Length of state cycles of random Boolean networks: an analytic study



2000 J. Phys. A: Math. Gen. 33 6699

(http://iopscience.iop.org/0305-4470/33/38/304)

J. Phys. A: Math. Gen. **33** (2000) 6699–6706. Printed in the UK PII: S0305-4470(00)11276-4

# Length of state cycles of random Boolean networks: an analytic study

Z Somogyvári and Sz Payrits

Department of Biophysics, KFKI Research Institute for Particle and Nuclear Physics of the Hungarian Academy of Sciences, H-1525 Budapest, PO Box 49, Hungary

Received 21 January 2000, in final form 6 June 2000

**Abstract.** In this paper we consider the mean length of transients and the length of state cycles in random Boolean networks. We present an approximate calculation of these quantities as a function of the size and connectivity of the network, where using an annealed approximation we derive a recursive formula for the length of steps of the system. Using the mean step length and an 'effective momentary' state space, we calculate an approximate formula for the probability distribution function (PDF) of the length of state cycles and transients. We compare this PDF with analytical results in special cases and with simulations by the Monte Carlo procedure.

## 1. Introduction

Random Boolean networks are typical examples of simple systems exhibiting complex behaviour and were first introduced and applied to biological networks by Kauffman [1]. They have been the subject of extensive research for several years and have become model systems of many complex biological networks. This type of network has been applied as an abstract model of numerous nonlinear complex systems: systems of interacting catalysts at the origins of life, self-regulatory genomic systems, coupled co-evolutionary ecosystems [2] and neural networks [3, 4]. The behaviour of cellular networks is also closely related to classical models of statistical physics such as the Ising model and percolation theory [5, 6]. Most previous studies have considered this system from the point of phase transition [7–9] and have recently extended this phenomenon to a random ordinary differential equation [10].

Focusing on the relationship between network structure and network dynamics, random Boolean networks deserve study in their own right. Initially, the statistical properties of trajectories of these networks were analysed using numerical simulations. Subsequent studies examined the length and number of attractors and their stability [11], the multivalley structure of the system [6, 12], the multifractality in the size of avalanches caused by an initial damage [13]; many of them also investigated generalizations of this system, such as ordered-structure automata [14, 15] or the inhomogeneous Kauffmann model [16, 17].

However, the scope of existing analytical results is rather restricted. The theory of random mappings [18] provides a good approximation for high connectivity and the large-system case [19]. On the other hand, Flyvbjerg and Kjær [20] provided exact solutions for most of the statistical properties for the lowest connectivity case, when each cell had only one input, at any system sizes. Between these two extremities, an annealed approximation by Derrida and Pomeau [21] gave quantitative predictions for the evolution of distance between two randomly chosen initial conditions. They predicted the critical number of input connections: $K_c = 2$.

0305-4470/00/386699+08$30.00  

This prediction agreed with numerical simulations and was later proven to be exact in the thermodynamical limit under given temporal constraints [22]. In spite of the long history of the topic, there are neither analytical results nor even approximations for the first simulated statistical properties of the system for cases other than the two extremes. The aim of this study is to provide an analytical approximation for the mean length of attractors and for the transient length at any connectivity parameters (even for finite system sizes) and to give a schema to understand the principles underlying the behaviour of this (structurally) simple but (dynamically) complex system.

## 2. Methods describing random Boolean networks

### *2.1. Structure*

A family of Boolean networks is characterized by the number of cells ($N$), and the number of input connections ($K$) of a cell that equals the mean number of output connections. In a representative of a family of these networks, $N$ Boolean functions (one for each cell) are chosen randomly with uniform distribution from the possible $2^{2^K}$ functions. During the construction of a representative net, $K$ cells are chosen randomly with uniform distribution (among the $N$) to provide the input for each function. Each element can be in either an *on* or *off* state and the discrete dynamics is given as the synchronous change of the states of all cells according to their respective Boolean functions (depending on the states of their respective $K$ inputs).

The network structure and Boolean functions were quenched after construction. Properties presented here are given as an average over all possible network structures; Boolean functions and initial conditions and are denoted by an overline.

### *2.2. Dynamics*

The state of the system is defined by an $N$-dimensional vector of Boolean functions' output values; thus the state space consists of an $N$-dimensional hypercube's vertices. The length of a step is given by the number of different values in two consecutive states.

For better understanding, let us follow the first step of the system. Due to the uniform distribution of initial conditions and Boolean functions the result of this first step is also uniformly distributed in state space. The probability density function (PDF) of these two states' Hamming distance (i.e. the PDF of first step length, $l_1$) is

$$p(\lambda_1 = l_1) = \binom{N}{l_1}\left(\frac{1}{2}\right)^N. \tag{1}$$

Comparing the first and second state, the number of changed outputs is $l_1$. Similarly to Derrida's calculations [21], the probability that all inputs of $m$ cells come from unchanged cells (if multiple connections are allowed) is

$$p(\mu = m) = \binom{N}{m}\left(\frac{N-l_1}{N}\right)^{Km}\left(1-\left(\frac{N-l_1}{N}\right)^K\right)^{N-m}. \tag{2}$$

This means that the inputs of $m$ cells remains unchanged, while $N-m$ cells have at least one changed input. Due to uniformly chosen Boolean functions, each of them changes its output with probability $\frac{1}{2}$. As the state of the system is determined by output values of cells, the PDF of the difference between the first state's successor and the second state's successor (i.e. the PDF of second step length, $l_2$) is

$$p(\lambda_2 = l_2) = \sum_{m=0}^{N} p(\mu = m)\left(\frac{1}{2}\right)^{N-m}\binom{N-m}{l_2}. \tag{3}$$

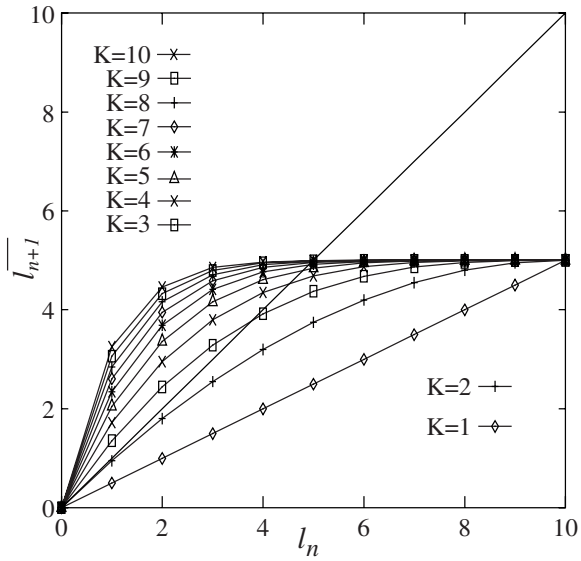


**Figure 1.** Evolution of step length in a return map-like representation. Mean length of the next step $\overline{l_{n+1}}$ is displayed as a function of previous step length $l_n$, for $N = 10$ and $K = 1$–10. For $K = 1$ or 2, only one fixpoint exists (at zero) and it is stable. For $K \geqslant 3$, zero becomes unstable but another stable fixpoint emerges. This fixpoint converges to $N/2$ as $N$ and $K$ tend to infinity.

Neglecting the correlation between consecutive steps and performing the summation over $m$, a recursive formula is derived for the successive step length:

$$p(\lambda_{n+1} = l_{n+1} \mid \lambda_n = l_n) = \binom{N}{l_{n+1}} \left(\frac{1}{2} + \frac{1}{2}\left(1 - \frac{l_n}{N}\right)^K\right)^{N-l_{n+1}} \left(\frac{1}{2} - \frac{1}{2}\left(1 - \frac{l_n}{N}\right)^K\right)^{l_{n+1}}. \quad (4)$$

Thus

$$\overline{p(\lambda_{n+1} = l_{n+1} \mid \lambda_n = l_n)} = \frac{N}{2}\left(1 - \left(1 - \frac{l_n}{N}\right)^K\right)^{l_{n+1}}. \quad (5)$$

Applying this mean approximation, equation (5) gives the evolution of step length during this random walk. The mean length of the next step as a function of the previous step is shown in figure 1. The step length shortens exponentially when $K = 1$, and converges exponentially to a finite value if $K \geqslant 3$. If $K = 2$ the decay is more slow, and proportional to $1/n$. This slow decay reflects the propensity of the system to sustain large fluctuations. This type of behaviour is closely related to the classical model of percolation [23]. Note that if $K$ and $N$ tend to infinity, equation (4) tends to the distribution in (1), referring to the uniform distribution of consecutive states in state space. In this case we return to the random mapping approximation.

A trajectory of the system is considered as a self-avoiding walk in a contractive state space, assuming that the system chooses a new state randomly from the current state space until it reaches an already visited state. The volume contraction refers to the 'freezing' of cells. 'Freezing' means that the state of a cell becomes stable from a moment of walk. The volume of the current state space was estimated backwards from the momentary step length. The effective momentary state space volume is

$$V_n = 2^{N(1-(1-\frac{l_n}{N})^K)}. \quad (6)$$

During the contraction, the number of reachable visited states also decreases. Let us denote as $b_n$ the number of reachable visited states in the $n$th step. We assume that their decrease is proportional to the state space decrease:

$$b_{n+1} = b_n \frac{V_{n+1}}{V_n} + 1. \tag{7}$$

$b_1 = 1$, thus

$$b_{n+1} = 1 + V_{n+1} \sum_{k=1}^{n} \frac{1}{V_k}. \tag{8}$$

By means of these quantities, we can construct the characteristic measures of the system's behaviour.

### *2.3. Characteristic measures*

The length of a self-avoiding walk (the number of steps required to reach an already visited place) equals the transient length plus the state cycle length ($tr + cl$), and its PDF is expressed by the following formula:

$$p(cl + tr = n) = \sum_{l_1=0}^{N} p(\lambda_1 = l_1) \frac{b_n}{V_n} \prod_{k=1}^{n-1} \left(1 - \frac{b_k}{V_k}\right) \tag{9}$$

while the PDF of the state cycle length ($cl$) is

$$p(cl = z) = \sum_{l_1=0}^{N} p(\lambda_1 = l_1) \sum_{k=z+1}^{M} \frac{1}{V_{k-z}} \prod_{i=1}^{k-1} \left(1 - \frac{b_i}{V_i}\right). \tag{10}$$

## 3. Results

When $K$ and $N$ tend to infinity, our calculations gave the same results as the random mapping approximation which is proven to be exact in the thermodynamical limit. The short approximative form in this case is

$$\overline{cl + tr} = 2\overline{cl} = \tfrac{1}{4}\sqrt{2\pi} 2^{\frac{N}{2}}. \tag{11}$$

Even in the case of finite connectivity and finite system size, our calculations for the length of the whole path and the cycle length agreed qualitatively with simulations, so the calculated decrease of step length can explain the observed phenomenon, see figures 2 and 3. In some cases ($K = 1, 2, 5$) they were even quantitatively acceptable approximations of simulations. Over the threshold for connectivity, the mean length of self-avoiding walks and cycle length rose exponentially as a function of $N$. In the case of $K = 1$ both measures remained at a very low value as in the simulations, but the relative error of computation was high. When $K = 2$ it rose much slower than exponentially, probably as a square-root function, very close to the simulated values. In the exponential regime, but close to the critical border, at $K = 3$ and 4, our calculations strongly overestimate the simulated results. This can be attributed to the approximation neglecting correlations of consecutive steps and using a mean-field-like approximation for the evolution of step length.

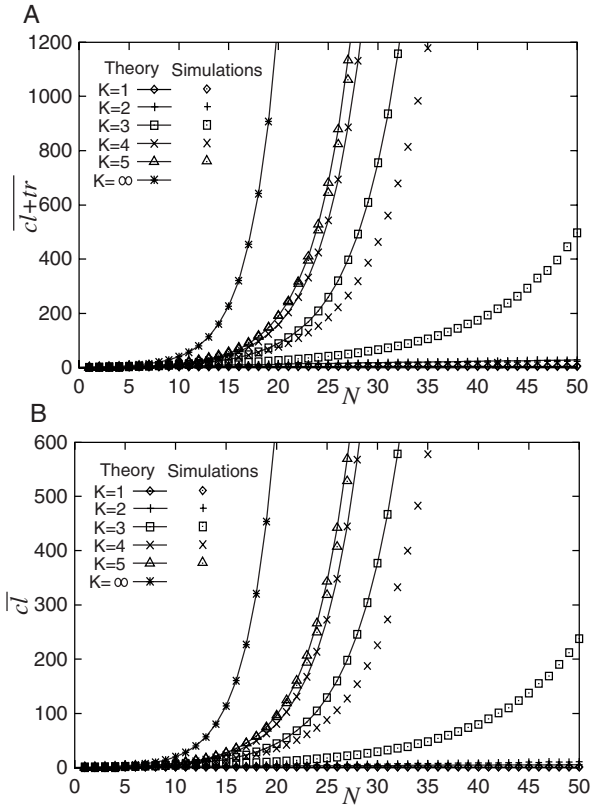


**Figure 2.** (*a*) Mean length of self-avoiding walks ($\overline{tr + cl}$) and (*b*) mean length of state cycles $\overline{cl}$ as a function of system size ($N$) and connectivity ($K$). The exponential increase for $K \geqslant 3$ refers to the 'chaotic' phase. Linked symbols: analytical calculations; unlinked symbols, numerical results from simulations. $K = \infty$ comes from random mapping approximation.

## 4. Discussion

Our approach has provided an analytical approximation of the mean length of state cycles and transients even at small system sizes, for any connectivity parameters. A major advantage of our approach is that it allowed us to calculate these long-examined properties. A major disadvantage of this approach is the low quantitative precision, especially in the exponential regime, close to the critical connectivity limit. By taking correlations between consecutive steps into consideration, this approximation could be improved, especially around the critical limit. In addition, an important future goal is to find short approximations of the presented measures, especially to check that cycle length increases as a square-root function in the case $K = 2$. Of the possible applications of this theory we would emphasize two main directions. Following Kürten's arguments [4], results for random Boolean networks are applicable for Hopfield-type neural networks and for attractor networks and are also appropriate for understanding the relationship between the structure and dynamics of these networks. In other words, the structure in 'real' space corresponds to the structure of a content-addressable memory in 'meaning' space. The 'chaotic' feature of this type of network is able to enhance the learning

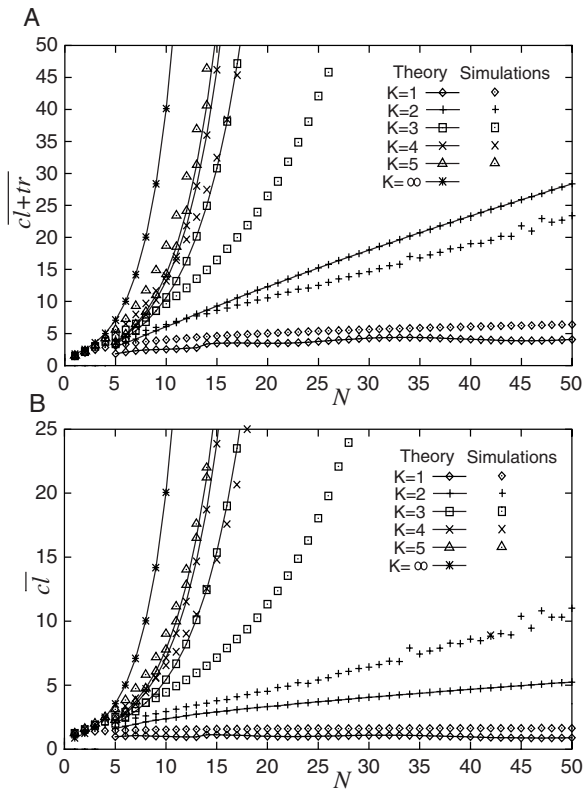


**Figure 3.** Zoom of 'frozen' phase. (*a*) Mean length of self-avoiding walks ($\overline{tr+cl}$) and (*b*) mean length of state cycles $\overline{cl}$ as a function of system size ($N$) and connectivity ($K$). In the case of $K = 2$, ($\overline{tr+cl}$) and ($\overline{cl}$) increase slowly but in the case of $K = 1$ both measures remain at a low value (same notation).

ability of an associative memory [24].

Recent findings have proved that numerous complex networks have random structure where the PDF of an element's connection number differs from the Gaussian-like distribution. These 'small-world' networks can be characterized by connectivity distribution following a power function. This property has been found in the networks of airlines, social connections, and the link structure of the World Wide Web [25]. By means of random Boolean network theory we could consider them as a dynamical system, and promote our understanding of these networks' behaviour.

## Acknowledgments

We would like to thank Máté Lengyel and Fülöp Bazsó and Krisztina Szalisznyó for their helpful comments on the manuscript. This work was supported by the National Scientific Research Foundation (OTKA) under grant number T025472.

## Appendix

For sake of accuracy we give the results in numerical forms in tables 1 and 2.

**Table 1.** The mean length of state cycles ($\overline{cl}$).

| $K$ | 1 | | 2 | | 3 | | 4 | | 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | sim. | calc. | sim. | calc. | sim. | calc. | sim. | calc. | sim. | calc. |
| 5 | 1.463 | 0.986 | 2.093 | 1.616 | 2.519 | 2.020 | 2.845 | 2.200 | 3.115 | 2.294 |
| 10 | 1.538 | 0.997 | 2.95 | 2.382 | 4.449 | 5.441 | 6.551 | 7.148 | 9.011 | 7.788 |
| 15 | 1.583 | 1.120 | 3.794 | 2.912 | 7.152 | 15.37 | 14.80 | 23.85 | 27.97 | 27.31 |
| 20 | 1.606 | 1.016 | 4.534 | 3.328 | 11.33 | 44.47 | 35.57 | 80.57 | 93.11 | 96.70 |
| 25 | 1.619 | 1.011 | 5.388 | 3.707 | 18.04 | 129.3 | 87.94 | 272.9 | 318.6 | 343.0 |
| 30 | 1.626 | 1.085 | 6.421 | 4.054 | 29.17 | 337.0 | 225.6 | 808.9 | | |
| 35 | 1.623 | 1.066 | 7.442 | 4.372 | 47.86 | | 578.0 | | | |
| 40 | 1.637 | 0.966 | 8.606 | 4.677 | 79.84 | | | | | |
| 45 | 1.648 | 0.885 | 10.37 | 4.974 | 137.7 | | | | | |
| 50 | 1.644 | 0.899 | 11.00 | 5.251 | 237.7 | | | | | |

**Table 2.** The mean length of self-avoiding walks ($\overline{tr + cl}$).

| $K$ | 1 | | 2 | | 3 | | 4 | | 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | sim. | calc. | sim. | calc. | sim. | calc. | sim. | calc. | sim. | calc. |
| 5 | 3.081 | 1.843 | 3.921 | 2.938 | 4.633 | 3.362 | 5.189 | 3.552 | 5.596 | 3.668 |
| 10 | 4.014 | 2.531 | 6.397 | 6.033 | 9.642 | 10.61 | 13.93 | 13.02 | 18.69 | 14.26 |
| 15 | 4.603 | 3.460 | 8.612 | 9.225 | 16.46 | 30.80 | 32.44 | 46.15 | 58.46 | 52.61 |
| 20 | 5.011 | 3.462 | 10.54 | 12.30 | 26.46 | 89.23 | 76.63 | 159.0 | 191.3 | 190.3 |
| 25 | 5.361 | 3.742 | 12.50 | 15.23 | 42.07 | 259.2 | 185.1 | 542.7 | 644.7 | 681.4 |
| 30 | 5.631 | 4.254 | 14.66 | 18.04 | 66.13 | 754.7 | 463.0 | | | |
| 35 | 5.847 | 4.346 | 16.80 | 20.71 | 106.2 | | 1178 | | | |
| 40 | 6.052 | 4.081 | 19.02 | 23.33 | 174.7 | | | | | |
| 45 | 6.227 | 3.867 | 21.78 | 25.89 | 292.8 | | | | | |
| 50 | 6.376 | 4.055 | 23.38 | 28.36 | 496.7 | | | | | |

## References


[1] Kauffman S A 1969 *Nature* **224** 177
[2] Kauffman S A 1989 *Origins of Order: Self-Organization and Selection in Evolution* (Oxford: Oxford University Press)
[3] Derrida B, Gardner E and Zippelius A 1987 *Eurphys. Lett.* **4** 167
[4] Kürten K E 1988 *J. Phys. A: Math. Gen.* **21** L615
[5] Derrida B 1987 *J. Phys. A: Math. Gen.* **20** L721
[6] Stauffer D 1987 *Phil. Mag.* B **56** 901
[7] Luque B and Solè R V 1995 *Phys. Lett.* A **196** 331
[8] Luque B and Solè R V 1997 *Phys. Rev.* E **55** 257
[9] Kürten K E 1988 *Phys. Lett.* A **129** 157
[10] Glass L and Hill C 1998 *Europhys. Lett.* **41** 599
[11] Kauffman S A 1984 *Physica* D **10** 145
[12] Derrida B 1987 *Phil. Mag.* B **56** 917
[13] Jan N 1988 *J. Phys. A: Math. Gen.* **21** L899
[14] Atlan H 1987 *Phys. Scr.* **36** 563
[15] Soulie F F, Robert Y and Tchuente M (ed) 1987 Nonlinear science: theory and applications *Automata Networks in Computer Science Theory and Applications* (Manchester: Manchester University Press) pp 20–34, chapter A.1
[16] Kürten K E 1989 *J. Physique* **50** 2313
[17] Kürten K E and Beer H 1997 *J. Stat. Phys.* **87** 929
[18] Harris B 1960 *Annu. Math. Stat.* **31** 1045

[19] Derrida B and Flyvbjerg H 1987 *J. Physique* **48** 971
[20] Flyvbjerg H and Kjær N J 1988 *J. Phys. A: Math. Gen.* **21** 1695
[21] Derrida B and Pomeau Y 1986 *Europhys. Lett.* **1** 45
[22] Hilhorst H J and Nijmeijer M 1987 *J. Physique* **48** 185
[23] Schulman L S and Seiden P E 1986 *Science* **233** 425
[24] Nara S and Davis P 1997 *Phys. Rev.* E **55** 826
[25] Albert R, Jeong H and Barabási A L 1999 *Nature* **401** 130