SANDIA REPORT

SAND2004-4185
Unlimited Release
Printed September 2004

Stability of Biological Networks as Represented in Random Boolean Nets

Alexander Slepoy, Marshall Thompson

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/ordering.htm

SAND2004-4185
Unlimited Release
Printed September 2004

Stability of Biological Networks as Represented in Random Boolean Nets

Alexander Slepoy and Marshall Thompson
Computational Materials and Molecular Biology
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0310

Abstract

We explore stability of Random Boolean Networks as a model of biological interaction networks. We introduce surface-to-volume ratio as a measure of stability of the network. Surface is defined as the set of states within a basin of attraction that maps outside the basin by a bit-flip operation. Volume is defined as the total number of states in the basin. We report development of an object-oriented Boolean network analysis code (Attract) to investigate the structure of stable vs. unstable networks. We find two distinct types of stable networks. The first type is the nearly trivial stable network with a few basins of attraction. The second type contains many basins. We conclude that second type stable networks are extremely rare.

This page intentionally left blank.

Contents

Figures

This page intentionally left blank.

Stability of Biological Networks as Represented in Random Boolean Nets

1 Biological Networks

Network formalism is particularly useful for representing relationships in complex systems. It is not surprising therefore that it has become an integral part of the theoretical investigations in systems biology. Complex highly-connected problems abound in living organisms: protein interaction networks, regulatory networks, evolutionary trees, chemical reaction networks, predator-prey networks all compose a small subset of biological applications.

Biological high-throughput techniques, such as micro-array experiments and gene knock-out studies, are generating large amounts of experimental data. Making sense of this information requires a better understanding of complex system dynamics.

It has long been believed that key properties of the system dynamics can be associated with the structure of the interaction network alone, without considering finer details of the interactions [1, 2]. Network representation is compact and permits rapid simulation. Ability to simulate long-time dynamics of an entire cell or an ecology may lead to greater understanding of such complex problems on the scales not otherwise accessible.

2 Random Boolean Networks

Random Boolean Networks (RBNs), also known as N-k network, belong to a class of random disordered networks proposed by Stuart Kauffman in 1969 as a model for genetic regulatory networks [1]. The time-dependent state of the system is a set of N binary variables $\sigma(t) = \{\sigma_i(t); \sigma_i = \pm 1\}$, equivalent to an Ising state[3]. The state evolves in discrete time steps by setting $\sigma_i(t)$ according to a Boolean rule $f(\sigma(t-1))$ based on a state of k other variables at time $t-1$. Such rules are randomly picked for each σ_i. Connections of variables to their k antecedents can be viewed as edges, in which case the rule set forms a network, where nodes are the state variables σ_i. The enumerated space of the possible RBN grows rapidly with N and k, since one can form $\binom{N}{k}^N$ graphs, with rule multiplicity of 2^{2^k} at each node.

The state space is finite consisting of 2^N possible states. The dynamic deterministically connects a state to its child state. If a dynamic step is viewed as an edge, and states appear as nodes, the resulting graph is a *state evolution diagram* (SED). For a given network, a SED is a tree that is rooted in a cycle. A cycle is composed of states that periodically repeat during the system evolution. The leaves of the tree (so called Garden-of-Eden states) are states that have no parent states. Each tree is called a basin of attraction. In fact, the state space is usually partitioned into several basins of attraction, each built of transients terminating in the common cycle. Figure 1 shows an example of several such state evolution diagrams.

Previous work investigates relationships between the overall structure of RBNs and some of their properties. Kauffman's original paper describes two distinct dynamic behaviors, the chaotic phase and the frozen phase. In chaotic networks, the number of attractors grows exponentially with the network size, while the number of attractors tends to stay fixed in frozen networks. Kauffman also described a transition phase, separating the chaotic and frozen dynamics, and observed many properties of the transition phase that parallel biological systems [1].

The dynamics of frozen phase networks were analyzed in detail at the level of $k = 1$ by Flyvbjerg and Kjær. They developed a set of analytical equations to predict the distribution of sizes of attraction basins in the infinite network size limit. In doing so, they also introduced the concept of relevant elements in a network [4]. Derrida and Pomeau then proposed the annealed approximation for RBNs, a statistical method for analysis of chaotic phase dynamics. Their methods proved effective for predicting cycle length distributions, number of attractors, and the values of N and k that cause a network to exhibit the transitional phase behavior [5].

Following this work, Bastolla and Parisi explored the limits of the annealed approximation. They formalized the definition of relevant elements and introduced some coarse reduction procedures to enable simulation of much larger networks [6, 7]. Bilke and Sjunnesson later expanded on the network simplification schemes proposed previously, developing an algorithm for identification and removal of the stable and non-influential elements of a network [8]. This development enabled full state space enumeration of networks as large as $N = 32$. The ability to fully examine these large networks lead to the discovery that the number of attractors in large transitional phase networks scales linearly with N, contrary to Kauffman's original findings, that the number of attractors grows as $\sqrt{N}$.

3 Stability of RBN

Stability of complex systems remains a subject of heated debate. Early work on RBN models demonstrated the onset of chaos at a critical connectivity $k = 2$. However, many biological systems, including protein interaction networks, exhibit connectivity that is significantly higher then 2 and yet retain high stability characteristics. We hope to address the subject of stability in our analysis.

We use and expand a notion of stability, originally proposed by Kauffman [1]. It is based on an idea that a most common perturbation to the system takes the form of a bit-flip. In such a case, a basin of attraction that maps into itself via all possible bit-flips for each of its member states is ultimately stable. Of course, such an extreme case is only possible when the system has only a single attractor. In fact, systems with very few attractors are almost trivially stable. Most generally, a basin may have a collection of states that map inside the basin by all bit-flips, some states that map both inside and out, and some states that map outside the basin only.

A sub-class of this may be defined when we consider all states that uniformly map inside the basin (through every possible bit-flip of a state) as volume states, while all the states that can be mapped outside the basin via a bit-flip, as surface states. This implies the hard definition of a stable state as a state that uniformly maps inside the basin due to any bit-flip. In such a case, the ratio of the surface states to the volume ones (SVR) can be viewed as a hard stability measure. A softer stability measure can be derived by redefining a stable state as one that maps inside the basin through some large fraction f of the available bit-flips. We would like to locate networks with stable characteristics by sampling, and understand the network properties associated with stability.

4 *Attract*: Object-Oriented Software Package for RBN Generation and Analysis

To conduct the search for stable networks, we have developed a software suite of analytic tools that investigates network dynamics in the context of RBN. *Attract* is a modular, object-oriented

application written in the C++ programming language that allows generation of random or custom boolean networks, simulates their time-state evolution, and provides a variety of static and dynamic analysis tools. Network connections may be generated randomly, with block diagonal connections, or may be read from a file. Rules may be selected at random, selected with a weighted probability of selecting a 1 or 0 for a particular rule, or read from a file. In addition to single network and rule generation, *Attract* contains various options for connection and rule sampling and enumeration. Sampling of connections and rules can be either directly executed over a given number of samples, or until the running averages of the surface-to-volume ratios of the basins of attraction converge to within a given percentage.

Attract also contains tools for analysis of the various statistics collected throughout the simulation. Quantities that the application can calculate include the SVR of each basin of attraction in its entirety, the SVR of the attractor cycles themselves, the percentage of random bit flips throughout the entire basin of attraction that result in states within the same basin, the percentage of random bit flips in the attractor cycle alone that result in states within the same basin [11], and the distribution of the relative sizes of all basins of attraction analyzed. All of these statistics can be printed to the screen or to a file in the form of a matrix, a list of ratios (for the surface/volume values), a human readable summary, or the various statistics can be written to *gnuplot* [10] format histogram data files. *Attract* can also print any stable networks (more than 2 basins of attraction, each with SVR < 1) that are found.

Our application has been developed for both parallel and serial execution. Distribution of the simulation across a parallel system allows a very large number of connection and rule samples to be examined, an ability that has proven invaluable given the extremely rapid growth of the rule space. The speed of simulation scales nearly perfectly with additional processors, as our initial parallel algorithm is a simple distribution of a network sampling job over all processors, and thus requires almost no communication between processors. Several additional scripts in the *Attract* package allow rapid and automatic graphing of histograms, network connections, and state progressions in *dot*, *neato* [9], and *gnuplot* [10].

We have tested the accuracy of our software by comparisons with existing literature on analytical properties of Kauffman networks [4]. For large networks with $k = 1$, the normalized basin weights obtained by our simulations satisfactorily approach the values given in the literature for networks of infinite size. This preliminary verification of *Attract* is sufficient for our present purposes.

5 Preliminary Results

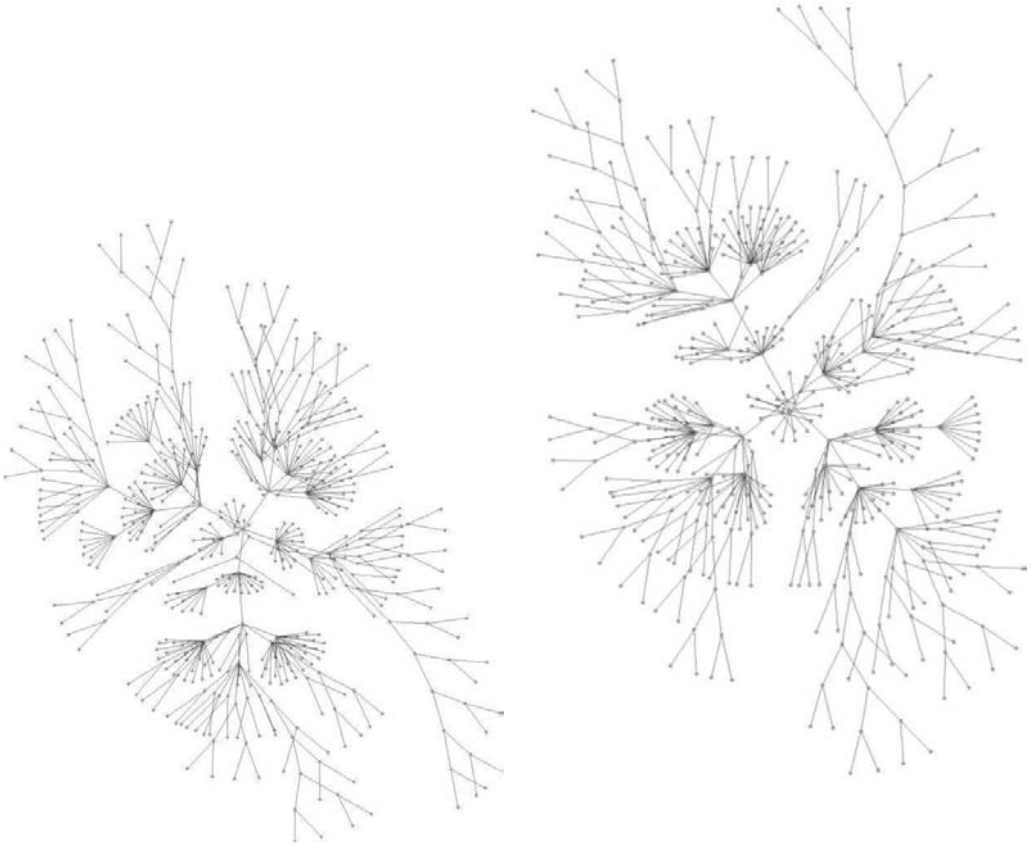
We used an extremely weak stability measure for the overall network, requiring at least one stable state per basin. Nevertheless, we find only a small number of networks stable according to this very weak definition. We have sampled 1×10^6 networks (out of a possible 7.49×10^{44}) of $N = 10$ at $k = 3$ for our preliminary analysis. Each sampling includes a complete enumeration of basins with the analysis of SVR. We exclude from consideration those networks in which the stability occurs due to a small number of basins, $n_b < 3$. For the purposes of this preliminary study, we have considered a network stable if each of its basins has $SVR < 1$; that is, each basin contains at least one state that is not a surface state. Of the networks sampled, 140 stable networks were found according to the hard stability requirement, and 8,186 stable networks found according to the soft stability requirement, where a non-surface state is defined as one in which the fraction of the state bit-flips that map outside the basin must be less than 10%.

Examination of the SVRs obtained from our first samples motivated a subsequent sampling of $N = 10$, $k = 2$ networks. To preclude the possibility that the stability in these networks was the result of too few basins instead of some inherent structural characteristic, networks with number of basins $n_b < 3$ were again excluded from consideration. Of the 1×10^6 networks sampled in this case (of a possible 3.74×10^{28}), 272 stable networks were found using the hard stability measure. With the soft stability requirement (again with the maximum fraction of surface-mapping bit-flips at 10%) 9,230 stable networks were found. As expected, the $k = 2$ sample located more stable networks than the $k = 3$ sample. This is due to the fact that smaller values of k drive network dynamics toward the frozen phase, where networks tend to be less dynamic and have much larger stable cores.

6 Future Directions

Since we have been able to locate a small number of networks that satisfy our stability criterion, we plan to take a closer look at their properties in an attempt to identify the causes for this rare condition. We plan to add static analysis tools for the network graph, as well as tools for functional analysis of the rules. The static analysis toolbox will include graph partition algorithms, loop identification and counting, and coloring methods. Current scripting interface will be converted to a graphical user interface to assist in building, running, and analysis of the networks. We will conduct further verification and testing procedures, and continue the parallelization effort to permit a distributed processing mode of a single network. This will lead to analysis of larger networks, which up to now have been prohibitive.

We hope to reach a maturity of understanding that will permit us to apply our analytic capability to the networks reverse engineered from biological systems via micro-array experiments. Currently, data of this sort is available, and we plan to use it to examine the stability properties of such networks.



(a) **(b)**

Figure 1. State evolution diagrams of an $N = 10$ network. The points in these graphs each represent a fixed state of the network. As each state is evolved one time step, an edge is created from the original state to the new state. Thus, a tree structure emerges, with the attractor cycle as the root and various transient states as radial branches. The state of the system can only travel in one direction (towards the center cycle of the graph). Each individual tree represents a single basin of attraction, as the deterministic nature of RBNs requires that each basin be completely disjoint of all other basins.

References

[1] Kauffman, S. A. (1969). Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets. *J. Theor. Biol., 22*, 437-467.

[2] Erdos P. & Renyi A. (1959). On random graphs. *Publicationes Mathematicae, 6*, 290-297.

[3] Ising E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift fr Physik, 31*, 253-258.

[4] Flyvbjerg, H., & Kjær, N. J. (1988). Exact Solution of Kauffman's Model with Connectivity One. *J. Phys. A. Math. Gen., 21*, 1695-1718.

[5] Derrida, B., & Pomeau, Y. (1986). Random Networks of Automata: A Simple Annealed Approximation. *Europhys. Lett., 1*, 45-49.

[6] Bastolla, U., & Parisi, G. (1998). Relevant Elements, Magnetization and Dynamical Properties in Kauffman Networks: a Numerical Study. *Physica D, 115*, 203-218.

[7] Bastolla, U., & Parisi, G. (1998). The Modular Structure of Kauffman Networks. *Physica D, 115*, 219-233.

[8] Bilke, S., & Sjunnesson, F. (2002). Stability of the Kauffman Model. *Phys. Rev. E, 65*, 016129-1 - 016129-5.

[9] Gansner, E. R., & North, S. C. (2000). An open graph visualization system and its applications to software engineering. *Software–Practice & Experience, 30*, 1203-1233.

[10] Gnuplot. http://www.gnuplot.info/.

[11] Wuensche,A. (2002). Basins of Attraction in Network Dynamics: A Conceptual Framework for Biomolecular Networks. In G.Schlosser and G.P.Wagner (Eds.), *Modularity in Development and Evolution* (pp. 288-311). Chicago: Chicago University Press.

DISTRIBUTION:

1 The University of New Mexico
Computer Science Department
Attn: Cristopher Moore
Farris Engineering Center,
Room 157
Albuquerque, NM 87131

7 MS 0196
M.A. Thompson, 9235

1 MS 0310
S.B. Martin, 9212

1 MS 0310
M.D. Rintoul, 9212

1 MS 0310
J.B. Aidun, 9235

5 MS 0310
A. Slepoy, 9235

1 MS 0316
P. Yarrington, 9230

1 MS 0318
G.S. Davidson, 9200

1 MS 0318
J.E. Nelson, 9209

1 MS 0321
W.J. Camp, 9200

1 MS 0321
R.W. Leland, 9220

1 MS 0370
S.A. Mitchell, 9211

1 MS 0885
G.S. Heffelfinger, 1802

1 MS 1109
R.J. Pryor, 9216

1 MS 1110
D.E. Womble, 9210

1 MS 1110
S.J. Plimpton, 9212

1 MS 1110
S.L.K. Rountree, 9215

1 MS 1111
B.A. Hendrickson, 9214

1 MS 1413
P.V. Dressendorfer, 1141

1 MS 9951
J.M. Faulon, 9212

3 MS 9018
Central Technical Files, 8945-1

2 MS 0899
Technical Library, 9616