EUROPHYSICS LETTERS 15 November 2003

Europhys. Lett., **64** (4), pp. 557–563 (2003)

Motif selection in a model of evolving replicators: The role of surfaces and limited transport in network topology

S. C. Manrubia[1] and J. F. Poyatos[1,2]

[1] *Centro de Astrobiología, CSIC-INTA - Ctra. de Ajalvir km. 4 28850 Torrejón de Ardoz, Madrid, Spain*

[2] *Structural and Computational Biology Programme, Spanish National Cancer Centre (CNIO) - Melchor Fdez. Almagro 3, 28029 Madrid, Spain*

(received 2 June 2003; accepted in final form 10 September 2003)

PACS. 89.75.-k – Complex systems.
PACS. 89.75.Fb – Structures and organization in complex systems.
PACS. 89.75.Hc – Networks and genealogical trees.

Abstract. – A system of replicators evolving on a surface is analysed. Population dynamics together with environmental constraints determine the network of interactions among species. We show that the mobility of individual elements, implemented as different values of the diffusion coefficient, plays a major role in shaping the topology of the emerging network by selecting simple network motifs during its evolution. This effect is fully quantified through the relative abundance of 3-node motifs and the clustering coefficient distribution. For large enough diffusion the behaviour crosses over to that of a well-stirred reactor, and parasites are able to disrupt the otherwise dynamically stable coexistence of species.

The study of catalytic networks has been an active field of research since the seminal work of Eigen and Schuster on hypercycles [1]. In this kind of system, a number of different chemical species interact cross-catalytically. If closed cycles exist, the system is self-maintained. These networks lie at the root of current theories describing the emergence and evolution of chemical organization. Systems of replicators are the basic constituent of evolving systems, as well. So far, most analyses have been performed in networks with fixed interactions among species, where the population numbers are the relevant dynamical variables. However, the interactions among species evolve as well, though at a larger time scale: Network architecture and population dynamics co-evolve to originate the structured systems that are found in Nature.

At present, the study of the topological architecture of networks of interacting elements is opening new avenues in our understanding of the dynamics of such collective systems [2, 3]. Biological and technological networks such as food webs, genetic networks, the World Wide Web, or those of social acquaintances, are far from being randomly organized. They indeed result from dynamical processes which might modify both the number and types of the nodes and the amount of connections present as the network evolves. Often, even the very nodes have an internal structure which changes in time. As complex evolving systems, networks are

thus subjected to selection principles, and their structural properties reflect the mechanisms and dynamical processes acting on them.

Different topological quantities have been used to characterize the architecture of networks, among them the clustering coefficient [4], the average distance between any pair of nodes, and the distribution of links per node. These statistics have permitted to establish a classification of networks and to recognize universal patterns in their structure [5, 6]. Recently, network motifs have been identified as relevant, though simple, units widely used in the construction of complex architectures [7–9]. A motif is a subgraph of the network formed by a small set of nodes (three or four, typically) with their interactions. The repetition of these simple patterns of interconnections in large networks seems to be related to the function to be fulfilled by the network (*e.g.*, information or energy transmission), and to the design principles which have shaped it. An additional interest of network motifs lies in the possibility for them to function as simple building blocks, able to replicate as a whole, and thus to behave effectively as a new level on which selection can act. In this way, it would be possible to get faster evolving and more complex networks in shorter time scales.

In this letter we study the architecture of a network which results from the dynamics of a system of simple catalytic replicators evolving on a surface. In contrast to other analysis which have identified universal statistical properties of certain networks, we will show how changes in environmental constraints (specifically, in the mobility of single elements) translate into different topological solutions. Our motivation to perform the present study under the conditions to be described is threefold. First, simple replicator dynamics taking place among a variable number of different species have been explored as model systems for a range of phenomena [1, 10–12]. In addition, replication is the crucial, minimal requirement for any system to undergo competition between akins and eventually Darwinian evolution. Second, the introduction of novelty in the form of replicators with new catalytic properties causes a continuous turnover of species, thus allowing the emergence and selection of interaction networks [13]. Third, surfaces play an extremely relevant role in stabilizing the dynamics of evolving replicator systems. In particular, they limit the detrimental action of parasites by restricting the mobility of the interacting agents [14, 15]. Considering together all these elements allows a study of the topology of networks evolving under endogenous design principles (consequence of the system's own dynamics) and further subjected to environmental constraints (or exogenous design principles, *e.g.*, the dimensionality of the space where the dynamics take place).

Our main result is that changes in the environment, here introduced in the form of a varying diffusion rate, significantly alter the topology of the interaction network. An increasing diffusion rate effectively interpolates between a situation where local processes are dominant and a mean-field situation equivalent to that taking place in a well-stirred reactor. In the latter limit, the system will eventually collapse due to the action of parasites. It will be shown that the relative abundance and success of three-node motifs and the clustering coefficient [4] reflect how the emerging network adapts to different environmental conditions.

The precise implementation of the dynamics, which is partially based on [13], goes as follows. We consider a two-dimensional lattice of size $N \times N$ with open boundary conditions. Initially, the lattice is empty. At each time step we add a single element belonging to a species i. Each species is characterized by its catalytic interactions with the other species. These interactions are defined through a sparse matrix $c(i, j)$, whose elements take value 0 or 1. If $c(i, j) = 1$, then i could potentially catalyse the production of j. This reaction becomes effective whenever two molecules of species i and j happen to meet at adjacent sites. Then, the catalytic reaction $i + j \to i + 2j$ takes place. If $c(i, j) = 0$ no interaction occurs. When a new species enters the system, there is a probability $p \ll 1$ for each element of the catalysis matrix to take value unity. Individual molecules decay at a rate δ and diffuse at a rate D.

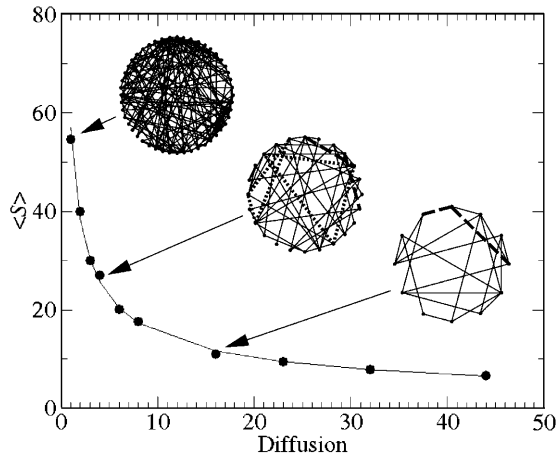

Fig. 1 – Decrease in the average number of species $\langle S \rangle$ coexisting at the statistically stationary state for increasing values of the diffusion coefficient. The networks show three particular cases for $N = 128$, $p = 0.05$ and $\delta = 0.005$ with $D = 1$, $D = 4$, and $D = 16$. Numerical interpolation returns a functional relation of the form $\langle S \rangle \propto D^{-0.57 \pm 0.01}$ (solid curve). Dashed lines stand for 2-cycles, dotted lines for 3-cycles. The largest network has seven 2-cycles and seven 3-cycles (not explictly shown). All our results have been averaged over 10^6 time steps.

Diffusion is modelled by means of the Toffoli-Margolus algorithm [16]. The elementary step of the algorithm consists in dividing the whole lattice in 2×2 squares. The states of the four sites are then moved clockwise or counterclockwise, the direction being chosen at random and with equal probability for each of the $N/2 \times N/2$ squares. The lattice is partitioned by starting in the $(x, y) = (1, 1)$ or in the $(x, y) = (2, 2)$ site, alternatively. The number of times that this elementary step is applied per reaction step coincides with the value of the diffusion coefficient D. It has been shown that, in its continuum version, this algorithm (designed in origin to accurately simulate lattice gas dynamics) exactly translates into a diffusion equation.

On the average, and during the initial transient, there will be $1/\delta$ molecules diffusing on the surface at a rate D. The state of the system changes once the first autocatalytic cycle appears. This transition occurs after a time of order $\tau \sim (N\delta/p)^2$, *i.e.*, the probability that two molecules catalyse each other's production and meet on the surface is the product of the cross-catalysis probability p^2/δ times the probability that they meet before decaying, $(\delta N^2)^{-1}$. At this point, the population of the two species involved grows until they almost fill the whole surface. This situation facilitates that new species establish, and the total diversity, measured as the number S of different species present, grows. Shortly after, a statistically stable equilibrium where the average number of species $\langle S \rangle$ keeps approximately constant is reached. Fluctuations in the amount of species present are narrow, and the distribution of S-values is approximately Gaussian. This result is in contrast with that obtained in a similar model where space was not made explicit [13], and large extinctions were observed. In the present case, it is the confinement of the dynamics to the surface that limits the action of parasites and allows, thanks to the slow diffusion of species, local extinctions driven by population dynamics to be counterbalanced by migration from adjacent areas [17].

At the statistically stationary state, and for low enough diffusion, the system has a rich spatial structure and high diversity. In fig. 1 we represent the typical decrease of $\langle S \rangle$ with increasing mobility, for other parameters fixed. The characteristic scale at which interactions between neighbors can take place grows proportionally to $\sqrt{D}$. For increasing D, the con-

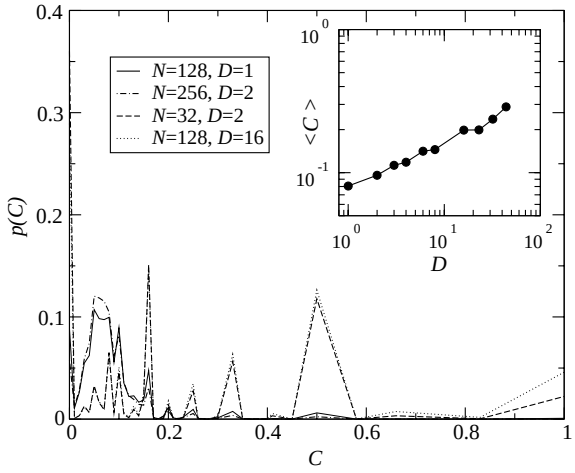

Fig. 2 – Distribution $p(C)$ for different parameters. Note the approximate invariance of the distribution $p(C)$ for constant D/N. The insert ($N = 128$) shows the variation in $\langle C \rangle$ as a function of D: more strongly connected networks are selected and spatial structure is lost in order to ensure global persistence. Other parameters as in fig. 1.

finement of particles to their spatial neighborhood is less and less efficient, and species which are kept apart for low D "see" each other for larger D. Eventually, parasites can have access to a large number of altruistic species [10], and the system becomes no longer sustainable. Restriction of movement to the surface can also be viewed as a kind of compartmentalization applied to the interacting elements. Thus, in a large enough system, independent evolution within subareas which are far away from each other becomes feasible.

It is important to emphasize the relevance of the restricted mobility for the architecture of the interactions. As inserts in fig. 1 we have represented the networks of potential interactions between species, that is the nodes and links in the matrix $c(i, j)$. Apparently, the average number of links per species decreases with increasing diffusion, an observation which might seem to contradict the fact that a greater degree of mixing implies contact with a higher number of species. This is a consequence of the difference between effective and potential interactions. When mobility is restricted, replicators will only effectively interact with their closest spatial neighborhood. The appearance of self-sustained subpopulations supports a richer diversity, increasing the opportunity for a newcomer to have more connections. But spatial isolation implies that most of such connections are only potential ones and the emergence of cliquishness among species is thus limited. This qualitative observation can be substantiated quantitatively: The clustering coefficient measures how important for the viability of the system is that the set of species connected to a given one is also linked, thus highlighting cliquishness in network structure.

We have calculated both the distribution of clustering coefficients $p(C)$ and its average $\langle C \rangle$ for each value of the diffusion D. Our results are represented in fig. 2 for different system sizes and D-values. We observe that $\langle C \rangle$ always increases for increasing D. This implies that the higher the mobility, the lower the number of areas where evolution proceeds in an approximately independent fashion. The distribution $p(C)$ shows how the weight of different C values is shifted from lower to higher clustering coefficients as D increases. In the example shown in the insert of fig. 2, parasites disrupt the system for $D \simeq 50$.

In order to further quantify the topological architecture of the matrix $c(i, j)$, we have

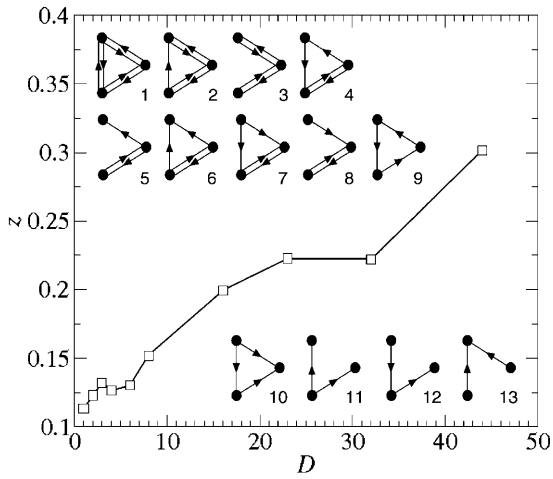

Fig. 3 – All thirteen 3-node directed motifs ranked according to accumulated population. The coefficient z is the ratio between the number of motifs with $\lambda = 0$ (motifs 10 to 13) and that of motifs with at least one cycle ($\lambda_{1\text{-}9} \geq 1$). For increasing D, parasites become increasingly abundant. See main text for details; parameters as in fig. 2.

calculated the average number of 3-node motifs present at stationarity for increasing D. Only because movement is restricted to the surface can the global network be analysed in terms of 3-species motifs: effective interactions take place among few different species. Our results are consistent with this assumption. We have first explored the "success" of each motif quantified by the fraction of the total population accumulated by the three involved species. Numerical simulations return a rank ordering of the motifs according to such criterion. This result is represented in fig. 3. The ranking obtained is independent of the external parameters of the system, and just reflects the growth rate of the populations interacting according to each motif. Such growth rate is proportional to the largest eigenvalue λ associated to each of the motifs [18]. We can think of it as an intrinsic property of the system set by the specific population dynamics implemented, and as such we expect it to be independent of diffusion. This is what is observed. For the first motif, with three 2-cycles and two 3-cycles, $\lambda_1 = 2$. For the second one, with two 2-cycles and one 3-cycle, $\lambda_2 = 1.6180$. The third one has two 2-cycles and $\lambda_3 = \sqrt{2}$, and the fourth motif has $\lambda_4 = 1.3247$, with one 2-cycle and one 3-cycle. Motifs from five to nine have one single cycle, which implies $\lambda_{5\text{-}9} = 1$. Purely parasitic motifs have no cycles, and their largest eigenvalue is thus $\lambda_{10\text{-}13} = 0$.

Environmental constraints are unable to modify the ability of different motifs to accumulate population. However, they play a main role as motif-selecting agents. By changing the amount of different motifs present, diffusion acts on the topology of the interactions network. First, we observe that the absolute number of motifs decreases for increasing D, simply reflecting the fast decrease in the average diversity $\langle S \rangle$. Second, it is remarkable that, on the average, the most abundant motifs (except for motif 5 which ranks fourth in topological abundance) are the parasitic ones with $\lambda = 0$. But the most relevant observation is that the type of motifs present is modified as D changes. We have measured the relative abundance z of parasitic *vs.* self-sustained motifs, $z = N(\lambda = 0)/N(\lambda \geq 0)$, where $N(\lambda)$ is the absolute number of motifs with the largest eigenvalue λ for each D. Increases in the mobility of the elements break apart the dense spatial core formed by species participating in self-sustained motifs. Then, parasites are able to reach the center of this core, and not only the periphery. For too high a

degree of mixing, the isolating effect of the surface is lost, the dynamics crosses over to that of a well-stirred reactor, and the system becomes dynamically unstable.

Confinement to the surface implies a strong suppression of long catalytic cycles. For relatively small values of the diffusion, cycles with more than three species are rarely observed and accumulate a small fraction of the population. It has been suggested that, in certain natural systems like the metabolism and the chemical reaction networks of planetary atmospheres, long cycles could have been selected against to shorten transition times and milden the response to external perturbations [12]. We have observed that the presence of 3- *vs.* 2-cycles increases as D increases. This is in agreement with our scenario where we have interpolated from interactions taking place in the two-dimensional space to those occurring in a mean-field–like environment, and where longer cycles could be partially sustained. Nonetheless, large cycles are more prone to support side reactions that would complicate or disrupt them. For this reason, it has been recently put forward that surfaces might have played a relevant role in the self-organization of prebiotic metabolisms, in particular helping organize simple (and more robust) cycles of the form here observed [19].

The distribution of population size of species and their lifetimes reach stable profiles at stationarity. In both cases, the probability of finding a species represented by a given number of particles and the lifetime of a species (typically until it accumulates too many parasites and is forced to extinction) behave as power law functions [17]. These results are qualitatively identical to those obtained in model ecosystems when the continuous arrival of new species keeps them out of equilibrium [20]. It is remarkable that, despite the strong structural differences between the networks analysed in this work and those of food webs (the former are basically dependent on cross-catalysis and the presence of cycles, while cross-predation is extremely rare in the latter), the quantities cited are almost identical in both systems. There is an additional observation which seems of interest: The analysis of 3-node motifs in ecological networks reveals that the most abundant one is the three-chain [7]. In our model, also this motif is the most abundant in the selected networks and for all the values of D analysed: it is found a hundred fold more frequently than motif 4 (see fig. 3), for instance. However, the three-chain motif ranks 12th according to its ability to sustain populations. We believe that network topology is accessory to function and other quantities (like population numbers) might help assess the real role played by a given structure.

We can compare the networks selected in our dynamical system with those generated by a simple growth model where no population dynamics is implemented. In the limit where the total number $\mathcal{S}$ of nodes is $\mathcal{S} \ll p^{-1}$, and in the absence of population dynamics, the obtained network would be equivalent to one of the models studied in [21], where a new node with a single connection is added per time step. Note that in our case most nodes would have no links for too small p, so they would be discarded, and nodes with more than one link would be very rare. The resulting network has a power law distribution of links per node, while our degree distributions are bell-shaped with an exponentially decreasing tail for increasing number of connections. In this sense, the initial distribution of links per species in the present model has little influence in the structure of the selected network.

In summary, the analysis of a simple model of a replicator network evolving on a surface has shown how the interplay between population dynamics and enviromental constraints is able to select non-trivial topologies. Quantitative and qualitative differences between the obtained networks in different situations (including the case when population dynamics is not implemented) have been highlighted. The results here presented speak for the selective power of dynamical mechanisms which are, in addition to functional requirements, an active design principle for network topology.

* * *

The authors thank J. Pérez-Mercader for support and U. Bastolla, C. Briones, and E. Lázaro for discussions. Support from the MCyT (Ramón y Cajal program) is acknowledged.

REFERENCES

[1] Eigen M. and Schuster P., *The Hypercycle: A Principle of Natural Self-Organization* (Springer, Berlin) 1979.
[2] Strogatz S. H., *Nature*, **410** (2001) 268.
[3] Newman M. E. J., *SIAM Rev.*, **45** (2003) 167.
[4] Consider the set of species linked to a given one. Suppose there are m such species. The clustering coefficient is the ratio between the number of links actually existing among these species and the maximal potential number of links (which equals $m(m-1)$) for a directed graph.
[5] Amaral L. A. N., Scala A., Barthélémy M. and Stanley H. E., *Proc. Natl. Acad. Sci. USA*, **97** (2000) 11149.
[6] Goh K.-I., Oh E., Jeong H., Kahng B. and Kim D., *Proc. Natl. Acad. Sci. USA*, **99** (2002) 1283.
[7] Milo R., Shen-Orr S., Itzkovitz S., Kashtan N., Chklovskii D. and Alon U., *Science*, **298** (2002) 824.
[8] Shen-Orr S. S., Milo R., Mangan S. and Alon U., *Nat. Gen.*, **31** (2002) 64.
[9] Lee T. I. *et al.*, *Science*, **298** (2002) 799.
[10] Hofbauer J. and Sigmund K., *Evolutionary Games and Replicator Dynamics* (Cambridge University Press, Cambridge) 1998.
[11] Happel R. and Stadler P. F., *J. Theor. Biol.*, **195** (1998) 329.
[12] Gleiss P. M., Stadler P. F., Wagner A. and Fell D. A., *Adv. Complex Syst.*, **4** (2001) 207.
[13] Jain S. and Krishna S., *Proc. Natl. Acad. Sci. USA*, **98** (2001) 543.
[14] Boerlijst M. C. and Hogeweg P., *Physica D*, **48** (1991) 17.
[15] Szabo P., Scheuring I., Czárán T. and Szathmáry E., *Nature*, **420** (2002) 360.
[16] Toffoli T. and Margolus N., *Cellular Automata Machines: A New Environment for Modeling* (MIT Press, Cambridge, Mass.) 1987, pp. 155-167.
[17] Manrubia S. C., Poyatos J. F. and Pérez-Mercader J., *ESA-SP*, **1** (2002) 518.
[18] Jain S. and Krishna S., *Proc. Natl. Acad. Sci. USA*, **99** (2002) 2055.
[19] Orgel L. E., *Proc. Natl. Acad. Sci. USA*, **97** (2000) 12503.
[20] Bastolla U., Laessig M., Manrubia S. C. and Valleriani A., *J. Theor. Biol.*, **212** (2001) 11.
[21] Krapivsky P. L., Redner S. and Leyvraz F., *Phys. Rev. Lett.*, **85** (2000) 4629.