Rapid Innovation Diffusion in Social Networks[*]

Gabriel E. Kreindler and H. Peyton Young[1]

November 21, 2013

Abstract. The diffusion of an innovation can be represented by a process in which agents choose perturbed best responses to what their neighbors are currently doing. Diffusion is said to be *fast* if the expected waiting time until the innovation spreads widely is bounded above independently of the size of the network. Previous work has identified specific topological properties of networks that guarantee fast diffusion. Here we apply martingale theory to derive *topology-free* bounds such that diffusion is fast whenever the payoff gain from the innovation is sufficiently high and the response function is sufficiently noisy. We also provide a simple method for computing an upper bound on the expected waiting time that holds for all networks. For example, under the logit response function it takes on average less than 80 revisions per capita for the innovation to diffuse widely in any network, provided that the error rate is at least 5% and the payoff gain (relative to the status quo) is at least 150%. Qualitatively similar results hold for other smoothed best response functions and populations that experience heterogeneous payoff shocks.

1 Local interaction topology and fast diffusion

New ideas and ways of doing things often spread through social networks. Agents adopt an innovation with increasing likelihood depending on the proportion of their friends and neighbors who have adopted it. The innovation in question might be a technological advance such as a new piece of software, a medical drug (Coleman, Katz and Menzel 1957), or a new hybrid seed (Griliches 1957). Or it might represent a social practice, such as contraception (Munshi and Myaux 2006), a novel form of employment contract (Young and Burke 2001), or a group behavior such as binge drinking (Kremer and Levy 2008).

In recent years such diffusion models have been extensively studied from both a theoretical and empirical standpoint. Some authors have highlighted the importance of the payoff gains from the innovation: larger gains lead to faster adoption (Griliches 1957, Bala and Goyal 1998). Others have pointed to the role that the network topology plays in the rate at which an innovation spreads (Blume, 1993, 1995; Bala and Goyal, 1998; Valente, 1995; 2005; Morris, 2000; Watts, 2002; Watts and Dodds, 2007; Jackson and Yariv 2007, Jackson and Rogers, 2007, Vega-Redondo, 2007; Golub and Jackson, 2010; Jackson, 2010, Montanari and Saberi 2010, Newton and Angus 2013).

A key topological feature of this nature is local clustering. Consider a group of individuals who have most of their interactions with each other as opposed to outsiders. Such a group is said to be cohesive (Morris, 2000) or close-knit (Young, 1998). Suppose also that the adoption process is stochastic: a given individual adopts with a probability that is increasing in the proportion of his neighbors who have adopted. This holds if the technology has positive network externalities and individuals choose best responses subject to a small utility shock. In this situation the topology of the network structure can profoundly influence the speed with which such a technology will spread. In a small close-knit group the innovation can gain a foothold relatively quickly, and maintain its hold even when those outside the group have not yet adopted. If everyone is contained in a sufficiently close-knit group of bounded size, the technology can spread quite rapidly through the whole population as it gains independent footholds among many groups scattered across the network. In such a case the expected waiting time until the innovation becomes widely adopted can be bounded independently of the network size. More precisely it can be shown that for a given small level of noise in the individual response functions, the

2

expected waiting time is bounded for all populations provided that the payoff gain from the innovation (relative to the status quo) is sufficiently large (Ellison 1993; Young 1998, 2011). In a similar vein, Montanari and Saberi (2010) employ a topological notion called "tilted cutwidth" to identify an even broader class of networks in which the convergence time is bounded.

We will show in this article that one can obtain qualitatively similar results without placing *any* topological restrictions on the network. The methods we use to establish this result are quite different from those in the prior networks literature, many of which rely on the parallel decomposition argument outlined above. Here we use a different approach that takes its inspiration from another branch of the learning literature, namely the analysis of noisy best response dynamics when agents draw samples uniformly at random from the *whole population*. When the population is large, such a process can be well-approximated by a deterministic (mean-field) dynamical system (McKelvey and Palfrey 1995, Sandholm, 2001, Blume and Durlauf 2003, Lopez-Pintado 2006, Jackson and Yariv 2007, Oyama, Sandholm and Tercieux 2012, Kreindler and Young 2013).

This set-up is quite different from network models, in which agents choose noisy best responses to a *fixed* set of neighbors. To see the distinction, consider the situation where agents can coordinate on one of two actions: A (the innovation) or B (the status quo). Assume for the moment that they choose pure (instead of noisy) best responses to the actions of their neighbors. Such a game will typically have a vast number of heterogeneous equilibria whose structure depends on the fine details of the network topology. By contrast, in a global interaction model with sampling there will typically be just three equilibria, two of which are homogeneous and stable and the other heterogeneous and unstable.

Now consider a (possibly noisy) best response process and an unboundedly large number of agents who interact globally. The expected motion can be represented by a deterministic, continuous-time dynamic of the form $\dot{x}_t = f(x_t) - x_t$ where x_t is the proportion of agents playing action A at time t, and $\dot{x}_t$ is the rate of change of this variable. Under fairly general assumptions $f(x)$ has a convex-concave shape as shown in Figure 1. Depending on the choice of parameters, such a curve will typically have either one or three rest points. In the latter case the middle one (the up-crossing) is unstable and the other two are stable. One can employ a mean-

field approach to study the effects of varying the payoffs and the distribution of sample sizes on the location of these rest points. The lower is the unstable rest point, the smaller the basin on attraction of the status quo equilibrium; hence the easier it is to escape that equilibrium (Jackson and Yariv 2007, López-Pintado 2006).

Of particular interest is the case where there is a single rest point and it lies above the halfway mark $(.5, .5)$. (See the right panel in Figure 1).[2] In this case the process moves from the initial state (all-*B*) to a majority playing *A* in a bounded length of time. In an earlier paper we examined the implications of this observation for the logit best response function (Kreindler and Young, 2013). The main result in that paper is that, if the payoff gain from the innovation and/or the noise level is sufficiently large, then there is a unique rest point lying above $(.5, .5)$ and the expected waiting time until a majority of the agents choose *A* is bounded above independently of the population size; moreover we give an explicit estimate of the expected waiting time. However, this argument depends crucially on the assumption that the agents interact globally, which ensures that the expected dynamics have the simple form shown in Figure 1. This allows one to employ standard stochastic approximation techniques to estimate the waiting time as a function of the payoffs, noise level, and sample size. In a network setting, by contrast, the underlying state space is vastly more complex and generally the unperturbed process will possess a large number of rest points, each with a different stochastic potential. Thus the approximation techniques from our earlier paper do not apply to this case.

The contribution of the present paper is to show that, in spite of the complications introduced by the network setting, we can derive an upper bound on the rate of diffusion that holds uniformly for all networks provided that payoff gain from the innovation is sufficiently high and the noise level is not too low. These results are established using martingale arguments rather than the stochastic approximation methods that are common in the prior literature. The bound we establish is easy to compute given the payoffs and the shape of the stochastic response function. Thus our approach provides a practical estimation method that can in principle be applied to empirical data.

---

[2] McKelvey and Palfrey (1995) show that in almost any normal form game with logit response function, for sufficiently large noise levels there is a unique rest point (a quantile equilibrium). Sandholm (2001) and Oyama et al (2011) examine the case when agents best-respond to random samples; they show that if there exists a $\frac{1}{k}$-dominant equilibrium (where k is a suitably chosen sample size) then the dynamics have a unique *stable* rest point.
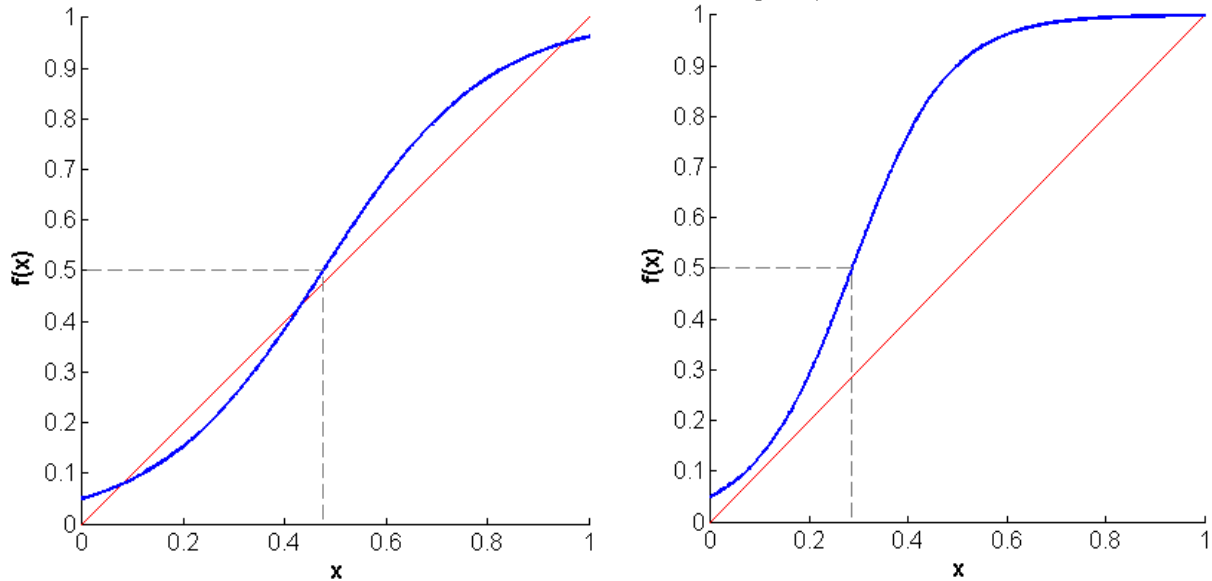
Figure 1 – The logit response function

The remainder of the paper is organized as follows. Section 2 introduces the adoption model with local interactions. In section 3 we establish the main results for regular networks, and in section 4 we show how to extend the arguments to general networks. In section 5 we show that the results are robust for a large family of stochastic best response functions other than the logit. We also show that the results can be interpreted in terms of payoff heterogeneity instead of noisy best responses.

2   The local interaction model

The model is expressed in terms of a stochastic process denoted $\Gamma(\alpha, \beta, G)$ that depends on three parameters: the potential gain α, the noise level β, and the interaction graph G. Consider a population of N agents numbered from 1 to N linked by an undirected graph G. Each agent chooses one of two available actions, A and B. Interaction is given by a symmetric 2×2 coordination game with payoff matrix

	A	B
A	a, a	c, d
B	d, c	b, b

(1)

where $a > d$ and $b > c$. This game has potential function

	A	B
A	$a - d$	0
B	0	$b - c$

Define the *normalized potential gain* of passing from equilibrium (B, B) to (A, A) as

$$\alpha = \frac{(a - d) - (b - c)}{b - c}.$$

Assume without loss of generality that the potential function achieves its strict maximum at the equilibrium (A, A), or equivalently $\alpha > 0$. Hence (A, A) is the risk-dominant equilibrium of the game; note that (A, A) may or may not also be the Pareto-dominant equilibrium. Standard results in evolutionary game theory imply that (A, A) will be selected in the long run (Kandori, Mailath and Rob 1993, Young 1993, Blume 2003).

An important special case is when interaction is given by a pure coordination game with payoff matrix

	A	B
A	$1 + \alpha, 1 + \alpha$	0,0
B	0,0	1,1

(2)

In this case we can think of B as the "status quo" and of A as the "innovation." The term $\alpha > 0$ is now the *payoff gain* of the adopting the innovation relative to the status quo. Note that the potential function of the pure coordination game is proportional to the potential function in the general case. It follows that for the logit model and under a suitable rescaling of the noise level, the two settings are in fact equivalent.

Payoffs are as follows. At the start of each time period every pair of agents linked by an edge interact once and they receive the one-shot payoffs from the game defined in (2). Thus each agent's payoff in a given period is the sum of the payoffs from his pairwise interactions in that period. Note that players with a high number of connections will, ceteris paribus, have higher payoffs per period. Formally, let $\pi(x_i, x_j)$ be i's payoff from interacting with j, when i plays x_i and j plays x_j. Letting N_i denote the set of i's neighbors, the total payoff for i from playing x_i is $\sum_{j \in N_i} \pi(x_i, x_j)$.

We posit the following revision process. At times $t = k/N$ with $k \in \mathbb{N}$, and only at these times, one agent is randomly (independently over time) chosen to revise his action.[3] When revising, an agent observes the actions currently used by his neighbors in the graph G. Assume that a *fraction* x of agent i's neighbors are playing A, then i chooses a noisy best response given by the logit model:

(3) $$\Pr(i \text{ chooses A} \mid x) = f(x; \alpha, \beta) = \frac{e^{\beta(1+\alpha)x}}{e^{\beta(1+\alpha)x} + e^{\beta(1-x)}},$$

where $1/\beta$ is a measure of the noise in the revision process. For convenience we will sometimes drop the dependence of f on β in the notation and simply write $f(x; \alpha)$, and in some cases we shall drop both α and β and write $f(x)$. Let $\varepsilon = 1/(1 + e^{\beta})$ denote the error rate at the start of the process when nobody has yet adopted the innovation A. This will be called the *initial error rate*. Both $1/\beta$ and ε measure the noise level of the process. Since ε is easier to interpret as the rate at which agents depart from best response at the outset, we shall sometimes express our results in terms of both β and ε.

Embedded in the above formulation is the assumption that the error rate depends only on the proportion of an agent's neighbors playing A or B, and it does not depend on the number of neighbors, that is, on the agent's degree. This follows naturally if we assume that in each period each agent experiences an idiosyncratic shock that affects his realized payoffs from *all* interactions during that period. In other words, an agent expects to obtain direct payoffs ϵ_A and

---

[3] The results in this article also hold under the following alternative revision protocol. Time is continuous, each agent has a Poisson clock that rings once per period in expectation, and an agent revises his action when his clock rings.

ϵ_B from playing A and B, respectively, in addition to the coordination game payoffs, from each of the agent's interactions. The logit model (3) results when the payoff shocks ϵ_A and ϵ_B are i.i.d. extreme value distributed (McFadden, 1976). This is one of the most commonly used models in the literature on stochastic learning and also in the literature on discrete choice (McFadden 1976, Anderson, Palma and Thisse 1992, Blume 1993, Blume 1995, Brock and Durlauf 2001). The key feature of logit for our results is that the probability of choosing one action decreases smoothly as a function of the payoff difference between the two choices. In section 5 we show that the main result in this paper remains true for a large family of smooth stochastic response functions that are qualitatively similar to the logit. The stochastic response functions we consider emerge in a setting where agents myopically best respond to the actions of their neighbors, and payoffs from playing each action are subject to random shocks.[4]

The above revision process fully describes the stochastic process $\Gamma(\alpha, \beta, G)$. The states of the process are the adoption vectors $x(t) = \left(x_i(t)\right)_{1 \leq i \leq N}$, where $x_i(t) = 1$ if agent i plays A at time t, and $x_i(t) = 0$ otherwise. The *adoption rate* is defined as $\bar{x}(t) = (1/N)\sum_{i=1}^{N} x_i(t)$. By assumption, the process starts in the *all-B* state, namely $x(0) = (0, ... ,0)$.

We now turn to the issue of speed of diffusion, measured in terms of the expected time until a majority of the population adopts action A. This measure is appropriate because it captures the time it takes for the system to escape the status quo equilibrium (B, B). Formally, define the random hitting time[5]

$$T(\alpha, \beta, G) = \min\left\{t : \bar{x}(t) \geq \frac{1}{2}\right\}.$$

More generally, given $p < 1$ one can consider the waiting time $T(\alpha, \beta, G, p)$ until $\bar{x}(t)$ is at least p for the first time. The method of analysis in this paper can be extended in a straightforward way to treat this case.

Fast diffusion is defined as follows.

---

[4] The response functions can also be viewed as a primitive element of the model that can be estimated directly from empirical data.

Definition 1. Given a family $\mathcal{G}$ of graphs, we say that the *family of processes* $\{\Gamma(\alpha, \beta, G) : G \in \mathcal{G}\}$ *displays fast diffusion* if there exists a constant $S = S(\alpha, \beta, \mathcal{G})$ such that for any $G \in \mathcal{G}$ the expected waiting time until a majority of agents play A under process $\Gamma(\alpha, \beta, G)$ is at most S independently of G, that is $ET(\alpha, \beta, G) < S$.

When the above conditions are satisfied we say that $\Gamma(\alpha, \beta, \mathcal{G})$ *displays fast diffusion*.

3 Topology-free fast diffusion in regular networks

In this section we establish our first result on sufficient conditions for fast diffusion in d-regular networks, and we provide an upper bound on the expected waiting time for a majority of the population to adopt. In the next section we shall show how to extend these arguments to general networks.

3.1 Fast diffusion in regular networks

In order to find sufficient conditions for fast diffusion, we begin by lower bounding the *expected* change in the adoption rate in any state where adopters constitute a weak minority.

A graph G is *d-regular* if every node in G has exactly d neighbors. Denote by $\mathcal{G}(d)$ the family of all d-regular graphs. Fix a payoff gain α, a noise level β and a graph $G \in \mathcal{G}(d)$, for some $d \geq 3$. Consider a state $x(t)$ of the process $\Gamma(\alpha, \beta, G)$ such that $\bar{x}(t) \leq 1/2$. For any $k \in \{0, 1, \dots, d\}$, denote by $q_k(t)$ the fraction of players who have exactly k neighbors currently playing A.

The expected next period adoption of an agent i who has k neighbors currently playing A is

$$Ex_i\left(t + \frac{1}{N}\right) = \frac{1}{N}f\left(\frac{k}{d}\right) + \frac{N-1}{N}x_i(t)$$

$$\Rightarrow Ex_i\left(t + \frac{1}{N}\right) - x_i(t) = \frac{1}{N}\left(f\left(\frac{k}{d}\right) - x_i(t)\right)$$

Note that each agent is chosen to revise with equal probability, hence the expected change in the population adoption rate is

$$(4) \qquad E\bar{x}\left(t + \frac{1}{N}\right) - \bar{x}(t) = \frac{1}{N}\left[\sum_{k=0}^{d} q_k(t) f\left(\frac{k}{d}\right) - \bar{x}(t)\right].$$

We wish to bound expression (4) from below, since this provides a lower bound on the expected change in the adoption rate for all configurations that have a minority of adopters. Let us begin by observing that the weights $q_k(t)$ "convexify" the set of points $F_d \equiv \left\{\left(\frac{k}{d}, f\left(\frac{k}{d}\right)\right) : 0 \le k \le d\right\}$. Denote by $f_d(x)$ the lower envelope of the convex hull of F_d. Figure 2 illustrates the construction.
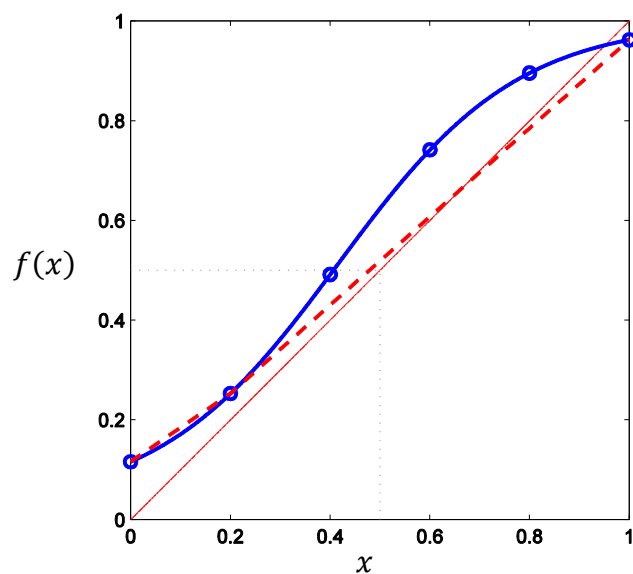

Figure 2 – The function $\boldsymbol{f_d}$ (red dashed line) is the lower envelope of the convex hull of the set $\boldsymbol{F_d}$ (blue circles).

For each player i, consider the fraction of i's neighbors who have adopted the innovation. The adoption rate can be expressed as the population average of this quantity. Let N_i denote the set of neighbors of i. We can then write

$$\bar{x}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{d} \sum_{j \in N_i} x_j(t)$$

$$= \sum_{k=0}^{d} \frac{k}{d} q_k(t).$$

Using this identity and Jensen's inequality for the convex function f_d we obtain

$$\sum_{k=0}^{d} q_k(t) f(k/d) \geq f_d(\bar{x}(t)).$$

Using identity (4) we obtain

(5)
$$E\bar{x}\left(t + \frac{1}{N}\right) - \bar{x}(t) \geq \frac{1}{N} [f_d(\bar{x}(t)) - \bar{x}(t)].$$

Given α, β and $d \geq 3$ define the quantities

$$\mu(\alpha, \beta, d) \equiv f_d(1/2) - 1/2$$

and

$$h(\beta, d) = \min\{\alpha : \mu(\alpha, \beta, d) \geq 0\}.$$

If the function f_d lies strictly above the identity function (the 45-degree line) on the interval $[0,1/2]$ or, equivalently, if $\mu(\alpha, \beta, d) > 0$, then the expected change in the adoption rate is positive for any state $x(t)$ with a weak minority of adopters. This implies diffusion for the family of d-regular graphs, as the following result shows. (The entire argument can be adapted to estimate the waiting time until a proportion $p < 1$ of players have adopted; here we have stated the results in terms of the target $p = 1/2$ for simplicity.)

Theorem 1. *There exist uniform lower bounds on the payoff gain and the noise level such that the expected waiting time until a majority of agents play A is bounded for all regular graphs of degree at least three, irrespective of the number of agents. Concretely, $\Gamma(\alpha, \beta, \mathcal{G}(d))$ displays fast diffusion whenever $\alpha > 0.83$ and $\varepsilon \geq 5\%$. Furthermore, given any $d \geq 3$, $\Gamma(\alpha, \beta, G(d))$ displays fast diffusion whenever $\alpha > h(\beta, d)$.*

Prior work in the literature on fast diffusion in evolutionary models has focused mainly on the topological properties of the graphs in the family $\mathcal{G}$ that guarantee fast diffusion (Ellison 1993, Young 1998; Montanari and Saberi, 2010). In contrast, Theorem 1 establishes *topology-free* bounds that guarantee fast diffusion for the entire family of d-regular graphs.

Proof. The proof proceeds in two steps. First, we show that the *expected* change in the adoption rate is strictly positive whenever the adoption rate is at most $1/2$. Secondly, we show that the expected waiting time until a majority adopts is bounded for all graphs in $\mathcal{G}(d)$.

Fix some $\alpha > h(\beta, d)$. By construction, $\mu(\alpha, \beta, d) = f_d(1/2) - 1/2 > 0$. In addition, the following lemma provides a positive lower bound for $f_d(x) - x$ for all $x \leq 1/2$. It follows from (5) that the expected change in the adoption rate is strictly positive for any state such that $\bar{x}(t) \leq 1/2$.

Lemma 1. *Let* $\mu_d = \mu(\alpha, \beta, d)$, *then* $f_d(x) - x \geq 2\mu_d(1 - x)$ *for all* $x \in [0, 1/2]$.

Proof. Because f is first convex and then concave, there exists $k^* < d/2$ such that for any $k \leq k^*$ we have $f_d(k/d) = f(k/d)$, and for any $x > k^*/d$ we have

$$f_d(x) = f(k^*/d)\frac{1-x}{1-k^*/d} + f(1)\frac{x - k^*/d}{1 - k^*/d}.$$

The right hand side is the equation of the line $L(x)$ passing through the points $(k^*/d, f(k^*/d))$ and $(1, f(1))$. In particular, we have that $f_d(x) \geq L(x)$ for all $x \in [0, 1]$.

We claim that $L(x) - x \geq 2\mu_d(1 - x)$ for all $x \in [0, 1/2]$. Note that by definition this holds with equality for $x = 1/2$. Moreover, $L(1) - 1 = f(1) - 1 < 0 = 2\mu_d(1 - 1)$, so the opposite inequality must hold for $x < 1/2$. This completes the proof of Lemma 1. ∎

The following claim provides explicit conditions that ensure that $\alpha > h(\beta, d)$.

Claim 1. If $\varepsilon = 1/(1 + e^\beta) > 5\%$ then $h(\beta, d) < 0.83$. It follows that when $\alpha > 0.83$ the expected change in the adoption rate is positive in any state with $\bar{x}(t) \leq 1/2$.

The proof of Lemma 1 implies that $\mu_d > 0$ provided that, for every $k \leq d/2$ we have

12

$$f(k/d) \frac{1/2}{1-k/d} + f(1) \frac{1/2 - k/d}{1-k/d} > \frac{1}{2}.$$

This inequality can be rewritten as

(6)
$$f\left(\frac{k}{d}\right) > 1 - f(1) + (2f(1) - 1)\frac{k}{d}.$$

To establish Claim 1, it suffices to show that inequality (6) holds whenever $\varepsilon = 1/(1 + e^\beta) > 5\%$ and $\alpha = 0.83$. We prove that more generally for all $x \in \left[0, \frac{1}{2}\right]$,

(7)
$$f(x) - (1 - f(1)) - (2f(1) - 1)x > 0.$$

Denote by x^* a point where the left hand side of (6) achieves its minimum on $\left[0, \frac{1}{2}\right]$. The first-order condition is $f'(x^*) = 2f(1) - 1$. Using the formula $f'(x) = \beta(\alpha + 2)f(x)(1 - f(x))$ we obtain the quadratic equation

$$f(x^*) - f(x^*)^2 = \frac{2f(1) - 1}{\beta(\alpha + 2)}.$$

Let $\gamma = (2f(1) - 1)/(\beta(\alpha + 2))$. When $\gamma > \frac{1}{4}$ there is no interior solution and $x^* = \frac{1}{2}$. When $\gamma \leq \frac{1}{4}$ we have

$$f(x^*) = \frac{1 - \sqrt{1 - 4\gamma}}{2}.$$

We need to check that inequality (7) holds at x^*. (Note that all terms depend on α and β, although we have suppressed them from the notation.)[6] It is straightforward to check numerically that the inequality holds for $\alpha = 0.83$ and $\beta = \log(19)$ (the solution to $\varepsilon = 5\%$). Figure 3 plots the function $u(\beta) = f(x^*) - (1 - f(1)) + (2f(1) - 1)x^*$; the function is positive for all $\beta \in [0, \log(19)]$.

This completes the proof of Claim 1 and the first step in the proof of Theorem 1.

---

[6] For x^* we use the explicit formula $f^{-1}(y) = \frac{1}{\alpha+2}\left(1 - \frac{1}{\beta}\log\left(\frac{1-y}{y}\right)\right)$.
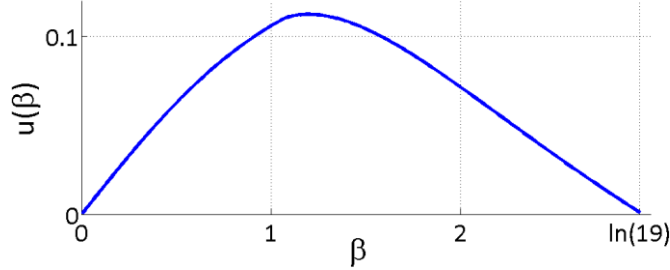
Figure 3 - The function $u(\beta)$

We now turn to the second step in the proof, namely that the expected waiting time to reach a majority of adopters is bounded for the family $\mathcal{G}(d)$ of d-regular graphs whenever $\alpha > h(\beta, d)$.

We already know that for any β and for any $\alpha \geq h(\beta, d)$, for any graph $G \in \mathcal{G}(d)$ and any state $x(t)$ of the process $\Gamma(\alpha, \beta, G)$ such that the adoption rate is at most one half, the expected change in the adoption rate is positive. By Lemma 1 we know that

(8)
$$E\left(\bar{x}\left(t + \frac{1}{N}\right)\right) - \bar{x}(t) \geq \frac{1}{N}\mu_d(1 - \bar{x}(t)) > 0.$$

We shall now establish a *uniform upper bound* on the expected waiting time for all d-regular graphs (of any size), expressed in terms of $\mu(\alpha, \beta, d)$.

Lemma 2. *Assume that* $\mu_d = f_d(1/2) - 1/2$ *is positive. Then for any* $N \geq 3$ *the expected waiting time until a majority of the population adopts the innovation satisfies*

$$ET(\alpha, \beta, G) < \frac{0.42}{\mu_d}.$$

Proof. Partition the states of the process $\Gamma(\alpha, \beta, G)$ according to their adoption rate. Specifically, let $X_m = \{x \in \{0,1\}^N : \bar{x} = m/N\}$. In any given period, the change in the adoption rate can only be $-1/N$, 0 or $1/N$. Given any state x, let $p(x)$ denote the probability that the adoption rate increases by $1/N$, and let $q(x)$ denote the probability that the adoption rate decreases by $1/N$. Thus for every $x_i \in X_m$ we have $p(x_i) = \Pr(x(t + 1/N) \in X_{m+1} \mid x(t) = x_i)$ and $q(x_i) = \Pr(x(t + 1/N) \in X_{m-1} \mid x(t) = x_i)$. By inequality (8), for any state x with $\bar{x} \leq 1/2$,

14

$$p(x) - q(x) \geq 2(1 - \bar{x})\mu_d.$$

For convenience, let $\mu_d^m = 2\mu_d(1 - m/N)$.

Given any $x_i \in X_m$ let $W(x_i)$ denote the expected waiting time until the process first reaches a state in X_{m+1} when starting from x_i. Let $W_m = \max\limits_{x_i \in X_m} W(x_i)$ denote the longest expected waiting time to reach X_{m+1} from a state in X_m. The waiting time to reach a majority of adopters satisfies

(9)
$$ET(\alpha, \beta, G) \leq W_0 + W_1 + \cdots + W_{\lfloor N/2 \rfloor}.$$

We now show that W_m is at most $1/(N\mu_d^m)$ for all $m < N/2$. We proceed by induction in m. By definition $q(\mathbf{0}) = 0$ where $\mathbf{0} = (0, \ldots, 0)$, hence $p(\mathbf{0}) \geq \mu_d^0$ and $W_0 \leq 1/(N\mu_d^0)$. (Recall that each time period lasts $1/N$.) Assume that $W_{m-1} \leq 1/(N\mu_d^{m-1})$; we shall show that $W_m \leq 1/(N\mu_d^m)$.

Consider some $x_i^+ \in X_m$ such that $W(x_i^+) = W_m$. From x_i^+ the process moves forward to a state in X_{m+1} with probability $p(x_i^+)$, it stays put with probability $1 - p(x_i^+) - q(x_i^+)$, and it moves backward to a state in X_{m-1} with probability $q(x_i^+)$. In the last case, it will take at most W_{m-1} periods in order for the process to return to X_m. It follows immediately that $W(x_i^+)$ satisfies

$$\left(p(x_i^+) + q(x_i^+)\right)W(x_i^+) \leq 1/N + p(x_i^+) \cdot 0 + q(x_i^+)(W_{m-1} + W_m).$$

Noting that $W(x_i^+) = W_m$ it follows that

$p(x_i^+)W_m \leq 1/N + q(x_i^+)W_{m-1}$.We know $p(x_i^+) - q(x_i^+) \geq \mu_d^m$, and the induction hypothesis states that $W_{m-1} \leq 1/(N\mu_d^{m-1})$. It follows that

$$\left(q(x_i^+) + \mu_d^i\right)W_m \leq \frac{\mu_d^m}{N\mu_d^m} + q(x_i^+)\frac{1}{N\mu_d^{m-1}}$$
$$\leq \frac{\mu_d^m + q(x_i^+)}{N\mu_d^m}.$$

15

The second inequality uses the fact that $\mu_d^{m-1} = 2\mu_d(1 - (m-1)/N) > 2\mu_d(1 - m/N) = \mu_d^m$. Simplifying we obtain $W_m \leq 1/(N\mu_d^m)$. This concludes the induction step. Using (9) and the inequalities that we have just proved we conclude that

$$ET(\alpha, \beta, G) \leq \frac{1}{2\mu_d} \cdot \frac{1}{N} \left(\sum_{m=0}^{\lfloor N/2 \rfloor} \frac{1}{1 - m/N} \right).$$

The second term is a discrete approximation of the integral

$$\int_0^{1/2} \frac{1}{1-x} dx = \log 2 \approx 0.69.$$

It is straightforward to show that this term is strictly less than 0.84 for any $N \geq 3$, and we obtain $ET(\alpha, \beta, G) < 0.42/\mu_d$. In fact, for any $N \geq 100$ this inequality improves to $ET(\alpha, \beta, G) < 0.35/\mu_d$. This concludes the proof of Lemma 2 and that of Theorem 1. ∎

3.2 Upper bound on the expected waiting time

We now show how to obtain a bound on the absolute magnitude of the waiting time for *any* regular graph, irrespective of size or degree, using a technique similar to one contained in the proof of Theorem 1.

Fix the payoff gain α and the noise level β and let G be a d-regular graph on N vertices. Given an adoption level $\bar{x}(t) \leq 1/2$, the expected change in the adoption rate is given by (4). We want to lower-bound this quantity *independently of d*. To this effect, let $f_{\min}(x)$ denote the lower envelope of the convex hull of the graph $\{(x, f(x)) : x \in [0,1]\}$. The function $f_{\min}(x) - x$ is convex, so it follows that for any d

(10) $$E\bar{x}\left(t + \frac{1}{N}\right) - \bar{x}(t) \geq \frac{1}{N}\left(f_{\min}(\bar{x}(t)) - \bar{x}(t)\right).$$

It is easy to show that $f_{\min}(x) - x \geq 2\mu(1-x)$, where $\mu = \mu(\alpha, \beta) \equiv f_{\min}(1/2) - 1/2$. (The proof is similar to that of Lemma 1.)

16

We can now apply Lemma 2 to establish a uniform upper bound on the expected waiting time, for any d and any d-regular graph (of any size), in terms of the shape of the function $f_{\min}$. Specifically, we have that for any α and β such that $\mu > 0$, and for any regular graph G we have

$$ET(\alpha, \beta, G) < \frac{0.42}{\mu}.$$

Figure 4 shows the expected waiting time to reach a majority of adopters, based on the above upper bound and on numerical simulations of the term $\mu = f_{\min}(1/2) - 1/2$. The expected waiting time is at most $100, 60$ and 40 revisions per capita for payoff gains α lying above the red, orange and green lines, respectively.

The figure shows, for example, that for $\varepsilon = 5\%$ and $\alpha = 1.10$, the expected waiting time is at most 60 revisions per capita, for regular graphs of arbitrary size.
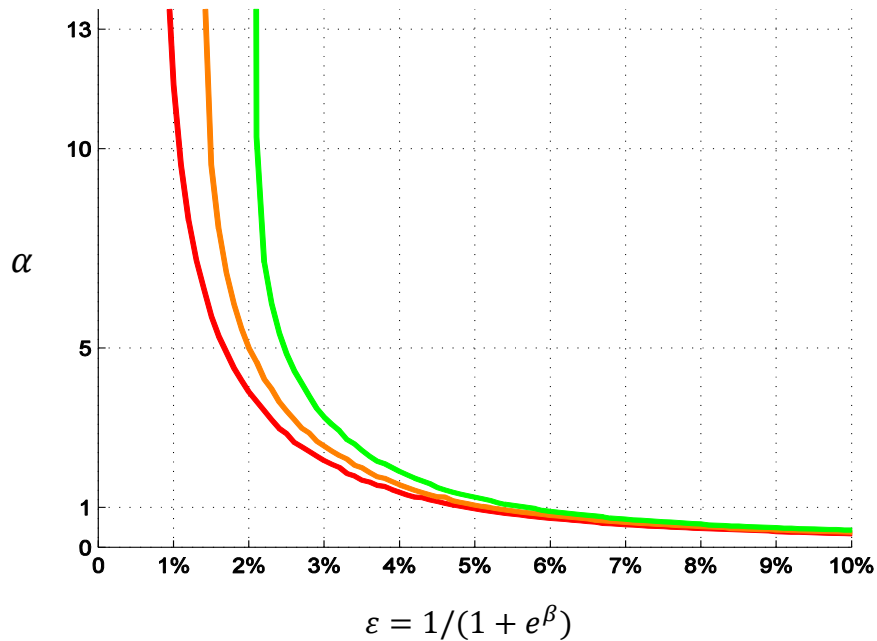

$$\varepsilon = 1/(1 + e^{\beta})$$

Figure 4 – Expected waiting times to reach a majority of adopters.
The expected waiting time is at most 100, 60 and 40 revisions per capita
for payoff gains above the red, orange and green lines, respectively.

17

4 General networks

In this section we shall show how our framework can be applied to more general families of graphs. We shall derive a similar result to Theorem 1, but the proof is more complex and relies on stopping time results in the theory of martingales. Consider a finite graph G, and let d_i denote the degree of agent i. Denote the average degree in the graph by $\bar{d} = (1/N) \sum_{i=1}^{N} d_i$. We will use the following measure of the adoption rate. For a state $x(t)$, define

$$\tilde{x}(t) = \frac{1}{N\bar{d}} \sum_{i=1}^{N} d_i x_i(t).$$

This is the probability of interacting with an adopter when placed at the end of a randomly chosen link in the graph (López-Pintado 2006, Jackson and Yariv 2007). In particular, adopters with higher degrees are weighted more heavily because they are more "visible" to other players. Note that the adoption rate $\tilde{x}(t)$ always lies in the interval $[0,1]$.

The definitions of the expected waiting time $ET(\alpha, \beta, G)$ and of fast diffusion introduced in section 2 extend naturally to the adoption rate measure $\tilde{x}(t)$. Concretely, a family $\mathcal{G}$ of graphs *displays fast diffusion* if for any graph $G \in \mathcal{G}$ the expected waiting time until $\tilde{x}(t) \geq 1/2$ is bounded independently of the size of G. Note that the analysis that follows carries through if we define fast diffusion relative to a threshold $p \in (0,1)$ other than $1/2$.

We shall begin by considering graphs with degrees bounded above by some integer $D \geq 3$. Denote by $\mathcal{G}^+(D)$ the family of all such graphs. Fix a state $x(t)$ such that $\tilde{x}(t) \leq \frac{1}{2}$. We propose to study the expected change in the adoption rate $E\tilde{x}(t + 1/N) - \tilde{x}(t)$.

Choose some individual i and let $n_i(t) \equiv \sum_{j \in N_i} x_j(t)$ be the number of adopters among i's neighbors at time t. The expected change in i's adoption rate is

$$Ex_i\left(t + \frac{1}{N}\right) - x_i(t) = \frac{N-1}{N} x_i(t) + \frac{1}{N} f\left(\frac{n_i(t)}{d_i}\right) - x_i(t)$$

$$= \frac{1}{N}\left(f\left(\frac{n_i(t)}{d_i}\right) - x_i(t)\right).$$

Summing across all agents we obtain

$$E\tilde{x}\left(t + \frac{1}{N}\right) - \tilde{x}(t) = \frac{1}{N\bar{d}}\sum_{i=1}^{N}\frac{1}{N}\left(d_i f\left(\frac{n_i(t)}{d_i}\right) - d_i x_i(t)\right).$$

Note that we can rewrite the adoption rate as

$$\tilde{x}(t) = \frac{1}{N\bar{d}}\sum_{i=1}^{N}d_i x_i(t)$$

$$= \frac{1}{N\bar{d}}\sum_{i=1}^{N}\sum_{j\in N_i}x_j(t)$$

(11)
$$= \frac{1}{N\bar{d}}\sum_{i=1}^{N}d_i\frac{n_i(t)}{d_i}.$$

Hence the expected change in the state variable can be written

(12)
$$E\tilde{x}\left(t + \frac{1}{N}\right) - \tilde{x}(t) = \frac{1}{N\bar{d}}\sum_{i=1}^{N}\frac{1}{N}\left(d_i f\left(\frac{n_i(t)}{d_i}\right) - d_i\frac{n_i(t)}{d_i}\right).$$

Let $X_D = \left\{\frac{k}{d} : 0 \le k \le d \le D \text{ and } d \ne 0\right\}$ and consider the set $F_D = \{(x, f(x)) : x \in X_D\}$. Let $\tilde{f}_D$ denote the lower envelope of the convex hull of the set F_D. Note that $f(x) \ge \tilde{f}_D(x)$ for all $x \in X_D$ so (12) implies that

$$E\tilde{x}\left(t + \frac{1}{N}\right) - \tilde{x}(t) \ge \frac{1}{N\bar{d}}\sum_{i=1}^{N}\frac{1}{N}\left(d_i\tilde{f}_D\left(\frac{n_i(t)}{d_i}\right) - d_i\frac{n_i(t)}{d_i}\right).$$

By definition $\tilde{f}_D$ is convex, and a fortiori so is $\tilde{f}_D(x) - x$. By applying Jensen's inequality and using (11) we obtain

(13)
$$E\tilde{x}\left(t + \frac{1}{N}\right) - \tilde{x}(t) \ge \frac{1}{N}\left(\tilde{f}_D(\tilde{x}(t)) - \tilde{x}(t)\right).$$

Thus, a sufficient condition for the expected motion to be positive is that $\tilde{f}_D$ lies strictly above the identity on the interval $[0, 1/2]$. The function $\tilde{f}_D(x) - x$ is decreasing on this interval, hence this condition is equivalent to

19

(14)
$$\mu(\alpha, \beta, D) \equiv \tilde{f}_D\left(\frac{1}{2}\right) - \frac{1}{2} > 0.$$

Note that $\tilde{f}_D(1/2)$ is a weighted average of $f(1)$ and $f(k/d)$, where k/d approximates the point on the interval $[0,1/2]$ where f is closest to the diagonal. Define the function

$$\tilde{h}(\beta, D) = \inf\{\alpha : \mu(\alpha, \beta, D) \geq 0\}.$$

Several key properties of the function $\tilde{h}(\beta, D)$ are summarized in the following result.

Proposition 1. $\tilde{h}(\beta, D) < D - 2$, *and* $\tilde{h}(\beta, D) < 0.83$ *for* $\varepsilon \geq 5\%$ ($\beta < \log(19)$).

Proof. The intuition for the first statement is that when $\alpha > D - 2$ it is a best response to play A whenever at least one neighbor plays A. Formally, for $\alpha \geq D - 2$ we have $f(k/d) \geq 1/2$ for all $0 < k \leq d \leq D$. It follows that the lower envelope of the convex hull of the set $F_d \equiv \left\{\left(\frac{k}{d}, f\left(\frac{k}{d}\right)\right) : 0 \leq k \leq d\right\}$ is given by the line that joins $(0, f(0))$ and $(1, f(1))$. This implies that $\mu(\alpha, \beta, D) = (f(0) + f(1))/2 - 1/2$, which is strictly positive. The last two statements of Proposition 1 follow from inequality (7) in the proof of Theorem 1 ∎

In order to study the family of *all* networks, we can consider the limit as D tends to infinity. We claim that $\tilde{h}(\beta) = \sup_D \tilde{h}(\beta, D)$ is finite for all β. Indeed, let $\tilde{f}_{\min}$ denote the lower envelope of the convex hull of the set $\{(x, f(x)) : x \in [0,1]\}$, and let

$$\mu(\alpha, \beta) = \tilde{f}_{\min}(1/2) - 1/2.$$

If we define $\tilde{h}(\beta) = \inf\{\alpha : \mu(\alpha, \beta) \geq 0\}$, then for all $\alpha > \tilde{h}(\beta)$ the expected change in the adoption rate is positive for any network as long as $\tilde{x}(t) \leq 1/2$.

The next result extends Theorem 1 to the family of all networks. It establishes the existence of a payoff threshold for fast diffusion, as well as an absolute bound on the expected waiting time.

Theorem 2. *Given a noise level $\beta > 0$, if the payoff gain α exceeds the threshold $\tilde{h}(\beta)$ then diffusion is fast for all graphs. Moreover with $\mu = \mu(\alpha, \beta)$, for any graph the expected waiting time until the adoption rate is at least $1/2$ satisfies*

$$ET(\alpha, \beta, G) < \frac{1}{\mu}.$$

Theorem 2 shows that for any graph the expected time until the adoption rate $\tilde{x}(t)$ reaches $1/2$ is *uniformly bounded* above as long as the payoff gain is greater than a threshold value that depends on the noise level. Moreover, the theorem provides an explicit bound on the expected waiting time that is easily computed and has a simple geometric interpretation.

We can improve on the threshold $\tilde{h}(\beta)$ in Theorem 2 if we restrict ourselves to graphs with degrees bounded by some number D. Specifically, for any integer $D \geq 3$, if the payoff gain exceeds $\tilde{h}(\beta, D)$ then diffusion is fast for all graphs with degrees bounded by D. Moreover, if $\mu = \mu(\alpha, \beta, D)$, the expected waiting time satisfies $ET(\alpha, \beta, G) < 1/\mu$ for every graph $G \in \mathcal{G}^+(D)$.

A notable feature of these results is that the bounds are topology-free: they do not depend on any of the network details. The only other result in the literature of this nature is due to Young (2011, Proposition 4), who shows that $\alpha > D - 2$ guarantees fast diffusion for all graphs in $\mathcal{G}^+(D)$. (The same bound arises in a number of other evolutionary models that are based on deterministic best-response processes; see in particular Ellison (1997), Morris (2000), Sandholm (2001) and Oyama et al (2011)). Proposition 1 shows that the bound $\tilde{h}(\beta, D)$ is better than this; indeed for many families of graphs it is much better.

Figure 5 plots the threshold $\tilde{h}(\beta, D)$ for several values of D. The figure also includes the threshold $\tilde{h}(\beta)$; combinations of α and β above this line have the property that diffusion is fast for the family of all finite graphs.
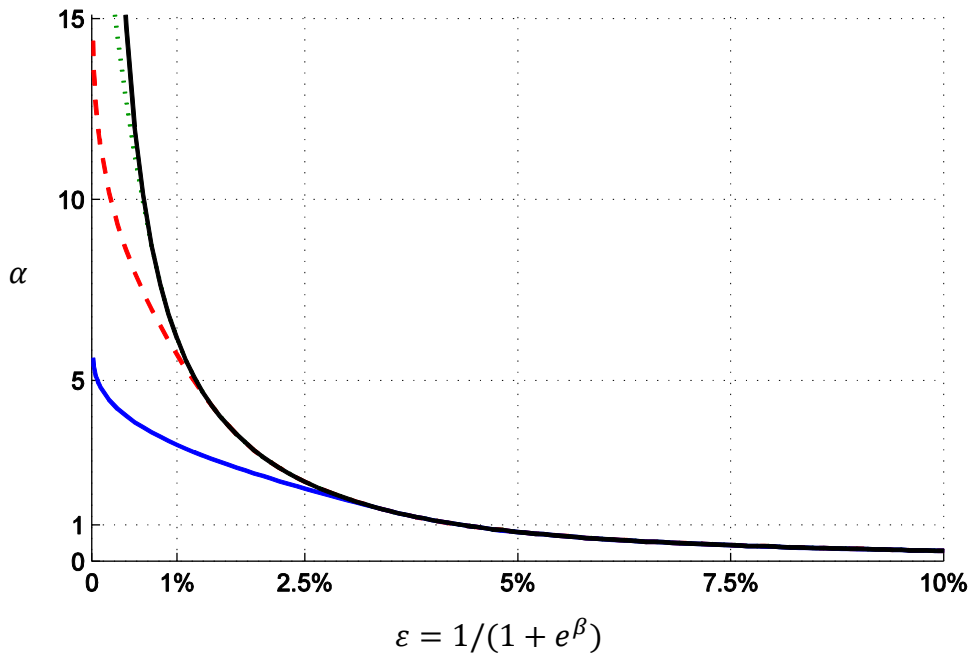
Figure 5 – Threshold $\tilde{h}(\beta, D)$ for $D = 10$ (blue, solid), $D = 25$ (red, dashed), $D = 50$ (green, dots) and $\tilde{h}(\beta)$ (black, solid).

For each pair (α, ε) and for any value of D, the term $\mu = \mu(\alpha, \beta, D)$ is uniquely determined by the shape of the function $f(\cdot; \alpha, \beta)$ and can be easily calculated. Table 1 presents the bounds on the expected waiting time established in Theorem 2. The bounds apply to all finite graphs, irrespective of maximum degree or size. For example, when $\varepsilon = 5\%$ and $\alpha = 2$, it takes at most 55 revisions per capita in expectation until the adoption rate $\tilde{x}(t)$ exceeds $1/2$.

Remark: Note that the waiting time upper bound in Theorem 2 cannot be lower than $2/\varepsilon$. Indeed, the points $(0, \varepsilon)$ and $\big(1, f(1)\big)$ are always part of the graph of f and $\tilde{f}_D$, which implies that μ is at most $\varepsilon/2$.

Table 1 – Upper bounds on the expected waiting time for any graph.

	ε			
	1%	2.5%	5%	10%
$\alpha = 1$	–	–	205	30
$\alpha = 2$	–	–	55	21
$\alpha = 4$	–	116	41	20
$\alpha = 8$	437	81	40	20

These numbers are expressed in terms of revisions per capita. The actual rate at which individuals revise will depend on the particular situation that is being modeled. For example, consider a new communication technology that is twice as good as the old technology when used by two agents who communicate with each other. Suppose that people review their decision about which technology to use once a week, and that they choose a best response nine times out of ten initially (when there are no adopters). The model predicts that the new technology will be widely adopted in less than seven months irrespective of the network topology.

Proof of Theorem 2. Inequalities (13) and (14) imply that for any $\tilde{x}(t) \leq 1/2$ we have

(15) $$E\tilde{x}(t + 1/N) - \tilde{x}(t) \geq \mu/N.$$

We want to estimate the stopping time $T = \min\{t : \tilde{x}(t) \geq 1/2\}$. Note that the process $\tilde{x}(t)$ stopped at T is a submartingale. Define the process $X(t) = \tilde{x}(t) - (t \cdot N)\mu/N$ and note that by (15) we know that $X(t)$ stopped at T is still a submartingale.

Doob's Optional stopping theorem says that if $X(t)$ is a submartingale and if the stopping time T satisfies $P(T < \infty) = 1$, $E|X(T)| < \infty$ and $E(X(T)|T > n)\Pr(T > n) \to 0$ as $n \to \infty$, then $EX(T) \geq EX(0)$ (Grimmett and Stirzaker 1992). In our case, note that there exists $p > 0$ such that with probability at least p we have $\tilde{x}(t + 1/N) - \tilde{x}(t) \geq p$, which implies that $E(T) < \infty$. This implies the three conditions of the theorem. Rewriting this result we obtain

$$E\tilde{x}(T) \geq \mu \cdot E(T).$$

Noting that $\tilde{x}(T) \leq 1$ we obtain $E(T) \leq 1/\mu$, as was to be proved.[7] ∎

Remark. When more is known about the specific degree distribution we can obtain even more precise results. To illustrate, suppose we are given a degree distribution $P = (p_d)_d$, where p_d denotes the fraction of agents in the network that have degree d. To be specific let us consider the case where P is described by the truncated power law $p_d \propto d^{-\gamma}$ for some $\gamma > 2$. (Such a network is said to be *scale-free*). Empirical studies of real networks show that they often resemble scale-free networks; examples include author citation networks, the World Wide Web, and the networks of sexual partners (Liljeros et al 2001). A network formation process that generates scale-free networks is the preferential attachment model, where a new agent added to the network is more likely to link with existing nodes that have a high degree (Barabási and Albert 1999). However, we want to stress that apart from the degree distribution, our results do not impose any constraints on the realized *topology* of the network.

Figure 6 plots a numerically simulated upper bound on the threshold for fast diffusion for the truncated distribution $p_d = \eta d^{-2.5}$ on the interval $5 \leq d \leq 5000$, where η is a normalizing constant. These simulations show that for small noise levels the threshold for the truncated power law is significantly lower compared to the threshold for all graphs.

---

[7] The method of proof for Theorem 1 is different because then we could establish the stronger bound $0.42/\mu$ instead of $1/\mu$.
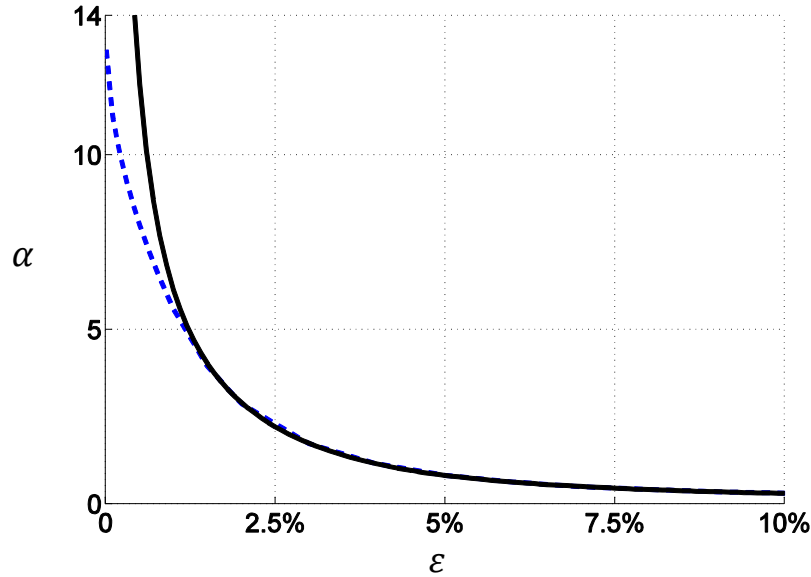
Figure 6 – The estimated bound (blue, dashed) for topology-free fast diffusion for a power-law degree distribution ($p_d \propto d^{-2.5}$ for $5 \leq d \leq 5000$). [8] Also shown is the threshold $\tilde{h}(\beta)$ (black, solid).

The estimated curve in Figure 6 has a very similar shape to the curves in Figure 5. In particular, fast diffusion is achieved when the noise level is 5% and the payoff gain is at least $\alpha = 0.81$. Changing the exponent and the parameters of the degree distribution yields qualitatively and quantitatively similar threshold functions.

5   Smooth stochastic best response functions

In this paper we have established the existence of a payoff gain threshold that ensures fast diffusion in networks. The bound is "topology-free" in the sense that it holds for *all* networks. In this section we shall show that our method for proving these results does not depend in any crucial way on the logit response function: results similar to Theorems 1 and 2 hold for a large family of response functions that are qualitatively similar to the logit, and that arise from idiosyncratic payoff shocks.Assume that at the start of each period each agent's payoffs from playing A and B are perturbed by independent payoff shocks ϵ_A and ϵ_B. These shocks alter the player's payoffs from all interactions during the period. In particular, agent i's payoff from an interaction with j when playing x_i is $\pi(x_i, x_j) + \epsilon_{x_i}$. His total payoff over the period is

---

[8] Empirical studies of scale-free networks typically find exponents between 2.1 and 4 (Barabási and Albert 1999).

$$\sum_{j \in N_i} \pi(x_i, x_j) + d_i \epsilon_{x_i},$$

where N_i is the set of i's neighbors, which has d_i elements.

Note that the shocks ϵ_A and ϵ_B are assumed to be constant for all of the agent's interactions in a given period; they are also independent across agents and across periods. Let us assume that the shocks are identically distributed with cumulative distribution function $\Theta(\epsilon; \beta)$ and density $\theta(\epsilon; \beta)$. The variance of the shocks is captured by the noise parameter β; as β tends to infinity the variance goes to zero. Let $\Theta_2(\Delta; \beta)$ denote the cdf of the random variable $\epsilon_A - \epsilon_B$.

Suppose that x is the proportion of adopters (i.e., A-players) in the neighborhood of a given agent i in a given period. The expected payoff difference from choosing A or B in a single interaction is

$$\Delta(x; \alpha) = (1 + \alpha)x - (1 - x).$$

If agent i has d_i neighbors, the payoff difference from choosing A versus B equals $d_i \Delta(x; \alpha)$.

Assume that whenever an agent revises his action, he chooses a (myopic) best response given his realized payoffs in that period. Thus from an observer's standpoint the probability that i chooses A is

$$
\begin{aligned}
f_\theta(x; \alpha, \beta) &= \Pr\big((1 + \alpha)d_i x + d_i \epsilon_A > d_i(1 - x) + d_i \epsilon_B\big) \\
&= \Pr\big(d_i(\epsilon_A - \epsilon_B) > -d_i \Delta(x; \alpha)\big) \\
&= \Pr\big((\epsilon_A - \epsilon_B) > -\Delta(x; \alpha)\big) \\
&= 1 - \Theta_2(-\Delta(x; \alpha); \beta).
\end{aligned}
$$

Note that this probability does not depend on the agent's degree.

If the shocks are drawn from an extreme-value distribution of the form $\Theta(\epsilon; \beta) = 1 - \exp(-\exp(\beta \epsilon))$, then $\epsilon_A - \epsilon_B$ is distributed according to a logistic distribution, and the resulting response function is the logit.

Another natural example is given by normally distributed payoff shocks. If the shocks are normally distributed with mean zero and standard error β^{-1}, the resulting response function is the probit choice rule (Myatt and Wallace 2003), given by

$$\Pr(A \mid x, \alpha, \beta) = \Phi\left(\frac{\beta\Delta(x; \alpha)}{\sqrt{2}}\right),$$

where Φ is the cumulative distribution function of the standard normal distribution. It will be seen that this response function, plotted in Figure 7, is very similar in shape to the logit. Moreover, the bound between slow and fast diffusion is qualitatively very similar to the other models.
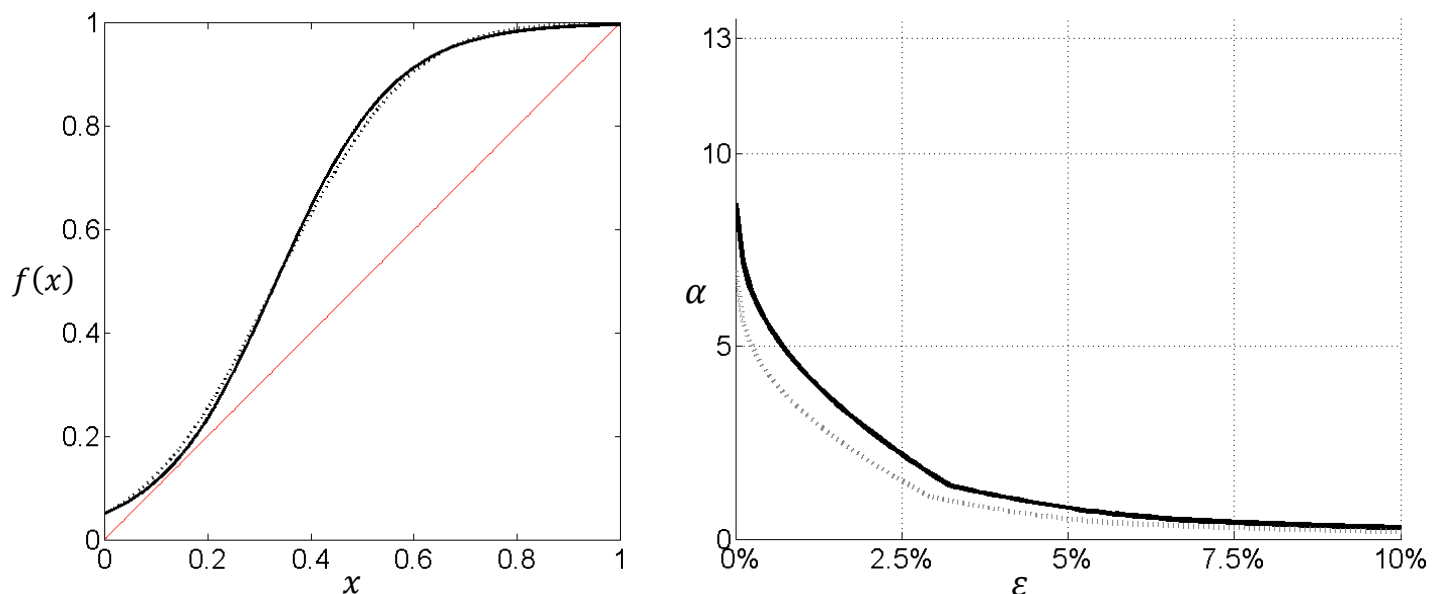
Figure 7 *Left Panel:* Logit response function (black solid) and normally distributed payoff shocks (gray dots). ($\alpha = 1$, $\varepsilon = 5\%$). *Right panel:* Thresholds for topology-free fast diffusion: logit (black solid) and normally distributed payoff shocks (gray dots). ($d = 15$)

In general, consider a family of densities $\theta(\epsilon; \beta)$ with parameter $\beta > 0$. Assume that θ satisfies the following conditions:

1. $\theta(\epsilon; \beta)$ is continuous; for every $\beta > 0$ it is quasi-concave in ϵ and symmetric around 0;

27

2. For any $\beta' > \beta$ the distribution given by $\theta(\cdot; \beta')$ second-order stochastically dominates the distribution given by $\theta(\cdot; \beta)$. Moreover, $\theta(\epsilon; \beta)$ tends to zero as β tends to zero, and it tends to the Dirac delta function $\delta(\epsilon)$ as β tends to infinity.

The first condition implies that $\epsilon_A - \epsilon_B$ is also quasi-concave and symmetric around zero. This implies that the function $\Theta_2(\Delta; \beta)$ is convex for $\Delta < 0$ and concave for $\Delta > 0$. Hence the response function f_θ has a convex-concave shape. The second condition says that the amount of payoff disturbances is arbitrarily large for small β, decreases in β and tends to zero as β tends to infinity. It follows that f_θ converges to random choice for β close to 0 and to the best response function for β large.

We are now in a position to extend Theorem 2 to general families of response functions. Let $\tilde{f}_{\theta,\min}$ denote the lower envelope of the convex hull of the set $\{(x, f_\theta(x)) : x \in [0,1]\}$, and let

$$\mu(\alpha, \beta, \theta) = \tilde{f}_{\theta,\min}\left(\frac{1}{2}\right) - \frac{1}{2}.$$

The next result extends Theorem 2 to families of payoff shocks that satisfy conditions 1 and 2.

Theorem 3. *Assume that agents experience independent and identically distributed payoff shocks with a density $\theta(\epsilon; \beta)$ that satisfies conditions 1 and 2. For any noise level $\beta > 0$, there exists a payoff threshold $\tilde{h}(\beta, \theta)$ such that whenever $\alpha > \tilde{h}(\beta, \theta)$, diffusion is fast for all graphs. Moreover, if $\mu = \mu(\alpha, \beta, \theta) > 0$, then for every graph G the expected waiting time until the adoption rate is at least $1/2$ is at most $1/\mu$.*

We now present a different characterization of response functions that are similar to the logit, which turns out to be equivalent to the one we have just described. A salient characteristic of the logit is that the probability of an error can be expressed as a decreasing function of the payoff difference between the two alternatives. As Blume (2003) has shown, the risk-dominant equilibrium remains stochastically stable for a large class of such response functions.

We begin by modeling the probability of making an error. An *error function* $\xi: \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow [0,1]$ takes as arguments the relative payoff difference $\Delta \geq 0$ and a noise parameter β. We assume that ξ has the following properties:

i) ξ is continuous, it is decreasing and convex in Δ and decreasing in β,

ii) $\xi(0;\beta) = 1$ and $\lim_{\Delta\to\infty} \xi(\Delta;\beta) = 0$ for any $\beta > 0$,

iii) $\xi(\Delta, 0) = 1$ and $\lim_{\beta\to\infty} \xi(\Delta,\beta) = 0$ for any $\Delta > 0$.

Recall that an agent's total payoff difference between choosing A and B is $d\Delta(x;\alpha)$, where x is the fraction of adopters in her neighborhood, d is her degree, and $\Delta(x;\alpha)$ is the expected payoff difference from a single interaction when the probability of interacting with an adopter is x. For example, when there are no adopters the payoff difference is equal to $-d$. Hence the term $\Delta(x;\alpha)$ may also be interpreted as the payoff difference between choosing A and B, *relative* to the state when there are no adopters. Denote the best response function by

$$BR(x;\alpha) = \begin{cases} 1, & \text{if } \Delta(x;\alpha) > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Let the probability that an agent chooses A be given by

$$f_\xi(x;\alpha,\beta) = BR(x;\alpha)\big(1 - \xi(|\Delta(x;\alpha)|;\beta)\big) + \frac{1}{2}\xi(|\Delta(x;\alpha)|;\beta).$$

f_ξ is a weighted average of the best response and an equal probability random choice. The weight on the random choice is equal to the error function evaluated at the relative payoff difference between choosing A and B. Because of our assumption that ξ is decreasing in Δ, the distance between f_ξ and the best response is decreasing in the payoff difference between the two actions.

The function f_ξ is qualitatively similar to the logit function. It is increasing, continuous, initially convex and then concave, and converges to the best response as the noise level β tends to infinity. Figure 8 plots three examples of functions f_ξ for $\alpha = 1$, normalized such that $\varepsilon = 5\%$. The logit response function is obtained from the error function $\xi_{logit}(\Delta,\beta) = 2/(1 + e^{\beta\Delta})$, and the exponential and power error functions are given by $\xi_{exp} = e^{-\beta|\Delta|}$ and $\xi_{pow} = (1 + \beta|\Delta|)^{-3}$.
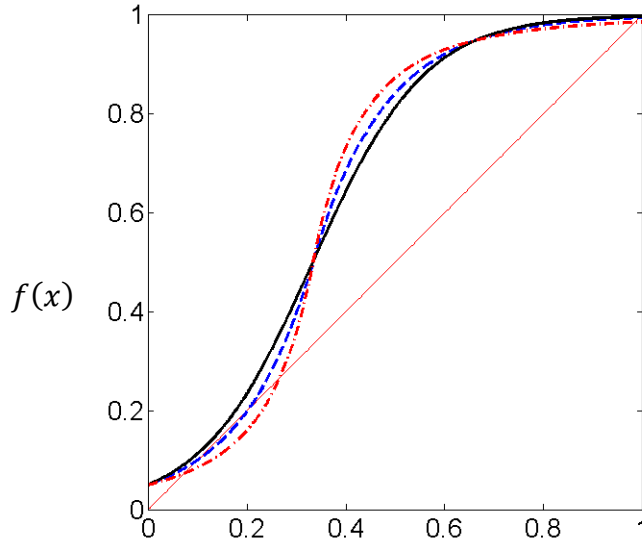
Figure 8 – Response functions: logit (black solid), exponential error function (blue dashed) and power error function (red dot-dash). ($\alpha = 1$, $\varepsilon = 5\%$)

The following result establishes that the characterization of families of response functions in terms of error functions is equivalent to the characterization using payoff disturbances.

Proposition 2. *There exists a one-to-one correspondence φ between error functions ξ and families of payoff disturbances π that satisfy conditions 1 and 2, such that ξ and $\pi = \varphi(\xi)$ generate the same family of response functions, namely for any $\beta > 0$*

$$f_\xi(x; \alpha, \beta) = f_\pi(x; \alpha, \beta).$$

The straightforward proof is omitted. It follows from Proposition 2 and Theorem 3 that the main message of Theorems 1 and 2 carries through for *any* error function ξ.

6 Conclusion

In this paper we have studied some of the factors that affect the speed of diffusion of innovations on social networks. The two main factors that we identify are the payoff gain of the innovation relative to the status quo and the amount of noise in the players' response functions.

As has been noted by a number of authors, including Griliches (1957) in his classic study of hybrid corn, larger payoff gains tend to increase the speed with which an innovation spreads.

This makes intuitive sense. A less obvious but equally important factor is the amount of noise or variability in the players' behavior. This variability can be variously interpreted as errors, experimentation, or unobserved payoff shocks. Under all of these interpretations, greater variability tends to increase the speed at which an innovation spreads. The reason is that higher variability makes it easier to escape from the initial low equilibrium. A particularly interesting finding is that different combinations of variability and payoff gain determine a threshold above which diffusion is fast in a 'strong' sense, namely the expected diffusion time is uniformly bounded irrespective of population size and interaction structure. These results are robust to quite general parameterizations of the variability in the system. For the logit, which is commonly used in empirical work, the waiting time is bounded (and quite small absolutely) if the initial error rate is at least 5% and the payoff gain from the innovation is at least 83% relative to the status quo.

Unlike previous results, a central feature of our analysis is that the bounds on waiting time are *topology-free*. The results apply to all networks irrespective of size and structure. In addition, our method of analysis extends to families of graphs with restrictions on the degree distribution. The virtue of this approach is that it yields concrete predictions that are straightforward to compute even when the fine details of the network structure are unknown, which is arguably the case in many real-world applications. Moreover, in practice social networks are constantly in flux, which makes predictions that depend on the specific network topology quite problematic. We conjecture that our framework can be extended to settings where the network co-evolves with players' choices, as in Jackson and Watts (2002) or Staudigl (2010) for example, but this issue will be left for future work.

References

Anderson, S., De Palma, A. and Thisse, J. (1992): Discrete Choice Theory of Product Differentiation, Cambridge, U.S.A.: MIT Press.

Bala, V. and Goyal S. (1998): Learning from Neighbours, The Review of Economic Studies, Vol. 65, No. 3, pp. 595-621.

Barabási, A.-L. and Albert, R. (1999): Emergence of Scaling in Random Networks, Science 15, Vol. 286 no. 5439 pp. 509-512.

Blume, L.E. (1993): The Statistical Mechanics of Strategic Interaction, Games and Economic Behavior, 5(3), 387–424.

Blume, L.E. (1995): The Statistical Mechanics of Best-Response Strategy Revision, Games and Economic Behavior, 11(2), 111–145.

Blume, L.E. (2003): How noise matters, Games and Economic Behavior, 44(2), 251–271.

Blume, L.E. and Durlauf, S. (2003), Equilibrium Concepts for Social Interaction Models, International Game Theory Review, Vol. 5, No. 3, 193–209.

Brock, W. A. and Durlauf, S. N. (2001): Interactions-based models, Handbook of Econometrics, in: Heckman, J.J. and Leamer, E.E. (eds.), Handbook of Econometrics, 1(5), chapter 54, 3297–3380.

Coleman, J., Katz, E. and Menzel, H. (1957), The Diffusion of an Innovation Among Physicians, Sociometry, Vol. 20, No. 4, pp. 253-270.

Ellison, G. (1993): Learning, Local Interaction, and Coordination, Econometrica, 61(5), 1047-71.

Ellison, G. (1997): Learning from Personal Experience: One Rational Guy and the Justification of Myopia, Games and Economic Behavior 19, 180-210.

Golub, Benjamin, and Matthew O. Jackson (2010), "Naïve Learning in Social Networks and the Wisdom of Crowds," American Economic Journal: Microeconomics, 2, 112-149.

Griliches, Z. (1957): Hybrid Corn: An Exploration in the Economics of Technological Change, Econometrica, Vol. 25, No. 4, pp. 501-522.

Grimmett, G. and Stirzaker, D., Probability and Random Processes, Oxford University Press, 1992.

Jackson, Matthew O. (2010), Social and Economic Networks. Princeton NJ: Princeton University Press.

Jackson, M. O., and Rogers, B. W. (2007): Meeting Strangers and Friends of Friends: How Random Are Social Networks? American Economic Review, 97(3): 890-915.

Jackson, M.O. and Watts, A. (2002): On the formation of interaction networks in social coordination games, Games and Economic Behavior 41: 265–291.

Jackson, M.O. and Yariv, L. (2007): Diffusion of Behavior and Equilibrium Properties in Network Games, American Economic Review, 97(2), 92-98.

Kandori, M., Mailath, G.J. and Rob, R. (1993): Learning, Mutation, and Long Run Equilibria in Games, Econometrica, 61(1), 29-56.

Kreindler, G.E. and Young, H.P. (2013): Fast convergence in evolutionary equilibrium selection, Games and Economic Behavior 80: 39-67.

Kremer, M. and Levy, D. (2008): Peer Effects and Alcohol Use among College Students, The Journal of Economic Perspectives, Vol. 22, No. 3, pp. 189-206.

Liljeros F., Edling C.R., Nunes Amaral L.A., Stanley H.E. and Aberg Y. (2001), The Web of Human Sexual Contacts, Nature 411, 907–908.

López-Pintado, D. (2006): Contagion and coordination in random networks, International Journal of Game Theory, 34(3), 371-381.

Myatt, D.P. and Wallace, C. (2003): A Multinomial Probit Model of Stochastic Evolution, Journal of Economic Theory, 113(2). pp 286-301. 2003

McFadden, D.L. (1976): Quantal Choice Analaysis: A Survey, NBER Chapters, in: Annals of Economic and Social Measurement, 5(4), 8-66.

McKelvey R.D. and Palfrey T.R. (1995) Quantal response equilibria for normal form games, Games and Economic Behavior, 10, 6-38.

Montanari, A. and Saberi, A. (2010): The spread of innovations in social networks, Proceedings of the National Academy of Sciences, 107(47), 20196-20201.

Morris, S. (2000): Contagion, Review of Economic Studies, 67(1), 57-78.

Munshi, K. and Myaux, J. (2006): Social norms and the fertility transition, Journal of Development Economics 80, 1-38.

Newton, J. and Angus, S.D. (2013), Coalitions, tipping points and the speed of evolution, mimeo.

Oyama, D., Sandholm, W. and Tercieux, O. (2011): Sampling best response dynamics and deterministic equilibrium selection, mimeo.

Sandholm, W.H. (2001): Almost global convergence to p-dominant equilibrium, International Journal of Game Theory, 30, 107-116.

Sandholm, W. H. (2010): Population Games and Evolutionary Dynamics, Cambridge MA: MIT Press.

Staudigl, M. (2010): On a General Class of Stochastic Co-Evolutionary Dynamics, Working Paper No 1001, Department of Economics, University of Vienna.

Valente, Thomas (1995), Network Models of the Diffusion of Innovations. Cresskill NJ: Hampton Press.

Valente, Thomas (2005), "Network models and methods for studying the diffusion of innovations," in Peter J. Carrington, John Scott, and Stanley Wasserman (eds.), Models and Methods in Social Network Analysis. Cambridge: Cambridge University Press.

Vega-Redondo, Fernando (2007), *Complex Social Networks*, Cambridge: Cambridge University Press.

Watts, D.J. (2002): A simple model of global cascades on random networks, Proceedings of the National Academy of Sciences, 99(9), 5766–5771.

Watts and Dodds (2007): Influentials, Networks, and Public Opinion Formation, Journal of Consumer Research, 34(4), 441-445.

Young, H.P. and Burke, M.A. (2001): Competition and Custom in Economic Contracts: A Case Study of Illinois Agriculture, The American Economic Review, Vol. 91, No. 3, pp. 559-573.

Young, H.P. (1993): The Evolution of Conventions, Econometrica 61:57-84.

Young, H.P. (1998): Individual Strategy and Social Structure: An Evolutionary Theory of Institutions, Princeton, NJ: Princeton University Press.

Young, H.P. (2011): The dynamics of social innovation, Proceedings of the National Academy of Sciences, 108(4), 21285–21291.