating the Average Degree in a Graph," *Siam Journal on Computing*, 35 (4), 964–984. [572,578,589]

FELLER, W. (2008): *An Introduction to Probability Theory and Its Applications,* Vol. 1. John Wiley & Sons. [591]

FOSTER, A., AND M. ROSENZWEIG (1995): "Learning by Doing and Learning From Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 103, 1176–1209. [570]

——— (2010): "Microeconomics of Technology Adoption," *Annual Review of Economic*, 2 (1), 395–424. [570]

GALE, D., AND S. KARIV (2003): "Bayesian Learning in Social Networks," *Games and Economic Behavior*, 45, 329–346. [571]

GARNETT, B. (2020): "Small Deviations of Sums of Independent Random Variables," *Journal of Combinatorial Theory, Series A*, 169, 105119. [572,578,589,590]

GOLUB, B., AND M. O. JACKSON (2010): "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2, 112–149. [571]

——— (2012): "How Homophily Affects the Speed of Learning and Best-Response Dynamics," *Quarterly Journal of Economics*, 1287 (1338). [571]

GRILICHES, Z. (1957): "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, 25, 501–522. [570]

JACKSON, M. O. (2008): *Social and Economic Networks*. Princeton, NJ: Princeton University Press. [571]

JACKSON, M. O., AND A. WATTS (2002): "The Evolution of Social and Economic Networks," *Journal of Economic Theory*, 106, 265–295. [572]

JACKSON, M. O., AND L. YARIV (2007): "Diffusion of Behavior and Equilibrium Properties in Network Games," *American Economic Review Papers and Proceedings*, 97, 92–98. [571]

KOHLER, H.-P. (1997): "Learning in Social Networks and Contraceptive Choice," *Demography*, 34, 369–383. [570,571]

KOHLER, H.-P., J. R. BEHRMAN, AND S. C. WATKINS (2001): "The Density of Social Networks and Fertility Decisions: Evidence From South Nyanza District, Kenya," *Demography*, 38, 43–58. [570,571]

KREINDLER, G. E., AND H. P. YOUNG (2013): "Fast Convergence in Evolutionary Equilibrium Selection," *Games and Economic Behavior*, 80, 39–67. [571]

——— (2014): "Rapid Innovation Diffusion in Social Networks," *Proceedings of the National Academy of Sciences*, 111, 10881–10888. [569-572,575,588]

LIGGETT, T. M. (1999): *Stochastic Interacting Systems: Contact, Voter, and Exclusion Process*. Berlin: Springer-Verlag. [578]

MÄS, M., AND H. H. NAX (2016): "A Behavioral Study of Noise in Coordination Games," *Journal of Economic Theory*, 162, 195–208. [571,573,589]

MONTANARI, A., AND A. SABERI (2010): "The Spread of Innovations in Social Networks," *Proceedings of the National Academy of Sciences*, 107, 20196–20201. [570]

MONTGOMERY, M. R., AND J. B. CASTERLINE (1996): "Social Influence, Social Learning, and New Models of Fertility," *Population Development Review*, 22, 151–175. [571]

MORRIS, S. (2000): "Contagion," *Review of Economic Studies*, 67, 57–78. [570]

MOSSEL, E., A. SLY, AND O. TAMUZ (2015): "Strategic Learning and the Topology of Social Networks," *Econometrica*, 83, 1755–1794. [571]

MOTWANI, R., AND P. RAGHAVAN (1995): *Randomized Algorithms*. Cambridge University Press. [586]

MUELLER-FRANK, M. (2013): "A General Framework for Rational Learning in Social Networks," *Theoretical Economics*, 8 (1), 1–40. [571]

MUELLER-FRANK, M., AND M. PAI (2016): "Social Learning With Costly Search," *American Economic Journal: Microeconomics*, 83 (109). [571]

MUNSHI, K., AND J. MYAUX (2006): "Social Norms and the Fertility Transition," *Journal of Development Economics*, 80, 1–38. [570,571]

PHELPS, C. E., AND C. MOONEY (1993): "Variations in Medical Practice Use: Causes and Consequences," in *Competitive Approaches to Health Care Reform*, ed. by R. J. Arnould, R. F. Rich, and W. White. Washington, DC: Urban Institute, 139–178. [571]

ROGERS, E. (2003): *Diffusion of Innovations* (Fifth Ed.). New York: Simon and Schuster. [570]

ROTH, G., AND W. H. SANDHOLM (2013): "Stochastic Approximations With Constant Step Size and Differential Inclusions," *SIAM Journal on Control and Optimization*, 51 (1), 525–555. [584,585]

RYAN, B., AND N. C. GROSS (1943): "The Diffusion of Hybrid Seed Corn in Two Iowa Communities," *Rural Sociology*, 8, 15–24. [570]

SKYRMS, B., AND R. PEMANTLE (2000): "A Dynamic Model of Social Network Formation," *Proceedings of the National Academy of Sciences of USA*, 97, 9340–9346. [572]

SOLAN, E., D. ROSENBERG, AND N. VIEILLE (2009): "Informational Externalities and Emergence of Consensus," *Games and Economic Behavior*, 66 (2), 979–994. [571]

UDRY, C., AND T. CONLEY (2001): "Social Learning Through Networks: The Adoption of New Agricultural Technologies in Ghana," *American Journal of Agricultural Economics*, 83, 668–673. [570]

VALENTE, T. (1995): *Network Models of the Diffusion of Innovations*. Cresskill, NJ: Hampton Press. [570]

——— (2005): "Network Models and Methods for Studying the Diffusion of Innovations," in *Models and Methods in Social Network Analysis*. Cambridge: Cambridge University Press. [570]

VEGA-REDONDO, F. (2007): *Complex Social Networks*. Cambridge: Cambridge University Press. [571,572]

WATTS, D. (2002): "A Simple Model of Global Cascades on Random Networks," *Proceedings of the National Academy of Sciences*, 99 (9), 5766–5771. [571]

WENNBERG, J., AND A. GITTELSOHN (1973): "Small Area Variations in Health Care Delivery," *Science*, 182, 1102–1108. [571]

YOUNG, H. P. (1998): *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton, NJ: Princeton University Press. [570]

———— (2009): "Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning," *American Economic Review*, 99 (5), 1899–1924. [572]

———— (2011): "The Dynamics of Social Innovation," *Proceedings of the National Academy of Sciences*, 108 (4), 21285–21291. [570]