On the dynamics of random boolean networks subject to noise: Attractors, ergodic sets and cell types

R. Serra, M. Villani, A. Barbieri, S.A. Kauffman, A. Colacci

www.elsevier.com/locate/yjtbi

On the dynamics of random boolean networks subject to noise: Attractors, ergodic sets and cell types

R. Serra, M. Villani, A. Barbieri, S.A. Kauffman, A. Colacci

PII: S0022-5193(10)00195-5
DOI: doi:10.1016/j.jtbi.2010.04.012
Reference: YJTBI 5961

To appear in: *Journal of Theoretical Biology*

Received date: 5 November 2009
Revised date: 21 March 2010
Accepted date: 10 April 2010

Cite this article as: R. Serra, M. Villani, A. Barbieri, S.A. Kauffman and A. Colacci, On the dynamics of random boolean networks subject to noise: Attractors, ergodic sets and cell types, *Journal of Theoretical Biology*, doi:10.1016/j.jtbi.2010.04.012

On the dynamics of random Boolean networks subject to noise: attractors, ergodic sets and cell types

R. Serra[a,b], M. Villani[a,b], A. Barbieri[a], S.A. Kauffman[c], A. Colacci[d]

[a] *Department of Social, Cognitive and Quantitative Sciences, University of Modena and Reggio Emilia, viale A. Allegri 9, I-42100 Reggio Emilia, Italy*
[b] *European Centre for Living Technology S. Marco 2940 30124 Venice, Italy*
[c] *Departments of Biochemistry and Mathematics, University of Vermont, Burlington, VT 05405 E-mail: Stuart.Kauffman@uvm.edu*
[d] *Center for Environmental Carcinogenesis and Risk Assessment Environmental Protection and Health Prevention Agency Emilia Romagna Region Viale Filopanti 20/22, 40126 Bologna, Italy*

Abstract

The asymptotic dynamics of random Boolean networks subject to random fluctuations is investigated. Under the influence of noise, the system can escape from the attractors of the deterministic model, and a thorough study of these transitions is presented. We show that the dynamics is more properly described by sets of attractors rather than single ones. We generalize here a previous notion of ergodic sets, and we show that the Threshold Ergodic Sets so defined are robust with respect to noise and, at the same time, that they do not suffer from a major drawback of ergodic sets. The system jumps from one attractor to another of the same Threshold Ergodic Set under the influence of noise, never leaving it. By interpreting random Boolean networks as models of genetic regulatory networks, we also propose to associate cell types to Threshold Ergodic Sets rather than to deterministic attractors or to ergodic sets, as it had been previously suggested. We also propose to associate cell differentiation to the process whereby a Threshold Ergodic Set composed by several attractors gives rise to another one composed by a smaller number of attractors. We show that this approach accounts for several interesting experimental facts about cell differentiation, including the possibility to obtain an induced pluripotent stem cell from a

Email addresses: `alessia.barbieri@gmail.com` (A. Barbieri), `skauffman@ucalgary.ca` (S.A. Kauffman), `a.colacci@istbiotech.it` (A. Colacci)

fully differentiated one by overexpressing some of its genes.

Keywords:
random Boolean networks, dynamical systems, noise, cell types, differentiation, pluripotent cells

1. Introduction

Random Boolean networks are one of the most thoroughly studied models of complex systems (Kauffman, 1969).They are particularly interesting due to their rich dynamical behaviour, which ranges from ordered to pseudochaotic as some parameters are varied. The critical surface in parameter space which separates the two regions, sometimes called the "edge of chaos", has also attracted considerable interest (Kauffman, 1995)

Besides their abstract properties, random Boolean networks (RBNs for short) have been initially devised as a model of genetic regulatory networks, and as such they can be compared with actual experimental data. The availability of genome-wide gene expression data has allowed interesting comparisons between the behaviour of these models and that of real cells in different cases, including the distribution of perturbations induced by gene knockout (Serra et al., 2006) and the time course of synchronized leukemia cells (Shmulevich et al., 2005). The most remarkable results so far are *i)* the demonstration that even Boolean models can describe significant quantitative features of biological systems and *ii)* the finding of indications in favour of the hypothesis that real cells might be operating at, or close to the edge of chaos.

The former fact can be surprising, given the crude approximations which are used, and in particular the fact that the expression levels of the various genes are forced to take one of two possible values, while in nature they are multiple valued. However, the above-mentioned results show that even Boolean models can provide an adequate description of some experimental data. There are indeed other models of gene regulation which take into account the fact that gene expression levels may be very different, as for example the models that make use of continuous variables (Kaneko, 2006). However, the use of simplified Boolean models has the unique advantage of allowing one to deal with very large networks[1].

[1]this advantage is shared by a peculiar class of continuous networks, i.e. Glass networks

Models based on ordinary differential equations provide a very rich and useful description of the phenomena of gene regulation, and they also allow one to explore the possible role of chaotic dynamics. For computational reasons they can be applied in those case where the number of genes (or of those which are really important) is limited. Note also that the validity of deterministic ODE models can be limited in those cases where the number of molecules of some gene product is very small, thus making the continuum approximation questionable.

We think that no single kind of model can capture the whole phenomenology associated to these complex systems, so it is appropriate to make use of more than one type and to check which phenomena are better captured by one or another. It is also extremely important to ascertain which phenomena are robust with respect to the modelling choices, so they appear in different models. It is interesting to observe that, at an abstract level, the overall picture of cell differentiation which comes out of the model we propose here shares some important features with important continuum models of the same phenomenon (Huang, 2009; Kaneko, 2006). This point will be better specified in the final section 6.

Section 2 of this paper briefly summarizes the main features of RBNs and reviews the results of these previous studies.

Concerning the biological interpretation of RBNs, one of us (S.A.K.) proposed many years ago to associate their attractors to the various cell types which can be found in multicellular organisms. This interpretation appears sound, since cell types correspond to different stable patterns of gene expression given the same genome. The "genome" corresponds to the topology and to the choice of the Boolean functions associated to the nodes, therefore the stable patterns should coincide with the network attractors.

Recently it has been proposed (Huang, 2009; Huang et al., 2009) that the association of cell types to attractors is in principle able to provide a dynamical description of two of the most interesting phenomena, i.e. cell differentiation and the development of cancer, which is richer and more convincing than the classical one.

It has also been observed (Ribeiro and Kauffman, 2007) that in order

(Glass and Hill, 1998); their maximum activation values are however equal for all the nodes, just like in RBNs (while the decay is exponential in time). Moreover, the behaviour of these networks is known to resemble that of RBNs in many important aspects.

to associate attractors and cell types one has to take into account the role of noise: biological hardware (wetware) is error prone, and some regulatory enzymes are present only in small amounts, so it is certainly possible that one active node switches to inactive, or vice versa (given that also inhibitory regulations may be subject to the same level of fluctuations).

A more detailed discussion of the presence and the significance of noise in biological systems will be given in the final section 6.

In Section 3 of this paper we provide an in-depth analysis of the behaviour of RBNs subject to random fluctuations.

The typical experiment is the following: for a given network (defined by a given set of connections and Boolean functions) we let the system evolve from a set of random initial conditions, so we can find a set of attractors[2]. Then we choose one attractor, flip a node and record the new attractor which is reached by the network. By "flip" here (and from now on in this paper) we mean changing the state of a node for a single time step - after which the network is back to its usual operations. We do this for all the nodes of all the states of the attractor cycle, and of course we repeat the experiment for various network realizations which share the same global parameters. In the end, for each network we provide a statistics of the jumps between attractors induced by single flips. In order to test the robustness of these results with respect to the level of noise, we also tested the behaviour under double flips.

On the basis of the results of Section 3 we see that the attractors of RBNs have several interesting features which form the basis for identifying them with cell types, but that they are unstable with respect to noise. One of the interesting results is that the probability that, by switching a node chosen at random, the network moves to a different attractor is rapidly decreasing with network size. So one might be tempted to ignore the problem of noise-induced transitions, claiming that they are relevant only for unrealistically small nets. But this is not justified, since larger nets have also a larger number of nodes, so the overall effect is that of making the probability of moving to a new attractor per unit time an increasing function of network size (as explained in detail in Section 3).

Noise has therefore to be taken seriously into account. In order to do so Ribeiro and Kauffman proposed to associate cell types to ergodic attractors

[2]For large networks this can be a subset of the complete set of attractors, but it is likely to include those with large basins of attraction.

(ES) rather than attractors. ESs are sets of attractors which can be reached from one to another via single flips, and which can never be abandoned under such dynamics (see Section 4 for a formal definition), so they are robust with respect to noise. Unfortunately the authors found that typically there is only one ES per network, a result which compromises the possibility to identify them with different cell types. In Section 4 we confirm this negative result by investigating a set of networks larger than that of Ribeiro and Kauffman.

However there is a way out of this impasse, which allows one to introduce sets of attractors which are robust with respect to noise and which are, at the same time, present in multiple realizations in a single RBN. In considering the transitions among attractors, a more careful consideration of the physics of the phenomenon leads us to exclude those transitions which are associated to very rare events, unlikely to take place during a cell lifetime. Following this line of reasoning, and generalizing a previous work of ours (Barbieiri et al., 2009), we introduce in Section 4 the notion of Threshold Ergodic Set (TES). While all the details are described in Section 4, the key idea is easy to synthesize: suppose that a particular transition from attractor A to B takes place only when a particular node (say, node 97) is flipped. This event can be too rare to be seriously taken into account. Depending upon the noise level, only those transitions which can take place from a certain number (say 4) of different positions need to be considered. In this case we set the threshold in such a way as to ignore all the transitions which take place only from 1, 2 or 3 different positions only.
Note that we are not considering multiple flips, but only single flips, ignoring those transitions which can take place in a too small number of different ways.

We show that by modifying the threshold the number of TESs changes: when the threshold is 0 the TES coincides with the usual ES, and we recover the result that there is almost always just a single exemplar per net. But by increasing the threshold the single TES breaks into smaller ones. Ultimately, when the threshold is high enough each TES is composed by a single attractor. We refer to TESs composed by two or more attractors as multi-attractor TESs, or multiplet-TESs, and to TESs composed by one attractor only as single-attractor TESs, or singlet-TESs. In the following we will sometimes refer for simplicity to a single-attractor TES as to an "attractor", although of course a TES is logically a set of attractors.

In Section 5 we then propose to associate cell types to TESs. This choice appears natural at this point and presents several advantages:

- TESs are robust to noise;
- there can be several TESs in a single network;
- a TES is a more general concept, which includes, by tuning the threshold, both attractors and ESs, i.e. the other two candidates which have been proposed in the literature to correspond to cell types.

We also show in Section 5 that the identification of cell types with TESs provides a satisfactory picture of some very interesting phenomena related to cell differentiation. In particular, we associate a pluripotent cell to a multi-attractor TES. The idea is that the physical system wanders among the different attractors which compose the TES, spending some time also in the transients between these different attractors. It is known that there are various degrees of differentiation, so we tentatively associate them to the number of different attractors. Therefore a pluripotent cell corresponding to a TES composed by L different attractors is less differentiated than one composed by Q attractors if $L > Q$. Completely differentiated cells correspond to single-attractor TESs.

By changing the threshold the number of TESs changes; in particular, when we increase its value from 0 we see that the initial TES (typically, only one) breaks into a small set of disjoint TESs, which in turn break if the threshold is further increased, as shown in Figure 2.

According to this approach, differentiation is associated to an increase of the threshold. This quantity has been introduced as a mathematical entity, and the question is open to which biological feature (if any) it can correspond. We tentatively suggest that the simplest answer is that increasing the threshold corresponds to lowering the level of noise, i.e. the probability per unit time that a node switches its state, and of course that lowering the threshold corresponds to an increase in the level of noise. This is intuitively plausible, given that the threshold is related to the number of different flips which can lead from an attractor to another, and is quantitatively analyzed in Section 5.

These theoretical considerations are consistent with the experimental observation that pluripotent cells express more genes than differentiated cells, at lower activity levels (Hu et al., 1997), a fact indicating the presence of higher noise levels. So our suggestion appears indeed coherent with known facts. Recent work by Yamanaka and by others (Takahashi and Yamanaka, 2006; Takahashi et al., 2007) have shown that it is possible to transform a

fully differentiated cell into a pluripotent one by overexpressing some of its genes. We have replicated this experiment in silico as follows: starting from a mother undifferentiated cell, described by a multi-attractor TES, we increase the threshold up to the point where at least one single-attractor TES appears (the completely differentiated cell). We then modify the state of one or a few nodes of this attractor to a constant 1 ("overexpression") and, without modifying the threshold any further, we see that the cell switches to a multi-attractor TES. This process exactly corresponds, at our abstract level, to the Yamanaka experiment, and further demonstrates the efficacy of the proposed picture.

The level of noise might be related to the efficacy of the various control and repair mechanisms which operate in the cell. This point is considered in Section 6 where we propose and discuss some hypotheses about the biological mechanisms which can affect the level of noise. These suggestions are tentative ones, and a major purpose of the present paper is that of stimulating further theoretical and experimental work concerning the biological meaning of the threshold and the analysis of the "gene expression noise" level.

So, in Section 6 we discuss also some possible experimental tests which are inspired by our model. Last but not least, we discuss some critical remarks and propose indications for further theoretical work.

2. A quick overview of random Boolean networks and of previous studies

There are excellent reviews of RBNs in literature (Aldana et al., 2003; Kauffman, 1993) so here we will very briefly summarize only some of their properties. A RBN is a directed graph with N nodes, which can assume binary values; $x_i(t) \in \{0, 1\}$ indicates the value of the i-th node ($i = 1...N$) at time t. Each node has k_{in} input connections; in the classical model used here, k_{in} is the same for all nodes and the input nodes are chosen randomly with uniform probability among the remaining nodes (prohibiting loops and multiple connections). To each node a Boolean function f_i is associated, which determines its value at time t from the values of its inputs at the previous time step. The Boolean function associated to a node is chosen randomly with uniform probability among all the possible Boolean functions of k_{in} arguments. Both the topology and the Boolean function associated to each gene do not change in time. The network dynamics is discrete and

synchronous, so fixed points and cycles are the only possible asymptotic states.

Considering the robustness with respect to small changes in initial conditions, the scaling of the cycle period with the system size and other properties, it is possible to distinguish three different dynamical regimes: ordered, critical and chaotic. The most interesting behaviour has been shown by nets in a critical regime, that is the zone of transition between order and chaos. Critical RBNs show an equilibrium between robustness and adaptiveness (Aldana et al., 2007) and it has been suggested that living organisms live in a dynamical regime close to the boundary between ordered and chaotic phases (Kauffman, 1993) For this reason in the present study we will concentrate on critical RBNs.

The dynamical regime of a RBN depends upon the average connectivity of the network and upon the magnetization bias p, which is the probability that the Boolean functions have the outcome 1. Many works have shown that if all Boolean rules are accepted the critical regime holds when $k_{in} = 2$ and $p = 0.5$, and we will focus our study on networks with these parameter values. Recent results support the view that biological genetic regulatory networks operate close to the critical region. In Serra et al. (2004) the authors considered several experiments where a single gene of *S. cerevisiae* has been knocked-out and the expression levels of all genes, in cells with a knocked-out gene have been compared with those in normal, wild-type cells. This experiment has been simulated in silico by comparing the evolutions of two RBNs which start from the same initial conditions, except for the fact that one gene is permanently fixed to the value 0 in the net which simulates knock-out. Introducing the notion of avalanche, which is the number of genes affected by the perturbation, and defining a threshold to binarize the experimental results, one finds that the distribution of avalanches in simulated RBNs accurately reproduce the experimental data. It was later shown by analytical calculations (Serra et al., 2007) that the distribution of avalanches depends upon a single parameter, which also determines whether a network is ordered, critical or chaotic. The good agreement with experimental data suggests that cells are operating in the critical regime or in the ordered region close to the critical boundary (Langton, 1992; Ramo et al., 2006; Serra et al., 2007). Similar conclusions have been obtained in a study which compares the time course of gene expression of $HeLa$ cells with the results of simulated RBNs (Shmulevich et al., 2005).

3. The influence of noise

The reasons why noise has to be taken into account have been extensively discussed in Section 1. So let us consider initially only a very low level of noise, i.e. the flip of a single node. In particular, we consider networks which are in an attractor cycle, and we switch the value of a node chosen at random for a single time step. After that, the network evolves according to its usual deterministic rules until a new attractor is found. Note that the presence of noise can induce jumps from an attractor to another.

Let A_i ($i = 1...M$) be the M attractors of a given network (under the action of the deterministic transition functions), and let A be the set of such attractors[3]. Let us now consider a network that, after a finite transient, is in attractor A_i. We say that A_j is directly reachable from A_i if there is (at least) a node such as the flip of that node at time t (when the system is in attractor A_i) has the effect of bringing (after a transient) the network to the attractor A_j. We also say that A_j is indirectly reachable from A_i if there exists a path from A_i to A_j as a result of more successive single bit flip. For example a first flip may lead the system from attractor A_i to attractor A_b and a second flip (not necessary of the same node) can lead the system from attractor A_b to attractor A_j. An attractor is defined to be reachable from another if it is either directly or indirectly reachable. In symbols, we represent the fact that A_j is reachable from A_i with an arrow: either $A_i \rightarrow A_j$ or $A_j \leftarrow A_i$.

In order to analyze the effects of noise, we generated critical noisy random Boolean networks ($k_{in} = 2$, $p = 0,5$) of different sizes and we characterised their dynamics by finding the whole set of its attractors, or a subset for large networks.

For each attractor found we perturb every node in every phase (one flip at time) and control the identity of the attractor the perturbation is leading to; this information is memorised in an adjacency matrix[4].

[3]From now on in this work, up to Section 5, the term "attractor" will be used uniquely to indicate the asymptotic cycles of the deterministic dynamics of the RBN or, occasionally, a single-state TES (see Section 4). Only in Section 6 the term will be referred to in a broader sense.

[4]For small networks (with $N \leqslant 20$) we test all possible initial conditions, so all the attractors are found. For large networks we explore a predefined numbers of different initial conditions (5000 for networks with N=100 and 10000 for networks with N=200 and N=1000). Note that in these latter networks it can sometimes happen that no attractor is found within the limits of our simulations. However, these occurrences are very few and

From the adjacency matrix we can obtain a graphical representation of the attractors and of their transitions. Figure 1. shows an example of such structure: each vertex represents an attractor of a network; there is an edge from vertex a to vertex b if there exists at least one node belonging to one state cycle of the attractor a that, if perturbed, leads the system to the attractor b. The edge labels correspond to the percentages of the number of attractor's perturbations that lead the system from one attractor to itself or to another one, calculated upon total number of possible attractor perturbations (equal to N times the attractor period).

FIGURE 1

Figure 1 Graphical representation of the transitions between attractors. Each vertex represents an attractor. There exist an edge from vertex a to vertex b if perturbing at least one node in one phase of the attractor a the system goes to attractor b. The edge labels indicates in percentage the frequency of transitions between two attractors. The vertex labels indicates in percentage the estimate values of the attractors' basins.

Analyzing the diagonal of the adjacency matrices, we found that the higher the network size, the higher is (on average) the percentage of perturbations that don't cause the system to leave the attractor; this is true both for the critical and ordered networks (see Table 1). Note also that, as it should be expected, for each size the ordered networks are less sensible to the noisy events.

This observation might lead one to dismiss the effect of noise as relevant only for small networks, but this would be misleading. Let us suppose for simplicity that in a cell each node has a certain probability of changing its state independently from the other nodes. Let b be the probability that a node chosen at random changes its state during the cell lifetime. Then we expect on average a total Nb flips during the cell lifetime. From Table 1 we see that, by enlarging N from 10 to 1000, the probability that the system is taken to another attractor when a particular node is flipped decreases about tenfold in the case of critical nets. But since there are 100 more nodes, the total number of expected transitions per unit time increases tenfold when N is increased from 10 to 1000. Therefore we come to the conclusion that large

are not considered in the adjacency matrix.

N	Ordered nets	Critical nets
10	12,5	19,2
20	10,5	18,5
100	5,1	10,3
200	2,65	6,5
1000	0,29	1,8

Table 1: Mean percentage of perturbations that lead the system to another state cycle (Q) in critical ($k_{in} = 2$, $p = 0.5$) and ordered ($k_{in} = 2$, $p = 0.7$) nets for different size networks.

networks are more susceptible to the transitions induced by flips than smaller ones. Note that his line of reasoning is robust with respect to the details of the calculations. If Q is the mean percentage of perturbations that leads the system in another state cycle, then it is suffices that the ratio Q(M)/Q(N) be smaller than the ratio N/M, with N>M.

Another interesting result that emerged from the simulations regards a behaviour that can be called "phase dependence". Briefly, perturbing a single node, the answer of the system could be different depending upon the attractor phase in which the node is perturbed. This behaviour involves both fixed nodes and oscillating nodes. This phenomenon is reminiscent of the fact that also real cells may behave differently if perturbed in different phases of their cell cycle.

In order to test the robustness of the previous results, we simulated a higher noise intensity by simultaneously flipping two different nodes (chosen randomly and independently). The results, which are shown in Table 2 show that there are no substantial differences between the number of transitions to a different attractor obtained with either one or two flips.

N	Single flip	Double flip
10	36,6	36,3
100	15,1	16,4

Table 2: Mean percentage of perturbations that lead the system to a different attractor in different size networks with different noisy intensity (single bit flip and double bit flip).

4. Ergodic Sets and Threshold Ergodic Sets

On the basis of the results of the previous section, we conclude that attractors (we recall that this term refers here to those of the deterministic dynamics of the RBN) are unstable with respect to random noise. This observation opens the question as to how one can properly characterize the asymptotic states of a noisy RBN.

A partial answer to this question has been provided in (Riberio and Kauffman, 2007) who propose to associate asymptotic states to sets of attractors, rather than to single ones. This is definitely a sound suggestion: under the influence of noise the system wanders among different attractors, so its asymptotic behaviour should be associated to such sets.

Let us first of all recall the definitions given in Section 3 of reachability of an attractor from another one. Let A_i $(i = 1...M)$ be the M attractors of a given network, and let A be the set of such attractors. We say that A_j is directly reachable from A_i if there is (at least) a node such as the flip of that node at time t (when the system is in attractor A_i) has the effect of bringing the network to the attractor A_j. We also define A_j to be indirectly reachable from A_i if there exist a path from A_i to A_j as a result of more successive single bit flip, and we define A_j to be reachable from A_i if it is either directly or indirectly reachable, and we represent reachability by a directed arrow.

Following Kauffman and Ribeiro we now define an ergodic set (ES) as a subset of A composed by attractors which are reachable from any other member of the ES, not necessarily in a single step. Also, the ergodic set is such that (one or more successive) single flips can't make the system leave the set itself. Formally:

$$ES \equiv \{A_i \in A | \exists A_j \in ES, A_j \rightarrow A_i; A_i \rightarrow A_k \Rightarrow A_k \in ES\}$$

A result already outlined in the previous study by Kauffman and Ribeiro (Ribeiro and Kauffman, 2007) and confirmed by several simulations we performed is that, for every size of network, the overwhelming majority of networks present only one ES (with very few exceptions having two ESs); therefore, if we associate ergodic sets to asymptotic states, the model loses the possibility to have many different asymptotic states (and therefore to represent different cell types).

It is however possible to extend the notion to a different set of attractors, keeping the nice properties of ESs with respect to noise, while at the same time allowing multiple asymptotic states. What it even more interesting is

that the extension of the notion of ergodic set seems well founded on the basis of the physics of noise-induced jumps between attractors.

Note that the transitions among attractors which are associated to the flip of just a single specific node could be too rare to occur with appreciable probability during a cell lifetime, and that the same remark applies to those transitions which are associated to the flip of a few nodes. In order to take into account only those transitions which have a reasonable chance to happen and extend in such a way the notion of ergodic set, we introduce the definition of Threshold reachability (T-reachability). We say that A_j is directly T_θ-reachable from A_i with threshold θ if there are at least a fraction θ of different flips, each one of which leads the system, when it is in attractor A_i, to attractor A_j. The extension of the previous notions to those of T_θ-indirect reachability and of T_θ-reachability are trivial. In symbols, we represent the fact that A_j is T_θ-reachable from A_i with threshold θ (directly or indirectly) by means of an arrow: either $A_i T_\theta \rightarrow A_j$ or $A_j \leftarrow T_\theta A_i$.

We can now introduce the concept of *Threshold Ergodic Set*. A Threshold Ergodic Set of attractors (T_θ-ergodic set, or briefly TES_θ) of the network is a subset of A composed by attractors which are T_θ-reachable from any other member of the TES_θ, not necessarily in a single step. Also, the TES_θ is such that (one or more successive) single flips cannot make the system leave the set itself[5]. Formally:

$$TES_\theta \equiv \{A_i \in A | \exists A_j \in TES_\theta, A_j T_\theta \rightarrow A_i; A_i T_\theta \rightarrow A_k \Rightarrow A_k \in TES_\theta\}$$

Within this definition, we can describe an ergodic set as a TES_θ with $\theta = 0$.

Let us now consider what happens when the threshold is varied. Note that the percentage of transitions between attractors given by the adjacency matrix between attractors (see also Figure 1) provides an estimate of the probability that such transition will occur, that is, the probability that a flip of an arbitrary state belonging to a cycle leads the system in another state cycle. So, in order to check which transitions are permitted, we can directly compare the values of the elements of the adjacency matrix with the threshold.

When the threshold is increased from zero to higher values, one typically observes two different phenomena: on the one hand, the initial single TES

[5]If there is no risk of ambiguity is possible to skip the θ symbol to brevity purposes, in these cases we use T instead of T_θ and TES instead of TES_θ

(recall that a TES_0 is just an ES) breaks up into disjoint TESs (see Figure 2), while on the other hand attractors which were not part of the initial TES can give rise to new ones. For very high values of the threshold all the TES_θ are composed by a single attractor (loosely speaking, "all the attractors of the RBN turn out to be TESs").

FIGURE 2

Figure 2. Graphical representations of the transitions between attractors of a network, referred to different value of threshold θ. In the first image $\theta = 0$ and the net has one TES. The second image show the $TES_\theta s$ of the same net with $\theta = 0.02$, whereas the third and the fourth have respectively $\theta = 0.03$ and $\theta = 0.04$.

Figure 2 indeed shows that there can be several TESs in the same network. The way in which the ratio η between the number of TES_θ and the total number of unperturbed networks' attractors NA ($\eta = TES_\theta/NA$) changes as a function of the threshold and of various network sizes is shown in Figure 3. Figure 4 shows the variation of η with respect to N for different fixed threshold values.

FIGURE 3

Figure 3. The fraction η of total possible TESs in the network over the total number of attractors N_A, with respect to the variation of the threshold θ: the higher is N, the lower are the values of θ able to separate all the attractors. All points are averages of 65 different networks, with the exception of the nets having $N = 1000$ for which only 20 networks are analysed.

FIGURE 4

Figure 4. The variation of η with respect to the variation of the net size N, for different fixed value of threshold.

5. Threshold ergodic sets and cell types

After elucidating the behaviour of TESs in the previous section, we are now ready to propose to associate cell types to TESs, rather than to ESs or attractors.

The reasons for this proposal, which have been already outlined in Section 1, are several.

First of all, TESs are attractors of the dynamics, so all the reasons in favour of associating cell types to attractors hold also for them. Contrary to attractors, however, TESs are sets of attractors and are stable with respect to the noise induced by bit flips.

Ergodic sets also share this property, but they have the undesirable properties of being just one or two per net, so they cannot be associated to several different cell types. It may be possible to remedy at least in part this problem by defining some nodes to be immune by noise, but the biological grounding of this hypothesis is nor completely clear. On the contrary, there can be several kinds of TESs in the same network, and moreover their introduction seems naturally grounded in the physics of the phenomena under study.

Recall indeed that the "gene expression noise" we refer to is not simply thermal motion, so the smallest noise event is the change of state of a single gene. This is unlikely to happen very frequently in cells, so it is definitely appropriate to ignore those transitions among attractors which are too rare. In a large cells population also rare transitions may occur, but in all the experimental studies on gene expression (e.g. on DNA microarrays) the results concern the behaviour of large populations, and are likely to be insensitive to the presence of a few "unusual" types.

As it was pointed out (Huang, 2009; Huang et al., 2009), identifying attractors with cell types presents several advantages with respect to the more conventional views, and all these advantages are retained in referring to TESs rather than to attractors (with the added benefit of robustness with respect to flips).

We will now show that this approach has also some major advantages, concerning in particular the important and intriguing phenomenon of cell differentiation. It is well known (Huang, 2009) that there some cells which are pluripotent, i.e. they can differentiate under proper stimuli giving rise to different cell types. It is also known that there is a sort of hierarchy of pluripotency, ranging from the totipotent cells (which can differentiate to any cell type) to various degrees of pluripotency and ultimately to fully

differentiated cell types.

We propose here to associate pluripotent cells to multi-attractor TESs, and to associate fully differentiated cells to single-attractor TESs. The reasons for this choice appear rather straightforward: in this approach pluripotent cells are described by a dynamical system that wanders among its attractors, and each attractor (or subset of attractors) describes a possible fate of cell differentiation.

Differentiation of a pluripotent cell, described by multi-attractor TES, should then be associated an increase in the threshold; as we have seen in Section 4 this leads to a breakup of the multiplet-TES into smaller ones, eventually leading to singlet-TESs.

The threshold has been considered so far as a mathematical device to induce change in the landscape of TESs, but as it has been discussed in Section 4 it has a deep physical meaning: a low threshold means that we have to take into account even those transitions which are led by a few flips, and this should be important when the noise level is high. In turn, high threshold means that we consider only those transitions among attractors which can take place in many different ways, and this corresponds to the case of low noise level. A crude, order-of-magnitude calculation is the following.

Let ϕ be the frequency of flips in the state of an arbitrary gene (we suppose this is the same for all the genes of the cell), and let us suppose for simplicity that all the attractors are fixed points (in this case the extension to oscillating states only complicates the formulae without adding any new insight). If the typical lifetime is T and N is the number of genes, then $NT\phi$ is the expected total number of flips in the whole cell lifetime. Suppose that the transition between attractor A and attractor B can be achieved by flipping q different nodes; then in a cell lifetime $qT\phi$ transitions of this kind are expected. In order to have a transition we need that this expected value be at least one, so we set $qT\phi = 1$ which implies $\theta = 1/NT\phi$ (recalling that $\theta = q/N$). Therefore the threshold θ scales as the reciprocal of the noise level ϕ^{-1}.

Given that here "noise" means change in the state of a node from active to inactive or vice versa, one is tempted to relate the level of noise to the efficiency of various control mechanisms which are at work in the cell. This aspect will be discussed in Section 6. Let us here just recall that the noise level in relatively undifferentiated cells is likely higher than that which is observed in fully differentiated ones (Hu, 1997), an observation which is coherent with our view of associating differentiation to an increase of the

threshold.

We also remark that it has been observed that from a population of pluripotent cells which are treated for differentiation, different cell types may form. This is exactly what we can expect on the basis of our approach, as it can be seen e.g. by looking at Figure 2 , where a pluripotent cell (the left most multiplet-TES) can break in different ways. While Figure 2 shows all the TESs of that RBN, a specific cell must of course choose one, so we expect to find a combination of different types.

A major experimental finding in recent years (Takahashi and Yamanaka, 2006; Yamanaka et al., 2007) has been the discovery of the possibility to revert a fully differentiated cell to a pluripotent one by overexpressing some of its transcription factors. We show here that the same can happen in our model in some particular cases. This is not a generic behaviour, but neither is it generic in real cells; what matters is the possibility to simulate with this model the experimental results.

Note that reverting from a differentiated cell to a pluripotent one would be trivial if one were allowed to change the threshold, but here we try to achieve the same goal simply by overexpressing some genes. In our Boolean model, overexpressing means that the activation of a gene which is oscillating (or constantly inactive) in a given attractor is kept constantly equal to 1.

We considered a network composed by 10 nodes, whose attractors and the corresponding transitions are represented in Figure 5.

FIGURE 5

Figure 5. One TES_0 composed by all the nodes (four different attractors). The rightmost attractor is called A_0, the others are A_1, A_2 and A_3 numbered counterclockwise.

By inspection of Figure 5 one can observe that, if $q = 0$, there is a single TES which comprises all the attractors. By increasing the threshold we reach the following situation:

FIGURE 6

Figure 6. One $TES_{0.11}$ composed by the red attractor A_0

Here there is a single TES, composed only by the rightmost attractor, which corresponds to a fully differentiated cell. Now, by fixing to 1 the value

of one of the genes of that attractor, we have *de facto* created a new network (indeed fixing to 1 the value of a gene is equivalent to modifying its Boolean function so that it is always 1 irrespective of its inputs). We then analyze the attractors of this perturbed network and their transitions which are shown in Figure 7.

FIGURE 7

Figure 7. The $TES_{0.11}$ composed by the red nodes related to the perturbed net. The rightmost attractor is called A'_0, the others are A'_1 .. A'_4 numbered counterclockwise.

Note that the transitions and their frequencies refer to the same high threshold value of Figure 6, but here a multiple-TES appears. This exactly corresponds, in our model, to the appearance of a pluripotent cell by over-expressing some genes of a fully differentiated one. Note also that most attractors of the new network correspond to those of the previous one, in particular $A_1 = A'_2$, $A_2 = A'_3$, $A_3 = A'_4$. So not only we have obtained a TES, but one which is very similar to the one which described the initial pluripotent cell.
Note that a similar result has been obtained by Yamanaka and Takahashi (Takahashi and Yamanaka, 2006): when the transcription factors of the fully differentiated fibroblasts are overexpressed, an induced pluripotent stem cell (iPS) is obtained which is similar but not identical to the original embryonic stem cell.

6. Discussion and indications for further work

The present work consists of two main parts, a mathematical analysis of the asymptotic dynamics of RBNs subject to noise, and a proposal to associate cell types to TESs. In this final section we give indications for further work in both directions, and we present a more detailed discussion of the role of noise and threshold in cell differentiation
Concerning the mathematical properties of noisy RBNs, one should note that the flip represents a particular type of transient perturbation, but in cells there can be also perturbations which last longer than a single time step. One can model these semipermanent perturbations by fixing the value of a node chosen at random to a given value, letting the network evolve to its new attractor, and then removing the constraint on the chosen node. Semipermanent perturbations can be associated for example to the exposure to some

environmental conditions or to some chemicals which are later removed, so the interesting questions concern the behaviour of the cell during exposure and after its removal. There can be also permanent perturbations, associated to never-ending exposure to chemicals or environmental conditions, and also to some kinds of mutations. The study of the behaviour of the model under permanent and semipermanent perturbations lies however outside the scope of the present paper.
Moreover, an interesting study might concern the effects of more frequent flips, that can be simulated by fixing a certain probability per unit time that a node is switched, without allowing the network to reach an attractor before a new flip takes place.
There are of course other interesting venues for further work, including e.g. the study of the behaviour of random threshold networks (Andrecut et al., 2009) rather than RBNs under the influence of noise and the analysis of the influence of the topology .

As far as the biological interpretation of the model is concerned, undifferentiated cells have been associated to systems which wander among a set of attractors under the action of noise, and differentiation to the freezing of these states in a smaller region of phase space.

Let us remark that, while in this study we made use of random Boolean networks subject to noise, the overall picture of cell types and differentiation can be much more general and applicable to different dynamical models of the gene regulatory networks (e.g. continuum models). And there are indeed some models of this kind (Kaneko, 2006) which share the idea that differentiation is a process where the portion of phase space available to the network gets smaller as it proceeds from a less differentiated to a more differentiated form.

There are of course important differences between continuous and discrete models, like for example whether the wandering in phase space is attributed to chaotic dynamics or to random fluctuations, but the general pictures are similar. And, as it was already pointed out in the Introduction, the fact that the same view holds in different kinds of models is an indication in favour of its validity.

Let us also observe that the present model deals with an isolated cell, while it is known that intercellular communication can play a key role in differentiation. It is however important to elucidate which mechanisms might be at work at the single cell level, so the present work can be seen as a first

step towards the building of a more comprehensive picture of differentiation[6].

Let us now consider in more detail the issue of cellular noise, which is crucial for our association of differentiation to modifications of the threshold which in turn, as it has been shown in section 4, corresponds to modifications of the noise level.

It has indeed been observed that the intracellular states of pluripotent or multipotent stem cells are heterogeneous and that in some cases gene expression levels slowly itinerate over several quasi-stable states (Chang et al., 2008; Furusawa and Kaneko, 2009). This is what should be expected on the basis of the model described here (a similar behaviour could also be expected in models where the wandering is attributed to chaotic itineracy).

A major point concerns the comparison of noise levels in undifferentiated vs differentiated cells. In this respect, there are experimental findings which suggest that changes in gene expression levels are indeed higher in the former case (Furusawa and Kaneko, 2009; Hayashi et al., 2008).

Moreover, it has been observed that in multipotent cells a higher number of genes is activated, albeit at a lower level than that of differentiated cells. This is indeed what should be expected by using the TES framework, where multipotent systems wonder through several asymptotic states (belonging to the same TES) with high levels of noise (i.e. low threshold values). In fact, it has to be observed that a low expression level (like that found in undifferentiated cells) implies a smaller number of mRNA molecules, and that the relative role of fluctuations is higher when the number of molecules is small. Turning this into a Boolean description, this corresponds exactly to a higher level of noise in undifferentiated cells, as we have supposed here.

An important issue concerns the mechanisms whereby a change in the level of noise can be achieved. While elucidating this point requires further study, the observation made above suggests that a possible way can be that of increasing the level of the signal with respect to that of the fluctuations. Although we do not know exactly in which way(s) this is done, one can suggest that, for expressed genes, this could be achieved by an increase of the speed with which they are transcribed and translated into proteins, and also by reducing the rate at which the proteins are degraded (this is controlled by specific enzymes and is therefore also under genetic control). Concerning

[6]an analysis of the behaviour of the attractors in a cellular automata model of coupled RBNs, simulating a tissue, has been presented elsewhere (Villani et al., 2006)

those genes which have to be silenced, a reduction of the noise level could be accomplished e.g. by increasing the rate of production of miRNAs which inhibit the expression of the targeted genes (Hornstein et al., 2006).

Let us point out that one of the major outcomes of our study is that of highlighting the importance of changes in the noise level to achieve differentiation. Note indeed that, according to our model, it is impossible to induce a transition to a more differentiated state by only activating or inhibiting a particular gene. This has been shown in several numerical experiments, and it can be explained by observing that a network with a gene which takes a constant value coincides with a RBN similar to the original one, where the only difference is that one Boolean function has been forced to take always the TRUE or FALSE value. Although this network is no longer completely random, if there are several genes this change has a minor effect, and so, if the threshold is 0, one almost always finds just a single TES (or rarely two) - and in this case no breakup of the original TES into more disjoint TESs is observed. The same observation can be made for the case where a few genes are forced to take a constant value.

Therefore, if our view is correct, a change of the threshold, i.e. of the noise level, necessarily accompanies the differentiation process. To the best of our knowledge this aspect has not been emphasized before; moreover, it suggests the importance of performing further experiments targeted to quantitatively analyze the way in which these changes occur (for example, by measuring the fluctuations in gene expression levels at various degress of differentiation).

Let us finally observe that also cancer is related to a change to a less differentiated state (Huang et al., 2009), so we hypothesize that this phenomenon could also be amenable to study by dynamical simulation and Threshold Ergodic Sets.

Acknowledgments

This work has been partially supported by the Italian MIUR-FISR project nr. 2982/Ric (Mitica).
Stimulating discussions with Sui Huang and Andrei Ribeiro are gratefully acknowledged. We also thank an anonymous referee for bringing to our attention important results on stem cells.

References

References

Aldana, M., Coppersmith, S., Kadanoff, L.P., 2003. Boolean dynamics with random couplings. in Kaplan, E., Marsden, J.E. & Sreenivasan, K.R. (Eds), Perspectives and problems in nonlinear science. A Celebratory Volume in Honor of Lawrence Sirovich. Springer Applied Mathematical Sciences Series.

Aldana, M., Balleza, E., Kauffman, S. A., Resendiz, O., 2007. Robustness and evolvability in genetic regulatory networks. Journal of Theoretical Biology. 245, 433-448.

Andrecut, M., Foster, D., Carteret, H., Kauffman, S.A., 2009. Maximal Information Transfer and Behavior Diversity in Random Threshold Networks. Journal of Computational Biology. 16 (7), 909-916., doi:10.1089/cmb.2009.0007.

Barbieri, A., Villani, M., Serra, R., Kauffman, S. A., Colacci, A., 2009. The influence of noise on the dynamic of random Boolean networks. In Morabito, C. (ed): Proceedings of Wirn 2009. Amsterdam: IOS Press (in press).

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., Huang, S., 2008. Transcriptome-wide noise controls lineage choice in mammalian protenitor cells. Nature. 453, 544-548.

Furusawa, C., Kaneko, K., 2009. Chaotic expression dynamics implies pluripotency: when theory and experiment meet. Biology Direct. 4 (1), 17., doi:10.1186/1745-6150-4-17.

Glass, L., Hill C., 1998. Ordered and disordered dynamics in random networks. Europhys. Lett., 41 (6), 599-604., doi: 10.1209/epl/i1998-00199-0.

Hayashi, K., de Sousa Lopes, S.M.C., Tang, F., Surani, M. A., 2008. Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. Cell Stem Cell. 3 (4), 391-40., doi:10.1016/j.stem.2008.07.027.

Hornstein E., Shomron N., 2006. Canalization of development by microRNAs. Nature Genetics. 38, Suppl:S20-S24., doi:10.1038/ng1803.

Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C., Enver, T., 1997. Multilineage gene expression precedes commitment in the hemopoietic system. Genes Dev. 11 (6), 774-85.

Huang, S., 2009. Reprogramming cell fates: reconciling rarity with robustness. Bioessays. 31 (5), 546-560., doi: 10.1002/bies.200800189.

Huang, S., Ernberg, I., Kauffman, S. A., 2009. Cancer Attractors: a systems view of tumors from a gene network dynamics and developmental perspective. Seminars in Cell & Developmental Biology. 20 (7), 869-876., doi:10.1016/j.semcdb.2009.07.003.

Kaneko, K., 2006. LIfe: an introduction to complex system biology. Springer-Verlag, Berlin, New York.

Kauffman, S. A., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. Journal of Theoretical Biology, 22 (3), 437-467., doi: 10.1016/0022-5193(69)90015-0.

Kauffman, S. A., 1993. The origins of order. Oxford University Press.

Kauffman, S. A., 1995. At home in the universe. Oxford University Press.

Langton, C.G., 1992. Life at the edge of chaos. In Artificial Life II, C.G. Langton, C.Taylor, J.D. Farmer, and S. Rasmussen (eds), Reading MA:Addison-Wesley, pp. 41-91.

Ramo, P., Kesseli, J., Yli-Harja, O., 2006. Perturbation avalanches and criticality in gene regulatory networs. Journal of Theoretical Biology. 242 (1), 160-164.

Ribeiro, A. S., Kauffman, S. A., 2007. Noisy attractors and ergodic sets in models of gene regulatory networks. Journal of Theoretical Biology. 247 (4), 743-755., doi:10.1016/j.jtbi.2007.04.020.

Serra, R., Villani, M., Semeria, A., 2004. Genetic network models and statistical properties of gene expression data in knock-out experiments. J. Theor. Biol. 227 (1), 149-157., doi:10.1016/j.jtbi.2003.10.018.

Serra, R., Villani, M., Graudenzi, A., Kauffman, S. A., 2006. On the distribution of small avalanches in random Boolean networks, in Ruusovori, P.

et al. (eds.): Proceedings of the 4th TICSP workshop on computational systems biology, 93-96. Tampere: Juvenes print.

Serra, R., Villani, M., Graudenzi, A., Kauffman, S.A., 2007. Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data. Journal of Theoretical Biology. 249 (3), 449-460., doi:10.1016/j.jtbi.2007.01.012.

Shmulevich, I., Kauffman, S.A., Aldana, M., 2005. Eukaryotic cells are dynamically ordered or critical but not chaotic. PNAS. 102 (38), 13439-13444., doi: 10.1073/pnas.0506771102.

Takahashi, K., Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblasts Cultures by Defined Factors. Cell. 126 (4), 663-676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., Yamanaka, S., 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell. 131 (5), 861-872.

Villani M., Serra, R., Ingrami, P., Kauffman, S.A., 2006. Coupled random boolean networks forming an artificial tissue. In El Yacoubi, S., Chopard, B. & Bandini, S. (eds) Berlin: Springer Lecture Notes in Computer Science 4173, 548-556.

Figure1

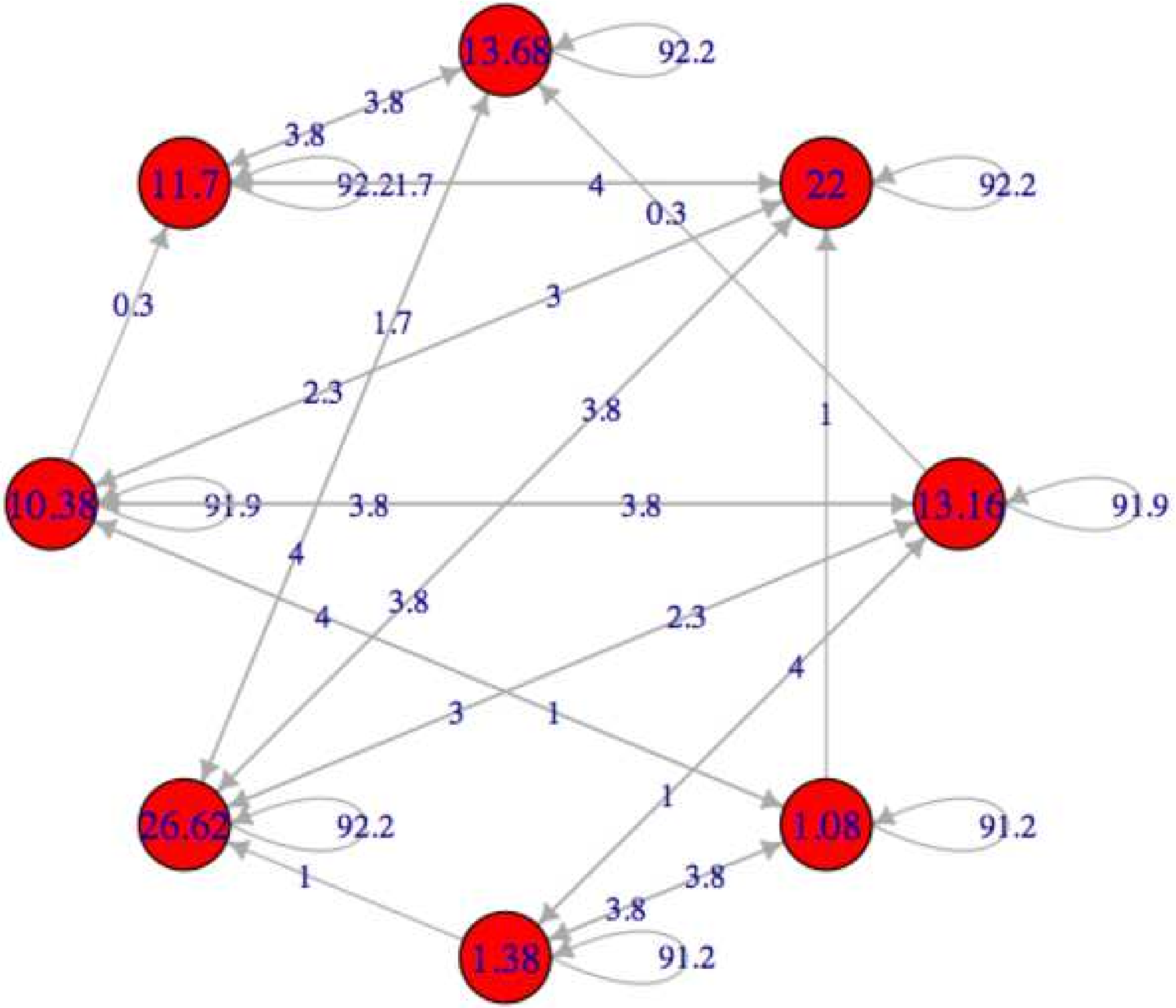
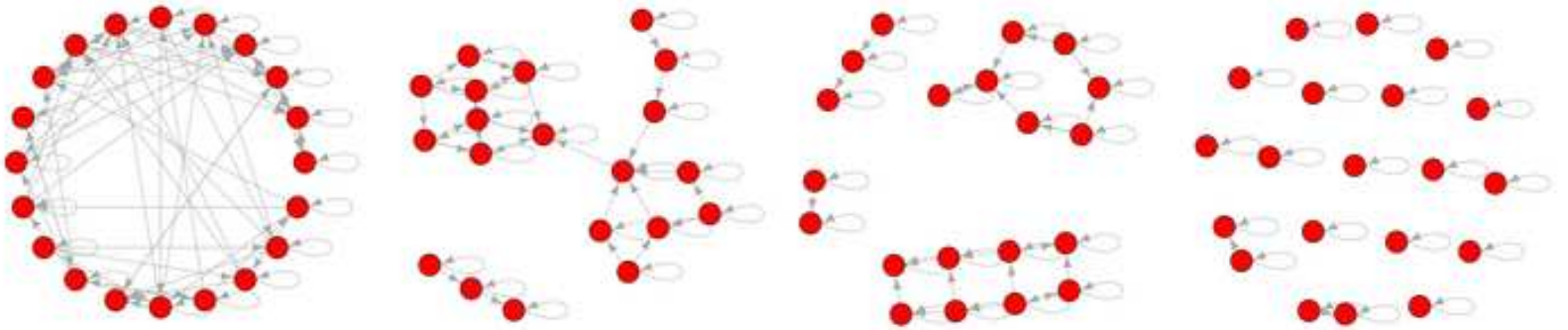

13.68
92.2
3.8
3.8
11.7
92.21.7
4
22
92.2
0.3
0.3
3
1.7
2.3
3.8
1
10.38
91.9
3.8
3.8
13.16
91.9
4
3.8
4
2.3
4
3
1
1
26.62
92.2
1.08
91.2
3.8
3.8
1
1.38
91.2

figure2



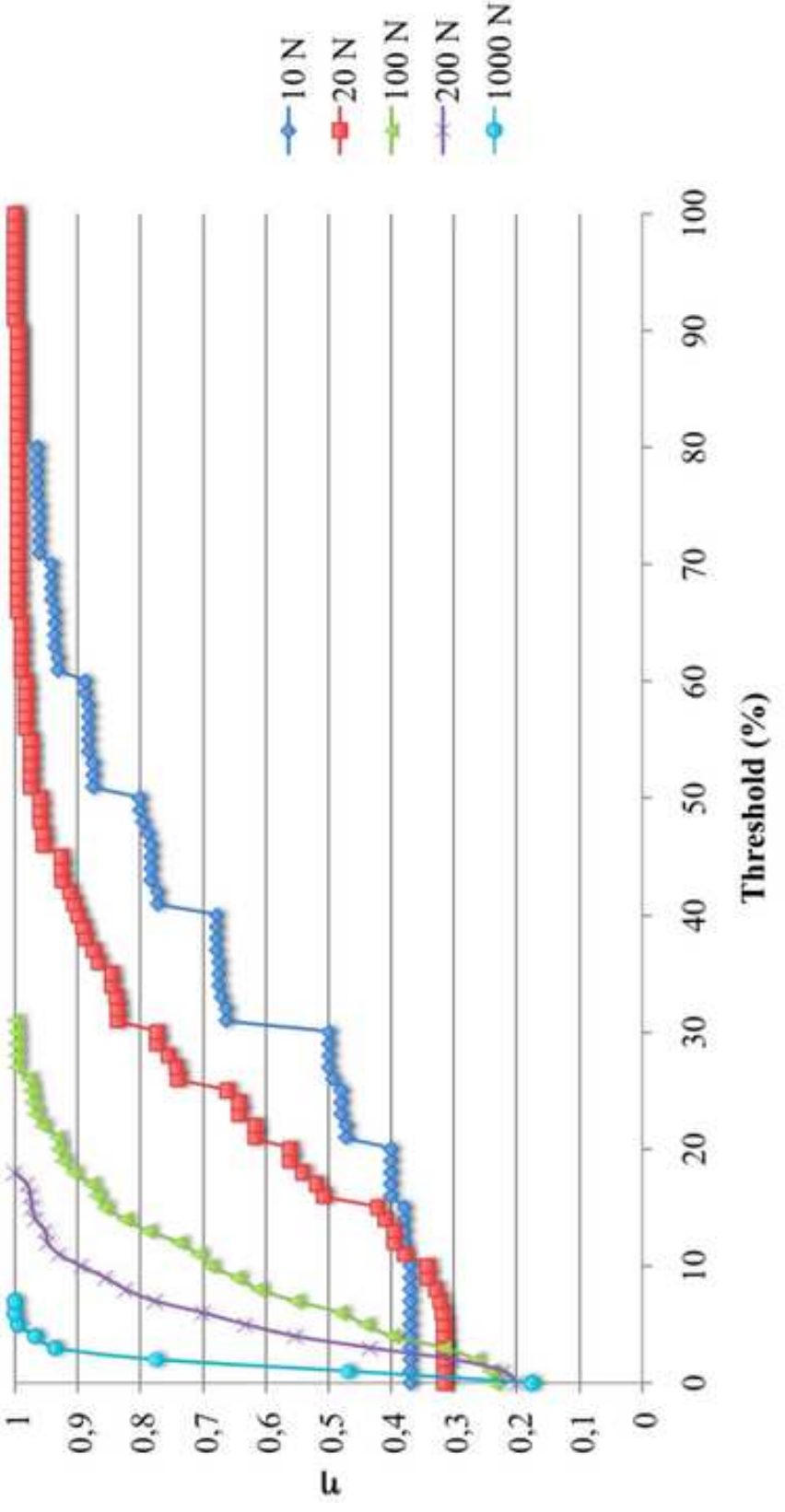


figure3

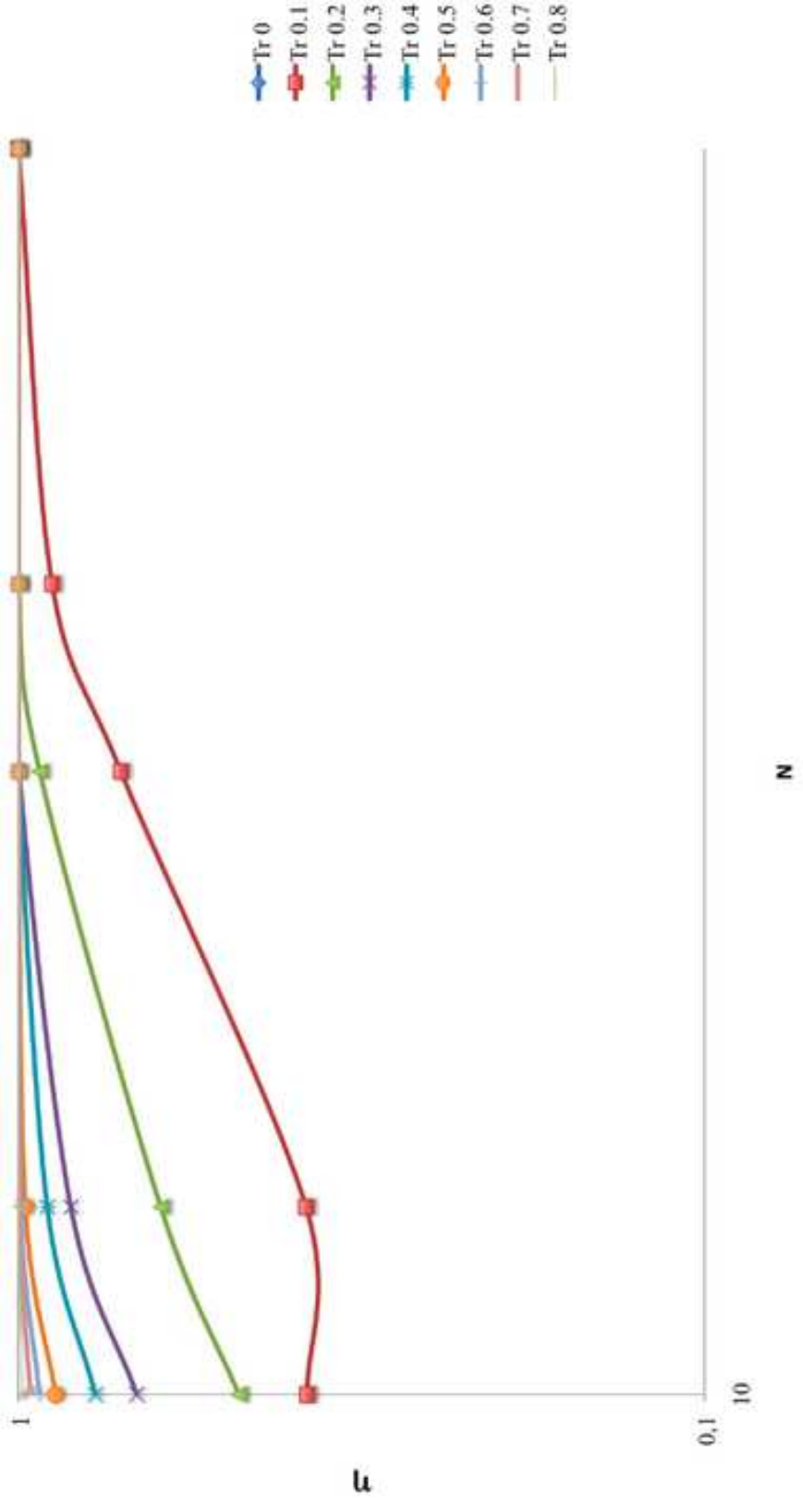


figure4

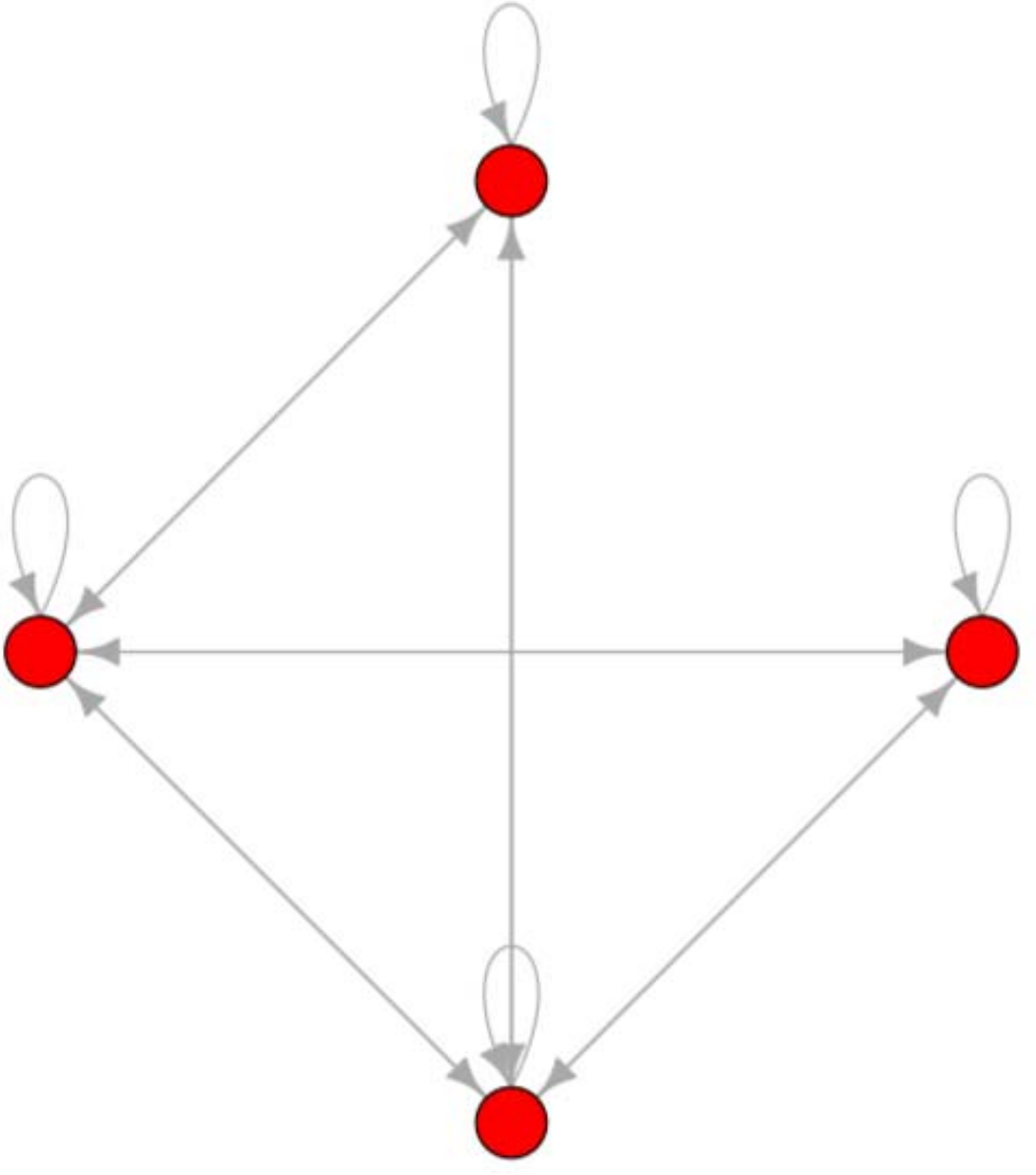


figure5

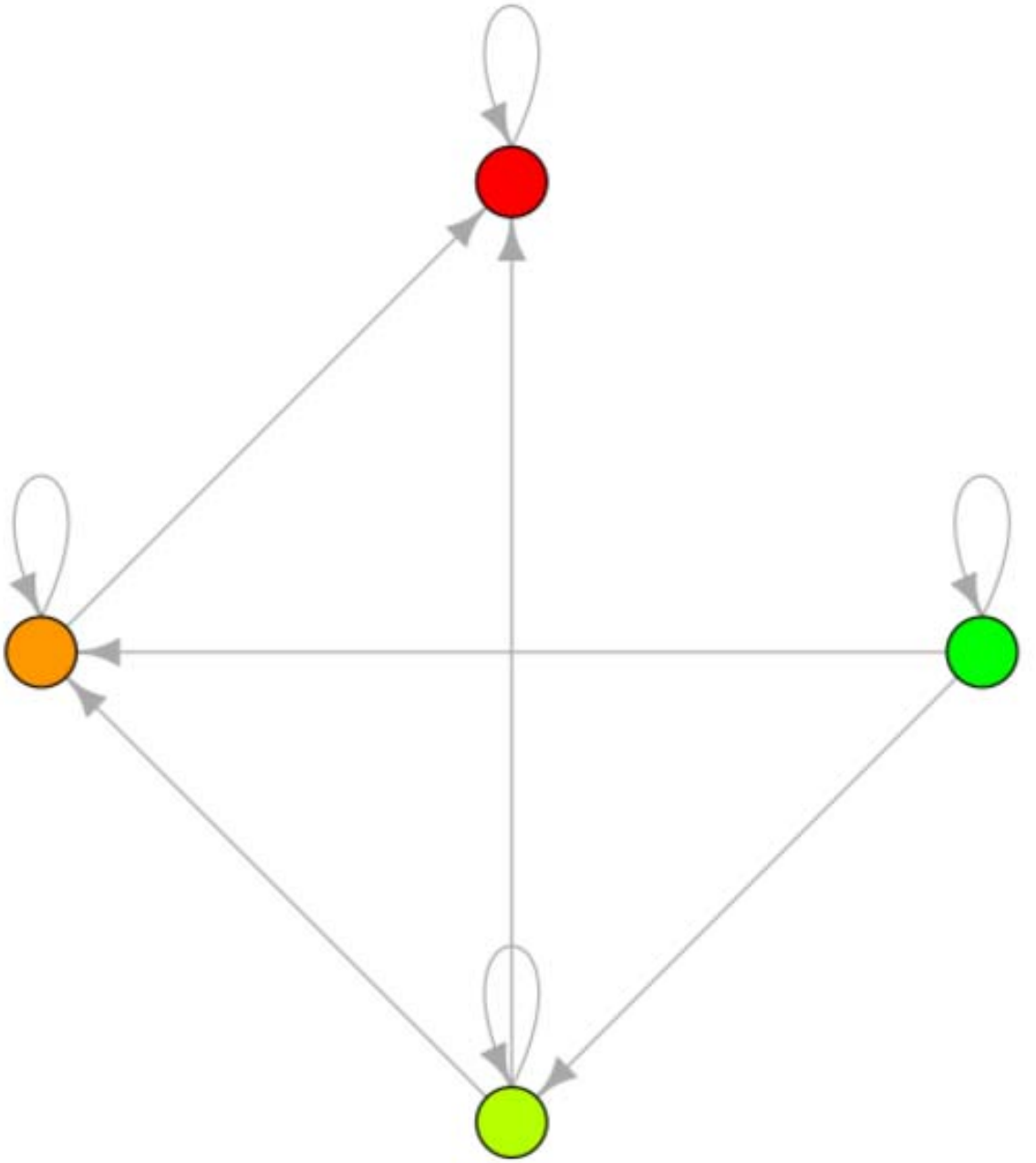


figure6

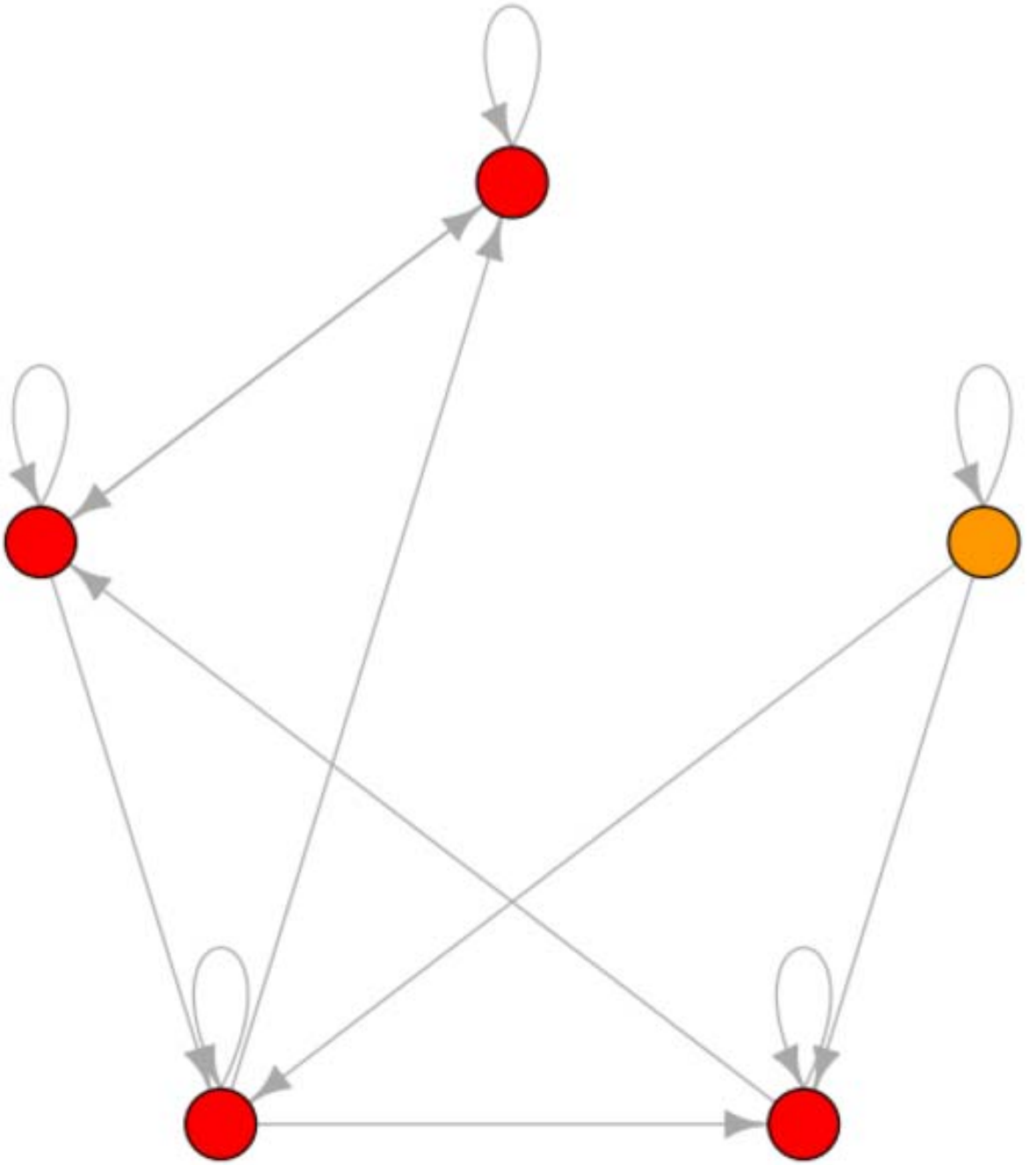


figure7

Table1

N	ORDERED NETS	CRITICAL NETS
10	12,5	19,2
20	10,5	18,5
100	5,1	10,3
200	2,65	6,5
1000	0,29	1,8

Table2

N	SINGLE FLIP	DOUBLE FLIP
10	36,6	36,3
100	15,1	16,4