

Variations as order parameter and complexity measure for random Boolean networks

Bartolo Luque¹, Fernando J Ballesteros² and Manuel Fernández¹

¹ Departamento de Matemática Aplicada y Estadística, Escuela Superior de Ingenieros Aeronáuticos, Universidad Politécnica de Madrid, Plaza Cardenal Cisneros 3, Madrid 28040, Spain

² Observatori Astronòmic, Universitat de València, Ed. Instituts d'Investigació, Pol. La Coma s/n, E-46980 Paterna, Valencia, Spain

E-mail: bartolo@dmae.upm.es

Received 15 July 2004, in final form 13 October 2004

Published 19 January 2005

Online at stacks.iop.org/JPhysA/38/1031

Abstract

Several order parameters have been considered to predict and characterize the transition between ordered and disordered phases in random Boolean networks, such as the Hamming distance between replicas or the stable core, which have been successfully used. In this work, we propose a natural and clear new order parameter: the temporal variance. We compute its value analytically and compare it with the results of numerical experiments. Finally, we propose a complexity measure based on the compromise between temporal and spatial variances. This new order parameter and its related complexity measure can be easily applied to other complex systems.

PACS numbers: 05.40.Fb, 02.50.-r, 02.70.Uu

1. Introduction

Complex systems are composed of many interacting elements, and usually present different regimes of behaviour. Most of them show an order/disorder transition with a clear frontier between the two regimes. Due to the great interest of most of these systems, many efforts have been made to determine analytically the transition between order and disorder. For this purpose, several order parameters have been proposed and through their study the boundary transition has been successfully determined in many cases.

Random Boolean networks (RBNs) are a case of a well-known complex system that presents a clear and well-studied transition between ordered and disordered phases. Although RBNs were proposed in the 1960s by Kauffman [1], a good phase transition characterization was not available until the 1980s. Several order parameters have been used in order to study and characterize the transition, i.e. the Hamming distance between two annealed replicas

[2], the stable core [3], or more recently, Boolean derivatives [4] and self-overlap between successive configurations [5].

In this work, we present a new order parameter that can be successfully used to determine the boundaries between ordered and disordered phases: the temporal variance. By using it, we can characterize the transition between order and disorder in random Boolean networks. We will also see how the compromise between temporal and spatial variances can be interpreted as a natural measurement of the system complexity. Temporal variance as order parameter and complexity as defined here can be easily applied to other complex systems.

2. Random Boolean networks

Random Boolean networks (RBNs) [1] are systems composed of a number N of automata ($i = 1, \dots, N$) with only two states available ($x_i(t) = 0$ or $x_i(t) = 1$). Each automaton i has associated other K automata i_1, i_2, \dots, i_k , that act as arguments for a Boolean function f_i that will be used to update the automaton state at each time step. That is, the automaton i will change its state x_i at each time step according to the rule

$$x_i(t+1) = f_i(x_{i_1}(t), x_{i_2}(t), \dots, x_{i_k}(t)). \quad (1)$$

Both f_i and the identity of its K inputs are initially randomly assigned to the automaton i , but maintained afterwards through the evolution of the system. To be more precise, the N Boolean functions are created by randomly generating outputs of value 1 with a probability p , and of value 0 with a probability $1 - p$. Thus p is called the bias of the network.

It is well known that RBNs, for a given value of p , show a transition between ordered and disordered phases, separated by a critical value K_c of the connectivity. In the ordered phase ($K < K_c$) the networks freeze in a pattern after a short transient where nearly all the automata remain in a completely frozen state. On the other hand, in the disordered phase ($K > K_c$) all patterns are lost and the automata appear to be in a completely disordered state. For instance, as we change the value of p the critical value K_c at which the transition takes place also changes and a ‘critical line’ appears in a $K-p$ phase diagram (figure 1, solid line). This line corresponds to [2, 6]:

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

3. Numerical variance in RBNs

In figure 1, three typical examples of RBN evolution are shown. All the networks have $N = 50$ automata. The black colour corresponds to automata in state 1 and the white colour in state 0. Each run contains 50 consecutive states, with time increasing upwards along the vertical axis. For $K = 3$, the value $p_c \approx 0.79$ (computed from equation (2)) acts as a critical value. For $0.9 = p > 0.79$ we have RBNs in ordered phase, and their dynamics is clearly ‘frozen’ (ordered). For $0.6 = p < 0.79$ RBNs are in disordered phase, as can be seen in the example of the figure. If $p = 0.6$ and p is fixed, the critical value can be found in $K_c = 2.08$. For $1 = K < 2.08$ we enter again the ordered zone and again the evolution of automata is ‘frozen’. Although in the ordered zone we can find automata with values changing periodically, generally order and disorder differ in the frostbite (or not) of most of their components. It seems natural to characterize both phases by just using a measure of the temporal variation of their states. The easiest and most direct measurement of this variation is the variance. In the ordered phase, a frozen automaton always has the same value, its temporal

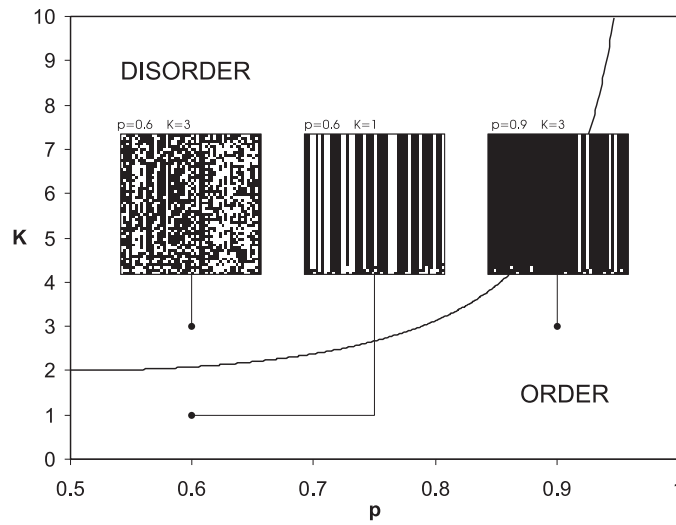


Figure 1. The boundary between the disordered and the ordered phase is shown in a K - p phase diagram. Three examples of networks each with $N = 50$ automata are shown. For $p = 0.6$ and $K = 3$ the network is in the disordered phase, as can be seen in the automata behaviour. If the connectivity is decreased (i.e. to $K = 1$) or the bias is increased (i.e. to $p = 0.9$) the network enters the ordered zone and the automata values, after a short transient, become frozen. Each run contains 50 consecutive states, time increasing upwards along the vertical axis.

variance is 0 and therefore the mean temporal variance of the network will be 0. As the system enters the disordered phase, the temporal variance increases. Hence the variance seems to be a natural candidate for an order parameter.

To measure numerically the temporal variance in a RBN, the system has to reach the stationary state, after a long enough transient. In the stationary state we store T iterations of the system, obtaining an $N \times T$ array of 0s and 1s. In any case, the proportion of 1s and 0s in the array will be p and $1 - p$.

Figure 2 shows the average results for simulations with $N = 10\,000$ and three possible connectivities: $K = 3, 4$ and 5 . For each K we have generated RBNs with p ranging from 0.5 to 1 with a step of 0.01. Each point represents the average of ten systems with identical K and p . For each K and p , after a transient of 250 iterations, we compute the variance of each automaton during the following $T = 250$ iterations by using

$$\sigma_n^2 = \frac{1}{T - 1} \sum_{t=1}^T (x_n(t) - \bar{x}_n)^2, \tag{3}$$

where \bar{x}_n is simply the average value of automaton n in T steps:

$$\bar{x}_n = \frac{1}{T} \sum_{t=1}^T x_n(t). \tag{4}$$

Finally, we define the temporal variance of the network as the average of temporal variances for all the automata:

$$\sigma_t^2 = \frac{1}{N} \sum_{n=1}^N \sigma_n^2. \tag{5}$$

As we can see, figure 2 shows a typical order/disorder phase transition diagram, proving that the experimental σ_t^2 clearly acts as an order parameter, marking the critical values of p for each K , in complete agreement with equation (1).

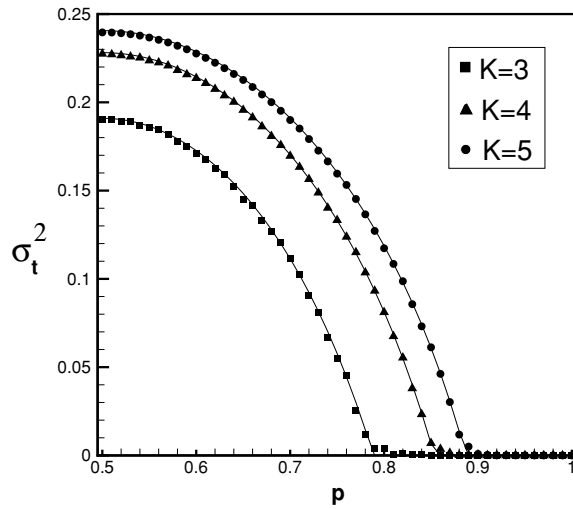


Figure 2. Temporal variance versus bias, for RBNs with $K = 3, 4$ and 5 . Each point represents the average of ten RBN simulations with identical K and p , with $N = 10\,000$. The solid line is the analytical calculation. The diagram shows clearly that the variance acts as an order parameter.

4. Analytical variance in RBNs

Can the temporal variance of the network defined in this way be determined analytically? A RBN is a set of N automata interacting in a strongly non-linear way. To calculate the temporal variance, we will work instead with a system of N decoupled Markov processes with the same statistical properties of RBNs through a mean-field approximation [7].

Previously we define $a_n(t + 1)$, the self-overlap in time $t + 1$ of an automaton n , as 1 if $x_n(t) = x_n(t + 1)$ and 0 otherwise. We will name self-overlap $a(t + 1)$ of a RBN as the average of $a_n(t + 1)$ over all automata at time $t + 1$. This average can also be written as

$$a(t + 1) = \frac{1}{N} \sum_{n=1}^N a_n(t + 1) = \frac{1}{N} \sum_{n=1}^N |1 - x_n(t + 1) - x_n(t)|. \tag{6}$$

It is fairly easy to find the equation that describes the evolution of $a(t + 1)$ in a mean-field approximation,

$$a(t + 1) = a^K(t) + (1 - a^K(t))(p^2 + (1 - p)^2), \tag{7}$$

where K is the connectivity of the network and p is the bias as previously defined. This equation forces $a(t)$ to evolve towards fixed points, $a(t) \rightarrow a$ as t tends to infinity, i.e. in the stationary state, a is a concrete value between 0 and 1, and depends only on K and p .

In an analogous way, we can define the ‘ ij -self-overlap’ $a_{nij}(t + 1)$ for the automaton n as a magnitude whose value is 1 when $x_n(t) = i$ and $x_n(t + 1) = j$, and 0 otherwise, for $i, j \in \{0, 1\}$. Similarly, the averaged value over all automata at time $t + 1$ is $a_{ij}(t + 1)$. And in the stationary state $a_{ij}(t) \rightarrow a_{ij}$. Obviously,

$$a = a_{11} + a_{00}, \tag{8}$$

and

$$a_{01} = a_{10} = \frac{1 - a}{2}. \tag{9}$$

It is easy to obtain analytical expressions for a_{00} and a_{11} by taking into account the fact that the automaton has probability p of being in state 1 and $1 - p$ in the state 0, i.e.,

$$a_{11} + a_{01} = p, \quad a_{00} + a_{10} = 1 - p. \tag{10}$$

From the above equations we have

$$a_{00} = \frac{a}{2} - \left(p - \frac{1}{2}\right), \quad a_{11} = \frac{a}{2} + \left(p - \frac{1}{2}\right). \tag{11}$$

Note that we can interpret these a_{ij} as joint probability distributions. By using them we can calculate the conditioned probabilities of transition from 1 to 0 or vice versa. Indeed, we can now calculate the conditional probabilities $p_{i|j}$ (probability of transition to the state i in $t + 1$, knowing that in t its state was j) as $p_{i|j} = a_{ji}/p_j$. We can interpret these probabilities as the transition matrix of a Markovian process,

$$\mathbf{T} = \begin{pmatrix} p_{0|0} & p_{1|0} \\ p_{0|1} & p_{1|1} \end{pmatrix} = \begin{pmatrix} \frac{a+1-2p}{2(1-p)} & \frac{1-a}{2(1-p)} \\ \frac{1-a}{2p} & \frac{a-1+2p}{2p} \end{pmatrix}, \tag{12}$$

which satisfies

$$\sum_{i=0}^1 p_{i|0} = 1, \quad \sum_{i=0}^1 p_{i|1} = 1, \tag{13}$$

where $p_{i|0}$, $p_{i|1}$ are the transition probabilities from the states 0, 1 to the state i , respectively.

Thus, we have now reduced the whole network to a set of mean-field automata evolving independently under Markovian conditions, all the effects of their interactions being encoded in a . Note that for $a = 1$ (the ordered phase), \mathbf{T} is the identity matrix. A RBN of N automata in ordered phase is equivalent to N Markov chains, pN of them having all their values equal to 1, and $(1 - p)$ of them with all their values equal to 0. Each one of these chains contributes to the average with a temporal variance of 0 (does not change), and therefore the average value of the variance is 0 in the ordered phase.

For $a \neq 1$ it is easy to show that the n th power of the stochastic matrix \mathbf{T} can be written as

$$\mathbf{T}^n = \frac{1}{A + B} \begin{pmatrix} B & A \\ B & A \end{pmatrix} + \frac{(1 - B - A)^n}{A + B} \begin{pmatrix} A & -A \\ -B & B \end{pmatrix}, \tag{14}$$

with

$$A = \frac{1 - a}{2(1 - p)}, \quad B = \frac{1 - a}{2p}. \tag{15}$$

Then in the disordered phase

$$\lim_{n \rightarrow \infty} \mathbf{T}^n = \frac{1}{A + B} \begin{pmatrix} B & A \\ B & A \end{pmatrix}, \tag{16}$$

with equilibrium probability vector $(1 - p, p)$, just as expected.

To compute the temporal variance for the Markov chain we simply need to weight properly the four possible two-time-step chains: 00, 01, 10 and 11. The average value of the case 01 is $\bar{x} = (0 + 1)/2 = 1/2$, and contributes with temporal variance: $\sigma_t^2 = \{(0 - 1/2)^2 + (1 - 1/2)^2\}/(2 - 1) = 1/2$, exactly the same as in case 10. The cases 00 and 11 will have temporal variance 0. Thus, the temporal variance of the network is

$$\sigma_t^2 = pp_{0|1} \frac{1}{2} + (1 - p)p_{1|0} \frac{1}{2} = \frac{1 - a}{2}. \tag{17}$$

Solid lines in figure 2 show the analytical result obtained through this approximation. It can be seen that the deduced variance fits perfectly with the values obtained with real RBNs, hence validating the method.

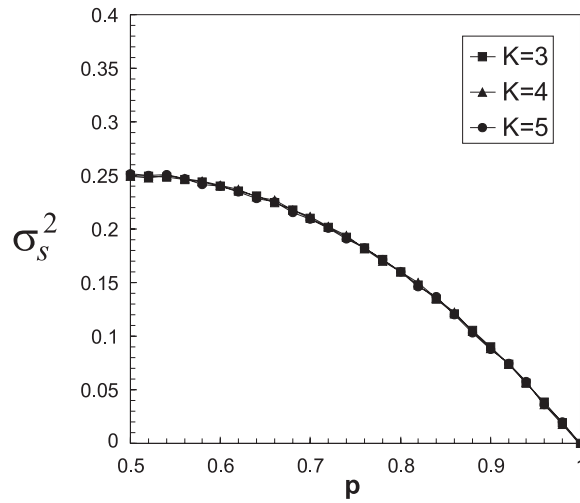


Figure 3. Spatial variance versus bias, for RBNs with $K = 3, 4$ and 5 . Each point represents the average of ten RBN simulations with identical K and p , with $N = 10000$.

5. Variance as a measurement of the complexity

Besides the measurement of temporal variance, we can also take into account the spatial variance, and measure it in a given iteration along all the N automata,

$$\sigma_s^2(t) = \frac{1}{N-1} \sum_{n=1}^N (x_n(t) - \bar{x}(t))^2, \quad (18)$$

where $\bar{x}(t)$ is the average value of all the automata at time t , which for values $N \gg 1$ is clearly equal to the bias:

$$\bar{x}(t) = \frac{1}{N} \sum_{n=1}^N x_n(t) = p. \quad (19)$$

The spatial term can be somewhat misleading as there is not a real Euclidean space in which the nodes are embedded. The nodes simply form a random network with no underlying space. In fact, sometimes it is said that RBNs are infinitely dimensional systems. Anyway, as we mark each automaton with a number, we are somehow ordering and embedding them in an artificial one-dimensional space. Of course, this ordering has no real meaning and it is just a labelling. But it is in this sense we use the word ‘spatial’ when we say spatial variance.

One can easily calculate the spatial variance for values $N \gg 1$,

$$\sigma_s^2(t) = p(1-p)^2 + (1-p)(0-p)^2 = p(1-p), \quad (20)$$

which is independent of K and t and depends only on the bias. Figure 3 shows the values of spatial variances for the same simulations as figure 2, which are in complete agreement with equation (20).

It is clear that the spatial variance σ_s^2 is a measurement of the diversity of possible configurations, reaching its maximum at $p = 0.5$ and its minimum at the extreme cases $p = 0$ or $p = 1$, where it is equal to 0. It is therefore a direct measure of spatial disorder, i.e. it is related to the spatial entropy of the system [5, 7].

Thus, because of this property, we propose a measure of the complexity of the system based on the variance. The complexity of network dynamics can be defined as a compromise

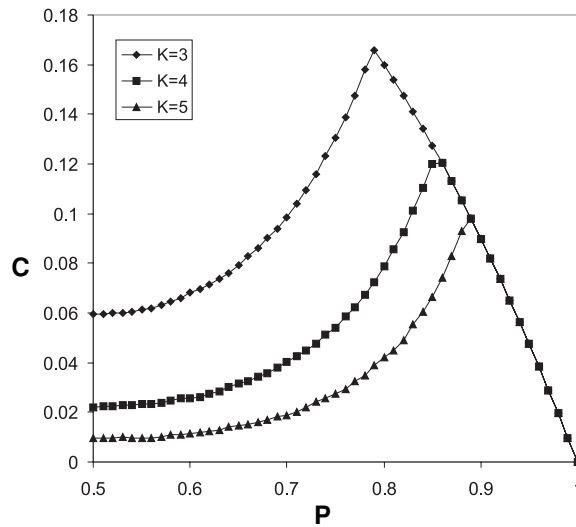


Figure 4. Spatial variance minus temporal variance as a measurement of complexity for RBNs with $K = 3, 4$ and 5 . Each point represents the average of ten RBN simulations with identical K and p , with $N = 10\,000$. It can be seen how complexity is maximum at the critical point between order and disorder.

between spatial disorder (maximum number of possible configurations) and temporal order. We propose as a measure of the complexity of the network, the difference between spatial and temporal variance:

$$C = \sigma_s^2 - \sigma_t^2 = p(1 - p) - \frac{1 - a}{2}. \tag{21}$$

This measure of complexity incorporates the main features of the intuitive notion of such a magnitude (see figure 4). It is minimum (in fact 0) for a completely crystallized system, when the output of all their Boolean functions is the same (that is, in a completely biased RBN, with $p = 0$ or $p = 1$). Its other minimum is at $p = 0.5$, where the system is more disordered, this minimum being smaller as we get deeper into the disordered zone (see figure 1) by increasing the value of K . The important point is that the complexity is maximum at the transition point between order and disorder, where $1 - a$ is equal to 0 for the first time.

Besides, as the critical p increases when K increases, the so-defined complexity is maximum at the critical point of $K = 2$, decreasing for larger K (note that for $K = 1$ RBNs are always ordered and there is no critical point).

6. Conclusions

Random Boolean networks are a classical well-studied example of complex systems with a clear transition between order and disorder. Several order parameters have been successfully applied to the study of RBNs, most of them of general use in other complex systems. In this work we present a more intuitive and easy to apply, new order parameter: the temporal variance, σ_t^2 . Temporal variance successfully points out the order/disorder transitions of the system and indeed behaves as expected in a classical order parameter. The election of this magnitude as order parameter is well based, and can easily be applied to other complex systems. We have also deduced analytically the variance for RBNs and checked that it is in complete agreement with numerical simulations.

As a bonus, we have also found that the difference between spatial and temporal variance can be reasonably used as a measure of complexity for random Boolean networks, as it reaches its maximum values at the boundary between order and disorder. Such a definition of complexity can easily be applied to many other physical systems.

Acknowledgments

The authors would like to thank Amelia Ortiz, Jose Olarrea and Marta Cordero for their help and the anonymous referees for their useful comments. This work has been supported by CICYT BFM2002-01812 (BL) and by MEC AYA 2003-08739-c02-0 (including FEDER) and Generalitat Valenciana GRUPOS 03/170 (FB).

References

- [1] Kauffman S A 1969 *J. Theor. Biol.* **22** 437
Kauffman S A 1974 *J. Theor. Biol.* **44** 167
Kauffman S A 1984 *Physica D* **10** 145
Kauffman S A 1990 *Physica D* **42** 135
Kauffman S A 1993 *The Origins of Order* (Oxford: Oxford University Press)
- [2] Derrida B and Pomeau Y 1986 *Europhys. Lett.* **1** 45
Derrida B and Pomeau Y 1986 *Europhys. Lett.* **2** 739
- [3] Flyvbjerg H 1988 *J. Phys. A: Math. Gen.* **21** L955
- [4] Luque B and Solé R V 2000 *Physica A* **284** 33
- [5] Luque B and Ferrera A 2000 *Complex Syst.* **12** 241
- [6] Luque B and Solé R V 1997 *Phys. Rev.* **55** 1
- [7] Wentian Li 1990 *J. Stat. Phys.* **60** 823
Wentian Li 1991 *Complex Syst.* **5** 381