

CHAOTIC MEAN FIELD DYNAMICS OF A BOOLEAN NETWORK WITH RANDOM CONNECTIVITY

MALIACKAL POULO JOY*, DONALD E. INGBER[†] and SUI HUANG[‡]

Vascular Biology Program Children's Hospital and Harvard Medical School 300 Longwood Ave, Boston, MA 02115, USA *mpjoy@unc.edu *joy.maliackal@childrens.harvard.edu †donald.ingber@childrens.harvard.edu ‡sui.huang@childrens.harvard.edu

> Received 18 March 2007 Accepted 22 March 2007

Random Boolean networks have been used as simple models of gene regulatory networks, enabling the study of the dynamic behavior of complex biological systems. However, analytical treatment has been difficult because of the structural heterogeneity and the vast state space of these networks. Here we used mean field approximations to analyze the dynamics of a class of Boolean networks in which nodes have random degree (connectivity) distributions, characterized by the mean degree k and variance D. To achieve this we generalized the simple cellular automata rule 126 and used it as the Boolean function for all nodes. The equation for the evolution of the density of the network state is presented as a one-dimensional map for various input degree distributions, with kand D as the control parameters. The mean field dynamics is compared with the data obtained from the simulations of the Boolean network. Bifurcation diagrams and Lyapunov exponents for different parameter values were computed for the map, showing period doubling route to chaos with increasing k. Onset of chaos was delayed (occurred at higher k) with the increase in variance D of the connectivity. Thus, the network tends to be less chaotic when the heterogeneity, as measured by the variance of connectivity, was higher.

Keywords: Boolean network; cellular automata; chaos; Lyapunov exponent; mean field theory.

PACS Nos.: 82.40.Bj, 02.50.-r, 05.45.Pq, 05.50.+q, 87.10.+e.

1. Introduction

Understanding the complexity exhibited by biological systems is one of the challenges of "post-genomic" era. Complex behaviors do not require complicated

^{*}Present address: Department of Pharmacology, University of North Carolina, 1139 Mary Ellen Jones Building CB# 7365, Chapel Hill, NC 27599, USA.

systems, as amply demonstrated in computational model systems of life, such as cellular automata or discrete networks.¹ A system composed of elements whose individual behaviors are constrained by simple rules that determine how these component parts influence each other, and thus establish a network, is a paradigm of such model systems. In this regard, Boolean networks, originally inspired by gene regulatory networks, provide a convenient and tangible approach to study complexity at the system-level dynamics of a multi-component system.

Random Boolean networks (RBN) were first introduced as a simplified model for large genomic networks by Kauffman.^{2,3} In RBN, a set of N nodes are randomly linked via Boolean functions, which determine the (binary) activity state value of a node. A large series of studies have demonstrated its utility as a model for complex bio-systems.^{4,5} Specifically, one interesting finding is that under certain circumstances the complexity of the network architecture afforded by the large number N of interacting elements collapses to generate simple, ordered and stable behavior in high-dimensional state space. Kauffman proposed that this principle underlies the spontaneous generation of order in complex (irregularly wired) networks, resulting in just a few stable states (relative to the fast number of combinatorially possible states of the network). This led to the central hypothesis that these stable, high-dimensional attractor states represent the observable cell types in the body.² Recently, using DNA microarrays for the simultaneous measurement of dynamics of the activity state of thousands of genes, experimental evidence has been obtained supporting this hypothesis.⁶

Conversely, random Boolean networks (RBN) and other discrete models can also serve to study how simple systems generate complex, behavior. RBN can be viewed as generalizations of cellular automata (CA) which were originally introduced by von Neumann in the 1940's and were extensively studied in the 1980's by Wolfram.^{1,7} They exhibit astonishing complex dynamics despite their simple system architecture. CAs are dynamical systems composed of cells (automata) which are arranged on a finite dimensional lattice and which can assume discrete (binary) state values, e.g., "ON" and "OFF" ("1" or "0"), that are updated in discrete space and time. The updating of the state value of a cell follows rules which depend upon the current states of the automata in a specified neighborhood. In one-dimensional automata with binary states and one neighbor on each side, 256 rules are possible. These basic CA rules were classified based on the dynamical behavior that they exhibit.⁷ Boolean networks represent a general case of CA in that the interaction between the cells (nodes) is not limited to the spatial neighbors. A set of CA rules that is symmetric with respect to and independent of the neighborhood, such as rule 126 (see below) can be easily applied to Boolean networks and are scalable in cases of non-uniform number of inputs (varying size of neighborhoods) without explicit specification.

Most works on RBN have been motivated by the question of how and when they generate ordered, rather than complex (chaotic) dynamics.^{3,8} Moreover, past studies often use the simplification of a uniform number of inputs (k) throughout the network. However, in biological networks, k is not fixed, but exhibit a broad distribution (which more specifically, have been proposed to follow a power-law distribution).⁹ Thus, we set out to study how chaotic dynamics of RBN is influenced by non-fixed k.

Here we are interested in a Boolean network model with a generalization of the basic CA rule 126 as per Wolfram's notation.⁷ By fixing the rule over all network nodes as a trade-off for varying k, we achieve some homogeneity that facilitate analysis. The rule 126 states that a given cell goes to the "OFF" (="0") state if the cells in its neighborhood (including itself) are all either "ON" or "OFF". Thus, extreme homogeneity of states of a neighborhood kills a cell. Rule 126 can be considered as a simple growth rule. Explicitly, in this CA, if the states are either "000" or "111" the new state of the central cell will be "0" and for all other configurations, namely "001", "010", "011", "100", "101", and "110", the new state will be "1". This rule is one of the 32 legal rules. Legal rules are the rules with reflection symmetry, and with the state "000" always giving a "0" output state. According to Wolfram's classification scheme, rule 126 is a chaotic rule, since it produces complex patterns and shows sensitivity to initial conditions. However, this chaotic behavior is not apparent when system behavior is monitored with the macroscopic variable c(t), or density. This aggregate quantity is the average number of "1"s of the system in a given state. In the case of rule 126, the density approaches a constant value for the simple case of one-dimensional CA with two neighbors.

In networks, such as biochemical or gene regulatory networks, interactions take place not only between the nearest neighbor cells (genes) on an array but also between distant ones, since there is no notion of an array. Thus, to incorporate non-local interactions, we move from the CA to the RBN formalism where there is no spatial order relationship between the cells (= nodes). Then the neighborhood of a node is expanded from the two nearest neighbors to any number of nodes in the network with which that node is connected, and rule 126 can be generalized as follows. The future state (at time t + 1) is "1" in all these cases except when the current states (at time t) of all neighbors, i.e., direct interaction partners, are either "0" or "1"; otherwise the future state will be "0". Rule 126 is a totalistic rule because it depends only on the total number of "1"s (or equivalently "0"s) among the neighbors, and hence can be easily scaled with varying numbers of inputs k("neighborhood size").

When the neighborhood has a regular structure, as in the case of CA, the dynamics of c(t) shows either a fixed point behavior or utmost a two period behavior, even for higher number of neighbors.^{10, 11} However, if one includes nonlocal interactions, as in the case of a Boolean network, then many interesting dynamics can occur for a system that uses rule 126. In this paper we consider a semilocal generalization of the neighborhood as the neighborhood includes the cell to be updated itself.¹¹ However, the results are qualitatively the same for strictly non-local or semi-local neighborhood (where cells to be updated are excluded or included, respectively). The structure of all the networks considered here are based on directed graphs.

The mean field theory of CAs in general have been described in detail by Gutowitz $et \ al.^{12}$ and the collective behavior of CA have been studied by Chate and Manneville using mean field approaches.¹⁰ Chaotic behavior in the temporal evolution of the density c(t) described by the mean field approximations of various CA rules has been studied by several authors.^{13–15} Mean field theory describes the whole system behavior in terms of one macroscopic variable, such as the density, c(t). Hence one can conveniently study chaos of the system by analyzing the density, c(t), instead of numerically evaluating chaos using microscopic variables, as in the case of the Derrida plot¹⁶ that has become standard in the study of large RBN. The Boolean system with rule 126 studied in this paper is chaotic and the time evolution of its density is also chaotic. We are specifically interested in rule 126 because it is one of the legal rules which can be generalized to higher numbers of neighbors. Moreover, this rule is a simple model of interacting populations of nodes with heterogeneity in the number of interaction partners of each cell. And recut and Ali¹⁷ have provided a mean field equation for the evolution of the time evolution of c of random Boolean networks with rule 126, for the case in which the number of inputs to a node is a constant, k. Matache and Heidel¹⁸ have studied the case with different ratios of neighborhood sizes. A case where the updating is asynchronous was analyzed in Ref. 19. Here we study the system with random distributions of input degree (connectivity), i.e., k is not fixed across the N nodes of the network, but has a normal or Gaussian distribution. Accordingly, the two control parameters of interest are the mean of the input degree, k, and the variance D for the distribution of k across the population of the N nodes. The order parameter of interest which characterizes chaotic behavior at the system level is the density c(t). A mean field equation for the time evolution of the density, as a function of Dand k is given which is a generalization of the case studied in Ref. 17. Bifurcation diagrams and Lyapunov exponents are computed for this Boolean network model and its dynamics is compared with the data obtained from the simulations.

2. Boolean Network Model and the Mean Field Theory

Consider N interconnected cells (nodes) and let cell i has k_i neighbors. Implementing rule 126, the state of cell i at time t + 1, $s_i(t + 1)$ is given by

$$s_i(t+1) = \begin{cases} 0, & \text{if } \sum_{j \in V_i} s_j(t) \mod k_i = 0\\ 1, & \text{otherwise} \end{cases},$$
(1)

where V_i is the neighborhood of the cell *i*. All cells are updated synchronously. The density $c(t) = 1/N \sum_{i=1}^{N} s_i(t)$ is the average of all $s_i(t)$, where i = 1, 2, ..., N and represents the fraction of the cells which are in state "1" at time *t*. The neighborhood V_i is defined by the network topology. The connections are determined before

the simulation and held fixed throughout the temporal evolution, i.e., the network structure is fixed for a particular simulation of the system. In other words, we are considering a quenched and not an annealed random Boolean network model in which the wiring diagram during the time evolution is changed.

2.1. Network construction

Most studies of RBN assume a constant number of inputs for all the nodes, i.e., the neighborhood size is the same for all the cells.^{1,2,10,17} This assumption of homogeneous k reduces complications in the analysis of the network behavior. In such a model each site has k number of neighbors, randomly picked from all the others. In all subsequent discussions, k_i of a node i includes the connection to the node itself, i.e., we deal with semi-local connectivity. In the broader context of variable k_i , a random network with homogeneous k can be said to have a distribution of input connectivity that is a delta function at k. The output connectivity is a Poissonian distribution for large N in such networks.

To construct networks with degree distributions that are not a delta function but exhibit a distribution with a specified mean degree k and variance D, we created a random list of N normally distributed integers with mean k and variance D, and assigned each of them randomly to the nodes and randomly connected the nodes iwith other nodes of the network k_i , which provides node i with k_i inputs. Multiple links between the two nodes are not allowed, while a link to a node itself is always made (semi-local property). We are considering only the semi-local case here. There is no significant difference in the system dynamics studied here when compared to the fully non-local case. For variance D = 0, this construction is equivalent to the random network with constant input connectivity (delta function distribution).

We also considered a version of the classic Erdös-Renyi (ER) random networks.²⁰ Here connections are introduced between all possible pairs of nodes with a specified probability so that the total number of links is Nk. Again, multiple links are not allowed and a link to a node itself is included. The input degree distribution for such a network approaches, like that for the output degree, a Poissonian as the system size N becomes large. The mean degree is k and the variance is also k for such a network.

2.2. Mean Field Approximation

Various collective phenomena and mean field theory of such cases for CA have been presented in Refs. 10 and 12. For a Boolean network with the rule 126 that has constant input connectivity, k, the evolution equation for the mean field is given by:¹⁷

$$c(t+1) = 1 - c(t)^{k} - (1 - c(t))^{k}.$$
(2)

Here the density of the system at time (t + 1), c(t + 1), is given in terms of the density c(t) at time t. The quantity c(t) represents the probability that a cell is in

1464 M. P. Joy, D. E. Ingber & S. Huang

state 1 at time t when $N \to \infty$. This has been generalized by Matache and Heidel¹⁸ for different sizes of neighborhood. When there are M_j cells with k_j inputs, the equation for the probability that a cell is in state 1 at time t + 1 is given by

$$c(t+1) = \sum_{j=1}^{J} \frac{M_j}{N} [1 - c_j(t)c(t)^{k_j - 1} - (1 - c_j(t))(1 - c(t))^{k_j - 1}], \qquad (3)$$

where J is the number of distinct neighborhood sizes and $c_j(t)$ is the probability that a cell with k_j inputs is in state 1, among M_j cells.

When the input connectivity is not a constant we can write the mean field equation by expanding Eq. (2) about the mean and neglecting higher order terms. Thus, one can write the c(t + 1) for a Boolean network with random connectivity distribution as

$$c(t+1) = 1 - c(t)^{k} - (1 - c(t))^{k} - \frac{D}{2} [(\log c(t))^{2} c(t)^{k} + (\log(1 - c(t)))^{2} (1 - c(t))^{k}], \qquad (4)$$

where D is the variance of the distribution of input degree across the nodes and k is its mean. Equation (4) can be obtained using Eq. (3) as well, by approximating the sum over a random distribution and neglecting higher order terms. We denote the right-hand side of Eq. (4) as the function F(c). Then the one-dimensional map of the system is given by,

$$c(t+1) = F(c(t)),$$
 (5)

which has only one maximum in the range $c \in (0, 1)$, and is bounded between 0 and 1. Since we assume in this model that the variance D of the degree distribution is small and finite so that we can neglect higher order terms, Eq. (4) is not a suitable approximation for networks with unbounded degree distributions such as scale free distributions. Note also that when D is very large compared to k, it is not well defined in the range (0, 1).

We now compare the prediction by the mean field model with the simulations of the dynamics of RBNs for different values of the mean input connectivity k and variance D. Representative plots of first and second return maps are presented in Figs. 1–3, for the mean connectivity k = 7. Similar results were obtained for other values of k and D. In Fig. 1(a) the function F and in Fig. 1(b) its second iterate, F(F), are plotted as a function of c(t), for the case D = 0. The calculated density of the network c(t + 1) at time t + 1 and c(t + 2) at time t + 2 as a function of the density c(t) at time t, in Figs. 1(a) and 1(b), respectively, are superimposed. Figure 2 shows the results for the case D = 1. In Fig. 3 the same plots are presented for the map with D = k, i.e., for a Boolean network with ER random graph structure. As evident from the plot of the map functions in Figs. 1–3, the map is a one-humped map, mapping the open unit interval $(0, 1) \rightarrow (0, 1)$. For D = 0, it reproduces the results of Ref. 17.



Fig. 1. Simulation of the Boolean Network and the mean field model. (a) first return map and (b) second return map, for k = 7 and D = 0. Dots represent results from the Boolean network simulations and the solid line using the function F of the mean field model.

Simulations for systems with different values of N gave similar results; the shown results are for N = 2000. Overall, there was a good agreement between the map model and the simulation results although for individual iterations, the mean field theory and simulations do not match exactly. This is not surprising since the mean field model is valid for a large ensemble of networks with $N \to \infty$.



Fig. 2. Same as in Fig. 1, for k = 7 and D = 1.

3. Analysis of the Map Model

To study the asymptotic stable dynamics of the map (5) we plotted bifurcation diagrams for c(t) with k as the control parameter and also calculated the Lyapunov exponent (LE). The LE quantifies the sensitive dependence on initial conditions.



Fig. 3. Same as in Fig. 1, for k = 7 and D = k. Here Boolean network is with an Erdös Renyi random graph structure.

It represents the average rate of the exponential divergence of nearby trajectories. If the LE is positive (the largest LE for higher dimensional systems) the system is chaotic. Calculation of LE, denoted here as λ , can be done by averaging $\log |\partial F/\partial c|$ over several iterations for several different initial conditions:

1468 M. P. Joy, D. E. Ingber & S. Huang

$$\lambda = \frac{1}{T} \sum_{t=1}^{T} \log \left| \frac{\partial F}{\partial c} \right|, \tag{6}$$

where T is the total number of iterations.²¹

The bifurcation diagrams with k as the bifurcation parameter (Figs. 4(a)– 6(a)) were created by plotting a few final iterates of the map after discarding transients, for each set of parameter values k and D. Figure 4(a) shows the bifurcation diagram for D = 0. As the mean connectivity k increases, the map dynamics goes from a single fixed point behavior to a two-periodic behavior. Further increase results in another period doubling to a period-4, and so on. Unlike in the case of CA, with RBN when k reaches a critical value k^* , the network exhibits chaotic behavior with respect to c(t). The system shows a period doubling route to chaos. The corresponding Lyapunov exponents as a function of k are shown in Fig. 4(b). For $k^* \approx 4.595$, LE becomes positive indicating that the map is chaotic with this value of k, in agreement with the bifurcation diagram in Fig. 4(a). Several windows of periodic behaviors are present in the bifurcation diagram after the onset of chaos as k increases, and accordingly, at these values of k, the LE is less than zero (Fig. 4(b)).

For the relevant ranges of parameters this is a unimodal map in the unit interval, and hence, it shows all the universality properties of such maps which are well



Fig. 4. (a) Bifurcation diagram and (b) Lyapunov exponents, λ , of the mean field model, for D = 0 and k = 2 to 25. For bifurcation diagrams we plot the 500 final iterates of the map, for each value of k. and for LE we take the average over 10000 iterations and many random initial conditions.



Fig. 5. (a) Bifurcation diagram and (b) Lyapunov exponents of the mean field model, for D = 1 and k = 2 to 25.



Fig. 6. (a) Bifurcation diagram and (b) Lyapunov exponents of the mean field model, for D = k and k = 2 to 25.

studied.²² Although one could increase the value of k indefinitely, due to finite precision of computers, not much relevant information can be extracted for larger k since the dynamics will appear as invariant under further change of k.



Fig. 7. Lyapunov exponent, λ , of the model in the (k, D) plane. Color represents the value of λ as denoted in the color bar.

Figure 5 shows the bifurcation diagram (Fig. 5(a)) of the system (5) for D = 1, with the corresponding LE plot (Fig. 5(b)), for k = 2 to 25. The result indicates that as D increases, the onset of chaos is delayed with respect to the value k. In this case, the chaotic regime begins at $k^* \approx 5.075$. Thus, one may argue that heterogeneity in network topology tend to maintain order in the system. Figure 6 presents similar plots as Figs. 4 and 5, but for k = D. The latter case corresponds to a distribution where the mean and variance are the same, such as in a Poisson distribution, and represents the Erdös-Renyi random graph. Simulation of the ER network also shows a good agreement with theory. For this case the chaos onset occurs when $k^* \approx 6.45$.

To illustrate the dependence of chaos onset on the two control parameters, Fig. 7, shows a plot of λ , in the (k, D) space. This figure clearly indicates that chaos onset occurs for higher value of k as we increase D. This is directly evident in Fig. 8, where k^* , is plotted as a function of D.

4. Conclusion

In this paper we presented a mean field approximation for the time evolution of the density, a macroscopic variable used to characterize system behavior of a network. We focus on the case of a random Boolean network with uniform functions, rule 126, but with varying input degrees which exhibit distributions with a specified mean and variance. Mean field equation agreed well with the simulation results of the model. The evolution of the mean field shows a period doubling route to chaos, as demonstrated in the bifurcation analysis and by calculating the Lyapunov exponents.



Fig. 8. Critical value k^* , at which the onset of chaos occurs as a function of the variance D.

These results suggest that even simple networks can produce chaotic behavior. Specifically, in the case of random networks with a simple rule for interactions the increasing average input degree of a node can lead to chaotic behavior. But the results also indicate that the extent of inhomogeneity with respect to input degree k, as reflected in the variance D, suppresses the onset of chaos. In other words, inhomogeneous connectivity may dampen the chaotic behavior of the Boolean network.

The model system used here was originally motivated by previous results on RBN with uniform input connectivity. The order parameter of the system, density c(t) is an aggregate quantity of the network states, and the Boolean function of nodes were uniform across the network, namely obeying rule 126. This model offers some general insights for how a class of relatively simple systems can generate complex behavior. The density c(t) may represent a global gene expression activity of a cell across the entire transcriptome of thousands of individual genes. These results suggest that such global activities may not "average out" even if it is contributed by thousands of genes. In fact, for instance, the cellular redox state has been shown to exhibit periodic fluctuations although it is contributed by genome-scale number of genes.²³ It will be interesting to also look at other generalizable CA rules with complex behavior, such as rule 22, in the context of RBN. Restricting the analysis to uniform rules enabled us to differentiate between complex behavior arising from the interaction structure (network topology) from that due to randomness in the choice of rules used as in conventional RBN.^{3,8,24}

Our result suggests that — given the limitation due to the Boolean function used — not simply the average connectivity of each gene but also the heterogeneity of input numbers among the genes, captured in the variance D, is a relevant parameter that influences global dynamics. This heterogeneity, which is ubiquitous in biological systems, may help suppress chaotic behavior. This would be consistent with the idea that biological networks operate in the 'ordered regime' (near the phase transition to chaos) which in turn provides high-dimensional stability to transcriptomes that may encode stable cell phenotypes. However, in molecular networks this heterogeneity is manifested in the form of scale-free topology.⁹ Also, most of the real networks show small-world property.²⁵ Thus, it is important to note that the mean field theory presented here is not adequate to describe scale free networks, because of the unbounded variance of degree distributions in such networks. However, our studies provide an indication that such systems may in general be less chaotic. Aldana and Kluzel⁸ recently demonstrated that scale-free networks with random Boolean functions are less chaotic and that the transition to chaos occurs below the power law exponent, $\gamma = 2.5$. Future studies on the role of degree heterogeneity needs to be extended on scale free networks, small-world networks and other random networks with various topological features.

Acknowledgments

We acknowledge the receipt of financial support from ARO (Army Research Office) project No. W911NF-04-1-0273 and AFOSR (Air Force Office of Scientific Research) Grant FA 9550-05-1-0078.

References

- 1. S. Wolfram, A New Kind of Science (Wolfram Media, Champaign, 2002).
- 2. S. A. Kauffman, J. Theor. Biol. 22, 437 (1969).
- 3. S. A. Kauffman, The Origins of Order (Oxford Univ. Press, Oxford, 1993).
- 4. J. A. de Sales, M. L. Martins and D. A. Stariolo, Phys. Rev. E 55, 3262 (1997).
- 5. S. Huang and D. E. Ingber, Exp. Cell Res. 261, 91 (2000).
- S. Huang, G. Eichler, Y. Bar-Yam and D. E. Ingber, *Phys. Rev. Lett.* 94, 128701 (2005).
- 7. S. Wolfram, Rev. Mod. Phys. 55, 601 (1983).
- 8. M. Aldana and P. Cluzel, Proc. Natl. Acad. Sci. USA 100, 8710 (2003).
- 9. R. Albert and A. L. Barabasi, Rev. Mod. Phys. 74, 47 (2002).
- 10. H. Chate and P. Manneville, Prog. Theor. Phys. 87, 1 (1992).
- 11. W. Li, J. Stat. Phys. 68, 829 (1992).
- 12. H. A. Gutowitz, J. D. Victor and B. W. Knight, Physica D 28, 18 (1987).
- 13. N. Boccara and M. Roger, J. Phys. A 25, L1009 (1992).
- 14. N. Boccara, O. Roblin and M. Roger, J. Phys. A 27, 8039 (1994).
- 15. N. Mousseau, Europhys. Lett. 33, 509 (1996).
- 16. B. Derrida and Y. Pomeau, *Europhys. Lett.* 1, 45 (1986).
- 17. M. Andrecut and M. K. Ali, Int. J. Mod. Phys. B 15, 17 (2001).
- 18. M. T. Matache and J. Heidel, *Phys. Rev. E* **69**, 056214 (2004).
- 19. M. T. Matache and J. Heidel, Phys. Rev. E 71, 026232 (2005).
- 20. B. Bollobas, Random Graphs (Academic Press, London, 1985).
- 21. J. C. Sprott, Chaos and Time-Series Analysis (Oxford Univ. Press, Oxford, 2003).
- 22. M. J. Feigenbaum, J. Stat. Phys. 21, 669 (1979).

- R. R. Klevecz, J. Bolen, G. Forrest and D. B. Murray, Proc. Natl. Acad. Sci. USA 101, 1200 (2004).
- 24. I. Shmulevich and S. A. Kauffman, Phys. Rev. Lett. 93, 048701 (2004).
- 25. D. J. Watts and S. H. Strogatz, Nature 393, 440 (1998).