Networks emerging in a volatile world*

Matteo Marsili
The International Centre for Theoretical Physics, Trieste

Fernando Vega-Redondo Bocconi University & IGIER

May 2016

Abstract

The individuals of a certain population play a coordination game with each of their partners in the social network. Over time, actions and links co-evolve with agents adjusting their actions and creating new links when they find it (myopically) optimal. The key feature of the model is that, as it indeed happens in many real-world environments, links become "obsolete" and consequently vanish at a certain rate. Our objective is to understand whether and how, under such link volatility, a densely networked society may still arise and persist.

The main contribution of the paper is two-fold. First, we fully characterize the limit behavior of the system at different time horizons (i.e. at the so-called long and ultralong runs). This characterization shows that the interplay between action choice and link creation may feed on each other to generate *sharp transitions*, *dense networks*, and *high coordination*. Second, we find that interesting path dependence and hysteresis arises in the long run (although not in the ultralong run). This has the interesing policy implication that even *small* and *temporary* interventions may have *large* effects that are quite *persistent*. .

JEL Classification nos.: C73, D83, D85.

---

*We thank George Ehrhardt for his collaboration in the early development of the project that led to this paper. We are also grateful to Bastian Westbrock, Stephanie Rosenkranz, and Marco van der Leij for suggestions on empirical evidence used to motivate our approach. Finally, we want to thank seminar participants at Boston University, Columbia, NYU, Michigan, Wisconsin, Princeton, and Berkeley whose comments and suggestions helped us improve earlier versions of this paper. Naturally, the usual caveat applies.

1 Introduction

High volatility, understood as rapid change in the environment, is an essential feature of modern economic systems. This applies, in particular, to the underlying social networks that channel the interaction among economic agents. Their links are generally subject to fast "obsolescence" and turnover, leading to a state of flux that impinges on any social phenomenon that is truly network-based.[1]

The starting point of this paper, therefore, is the notion that socio-economic networks can hardly be conceived (i.e. modelled) as either static or in equilibrium. Rather, they must be regarded as entities in continuous change, often having large stochastic forces impinge on them. The issue then arises as to whether, and how, under these circumstances, a widely networked system may arise and remain in place, thus providing the overall structure required for socioeconomic activity. And it is important to stress that the relevance of this question does not lie only in the fact that much economic activity is performed through direct inter-agent contacts. An additional important feature of networks is that they underlie the unfolding of global phenomena – such as information diffusion or overall coordination/synchronization – that are often a largely unintended by-product of agent interaction. Social networks are, in other words, the channel of link-mediated "externalities" that play an important role in economic performance.[2] The question, therefore, of whether agent interaction can be sustained at relatively high levels in the presence of sustained volatility appears interesting and nontrivial in quite a number of different respects.

Our model is abstract and purposefully simple in order to highlight the basic forces at work and the main methodological features of our approach. It couples the *network dynamics* – by which nodes adjust their links – to the *behavioral dynamics* – by which nodes modify their actions. In a nutshell, their combined operation can be described as follows: agents find it worthwhile to form (costly) links only with those displaying the same actions, while they also change these actions to coordinate as best as they can with their current partners. Under the aforementioned conditions, the problem posed by volatility is quite transparent. On the one hand, it leads to a decay of the "social fabric" that is crucially needed to preserve a high level of coordination. And, if such coordination deteriorates, new links cannot be established at a fast enough rate to compensate for that decay. This brings in the risk of a self-reinforcing collapse of the social network and, conversely, requires the onset of a virtuous circle if a dense social network

[1] By way of illustration, the Section 2 discusses evidence on two paradigmatic instances: inter-firm alliances and scientific collaborations, for which links have indeed been found to display relatively short lifetimes.

[2] Many of the network phenomena that will be listed in Section 2 (e.g. job-market contacts or interfirm cooperation) generally involve behavioral/networking decisions whose *external effects* on others are not fully taken into account by decision makers. This also happens when the social network acts as "social collateral," a phenomenon that has been recently studied (theoretically and empirically) by Karlan, Mobius, Rosenblat, and Szeidl (2008). In their empirical leading application, they show that the network of friendship and family relationship in two low-income Peruvian communities provides important (but, presumably, largely unintended) support for the interpersonal credit arrangements that represent an important underpinning of the economic life of those communities.

is to rise from scratch.

We obtain a complete characterization of how the dynamics of the system evolves in different time scales. This in turn leads to a number of predictions and interesting insights. First, we find that the mutual feedback between the "success to connect" and the "success to coordinate" leads to a positive (negative) change in both connectivity and coordination as volatility falls (rises) throughout. This effect, intuitive as it is, traces a continuous relationship except at certain specific thresholds for volatility. At these thresholds – one for an upward transition and another for a downward one – there is a drastic (de)escalation of the aforementioned feedbacks. This then brings about a sharp discontinuous change, up or down, in the density of the social network. Such abrupt transitions mark a clear separation between a networked society – in which a significant fraction of the individuals are connected by a spanning, though volatile, network – and a largely disconnected one where individuals are fragmented in many small components.

We also find that the nature of the aforementioned transitions may be affected by the time horizon considered by the analysis.[3] In a shorter time scale – what we call the long run – initial conditions essentially anchor the dynamics. This implies, in particular, that multiple stable situations are possible. Specifically, we conclude that, within a certain parameter range, either of the two aforementioned situations are stable: a densely networked or a sparsely connected society. But if one adopts the perspective that has been labelled the ultralong run (cf. Binmore, Samuelson and Vaughan (1995)), the situation drastically changes. For, in this time scale, initial conditions do *not* matter at all and a unique prediction obtains for essentially all levels of volatility. As we shall discuss, the effect of the time horizon on the predictions of the model has important implications for policy assessment. In the long run (which indeed can be arbitrarily long if the population is large enough), small and *temporary* policies can have substantial and *persistent* effects. Instead, in the ultralong run, the outcome can be affected significantly only if the policy intervention is permanent.

The rest of the paper is organized as follows. First, in Section 2 we discuss some empirical evidence on the relevance of volatility in socio-economic networks, and discuss as well how our paper relates and contributes to different strands of the literature. Then, Section 3 presents the basic theoretical framework. It describes in detail both the subprocesses of link formation and action adjustment, and also explains the different time scales to be considered in the analysis. Next, in Section 4, we prove the ergodicity of the process, characterize its invariant distribution, and highlight the most prominent features of the unique configurations that, for large populations, absorb most of the weight in

[3] As we shall explain in detail, from a mathematical viewpoint, the contrast between the two horizons is rooted in the fact that, in our model, the limit of large population sizes does not commute with the limit of infinite times. But, in fact, a remarkable technical insight that will be gathered from our analysis is that such contrast between different time scales can be equivalently obtained from a careful study the critical points of the invariant distribution of the process (in particular, a comparison of its local and global maxima).

the ultralong run. In Section 5, we turn to studying the long run, i.e. the shorter time scale where multiplicity is possible and initial conditions may play a decisive role. The main body of the paper concludes in Section 6 with a summary of our approach and a review of its main conclusions. For the sake of readability, the formal proofs of the results are gathered in an Appendix.

2 Empirical evidence and related literature

There is already a quite large body of theoretical and empirical research in economics and other social sciences that underscores the role played by social networks in a wide number of different contexts. By way of illustration, a very partial sample of this literature include the following phenomena: job search (Granovetter (1974), Calvó-Armengol and Jackson (2004)); informal insurance (Murgai *et al.* (2002), Fafchamps and Lund (2003)); trade arrangements (Kirman *et al.* (2000), Kranton and Minehart (2001)); the operation of firms and other organizations (Krackhardt and Hanson (1993), van Alstyne (1997)); industrial districts (Saxenian (1994), Castilla *et al.* (2000)); scientific collaboration (Newman (2001), Grossman (2002)), and inter-firm alliances (Hagedoorn (2002), Kogut *et al.* (2007)).

Anecdotal evidence suggests that, in many of these contexts, links are far from permanent and hence network volatility is an important force at work. To provide, however, a precise and quantitative assessment of the phenomenon, let us focus on the last two contexts listed above: scientific collaboration and inter-firm alliances. Starting by the latter, one can interpret the concept of inter-firm alliance quite widely as encompassing a number of different possibilities, e.g. research partnerships, cooperative market agreements, joint ventures, etc.. There has been much empirical research in each of these respects, which is well summarized, for example, by the aforementioned paper of Hagedoorn (2002) for research partnerships or by Kogut *et al.* (2007) for joint ventures. (Both of these papers cover a similar period starting around 1960, when inter-firm alliances of different kinds started to be an important and fast-growing phenomenon in many industries). But, to focus on the specific issue of link volatility in such inter-firm networks, two further papers worth highlighting are Harrigan (1988) and Park and Russo (1996). As we explain next, both provide empirical evidence that can be used to suggest some quantitative proxy for it.

Specifically, Harrigan (1988) studies 895 alliances (including both those equity-based and not) from 1924 to 1985 in a wide range of different economic sectors (see Table 1 in her paper). She concludes that the average life-span of the alliance is relatively short, 3.5 years, with a standard deviation of 5.8 years and 85% lasting less than 10 years. On the other hand, Park and Russo (1996) consider the narrower notion of (equity-based) joint venture, and focus on 204 such alliances among firms in the electronic industry for the period 1979-1988. Accounting for both outright failures and other routes to termination, they obtain (see Table 1 in their paper) a half-life of less than five years – i.e. less than half of them remain active beyond that period. On the other hand, among those joint ventures that last less than 10 years (2/3 of the total), the average lifetime turns out to be 3.9 years.

Turning now to networks of scientific collaboration, these has been extensively studied for disciplines such as physics, biomedical research, computer science, or mathematics (see e.g. the aforementioned papers by Newman (2001) and Grossman (2002)). Recently, the analysis has been extended by Goyal, van der Leij, and Moraga-González (2006) to the field of economics. Based on the data set used in the latter paper, Marco van der Leij (private communication) has provided us with the following approximate way of assessing "link volatility" among economists who collaborated in writing papers during the period 1970-2000. Let us first operationalize the duration of a link between two researchers as the time elapsed between their last and first *joint* publication. Identify, on the other hand, the length of the academic life of a researcher (which obviously bounds the length of any collaboration) with the time elapsed between his/her first and last published papers. Then, a natural proxy for overall volatility results from comparing the averages of those two variables: actual duration of links and their maximum possible duration. Restricting consideration to economists with a relatively long academic career (at least 15 years), the average link duration among the almost 6000 remaining economists turns out to be 3.5 years. This is to be compared with an average maximum duration of 23 years, indeed suggesting relatively high levels of volatility in the pattern of collaboration taking place among economists during that period.

Now we address the relationship of the present paper to the theoretical network literature. First we briefly refer to the part of this literature that has studied network formation alone, then to that part which has focused on the study of games in networks, and finally we discuss some of the more recent papers that have integrated both network formation and play.

The formal study of network formation problems in economics was initiated by the seminal papers of Jackson and Wolinsky (1996) and Bala and Goyal (2000). Since then, there has been a fast-growing literature that, to put it very schematically, has been addressing the problem from a static (strategic) viewpoint or/and a dynamic (learning) perspective.[4] Mostly developed outside of economics, there is also a recent and influential literature on network formation that adopts an essentially algorithmic viewpoint and emphasizes the complexity and topological properties of the resulting networks. The so-called *small-world networks* introduced by Watts and Strogatz (1999) or the *scale-free networks* studied by Barabási and Albert (1999) inaugurated this line of research.[5]

A second branch of the network literature has been concerned with understanding how local interaction (as modelled by some underlying network) affects play in games. Much of this research has focused on either Prisoner's Dilemma games (Eshel, Shaked and Samuelson (1998)) or coordination games (Blume (1993), Ellison (1993), Young (1998)). While in the first case the issue

[4] See Jackson (2005) for an extensive survey.

[5] Newman (2003) and Vega-Redondo (2007) provide partial surveys of this literature. On the other hand, for an interesting paper that combines ideas from the study of complex networks with purposeful and local networking behavior, see Jackson and Rogers (2007).

is whether local interaction may facilitate cooperation, in the second setup the question is how it affects equilibrium selection – specifically, the possible conflict between efficiency and risk dominance.

Finally, we refer to some recent papers that integrate both network formation and behavioral adjustment into a common coevolutionary setup. When local interaction is assumed to take place through a coordination game – e.g. in the papers by Jackson and Watts (2002), Ely (2003), or Goyal and Vega-Redondo (2005) – the setup considered bears close similarities to our present framework. The main point of contrast lies in the role played by noise in each case. The aforementioned papers follow the received evolutionary literature in conceiving noise as "infinitesimal" – i.e. a small stochastic perturbation of the core dynamics. Instead, in our setup, volatility is taken to operate at a pace commensurable with other components of the process, and this affects the analysis very significantly. For, as we have explained, it plays a key role in the strong behavior/network interplay – e.g. the reciprocal feedback between the "success to connect" and the "success to coordinate" – that characterizes the dynamics of the model.

3 Theoretical framework

3.1 Description of the model

Let there be a certain population of agents, $P = \{1, 2, ..., N\}$, who interact bilaterally over time as specified by the evolving social network. Time is modelled continuously, with $t \in [0, \infty)$. At any t, the state of the system $\omega(t) = (\boldsymbol{\alpha}(t), G(t))$ consists of two items: a strategy profile and a network.

The profile $\boldsymbol{\alpha}(t) = (\alpha_i(t))_{i \in P} \in A^P$ specifies the action $\alpha_i(t) \in A$ chosen by each agent i in her interaction with everyone of her neighbors in the network $G(t)$. This interaction is carried out through a certain symmetric coordination game for which the set $A = \{a_1, a_2, ..., a_q\}$ is the strategy space and the payoffs are given by the function $\psi(a_l, a_m)$ specifying the *payoff flow* earned by a player who chooses a_l when the partner chooses a_m. More compactly, payoffs can be represented by a (square) matrix $W \equiv (w_{lm})_{l,m=1}^{q}$ in which each $w_{lm} \equiv \psi(a_l, a_m)$. The game is assumed to be one of coordination, so that $w_{ll} > w_{ml}$ for all $l \neq m$.

The second component of the state prevailing at any given time t is the network, which is described by a symmetric adjacency matrix $G(t) = (g_{ij}(t))_{i,j \in P} \in \{0, 1\}^{N^2}$. An *undirected* link between i and j exists at t if $g_{ij}(t) \equiv g_{ji}(t) = 1$, while this link does not exist if $g_{ij}(t) \equiv g_{ji}(t) = 0$. (For simplicity, we make $g_{ii}(t) = 0$ for all i.) Each player i is taken to play the coordination game with each of her neighbors $j \in \mathcal{N}_i(t) \equiv \{k \in P : g_{ik}(t) = 1\}$. Such accumulated interaction yields her total *gross* payoffs. Her *net* payoffs, on the other hand, are obtained by simply subtracting the cost of links from the gross payoffs.

Formally, given the network $G(t)$ and action profile $\boldsymbol{\alpha}(t)$ prevailing at any t, the total *gross payoff* $\pi_i(\omega(t))$ is given by:

$$\pi_i(\omega(t)) = \pi_i(\boldsymbol{\alpha}(t), G(t)) = \sum_{j \in \mathcal{N}_i(t)} \psi(a_i(t), a_j(t)),$$

i.e. the sum of all bilateral payoffs obtained by i when she chooses the *same* action $a_i(t)$ against the actions $a_j(t)$ chosen by each of her neighbors. Denoting by $c > 0$ the cost flow of maintaining each link per unit of time, the net payoffs $\hat{\pi}_i(\omega(t))$ are simply given by:

$$\hat{\pi}_i(\omega(t)) = \pi_i(\omega(t)) - c\,|\mathcal{N}_i(t)|$$

where $|\mathcal{N}_i(t)|$ stands for the cardinality of the set $\mathcal{N}_i(t)$, i.e. the number of neighbors of i at t.

Over time, agents adjust both their links and their behavior. Mathematically, the induced dynamics is described by a continuous time Markov process with states $\omega(t) \in \Omega$. It is completely determined, therefore, by the rates $\rho(\omega \to \omega')$ governing all possible transitions $\omega \to \omega'$. These transitions pertain to adjustments that involve (a) link revision, (b) action revision, (c) volatility. We now describe each of these in turn.

(a) *Link revision:* At a certain positive rate η, each agent i receives a link revision opportunity. When such an opportunity arrives at some t, another agent j is randomly chosen in the population (all with the same probability). The main idea we want to capture is that the link ij is formed (when nonexisting) or kept (when already existing) if, and only if, it induces a positive change in the net flow of payoffs; or, equivalently, if the gross payoff entailed is larger than the linking cost c. To make things simple, we shall posit, in particular, that the gross payoff earned by each of the two agents i and j exceeds the linking cost if, and only if, both agents are coordinated in the same action, i.e.

$$\min\{\psi(a_i(t), a_j(t)), \psi(a_j(t), a_i(t))\} > c \Leftrightarrow a_i(t) = a_j(t).$$

This condition would be satisfied if, for example, the linking cost c is relatively small and the game is of so-called pure coordination, i.e. both players obtain a given positive payoff if they coordinate on the same strategy (or action) while obtain a zero payoff otherwise. Or, more generally, it would apply (if c is small) to any coordination game where in case of miscoordination, at least one agent obtains a non-positive payoff.

Thus, to sum up, when player i receives an opportunity to revise her linking status with some player j, we posit that the link between them will be formed (i.e. created or kept) if, and only if, both agents are currently displaying the same action. Conceptually, this formulation is based on the standard assumption of evolutionary models that agents are guided by myopic considerations – or, as sometimes indicated, that they have static expectations. Mathematically, it implies that, for all pairs of states $\omega = (\boldsymbol{\alpha}, G)$ and $\omega' = (\boldsymbol{\alpha}, G')$, the rate of change from the former to the latter is given by

$$\rho(\omega \to \omega') = \frac{\eta}{N-1} \tag{1}$$

if, for some particular pair ij, we have $\alpha_i = \alpha_j$ and $g_{ij} = 0$ in state ω, while in state ω' we have $g'_{kl} = 1$ if $kl = ij$ and $g'_{kl} = g_{kl}$ otherwise.

(b) *Action revision*: At a certain positive rate ν, each agent i receives an action revision opportunity. When such an opportunity arrives at some t, we assume that she simply best-responds to the current situation. That is, agent i chooses an action $a_i^* \in A$ such that[6]

$$\hat{\pi}_i((a_i^*, (\alpha_j(t))_{j\neq i}), G(t)) \geq \hat{\pi}_i((a_i, (\alpha_j(t))_{j\neq i}), G(t))$$

for all $a_i \in A$. Let $B((\alpha_j(t))_{j\neq i}), G(t))$ denote the set of actions that satisfy the above inequality. If this set is not a singleton, we shall just assume, for simplicity, that all the actions in that set are chosen with equal probability. Again, such a best-response adjustment rule is typical in evolutionary game theory, and can be motivated by the assumption that agents hold static expectations and/or behave myopically.

Mathematically, it implies that, for all pairs of states $\omega = (\boldsymbol{\alpha}, G)$ and $\omega' = (\boldsymbol{\alpha}, G')$, the rate of change from the former to the latter is given by

$$\rho(\omega \to \omega') = \frac{\nu}{|B((\alpha_j)_{j\neq i}), G)|}, \tag{2}$$

if $G = G'$ and there is some particular agent i such that $\hat{\pi}_i((\alpha'_i, (\alpha_j)_{j\neq i}), G) \geq \hat{\pi}_i((a_i, (\alpha_j)_{j\neq i}), G)$ and $\alpha'_j = \alpha_j$ for all $j \neq i$. It is worth noting the implications of our formulation for the following two particular cases. First, if a state ω has all agents being perfectly coordinated on the same action as all their respective neighbors, then no action adjustment can take place. Thus, in this case, any change must concern the network alone. Second, if in state ω a particular agent i is isolated (i.e. has no links), then a transition to another ω' where this agent displays *some* different action occurs at the rate $\nu(1 - 1/q)$.

(c) *Volatility:* Existing links decay at a rate λ, which for simplicity is taken to be constant and exogenous. This component of the process is to be conceived as capturing unmodelled environmental change that destroys the value or feasibility of existing links. It is different from the endogenous process of link removal explained in (a) above. It is akin to the exogenous noise (sometimes called mutation) that is often considered by evolutionary game theory. But in contrast with most of the received evolutionary literature, a key difference is that such noise is taken to be of a significant magnitude and interacts with the rest of the components of the process in a comparable time scale. The importance of allowing for such significant noise in modelling modern economic systems was already stressed and motivated in Sections 1 and 2.

Mathematically, such a formalization of volatility implies that, for all pairs of states $\omega = (\boldsymbol{\alpha}, G)$ and $\omega' = (\boldsymbol{\alpha}, G')$, the rate of change from the former to the latter is given by

$$\rho(\omega \to \omega') = \lambda, \tag{3}$$

[6] Note that, since only an action change is considered at this juncture, whether the agent uses gross or net payoffs to evaluate its performance is inessential.

if, for some particular pair ij, we have $g_{ij} = 1$ in state ω, while in state ω' we have $g'_{kl} = 0$ if $kl = ij$ and $g'_{kl} = g_{kl}$ otherwise.

The three kind of transitions whose rates are given by (1), (2), and (3) exhaust all possibilities. Every other conceivable transition, in other words, occurs at a zero rate given the rules that govern the three different components of the process: link revision, action revision, and volatility.

The speeds (or rates) at which these subprocesses operate is modulated by the three parameters of the model: η, λ, and ν, respectively. The rate ν that governs action revision will be found to play a very subsidiary role in the analysis – e.g. it does not enter the characterization of the invariant distribution of the process, which is independent of ν. Concerning the other two parameters, on the other hand, it is clear that only the ratio η/λ matters. This allows us to normalize $\lambda = 1$ (by, say, a suitable choice of time units) and thus focus on η as the single key parameter of the model. With this normalization, if η $(= \eta/\lambda)$ is small, the process is to be conceived as dominated by volatility (i.e. a *relatively* high λ). In this case, therefore, the typical time between the arrivals of link-creation opportunities is long compared to the lifetime of links. Conversely, high values of η correspond to a context with low volatility, so the links that are created endure long enough to have a significant effect on the ensuing evolution of the process.

3.2 Analysis of the model: two time scales

Our analysis of the model will focus on the asymptotic behavior of the induced dynamics for large populations, i.e. we shall study the model in the limit $t \to \infty$ and $N \to \infty$. As we already advanced in Section 1, the order in which these limits is taken has a particular significance for the dynamics of the model, for it produces a sharp distinction between two different regimes that we call the "long run" and the "ultralong run."[7] Before analyzing these two cases in the next sections, it is worth elaborating further on the differences between them and how the contrast arises.

The *ultralong run* represents the time scale where, given any arbitrarily large but *finite* population, one is ready to let the dynamics proceed for arbitrarily long times. Formally, it corresponds to the limit operation being taken first in time and then (since the population is assumed large) in population size. Because of the ergodicity of the process (cf. Proposition 1 below), the model delivers in this case the unique (probabilistic and frequentist)[8] prediction embodied by the invariant distribution. We shall find that, for large populations, the invariant distribution concentrates its mass around the configurations where the

[7] As mentioned, an analogous dichotomy has been proposed by Binmore, Samuelson, and Vaughan (1995) in their analysis of evolutionary models.

[8] The so-called *Ergodic Theorem* asserts that if μ denotes the invariant distribution of the process $\{\omega(t)\}$ then, for every state $\hat{\omega}$, $\lim_{t\to\infty} \Pr(\omega(t) = \hat{\omega}) = \mu(\hat{\omega})$. On the other hand, a frequentist consequence of this theorem is that if $\delta_{\hat{\omega}}(t)$ stands for the indicator function which is 1 if $\omega(t) = \hat{\omega}$ and 0 otherwise, then $\lim_{t\to\infty} \frac{1}{t}\int_0^t \delta_{\hat{\omega}}(s)\, ds = \mu(\hat{\omega})$ almost surely.

invariant distribution achieves a global maximum. Such configurations, therefore, can be regarded as the ultralong-run prediction of the model.

In contrast, in order to identify the shorter time scale we call the *long run* the aforementioned limit operations must be taken in the reverse order. That is, for any arbitrarily large but finite time horizon, the population is assumed to be so large (virtually infinite) that the finite-population noise that is crucial for the convergence to the ergodic distribution of the original process is ruled out by construction. Mathematically, it turns out that such a long-run time scale can be suitably studied through the so-called mean-field dynamics, i.e. a *deterministic* dynamical system where the original stochastic process is replaced by the expected law of motion of the relevant aggregate variables. Over this time-scale, therefore, the outcome selected by the dynamics depends on the initial conditions of the process, in contrast with the ultralong run prediction which is independent of initial conditions.

Mathematically, we shall find out that the different (robust) predictions of the mean-field dynamics (i.e. its asymptotically rest points) happen to coincide with the local maxima of the invariant distribution of the (finite-population) stochastic process. This, in turn, leads to a sharp integration of the long and ultralong run analyses in a common setup. For, in brief, the different time scales of the process then simply arise from the dichotomy between local and global maxima of the unique invariant distribution of the original stochastic process.

And conceptually, the contrast between the two time scales underscores the point that, *even asymptotically*, there are different horizons in which the predictions of the model can be assessed. If the population is very large, the long run is the most appropriate and thus initial conditions may matter crucially. Instead, if enough noise persists in the system because the population is not too large, the analysis should focus on the ultralong run and largely ignore the influence of initial conditions. In the latter case, no significant effect can be expected from a parameter change if it is not permanent. Instead, in the former case, an even temporary change in a parameter (possibly due, say, to a brief policy intervention) can have persistent effects by "freeing" the system from the weight of a bad history. An elaboration on this idea will be undertaken at the end of Section 5, once the formal analysis of the model has been completed.

4 The ultralong run

The first point to note is that the Markov process $\{\omega(t)\}$ with law of motion given by (a)-(c) is ergodic. This is the content of the following result, which also indicates that any state with links connecting agents choosing different actions is transient and thus is assigned a vanishing asymptotic probability by the invariant distribution.

Proposition 1 *Let* $\hat{\Omega} \equiv \{\omega = (\boldsymbol{\alpha}, G) \in \Omega : \forall i, j \in P,\ [g_{ij} = 1 \Rightarrow \alpha_i = \alpha_j]\}$ *be the set of all states where links exist only between agents exhibiting the same action. The process* $\{\omega(t)\}$ *has a unique invariant distribution* μ *with* $\mu(\hat{\Omega}) = 1$.

Proof: See the Appendix.

The above result is a simple consequence of the fact that all links decay at a constant rate. Thus, the empty network can be reached from any state, and from an empty network only states in $\hat{\Omega}$ can be reached through subsequent link adjustment. This in turn implies that, independently of initial conditions, a single recurrent class in $\hat{\Omega}$ must eventually absorb all paths of the stochastic process with probability one. Thus, as anticipated above, all configurations where an agent is in a situation different from those two explicitly covered in our specification of action revision dynamics are transient. That is, asymptotically, each agent will either be isolated or linked to agents all adopting his/her same action.

Next, in Subsection 4.1, we characterize the invariant distribution for finite population size N and then in Subsection 4.2 obtain some induced measures on interesting aggregates. Finally, in Subsection 4.3, we study the situation when the population size N grows large.

4.1 The invariant distribution

To characterize the invariant distribution established by Proposition 1, it is enough to identify *one* probability distribution on Ω that verifies the stationarity conditions embodying invariance. For, by ergodicity, there can be *only one* such probability distribution. So let $\rho(\omega \to \omega')$ denote the rate at which a transition from states ω to ω' occurs at any point in time. Then it is clear that any probability distribution μ that satisfies the following so-called "detailed balance condition" (see Gardiner (2004)):

$$\mu(\omega')\rho(\omega' \to \omega) = \mu(\omega)\rho(\omega \to \omega') \qquad (\forall \omega, \omega' \in \Omega)$$

is stationary (or invariant) under the contemplated process. Building upon this condition, we can then show that the (unique) probability distribution that remains invariant is that specified in the following result.

Proposition 2 *The invariant probability measure μ is given by*

$$\mu(\omega) = \begin{cases} \mu_{g=0} \prod_{i,j \in P,\, i<j} \left(\frac{2\eta}{N-1}\right)^{g_{ij}} & \text{if } \omega = (\alpha, G) \in \hat{\Omega} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $\mu_{g=0} \equiv \left[\sum_{\omega \in \hat{\Omega}} \prod_{i<j} (2\eta/(N-1))^{g_{ij}}\right]^{-1}$ is a normalizing constant (interpretable as the aggregate probability of all states associated to an empty network).

Proof: See the Appendix.

4.2 Class size and degree distributions

Having characterized through the invariant distribution μ the limit behavior of the process, we are now interested in "extracting" from it some of the key properties it imposes on the corresponding social structure. We shall start by deriving the distribution induced over the number of agents displaying each of the q possible actions. This distribution – which we call the class size distribution – gives us a measure of the extent to which the population settles or not in some dominant behavior (or social norm). As we shall explain, it also induces uniquely key properties of the prevailing social network.

Denote by $\mathbf{N} = (N_1, ..., N_q)$ a generic population profile specifying the number N_r of agents displaying each action a_r, with $\sum_{r=1}^{q} N_r = N$. Correspondingly, let $\hat{\Omega}(\mathbf{N}) \equiv \{\omega = (\boldsymbol{\alpha}, G) \in \hat{\Omega} : |\{i \in N : \alpha_i = a_r\}| = N_r, \ r = 1, ..., q\}$ represent the set of states where there are exactly N_r agents displaying each action a_r. Then, the *class size distribution* ζ is defined by

$$\zeta(\mathbf{N}) = \mu(\hat{\Omega}(\mathbf{N})).$$

That is, for each profile $\mathbf{N}$, the value $\zeta(\mathbf{N})$ is simply obtained by adding the probability associated to all possible states that are consistent with such a vector $\mathbf{N}$. By relying on the factorized form of μ in (4), we arrive at the specific expression for ζ that is stated in the following result.

Proposition 3 *For each* $\mathbf{N} = (N_1, ..., N_q)$, $\sum_{r=1}^{q} N_q = N$, *the probability* $\zeta(\mathbf{N})$ *induced by the invariant distribution* μ *is given by*

$$\zeta(\mathbf{N}) = \mu_{g=0} \frac{N!}{\prod_{r=1}^{q} N_r!} \prod_{r=1}^{q} \left[\left(1 + \frac{2\eta}{N-1}\right)^{\frac{1}{2}N_r(N_r-1)} \right] \tag{5}$$

where $\mu_{g=0}$ *is the normalization constant in (4).*

Proof: See the Appendix.

Likewise, as a corollary to the previous results, it is possible to rely on the form of the invariant distribution (4) to derive exhaustive information on the statistics of the social network.

Corollary 4 *Let* $\mathbf{P} = (P_1, ..., P_q)$ *be any given partition of agents in action classes, with* $P_r = \{i \in N : \alpha_i = a_r\}$, $r = 1, ..., q$.

(i) For any non-transient state $\omega = (\boldsymbol{\alpha}, G)$ *consistent with* $\mathbf{P}$, *the social network* G *is composed of* q *disjoint subnetworks* G_r *for each sub-population* P_r, $r = 1, \ldots, q$.

(ii) For each $r = 1, \ldots, q$, *the probabilistic ensemble of subnetworks* G_r *consistent with* $\mathbf{P}$ *that are induced by the invariant distribution (4) defines*

an Erdös-Rényi random network ensemble[9] $G(N_r, p_r)$ *with* $N_r = |P_r|$ *and* $p_r = 2\eta/(N-1+2\eta)$.

The proof, detailed in the Appendix, relies on Proposition 1 (which implies that no link between agents displaying different behavior is possible in the ultralong run) and on the factorized form of the invariant distribution established by Proposition 2 (which implies statistical independence among different links involving agents $i, j \in P_r$ in the same partition).

The previous results represent a key starting point of our analysis. For it turns out that the induced measure $\boldsymbol{\zeta}$ over class sizes that is characterized in Proposition 3 underlies most of the asymptotic properties of interest of our process. For example, by relying on Corollary 4, it determines the properties of the underlying social network, such as the *degree distribution*. To see this, recall that the degree of an agent i is defined as the number $k_i = |\{j : g_{ij} = 1\}|$ of agents she is connected to, and the degree distribution simply specifies the probabilities that any given agent is connected to some k other agents, with $k = 0, 1, 2, ..., N-1$. Then, given any profile of class sizes, $\mathbf{N} = (N_1, ..., N_q)$, Part (ii) of the above Corollary indicates that the conditional probabilities for each degree k in each action class $r = 1, 2, ..., q$ takes the following binomial form:

$$\Pr_\mu\{k_i = k | \alpha_i = a_r\} = \binom{N_r - 1}{k} \left(\frac{2\eta}{N-1+2\eta}\right)^k \left(1 - \frac{2\eta}{N-1+2\eta}\right)^{N_r-1-k} \tag{6}$$

This implies, therefore, that the conditional probability distributions over degrees prevailing within every action class indeed depend solely on the distribution ζ over class-size profiles $\mathbf{N}$.

4.3 Large populations

In this subsection, the focus is on the regularities displayed by the ultralong run behavior of the process when the population is very large, i.e. in the limit $N \to \infty$. As explained above, we want to focus on the induced size distribution of the different action classes, since this determines the underlying social structure. Here, however, rather than considering for each action a_r the total number N_r of agents who display it (which would grow to infinity), we turn our attention to their *frequency* $n_r \equiv N_r/N \in [0, 1]$. Thus, as a counterpart of Proposition 3, our primary objective will be to characterize the main properties of the invariant distribution over profiles $\mathbf{n} = (n_1, ..., n_q)$.

Before turning to the formal analysis of the problem, it may be useful to anticipate informally its essential features. The key conclusions are established by Proposition 6, which characterizes the *local maxima* of the invariant distribution for large N. To understand why such a characterization is important, we

[9] We recall that the Erdös-Rényi random-network ensemble $G(N, p)$ assigns to each network g of N nodes a probability equal to $p^{K(g)}(1-p)^{N-K(g)}$, where $K(g)$ is the total number of links on g.

must refer to Lemma 5. This auxiliary result indicates that, for large N, the invariant distribution associates essentially all the probability mass to its global maxima. This suggests that, in order to assess what configuration dominates in the ultralong run, it is enough to compare the value of the invariant distribution μ at its different local maxima.

More specifically, Proposition 6 shows that when η, the rate of link creation, is very low or very high the situation is clear-cut: In these cases, μ displays unique *global* maxima representing configurations where, respectively, either all actions are equally represented (and the network connectivity is low) or there is one action that becomes dominant (and overall connectivity is high). In an intermediate range of values of η, however, those two type of configurations correspond to distinct local maxima, so a comparison between them is in order. In this respect, we shall see that there is a certain threshold η^* such that when $\eta < \eta^*$ the global maximum is achieved in configurations of the first kind, while if $\eta > \eta^*$ those of the second type obtain. This, therefore, implies a sharp transition between the two regimes, as η crosses the threshold.

As the preceding discussion explains, our first step in the analysis involves showing that, for large N, the maxima of the invariant distribution absorb most of the probability mass. This follows directly from the fact that the distribution (5) is arbitrarily well approximated by an expression of the form

$$\zeta(\mathbf{N}) \simeq e^{-Nf(\mathbf{n})} \tag{7}$$

for some function $f : \Delta^{q-1} \to \mathbb{R}$ of $\mathbf{n} = (n_1, n_2, ..., n_q)$, where we recall that $n_r \equiv N_r/N$ is the fraction of agents adopting action a_r and Δ^{q-1} is the $(q-1)$-dimensional simplex. The crucial implication of the above expression is that it allows a sharp *decoupling* of the effect of population size N and that channeled through class frequencies $\mathbf{n}$. This is indeed what allows us to conduct an effective analysis of the limit scenario where N is taken to grow arbitrarily large. The following result states the indicated conclusion more precisely, and also provides a specific form for the function f.

Lemma 5 *Given any* $\mathbf{n}$*, and any sequence of* $\mathbf{N}$ *such that* $\mathbf{N}/N \to \mathbf{n}$ *as* $N \to \infty$*, the limit invariant probability measure satisfies*

$$\begin{aligned} -\lim_{N\to\infty} \frac{1}{N} \log \zeta(\mathbf{N}) &= f(\mathbf{n}) \\ &= f_0 + \sum_{r=1}^{q} \left[n_r \log n_r - \eta n_r^2 \right] \end{aligned} \tag{8}$$

for some constant f_0*.*

Proof: The proof – an iterated application of Stirling's formula – can be found in the Appendix.

Lemma 5 implies that, for large N, the configurations around which the invariant distribution concentrates its mass must be minima of the function

f. Motivated by this observation, we now address the following minimization problem:

$$\min_{\mathbf{n}\geq\mathbf{0}} f(n_1, n_2, ..., n_q) \tag{9}$$

$$\text{s.t.} \quad \sum_{r=1}^{q} n_r = 1. \tag{10}$$

As a first step in finding the solution to such a constrained optimization problem, let us focus on the following necessary First-Order Conditions (FOC):

$$n_r \, e^{-2\eta n_r} = e^{\beta - 1} \qquad (r = 1, ..., q) \tag{11}$$

where β is a Lagrange multiplier enforcing the normalization constraint (10). In combination with suitable second-order conditions (see below), those conditions characterize all local minima of f – which, of course, may fail to be global minima. Local minima, however, are of interest for two reasons. First, of course, they are the sole candidates to minimize f globally. Second, as outlined in Subsection 3.2, they will be seen to be of interest in themselves, since they turn out to be the relevant predictions in a time scale shorter than that associated to ergodicity.

The properties of the function $h(x) = xe^{-2\eta x}$ appearing in (11) readily imply a stark conclusion: there are only two possible values for each n_r – which we shall denote by n_+ and n_- ($n_+ \geq \frac{1}{2\eta} \geq n_-$) – that can be part of a solution to the above optimization problem. (This follows from the fact that $h(x)$ has a unique local maximum at $x = \frac{1}{2\eta}$, and $\lim_{x\to\infty} h(x) = h(0) = 0$.) Consequently, the set of actions can be divided into two categories alone. On the one hand, there are the relatively dominant ones, whose *common* frequency is n_+. And on the other hand, we have the actions that are in relative minority, whose common frequency is n_-. It becomes possible, therefore, to classify every solution of the first order conditions (11) by the number L_+ of components with $n_r = n_+$, the number of components with frequency n_- being $L_- = q - L_+$.

Furthermore, (11) implies that $n_+ e^{-2\eta n_+} = n_- e^{-2\eta n_-}$. This, in combination with the normalization condition $L_+ n_+ + (q - L_+) n_- = 1$, allows one to determine suitable values of n_+ (and n_-) for all L_+ and η. With a little algebra, we find that n_+ is implicitly defined by the solution(s) of the following equation:

$$n_+ = \left[L_+ + (q - L_+) e^{-2\eta \frac{q n_+ - 1}{q - L_+}} \right]^{-1} \tag{12}$$

whereas $n_- = (1 - L_+ n_+)/(q - L_+)$ is again given by normalization.

Summarizing, the set of possible solutions of the FOC can be parametrized by the number L_+ of dominant components with $n_r = n_+$, where n_+ is a solution of (12). Local minima are the solutions in this set that also satisfy the relevant second-order conditions of the minimization problem. Then, by evaluating the function f at those local minima, one can further select those that are global minima. The following proposition provides a detailed analysis of the situation, including a full characterization of the local and global minima associated to the different values of q and η.

Proposition 6 *Given $q \geq 2$ (the number of actions), there exist two thresholds, $\check{\eta} \leq \hat{\eta}$, for the link-formation rate η such that the structure of local and global minima of f can be characterized as follows:*

(i) The uniform configuration $\mathbf{n} = (1/q, ..., 1/q)$ is a local minimum (thus $L_+ = 0$) if, and only if, $\eta \leq \hat{\eta}$.

(ii) For each action a_r $(r = 1, \ldots, q)$ there exist a unique local minimum with action a_r being the single dominant action if and only if $\eta \geq \check{\eta}$. The corresponding configuration $\mathbf{n} = (n_1, n_2, ..., n_q)$ has

$$n_r = n_+ \geq \frac{1}{q}, \quad n_{r'} = n_- = \frac{1 - n_+}{q - 1}, \quad \forall r' \neq r$$

where n_+ is the (single) solution of (12) with $L_+ = 1$ for which n_+ increases with η.

(iii) No other configurations are local minima – in particular, none exists with more than one dominant action (i.e. with $L_+ \geq 2$).

(iv) There is a value η^* given by

$$\eta^* = \frac{q-1}{q-2} \log(q-1), \qquad \check{\eta} \leq \eta^* \leq \hat{\eta} \tag{13}$$

such that for $\eta < \eta^*$ the uniform local minimum specified in (i) is the unique global minimum, while for $\eta > \eta^*$ the local minima with one dominant action specified in (ii) are the sole global minima.

(v) For $q = 2$, $\hat{\eta} = \eta^* = \check{\eta} = 1$ whereas $\check{\eta} < \eta^* < \hat{\eta} = q/2$ for $q > 2$.

Proof: See the Appendix.

The combination of Parts *(i)-(iii)* of Proposition 6 establishes the existence of three different ranges for the parameter η in which qualitatively different situations are possible.

1. First, in the high (relative)[10] volatility region $\eta \in (0, \check{\eta})$, the function f attains its unique local minimum at the symmetric configuration where all q actions display the same frequency. This configuration thus defines the global maximum of the invariant distribution and thus represents the unique prediction of the model in the infinite-time limit.

[10] Note that, since the volatility rate λ has been normalized to one, the value $1/\eta$ can be regarded as the relative volatility rate.

2. A polar situation arises for low volatility, i.e. $\eta \in (\hat{\eta}, \infty)$. In this case, the local minima of f are given by all of the q configurations where *one* of the actions becomes dominant. Each of these configurations also defines a global maximum of the invariant distribution of the process, and thus all of them represent alternative ultralong-run outcomes, although with fully equivalent implications.[11]

3. Finally, a significantly different state of affairs obtains for the intermediate parameter range where $\eta \in (\check{\eta}, \hat{\eta})$. In this case, configurations of the two polar kinds described above are *a priori* possible, since both of them embody local minima of the function f.

Figure 1 illustrates the above set of conclusions by depicting the minima of f as a function of η, for $L_+ = 0$ and $L_+ = 1$. These two types of configuration reflect a different way of striking the balance between link creation and link destruction for a given action class, depending on its size. When the class size is relatively small/large, the rate at which link creation absorbs formerly isolate nodes (and fixes their action) is similarly small/large. But this inflow is matched by a correspondingly small/large rate at which nodes become isolates and then may change actions. These heuristic considerations indeed suggest that configurations such as those above where class size is either uniform ($L_+ = 0$) or there is a dichotomy between large and small classes ($L_+ = 1$) may be stable – even both of them simultaneously – for suitable values of the parameters.

Figure 1 also illustrates the second-order requirements that allow us to distinguish the minima of f (whose corresponding value is depicted by solid lines) from its saddle points (given by dashed lines, which are also shown for $L_+ = 2$). An intuitive understanding of why solutions with $L_+ > 1$ are not (local) minima of f can be gained with the following argument. Imagine a situation with $s = L_+ > 1$ actions being dominant (e.g. $n_r = n_+$ for $r \leq s$). Now suppose that, by a random fluctuation, one of the dominant classes, say the one for a_1, acquires slightly more population mass then the others ($n_1 > n_+$). Then this class will recruit slightly more members from the pool of isolated agents, in comparison with other components with dominant actions. (Just recall that the probability for successful link formation by way of random matching for an agent choosing action a_r is proportional to the current density n_r.) For the same reason, the density of links in class 1 will be slightly higher, making it less likely that agents in this component become isolated. The combined effect is then that the initial fluctuation will grow larger as time proceeds, eventually leading to a configuration with $L_+ = 1$, i.e. a configuration where class 1 is uniquely dominant.

[11] Since these q alternative configurations are all isomorphic (only the identity of the dominant action is changed), we can naturally think of them as essentially capturing a unique prediction. Indeed, in principle, one can conceive of a time scale even beyond the ultralong run in which no single action stands out from the rest, and the system undergoes transitions across those isomorphic configurations. Such inordinately large time-scales are beyond our present interest.

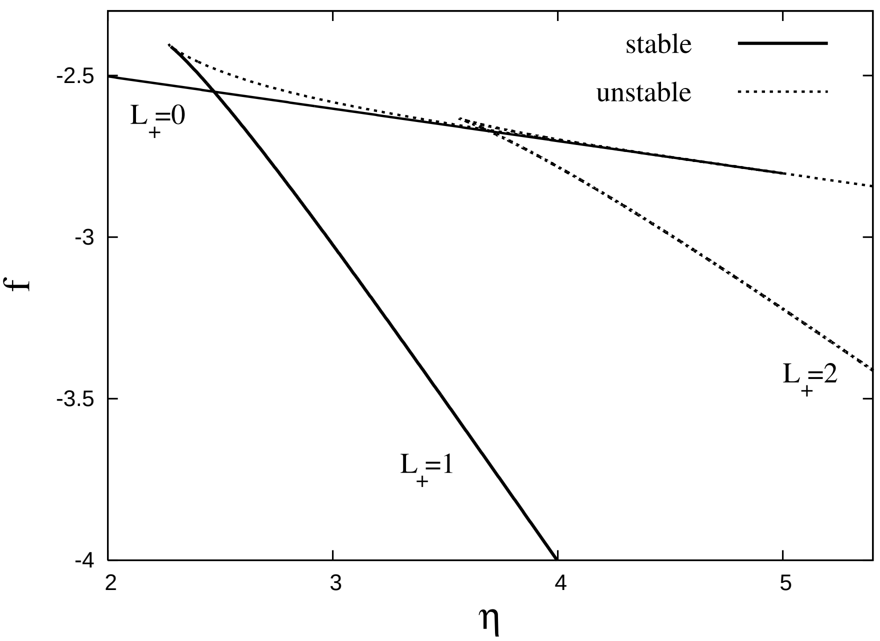

Figure 1: Solid lines trace the value of f at its local minima (cases $L_+ = 0, 1$), as a function of η, for $q = 10$. The intersection of these curves identify the point η^* given by (13) which separates the regions with two different kind of ultralong-run predictions. Dashed curves correspond to saddle points for the cases with $L_+ = 0, 1, 2$.

Let us now turn to Parts *(iv)-(v)* of Proposition 6, which concern the characterization of the *global* minima of the function f. Outside of the interval $[\check{\eta}, \hat{\eta}]$, the issue is simply settled by the characterization of the local minima, since these are unique (up to action relabelling). Instead, within that interval, one needs to evaluate the value of f for each local minima, which gives rise to the finding described in part *(iv)* of the result. It states that there is a threshold value η^* such that for $\eta < \eta^*$, the invariant distribution is dominated by the symmetric configuration, whereas for $\eta > \eta^*$ the asymmetric outcome (with $L_+ = 1$) prevails. Since, as we have explained, such global minima define the ultralong-run prediction of the model, we conclude that whether the symmetric or asymmetric configurations are selected in the ultralong run solely depends on whether η lies below or above the threshold η^* given in (13). In fact, as one can check by direct substitution of $\eta = \eta^*$ in (12), the fractional size of the dominant class in the asymmetric solution ($L_+ = 1$) takes the value $n_+ = 1 - 1/q$. This implies that as η comes to exceeding η^* the population frequencies experience a sharp upward discontinuity.

At this point, it is worth emphasizing that the solutions with $L_+ = 0$ and $L_+ = 1$ not only differ in the (a)symmetry of their action frequencies but, as importantly, in the characteristics of the corresponding networks. The solution with $L_+ = 0$ induces a sparsely connected network, which is in turn fragmented into many components of insignificant size. Instead, the solution with $L_+ = 1$ yields a much more connected network, where a giant component exists that encompasses a significant (i.e. nonvanishing) fraction of all nodes. These conclusions follow from a combination of the following observations:

(i) within each class r, the corresponding subnetwork is an Erdös-Rényi random graph (recall (6)) whose binomial degree distribution converges, for large N, to a Poisson distribution given by $\Pr\{k_i = k | \alpha_i = a_r\} = \frac{(2\eta n_r)^k}{k!} e^{-2\eta n_r}$ and average degree $2\eta n_r$;

(ii) the values for n_- and n_+ satisfy $2\eta n_- < 1 < 2\eta n_+$;

(iii) a giant component exists almost surely in Erdös-Rényi random graphs if, and only if – cf. Erdös and Rényi (1960) or Newman *et al.* (2001) – the average degree is larger than one, i.e. $2\eta n_r > 1$.

The above considerations entail that the sharp discontinuity exhibited by the utralong-run profile $\mathbf{n}$ at η^* is mirrored by a substantial change in the connectivity of the social network. That is, for $\eta < \eta^*$ the underlying social network displays relatively low connectivity and is highly fragmented, while for $\eta > \eta^*$ the social network enjoys a significantly higher connectivity and includes a giant component. This in turn shows that, in our model, the *ability to coordinate* and the *ability to connect* are two sides of the same coin.

To gain a stark manifestation of the former point, Figure 2 depicts the behavior of the average network degree z prevailing in the ultralong run, i.e. the corresponding expected number of neighbors of a given agent. This is the population average of the average degrees $z_r = 2\eta n_r$ for the different action classes, as given by

$$z = \sum n_r z_r = 2\eta \left[L_+ n_+^2 + (q - L_+) n_-^2 \right]. \tag{14}$$

We observe that z displays a sharp upward discontinuity when η turns higher than η^* and the system then shifts from a symmetric to an asymmetric configuration. Below η^*, the size distribution is uniform across classes and the common average degree is relatively low: $z = 2\eta/q < 1$, since $\eta < \hat{\eta}(\equiv q/2)$. Instead, when $\eta > \eta^*$, the induced average degree z experiences an upward shift that builds upon the rise of an action class r that becomes relatively large. Its larger size n_+ can then support an average degree $z_r = 2\eta n_+ > 1$, which in turn spans a giant component, as explained above. Finally, a further understanding of the situation is provided by Figure 3, which describes how the threshold η^* varies with q, this value being bracketed by the two thresholds – $\check{\eta}$ and $\hat{\eta}$ (also functions of q) – that define the boundaries of the parameter region where multiple solutions coexist.

[Insert Figures 2 and 3 about here]

Despite the sharp prediction afforded by the model in the infinite-time limit (i.e. the ultralong run), Proposition 6 indicates that, within the range $\eta \in (\check{\eta}, \hat{\eta})$, the global minima of the function f coexist with another local minima that, locally, should also tend to strongly "attract" the configurations in its vicinity when N is very large. Heuristically, as the population grows, the expected time T needed for the system to escape a neighborhood of a local minimum of f, say $\mathbf{n}^*_{\text{loc}}$, must grow exponentially in N. For, in essence, the most likely transitions away from $\mathbf{n}^*_{\text{loc}}$ must go through a certain saddle point $\mathbf{n}_{\text{saddle}}$ and thus[12]

$$T \propto e^{N[f(\mathbf{n}_{\text{saddle}}) - f(\mathbf{n}_{\text{loc}})]} = \frac{e^{-N[f(\mathbf{n}_{\text{loc}})]}}{e^{-N[f(\mathbf{n}_{\text{saddle}})]}} \tag{15}$$

i.e. T is of the same order as the ratio of probabilities assigned to $\mathbf{n}_{\text{saddle}}$ and $\mathbf{n}_{\text{loc}}$ by the invariant distribution. Therefore, even though the process will ultimately converge to a global minimum of f in the *ultralong run*, it may remain confined close to a local minimum for an arbitrarily long time for large N. So, within this time horizon, which we refer to as the *long run*, initial conditions may well matter.

As it turns out, the intuitive description of the situation just outlined is essentially correct and can be made precise in a number of different ways. In the next section, we choose to do so through the route afforded by the so-called mean-field analysis of the stochastic process.

[12] Our informal reasoning here is based on the fact that the induced process on the variable $\mathbf{n}$ generates almost continuous paths, for large populations. For a formal discussion of these matters, see Section 5 below.

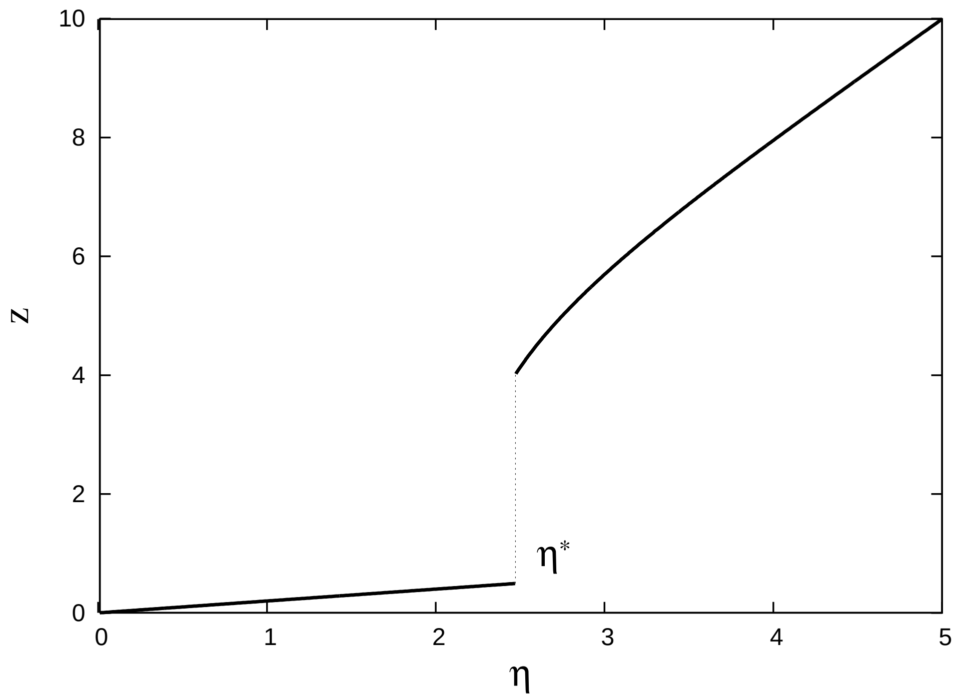

Figure 2: The average degree z prevailing in the ultralong run, as an increasing function of the rate η, for $q = 10$. The point of discontinuity arises at the value $\eta = \eta^*$ specified in (13).

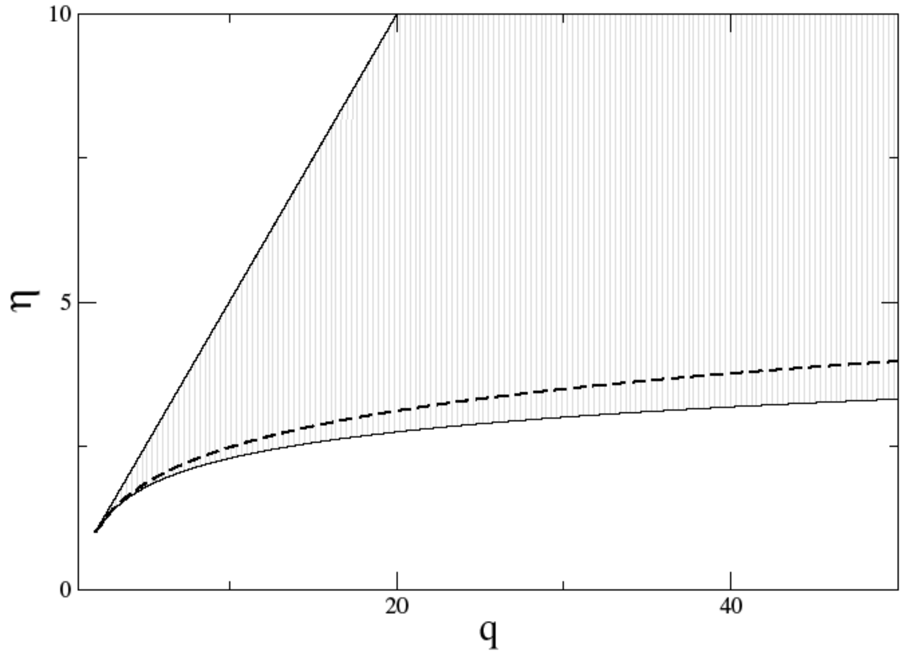

Figure 3: The value of η^* as a function of q (given by the dashed curve), which lies between the corresponding values of $\check{\eta}$ (lower solid curve) and $\hat{\eta} = q/2$ (upper solid curve) specified in Proposition 6.

5 The long run: a mean-field approach

As explained, we want to identify the long run behavior of the process as that which obtains (with high probability) for large but finite times. Formally, such long-run behavior may be accessed by taking *first* the limit of infinite population size ($N \to \infty$) and then focusing on the large-time behavior ($t \to \infty$). This allows one to trace the dynamics of population frequencies through a deterministic continuum-time dynamics: the Mean-Field Dynamics (MFD). In mathematical terms, the MFD is simply given by a system of ordinary differential equations that reflects *expected motion*. And, in essence, the assumption that justifies this approach (conceived as an approximation) is the assumption that the population is so large that, within *any finite* time horizon T under consideration, any finite-system randomness can be largely ignored.

5.1 Mean-field dynamics

The mean-field dynamics is defined on an augmented description of the population frequencies considered in Section 4, under the maintained assumption that the process has already abandoned transient states. Thus, relying for simplicity on analogous notation, the states of the system consist of the current vector of population *frequencies* $\mathbf{n} \equiv [(n_{r,k})_{k=0}^{\infty}]_{r=1}^{q}$ specifying the frequency $n_{r,k}$ of agents in the population that choose each possible action r and have every possible degree k. The set of all such vectors – also called *population states* here – will be denoted by Φ. Then, the *mean-field dynamics* (MFD) is defined as the continuous-time ODE on Φ whose induced paths satisfy, for each r and k, the equation

$$\dot{n}_{r,k} = \lim_{\Delta t \to 0} \frac{E[\Delta n_{r,k} | \; \mathbf{n}(t) = \mathbf{n}]}{\Delta t} \equiv F_{r,k}(\mathbf{n}) \tag{16}$$

where $E[\Delta n_{r,k}| \; \mathbf{n}(t) = \mathbf{n}]$ represents the conditional expected change in each $n_{r,k}$ during the infinitesimal time interval $(t, t+\Delta t)$. The form of this dynamics is provided by the following result.

Proposition 7 *The vector field $F(\mathbf{n}) \equiv [F_{r,k}(\mathbf{n})]_{r,k}$ of the mean-field dynamics is, up to order $\mathcal{O}(1/N)$, given by:*

$$F_{r,k}(\mathbf{n}) = 2\eta n_r n_{r,k-1} - 2\eta n_r n_{r,k} + (k+1) n_{r,k+1} - k n_{r,k} \quad (k > 0) \tag{17}$$

$$F_{r,0}(\mathbf{n}) = -2\eta n_r n_{r,0} + n_{r,1} + \nu \sum_{s=1}^{q} [n_{s,0} - n_{r,0}] \tag{18}$$

where

$$n_r \equiv \sum_k n_{r,k}. \tag{19}$$

Proof: See the Appendix.

It may be useful, at this point, to provide a concise overview of the derivation of each $F_{r,k}$ carried out in detail in the Appendix. First, consider expression

(17), which applies to all positive degrees k. In it, the terms proportional to η reflect the effect of the link creation subprocess on the population density, whereas all the other terms ("proportional" to $\lambda = 1$) embody the link decay subprocess. Either subprocess can increase or decrease each $n_{r,k}$, which is why each of them is responsible for two terms with opposite sign. In contrast, nodes with degree $k = 0$ (i.e. isolates) must be dealt with separately in expression (18) both because some of the terms considered before do not apply in this case (i.e. there are no nodes with degree $k = -1$), and also because there is an additional exchange subprocess that governs the bidirectional flows across different action classes $r \leftrightarrow s$. This subprocess occurs at the rate ν at which an (isolated) node receives an action revision opportunity.

It is intuitive that the MFD should represent a good (probabilistic) approximation of the behavior of the stochastic process for large N. Indeed, this can be rigorously confirmed by an adaptation of existing results from the modern literature on so-called stochastic approximation theory – see, specifically, the work by Benaïm and Weibull (2003*a*), which focuses on stochastic evolutionary dynamics.[13] The essential feature shared by our framework and theirs is that, whenever an adjustment event takes place, it can only involve (with full probability) a *bounded* change in the characteristics of a *finite* number of nodes. This allows one to reproduce all the basic steps in their analysis,[14] and claim that the MFD (16) approximates the original stochastic process in the following two ways:

1. Let $\{\mathbf{x}(t) \in \Phi\}_{t\geq 0}$ be the stochastic process induced by the model on the set of population states. Let $\{\mathbf{n}(t, \mathbf{x}_0)\}_{t\geq 0}$ be a solution of the ODE (16) with initial conditions $\mathbf{n}(0, \mathbf{x}_0) = \mathbf{x}_0$. Then, for any time horizon T, and any given $\varepsilon, \delta > 0$, there exists some lower bound $\hat{N}$ on population size such that if $N \geq \hat{N}$, then $\Pr\left[\max_{0\leq t\leq T} \|\mathbf{x}(t) - \mathbf{n}(t, \mathbf{x}(0))\| \geq \varepsilon\right] \leq \delta$, where $\|\cdot\|$ stands for the sup norm.

[13] Refer, in particular, to Section 6 of Benaïm and Weibull (2003*a*), in which they study stochastic processes where, as in the present case, the arrival of adjustment opportunities is governed by independent Poisson clocks. It must be mentioned, however, that the framework originally studied by Benaïm and Weibull presumed that the approximating vector field (i.e. the corresponding transition probabilities) can be defined independently of population size. But in many applications (the present one being an example), this is not the case. This issue has been addressed in a later paper by these same authors (Benaïm and Weibull (2003*b*)), where they generalize their previous analysis in this direction.

[14] There are, however, two differences. The first is a minor and inconsequential adaptation in the precise statement of their Lemma 2. In our case, the value of Γ becomes $\Gamma = (2 + \|F\|)$, which simply reflects that a link adjustment can affect four (rather than just two) dimensions of the state. A second difference concerns the fact that Benaïm and Weibull (2003*a*,*b*) presume that the vector field is finite-dimensional. In our case, this would require bounding the support of the degree distribution by some $\bar{k}$, arbitrarily large but finite. This could be done, for example, by considering a variant of the model where no agent can support more than $\bar{k}$ links, so that if a linking opportunity arrives to an agent with that many links it cannot be materialized. In practice, however, it is easy to see that our model (in particular, as it pertains the stationarity and stability of its equilibria) would be continuous in such a parameter $\bar{k}$ at infinity. For simplicity, therefore, we choose to maintain our assumption throughout that $\bar{k} = \infty$, thus dispensing with a parameter of minor interest.

2. Let $\mathbf{n}^*$ be an asymptotically stable state of the ODE (16). There exists some U, an open (Borel) subset of Φ that includes $\mathbf{n}^*$, such that for any time horizon T, and any given $\varepsilon > 0$, there is some lower bound $\hat{N}$ on population size such that, if $N \geq \hat{N}$, the random variable $\tau(U) \equiv \inf\{t \geq 0 : \mathbf{x}(0) \in U,\ \mathbf{x}(t) \notin U\}$ giving the first exit time from U satisfies $\Pr[\tau(U) \geq T] \geq 1 - \varepsilon$.

The preceding statements specify two related forms in which it can be formally argued that the mean-field dynamics is a good approximation of the stochastic process in the long run. Verbally, the first one asserts that the population path induced by the stochastic process and that resulting from the Mean-Field Dynamics (MFD) are arbitrarily close, for an arbitrarily long period of time and with an arbitrarily high probability, provided the population is large enough. The second statement, on the other hand, focuses on asymptotically stable states of the MFD and claims that any one of them is a robust long-run prediction in the following sense: whenever the process starts close to it, the ensuing path is very likely to remain also close for a very long period of time if the population is large. Both conclusions, 1 and 2, allow us to view the MFD as a suitable description of the process in the limit $N \to \infty$. This in turn is a reflection of the fact that, when the population is sufficiently large, the "ergodicity-inducing" noise displayed by the finite-population process can be largely ignored for large but finite times (i.e. in the long run) .

In view of the preceding discussion, our objective here will be to characterize the dynamic paths of (16). A first preliminary observation pertaining to its stationary points is contained in the following result.

Proposition 8 *Density profiles of the form*

$$n_{r,k} = n_r \frac{(2\eta n_r)^k}{k!} e^{-2\eta n_r}, \tag{20}$$

where $\mathbf{n} = (n_1, \ldots, n_q)$ *are solutions of the first order conditions for the minimization of* $f(\mathbf{n})$ *in (8), are stationary points of the MFD, i.e.* $F_{r,k}(\mathbf{n}) = 0$ *for all* $r = 1, \ldots, q$ *and* $k \geq 0$*.*

The proof proceeds by straightforward substitution, using (11). This proposition implies that the ultralong run configuration derived in the previous section is a stationary point of the dynamics. It does not, however, specify whether the dynamical paths converge to it or not. Indeed, in principle, any extreme or saddle point of the function f is a possible candidate for the asymptotic behavior of the MFD.

We will now turn to the issue of characterizing the dynamical paths of the MFD. To simplify matters, we choose to focus on the case where the drift among different actions undertaken by isolated agents proceeds at a much slower pace than the change of the network, so that both dynamics can be effectively decoupled. At a heuristic level, this can be motivated by the observation that such action changes are essentially payoff-irrelevant in the short run and thus (myopic) agents have a weak incentive to implement them. Formally, it will be captured by making $\nu \searrow 0$ – that is, by assuming ν is infinitesimally small but positive. As we shall explain below, numerical simulations show that this simplifying assumption does not affect the predictions of the model in any essential way.

Mathematically, the analysis of such a limit scenario may be carried out by introducing a slow time variable

$$\tau = \nu t \tag{21}$$

that tracks the evolution of the class sizes n_r. Then, by taking the limit $\nu \to 0$, we formalize the idea that the dynamics of link adjustment within each class r operates infinitely faster than the dynamics of its corresponding size. To see this, note (by directly taking the sum of (16) over $k \geq 0$) that the evolution of each aggregate frequency n_r is given by

$$\frac{dn_r}{dt} = \nu \sum_{s=1}^{q} \left[n_{s,0} - n_{r,0} \right], \tag{22}$$

which is proportional to ν and thus becomes very slow as $\nu \to 0$. In other words, the class size $n_r(t) = \bar{n}_r(\tau)$ is in effect a function of the slow time variable. In contrast, the dynamics of each constituent $n_{r,k}$ in (17)-(18) has terms which do not vanish as $\nu \to 0$. We also note that if one neglects terms proportional to ν in (17)-(18), populations evolve in an independent fashion within each action class $r = 1, \ldots, q$.

The previous discussion indicates that, when ν is infinitesimally small, the (fast) network dynamics for given class sizes and the (slow) dynamics of class sizes can be studied separately. This is what we respectively do in the next two subsections.

5.2 The network dynamics

In this section, we posit that class sizes evolve slowly as specified by some given functions $n_r(t)$, and then find a solution of the ODE (16) under the following conditions:

$$n_{r,k}(t=0) = n_{r,k}^{(0)} \quad \forall\, k \geq 0 \tag{23}$$

$$\sum_{k \geq 0} n_{r,k}(t) = n_r(t). \tag{24}$$

To carry out the analysis, it is convenient to describe the degree distribution prevailing within each class r through its Poisson representation given by

$$n_{r,k}(t) = \int_0^\infty dx \frac{x^k}{k!} e^{-x} f_r(x,t), \qquad \forall k \geq 0, \qquad (r = 1, \ldots, q) \tag{25}$$

where f_r is a *generalized function*[15] defined over the positive reals, for all $t \geq 0$. The transformation from $f_r(x,t)$ to $n_{r,k}(t)$ leading to this representation is invertible.[16] What makes it particularly useful is that a system of infinitely many coupled ODE's can then be turned into a single partial differential equation, as indicated in the following result.

Proposition 9 *Let $\phi_r(x)$ be such that $n^{(0)}_{r,k} = \int_0^\infty \frac{x^k}{k!} e^{-x} \phi_r(x)$ for all $k \geq 0$. Let $f_r(x,t)$, with $t \geq 0$, be a solution of the partial differential equation (PDE)*

$$\frac{\partial}{\partial t} f_r = \frac{\partial}{\partial x} \left[(x - 2\eta n_r) f_r \right] \tag{28}$$

with initial condition $f_r(x,0) = \phi_r(x)$ and such that

$$\int_0^\infty dx f_r(x,t) = n_r(t), \qquad \forall t. \tag{29}$$

Then $n_{r,k}(t)$ given by (25) is a solution of the system of ODE (16) with initial conditions given by (23).

Proof: See the Appendix.

The PDE (28) is solved with the method of characteristics. We look, that is, for a function $\xi_r(t)$ such that along the *characteristic* trajectories $(x,t) =$

[15] Generalized functions are discussed in detail in Kolmogorov and Fomin (1999, Ch. IV). In brief, the function $f_r(x,t)$ does not have the standard meaning of mapping a real value x to a real value $f_r(x,t)$. Rather, it defines a linear transformation through the integral in (25) that associates a number $n_{r,k}(t)$ to functions of the form $x^k e^{-x}/k!$. Hence, the action of $f_r(x,t)$ is given by its behavior inside an integral. By way of illustration, consider the distribution with $n_{r,k} = 1$ if $k = h$ and $n_{r,k} = 0$ otherwise. This case corresponds to the (generalized) function $f_r(x,t) = (-1)^h e^x \frac{d^h}{dx^h} \delta(x)$, where $\delta(x)$ is the Dirac's delta function.

[16] A constructive way to achieve the inverse transformation, i.e. to derive $f_r(x,t)$ from $n_{r,k}(t)$, is through the expansion of $f_r(x,t)$ in orthogonal polynomials – see Kolmogorov and Fomin (1999, Ch. IV). The appropriate polynomials, for functions defined for $x \in [0,\infty)$, are Laguerre polynomials $L_m(x) = \sum_{k=0}^m \ell_{m,k} x^k$, with $\ell_{m,k} = \binom{m}{k} \frac{(-1)^k}{k!}$. The expansion reads

$$f_r(x,t) = \sum_{m=0}^\infty c_{m,r}(t) L_m(x). \tag{26}$$

In order to derive the coefficients $c_{m,r}(t)$, multiply (25) by $k!\ell_{n,k}$ and sum over $k = 0,\ldots,n$. Using (26) and the orthonormality of Laguerre polynomials, i.e. $\int_0^\infty dx e^{-x} L_m(x) L_n(x) = \delta_{n,m}$, we find

$$\sum_{k=0}^m k!\ell_{k,m} n_{k,r}(t) = \int_0^\infty dx e^{-x} L_m(x) f_r(x,t) = c_{r,m}(t). \tag{27}$$

A combination of the coefficients $c_{r,m}(t)$ computed in this manner and (26) yields $f_r(x,t)$ in terms of $n_{k,r}(t)$, which is the desired inverse transformation. In particular, this procedure allows one to compute the initial condition $f_r(x,0)$ from the initial class profile $n_{r,k}(t=0)$. As an illustrative example, for initial conditions of the form $n_{r,k}(t=0) = a_r(1-p_r)^k$, (27) yields the coefficients $c_{r,m}(0) = ap^m$. Then, by (26) and some algebra, one finds $f_r(x,0) = \frac{a}{1-p} e^{-px/(1-p)}$.

$(\xi_r(t), t)$ the evolution of f_r is described by the simple ODE :

$$\frac{d}{dt} f_r(\xi_r(t), t) - f_r(\xi_r(t), t) = \frac{\partial}{\partial t} f_r - (x - 2\eta n_r) \frac{\partial}{\partial x} f_r - f_r = 0. \tag{30}$$

The solution thus obtained can then be used to solve for the degree distribution in each class r, under the assumption that the speed at which the class sizes $n_r(t)$ change is much lower than that for network adjustment – i.e., specifically, we consider the limit $\nu \to 0$ with $\tau = \nu t$ finite. The derivations involved in the whole procedure are described in the Appendix, as part of the proof leading to the following result.

Proposition 10 *Let $\tau = \nu t > 0$ be given and let $\bar{n}_r(\tau) = n_r(\tau/\nu)$. Then, in the limit $\nu \to 0$, the degree distribution induced by (16) converges in each component to the Poisson distribution with mean $2\eta\bar{n}_r(\tau)$, i.e.*

$$\lim_{\nu \to 0} n_{r,k}(t = \tau/\nu) = \bar{n}_r(\tau) \frac{[2\eta\bar{n}_r(\tau)]^k}{k!} e^{-2\eta\bar{n}_r(\tau)} \tag{31}$$

The previous result indicates that the network topology eventually converges to that of a Poisson random network if isolated agents change their action very slowly and thus class sizes also change slowly. This, in essence, is the MFD counterpart of Corollary 4 obtained in the ultralong run, since the degree distributions of Erdös-Rényi random graphs converge to a Poisson law for large N.

5.3 The dynamics of class sizes

Now we turn to studying the evolution of class sizes, still under the simplifying assumption that their change proceeds at a much slower rate than that at which the network adjusts. Then, we can rewrite the dynamics of the population sizes $\bar{n}_r(\tau) = n_r(t = \tau/\nu)$ in the slow time variable as follows:

$$\frac{d\bar{n}_r}{d\tau} = \sum_{s=1}^{q} [n_{s,0} - n_{r,0}] \tag{32}$$

and, by virtue of the results of the previous subsection, use $n_{r,0} \cong \bar{n}_r(\tau)e^{-2\bar{n}_r(\tau)}$ as an accurate approximation for small ν. This will allow us to establish quite readily that the function $f(\cdot)$ used in Subsection 4.3 to characterize the invariant distribution of the underlying stochastic process acts as a Lyapunov function for the MFD. That is, the value of this function decreases monotonically along any trajectory of the dynamical system. This, in turn, implies that the local minima of this function define configurations that are locally stable, i.e. all trajectories that start close enough to any one of them eventually converge to it. Formal statements of these two conclusions follow.

Proposition 11 *Let* $\tau = \nu t$ *and consider any induced path on class sizes* $(\bar{n}_1(\tau), ..., \bar{n}_q(\tau))$ *obtained as* $\nu \to 0$. *Then, the function* f *defined in Proposition 5 satisfies* $df(\bar{n}_1(\tau), ..., \bar{n}_q(\tau))/d\tau \leq 0$, *where it may hold with equality only if* $(\bar{n}_1(\tau), ..., \bar{n}_q(\tau))$ *is a critical point of* f *(i.e. satisfies the FOC (11)).*

Corollary 12 *Under the conditions considered in Proposition 11, the local minima of the function* f *are asymptotically stable configurations of the MFD.*

The above corollary implies that the process can be predicted to spend an arbitrary long time[17] in any small neighborhood of a local minima of the function f with high probability. In this sense, we can view each of the local minima of f as an alternative long-run prediction if the process starts close to it. This provides a formal basis for our former heuristic comparison of *local* versus *global* minima of the function f as a reflection of the dichotomy between the long- and ultralong-run outcomes.

To illustrate sharply the contrast induced by these two time horizons, it is again useful to focus on how the alternative possibilities translate into different levels of overall connectivity. First, let us consider the *long-run* average degree predicted by the model, through the asymptotically stable configurations of the MFD. At any such configuration $\mathbf{n}$ we can compute the average degree z as in (14), relying on the fact that the stable equilibria of the MFD yield aggregate configurations that have exactly the same structure as the minima of the function f. Thus, associated to the thresholds $\check{\eta}$ and $\hat{\eta}$ contemplated in Proposition 6, the model predicts three different regions in η where qualitatively very different long-run levels of network connectivity can materialize. The situation is illustrated in Figure 4, which depicts the situation for $q = 10$ and can be usefully compared with Figure 2 . In essence, the difference here is that the upper and lower sections of the curve in Figure 2 are "stretched" in Figure 4, extending the curve to the left and right respectively of the former discontinuity at the threshold η^*. This, in turn, opens up the parameter region $(\check{\eta}, \hat{\eta})$ $(= (2.28, 5)$ for $q = 10)$ where long-run multiplicity arises, allowing in turn for two significantly different levels of network connectivity. The same conclusions, of course, could have been derived in Subsection 4.3 from the analysis of the local minima of the function f (or, equivalently, of the local maxima of the invariant distribution).

[Insert Figure 4 about here]

Figure 4 also depicts the outcome of numerical simulations, which were conducted for a population size $N = 1000$, the indicated value of $q = 10$, and two values of $\nu = 1, 10$. We find that the asymptotically stable configurations singled

[17] As suggested in (15), the expected time spent in the neighborhood of a local minimum $T \sim e^{cN}$ grows exponentially with population size N, where $c = f(\mathbf{n}_{\text{saddle}}) - f(\mathbf{n}_{\text{loc}}) > 0$ is given by the difference between the values of the function f at a suitable saddle point and the local minimum in question. See Benaïm and Weibull (2003a) for a detailed and rigorous study of this issue.

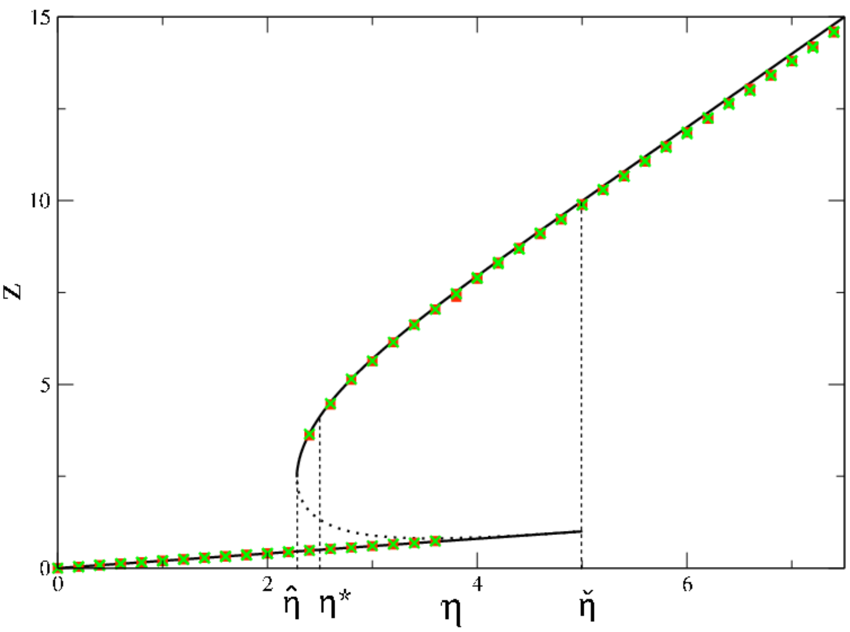

Figure 4: The solid curve represents the mean connectivity induced by the theoretical mean-field model (as given by (14)) against the rate of link formation, for a value of $q = 10$. Symbols, on the other hand, stand for observations obtained by numerical simulation. The simulations were conducted for two different values $\tilde{\nu}$ – i.e. $\tilde{\nu} = 1$ (squares) and $\tilde{\nu} = 10$ (crosses). They are performed by implementing a series of gradual changes in η (starting from both high and low values for it), letting the system equilibrate at each stage before proceeding to record the situation. In the simulations, the low equilibrium becomes unstable below the theoretically predicted value of 5 because fluctuations have significant effects close to the transition point due to the finiteness of the population. Such finiteness also explains the slight downward deviations observed along the upper branch for high values of η.

out by the mean-field theory indeed act as strong local attractors of the simulations. It is interesting to observe that only a moderately large population size is sufficient to achieve a good predictive performance of the theoretical model. We also find that the predictions are essentially unaffected by the specific values of ν considered, even though these values differ in an order of magnitude. The latter is in consonance with the minor role played by this parameter in the core of our analysis – e.g. in the determination of the invariant distribution of the underlying stochastic process, which is independent of ν (cf. (4)).

In sum, we find that the long-run behavior of the system displays many of the key features that were already highlighted in our analysis of the ultralong run. Namely, there is a positive cross-reinforcing effect between (endogenous) action homogeneity and network connectivity, and both increase – at some point discontinuously – as the rate of link creation η rises. Important differences, however, transpire when such discontinuous transitions are studied in each of the two time scales. In the ultralong run, the transitions are reversible around the *single* threshold η^*. This, as we have seen, is a mere reflection of the fact that, in the infinite-time limit, the *ergodic* behavior of the process is independent of initial conditions.

Instead, within the shorter time scale that we have called the long run, the situation is qualitatively different. For, within the parameter range limited by the *two* thresholds $\check{\eta}$ and $\hat{\eta}$, the system can settle in either a high- or a low-connectivity configuration, depending on the initial conditions. This implies, in particular, that the sizable changes triggered by gradual changes in η around those thresholds are *not* reversible in the long run. The system, therefore, can be said to display hysteresis in such a time scale. And, as advanced, this opens up a host of interesting policy issues which we can only briefly illustrate here.[18]

Think, for example, of the rate η as responding to some policy measures that, say, enhance the agents' ability or desire to meet each other.[19] Then, any such measure that maintains the value of η above $\check{\eta}$ (possibly, just slightly so) during a relatively brief span time will have consequences that persist during the time horizon we have labelled the long run. This is the case even if the base (intervention-free) value of η were below η^*. In this case, the policy consequences of such a short-run policy will be unraveled in the ultralong run but may nevertheless bear fruit for what may be (if the population is large) a long time. And in case the base value of η lies above η^*, such a policy can be seen as an effective way of speeding up the transition to the superior ultralong-run situation.

[18] To make our ensuing discussion fully precise, we would need to build on the analysis undertaken in this paper to obtain a specific assessment of the average time lengths displayed by the long-run behavior (as a function of population size) as well as a determination of the average time spans required for the transitions to long-run equilibria to materialize from any point in their basin of attractions. Due to space limitations, however, we choose to keep the discussion at an intuitive level.

[19] To fix ideas even further, some of the progams established by the European Union to finance and stimulate the visits, exchanges, or workshops among European researchers based in distant locations (geographically or/and culturally) is a good case in point.

6 Summary and conclusion

In this paper we have studied the evolution of large and networked social systems whose inter-agent connections are subject to potentially high volatility. We have argued that such volatility is important in many social contexts, and also calls for a methodological approach different from that pursued by the recent evolutionary literature in economics. To highlight the main issues involved, our model embodies a stylized description of three forces that appear to be at work in many social networks:

(a) the establishment of links depends on some suitable form of behavioral affinity or compatibility;
(b) existing links decay over time and this process limits the extent of social connectivity that can be sustained;
(c) links among individuals tend to coordinate their actions, while social isolation leads to behavioral drift.

The asymptotic dynamics of the induced ergodic process has been studied within different time horizons. First, we considered the infinite-time limit (what is sometimes labelled the ultralong run), where the evolution of the system is captured by the unique invariant distribution. This distribution was fully characterized, in turn delivering many of the properties of interest of the model, e.g. the extent of behavioral coordination and the degree of connectivity of the underlying social network. We found, in particular, that there is a reciprocal feedback between coordination and connectivity that induces a sharp (discontinuous) transition in both respects at a certain volatility threshold. Thus, in our model, a society with sparse bilateral interactions does not turn continuously into a densely networked one as volatility falls, but this change occurs discontinuously.

Our analysis of the invariant distribution also suggested the existence of a different time frame for the analysis of the process, which we have called the long run. It refers to long but finite time horizons where, if the population is large enough, the evolution of the process can be well approximated by a deterministic (so-called mean-field) dynamics. In this shorter perspective, there is a certain range for volatility where the initial conditions largely shape the evolution of the process, thus allowing for multiple possible outcomes tailored to history. Again in this case we have discontinuous transitions triggered by slight changes in volatility, but these now display local irreversibility (i.e. hysteresis).

The main contribution of our model and analysis is two-fold. First, at a *conceptual level*, it highlights the role played by volatility in economic contexts and studies its implications in a social-network setup where there is a strong interplay between network formation and behavioral adjustment. Second, at a *methodological level*, it illustrates the use of mathematical techniques that allow an exhaustive analysis of the induced dynamics and provide an integrated analysis of the different time scales inherent to the process.

Despite the abstract nature of our model, we believe that much of what is learned from it can be applied to the study of high-volatility phenomena in a rich variety of socio-economic contexts. This pertains, in particular, to the interesting range of problems (such as inter-firm alliances or academic collaboration, discussed in Section 2) where some versions of (a)-(c) above appear to be key mechanisms at work. The model also delivers some insights that are policy relevant if one is interested in, say, raising the level of network-based activity in those cases. Specifically, we have highlighted that small and temporary policies can have quite persistent effects by triggering the self-reinforcing and stabilizing effects that underlie network formation.

Appendix

Proof of Proposition 1: The argument is decomposed in two claims. First, we argue that the subset $\hat{\Omega} \equiv \{\omega = (\boldsymbol{\alpha}, G) \in \Omega : \forall i, j \in P,\ [g_{ij} = 1 \Rightarrow \alpha_i = \alpha_j]\}$ is included in a single recurrent class. Second, we show that all states $\omega \notin \hat{\Omega}$ are transient. The combination of these two claims implies that the Markov process has a unique invariant distribution μ with $\mu(\hat{\Omega}) = 1$.

To establish the first claim, it is enough to prove that, for any two states $\omega, \omega' \in \hat{\Omega}$, it is possible to find a finite sequence of transitions (jointly occurring with some positive probability, bounded away from zero) that leads from one to the other in finite time. To construct it, note that there is positive probability for an initial sequence of transitions from the original state ω to some other state $\tilde{\omega}$ that consists of the empty network (i.e. with no links) and an action profile where every agents chooses the same action as in ω'. Indeed, such a transition takes place if all links in ω vanish in sequence and then all nodes receive (also in sequence) an action-revision opportunity and switch to the action they choose in ω'. Next, a sequence of transitions from $\tilde{\omega}$ to ω' can also occur with positive probability since $\omega' \in \hat{\Omega}$. Specifically, starting from the state $\tilde{\omega}$, all links present in ω' can be simply added sequentially through a suitable chain of link-creation opportunities. By construction, these links will be formed since the agents involved at each step display the same action. We conclude, therefore, that the desired transition from ω to ω' exists, as desired.

Now we show that all states in $\Omega \backslash \hat{\Omega}$ are transient. On the one hand, a straightforward variation of the previous argument indicates that, starting at any given $\omega \notin \hat{\Omega}$, the process must enter the set $\hat{\Omega}$ in some finite time with probability bounded above zero. Thus, to complete the argument, we simply need to show that from any state $\omega' = ((\alpha'_i)_{i \in P}, (g'_{ij})_{i,j \in P}) \in \hat{\Omega}$, any one-step transition leads to a state $\omega'' \in \hat{\Omega}$ with probability one. We need to consider only six possibilities:

i) The transition involves a link-creation opportunity between some i and j, the link ij is not in place (i.e. $g'_{ij} = 0$), and $\alpha'_i \neq \alpha'_j$. In this case, the link ij will not be formed and no change in the state is produced, i.e. $\omega'' = \omega' \in \hat{\Omega}$.

ii) The transition involves a link-creation opportunity between some i and j, the link ij is not in place (i.e. $g'_{ij} = 0$), and $\alpha'_i = \alpha'_j$. In this case, the link ij is formed and the new state $\omega'' \in \hat{\Omega}$.

iii) The transition involves a link-creation opportunity between some i and j and the link ij is in place (i.e. $g'_{ij} = 1$). In this case, the link ij is maintained and no change in the state is produced, i.e. $\omega'' = \omega' \in \hat{\Omega}$.

iv) The transition involves an action-revision opportunity for some i and this node has some links. Since $\omega' \in \hat{\Omega}$, we know that $\alpha'_j = \alpha'_i$ for all j such that $g'_{ij} = 1$. Thus, by the conditions required from action revision, we must have $\alpha''_i = \alpha'_i$ and again no change in the state is produced, i.e. $\omega'' = \omega' \in \hat{\Omega}$.

v) The transition involves an action-revision opportunity for some i and this node has no links. Then, independently of whether $\alpha''_i = \alpha'_i$ or $\alpha''_i \neq \alpha'_i$ (both occurring with positive probability) we have $\omega'' \in \hat{\Omega}$ since no new links are formed.

vi) The transition involves the destruction of an existing link. Given that all links in ω'' also exist in ω' and the action profile is the same in both cases (i.e. $\alpha' = \alpha''$), it follows that $\omega'' \in \hat{\Omega}$.

Since all six possibilities lead to some $\omega'' \in \hat{\Omega}$, the desired conclusion follows. This establishes the second claim and thus completes the proof of the proposition. ■

Proof of Proposition 2: Clearly, a sufficient condition for the stationarity of the distribution μ defined in (4) is given by the following "detailed-balance" equalities (see Gardiner (2004)):

$$\mu(\omega)\ \rho(\omega \to \omega') = \mu(\omega')\ \rho(\omega' \to \omega) \qquad (\omega, \omega' \in \Omega). \tag{33}$$

To verify the above equalities, first notice that, for any pair of states ω', ω, either $\rho(\omega' \to \omega)$ and $\rho(\omega \to \omega')$ are both zero, or they are both non-zero. Hence we only need to check (33) for states ω, ω' across which a direct transition is possible with positive probability, i.e. with $\rho(\omega \to \omega') > 0$. These possible transitions must fall into two categories: link adjustment alone (with actions remaining fixed), or action adjustment alone (with links unchanged). We address each of them in turn.

For the first case (link adjustment), we need to consider any two states, $\omega = (\boldsymbol{\alpha}, G)$ and $\omega' = (\boldsymbol{\alpha}', G')$, such that $\boldsymbol{\alpha} = \boldsymbol{\alpha}'$ and $g_{ij} = g'_{ij}$ for all $i, j \in P$ except for one pair, k and ℓ, such that $\alpha_k = \alpha_\ell$ and, say, $g_{k\ell} = 1$ but $g'_{k\ell} = 0$. Then, from (4), we have:

$$\frac{\mu(\omega)}{\mu(\omega')} = \frac{2\eta}{N-1}.$$

On the other hand, note that the rate $\rho(\omega' \to \omega) = 2\eta/(N-1)$ since a link-creation opportunity arrives to either k or ℓ at the rate 2η and the link $k\ell$ is actually created if the agent who receives the opportunity (either k or l) meets the other one – an event with probability $1/(N-1)$. On the other hand, the rate $\rho(\omega' \to \omega) = 1$, since each existing link vanishes at the rate $\lambda = 1$. Combining these considerations, we arrive at the following conclusion:

$$\frac{\mu(\omega)}{\mu(\omega')} = \frac{2\eta}{N-1} = \frac{\rho(\omega' \to \omega)}{\rho(\omega \to \omega')},$$

which obviously implies the detailed-balance condition (33).

For action revision, we need to consider any two states, $\omega = (\boldsymbol{\alpha}, G)$ and $\omega' = (\boldsymbol{\alpha}', G')$, such that $G = G'$ and, for some $k \in P$, $\alpha_i = \alpha_i'$ for all $i \neq k$ whereas $\alpha_k \neq \alpha_k'$. The rate of this transition is non-zero only if agent k has no links ($g_{ik} = 0$ for all $i \in P/\{k\}$) and, in this case, $\rho(\omega \to \omega') = \rho(\omega' \to \omega) = \nu$. Clearly the distribution μ defined in (4) satisfies (33) for pairs of states of this type.

The previous considerations establish that the probability distribution μ is invariant. On the other hand, ergodicity implies that it is the unique such invariant distribution, which completes the proof. ■

Proof of Proposition 3: Given any state ω, denote by $\mathbf{P}(\omega) = (P_1(\omega), ..., P_q(\omega))$ the induced partition of agents in action classes, so that $P_r(\omega) \equiv \{i \in N : \alpha_i = a_r\}$ for each $r = 1, 2, ..., q$. A first preliminary observation is that, for every $\omega \in \hat{\Omega}$, we can simply write:

$$\mu(\omega) = \mu_{g=0} \prod_{r=1}^{q} \prod_{i,j \in P_r(\omega),\, i<j} \left(\frac{2\eta}{N-1} \right)^{g_{ij}}.$$

Next, we compute, for any given $\mathbf{P}$,

$$\begin{aligned}
\mu(\hat{\Omega}(\mathbf{P})) &= \sum_{\omega \in \hat{\Omega}: \mathbf{P}(\omega) = \mathbf{P}} \mu(\omega) \\
&= \mu_{g=0} \prod_{r=1}^{q} \left\{ \prod_{i,j \in P_r(\omega),\ i<j} \left[\sum_{g_{ij}=0,1} \left(\frac{2\eta}{N-1} \right)^{g_{ij}} \right] \right\} \\
&= \mu_{g=0} \prod_{r=1}^{q} \left\{ \prod_{i,j \in P_r(\omega),\ i<j} \left(1 + \frac{2\eta}{N-1} \right) \right\} \\
&= \mu_{g=0} \prod_{r=1}^{q} \left\{ \left(1 + \frac{2\eta}{N-1} \right)^{\frac{1}{2}|P_r|(|P_r|-1)} \right\}.
\end{aligned}$$

And, finally, adding over all partitions $\mathbf{P}$ that are consistent with each given $\mathbf{N}$, we obtain:

$$\mu(\hat{\Omega}(\mathbf{N})) = \mu_{g=0} \sum_{\{P:\ |P_r| = N_r\ (r=1,..,q)\}} \left\{ \prod_{r=1}^{q} \left[\left(1 + \frac{2\eta}{N-1} \right)^{\frac{1}{2}N_r(N_r-1)} \right] \right\}$$

which is readily seen to be equal to (5), as desired. ■

Proof of Corollary 4: A fixed partition $\mathbf{P}$ of agents corresponds to a specific action profile $\boldsymbol{\alpha}$ for agents and thus can be uniquely associated to a subset of non-transient states consistent with it that is given by $\hat{\Omega}(\mathbf{P}) \equiv \{\omega = (\boldsymbol{\alpha}, G) \in \hat{\Omega} : \{i \in N : \alpha_i = a_r\} = P_r,\ r = 1, ..., q\}$. In view of Proposition 1, the social network G prevailing at any given state $\omega = (\boldsymbol{\alpha}, G) \in \hat{\Omega}(\mathbf{P})$ can be decomposed into q disjoint subnetworks $(G_1, \ldots, G_q)$, with the nodes in each $G_r = (g_{r,ij})_{i,j \in P_r}$ belonging to the respective class P_r. This shows item *(i)* in the statement of the result. In order to show *(ii)*, write any state $\omega \in \hat{\Omega}(\mathbf{P})$ as $\omega = [(\alpha_i)_{i \in P_r}, G_r]_{r=1}^q$ and note that such a state can be equivalently identified by the q disjoint networks $G = (G_1, \ldots, G_q)$ induced by it. Thus, with slight abuse of notation, the invariant distribution μ can be defined on the set of all such G and, correspondingly, for any particular $\hat{G}_r$, specify the marginal probability

$$\mu(\hat{G}_r) = \mu(\{G = (G_1, ..., G_q) : G_r = \hat{G}_r\}).$$

Then, it is immediate from the form of μ specified in (4) that, for any collection of class-based networks $G = (G_1, ..., G_q)$, we have:

$$\mu(G) = \prod_{r=1}^{q} \mu(G_r)$$

so that the invariant distribution factorizes the partition $\mathbf{P}$ and, therefore, the corresponding distributions induced on each action class are stochastically independent. More specifically, from (4) we find that, for any given class network G_r $(r = 1, 2, ..., q)$,

$$\mu\left(G\right) = \frac{1}{Z_r} \prod_{i<j \in P_r} \left(\frac{2\eta}{N-1}\right)^{g_{r,ij}} \tag{34}$$

with

$$Z_r = \sum_{g^{(r)}} \prod_{i<j \in P_r} \left(\frac{2\eta}{N-1}\right)^{g_{r,ij}} = \prod_{i<j \in P_r} \sum_{g_{ij}^{(r)}=0,1} \left(\frac{2\eta}{N-1}\right)^{g_{r,ij}} \tag{35}$$

being a normalization constant. This factorized form over the different possible links ij between nodes in P_r corresponds precisely to the definition of the Erdös-Rényi random network $G(N_r, p_r)$ (see Bollobás 2001), in which a graph of N_r nodes is built by drawing independently each possible link with probability p_r satisfying

$$\frac{p_r}{1-p_r} = \frac{2\eta}{N-1}$$

or $p_r = 2\eta/(N-1+2\eta)$. ■

Proof of Lemma 5: We rely on repeated applications of Stirling's formula ($k! \simeq (k/e)^k$ for large k) to write:

$$\begin{aligned} \log \mu(\hat{\Omega}(\mathbf{N})) &\simeq \log \mu_{g=0} - \sum_{r=1}^{q} \left[N_r \log(N_r/N) - \frac{N_r(N_r-1)}{2} \log\left(1 + \frac{2\eta}{N-1}\right)\right] \\ &\simeq \log \mu_{g=0} - N\left[f(\mathbf{N}/N) + O(1/N)\right]. \end{aligned} \tag{36}$$

where f is defined in (8). The proof is completed by showing that $N^{-1}\log\mu_{g=0}\to f_0$, for some finite constant f_0. This is done by using the normalization identity

$$-\log\mu_{g=0}=\log\left\{\sum_{\mathbf{N}}\frac{N!}{\prod_{r=1}^{q}N_r!}\prod_{r=1}^{q}\left[\left(1+\frac{2\eta}{N-1}\right)^{\frac{1}{2}N_r(N_r-1)}\right]\right\}\tag{37}$$

where the sum runs on all sets of positive integers $\mathbf{N}=(N_1,\ldots,N_q)$ such that $N_1+N_2+\ldots+N_q=N$. We easily derive a lower bound $-\log\mu_{g=0}\geq N\log q$ by setting $\eta=0$ and observing that the sum becomes the multinomial expansion of q^N. For the upper bound, we use $N_r(N_r-1)\leq N(N-1)$ in the exponent of (37) so that $-\log\mu_{g=0}\leq\frac{qN(N-1)}{2}\log\left(1+2\eta/(N-1)\right)+N\log q$. Using $\log(1+x)\leq x$ we finally arrive at

$$\log q\leq-\frac{1}{N}\log\mu_{g=0}\leq q\eta+\log q,$$

which implies that f_0 is finite, as claimed. ■

Proof of Proposition 6: Let $\mathbf{n}^*$ be a configuration that satisfies the FOC (11). In order for it to be a minimum of f, we need to check that f increases along all directions on the simplex Δ^{q-1} around $\mathbf{n}^*$ (second-order conditions). For an infinitesimal perturbations $\mathbf{n}=\mathbf{n}^*+\boldsymbol{\varepsilon}\in\Delta^{q-1}$, the vector $\boldsymbol{\varepsilon}=(\varepsilon_1,...,\varepsilon_q)\in\mathbb{R}^q$ satisfies $\sum_{r=1}^{q}\varepsilon_r\equiv 0$, and the change in the value of f to leading order is:

$$f(\mathbf{n})\simeq f(\mathbf{n}^*)+\frac{1}{2}\sum_{r=1}^{q}\frac{1-2\eta n_r^*}{n_r^*}\varepsilon_r^2+O(\boldsymbol{\varepsilon}^3).\tag{38}$$

A sufficient condition for $\mathbf{n}^*$ to be a minimum is that $f(\mathbf{n})-f(\mathbf{n}^*)\geq 0$ for all possible vectors $\boldsymbol{\varepsilon}$ which satisfy $\sum_{r=1}^{q}\varepsilon_r\equiv 0$.

Let us divide our discussion into three different cases: $L_+=0$, $L_+=1$, $L_+>1$. We address each of them in turn.

In the case $L_+=0$, by symmetry, we must have $n_r^*=n_-=1/q$ for all $r=1,...,q$. However, the second term in (38) is positive only for $\eta\leq q/2$. Hence, the symmetric ($L_+=0$) solution satisfy the second order conditions only for $\eta\leq\hat{\eta}\equiv q/2$, which proves *(i)* and part of *(v)*.

Next, we consider the case with $L_+=1$. Let $n_1=n_+$ (i.e. the larger class adopts action a_1) and $n_r=n_-<n_+$ for $r>1$. Notice that, in view of the FOC (11), $2\eta n_-<1<2\eta n_+$. The first of these inequalities ($2\eta n_-<1$) implies that, for all vectors $\boldsymbol{\varepsilon}$ with $\epsilon_1=0$ (and $\sum_{r>1}\varepsilon_r=0$), $f(\mathbf{n}^*+\boldsymbol{\varepsilon})-f(\mathbf{n}^*)>0$. The remaining linearly independent perturbations $\boldsymbol{\varepsilon}$ have $\varepsilon_1\neq 0$ and $\varepsilon_r=-\varepsilon_1/(q-1)$ for $r>1$, which are indeed orthogonal to all vectors $\boldsymbol{\varepsilon}=(0,\varepsilon_2,\ldots,\varepsilon_q)$, with $\sum_{r>1}\varepsilon_r=0$. For these perturbations, the variation of f can be seen to take the form

$$\frac{1}{2}\sum_{r=1}^{q}\frac{1-2\eta n_r^*}{n_r^*}\varepsilon_r^2 = \left[\frac{1}{n_+(1-n_+)}-\frac{2\eta q}{q-1}\right]\varepsilon_1^2.$$

$$= 2\frac{qn_+-1}{q-1}\left(\frac{dn_+}{d\eta}\right)^{-1}\varepsilon_1^2. \tag{39}$$

The last expression is confirmed by taking the derivative of (12) with respect to η. This yields:

$$\frac{dn_+}{d\eta} = 2\frac{n_+(1-n_+)}{q-1}\left[qn_+ - 1 + \eta q\frac{dn_+}{d\eta}\right]$$

which is easily solved for $\frac{dn_+}{d\eta}$. Then, (39) completes the proof of *(ii)*, since it shows that a solution with $L_+ = 1$ is a minimum if and only if n_+ increases with η.

In order to prove that solutions of the FOC with $L_+ > 1$ are not minima, it is sufficient to find a vector $\boldsymbol{\varepsilon}$ along which f decreases. Thus suppose that $L_+ > 1$ and let $n_r = n_+$ for $2 \leq r \leq L_+$ with $n_r = n_-$ otherwise. Along the direction $\boldsymbol{\varepsilon} = (\epsilon_1, -\epsilon_1, 0, \ldots, 0)$ the variation in f is given by

$$f(\mathbf{n}) - f(\mathbf{n}^*) = -\frac{2\eta n_+ - 1}{n_+}\epsilon_1^2 + O(\boldsymbol{\varepsilon}^3)$$

This is negative because $2\eta n_+ > 1$, which proves *(iii)*.

In order to prove *(iv)* we first observe that

$$\Delta f \equiv f(\mathbf{n}_s^*) - f(\mathbf{n}_a^*) \tag{40}$$

$$= -\log q - n_+\log n_+ - (1-n_+)\log\left[\frac{1-n_+}{q-1}\right] + \eta\frac{(qn_+-1)^2}{q(q-1)} \tag{41}$$

where $\mathbf{n}_s = (1/q, \ldots, 1/q)$ is the symmetric solution and $\mathbf{n}_a$ is one of the q asymmetric ones. The expression for Δf defines an increasing function of η. Indeed,

$$\frac{d\Delta f}{d\eta} = \frac{\partial \Delta f}{\partial \eta} + \frac{\partial \Delta f}{\partial n_+}\frac{dn_+}{d\eta}$$

but, since n_+ satisfies the FOC, the second term vanishes and the first is positive since $n_+ > 1/q$. Therefore there is a single value η^* for which $\Delta f = 0$. Direct substitution shows that the pair of values $n_+ = 1 - 1/q$ and η^* given in (13) simultaneously solve the equations $\Delta f = 0$ and (12) with $L_+ = 1$. This completes the proof of *(iv)*. Finally, in order to prove *(v)*, we first argue that for $q = 2$ and $\eta = \check{\eta} = q/2$, there is no solution of (12) with $L_+ = 1$ apart from $n_+ = 1/q$. To see this, note that when particularized to $q = 2$, $L_+ = 1$, and $\eta = q/2$, condition (12) takes the simple form $2n_+ - 1 = \tanh(2n_+ - 1)$, which has the unique solution $n_+ = 1/q$. On the other hand, the existence of a second solution $n_+ > 1/q$ for (12) when $q > 2$ is guaranteed by the existence of the asymmetric solution at $\eta^* < \check{\eta} = q/2$, which has been shown explicitly above. This concludes the proof. ■

Proof of Proposition 7: Let $N_{r,k} \equiv n_{r,k}N$ stand for the *total number* of nodes displaying each action r and degree k. First, we consider the evolution of these variables for $k > 0$. In this case, the magnitudes $N_{r,k}$ change over time solely due to link creation and link destruction. In any time interval $[t, t+\Delta t]$ of infinitesimal length Δt, link creation opportunities arrive independently to each node with probability $\eta\Delta t$, while each existing link is destroyed with probability $\lambda\Delta t = \Delta t$ (recall the normalization $\lambda = 1$). Then, we now argue that expected change $E[\Delta N_{r,k}]$ over that infinitesimal time interval is given, up to terms of order $\mathcal{O}(1/N)\Delta t$, by the following expression:

$$\begin{aligned} E[\Delta N_{r,k}] = N\eta \left\{ 2n_{r,k-1} \sum\nolimits_{k'} n_{r,k'} - 2n_{r,k} \sum\nolimits_{k'} n_{r,k'} \right\} \Delta t \\ + N \left\{ (k+1)n_{r,k+1} - kn_{r,k} \right\} \Delta t \qquad (k > 0). \end{aligned} \tag{42}$$

The first bracketed term of (42) concerns events of link creation. These affect $N_{r,k}$ through five possible routes:

(i) Some node counted in $N_{r,k-1}$ is selected for link creation and then meets another node counted in $N_{r,k-1}$ as well. This occurs with probability $\eta\Delta t\ N\ n_{r,k-1} \left[(N_{r,k-1} - 1)/(N-1)\right] = N\eta \left[n^2_{r,k-1} + \mathcal{O}(1/N)\right] \Delta t$.

(ii) Some node counted in $N_{r,k-1}$ is selected for link creation and then meets another node counted in $N_{r,k'}$ for k' such that $k \neq k' \neq k-1$, or *vice versa.* This occurs with probability $2\eta\Delta t\ N\ n_{r,k-1} \left[(\sum_{k\neq k'\neq k-1} N_{r,k'})/(N-1)\right] = 2N\eta \left[n_{r,k-1}(\sum_{k\neq k'\neq k-1} n_{r,k'}) + \mathcal{O}(1/N)\right] \Delta t$.

(iii) Some node counted in $N_{r,k}$ is selected for link creation and then meets another node counted in $N_{r,k}$ as well. This occurs with probability $\eta\Delta t\ N\ n_{r,k} \left[(N_{r,k} - 1)/(N-1)\right] = N\eta \left[n^2_{r,k} + \mathcal{O}(1/N)\right] \Delta t$.

(iv) Some node counted in $N_{r,k}$ is selected for link creation and then meets another node counted in $N_{r,k'}$ for k' such that $k \neq k' \neq k-1$, or *vice versa.* This occurs with probability $2\eta\Delta t\ N\ n_{r,k} \left[(\sum_{k\neq k'\neq k-1} N_{r,k'})/(N-1)\right] = 2N\eta \left[n_{r,k}(\sum_{k\neq k'\neq k-1} n_{r,k'}) + \mathcal{O}(1/N)\right] \Delta t$.

(v) Some node counted in $N_{r,k}$ is selected for link creation and then meets another node counted in $N_{r,k-1}$, or *vice versa.* This occurs with probability $2\eta\Delta t\ N\ n_{r,k} \left[N_{r,k-1}/(N-1)\right] = 2N\eta \left[n_{r,k}n_{r,k-1} + \mathcal{O}(1/N)\right] \Delta t$.

Now let us determine what is the induced change in $N_{r,k}$ for each of the above possibilities. For (i), $\Delta N_{r,k} = 2$ since the link created brings in two new nodes to the set of those that display action a_r and have degree k; for (ii), $\Delta N_{r,k} = 1$ since only one new node is added to that set; for (iii), $\Delta N_{r,k} = -2$ since the link created has the two connecting nodes increase their degree to $k+1$ and thus they abandon the set in question; for (iv), $\Delta N_{r,k} = -1$ since the connecting node that originally had degree k then has degree $k+1$ and thus abandons the set; for (v), $\Delta N_{r,k} = 0$ since the entry of one node in the set is exactly cancelled by the exit of one other node. Bringing together all these considerations, the first bracketed term of (42) readily obtains.

Let us now address the second bracketed term of (42). In this respect, simply note that a single link of some node with k' links is removed with a probability $\Delta t\ k' N_{r,k'} = N\ k' n_{r,k'}\ \Delta t$. Then, the expression for that term simply follows from the observation that the removal of one link from a node with degree $k+1$ induces $\Delta N_{r,k} = 1$ (i.e. increases $N_{r,k}$ by one), while if it affects a node with degree k it leads to $\Delta N_{r,k} = -1$.

Next, we consider the case where $k = 0$ and compute the expected change in the numbers $N_{r,0}$ for each r. Unlike for the case with $k > 0$, the dynamics is now affected by action adjustment, thus giving rise to the following expression:

$$E[\Delta N_{r,0}] = N\eta \left\{-2n_{r,0} \sum\nolimits_{k'} n_{r,k'}\right\} \Delta t + N n_{r,1} \Delta t + N\nu \sum_{s=1}^{q} [n_{s,0} - n_{r,0}]\, \Delta t. \tag{43}$$

The first two terms in (43) are just as before – reflecting link creation and link destruction, respectively – except that they now can operate only in one direction: neither isolate nodes can lose any links, nor link creation can lead a node to become an isolate node. The third term, on the other hand, embodies the process unfolding at the rate ν that makes isolate nodes move across the set of possible actions.

Expressions (42)-(43) give the expected change in the absolute numbers of nodes $N_{r,k}$ displaying each action r and degree k. The corresponding change in frequencies $n_{r,k}$ obviously satisfies $E[\Delta N_{r,k}] = N\ E[\Delta n_{r,k}]$, thus yielding (17)-(18), as desired. ■

Proof of Proposition 9: For each r and $k > 0$, the representation of $F_{r,k}(\mathbf{n})$ in terms of $f_r(x,t)$ reads:

$$\begin{aligned} F_{r,k}(\mathbf{n}) &= \int_0^\infty dx f_r(x,t)(x - 2\eta n_r) \left[\frac{x^k}{k!} - \frac{x^{k-1}}{(k-1)!}\right] e^{-x} \\ &= -\int_0^\infty dx f_r(x,t)(x - 2\eta n_r) \frac{\partial}{\partial x} \left[\frac{x^k}{k!} e^{-x}\right] \\ &= \int_0^\infty dx \frac{x^k}{k!} e^{-x} \frac{\partial}{\partial x} \left[f_r(x,t)(x - 2\eta n_r)\right] \end{aligned}$$

where, in the last passage, we have integrated by parts using the fact that $f_r(x,t)(x - 2\eta n_r)\frac{x^k}{k!} e^{-x}$ vanishes both at $x = 0$, for all $k > 0$, and at $x \to \infty$. Therefore the ODE $\dot{n}_{r,k} = F_{r,k}(\mathbf{n})$ requires that, for all r and $k > 0$,

$$0 = \int_0^\infty dx \frac{x^k}{k!} e^{-x} \left[\partial_t f_r - \partial_x \left(x - 2\eta n_r\right) f_r\right]. \tag{44}$$

If (44) holds for all $k > 0$, then it means that the term in square brackets vanishes, i.e. f_r satisfies (28). Thus, to complete the proof, it is enough to note that, since the transformation to the Poisson representation is invertible, for any initial conditions specifying the $n_{r,k}(0)$ one can identify a suitable initial condition $\phi_r(x)$ for each f. ■

Proof of Proposition 10: First, we determine the characteristic paths $(\xi_r(t), t)$ along which the PDE (28) is equivalent to the ODE (30). These characteristic trajectories are obtained from a solution to the equation

$$\frac{d\xi_r(t)}{dt} = 2\eta n_r(t) - \xi_r(t), \quad \xi_r(0) = x_0,$$

i.e.

$$\xi_r(t) = x_0 e^{-t} + 2\eta \int_0^t ds e^{s-t} n_r(s) \equiv e^{-t} x_0 + \chi_r(t). \tag{45}$$

where the last equality defines

$$\chi_r(t) \equiv 2\eta \int_0^t ds n_r(s) e^{s-t}. \tag{46}$$

On the characteristic path starting from the initial condition $(x_0, t = 0)$, the function f_r satisfies $\frac{df_r}{dt} = f_r$, i.e. $f_r(\xi_r(t), t) = \phi_r(x_0)e^t$. Inverting (45) we obtain $x_0 = e^t[x - \chi_r(t)]$ and therefore:

$$f_r(x,t) = e^t \phi_r \left\{ e^t [x - \chi_r(t)] \right\}. \tag{47}$$

For $x \geq \chi_r(t)$, this solution is related to values of the initial condition $\phi_r(x_0)$ with $x_0 \geq 0$. Instead, for $x < \chi_r(t)$, the initial condition is associated to negative values of x_0, for which ϕ_r is not defined. We therefore set the values of $\phi_r(x_0)$ for $x_0 < 0$ in such a way as to satisfy the condition (29). In order to do this, integrate (47) over $x \in [0, \infty)$:

$$\begin{aligned} n_r(t) &= \int_0^\infty dx e^t \phi_r \left\{ e^t [x - \chi_r(t)] \right\} = \int_{-e^t \chi_r(t)}^\infty dz \phi_r(z) && (48) \\ &= \int_{-e^t \chi_r(t)}^0 dz \phi_r(z) + n_r(0) && (49) \end{aligned}$$

where we have changed variable to $z = e^t[x - \chi_r(t)]$ and split the integrals for $z \geq 0$ and $z < 0$. Taking now the derivative with respect to t on both sides, we get

$$\frac{dn_r}{dt} = \phi_r(-e^t \chi_r) \frac{d}{dt} \left[e^t \chi_r(t) \right] = 2\eta \phi(-e^t \chi_r) e^t n_r(t) \tag{50}$$

or

$$\phi(-e^t \chi_r(t)) = \frac{e^{-t}}{2\eta} \frac{d}{dt} \log n_r(t).$$

This implicitly defines the function $\phi_r(x_0)$ for negative x_0 in terms of the function $n_r(t)$. Notice that $\chi_r(t)$ is also fully specified in terms of $n_r(t)$ and that

$e^t\chi_r(t)$ is an increasing function of t as long as $n_r > 0$, so the procedure yields a unique selection.

The solution of the fraction $n_{r,k}$ of nodes in class r with degree k can then be expressed, with the same change of variable as before, as

$$n_{r,k}(t) = \int_{-e^t\chi_r}^{\infty} dz\phi_r(z) \left\{ \frac{[\chi_r + (z-\chi_r)e^{-t}]^k}{k!} e^{-\chi_r+(\chi_r-z)e^{-t}} \right\} \tag{51}$$

where $\chi_r(t)$ is a function of $n_r(t)$.

Now, we specialize this solution to the case where the evolution of n_r takes place at a speed much lower than that for $n_{r,k}$. To this end, let the population in component r be a smooth function of the time variable $\tau = \nu t$, i.e. $n_r(t) = \bar{n}_r(\tau)$. Then, by (46), for $t = \tau/\nu$ and small ν:

$$\chi_r(t) = 2\eta\bar{n}_r(\tau) - 2\eta\nu\frac{d\bar{n}_r}{d\tau} + O(\nu^2).$$

Likewise the term $\chi_r+(z-\chi_r)e^{-t} \to 2\eta\bar{n}_r(\tau)$ in the limit $\nu \to 0$ with $\tau = \nu t > 0$ given, and hence the term in braces of (51) converges uniformly to $\frac{(2\eta n_r)^k}{k!}e^{-2\eta n_r}$ in the same limit. Since, by (48), the remaining integral on z is exactly $\bar{n}_r(\tau)$, the desired conclusion follows. ■

Proof of Proposition 11: The derivative of f with respect to τ is given by

$$\begin{aligned} \frac{df}{d\tau} &= \sum_r \left[1 + \log \bar{n}_r - 2\eta\bar{n}_r\right] \frac{d\bar{n}_r}{d\tau} && (52) \\ &= \sum_r \left[1 + \log n_{r,0}\right] \sum_s \left[n_{s,0} - n_{r,0}\right] && (53) \\ &= -\frac{1}{2}\sum_{r,s} \left[\log n_{s,0} - \log n_{r,0}\right]\left[n_{s,0} - n_{r,0}\right] \leq 0 && (54) \end{aligned}$$

where we use (32) and also $n_{r,0} = \bar{n}_r e^{-2\eta\bar{n}_r}$, the latter obtained from Proposition 10 by particularizing (31) to $k = 0$. On the other hand, in the last line we split the sum $\sum_{r,s}[\ldots] = \frac{1}{2}\sum_{r,s}[\ldots] + \frac{1}{2}\sum_{r,s}[\ldots]$, interchange the indices in the second term, and recombine the resulting expression. Given that $\log x$ is an increasing function, the expression (54) is non-negative, and it is zero only if $n_{r,0} = n_{s,0}$ for all $r, s = 1, \ldots, q$. This establishes the required property for time derivative of f in the slow time variable τ of the MFD. ■

References

[1] van Alstyne, M. (1997): "The state of network organization: a survey in three frameworks," *Journal of Organizational Computing and Electronic Commerce* 7, 83-151.

[2] Bala, V. and S. Goyal (1999): "A non-cooperative model of network formation," *Econometrica* **68**, 1181-1230.

[3] Barabási, A.-L. and R. Albert (1999):"Emergence of scaling in random networks", *Science* **286**, 509-12.

[4] Benaïm, M. and J. Weibull (2003*a*): "Deterministic approximation of stochastic evolution in games," *Econometrica* **3**, 873-903.

[5] Benaïm, M. and J. Weibull (2003*b*): "Deterministic approximation of stochastic evolution in games – a generalization," *mimeo*, Université de Neuchatel and Boston University.

[6] Binmore, K., L. Samuelson, and R. Vaughan (1995): "Musical Chairs: Modelling Noisy Evolution," *Games and Economic Behavior* **11**, 1-35.

[7] Blume, L. (1993): "The statistical mechanics of strategic interaction", *Games and Economic Behavior* **4**, 387-424.

[8] Bollobás, B. (2001): *Random Graphs* (2nd. Edition; First Edition 1985), Cambridge (U.K.): Cambridge University Press.

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

[10] Castilla, E. J., H. Hwang, E. Granovetter, and M. Granovetter (2000): "Social Networks in Silicon Valley," in C.-M. Lee, W. F. Miller, M. G. Hancock, and H. S. Rowen, editors, *The Silicon Valley Edge*, Stanford: Stanford University Press.

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

[12] Ely, J. (2003): "Local Conventions," *Berkeley Electronic Press Journals, Advances in Theoretical Economics* **2**, 1, 1.

[13] Erdös, P. and A. Rényi (1960): "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17-61.

[14] Eshel, I., L. Samuelson, and A. Shaked (1998): "Altruists, egoists, and hooligans in a local interaction model," *American Economic Review* **88**, 157-179.

[15] Fafchamps, M. and S. Lund (2003): "Risk-sharing networks in rural Philippines," *Journal of Development Economics* **71**, 261-87.

[16] Gardiner (2004), C.W. : *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, Berlin: Springer-Verlag.

[17] Goyal, S. and F. Vega-Redondo (2005): "Network formation and social coordination," *Games and Economic Behavior* **50**, 178-207.

[18] Goyal S., M. van der Leij, J. L. Moraga-González (2006): "Economics: an emerging small world?," *Journal of Political Economy* **114**, 403-412.

[19] Granovetter, M. (1974): *Getting a Job: a Study on Contacts and Careers*, Chicago: Chicago University Press.

[20] Grossman, J. (2002): "The evolution of the mathematical research collaboration graph," *Congressus Numerantium* **158**, 201-212.

[21] Hagedoorn, J. (2002): "Inter-firm R&D partnerships: an overview of major trends and patterns since 1960", *Research Policy* **31**, 477-492.

[22] Harrigan, K.R. (1988): "Strategic alliances and partner asymmetries," *Management International Review* **28**, 53-72.

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

[24] Jackson, M. O. and A. Wolinsky (1996): "A Strategic model of economic and social networks," *Journal of Economic Theory* **71**, 44-74.

[25] Jackson, M. (2005): "A survey of network formation models: stability and efficiency," in G. Demange and M. H. Wooders (eds.), *Group Formation in Economics: Networks, Clubs, and Coalitions*, Cambridge: Cambridge University Press.

[26] Jackson, M. O. and B. Rogers (2007): "Meeting Strangers and Friends of Friends: How Random are Socially Generated Networks?" *American Economic Review* **97**, 890-915.

[27] Karlan, D., M. Möbius, T. Rosenblat, and A. Szeidl (2008): "Trust and social collateral," forthcoming in *Quarterly Journal of Economics.*

[28] Kirman, A., D. K. Herreiner, and G. Weisbuch (2000): "Market organization and trading relationships," *Economic Journal* **110**, 411-436.

[29] Kogut, B., P. Urso, and G. Walker (2007): "The Emergent Properties of a New Financial Market: American Venture Capital Syndication from 1964 to 2004," Management Science **53**, 1181-98.

[30] Kolmogorov, A.N. and S. V. Fomin (1999): *Elements of the Theory of Functions and Functional Analysis,* New York: Dover Publications.

[31] Krackhardt, D., and J. R. Hanson (1993): "Informal networks: the company behind the chart," *Harvard Business Review* **71**, 104-11.

[32] Kranton, R. and D. Minehart (2001): "A theory of buyer-seller networks," *American Economic Review* **91**, 485-508.

[33] Murgai, R., P. Winters, E. Sadoulet, and A. de Janvry (2002): "Localized and incomplete mutual insurance," *Journal of Development Economics* **67**, 245-274.

[34] Newman, M. E. J. (2001): "The structure of scientific collaboration networks," *Proceedings of the National Academy of Sciences USA* **98**, 404-409.

[35] Newman, M. E. J. (2003):"The structure and function of complex networks", *SIAM Review* **45**, 167-256.

[36] Newman, M.E., S.H. Strogatz, and D.J. Watts (2001): "Random graphs with arbitrary degree distributions and their applications," *Physical Review E* **64**, 026118.

[37] Park, S.H. and M.V. Russo (1996): "When competition eclipses cooperation: an event history analysis of joint venture failure," *Management Science* **42**, 875-90.

[38] Saxenian, A. (1994): *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Cambridge, Mass.: Harvard University Press.

[39] Vega-Redondo, F. (2007): *Complex Social Networks*, Econometric Society Monograph Series, Cambridge: Cambridge University Press.

[40] Watts D.J. and S. H. Strogatz (1998): "Collective dynamics of 'small-world' networks", *Nature* **393**, 440-42**.**

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