Clustering, Cooperation, and Search in Social Networks*,†

Fernando Vega-Redondo
Universidad de Alicante,
Instituto Valenciano de Investigaciones Económicas, and
University of Essex

Matteo Marsili
Abdus Salam International Centre for Theoretical Physics, Trieste

Frantisek Slanina
Inst. of Physics, Academy of Sciences of the Czech Republic, Prague

September 2004

Abstract

The paper discusses the important role of clustering in the evolution of social networks, as it affects not only the incentives of players to cooperate but also their ability to search for fresh opportunities. Depending on the volatility of the environment and the social convention in place, we show that network clustering endogenously adapts to achieve and maintain, to the extent possible, a state of high social connectivity in the long run.

JEL Classification nos.: C71, D83, D85

---

* **Acknowledgement**: This work was supported by Spanish MEC (SEJ2004-02170), the European Union (HPRN-CT-2002-00319), and the Grant Agency of the Czech Republic (202 01 1091).

† Vega <vega@merlin.fae.ua..es>; Marsili <marsili@ictp.trieste.it>; Slanina <slanina@fzu.cz>.

1 Introduction

There is a recent burst of interest in the study of networks that springs from a wide diversity of fields: economics, sociology, biology, mathematics, or physics. Despite their common object of concern, the focus and methodology is markedly different in each case.

Thus, in economics and other social sciences the social network is often taken to be endogenous and the objective is to understand how payoffs and the entailed agents' incentives might explain their pattern of links and the induced outcomes.[1] The problem is frequently formalized as a network-formation game, the focus then being on how some suitable notion of equilibrium may define the range of (strategically) stable networks. Implicitly, this presumes that the underlying environment is rather stable and/or simple, which allows agents to know (or eventually learn) their relevant range of possibilities. In a sense, therefore, this approach abstracts from the variability and complexity that appears to be a key feature of most interesting setups in the real world.

By contrast, the approach pursued by other disciplines (e.g. the growing literature on so-called "complex networks")[2] is of a somewhat polar nature. Their main concern has been to unveil the large-system regularities that arise from simple mechanisms of network formation, aiming to understand as well their implications on different measures of performance, e.g. diffusion, search, or adaptability. Applied to social setups, however, this approach lacks the "micro foundation" on agents' incentives that economists generally insist is essential to truly understand socio-economic phenomena.

The above considerations suggest that fruitful insights may be gained from integrating the former two approaches. In a nutshell, the basis of this effort should be the recognition that both incentives and complexity are essential ingredients (for us, analysts, as well as for the agents) in any proper understanding of the process of network formation in most real-world environments. Here, we illustrate the potential of such integration by reflecting on a particular issue: the (endogenous) role of network clustering in supporting the conflicting requirements for cooperation and search in a volatile environment.

The insights obtained in this respect build upon research conducted by Vega-Redondo (2002), and Marsili, Vega-Redondo, and Slanina (2004), thereafter referred to as VR and MVS respectively. The common setup consists of a given population of agents who continuously search for valuable new links (fresh payoff opportunities), at the same time as the links already in place decay stochastically and eventually vanish due to volatility. The struggle between these two forces, search and volatility, generates a tension reminiscent of what

[1] See, for example, Aumann and Myerson (1988), Jackson and Wolinsky (1996), Bala and Goyal (2000) or, for a selected overview, the papers collected in Dutta and Jackson (2003). There is also, however, a very large sociological literature that places less emphasis on individual incentives underlying network formation and more on the performance and payoffs associated to given network architectures and corresponding network positions. Granovetter (1973) or Burt (1982) are good examples of this approach.

[2] See, for example, Watts and Strogatz (1998), Barabasi and Albert (1999), the monograph by Dorogovtsev and Mendes (2003), or the excellent survey paper by Newman (2003).

evolutionary biologists label the Red Queen Principle (see van Valen (1973)): "... it takes all the running you can do, to keep in the same place" (Carrol (1872)). We find that the resolution of such a tension crucially depends on two exogenous features of the situation: (a) the nature of search (i.e. its intensity and depth); (b) the prevailing social norm (i.e. whether opportunism is deterred by bilateral or multilateral punishments). We also stress interesting implications on the architecture of the evolving social network, as it adapts endogenously to changes in volatility. In particular, our main conclusion is that the impact of increasing volatility on clustering depends crucially on the social convention in place as well as the characteristics of search – specifically, we find that clustering grows with volatility if the social convention is multilateral and search is long-range, while the opposite happens if the convention is purely bilateral and search is short range. This points to an interesting interplay among the social convention in place, the exogenous environmental conditions, and the endogenous architecture of the social network.

2 The theoretical framework

Let time be modelled discretely, with $t = 0, 1, 2, ...$ indexing the consecutive time periods. The population of players, N, is finite. At every t, each agent $i \in N$ supports a certain number of links that connect her to a set of other agents, $A_i(t) \subset N \backslash \{i\}$. To simplify matters, let us suppose that the number of links that can be supported by any given agents is at most m, a parameter of the model. Given the set of links $\{A_i(t)\}_{i \in N}$ supported by all agents at t, the prevailing social network $g(t)$ is defined as follows: $g(t) = \{(i, j) \in N \times N : j \in A_i(t) \vee i \in A_j(t)\}$. By construction, therefore, $g(t)$ is **undirected** so that $(i, j) \in g(t) \Leftrightarrow (j, i) \in g(t)$, and we may simply use the notation ij to refer to the corresponding link. For any agent i, the set $N_i(t) \equiv \{j \in N : ij \in g(t)\}$ is called her **neighborhood** at t.

Every agent i is assumed to play an infinite repetition of a Prisoner's Dilemma (PD) game with each of her neighbors $j \in N_i(t)$. The payoffs of the stage PD game are ij-**idiosyncratic**, as given by the table:

$i \backslash j$	C	D
C	ζ_{ij}, ζ_{ij}	$\zeta_{ij} - 1, \zeta_{ij} + 1$
D	$\zeta_{ij} + 1, \zeta_{ij} - 1$	$0, 0$

(1)

where $1 > \zeta_{ij} (= \zeta_{ji}) > 0$. As customary, C and D are labelled as "Cooperate" and "Defect," respectively. Thus, the payoff ζ_{ij} obtained by both players if they jointly cooperate is ij-specific and will change over time in the dynamic ("volatile") environment to be considered later on.

The social dynamics consists of three main components: volatility, search, and link removal. We describe each of them in turn.

1. **Volatility**: At every t, each of the links $ij \in g(t-1)$ prevailing during the preceding period has its corresponding PD cooperation payoff $\zeta_{ij}(t-1)$ subject

to an independent random redraw with probability ε. Thus, with probability ε the new payoff $\zeta_{ij}(t)$ is obtained through a random, say uniform, draw from some non-negative real interval $[\underline{\zeta}, \overline{\zeta}]$. With the complementary probability $1-\varepsilon$, there is payoff persistence, i.e. $\zeta_{ij}(t) = \zeta_{ij}(t-1)$. The value of ε is taken to be a measure of the volatility of the environment.

2. Search: At every t, every agent i enjoys the possibility of supporting new links, this opportunity being obtained through either "local" or "global" search.

- With probability p, search is local, which means that it is restricted to lie within the component of i in $g(t-1)$, denoted by $C_i(t-1)$.[3] Two parameters determine the depth and intensity of local search. The depth θ (≥ 2) specifies the geodesic distance from i at which this player can find new potential partners. On the other hand, the intensity is parametrized by σ (≥ 1), which specifies the maximum number of different players that can be "sampled" by i for new payoff opportunities. Given these parameters, we postulate that any searching player makes σ distinct explorations of new linking opportunities by "moving" along the network at most θ steps away. If the player j thus "met" is not already a neighbor $(j \notin N_i(t-1))$, she obtains an independent payoff draw for each of them, again uniformly on the contemplated interval.
- With probability $(1-p)q$, the search is global, which means that player i obtains one new draw for some arbitrarily chosen player $j \in N \backslash N_i(t-1)$.

Once completed her search (local or global) of new payoff opportunities, player i is entitled to create at most a single new link, possibly dispensing with one of the pre-existing links if she already supports m of them. For simplicity, we assume that the player ends up supporting the admissible collection of links with maximum payoffs, thus ignoring at this stage all strategic considerations.[4]

3. Link removal: Let $\tilde{g}(t)$ denote the network resulting from search at t. For every $ij \in \tilde{g}(t)$, the two players involved, i and j, are assumed to verify, prior to actual play, whether their link is worth keeping. Assuming that only links that support cooperation are judged worthwhile, we are interested in contrasting two possibilities. These are associated to different conventions governing agents' strategic interaction.

(i) Bilateral Convention (BC): In this case, cooperation between i and j must be supported bilaterally in the infinitely repeated PD played by them. This means that the cooperation payoff $\zeta_{ij}(t)$ must satisfy $\zeta_{ij}(t) \geq \frac{1-\delta}{\delta}$ where, as customary, δ denotes the (common) factor at which players discount their stage payoff in the repeated game.

[3] The component $C_i(t-1)$ consists of all those other nodes j for whom there is a network path from i to j in $g(t-1)$.

[4] Allowing players to factor strategic considerations into their decisions of link creation has no significant effect in the analysis.

(ii) **Multilateral Convention** (MC): Under this convention, any unilateral deviation from cooperation by a player with one of her neighbors will be punished by any of her other neighbors as soon as the latter get to learn about it. More specifically, let us suppose that, in every round of the PD games played "in parallel" by each pair of connected players, not only this interaction takes place but they also share any other strategic information they might have – most importantly, information on whether *any* other agent has *unilaterally* deviated from cooperation with some of their partners in a preceding period. Then, one can show (cf. VR) that the relevant condition for cooperation between i and j (under the simplifying assumption that any deviation can affect only one link at a time) is given by the following requirement:

$$\zeta_{ij}(t) \geq \frac{1-\delta}{\delta} - \sum_{k \in N_i \setminus \{j\}} \left(\zeta_{ik}(t) + \frac{1-\delta}{\delta} \right) \delta^{d^i(j,k)-1}, \qquad (2)$$

where $d^i(j,k)$ is the length of the shortest path joining j and k which does *not* involve player i. The interpretation of this i-*excluding distance* between j and k is straightforward: it is the number of steps (and therefore periods, in the repeated game) which would be required for any information held by j (or k) to reach k (or j) without the concourse of player i. Note, of course, that the requirement of cooperation under MC is never stronger than under BC, and may be much weaker if the average distance between neighbors (excluding the agent in question) is low. In the language of the sociologist James Coleman (1990), this underscores the importance of "social closure" (or network clustering, using the network terminology introduced below) in supporting cooperation multilaterally when bilateral incentives may be too weak to do so.[5]

Under whatever convention applies, BC or MC, all links in $\tilde{g}(t)$ that are not anticipated to support cooperation are removed, thus leading to the actual network $g(t)$ ($\subset \tilde{g}(t)$) prevailing at t.

3 Analysis

We want to contrast the implications of BC and MC, the two different social conventions considered above. This comparison has been carried out in VR and MVS under polar search conditions, both through numerical simulations and mean-field analysis. (Mean field analysis is a common approach used in statistical physics that approximates the dynamics of the system by its expected motion, under the assumption that the forces impinging on each node

[5] Within the economic literature, the present considerations are reminiscent of the well-known work on multimarket collusion pioneered by Bernheim and Whinston (1990) – see also Spagnolo (1999) – or the historical research conducted by Greif (1993) on early trading institutions.. Related work, with a different perspective, can also found in Haag and Lagunoff (2000) or Annen (2003)

are stochastically independent). The main insights delivered by both numerical simulations and mean-field analysis essentially coincide. Here, due to space limitations, we restrict attention to the numerical results, referring the interested reader to VR and MVS for a detailed discussion of the analytical approach.[6]

Our main concern is to understand how a *rise* in the key environmental parameter, the effective (normalized) rate of volatility ε/p, affects the network architecture and the ability of the society to maintain a high level of interaction. First, Figure 1 depicts the situation under MC in a context with high depth and intensity of search – specifically, we make $\theta = \sigma = n$, where n is the (large but finite) population size.[7] This diagram traces long run magnitudes obtained as volatility is increased very gradually and, at each step, the system is let to stabilize around well-defined long-run values. The upper panel shows that, as expected, rising volatility has a detrimental effect on long-run network connectivity. It is also interesting to note that the (negative) effect is of decreasing (absolute) magnitude as volatility rises. But the more interesting observation reported in VR is that, as the volatility rate ε grows, the architecture of the social network endogenously adjusts so that the average distance between neighbors of a randomly selected node, i.e. $\frac{1}{n}\sum_{i\in N}\left(\frac{1}{|N_i||N_i-1|}\sum_{j,k\in N_i} d^i(j,k)\right)$, correspondingly decreases.[8] To cast this conclusion in a way that may be conceptually related to the customary notion of clustering, the lower panel of Figure 1 traces the change in the *inverse* of the aforementioned average distance, a magnitude that, for the sake of a better name, we simply label the *generalized clustering* of the network. Its role in supporting cooperation should be clear in view of (2). It is remarkable that such a cooperation-enhancing feature of the social network should emerge in the long run as the result of cumulative but unintended by-product of agents' independent adjustments.

In contrast, Figure 2 presents the counterpart of the above results when the underlying social convention is BC. The differences with the scenario governed by MC are quite sharp and interesting. The upper panel shows that the detrimental effect of volatility on network connectivity is already significant at volatility rates that are lower than before by a full order of magnitude. Moreover, these effects become of increasing magnitude as volatility grows, eventually leading to an abrupt transition to a social network with very low average connectivity as the volatility rate ε surpasses a certain threshold. The responsibility for this contrasting state of affairs is to be found in the absence of *strategic* network effects that, in this scenario, no longer can be relied upon to act as

[6] The gist of our conclusions does not depend on a particular choice of parameter values. The simulations reported here were obtained under the following parameters: the discount factor $\delta = 3/4$; the relative rates of global vs. local search $q/p = 0.1$; and the payoff support $[\underline{\zeta}, \overline{\zeta}] = [0.1, 0.35]$. Note that the latter implies that there is only a small (but positive) probability that a random payoff draw allows for cooperation to be supported bilaterally.

[7] In VR, the bulk of the numerical simulations were conducted for $n = 100$, while in MVS the population size was increased an order of magnitude to $n = 1000, 2000$. In neither paper were significant effects discerned concerning population size.

[8] When no path exists between two neighbors of a given node, the entailed "infinite" distance was truncated to n, the size of the population.

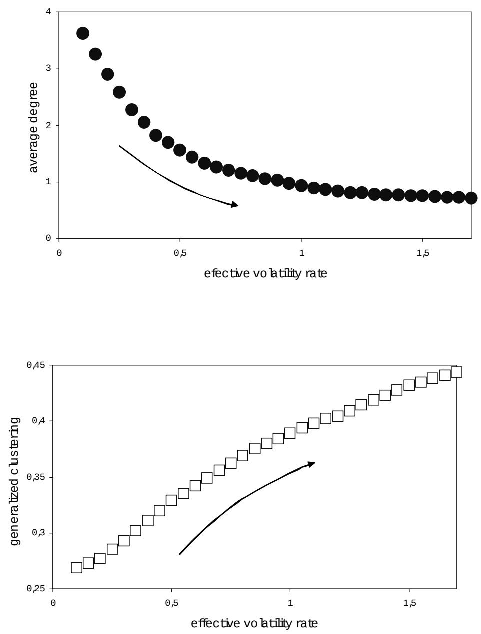

Figure 1: : Long-run behavior of the process studied in VR operating under MC, as volatility gradually increases and the system stabilizes to corresponding long-run values.

a "buffer" against volatility. This is indeed confirmed by the lower panel of Figure 2 that shows in a clear way the irrelevance of these effects in the present scenario. Since they no longer provide any advantage under BC, generalized clustering remains largely unchanged, as volatility increases, provided the social network remains significantly connected. At the point of the transition, however, it falls abruptly as a by-product of the general increase in distances (in particular, between neighbors) brought about by the sharp decrease in overall network connectivity.

The above discussion stresses the potentially beneficial role of network clustering in supporting cooperation when the convention governing strategic play is multilateral rather than bilateral. In general, however, it is intuitive that there could be a countervailing negative effect of network (generalized) clustering on the effectiveness of search. For example, if many of an agent's neighbors are close to each other, then local search typically induces substantial redundancies: the exploration of new linking options via neighbors will often lead to agents who are already neighbors. These considerations did not have any significant bearing on our former discussion because of the extreme parameter configuration used concerning the intensity and depth of search – specifically,

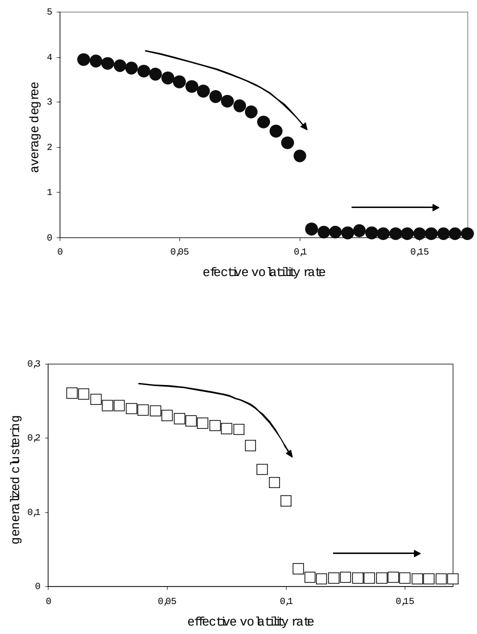

Figure 2: : Long-run behavior of the process studied in VR operating under BC, as volatility gradually increases and the system stabilizes to corresponding long-run values.

we posited $\theta = \sigma = n$, so that search is effectively global, i.e. a searching agent samples a fresh opportunity with every other player in the component who is not yet a neighbor. However, if search is substantially less intense and deep (i.e. if the parameters θ and σ attain much lower values), matters should be expected to turn out markedly different .

To shed light on this issue in the starkest fashion, let us contrast the previous search scenario with the polar context where $\theta = 2$ and $\sigma = 1$ (i.e. only second neighbors may become new fresh neighbors and just one agent is sampled in every search opportunity). Furthermore, suppose for concreteness that the convention in place does not allow for strategic multilateral effects, i.e. BC applies. This is essentially the context studied by MVS that can be regarded as a particular case of the above general setup for low-intensity and short-range search. As volatility increases gradually, the numerical results obtained in the simulations are as depicted in Figure 3.[9]

[9] Figure 3 is intended to be only suggestive of the main ideas underlying MVS and abstract from some of the interesting features that arise in the analysis. For example, we found that there is some hysteresis (or dependence of initial conditions) around the point of transition, as is characteristic of (first-order) phase transitions in models of statistical physics. That is, the threshold on volatility that marks the transition from a dense to a sparse network as volatility

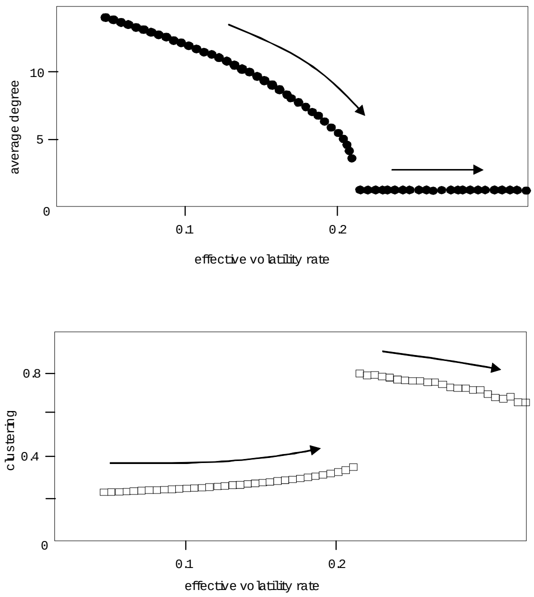

Figure 3: : Long-run behavior of the process studied in MVS operating under BC, as volatility gradually increases and the system stabilizes to corresponding long-run values.

First, its upper panel shows that, as volatility grows, network connectivity falls, again in an accelerating (and eventually abrupt) fashion, in line with what was observed in Figure 2 above. The interesting point of contrast, however, is found in the lower panel of Figure 3. There we depict the effect of increasing volatility on the usual measure of network clustering, defined as the probability that two randomly selected neighbors of any given node be neighbors themselves.[10] We observe that an inherent feature of any long-run state with high connectivity is its low clustering. Why is this so? Because, as explained, high clustering leads to redundancy in search in the present case – i.e. the neighbor of a neighbor is likely to be *already* a neighbor. This, in turn, reduces the effectiveness of search in offsetting volatility by the creation of *new* links. Since

---

rises is higher than the one that triggers the opposite transition when volatility falls.

[10] Note that a network has a maximum clustering of one if, and only if, its *generalized* clustering (as defined above) is also one. However, low levels of clustering may be consistent with much higher levels of generalized clustering if neighbors tend to be connected by short, albeit not direct, paths. Formally, note that clustering can be defined by $\frac{1}{n}\sum_{i\in N}\left\{\frac{1}{|N_i||N_i-1|}\sum_{j,k\in N_i}\left[\tilde{d}^i(j,k)\right]^{-1}\right\}$, where $\tilde{d}^i(j,k)=1$ if j and k are neighbors or $\tilde{d}^i(j,k)=\infty$ otherwise. This suggests that generalized clustering could have been alternatively defined as $\frac{1}{n}\sum_{i\in N}\left\{\frac{1}{|N_i||N_i-1|}\sum_{j,k\in N_i}\left[d^i(j,k)\right]^{-1}\right\}$, a modification that should deliver the same qualitative insights as reported in Figures 1 and 2.

we have $\theta = 2$, a network with low clustering may be seen as representing a situation where neighbors are "far apart," i.e. more than the one step away required for local search. In this light, therefore, we find a sharp contrast between the situation depicted in Figures 2 and 3. In both of them, the social norm BC applies so that strategic network effects are absent and thus clustering has no behavioral implications. However, while in the first case clustering (or neighbor distances) have no effect on search (which is both extensive and long-range), in the second scenario (where the opposite applies) this effect is crucial and has a strong bearing on the sustainability of a high-connectivity state as volatility increases. Only when volatility rises beyond a certain point and a connected network can no longer be maintained does clustering increase, a manifestation of the fact that the network becomes fragmented in a large number of small cliques.

We started this paper by putting forward the view that a much richer understanding of social networks ought to follow from an integration of the concerns for incentives and complexity that characterize, respectively, the socio-economic and complex-system literatures. We have briefly illustrated the potential of such an integration by focusing attention on the important specific issue of how the independent adjustment of agents endogenously shapes the clustering of the social network. In essence, our conclusion has been that the long-run outcome may be understood as striking a compromise between two conflicting aims: (a) the incentives of agents to behave cooperatively; (b) their ability to search effectively in a complex and volative environment.

References

[1] Annen, K. (2003): "Social capital, inclusive networks, and economic performance", Journal of Economic Behavior and Organization **50**, 449-63.

[2] Aumann, R. and R. Myerson (1988): "Endogenus formation of links between players and coalitions: an application to the Shapley Value" in The Shapley Value, A. Roth (ed.), 175-91, Cambridge University Press.

[3] Bala V. and S. Goyal (2003): "A noncooperative model of network formation", Econometrica **68**, 1181-1229.

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

[5] Bernheim, D. and M. Whinston (1990): "Multimarket contact and collusive behavior," Rand Journal of Economics **21**, 1-26.

[6] Burt, R.S. (1982): Towards a Structural Theory of Social Action, New York: Academic Press.

[7] Carrol, L. (1872): Through the Looking Glass and What Alice Found There, London: Macmillan.

[8] Coleman, J. S. (1988): "Social capital in the creation of human capital," American Journal of Sociology **94**, S95-S120.

[9] Dorogvtsev, S.N. and J.F.F. Mendes (2003): Evolution of Networks, Oxford: Oxford University Press

[10] Dutta, B. and M.O. Jackson (2003): Networks and Groups: Models of Strategic Formation, Berlin: Springer Verlag.

[11] Granovetter, M.S. (1973): "The strength of weak ties," American Journal of Sociology **78**, 1360-80.

[12] Greif, A. (1993): "Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition," American Economic Review **83**, 525-48.

[13] Haag, M. & R. Lagunoff (2000): "Social norms, local interaction, and neighborhood planning," mimeo, Department of Economics, Georgetown University.

[14] Jackson, M. and A. Watts (2002): "On the formation of interaction networks in social coordination games," Games and Economic Behavior **41**, 265-91.

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

[16] Marsili M., F. Vega-Redondo & F. Slanina (2004): "The rise and fall of a networked society: a formal model," Proceedings of the National Academy of Sciences of USA **101**, 1439-43.

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

[18] Spagnolo, G. (1999): "On interdependent supergames, multimarket contact, and collusion", Journal of Economic Theory **89**, 127-39.

[19] Van Valen, L. (1973): "A new evolutionary theory," Evolutionary Theory **1**, 1-30.

[20] Vega-Redondo, F. (2002): "Building Up Social Capital in a Changing World," WP-AD 2002-26, Instituto Valenciano de Investigaciones Económicas.

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