The detection of intermediate-level emergent structures and patterns

M. Villani[1,2], A. Filisetti[1,3], S. Benedettini[1], A. Roli[1,4], D. Lane[1,5] and R. Serra[1,2]

[1] European Centre for Living Technology, Ca' Minich, S. Marco 2940, 30124 Venezia, Italy
[2] Dept. of Physics, Informatics and Mathematics, University of Modena e Reggio Emilia, v. Campi 213b, 41125 Modena, Italy
[3] Energy and Environment Interdepartmental Center for Industrial Research, University of Bologna, Italy
[4] DISI Alma Mater Studiorum University of Bologna Campus of Cesena, Via Venezia 52, I-47521 Cesena, Italy
[5] Dept. of Communication and Economics, University of Modena e Reggio Emilia, v. Allegri 9, 41121 Reggio Emilia, Italy

marco.villani@unimore.it

Abstract

Artificial life is largely concerned with systems that exhibit different emergent phenomena; yet, the identification of emergent structures is frequently a difficult challenge. In this paper we introduced a system to identify candidate emergent mesolevel dynamical structures in dynamical networks. This method is based on an extension of a measure introduced for detecting clusters in biological neural networks; its main novelty in comparison to previous application of similar measures is that we used it to consider truly dynamical networks, and not only fluctuations around stable asymptotic states. The identified structures are clusters of elements that behave in a coherent and coordinated way and that loosely interact with the remainder of the system. We have evidence that our approach is able to identify these "emerging things" in some artificial network models and in more complex data coming from catalytic reaction networks and biological gene regulatory systems (*A.thaliana*). We think that this system could suggest interesting new ways in dealing with artificial and biological systems.

Introduction

Artificial life is largely concerned with systems that exhibit different emergent phenomena, life itself being one of the most intriguing ones. Yet defining emergence is a controversial issue, since it is deeply related to the relationship between the observer and the observed system. We will not enter here this debate, but we rather want to stress an aspect of emergence that is often overlooked, i.e. its intermediate-level characteristics.

Most discussions of emergence, as well as its existing theories and models, take into account a two-level system, and describe the bottom-up features of the phenomenon. For example, take the well-known Benard-Marangoni hexagonal convection pattern (Haken H. 2004) that is generated when the heat flow exceeds a certain threshold: here the microscopic level is that of the water "particles" and the macroscopic one is that of the hexagonal convection cells (in this case, the hierarchy of levels is related to their characteristic dimension). There is indeed a further upper level, i.e. that of the apparatus where the phenomenon takes place; this uppermost level is necessary, and indeed it determines some major features of the phenomenon, as it can be seen e.g. by replacing the free surface with a metallic plate, thereby changing the pattern from hexagonal cells to cylindrical rolls. However the uppermost level is not affected by what happens at the lower levels and it therefore just provides the fixed boundary conditions that allow the establishment of the emergent patterns.

However, at a close look one finds that most emergent phenomena take place at levels that can be regraded as intermediate between pre-existing levels, that are in turn affected by the appearance of the intermediate emergent pattern. This topic is strictly related to the concept of emergence of hierarchies (Salthe, 1985) (Emmeche et al, 1997). Here we focus on the so-called "sandwiched" emergent phenomena, which appear in several fields such as physics, biology and social science (Lane et al, 2009). The most striking case is likely to be that of the formation of organs and tissues in multicellular organisms. Multicellularity predates the formation of organs, so the microscopic and macroscopic levels, i.e. cells and organism, were already in place when organs appeared. However, one they were formed, both organisms and cells were modified. Other examples of sandwiched emergence include the formation of clouds in physics and that of political factions, within parties, in social science, but there are actually very many. Indeed, once the importance of mesolevel emergence has been appreciated, it becomes difficult to find truly two-level systems in the sense defined above.

While in some cases it may be simple to identify the emergent structures or patterns, this is not always the case. Take for example a network of nodes that lacks any explicit all-encompassing spatial regularity, like e.g. a model of a genetic regulatory network with random connections, or a random chemical reaction network. While in spatially regular systems the appearance of regular patterns (like in the Benard case) or of clusters of nodes may be easy to find, in random systems that is by far more difficult.

In real genetic networks a lot of effort has been devoted to identifying frequently occurring motifs, i.e. small connection patterns that are much more frequent that what might be

expected if the network had been completely random; their high relative frequency can be regarded as a hint to the fact that they might have been selected by evolution due to the usefulness of the functions they perform. Indeed, the search for relevant connection patterns in complex networks is an important research topic. However, these approaches are mainly concerned with features that are directly related to the network topology, while here we want to look for structures and patterns that can be observed while looking at the dynamics of the system.

So, in order to escape from a merely topological view, we consider different subsets of the system, looking for those whose elements appear to be well coordinated among themselves and have a weaker interaction with the rest (Mesolevel Dynamical Structures, or MDS, in the following). For each subset of elements we will measure its so-called cluster index, a measure based on information theory that has been proposed by Tononi and Edelman (Tononi et al. 1998). After a suitable normalization procedure we rank the various subsets in order to identify those that are good candidates for the role of partially independent "organs" (note that they not necessarily exist in any network).

The approach

For the sake of definiteness, let us consider a system U, our "universe" that is a network of N elements that can change their state in discrete time, taking one of a finite number l of discrete values. The value of element i at time $t+1$, $x_i(t+1)$, will depend in a deterministic way upon the values of a fixed set of input elements at time t, possibly including the i-th (self-loops are not prohibited).

We will consider the systems' behaviors after an adequate relaxation time, in order to observe its asymptotic states. Given this quasi-equilibrium hypothesis we can estimate the entropy of each element from a long series of states by taking its frequencies f_v of observed values as proxies for probabilities, so:

$$H_i = -\sum_{v=1}^{m} f_v \log f_v \qquad [1]$$

where the sum is taken over all the possible values an element can take. Of course, the average entropy of the whole system is the average of H_i taken over all the elements.

In case of a fixed point attractor $H_i=0$ for every element since each node takes its value with frequency one. In order to apply entropy-based methods, Edelman and Tononi considered a system subject to gaussian noise around an equilibrium point. However nonlinear systems can carry several different attractors, each attractor revealing a particular way of functioning of the system itself: so the composition of all these asymptotic behaviors should help us in finding the parts of the system able to dynamically support them. Our "long data series" therefore will be composed by several repetitions of a single attractor, followed by repetitions of another one, etc. (ignoring the short transients

between the attractors)[1], the number of repetitions reflecting the nature of the system we are analyzing. There are several different strategies to estimate these attractors' weights: in case of noisy systems a possibility is that of using the persistence time of the systems in each of them (Villani and Serra, 2013), whereas deterministic systems might be analyzed by weighting attractors with their basins of attraction. Given the nature of the cases of this work in the following we opt here for this second choice.

Now let us look for interesting sets of nodes (clusters, from now on). A good cluster should be composed by nodes (i) that possess high integration among themselves and (ii) that are more loosely coupled to other nodes of the system. The measure we define, called the cluster index, provides a value that can be used to rank various candidate clusters (i.e., emergent intermediate-level sets of coordinated nodes).

The cluster index

Following Edelman and Tononi (Tononi et al. 1998), we will define the cluster index $C(S)$ of a set S of k elements, as the ratio of a measure of their integration $I(S)$ to a measure of the mutual information $M(S;U-S)$ of that cluster with the rest of the system.

The integration is defined as follows: let $H(S)$ be the entropy (computed as before) of the elements of S. This means that each state is a vector of k elements, and that the entropies are computed by counting the frequencies of the k-dimensional vectors. Then:

$$I(S) = \sum_{j \in S} H(x_j) - H(S) \qquad [2]$$

So $I(S)$ measures the deviation from statistical independence of the k elements in S, by subtracting the entropy of the whole subset to the sum of the single-node entropies. The mutual information of S to the rest of the world $U-S$ is also defined by:

$$M(S;U-S) \equiv H(S) + H(S|U-S) =$$
$$= H(S) + H(U-S) - H(S,U-S) \qquad [3]$$

where, as usual, $H(A|B)$ is the conditional entropy and $H(A,B)$ the joint entropy. Finally, the cluster index C(S) is defined by:

$$C(S) = \frac{I(S)}{M(S;U-S)} \qquad [4]$$

The cluster index vanishes if $I=0$, $M\neq0$, and is not defined whenever $M=0$. These cases, in which S is statistically independent from the rest of the system, can nevertheless be diagnosed in advance: the 0/0 form does not provide any information, whereas $I(S)/0$ form - with $I(S)\neq0$ - points to statistical independence of S from the rest of the system, and calls for a separate analysis.

---

[1] Note that – given the nature of the average computation – the particular order of the data vectors on the series do not alter the analysis; in addition the data series can be composed by states belonging to different attractors

C(S) scales with the size of the subsystem, so a loosely connected subsystem may have a larger index than a more coherent, smaller one: to compare the indices of the various candidate clusters it is therefore necessary to normalize their cluster indexes, for example by comparing them with those of subsystems having same size, but belonging to a non-clustered homogeneous system (a "null system").

The definition of the "null system" is critical: it could be problem-specific, but we prefer a simple solution which is fairly general: given a series of discrete vectors, we compute the frequency of each symbol and generate a new random series where each symbol has a probability of appearing equal to that of the original series. This random null hypothesis is easy to calculate, related to the original data and parameter-free; moreover it satisfies the requirements set by Tononi of homogeneity and cluster-freeness.

The "null system" therefore provide us with a null hypothesis and allows us to calculate a set of normalization constants, one for each subsystem size. For each subsystem size, we compute average integration $\langle I_h \rangle$ and mutual information $\langle M_h \rangle$ (subscript h stands for "homogeneous"); we can then normalize the cluster index value of any subsystem S using the appropriate normalization constants dependent on the size of *S*:

$$C'(S) = \frac{I(S)}{\langle I_h \rangle} \bigg/ \frac{M(S;U-S)}{\langle M_h \rangle} \qquad [5]$$

In order to compute a statistical significance index (T_c in the following) we apply this normalization to both the cluster indexes in the analyzed system and in the null system:

$$T_c(S) = \frac{C'(S) - \langle C'_h \rangle}{\sigma(C'_h)} \qquad [6]$$

where $\langle C'_h \rangle$ and $\sigma(C'_h)$ are respectively the average and the standard deviation of the population of normalized cluster indices with the same size of *S* from the null system (Benedettini 2013). Finally we use T_c to rank the obtained clusters.

Results

The cluster index has been introduced by Tononi (Tononi et al. 1998) for quasi-static systems; in the previous section we have shown how it could be extended to nonlinear dynamical systems, and in the following we will show the result of the application of this ranking method to some relevant systems, including generic models of gene regulatory networks, models of sets of catalytic chemical reactions and models of specific regulatory networks (*A.thaliana*). The method draws our attention on the subsets of the analyzed system that are highly functionally correlated and that could represent possible candidates MDSs. In the end we will also comment on the fact that our method, although not yet fully developed, outperforms usual correlation techniques.

Boolean networks

The case study we are going to examine consists of three synchronous deterministic Boolean networks (BNs), described in Fig.1. BNs are an important framework frequently used to model genetic regulatory networks (Kauffman, 1993) (Kauffman, 1995), also applied to relevant biological data (Serra et al. 2004) (Shmulevich et al. 2005) (Villani et al. 2007) and processes (Serra et al. 2010) (Villani et al. 2011). The aim of this case study is to check whether CI analysis is capable of recognizing special topological cases, such as causally (in)dependent subnetworks and oscillators, where the causal relationships are more than binary. Note that given this "more than binary" nature in all the following cases, traditional analyses based on correlation between pairs of variables might fail.. For example the computation of Pearson correlation coefficients of the networks of this section does not lead to identify related variables, given that only diagonal elements take non negligible values.
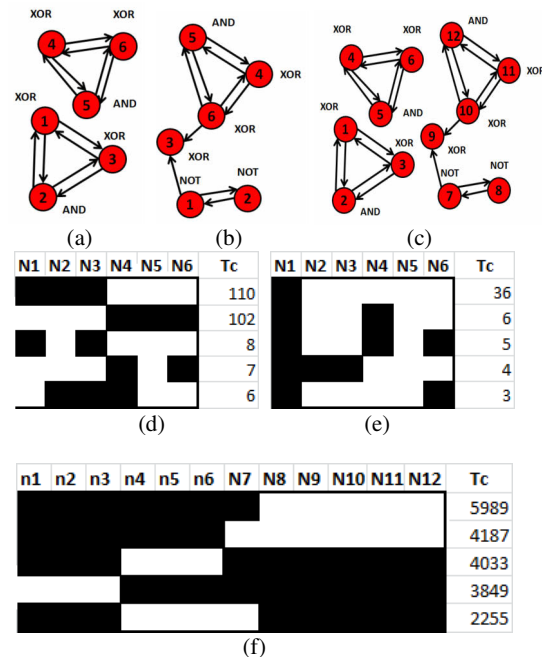

Figure 1 (a) independent Boolean networks (BN1); (b) interdependent networks (BN2); (c) a system composed by the merging of both the previous networks (BN3). Beside each boolean node there is the boolean function the node is realizing. The second part of the figure shows the matrixes illustrating the elements belonging to the clusters (white on figures) and the corresponding T_c values, for (d) BN1, (e) BN2 and (f) BN3 systems

CI analysis is able to correctly identify the two subnetworks of BN1 (first and second rows). The analysis clusters together 5 of 6 nodes of BN2: those already clustered in BN1, plus nodes 1 and 2 (which negates each other - figure 1b) and the node that compute the XOR of the signal coming from the two just mentioned groups. Indeed, all these nodes are needed in order to correctly reconstruct the BN2 series. The analysis

is able to identify all MDSs also when all the series are merged together (figure 1f, where the top two clusters correspond respectively to the 5 nodes already recognized in BN2 and to the whole BN2 system, while the third and fourth rows correspond to the independent subgraphs of BN1 - see (Villani et al., 2013) for details). Experiments performed using asynchronous update yielded essentially the same results with respect to both CI and correlation analyses.

We would like to point out that CI analysis does not require any knowledge about system topology or dynamics. This information is normally unavailable in real cases; on the other hand, our methodology just needs a data series.

Perturbing a catalytic reactions system

It is widely believed that the origin of life required the formation of sets of molecules able to collectively self-replicate (Carletti et al., 2008) (Filisetti et al., 2010) (Ganti, 2003) (Luisi et al., 2006) (Mansy et al., 2008) (Rasmussen et al., 2003) (Stano and Luisi, 2010) (Szostak et al., 2001), a phenomenon that may play an important role also in future bio-technological systems (Solè et al., 2007). There are many efforts to identify the dynamical cores of these systems, mainly based on static properties of the reaction networks they are forming (Farmer et al., 1986) (Hordijk et al., 2010) (Kauffman, 1986): in this work we present a first attempt toward a dynamical detection of these systems.
We use a simple system (inspired by a model (Filisetti et al. 2011a) (Filisetti et al. 2011b) (Filisetti et al. 2011c) (Farmer et al., 1986) originally due to Kauffmann (Kauffmann, 1993) (Kauffmann, 1995)) where there are two distinct reaction pathways, a linear reactions chain (CHAIN) and an autocatalytic set of molecular species (ACS) (see figure 2): both reactions pathways occur in an open well-stirred chemostat (CSTR) with a constant influx of feed molecules and a continuous outgoing flux of all the molecular species proportional to their concentration. The dynamics of the system is described adopting a deterministic approach whereby the reaction scheme is translated in a set of Ordinary Differential Equations (ODE) integrated by means a fourth-order Runge-Kutta method (Young and Gregory, 1988).

The main entities of the model are molecular species ("polymers") represented by linear strings of letters A and B, forming together a catalytic reactions system composed of 6 distinct condensation reactions in which two species are glued to create a longer species. The reactions occur only in presence of a specific catalyst, since spontaneous reactions are assumed to occur too slowly to affect the system behavior. Accordingly, in the following the reaction scheme is presented:

According to the three molecular nature of the condensation reaction, reactions occur in 2 two steps: in the former the catalyst binds the first substrate forming a molecular complex, while in latter the molecular complex binds the second substrate releasing the product and the catalyst. The "food set" of the linear chain (BAB→ABBBA→BBBABA→BAAB) is formed by the species ABB, BBA, BBB, ABA, BAA, B, whereas the food set of and the autocatalytic cycle (AABBA→ AAAA→AAAB→AABBA) is formed by the species BA, AAB, AAA, A, AB, AA. Besides, an independent molecular species BB not involved in any reactions has been introduced as control species (figure 2).

The asymptotic behavior of this kind of systems is a single fixed point (Vasas et al., 2012), due to the system feedback structure. In order to apply our analysis we need to observe the feedbacks in action, therefore we perturb the concentration of some molecules in order to trigger a response in the concentration of (some) other species. So we temporarily set to zero the concentration of some species (in the example of fig.2 of the species ABBBA, BBBABA, AABBA, AAAA, AAAB) after the system reached its stationary state[2]: in order to analyze the system response to perturbations we use a 3-level coding, where for each species the digit '0'-'1'-'2' stand respectively for "concentration decreasing", "no change" and "concentration increasing"[3].
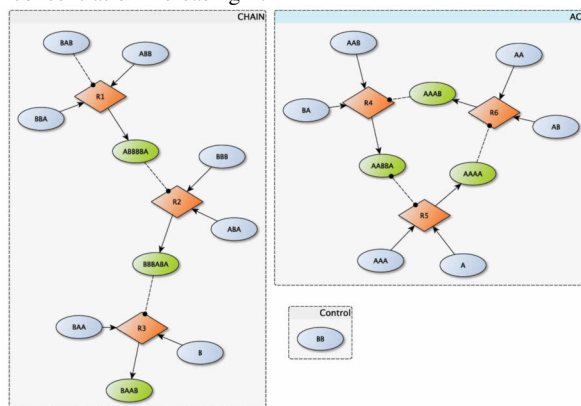

Figure 2 The chemical system under analysis. Circular nodes depict chemical species, the blue ones stand for those injected on the CSTR (food species) and the green ones represent the more complex species built by specific concatenations of the food species, see reaction scheme in the text. Diamond shapes represent reactions where incoming arrows go from substrates to reactions and outgoing arrows go from reactions to products. Dashed lines indicate the catalytic role of a particular molecular species within the specific reaction context. The kinetic constants of all present reactions have the same value k_{dir}=0.0025 s^{-1}mol^{-1}); the incoming concentration of each food species is 1.0 mol, whereas each second the 2% of the CSTR volume is renewed

---

[2] In this example the analyzed data series starts from second 200, in order to avoid the initial transient

[3] In such a way we can abstract from the different concentration present on the system, a species being constant if its concentration change from previous time instant is below the threshold of 0.1%

The results clearly indicate the presence of two distinct systems of size 3 (the second and third rows in fig.4a) that correspond to CHAIN and ACS. Note that the leave of CHAIN (BAAB) is not strongly affected by the zeroing of ABBBA species (because the perturbation of this species, root of the linear chain, affects only in a limited manner the following species BBBABA, whose change in turn even lesser affects the concentration of species BAAB…): this attenuation process induces a dynamical hierarchy on CHAIN system, which allows the finer subdivision highlighted by the first row of fig.4a. This phenomenon is absent on ACS, a more homogeneous system where no roots are present.
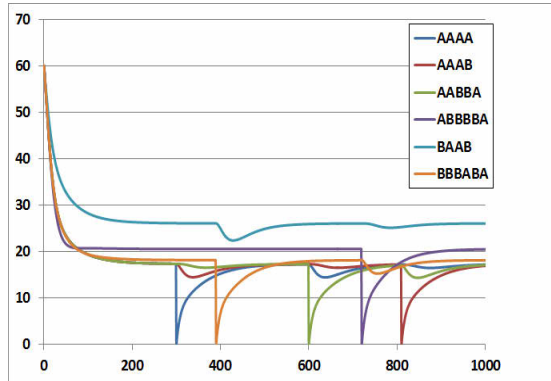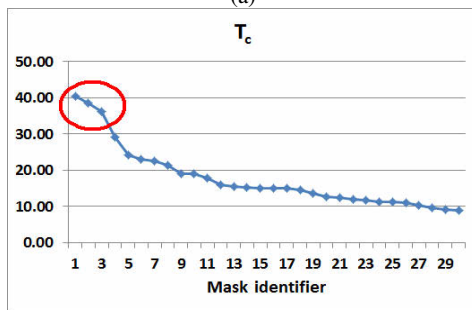

Figure 3 The chemical system trajectory, including the performed perturbations (only products are analyzed)

(a)

(b)

Figure 4 (a) The masks resulting from the chemical system analysis and (b) their corresponding Tc values. Note that the three masks whose Tc values outperform the other ones correctly identify the system's components (see text for details)

Arabidopsis thaliana

It is possible to expand the analysis to BN derived from biological data of specific living beings. In this work we take advantage from the available data of the gene regulatory network shaping the developmental process of Arabidopsis thaliana: although the whole network is largely unknown, a certain subsystem has been identified as responsible for the floral organ specification. We will not enter here a discussion about the merits and limits of this simplified model, but we will take it "for granted" and we will apply our method to test whether it can discover significant MDSs.

The network is modeled by means of a BN described in (Chaos et al., 2006), which has 15 nodes and 10 different attractors (all fixed points): we therefore build a data series containing a number of repetitions of these attractors in proportion to their basins of attraction. In doing so it is possible to note that genes LUG and CLF are constantly active in all the attractors: this particular feature introduces a particular "noise" on CI analysis, by adding spurious cluster among the first positions. Indeed, it is possible to analytically demonstrate that the addition of constant nodes in clusters with high T_c leads again to other clusters with high T_c values: these additions nevertheless do not have particular biological meanings (the added elements do not introduce any variation), so the corresponding clusters can be memorized as "not significant".

The analysis clearly groups genes UFO and AP3, present alone on the best significant cluster and in all the following 20 most significant clusters. Note that the second significant cluster includes gene WUS: indeed, for biologists (Lenhard et al., 2001) (Lohman and others 2001) UFO and WUS are key inputs for determining the specific time and site where the combinations of gene activities considered in the developmental process are established, whereas AP3 is an important transcription factor. So, our analysis perceives the combination of a "sensor" (UFO) and of an influential "signaler" (AP3) as a single powerful dynamical engine, whose action can be tuned by WUS gene, demonstrating that it could highlight biologically interesting functional relationships.
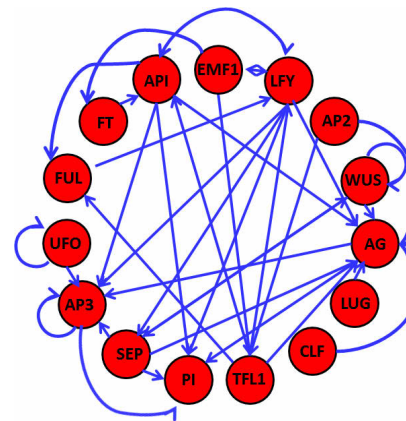

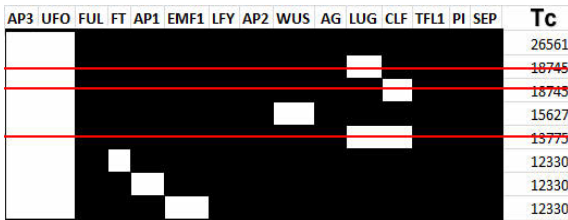**Figure 5** Floral network of A. thaliana (from (Chaos et al., 2006))

Figure 6 Matrix illustrating the elements of the clusters (white in figure) identified by our analysis and their corresponding T_c values. Genes LUG and CLF are always constant along all the attractors and therefore their insertion in "active" MDSs can be excluded a priori (it is possible to analytically demonstrate that the addition of constant nodes in already existing clusters leads to cluster with high T_c values – but these additions do not seem have particular biological meanings)

Conclusions

In this paper we introduced a system to identify candidate emergent mesolevel dynamical structures in dynamical networks. The main novelty of the present work, in comparison to previous application of the cluster index and of similar measures (Tononi et al. 1998) is that we used it to consider truly dynamical networks, and not only fluctuations around stable asymptotic states.

Future works will consider the application of the method to other important natural and artificial networks, an improved understanding of its working and the use of entropies taken at different times.

As examples of application we used time series of simple artificial systems and more complex data coming from catalytic reaction networks and biological gene regulatory systems (*A.thaliana*). The analysis performed by our system was able to identify correctly the MDSs, and we think it could suggest interesting new ways in dealing with artificial and biological systems.

Future work will consider the application of the method to other important natural and artificial networks, with the aim of deepen our understanding of its working principles and assessing its analysis power. In addition, we also plan to extend the definition of cluster index so as to take into account time relationships, for example by using of entropies taken at different times.

References

1. Benedettini S. Identifying mesolevel dynamical structures ECLT (European Center for Living Technologies) technical report, Venice 2013

2. Carletti, T., Serra, R., Poli, I., Villani, M. & Filisetti, A. (2008): Sufficient conditions for emergent synchronization in protocell models. *Journal of Theoretical Biology* - Elsevier Ltd. v. 254 pp.741–751

3. Chaos A., Aldana m., Espinosa-Soto C., Ponce de Leon B.G., Garay Arroyo A., Alvarez-Buylla E.R. (2006) From Genes to Flower Patterns and Evolution: Dynamic Models of Gene Regulatory Networks *J Plant Growth Regul* 25:278–289

4. Emmeche, C., Køppe, S., Stjernfelt, F. (1997) Explaining emergence: Towards an ontology of levels. Journal for General Philosophy of Science, 28:83-119

5. Farmer, J., Kauffman, S.A., Packard N. (1986) Autocatalytic replication of polymers. Physica D: Nonlinear Phenomena. 220, 50–67

6. Filisetti A., Serra R., Carletti T., Villani M., Poli i. (2010) Non-linear protocell models: Synchronization and Chaos Eur. Phys. J. B 77, 249–256

7. Filisetti A., Graudenzi A., Serra R., Villani M., De Lucrezia D., Füchslin R.M., Kauffman S.A., Packard N., Poli I. (2011a) A stochastic model of the emergence of autocatalytic cycles *Journal of Systems Chemistry*, 2:2 doi:10.1186/1759-2208-2-2

8. Filisetti A., Graudenzi A., Serra R., Villani M., Füchslin R.M., Packard n., Kauffman S.A.. Poli I (2011b) A stochastic model of autocatalytic reaction networks *Theory in Biosciences* v.130 Springer Berlin/Heidelberg DOI 10.1007/s12064-011-0136-x

9. Filisetti, A., Graudenzi, A., Serra, R., Villani, M., De Lucrezia, D., and Poli, I. (2011c). The role of energy in a stochastic model of the emergence of autocatalytic sets. In *Advances in Artificial Life ECAL* 2011 Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems, H. Bersini, P. Bourgine, M. Dorigo, and R. Doursat, eds. (MIT Press, Cambridge, MA), pp. 227-234.

10. Ganti, T.: *Chemoton Theory, Vol. I: Theory of fluyd machineries: Vol. II: Theory of living system*. New York: Kluwer Academic/Plenum. (2003)

11. Haken H. (2004) *Synergetics* Springer-Verlag Berlin Heidelberg

12. Hordijk, W., Hein J., Steel m. (2010) Autocatalytic Sets and the Origin of Life. *Entropy*. 12, 7, 1733–1742

13. Kauffman, S.A. (1986) Autocatalytic sets of proteins. *J Theor Biol*. 119, 1, 1–24

14. Kauffman, S.A., 1993. *The Origins of Order*. Oxford University Press, Oxford.

15. Kauffman, S.A., 1995. *At Home in the Universe*. Oxford University Press, Oxford.

16. Lane, D., van der Leeuw, S., Pumain, D., West, G. eds. (2009) Complexity perspectives on innovation and social change. Springer-Verlag Berlin Heidelberg.

17. Lenhard M, Bonhert A, Jurgens G, Laux T. (2001) Termination of stem cell maintenance in Arabidopsis floral meristems by interactions between WUSCHEL and AGAMOUS. *Cell* 105:805–808

18. Lohmann JU, Hong RL, Hobe M, Busch MA, Parcy F, others. (2001) A molecular link between stem cell regulation and floral patterning in Arabidopsis. *Cell* 105:793–803.

19. Luisi, P.L. Ferri F., Stano P. (2006) Approaches to semi-synthetic minimal cells: a review. *Naturwissenschaften*. 93, 1, 1–13

20. Mansy, S.S., Schrum J.P., Krishnamurthy M., Tobé S., Treco1 D.A., Szostak J.W. (2008) Template-directed synthesis of a genetic polymer in a model protocell. *Nature*. 454, 7200, 122–125

21. Rasmussen, S. Chen L, Nilsson M, Abe S. (2003) Bridging nonliving and living matter. *Artificial life*. 9, 3, 269–316

22. Salthe, S.N. (1985) Evolving Hierarchical Systems: Their Structure and Representation. Columbia University Press.

23. Serra, R., Villani, M. & Semeria A. (2004): Genetic network models and statistical properties of gene expression data in knock-out experiments *Journal of Theoretical Biology* 227: 149-157

24. Serra R., Villani M., Barbieri B., Kauffman S.A., Colacci A. (2010) On the dynamics of random boolean networks subject to noise: attractors, ergodic sets and cell types *Journal of Theoretical Biology* - Elsevier Ltd. v.265 pp.185-193

25. Shmulevich, I., Kauffman, S.A., Aldana, M., 2005. Eukaryotic cells are dynamically ordered or critical but not chaotic. *Proc. Natl Acad. Sci.* 102, 13439–13444.

26. Solé, R.V., Munteanu A, Rodriguez-Caso C, Macía J. (2007) Synthetic protocell biology: from reproduction to computation. Philos Trans R Soc Lond B Biol Sci. 362, 1486, 1727–1739 (2007)

27. Stano, P., Luisi, P.L. (2010) Achievements and open questions in the selfreproduction of vesicles and synthetic minimal cells. *Chemical communications* (Cambridge, England). 46, 21, 3639–53

28. Szostak, J.W. Bartel D.P., Luisi P.L. (2001) Synthesizing life. *Nature*. 409, 6818, 387–390

29. Tononi G., McIntosh A.R., Russell D.P., Edelman G.M. (1998) Functional Clustering: Identifying Strongly Interactive Brain Regions in Neuroimaging Data *NEUROIMAGE* 7, 133–149

30. Vasas V., Fernando C., Santos M., Kauffman S.A., Szathmáry E. (2012) Evolution before genes *Biology Direct*, 7:1

31. Villani M., Serra, R., Graudenzi, A. & Kauffman, S.A. (2007): Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data. J. *Theor. Biol.* 249 : 449-460

32. Villani M, Barbieri A, Serra R (2011) A Dynamical Model of Genetic Networks for Cell Differentiation. *PLoS ONE* 6(3): e17703. doi:10.1371/journal.pone.0017703

33. Villani M., Serra R. (2013) On the dynamical properties of a model of cell differentiation *EURASIP Journal on Bioinformatics and Systems Biology*, 2013:4 – Springer

34. Villani M., Benedettini S., Roli A., Lane D., Poli I., Serra R. (2013) Identifying emergent dynamical structures in network models *Proceeding of WIRN 2013*, submitted

35. Young D.M., Gregory R.T. (1988) A Survey of Numerical Mathematics *Dover Publications, Incorporated*