Original Paper

Adaptive Behavior

Adaptive Behavior
18(5) 416–427

DOI: 10.1177/1059712310384282
adb.sagepub.com

Learning to coordinate in complex networks

Sven Van Segbroeck[1,2], Steven de Jong[1,3], Ann Nowé[1], Francisco C Santos[4] and Tom Lenaerts[1,2]

Abstract

Designing an adaptive multi-agent system often requires the specification of interaction patterns between the different agents. To date, it remains unclear to what extent such interaction patterns influence the dynamics of the learning mechanisms inherent to each agent in the system. Here, we address this fundamental problem, both analytically and via computer simulations, examining networks of agents that engage in stag-hunt games with their neighbors and thereby learn to coordinate their actions. We show that the specific network topology does not affect the game strategy the agents learn on average. Yet, network features such as heterogeneity and clustering effectively determine how this average game behavior arises and how it manifests itself. Network heterogeneity induces variation in learning speed, whereas network clustering results in the emergence of clusters of agents with similar strategies. Such clusters also form when the network structure is not predefined, but shaped by the agents themselves. In that case, the strategy of an agent may become correlated with that of its neighbors on the one hand, and with its degree on the other hand. Here, we show that the presence of such correlations drastically changes the overall learning behavior of the agents. As such, our work provides a clear-cut picture of the learning dynamics associated with networks of agents trying to optimally coordinate their actions.

Keywords

Complex networks, evolutionary game theory, learning automata

1 Introduction

In multi-agent systems (MAS), an agent's success is often not solely the result of its own actions, but depends also on the actions of other agents. Moreover, agents often need to coordinate their actions to solve a particular task (Kraus, 1997). Game theory (Binmore, 1991; Osborne & Rubinstein, 1994) has proven itself an excellent framework to analyze both issues, representing the problem of coordination between agents in terms of simple games such as the stag-hunt game (Skyrms, 2004). In this game, two interacting agents have to choose simultaneously to hunt either stag or hare. Hunting stag yields the highest possible payoff, provided that the other agent cooperates and hunts stag as well. In case the other agent does not to cooperate (i.e., he defects) and chooses to hunt hare, the best reply is to do the same and to hunt hare. In other words, agents obtain a high payoff by coordinating their actions, but may end up in a suboptimal situation when taking the safest bet.

As such, the stag-hunt game represents, in a simplified manner, coordination problems like those encountered in swarm robotics (Dorigo et al., 2005) or the evolution of language (Christiansen & Kriby, 2003). Swarm robotics research investigates how the collective

[1]COMO, Vrije Universiteit Brussel, Brussels, Belgium.
[2]MLG, Université Libre de Bruxelles, Brussels, Belgium.
[3]DKE, Universiteit Maastricht, Maastricht, Netherlands.
[4]CENTRIA, Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal.

Corresponding author:
Tom Lenaerts, MLG, Département d'Informatique, Faculté des Sciences -Université Libre de Bruxelles, Boulevard du Triomphe - CP 212, 1050 Brussels, Belgium
Email: tlenaert@ulb.ac.be

behavior of social insects can be exploited for designing multi-robot systems. A very interesting feature of these insects is their remarkable capacity for coordination, even without any central authority, allowing them to solve tasks that cannot be handled by one insect alone (e.g., carrying heavy prey). The field of language evolution, on the other hand, studies how language conventions emerge in a population of agents, leading to the formation of a common vocabulary and grammar. In this domain, agents learn from observations or from interaction with other agents to coordinate on a particular signaling system that allows them to communicate in a correct manner.

In many MAS, not all agents are equally likely to interact with each other. Instead, they interact along the links of complex networks, with applications such as network routing (Johnson et al., 2001; Perkins & Royer, 1999) and sensor networks (Akyildiz, Su, Sankarasubramaniam, & Cayirci, 2002). In recent years, it has been recognized that the specific structure of networks plays an important role in the overall behavior of many dynamical processes that take place on them, with examples ranging from disease spreading (May, 2006) to opinion formation (Sood, Antal, & Redner, 2007), and evolutionary game dynamics (Szabó & Fáth, 2007).

Here, we investigate how, and under which conditions, the structure of the interaction network influences the learning dynamics of the agents when they engage in stag-hunt games with their neighbors in the network. In contrast to previous work (Santos, Pacheco, & Lenaerts, 2006b), we focus on a learning scheme that relies purely on trial-and-error (individual-based learning), rather than imitation (social learning). Specifically, we implement each agent as a learning automaton (Narendra & Thathachar, 1989), which uses only its game payoff to improve its game strategy. This study provides theoretical insight into how particular network features affect individual-based learning, and is a first step toward a better understanding of the similarities with, and differences from, what is known from social learning.

Given this setup, we first examine how network features such as clustering and heterogeneity affect the learning process. The amount of clustering in a network is typically represented by its clustering coefficient, which refers to the degree to which neighbors of a certain node are also connected to each other. A prototypical situation that involves clustered networks are MAS embedded in spatial environments, such as sensor networks. Heterogeneity, on the other hand, is here defined as the differences in connectivity among the agents, and corresponds to the variance in the network's degree distribution. This variance is typically small in spatial MAS, due to physical limitations, but may be sizable in problems such as network routing where some nodes may have a significantly more important role than others. We show that neither clustering nor heterogeneity influence the average strategy obtained at the global level, that is, over all agents in the network, in contrast to what has been observed when using social learning (Santos et al., 2006b). Yet, at the level of the individual agents (the local level), the network topology does play an important role, namely (a) clusters of agents with similar strategies emerge depending on the clustering coefficient of the network, and (b) heterogeneity induces variation in the learning speed of the agents in the system. Agents occupying high degree nodes converge faster than those occupying low degree nodes.

Clusters of agents with similar strategies may form spontaneously, but may also arise when agents are allowed to adjust both their strategy and their neighbor relations (Santos, Pacheco, & Lenaerts, 2006a; Van Segbroeck, Santos, Lenaerts, & Pacheco, 2009; Van Segbroeck, Santos, Nowé, Pacheco, & Lenaerts, 2008). Such networks, whose structure is adaptive, occur frequently in MAS, for instance in applications of swarm intelligence (Bonabeau, Dorigo, & Theraulaz, 1999). Indeed, cooperative agents that have the ability to choose their neighbors in the network will look for other cooperative agents to interact with, thereby creating correlations between the strategies of connected agents. We show that the presence of such correlations changes the spectrum of games in which the agents in the network become fully cooperative. A similar observation is obtained when correlating the strategy of the agents with their connectivity, highlighting another important feature of adaptive networks. Since cooperative agents will most likely attract more connections than defective agents, their degree in the network will naturally increase. Here, we demonstrate that the average level of cooperation in the MAS increases under such constraints.

Taken together, our current work justifies both analytically and in simulations why the specific structure of static networks does not affect the global average learning behavior in a typical MAS, and why the presence of correlations that emerge naturally in adaptive networks modifies the collective learning dynamics.

2 Materials and methods

We consider a network of constant size N. Each node in the network is occupied by exactly one agent. Agents may interact with each other only if they are directly connected in the network. Each agent learns from its interactions, adjusting its strategy accordingly.

2.1 Interactions between agents

Interactions are modeled as symmetric two-player games defined by the payoff matrix

$$\begin{array}{c} \\ C \\ D \end{array}\begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} R, R & S, T \\ T, S & P, P \end{pmatrix} \end{array}. \tag{1}$$

An agent who chooses to cooperate (C) with someone who defects (D) receives the sucker's payoff S, whereas the defecting agent gains the temptation to defect, T. Mutual cooperation (defection) yields the reward R (punishment P) for both agents.

We enter the realm of the stag hunt (Skyrms, 2004) or coordination game whenever these payoff values fulfill the condition

$$S < P < T < R \tag{2}$$

(Macy & Flache, 2002). As in all social dilemmas, mutual cooperation (R) is preferred over mutual defection (P) and over an equal probability of unilateral cooperation and defection ($2R > T + S$). In a well-mixed population the game dynamics will end in either of the two stable, pure Nash equilibria: (C, C) and (D, D). The equilibrium corresponding to mutual cooperation is always Pareto dominant ($P < R$). The defective equilibrium is, however, strictly risk dominant when $R + S < T + P$. We will treat only this social dilemma in this work.

We adopt the common convention (e.g., Santos et al., 2006b) of normalizing the payoff values so that the difference between the payoffs associated with the two Nash equilibria becomes 1, taking $R = 2$ and $P = 1$. Fixing the value of T to 1.5 (other values produce equivalent results) leaves only one parameter S to tune the game. When cooperation and defection are equally likely, the expected payoff $E[\pi_C]$ for cooperating is $\frac{R}{2} + \frac{S}{2}$, whereas the expected payoff $E[\pi_D]$ for defecting equals $\frac{T}{2} + \frac{P}{2}$. It is easy to see that in this case, $E[\pi_C] > E[\pi_D]$ for $S > 0.5$, and $E[\pi_C] < E[\pi_D]$ for $S < 0.5$. Hence, the cooperative equilibrium is strictly risk dominant when $S > 0.5$, whereas the defective equilibrium is strictly risk dominant when $S < 0.5$. When $S = 0.5$, neither cooperation nor defection are strictly risk-dominant, meaning that both Nash equilibria are equally likely to be selected.

2.2 Learning

At each discrete time step t, a randomly selected agent i interacts with a random neighbor j. Each of the two agents chooses its action (C or D) based on its strategy. The strategy of agent i (respectively, j) is encoded as a single time-dependent probability $p_i(t)$ (respectively, $p_j(t)$) to cooperate. Each agent uses the payoff it receives upon interaction to update its strategy. The strategy of agent i is updated according to the rule

$$p_i(t+1) = \begin{cases} p_i(t) + \lambda\beta_i(t)[1 - p_i(t)] & \text{when } i \text{ chooses } C \text{ at time } t \\ p_i(t) - \lambda\beta_i(t)\,p_i(t) & \text{when } i \text{ chooses } D \text{ at time } t \end{cases} \tag{3}$$

and that of agent j similarly. $\beta_i(t) \in [0, 1]$ denotes the feedback agent i obtains at time t. This feedback is given by the game payoff the agent received, divided by the maximum possible payoff value ($R = 2$ in our case, see Equation 2). Equation 3 shows that when an agent chooses to cooperate, its probability to cooperate in the future will increase proportionally to the obtained feedback. Similarly, an agent who chooses to defect will become more likely to defect in the future. The parameter λ, known as the learning rate, specifies the immediate impact of the feedback on the agent's strategy. Agents that learn in this way are essentially finite-action learning automata with a reward-inaction update scheme (Narendra & Thathachar, 1989), which are known to converge to an equilibrium point, that is, a pure Nash equilibrium. Learning automata belong to the class of policy iteration approaches to reinforcement learning (Sutton & Barto, 1998). Policy iteration approaches manipulate the strategies directly, in contrast to value iteration approaches that indirectly represent the strategies via the optimal value function. A key feature of learning automata is that the action selection and policy evaluation are all integrated into one mechanism. More elaborate approaches exist that include forms of exploration. Yet in all cases, they learn and explore purely starting from their current individual state and feedback. As such the automata selected here represent a base-mechanism which can be extended into multiple other policy iteration schemes that separate action selection from the policy representation and policy evaluation. As long as the approach guarantees convergence to a Nash equilibrium, we expect that these extensions will not change the conclusions we draw here.

2.3 Network topology

The first aspect of individual-based learning in networks that we investigate is the direct impact of the network topology. As explained in the introduction, we address two distinct network features: clustering and heterogeneity.

To determine the effect of clustering, we define networks with a varying clustering coefficient

C (Watts & Strogatz, 1998). C is defined as the average of all local clustering coefficients C_i $(0 < i < N)$, which are in turn given by the number of edges between the neighbors of node i divided by the maximum possible number of edges between those neighbors. We start with regular ring lattices of degree Z (with $Z \geq 2$). The nodes of such networks are placed on a ring and connected with their Z nearest neighbors ($\frac{Z}{2}$ on each side). Many neighbors of each node are therefore also neighbors of each other, resulting in a high clustering coefficient. We now gradually decrease this clustering coefficient by rewiring an increasing fraction f of the total number of edges (Santos, Rodrigues, & Pacheco, 2005; Szabó & Vukov, 2004), ultimately leading to a random network whose clustering coefficient approaches 0 (see the inset of Figure 2b). In order to avoid changing the degree distribution of the network, we apply the rewiring procedure introduced by Maslov and Sneppen (2002). The algorithm repeatedly swaps the ends of two randomly selected edges, provided no duplicate edges arise, until the necessary number of edges have been rewired (see Figure 1). We ensure that edges are rewired only once during the entire procedure.

We determine the influence of heterogeneity by comparing the learning process in homogeneous random networks (Santos et al., 2005) with that in random scale-free networks (Albert & Barabási, 2002). Homogeneous random networks represent the extreme case where all nodes have exactly the same number of neighbors, evidenced by a degree distribution that has one single peak. They are random in the sense that edges connect nodes that are chosen randomly, to the extent that the given degree distribution remains satisfied. We generate such networks by randomizing regular ring lattices, applying the algorithm described above for $f = 1$. Scale-free networks form the other extreme case in terms of heterogeneity, having a degree distribution that follows a power law. Several algorithms to construct such networks have been proposed over the years, the most famous one probably being that of Barabási and Albert (1999). The specific construction algorithm defines the exponent γ associated with the degree distribution of the resulting networks. Here, we adopt the configuration model of Molloy and Reed (1995) to construct random scale-free networks with $\gamma = 3$, a typical exponent of many real-life scale-free networks (Albert & Barabási, 2002). The randomness of the networks allows us to study pure heterogeneity effects, detached from any other structural network features associated with a particular construction algorithm.

2.4 Correlations in the network

The next part of our investigation concerns distinct connectivity patterns between agents with different initial strategies. First, we introduce correlations between the initial strategies of these agents (which we refer to as *strategy–strategy correlations*) and study how they affect the learning process. Afterwards, we examine the impact of correlations between the initial strategy of an agent and its degree (*strategy–degree correlations*).

In order to isolate the effect of strategy–strategy correlations, we use homogeneous random networks, thereby excluding specific network features like clustering and heterogeneity from the analysis. The strategy of an agent is determined by a parameter $\delta \in]0, 0.5[$. Initially, 50% of the agents in the network cooperate with a high probability, adopting strategy $p_C = 1 - \delta$. The other 50% adopt strategy $p_C = \delta$ and have therefore a high probability of defecting. Here, we take $\delta = 0.2$. Yet, the results remain qualitatively the same

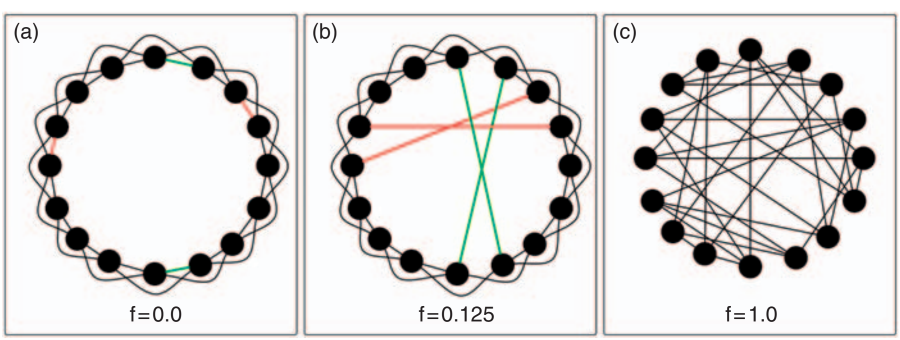


Figure 1. Construction of networks with varying clustering coefficient CC. (a) We start with a regular ring lattice. The neighbors of each node are often also connected to each other, resulting in a high value of CC. (b) Select a fraction f of the edges. Group the selected edges into pairs, and swap the ends of each pair (keeping the degree of each node fixed). This rewiring procedure destroys part of the original clustering. (c) Ultimately, that is, for $f = 1$, we obtain what we call a homogeneous random network, which exhibits an extremely low degree of clustering.

for any value of $\delta \in]0, 0.5[$ (Note that δ should strictly be larger than 0 in order to avoid agents sticking to their initial strategy at any time).

We start by distributing the two types of agents randomly in the network. We then apply a rewiring procedure similar to the one discussed above (i.e., for tuning the clustering coefficient). However, instead of taking all possible links into account for rewiring, we consider only those links connecting agents with a different strategy. By swapping a fraction f_{cd} of those links, connecting the cooperative agents with each other and the defective agents with each other, the initial strategies of connected agents become correlated. When $f_{cd} = 0$, on average 50% of the neighbors of a cooperator (defector) will be other cooperators (defectors). When $f_{cd} = 1.0$, cooperators will only be connected to other cooperators, and defectors to other defectors.

We study strategy–degree correlations by considering random scale-free networks in which agent strategies are initialized in one of three manners, namely (a) positive, (b) negative, and (c) no strategy–degree correlations. In each of the three cases, we first produce the random scale-free network of N nodes and then independently produce N probabilities $p_i(0)$, one for every agent i, from a uniform distribution $U(0, 1)$. In case (a), we place these N probabilities in decreasing order. The node to which the highest (i.e., most cooperative) $p_i(0)$ value is assigned is selected from the existing network with a probability proportional to its degree. The second highest $p_i(0)$ value is assigned similarly, and so on until all probabilities are assigned to nodes in the network. In this way, agents in high degree nodes tend to be more cooperative than those in low degree nodes. We use the same initialization strategy for case (b), but this time we order the N initial strategies in increasing order. As a consequence, agents with many neighbors tend to be less cooperative than those that have only few. Finally, in case (c), the N strategies are randomly distributed in the network, so that strategy–degree correlations are absent.

3 Results and discussion

3.1 Learning in complete networks

Before addressing the problem of learning in networks with a specific topology, we investigate the learning process in complete networks to use as a comparison to ground the other results. The expected probability that an agent j cooperates at time t is given by

$$\langle p \rangle_t = \frac{1}{N} \sum_{i=1}^{N} p_i(t). \tag{4}$$

A random neighbor of j will cooperate at time t with approximately the same probability. The two agents interact and update their strategies according to Equation 3. The expected change in $\langle p \rangle$ resulting from this update is given by

$$\begin{aligned}\Delta \langle p \rangle &= 2\, \langle p \rangle_t\, \lambda\, E[\beta_C]\big(1 - \langle p \rangle_t\big) - 2\big(1 - \langle p \rangle_t\big)\lambda\, E[\beta_D]\langle p \rangle_t \\ &= \langle p \rangle_t (1 - \langle p \rangle_t)[2\, \lambda(E[\beta_C] - E[\beta_D])]\end{aligned} \tag{5}$$

where $E[\beta_C]$ ($E[\beta_D]$) denotes the expected feedback an agent obtains when cooperating (defecting). Note the correspondence between Equation 5 and the standard replicator equation from evolutionary game theory (Hofbauer & Sigmund, 1998).

Equation 5 implies that cooperation will become more likely when $E[\beta_C] > E[\beta_D]$, whereas the opposite will be true when $E[\beta_C] < E[\beta_D]$. The transition, that is, $E[\beta_C] = E[\beta_D]$, occurs for

$$\langle p \rangle^* = \frac{P - S}{R - S - T + P} \tag{6}$$

Hence, when $\langle p \rangle_0 > \langle p \rangle^*$ (respectively, $\langle p \rangle_0 < \langle p \rangle^*$), all agents will learn to cooperate (respectively, defect) 100% of the time.

3.2 Global effects of the network topology

As a first step to investigate the effect of the network topology, we simulate the learning process in regular ring lattices, homogeneous random networks, as well as random scale-free networks. Comparison of the results for regular ring lattices with those for homogeneous random networks demonstrates the effect of the clustering coefficient. By comparing homogeneous random networks with scale-free networks, we address the role of heterogeneity.

We simulate the learning process on each type of network for different values of the sucker's payoff $S \in [0, 1]$. At the start of each simulation, we initialize the strategy $p_i(0)$ of each agent i to 0.5, making cooperation and defection equally likely. Each simulation continues until a stationary regime has been reached, that is, until the average strategy of the agents changed less than 10^{-5} during the last 10^6 time steps. For each value of S, we run 1, 000 independent simulations and calculate the strategy the agents learn on average.

Figure 2a shows that this strategy does not depend on the specific structure of the interaction network. Indeed, in each of the three types of networks, a transition from full defection to full cooperation occurs at $S = 0.5$. This transition corresponds exactly to the one predicted by Equation 6 for $\langle p \rangle_0 = 0.5$. This means that on average the network does not change the probability

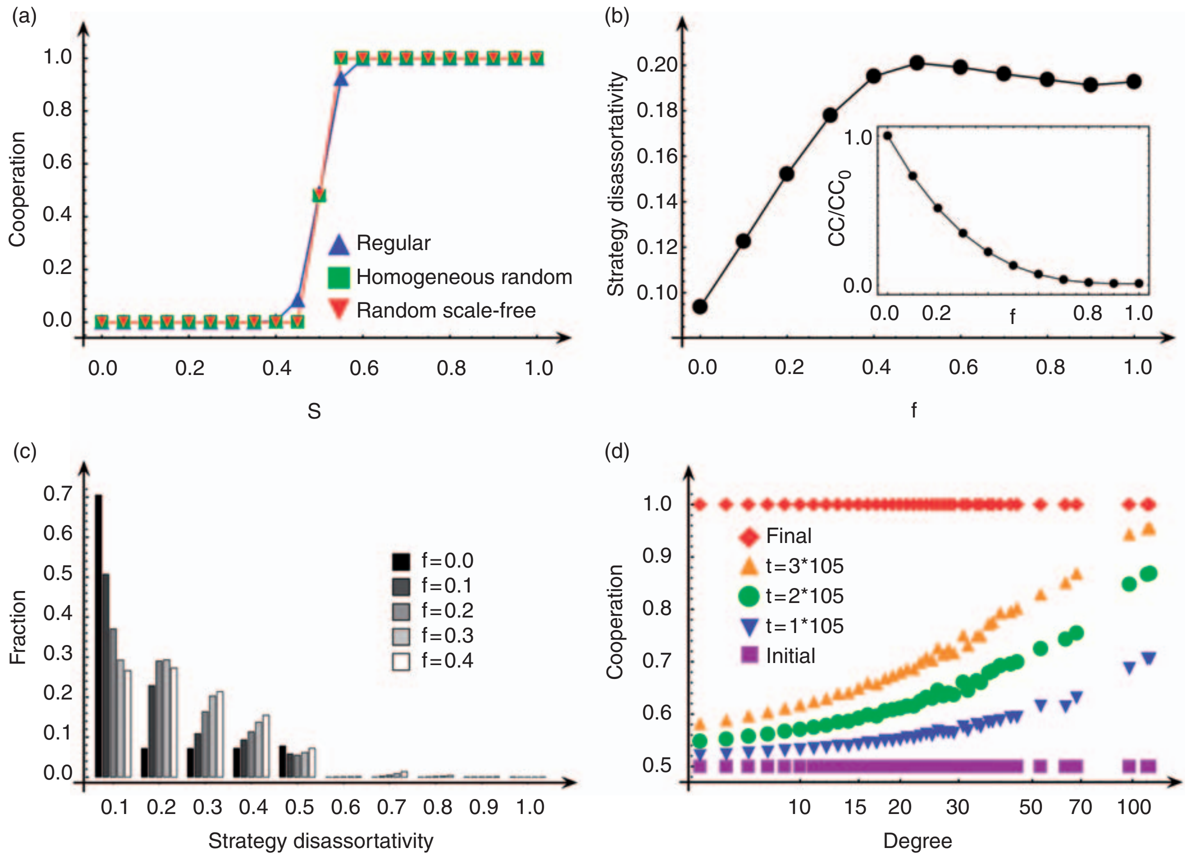


Figure 2. Global and local behavior in networks. In all four panels, we use $p_i(0) = 0.5$, $\lambda = 0.01$, $N = 10^3$, and $Z = 8$ as parameter values. (a) Average strategy of the agents as a function of the sucker's payoff S for different network structures. The specific type of network does not affect the critical value of S at which the transition from full defection to full cooperation occurs. (b) Average strategy disassortativity $\mathcal{D}$ for $S = 0.5$ in networks created from regular ring lattices by rewiring a fraction f of the edges. The inset shows the clustering coefficient C of the resulting networks, divided by the clustering coefficient C_0 of the original lattice. The strategy an agent learns is more similar to that of its neighbors when the network has a higher clustering coefficient. (c) Fraction of agents with individual strategy disassortativity $\mathcal{D}_i$ within the ranges specified by the x-axis, using the same settings as in (b). (d) Average strategy of the agents as a function of their position (degree) in a random scale-free network, and this at different points in (simulation) time t, given $S = 0.7$. Agents that occupy high degree nodes will converge much faster than those occupying low degree nodes.

that an agent perceives either cooperation or defection, so that the learning process in any of the three types of networks becomes equivalent to that in a complete network. In other words, wiring agents in a different way will not have any direct impact on the average strategy of the agents in the network.

3.3 *Local effects of the network topology*

The individual strategies that make up the average behavior exhibit some remarkable features. The first one concerns the final distribution of agent strategies in networks with varying clustering coefficient, which are created following the methodology described in Section 2.3. The value of the sucker's payoff S is set to 0.5, so that neither cooperation nor defection dominates. Running simulations with this parameter setting will produce networks that have on average 50% fully cooperative agents and 50% fully defective agents. Although the actual clustering coefficient does not influence the final fraction of each strategy (see Figure 2a), we observe that it does play an important role in the position of cooperators and defectors in the network. We visualize this effect by measuring the average difference in strategy between each agent i and its neighbors. We refer to this value as the *strategy disassortativity* of the agent, which is defined as

$$D_i(t) = \frac{1}{N_i} \sum_{j \in N(i)} |p_i(t) - p_j(t)|, \tag{7}$$

where $N(i)$ (respectively, N_i) denotes the neighbors (respectively, number of neighbors) of agent i. This value is similar to the index of dissimilarity defined in the context of social segregation literature

(Schelling, 1971, 1978). Figure 2b shows the average strategy disassortativity

$$\mathcal{D} = \frac{1}{N}\sum_{i=1}^{N} D_i \tag{8}$$

of all agents in the network at the end of the simulations, as a function of the fraction of edges f that are rewired in a regular ring lattice (see Section 2.3). When f is small, the strategy of the agents is very similar to that of their neighbors, which means that clusters of cooperators and defectors emerge. This is nicely illustrated in Figure 2c, which shows the distribution of the strategy disassortativity of all agents in the network. When $f = 0$, the strategy disassortativity of almost 70% of the agents is smaller than 0.1. Rewiring more and more edges inhibits the spontaneous formation of clusters, a result that is clearly visible in both Figures 2b and 2c. This phenomenon is correlated with the impact of the rewiring procedure on the clustering coefficient C of the network. The inset of Figure 2b shows that C drops very rapidly when f increases from 0 to 0.5, but remains almost constant afterwards. The average strategy disassortativity $\mathcal{D}$ follows the same trend.

The second local effect of the network topology concerns the impact of heterogeneity on the learning speed of different agents. Since high degree nodes engage in more interactions than the low degree nodes, they will converge much faster to an equilibrium. This is nicely illustrated in Figure 2d, where we show the average strategy of nodes with given degree at different points in simulation time. Initially, all agents cooperate 50% of the time and defect 50% of the time. For $S = 0.7$, the hubs will learn to cooperate quickly and have almost converged after 3×10^5 iterations. At this point in the simulation, the low degree nodes still have a very long way to go. On average, the stationary regime is reached after 5×10^6 iterations. When $S < 0.5$, a similar phenomenon occurs. In this case, the high degree nodes will learn to defect much more rapidly than the low degree nodes.

Thus, both network clustering and heterogeneity induce specific effects at the individual level which are not visible when considering only the average behavior in a MAS. Spontaneous clustering of agents with similar strategies does not only occur in certain static network structures, but is also a typical characteristic of adaptive networks (Santos et al., 2006a; Van Segbroeck, Santos, Lenaerts et al., 2009; Van Segbroeck, Santos, Nowé et al., 2008), as explained in the introduction. In essence, network adaptivity introduces two types correlations in the network, namely (a) strategy–strategy correlations, and (b) strategy–degree correlations. In the next subsection, we investigate how each of these correlations affects the overall game behavior of agents in a network.

3.4 Influence of strategy correlations in the network

As discussed in Section 2.4, we study strategy–strategy correlations by considering homogeneous random networks with two types of agents: those that initially cooperate most of the time and those that initially defect most of the time. Strategy–strategy correlations arise by rewiring a fraction f_{cd} of the edges between agents with different initial strategies. Rewiring more and more of these edges segregates the network into two sparsely connected groups: one containing the cooperative agents and the other containing the defective ones. As a result of this segregation, the expected feedback of the agents in the cooperative group will differ from that of the agents in the defective group. As a consequence, agents in the cooperative group will remain cooperative for smaller values of S than usual, whereas those in the defective group will remain defective for larger values of S than usual (see Figure 3a).

We can again make an analytical approximation of the critical value of S at which each of these transitions occurs. In the previous subsection, we have seen that Equation 5 provides an accurate description of the learning process in case all agents have the same initial strategy, irrespective of the network topology. This is, however, no longer true when the network is segregated into two groups of agents with a fundamentally different strategy. Indeed, in that case we can no longer estimate the probability of receiving cooperation by the global average $\langle p \rangle$. Instead, the agents in the first group receive cooperation with probability $\langle p \rangle^1$ and those in the second group with probability $\langle p \rangle^2$. As such, the following two equations describe the learning process of the MAS

$$\begin{aligned}\Delta\langle p\rangle^1 &= \langle p\rangle_t^1\left(1-\langle p\rangle_t^1\right)\left[2\,\lambda\left(E\left[\beta_C^1\right]-E\left[\beta_D^1\right]\right)\right]\\ \Delta\langle p\rangle^2 &= \langle p\rangle_t^2\left(1-\langle p\rangle_t^2\right)\left[2\,\lambda\left(E\left[\beta_C^2\right]-E\left[\beta_D^2\right]\right)\right]\end{aligned} \tag{9}$$

where $E[\beta_C^1]\left(E[\beta_C^2]\right)$ denotes the expected feedback an agent in the first (second) group obtains when cooperating. Similarly, $E[\beta_D^1]$ and $E[\beta_D^2]$ denote the expected feedback for defection in each of the two groups. Note that the two equations are coupled, in the sense that the expected feedback values depend on both $\langle p \rangle^1$ and $\langle p \rangle^2$. The strength of this coupling depends on the value f_{cd}. When f_{cd} is sufficiently large, we can treat the two equations as if they were independent of each

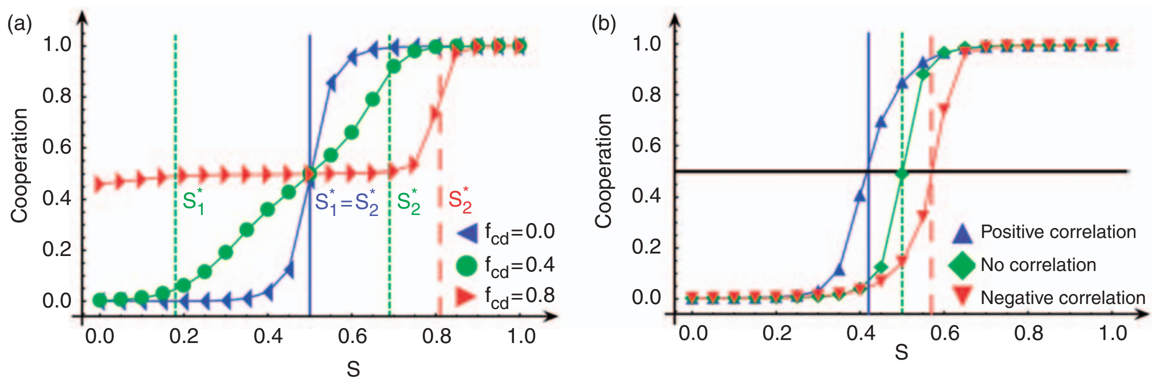


Figure 3. Correlation effects. In both panels, we use $\lambda = 0.0025$, $N = 10^3$, and $Z = 8$ as parameter values. (a) The effect of strategy–strategy correlations. We distinguish two types of agents: those that are initially mostly cooperative and those that are initially mostly defective. The correlation factor f_{cd} determines the level of segregation between these two groups of agents. As soon as $f_{cd} > 0$, agents in the cooperative (defective) group learn to cooperate (defect) even when $S < 0.5$ ($S > 0.5$). $S_1^*(S_2^*)$ corresponds to the analytical approximation of the critical S value for agents in the cooperative (defective) group. Note that S_1^* for $f_{cd} = 0.8$ lies outside the frame to the left. (b) Introducing correlations between the initial strategies of the agents and their degree shifts the critical value of S where the transition from full defection to full cooperation takes place. Having more cooperative (defective) agents in high degree nodes increases the basin of attraction of the cooperative (defective) equilibrium.

other, as illustrated in Figure 3a. The vertical lines in this figure represent the values $S_1^*(S_2^*)$ at which $E[\beta_C^1] = E[\beta_D^1](E[\beta_C^2] = E[\beta_D^2])$, assuming that the agents interact with a member of the first group with probability $\frac{1}{2} + \frac{1}{2} f_{cd}(\frac{1}{2} - \frac{1}{2} f_{cd})$, and with one of the second group otherwise. Figure 3a shows that the accuracy of S_1^* and S_2^* improves when increasing f_{cd}. When considering for instance $f_{cd} = 0.8$, the analytical prediction fits perfectly with the simulations. When f_{cd} is too small, the two groups of agents are too much intertwined, resulting in rather poor analytical predictions.

3.5 Influence of strategy–degree correlations in the network

Finally, we investigate the effect of strategy–degree correlations. We address this issue by considering random scale-free networks with initial agent strategies selected from a uniform distribution $U(0, 1)$. Correlations between strategy and degree are introduced following the methodology discussed in Section 2.4. In the case of positive strategy–degree correlations, the high degree nodes are initially more cooperative than the low degree ones. As a consequence, the expected feedback of the agents becomes degree dependent. Based on this observation, one would expect that the learning process can only be described analytically by considering one equation for each degree present in the network. We can, however, calculate analytically the average probability that a cooperative action occurs during each interaction:

$$\langle p \rangle = \sum_j P(j) \left[\sum_k \frac{kP(k) \frac{\langle p \rangle_j + \langle p \rangle_k}{2}}{\sum_s sP(s)} \right]. \tag{10}$$

$P(j)$ denotes the probability of selecting a node with degree j, which depends on the degree distribution of the network. The agent associated with this node interacts with a randomly selected neighbor, who is of degree k with probability proportional to $kP(k)$ (Pastor-Satorras & Vespignani, 2001). The average probability that a cooperative action takes place during this interaction is given by $\frac{\langle p \rangle_j + \langle p \rangle_k}{2}$, where $\langle p \rangle_i$ denotes the average strategy of nodes with degree i. We now calculate the values $P(i)$ and $\langle p \rangle_i$ observed in the simulations. Plugging the resulting value for $\langle p \rangle$ into Equation 5 gives us an estimate of the critical S value at which cooperation and defection are equally likely. This value, indicated by the vertical line in Figure 3b, shows a remarkable fit with the simulation results. The presence of positive strategy–degree correlations increases $\langle p \rangle$, so that the transition from full defection to full cooperation shifts in favor of the cooperative strategy. In the case of negative strategy–degree correlations, similar arguments explain why the transition from defection to cooperation shifts in favor of the defective strategy.

4 Conclusion

In this work, we have examined the role of network clustering and heterogeneity in MAS where the agents, located at the nodes of the network, learn to coordinate their actions in a stag-hunt game. The results show that neither of these topological features affects the average level of cooperation in the system. Moreover, the outcome of the learning process in networks with different topologies is equivalent to that in complete networks, for which we derived analytical predictions.

However, closer inspection of the learning dynamics reveals that network clustering and heterogeneity induce specific phenomena at the level of the individual agents that do not occur in complete networks: (a) heterogeneity alters the convergence speed of the agents, that is, agents located in the hubs of the network tend to converge much faster than those located in sparsely connected nodes, and (b) agents with similar strategies tend to cluster depending on the clustering coefficient of the network.

The first feature suggests that the initial environment perceived by the agents in high degree nodes defines to a large extent the overall final behavior of the MAS. The agents in the hubs are by far the first to converge, which gives them the opportunity to change the environment of the other agents and, as such, alter their learning process. This issue may have important implications in situations where cultural evolution is at work as well, that is, when individuals do not learn by trial-and-error only, but also by imitation (Boyd & Richerson, 2005). If individual-based learning is more costly than social learning, then the agents in the network may prefer to imitate the behavior of their neighbors, rather than having to learn it themselves. Depending on the network structure, there may be an optimal distribution of the two learning schemes in the population, minimizing the total learning cost without compromising the population's capacity to adapt to changing environmental conditions (Rendell et al., 2010).

The second feature implies that although the clustering coefficient does not affect the average fraction of cooperation in the MAS, it does change the actual payoffs the agents receive. Indeed, cooperative agents will collect a higher payoff in networks with a higher clustering coefficient, because such networks allow them to form clusters with other cooperative agents, protecting them from interactions with defective agents. Hence, it will be easier for agents to coordinate their actions when a certain amount of clustering is introduced in the interaction topology. Without clustering, agents seem to be more exposed to a mixed environment and find it more difficult to agree upon their actions. Such effects have also been observed in the Talking Heads experiment (Steels, 1999). A common language evolved more easily when the agents were forced to remain in the same location for a while, before migrating to other groups in other locations.

In addition, we know from previous research (e.g., Santos et al., 2006a) that clusters of cooperative and defective agents, as observed in networks with high clustering coefficient, are known to emerge in MAS where the agents have the ability to individually choose their interaction partners. For instance, when agents decide to change partner based on the outcome of the game, all agents look for cooperative agents to interact with, while trying to avoid contact with defective agents. Together, cooperators may form clusters in order to produce strong islands in a highly competitive environment. On the other hand, the fact that all agents try to interact with cooperators creates correlations between agent strategy and degree. From a learning perspective, adaptive network structures are important since they alter the feedback the agents receive about their behavior in the current environment. By simply introducing the types of correlations one can expect in MAS with adaptive networks, we showed here that the final level of cooperation indeed changes. Furthermore, in this case, we are also able to predict analytically where the transition from cooperation to defection will take place. Even though the present work already suggests what can be expected from learning in adaptive networks, we leave the explicit analysis to future work. At that point, differences in timescales between network dynamics and learning dynamics will require intensive examination.

For now, the takeaway message is that topological differences do not produce a different final behavior in coordination games when agents are learning individually. Correlations, on the other hand, which can be produced by different mechanisms such as reciprocity and partner choice (Hammerstein, 2003), may effectively alter the way agents perceive the game and, as a consequence, shift the final outcome.

Acknowledgments

SVS, SdJ, and AN acknowledge financial support from F.W.O. Belgium. TL acknowledges support from F.N.R.S. Belgium. FCS acknowledges the financial support from FCT Portugal. We also want to thank Karl Tuyls for reviewing a draft of the manuscript. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Akyildiz, I., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). A survey on sensor networks. *IEEE Communications Magazine*, *40*(8), 102–114.

Albert, R., & Barabási, A. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*, 47–97.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512.

Binmore, K. G. (1991). Fun and games: A text on game theory. Lexington: D.C. Heath.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: From natural to artificial systems*. New York: Oxford University Press.

Boyd, R., & Richerson, P. (2005). *The origin and evolution of cultures*. New York: Oxford University Press.

Christiansen, M., & Kriby, S. (2003). *Language evolution*. Oxford: Oxford University Press.

Dorigo, M., Tuci, E., Gross, R., Trianni, V., Labella, T., & Nouyan, S. et al (2005). The swarm-bots project. *Swarm Robotics*, *3342*, 31–44.

Hammerstein, P. (Ed.) (2003). *Genetic and cultural evolution of cooperation*. Cambridge, MA: MIT Press.

Hofbauer, J., & Sigmund, K. (1998). *Evolutionary games and population dynamics*. Cambridge: Cambridge University Press.

Johnson, D., Maltz, D., & Broch, J. (2001). DSR: The dynamic source routing protocol for multi-hop wireless ad hoc networks. In C. E. Perkins (Ed.), *Ad hoc Networking* (Vol 5, pp. 139–172). Boston: Addison-Wesley Longman.

Kraus, S. (1997). Negotiation and cooperation in multi-agent environments. *Artificial Intelligence*, *94*(1-2), 79–98.

Macy, M., & Flache, A. (2002). Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences of the USA*, *99*(3), 7229–7236.

Maslov, S., & Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science*, *296*, 910–913.

May, R. M. (2006). Network structure and the biology of populations. *Trends in Ecology & Evolution*, *21*(7), 394–399.

Molloy, M., & Reed, B. (1995). A critical point for random degree graphs with a given degree sequence. *Random Structures and Algorithms*, *6*, 161–180.

Narendra, K., & Thathachar, M. (1989). *Learning automata: An introduction*. Boston, MA: Prentice-Hall International.

Osborne, J. O., & Rubinstein, A. (1994). *A course in game theory*. Cambridge, MA: MIT Press.

Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Physical Review Letters*, *86*(14), 3200–3203.

Perkins, C., & Royer, E. (1999). Ad-hoc on-demand distance vector routing. In *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications*, *2*, 90–100.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., & Feldman, M. W. et al (2010). Why copy others? Insights from the social learning strategies tournament. *Science*, *328*(5975), 208–213.

Santos, F. C., Pacheco, J. M., & Lenaerts, T. (2006a). Cooperation prevails when individuals adjust their social ties. *PLoS Computational Biology*, *2*(10), 1284–1291.

Santos, F. C., Pacheco, J. M., & Lenaerts, T. (2006b). Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proceedings of the National Academy of Sciences of the USA*, *103*, 3490–3494.

Santos, F. C., Rodrigues, J. F., & Pacheco, J. M. (2005). Epidemic spreading and cooperation dynamics on homogeneous small-world networks. *Physical Review E*, *72*, 056128.

Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*, 143–186.

Schelling, T. (1978). *Micromotives and macrobehavior*. New York: Norton.

Skyrms, B. (2004). *The stag hunt and evolution of social structure*. Cambridge, UK: Cambridge University Press.

Sood, V., Antal, T., & Redner, S. (2007). Voter model on heterogeneous networks. *Physical Review E*, *77*, 041121.

Steels, L. (1999). *The talking heads experiment: Words and Meanings*. Antwerpen: Special pre-edition for LABORATORIUM.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge MA: MIT Press.

Szabó, G., & Fáth, G. (2007). Evolutionary games on graphs. *Physics Report*, *446*(4–6), 97–216.

Szabó, G., & Vukov, J. (2004). Cooperation for volunteering and partially random partnerships. *Physical Review E*, *69*(3), 036107.

Van Segbroeck, S., Santos, F. C., Lenaerts, T., & Pacheco, J. M. (2009). Reacting differently to adverse ties promotes cooperation in social networks. *Physical Review Letters*, *102*, 058105.

Van Segbroeck, S., Santos, F. C., Nowé, A., Pacheco, J. M., & Lenaerts, T. (2008). The evolution of prompt reactions to adverse ties. *BMC Evolutionary Biology*, *8*, 287.

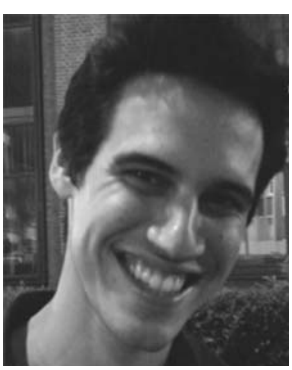
Watts, D. J., & Strogatz, S. (1998). Collective dynamics of small-world networks. *Nature*, *393*, 440–442.

About the Authors

Sven Van Segbroeck (Antwerp, Belgium, 1982) obtained recently his Ph.D. in Computer Science at Vrije Universiteit Brussel (Belgium), where he also finished his graduation in both Mathematics and Computer Science. As of October 2010, he works at Université Libre de Bruxelles (Belgium), as a researcher of the Fund for Scientific Research of the Belgian Francophone community (F.R.S.-FNRS). His research covers several topics in the field of Complex Adaptive Systems, ranging from learning dynamics and evolutionary game dynamics to the dynamics of infectious diseases.

Steven de Jong (Heerlen, The Netherlands, 1981) studied Knowledge Engineering and Artificial Intelligence in Maastricht (The Netherlands), where he obtained his Ph.D. in Artificial Intelligence in 2009 with a thesis entitled "Fairness in Multi-Agent Systems". He spent a year at the Vrije Universiteit Brussel (Belgium) as a postdoctoral researcher. He is currently employed at the Department of Knowledge Engineering of Maastricht University (The Netherlands) as a postdoctoral researcher in adaptive multi-agent systems, evolutionary game theory and swarm intelligence.

Ann Nowé (Kortrijk, Belgium, 1965) received the M.S. degree from Universiteit Gent, Belgium, in 1987, where she studied mathematics with a minor in computer science, and the Ph.D. degree from Vrije Universiteit Brussels (VUB), Belgium, in collaboration with Queen Mary and Westfield College, University of London, U.K., in 1994. The subject of her dissertation was located in the intersection of computer science (AI), control theory (fuzzy control), and mathematics (numerical analysis, stochastic approximation). Currently she is a full professor at the VUB and co-head of the Computational Modeling lab. Her major area of interest is machine learning, in particular reinforcement learning and learning in multi-agent systems.



Francisco C. Santos (Lisbon, Portugal, 1981) studied Theoretical Physics at the University of Lisbon (Portugal) and earned his Ph.D. in Computer Science at the Université Libre de Bruxelles (Belgium, 2007). After 3 years as an associate researcher in Brussels, he is currently an associate researcher at the Computer Science Department of the New University of Lisbon (Portugal). His interests span several aspects of complex adaptive systems, from the structure of social networks and the dynamics of human cooperation, to the evolution of social norms and reputation-based systems.

Tom Lenaerts (Brasschaat, Belgium, 1972) studied Computer Science at the Vrije Universiteit Brussel (Belgium), where he also obtained his Ph.D. in 2003. Afterwards he held different postdoctoral positions, performing research on different topics, ranging from evolutionary game theory to protein bioinformatics research. Since 2008, he is associate professor at the Université Libre de Bruxelles (Belgium) where is co-heading the Machine Learning group in the Department of Computer Science. His current research interests cover different interdisciplinary topics, including the dynamics of learning and evolution, the evolutionary dynamics of chronic myeloid leukemia, and the mechanisms of protein allostery within signal transduction.