2 January 1995

PHYSICS LETTERS A

ELSEVIER

Physics Letters A 196 (1995) 331–334

Phase transitions and antichaos in generalized Kauffman networks

Ricard V. Solé, Bartolo Luque

Complex Systems Research Group, Department de Fisica i Enginyeria Nuclear, Universitat Politécnica de Catalunya, Sor Eulàlia d'Anzizu s/n. Campus Nord, Mòdul B5, 08034 Barcelona, Spain

Received 11 July 1994; revised manuscript received 28 September 1994; accepted for publication 24 October 1994
Communicated by A.P. Fordy

Abstract

The critical properties of Kauffman networks with different input connectivities are studied. A general condition for stability is derived and used as a constraint (the antichaos constraint) in the calculation of the probability distribution of connections in a stable network. Some consequences for genome stability are outlined.

Random Boolean networks (RBN) [1–3] were proposed by Kauffman as simple (yet reasonable) models of genetic systems. In these networks, a set of N binary elements is used. The state of each element at a given time step t, is given by $S_i(t)\in\{0,1\}$ $(i=1, ..., N)$. The dynamical state of each $S_i(t)$ is updated (synchronously) by means of a Boolean function Λ_i. Each element receives inputs from exactly K elements, and we have a dynamical system defined from

$$S_i(t+1)=\Lambda_i[S_{i_1}(t), S_{i_2}(t), ..., S_{i_K}(t)] . \quad (1)$$

Using this theoretical framework, several methods of statistical mechanics can be applied [4–7]. A very important result of these studies was the existence of a phase transition in the dynamical behaviour of RBN [4]. Using K as a key parameter, it was shown that for $K_c=2$ a phase transition occurs separating the so-called chaotic ($K>K_c$) and frozen ($K<K_c$) phases. Here by "chaos' we do not mean low-dimensional deterministic chaos but a phase where damage spreading takes place (i.e. the appearance of changes in global dynamical behaviour caused by transiently altering the activity of a single binary variable). At K_c, nets crystallize spontaneous order, and several properties involving the number of attractors, cycle lengths or the stability against minimal and structural perturbations seem to be in agreement with what one can expect about the real genome [1,8,9].

The critical point K_c was analytically determined by Derrida and his colleages [4–6] in a set of remarkable theoretical studies. Using the so-called annealed approximation, they analysed the evolution in the overlaps of two randomly chosen RBN of a given connectivity K. The connections and the Boolean functions are chosen at random among the elements of the set $\mathcal{F}_k(N)$ of all the Boolean functions of connectivity K. Using two configurations, say

$$C_1(t)\equiv(S_1^{(1)}(t), ..., S_N^{(1)}(t)) ,$$

$$C_2(t)\equiv(S_1^{(2)}(t), ..., S_N^{(2)}(t)) ,$$

which are also randomly taken from the set $\mathscr{C}(N)$ of all the possible N-strings (clearly $\#\mathscr{C}(N)=2^N$), we have $Na_{12}(t)$ common spins; let $Na_{12}(t+1)$ be the net overlap after one iteration of $C_1(t)$ and $C_2(t)$ under (1). Then a new set of connections and Boolean functions is again taken from $\mathcal{F}_K(N)$, and a new iteration is performed over $C_1(t+1)$ and $C_2(t+1)$. A

0375-9601/95/$09.50
SSDI 0375-9601(94)00876-0

general recurrence relation can be derived [4,5] leading to

$$a_{12}(t+1)=\tfrac{1}{2}\{1+[a_{12}(t)]^K\}, \quad (2)$$

which can be transformed (by means of the Hamming distance) into the following one-dimensional map,

$$D_{t+1}=F_K(D_t)=\tfrac{1}{2}[1-(1-D_t)^K], \quad (3)$$

which gives us a measure of how two initially close configurations $\{C_1(t), C_2(t)\}$ separate in time. In the frozen phase, the fixed point $D^*=0$ is stable (i.e. two given similar configurations converge with time and we have $\partial_D F_K(D^*)<1$) and in the chaotic one D^* is unstable and two initially close conditions diverge to a finite distance.

The previous results are valid for a particular situation: a theoretical problem where the number K is fixed to a given integer value. But in a more realistic situation, say in the genome, the connectivity among elements is clearly different. Some genes are control elements and as a consequence are highly connected. Some others are always necessary (in all cell types) and are weakly controlled.

Following the previous procedure, we can ask which type of condition over the distances is obtained if an arbitrary set of inputs is introduced. More specifically, let us now suppose that each S_i receives $K_i \in \{1, 2, ..., \#K\}$ inputs. In this case, the dynamics will be defined through the evolution equations

$$S_i(t+1)=A_i[S_{i_1}(t), S_{i_2}(t), ..., S_{i_{K_i}}(t)]. \quad (4)$$

Now we use again two randomly chosen initial configurations $C_i(t)\in \mathscr{C}(N)$ and two Boolean functions, now chosen from the new set,

$$S(N, \#K)=\bigcup_{K_i=1}^{\#K} \mathscr{F}_{K_i}(N),$$

$\#K\in\mathbb{Z}$ being the maximum connectivity ($\#K\approx 10$ for the genome [8,10]).

In this Letter we want to address two questions involving this approximation: (1) which critical condition is obtained separating the two zones in phase space and (2) at the critical point, which kind of probability distribution $f(K_i)$ over the input connectivities is expected to appear.

Following Derrida and Poneau's [5] it can be shown that for these generalized Kauffman networks (GKN), the overlap evolves as

$$a_{12}(t+1)=\frac{1}{2}\left(1+\sum_{K_i=1}^{\#K} f(K_i)a_{12}(t)^{K_i}\right), \quad (5)$$

or, in terms of D_t,

$$D_{t+1}=G(D_t)=\frac{1}{2}\left(1-\sum_{K_i=1}^{\#K} f(K_i)(1-D_t)^{K_i}\right). \quad (6)$$

This equation can be further generalized by introducing the bias p in the sampling of Boolean functions. Here p is the probability of a function $A_i\in\mathscr{S}(N, \#K)$ to yield a 1 for any input configuration. Using the bias, we obtain

$$D_{t+1}=4p(1-p)G(D_t). \quad (7)$$

Both equations (6) and (7) were first introduced in Ref. [4]. In order to analyse the stability properties of (7), we study the marginal stability of the fixed point D^*. This condition (which defines the phase transition point) gives

$$\left|4p(1-p)\frac{\partial G(D^*)}{\partial D^*}\right|=1 \quad (8)$$

and this leads to

$$2p(1-p)\sum_{K_i=1}^{\#K} f(K_i)K_i=1, \quad (9)$$

which gives us a relationship between the mean connectivity $\langle K\rangle$ and the bias,

$$\langle K\rangle=\frac{1}{2p(1-p)}. \quad (10)$$

This critical condition leads to $K_c=2$ when $p=0.5$, as in standard RBN. Now we can see that order is guaranteed even for high mean connectivities as long as p becomes small enough. For a genomic regulation system at the transition point, this means that as regulatory circuits increase in complexity (and so $\langle K\rangle$ grows) the bias in the Boolean functions is reduced, following the critical line.

The critical condition (10) can be used in order to obtain the probability distribution $F(K_i)$ of input connectivities by means of a variational principle. As it is well known from the studies of Jaynes [11] and Haken [12], unbiased estimates of $f(K_i)$ can be obtained in agreement with all the possible (macroscopic) knowledge available about the system. The maximum entropy formalism (Maxent) has been mainly applied to equilibrium systems, but extensions to nonequilibrium systems as lasers [12] or energy flow networks in ecosystems [13] have also been obtained.

The standard procedure starts from the Boltzmann entropy, defined as

$$S(\{f(K_i)\}) = -\sum_{K_i=1}^{\#K} f(K_i)\log[f(K_i)]\,. \tag{11}$$

The problem is to find the extremum of (11) under an adequate set of constraints, by using the method of Lagrange multipliers. One constraint is obviously the normalization condition,

$$\sum_{K_i=1}^{\#K} f(K_i) = 1\,, \tag{12}$$

and the second one will be the critical condition defining the phase transition point,

$$\sum_{K_i=1}^{\#K} f(K_i)K_i = \frac{1}{2p(1-p)}\,. \tag{13}$$

Let us justify our choice. In recent theoretical studies, Kauffman and other authors [14–16] have conjectured that complex behavior in nature will appear on or near phase transition points. For RBN at such a critical point (K_c) a compromise between "stability" and "evolvability" [9] is reached. Here by stable we mean statistic stability: Boolean nets are stable, for $K=K_c$, against minimal perturbations. Such a stability is necessary in order to guarantee robustness, as we expect for a biological system as the genome. It was shown [9] that for K_c also the conditions for evolvability were optimum. And so the constraint (13) plays an important role. Following Kauffman [16] we call it the "antichaos constraint". By antichaos we indicate the spontaneous emergence of the previous properties close to the instability point.

On the other hand, the number of genes in an evolving genome grows by random duplications of other genes [10,17] as well as as a consequence of other processes where retroviruses and other moveable genetic elements [10] are involved. Random introduction of new elements takes place together with a change in the distribution of connectivities $f(K_i)$. The macroscopic constraint operates over the new set, and $f(K_i)$ is adequately changed. The random character of this process makes possible the application of the Maxent formalism.

Using (12) as a macroscopic constraint, we can solve the variational equation:

$$\delta\left(S(\{f(K_i)\}) - (\lambda-1)\sum_{K_i=1}^{\#K} f(K_i) - \beta\sum_{K_i=1}^{\#K} f(K_i)K_i\right) = 0\,, \tag{14}$$

which leads to the canonical distribution [12],

$$f(K_i) = \exp(-\lambda-\beta K_i)\,. \tag{15}$$

Here $\{\lambda,\beta\}$ are the Lagrange multipliers, which can be calculated from the constraints (12) and (13). So we get

$$f(K_i) = \frac{1}{Z}\exp(-K_i/\langle K\rangle)\,,$$

or, more specifically,

$$f(K_i) = \frac{1}{Z}\exp[-2p(1-p)K_i]\,, \tag{16}$$

where as usual Z is the partition function,

$$Z = \sum_{K_i=1}^{\#K} \exp[-2p(1-p)K_i]\,. \tag{17}$$

Eq. (16) gives us a general condition for the expected distribution of input connectivities. Most of

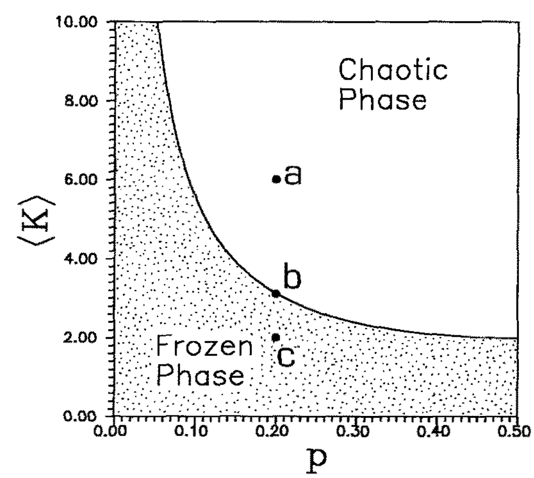

Fig. 1. Phase space of the GKN, obtained from Eq. (10). Two phases are indicated. Three particular points in this phase space are indicated. Their dynamics is shown in Fig. 2.

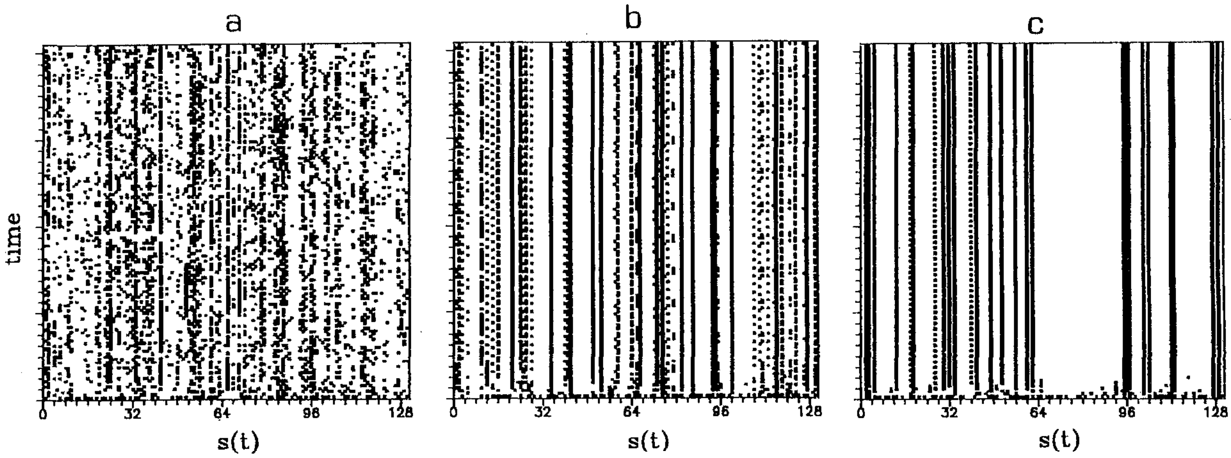

Fig. 2. Dynamics of a $N=130$ GKN in three different points of the phase space (as indicated in Fig. 1). The Boolean functions are randomly chosen from $\mathscr{S}(130, 10)$ following the distribution given by (16). Here $p=0.2$ and (a) $\langle K\rangle=6$, (b) $\langle K\rangle=3.1$ and (c) $\langle K\rangle=2$.

the elements receive a small number of inputs and a small fraction of them is controlled by a large number of input elements. Using this probability distribution, we can simulate our system in different parts of the parameter space. In Fig. 2 we can see three examples of (a) frozen ($\langle K\rangle=2, p=0.2$), (b) critical ($\langle K\rangle=3.1$, $p=0.2$) and (c) chaotic ($\langle K\rangle=6$, $p=0.2$) dynamics. Here a $N=130$ network was used. The points in the parameter space are indicated in Fig. 1. As we can see, the observed dynamics agrees with the typical situations obtained in standard RBN.

This result, if true, can give us some information concerning the problem how many genes are influenced by K_i other genes. It seems clear from studies in molecular genetics that a more or less small set of genes (oncogenes or genes involved in developmental processes) is linked with a large number of elements [8,10,17]. Many genes, on the other hand, are weakly linked with other parts of the genome. Our approach can be useful in future studies in order to give theoretical estimates of which frequencies can be expected. On the other hand, the phase space shown in Fig. 1 provides us with an image of what we can expect in relation with the bias of the random Boolean functions and the degree of connectivity $\langle K\rangle$ which is, in fact, a measure of how strongly interconnected is the genome and so how complex is the regulation involved.

The general biological consequences, as the application to neural networks [18] or the immune system (where the connectivity and size of the network have been shown to be an emergent property [19]) will be analysed in a further paper.

References

[1] S.A. Kauffman, J. Theor. Biol. 22 (1969) 437.
[2] S.A. Kauffman, J. Theor. Biol. 44 (1974) 167.
[3] S.A. Kauffman, Physica D 10 (1984) 145.
[4] B. Derrida and Y. Pomeau, Europhys. Lett. 1 (1986) 45.
[5] B. Derrida and Y. Pomeau, Europhys. Lett. 2 (1986) 739.
[6] B. Derrida and G. Weisbuch, J. Phys. (Paris) 47 (1986) 1297.
[7] G. Weisbuch and D. Stauffer, J. Phys. (Paris) 48 (1987) 11.
[8] S.A. Kauffman, The origins of order (Oxford Univ. Press, Oxford, 1993).
[9] S.A. Kauffman, Physica D 42 (1990) 135.
[10] B. Lewin, Genes (Wiley, New York, 1991).
[11] E.T. Jaynes, Phys. Rev. 106 (1957) 620.
[12] H. Haken, Information and selforganization (Springer, Berlin, 1988).
[13] J. Wagensberg, A. García and R.V. Solé, in: Maximum entropy and Bayesian methods, eds. W.T. Grandy and L.H. Schick (Kluwer, Dordrecht, 1991).
[14] C.G. Langton, Physica D 52 (1990) 12.
[15] J.P. Crutchfield and K. Young, Phys. Rev. Lett. 63 (1989) 105.
[16] S.A. Kauffman, Sci. Am. 265 (1991) 78.
[17] B. Alberts, D. Bray, J. Lewis, K. Roberts and J.D. Watson, Molecular biology of the gene (Garland, New York, 1993).
[18] K.E. Kurten, Phys. Lett. A 129 (1992) 157.
[19] R.J. De Boer and A. Perelson, J. Theor. Biol. 149 (1991) 381.