The Dynamics of Social Innovation

H. Peyton Young
University of Oxford

Sackler Colloquium on the Dynamics of Social, Political, and Economic Institutions

Irvine California
December 3, 2010

Social norms are the building blocks of social institutions, but the mechanisms that induce norm shifts are complex and not well understood. I argue that changes in norms are typically caused by one of three mechanisms: they can be triggered by idiosyncratic changes in individual perceptions and expectations, by changes in general external conditions (such as prices or technology), and by deliberate experimentation with novel ideas. While these mechanisms differ in some respects, they have similar dynamic effects: change tends to be slow and fitful at first, then accelerates rapidly once a critical threshold is crossed. Of particular importance to norm diffusion is the social network through which agents obtain information. We show how the topology of the network determines the ease with which norms shift, and formulate conditions under which the rate of norm displacement is rapid and independent of network size. The analysis also shows why norm shifts tend to occur in large jumps rather than through small incremental improvements.

Acknowledgements. I wish to thank Gabriel Kreindler for suggesting a number of improvements and for running the simulations reported here. The paper also benefited from comments by Tomas Rodriguez Barraquer and participants at the Sackler Colloquium on the Dynamics of Social, Political, and Economic Institutions, Irvine, December 3-4, 2010.

1. The diffusion of social innovations in social networks

Social institutions are forms of capital that, together with physical and human capital, determine the rate of economic growth. Notable examples include private property rights and legal systems to protect them, accounting methods, forms of corporate governance, and the terms of economic contracts. Yet, in contrast with the literature on technological progress, relatively little is known about the ways in which new institutions are created and how they become established within a given social framework. In this paper I discuss one approach to this problem using methods from evolutionary game theory.

In the abstract, an institution can be viewed as a set of rules that structure a given type of interaction between individuals (North, 1990). These rules can be simple coordination devices, such as which hand to extend in greeting or who goes through the door first. Or they can be very elaborate, such as rituals of courtship and marriage, cycles of retribution, performance criteria in employment contracts, or litigation procedures in the courts. We would like to know how a particular set of rules becomes established as common practice, and what process describes the displacement of one set of rules by another.

The viewpoint we shall adopt here is that new norms and institutions are introduced through a process of invention or discovery at the local level. This is followed by an extended period in which people try to coordinate using the new rules, possibly incorporating refinements along the way. Under certain conditions that we shall discuss below, this trial-and-error process will eventually trigger a general change in expectations and behaviors that establishes the new institution within society at large. However, it can take a

very long time for this to happen even though the new way of doing things is superior to the status quo.

One reason for this inertia is lack of information: it may not be immediately evident that an innovation actually is superior to the status quo, due to the small number of prior instances and variability in their outcomes. Thus it may take a long time for enough information to accumulate before it becomes clear that the innovation is superior (Young, 2009). A second reason is that an innovation as initially conceived may not work very well in practice; it must be refined over time through a process of learning by doing (Arrow, 1962). A third reason is that innovations often exhibit increasing returns. Indeed this feature is especially important for *social* innovations, which by definition require coordinated change in expectations and behaviors by multiple individuals. For example, an individual who invents a new form of legal contract cannot simply institute it on his own: first, the other parties to the contract must enter into it, and second, the ability to enforce the contract will depend on its usage in society more generally. Institutions exhibit strongly increasing returns precisely because of their function as coordination technologies.[1]

This increasing returns phenomenon has important implications for the dynamics governing institutional change. Of particular relevance is the social network through which individuals communicate. The reason is that, when a social innovation first appears, it will typically gain a foothold in a relatively small subgroup of individuals that are closely linked by geography or social connections. Once the new way of doing things has become firmly established

[1] Some forms of technological innovation also exhibit increasing returns due to network externalities (Katz and Shapiro,1985; Arthur, 1989).

within a local social group, it propagates to the rest of society through the social network. Thus a key determinant of the speed with which institutional change occurs is the *network topology*, and in particular the extent to which interactions are 'localized.'

Ellison (1993) was the first to study this issue for simple network structures such as circles and grids; subsequently the analysis was extended by Young (1998, 2003), Morris (2000), and Montanari and Saberi (2010) among others. Broadly speaking the aim of this literature is to characterize the rate of convergence of various learning processes as a function of key topological properties. For example, Young (1998, 2003) shows that log linear learning results in rapid diffusion when the network is composed of small close-knit enclaves or clusters. Montanari and Saberi (2010) extend this approach by exhibiting a topological measure called "tilted cutwidth" that allows one to characterize the rate of convergence in a wide range of networks as a function of network size. Morris (2000) provides necessary and sufficient conditions under which an innovation spreads throughout the network by deterministic best-response dynamics. This does not address the issue of speed of convergence directly; rather it identifies necessary and sufficient conditions for the feasibility of convergence when learning is deterministic.

The contribution of this paper is to show that it is not only the network topology that determines how fast an innovation spreads. In fact, the speed depends on the interaction between three complementary factors: i) the *payoff gain* represented by the innovation in relation to the status quo, ii) the degree of *rationality* of the agents (the probability with which they choose a best response given their information), and iii) the presence of *small autonomous enclaves* where

the innovation can gain an initial foothold. (This latter condition is related to Montanari and Saberi's concept of tilted cutwidth as well as to Young's concept of close-knittedness.) We also employ a somewhat subtle notion of what it means for an innovation to 'spread quickly'. Namely, we ask how long it takes in expectation for a high proportion of agents to adopt the innovation and stick with it with high probability. This latter condition is needed because, in a noisy best response process, it is quite possible for an innovation to spread initially, but then go into reverse and perhaps even die out. This is presumably not what we have in mind when we say that an innovation spreads successfully.

The main results of the paper can be summarized as follows. First we distinguish between fast and slow rates of diffusion. Roughly speaking, an innovation *spreads quickly* in a given class of networks if the expected waiting time to reach a given level of penetration (say 99%) and stay at that level with high probability (say 99%) is bounded above independently of the number of agents; otherwise the innovation *spreads slowly*. Whether it spreads quickly or slowly depends on the particular learning rule used, the degree of rationality of the agents, the gain in payoff from the innovation, and certain topological properties of the network. In much of the paper we shall focus on log linear learning to illustrate these points. Similar results hold for other noisy learning processes.

Among the key findings are the following.

1. If the agents' level of rationality is too low, the waiting time to spread successfully is very long (in fact it may be infinite) because there is too much noise in the system for a substantial proportion of the agents to *stay coordinated* on the innovation once it has spread initially. However, if the level of rationality is too high, it takes an exponentially long time in expectation for the innovation to gain a foothold anywhere. Hence only for intermediate levels of rationality can one expect the waiting time to be fast in an absolute sense and to be bounded independently of the number of agents.

2. Certain topological characteristics of the network promote fast learning. In particular, if the agents fall into small enclaves that are mainly connected with each other as opposed to outsiders, then learning will be fast (assuming the level of rationality is not too low and not too high). But not everyone has to be contained in a small enclave for this to be true: it suffices that the innovation be able to spread by contagion from a *subset* of enclaves where it can gain an initial foothold.

3. For convergence to be fast, it is not necessary for the agents to be contained in enclaves that are small in an absolute sense; it suffices that everyone be contained in a subgroup of bounded (possibly large) size that has a sufficiently high proportion of its interactions with other members of the group as opposed to outsiders. Various natural networks have this property, including those in which agents are embedded more or less uniformly in a finite Euclidean space, and are neighbors if and only if they are within some specified distance of one another. (This follows from Proposition 2 below; a similar result is proved by Montanari and Saberi (2010).)

4. The payoff gain from the innovation relative to the status quo – the size of the advance -- has an important bearing on the absolute speed with which it spreads. If the advance is sufficiently large, *no* special topological properties of the network are required for fast learning: it suffices that the maximum degree is bounded.

5. An innovation that leads to a small advance will tend to take exponentially longer to spread than an innovation with a large advance. This implies that successful innovations will often occur in big bursts, because a major advance may overtake prior attempts at small advances, which take too long relative to the former. This idea is illustrated by example toward the end of the paper.

2. The model

Let Γ be a graph with vertex set V and edge set E, where the edges are assumed to be undirected. Thus E is a collection of unordered pairs of vertices $\{i, j\}$ where $i \neq j$. Assume that there are n vertices, which we shall sometimes refer to as *nodes*. Each edge $\{i, j\}$ has a weight $w_{ij} = w_{ji} > 0$ which we shall interpret as a measure of the *mutual influence* that i and j have on one another. For example, w_{ij} may increase the closer that i and j are to each other geographically. Since we can always assume that $w_{ij} = w_{ji} = 0$ whenever $\{i, j\}$ is *not* an edge, the graph Γ is completely specified by a set V consisting of n vertices, and a symmetric nonnegative n x n matrix of weights $W = (w_{ij})$ where $w_{ii} = 0$ for all i.

Assume that each agent has two available choices, A and B. We shall think of B as the *status quo behavior* and A as the *innovative behavior*. The *state* of the

evolutionary process at time t is a vector $x^t \in \{A, B\}^n$, where x_i^t is i's choice at time t. Let G be a symmetric two person game with payoff function $u(x, y)$, which is the payoff to the player who chooses x against an opponent who chooses y. The payoff matrix can be written as follows:

$$\begin{array}{ccc} & A & B \\ A & a, a & c, d \\ B & d, c & b, b \end{array} \qquad (1)$$

We assume that $a > d$ and $b > c$, hence this is a coordination game with pure equilibria *(A, A)* and *(B, B)*.

The two-person game G induces an n-person *network game* on Γ as follows: the payoff to individual i in any given period results from playing the game G against each of his neighbors once, where i's current strategy x_i^t is *unconditional* on which neighbor he plays. The payoff from any given match is weighted according to the influence of that neighbor. For each agent i, let N_i denote the set of i's neighbors, that is, the set of all vertices j such that $\{i, j\}$ is an edge. Thus the payoff to i in state x is

$$U_i(x) = \sum_{j \in N_i} w_{ij} u(x_i, x_j). \qquad (2)$$

To give this a concrete interpretation, suppose that B is a form of contractual negotiation that relies solely on a verbal understanding and a handshake, whereas A requires a written agreement and a witness to the parties' signatures. If one side insists on a written agreement (A) while the other side views a verbal

understanding as appropriate (B), they will fail to coordinate because they disagree on the basic rules of the game. Note that this meta-game (agreeing on the rules of the game) can be viewed as a pure coordination game: if the parties fail to coordinate there are no payoff gains relative to the status quo. This situation is shown below, where α is the added benefit from having a written agreement.

$$\begin{array}{cc|cc} & & A & B \\ A & & 1+\alpha,\ 1+\alpha & 0,\ 0 \\ B & & 0,\ 0 & 1,\ 1 \end{array} \qquad (3)$$

This is the set-up we shall use throughout the paper.[2]

2.1. Learning

How do agents adapt their expectations and behaviors in such an environment? Several models of the learning process have been proposed in the literature; here we shall discuss two of the main contenders. Suppose that each agent updates at random times according to the realization of a Poisson arrival process with unit expectation. The processes are assumed to be independent among agents. Thus the probability is zero that two agents update simultaneously, and each agent updates once per unit time period in expectation. We shall think of each update

[2] This formulation is fairly general, because given any symmetric *2 x 2* game *G* we can rescale the payoffs so that the game has a potential function of this form (assuming *A* has higher potential than *B*). Most of our results hold with this interpretation of α, namely, that it is the gain in potential of *A* relative to *B*.

as a tick of the clock, where the ticks are denoted by *t = 1, 2, 3, …..* Each of these will be called a *period*.

Denote the state of the process at the start of period *t* by x^{t-1}. Let $\varepsilon > 0$ be a small error probability. Suppose that agent *i* is the unique agent who updates in period *t*. Assume that, with probability $1-\varepsilon$, *i* chooses an action *w* that maximizes $U_i(w, x_{-i}^{t-1})$, and with probability ε he chooses *A* or *B* uniformly at random. This is known as the *uniform error model* (Kandori, Mailath, and Rob, 1993; Young, 1993).

An alternative approach is the *log linear model* suggested by Blume (1993, 1995). Given a real number $\beta \geq 0$, assume that agent *i* chooses *A* with probability

$$\frac{e^{\beta U_i(A, x_{-i}^{t-1})}}{e^{\beta U_i(A, x_{-i}^{t-1})} + e^{\beta U_i(B, x_{-i}^{t-1})}}. \tag{4}$$

In other words, the log probability of choosing *A* minus the log probability of choosing *B* is β times the difference in payoff; hence the term 'log linear learning.'[3] The parameter β measures the *rationality* of the agent: the larger β is, the more likely it is that he chooses a best reply given the actions of his neighbors. In what follows we shall focus mainly on this case since it is particularly easy to work with analytically. The game will be specified by the size of the *advance* α, and the log learning process by the level of *rationality* β.

2.2. The speed of diffusion

[3] This model is also standard in the discrete-choice literature (McFadden, 1973). In statistical mechanics it is known as *heat bath* or *Glauber dynamics*.

Let $\vec{B}$ be the initial state in which everyone is following the status quo behavior *B*. How long does it take for the innovative behavior *A* to become widely adopted? One criterion would be the expected waiting time until the first time that everyone plays *A*. Unfortunately this definition is not satisfactory, because of the noise in the learning process. To understand the nature of the problem, consider a situation in which β is close to zero, so that the probability of playing *A* is only slightly larger than the probability of playing *B*. If society consists of *n* agents, the expected waiting time until the first time that all of them play *A* is on the order of 2^{-n} . But no matter how long we wait, the probability is high that a sizable proportion of the population will be playing *B* at any given future time. Thus the expected waiting time until everyone first plays *A* is not the relevant concept. This difficulty arises for any noisy learning process: if the noise is too large it is unlikely that everyone is playing *A* in any given period.

We are therefore led to the following definition, which was first proposed in Young (1998). Given a noisy learning process P^ε on a graph Γ (not necessarily log linear learning), for each state x let $\alpha(x)$ denote the *proportion* of agents playing *A* in state x. Given a target level of *penetration* $0 < p < 1$, define

$$T(P^\varepsilon, \Gamma, G, p) = E[\min \{t : \alpha(x^t) \geq p \,\&\, \forall t' \geq t, P(\alpha(x^{t'}) \geq p) \geq p] \ . \qquad (5)$$

In other words, $T(P^\varepsilon, \Gamma, G, p)$ is the expected waiting time until at least *p* of the agents are playing *A*, and the probability is at least *p* that at least this proportion plays *A* at all subsequent times. The waiting time depends on the learning process P^ε (including the specific level of noise ε), as well as on the graph Γ and

the game G. In particular, the higher the value of *p*, the smaller the noise must be or the waiting time as defined in (5) will be infinite.

To distinguish between fast and slow learning as a function of the number of agents, we shall consider families of networks of different sizes, where the *size* of a network is the number of nodes (equivalently, the number of agents).

Fast versus slow learning. Given a family of networks $\mathscr{G}$ and an innovation advance $\alpha > 0$, learning is *fast* for $\mathscr{G}$ and α if, for every $p < 1$ there exists $\varepsilon_p > 0$ such that for all $\varepsilon \in (0, \varepsilon_p)$,

$$T(P^\varepsilon, \Gamma, \alpha, p) \text{ is bounded above for all } \Gamma \in \mathscr{G}.^4 \qquad (6)$$

Otherwise learning is *slow,* that is, there is an infinite sequence of graphs $\Gamma_1, \Gamma_2, ..., \Gamma_n, ... \in \mathscr{G}$ such that $\lim_{n\to\infty} T(P^\varepsilon, \Gamma_n, \alpha, p) = \infty$.

2.3. Autonomy

In this section we describe a general condition on families of networks that guarantees fast learning. Fix a network $\Gamma = (V, W)$, a learning process P^ε, and an advance $\alpha > 0$. Given a subset of vertices $S \subseteq V$, define the *restricted learning process* $\tilde{P}_S^\varepsilon$ as follows: all nodes $i \notin S$ are held fixed at strategy *B* while the nodes in *S* update according to the process P^ε. Let $(\vec{A}_S, \vec{B}_{V-S})$ denote the state in which every member of *S* plays *A* and every member of $V - S$ plays *B*.

[4] For log linear learning, the definition is that the waiting time is bounded above for all networks in the family whenever the response parameter β is sufficiently large.

The set S is *autonomous for* $(P^\varepsilon, \Gamma, \alpha)$ if and only if $(\vec{A}_S, \vec{B}_{V-S})$ is stochastically stable for the restricted process $\tilde{P}_S^\varepsilon$.[5]

Proposition 1. *Given a learning process* P^ε, *a family of networks* $\mathcal{G}$, *and an innovation with advance* $\alpha > 0$, *suppose that there exists a positive integer s such that for every* $\Gamma \in \mathcal{G}$, *every node of* Γ *is contained in a subset of size at most s that is autonomous for* $(P^\varepsilon, \Gamma, \alpha)$. *Then learning is fast.*

Concretely this means that given any target level of penetration $p < 1$, there is an upper bound on the noise, $\varepsilon_{p,\alpha}$, such that for any given ε in the range $0 < \varepsilon < \varepsilon_{p,\alpha}$, the expected waiting time until at least p of the agents play A (and continue to do so with probability at least p in each subsequent period) is bounded above independently of the number of agents n in the network. This differs from the results of Montanari and Saberi (2010), who establish conditions under which the waiting time to reach all-*A* with high probability is bounded as a function of *n provided that the noise level is arbitrarily small* (which implies that the absolute waiting time is arbitrarily large). The proof of Proposition 1 is accomplished by a coupling argument similar to that in Young (1998, Chapter 6).

2.4. Autonomy and close-knittedness

[5] A state is *stochastically stable* if it has nonvanishing probability in the limit as the noise goes to zero (Foster and Young, 1990).

The autonomy condition has a particularly natural topological interpretation under the log linear learning model. Let $\Gamma = (V, W)$ be a graph and $\alpha > 0$ the size of the advance.

For every subset of vertices $S \subseteq V$ let

$$d(S) = \sum_{i \in S,\, j \in V} w_{ij} \,. \tag{7}$$

Further, for every nonempty subset $S' \subseteq S$ let

$$d(S', S) = \sum_{\{i,j\}:\, i \in S',\, j \in S} w_{ij} \,. \tag{8}$$

In other words, $d(S)$ is the weighted sum of edges with at least one end in S whereas $d(S', S)$ is the weighted sum of edges with one end in S' and the other end in S. Given any real number $r \in (0, 1/2]$, we say that the set S is *r-close-knit* if

$$\forall S' \subseteq S, S' \neq \varnothing, \quad d(S', S) / d(S') \geq r \,.^{6} \tag{9}$$

S is *r*-close-knit if no subset has 'too many' interactions with outsiders, where 'too many' means 'more than $1 - r$ of its interactions'. This implies in particular that no individual $i \in S$ has more than $1 - r$ of its interactions with outsiders. (The set S is said to be *r-cohesive* if this holds for all $i \in S$; see Morris (2000).)

[6] Montanari and Saberi (2010) employ a different concept in their analysis, namely the isoperimetric function of the graph, which is defined for each positive integer k as the minimum of $d(S, V - S)/|S|$ over all subsets S of size k.

One consequence of close-knittedness is that the "perimeter" of S must not be too large relative to its "area." Specifically, let us define the *perimeter* and *area* of any nonempty set of vertices $S \subseteq V$ as follows:

$$peri(S) = d(S, V - S), \ area(S) = d(S, S). \tag{10}$$

Next observe that $d(S) = 2area(S) + peri(S)$. It follows from (9) with $S' = S$ that

$$peri(S) / area(S) \leq (1/r) - 2. \tag{11}$$

It is straightforward to show the following:

Proposition 2. *Given a graph* Γ *and innovation advance* $\alpha > 0$, S *is autonomous for* α *under log linear learning if and only if* S *is r-close-knit for some* $r > 1/(\alpha + 2)$.

Corollary 2.1. If S *is autonomous for* α *under log linear learning, then* $peri(S) / area(S) < \alpha$.

A family of graphs $\mathscr{G}$ is *close-knit* if for *every* $r \in (0, 1/2)$ there exists a positive integer *s(r)* such that, for every $\Gamma \in \mathscr{G}$, every node of Γ is in an *r*-close-knit set of cardinality at most *s(r)* (Young, 1998, Chapter 6).

Corollary 2.2. *Given any close-knit family of graphs* $\mathscr{G}$, *log linear learning is fast for all* $\alpha > 0$.

2.5. Examples

Consider n nodes located around a circle, where each node is linked by an edge to its two immediate neighbors and the edge-weights are one. (This is the situation originally studied by Ellison (1993).) Any set of s consecutive nodes has area $s - 1$ and perimeter 2. It follows that for any $\alpha > 0$, log linear learning is fast.

Next consider a two-dimensional regular lattice (a square grid) in which every vertex has degree 4 (see figure 1). Assume that each edge has weight 1. The shaded region in figure 1 is a subset of nine nodes that is 1/3-close-knit. Hence it is autonomous whenever $\alpha > 1$.

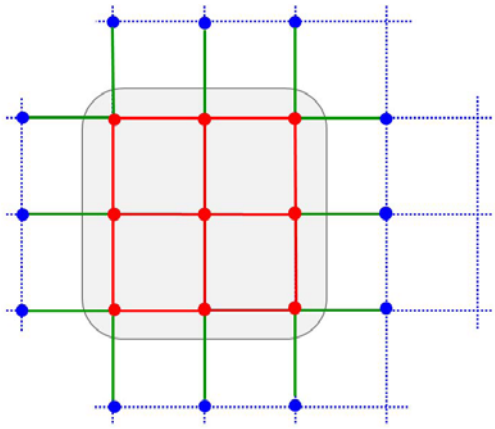

Figure 1. A two-dimensional lattice with a subset of nine vertices that is autonomous for any $\alpha > 1$ under log linear learning.

More generally, observe that any square S of side m has $2m(m-1)$ internal edges and m^2 vertices, each of degree 4, hence

$$d(S,S)/d(S) = 2m(m-1)/4m^2 = 1/2 - 1/2m. \qquad (12)$$

Furthermore it is easily checked that for every nonempty subset $S' \subseteq S$,

$$d(S',S)/d(S') \geq 1/2 - 1/2m. \qquad (13)$$

Therefore every square of side m is $(1/2 - 1/2m)$-close-knit, hence is autonomous for all $\alpha > 2/(m-1)$. A similar argument holds for any regular *d*-dimensional regular lattice: given any $\alpha > 0$ every sufficiently large sublattice is autonomous for α, and this holds independently of the number of vertices in the full lattice.[7]

Notice that in these examples fast learning does not arise because neighbors of neighbors tend to be neighbors of one another. In fact, a *d*-dimensional lattice has the property that *none* of the neighbors of a given agent are themselves neighbors. Rather, fast learning arises from a basic fact of euclidean geometry: the ratio of "surface" to "volume" of a *d*-dimensional cube goes to zero as the cube becomes arbitrarily large.

A *d*-dimensional lattice illustrates the concept of autonomy in a very transparent way, but it applies in many other situations as well. Indeed one could argue that many real-world networks are composed of relatively small autonomous groups, either because people tend to cluster geographically, or because they tend to interact with people of their own kind (homophily), or for both reasons.

[7] Montanari and Saberi (2010, Proposition 5) establish a similar result for this class of networks but with a different notion of waiting time.

To understand the difference between a network with small autonomous groups and one without, consider the pair of networks in Figure 2. The top panel shows a tree in which every node other than the end-nodes has degree 4, and there is a "hub" (not shown) that is connected to all the end-nodes. The bottom panel shows a graph with a similar overall structure in which every node other than the hub has degree 4; however, in this case everyone (except the hub) is contained in a clique of size 4. In both networks all edges are assumed to have weight 1.

Suppose that we begin in the all-*B* state in both networks, that agents use log linear learning with $\beta = 1$, and that the size of the advance is $\alpha > 2/3$. Let each network have *n* vertices. It can be shown that the waiting time to reach at least 99% A (and stay there with probability at least 99%) is unbounded in *n* for the network in the top panel, whereas it is bounded independently of *n* for the network in the bottom panel. In fact, simulations show that it takes less than 25 periods (on average) for A to penetrate to the 99% level independently of *n*. The key difference between the two situations is that the innovation can establish a toehold in the cliques relatively quickly, which then causes the hub to switch to the innovation also.

Note, however, that fast learning in the network with cliques does *not* follow from Proposition 2, because not every node is contained in a clique. In particular, the hub is connected to all of the leaves, the number of which grows with the size of the tree, so it is not in an *r*-close-knit set of bounded size for any given $r < 1/2$. Nevertheless learning is fast: any given clique adopts *A* with high probability in bounded time, hence a sizable proportion of the cliques linked to the hub switch

to A in bounded time, and then the hub switches to A also. In other words, fast learning occurs through a combination of autonomy and contagion, a topic that we explore in detail in the next section.

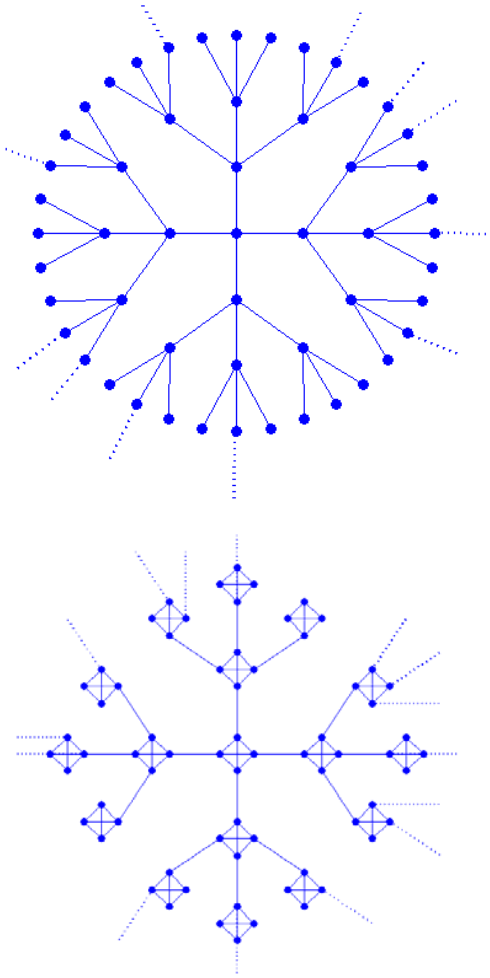

Figure 2. Two networks of degree 4, except for a hub (not shown) that is connected to every end-node (dashed lines). All edge-weights equal 1.

3. Autonomy and contagion

Contagion expresses the idea that once an innovation has become established for some core group, it spreads throughout the network via the best reply dynamic. Morris (2000) was the first to study the properties of contagion in the setting of local interaction games, and to formulate graph-theoretic conditions under which contagion causes the innovation to spread throughout the network. While contagion by itself may not guarantee fast learning in a stochastic environment, a combination of autonomy and contagion does suffice. The idea is that autonomy allows the innovation to gain a foothold somewhere in a reasonably short period of time, after which contagion assures that the process will spread through the rest of society fairly quickly.

Consider a subset *S* of nodes, all of which are playing *A*, and choose some $i \notin S$. Let α be the size of the advance of *A* relative to *B*. Then *A* is a strict best response by *i* provided that

$$(1+\alpha)\sum_{j\in S} w_{ij} > \sum_{j\notin S} w_{ij} \ . \tag{14}$$

Letting $r = 1/(\alpha+2)$, we can write this as follows

$$\sum_{j\in S} w_{ij} > r\sum_{j\neq i} w_{ij} = rd_i \ . \tag{15}$$

Recall that for any vertex *i* and subset of vertices *S*, $d(i,S) = \sum_{j\in S, j\neq i} w_{ij}$ is the total weight on the edges linking *i* to a member of *S*. Given a graph $\Gamma = (V,W)$, a real number $r \in (0,1/2)$, and a subset of vertices *S*, define the *first r-orbit* of *S* as follows

$$O_r^1(S) = S \cup \{i \notin S : d(i,S) > rd_i\}. \qquad (16)$$

Similarly, for each integer $k > 1$ recursively define the k^{th} *r-orbit* by

$$O_r^k(S) = O_r^{k-1} \cup \{i \notin O_r^{k-1} : d(i,O_r^{k-1}) > rd_i\}. \qquad (17)$$

Now suppose that $r = 1/(\alpha + 2)$. Suppose also that everyone in the set S is playing A. If the learning process is deterministic best response dynamics, and if everyone updates once per time period, then after k periods everyone in the k^{th} r-orbit of S will be playing A. Of course, this does not show that log linear learning with asynchronous updating will produce the same result. The key difficulty is that the core set S of A-players might unravel before contagion converts the other players to A. However, this problem can be avoided if: *i)* the core set S reaches the all-A state within a bounded period of time; *ii) S* is autonomous and hence its members *continue to play A with high probability.* These conditions are satisfied if the core set is the union of autonomous sets of bounded size. We therefore have the following result.

Proposition 3. *Let $\mathcal{G}$ be a family of graphs of bounded degree. Suppose that there exist positive integers s, k and a real number $r \in (0,1/2)$ such that, for every $\Gamma = (V,W) \in \mathcal{G}$, there is a subset of vertices S such that: i) S is the union of r-close-knit sets of size at most s; and ii) $O_r^k(S) = V$. Then log linear learning is fast on $\mathcal{G}$ whenever $\alpha > (1/r) - 2$.*

We illustrate with a simple example. Let the network consist of a circle of n agents (the rim) plus a central agent (the hub). Each agent on the rim is adjacent to the hub and to its two immediate neighbors on the rim. Note that the hub is

not contained in an r-close-knit set of bounded size for any $r < ½$. However, for every $r < ½$, the hub is in the first r-orbit of the rim. Moreover, for every $r < 1/3$, there is an r-close-knit set of bounded size that consists of rim nodes; namely, choose any sequence of k adjacent rim nodes where $k > 1/(1 - 3r)$. It follows from Proposition 3 that learning is fast for this family of graphs whenever $\alpha > 1$.

Fast learning says that the waiting time is bounded for a given family of networks, but it does not specify the size of the bound concretely. Actual examples show that the waiting time can be surprisingly short in an absolute sense. Consider an innovation with advance $\alpha = 1$, and suppose that all agents use log linear learning with $\beta = 1$. Figure 3 shows the expected waiting time to reach the 99% penetration level for two families of networks: circles where agents are adjacent to their nearest four neighbors, and two-dimensional lattices. (Thus in both cases the networks are regular of degree 4.) The expected waiting time is less than 25 periods in both situations. In other words, almost everyone will be playing A after just 25 revision opportunities per individual.

Notice that this waiting time is substantially shorter than it takes for a given individual to switch to A when his neighbors are playing B. Indeed, the probability of such a switch is $e^0/(e^0 + e^{4\beta}) \approx e^{-4\beta} \approx 0.018$. Hence, in expectation, it takes about $1/.018 = 54$ periods for any given agent to adopt A when none of his neighbors has adopted. Yet it takes only about half as much time for nearly everyone to adopt. The reason, of course, is that the process is speeded up by contagion. The rate at which the innovation spreads results from a combination of autonomy and contagion.

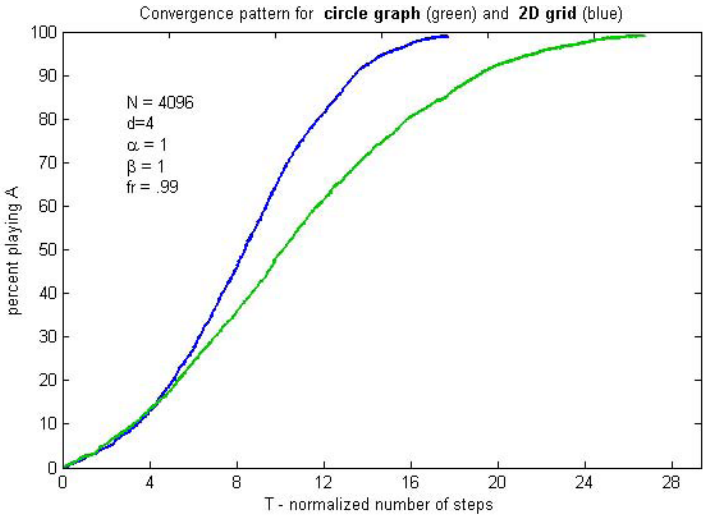

Figure 3. Simulated waiting times to reach 99% *A* starting from all-*B* for circles and two-dimensional lattices. Time periods represent the expected number of updates per individual.

4. Bursts of innovation

We have seen that the speed with which innovations spread in a social network depends crucially on the interaction between three features: the size of the advance α, the degree of rationality β, and the existence of autonomous groups that allow the innovation to gain a secure foothold. The greater the advance from the innovation relative to the status quo, the more rapidly it spreads for any given topology, and the more that people are clustered in small autonomous groups the more rapidly the innovation spreads for any given size of advance. Furthermore the degree of rationality must be at an intermediate level for the rate of spread to be reasonably fast: if β is too high it will take a very long time before anyone even experiments with the innovation, whereas if β is too low

there will be so much random behavior that even approximate convergence will not take place.

In this section we examine the interaction between the network topology and the size of advance in more detail, and show that it leads to an important qualitative prediction about the spread of social innovations. Namely, they tend to occur in large bursts rather than through small incremental improvements. The reason is that a small improvement takes a much longer time to gain an initial foothold than does an innovation that results in a major gain.

To illustrate this point concretely, fix a network Γ and assume for simplicity that all edge-weights equal 1. Consider an individual i with d_i neighbors. The probability that i tries A, given that none of his neighbors is currently using A, equals $1/(1+e^{\beta d_i}) \approx e^{-\beta d_i}$. (Notice that this expression does not depend on the size of the advance α.) This highlights the complementary nature of β and d_i: together they measure the inertia that results from increasing returns and rationality. The more neighbors an agent has, the more irrational it is to try out something new, because the expected loss in payoff is very large.[8]

Now consider one of i's neighbors, say j, whose degree is d_j. If i was the sole adopter in period t, then j's probability of adopting in period $t + 1$ is approximately equal to

[8] One could argue, of course, that a forward-looking rational agent might deliberately try out an innovation in anticipation of future gains in the event that his neighbors adopt later on. This leads to a more complex model that we shall not investigate here. For related work on this issue see Ellison (1997).

$$(1-1/e)(1/e)e^{-\beta[(d_j-2-\alpha)]_+} \text{ if } \alpha < d_j - 2 \text{ and } \beta \text{ is large.} \qquad (18)$$

This follows from the assumption that each agent updates according to a Poisson process with expectation one: the probability is *(1 - 1/e)* that *i* does not update in period *t + 1* and that *j* does update. Note, however, that if $\alpha > d_j - 2$, the probability is fairly high that *j* will adopt – indeed the probability is close to one when β is large.[9]

Thus the probability that any given edge $\{i, j\}$ becomes coordinated on *A* within two periods is $O(e^{-\beta[(d_j-2-\alpha]_+}e^{-\beta d_i})$. Furthermore the set $\{i, j\}$ is autonomous if α is sufficiently large, specifically, it suffices that $\alpha + 2 > d_i + d_j$. Assume now that Γ contains no isolated vertices and that the degrees are bounded above by some positive integer *d*. The preceding shows that if $\alpha > 2d - 2 \geq d_i + d_j - 2$ then every edge is autonomous. One can extend this line of argument to show that a tree T consisting of *m* vertices and *m-1* edges is autonomous provided that $\alpha > (1-1/m)d - 2$.[10] Now suppose that $\alpha > d - 2$. If the network is connected and n is sufficiently large, then every vertex is in an autonomous set, namely a tree with a sufficiently large number of vertices. Hence we can apply Proposition 2 to conclude the following.[11]

---

[9] Given that exactly one neighbor of *j* plays *A*, *j* adopts *A* with probability $e^{\beta(1+a)}/[e^{\beta(1+a)} + e^{\beta(d_j-1)}]$, which is approximately equal to $e^{-\beta[(d_j-2-\alpha]}$ when $\alpha < d_j - 2$, and is *O(1)* when $\alpha > d_j - 2$.

[10] A tree *T* with *m* vertices necessarily has $m - 1$ edges. *T* is autonomous for α if the all-*A* state on *T* has higher potential than the all-*B* state on *T* conditional on the vertices outside of *T* being held fixed at *B*. This holds if $(1+\alpha)(m-1) > \sum_{i \in T} d_i - (m-1) \geq md - m + 1$, which implies that $\alpha > (1-1/m)d - 2$.

[11] I am indebted to Gabriel Kreindler for suggesting this formulation of the result.

Proposition 4. *Let $\mathscr{G}$ be a family of networks with no isolated vertices and bounded degree $d > 0$. If $\alpha > d - 2$, then learning is fast. In particular, given any penetration level $p < 1$ and any sufficiently large β, the expected waiting time $T_{p,\beta}$ to reach p is bounded above independently of the number of agents.*

We now illustrate how the speed of learning varies with α in a concrete case. Let $\mathscr{G}$ be the family of two-dimensional lattices where each agent has degree four. Figure 4 shows the simulated waiting time to reach the target $p = .99$ as a function of α. For small values of α the waiting time grows exponentially and is many thousands of periods long, whereas for large values of α (e.g. a > 1) the waiting time is less than 20 periods.

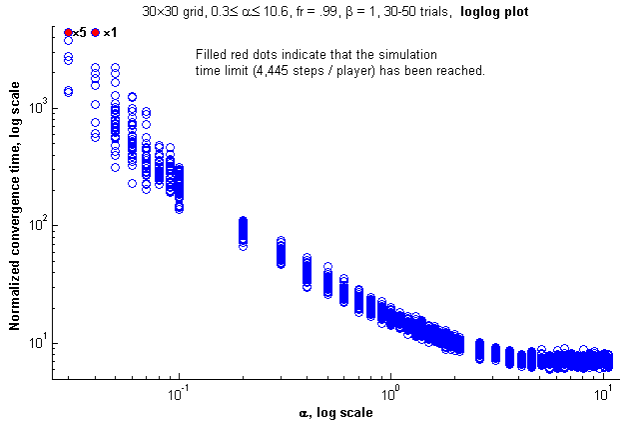

Figure 4. Expected number of periods to reach 99% playing *A* as a function of the size of advance α *(β= 1)*. Estimated from 10,000 simulations starting from the all-*B* state.

To understand the implications of this relationship, suppose that each successive innovation leads to an advance of size α. Let $T(\alpha)$ be the expected waiting time for such an innovation to spread (for a given level of penetration p and level of

rationality β). Assume that each new advance starts as soon as the previous one has reached the target *p*. Then the average rate of advance per period is $r(\alpha)$, where $(1+r(\alpha))^{T(\alpha)} = 1+\alpha$, that is,

$$r(\alpha) = (\ln(1+\alpha)/T(\alpha)) - 1. \qquad (19)$$

The inverse function *f(r)* mapping r to α is shown in Figure 5 for a two-dimensional lattice and log linear learning with $\beta = 1$.

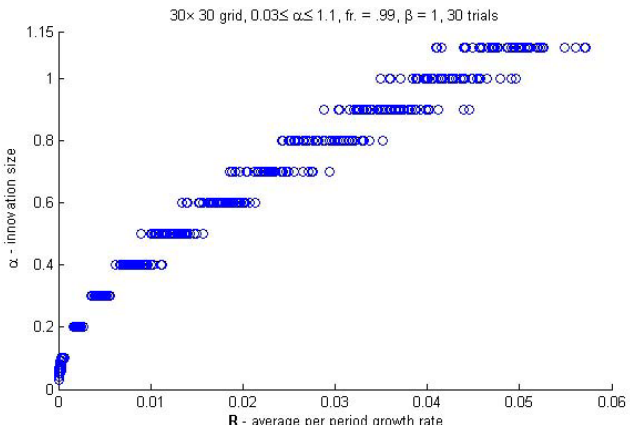

Figure 5. The level of innovative advance $\alpha = f(r)$ required to achieve an average growth rate of r per period. Log linear learning with $\beta=1$ on a 30 x 30 two-dimensional lattice.

The figure shows that to achieve a 1% average growth rate per period requires innovative bursts of size at least 40%. A 3% growth rate requires innovative

bursts of at least 80%, and so forth.[12] Of course, these numbers depend on the topological properties of the grid and on our assumption that agents update using a log linear model. Different network structures and different learning rules may yield different results.

Nevertheless, this example illustrates a general phenomenon that we conjecture holds across a range of situations. Institutional change that involves a series of small step-by-step advances may take a very long time to spread as compared to a change of comparable magnitude that occurs all at once. The basic reason is that it takes much longer for a small advance to gain a secure foothold in an autonomous group: the group must be quite large and/or it must be quite interconnected to prevent the small advance from unraveling. Furthermore, under a small advance there are fewer pathways through that allow contagion to complete the diffusion process. The point is that social innovations are technologies that facilitate – and require -- coordination with others to be successful. It is this feature that makes social change so difficult, and that favors large advances when change does occur.

References

[12] The graph assumes innovative advances of constant size α. However, the concave nature of the function *f(r)* mapping *r* to α implies that to achieve a given rate of growth *r* requires at least some bursts of size greater than or equal to *f(r).*

Arrow, Kenneth J. (1962), "The economic implications of learning by doing," *Review of Economic Studies*, 29: 155-173.

Arthur, W. Brian (1989), "Competing technologies, increasing returns, and lock-in by historical events," *Economic Journal* 99: 116-131.

Blume, Lawrence E. (1993), "The statistical mechanics of strategic interaction," *Games and Economic Behavior*, 4: 387-424.

Blume, Lawrence E. (1995), "The statistical mechanics of best-response strategy revision," *Games and Economic Behavior*, 11: 111-145.

Blume, Lawrence E. and Steven N. Durlauf (2000), "The interactions-based approach to socio-economic behavior," in *Social Dynamics*, edited by Steven N. Durlauf and H. Peyton Young, Washington DC: The Brookings Institution.

Coleman, James S., Elihu Katz, and Herbert Menzel (1957), *Medical Innovation: A Diffusion Study*. New York: Bobbs Merrill.

Ellison, Glenn (1993), "Learning, local interaction, and coordination," *Econometrica*, 61: 1046-1071.

Ellison, Glenn (1997), "Learning from personal experience: one rational guy and the justification of myopia," *Games and Economic Behavior* 19:180-210.

Dean P. Foster and H. Peyton Young (1990,) "Stochastic evolutionary game dynamics," *Theoretical Population Biology* 38: 219-232.

Goyal, Sanjeev (1996), "Interaction structure and social change," *Journal of Institutional and Theoretical Economics*, 152: 472-495.

Kandori, Michihiro, George Mailath, and Rafael Rob (1993), "Learning, mutation, and long run equilibria in games," *Econometrica* 61: 29-56.

Katz, Michael L., and Carl Shapiro (1985), "Network externalities, competition, and compatibility," *American Economic Review* 75: 424-440.

McFadden, Daniel (1973), "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, edited by P. Zarembka, New York: Academic Press.

Montanari, Andrea, and Amin Saber (2010), "The spread of innovations in social networks," *Proceedings of the National Academy of Sciences of the USA*, :20196-20201

Morris, Stephen (2000), "Contagion," *Review of Economic Studies*, 67: 57-78.

North, Douglass (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge UK: Cambridge University Press.

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

Young, H. Peyton (1993), "The evolution of conventions," *Econometrica*, 61: 57-84.

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

Young, H. Peyton (2006), "The diffusion of innovations in social networks," in *The Economy as a Complex Evolving System III*, edited by Lawrence E. Blume and Steven N. Durlauf, Oxford University Press.