Contagion and coordination in random networks*

Dunia López-Pintado†

September 9, 2005

Abstract

We study the problem of spreading a particular behavior among agents located in a random social network. In each period of time, neighboring agents interact strategically playing a 2×2 coordination game. Assuming a myopic-best response dynamics, we show that there exists a threshold for the degree of risk dominance of an action such that below that threshold, contagion of the action occurs. This threshold depends on the connectivity distribution of the network. Based on this, we show that the well-known (and extremely popular in epidemiology) scale-free networks do not properly support this type of contagion, which is better accomplished by a more intermediate variance network. (***JEL Numbers***) C73, O31, O33, L14.

Keywords: contagion, coordination games, scale-free networks, mean-field theory.

*This paper is based on part of my Ph.D. thesis at the University of Alicante. I am most grateful to Peter Dodds and Fernando Vega-Redondo for valuable comments and suggestions. Thanks are also due to Matt Jackson, Juan Moreno-Ternero, Stephen Morris, Duncan Watts and the audience at the 2005 SAET Conference in Vigo. Financial support from the McDonnell Foundation under the form of a post-doc fellowship is gratefully acknowledged.

†Address for correspondence: Dunia López-Pintado. ISERP, 8th floor, International Affairs Building, 420 West 118th St., Columbia University, Mail Code: 3355. New York, NY 10027 (USA). e-mail: dl2244@columbia.edu. Tlf: (1) 212 854 0367. Fax: (1) 212 854 7998

1 Introduction

There is a wide range of situations where agents coordinate their decisions and form conventions. Some paradigmatic examples are as diverse as the choice of a language, the adoption of network goods (e.g., fax machines, cell phones, E-mail accounts etc.) or even the selection of which TV show to watch. In all of these examples it is clear that an agent's decision depends on the decisions of the subset of the population with whom he interacts. Therefore, the population constitutes a social network where agents benefit from coordinating with their neighbors.

A question that has drawn a lot of attention in the field of economics has been to determine the properties that lead to the selection of one particular convention. Or by the same token, to understand what makes some conventions more vulnerable than others. This was originally analyzed by Kandori, Mailath and Rob (1993) and Young (1993) in a setting where individuals were randomly matched to play a game every period. The dynamics considered in these studies is a myopic-best response with a small mutation probability. The common result is that the unique stationary stable state is such that the risk dominant equilibrium is played. Building on top of this literature there are several papers that introduce the idea that there is local interaction and thus agents only play with a subset of the population. For instance, the work by Ellison (1993) considers individuals located in a circle and interacting with the individuals to their right and left. Furthermore, Blume (1995), Morris (2000) and Young (2003), among others, have developed techniques to study coordination games in more general network structures.

Following this last approach, this paper analyzes the stability of conventions in a social network. The dynamics is such that, every period, agents play a 2×2 coordination game with each neighbor. Players are assumed to revise using a deterministic myopic-best response. In other words, a player will choose the action that maximizes his current payoff given the proportion of neighbors choosing each action in the previous period. Assuming that there is an incumbent action played by all agents in the population (convention), the aim of this work is to answer the following question: under which conditions a new action that is initially played by a small seed can take off and spread to a significant fraction of the population? Whenever this property holds we say that there is *contagion*.

One feature that distinguishes our approach from previous work in economics is the way the social network is modelled. We study *random networks*, that is, networks that have been generated by a random process of link formation. This framework provides a natural setting for the analysis of complex networks since it allows for intricate structures with significant heterogeneity among its agents to arise. Moreover, random networks represent a tractable benchmark case which can be studied through the statistical analysis of some large-scale regularities. One fundamental property that characterizes a random network in our setting is the *connectivity distribution* where the connectivity of an agent refers to the number of direct connections he has. Two additional properties have obtained a great deal of attention in the literature of networks. First, the *clustering coefficient* which describes with which probability two neighbors of an agent are also neighbors of each other. Second, the *average path length* which describes the average geodesic distance across pairs of nodes in the network. It is well known that random networks have low clustering coefficient (precisely because of the randomness of the link formation process) and short average path lengths. Although low clustering is crucial for our results to hold, many real-world networks do not satisfy this property, therefore, this analysis can only be considered as a starting point useful for guiding further research.[1]

Although our results apply to any random network, for most of our discussions we focus on the comparison between: *homogeneous*, *exponential* and *scale-free* networks. An homogeneous network is such that all nodes have approximately the same connectivity. In an exponential network the upper tail of the connectivity distribution decreases exponentially and thus, despite there is some variance, the average connectivity is representative of the connectivity of most nodes in the network. Finally, a scale-free network has a power-law connectivity distribution of the form $P(k) \sim k^{-\gamma}$ where γ ranges between 2 and 3 which implies that each node has a statistically significant probability of having a very large number of connections compared to the average connectivity. Thus, there is no representative connectivity in the network which is the reason why they are called scale-free. The interest in the study of scale-free networks is enhanced by the empirical evidence that many paradigmatic examples of complex networks such as the WWW, the Internet and the human sexual contact network, among others, are characterized by scale-free connectivity properties (see e.g., Barabási et al.; 2000, Faloutsos et al.; 1999 and Lijeros et al.; 2001).

[1] The empirical literature on complex networks shows that most real-world networks are "small worlds", i.e., exhibit high clustering and short average path lengths.

In this paper we characterize, in terms of the degree of risk dominance, when there is contagion of an action in the population, and state how this depends on the connectivity distribution of the network. We study the implications of this general result to homogeneous, exponential and scale-free networks and find that contagion is *easier* in exponential than in scale-free or homogeneous networks. This would indicate that, the well-known fact in epidemiology models (i.e., that scale-free networks are optimal for contagion purposes) does not hold when instead of a physical contagion, of for example an infectious disease, we are concerned with the contagion of a social-economic behavior.

The results obtained in this paper are derived using a mean-field approximation of the dynamics. This approach, standard in statistical physics, has been used in a wide range of problems dealing with complex systems since it provides a good approximation of the qualitative features of stochastic dynamics.

The rest of the paper is organized as follows. Section 2 contains the model. The mean-field theory is described in Section 3, whereas in Section 4 the results of the paper are presented. In Section 5 some simulations of the original dynamics are provided as a test for the mean-field approximations. Section 6 concludes with a discussion and some final remarks.

2 The model

2.1 Social network

Consider a finite population where individuals interact with each other and form a social network. Let $\Gamma \equiv (N, L)$ represent the network where $N = \{1, 2, ..., i, ..., n\}$ is the set of individuals in the population (i.e., the nodes) and $L \subseteq N \times N$ is the set of pairwise interactions among them (i.e., the links). Networks can be directed or undirected. Here, to simplify matters, we consider undirected networks. This implies that if $(i, j) \in L$ then $(j, i) \in L$. We rule out the existence of reflexive links, that is, $(i, i) \notin L$ for any given $i \in N$.

Let N_i be the neighborhood of i, i.e., the set of individuals with whom i is directly linked. Formally,

$$N_i = \{j \in N, \text{ s.t. } (i, j) \in L\}.$$

In addition, let $k_i \equiv |N_i|$ be the number of neighbors of i, often referred as his connectivity (or degree). The connectivity can differ across individuals in the population. Therefore, one of the key characteristics of any network is the connectivity distribution. For any network Γ, the connectivity distribution $P(k)$ displays for each $k = 0, 1, ..., n-1$ the fraction of nodes with connectivity k. Specifically,

$$P(k) = \frac{1}{n} \left| \{ i \in N \text{ s.t. } k_i = k \} \right| .$$

We now introduce three specific examples of connectivity distributions which will play a prominent role in our discussions. An homogeneous network is such that all nodes have the same connectivity, say $\overline{k}$ (i.e., $P(\overline{k}) = 1$). An exponential network is such that the connectivity distribution is an exponential function (i.e., $P(k) \sim e^{-k}$) and thus although there is some variance in the connectivity of nodes, the average connectivity is representative of the connectivity of most individuals in the population. A scale-free network exhibits a power-law connectivity distribution (i.e., $P(k) \sim k^{-\gamma}$, where γ ranges between 2 and 3) which implies that there is a large variance in the connectivity of nodes. In fact, the variance of a scale-free connectivity distribution tends to infinity as the size of the population increases.

We assume that the network Γ is a realization of a *random network*. A random network is a random variable characterized by the following two items: (i) a family of possible networks Π and (ii) a probability distribution $\widetilde{P}(\Gamma)$ (with support Π) indicating the ex-ante probability with which each network $\Gamma \in \Pi$ is selected. Throughout the paper we assume random networks characterized by a connectivity distribution. To be more precise, given a connectivity distribution $P(k)$, let the set Π contain all the networks with connectivity distribution $P(k)$ and let $\widetilde{P}(\Gamma)$ be a probability distribution which selects uniformly at random anyone of the networks in Π.[2]

Finally, we assume that the network has a *giant component* which means that there exists a unique connected subgraph of the network whose relative size remains bounded above zero as the number of nodes tends to infinity. A sufficient condition for this to hold in random networks with arbitrary connectivity distribution is $\langle k \rangle > 2$, where $\langle k \rangle$ denotes the average connectivity in the network.[3]

[2]It is generally difficult to define mechanisms for constructing random networks with arbitrary connectivity distributions. However, the so-called configuration model proposed by Bender and Cannfield (1978) is a stochastic process of network formation which provides a good approximation of a random network.

[3]The condition for the existence of a giant component in random networks, first derived by Mollow and Reed (1995), is $\langle k^2 \rangle - 2\langle k \rangle > 0$ where $\langle k^2 \rangle$ is the second order moment of $P(k)$.

2.2 Dynamics

We consider a dynamics that describes the evolution of players' choices through time. At time t, each agent plays a 2×2 game with each neighbor and chooses an action from the space $S = \{0, 1\}$. It is our assumption that an agent cannot make contingent his action on his neighbor. This assumption is natural in this context since otherwise the behavior of an agent would be independent of the network structure.

Payoffs from each interaction in each period are given by a function $\pi(s, s')$ where $s, s' \in S$, and they are summarized in the following symmetric matrix

$i \backslash j$	1	0
1	d	e
0	f	b

(1)

We assume that $d > f$ and $b > e$. This implies that the game is a coordination game (whose Nash equilibria are $(0, 0)$ and $(1, 1)$).

Player i's payoff from playing $s_i \in \{0, 1\}$ when the strategy profile of the remaining players is s_{-i} is given by

$$\Pi_i(s_i, s_{-i}) = \sum_{j \in N_i} \pi(s_i, s_j)$$

Thus, an individual's payoff is simply the sum of the payoffs obtained across the bilateral games in which he is involved.

The dynamics can then be described as follows. At each time t, with a certain probability a player, say i, revises his strategy and chooses a myopic-best response. This implies that, if at time t the proportion of neighbors choosing 1 is higher than $q \equiv \frac{b-e}{d-f+b-e}$, then i's best response is to choose 1. Otherwise i chooses 0. For simplicity and without loss of generality, let us assume that if the proportion of neighbors choosing 1 equals q, action 0 is chosen. The value q is often called the *degree of risk dominance* of action 1. Notice that q determines a lower bound for the fraction of individuals that must be choosing 1 in order to make action 1 preferred to action 0. If $q \leq 1/2$ action 1 is risk dominant.[4] Also, the higher the degree of risk dominance of action 1 the lower the value of q.

[4] An action is risk dominant in a 2×2 game if it is a best response to a mixed strategy played by the opponent that assigns equal probability to both actions.

Our aim in this paper is to obtain the conditions under which a small seed of agents adopting a certain action can spread to a significant fraction of the population. We also want to determine how this depends on the properties of the social network. To fix ideas, consider that action 0 is the incumbent (or default) action. That is, initially the whole population chooses action 0. Notice that this state, as well as the state where all individuals are choosing action 1, is a stationary state of the dynamics. Now assume that a small seed of individuals switch to action 1. Given the dynamics described above we want to know if the action will spread significantly or if it would eventually disappear.

3 Mean field theory

The study of this stochastic process (an infinite Markov chain) is hard to solve for explicitly. This is why we do a series of approximations. We first approximate the random discrete time dynamics by its continuous counterpart where individuals revise their strategy at a certain rate every period. Furthermore, as will become clear later in the paper, some local variables relevant for the dynamics are substituted by their average values. This approach, denoted by mean-field approach, has been widely used in other areas of science since it describes successfully the main qualitative features of the stochastic dynamics.

Several assumptions make the mean-field approach reasonable. First, the population must be sufficiently large (i.e., $n \to \infty$). Second, the connectivity of linked nodes must be stochastically independent and third, there should be absence of network clustering. The two last assumptions reflect a situation where the induced network displays "no structure", i.e., the characteristics of any given node is unaffected of structural correlations. It is worth emphasizing that these features are approximately true in random networks. Finally, we make the additional assumption that the giant component encompasses all the nodes in the network. In other words, the network is connected, i.e., there is a path connecting any two nodes from the network.

Denote by $\rho_k(t)$ the proportion of agents with k links that are choosing action 1 at time t. The probability that any given link points to an agent choosing 1 is denoted by $\theta(\{\rho_k(t)\}_k)$ and can be calculated as

$$\theta(t) = \frac{1}{\langle k \rangle} \sum_k kP(k)\rho_k(t) \tag{2}$$

where the average connectivity can be computed as

$$\langle k \rangle = \sum_k kP(k).$$

Thus, the probability that an individual with k links has exactly k_1 neighbors choosing 1, follows a binomial distribution with a pdf given by

$$\binom{k}{k_1} \theta^{k_1}(1-\theta)^{(k-k_1)}$$

Note that $\theta(t)$ is used as a mean-field parameter since it is considered the same for all nodes independently on their connectivity or position in the network.

An individual with k neighbors and with k_1 of them choosing 1, chooses action 1 with probability

$$P_q(s=1 \mid k_1, k) = \begin{cases} 1 \text{ if } k_1/k > q \\ 0 \text{ if } k_1/k \leq q \end{cases} . \tag{3}$$

Also denote as $P_q(s=0 \mid k_1, k)$ the probability that an individual with k neighbors and such that k_1 of them are choosing action 1, chooses action 0. It is straightforward to see that $P_q(s=1 \mid k_1, k) + P_q(s=0 \mid k_1, k) = 1$.

Given $\theta(t)$ an agent with connectivity k chooses action 1 at a rate

$$rate(1 \mid k, \theta(t)) = \sum_{k_1=0}^{k} \lambda P_q(s=1 \mid k_1, k) \binom{k}{k_1} \theta(t)^{k_1}(1-\theta(t))^{(k-k_1)}$$

and chooses action 0 at a rate

$$rate(0 \mid k, \theta(t)) = \sum_{k_1=0}^{k} \lambda P_q(s=0 \mid k_1, k) \binom{k}{k_1} \theta(t)^{k_1}(1-\theta(t))^{(k-k_1)}$$

where $\lambda > 0$ is the rate at which an individual revises his strategy.

Then, the law of motion of the system may be written as

$$\frac{d\rho_k(t)}{dt} = -\rho_k(t) rate(0 \mid k, \theta(t)) + (1-\rho_k(t)) rate(1 \mid k, \theta(t)). \tag{4}$$

Roughly speaking, equation (4) says the following: the variation of the relative density of agents with connectivity k choosing 1 at time t, equals the proportion of agents with connectivity k choosing 0 that switch to 1 at time t, minus the proportion of agents with connectivity k choosing 1 that switch to 0 at time t.

After imposing the stationary condition $\frac{d\rho_k(t)}{dt} = 0$ we obtain the equation valid for the behavior of the system at large times,

$$\rho_k(\theta) = \sum_{k_1=0}^{k} P_q(s=1 \mid k_1, k) \binom{k}{k_1} \theta^{k_1}(1-\theta)^{(k-k_1)} \tag{5}$$

Upon replacing equation (5) in equation (2) we obtain that

$$\theta = H_q(\theta) \tag{6}$$

where

$$H_q(\theta) \equiv \frac{1}{\langle k \rangle} \sum_k kP(k) \sum_{k_1=0}^{k} P_q(s=1 \mid k_1, k) \binom{k}{k_1} \theta^{k_1}(1-\theta)^{(k-k_1)}.$$

The solutions of equation (6) are the stationary values of θ. Given θ we can also compute the fraction of players choosing 1 in the stationary states as $\rho = \sum_k P(k)\rho_k(\theta)$.

4 The results

In this section we present the main results of the paper. For concreteness, we define first the concept of *contagion*.

Consider the mean-field dynamics described above and a network with connectivity distribution $P(k)$. Then, we say that there is **contagion** of action 1 if, starting at an initial state with an infinitesimally small fraction of agents choosing 1, the mean-field dynamics converges to a stable state with a positive fraction of agents choosing 1.

Notice that in the previous definition we do not specify what will be the size of the final fraction of agents choosing 1 since this is often hard to compute. The extreme case, however, where the whole population eventually adopts action 1 will be referred as **global contagion**. We will come back to this issue in Section 5 where we determine throug simulations, the number of agents choosing action 1 in the long run of the dynamics for some simple networks.

We now present Theorem 1 followed by some corollaries.

Theorem 1 *Given a random network with connectivity distribution $P(k)$, there exists a threshold for the degree of risk dominance of action 1, $q^* \in [0,1]$, such that contagion occurs if and only if $q < q^*$.*

Proof: Using the mean-field stationary equations derived above we can find the stationary values of θ as implicitly determined by equation (6). Replacing equation (3) in equation (6) we obtain the following fixed point equation,

$$\theta = H_q(\theta) = \frac{1}{\langle k \rangle} \sum_k kP(k) \sum_{k_1=[kq]+1}^{k} \binom{k}{k_1} \theta^{k_1} (1-\theta)^{(k-k_1)}$$

where $[x]$ stands for the highest integer smaller or equal to x.
As illustrated in Figure 1, for any given $q \in [0,1]$ the states corresponding to $\theta = 0$ and $\theta = 1$, i.e., the states where all players are choosing the same action (either 0 or 1), are stationary. Thus, to proof Theorem 1 we must determine which values of q make the state $\theta = 0$ unstable. This is equivalent to computing the values of q satisfying the following condition,

$$H_q'(0) \geq 1$$

where

$$\begin{aligned} H_q'(0) &= \frac{1}{\langle k \rangle} \sum_{k \geq 1} kP(k) \binom{k}{[kq]+1} (0)^{[kq]} \\ &= \frac{1}{\langle k \rangle} \sum_{k \geq 1}^{[1/q]} kP(k) \binom{k}{1} \\ &= \frac{1}{\langle k \rangle} \sum_{k \geq 1}^{[1/q]} k^2 P(k). \end{aligned}$$

We can thus implicitly determine the threshold as follows,

$$q^* = \arg \min_{q \in [0,1]} \frac{1}{\langle k \rangle} \sum_{k \geq 1}^{[1/q]} k^2 P(k) \tag{7}$$

such that

$$\frac{1}{\langle k \rangle} \sum_{k \geq 1}^{[1/q]} k^2 P(k) \geq 1$$

$\square$

Insert Figure 1 about here

The intuition behind the contagion condition is the following. Assume for simplicity that there is only one agent choosing action 1 in the initial state of the dynamics. We define as a **vulnerable node**, a node that prefers action 1 to action 0 even in the case where only one of its neighbors is choosing 1. Formally, given q and the network Γ the set of vulnerable nodes is

$$V(\Gamma) = \{i \in N \text{ s.t. } q < \frac{1}{k_i}\}.$$

Given that the network is random, it can locally be considered as a tree (i.e., a network with no cycles). This in particular implies that the set of vulnerable nodes are crucial for the contagion process to take off, since the chances that a node has two of its neighbors choosing action 1 is very small. Notice that, if the network is a tree this probability is precisely 0. Therefore, at least in the initial stages of the dynamics, the only nodes that switch to action 1 are the vulnerable nodes. Consequently, in order to have contagion the subgraph formed by the vulnerable nodes needs to be relatively large and, more importantly, well connected. To phrase it in a different way, the set of vulnerable nodes must form a giant component. Note that this will crucially depend on $P(k)$ and q as the contagion threshold provided by Theorem 1 indicates.

Several interesting points follow from Theorem 1. Note that action 1 must be risk dominant in order for contagion to occur.[5] Furthermore, if there exists a maximum connectivity in the network, say $k_{\max}$, then, for q sufficiently low (e.g., $q < \frac{1}{K_{\max}}$) there is global contagion. That is, action 1 will eventually be played by the whole population. This observation is trivial since, if $q < \frac{1}{K_{\max}}$ all nodes are vulnerables.

One of the main purposes of this paper is to compare the threshold obtained for different connectivity distributions. More specifically, we would like to consider networks with the same density of links (i.e., average connectivity) but distributed differently (i.e., different variance). To begin addressing this issue let us consider the following simple example.

Assume that the connectivity of individuals can take only two values which distance from the average $a \in [0, \langle k \rangle]$. Specifically, given $a \in [0, \langle k \rangle]$ the connectivity distribution is

$$P_a(k) = \begin{cases} \frac{1}{2} \text{ if } k = \langle k \rangle + a \\ \frac{1}{2} \text{ if } k = \langle k \rangle - a \end{cases}$$

These distributions have all the same average connectivity. Also note that half of the population has connectivity $\langle k \rangle + a$ whereas the other half has connectivity $\langle k \rangle - a$.

[5] This follows from the fact that if action 1 is not risk dominant (i.e., $q > \frac{1}{2}$) a necessary condition for contagion is $\frac{P(1)}{\langle k \rangle} \geq 1$ which never holds for random networks with a giant component.

Using equation (7) we can compute the threshold for different values of a. For instance if $a \leq \langle k \rangle - \sqrt{2\langle k \rangle}$ then $q^*(a) = \frac{1}{\langle k \rangle - a}$ whereas, if $a > \langle k \rangle - \sqrt{2\langle k \rangle}$ then $q^*(a) = \frac{1}{\langle k \rangle + a}$. It is straightforward to show that $q^*(a)$ is maximized when $a^* = \langle k \rangle - \sqrt{2\langle k \rangle}$. Consequently, the contagion threshold reaches its maximum in a network with an intermediate variance.

An intuition for this result is the following. On the one hand, in order to induce contagion the set of nodes with low connectivity ($k = \langle k \rangle - a$) have to be vulnerables, or equivalently $q^*(a) \leq \frac{1}{\langle k \rangle - a}$. Note that, the higher the variance the lower the connectivity of the vulnerable nodes and thus the higher the upper bound for $q^*(a)$. However, as discussed previously, to enhance contagion the set of vulnerables must form a giant component and this can only hold if their connectivities are not too low (i.e., if the connectivity variance is not too high). As the variance increases these two opposite effects take place and this is precisely what makes the contagion threshold non monotonic with respect to the variance.

As the previous example shows there is no trivial relationship between the contagion threshold and the variance of the connectivity distribution. The previous example, however, focuses on a very particular and unrealistic class of networks. We therefore study next if this result extends to more interesting types of networks. In what follows, we compare the contagion threshold of an homogeneous, an exponential and a scale-free network. The standard result in epidemiology is that scale-free networks are more vulnerable to epidemics than exponential and homogeneous networks. Nevertheless, this is not true for the contagion process described in this paper. To show this, let us first determine explicitly the contagion threshold for homogeneous networks.

Corollary 2 *The contagion threshold for a random network with an homogeneous connectivity distribution is* $q^* \simeq \frac{1}{\langle k \rangle}$.

In particular, this corollary implies that q^* is decreasing with respect to the average connectivity. In other words, networks with high density are more resistant to the adoption of a new action.[6] To show this, observe that the connectivity distribution of an homogeneous network is

$$P(k) = \begin{cases} 1 \text{ if } k = \overline{k} \\ 0 \text{ if } k \neq \overline{k} \end{cases} .$$

[6] This, ofcourse, is subject to the condition $\langle k \rangle > 2$ which guarantees the existence of a giant component in the network.

Notice that, in this case, equation (7) can be simplified as follows

$$\frac{1}{\langle k\rangle}\sum_{k\geq m}^{[1/q]} k^2 P(k) = \begin{cases} \overline{k} \text{ if } [1/q] \leq \overline{k} \\ 0 \text{ if } [1/q] > \overline{k} \end{cases}$$

Therefore, the threshold is implicitly determined by the condition

$$[1/q^*] = \overline{k}$$

which implies that $q^* \simeq \frac{1}{\overline{k}}$.

It is obvious that in homogeneous networks whenever the degree of risk dominance of action 1 is below the threshold there is global contagion whereas, if the degree of risk dominance is above the threshold action 1 rapidly vanishes.

Let $m \geq 1$ and consider a scale-free network with connectivity distribution $P_{SF}(k) \propto k^{-2.5}$ for $k \geq m$ and $P_{SF}(k) = 0$ otherwise. Also consider an exponential network with connectivity distribution $P_E(k) \propto e^{-k/2m}$ for $k \geq m$ and $P_E(k) = 0$ otherwise. Finally, consider an homogeneous network such that $P_H(k) = 1$ if $k = 3m$. Note that in all cases the average connectivity equals $3m$. Then, to obtain the condition for contagion we must substitute $P_{SF}(k)$, $P_E(k)$ and $P_H(k)$ in equation (7). We show, using numerical computations, the following corollary.

Corollary 3 *Consider an homogeneous, an exponential and a scale-free random network with connectivity distributions $P_H(k)$, $P_E(k)$ and $P_{SF}(k)$ respectively. Then,*

$$q_H^* \leq q_{SF}^* \leq q_E^*.$$

The results are summarized in Figure 2 where we represent the contagion threshold as a function of the average connectivity for the homogeneous, exponential and scale-free networks introduced above (green, red and blue curve respectively). Notice that, as in the previous example, the network with an intermediate variance is the most successful for contagion. The reason for this again has to do with the existence of a giant component of vulnerable nodes. This result differs from what standard contagion models proposed to describe epidemics have shown. Also notice that the higher the average connectivity of the network the lower the threshold.

Insert Figure 2 about here

5 Simulations

In this section we run simulations to test the validity of the mean-field approximations. We concentrate on the simple example presented above where the connectivity distribution only takes two values. That is,

$$P_a(k) = \begin{cases} \frac{1}{2} \text{ if } k = \langle k \rangle + a \\ \frac{1}{2} \text{ if } k = \langle k \rangle - a \end{cases}$$

We generate four random networks each with $n = 1000$ nodes and an average connectivity of $\langle k \rangle = 20$. The difference among these networks is a, which takes the values $8, 11, 15$ and 17. Notice that the higher a the higher the variance of the connectivity distribution.

We consider the discrete version of the continuous dynamics used to derive the theoretical results. In this respect we assume that in every period one (and only one) agent is chosen to revise his action. Note that the definition of contagion cannot be described in the exact same way as in the infinite population case used for the mean-field approximations. We therefore simply look at the fraction of individuals choosing action 1 in the long run of the dynamics and compare the results for the different networks generated. We fix $q = \frac{1}{10}$, which means that an agent would choose action 1 if and only if the fraction of neighbors choosing 1 is above $\frac{1}{10}$. As illustrated in Figure 3, we find how the fraction of agents choosing 1 changes as a function of the periods of the dynamics for the four networks generated (i.e., $a = 8, 11, 15$, and 17; yellow, red, blue and green curve respectively). The data are the average of 100 simulations. For each simulation the initial condition is such that agents are choosing 1 in period $t = 1$ with probability 0.01. Thus, in the initial period approximately 10 agents are choosing action 1.

Insert Figure 3 about here

Notice that, similarly to what was already shown in the theoretical approach, the diffusion of action 1 is larger for networks with an intermediate variance (i.e., for $a = 11$ and, although significantly less, for $a = 15$). Notice that, all networks except the one where $a = 8$ have 500 vulnerable nodes. This is why there is no diffusion at all in the case $a = 8$ whereas, there is at least some diffusion in all the remaining cases. Furthermore, the network with the highest variance (i.e., $a = 17$) has very little diffusion since the connectivity of the vulnerable nodes is too small for them to form a giant component. We can conclude that the network with largest diffusion is the one satisfying that it is the closest to the homogeneous case (i.e., where $a = 0$), under the restriction that the low connectivity nodes are vulnerables. In this case, this corresponds to the network where $a = 11$.

6 Conclusion and discussion

In this paper we have studied the contagion of a behavior with social-economic relevance in a population triggered by a small initial seed. We have analyzed this phenomenon using a simple binary-decision model where individuals play a coordination game with their neighbors. The interaction structure is modelled as a random network with an arbitrary connectivity distribution. We provide a threshold for the degree of risk dominance of a certain behavior that indicates whether the behavior spreads and becomes persistent or does not spread and vanishes.

We assume a contagion dynamics analogous to the dynamics proposed in previous works in the field of economics (see e.g., Morris; 2000, Young; 2003, among others). The main difference with this literature relies on the types of networks considered. For instance, Morris (2000) assumes a network with a known deterministic shape and a high degree of regularity, whereas we consider networks generated randomly and therefore complex. The main result in Morris (2000) is that the properties of *low neighborhood growth* and *uniformity* favor the contagion of a new behavior. The types of networks we consider are random and thus have by definition a short average path length which in particular implies a high neighborhood growth. Therefore, the previous result does not apply to our networks. Moreover, we concentrate on analyzing how the contagion threshold depends on the connectivity distribution of the network, something that had not been addressed before in this context.

There is also a vast literature in epidemiology and, in recent years, statistical physics that analyzes the phenomenon of the contagion of a disease (or computer virus) in a population (or in the Internet). Some instances of this literature, as for example the paper by Pastor-Satorrás and Vespignani (2001), also model the network as random and use mean-field theory to derive the results. Nevertheless, in social contexts the diffusion of behavior often exhibits features that do not match well those of the epidemic models. In the epidemic diffusion, the transmission of an infection depends on the absolute number of infected neighbors whereas in social contagion there are coordination effects involved, and thus relative numbers are also important. In other words, in order to adopt a behavior the number of non-adopters is also relevant. The present paper incorporates these considerations and obtains results different to the ones described in the literature of epidemiology. For example, Pastor-Satorrás and Vespignani (2001) show that scale-free networks are the best structures for spreading a disease whereas, in our framework, this is better accomplished by networks with lower variance in the connectivity of agents (such as exponential networks).

There are a few other studies about coordination effects on networks that attempt to reconcile the approaches used in economics and epidemiology. For instance, the work by Watts (2002) studies, using percolation theory instead of mean-field, how diffusion waves (unidirectional processes of propagation) advance in a large population.[7] Although in Watts (2002) the behavior of individuals is also characterized by a threshold determining when to adopt or not, he provides no micro-foundation of how these behavioral rules arise. Moreover, Watts (2002) does not focus on the effect that the variances of the networks has on the predictions, which is the main question addressed in our paper. Finally, in López-Pintado (2004) a general family of contagion processes are studied on random networks. However, in contrast to what we do in this paper, the transition rates (from non-adopter to adopter and vice versa) are not symmetric and thus the model proposed in López-Pintado (2004) is closer to the standard susceptible-infected-susceptible models studied in epidemiology.

We would like to conclude by briefly summarizing some additional literature that relates network structure to individual behavior in other contexts. There is a set of studies dealing with markets and networks that explains how the complex structure of bilateral trades and relationships affects the market outcomes and specifically price dispersion (see e.g., Corominas-Bosch; 2005, Kakade et al.; 2005 and Kranton and Minehart; 2001). Other studies, analyze how information propagates through a network. For instance, the work by Calvó-Armengol and Jackson (2004) studies the labor market and proposes a model where agents obtain information about jobs through a network of connections. Other examples are the study of the provision of local public goods (see e.g., Bramoullé and Kranton; 2005) or the study of how network structure influences criminal activity (see e.g., Calvó-Armengol and Zenou; 2004). The fact that there is a wide variety of settings where network structure

[7] In the work by Watts (2002) the analytical results are derived using percolation theory instead of mean-field theory which is appropriate for unidirectional processes of propagation. The advantage of the mean-field approach, however, is that it can be easily extended to account for more general contagion processes (see e.g., López-Pintado and Watts; 2005 and Dodds, López-Pintado and Watts; 2005).

is a crucial determinant of behavior makes it clear that this is an important area for further study. Furthermore, the study of complex networks is a rather interdisciplinary field of research where areas such as biology, physics and social sciences share interests, insights and methodological tools. We think that this feature should be exploited since lessons learned in one field could be useful for another. This paper is an attempt to move in this direction.

References

[1] Barabási A, Albert R, Jeong H (2000) Scale-free characteristics of random networks: the topology of the World-Wide Web. Physica A 281:69- 77

[2] Bender E A, Canfield E R (1978) The asymptotic number of labelled graphs with given degree sequences. Journal of Combinatorial Theory A 24:296-307

[3] Blume L E (1995) The statistical mechanics of best-response strategy revision. Games and Economic Behavior 11: 111-145

[4] Bramoullé Y, Kranton R (2003) Public goods in networks: how social networks can shape social learning and innovation. Mimeo, University of Maryland

[5] Calvó-Armengol A, Jackson M O (2004) The effects of social networks on employment and inequality. American Economic Review 94:426-454

[6] Calvó-Armengol A, Zenou Y (2004) Social networks and crime decisions: the role of social structure in facilitating delinquent behavior. International Economic Review 45:935-954

[7] Coromina-Bosch M (2004) Bargaining in a network of buyers and sellers. Journal of Economic Theory 115:35-77

[8] Dodds P S, López-Pintado D, Watts D J (2005) Social influence and random networks. Mimeo, Columbia University

[9] Ellison G (1993) Learning, local interaction, and coordination. Econometrica 61:1047-1071

[10] Erdos P, Renyi A (1959) On random graphs. Pub. Math. Debrecen 6:290-297

[11] Faloutsos M, Faloutsos P, Faloutsos C (1999) On power-law relationships of the Internet topology. Computer Communications Review 29:251-262

[12] Kandori M, Mailath G J, Rob R (1993) Learning, mutations and long run equilibria in games. Econometrica 61:29-56

[13] Kakade S M, Kearns M, Ortiz L E, Pemantle R, Suri S (2005) Economic properties of social networks. Mimeo, University of Pennsylvania

[14] Kranton R, Minehart D (2001) A thoery of buyers-seller networks. American Economic Review 91:485-508

[15] Lijeros F, Edling C R, Amaral L A N , Stanley H E, Aberg Y (2001) The web of human sexual contacts. Nature 411:907-908

[16] López-Pintado D (2004) Diffusion in complex social networks. Working paper IVIE WP-AD 2004-33, University of Alicante

[17] López-Pintado D, Watts D J (2005) Social influence and collective dynamics. Mimeo, Columbia University

[18] Mollow M, Reed B (1995) A critical point for random graphs with a given degree sequence. Random structures and algorithms 6:161-179

[19] Morris S (2000) Contagion. Review of Economic Studies 67:57-78

[20] Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. Physical Review Letters 86:3200-3203

[21] Young H P (1993) The evolution of conventions. Econometrica 61:57-84

[22] Young H P (2003)The diffusion of innovation in social networks. In Blume L, Durlauf S (Ed), The economy as a complex evolving system, vol. III, Oxford University Press, forthcoming

[23] Watts D J (2002) A simple model of global cascades on random networks. Proceedings of National Academy of Science 99:5766-5771

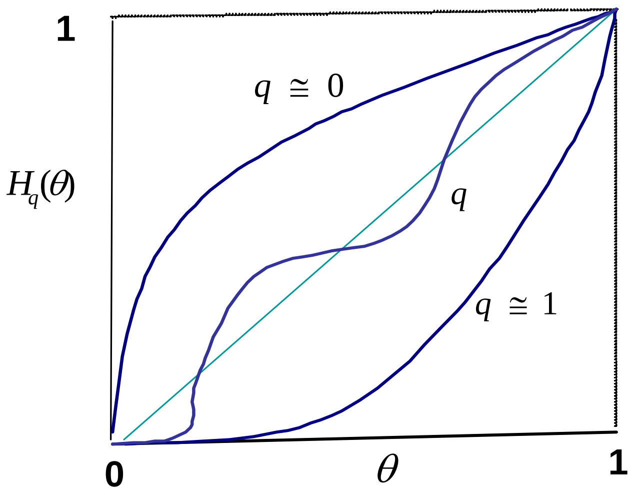

Figure 1: Representation of $H_q(\theta)$ for three values of q. It is illustrated how the shape of $H_q(\theta)$ is concave for low values of q, convex for high values of q, and might be neither concave nor convex for intermediate values of q.

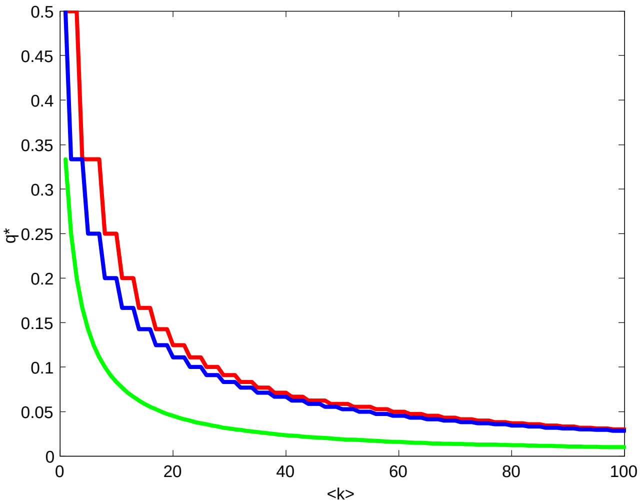

Figure 2: Values of the contagion threshold (q^*, ordinate) for an homogeneous (green), exponential (red) and scale-free (blue) network, as a function of the average connectivity ($\langle k \rangle$, abscissa).

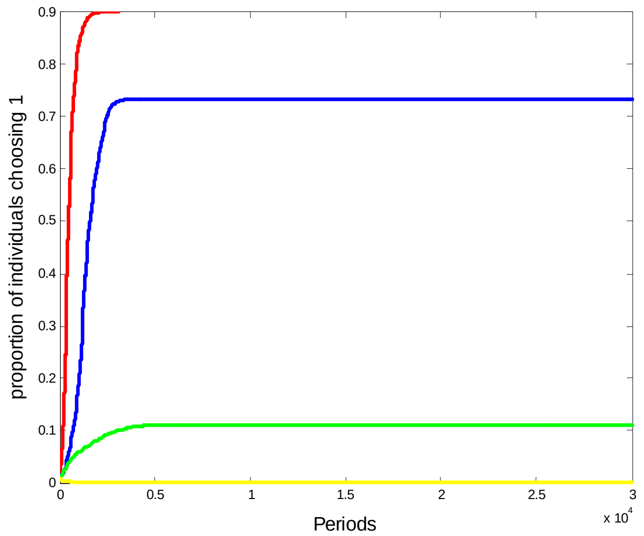

Figure 3: Proportion of individuals choosing 1 (ordinate) as a function of the periods (abscissa) for networks with $a = 8$ (yellow), $a = 11$ (red), $a = 15$ (blue) and $a = 17$ (green).