

Are self-organizing biochemical networks emergent?

C. Malaterre

Université Paris 1 – Panthéon Sorbonne, Institut d'Histoire et de Philosophie des Sciences et Techniques, France

Abstract. Biochemical networks are often called upon to illustrate emergent properties of living systems. In this contribution, I question such emergentist claims by means of theoretical work on genetic regulatory models and random Boolean networks. If the existence of a critical connectivity K_c of such networks has often been coined “emergent” or “irreducible”, I propose on the contrary that the existence of a critical connectivity K_c is indeed mathematically explainable in network theory. This conclusion also applies to many other types of formal networks and weakens the emergentist claim attached to bio-molecular networks, and by extension to living systems.

1. INTRODUCTION

Biochemical networks are often called upon to illustrate the irreducibility of living systems to physico-chemical phenomena, and thereby their emergent nature: for instance, biochemical systems are said to be irreducibly complex (Behe, 1996); it is also claimed that biochemical networks display self-organizing emergent phenomena (Boogerd, Bruggeman, Richardson, Stephan & Westerhoff, 2005), and this can be the case for metabolism (Cornish-Bowden, Cardenas, Letelier, Soto-Antrade & Abarzua, 2004) or for immune systems (de Boer & Perelson, 1991) among others. In such cases, properties at the level of the network as a whole are claimed to be emergent from the properties of its molecular constituents, in the sense that the former cannot be deduced nor predicted from the latter. The emergent status of such molecular networks appears rooted in structural and dynamical analyses of complex systems. In this contribution, I propose to consider in more details the modeling of genetic regulatory pathways by random Boolean networks: this modeling gives rise to most surprising phenomena at the level of the network itself, often referred to as “emergent” from the properties of its constituents. Are the appearances of such properties irreducible in the sense that they can only be observed and not predicted from previous knowledge? Can they be deduced in one way or another from the properties of the constituents? Can they be mathematically explained? I will first describe random Boolean networks and their use to model genetic regulatory pathways. I will then focus on the “emergent” status of a specific property: the so-called “critical connectivity” displayed by these networks. I will argue that, despite the fact that this property is often witnessed with “natural piety”, it can indeed be mathematically explained from basal properties within the framework of Boolean network theory. I conclude that such theoretical results considerably weaken the “emergent” claim attached to complex molecular networks and shift the burden of proof onto emergentism.

2. GENETIC REGULATORY PATHWAYS AND RANDOM BOOLEAN NETWORKS

The idea that genes are acting onto one another, sometimes in fairly tricky ways, is not new and can be traced back to the lactose operon discovery of Jacob and Monod in the 1960's. The realization however of the daunting sizes of such interacting genetic systems is more recent and stems from the massive sequencing projects undertaken in the past decade. The genome is now more than ever considered to be a gigantic network of reflexively interacting genes. Other biomolecular systems are also now frequently described in terms of complex networks of organic molecules, be it for instance the phenomenon of carcinogenesis or the description of metabolic pathways. What is more, the computational study of formal networks has led to the discovery of interesting features of such networks, features that are frequently coined "emergent" in so far as they appear unexpectedly during computer simulation runs and do not seem to be deducible from the basal properties of the network components themselves: stability features, chaotic and disorderly properties, sudden "phase" transitions and so forth. And it was not long before such "emergent" formal systemic patterns were transferred to natural realm of biological systems. One of the most studied types of formal networks is the class of so-called random Boolean networks (RBNs). RBNs were originally developed by Stuart Kauffman in the late 1960s to model genetic regulatory networks (Kauffman, 1969; Kauffman, 1993).

A random Boolean network consists of nodes that are Boolean in the sense that they can take a "zero" or "one" value, and connections between nodes. Let us imagine an RBN with N nodes: the value of each node is determined by a logical function that is randomly attributed to the node and which takes inputs from K other nodes also chosen randomly. It is this random structure that determines how the network behaves in successive discrete increments of time. RBNs can be pictured as a highly simplified model of a genetic system in the following way: nodes represent genes that are "active" if the node has value "one" (inactive otherwise) and their activity is triggered by the activity/inactivity of other related genes. In the case when the genetic system is the complete genome of a cell, the corresponding RBN provides a model for studying the dynamical behavior of the complete cell, this behavior being given by the succession of states of the cell's full set of genes.

3. CRITICAL CONNECTIVITY: AN "EMERGENT" PROPERTY OF RBNs?

To study this dynamical behavior, Kauffman ran simulations of such RBNs on computers and carried out different runs of numerical studies (for instance Kauffman, 1974; Kauffman, 1984). This is how he came to discover patterns of behaviors among types of RBNs and postulate the existence of a critical connectivity K_c of RBNs, a most interesting and surprising property that has subsequently often been termed "emergent". What is this surprising critical connectivity?

Numerical simulations typically consist in the random construction of a Boolean network subject to some constraints such as the number of nodes N or the number of input connections K , each node being randomly initialized to zero or one. A run n then consist in the calculation of the Boolean functions of all nodes based on inputs from run $n - 1$. The RBN being constituted of a finite number of nodes, each of them with two

possible states, it has itself also a finite number of states, namely 2^N . Therefore any RBN will eventually return to a state it has already encountered previously, thereby entering into a cycle called the “limit cycle”. During numerical simulations, RBNs can behave in different characteristic ways: for instance, long or short limit cycles, same limit cycles despite differing initializations of certain nodes, numerous nodes entering a stable state during the limit cycle, etc. Some of these behaviors can be qualified as being more orderly or predictable than others: for instance, a short limit cycle is more predictable than a long one; robustness to differing initialization parameters offers also more predictability, and so forth. Hence the search for parameters that would help identify orderly RBNs from non orderly ones. This is where the critical connectivity comes into play. Indeed, after many numerical simulations, Kauffman found out that the degree of interconnection of the nodes played a critical role: RBNs where the number of inputs to each node $K \leq 2$ were more orderly than those where $K > 2$. The borderline $K = 2$ came to be called the critical connectivity K_c . From this, Kauffman made a certain number of conjectures: for instance, he postulated that for $K \leq 2$ most nodes would stabilize quickly and that the size of the limit cycle would be smaller than $N^{1/2}$, and that for $K > 2$ many nodes would not stabilize at all.

For Kauffman, if the state of a cell can be modeled using an RBN, then the behavior of the RBN before entering its limit cycle would represent the embryonic cell during its differentiation phase, and the behavior of the RBN along the limit cycle would represent the cell’s replication cycle. In order to have stable and robust cells, one would therefore look for criteria that would make the RBN somehow orderly and robust, but not excessively rigid, hence the choice of a connectivity of 2, borderline between order and chaos, and often named therefore “edge of chaos”. It is properties of RBNs like the critical connectivity that are often used as examples of “emergent” phenomena in complex systems, this emergence then percolating into systems biology. Yet, to which extent is this critical connectivity “emergent” in the sense of being “unpredictable”?

4. MATHEMATICAL ANALYSES (I): THE “ANNEALED APPROXIMATION”

Interestingly, there exist mathematical analyses of this “emergent” property. One of the earliest approaches, quite known among network theorists, is the one proposed by Derrida and Pomeau in the 1980’s : their mathematical approach consists in what they called an “annealed approximation” of Kauffman’s RBNs (Derrida & Pomeau, 1986). Instead of considering, for a given RBN, the Boolean functions of the nodes and the connections between the nodes to be fixed (or “quenched”) during the successive time-increments, they proposed to randomly reassign functions and connections at every time-step. This enabled them to derive and solve a recursion formula which describes the time evolution of the gap between two different initial states C^1 and C^2 of a given network, i.e. the number of nodes with a different value: when this gap, called the “Hamming distance”, decreases over time, the two network states converge towards one another since they have more and more nodes that have the same value. This can also be interpreted in the sense that the introduction of a small perturbation into C^1 so as to give C^2 would not impact the end result, the two configurations resulting ultimately in the same network state. Within this “annealed model”, Derrida and Pomeau proved analytically that the convergence of such two states depended on the average connectivity K of the network being below

the critical value $K_c=2$. In addition, they proved that, in the thermodynamic limit when $N \rightarrow \infty$, the “annealed model” they had introduced and the original “quenched” one show similar behaviors: more specifically, they proved that, for any time t , when $N \rightarrow \infty$, the behavior of a given node is asymptotically the same in both models. Also, further similarities between the “annealed” and the “quenched” models were subsequently analytically proven (e.g. Hillhorst & Nijmeijer, 1987).

As a consequence, when the analytical proof of the existence of a critical connectivity $K_c = 2$ in the case of the “annealed” model, and the proofs of shared asymptotic behaviors between the “annealed” and “quenched” models, are taken together, they establish a strong argument in favor of a mathematical explanation of the existence of a critical connectivity $K_c = 2$ in the case of the “quenched” model. Now, if the existence of such critical value can be mathematically assessed and its value also determined, then the “surprising” property of random Boolean networks to exhibit threshold behaviors at $K_c = 2$ is effectively *explained* on the basis of the mathematical premises describing the model at the level of its constituents, and thereof should no longer be called “emergent”.

Interestingly also, very similar mathematical analyses concerning the existence of critical threshold values in different types of networks other than RBNs have been elaborated on the basis of the “annealed approximation”. This is for instance the case of random networks with multiple states for which it was established that $K_c = 1/(1 - 1/S)$ where S is the number of states; when $S = 2$, i.e. Boolean case, one finds back the value $K_c = 2$ (Solé, Luque & Kauffman, 1999). And this is also the case of random spin threshold networks (Rohlf & Bornholdt, 2002). Furthermore, some theoretical work can act as prediction and serve as guideline for simulations: the “annealed approximation” for instance makes it possible to devise new calculation methods and predict other properties of RBNs such as the distribution of periods and weights of attraction basins in cases of a finite K (Bastolla & Parisi, 2006) or necessary conditions for RBN stability (Solé & Luque, 1995).

5. MATHEMATICAL ANALYSES (II): FURTHER PROOFS

Despite all these analyses, one may still argue that such mathematical results are only approximations that are valid in the thermodynamic limit, and that they do not constitute any rigorous deductive proof of properties of “quenched” RBNs. And that, as a consequence, the critical connectivity of “quenched” RBNs is still not mathematically proven (or deduced) and remains emergent.

Yet, there exist further mathematical analyses in the form of deductive proofs that do establish rigorous theorems on the topic of the critical connectivity. For instance, Luczak & Cohen (1991) proved that “quenched” RBNs with $K = 2$ exhibit certain forms of stable behavior in the sense that a large fraction of such networks stabilize quickly, and exhibit small limit cycles. In addition, Lynch (1995) rigorously established that the average limit cycle size when $K = 2$ grows faster than any polynomial in N . And, in the case of networks with $K > 2$, Lynch also rigorously established that such networks are not as stable as networks with ≤ 2 (2007). In addition, analytical solutions concerning other properties of RBNs have been proposed in the special case $K = 1$, providing a complete analytical study of this particular case (Flyvbjerg & Kjaer, 1988). Complete analytical solutions have also been developed for the special case $K \rightarrow \infty$ (Derrida & Flyvbjerg, 1987). In addition, specific properties such as the size of the stable

core have been analytically deducted and solved for special cases $K = 1, 2, 3, 4$ or 5 (Flyvbjerg, 1988).

In other words, even if finding more general analytic solutions is still an open issue (Gershenson, 2004), specific mathematical proofs have been established in a certain number of well-studied cases. Of course, not all questions are resolved: for instance, there is so far no proof of a long-term fully chaotic behavior of “quenched” RBNs with $K > 2$. Yet there is now a substantial body of mathematical work that makes at least two points. First, even if the critical connectivity of RBNs is not rigorously deduced in the case of “quenched” networks, it is at least *mathematically explained*, thanks to a rigorous proof in the case of the “annealed” model and a convergence proof of the “annealed” and “quenched” models in the thermodynamic limit. Second, some mathematical proofs have *rigorously established specific properties* related to that of critical connectivity and have shown that the picture is likely to be more complicated than initially supposed.

6. IMPLICATIONS

The mathematical proofs I referred to above in the specific case of random Boolean networks point to, at least, three tentative conclusions that weaken the emergentist stance often advocated in the study of molecular biological networks.

First, it is not because a phenomenon in biology is to be modeled by a complex network of interacting entities that it is a potential good candidate for emergence. Mathematical or analytical proofs of complex networks’ properties do exist. Of course, they may be austere and probably not as easy to account for as computer simulation runs. They are also rare compared to the overwhelming amount of simulation results. Yet, such proofs do exist and show that intriguing and initially so-called “emergent” properties of complex networks can be and have been mathematically studied, and as such should not longer be called “emergent” in any sense but “epistemically surprising”.

Second, if “emergence” there is in networks theory and molecular networks, it is more likely to be under a soft epistemic form rather than with a strong ontological commitment. Indeed, if some initially “emergent” properties, such as the critical connectivity of RBNs, can later be deduced from the nodes’ properties and network theory, why not others, why not all? The problem here stems from the daunting number of surprising properties that can be displayed by complex networks and that are most often revealed by computer simulations. It is simply an impossible task to find analytical proofs for every single one of them. On the other hand, it appears possible to find analytical proofs for some when proper attention has been given to them.

Third, such cases shift the burden of proof to emergentists: what needs to be found is not a “thought-to-be” emergent property but a “proven-to-be” emergent property, i.e. a property supported by a mathematical proof according to which this property cannot be analytically solved. This would amount in particular to finding new impossibility theorems or relying on an existing one. Yet, such theorems, as Gödel’s, are rare in mathematics, to say the least.

7. CONCLUSION

If emergence appears to percolate from the study of formal networks into the investigation of molecular biological networks, it is more likely to be in the sense of “epistemic

surprise” than anything else. As in the case of RBNs applied to the modeling of genetic regulatory pathways, mathematical analyses of so-called “emergent” properties can be found. Therefore, if emergence there is in biological molecular networks, such claims need in turn to be supported by mathematical “impossibility” proofs. Incidentally, this shifts the burden of proof to emergentists and has implications outside of molecular biology. This also considerably weakens the emergentist stance: if emergence there is, it is more likely to be due to the huge amount of epistemically surprising properties displayed by complex networks rather than to the intrinsic nature of such properties. In this case, emergence might be everywhere yet one may wonder what is really left of it.

Acknowledgments

Earlier versions of this paper were presented at the 2008 *PSA* biannual meeting in Pittsburgh as well as at the 2008 *APA Eastern* annual meeting in Philadelphia. I would like to thank the audiences at these meetings for their questions. In particular, I am indebted to James Lynch for his most helpful comments and his mathematical insights. My thanks also go to the organizers of the CNRS interdisciplinary program “*Origines des planètes et de la vie*” for stimulating discussions.

References

- Bastolla U. and G. Parisi (2006), “Closing probabilities in the Kauffman model: An annealed computation”, *Physica D*, 98: 1-25, available on <http://arxiv.org/abs/cond-mat/9510137>.
- Behe, Michael J. (1996), *Darwin's Black Box: The Biochemical Challenge to Evolution*, The Free Press.
- Boer, Rob J. de and Alan S. Perelson (1991). “Size and connectivity as emergent properties of a developing immune network”, *Journal of Theoretical Biology*, 149(3): 381-424.
- Boogerd, F.C., F.J. Bruggeman, R.C. Richardson, A. Stephan and H.V. Westerhoff (2005), “Emergence and its place in nature: a case study of biochemical networks”, *Synthese*, 145:131-164.
- Cornish-Bowden, Athel, Maria Luz Cardenas, Juan-Carlos Letelier, Jorge Soto-Antrade and Flavio Guinez Abarzua (2004), “Understanding the parts in terms of the whole”, *Biology of the Cell*, 96:713-717.
- Derrida, B. and H. Flyvbjerg (1987), “The random map model: A disordered model with deterministic dynamics”, *Journal de Physique*, 48:971-978.
- Derrida, Bernard and Yves Pomeau (1986), “Random networks of automata: a simple annealed approximation”, *Europhysics Letters* 1: 45-49.
- Flyvberg H. (1988), “An order parameter for networks of automata”, *Journal of Physics A: Mathematical and General*, 21: L955-L960.
- Flyvberg H. and N.J. Kjaer (1988), “Exact solution of Kauffman's model with connectivity one”, *Journal of Physics A: Mathematical and General*, 21: 1695-1718.
- Gershenson, Carlos (2004), “Introduction to Random Boolean Networks”, in Bedau, M., P. Husbands, T. Hutton, S. Kumar, and H. Suzuki (eds.) *Workshop and Tutorial Proceedings, Ninth International Conference on the Simulation and Synthesis of Living Systems (ALife IX)*, Cambridge: MIT Press, 160-173.

- Hilhorst H. J. and M. Nijmeijer (1987), "On the approach of the stationary state in Kauffman's random boolean network", *Journal de Physique*, 48: 185-191.
- Kauffman, Stuart A. (1969), "Metabolic stability and epigenesis in randomly constructed genetic nets", *Journal of Theoretical Biology*, 22(3): 437-467.
- Kauffman, Stuart A. (1974), "The large scale structure and dynamics of gene control circuits: An ensemble approach", *Journal of Theoretical Biology*, 44(1): 167-190.
- Kauffman, Stuart A. (1984), "Emergent properties in random complex automata", *Physica D: Nonlinear Phenomena*, 10(1-2): 145-156.
- Kauffman, Stuart A. (1993), *The Origins of Order: Self-Organization and Selection in Evolution*, New York: Oxford University Press.
- Luczak T. and J. E. Cohen (1991), "Stability of vertices in random Boolean cellular automata", *Random Structures and Algorithms* 2: 327-334.
- Lynch, James (1995), "On the threshold of chaos in random Boolean cellular automata", *Random Structures and Algorithms* 6:239-260.
- Lynch, James (2007), "Dynamics of Random Boolean Networks", in Mahdavi, K., R. Culshaw, and J. Boucher (eds.), *Current Developments in Mathematical Biology*, World Scientific Publishing Co., 15-38.
- Rohlf, Thimo and Stefan Bornholdt (2002), "Criticality in Random Threshold Networks: Annealed Approximation and Beyond", *Physica A*, 310:245-259, available on <http://arxiv.org/abs/cond-mat/0201079>.
- Solé, Ricard, Bartolo Luque and Stuart Kauffman (1999), "Phase Transition in Random Networks with Multiple States", submitted to *Physical Review E*, available on <http://arxiv.org/abs/adap-org/9907011>.