Dynamical Properties of a Gene-Protein Model

[1] Department of Physics, Informatics and Mathematics,
Modena and Reggio Emilia University, Modena, Italy
{marco.villani,rserra}@unimore.it
[2] European Centre for Living Technology, Ca' Foscari University, Venice, Italy

Abstract. A major limitation of the classical random Boolean network model of gene regulatory networks is its synchronous updating, which implies that all the proteins decay at the same rate. Here a model is discussed, where the network is composed of two different sets of nodes, labelled G and P with reference to "genes" and "proteins". Each gene corresponds to a protein (the one it codes for), while several proteins can simultaneously affect the expression of a gene. Both kinds of nodes take Boolean values. If we look at the genes only, it is like adding some memory terms, so the new state of the gene subnetwork network does no longer depend upon its previous state only.

In general, these terms tend to make the dynamics of the network more ordered than that of the corresponding memoryless network. The analysis is focused here mostly on dynamical critical states. It has been shown elsewhere that the usual way of computing the Derrida parameter, starting from purely random initial conditions, can be misleading in strongly non-ergodic systems. So here the effects of perturbations on both genes' and proteins' levels is analysed, using both the canonical Derrida procedure and an "extended" one. The results are discussed. Moreover, the stability of attractors is also analysed, measured by counting the fraction of perturbations where the system eventually falls back onto the initial attractor.

Keywords: Gene-protein model · Generic properties · Memory effect
Dynamical regimes

1 Introduction

Random Boolean models of genetic regulatory networks (RBNs) are very well-known and, in spite of their long age, they still provide useful descriptions of important observational and experimental results [8, 12–17]. A major limitation of the classical RBN model is its synchronous updating: from a physical viewpoint, this amounts at assuming that all the proteins decay at equal rates: this unrealistic assumption allows one to write the gene activation pattern at time t + 1 as a function of that pattern at time t, forgetting the previous history. Asynchronous updating has been sometimes proposed (one gene at each time step), but this also leads to difficult interpretations, due to the relatively large typical protein decay time and to the very large number of genes. Other interesting "intermediate" update strategies have also been proposed [5, 19].

Some properties of RBNs are robust with respect to the updating strategy, but in general there is no guarantee that this is the case. In particular, one should be very careful when dealing with the networks' dynamical properties. We have been particularly interested in the response of genetic networks to perturbations like gene knock-out and we have shown that, if the RBN model is chosen, the distribution of avalanches in gene expression levels in S. Cerevisiae that follows a single knock-out provides information about the dynamical regime of the biological network [8, 16]. This result is particularly relevant, given the importance of the "criticality hypothesis", which states that biological systems should preferentially be found in dynamically critical states [13]. If we are indeed interested in biological genetic networks, such issues should be addressed in a way that does not critically depend upon the unrealistic assumption of synchronicity: different updating schemes should be considered, privileging whenever possible those that are closer to what we know about the behaviour of real gene regulatory networks.

In order to do so, while retaining the simplifications related to the use of Boolean variables and to the "generic" approach of RBNs, we introduced the GPBN model (Gene-Protein Boolean Network), where the network is composed of two different sets of nodes, labelled G and P with reference to "genes" and "proteins" [9–11]. It is now well-established that proteins are not the only genetically-encoded products which can influence the effective expression level of other genes (think for example of miRNAs [2, 3]). However, in order to simplify the model description, we will call here "proteins" all the products of gene activation that are able to influence the expression of other genes.

Each gene corresponds to a protein (the one it codes for), while several proteins can simultaneously affect the expression of a gene. Both kinds of nodes take Boolean values: the state at time t + 1 of a G node depends upon the state of a fixed set of P nodes at the same time, while the state at time t + 1 of a P node depends upon the state of its corresponding G node at time t. Once a P node is set active (its state is 1), it remains active for at least a fixed number of steps. If a new activation signal comes in before decaying, the counter is reset. If no activation signal arrives, the P node is set to 0 at the end of its "lifespan". If we look at the genes only, it is like adding some memory terms, so the new state of the network is no longer "Markovian", i.e. it does no longer depend upon the previous state only.

This model has been thoroughly studied and its properties have been described elsewhere [9, 11]. In those papers the usual definition of dynamical criticality, based on the value of the so-called Derrida parameter, had been used. We have recently shown some limitations related to the use of that single measure to characterize critical states in RBNs [4]. In particular, the choice of a completely random initial state in the computation of the Derrida parameter has been criticized and a different measure ("extended Derrida parameter") has been proposed [18].

This prompted a more thorough analysis of the dynamics of GPBNs, whose main features are presented in this paper.

The paper is organized as follows: in Sect. 2 the GBPN model is described, while in Sect. 3 the measures of dynamical criticality are discussed and the extended Derrida parameter is introduced. In Sect. 4 the results obtained by simulating GPBNs are shown and discussed, paying particular attention to the similarities and differences between the "canonical" (i.e. standard) and the extended Derrida procedures.

marco.villani@unimore.it

A different way to evaluate the robustness of the network behaviour, based upon perturbations of its dynamical attractors, is also presented. Critical discussion and suggestions for further research are summarized in Sect. 5.

2 The GPBN Model

A GPBN model [9–11] is a bipartite oriented graph containing two types of Boolean nodes: the G nodes, which represent the genes set, and the P nodes, which represent the set of proteins (or, in general, gene products). A G node can be active or inactive (producing or not its protein), whereas a P node describes the presence (or absence) of a protein within the system. There are two types of links: *synthesis links*, which go from a G node to only one P node, and *transcriptional regulation links*, from a P node to one or more G nodes.

As usual in RBNs, time evolves in discrete steps. Note that the state at time t + 1 of the GPBN model is determined by its state at time t, and the update is formally synchronous. However, due to the presence of the P nodes, the updating of the gene subnetwork is not synchronous, i.e. the states of G nodes at time t + 1 are not determined by their states at the previous time step.

Each G node, say the j-th, produces its protein when active (synthesis link) and a G node is driven by the action of its k inputs (k being the number of its transcriptional regulation links, coming from P nodes), according to a fixed Boolean function f_j associated to it (f_j: $\{0, 1\}^k \rightarrow :\{0, 1\}$).

The topology of the transcriptional links is random, and so is the choice of the Boolean functions: each f_j is generated by assigning at random to each of its 2^k possible inputs an output equal to 1 with probability p (the so-called bias of the set of Boolean functions), 0 otherwise.

To each P node, say the i-th, an integer non-negative variable h_i is also associated (its *decay phase*) which can change in time and which represents its residual lifetime. The maximum value of h_i is the *decay time dt_i* of node i, representing the lifespan of the protein, once activated (i.e. just synthesized). When a P node is activated, its decay phase h_i takes the value dt_i and it is later decreased by 1 at each time step, until it ends in 0 (unless the same node is not activated again in that time interval). When the incoming G node is active, then the corresponding P node resets its decay phase to the decay time. As long as the decay phase takes a nonzero value, the P node has a regulation role on its outgoing links (i.e. its value in the transition function is 1).

The decay time of each node is taken randomly with uniform probability between 1 and a parameter defined as *maximum decay time* (MDT); note that when MDT is equal to 1 the GPBN is identical to the corresponding RBN (i.e. the one with the same topology and the same activation functions). If the value of a G node is 1 at time t then the value of the corresponding P node will be 1 at time t + 1 and its decay phase will be set to dt_i, otherwise the decay phase of the P node is decremented by one unit (in case of $dt_i = 0$, the activation of P is set to 0). On the other hand, the value of the G node at time t is immediately determined by its function f_j, which depends on the states of its incoming P nodes at time t.

3   Dynamical Regimes

The asymptotic states of finite RBNs are periodic cycles; fixed points correspond to cycles with unitary period. Different dynamical regimes have been observed in RBNs [1, 13, 14], classified as disordered (sometimes called "chaotic", although all the attractors are indeed periodic), ordered or critical depending upon the length of their periods and the sensitive dependence upon initial conditions. In chaotic networks the cycle length sharply increases with the network size, and nearby initial states are likely to lead to different attractors, while in ordered systems the typical cycle length shows a polynomial dependence upon the number of nodes, and basins of attraction are quite regular. Given the random nature of these systems, the analysis usually concerns families of networks built by keeping fixed some parameters, like e.g. the number of nodes, the average number of connections per node and/or the average bias of the Boolean functions, while changing in different network realizations the topology of connections and the transition functions. Critical networks are those whose parameters lie on (or close to) the manifolds that separate regions in parameter space with ordered behaviours from the chaotic regions. It is important to stress that these terms refer to the typical features of networks with those parameters, while a single network realization can behave in a way very different from the typical ones. Large deviations from typical behaviours can easily be found in critical networks [15].

The asymptotic dynamics can be identified by means of the so-called dynamical Derrida parameter λ [6, 7], which measures the tendency of a temporary perturbation to vanish, to persist or to spread through the entire system: so, ordered, critical and chaotic dynamical regimes correspond respectively to $\lambda < 1$, $\lambda \approx 1$ and $\lambda > 1$.

This parameter can be determined by analysing a plot of the average distance between two states at time $t + 1$ versus their distance at time t (the Derrida plot) and by looking at the slope of the tangent to the curve in the limit of small initial distances.

Different (static) measures of the dynamical properties have also been proposed, based on an analysis of the properties of the set of Boolean functions rather than on actual simulations: they are discussed in depth in [18] alongside with their relationships with the dynamical Derrida parameter, described above, which is the only such measure considered in this paper.

Another important remark raised in [18] concerns the dependency of the dynamical Derrida parameter from the set of initial conditions. The usual recipe is that of choosing a fully random initial state, and of considering the time behaviour of its perturbed states. While this is entirely reasonable in ergodic systems (where all accessible states are equiprobable over a long period of time), RBNs with a small number of connections per node are strongly non-ergodic [20], so it may easily happen that such purely random states are never encountered in the life of the cell modelled by the Boolean genetic network.

It seems therefore physically much more appropriate to determine the dynamical Derrida parameter while limiting the set of allowed initial states only to those states that are the successors of some other states. The initial state might be found by starting the network simulation from a purely random state, letting it evolve for T_{ev} steps ($T_{ev} \geq 1$) and by choosing the state that has been reached as the initial state for computing the

Derrida parameter. When the set of allowed initial states is limited in this way, we refer to an "extended Derrida approach", or to an "extended Derrida parameter", to distinguish it from the canonical one.

Note also that different types of perturbations are possible: in GPBNs the initial perturbation could affect G nodes, P nodes, or both. In our approach a perturbation of a P node can correspond either (i) to an activity change from 0 to 1, with a decay phase h_i randomly chosen within the range $[1, dt_i]$ or (ii) to an activity change from 1 to 0, with $h_i = 0$. A perturbation of a G node can correspond (i) to an activity change from 0 to 1, followed by the appropriate effect on the protein or (ii) to an activity change from 1 to 0 – in this case, the G node is not producing its protein, and the P node reduces its decay phase by one.

4   Results

It had already been observed in [9, 11] that, as it might be apriori expected, the presence of a memory term tends to make the dynamical behaviour "more ordered". This can be shown by comparing the behaviour of networks with MDT $\neq$ 1 with those of the corresponding network with MDT = 1 (that are identical to the corresponding RBNs). The comparison can be made for different dynamical behaviours, in this paper we will report results concerning networks that are critical if MDT = 1. Three sets of parameters, all corresponding to critical behaviours, will be discussed: [k = 2, p = 0.5], [k = 3, p = 0.21], [k = 3, p = 0.79]. The fact that two different cases are chosen for k = 3 is due to the fact that in GPBN the 0–1 symmetry of RBNs no longer holds.
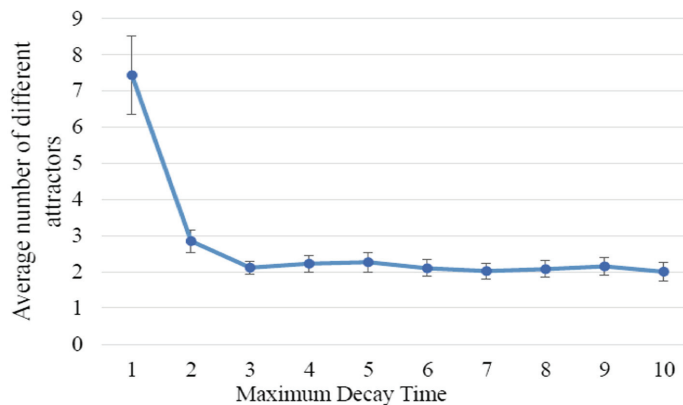

Fig. 1. Number of different attractors vs. maximum decay time (MDT, ranging from 1 to 10); each point represents the average of 1000 different networks (case [k = 2, p = 0.5]) with 100 G-P node pairs. For each network 100 runs with different initial conditions are performed, until an attractor (with period lower than 1000 time steps) is reached or until the sum of the transient time exceed 10000 time steps

The stabilizing effect of memory can be seen in Fig. 1, where the number of different attractors versus the maximum decay time is shown to decrease sharply even with a short memory term [9].

Let us now turn to the dynamical regime, as determined by the Derrida procedure. As discussed in Sect. 3, perturbations can be performed either on G or on P nodes. Let us first consider this latter case. In all the simulations described here below the perturbations can be either up (i.e. setting equal to one the value of a P node which is 0) or down, depending on the not perturbed activity of the chosen P node. In each simulation series we create 50 different networks with 100 G-P node pairs, 100 different initial conditions for each network. In order to allow an easier series comparison we consider the decay time of each P node being exactly equal to MDT.[1]

In Fig. 2 the behaviour of the Derrida parameter for the critical case k = 2, p = 0.5 is shown. The two curves refer to the G-node and to the P-node subnetworks. Very large values of MDT have also been considered, and it is shown that the network remains critical notwithstanding the memory term.
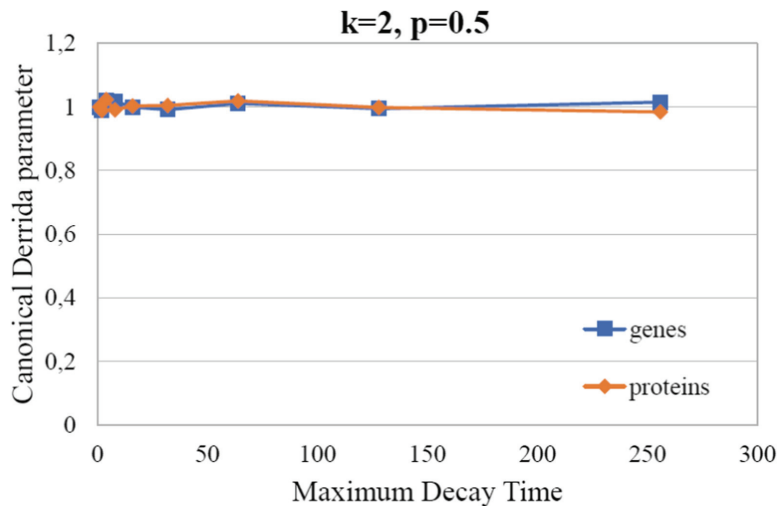

Fig. 2. Canonical Derrida parameter vs MDT (MDT $\in$ {1, 2, 4, 8, 16, 32, 64, 126, 256}), case k = 2, p = 0.5. The two curves refer to the G-node and to the P-node subnetworks, subject to a P-node perturbation

In Fig. 3 the same parameter is shown for the two cases with k = 3. While the G-node subnetwork remains critical, here the effect of the memory term on the P subnetwork is neither that of leaving it critical, nor that of always bringing it in the

---

[1] Subsequent simulation series where the decay time of each node is randomly chosen (with uniform probability) in [1, MDT] show that the main effect of choosing the decay times randomly with uniform probability between 1 and MDT is that of slightly soften the shape of the curves, without altering their behavior (data not shown).
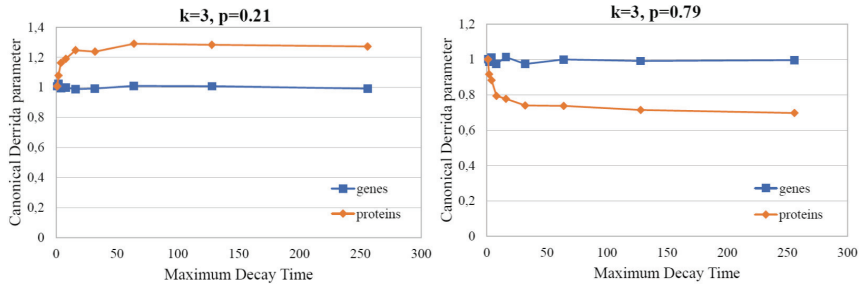
Fig. 3. Canonical Derrida parameter vs MDT, case k = 3: left p = 0.21, right p = 0.79. The two curves refer to the G-node and to the P-node subnetworks, subject to a P-node perturbation

ordered region; this happens for the case with high bias, while the Derrida parameters becomes larger than one in the low-bias case.

This behaviour may seem surprising (but see the comments in Sect. 5), therefore it is interesting to consider also the extended Derrida parameter described in Sect. 3. The results are shown in Figs. 4 and 5.
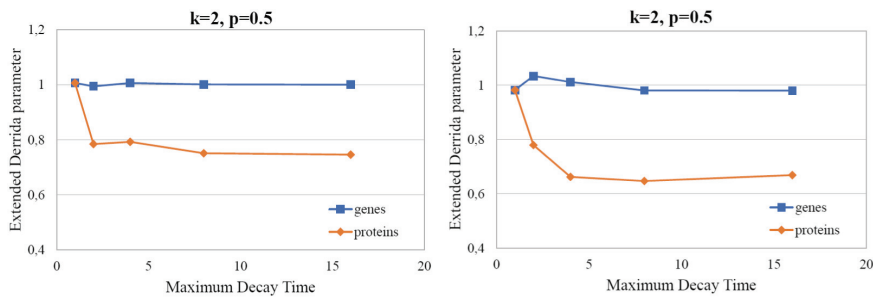

Fig. 4. Extended Derrida parameter vs maximum decay time for the case k = 2, p = 0.5; left $T_{ev} = 1$, right $T_{ev} = 3$. The two curves refer to the G-node and to the P-node subnetworks, subject to a P-node perturbation
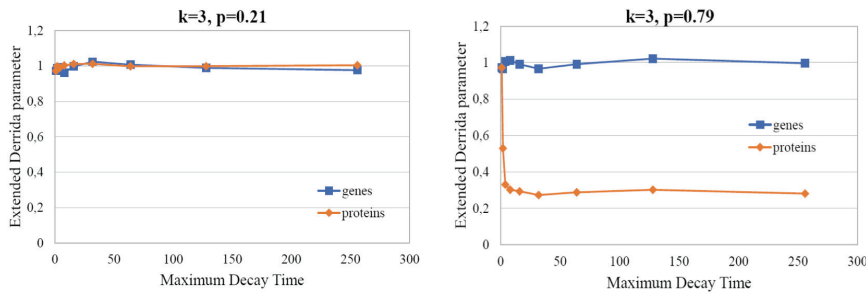

Fig. 5. Extended Derrida parameter vs maximum decay time for the case k = 3; Left p = 0.21, right p = 0.79. In both cases $T_{ev} = 3$. The two curves refer to the G-node and to the P-node subnetworks, subject to a P-node perturbation

Note that, while the G subnetwork remains critical, the behaviour of the P subnetwork is different from that of the canonical Derrida parameter. In the k = 2 case, it is more ordered ($\lambda < 1$ even for values of MDT slightly larger than 1) while it was critical in Fig. 2. In the k = 3, low-bias case the network is critical, while it was supercritical in Fig. 3. Only in the case of k = 3 with low bias the two behaviours are at least qualitatively the same. It should also be observed that the length of the time window T_{ev} may affect the outcomes: for example, by choosing it equal to one in the same case as that of Fig. 5 left, one would have concluded that the P subnetwork is slightly supercritical (data not shown here).

In order to complete the description of the model behaviours, let us now consider the results that have been obtained by perturbing the gene subnetwork (recall that all the previous ones referred to perturbations of P nodes). As it can be seen from Fig. 6 below, in all the cases both subnetworks are ordered even for values of MDT larger than 1.
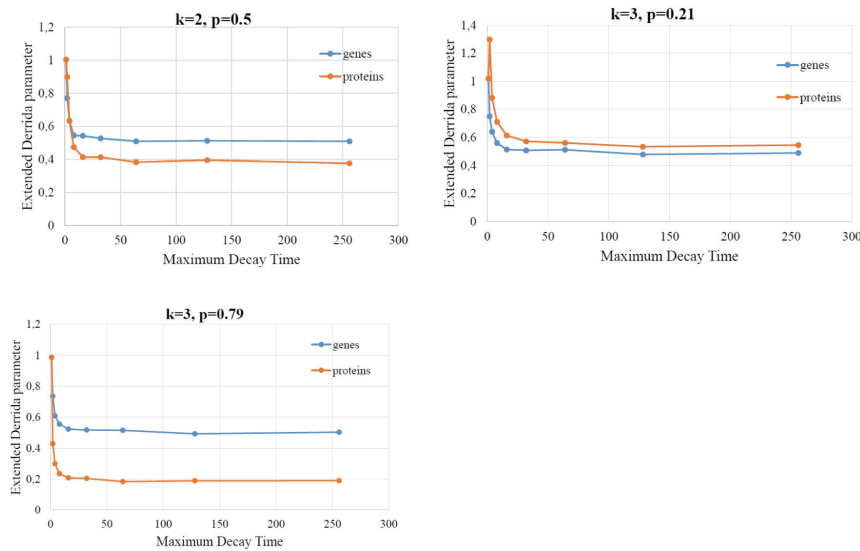

Fig. 6. Extended Derrida parameter vs maximum decay time for the cases k = 2 and p = 0.5, k = 3 and p = 0.21, k = 3 and p = 0.79. In all cases $T_{ev} = 1$. The curves refer to the G-node and to the P-node subnetworks, subject to a G-node perturbation

The dynamical regimes of GPBNs have been analysed so far by using canonical or modified Derrida methods, i.e. the discrete analogues of Lyapunov exponents. A major interest concerns the robustness of networks of this kind, and in order to characterize this property a different measure, independent of T_{ev} or of any similar parameter, is given by the fraction of perturbations that, starting from an attractor cycle, end in the same attractor.

These data are shown in Fig. 7. As it is expected, the fraction of perturbations that fall back onto the initial attractor decreases as the intensity of the perturbation increases. This fraction increases when a memory term is added and, like in the other cases described above, the effect is observed for small values of the maximum decay time, while further increases of MDT do not lead to any appreciable change.
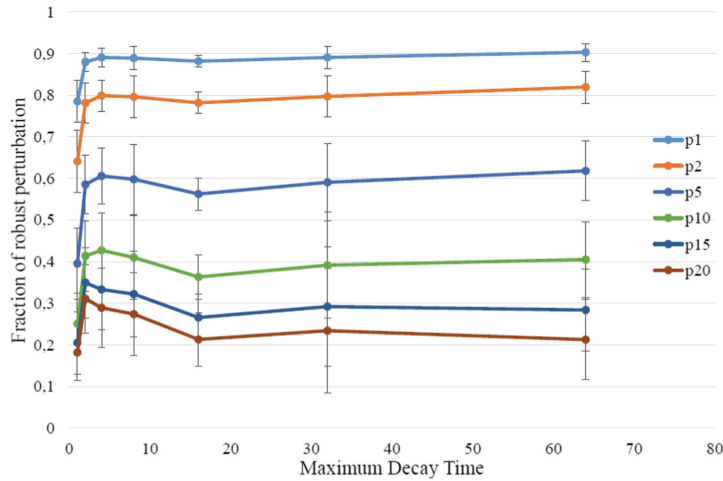

Fig. 7. The fraction of perturbations that came back to the starting attractor by varying MDT, if perturbing 1, 2, 5, 10, 15 or 20 P-nodes. Each point is the average of 50 different systems with 100 GP nodes: in each system the attractors are identified by using 100 random initial conditions; all states of the so sampled attractors are perturbed. In these experiments, we considered the same decay time for each P node.

5   Conclusion

The GPBN model of genetic regulatory systems maintains the abstraction level of the RBN framework and at the same time allows an explicit modelling of time delay effects.

It is of course extremely interesting to compare abstract-level models with real-world data. It has indeed been possible to show that RBNs can properly describe the distribution of perturbations in gene expression levels induced by single knock-outs in S. Cerevisiae [15, 16]. However, the techniques used for this purpose do not allow one to test the behaviour of the model when the perturbation affects several genes at the same time – a situation that is much more frequently encountered in experiments, like those related to the effects of drugs or contaminants. In these cases the comparison of model behaviour and experimental data should concern the time behaviour of the perturbation after the initial shock, but time-course data cannot be properly compared to RBNs because of their unrealistic synchronous updating. On the contrary, the introduction of memory terms in GPBNs should make it possible to deal also with

time-course data following a multiple initial perturbation, thus greatly increasing the wealth of experimental data available for testing the appropriateness of the abstract framework.

The kind of memory that has been introduced has different effects in case of information transmission from G to P nodes or from P to G nodes, and pose some interesting questions about the correct way of measuring of the system dynamical regimes through Derrida-like procedures. Anyway, the robustness of the system's attractors can constitute a sort of global measure related to its general "degree of order". In the future it will be interesting to analyse a Derrida parameter modified in a way different from those of Sect. 4, i.e. computed by allowing as initial states only those that belong to an attractor.

In order to understand the behaviour of the GPBN model when P nodes are perturbed, it will be interesting to consider separately the effects of up and down perturbations. Indeed, the impacts of "up" and "down" perturbations of P nodes are likely to have different intensities. The effect of a "down" perturbation, i.e. the disappearance of a protein, should typically die out quite rapidly, as the rest of the nodes resynthesize that protein. On the other hand, the impact of an "up" perturbation is likely to last longer, i.e. for a number of steps equal to its phase. Investigating the effects of the two types of perturbations by canonical and modified Derrida parameters may therefore provide important clues about the properties of the model.

References

1. Aldana, M., Coppersmith, S., Kadanoff, L.P.: Boolean dynamics with random couplings. In: Kaplan, E., Marsden, J.E., Sreenivasan, K.R. (eds.) Perspectives and Problems in Nonlinear Science, pp. 23–89. Springer, New York (2003). https://doi.org/10.1007/978-0-387-21789-5_2
2. Bartel, D.P.: MicroRNAs: genomics, biogenesis, mechanism, and function. Cell **116**, 281–297 (2004)
3. Bartel, D.P.: MicroRNAs: target recognition and regulatory functions. Cell **136**(2), 215–233 (2009)
4. Campioli, D., Villani, M., Poli, I., Serra, R.: Dynamical stability in random Boolean networks In: Apolloni, B., Bassis, S., Esposito, A., Morabito, F.C. (eds.) Frontiers in Artificial Intelligence and Applications, WIRN, vol. 234, pp. 120–128. IOS Press, Amsterdam (2011)
5. Darabos, C., Giacobini, M., Tomassini, M.: Generalized Boolean networks: how spatial and temporal choices influence their dynamics. In: Handbook of Research on Computational Methodologies in Gene Regulatory Networks, pp. 429–449. IGI Global, Hershey (2010)
6. Derrida, B., Pomeau, Y.: Random networks of automata: a simple annealed approximation. Europhys. Lett. **1**(2), 45–49 (1986)
7. Derrida, B., Weisbuch, G.: Evolution of overlaps between configurations in random Boolean networks. J. Phys. **47**, 1297–1303 (1986)
8. Di Stefano, M.L., Villani, M., La Rocca, L., Kauffman, S.A., Serra, R.: Dynamically critical systems and power-law distributions: avalanches revisited. In: Rossi, F., Mavelli, F., Stano, P., Caivano, D. (eds.) WIVACE 2015. CCIS, vol. 587, pp. 29–39. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-32695-5_3

marco.villani@unimore.it

9. Graudenzi, A., Serra, R., Villani, M., Damiani, C., Colacci, A., Kauffman, S.A.: Dynamical properties of a Boolean model of gene regulatory network with memory. J. Comput. Biol. **18**, 1291–1303 (2011)

10. Graudenzi, A., Serra, R.: A new model of genetic network: the gene-protein network. In: Serra, R., Poli, I., Villani, M. (eds.) Artificial Life and Evolutionary Computation, pp. 283–291 (2009)

11. Graudenzi, A., Serra, R., Villani, M., Colacci, A., Kauffman, S.A.: Robustness analysis of a Boolean model of gene regulatory network with memory. J. Comput. Biol. **18**(4), 559–577 (2011)

12. Kauffman, S.A.: Homeostasis and differentiation in random genetic control networks. Nature **224**, 177–178 (1969)

13. Kauffman, S.A.: The Origins of Order: Self Organization and Selection in Evolution. Oxford University Press, Oxford (1993)

14. Kauffman, S.A.: At Home in the Universe. Oxford University Press, New York (1995)

15. Serra, R., Villani, M., Semeria, A.: Genetic network models and statistical properties of gene expression data in knock-out experiments. J. Theor. Biol. **227**, 149–157 (2004)

16. Serra, R., Villani, M., Graudenzi, A., Kauffman, S.A.: Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data. J. Theor. Biol. **246**, 449–460 (2007)

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

18. Villani, M., Campioli, D., Damiani, C., Roli, A., Filisetti, A., Serra, R.: Dynamical regimes in non-ergodic random Boolean networks. Nat. Comput. **16**(2), 353–363 (2017)

19. Darabos, C., Tomassini, M., Giacobini, M.: Dynamics of unperturbed and noisy generalized Boolean networks. J. Theor. Biol. **260**(4), 531–544 (2009)

20. Roli, A., Villani, M., Filisetti, A., Serra, R.: Dynamical criticality: overview and open questions. J. Syst. Sci. Complex **30**, 1–17 (2017)