



Dynamical properties and path dependence in a gene-network model of cell differentiation

Michele Braccini¹ · Andrea Roli^{1,4} · Marco Villani^{2,4} · Roberto Serra^{2,3,4}

Published online: 2 November 2020
© The Author(s) 2020

Abstract

In this work, we explore the properties of a control mechanism exerted on random Boolean networks that takes inspiration from the methylation mechanisms in cell differentiation and consists in progressively freezing (i.e. clamping to 0) some nodes of the network. We study the main dynamical properties of this mechanism both theoretically and in simulation. In particular, we show that when applied to random Boolean networks, it makes it possible to attain dynamics and path dependence typical of biological cells undergoing differentiation.

Keywords Boolean networks · Cell differentiation · Path dependence · Epigenetics · Methylation

1 Introduction

The co-existence of different cell types, which share the same genome, in a multicellular organism, and the processes of progressive cell differentiation raise challenging theoretical issues. The possibility that the same set of genes, with the same type of reciprocal influences, gives rise to very different phenotypes is associated to the fact that not all the genes are active in every cell type, i.e. to the existence of different stable activation patterns of the same set of genes. An important question which naturally raises concerns how these stable patterns are determined, among the huge amount of states which are possible in principle. It is well-known that active genes interfere with the activation of other genes, through proteins or other gene products like, e.g. miRNA. Therefore, it is natural to associate the stable patterns, i.e. the cell types, to the attractors of the dynamical system which describes the

complex web of interactions among genes, RNA, proteins, other gene products and other molecules in the cell.

There are beautiful models of various parts of this system; for example, some describe in detail the steps which lead from DNA to mRNA to ribosomes, and some describe the details of the regulation processes. They are very well suited to describe in-depth the translation and transcription processes, but here we are rather interested in global activation patterns of thousands of heterogeneous entities. Therefore, it is mandatory to introduce some simplifications to keep the model manageable and meaningful. One interesting possibility (pioneered by Stuart Kauffman about 50 years ago) is that of simplifying the picture to that of a network of interacting genes only, without explicitly taking into account the other entities—whose influence is subsumed by the interaction rules among the genes.

Following this approach, we will consider a dynamical system of n “genes”, which can take different activation values. The time change of the activation of a gene is ruled by a differential (or difference) equation; the rules can be different for different genes. As time passes, the system will tend to one of its stable steady states—to be associated to a cell type.

This framework is attractive, due to its simplicity, yet it has to be reconciled with the crucial process of cell differentiation, i.e. the development of pluripotent types into fully differentiated ones, through various intermediate stages. If we associate a pluripotent type with a stable or metastable attractor, then it is necessary to understand what can make this

Communicated by Tomas Veloz.

✉ Michele Braccini
m.braccini@unibo.it

¹ Department of Computer Science and Engineering, Campus of Cesena - Alma Mater Studiorum Università di Bologna, Cesena, Italy

² Department of Physics, Informatics and Mathematics, Università di Modena e Reggio Emilia, Modena, Italy

³ Institute for Advanced Study (IAS), University of Amsterdam, Amsterdam, The Netherlands

⁴ European Centre for Living Technology, Venice, Italy

attractor unstable. One might be tempted to identify the process of cell maturation with transients, while attractors would be associated to mature cell types only. However, pluripotent intermediate types may be long lived, so it is difficult to devise satisfactory models along this line. It seems more appropriate to associate attractors also to the long-lived intermediate states, but in this case it is necessary to identify mechanisms which can actually modify the dynamical system, making these attractors unstable and driving the system towards the other attractors.

This is possible since the cell is not a fully autonomous system, but it is coupled with the environment in which it lives. Various molecules can indeed interfere with the gene expression mechanisms, switching on or off single genes or groups of genes. However, one has to resist the temptation of simplistic explanations based only upon detailed properties of external influences, because the gene network (or rather, the whole dynamical system it represents) is always present. Switching one gene on or off does not only lead to the synthesis of a specific protein or gene product, but it may also affect the expression of other genes. The activation patterns must be coherent with the whole dynamics of the network Huang and Kauffman (2013).

In this paper, we explore the behaviour of a control mechanism, which affects the attractors of a gene network and their stability. This mechanism, which is simple enough to be amenable to large-scale simulation, is directly inspired by biological observations and experimental knowledge about the effects of DNA methylation; refer to Sect. 2 for a brief introduction on biological methylation and its role in the gene expression process.

While the model is described in detail in Sect. 2, let us mention here that we will make the simplifying assumption that nodes are either off or on, i.e. that they take Boolean values, neglecting intermediate values. Moreover, let us suppose that the time evolution of the network is synchronous and deterministic in discrete steps. The attractors of any finite network are therefore either cycles or fixed points (which can be regarded as period-1 cycles). The silencing effects of methylation will therefore be modelled by keeping constantly equal to zero the activation of some nodes (which will be called “externally frozen”, briefly *e-frozen*, or sometimes “blocked”). The network is assumed to be in one of its attractors at time $t < 0$, while at time $t = 0$ some nodes are externally frozen. The clamping to zero of a node may affect other nodes, inducing in time a propagation of freezing (i.e. an increase in the number of constant nodes) and the attractor landscape will be modified.

In Sect. 3, we will present a simplified, mean-field description of this process of propagation in time of freezing. In Sect. 4, we will describe the results of several simulations of the effects of freezing on the attractor landscape, investigating the effects of the choice of some key parameters. A

particularly important question concerns the importance of the order (in time) of gene silencing: if gene *A* is silenced at a certain time step, and gene *B* is silenced at a later stage, will the attractors and their basins be the same as in the case where silencing of *B* precedes that of *A*? Interestingly, one can prove that the answer is (sometimes) no, so the system described here can show a strong form of *path dependence*. This is described in detail in Sect. 5. The final section is devoted to a discussion of the results and of their biological implications.

2 Methylation model

Epigenetic mechanisms significantly contribute to the determination and maintenance of cell fates in biological organisms. Methylation, in particular, can occur to both DNA regions and proteins. DNA methylation typically occurs at *CpG sites*¹ changing the activity of DNA sequences, e.g. blocking gene promoter activities and hence their ability to induce transcription, without requiring mutations. Differently, histone methylations change the degree of compactness of the chromatin by adding methyl groups to histone proteins: proteins on which DNA of eukaryotic cells wrap around to form structural units called nucleosome. Changes the degree of compactness of chromatin have implications on the expression of genes belonging to the DNA regions involved: condensed DNA (*heterochromatin* condition) prevent transcription by polymerases, while loosely packed regions (*euchromatin* condition) allow transcription. Differential methylation is therefore a mechanism exploited by biological cells to modulate gene regulation and expression during development and differentiation.

Taking into account histone methylation, we observe that although its effects depend on the particular positions on histones on which it acts, it most often leads to heterochromatin conditions (Gilbert and Barresi 2016; Perino and Veenstra 2016; Schuettengruber and Cavalli 2009). In addition, along lineages, the attained configurations of DNA methylation are inherited and progressively extended as cells become more specialised Kim and Costello (2017). Therefore, methylation has a prominent role in the maintenance and stabilisation of the attained gene expressions that ultimately characterise the identities of the various cell states. It is also worth mentioning that methylation is tightly regulated by complex interactions, and that its dysregulation can be the *causa prima* of a lot of disorders, from cognitive, neurological and chronic diseases to cancer. It is precisely due to the complexity of these mechanisms that the adoption of models can support the analysis

¹ Sites in which a cytosine nucleotide is followed by a guanine nucleotide.

of the role of methylation, and more in general of epigenetics, in (patho)physiological processes.

Several works making use of mathematical models of specific aspects of epigenetic processes are present in the literature (Bull 2014; Miyamoto et al. 2015; Turner et al. 2017, 2013); however, to the best of our knowledge, there is no systematic attempt to study the capabilities of reproducing differentiation dynamics by methylation-like mechanisms in modelling. A noteworthy model of differentiation, based on an entirely different mechanism (i.e. intracellular noise), had been presented elsewhere (Serra et al. 2010; Villani et al. 2011; Villani and Serra 2013). It is likely that the mechanism described here might flank the previous one in giving rise to the complex phenomena of differentiation. In a previous work conducted by us Braccini et al. (2019), we began to investigate the effects of a simplified model of methylation in the dynamics of Boolean networks. Boolean networks (BNs) are a well-known model of biological genetic regulatory network introduced by Kauffman (1969), frequently employed for the investigation of the causes of the rich dynamics arising in complex systems, such as biological cells. Indeed, despite their simplifications, they proved to be suitable systems to represent the dynamics of biological GRNs to many level of abstractions Graudenzi et al. (2011), Serra et al. (2006), Serra et al. (2007), Shmulevich et al. (2005).

Formally, BNs are discrete-time and discrete-state dynamical systems. Following their original formulation, they can be represented by a directed graph with n nodes each having associated a Boolean variable x_i , $i = 1, \dots, n$ and a Boolean function $g_i = (x_{i_1}, \dots, x_{i_k})$ which depends on k other nodes. One among the most prominent classes of BNs is that of random BNs (RBNs), in which functions and connections are chosen according to pre-defined distributions. A special case is the one in which nodes receive exactly k distinct inputs chosen at random (avoiding self-loops) and Boolean functions are defined by choosing for each of the 2^k entries of the truth table value 1 with probability p , being p called the *bias*. RBNs exhibit a phase transition between order and chaos depending on the values of k and p Derrida and Pomeau (1986), Bastolla and Parisi (1997). For $2p(1-p)k < 1$, RBNs have on average an ordered behaviour, whilst for $2p(1-p)k > 1$ the networks show extreme sensitivity to initial conditions and very long cyclic attractors, which denote a chaotic behaviour. A *critical* regime is attained for $2p(1-p)k = 1$. Critical RBNs have proven to show properties typical of real cells Nykter et al. (2008), Roli et al. (2018), Serra et al. (2007), Serra et al. (2004), Villani et al. (2018).

The abstract methylation model to which we will refer has been introduced in our work mentioned above and it will be summarised in its main concepts in the following. It is inspired by the idea of progressive methylation of chromatin along the development and differentiation of biological

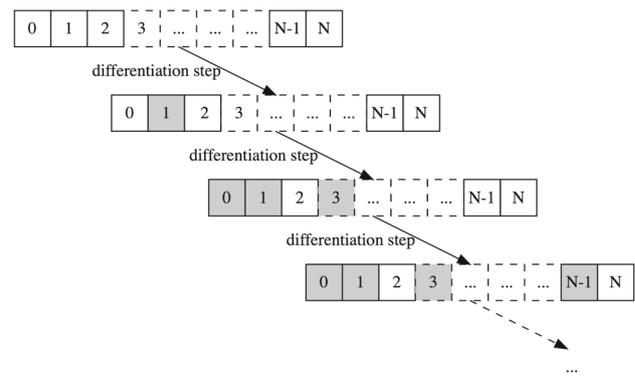


Fig. 1 Schematic representation of the methylation-inspired model. Grey nodes represent frozen nodes (blocked to value 0, regardless of the actual values of their inputs). The specific expression patterns of nodes in the ensembles that represent the state of the BN over time have no other meaning than to exemplify the increase in methylated nodes, characteristic of the methylation process introduced

cells. So, by analogy with the heterochromatin status, the methylation process in Boolean networks has been modelled by blocking the expression of some BN nodes to value 0; these nodes will be referred to as e-frozen in the following (this process is sketched in Fig. 1). This model is based on the hypothesis that—even though it is not the only phenomenon in place—the progression of frozen nodes imposes the arrow of time of the differentiation process. A differentiation model, and therefore also this methylation-based one, should accommodate, at least some, properties related to the differentiation phenomenology. We have already shown that the model in question can give rise, progressively, to reduced alternatives and an increment in stability along the differentiation process—representing cell types by means of the system's attractors Huang et al. (2005), Huang et al. (2009), Huang and Ingber (2000). Moreover, we assessed its ability to maintain diversity in terms of possible asymptotic states originating from different combinations of frozen nodes, during all the differentiation process described by the progressive freezing itself.

In this work, we provide a theoretical prediction—supported by a mean-field model—on the progression of the number of fixed genes as a consequence of perturbations in terms of externally frozen genes. The model predictions will be validated through simulations of ensembles of RBNs. This step highlights the twofold role of the model: its applicability in the determination of prediction in finite-size model's instances and its role as a means to generalise the obtained results.

Furthermore, we experimentally investigate, by employing ensembles of RBNs, the presence of a stronger version of the classical *path dependence*, i.e. the possibility of reaching different asymptotic states (cell types) by freezing the same set of nodes (genes), but in different temporal orders.

2.1 Related models

Since the seminal works by Kauffman (1969), Kauffman (1993), Boolean networks have become one among the most used models of GRNs Albert and Thakar (2014), as also proven by the availability of Boolean models in the Cell Collective repository Helikar et al. (2012), hence our first motivation for studying the effect of a methylation-like mechanism in RBNs. Besides this, BNs provide a suitable level of description for cellular dynamics, as they both make it possible to reify the notion of gene activity, and enable to apply the ensemble approach Kauffman (2004) by in silico simulations of BN models with given characteristics.

However, several variants of BNs exist and alternative or complementary models are available De Jong (2002). An overview of these models is out of the scope of this contribution; anyway, here we provide a succinct list of the ones we may suggest as a starting point for the interested reader. Apart from BNs subject to asynchronous dynamics (Harvey and Bossomaier 1997; Di Paolo 2000) and mixed dynamics Darabos et al. (2007), a prominent variant of classical BNs is that of Probabilistic BNs Shmulevich and Dougherty (2010), which extend the original model by allowing for multiple functions per node, each associated to a probability of being chosen for the update. Multi-valued logic networks have also been proposed Thomas et al. (1995). Nonlinearity can be still preserved also in models including continuous variables, as in piecewise linear differential equations (Glass and Pasternack 1978; Kappler et al. 2003; Roli et al. 2010). We conclude this summary by mentioning a couple of recent BN extensions that allow for formal verification of some dynamical properties by means of modal logics. The first is that of Reactive BNs Figueiredo and Barbosa (2018), which introduces the notion of reactive frames Gabbay and Marcelino (2009a) into BNs. The second one builds upon Abstract BNs Yordanov et al. (2016)—whereby update functions might be partially known—and provides a model checking tool for the verification of network dynamical properties Goldfeder and Kugler (2018).

3 A mean-field description of the spreading of freezing

The external freezing of some nodes can modify the dynamics of the network, since the values of other nodes may be affected. Moreover, at successive time steps, more nodes will be frozen in cascade (locked at either 0 or 1 as a consequence of the external freezing), so there will be a spreading of freezing in the system. In order to understand the main features of this process a simple, mean-field model can be studied. Of course, specific networks can behave in a way very different from the average one.

Note that we will consider here only RBNs with two inputs per node, which suffice to provide information about the most relevant properties of the model. Generalisations are formally straightforward but calculations can become complicated. We will also assume that all the Boolean functions are allowed, with the same probability, so there is a symmetry between 0's and 1's. As discussed above, these assumptions imply that the network starts in a critical state.

Let us suppose to start in attractor A at time $t < 0$: There is an initial set V_i of nodes which can take different values in A , and a set F_i of fixed nodes in A . The union of these two sets is of course the whole set of nodes $V_i \cup F_i = U$. Note that here we do not care about the value of a fixed node, but we focus on the distinction between fixed and oscillating nodes. Let f_i and v_i be the fractions of fixed and variable nodes in A , with

$$f_i + v_i = 1 \quad (1)$$

At time $t = 0$ some nodes are externally blocked to 0. If they are constant nodes, the balance of fixed versus variable nodes does not change, so we will consider only those nodes which were previously variable. So at time $t = 0$ let a fraction b_0 of nodes (which were in V) be externally blocked to 0, and let B_0 be their set. At time $t = 0$, the total fraction of constant nodes jumps to $f_0 = f_i + b_0$, while the fraction of variable nodes shrinks to $v_0 = v_i - b_0$. Summarising, at time 0:

$$\begin{aligned} v_0 &= v_i - b_0 \\ f_0 &= f_i + b_0 \end{aligned} \quad (2)$$

In the evolution from $t = 0$ new nodes can become constant due to spreading of freezing. Now we consider how many nodes will be frozen at the following time step $t = 1$ (excluding from this count those which are already constant at time 0). Each node has two inputs, therefore a randomly chosen node in V can become frozen *iff* one of the following conditions apply:

- (i) its two inputs were in V and are now both in B , due to novel freezing; or
- (ii) only one input is in B due to novel freezing, while the other was already in F due to the network dynamics; or
- (iii) only one input is in B due to novel freezing, while the other one is variable, and the Boolean functions are such that the daughter node which was in V now becomes frozen (i.e., when the function of the node is *canalizing*² in that particular value assumed by the frozen parent).

² A Boolean function is called *canalizing* if the output value is determined by only one input value—e.g. in the OR function one input equal to 1 is enough to have 1 as output.

Note that it is also possible that the daughter of two nodes in V be in F , and that this might be altered by the blocking of one node at $t = 0$. The probability of this event depends quadratically upon v_i , which decreases along the freezing process. Therefore, we decided to neglect this term.

The probabilities associated to the previous situations are the following

- (i) two parents move to B : b_0^2 ;
- (ii) one moves to B , the other was in F : $2b_0f_i = 2b_0(f_0 - b_0)$. The factor 2 is due to the fact that the node originally in V can be either parent;
- (iii) one moves to B , the other was in V : $2\eta b_0 v_0 = b_0 v_0$ where the spreading constant η is the probability that a randomly chosen node becomes frozen (at time $t + 1$ etc.) if one and only one of its parent nodes becomes frozen (at time t) while the other parent node was not frozen at time t . In the case $k = 2$, all the Boolean functions allowed with equal probability it turns out that $\eta = 1/2$. Therefore $2\eta b_0 v_0 = b_0 v_0$.

Of course, in order to have a change in the fraction of fixed nodes it is necessary that the target node was a variable one (an event whose probability is v_0), so the above values must be multiplied by v_0 . Therefore, summing the contributions from (i) to (iii), the fraction of newly frozen nodes at time $t = 1$ is

$$\begin{aligned}
 b_1 &= v_0(b_0^2 + 2b_0f_0 - 2b_0^2 + b_0v_0) \\
 &= b_0v_0(1 + f_0 - b_0)
 \end{aligned}
 \tag{3}$$

where one has taken into account that $f_0 + v_0 = 1$. Summarising, at time $t = 1$

$$\begin{cases}
 b_1 = b_0v_0(1 + f_0 - b_0) \\
 v_1 = v_0 - b_1 \\
 f_1 = f_0 + b_1
 \end{cases}
 \tag{4}$$

At the following time step, the situation is basically the same, except that now the system starts from the values of f and v given by Eq. 4, and the newly frozen nodes are b_1 . Applying the same procedure one gets

$$\begin{cases}
 b_2 = b_1v_1(1 + f_1 - b_1) \\
 v_2 = v_1 - b_2 \\
 f_2 = f_1 + b_2
 \end{cases}
 \tag{5}$$

This is true in general, so

$$\begin{cases}
 b_{k+1} = b_k v_k (1 + f_k - b_k) \\
 v_{k+1} = v_k - b_{k+1} \\
 f_{k+1} = f_k + b_{k+1}
 \end{cases}
 \tag{6}$$

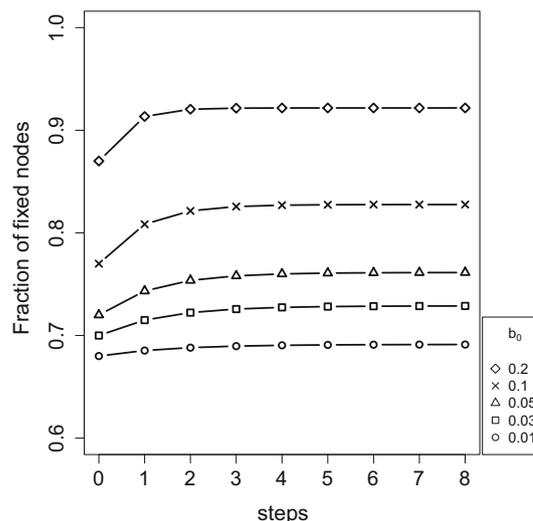


Fig. 2 Typical progression of fixed nodes (f_k) as predicted by the model. We suppose that the network has initially 67% of blocked nodes (which is the average value found experimentally for 100 nodes networks) and we externally block to 0 a fraction $b_0 \in \{0.01, 0.03, 0.05, 0.1, 0.2\}$ of nodes. We can observe that the values tends to finite asymptotic values strictly lower than 1

This formula can be applied at the following time steps. However, dealing with finite networks it is necessary to observe that nothing changes if less than one node has to be added, so the freezing necessarily halts when $b_k < 1/N$. It is interesting to note that the freezing dynamics described above tends to finite asymptotic values, as shown in Fig. 2.

4 Simulation results against model predictions

We set up the following in silico experimental setting to compare the RBNs dynamics against the model predictions concerning the spreading of freezing induced by external perturbations—in terms of nodes clamped to 0 values—to system’s attractors.

For assessing the freezing progression in RBNs, we considered different experimental conditions: ensembles of RBNs with 100 nodes in critical dynamical regime— $k = 2$, bias= 0.5—and with $b_0 \in \{1, 3, 5, 10\}$. In accordance with the methylation-inspired model previously formulated, we used a synchronous updating scheme for state update of BNs. For each parameter configuration, we drew 100 samples (RBNs) and for each of them we started from one of its attractor (let’s call it A_0), chosen at random; with reference to the mean-field model, this represents the situation at $t < 0$. Then, we clamp to value 0 a number b_0 of nodes; otherwise, if the attractor does not present a number of variable nodes greater than or equal to b_0 we replace the network with another one that satisfies this requirement. After having

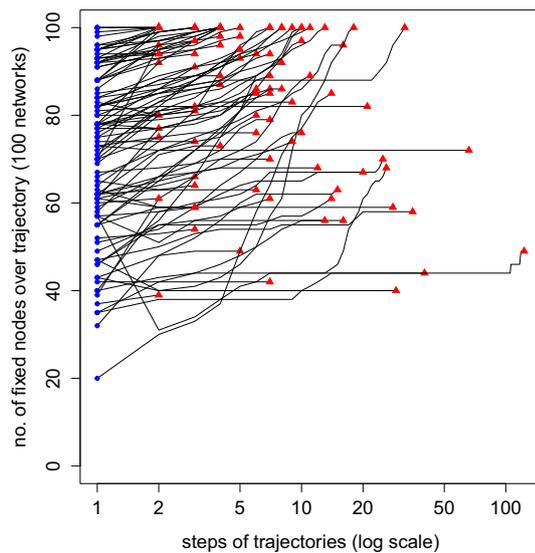


Fig. 3 Progression of fixed nodes in trajectories of 100 RBNs in the case of no. 3 externally frozen nodes. Blue dots denote the number of fixed nodes in the starting attractor ($t < 0$), while the red triangles show the same for the final attractor: the fixed nodes count includes the e-frozen nodes

perturbed the attractor A_0 ³ we follow its trajectory recording the number of nodes that are fixed on a value (regardless of whether they are 0 or 1) and which remain so until the dynamic relaxes to its new attractor. In this way, we were able to keep track of the trend of the fixed nodes along all 100 trajectories, as shown in Fig. 3 (see also Figs. 15, 16, 17 in Appendix): with the blue dots we denote the fixed nodes at $t < 0$, while with the red triangles we denote the final number of frozen nodes in the new stable dynamic condition.

The figures regarding the progression of fixed nodes show considerable variability. As predicted by the mean-field model, in each replica we observe the tendency to converge to a value smaller than the maximum, i.e. 100 nodes for these experiments. However, to perform a comparison between the experimental and theoretical results, we must take into consideration the progression of the number of the mean fixed nodes predicted by the model and compare it—step by step—with the average of the experimental values of fixed nodes. This comparison has been done for all the experimental conditions reported above. The steps reported, including standard deviation and standard error of the mean, are presented for the number of steps for which at least 30 samples have been found. The results for $b_0 = 3$ are shown in Fig. 4 (see also Figs. 12, 13, 14 in Appendix for the other comparisons). As starting condition for computing the model predictions, we have used an initial value of fixed nodes equal to the average value of the fraction of fixed nodes found in

³ We choose the last state in ascending lexicographic order, to be almost sure to perturb the attractor's state also in cyclic attractors effectively.

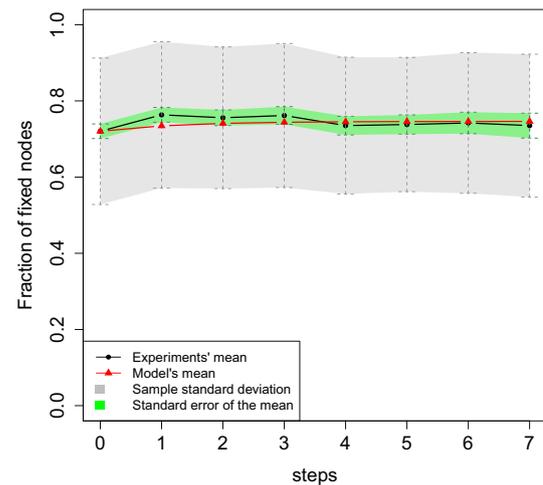


Fig. 4 Comparison between the mean-field model predictions of fixed nodes (*model's mean*) and the mean value of the fraction of fixed nodes found in the experiments (*Experiments' mean*) for the case of $b_0 = 3$. In the plot, only the steps in which there was a number of samples equal to or greater than 30 have been reported. For these steps also the sample standard deviation (grey area) and the standard error of the mean (green area) have been computed and reported

random attractors chosen as initial condition in our ensembles of RBNs at time $t < 0$ (i.e. in their wild condition). In particular, we used $f_0 \in \{0.6866, 0.6903, 0.6419, 0.6424\}$ for values of $b_0 \in \{1, 3, 5, 10\}$, respectively.

The comparison shows, in general, a high agreement between the values predicted by the model and the experimental results. On the one hand, as pointed out in the previous sections, this supports the validity of the mean-field model for the prediction of fixed nodes spreading and allows us to generalise the results obtained even for larger networks in size, which would be otherwise computationally too expensive to simulate. On the other hand, these results provide further support to the validity of this specific methylation-inspired mechanism as a driving mechanism for reproducing differentiation phenomenology in Boolean networks. Indeed, a bounded spreading of fixed nodes in response to externally frozen genes is a necessary condition for the existence of a progression of different (meta)stable activation patterns of gene expressions, and so for the existence of differentiation lineages.

5 Path dependence

The notion of *path dependence* is expressed as the property of attaining different outcomes from an initial common condition, depending on the path taken by the process Desjardins (2011). The role of path dependence in biological systems has been thoroughly discussed (see, e.g. Longo (2018), Szathmáry (2006)) and is a core ingredient in cell

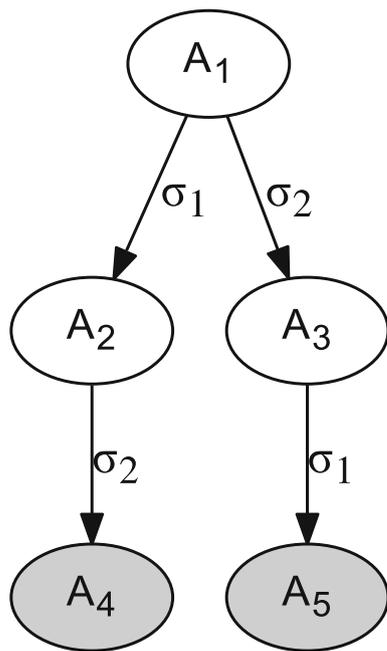


Fig. 5 A graphical representation of the specific formulation of the path dependence taken into account throughout this work. σ_1 and σ_2 labels represent different sets of freezing genes. We ascertain the presence of path dependence if the attractors A_4 and A_5 , reached by different temporal sequences of the same freezing sets of genes, are different

differentiation Huang (2009). In this work, we focus on a specific definition of path dependence that refers to the capability of reaching different attractors under different freezing sequences of the *same* genes, i.e. the attractor reached by freezing first gene σ_1 and then gene σ_2 can be different than the attractor reached by first freezing σ_2 and then σ_1 . In a sense, this is a stronger version of the classical path dependence, as it refers to the possibility of reaching a different outcome by acting on the same set of control variables, but in different temporal order. This property is particularly important because it makes it possible to manoeuvre a limited number of control variables to lead the system towards target asymptotic states.

To assess to what extent RBNs exhibit path dependence as a function of the order of gene freezing, we run experiments in which two genes (or two small groups of genes) σ_1 and σ_2 are successively frozen in both orders starting from the same state, and the final attractor reached is compared. A graphical representation of this process is depicted in Fig. 5: let us suppose to select two sets σ_1 and σ_2 of nodes to be externally frozen; starting from a state in attractor A_1 ⁴ we freeze the nodes in σ_1 (resp. σ_2) and let the network relax to the asymp-

⁴ For both the freezing sequences, we take the same state in the attractor, so as to avoid the variance caused by starting from different attractor phases. In particular, we choose the first state in ascending lexicographic order.

Table 1 Parameters of the experiments for assessing path dependence

$n = 100, k = 2, \text{bias} = 0.5$
$ \sigma_i \in \{1, 3\}, i = 1, 2$
100 randomly sampled pairs $\langle \sigma_1, \sigma_2 \rangle$
$m \in \{10, 40, 70\}$

totic condition represented by attractor A_2 (resp. A_3); once in A_2 (resp. A_3), the other group of nodes is frozen and the final attractor A_4 (resp. A_5) is stored. If the two final asymptotic states are different, then we can say that the network shows path dependence under the condition that the starting attractor is A_1 and the pair of sets to be frozen is $\langle \sigma_1, \sigma_2 \rangle$. By replicating this comparison with random samples of σ_1 and σ_2 , we can estimate the tendency of a RBN to exhibit path dependence. The comparison between two asymptotic states reached after freezing two different groups of nodes requires some care, because one cannot simply compare the two attractors *tout court*, i.e. on all the nodes, for they are anyway very likely to differ in the frozen nodes by construction (they might coincide only in the rare case in which in both the attractors A_2 and A_3 , reached by, respectively, freezing the two different groups of nodes σ_1 and σ_2 , all the nodes in σ_1 are constant at 0 in A_3 and all the nodes in σ_2 are constant at 0 in A_2). Furthermore, a comparison based on the set of nodes complementing $\sigma_1 \cup \sigma_2$ would miss an important modelling assumption: The observed cell types are a function of a limited number of coding genes, which are in turn controlled by the remaining part of the network (Borriello et al. 2018; Espinosa-Soto et al. 2004; Fumiã and Martins 2013). Therefore, we compared the asymptotic states on the basis of the values assumed by the nodes in a subset M , that we name *attractor projection* of size $m = |M|$. Referring to Fig. 5, $\text{proj}(A_4|M) = \text{proj}(A_5|M)$ if the network states of the two attractors restricted to the nodes in M are the same. The nodes to be frozen are chosen outside M and excluding those nodes always at 0 along the attractor.

The assessment of the general tendency towards path dependence of RBNs has been undertaken in different experimental conditions, summarised in Table 1. For each parameter configuration, 1000 RBNs were generated and for each of them the experiment has been replicated for 100 random samples of σ_1 and σ_2 and from a different initial number of externally frozen nodes. In this way, it is possible to ascertain the progressive tendency of exhibiting path dependence, while the number of externally frozen nodes is incremented. A representative example of the results we obtained is depicted in Fig. 6—the results of the other configurations are qualitatively the same and can be found in the Appendix. For each RBN, the fraction of pairs $\langle \sigma_1, \sigma_2 \rangle$ that led to differing asymptotic states (y-axis) is shown w.r.t. the number of initially frozen nodes (x-axis). The area below the

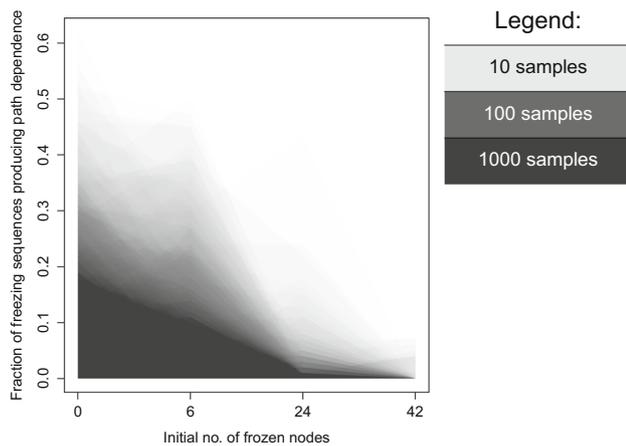


Fig. 6 Pictorial view of the results concerning path dependence for the case of **40 nodes** chosen for the projection and $|\sigma_1| = |\sigma_2| = 3$. The figures concerning the other cases are reported in Appendix. For each RBN, the fraction of pairs (σ_1, σ_2) that led to differing projected asymptotic states (y-axis) is shown w.r.t. the number of initially frozen nodes (x-axis). The area below the segmented line is coloured so as to have a pictorial view of the density of this property across RBNs: the darker the colour, the higher the fraction of RBNs with that amount of path dependence

segmented line is coloured so as to have a pictorial view of the density of this property across RBNs: the darker the colour of a point, the higher the fraction of RBNs with that amount of path dependence. We first observe the darker area, corresponding to the majority of replicas: The impact of path dependence is about the 20% in the case without initially frozen nodes and decreases linearly to a smaller yet non-negligible fraction. This trend is qualitatively the same by considering the distribution of this fraction across the RBNs; nevertheless, we observe that path dependence is still quite intense even for about one fourth of initially frozen nodes in a considerable fraction of replicas.

6 Conclusion

In the previous sections, we have shown that RBNs subject to (possibly progressive) external clamping of some nodes to a fixed value exhibit the properties of (i) keeping the avalanche of frozen nodes bounded—so that on average the asymptotic fraction of frozen nodes is less than 1—and (ii) allowing the possibility of reaching two different asymptotic states by exerting the external freezing of the same genes but in different temporal order. The first property guarantees that in general the network has still multiple fates after e-freezing, i.e. the external intervention can constrain and control the future asymptotic states but does necessarily imply a final unique asymptotic state. The second property—i.e. strong path dependence—shows that it is possible to reach different fates by simply permuting the sequence of external node

freezing. This overall outcome supports the use of external freezing in analogy with methylation mechanisms in real cells, enabling us to address some questions on cell differentiation by looking for generic properties in RBNs. For example, one may ask how many control actions are available to reach fates with given characteristics, or, conversely, whether there exist fates that can be reached by specific or fragile sequences of external freezing interventions. Future work is indeed planned to assess in more detail the controllability of RBNs under e-freezing in terms of control theory.

A question may arise as to whether the choice of externally freezing the nodes to 0 might impact the results. In RBNs with $k = 2$ and Boolean functions uniformly distributed, the average distribution of 0s and 1s is symmetrical, therefore the choice of either 0 or 1 simply breaks the symmetry towards one of the two values but the results are the same. Of course, results might differ if the RBNs considered have a different number of inputs per node and an asymmetric distribution of Boolean functions, but they are not expected to undergo qualitative changes.

In addition, we remark that our results are the outcome of models and experiments characterised by the hypothesis that networks are random, therefore one may ask to what extent these results can be valid also for real cells, which are supposed to be characterised by genes with non-random relations. First of all, some questions can be addressed in terms of *generic properties*—in the frame of the ensemble approach Kauffman (2004)—that can be investigated also in random models. Second, if some properties are generally found in random networks, even if with a mild tendency, it means that selection could easily exploit and enhance those properties during evolution. So ensembles of networks that are the result of evolutionary processes, or have different architectures than the one tested here (different degree of connectivity, scale-free, modular, etc.), or are subject to other updating mechanisms (asynchronous), or can accommodate some kind of expression noise (extrinsic and intrinsic ones), could better match the statistical features of real cells. On the other hand, the simple control mechanism we have discussed in this work may be also applied in the case of Boolean models of real gene regulatory networks. Furthermore, other models or formalisms that can be applied to the study of cellular differentiation can extend our proposed model. For this purpose, reactive graphs Gabbay and Marcelino (2009b), or more specifically reactive Boolean networks Figueiredo and Barbosa (2019), can assume the role of the control module and so represent the causal interactions that ultimately produces the progression of e-frozen nodes—currently externally controlled—that drive the differentiation process. Future work is planned in these directions.

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Pictorial view of the results concerning path dependence. For each RBN, the fraction of pairs (σ_1, σ_2) that led to differing projected asymptotic states (y-axis) is shown w.r.t. the number of initially frozen nodes (x-axis). The area below the segmented line is coloured so as to have a pictorial view of the density of this property across RBNs: the darker the colour, the higher the fraction of RBNs with that amount of path dependence (Figs. 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17).

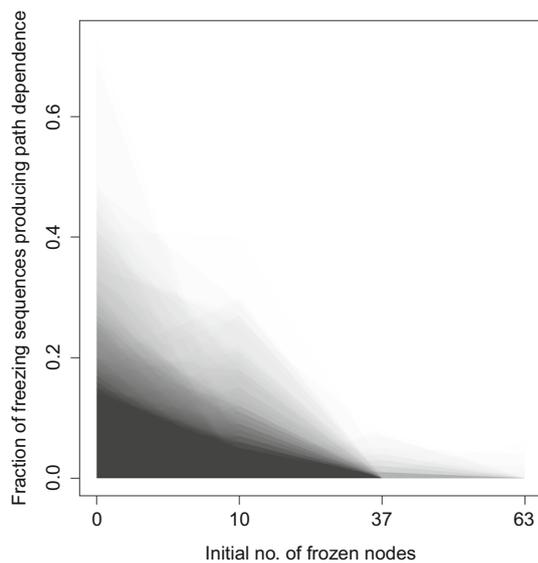


Fig. 7 10 nodes, $|\sigma_1| = |\sigma_2| = 3$

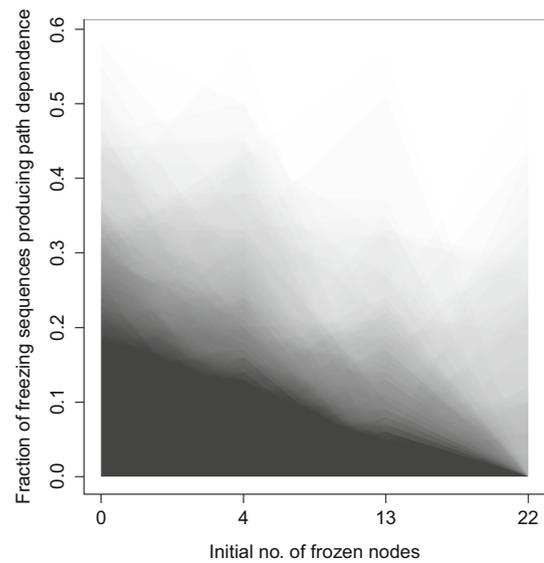


Fig. 8 70 nodes, $|\sigma_1| = |\sigma_2| = 3$

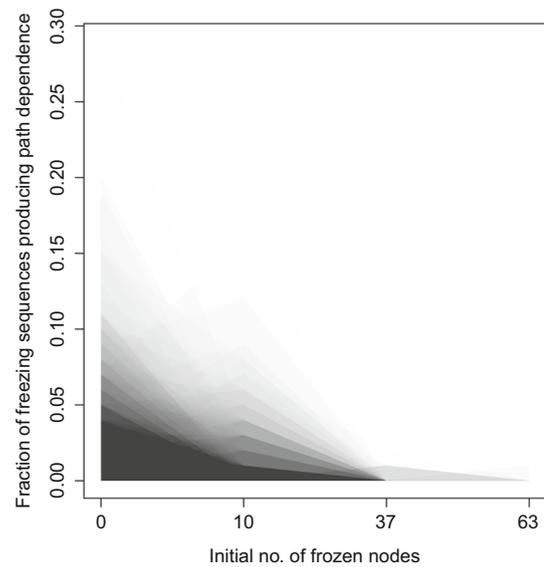


Fig. 9 10 nodes, $|\sigma_1| = |\sigma_2| = 1$

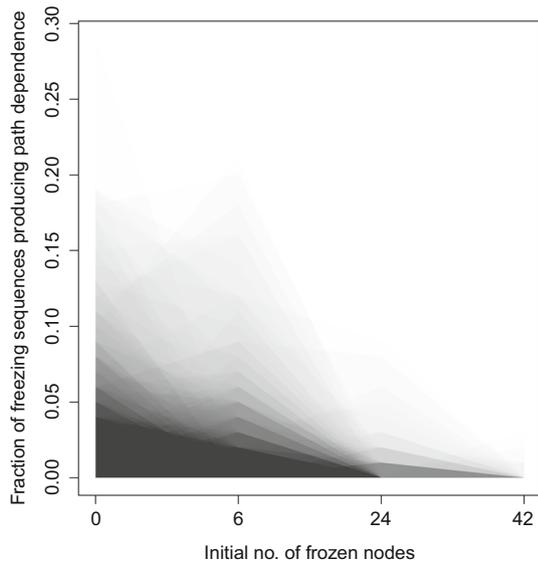


Fig. 10 40 nodes, $|\sigma_1| = |\sigma_2| = 1$

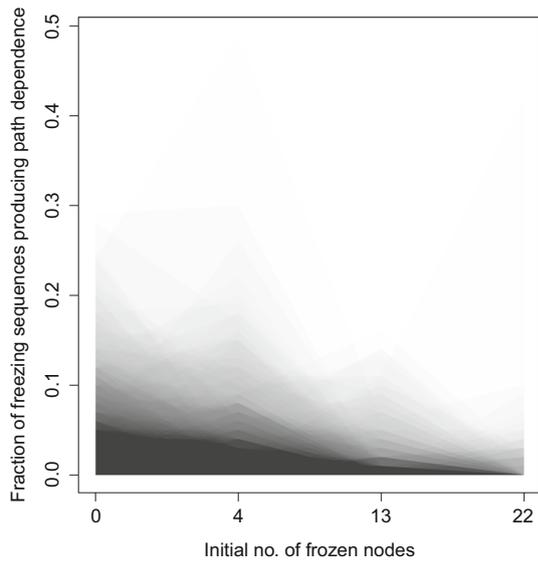


Fig. 11 70 nodes, $|\sigma_1| = |\sigma_2| = 1$

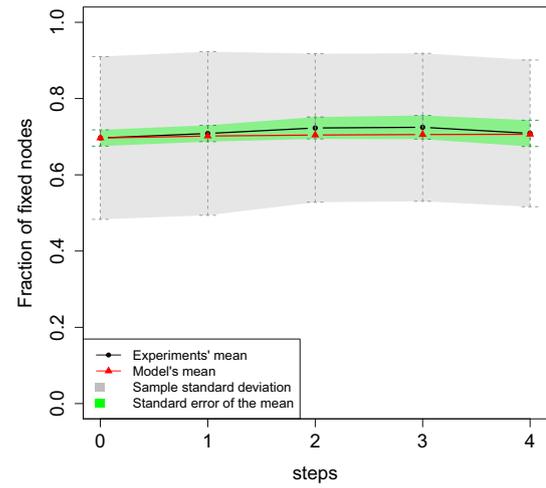


Fig. 12 Comparison between the mean-field model predictions of fixed nodes (*model's mean*) and the mean value of the fraction of fixed nodes found in the experiments (*Experiments' mean*) for the case of $\mathbf{b}_0 = \mathbf{1}$

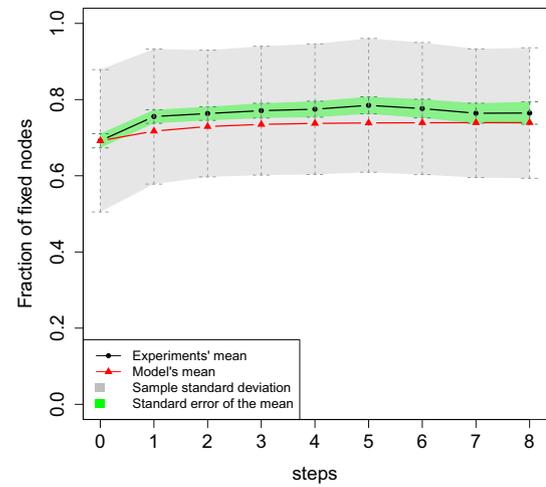


Fig. 13 Comparison between the mean-field model predictions of fixed nodes (*model's mean*) and the mean value of the fraction of fixed nodes found in the experiments (*Experiments' mean*) for the case of $\mathbf{b}_0 = \mathbf{5}$

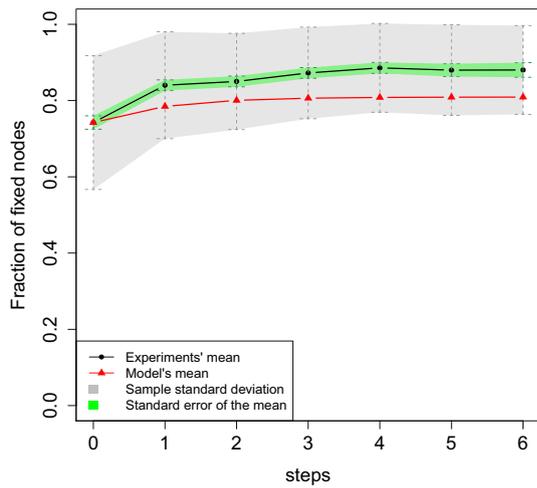


Fig. 14 Comparison between the mean-field model predictions of fixed nodes (*model's mean*) and the mean value of the fraction of fixed nodes found in the experiments (*Experiments' mean*) for the case of $b_0 = 10$

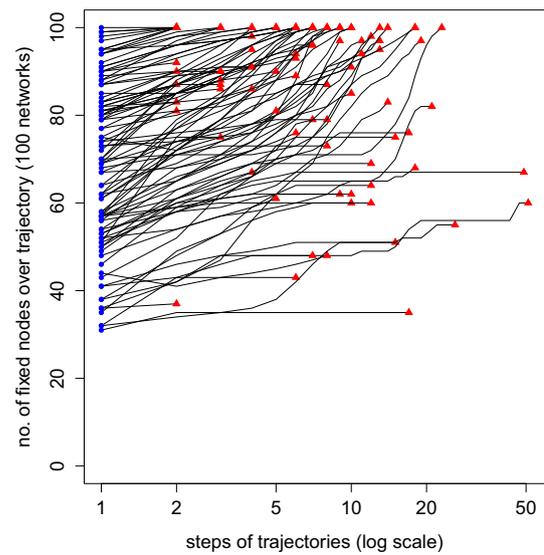


Fig. 16 Progression of the number of fixed nodes in 100 trajectories after externally freezing **5 nodes**. Blue dots indicate the number of fixed nodes in the starting attractor, while the red dots show the same for the final attractor: the fixed nodes count includes the externally frozen nodes

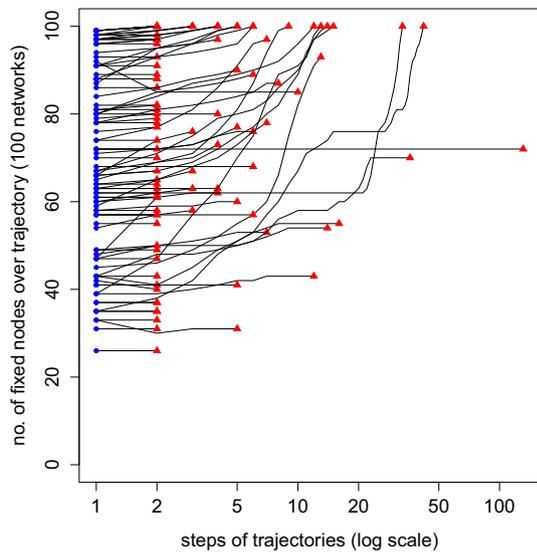


Fig. 15 Progression of the number of fixed nodes in 100 trajectories after externally freezing **1 node**. Blue dots indicate the number of fixed nodes in the starting attractor, while the red dots show the same for the final attractor: the fixed nodes count includes the externally frozen nodes

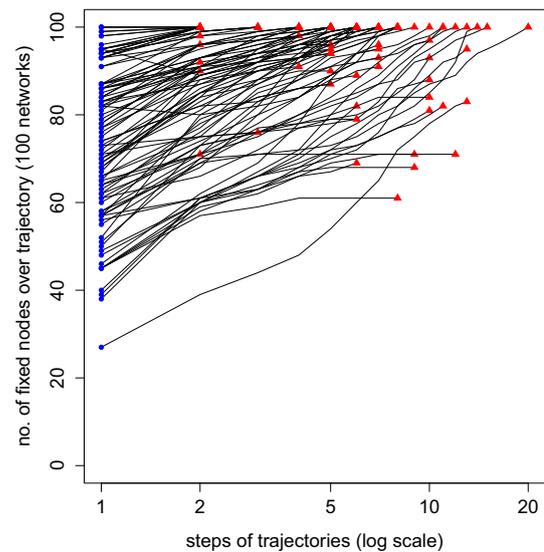


Fig. 17 Progression of the number of fixed nodes in 100 trajectories after externally freezing **10 nodes**. Blue dots indicate the number of fixed nodes in the starting attractor, while the red dots show the same for the final attractor: the fixed nodes count includes the externally frozen nodes

References

- Albert R, Thakar J (2014) Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions. *Wiley Interdiscip Rev Syst Biol Med* 6(5):353–369
- Bastolla U, Parisi G (1997) A numerical study of the critical line of Kauffman networks. *J Theor Biol* 187(1):117–133
- Borriello E, Walker SI, Laubichler MD (2018) A unified, mechanistic framework for developmental and evolutionary change. *arXiv preprint arXiv:1809.02331*
- Braccini M, Roli A, Villani M, Montagna S, Serra R (2019) A simplified model of chromatin dynamics drives differentiation process in Boolean models of GRN. In: *The 2019 conference on artificial life*, vol 31, pp 211–217. https://doi.org/10.1162/isal_a_00163. https://www.mitpressjournals.org/doi/abs/10.1162/isal_a_00163
- Bull L (2014) Evolving boolean regulatory networks with epigenetic control. *Biosystems* 116:36–42
- Darabos C, Giacobini M, Tomassini M (2007) Semi-synchronous activation in scale-free Boolean networks. In: *European conference on artificial life*. Springer, pp 976–985
- De Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 9(1):67–103
- Derrida B, Pomeau Y (1986) Random networks of automata: a simple annealed approximation. *Europhys Lett* 1:45–49
- Desjardins E (2011) Reflections on path dependence and irreversibility: lessons from evolutionary biology. *Philos Sci* 78(5):724–738
- Di Paolo E (2000) Searching for rhythms in asynchronous random Boolean networks. In: *Artificial life VII: proceedings of the seventh international conference on artificial life*, vol 7. MIT Press, p 73
- Espinosa-Soto C, Padilla-Longoria P, Alvarez-Buylla E (2004) A gene regulatory network model for cell-fate determination during arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. *Plant Cell* 16(11):2923–2939
- Figueiredo D, Barbosa L (2018) Reactive models for biological regulatory networks. In: *International symposium on molecular logic and computational synthetic biology*. Springer, pp 74–88
- Figueiredo D, Barbosa LS (2019) Reactive models for biological regulatory networks. In: Chaves M, Martins MA (eds) *Molecular logic and computational synthetic biology*. Lecture notes in computer science. Springer, Cham, pp 74–88. https://doi.org/10.1007/978-3-030-19432-1_5
- Fumiã H, Martins M (2013) Boolean network model for cancer pathways: predicting carcinogenesis and targeted therapy outcomes. *PLoS ONE* 8(7):e69008
- Gabbay D, Marcelino S (2009) Modal logics of reactive frames. *Stud Log* 93(2–3):405
- Gabbay DM, Marcelino S (2009) Modal logics of reactive frames. *Stud Log* 93(2):405. <https://doi.org/10.1007/s11225-009-9214-1>
- Gilbert SF, Barresi MJF (2016) *Developmental biology*, XI edn., chapter 3. Sinauer Associates Inc, pp 50–52
- Glass L, Pasternack J (1978) Stable oscillations in mathematical models of biological control systems. *J Math Biol* 6(3):207
- Goldfeder J, Kugler H (2018) Temporal logic based synthesis of experimentally constrained interaction networks. In: *International symposium on molecular logic and computational synthetic biology*. Springer, pp 89–104
- Graudenzi A, Serra R, Villani M, Damiani C, Colacci A, Kauffman SA (2011) Dynamical properties of a boolean model of gene regulatory network with memory. *J Comput Biol* 18(10):1291–303
- Harvey I, Bossomaier T (1997) Time out of joint: attractors in asynchronous. In: *Fourth European conference on artificial life*, vol 4. MIT Press, p 67
- Helikar T, Kowal B, McClenathan S, Bruckner M, Rowley T, Madrahimov A, Wicks B, Shrestha M, Limbu K, Rogers JA (2012) The cell collective: toward an open and collaborative approach to systems biology. *BMC Syst Biol* 6(1):96
- Huang S (2009) Non-genetic heterogeneity of cells in development: more than just noise. *Development* 136(23):3853–3862
- Huang S, Eichler G, Bar-Yam Y, Ingber D (2005) Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett* 94(128701):1–4
- Huang S, Ernberg I, Kauffman S (2009) Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. In: *Seminars in cell & developmental biology*, vol 20. Elsevier, pp 869–876
- Huang S, Ingber DE (2000) Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Exp Cell Res* 261(1):91–103
- Huang S, Kauffman S (2013) How to escape the cancer attractor: rationale and limitations of multi-target drugs. In: *Seminars in cancer biology*, vol 23. Elsevier, pp 270–278
- Kappler K, Edwards R, Glass L (2003) Dynamics in high-dimensional model gene networks. *Sig Process* 83(4):789–798
- Kauffman S (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 22(3):437–467
- Kauffman S (1993) *The origins of order: self-organization and selection in evolution*. Oxford University Press, Oxford
- Kauffman S (2004) A proposal for using the ensemble approach to understand genetic regulatory networks. *J Theor Biol* 230:581–590
- Kim M, Costello J (2017) Dna methylation: an epigenetic mark of cellular memory. *Exp Mol Med* 49(4):e322
- Longo G (2018) How future depends on past and rare events in systems of life. *Found Sci* 23(3):443–474
- Miyamoto T, Furusawa C, Kaneko K (2015) Pluripotency, differentiation, and reprogramming: a gene expression dynamics model with epigenetic feedback regulation. *PLoS Comput Biol* 11(8):e1004476
- Nykter M, Price N, Aldana M, Ramsey S, Kauffman S, Hood L, Yli-Harja O, Shmulevich I (2008) Gene expression dynamics in the macrophage exhibit criticality. In: *Proceedings of the national academy of sciences, USA*, vol 105, pp 1897–1900
- Perino M, Veenstra G (2016) Chromatin control of developmental dynamics and plasticity. *Dev Cell* 38(6):610–620
- Roli A, Vernocchi F, Serra R (2010) Continuous network models of gene expression in knock-out experiments: a preliminary study. In: *Artificial life and evolutionary computation*. World Scientific, pp 293–301
- Roli A, Villani M, Filisetti A, Serra R (2018) Dynamical criticality: overview and open questions. *J Syst Sci Complex* 31(3):647–663
- Schuettengruber B, Cavalli G (2009) Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* 136(21):3531–3542
- Serra R, Villani M, Barbieri A, Kauffman S, Colacci A (2010) On the dynamics of random Boolean networks subject to noise: attractors, ergodic sets and cell types. *J Theor Biol* 265(2):185–93
- Serra R, Villani M, Graudenzi A, Kauffman SA (2006) On the distribution of small avalanches in random Boolean networks. In: Ruusovori P et al (eds) *Proceedings of the 4th TICSP workshop on computational systems biology*. Juvenes Print, Tampere, pp 93–96
- Serra R, Villani M, Graudenzi A, Kauffman SA (2007) Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data. *J Theor Biol* 246(3):449–460
- Serra R, Villani M, Semeria A (2004) Genetic network models and statistical properties of gene expression data in knock-out experiments. *J Theor Biol* 227:149–157

- Shmulevich I, Dougherty E (2010) Probabilistic Boolean networks: the modeling and control of gene regulatory networks. *SIAM*, vol 118
- Shmulevich I, Kauffman SA, Aldana M (2005) Eukaryotic cells are dynamically ordered or critical but not chaotic. *PNAS* 102(38):13439–13444
- Szathmáry E (2006) Path dependence and historical contingency in biology. In: *Understanding change*. Springer, pp 140–157
- Thomas R, Thieffry D, Kaufman M (1995) Dynamical behaviour of biological regulatory networks —I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull Math Biol* 57(2):247–276
- Turner AP, Caves LS, Stepney S, Tyrrell AM, Lones MA (2017) Artificial epigenetic networks: automatic decomposition of dynamical control tasks using topological self-modification. *IEEE Trans Neural Netw Learn Syst* 28(1):218–230
- Turner AP, Lones MA, Fuente LA, Stepney S, Caves LS, Tyrrell AM (2013) The incorporation of epigenetics in artificial gene regulatory networks. *BioSystems* 112(2):56–62
- Villani M, Barbieri A, Serra R (2011) A dynamical model of genetic networks for cell differentiation. *PLoS ONE* 6(3):e17703
- Villani M, La Rocca L, Kauffman S, Serra R (2018) Dynamical criticality in gene regulatory networks. *Complexity* 2018
- Villani M, Serra R (2013) On the dynamical properties of a model of cell differentiation. *EURASIP J Bioinf Syst Biol* 2013(1):4. <https://doi.org/10.1186/1687-4153-2013-4>
- Yordanov B, Dunn SJ, Kugler H, Smith A, Martello G, Emmott S (2016) A method to identify and analyze biological programs through automated reasoning. *NPJ Syst Biol Appl* 2(1):1–16

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.