The Speed of Innovation Diffusion

Itai Arieli[*]   Yakov Babichenko[†]   Ron Peretz[‡]

H. Peyton Young[§]

December 17, 2018

Abstract

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, however, 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 [33, 2014] demonstrated such a bound for regular networks when agents choose between two options: the innovation and the status quo. Our

[*]Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology, iarieli@tx.technion.ac.il.

[†]Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology, yakovbab@technion.ac.il.

[‡]Department of Economics, Bar Ilan University, ron.peretz@biu.ac.il.

[§]London School of Economics and Nuffield College, Oxford, peyton.young@economics.ox.ac.uk.

bound holds for directed and undirected networks of arbitrary size and degree distribution, and for multiple competing innovations with different payoffs.

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 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

2

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 [33, 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 contribution of this paper is to substantially generalize this result by establishing an upper bound on the expected waiting time that holds for networks of any size and degree distribution, whether directed or undirected. We also generalize the analysis by considering multiple competing innovations instead of a single innovation versus the status quo, which is the usual assumption in the literature.

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 [47, 1943] demonstrated that farmers' decisions to adopt an agricultural innovation — hybrid corn – were strongly influenced by the adoption decisions of their neighbors. Subsequently Griliches [26, 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 [14, 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

---

[1]There is an extensive empirical literature on innovation diffusion in networks. See among others Valente [51, 1995]; Foster and Rosenzweig [20, 1995]; Kohler [30, 1997]; Kohler et al. [31, 2001]; Udry and Conley [50, 2001]; Rogers [46, 2003]; Munshi [43, 2004].

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 [16, 1993]; Young [56, 1998]; and [57, 2009]; Montanari and Saberi [39, 2010]). Experimental studies of games played on networks are consistent with these predictions (Centola et al. [12, 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 [38, 2000]; Banerjee et al. [6, 2013]). These are instances of threshold models, in which a given individual adopts once a sufficient number of his neighbors have adopted (Watts [54, 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; 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. A key 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 behaviour of subjects who are embedded in

---

[2]Targeting small close-knit groups is a strategy that policy makers sometimes use to try to overturn inferior or harmful practices, such as foot binding in China and female genital mutilation practices in Africa (Mackie [35, 1996]).

[3]See Banerjee [5, 1992]; Bikchandani et al. [7, 1992]; Ellison and Fudenberg [17, 1993]; Bala and Goyal [4, 1998]; Jackson [29, 2008]; Solan et al. [49, 2009]; Golub and Jackson

different types of networks has been investigated experimentally by Gale and Kariv [21, 2003] and Maes and Nax [36, 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 adopt 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 net-

[23, 2010] [24, 2012]; Acemoglu et al. [1, 2011]; Mueller-Frank [41, 2013]; Mueller-Frank and Pai [42, 2016]; Mossel et al. [40, 2015].

[4]David [15, 1993] and Arthur [3, 1994].

[5]See Bongaarts and Watkins [9, 1996]; Montgomery and Casterline [37, 1996]; Kohler [30, 1997]; Kohler et al. [31, 2001]; Munshi and Myaux [43, 2006].

[6]See Wennberg and Gittelsohn [55, 1973]; Phelps and Mooney [45, 1993]; Chandra and Staiger [13, 2007]; Burke et al. [10, 2007], [11, 2010].

work game in which individuals repeatedly play a coordination game against their neighbors (Blume [8, 1993]; Jackson and Yariv [28, 2007]; Jackson [29, 2008]; Vega-Redondo [53, 2007]; Golub and Jackson [23, 2008]; [23, 2010]). 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. Given the payoff parameters and the distribution of random shocks, we 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.

A similar approach was taken by Kreindler and Young [32, 2013], [33, 2014]); indeed the present paper was inspired by their work. Using martingale theory, they show that when the perturbations are sufficiently noisy and/or the payoff gain from the innovation is sufficiently large, the expected time it takes to reach a given target level of adoption (say one-half) can be bounded above for all regular undirected networks. In the case of irregular networks, they establish a similar bound for reaching a given degree-weighted adoption rate, but not the adoption rate as such.[7]

Here we employ different methods to establish a much more general bound on the waiting time to reach a given target (weighted or unweighted) 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

---

[7]Ellison, Fudenberg, and Imhof [18, 2016] establish criteria for rapid convergence in more general games using Lyapunov functions. Using different techniques, Arieli and Young [2, 2016] bound the rate of convergence of stochastic learning dynamics when agents interact globally instead of locally.

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 [19, 22].

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.

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. 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 A, or the status quo B. 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 coordi-

---

[8]There is a large literature on the dynamics of link formation in social networks. See among others Skyrms and Pemantle [48, 2000]; Jackson and Watts [27, 2002]; Goyal and Vega-Redondo [53, 2007]; Jackson [27, 2008].

nation payoff that results from coordinating with others. For an interaction with a single neighbour, we represent the payoffs by the following 2×2 matrix

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

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 $c < a$ 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 n-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 $s(0) = (0,\dots,0)$. We suppose that every agent receives updating opportunities according to a Poisson arrival process with rate one per time period, and suppose that 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 $u_i(1) = a + c x_i(t)$ if he chooses action 1, and $u_1(0) = c(1 - x(t))$ 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 $e_i(t)$ with c.d.f. $F(e)$. For ease of interpretation we shall assume that F has a continuous density $f(e)$ that is symmetrically

8

distributed about the origin, and 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 distributed according to the extreme value distribution. Moreover, 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 [36, 2016]).

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

$$Pr[s_i(t) = 1 | x_i(t)] = Pr[a + e_{it} + cx_i(t) > c(1 - x_i(t))] \tag{1}$$
$$= Pr[e_{it} > c - a - 2cx_i(t)] = F(a - c + 2cx_i(t)),$$

where the latter holds because by assumption $F(e) = 1 - F(-e)$. We shall call this the *response function* and write

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

In particular, $r(0) = F(a - c)$ and $r(1) = F(a + c)$. To illustrate, suppose that $F(e)$ is the standard cumulative normal distribution Φ as shown in Figure 1.

More generally, suppose that F has a density $f(e)$ that is symmetric about the origin and is non-increasing in $|e|$, that is, smaller errors are at least as probable as larger errors. Then $r(x)$ is increasing and convex-concave with inflection point 0.5.

We define *the supporting line $L(x)$* (see Figure 2) to be the unique line
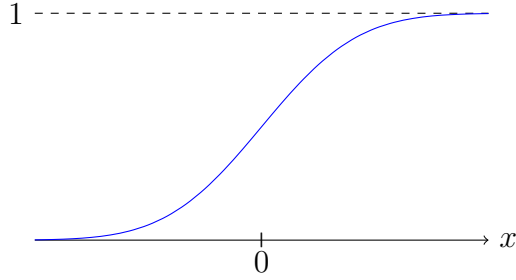
9

Figure 1: The CDF of the standard normal distribution.

that is the solution to the following maximization problem:

$$\max L(0)$$

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

In this paper we restrict attention to the case where $L(0) > 0$. (For a discussion of this assumption see Section 3.1). The *slope* of L is denoted by $1 - \alpha := L(1) - L(0) < 1$, and its *fixed point* is denote by $p = \frac{L(0)}{\alpha}$.

3   A General Result on Waiting Times

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

$$T_q(F, P(t)) = \mathbb{E}[\min\{t \; : \frac{1}{n} \sum_{i \in [n]} s_i(t) \ge q, \; s(0) = (0, \ldots, 0)\}].$$
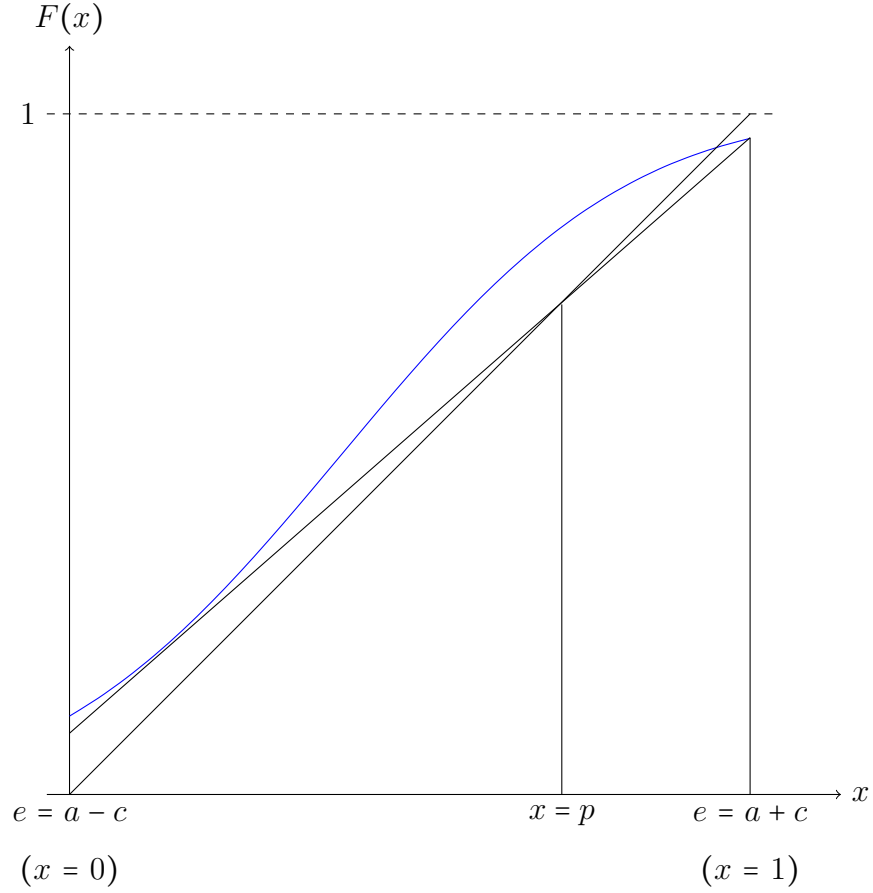
We can now state our main result:

10

Figure 2: Supporting line $L(x)$ and response function $r(x)$ based on the normal distribution.

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

$$T_q(F, P(t)) \leq \frac{7.2}{p\alpha}\left(1 + \ln\left(\frac{p}{p-q}\right)\right) = \frac{7.2}{L(0)}\left(1 + \ln\left(\frac{p}{p-q}\right)\right). \qquad (3)$$

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 [33, 2014]. Using martingale methods they derive the following upper bound on the expected waiting time until a majority have adopted and $P(t)$ is any regular, undirected network, namely[9]

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

The same bound holds for irregular networks provided that the target q is expressed as the degree-weighted proportion of adopting agents. This is quite different than 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. By contrast, Theorem 1 holds for any weighting of the agents and any (possibly asymmetric) amount of influence between the agents. Note that our bound depends logarithmically on $1/p - q$ whereas the martingale bound in [33, 2014] is proportional to $1/p - q$ which grows much more rapidly as q approaches p.

3.1   Discussion of the condition $L(0) > 0$

A key assumption underlying Theorem 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 too small (given c), or the variance of the distribution F is too small. The first case is shown in Figure 3. By shifting to the right by an amount b, the intercept becomes positive, as shown in the right panel. The second case is illustrated in Figure 4. 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 small enough, as shown in the right panel of Figure 4.

---

[9]For the case $q = \frac{1}{2}$ Kreindler and Young [33, 2014] show that the numerator in (4) can be replaced by 0.42.
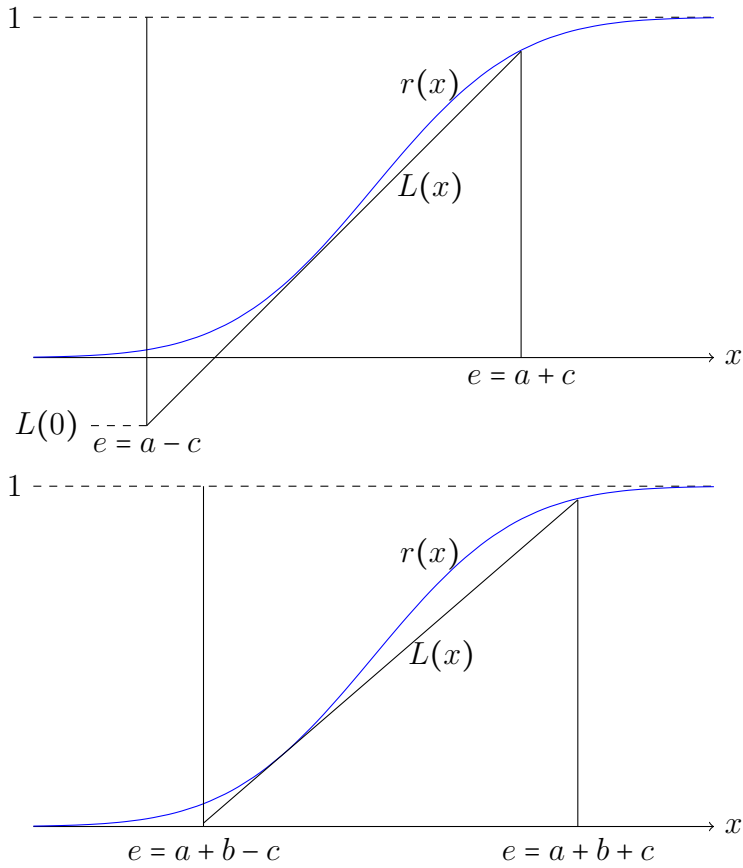
Figure 3: Increasing the payoff advantage of the innovation by an amount b increases the intercept $L(0)$.

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 choosing 1. The idea is to analyse 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 $L(x) \leq r(x)$, the expected time it takes the linear dynamic to reach the threshold q is at least as long as the expected time under the actual dynamic.
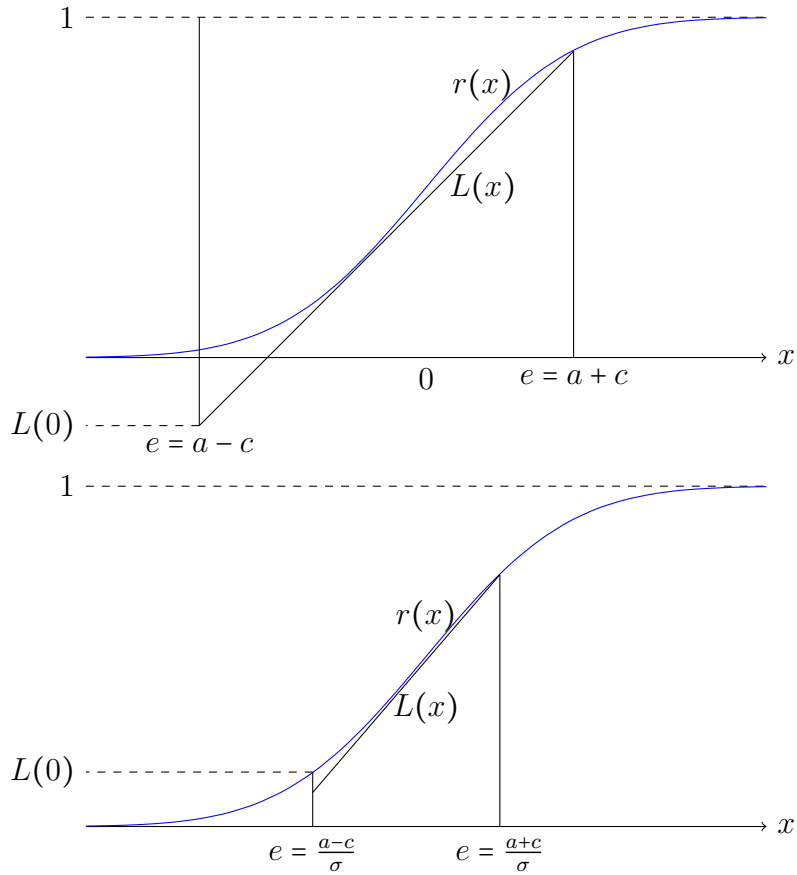
13

Figure 4: Increasing the standard deviation of the error distribution by a factor σ increases the intercept $L(0)$.

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

An advantage of analysing the linear dynamic is that it can be viewed as an *imitation dynamic*. Namely, at each time $t \in \mathbb{N}$ an agent i is chosen uniformly at random. With probability α agent i 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 $P_i(t)$.
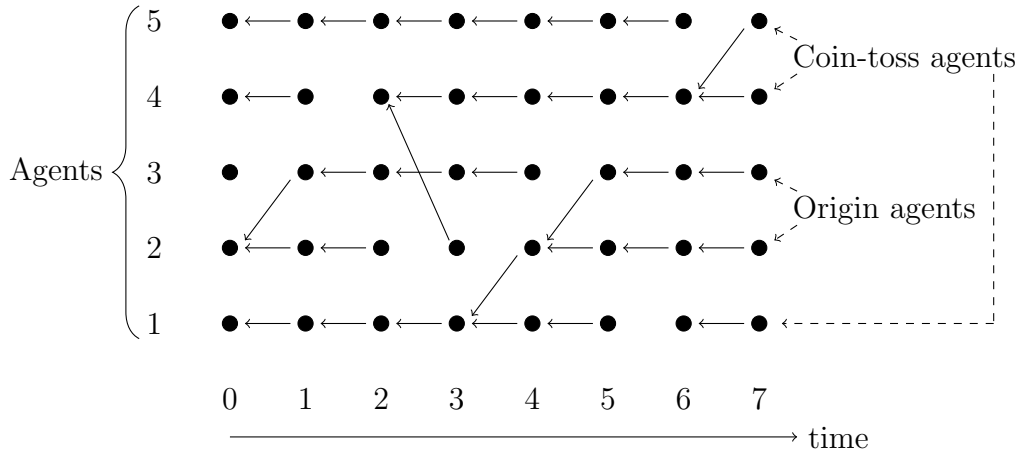

Figure 5: The imitation forest of the following history. At time $t = 1$, agent 3 imitates agent 2. At time $t = 2$, agent 4 tosses a coin. At time $t = 3$, agent 2 imitates agent 4. At time $t = 4$, agent 2 imitates agent 1. At time $t = 5$, agent 3 imitates agent 2. At time $t = 6$, agent 1 tosses a coin. At time $t = 7$, agent 5 imitates agent 4.

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 5.

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

15

root. Similarly, an *origin agent* it is one that belongs to a tree with a origin root.

This set-up bears some resemblance to the "voter model" (see for example Liggett [34]), but there is a crucial difference. In the voter model a randomly drawn agent imitates a randomly drawn neighbor. 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.

The proof consists of two key steps.

Step 1 First we show that the proportion of coin-toss players 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 players at each time t. We then apply Markov's inequality to show that *with high probability* the proportion of coin-toss players is large at all sufficiently large times t.

Step 2 At each point in time the proportion of players 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 [19] and its subsequent improvement by Garnett [22].

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

$$\mathbf{P}\left(\sum_{i=1}^{k} \beta_i c_i \geq \left[\sum_i \beta_i\right] p\right) \geq 0.14p.$$

This corollary allows us to place a nonzero lower bound on the probability that a given proportion are choosing 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 imitation dynamic. Since the Poisson clock has a rate of 1, the discrete process is m times slower then the actual continuous process. We shall prove that for the discrete imitation dynamic we have

$$T_q(F, P(t)) \leq \frac{7.2m}{p\alpha} \left(1 + \ln\left(\frac{p}{p-q} \right) \right).$$

First we analyze the diffusion of coin-toss agents. We define the process as follows. The initial state is $y(0) = (y_i(0))_{i \in [m]} = (0, ..., 0)$. In each period we choose player $i \in [m]$ at random. With probability α we set $y_i(t) = 1$. With probability $1 - \alpha$ we set $y_i(t) = y_j(t-1)$ where j is drawn according to $P_i(t)$.

Lemma 2. *The probability of an agent to be 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. It holds for $t + 1$ by the

following equation:

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

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

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

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

$\square$

Proof of Theorem 3.1. From Lemma 2 we get that the expected number of coin-toss agents at time t is given by $\mathbb{E}[\sum_i y_i(t)] = m - m(1 - \frac{\alpha}{m})^t$. We set a threshold of $\frac{q}{p}m$ for the diffusion of the coin-toss players. We use Markov inequality to bound the probability of small fraction (smaller than $\frac{q}{p}$) of coin-toss agents.

$$\mathbf{P}\left[\sum_i y_i(t) \le \frac{q}{p}m\right] = \mathbf{P}\left[\sum_i(1 - y_i(t)) \ge (1 - \frac{q}{p})m\right]$$
$$\le \frac{\mathbb{E}[\sum_i(1 - y_i(t))]}{(1 - \frac{q}{p})m} = \frac{(1 - \frac{\alpha}{m})^t}{1 - \frac{q}{p}} \tag{5}$$

At each point in time t, let $R(t) = \{r_1(t), ..., r_k(t)\} \subset [m] \times [t]$, $k = k(t)$, be the relevant coin-toss roots, and denote by $c_1(t), ..., c_k(t) \in \{0, 1\}$ their realized actions (that have been realized in the corresponding stages). Note that for any fixed t, $c_1(t), ..., c_k(t)$ are i.i.d. *Bernoulli*(p) random variables. For convenience of notations we also set $c_0 = 0$. Namely, $r_1, ..., r_k$ are the coin toss roots of $\{(i, t) : i \in [m]\}$ in the imitation forest (see Figure 5). Let $f = f(t) : [m] \to \{0, 1, ..., k\}$ be the function that associates each agent with its relevant 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 $f(i) = j$. If i is an origin agent at time t we

18

set $f(i) = 0$. Finally, for $j \in [k]$ we let $\beta_j(t) = |\{i : f(i) = j\}|$ be the number of appearances of the coin-toss j, with the corresponding action $c_j(t)$. Note that $\sum_{i \in [m]} y_i(t) = \sum_{j \in [k]} \beta_j(t)$.

The original imitation dynamic $s(t)$ can be written as $s_i(t) = c_{f(i)}$, and therefore $\sum_i s_i(t) = \sum_{j \in [k]} \beta_j(t) c_j(t)$. In order to bound T_q from above, we define a stopping time

$$\tau := \min\{t : \sum_i y_i(t) > \tfrac{q}{p}m\},$$

and an event

$$\mathcal{E} = \{\sum_j \beta_j(\tau) c_j(\tau) \geq \sum_j \beta_j(\tau) p\}.$$

Note that $\mathcal{E}$ implies that $\tau \geq T_q$. Since τ is measurable w.r.t. the imitation process and $\{c_1(\tau), \ldots, c_{k(\tau)}(\tau)\}$ are i.i.d. $Bernoulli(p)$ conditioned on the imitation process, and by Lemma 1, we have

$$\mathbf{P}(\mathcal{E}|\tau) \geq 0.14p. \tag{6}$$

We next show that

$$\mathbb{E}[\tau] \leq \frac{m}{\alpha}\left(1 + \ln\left(\frac{p}{p-q}\right)\right). \tag{7}$$

Indeed,

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

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

Where, the first inequality follows from (5), 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 events $\mathcal{E}_1, \mathcal{E}_2, \cdots$ in the

above fashion. Namely, τ_{i+1} is the first time at which there are more than $\frac{q}{p}m$ 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 $T_q \leq \tau_{i_*}$. 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 (6) and (7) 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,

$$\mathbb{E}[T_q] \leq \mathbb{E}[\tau_{i_*}] \leq \frac{1}{0.14p}\frac{m}{\alpha}\left(1 + \ln\left(\frac{p}{p-q}\right)\right) = \frac{7.2m}{p\alpha}\left(1 + \ln\left(\frac{p}{p-q}\right)\right).$$

$\square$

5   Multiple Technologies

So far we have analyzed the case where there are two competing technologies: the status quo and the innovation. In practice, however, 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 $A = \{1, \ldots, k\}$ consists of k alternatives, including the status quo. We shall assume that the payoff matrix U has the following form:

		1	2	$\cdots$	k
	1	$a_1 + c$	a_1	$\cdots$	a_1
$U =$	2	a_2	$a_2 + c$	$\cdots$	a_2
	$\vdots$				
	k	a_k	a_k	$\cdots$	$a_k + c$

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 e_j be the idiosyncratic payoff shock from playing alternative j. We assume that the shocks $(e_j)_{j=1,\ldots,k}$ are nonatomic, i.i.d., and have mean zero.

Given a distribution $\vec{x} = (x_1, \ldots, x_k) \in \Delta(A)$ let $r_i(\vec{x})$ be the probability that technology i is the best-reply alternative that is,

$$r_i(\vec{x}) = \mathrm{P}((U\vec{x})_i \geq (U\vec{x})_j \text{ for all } j \in A).$$

Let $\vec{r}(\vec{x}) = (r_1(\vec{x}), \ldots, r_k(\vec{x}))$ denote the *multidimensional response function.*

For every value $x \in [0, 1]$ of adopters of the superior technology let

$$\rho(x) = \min_{\vec{x} \in \Delta^k \text{ s.t. } x_1 = x} r_1(\vec{x}). \tag{8}$$

Thus we have replaced the function r of the two technologies case with the function ρ which bounds the probability of adopting technology 1 as a function of the proportion of adopters of 1. We can now approximate ρ using a linear function L as described above. We note that for every vector of proportions $\vec{x}$,

$$r_1(\vec{x}) \geq \rho(x_1) \geq L(x_1),$$

hence we can use Theorem 3.1 to approximate the waiting time for a proportion of $q < p$ adopters of the superior technology 1.

Deriving the function ρ explicitly from the multidimensional response function $\vec{r}$ is not always straightforward. In many applications, however, as we shall next show, ρ can be easily derived from $\vec{r}$.

Lemma 3. *If the payoff shock CDF F is log concave, then for every $x \in [0, 1]$,*

$$\rho(x) = \vec{r}(x, 1 - x, 0, \ldots, 0).$$

21

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 $1 - x$.

Before providing the proof note that

$$U\vec{x} = (a_1 + cx_1, a_2 + cx_2, \ldots, a_n + cx_n).$$

Proof. Consider the case where the realized payoff shock for technology 1 equals e_1. In this case the payoff from using technology 1 is $u_1 = a_1 + cx + e_1$. It is sufficient to show that across all $\vec{x} \in \Delta^k$ with $x_1 = x$ the vector that minimizes the probability

$$P\big((U\vec{x})_j + e_j \leq u_1 \text{ for } j = 2, \ldots, k, \big). \tag{9}$$

is $(x, 1 - x, 0, \ldots, 0)$. Note that the probability in (9) equals

$$F(u_1 - a_2 - cx_2)\cdots F(u_1 - a_k - cx_k). \tag{10}$$

We wish to find $\vec{x} \in \Delta^k$ that minimizes (10) subject to $x_1 = x$. It follows from the log concavity of F that the product in (10) is also log concave. Since the domain, $\{\vec{x} \in \Delta^k : x_1 = x\}$, 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 $(x, 1-x, 0, \ldots, 0)$. 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 since $a_j < a_2$ we have $u_1 - a_j > u_1 - a_2$. Since the function $h(z) = \ln(F(z))$ is concave, the slope

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

22

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

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

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

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

This concludes the proof of the lemma. $\qquad\square$

6 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 (or Lyapunov) function and hold only for regular networks with symmetric interactions, as in Kreindler and Young [33, 2014]. We have also established waiting time bounds for the diffusion of multiple innovations instead of just a single innovation, which is the usual assumption 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 (Maes and Nax [36, 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

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)).

We do not claim that the constant of proportionality in (3) is best possible, nor do we claim that the topology of the network does not affect the waiting time. Indeed it seems reasonable to suppose that some types of networks converge more quickly than others. The usefulness of the 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.

A   Feige's Inequality and Lemma 1

The following theorem is due to [19].

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

$$Pr[X < E[X] + 1] \ge \alpha > 0,$$

for some $\alpha \ge 1/13$.

Garnett [22] improved upon the constant α and show that $\alpha \ge 0.14$. We next prove Lemma 1 from the above theorem.

Proof of Lemma 1. Let $c_1, \ldots, c_k$ be i.i.d. Bernoulli(p), and let $\beta_1, \ldots, \beta_k$ be positive real numbers. We can assume without loss that $\beta_i \le \frac{1}{1-p} \ \forall i \in [k]$ and

that $\beta_1 = 1/(1-p)$. Let $X_i = \beta_i(1-c_i)$, $m_i = c_i(1-p)$, $X = X_2 + ... = X_k$, and $m = m_2 + ... + m_k$. We must show that $Pr(\sum_{i \in [k]} \beta_i c_i \geq \sum_{i \in [k]} \beta_i p) = P(X_1 + X \leq m + 1) \geq ap$. Indeed, by the above theorem,

$$P(X_1 + X \leq m + 1) \geq P(X_1 = 0)P(X < m + 1) \geq ap.$$

$\square$

References

[1] Acemoglu, D., Dahleh M.Z., Lobel, I, Ozdaglar, A. 2011. "Bayesian Learning in Social Networks," *Review of Economic Studies* 78: 1201-1236.

[2] Arieli, Itai, and H. Peyton Young. 2016. "Stochastic Learning Dynamics and Speed of Convergence in Population Games." *Econometrica* 84, no. 2: 627-676.

[3] Arthur W. Brian. 1994. "Increasing Returns and Path Dependence in the Economy" (Univ of Michigan Press, Ann Arbor, MI).

[4] V. Bala and S. Goyal. 1998. "Learning from Neighbours," The Review of Economic Studies, vol. 65, no. 3: 595-621.

[5] Banerjee, A. 1992. "A Simple Model of Herd Behavior", *The Quarterly Journal of Economics,* 107:797-817.

[6] Banerjee, A., Chandrasekhar A.G., Duflo E., and Jackson M.O. 2013. "The Diffusion of Microfinance." *Science* 341, Issue 6144.

[7] Bikhchandani, J. S., Hirschleifer, D. and I. Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades", *Journal of Political Economy,* 100: 992-1026

[8] Blume, L. E., 1993. "The statistical-mechanics of strategic interactions." *Games and Economic Behavior* 5, 387-424.

[9] Bongaarts, J. and S. C. Watkins. 1996. "Social interactions and contemporary fertility transitions," *Population and Development Review* 22(4): 639-682.

[10] Burke, Mary A, Fournier GM, Prasad K. 2007. "The Diffusion of a Medical Innovation: Is Success in the Stars?" *Southern Economic Journal* 73: 588-603.

[11] Burke, Mary 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.

[12] Centola, D. and A. Baronchelli. 2015. "The Spontaneous Emergence of Conventions: An Experimental Study," *PNAS* 112(7): 1989-1994.

[13] 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.

[14] Coleman, James, Elihu Katz, and Herbert Menzel. 1957. "The Diffusion of an Innovation Among Physicians." *Sociometry* 20, no. 4: 253-270.

[15] David PA. 1985. "Clio and the Economics of QWERTY." *American Economic Review Papers and Proceedings,* 75:332-337.

[16] Ellison, Glenn. 1993. "Learning, Social Interaction, and Coordination." *Econometrica* 61: 1047-1071.

[17] Ellison, Glenn. and Fdenberg D. 1993. "Rules of Thumb for Social Learning." *Journal of Political Economy,* 101:612-643.

[18] Ellison, Glenn., D. Fdenberg, and L. Imhof 2016. "Fast Convergence in Evolutionary Models: A Lyapunov Approach." *Journal of Economic Theorey,* 161: 1-36.

[19] Uri Feige. 2006. 'u'On sums of independent random variables with unbounded variance, and estimating the average degree in a graph," *Siam Journal on Computing* vol. 35, no. 4, pp. 964-984.

[20] Foster A. and rosenzweig M. 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy,* 103: 1176-1209.

[21] Gale D. and Kariv S. 2003. "Bayesian Learning in Social Networks." *Games and Economic Behavior* 45: 329-346.

[22] Garnett B., 2018. "Small Deviations of Sums of Independent Random Variables" working paper.

[23] Golub B. and M. O. Jackson. 2010. "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics,* 2: 112-149.

[24] Golub B., M. O. Jackson. 2012. "How Homophily Affects the Speed of Learning and Best-Response Dynamics." *Quarterly Journal of Economics* 1287-1338.

[25] Goyal S, Vega-Redondo F. 2005. "Learning, Network Formation, and Coordination." *Games and Economic Behavior* 50:178-207.

[26] Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* 25, no. 4: 501-522.

[27] Jackson MO, Watts A. 2002. "The evolution of social and economic networks." *Journal of Economic Theory* 106:265-295.

[28] Jackson MO, Yariv L. 2007. "Diffusion of Behavior and Equilibrium Properties in Network Games. *American Economic Review Papers and Proceedings,* 97:92-98.

[29] Jackson MO. 2008. Social and Economic Networks. Princeton NJ: *Princeton University Press.*

[30] Kohler H-P. 1997. "Learning in Social Networks and Contraceptive Choice." *Demography* 34:369-383.

[31] Kohler H-P, Behrman JR, Watkins SC. 2001. "The Density of Social Networks and Fertility Decisions: Evidence from South Nyanza District, Kenya." *Demography* 38:43-58.

[32] Kreindler, Gabriel E., and H. Peyton Young. 2013. "Fast Convergence in Evolutionary Equilibrium Selection," *Games and Economic Behavior* 80, 39-67.

[33] Kreindler, Gabriel E., and H. Peyton Young. 2014. "Rapid Innovation Diffusion in Social Networks." *Proceedings of the National Academy of Sciences* 111, no. 3: 10881-10888.

[34] Liggett, T.M. 1999. "Stochastic Interacting Systems: Contact, Voter, and Exclusion Process." *Berlin: Springer-Verlag.*

[35] Mackie, Gerry. 1996. "Ending footbinding and infibulation: A convention account." *American Sociological Review* 61 (6): 999-1017.

[36] Maes M. and Nax HH. 2016. "A Behavioral Study of Noise in Coordination Games." *Journal of Economic Theory* 162: 195-208.

[37] Montgomery MR, Casterline JB. 1996. "Social Influence, Social Learning, and New Models of Fertility." *Population Development Review* 22: 151-175.

[38] Morris S. 2000. "Contagion." *Review of Economic Studies* 67:57-78.

[39] Montanari A, Saberi A. 2010. "The Spread of Innovations in Social Networks." *Proceedings of the National Academy of Sciences* 107:20196-20201.

[40] Mossel, Elchanan, Allan Sly, and Omer Tamuz. 2015. "Strategic learning and the topology of social networks." *Econometrica* 83, no. 5 : 1755-1794.

[41] Mueller-Frank M, 2013. "A General Framework for Rational Learning in Social Networks." *Theoretical Economics* 8(1): 1 -40.

[42] Mueller-Frank M, Pai M. 2016. "Social Learning with Costly Search." *American Economic Journal: Microeconomics* 83-109.

[43] Munshi K. 2004. "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution." *Journal of Development Economics* 73: 185-215.

[44] K. Munshi and J. Myaux. 2006. "Social Norms and the Fertility Transition," *Journal of Development Economics* vol. 80, pp. 1-38.

[45] Phelps CE, Mooney C. 1993. "Variations in Medical Practice Use: Causes and consequences." In Competitive Approaches to Health Care Reform, ed. RJ Arnould, RF Rich, W. White, pp. 139-178. Washington DC: Urban Institute.

[46] Rogers, Everett. 2003. "Diffusion of Innovations." 5th edition. New York: Simon and Schuster.

[47] Ryan, Bryce, and Neal C. Gross. 1943. "The Diffusion of Hybrid Seed Corn in Two Iowa Communities." *Rural Sociology* 8, no. 1 : 15-24.

[48] Skyrms B, Pemantle R. 2000. "A Dynamic Model of Social Network Formation." *Proc NatlAcad Sci* USA 97:9340-9346.

[49] Solan, Eilon, Dinah Rosenberg, and Nicolas Vieille. 2009. "Informational Externalities and Emergence of Consensus." *Games and Economic Behavior* 66, no. 2 : 979-994.

[50] Udry C. and Conley T. 2001. "Social Learning through Networks: The Adoption of New Agricultural Technologies in Ghana." *American Journal of Agricultural Economics* 83: 668-673.

[51] Valente, Thomas. 1995. "Network Models of the Diffusion of Innovations," Cresskill NJ: Hampton Press.

[52] Valente, Thomas. 2005. "Network Models and Methods for Studying the Diffusion of Innovations," in Models and Methods in Social Network Analysis, Cambridge, Cambridge University Press.

[53] Vega-Redondo, Fernando. 2007. "Complex Social Networks," Cambridge: Cambridge University Press.

[54] Watts, D. 2002. "A Simple Model of Global Cascades on Random Networks," *Proceedings of the National Academy of Sciences,* vol. 99, no. 9, p. 5766-5771.

[55] Wennberg, J., and A. Gittelsohn. 1973. "Small Area Variations in Health Care Delivery," *Science* 182: 1102-1108.

[56] Young, H. Peyton. 1998. "Individual strategy and social structure: An evolutionary theory of institutions." Princeton NJ: Princeton University Press.

[57] Young, H. Peyton. 2009. "Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning." *American Economic Review* 99, no. 5 (2009): 1899-1924.

[58] Young, H. Peyton. 2011. "The Dynamics of Social Innovation." Proceedings of the National Academy of Sciences 108, no. 4 : 21285-21291.