Life at the "edge of chaos" in a genetic model

Y. Grondin, D. J. Raine

Centre for Interdisciplinary Science, Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK iscience@le.ac.uk

and V. Norris

Assemblages Moléculaires: Modélisation et Imagerie SIMS, FRE CNRS 2829, Faculté des Sciences et Techniques de Rouen, F76821 Mont Saint Aignan, France &

Epigenomics Project, genopole®, 93 rue Henri-Rochefort, 91000 Evry, France

Genetic regulatory networks are often represented as Boolean networks and characterised by average parameters such as internal homogeneity (the probability that a node outputs a 1). Here we present a different formalism in which the nodes interact through positive and negative links with the state of the nodes determined by a single, general logical function. The main parameter of the system is in this case the proportion of negative links. At a critical value of the proportion of negative links the networks display a phase transition from a periodic to a chaotic regime The proportion of negative links in the bacterium Escherichia coli corresponds in our model to a range where the network behaves at the edge of chaos.

1 Introduction

The genetic regulatory network of a cell is a complex dynamical system yielding a wide diversity of living cells and organisms. Specific variations at any small or large scale in the level of expression of the genes, in their timing of action and in the architecture of the network are amongst the factors responsible for such diversity [1]. Boolean networks, a general form of cellular automata, have long been used to study the dynamical properties of such biological systems at both small and large scale [2,3]. For example, Boolean networks are used to represent the genetic regulatory networks inferred from mRNA data [4] and as models of cell differentiation [3,5].

Here, we use a variant of Boolean networks to model genetic regulation and to study properties related to architecture and dynamics. In the Boolean network introduced by Kauffman [3], also called the NK-model, the state of each of N nodes is represented by a Boolean variable (ON or OFF) where the output value is determined by a Boolean function, or transition function, that has for input the K connections to that node. The model is characterised by parameters such as the internal homogeneity, p, that is the probability that an output is ON [3,6], and the size of the stable core given by the number of stable nodes, that is the nodes that have a constant state independent of the initial state [7]. This model displays a complex dynamics. A major result of the NK-model [3] is the phase transition occurring at K = 2 between a crystallised phase (K < 2) and a chaotic phase (K > 2). For K = 2, the network is said to operate at the "edge of chaos" [3].

From the point of view of genetic regulation, the traditional representation of Boolean networks does not capture adequately the regulatory mechanism of the genes. We choose a different representation by assigning to the links either a positive or a negative regulatory effect in a way similar to the activating or inhibiting effect of regulatory genes on those that they regulate. Although this could be modelled in the framework of classical Boolean networks, our results differ in many aspects and, allow comparisons to real biological data.

We first introduce the genetic regulatory model and then show that a critical regime occurs for two ranges of the proportion of inhibitor links, μ . Further analysis of this critical regime shows that the network behaves at the edge of chaos. Though this transition is usually observed for variation of the internal homogeneity [6,8], our study shows that μ and p are different parameters. Furthermore, we find that the fraction of negative links in this critical regime corresponds to that found in the transcriptional network of the bacterium Escherichia coli, suggesting that it too behaves at the edge of chaos.

2 The Model

2.1 Architecture of the model

We consider directed networks where the agents, or nodes, represent the cellular machinery of gene regulation and the links represent the regulating influence of the agents on each other. The links, which are fixed, can either have an activating or inhibiting effect on the nodes to which they are connected. A network is represented by its adjacency matrix A, with elements aij given by

 $a_{ij} = \begin{cases} 0, \text{ if there is no link from the node } j \text{ to the node } i; \\ 1, \text{ if node } j \text{ is connected and directed to node } i \text{ and acts as an inducer on } i; \end{cases}$

-1, if node *j* is connected and directed to node *i* and acts as a repressor on *i*.

We denote by μ the proportion of repressors in the network, that is the number of negative links in the matrix *A* as a fraction of the number of non-zero links. The mean connectivity of the network is $k = \langle \Sigma a i j \rangle$.

2.2 Dynamics

As for Boolean networks, each agent of the network is characterised by its binary state. The configuration of the network at any one time is given by a vector S(t), where the element $s_i(t)$ is the state of agent *i* at time *t*, such that $s_i(t) = 0$ if the agent is OFF and $s_i(t) = 1$ otherwise.

The dynamics of the network is provided by a simple rule in which the state of the nodes at a given time step depends only on the configuration of the network at the previous time. This rule states that a node is ON if the number of active positive incoming links to the node is greater than the number of active negative ones. Furthermore, only the nodes that are ON can exert their control over the other nodes: that is a node that is OFF does not exert any control on other nodes whatever its outgoing links may be. This translates to the following expression in which a node *i* is ON at t + 1 if

$$\sum_{j} a_{ij} s_j(t) > 0 \tag{1}$$

and otherwise the node is OFF. A consequence is that if all of the nodes connecting a node i are OFF, this node does not receive signals; this node is OFF by default. The nodes that remain OFF in this way during the simulation belong to the inactive core of the network. The other nodes are part of the active core.

Although the activation function in equation (1) could be expressed as a Boolean function, this would neither be the most appropriate approach, nor the simplest. The function given in (1) is a more general function than the standard activation-inhibition functions, where a node is activated if there is at least an activator but no inhibitor and not activated otherwise. Indeed, expression (1) carries a simple symmetrical cooperative effect between activators and between inhibitors: the more

activator the more likely a node will be activated and, conversely, with the more inhibitors the more likely a node will be inhibited. Finally, observations suggest that the binding of a regulator to specific DNA sequences may not be binary, that is designed to bind or not. On the contrary, there are apparently thresholds conditioned, for example, by the specificity of the sequence or the concentration of the regulator [13]. Nevertheless for simplicity in this paper we set the threshold to zero.

2.3 External input

We consider that the networks are not autonomous, that is they cannot exhibit a dynamics in the absence of a constraining environment. A subset of the nodes is chosen to receive an external input, or signal, which are subsequently considered as input nodes. There are several possibilities for the choice of those nodes: the nodes without incoming links, a subset of those, a subset of any nodes, etc. However, for simplicity, the subset of input nodes is chosen at random regardless of the connectivity of the nodes. The chosen nodes remain ON at any time regardless the value of equation (1), which provides a clamping effect on the network.

3 Result

In the following, we consider two different network architectures: a random network in which links are formed between pairs of nodes at random [10] and a power-law network [14]. These networks differ in many aspects such as in their degrees of clustering and diameter [10]. The networks presented below are constructed with a number of nodes and a mean connectivity of the order of magnitude that is observed in typical bacterial models, that is of the order of 1000 nodes and a small mean connectivity, in view of the sparseness of the regulatory networks in cellular organisms [15].

At the start of a simulation, all the nodes of the network are set to be OFF. A number of nodes, set to I = 50 in the following, is selected at random with equal probability, to receive an external input. The state of each node is then repeatedly updated until either an attractor is found, which occurs after a transient phase, or until a maximum number of set time steps, L, is reached.

3.1 Phases

For μ varying between 0 and 1, simulations show that for random networks with N = 1500 nodes and k = 8.0 the model displays three different behaviours characterised by the nature of the attractor. The network is (i) crystallised if the attractor is a fixed point, (ii) periodic if the attractor cycles over a length of time < L and (iii) considered to be chaotic if no period is found, that is the network has a period > L. This relates to the fact that for large scale networks a period may not be reached in a reasonable computational time, despite the fact that in a finite deterministic model a period must exist. The three observed behaviours are characteristic of classical Boolean networks and cellular automata [3,6,16].



Figure 1. Fraction of a given attractor according to the fraction of negative links. The curves represent the fraction of static attractors (cross), periodic attractors (open circle) and chaotic attractors (open square) according to the proportion of negative links. Each point gives the fraction of network of a given attractor over 500 repeats. For each repeat, a random network of N = 1500, k = 8.0 is generated and a random subset of 50 nodes chosen as external inputs. Each simulation runs for a maximum of 105 time steps beyond which an attractor is considered chaotic.



Figure 2 (left): Illustration of local and propagating structures. In both cases, the nodes A and C compose the neighbourhood of the node B. (a) The oscillation of the state of node a spreads locally to its neighbourhood whilst in (b) the oscillation spreads outside the neighbourhood.

Figure 3 (right): Transition in a single network for variation of the proportion of negative links. A network of 1500 nodes is initially constructed with $\mu = 0.3$. The proportion of negative links is then increased and the number of nodes of variable states and the number of cyclic patterns formed from the nodes of variable state at the steady state is recorded. The variation of the number of cyclic motifs is given in (a) and the number of nodes of variable state in (b).

As shown in Fig. 1 for $L = 10^5$, the probability of finding a particular attractor depends on the value of μ . Thus, for $\mu < 0.27$ and $\mu > 0.78$, the network is likely to have a have a fixed point attractor while for $0.27 < \mu < 0.38$ and $0.65 < \mu < 0.78$ the network is more likely to reach a periodic attractor with a peak at $\mu \sim 0.35$ and $\mu \sim 0.70$. Finally, for $0.38 < \mu < 0.65$ the network is most likely to be in the chaotic regime with a symmetry in the distribution at $\mu \sim 0.53$.

Networks with a power-law distribution of the degree of connectivity show identical results (data not shown), which suggests that the architecture does not influence the behaviour of the model in this respect.

Further information is required to determine whether the periodic behaviour of the network has properties similar to that of cellular automata at the edge of chaos. In cellular automata, the probability for a node to be ON depends on a parameter λ , that is identical to the internal homogeneity of Boolean networks [16]. For λ close to 0 the system presents no activity after a very short number of steps. For λ around 0.2 some oscillatory states will persist either locally or propagating through the system in what are defined either as local or as propagating structures, respectively [16]. For λ around 0.3 those structures start to interact in complex patterns and when λ reaches 0.5 the system has become chaotic [16]. Here, a local structure is characterised by a periodic pattern confined to its neighbourhood (a node and the nodes it is directly connected to), while a propagating structure is a pattern that travels across neighbourhoods. In the network model, these structures are characterised by specific motifs. For example, the oscillatory state of node B generated by the structural motif in Fig.2(a) can spread only to its neighbouring node C, and similarly, the oscillatory state of node C can spread only to its neighbouring node B. The oscillatory pattern formed by the variation of the states of the nodes B and C is therefore local. Introducing a series of nodes between the nodes B and C, as shown in Fig. 2(b), allows the oscillatory states to propagate beyond its neighbourhood. The oscillatory pattern in the neighbourhood of node B is then a propagating structure.

To test whether the network behaves at the edge of chaos we need to look first at the number of local and propagating structures observed in a given network as the proportion of negative links increases [16]. This is equivalent to looking at the number of cyclic motifs in the part of the network formed by only the nodes of variable state. In addition, we need to check whether these structures are interacting with each other, that is, whether these structures can be connected by paths between the nodes of variable state. The number of cyclic motifs for a network of 1500 nodes and k = 8.0 is shown in Fig. 3(a) for a given network, and a given set of input nodes, as μ increases from 0.3 to 0.46. The number of cyclic motifs is measured by investigating the network made up of the nodes of variable state. First, we list all the nodes that are connected to a chosen starting node. Those nodes become starting nodes and the nodes connected to them are subsequently included in the list. This is repeated until no new node is added to the list. We then extract the nodes that appear at least twice in the list as they are either part of parallel pathways, as in a feed-

forward loop for example [17], or they belong to a cycle, or both. Each of the selected nodes is then considered as the potential start of a cyclic pathway with the condition that once any such starting node has been identified as part of a cycle it cannot be part of any other cycles. This tends to under-estimate the number of cycles although it does not impair the result as shown below.

We also show in Fig. 3(b) the number of nodes of variable state forming the network. Thus for a small proportion of negative links ($\mu \sim 0.3$), where the network is likely to be crystallised, the number of cyclic motif is small (Fig. 4(a)). As the proportion of negative links increases to about $\mu \sim 0.39$, the number of cycles rises by one order of magnitude. Beyond this value where the network is likely to be chaotic (Fig. 1), the number of cyclic motifs varies as *N*, despite the under-estimation of the number of cyclic patterns. Note that all these structures are propagating: the probability of having a simple local structure such that of Fig. 2(a) is proportional to $k N^{-2}$, hence close to zero in a network of small connectivity. Figure 3(b) shows also that the size of the network made of nodes of variable states varies similarly to the number of cyclic motifs. Finally, the measured number of independent components is less than 3, meaning that the propagating structures are interacting with each other. This demonstrates that for a range of μ , the network behaves at the edge of chaos.

4. Discussion

Classically, the behaviour of Boolean networks is affected by the bias introduced by the internal homogeneity parameter, p [3,6]. In the present model, the behaviour of the network is similarly affected by the proportion of negative links, μ . However, the parameters p and μ are noticeably different, this for the two following reasons. First, the value of p calculated from expression (1) for given values of μ does not equal μ . For example, in a structural motif where one incoming link is positive and another one is negative, that is $\mu = 0.5$, the probability for a node to be ON according to (1) is p = 0.25. Second, in the classical case, each value of p corresponds to a set of Boolean functions, whereas in our model each value of μ corresponds to a specific pcalculated according to (1), but also to a Boolean function. Expression (1) determines the only possible Boolean function for a given value of μ and a given number of incoming links, ensuring that the logic of the transition functions is constant over the range of μ . This has the modelling advantage that μ together with (1) provides a control parameter over the internal homogeneity of the system.

The rather different formalism adopted in the representation of the interactions between genes allows us to make comparisons with real genetic regulatory networks. For instance in RegulonDB, the transcriptional network of the bacterium Escherichia coli is described in term of the activating, inhibiting or dual function of regulators on the genes they regulate [18]. Considering the dual effect as a neutral one, the proportion of negative links to that of the total number of links, excluding those with a dual effect, gives $\mu = 0.4$. This corresponds in our model to a range at which the network is almost as likely to be periodic as to be chaotic, that is, it operates at the

edge of chaos. This suggests that, similarly to the model, the transcriptional network of E. coli operates at the edge of chaos. Furthermore, the value of μ at which the networks are more likely to be periodic spans a small range, meaning that the value of μ for real systems may not be unique. However, because of the need for both robustness and adaptability we do expect a fine tuning of this value. Indeed, a small increase in μ and the networks may become too sensitive to perturbations, while conversely a small decrease in μ and the networks may become unresponsive to a change of vital signals. Note that such direct comparison would not be possible using the internal homogeneity as the key parameter.

Bibliography

- [1] Adam S. Wilkins, The evolution of developmental pathways. (Sinauer Associates, 2002).
- [2] D. Thieffry and D. Romero, Biosystems 50 (1), 49 (1999).
- [3] S. A. Kauffman, The origins of order: Self-Organization and Selection in Evolution. (Oxford University Press, Oxford, 1993).
- [4] S. Liang, S. Fuhrman, and R. Somogyi, Pac. Symp. Biocomput., 18 (1998); T. Akutsu, S. Miyano, and S. Kuhara, Bioinformatics 16 (8), 727 (2000); S. Martin, Z. Zhang, A. Martino et al., Bioinformatics (2007).
- [5] S. Huang, J. Mol. Med. 77 (6), 469 (1999).
- [6] G. Weisbuch and D. Stauffer, J. Physique 48 (1), 11 (1987).
- [7] S. Bilke and F. Sjunnesson, Phys. Rev. E. 65, 016129 (2002).
- [8] L. de Arcangelis and D. Stauffer, J. Physique 48 (11), 1881 (1987).
- [9] R. Albert and A.-L. Barabási, Phys. Rev. Lett. 84 (24), 5660 (2000).
- [10] R. Albert and A.-L. Barabasi, Rev. of Mod. Phys. 74 (1), 47 (2002).
- [11] D. J. Watts and S. H. Strogatz, Nature 393 (6684), 440 (1998).
- [12] B. Yuan, K. Chen, and B-H. Bang, cond-mat/0408391 (v1) (2004).
- [13] U. Gerland, J. D. Moroz, and T. Hwa, Proc. Natl. Acad. Sci. U S A 99 (19), 12015 (2002).
- [14] A.-L. Barabasi and R. Albert, Science 286 (5439), 509 (1999).
- [15] N. Guelzim, S. Bottani, P. Bourgine et al., Nat. Genet. 31 (1), 60 (2002).
- [16] C. G. Langton, Phys. D 2 (1-3), 120 (1986).
- [17] R. Milo, S. Shen-Orr, S. Itzkovitz et al., Science 298 (5594), 824 (2002).
- [18] H. Salgado, S. Gama-Castro, A. Martinez-Antonio et al., Nucleic Acids Res. 32 (Database issue), D303 (2004).

This work was performed using the University of Leicester Mathematical Modelling Centre's Supercomputer which was purchased through the EPSRC strategic equipment initiative