

Structural Circuits and Attractors in Kauffman Networks

K.A. Hawick, H.A. James and C.J. Scogings
 Computer Science, Institute for Information and Mathematical Sciences,
 Massey University,
 Albany, North Shore 102-904, Auckland, New Zealand
 {k.a.hawick, h.a.james, c.scogings}@massey.ac.nz
 Tel: +64 9 414 0800 Fax: +64 9 441 8181

Abstract

There has been some ambiguity about the growth of attractors in Kauffman networks with network size. Some recent work has linked this to the role and growth of circuits or loops of boolean variables. Using numerical methods we have investigated the growth of structural circuits in Kauffman networks and suggest that the exponential growth in the number of structural circuits places a lower bound on the complexity of the growth of boolean dependency loops and hence of the number of attractors. We use a fast and exact circuit enumeration method that does not rely on sampling trajectories. We also explore the role of structural self-edges, or self-inputs in the NK-model, and how they affect the number of structural circuits and hence of attractors.

Keywords: Kauffman networks; Random boolean functions; Circuit enumeration; Loops; Attractors.

1 Introduction

Random Boolean Network (RBN) models are effectively a generalisation of the 1-dimensional Cellular Automata model [1]. Kauffman's NK-Model [2] of an N-node network with K-inputs to a boolean function residing on each node has found an important role in the study of complex network properties. RBNs have found important applications in biological gene regulatory networks [3] but also in more diverse areas such as quantum gravity through their relationship with ϕ^3 -networks [4, 5]. RBNs have many interesting properties [6] and have been amenable to various analyses [7] including mean-field theory. They continue to be an important and interesting tool in studying biological and artificial life problems.

A key property of RBNs is the now well established existence of a frozen phase and a chaotic phase [8, 9] and the critical phase transition lies at the integer value of connectivity $K_c = \frac{1}{2p(1-p)} = 2$ for unbiased networks with a mean boolean function output value of $p = 0.5$. It is therefore of most interest to study RBNs at or around this critical value.

The Random Boolean Network or graph G is expressed as a four-tuple $\mathbf{G} = (V, E, F, x)$ and has $N = |V| = |F| = |x|$ nodes or vertices, and $N_e = |E|$ directed edges or arcs, which express the K_i inputs for node i . The Kauffman NK-Network is constructed with fixed $K = 1, 2, 3, ..$ and a boolean function f_i of K_i inputs is assigned to each node. All the nodes of the network carry boolean variables x_i which may be initialised

randomly and which are updated (usually, but not necessarily) synchronously so that:

$$x_i(t) \leftarrow f_i(x_j(t-1)), j = 1, 2, \dots, K_i \quad (1)$$

The NK-network model assigns the K_i inputs for node i randomly and with uniform probability across all nodes. Even for a large network there is still a non-zero probability of assigning a node as one of its own inputs. In the case of $K_i > 1$ there is also a possibility of assigning a node j as an input of i more than once. These self-edges or multiple edges can have a subtle effect on the behaviour of the NK-network model.

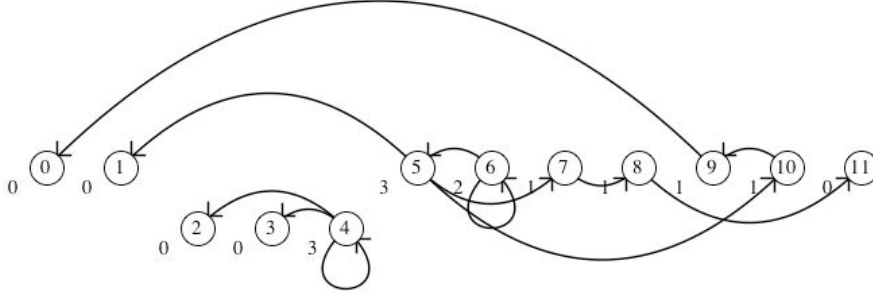


Figure 1: 12 Node Network with $K = 1$ inputs, showing the output degree of each node.

Figure 1 shows a small example Random Boolean Network of 12 nodes, each with $K = 1$ inputs. The out-degree varies depending upon the number of other nodes that depend upon each node, and for this network self-arcs are allowed. This example is fragmented into two independent components and there are interesting structural changes in the component composition of RBNs between $K = 1$ and $K = 2$.

Work has been carried out on a number of different update mechanisms for boolean networks including asynchronous algorithms [10]. In this paper we consider only synchronous updating where all nodes execute their boolean function once, together, and at every time step. A significant body of work has now been carried out on the roles of different sub-classes of boolean functions including the so called canalizing functions [11] and in particular the effect of the frozen or fixed-value boolean functions on particular elements of the network.

A consequence of the boolean functions in RBNs is the formation of attractor basins [12]. These are observed in RBN models whereby diverse initial starting conditions will still lead to statistically similar behaviour. The state of the network falls into attractor cycles whereby a chain of interdependence of nodes (via their boolean functions) leads to the network periodically repeating its state. The number and length of these periods or attractors is of great importance in understanding the behaviour of the NK-model and associated application problems. This can be seen quantitatively by tracking a metric such as changes in the normalised Hamming distance between the network’s successive boolean states.

Of particular recent interest in the literature has been the uncertainty concerning the number of attractors [13, 14] and how their number and lengths varied with the size of the network. Scaling was initially believed to be $O(\sqrt{N})$ [15]. It was later reported as linear [16], and then as “faster than linear” [17] and subsequently as “stretched exponential” in [18, 19] but is now known to be faster than any power law [20].

A recent review of the RBN model [7] discusses the attractor behaviour in terms of the loops of boolean variable states that form and several exact results concerning these loops have been obtained [21]. Important observations concern the distribution of components with particular sub-classes of possible boolean functions. These “relevant elements” are defined as those nodes that are not frozen and that control at least one other relevant element in the system [19]. A number of important results have been obtained using particular sub-classes of the possible boolean functions. Drossel et al. have considered networks with only non-fixed

boolean functions thus making all elements relevant and have therefore shown the equivalence of $K = 1$ and $K = 2$ networks under appropriate restrictions on the boolean functions used.

In this paper we use numerical methods to investigate the role that structural circuits play in the complex structure of the network and the resulting attractor behaviour of RBNs. Recent work in the literature has used trajectory sampling. The combinatorics of RBN models means that the number of boolean functions grows as 2^{2^K} and a rapid growth in the number of possible network states with network size. Taking limited numbers of sample trajectories through this state space can lead to very misleading results. Numerical sub-sampling of attractor trajectories seems to be the main difficulty behind obtaining a good understanding of attractor scaling. We focus on the structural properties of RBNs including the number and length distribution of elementary circuits and of components. We compute these properties exactly using brute force enumeration techniques for a range of network sizes and connectivities. Our statistical sampling is only over different randomly configured networks, not over attractor trajectories.

In [22] we described the D Code we developed to simulate very large-scale Random Boolean Networks. In this paper we exploit this capability to study the cluster and monomer populations in large systems for which even an $O(n^2)$ cluster labelling algorithm is feasible. However, in the study of circuits we are severely limited by the time complexity of even the best circuit enumeration algorithm we have been able to find (see section 3).

As we discuss in section 2, there is a close relationship between the number of elementary circuits in the underpinning structural network and the number of attractors that can be supported in an associated RBN. We also analyse the role that self-edges play in the structural networks and the consequences for both the number of circuits and disconnected components. We present some results for various network sizes and connectivities, with and without self-edges in section 4 and draw some conclusions on scaling with N and K in section 5.

2 Attractor Numbers

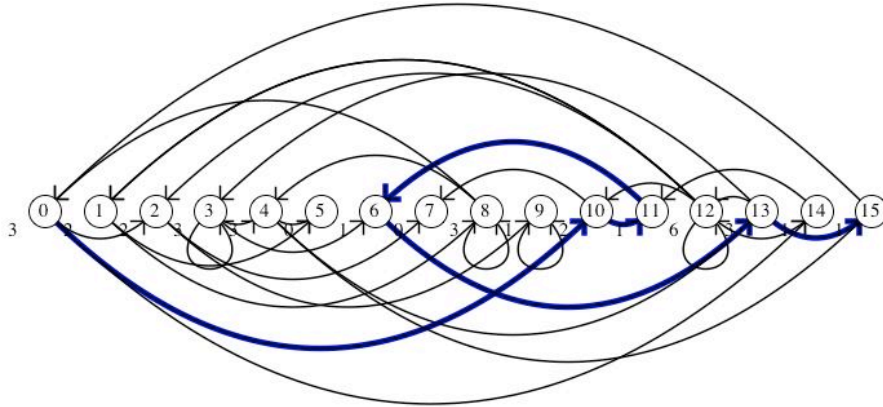


Figure 2: 16 Node Network with $K = 2$ inputs, showing the output degree of each node and one of the circuits in the graph, connecting node 0 to node 15. This network allows self-arcs.

Figure 2 shows a network with $N = 16$, $K = 2$. The construction algorithm has allowed self-arcs - in other words the inputs for each node have been chosen according to a flat uniform distribution so they can connect to themselves. The consequent self-edges allow self-inputs in the corresponding RBN. These are known to play a vital role in supporting the number of attractors. A self-input or “self-ancestor” in the input dependence

chain of boolean variables anchors the periodic or attractor behaviour [7] of RBNs.

We felt intuitively that the presence of structural circuits would also be vital to the periodic attractor behaviour. Figure 2 shows one such circuit or loop in the network structure. In fact, exact enumeration (as shown in figure 3) indicates that there are 22 circuits composed as follows: 4 of length 1; 2 of length 3; 4 of length 5; 2 of length 6; 6 of length 8; and 4 of length 10. If self-edges are disallowed we would obtain a higher number of circuits present in the network.

0	10	11	6	13	3	4	15	0		1	8	4	13	12	1	
0	10	11	6	13	12	1	8	0		1	8	4	13	12	1	
0	10	11	6	13	12	1	8	4	15	0	3	3				
0	10	11	6	13	12	1	8	0		3	4	13	3			
0	10	11	6	13	12	1	8	4	15	0	3	6	13	3		
0	10	11	6	13	15	0					6	13	12	10	11	6
0	14	11	6	13	3	4	15	0		6	13	12	14	11	6	
0	14	11	6	13	12	1	8	0		8	8					
0	14	11	6	13	12	1	8	4	15	0	9	9				
0	14	11	6	13	12	1	8	0		12	12					
0	14	11	6	13	12	1	8	4	15	0						
0	14	11	6	13	15	0										

Figure 3: 22 Circuits found in the network shown in figure 2 which has 16 nodes and 32 arcs and allows self-arcs. Note there are repeated circuits due to the presence of a multiple-arc connecting nodes 12 and 1.

As Drossel et al. have shown there are definite relationships between the number of attractors and the number of loops. Qualitatively summarizing, the number of structural circuits provides a lower bound on the number of possible attractors. It therefore gives insight into the controversy over the number of attractors in RBNs to consider the exactly enumerated number and distribution of circuits in the underlying networks.

An elementary circuit is a closed path along a subset of the edges of the graph such that no node, apart from the first and last, appears twice. The number of elementary circuits for a fully connected graph is bounded by $\sum_{i=1}^{N-1} C_{N-i+1}^N (N-i)!$, [23]. This expression represents the limit for the number of structural circuits in an NK-network when $K \rightarrow N$.

3 Graph Algorithms

RBNs can be represented in a number of different data structures in computer simulation programs. We used a neighbour list approach [24] and have experimented with various structures for boolean variables and boolean functions using the D programming language [22]. D is essentially a systems-oriented programming language derived from C and C++. It is a good platform for custom simulation codes that must be efficient in both time and space to tackle problems with high computational complexity.

The problem of component labelling or clustering is well known and we used a simple colour propagation algorithm [25] which was readily re-implemented as part of our custom RBN code.

Various algorithms have been formulated to count the circuits in a graph but these either use infeasible amounts of memory or are time exponential [26, 27] with a time bound of $O(N.e(c+1))$. We count circuits using a variation of Johnson's algorithm [28] implemented in D. For graphs of N vertices, e edges, c circuits and 1 fully connected component, Johnson's algorithm is time bounded in time by $O((N+e)(c+1))$ and space bounded by $O(N+e)$. Unlike Johnson's algorithm our code copes with partially connected graphs without resorting to the need to treat each of the possible $N_c > 1$ components separately [29]. This is still a

highly expensive process since the number of circuits c itself grows very rapidly with (N, e) .

In the graph literature the term loop is unfortunately sometimes used to describe a self-edge or a circuit of length 1. In the NK-networks we study the number of self-edges is much less than N , even for low K . However we do count them and observe the effect of allowing them in the number of possible circuits and their length histograms.

On a modern compute server with 4GBytes memory and a speed of 2.66GHz, we found it was entirely feasible to enumerate circuits exactly in networks of up to $N \approx 100$ for $K = 1, 2$. Smaller networks were required for higher K . We were able to count components quite easily up to networks of around $N \approx 20,000$. We were able to exploit the near-linear speed-up of parallel job-farming to average our exact enumeration/counting results over many independently generated networks.

4 Numerical Results

In this section we present results for the number of component clusters; the number of monomers and the number and length distribution of circuits. The sample numbers are shown but typically these are over 100,000 independently generated NK-networks for $K=1,2$ and 100 for $K=3,4$ and 10 for $K=5,6$ and higher.

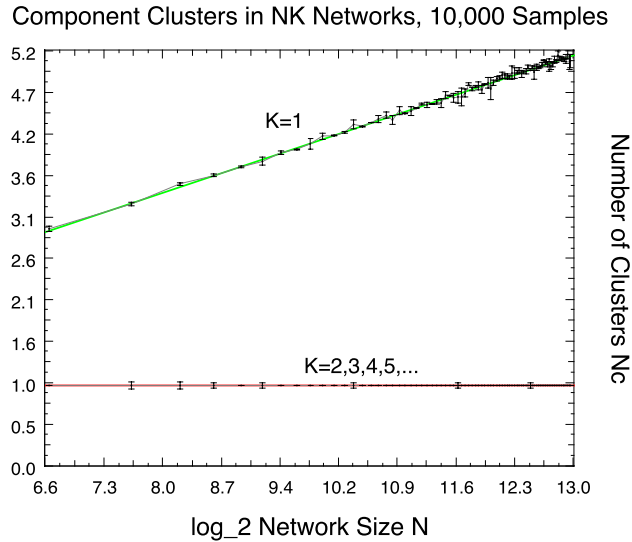


Figure 4: Number of Cluster Components for Kauffman Networks of $K = 1, 2, 3, 4, 5, 6$.

Figure 4 shows the variation of the mean number of clusters in NK-networks. For $K \geq 2$ the system is dominated by a single giant cluster, although even in large networks it is still possible for a single isolated monomer to exist. The figure shows that for $K = 1$ that the number of component clusters $N_c \approx 0.345 \log_2 N + 0.65$, showing that a range of different cluster sizes are co-existing even in arbitrarily large networks.

It is instructive to consider what size distribution makes up the number of components. Figure 5 shows the cluster size distribution averaged over one million sample networks of size $N = 256$. Interestingly this shows a relationship between the mean population $\langle P \rangle$ for cluster component size s such that $\log \langle P \rangle \approx As^{-0.89}$. This relationship appears to hold true for arbitrarily large network sizes, but only for sizes up to one half of the total network size. For $K > 1$ the size distribution is completely dull - being just one single cluster of size N .

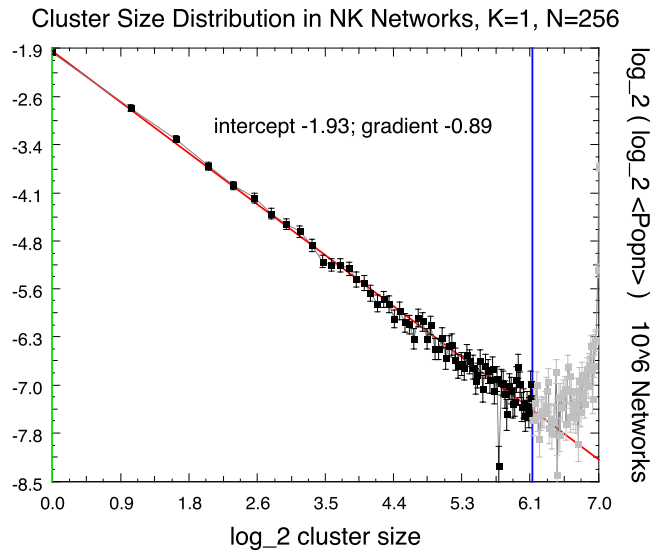


Figure 5: The cluster size distribution for a $K = 1$ network, generated using the algorithm described. Using $N = 256$ and 1,000,000 sample networks. Note the plot does not show the single giant component that occurs for $K > 1$.

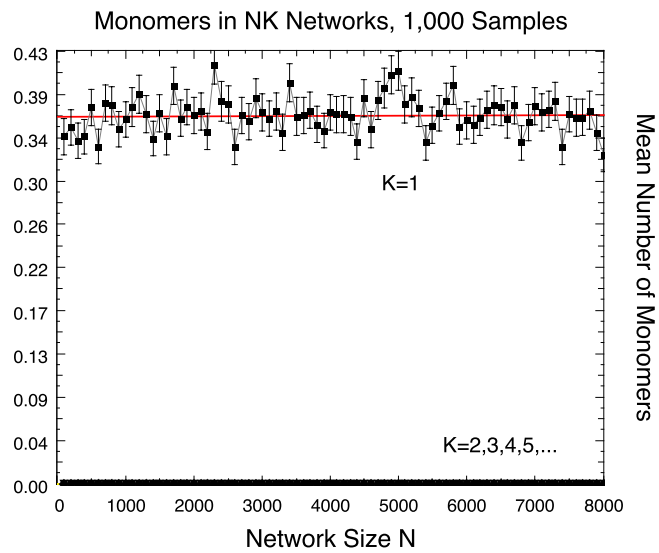


Figure 6: Number of Monomers for Kauffman Networks of $K = 1, 2, 3, 4, 5, 6$.

One might expect that in those regimes where there are multiple components, single isolated monomers or other very small-sized components dominate the component distribution. However the number of monomers is almost entirely flat, independent of the network size. In the case of $K = 1$ it is small but definitely non-zero. In the case of $K > 1$ the number of monomers is almost completely zero on average. Figure 6 shows the variation in the number of monomers over the same networks sampled for components in figures 4 and 5. This suggests an interesting cluster size multi-scale behaviour for $K = 1$ and in particular that the cluster composition is influenced by two competing effects. As the network grows in size there are more nodes and therefore there is a growing possibility of some being disconnected, but conversely there are also more arcs available in the network and therefore a higher probability of each node being connected to some other node. The number of monomers is affected surprisingly little by whether the network is allowed to have self-edges or not.

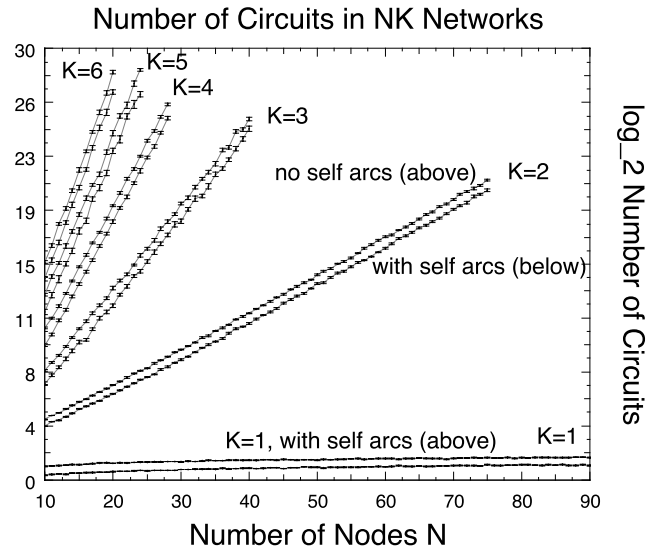


Figure 7: Growth of the number of circuits with network size for Kauffman Networks of $K = 1, 2, 3, 4, 5$.

Unlike the number of monomers and components, the number of elementary circuits is significantly affected by the presence of self-edges in the network. Figure 7 shows the growth of the number of circuits with network size in NK-networks. It is clear that for $K > 1$ the number of circuits grows very rapidly - much faster than any power law. This would appear to confirm the present view that the related number of attractors in an RBN will grow at least this fast. A least-squares fit again reveals that the number of circuits or loops varies as $N_L \approx A_L e^{bN}$. It is not entirely clear from our data what the exact relationship between exponent b and K is. While there is clearly a positive monotonic relationship, our data are not good enough to distinguish b linear with K or $\log K$ or some power law in K .

It is interesting that for $K > 1$ the presence of self-edges in the network considerably lowers the number of circuits present. It does not appear to influence the value of b or its relationship with K . However for $K = 1$ networks, the presence of self-edges actually raises the number of circuits.

This is intriguing since the self-arcs are vital for loops in RBNs but this entirely structural behaviour crossover occurs exactly at the RBN critical K_c value.

Figure 8 shows the growth of the number of circuits with network size in Kauffman networks for the special case of $K = 1$. This definitely does not exhibit the same relationship as for $K > 1$. Even over the relatively small network sizes of $N = 10, \dots, 100$ the data are consistent with a power law $N_L \approx N^x$ where $x \approx$

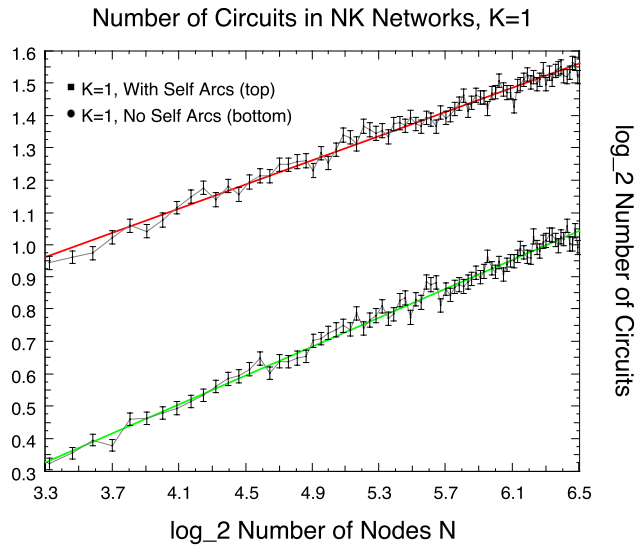


Figure 8: Growth of the number of circuits with network size for Kauffman Networks of $K = 1$.

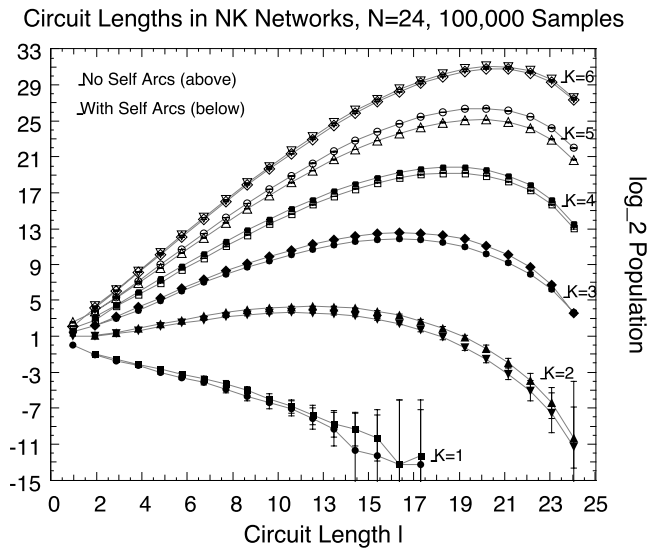


Figure 9: Circuit length distribution in a 24 Node Kauffman Network for $K = 1, 2, 3, 4, 5, 6$ with and without self-edges allowed in the network.

0.225 ± 0.003 for a network with no self-edges, and $x \approx 0.187 \pm 0.003$ when self-edges are present. Although even with 100,000 sample networks there is still a sizeable spread in the mean number of circuits (shown by the uncertainty bars on the plot), the high quality nature of the numerical fit suggests these split behaviours are significant.

Figure 9 shows the distribution of circuit lengths in 24 node NK-networks for $K = 1, 2, 3, 4, 5, 6$. This is again averaged over many samples (100,000 for $K = 1, 2$; 100 for $K = 2, 3$; and 10 for $K = 5, 6$). The distribution shows again the split behaviour for the cases $K > 1$ for which the self-edges lower the number of circuits of each length, and for $K = 1$ for which the self-edges raise the number of circuits. The shape of the distribution itself is quite revealing. For $K > 1$ there will be circuits of lengths up to the Hamiltonian circuit length of $L_H \equiv N$, with a modal value at some lesser length that increases with K . For the case of $K = 1$ however, the modal length is always unity and the maximum circuit length is truncated (perhaps only in the limit of large N ?) to $N/2$. As the uncertainty bars on the plots show, very large samples are needed to extract smooth mean values the longer length parts of the distributions for $K = 1, 2$.

5 Discussion and Conclusions

We have explored several of the structural properties of the NK-network model and have found that the number of circuits grows faster than any power law with network size. This confirms that the number of attractor loops in Random Boolean Networks should also grow faster than any power law. This behaviour also gives some insights into why particular lengths of attractors should form, based upon the shape of the circuits length distribution.

We have identified some intriguing structural behaviour between the values $1 < K < 2$ where the circuit length distribution function exhibits a change over from exponential decay to growth towards a non-unit modal value. Our data appears to show that the network self-edges or RBN nodes with self-inputs have a decisive role to play in influencing the location of the phase transition and hence the number of circuits and hence attractors present. The structural components, which for $K > 1$, are completely dominated by the giant component and are insensitive in number to the presence or absence of self-edges. For $K = 1$ however, disconnected components of sizes up to half the network size are present, and not just monomers. It appears in fact that for $K = 1$ the self-edges dominate the circuit size distribution and are the most prevalent loop type present.

We observe that the number of circuits displays some dependence on the presence of multiple-edges. We are investigating this more thoroughly, but preliminary data suggests that the number of multiple-edges grow logarithmically with N and of course can only affect (by definition) NK-networks with $K > 1$. Eliminating multiple-edges like self-edges from the network generation model does change the connection distribution probability away from a flat uniform one.

We have determined some of the growth behaviours for monomers, components and circuits. We might expect the results we have found to hold well, on average for finite practicably sized networks as well as for the large N limit. We are investing more computational effort into studies of higher K values to investigate the exponent dependence on K for the number of circuits.

References

- [1] Wolfram, S.: Theory and Applications of Cellular Automata. World Scientific (1986)
- [2] Kauffman, S.A.: The Origins of Order. Oxford University Press (1993)

- [3] Kauffman, S., Peterson, C., Samuelsson, B., Troein, C.: Random boolean network models and the yeast transcriptional network. *Proc. Natl. Acad. Sci. USA* **100** (2003) 14796
- [4] Baillie, C.F., Johnston, D.: Damping 2d quantum gravity. *Physics Letters B* **326** (1994) 51–56
- [5] Baillie, C., Hawick, K., Johnston, D.: Quenching 2d quantum gravity. *Physics Letters B* (1994) 284–290
- [6] Gershenson, C.: Introduction to random boolean networks. Available at arXiv:nlin/0408006v3 (2004)
- [7] Kadanoff, L., Coppersmith, S., Aldana, M.: Boolean dynamics with random couplings. In Kaplan, E., Marsden, J., Sreenivasan, K., eds.: *Perspectives and Problems in Nonlinear Science*. Springer (2003)
- [8] Derrida, B., Pomeau, Y.: Random networks of automata: A simple annealed approximation. *Europhys. Lett.* **1** (1986) 45–49
- [9] Derrida, B., Stauffer, D.: Phase transitions in two-dimensional Kauffman cell automata. *Europhys. Lett.* **2** (1986) 739
- [10] Harvey, I., Bossomaier, T.: Time out of joint: Attractors in asynchronous random boolean networks. In Husbands, P., Harvey, I., eds.: *Proc Fourth European Conference on Artificial Life (ECAL97)*, MIT Press (1997) 67–75
- [11] Szejka, A., Drossel, B.: Evolution of canalizing boolean networks. *The European Physical Journal B* **56** (2007) 373–380
- [12] Wuensche, A.: Discrete dynamical networks and their attractor basins. In R. Standish, e.a., ed.: *Proc. Complex Systems 1998*, UNSW, Sydney, Australia (1998) 3–21
- [13] Drossel, B.: On the number of attractors in random boolean networks. Technical Report arXiv.org:cond-mat/0503526, Institute fur Festkorperphysik, TU Darmstadt (2005)
- [14] Drossel, B., Mihaljev, T., Greil, F.: Number and length of attractors in a critical Kauffman model with connectivity one. *Phys. Rev. Lett.* **94** (2005) 088701
- [15] Kauffman, S.A.: Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology* **22** (1969) 437–467
- [16] Bilke, S., Sjunnesson, F.: Stability of the Kauffman model. *Phys. Rev. E* **65** (2001) 016129
- [17] Socolar, J.E.S., Kauffman, S.A.: Scaling in ordered and critical random boolean networks. *Phys. Rev. Lett.* **90** (2003) 068702–1
- [18] Bastolla, U., Parisi, G.: Relevant elements, magnetization and dynamical properties in Kauffman networks: A numerical study. *Physica D* **115** (1998) 203–218
- [19] Bastolla, U., Parisi, G.: The modular structure of Kauffman networks. *Physica D* **115** (1998) 219–233
- [20] Samuelsson, B., Troein, C.: Superpolynomial growth in the number of attractors in Kauffman networks. *Phys. Rev. Lett.* **90** (2003) 098701–1
- [21] Flyvbjerg, H., Kjaer, N.J.: Exact solution of Kauffman’s model with connectivity one. *J. Phys. A: Math. Theor.* **21** (1988) 1695–1718
- [22] K.A.Hawick, H.A.James, C.J.Scogings: Simulating large random boolean networks. Technical Report CSTN-039, Information and Mathematical Sciences, Massey University, Albany, North Shore 102-904, Auckland, New Zealand (2007)

- [23] Harary, F., Palmer, E.M.: Graphical Enumeration. New York, Academic Press (1973)
- [24] H.A.James, K.A.Hawick: Computational data structures for lattice-based small-world simulations. Technical report, Institute for Information and Mathematical Sciences, Massey University (2005)
- [25] Dewar, R., Harris, C.K.: Parallel computation of cluster properties: application to 2d percolation. *J. Phys. A Math. Gen.* **20** (1987) 985–993
- [26] Tiernan, J.C.: An efficient search algorithm to find the elementary circuits of a graph. *Communications of the ACM* **13** (1970) 722–726
- [27] Tarjan, R.: Enumeration of the elementary circuits of a directed graph. *SIAM Journal on Computing* **2** (1973) 211–216
- [28] Johnson, D.B.: Finding all the elementary circuits of a directed graph. *SIAM Journal on Computing* **4** (1975) 77–84
- [29] K.A.Hawick, H.A.James: A fast code for enumerating circuits and loops in graphs. Massey University Technical Note (2005)