Pseudo-attractors in Random Boolean Network Models and Single-Cell Data

Marco Villani[1,2], Gianluca D'Addese[1], Stuart A. Kauffman[3] and Roberto Serra[1,2,4]

[1]Department of Physics, Informatics and Mathematics, Modena and Reggio Emilia University, 41125, Modena (IT)
[2]European Centre for Living Technology, 30123, Venice (IT)
[3]Institute for Systems Biology, Seattle, WA 98109, USA
[4]Institute of Advanced Studies, University of Amsterdam, 1012 GC Amsterdam, (NL)
Corresponding author: marco.villani@unimore.it

Abstract

In this extended abstract two novel concepts are defined in the study of Random Boolean Networks, i.e. those of "pseudo-attractors" and "common sea", and it is shown how their analogues can be measured in experimental data on gene expression in single cells.

Introduction

Random Boolean Networks (RBNs for short, see [1] for a recent review) have been widely studied as abstract models of complex systems, thanks to the possibility of tuning their behaviors from ordered to pseudo-chaotic. They are generic models, which can however be used also to describe important biological phenomena, in particular those concerning gene expression (indeed, RBNs were originally introduced [2] as strongly simplified models of gene regulatory networks).

A RBN is a time-discrete, Boolean deterministic dynamical system where the overall state of a given network of N nodes, $X(t+1) \in \{0,1\}^N$, is uniquely determined by the previous state X(t), given the connection topology and the transition functions at each node. Both connections and transitions functions are chosen at random according to some probability distribution. While the success of this model led to the introduction of several variants, here we will consider the "classical" case, where updating is synchronous: dynamical attractors of finite networks of this kind can be either fixed points or limit cycles, but the oscillations of the latter are largely due to the choice of synchronous updating. While this is a clear choice, it limits the biological plausibility of RBNs to describe gene regulatory networks, since it requires simultaneous forgetting of the previous states of all the nodes. Different updating schemes (e.g. asynchronous) have been proposed [1,3,4], but none can claim undisputed plausibility. In particular, cyclic attractors are fragile if the updating scheme is changed, while on the other hand point attractors are conserved. Moreover, cyclic attractors do not seem to be the analogue of the cell cycle, so experimental data on gene expression do not show this type of time dependence.

While different alternatives have been proposed, the usual recipe to interpret the biological significance in multicellular organisms of RBNs' attractors is that of regarding them as the analogue of cell types. We therefore generalize this approach to pseudo-attractors. In real cells, the analogue of the CS is then the set of genes which take the same value in every cell type.

Pseudo-attractors and common sea

As anticipated, the identification of which genes "take the same value" in different cyclic attractors requires some care, since cycles in RBNs depend to a large extent upon the choice of synchronous updating, which does not have a sound biological basis [5]. Synchronous RBNs are Markovian systems, whose state $\boldsymbol{X}(t+1)$ depends upon $\boldsymbol{X}(t)$, forgetting the previous states of all the nodes of the network. The action of a gene on the activation of other genes takes place through the action of its corresponding protein; therefore, the notion of a single time step corresponds to assuming a common decay time of the different regulatory proteins, which is not supported by biological data.

By following [6] we therefore define, for each N-dimensional attractor, a corresponding constant N-dimensional "pseudo-attractor", in which each component assumes the value 1 if its time average in the dynamic attractor is $\geq\theta$ (in the following we suppose θ=0.5) and take the value 0 otherwise. As a consequence, the relationship between dynamical attractors and pseudo-attractors is not injective, and it qualitatively corresponds to a kind of coarse graining in phase space.

The "common sea" (CS) is then defined as the set of nodes which take the same value in all the pseudo-attractors of a given network realization, while the set of all the other nodes is called the "specific part" (SP). Note that the concept of CS differs from existing ones like the "frozen sea" [7] in that it is based on pseudo-attractors (so that also oscillating nodes can belong to it) and it requires that the nodes take the same value in all the pseudo-attractors.

We studied the properties of the CS and the SP by simulating RBNs which belong to different ensembles, generated with different parameter values. The following results have been presented in [6], so we will avoid to continuously refer to it. Most simulations concern dynamically critical networks (i.e. those whose parameters take values which separate the regions of ordered behaviors from the pseudo chaotic ones) which are particularly interesting, for reasons discussed at length in the literature [7,8], which will not be reviewed in this extended abstract.

It turns out that the fraction of nodes belonging to the CS of critical networks increases as the overall size of the network

(N) is increased, and that it comprises the majority of the nodes. This may look surprising but a simplified mean-field calculation shows that it should indeed be expected. An interesting result comes from the comparison of dynamically critical networks with different average number of connections per node (k) and different biases (b) of the Boolean functions. Indeed, dynamical criticality imposes a relationship between k and b, so it is possible to consider ensembles of networks with different pairs of values. Perhaps surprisingly, simulations show that the criticality condition does not suffice to determine the size of the CS. It appears that the larger the bias, the larger the CS.

It is also interesting to observe the internal organization of the common sea and of the specific part. For example, once the CS of a given network realization is identified, we can look at the topology of the network that is composed of its nodes only. If we identify the subparts of the CS with its weakly connected components (WCCs), then there is often (in 70% of the cases) only one subpart per network realization, but in other cases, there are more than one, although one usually finds a dominant subpart that comprises many more nodes than the others. If we perform a similar analysis on the specific part, we often find a more evenly distributed situation, with more fragments of similar size. It should be emphasized that these are not completely independent parts, and that some changes in one WCC (for example, the knock-out of a gene) can affect the values of nodes in other WCCs.

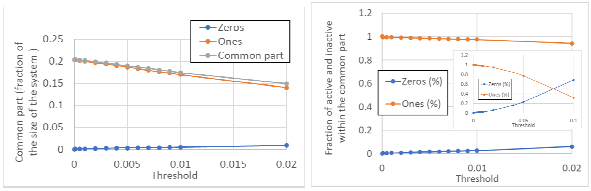
The presence of a large CS obviously limits the maximum possible distance between pairs of attractors, whose distribution turns out to be unimodal.

A Preliminary Look at Single-Cell Data

While we do not aim at an in-depth comparison of the behavior of the models with experimental data, we suggest that looking at experimental data through the lenses of our models can lead to new insights and new questions. To show how this might work, we have performed a preliminary analysis of an important experimental data set concerning the expression levels of human single cells [9]. Although these data are very noisy, since many different exemplars of each type are available, it is possible to aggregate all the contributions into a single profile that then constitutes the "average profile" of the cell type. Since these averages are real valued, binarization is necessary to compare them with pseudo-attractors: a simple way in which binarization can be achieved is by rescaling the values for each gene so to match the [0,1] interval, and by comparing the rescaled values with a fixed threshold ζ.

The approach is simple and it could certainly be refined: however, it is very interesting to observe that the notions of a common sea and specific parts, which have been defined here in a model system, can also be applied to experimental data, as shown e.g. in fig.1.

It remains to investigate possible independent criteria to determine the correct threshold values (possibly different for different genes [10]), an activity that we leave for future work. Here we can note that the approach potentially allows to identify new constraints that simulation models must satisfy in order to correctly interpret experimental data.



(a) (b)

Figure 1 (a) Size of the common part in the Human Cell Landscape data, as the threshold ζ varies. (b) Fraction of active ("Ones") and inactive ("Zeros") genes out of the total of genes belonging to the common sea (in the insert the threshold ζ reaches the value 0.1).

Moreover, the rescaled profiles can be used to compute the distribution of distances (Hamming distances) between cell types – a quantity which can be computed also for our simulated models.

We believe that this kind of comparisons will prove really fruitful to improve both theory and experiment, as it provides new constraints on the acceptable parameters of the model as well as new quantities which are worth measuring. A thorough quantitative comparison between theoretical models and experimental data will be the subject of further work.

Acknowledgments. Useful discussions with Andrea Roli and Michele Braccini are gratefully acknowledged

References

[1] Schwab, J.D.; Kuehlwein, S.D.; Ikonomi, N.; Kuehl, M.; Kestler, H.A. (2020) Concepts in Boolean network modeling: What do they all mean? Comput. Struct. Biotechnol. J., 18:571-582

[2] Kauffman, S.A. (1969) Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets. Journal of Theoretical Biology 22:437–467

[3] Harvey, I.; Bossomaier, T. (1997) Time out of Joint: Attractors in Asynchronous Random Boolean Networks. In Husbands, P., Harvey, I., editors, *Proceedings of ECAL 97*, pages 67–75. MIT Press, Cambridge, MA

[4] Darabos, C,, Tomassini, M., Giacobini, M. (2009) Dynamics of Noisy and Unperturbed Generalized Boolean Networks. ArXiv 0909.5297v1, DOI:10.1016/j.jtbi.2009.06.027

[5] Gershenson, C. Introduction to Random Boolean Networks. arXiv 2004, arXiv:nlin/0408006

[6] Villani, M.; D'Addese, G.; Kauffman, S.A.; Serra, R. Attractor-Specific and Common Expression Values in Random Boolean Network Models (with a Preliminary Look at Single-Cell Data). Entropy 2022, 24, 311. https://doi.org/10.3390/e24030311

[7] Kauffman, S.A. At Home in the Universe (1995) The Search for Laws of Self-Organization and Complexity; Oxford University Press: Oxford, UK

[8] Roli, A.; Villani, M.; Filisetti, A.; Serra, R. Dynamical Criticality: Overview and Open Questions. J. Syst. Sci. Complex. 2018, 31, 647–663. https://doi.org/10.1007/s11424-017-6117-5

[9] Han, X.; Zhou, Z.; Fei, L.; Sun, H.; Wang, R.; Chen, Y.; Chen, H.; Wang, J.; Tang, H.; Ge, W.; et al. (2020) Construction of a Human Cell Landscape at Single-Cell Level. Nature, 581:303–309

[10] Müssel C.; Schmid F.; Blätte T.J.; Hopfensitz M.; Lausser L.; Kestler H.A.; BiTrinA—multiscale binarization and trinarization with quality analysis, Bioinformatics, 2016, V.32, Issue 3 465–468, https://doi.org/10.1093/bioinformatics/btv591