

Contagion and coordination in random networks

Dunia López-Pintado

Accepted: 27 June 2006 / Published online: 2 August 2006
© Springer-Verlag 2006

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 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 scale-free networks do not always properly support this type of contagion, which is better accomplished by more intermediate variance networks.

JEL Classification C73 · O31 · O33 · L14

Keywords Contagion · Coordination games · Scale-free networks · Mean-field theory

1 Introduction

In this paper we present a simple model in which individuals located in a social network play a 2×2 coordination game with each neighbor. We provide a threshold (the *contagion threshold*) for the degree of risk dominance of a certain action that determines when the action spreads to a significant fraction of the population and becomes persistent. One feature that distinguishes our approach from previous work in economics is the way in which the social network is modeled (see Ellison 1993; Morris 2000; Young 1998, among others).

D. 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

More precisely, in this paper we use mean-field theory to study *random networks* (networks that have been generated by a random process of link formation).¹ A random network is characterized by its *connectivity distribution*, where the connectivity of an agent refers to the number of links he has. We obtain the contagion threshold for random networks with different connectivity distributions and show that, unlike standard epidemiology models, networks with intermediate variance in the connectivity distribution can be optimal for this diffusion process.

This paper is closely related to Morris (2000). The contagion process in the two papers is the same and their major focus, namely the characterization of the contagion threshold, is also similar. However, the generalizability of Morris's results only allows us to compute precisely the contagion threshold in networks with some recurrent pattern. Therefore, the main advantage of this work with respect to Morris (2000) is that here knowledge of the connectivity distribution of the network is enough to compute the exact value of the contagion threshold for random networks, provided we use the mean-field approach. Some of the results found in Morris (2000) also hold in our model. For instance, an action can only spread if it is risk dominant. Moreover, in an homogeneous network case (i.e. where all nodes have the same connectivity) the threshold equals the inverse of the connectivity. Another related work is Watts (2002) that analyzes, using percolation theory instead of mean-field theory, how diffusion waves (unidirectional processes of propagation) advance in a large population. Also, Jackson and Yariv (2006) propose a similar model and analyze, through a slightly stronger mean-field approximation of the dynamics, the size of the fraction of initial adopters required to obtain diffusion.

2 The model

Consider a finite population of individuals $N = \{1, 2, \dots, i, \dots, n\}$ that interact with each other to form a social network $\Gamma = (N, L)$, where $(i, j) \in L$ means that i and j are linked in the social network. We consider undirected networks i.e. $(i, j) \in L$ if and only if $(j, i) \in L$. 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 $k_i = |N_i|$ be the number of neighbors of i , often referred as his connectivity (or degree). The connectivity can differ across individuals in the population. The connectivity distribution $P(k)$ displays for each $k = 0, 1, \dots, n - 1$ the fraction of nodes with connectivity k . More precisely, $P(k) = \frac{1}{n} |\{i \in N \text{ s.t. } k_i = k\}|$. Denote by Π to the set of networks with connectivity distribution $P(k)$. A random network Γ characterized by $P(k)$ is simply a realization of a random variable that selects uniformly at random one of the networks in Π .

¹ The mean-field approach is a standard tool in statistical physics (see e.g., Goldenfeld 1992; Pastor-Satorrás and Vespignani 2001, among others) which is recently being applied in economic studies on networks (see e.g. Vega-Redondo 2006; Jackson 2006 for updated and detailed reviews).

We consider the following dynamics to describe 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 $S = \{0, 1\}$. The assumption that an agent cannot make his action contingent on his neighbor's action is natural in this context. 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:

j	1	0
i	1	0
	d	e
	f	b

(1)

We assume that $d > f$ and $b > e$. This implies that the game is a coordination game [whose strict 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 all the bilateral games in which he is involved.

With a certain probability agents are chosen each period to revise their strategy. If agent i is chosen, then he selects the action that maximizes his benefits given the action of others in the previous period (a myopic best response). Therefore, if at time t the proportion of his neighbors choosing 1 is higher than $q = \frac{b-e}{d-f+b-e}$, then i 's best response is to choose 1. Otherwise i chooses 0. Let us also assume that if the proportion of neighbors choosing 1 equals q , action 0 is chosen. The value q , namely the *degree of risk dominance* of action 1, specifies 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.² Also, the more risk dominant action 1 is the lower the value of q .

The dynamics outlined above defines a Markov process over the set of possible states S^n . In what follows, we assume that action 0 is the incumbent (or default) action. Our aim is to obtain the conditions under which a small seed of agents adopting action 1 can spread to a significant fraction of the population. We want to study how this depends on the properties of the social network.

3 Mean-field theory

We consider the following approximations which allow us to derive analytical results. First, the stochastic dynamics is substituted by deterministic dynamics in continuous time. This approximation is appropriate when dealing with large populations as described by Benaim and Weibull (2003) who show that if the

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

deterministic population flow remains forever in some subset of the state space, then the stochastic process will remain in the same subset space for a very long time with a probability arbitrarily close to one, provided that the population is large enough. Thus, hereafter we assume that the population is infinite and therefore, the network is characterized by a connectivity distribution $P(k)$ with an infinite support, that is, where $k = \{1, 2, \dots\}$. Second, we assume no spatial correlation of the set of adopters across time, namely, an *homogeneous mixing hypothesis*. To put it differently, the model that we analyze with the mean-field equations is analogous to one where the random network is generated every period, although the connectivity of each individual remains constant. We believe that the qualitative results of this alternative model coincide with the results of the original model where the network is fixed throughout the dynamics but has been generated by a random process. The main reason why random networks are suitable for these types of mean-field approximations is that the characteristics of any given node in random networks is unaffected by structural correlations.

Now consider the following notation. Let $\rho_k(t)$ be the proportion of agents with k links that are choosing action 1 at time t . Notice that $\frac{kP(k)}{\langle k \rangle}$ is the probability that a link points to a node with connectivity k . Thus, the probability that any given link points to a an agent choosing 1, denoted by $\theta(t)$, can be calculated as

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

where the average connectivity is $\langle k \rangle = \sum_{k \geq 1} kP(k)$. Therefore, the probability that an individual with k links has exactly k_1 neighbors choosing 1 equals $\binom{k}{k_1} \theta^{k_1} (1 - \theta)^{(k-k_1)}$. Note that $\theta(t)$ is a mean-field parameter since it is considered the same for all nodes independently of their connectivity or position in the network. This simplification, a consequence of the homogenous mixing hypothesis, makes the analysis tractable. Then, an individual with k neighbors and with k_1 of them choosing 1 adopts action 1 with a probability denoted by $P_q(1 | k_1, k)$ and action 0 with probability $1 - P_q(1 | k_1, k)$. Given that our model assumes that individuals use a deterministic myopic best response these probabilities are degenerated. Specifically, $P_q(1 | k_1, k) = 1$ if $k_1/k > q$ and $P_q(1 | k_1, k) = 0$ otherwise.

Let $\lambda > 0$ be the rate at which an individual revises his action. If this individual has connectivity k , he chooses action 1 at an overall rate

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

where the rate of choosing 0 is denoted by $rate(0 | k, \theta(t))$ and equals $\lambda - rate(1 | k, \theta(t))$.

Hence, for each $k \geq 1$ the dynamical mean-field equation may be written as

$$\frac{d\rho_k(t)}{dt} = -\rho_k(t)rate(0 | k, \theta(t)) + (1 - \rho_k(t))rate(1 | k, \theta(t)). \tag{3}$$

Equation (3) 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 . Substituting the value for $rate(0 | k, \theta(t))$ in Eq. (3) we obtain that

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

The stationary condition $\frac{d\rho_k(t)}{dt} = 0$ implies that $\rho_k = \frac{1}{\lambda}rate(1 | k, \theta)$. Then, replacing ρ_k in Eq. (2) we obtain that

$$\theta = H_q(\theta), \tag{5}$$

where $H_q(\theta) = \frac{1}{\langle k \rangle} \sum_{k \geq 1} kP(k) \sum_{k_1=0}^k P_q(1 | k_1, k) \binom{k}{k_1} \theta^{k_1} (1 - \theta)^{(k-k_1)}$. The solutions of equation (5) are the stationary values of θ . Given θ , we can also compute the fraction of players choosing 1 in the stationary state of the dynamics as $\rho = \sum_{k \geq 1} P(k)\rho_k$.

4 The results

Consider the mean-field dynamics described above and a network with connectivity distribution $P(k)$. 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 action 1. We have the following result:

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

Proof The stationary values of θ are implicitly determined by equation (5). Replacing the value of $P_q(1 | k_1, k)$ in equation (5) we obtain the fixed point equation

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

where $[x]$ stands for the highest integer smaller or equal to x .

Also, notice that the derivative of $\theta(t)$ is equal to $\frac{d\theta(t)}{dt} = \frac{1}{\langle k \rangle} \sum_{k \geq 1} kP(k) \frac{d\rho_k(t)}{dt}$ and replacing the value of $\frac{d\rho_k(t)}{dt}$ given by Eq. (4) we find that $\frac{d\theta(t)}{dt} = -\lambda\theta(t) + \lambda H_q(\theta(t))$ which therefore implies that $\frac{d\theta(t)}{dt} > 0$ if and only if $H_q(\theta(t)) > \theta(t)$. Notice that, 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. We want to study when starting from a infinitesimally small value $\theta_0 \simeq 0$, the mean-field dynamics converges to a state where there is a positive fraction of individuals choosing 1 (i.e. where $\theta > 0$). Thus, we must determine which values of q make the state $\theta = 0$ unstable (see Fig. 1). This corresponds with the values of q such that $H'_q(0) > 1$. Notice that

$$H'_q(\theta) = \frac{1}{\langle k \rangle} \sum_{k \geq 1} kP(k) \sum_{k_1=[kq]+1}^k \binom{k}{k_1} (k_1 \theta^{(k_1-1)} (1 - \theta)^{(k-k_1)} - (k - k_1) \theta^{k_1} (1 - \theta)^{k-k_1-1}.$$

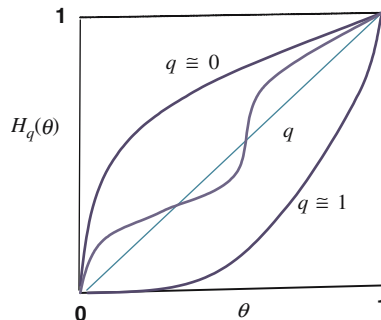
If we substitute $\theta = 0$ in the previous equation the only non-zero term of the second sum is obtained for the value $k_1 = 1$. This implies that only when $[kq] = 0$ (or equivalently $k \leq [1/q]$) the second sum is different than zero. Thus,

$$H'_q(0) = \frac{1}{\langle k \rangle} \sum_{k \geq 1}^{[1/q]} kP(k) \binom{k}{1} = \frac{1}{\langle k \rangle} \sum_{k \geq 1}^{[1/q]} k^2 P(k)$$

We can therefore implicitly determine the contagion threshold q^* as follows

$$q^* = \arg \min_{q \in \mathbb{Q}} \frac{1}{\langle k \rangle} \sum_{k \geq 1}^{[1/q]} k^2 P(k) \tag{6}$$

Fig. 1 $H_q(\theta)$ for three values of q . 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



where Q is defined as the set of $q \in [0, 1]$ such that $\frac{1}{\langle k \rangle} \sum_{k \geq 1}^{[1/q]} k^2 P(k) \geq 1$. \square

Several interesting points follow from Theorem 1. Note that action 1 must be risk dominant for contagion to occur.³ Also, if there exists a maximum connectivity in the network, say k_{\max} , and the network is connected (i.e. there exists a path connecting any two pairs of nodes) then for q sufficiently low (i.e. $q < \frac{1}{K_{\max}}$) there is global contagion. In other words, action 1 will eventually be played by the whole population.

To continue we would now like to determine which types of networks, in terms of their connectivity distributions, support contagion better.

Proposition 2 *Let $P(k)$ and $\widehat{P}(k)$ be two connectivity distributions with the same average connectivity and where $P(k)$ Second Order Stochastic Dominates (SOSD) $\widehat{P}(k)$. Let \overline{M} be the minimum integer such that*

$$\sum_{k \geq 1}^M k^2 \widehat{P}(k) \geq \sum_{k \geq 1}^M k^2 P(k) \tag{7}$$

for all $M \geq \overline{M}$. Then, if

$$\frac{1}{\langle k \rangle} \sum_{k \geq 1}^{\overline{M}} k^2 \widehat{P}(k) \leq 1 \tag{8}$$

the contagion threshold is higher for $\widehat{P}(k)$ than for $P(k)$. That is, $q^*(\widehat{P}) \geq q^*(P)$.

Proof Given that $P(k)$ SOSD $\widehat{P}(k)$ then $\langle k^2 \rangle_{\widehat{P}} > \langle k^2 \rangle_P$, where $\langle k^2 \rangle_P$ stands for the second order moment of P .⁴ This implies that \overline{M} exists. By assumption

$$\frac{1}{\langle k \rangle} \sum_{k \geq 1}^{\overline{M}} k^2 \widehat{P}(k) \leq 1 \tag{9}$$

which implies that

$$\frac{1}{\langle k \rangle} \sum_{k \geq 1}^{\overline{M}} k^2 P(k) \leq 1 \tag{10}$$

Let $M^* = [1/q^*]$ and $\widehat{M}^* = [1/\widehat{q}^*]$, where q^* and \widehat{q}^* are the contagion thresholds for the distributions $P(k)$ and $\widehat{P}(k)$ respectively. Equations (9), (10) and

³ This follows from the fact that if action 1 is not risk dominant (i.e. $q > \frac{1}{2}$) a necessary condition for contagion derived from the characterization of the contagion threshold given by Theorem 1 is $\frac{P(1)}{\langle k \rangle} \geq 1$ which never holds when $\langle k \rangle > 1$.

⁴ This is a consequence of the following property (see e.g. Mas-collel et al. 1995): If $P(k)$ and $\widehat{P}(k)$ have the same average connectivity and $P(k)$ SOSD $\widehat{P}(k)$ then $\sum_{k \geq 1} u(k)\widehat{P}(k) \geq \sum_{k \geq 1} u(k)P(k)$ for any convex function $u(k)$.

the characterization of the contagion threshold given by Theorem 1 imply that $M^*, \widehat{M}^* > \overline{M}$. And, given that for all $M \geq \overline{M}$ Eq. (7) holds, $M^* \geq \widehat{M}^*$ which in particular implies that $\widehat{q}^* \geq q^*$. \square

Therefore, under certain conditions, broader connectivity distributions are better than more homogeneous ones for contagion purposes (i.e. have a higher contagion threshold). However, these conditions also suggest that the relationship between the contagion threshold and the connectivity variance of the network is non-monotonic. To further illustrate this point, it is useful to consider some examples.

5 Examples

We compare the contagion thresholds of homogeneous, exponential and scale-free networks. Specifically, 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, where \propto means equal up to a multiplicative constant.⁵ The second network considered is an exponential network with connectivity distribution $P_E(k) \propto e^{-k/2m}$ for $k \geq m$ and $P_E(k) = 0$ otherwise. Finally, we consider a homogeneous network $P_H(k) = 1$ if $k = 3m$. Notice that the three networks introduced above have an average connectivity equal to $3m$ whereas the variance is higher in the scale-free network than in the exponential network, and higher in the exponential network than in the homogenous network. More precisely $P_H(k)$ SOSD $P_E(k)$ and $P_E(k)$ SOSD $P_{SF}(k)$.⁶

Given the characterization of the contagion threshold given by Theorem 1, and through numerical computations, we find that $q^*(P_H) \leq q^*(P_{SF}) \leq q^*(P_E)$. This result, illustrated in Fig. 2, indicates that the network with an intermediate variance (the exponential network) is the one with the highest threshold, which implies that contagion of action 1 is easier in the exponential network than in any of the other two networks.⁷ The intuition for such a result is the following:

⁵ 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. Barabasi et al. 2000).

⁶ To calculate the average connectivity we approximate the connectivity distribution $P(k)$ by a continuous distribution where $k \in [m, +\infty)$. First, we compute the multiplicative constant. For example, in the exponential network case, $P(k) = Ce^{-k/2m}$ and we can solve for C in the equation $\int_m^{+\infty} Ce^{-k/2m} dk = 1$. Notice that, once we know C , we can easily compute the average connectivity as $\int_m^{+\infty} kP(k)dk$. Furthermore, to show that $P(k)$ SOSD $\widehat{P}(k)$ we simply apply the condition that for all $x \geq 0$, $\int_0^x \widehat{F}(k)dk - \int_0^x F(k)dk \geq 0$, where $\widehat{F}(k)$ and $F(k)$ are the cumulative distribution functions of $\widehat{P}(k)$ and $P(k)$ respectively (see Mass-Colell et al. 1995).

⁷ Another simple example that leads to the same conclusion is to consider a family of connectivity distributions $P_a(k)$ with the same average connectivity $\langle k \rangle$ where, given $a \in [0, \langle k \rangle]$ half of the population has connectivity $\langle k \rangle - a$ and the other half has connectivity $\langle k \rangle + a$. Then $P_{a_1}(k)$ SOSD $P_{a_2}(k)$ if and only if $a_1 < a_2$. It is rather straightforward to show that the contagion threshold is highest for an intermediate variance connectivity distribution, specifically when $a^* = \langle k \rangle - \sqrt{2\langle k \rangle}$.

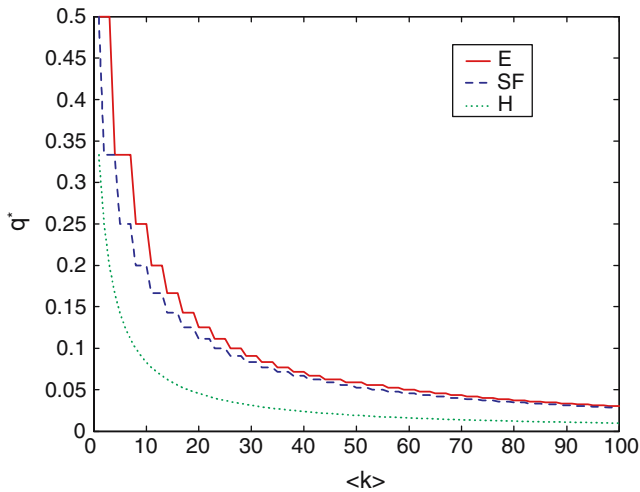


Fig. 2 The contagion threshold (q^* , ordinate) for the exponential (E), the scale-free (SF) and the homogenous (H) networks, as a function of the average connectivity ($\langle k \rangle$, abscissa)

on the one hand, although higher connectivity nodes (which are more frequent in broader networks) tend to enhance diffusion of the new action once they become adopters, they are less likely to adopt (since they require more neighbors to adopt) which offsets their relevance for the diffusion process. Moreover, low connectivity nodes are important since they are typically the initial adopters of the diffusion process. However, they must be well connected and form a giant cluster in order for the action to take off. This implies that their connectivity cannot be too low. Consequently, networks with intermediate variance (where the connectivity of the lowest connectivity nodes are not so low) are best for diffusion purposes.

Besides the analytical (and numerical) results described above we have also run simulations of the stochastic dynamics to compare their predictions with those based on the mean-field version of the model. To perform these simulations we have generated three random networks: homogenous, exponential and scale-free; each with $n = 1,000$ nodes and an average connectivity of $\langle k \rangle = 9$. We consider the discrete version of the continuous time dynamics used to derive the theoretical results for the particular value of $q = \frac{1}{6}$. 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 applied in the exact same way as in the mean-field dynamics since we now have a finite population. We therefore simply compute the fraction of individuals choosing action 1 in the long run and compare the results for the three different networks generated. In Fig. 3 we represent the number of agents choosing 1 (ordinate) as a function of the period (abscissa) for the homogenous, exponential and scale-free networks. We decided to stop at $t = 3 \times 10^4$ since the dynamics seem to reach a stationary state before that. The data are the average of 100 simulations. For each simula-

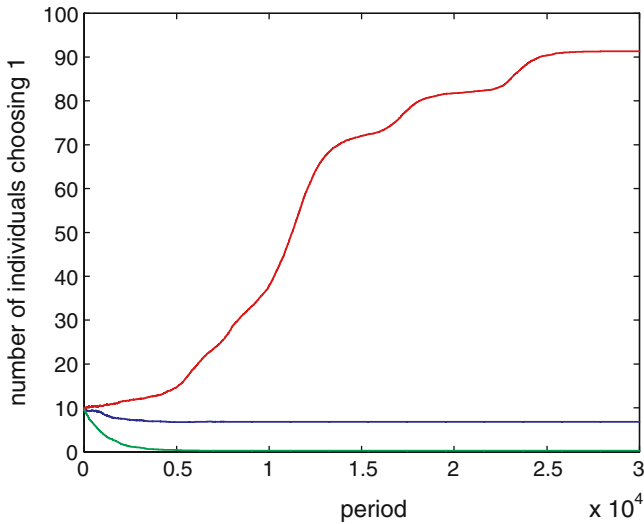


Fig. 3 Proportion of individuals choosing 1 (ordinate) as a function of the periods (abscissa) for the exponential (highest curve), the scale-free (intermediate curve) and the homogeneous (lowest curve) networks, when $q = 1/6$

tion, the initial condition is such that agents are choosing 1 in period $t = 1$ with probability 0.01.

Notice that action 1 spreads significantly only for the exponential network which would indicate that the contagion threshold is above $q = 1/6$ for the exponential network, but below $q = 1/6$ for the scale-free and homogenous networks. Therefore, as suggested by the mean-field theory, the exponential network has a higher contagion threshold than the scale-free and homogenous networks.

Acknowledgement I am most grateful to Peter Dodds and Fernando Vega-Redondo for valuable comments and suggestions. Thanks are also due to Toni Calvó-Armengol, Matt Jackson, Alison Binkowski, Juan Moreno-Ternero, Stephen Morris, Duncan Watts and the audience at the 2005 SAET Conference in Vigo. This work was made possible thanks to a postdoctoral fellowship associated to the program on robustness in social processes supported by the James S. McDonnell Foundation.

References

- 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
- Benaim M, Weibull J (2003) Deterministic approximation of stochastic evolution in games. *Econometrica* 71:873–903
- Ellison G (1993) Learning, local interaction, and coordination. *Econometrica* 61:1047–1071
- Goldenfeld N (1992) Lectures on phase transitions and the renormalization group. Addison-Wesley, New York
- Jackson MO (2006) The economics of social networks. In: Proceedings of the 9th World congress of the econometric society. Cambridge University Press, Cambridge (Forthcoming)

- Jackson MO, Yariv L (2006) Diffusion on social networks. *Économie Publique* (Forthcoming)
- Mass-Colell A, Winston M, Green J (1995) *Microeconomics theory*. Oxford University Press, Oxford
- Morris S (2000) Contagion. *Rev Econ Stud* 67:57–78
- Pastor-Satorrás R, Vespignani A (2001) Epidemic spreading in scale-free networks. *Phys Rev Lett* 86:3200–3203
- Vega-Redondo F (2006) Diffusion, search and play in complex social networks. *Econometric Society Monograph Series*. Cambridge University Press, Cambridge (Forthcoming)
- Young HP (1998) *Individual strategy and social structure: an evolutionary theory of institutions*. Princeton University Press, Princeton
- Watts DJ (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci* 99:5766–5771