

Rapid Innovation Diffusion in Social Networks

Gabriel E. Kreindler^a and H. Peyton Young^b

April 12, 2014

Abstract. Social and technological innovations often spread through social networks as people respond to what their neighbors are doing. Previous research has identified specific network structures, such as local clustering, that promote rapid diffusion. Here we derive bounds that are independent of network structure and size, such that diffusion is fast whenever the payoff gain from the innovation is sufficiently high and the agents' responses are sufficiently noisy. We also provide a simple method for computing an upper bound on the expected time it takes for the innovation to become established in any finite network. For example, if agents choose log-linear responses to what their neighbors are doing, it takes on average less than 80 revision periods 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.

\body

1 Local interaction topology and fast diffusion

New ideas and ways of doing things often spread through social networks. People tend to adopt an innovation with increasing likelihood depending on the proportion of their friends and neighbors

^a Department of Economics, Massachusetts Institute of Technology.

^b Department of Economics, University of Oxford, and the Institute for New Economic Thinking at the Oxford Martin School.

who have adopted it. The innovation in question might be a technological advance such as a new piece of software, a medical drug (1), or a new hybrid seed (2). Or it might represent a social practice, such as contraception (3), a novel form of employment contract (4), or a group behavior such as binge drinking (5).

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 ((2), (6)). Others have pointed to the role that the network topology plays in the rate at which an innovation spreads ((7), (8), (6), (9), (10), (11), (12), (13), (14), (15), (16), (17), (18), (19)).

A key finding of this literature is that the amount of local clustering has a significant effect on the speed with which the innovation spreads. Figure 1 presents three examples of simple networks that differ in the amount of local clustering. Panel (a) shows a randomly generated network in which the neighborhoods of nearby agents have relatively little overlap. In such a network the expected waiting time until the innovation spreads grows exponentially in the network size unless the payoff gains from the innovation are very large ((11), (18)). Panel (b) is a ring in which agents are connected to nearby agents. In this case the expected waiting time until the innovation becomes widely adopted is bounded independently of the network size ((20)). The intuition for this result is that the innovation gains an initial foothold relatively quickly within small groups of adjacent agents on the ring, and it then maintains its hold even when those outside the group have not yet adopted. Thus the innovation tends to diffuse quite rapidly as it gains independent footholds among many groups scattered across the network. A similar logic applies to the lattice in panel (c); in this case the local footholds consist of squares or rectangles of agents.

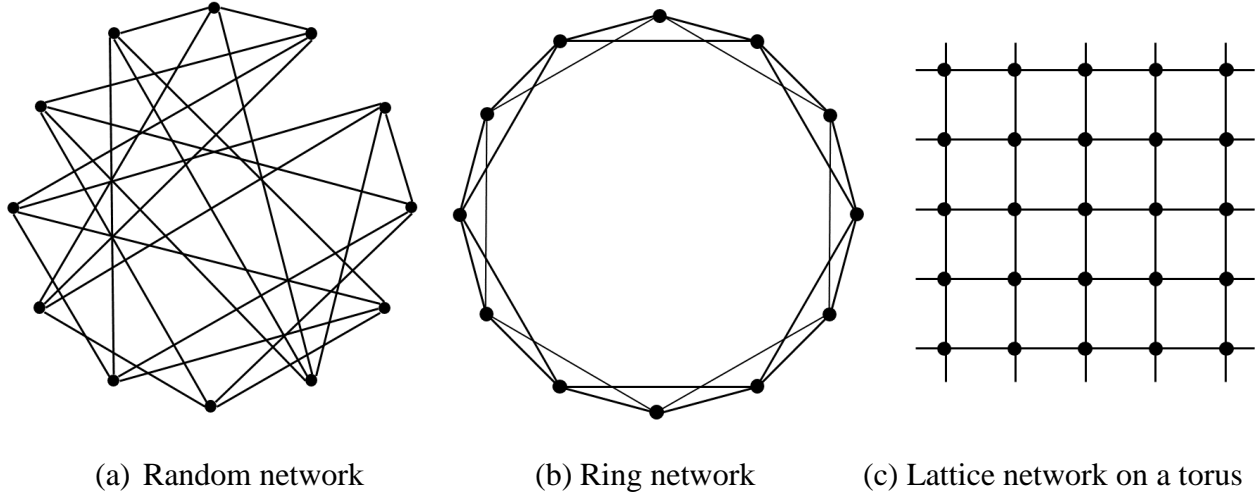


Figure 1 – 4-regular networks with different topologies.

The contribution of this article is to show that innovations can in fact spread quite rapidly in any finite network (as long as it does not have isolated vertices). The methods we use to establish our results are quite different from those in the prior networks literature, many of which rely on the “independent footholds” argument outlined above. Instead we take our inspiration from another branch of the learning literature, namely the analysis of dynamics when agents choose noisy best responses to samples drawn at random from the whole population. If the population is large, such a process can be well-approximated by a deterministic (mean-field) dynamical system ((21), (22), (23), (24), (14), (25), (26)).

This set-up is 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 2. 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 of attraction of the status quo equilibrium; hence the easier it is to escape that equilibrium ((14), (24)).

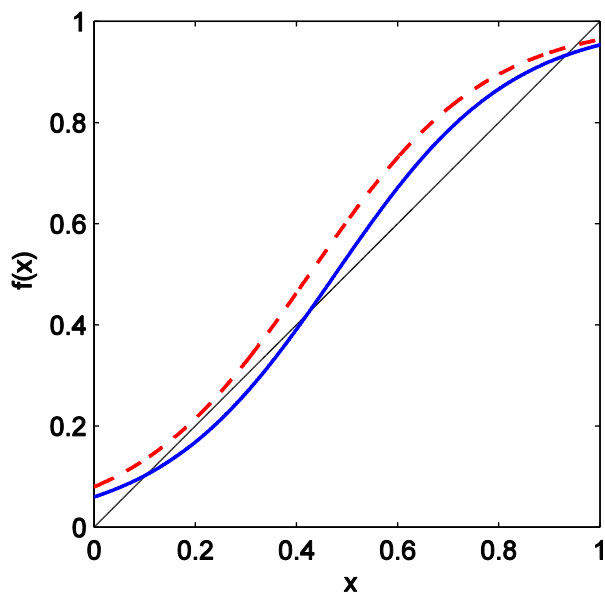


Figure 2 – The logit response function.

Of particular interest is the case where there is a single rest point and it lies above the halfway mark $(.5, .5)$.^{*} (See the red dashed curve in Figure 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 (26). 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

^{*} In almost any normal form game with the logit response function, for sufficiently large noise levels there is a unique rest point (a quantile equilibrium) (20). In the case when agents best-respond to random samples, 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 ((21), (24)).

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 2. 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 shape of the stochastic response function. Thus our approach provides a practical estimation method that can in principle be applied to empirical data.

We would like to emphasize that our results do not imply that network topology does not matter at all. Indeed, the diffusion rate will be faster than the upper bound that we establish, for classes of networks identified by previous research ((27), (14)). Thus, our results are particularly useful in situations where the exact topology of the network is not known or is changing over time.

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 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 . We assume

throughout this paper that each agent is connected to at least one other agent, i.e. the graph G does not have isolated nodes. Each agent chooses one of two available actions, A and B . Interaction is given by a symmetric 2×2 coordination game with payoff matrix

$$[1] \quad \begin{array}{cc} & \begin{array}{c} A \\ B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{array}{cc} a, a & c, d \\ d, c & b, b \end{array} \end{array}$$

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

$$\begin{array}{cc} & \begin{array}{c} A \\ B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{array}{cc} a - d & 0 \\ 0 & b - c \end{array} \end{array}$$

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 ((28), (29), (30)).

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

$$[2] \quad \begin{array}{cc} & \begin{array}{c} A \\ B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{array}{cc} 1 + \alpha, 1 + \alpha & 0, 0 \\ 0, 0 & 1, 1 \end{array} \end{array}$$

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.[†] 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:

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

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 $\varepsilon = f(0; \alpha, \beta) = 1/(1 + e^\beta)$ 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

[†] 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.

neighbors, that is, on the agent's degree. In section 5 we show that 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. The logit is one of the most commonly used models of this type ((31), (32), (7), (8), (33)). 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.[‡]

The above revision process fully describes the stochastic process $\Gamma(\alpha, \beta, G)$. The states of the process are the adoption vectors $x(t) = (x_i(t))_{1 \leq i \leq N}$, 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 $x(0) = (0, \dots, 0)$.

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 . Formally, define the random hitting time

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

This measure is appropriate because it captures the time it takes for the system to escape the status quo equilibrium (B, B) . 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.

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

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

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 $\Gamma(\alpha, \beta, G)$ 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. Throughout this section, $d \geq 1$. Denote by $\mathcal{G}(d)$ the family of all d -regular graphs. Fix a payoff gain α , a noise level β and a graph $G \in \mathcal{G}(d)$. 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

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

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

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

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. We begin by expressing $\bar{x}(t)$ as the population average of the fraction of i 's neighbors who have adopted the innovation. Let N_i denote the set of neighbors of i . We can then write

$$\begin{aligned} \bar{x}(t) &= \frac{1}{N} \sum_{j=1}^N x_j(t) = \frac{1}{N} \sum_{i=1}^N \frac{1}{d} \sum_{j \in N_i} x_j(t) \\ &= \sum_{k=0}^d q_k(t) \frac{k}{d}. \end{aligned}$$

Note that the weights $q_k(t)$ “convexify” the set of points $F_d \equiv \left\{ \left(\frac{k}{d}, f\left(\frac{k}{d}\right) \right) : 0 \leq k \leq d \right\}$. Denote by $f_d(x)$ the lower envelope of the convex hull of F_d . Figure 3 illustrates the construction.

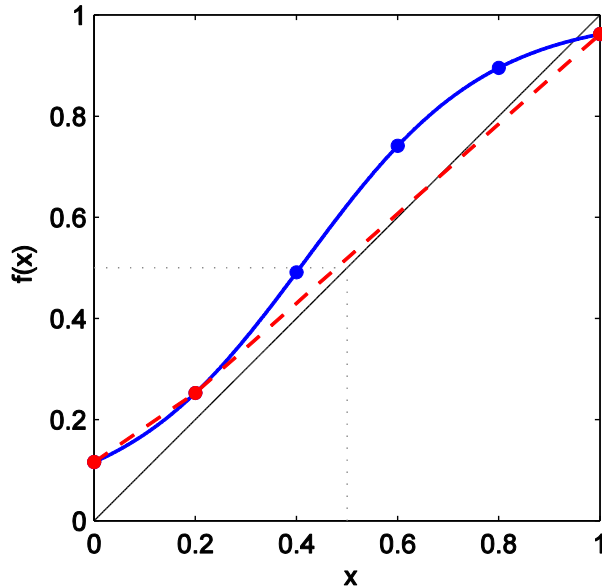


Figure 3 –The function f_d (red dashed line) is the lower envelope of the convex hull of the set F_d (filled dots).

We now apply Jensen's inequality for the convex function f_d to obtain

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

Using identity [4] we obtain

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

Given α, β and $d \geq 1$ define the quantities

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

and

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

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 one, 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 $d \geq 1$, $\Gamma(\alpha, \beta, \mathcal{G}(d))$ 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 ((20), (27), (18)). 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 $\mu_d \equiv \mu(\alpha, \beta, d)$, 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 Claim 1 relies on elementary calculus using the logit formula, and can be found in the Supporting Information document.

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 $\alpha > h(\beta, d)$, 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

$$[6] \quad E\bar{x}\left(t + \frac{1}{N}\right) - \bar{x}(t) \geq \frac{2}{N}\mu_d(1 - \bar{x}(t)) > 0.$$

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 $\mu_d = \mu(\alpha, \beta, d)$.

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*

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

The proof of Lemma 2 can be found in the SI Appendix. It is based on inequality [6] and a careful accounting of the worst-case scenario expected waiting time to go from a state with adoption rate m/N to any state with adoption rate $(m + 1)/N$, where $m \in \{0, 1, \dots, \lfloor N/2 \rfloor\}$.

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

$$[7] \quad E\bar{x}\left(t + \frac{1}{N}\right) - \bar{x}(t) \geq \frac{1}{N}\left(f_{\min}(\bar{x}(t)) - \bar{x}(t)\right).$$

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.)

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 $\mu = \mu(\alpha, \beta) > 0$, 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 blue, red 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.

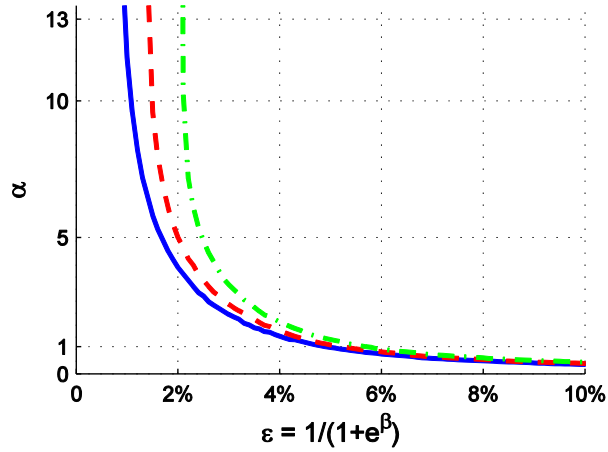


Figure 4 – Expected waiting times to reach a majority of adopters. The expected waiting time is at most 100, 60 and 40 revisions per capita for payoff gains above the blue (solid), red (dash) and green (dash dot) lines, respectively.

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 . We assume that $d_i \geq 1$ for all i , i.e. there are no isolated nodes. 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 adoption. 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 G ((24), (14)). 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 $[0,1]$, and it reduces to the usual definition when G is regular.

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 $D \geq 1$. 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, namely $E\tilde{x}(t + 1/N) - \tilde{x}(t)$.

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

$$\begin{aligned} E x_i \left(t + \frac{1}{N} \right) - x_i(t) &= \frac{1}{N} f \left(\frac{n_i(t)}{d_i} \right) + \frac{N-1}{N} x_i(t) - x_i(t) \\ &= \frac{1}{N} \left(f \left(\frac{n_i(t)}{d_i} \right) - x_i(t) \right). \end{aligned}$$

Taking the average across all agents we obtain

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

We can rewrite the adoption rate as

$$\begin{aligned} \tilde{x}(t) &= \frac{1}{N\bar{d}} \sum_{j=1}^N d_j x_j(t) = \frac{1}{N\bar{d}} \sum_{i=1}^N \sum_{j \in N_i} x_j(t) \\ &= \frac{1}{N\bar{d}} \sum_{i=1}^N d_i \frac{n_i(t)}{d_i}. \end{aligned}$$

[8]

Hence the expected change in the state variable can be written

$$[9] \quad E\tilde{x}\left(t + \frac{1}{N}\right) - \tilde{x}(t) = \frac{1}{N\bar{d}} \sum_{i=1}^N \frac{1}{N} \left(d_i f\left(\frac{n_i(t)}{d_i}\right) - d_i \frac{n_i(t)}{d_i} \right).$$

Let $X_D = \left\{ \frac{k}{d} : 0 \leq k \leq d \leq D \text{ and } d \neq 0 \right\}$ 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 $f(x) \geq \tilde{f}_D(x)$ for all $x \in X_D$ so [9] implies that

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

By definition \tilde{f}_D is convex, and a fortiori so is $\tilde{f}_D(x) - x$. By applying Jensen's inequality and using [8] we obtain

$$[10] \quad E\tilde{x}\left(t + \frac{1}{N}\right) - \tilde{x}(t) \geq \frac{1}{N} \left(\tilde{f}_D(\tilde{x}(t)) - \tilde{x}(t) \right).$$

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

$$[11] \quad \tilde{\mu}(\alpha, \beta, D) \equiv \tilde{f}_D\left(\frac{1}{2}\right) - \frac{1}{2} > 0.$$

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 identity function. Define the function

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

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

Proposition 1. $\tilde{h}(\beta, D) \leq \max(D - 2, 0)$, and $\tilde{h}(\beta, D) < 0.83$ for $\varepsilon \geq 5\%$ ($\beta < \log(19)$).

Proof. The intuition for the first statement is that when $\alpha > \max(D - 2, 0)$ it is a best response to play A whenever at least one neighbor plays A . Formally, for $\alpha \geq \max(D - 2, 0)$ 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 $\tilde{\mu}(\alpha, \beta, D) = (f(0) + f(1))/2 - 1/2$, which is strictly positive. The last two statements of Proposition 1 follow from Claim 1 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 $\sup_D \tilde{h}(\beta, D)$ is finite for all β . Indeed, let f_{\min} denote the lower envelope of the convex hull of the set $\{(x, f(x)) : x \in [0, 1]\}$, and let

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

If we define $\tilde{h}(\beta) \equiv \inf\{\alpha : \tilde{\mu}(\alpha, \beta) \geq 0\}$, 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 $\tilde{\mu} = \tilde{\mu}(\alpha, \beta)$, for any graph the expected waiting time until the adoption rate is at least $1/2$ satisfies*

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

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 $D \geq 1$, if the payoff gain exceeds $\tilde{h}(\beta, D)$ then diffusion is fast for all graphs with degrees bounded by D . Moreover, if $\tilde{\mu} = \tilde{\mu}(\alpha, \beta, D)$, the expected waiting time satisfies $ET(\alpha, \beta, G) < 1/\tilde{\mu}$ 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 (34), (11), (22) and (25)). 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.

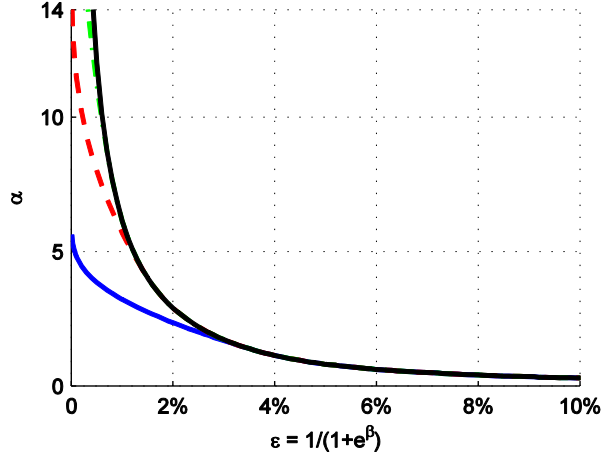


Figure 5 – Threshold $\tilde{h}(\beta, D)$ for $D = 10$ (blue, solid), $D = 25$ (red, dashed), $D = 50$ (green, dash dot) and $\tilde{h}(\beta)$ (black, solid).

For each pair (α, ε) and for any value of D , the term $\tilde{\mu} = \tilde{\mu}(\alpha, \beta, D)$ 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 $(1, f(1))$ are always part of the graph of f and \tilde{f}_D , which implies that $\tilde{\mu}$ is at most $\varepsilon/2$.

[Table 1 here]

The numbers in Table 1 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. Inequality [10] implies that for any $\tilde{x}(t) \leq 1/2$ we have

$$[12] \quad E\tilde{x}(t + 1/N) - \tilde{x}(t) \geq \tilde{\mu}/N.$$

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 $X(t) \equiv \tilde{x}(t) - (t \cdot N)\tilde{\mu}/N$ and note that by [12] 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 $E(X(T)|T > n) \Pr(T > n) \rightarrow 0$ as $n \rightarrow \infty$, then $EX(T) \geq EX(0)$ (35). 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

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

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

Remark. The analysis so far corresponds to a worst-case analysis over all graphs with degrees below a certain D . When the degree distribution is known, this information can be used in conjunction with identity [9] to derive a more precise payoff threshold for fast diffusion. 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 (36). 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 (37). 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.

[§] In the proof of Theorem 1 we use a different method in order to establish the stronger bound $0.42/\mu$ instead of $1/\mu$.

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.

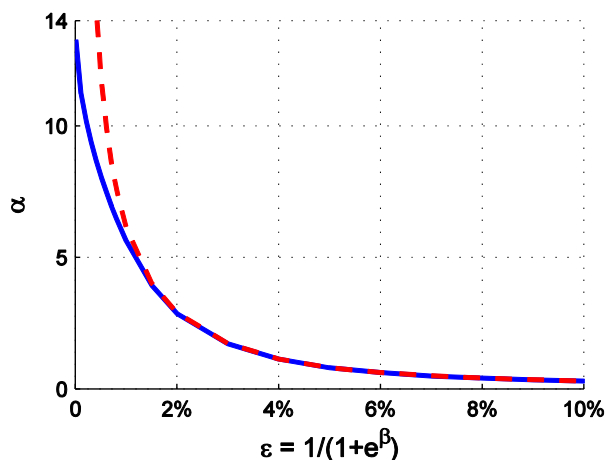


Figure 6 – The estimated bound (blue, solid) for topology-free fast diffusion for a power-law degree distribution ($p_d \propto d^{-2.5}$ for $5 \leq d \leq 5000$).** Also shown is the threshold $\tilde{h}(\beta)$ (red, dashed).

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.

** Empirical studies of scale-free networks typically find exponents between 2.1 and 4 (37).

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

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

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

We stress 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. If agent i has d_i neighbors, the payoff difference from choosing A versus B equals $d_i \Delta(x; \alpha)$, where $\Delta(x; \alpha)$ denotes the expected difference from a single interaction, namely

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

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

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

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* (38), 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.

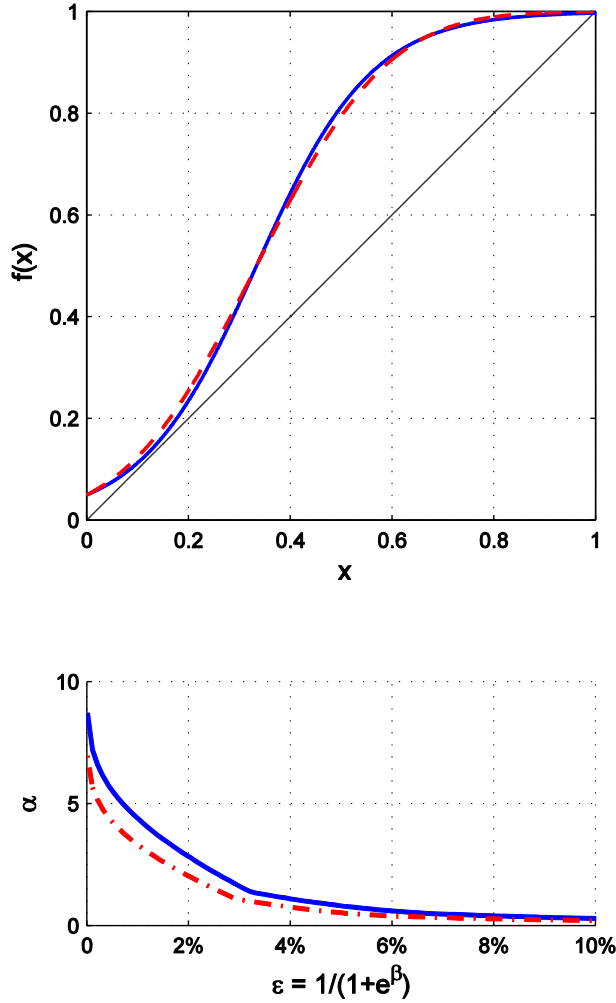


Figure 7 *Top Panel:* Logit response function (blue, solid) and normally distributed payoff shocks (red, dash). ($\alpha = 1$, $\epsilon = 5\%$) *Bottom panel:* Thresholds for topology-free fast diffusion: logit (blue, solid) and normally distributed payoff shocks (red, dash). ($d = 15$)

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;
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

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

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 drawn from 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 $\tilde{\mu} = \tilde{\mu}(\alpha, \beta, \theta) > 0$, then for every graph G the expected waiting time until the adoption rate is at least $1/2$ is at most $1/\tilde{\mu}$.*

In the SI Appendix we present a different characterization of response functions that are similar to the logit, in terms of *error functions*. 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. The risk-dominant equilibrium remains stochastically stable for a large class of such response functions (30). We show that there is a one-to-one correspondence between decreasing, convex error functions, and families of payoff disturbances. It follows that the main message of Theorems 1 and 2 carries through for a general class of error functions.

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 in his classic study of hybrid corn (2), 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 (39) or (40) for example, but this issue will be left for future work.

7 Acknowledgements

We thank Glenn Ellison and participants in the 2014 Sackler Colloquium, the MIT theory workshop and the 2013 Brown University Mini-Conference on Networks for constructive suggestions. This research was sponsored by the Office of Naval Research, grant N00014-09-1-0751 and the Air Force Office of Scientific Research grant #FA9550-09-1-0538.

8 References

- 1 J. Coleman, E. Katz and H. Menzel, "The Diffusion of an Innovation Among Physicians," *Sociometry*, vol. 20, no. 4, pp. 253-270, 1957.
- 2 Z. Griliches, "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, vol. 25, no. 4, pp. 501-522, 1957.
- 3 K. Munshi and J. Myaux, "Social norms and the fertility transition," *Journal of Development Economics*, vol. 80, pp. 1-38, 2006.
- 4 H. Young and M. Burke, "Competition and Custom in Economic Contracts: A Case Study of Illinois Agriculture," *The American Economic Review*, vol. 91, no. 3, pp. 559-573, 2001.
- 5 M. Kremer and D. Levy, "Peer Effects and Alcohol Use among College Students," *The Journal of Economic Perspectives*, vol. 22, no. 3, pp. 189-206, 2008.
- 6 V. Bala and G. S., "Learning from Neighbours," *The Review of Economic Studies*, vol. 65, no. 3, pp. 595-621, 1998.
- 7 L. Blume, "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, vol. 5, no. 3, pp. 387-424, 1993.
- 8 L. Blume, "The Statistical Mechanics of Best-Response Strategy Revision,," *Games and Economic Behavior*, vol. 11, no. 2, p. 111-145, 1995.
- 9 T. Valente, *Network Models of the Diffusion of Innovations*, Cresskill NJ: Hampton Press, 1995.
- 10 T. Valente, "Network models and methods for studying the diffusion of innovations," in *Models and Methods in Social Network Analysis*, Cambridge, Cambridge University Press, 2005.
- 11 S. Morris, "Contagion," *Review of Economic Studies*, vol. 67, no. 1, pp. 57-78, 2000.
- 12 D. Watts, "A simple model of global cascades on random networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, p. 5766-5771, 2002.
- 13 D. Watts and P. Dodds, "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, vol. 34, no. 4, pp. 441-445, 2007.
- 14 M. Jackson and L. Yariv, "Diffusion of Behavior and Equilibrium Properties in Network Games," *American Economic Review*, vol. 97, no. 2, pp. 92-98, 2007.
- 15 F. Vega-Redondo, *Complex Social Networks*, Cambridge: Cambridge University Press, 2007.
- 16 B. Golub and M. Jackson, "Naïve Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, vol. 2, pp. 112-149, 2010.
- 17 M. O. Jackson, *Social and Economic Networks*, Princeton NJ: Princeton University Press, 2010.
- 18 A. Montanari and A. Saberi, "The spread of innovations in social networks," *Proceedings of the National Academy of Sciences*, vol. 107, no. 47, pp. 20196-20201, 2010.
- 19 J. Newton and S. Angus, "Coalitions, tipping points and the speed of evolution," *University of Sydney School of Economics Working Paper 2013-02*, 2013.
- 20 G. Ellison, "Learning, Local Interaction, and Coordination," *Econometrica*, vol. 61, no. 5, pp. 1047-1071, 1993.
- 21 R. McKelvey and T. Palfrey, "Quantal response equilibria for normal form games," *Games and Economic Behavior*, vol. 10, pp. 6-38, 1995.

- 22 W. Sandholm, "Almost global convergence to p-dominant equilibrium," *International Journal of Game Theory*, vol. 30, pp. 107-116, 2001.
- 23 L. Blume and S. Durlauf, "Equilibrium Concepts for Social Interaction Models," *International Game Theory Review*, vol. 5, no. 3, p. 193–209, 2003.
- 24 D. López-Pintado, "Contagion and coordination in random networks," *International Journal of Game Theory*, vol. 34, no. 3, pp. 371-381, 2006.
- 25 D. Oyama, W. Sandholm and O. Tercieux, "Sampling best response dynamics and deterministic equilibrium selection," forthcoming in *Theoretical Economics*, 2014.
- 26 G. Kreindler and H. Young, "Fast convergence in evolutionary equilibrium selection," *Games and Economic Behavior*, vol. 80, pp. 39-67, 2013.
- 27 H. Young, *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton, NJ: Princeton University Press, 1998.
- 28 M. Kandori, G. Mailath and R. Rob, "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica*, vol. 61, no. 1, pp. 29-56, 1993.
- 29 H. Young, "The Evolution of Conventions," *Econometrica*, vol. 61, pp. 57-84, 1993.
- 30 L. Blume, "How noise matters," *Games and Economic Behavior*, vol. 44, no. 2, p. 251–271, 2003.
- 31 D. L. McFadden, "Quantal Choice Analysis: A Survey," in *Annals of Economic and Social Measurement*, Volume 5, number 4, National Bureau of Economic Research, Inc., 1976, pp. 363-390.
- 32 S. Anderson, A. De Palma and J. Thisse, *Discrete Choice Theory of Product Differentiation*, Cambridge, U.S.A: MIT Press, 1992.
- 33 W. A. Brock and S. N. Durlauf, "Interactions-based models," in *Handbook of Econometrics*, 1(5), chapter 54., 2001, p. 3297–3380.
- 34 G. Ellison, "Learning from Personal Experience: One Rational Guy and the Justification of Myopia," *Games and Economic Behavior*, vol. 19, pp. 180-210, 1997.
- 35 G. Grimmett and D. Stirzaker, *Probability and Random Processes*, Oxford University Press, 1992.
- 36 F. Liljeros, C. Edling, L. Nunes Amaral, H. Stanley and Y. Aberg, "The Web of Human Sexual Contacts," *Nature*, vol. 411, p. 907–908, 2001.
- 37 A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509-512, 1999.
- 38 D. Myatt and C. Wallace, "A Multinomial Probit Model of Stochastic Evolution," *Journal of Economic Theory*, vol. 113, no. 2, pp. 286-301, 2003.
- 39 M. Jackson and A. Watts, "On the formation of interaction networks in social coordination games," *Games and Economic Behavior*, vol. 41, p. 265–291, 2002.
- 40 M. Staudigl, "On a General Class of Stochastic Co-Evolutionary Dynamics|," Working Paper No 1001, Department of Economics, University of Vienna, 2010.