A Fisher Information Study of Phase Transitions in Random Boolean Networks

X. Rosalind Wang[1], Joseph T. Lizier[1,2] and Mikhail Prokopenko[1]

[1]CSIRO Information and Communications Technology Centre, PO Box 76, Epping, NSW 1710, Australia
[2]School of Information Technologies, The University of Sydney, NSW 2006, Australia
Rosalind.Wang@csiro.au

Abstract

We study the order-chaos phase transition in random Boolean networks (RBNs), which have been used as models of gene regulatory networks. In particular we seek to characterise the phase diagram in information-theoretic terms, focussing on the effect of the control parameters (activity level and connectivity). Fisher information, which measures how much system dynamics reveal about its parameters, offers a natural interpretation of the phase diagram in RBNs. We report that this measure is maximised near the critical state in the order-chaos phase transitions in RBNs, since this is the region where the system is most sensitive to its parameters. Furthermore, we use this study of RBNs to clarify the relationship between Shannon and Fisher information measures.

Introduction

Random Boolean Networks (RBNs) (Kauffman, 1993) have typically been used by Artificial Life researchers as discrete dynamical network models (e.g., models of Gene Regulatory Networks) with a large sample space available. In particular, RBNs exhibit a well-known phase transition from ordered to chaotic dynamics, with respect to average connectivity or activity level.

Recently, there have been several attempts to study the order-chaos phase transitions of RBNs using information theory.[1] Ribeiro et al. (2008) measured mutual information within random node pairs as a function of connectivity in the network, finding a maximum near the critical point. Rämö et al. (2007) measured the uncertainty (entropy) in the size of perturbation avalanches as a function of an order parameter, and also found a maximum near the critical point. Lizier et al. (2008a) studied the information storage and transfer components of the computation conducted by each node in RBNs. The authors found maxima of these computational quantities just inside the ordered and chaotic sides of the critical point respectively.

While all of these studies provide useful findings regarding the nature of the phase transition, none provide a generic measure that can directly, reliably, and information-theoretically locate the critical point in other systems. For example, the study of perturbation avalanches in (Rämö et al., 2007) is not applicable to systems in which we cannot interfere. The measure of pairwise mutual information (Ribeiro et al., 2008) can be imagined to be maximised for trivial short-periodic behaviour as well as complex behaviour at critical point. And while our previous work (Lizier et al., 2008a) certainly characterises how the RBNs' computation is made up of both information storage and transfer, none of the measures of computation examined were maximised precisely at the critical point in finite-sized systems. In this study we aim to provide a preliminary analysis (in the context of RBNs) of a phase diagram in information-theoretic terms, aiming for the analysis to be generically applicable to other phase transitions. The search for generic tools motivates our study and we use information theory that allows us to analyse and compare critical behaviours across different domains.

Phase transitions are often related to symmetry breaking and self-organisation (Polani, 2007). For instance, Jetschke (1989) defines a system as undergoing a self-organising transition if the symmetry group of its dynamics changes to a less symmetrical one (e.g., a subgroup of the original symmetry group). An example may be given by a ferromagnetic system undergoing a second-order phase transition: (i) in the high-temperature phase the system has no net magnetisation, is 'disordered' and has a complete rotational symmetry (isotropy); (ii) at low temperature, the system becomes 'ordered', and the net magnetisation defines a preferred direction in space (anisotropy), breaking rotational symmetry. The low-temperature ordered phase is therefore *less* symmetrical and can be fully described by an order parameter — the magnetisation vector (Parwani, 2001).

In explaining non-equilibrium structures that spontaneously self-organise in nature, Synergetics (Haken, 1983) — a theory of pattern formation in complex systems — also employs order parameters. When energy or matter flows into a system typically describable by many variables, it may move far from equilibrium, approach a threshold (that can

[1]We note the study of entropy and mutual information between node inputs and outputs by Oosawa and Savageau (2002), though this study did not consider the phase transition in RBNs.

be defined in terms of some control parameters, e.g., the strength of interactions within the system, or the correlation length), and undergo a phase transition. At this stage, the behaviour of the overall system can be described by only a few order parameters (degrees of freedom) that characterise newly formed patterns. In other words, the system becomes low-dimensional as some dominant variables "enslave" others, making the whole system act in synchrony. By varying control parameters (e.g., the strength of interactions within the system) one may trigger phase transitions.

At this stage we would like to highlight the role of (Shannon) information: "a macroscopic description allows an enormous compression of information so that we are no more concerned with the individual microscopic data but rather with global properties" (Haken, 2006). A canonical example is a laser: a beam of coherent light created out of the chaotic movement of particles. Rather than using a large amount of information describing the states of individual atoms, only a single quantity (e.g., the phase of the total light field) is needed, achieving compression of information. Hence, a consensus is reached among the individual parts of the system, indicated by the compression of information, and only one or a few variables have to be guided or controlled (Prokopenko, 2009). In addition, in a vicinity of phase transitions, the information of the order parameters changes dramatically whereas the information of the enslaved modes does not (Haken, 2006).

These insights suggest the use of Fisher information (Fisher, 1922), which measures the amount of information that an observable random variable carries about an unknown parameter. Intuitively, if this unknown parameter can be estimated well using the observable random variable, then Fisher information carried by these observations with respect to this parameter must be high. Otherwise, if the parameter cannot be well-estimated using the observations, the corresponding Fisher information must be low. The application of Fisher information to measure the information that system dynamics contain about control parameters during a phase transition is quite natural. One could expect this quantity to be maximised near the critical point where system dynamics are most sensitive to control parameters.

Our main goal then is to obtain a phase diagram of RBNs in information-theoretic terms using Fisher information. Furthermore, since some studies of Fisher information discuss its connections to (derivatives of) Shannon information, we intend to clarify the relationship between Shannon and Fisher information, using RBNs.

We begin this paper with overviews of RBNs, Fisher information and Shannon information. This is followed by a discussion of how to apply Fisher information to RBNs. We then present the phase diagram of RBNs in terms of Fisher information about the control parameter, demonstrating that this quantity is indeed maximised near the critical point in the order-chaos phase transition in RBNs. Finally, we provide quantitative clarification regarding the relationship between Fisher and Shannon information measures using RBNs as an example.

Random Boolean Networks

Random Boolean Networks is a class of generic discrete dynamical network models. They are particularly important in artificial life, since they were proposed as models of gene regulatory networks by Kauffman (1993). See also Gershenson (2004a) for another thorough introduction to RBNs.

An RBN consists of N nodes in a directed *network*. The nodes take *boolean* state values, and update their state values in time as a function of the state values of the nodes from which they have incoming links. The network topology (i.e. the adjacency matrix) is determined at *random*, subject to whether the in-degree for each node is constant or stochastically determined given an average in-degree $\overline{K}$ (giving a Poisson distribution). It is also possible to bias the network structure, e.g., toward scale-free degree distribution (Aldana, 2003). Given the topology, the deterministic boolean function or lookup table by which each node computes its next state from its neighbours is also decided at *random* for each node, subject to a probability r of producing outputs of "1" (the *bias*). Note that r close to 1 or 0 gives low activity, whereas r close to 0.5 gives the highest activity for any $\overline{K}$. The nodes here are heterogeneous agents: there is no spatial pattern to the network structure (indeed there is no inherent concept of locality), nor do the nodes have the same update functions. (Though, of course either of these can arise at random). Importantly, the network structure and update functions for each node are held static in time ("quenched"). In classical RBNs (CRBNs), the nodes all update their states synchronously.[2]

The synchronous nature of CRBNs, their boolean states and deterministic update functions give rise to a global state space for the network as a whole with deterministic transient trajectories ultimately leading to either fixed or periodic attractors in finite-sized networks (Wuensche, 1997). Effectively, the transient is the period in which the network is *computing* its steady state attractor.

RBNs are known to exhibit three distinct phases of dynamics, depending on their parameters: ordered, chaotic and critical. At relatively low connectivity (i.e., low degree K) or activity (i.e., r close to 0 or 1), the network is in an ordered phase, characterised by high regularity of states and strong convergence of similar global states in state space. Alternatively, at relatively high connectivity and activity, the network is in a chaotic phase, characterised by low regularity

[2]While there has been some debate about the best updating scheme to model GRNs (Darabos et al., 2007), the relevant phase transitions are known to exist in all updating schemes, and their properties depend more on the network size than on the updating scheme (Gershenson, 2004b). As such, the use of CRBNs is justified for ensemble studies such as ours (Gershenson, 2004c).

of states and divergence of similar global states. In the critical phase (the *edge of chaos* (Langton, 1990)), there is percolation in nodes remaining static or updating their values, and uncertainty in the convergence or divergence of similar macro states. This phase transition is typically quantified using a measure of sensitivity to initial conditions, or damage spreading. Following Gershenson (2004c), we take a random initial state A of the network, invert the value of a single node to produce state B, then run both A and B for many time steps (enough to reach an attractor is most appropriate). We then use the Hamming distance:

$$D(A,B) = \frac{1}{N}\sum_{i=1}^{N}|a_i - b_i|, \quad (1)$$

between A and B at their initial and final states to obtain a convergence/divergence parameter δ:

$$\delta = D(A,B)_{t\to\infty} - D(A,B)_{t=0}. \quad (2)$$

(Note $D(A,B)_{t=0} = 1/N$). Finding $\delta < 0$, implies the convergence of similar initial states, while $\delta > 0$ implies their divergence.[3] For fixed r, the critical value of $\overline{K}$ between the ordered and chaotic phases is (Derrida and Pomeau, 1986):

$$K_c = \frac{1}{2r(1-r)}. \quad (3)$$

For finite-sized networks the standard deviation of δ peaks slightly inside the chaotic regime, indicating the widest diversity of networks for those parameters (Gershenson, 2004b). Indeed, the standard deviation is used as a guide to the relative regions of dynamics in finite-sized networks by Rämö et al. (2007), and the indicated shift of the critical point towards the chaotic regime at these finite sizes is reflected by other measures, e.g. (Ribeiro et al., 2008).

Much has been speculated on the possibility that gene regulatory and other biological networks function in (or evolve to) the critical regime (see Gershenson (2004a)). It has been suggested that computation occurs more naturally with the balance of order and chaos there (Langton, 1990), possibly with information storage, propagation and processing capabilities maximised (Kauffman, 1993). Indeed, our previous work has indicated that both information storage and coherent (single-source) information transfer are maximised near the critical state, just within the ordered and chaotic regimes respectively (Lizier et al., 2008a). Because of the importance of the critical state, identifying its precise location is a crucial task, particularly in other systems where analytical solutions are not possible. We look to information theory to address this question.

[3]Typically an order parameter is 1 in the extreme ordered phase, and 0 in the extreme disordered phase. Here, δ is a proxy to this, with negative values representing the ordered phase and positive values representing the chaotic phase.

Fisher Information

Information theory (MacKay, 2003) is an increasingly popular framework for the study of complex systems and their phase transitions (Prokopenko et al., 2009). In part, this is because complex systems can be viewed as distributed computing systems, and information theory is a natural way to study computation, e.g. Lizier et al. (2008b). Information theory is applicable to any system, provided that one can define probability distribution functions for its states. This is a particularly important characteristic since it means that information-theoretic insights can be directly compared across different system types. It is for these reasons that we seek an information-theoretic characterisation of the phase transition in RBNs.

Fisher information (Fisher, 1922) is a way of measuring the amount of information that an observable random variable X has about an unknown parameter θ, upon which the likelihood function of θ depends. Let $p(x|\theta)$ be the likelihood function of θ given the observations x. Then, Fisher information can be written as:

$$F(\theta) = \int_x \left(\frac{\partial \ln(p(x|\theta))}{\partial\theta}\right)^2 p(x|\theta)dx, \quad (4)$$

where $\ln(p(x|\theta))$ is the log-likelihood of θ given x. Thus, Fisher information is not a function of a particular observation, since the random variable X has been averaged out.

Fisher information can be reduced to:

$$F(\theta) = -\int_x \left(\frac{\partial^2 \ln(p(x|\theta))}{\partial\theta^2}\right) p(x|\theta)dx, \quad (5)$$

if $\ln(p(x|\theta))$ is twice differentiable with respect to θ and if the regularity condition:

$$\int \frac{\partial^2}{\partial\theta^2} p(x|\theta)dx = 0 \quad (6)$$

holds. In this paper we use Equation 4, since the regularity condition (Equation 6) does not necessarily hold.

The discrete form of Fisher information is:

$$F(\theta) = \sum_{x_j} p_{x_j}\left(\frac{\Delta \ln(p_{x_j})}{\Delta\theta}\right)^2, \quad (7)$$

where $\Delta\ln(p_{x_j}) = \ln(p'_{x_j}) - \ln(p_{x_j})$ and $p_{x_j} = p(x_j|\theta)$, $p'_{x_j} = p(x_j|\theta+\Delta\theta)$. In this case, $p(x)$ is a discrete probability distribution function, such that $x \in \{x_1, \ldots, x_D\}$, where D is the number of states for the variable X. For example, for a boolean network, $x \in \{0,1\}$.

Fisher information has been extensively used in many fields of science. Frank (2009) argued that Fisher information may be used as the intrinsic metric of natural selection and evolutionary dynamics. Brunel and Nadal (1998) showed that in the context of neural coding, the mutual information between stimuli applied to neurons and neuronal

activity can be characterised by Fisher information. In computer science, Ganguli et al. (2008) studied short term memory in discrete time neural networks by using a criterion based on Fisher information.

We are interested in two aspects of Fisher information. Firstly, it is a measure of the ability to estimate a parameter, making it an important aspect of parameter estimation in statistics (Frieden, 1998). Secondly, it is related to the fundamental quantity of information theory, Shannon information that measures system's uncertainty.

Shannon Information

Shannon Information (Shannon, 1948) was originally developed for reliable transmission of information from a source X to a receiver Y over noisy communication channels. Put simply, it addresses the question of "how can we achieve perfect communication over an imperfect, noisy communication channel?" (MacKay, 2003). When dealing with outcomes of imperfect probabilistic processes, it is useful to define the information content of an outcome x, which has the probability $P(x)$, as $\log_2 \frac{1}{P(x)}$. Crucially, improbable outcomes convey more information than probable outcomes. Given a probability distribution P over the outcomes $x \in \mathcal{X}$ (a discrete random variable X representing the process, and defined by the probabilities $P(x) \equiv P(X = x)$ given for all $x \in \mathcal{X}$), the average Shannon information content of an outcome is determined by

$$H(X) = -\sum_{x \in \mathcal{X}} P(x) \log_2 P(x), \tag{8}$$

We note the information is measured in bits, and henceforth omit the logarithm base 2. This quantity is known as *(information) entropy*, and may be contrasted with Fisher information in Equation 7.

Intuitively, Shannon information measures the amount of freedom of choice (or the degree of randomness) contained in the process — a process with many possible outcomes has high entropy. This measure has some unique properties that make it specifically suitable for measuring "how much 'choice' is involved in the selection of the event or of how uncertain we are of the outcome?" (Shannon, 1948). In answering this question, Shannon suggested the entropy function $-k \sum_{i=1}^{n} P(x_i) \log P(x_i)$, where a positive constant k represents a unit of measure.

In this paper we consider the entropy defined in terms of the probability distribution of the states of each node with respect to some parameter θ.[4] Here the probabilities $p(x_j^i|\theta)$ are defined for each possible state x_j^i for each node i (given θ), and Shannon entropy

$$H(X^i|\theta) = -\sum_{x_j} p(x_j^i|\theta) \log p(x_j^i|\theta)$$

is subsequently also defined for each node i given θ, measuring the diversity of system's states. Then this quantity is averaged across the network given θ,

$$H(\theta)_{RBN} \triangleq \langle H(X^i|\theta) \rangle_i. \tag{9}$$

Fisher Information for RBNs

We aim to study Fisher information $F(r)$ in RBNs as a function of the probability r of each node producing an output of "1". When changing r, the total number of 1s and 0s in the logic tables (which each node uses to compute its next state from its neighbours) would change. So when we calculate $p(x|r)$ and $p(x|r + \Delta r)$ for each r, some nodes in the network with $\theta = r + \Delta r$ would have different logic tables. Therefore, we will produce two sets of results when calculating Fisher information: one where we take into account all the nodes in the network, and one where we ignore those nodes that have their logic table changed. This will allow us to see whether the changes in dynamics are mostly constrained to the nodes whose logic tables have changed, or whether the alterations to their logic genuinely cause changes to the dynamics of the *whole* network and allow insights into r from across the network.

To find Fisher information for the networks, Equation 7 is used since the RBN has nodes with discrete states 0 and 1. If we applied this equation to the RBN as a whole, the likelihood function $p(\mathbf{x}|r)$ is a joint distribution over all nodes $\mathbf{X}$ in the network. This means that for an RBN of 100 nodes, there are 2^{100} possible joint states, which makes a calculation of Fisher information for the joint state of the RBN impractical. Furthermore, since the RBN is not a directed acyclic graph, and its nodes are not independent and identically distributed (i.i.d.), we can not write the likelihood function as a product of the individual nodes. An alternative would be to apply Equation 7 to the single node states x, computing the $p(x|r)$ by combining observations of all nodes in the RBN. This is undesirable though, since the nodes are heterogeneous agents with very different dynamics. Instead, we chose to study the average Fisher information of the individual nodes:

$$F(r)_{RBN} \triangleq \langle F_i(r) \rangle \tag{10}$$

where $F_i(r)$ is the Fisher information of the i-th node of the RBN calculated using Equation 7.

We model the RBNs using enhancements to Gershenson's RBNLab software (http://rbn.sourceforge.net). When approximating an infinitely-sized network with a finite one, the risk is to run the dynamics for too many time steps and reach a periodic or fixed attractor (inevitable for finite-sized

[4] We note the alternative view used elsewhere of information in networks as that contained in the degree distribution amongst nodes (Solé and Valverde, 2004; Bianconi, 2008; Piraveenan et al., 2009).

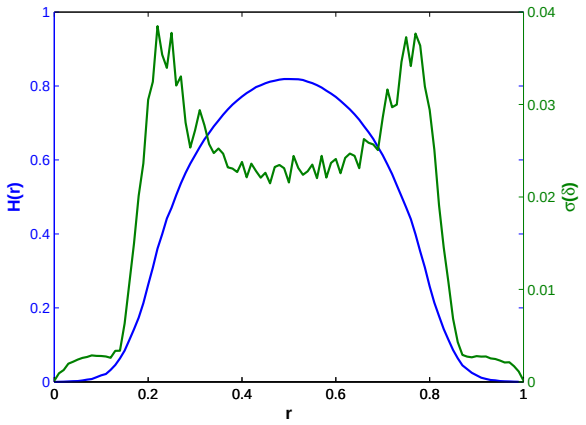

Figure 1: (blue) Average Shannon information $H(r)$ and (green) standard deviation of the convergence/divergence parameter δ versus the bias of the network r. The RBNs here have network size $N = 250$, and average network connectivity $\overline{K} = 4.0$.

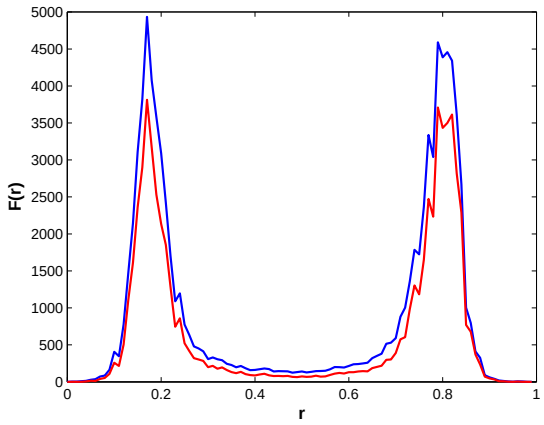

Figure 2: Average Fisher information $F(r)$ versus the bias of the network, r, for networks of size $N = 250$ and average connectivity $\overline{K} = 4.0$. The blue curve shows the Fisher information if we take into account all the nodes in the network, the red curve shows the Fisher information if we ignore those nodes whose logic table has changed due to the change in parameter r.

RBNs). In order to avoid this, for each simulation run starting from an initial randomised state, we ignore a short initial transient of 30 steps to allow the network to settle into the main phase of the computation, and then stop the computation after 400 time steps.

In order to properly sample the dynamics of each node in each RBN and generate enough data for the information theoretic calculations, many repeat runs from random initial states are required for each network (250 were used).

We thus calculate $p(x^i|r)$ of each node i in a given RBN over all the repeat runs. This likelihood of each node is used to calculate the Fisher information at node i, thus giving us the average Fisher information of the network, $F(r) = \langle F(r)_{RBN} \rangle$. Similarly, we averaged the entropy measurements $H(r) = \langle H(r)_{RBN} \rangle$ over the network realisations for each r.

It should be noted, that for many nodes, it often happens that p_x and/or $p'_x = 0$ because a node may exhibit either all 0s or all 1s, especially when r of the network is heavily biased towards 0 or 1. In these cases, if $p_{x_j} = 0$, we set the corresponding individual terms in Equation 7 $p_{x_j} \left(\frac{\Delta \ln(p_{x_j})}{\Delta r} \right)^2 = 0$, where j is the state of the node i. If $p'_{x_j} = 0$, we write the respective terms as (Frank, 2009):

$$p_{x_j} \left(\frac{\Delta \ln(p_{x_j})}{\Delta r} \right)^2 = \frac{1}{p_{x_j}} (p'_{x_j} - p_{x_j})^2,$$

yielding: $p_{x_j} \left(\frac{\Delta \ln(p_{x_j})}{\Delta r} \right)^2 = p_{x_j}$.

Results and Discussions

We focus on RBNs with $N = 250$ nodes and average connectivity of $\overline{K} = 4.0$, while altering the bias in the network r. $\overline{K} = 4.0$ was chosen because, with it held constant, RBNs at low and high values of r exhibit ordered behaviour and RBNs at mid-range values of r exhibit chaotic behaviour.

Figure 1 shows two baseline measures for studying the phase transition. The green curve shows the standard deviation of the convergence/divergence parameter δ as it changes over r. As discussed earlier, this is a typical parameter used to study this phase transition, and the standard deviation is known to reflect the shift of the edge of chaos in finite-sized networks. We can see that there are two separate peaks in this curve, representing the edge of chaos for this finite-sized RBN. This is expected, since the probability distribution function is symmetrical about $r = 0.5$, where there is no bias between choosing a 0 or a 1. These two peaks occur at $r = 0.22$ and 0.77, which as expected are 'inside' the theoretical edge of chaos of an infinite-sized RBN at $r = 0.147$ and 0.853 as found using Equation 3. The blue curve shows the average Shannon information $H(r)$ through this phase transition. $H(r)$ exhibits a bell shaped curve with maximum at $r = 0.5$; this is as expected since the level of activity in the network should be maximum when there is no bias. This result aligns with the previous study of the entropy through the phase transition in RBNs as a function of $\overline{K}$ while holding r constant (Lizier et al., 2008a) .

Now, we examine the phase transition with respect to r using Fisher information $F(r)$. Figure 2 shows the average $F(r)$ calculated in two scenarios: the blue curve shows $F(r)$

when we take into account observations from all nodes in the network, the red curve shows $F(r)$ when we ignore observations from those nodes that have their logic table changed from $p(x|r)$ to $p(x|r + \Delta r)$.

It can be seen from this plot that $F(r)$ has two peaks almost mirrored about $r = 0.5$. These peaks occur approximately at the phase transition between the chaotic phase and the ordered phase for RBN with $\overline{K} = 4.0$ as shown in Figure 2, while $F(r)$ away from the phase transition r has values at least one order of magnitude smaller than the peaks. This indicates that close to the phase transition, there is a large increase in the information in the state distribution of the nodes about the parameter r. On the other hand, deep inside the ordered and chaotic phases, the state distribution of the nodes indicates little about r, other than that the network is in one of these phases.

Certainly the blue curve for $F(r)$ is consistently higher. This curve includes Fisher information from the nodes whose logic tables were changed, and these nodes obviously carry a significant amount of information about the r parameter. Crucially though, there is little difference between the two curves, and both have peaks at $r = 0.17$ and $r = 0.79$. Were the curves identical, this would imply the amount of information about r in the changed nodes did not differ from that in the unchanged nodes, and the average $F(r)$ was not affected. A small quantitative difference indicates that the nodes with changed logic tables retain more information about r. Nevertheless, the information diffuses through the whole network, making the curves quite similar here.

Some studies on Fisher information discuss the relationship between Fisher and Shannon information. Frank (2009) proposed the interpretation that Fisher information is equivalent to the acceleration of Shannon information, i.e. the second derivative of $H(X|\theta)$ with respect to θ. This was shown under the assumption that the outer (or averaging) term $p(x|\theta)$ holds constant while differentiating $H(X|\theta)$, thus differentiating $\log p(x|\theta)$ only. The equivalence between Fisher and acceleration of Shannon information also requires that the regularity condition in Equation 6 holds. However, this is not always the case, and here we now describe our observation of more similarity between Fisher information and *first* derivative of Shannon information.

Figure 3 shows the derivatives of Shannon information $H(r)$ versus network bias r for RBNs with average connectivity of $\overline{K} = 4.0$: the square of the first derivative of Shannon information, $(\frac{d}{dr}H)^2$, is shown in blue and the second derivative, $\frac{d^2}{dr^2}H$, is shown in green. In comparison with Figure 2, we can see that Fisher information for RBNs is more qualitatively similar in shape to the square of rate of change of Shannon information than the acceleration of Shannon information. However, there is a difference in their orders of magnitude, an explanation for which is presented in the Appendix. In general, this is because in finding $F(\theta)$, we first differentiate and then square and average the values, while for $(\frac{dH}{d\theta})^2$ we average and then differentiate and square the values. Furthermore, the peaks for $(\frac{dH}{d\theta})^2$ occur at $r = 0.21$ and $r = 0.79$, coinciding with the Fisher information peaks shown in Figure 2. This shows that for RBNs, the regularity condition of Equation 6 does not hold, and Fisher information is not equivalent to the acceleration of Shannon entropy.

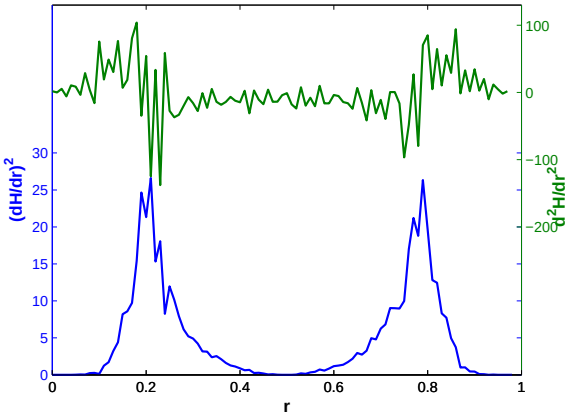

Figure 3: Derivatives of Shannon Information, $H(r)$, for the same networks as Figure 2 ($\overline{K} = 4.0$). (blue) First derivative of H squared. (green) Second derivative of H.

Let r_{max} denote the maximum Fisher information that occurs with respect to r for fixed $\overline{K}$. Formally, r_{max} for every $\overline{K}$ is set to the global maxima of $F(r)$ in two regions: $0 \leq r \leq 0.5$ and $0.5 \leq r \leq 1$. For example, r_{max} correspond to the peaks shown in Figure 2. We now examine the values of r_{max} as a function of $\overline{K}$. To reiterate, each $F(r)$ is an average of Fisher information $F(r)_{RBN}$ over 250 networks, yielding r_{max} values for both regions. Repeating the experiment 10 times with different 250 networks allows us to average these r_{max} values over 10 runs. Figure 4 shows the plot of r_{max} versus $\overline{K}$ for $\overline{K} = 2.0$ to $\overline{K} = 10.0$. The blue curve shows the r_{max} computed over all nodes in the network, and the red curve corresponds to the case when those nodes that have their logic table changed were ignored. As we can see from the figure, there is very little difference between the two r_{max} curves. In alignment with the findings for Figure 2, we see that the changes to the logic tables of a few nodes genuinely cause the effect of changes in r to diffuse throughout the network.

The green dashed curve in Figure 4 shows the theoretical critical phase (edge of chaos) of the RBNs, generated using Equation 3. We can see from the figure that the phase diagram obtained by maximising Fisher information generally follows the same shape, but is bounded by the theoretical curve for critical K_c versus r. This is because the theoretical curve corresponds to an RBN with an infinite size, while the phase diagram based on the maximum Fisher information is

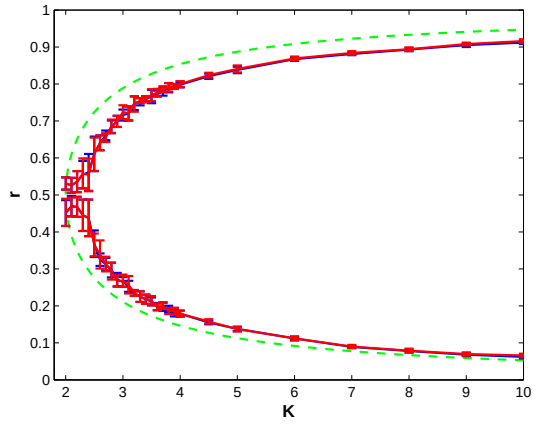

Figure 4: Phase diagram of r_{max} where the maximum Fisher information, $F(r)$, occurs with respect to r for fixed $\overline{K}$, as a function of $\overline{K}$. Blue: when all the nodes in the network were taken into account; Red: when those nodes whose logic table has changed due to the change in parameter r were ignored. The error bars on the curves show the standard deviation of r_{max}. The green dashed line is the theoretical curve for critical K_c versus r.

for a finite size RBN. As pointed out previously, for finite-size networks the critical point is known to shift towards the theoretically chaotic region, and the maximum Fisher information certainly reflects this.

Indeed, these finite-size effects also partly explain why the loci of the divergent maxima of Fisher information do not meet as $K \to 2$. For $r = 0.5$, the phase transition with respect to K shifts towards the chaotic regime at around $K \approx 2.5$ in these finite size RBNs rather than the theoretical 2.0. Our experimental curve(s) should converge/diverge at around $K \approx 2.5$. The fact that they do not converge is an artifact of our explicit search for two maximum values of $F(r)$ for $0 \leq r \leq 0.5$ and $0.5 \leq r \leq 1$.

Conclusion

In this paper, we contrasted Fisher information and Shannon information in the context of Random Boolean Networks (RBNs). RBNs are known to exhibit three distinct phases of dynamics, depending on their parameters: ordered, chaotic and critical, and we analysed the phase diagram of RBN dynamics interpreted in information-theoretic terms.

Both the activity level r and average connectivity K play the role of control parameters, and the phase diagram is obtained by plotting (K, r) points that separate the ordered and chaotic phases. If δ was used as a proxy to an order parameter, the critical (K, r) points are those where δ changes sign. Information-theoretically, Shannon information $H(r)$ which measures (globally) the diversity of RBN's states given the parameter r, is minimal in the ordered phase and maximal in the chaotic phase. However, it does not identify the precise location of the critical points. On the other hand, Fisher information about the control parameters has maxima at the critical (K, r) points. This is because $F(r)$ measures (locally) the amount of information that RBN dynamics carry about the parameter r, and these dynamics are most sensitive to the control parameter near the critical point.

Our analysis showed that an information-theoretic interpretation of the phase diagram (K with respect to r) reveals expected phases (ordered, chaotic and critical) as well as symmetry breaking (slightly obscured by finite-size effects). In addition, the comparison between Fisher information $F(r)$ and a square of a first derivative of Shannon information $H(r)$ uncovered their strong qualitative similarity, albeit separated by an order of magnitude. The analysis shed more light on connections between Fisher information and (derivatives of) Shannon information, and provided a means for further rigorous information-theoretic studies of phase transitions in complex networks.

Acknowledgements

We like to thank the High Performance Computing and Communications Centre (http://www.hpccc.gov.au/) for the use of their supercomputer clusters in performing the experiments for this paper.

Appendix

It can be seen from Figure 2 and 3 that the magnitude of the $F(r)$ is much higher than $(\frac{dH}{dr})^2$, in fact, the peak for $F(r)$ is approximately 200 times that of $(\frac{dH}{dr})^2$'s peak. This is due to the order of averaging and differential in the two calculations. To illustrate this, let us take one simple example, where the variable x has two states $\{0, 1\}$ the probabilities of which depend on some parameter θ:

$$\begin{aligned} \text{Let}: \quad & p(0|\theta) = 0.5 && p(1|\theta) = 0.5 \\ & p(0|\theta + \Delta\theta) = 0.3 && p(1|\theta + \Delta\theta) = 0.7 \\ & \Delta\theta = 0.01 \end{aligned}$$

Now, using Equation 8, we can find the Shannon information:

$$H(X|\theta) = -(0.5 \log_2 0.5 + 0.5 \log_2 0.5) = 1,$$
$$H(X|\theta + \Delta\theta) = -(0.3 \log_2 0.3 + 0.7 \log_2 0.7) = 0.8843.$$

Thus, the first derivative squared in this case is:

$$\left(\frac{dH(X|\theta)}{d\theta}\right)^2 = \left(\frac{H(X|\theta + \Delta\theta) - H(X|\theta)}{\Delta\theta}\right)^2 = 133.86.$$

Using Equation 7, we can find the Fisher information:

$$F(\theta) = 0.5 \left(\frac{\ln 0.3 - \ln 0.5}{\Delta\theta}\right)^2 + 0.5 \left(\frac{\ln 0.7 - \ln 0.5}{\Delta\theta}\right)^2$$
$$= \frac{0.5(-0.5108)^2 + 0.5(0.33647)^2}{(0.01)^2} = 6965.$$

Here, we can see that $F(\theta)$ is 50 times larger than $(\frac{dH}{d\theta})^2$. This shows that while at the first glance, the values of $F(\theta)$ and $(\frac{dH}{d\theta})^2$ should be similar, there is actually one to two orders of magnitudes difference between them.

References

Aldana, M. (2003). Boolean dynamics of networks with scale-free topology. *Physica D*, 185(1):45–66.

Bianconi, G. (2008). The entropy of randomized network ensembles. *Europhysics Letters*, 81:28005.

Brunel, N. and Nadal, J.-P. (1998). Mutual information, Fisher information and population coding. *Neural Computation*, 10(7):1731–57.

Darabos, C., Giacobini, M., and Tomassini, M. (2007). Semi-synchronous activation in scale-free boolean networks. In Almeida e Costa, F., Rocha, L. M., Costa, E., Harvey, I., and Coutinho, A., editors, *Proceedings of the 9th European Conference on Artificial Life (ECAL 2007), Lisbon, Portugal*, volume 4648 of *Lecture Notes in Artificial Intelligence*, pages 976–985, Berlin / Heidelberg. Springer.

Derrida, B. and Pomeau, Y. (1986). Random networks of automata: a simple annealed approximation. *Europhysics Letters*, 1(2):45–49.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society, A*, 220:309–368.

Frank, S. A. (2009). Natural selection maximizes Fisher information. *Journal of Evolutionary Biology*, 22:231–244.

Frieden, B. R. (1998). *Physics from Fisher Information*. Cambridge University Press, Cambridge, United Kingdom.

Ganguli, S., Huh, D., and Sompolinsky, H. (2008). Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences*, 105(48):18970–18975.

Gershenson, C. (2004a). Introduction to random boolean networks. In Bedau, M., Husbands, P., Hutton, T., Kumar, S., and Suzuki, H., editors, *Proceedings of the Workshops and Tutorials of the Ninth International Conference on the Simulation and Synthesis of Living Systems (ALife IX), Boston, USA*, pages 160–173.

Gershenson, C. (2004b). Phase transitions in random boolean networks with different updating schemes. arXiv:nlin/0311008.

Gershenson, C. (2004c). Updating schemes in random boolean networks: Do they really matter? In Pollack, J., Bedau, M., Husbands, P., Ikegami, T., and Watson, R. A., editors, *Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems (ALife IX), Boston, USA*, pages 238–243, Cambridge, USA. MIT Press.

Haken, H. (1983). *Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology*. Springer-Verlag, 3rd rev. enl. ed., New York.

Haken, H. (2006). *Information and Self-Organization: A Macroscopic Approach to Complex Systems*. Springer-Verlag, Berlin Heidelberg.

Jetschke, G. (1989). *Mathematik der Selbstorganisation*. Vieweg, Braunschweig.

Kauffman, S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York.

Langton, C. G. (1990). Computation at the edge of chaos: phase transitions and emergent computation. *Physica D*, 42(1-3):12–37.

Lizier, J. T., Prokopenko, M., and Zomaya, A. Y. (2008a). The information dynamics of phase transitions in random boolean networks. In Bullock, S., Noble, J., Watson, R., and Bedau, M. A., editors, *Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems (ALife XI), Winchester, UK*, pages 374–381, Cambridge, MA. MIT Press.

Lizier, J. T., Prokopenko, M., and Zomaya, A. Y. (2008b). Local information transfer as a spatiotemporal filter for complex systems. *Physical Review E*, 77(2):026110.

MacKay, D. J. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge.

Oosawa, C. and Savageau, M. A. (2002). Effects of alternative connectivity on behavior of randomly constructed boolean networks. *Physica D*, 170(2):143–161.

Parwani, R. R. (2001). Complexity: A web-book. http://staff.science.nus.edu.sg/~parwani/complexity.html.

Piraveenan, M., Prokopenko, M., and Zomaya, A. Y. (2009). Assortativeness and information in scale-free networks. *European Physical Journal B*, 67(3):291–300.

Polani, D. (2007). Foundations and formalizations of self-organization. In Prokopenko, M., editor, *Advances in Applied Self-organizing Systems*, pages 19–37. Springer, London.

Prokopenko, M. (2009). Book review: Information and self-organization: A macroscopic approach to complex systems (H. Haken). *Artificial Life*, 15(3):377–383.

Prokopenko, M., Boschetti, F., and Ryan, A. J. (2009). An information-theoretic primer on complexity, self-organization, and emergence. *Complexity*, 15(1):11–28.

Rämö, P., Kauffman, S., Kesseli, J., and Yli-Harja, O. (2007). Measures for information propagation in boolean networks. *Physica D*, 227(1):100–104.

Ribeiro, A. S., Kauffman, S. A., Lloyd-Price, J., Samuelsson, B., and Socolar, J. E. S. (2008). Mutual information in random boolean models of regulatory networks. *Physical Review E*, 77(1):011901–10.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, July, October.

Solé, R. V. and Valverde, S. (2004). Information theory of complex networks: Onevolution and architectural constraints. In Ben-Naim, E., Frauenfelder, H., and Toroczkai, Z., editors, *Complex Networks*, volume 650 of *Lecture Notes in Physics*, pages 189–207. Springer, Berlin / Heidelberg.

Wuensche, A. (1997). *Attractor Basins of Discrete Networks*. PhD thesis, The University of Sussex.