

Published in IET Systems Biology
 Received on 29th November 2007
 Revised on 20th May 2008
 doi: 10.1049/iet-syb.2007.0070



Inference of Boolean networks under constraint on bidirectional gene relationships

G. Vahedi¹ I.V. Ivanov² E.R. Dougherty^{1,3}

¹Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA

²Department of Veterinary Physiology and Pharmacology, Texas A&M University, College Station, TX 77843, USA

³Translational Genomics Research Institute, 400 North Fifth Street, Suite 1600, Phoenix, AZ 85004, USA

E-mail: golnaz@tamu.edu

Abstract: The coefficient of determination (CoD) has been used to infer Boolean networks (BNs) from steady-state data, in particular, to estimate the constituent BNs for a probabilistic BN. The advantage of the CoD method over design methods that emphasise graph topology or attractor structure is that the CoD produces a network based on strong predictive relationships between target genes and their predictor (parent) genes. The disadvantage is that spurious attractor cycles appear in the inferred network, so that there is poor inference relative to the attractor structure, that is, relative to the steady-state behaviour of the network. Given steady-state data, there should not be a significant amount of steady-state probability mass in the inferred network lying outside the mass of the data distribution; however, the existence of spurious attractor cycles creates a significant amount of steady-state probability mass not accounted for by the data. Using steady-state data hampers design because the lack of temporal data causes CoD design to suffer from a lack of directionality with regard to prediction. This results in spurious bidirectional relationships among genes in which two genes are among the predictors for each other, when actually only one of them should be a predictor of the other, thereby creating a spurious attractor cycle. This paper characterises the manner in which bidirectional relationships affect the attractor structure of a BN. Given this characterisation, the authors propose a constrained CoD inference algorithm that outperforms unconstrained CoD inference in avoiding the creation of spurious non-singleton attractor. Algorithm performances are compared using a melanoma-based network.

1 Introduction

Various models have been proposed for gene regulatory networks [1] and inference from gene expression data has played a major role. Perhaps the key issue concerning network inference is the large space of networks from which a model must be selected in relation to the amount of data typically available. This dimensionality problem drives inference in two directions: (1) towards coarse-grained models that require less data for inference [2], and (2) application of biological constraints [3]. This paper concerns Boolean networks (BNs) and several inference methods have been proposed [4–7]. These methods generally assume time-course data; however, here we are concerned with inference from time-independent data, the kind of data one typically obtains from microarray studies

involving human subjects. In this context, it is generally assumed that the data come from the steady state of the network. An example of inference of gene regulatory networks from steady-state data using Bayesian networks can be found in [8].

The long-run behaviour of a BN is characterised by its attractor cycles: given an initial state, the network will transition into a cycle of states and remain cycling forever. The attractor cycles are associated with functional states on physiological timescales and cell types on developmental timescales [9, 10]. In the absence of time-course data, we assume that we are sampling from the steady state. Furthermore, we assume that the data states represent singleton attractors.

One method proposed for inference of BN from steady-state data involves the coefficient of determination (CoD) [11]. Given a set of predictor variables and a target variable to be predicted, the CoD measures the relative decrease in prediction error when using the predictor variables in comparison to using the best estimate of the target in the absence of knowledge concerning other variables. The CoD was the first method used to infer probabilistic Boolean networks (PBNs), each of which is constructed from a collection of BNs [12].

A fundamental issue is that, without time-course data, the CoD cannot provide information on the direction of prediction. This problem manifests itself in the situation where, if gene *a* is a high-CoD predictor of gene *b*, then gene *b* is typically a high-CoD predictor of gene *a*. We refer to this situation as a bidirectional relationship between genes *a* and *b*. Not only do bidirectional relationships impact the inference process by producing spurious connections in the regulatory graph of the network, thereby deleteriously affecting the graph topology of the inferred network, they also affect the attractor structure to the extent that the inferred network possesses spurious attractor cycles. The problem is sufficiently troublesome that it has suppressed use of the CoD inference method. The inference methods that have taken its place are primarily based on the attractor structure [13, 14] or graph topology [15, 16]. In the methods that emphasise attractor structure, the key concern is that the inference algorithm produces an attractor structure close to that of the true network, and in the methods that emphasise graph topology, the key concern is the agreement between the graph connections, for instance, as measured by the Hamming distance between the regulatory graphs [16]. While these methods tend to achieve their goals, they pay only secondary concern for the functional relations among genes in the network, that is, the ability to predict the value of a specific target gene based on the values of a set of regulatory genes for that gene.

In this paper, we will accomplish both goals in network design: preservation of attractor structure and connectivity based on strong gene prediction. To accomplish this aim, we investigate the bidirectional effects for BNs with connectivity $K = 1$ and $K = 2$, the connectivity of a BN being the maximum number of variables allowed for a Boolean function. It should be noted that when we set the connectivity to a fixed value K , genes in the network will have at most K predictors. In contrary to random BNs, in this work genes are not required to have identical number of predictors. Kauffman predicted low connectivity for gene regulatory networks [10]. We assign the maximum connectivity of $K = 3$ in this work, although we investigate bidirectional effects for BNs with connectivity $K = 1$ and $K = 2$. As a consequence of our analysis, we propose a novel constrained CoD-based inference algorithm that performs significantly better than unconstrained CoD inference relative to the attractor structure.

We should point out that two genes with bidirectional relationship are a special case of relevant nodes. Relevant nodes are those genes that can influence themselves via a loop of connections [17]. Genes with bidirectional relationships are relevant nodes in which the length of the corresponding loop is 2. A significant body of work on random BNs investigates the effect of relevant nodes on the attractor structure of BNs [18–20]. However, in the random BNs considering relevant nodes the number of inputs is the same for all nodes [19], whereas the analytical investigations in this work consider the general case where the number of inputs for every node is not the same and study the effect of bidirectional relationships on the attractor structure of BNs.

We will begin by briefly reviewing BNs and defining the bidirectional relationship among two genes of a network. We then investigate the effect of such relationships on the attractor structure of specific classes of BNs. After discovering how the bidirectional relationships influence the attractor structure of a BN, and providing estimates of encountering such relationships and, particular, attractor structures, we discuss CoD-based inference. We propose a novel algorithm that mitigates bidirectional relationships and provides simulation results to confirm our analysis. Finally, we present an application of the proposed algorithm to melanoma gene expression data and compare its performance to unconstrained CoD inference procedures.

2 Systems and methods

2.1 Bidirectional relationships

A BN $G(V, F)$, [9], consists of a sequence $V = \{x_i\}_{i=1}^n$ of n nodes, where $x_i \in \{0, 1\}$, and a sequence of Boolean functions $F = \{f_1, \dots, f_n\}$. In gene regulatory modelling, x_i represents the expression level of gene i , which can be either active (1) or inactive (0). As is commonly done, we will mix terminology by referring to the nodes as genes. The set of Boolean functions F represents the regulatory rules among genes. At time step $t + 1$, the expression of gene x_i , called target gene, is predicted by the expression of a set, W_i , of genes at time step t , based on the regulatory function f_i . The set of genes $W_i = \{x_{i_1}, \dots, x_{i_{k_i}}\}$ is called the predictor set of x_i . The function f_i is called the predictor function of x_i . We assume that there are no non-essential genes in a predictor set, meaning that the predictor function requires the full set as input. The cardinality, $|W_i|$, of W_i is called the connectivity of x_i and the maximum order of the genes is called the connectivity of the network.

A state of a BN at time t is a vector $[x_1(t), \dots, x_n(t)]$ of gene values. The possible states of the BN form its state space. Given an initial state, the network will eventually enter a set of states through which it will repeatedly cycle forever. Each such set is called an attractor cycle, and a singleton attractor is an attractor cycle of length 1. The

attractor cycles are mutually disjoint. The set of all states that transition into an attractor cycle is called the basin of that cycle. The family of basins partitions the state space of the BN.

Our particular interest is with how genes that are predictors of each other affect the attractor structure. As noted in the Introduction, when such pairs arise on account of network inference, they can lead to the existence of certain attractor structures. This motivates the following definition.

Definition 1: The genes x_i and x_j in a BN are said to have a bidirectional relationship if $x_i \in W_j$ and $x_j \in W_i$. The relationship is said to be of connectivity n if $|W_i| = |W_j| = n$.

To say that x_i and x_j have a bidirectional relationship of connectivity n is to say that they have a bidirectional relationship and each has connectivity n . Alternatively, one might have defined the relationship to be of connectivity n if n is the maximum of $|W_i|$ and $|W_j|$, or to be of connectivity (m, n) if $|W_i| = m$ and $|W_j| = n$, the rationale behind the first alternative being to bound the complexity of the predictor relations and the second being to specify directly the predictor-set cardinalities. We have defined order n as we have because it characterises the most complex case when one of the predictor sets has cardinality n . It is this maximum complexity that interests us.

We will investigate the effect of bidirectionality on the attractor structure, provide estimates of how often such bidirectional relationships happen, and derive a lower bound estimate for the probability of a BN with such relationships having at least one non-singleton attractor cycle. We first consider connectivity 1 and show that there is at least one non-singleton attractor cycle in the BN. Next we consider connectivity 2. There we will see that even such a minimal increase of the cardinality of the predictor sets complicates the analysis of the attractor structure. Proofs of the propositions can be found in the Appendix of the paper.

2.2 Connectivity-1 bidirectionality

Proposition 1: If there are two genes in a BN having a bidirectional relationship of connectivity 1, then the BN has at least one non-singleton attractor cycle.

2.3 Connectivity-2 bidirectionality

Proposition 2: If a BN possesses a pair of genes that have a bidirectional relationship of connectivity 2, then at least $1/8$ of the states in its state space cannot be singleton attractors of the network.

While the preceding two propositions explain why BNs with bidirectional relationships possess significant numbers of states that cannot be singleton attractors, one should not

jump to the conclusion that the presence of bidirectional relationships is equivalent to the absence of singleton attractors in a BN. The next proposition explains why the bidirectional relationship of connectivity 2 is not a sufficient condition for the absence of singleton attractors.

Proposition 3: In the class of BNs possessing a pair of genes with bidirectional relationship of connectivity 2, there exists a BN with at least one singleton attractor cycle.

The above considerations show that while the bidirectional relationship between pairs of genes in a BN is an important condition for the presence of non-singleton attractors, it is not a sufficient one, especially in the case of predictor sets of cardinality greater than 1. In the next two sections, we consider the issue of determining a lower bound for the probability of having a non-singleton attractor in a BN with $x_2 = f_2(x_1, x_3)$ and $x_1 = f_1(x_2, x_4)$. To address this we focus our attention on a specific subclass of BNs on n genes. First, we investigate the probability of having bidirectional relationships in an arbitrary BN from that subclass. Then, we calculate a lower bound for the above mentioned probability.

2.4 Probability of forming bidirectional relationships

Consider the class \mathcal{B} of BNs on n genes such that, for $i = 1, 2, \dots, n$, $|W_i| = 2$, $x_i \in W_i$, and both variables in W_i are essential ones for the corresponding predictor function. Since both genes are essential, the predictor function may not be (1) constant, (2) equal to one of the input genes or (3) equal to the complement of one of the input genes. Owing to this assumption, six functions are excluded from the 16 possible functions (in the connectivity-2 case), so that there are ten possible predictor functions for each gene in the BN. Thus $|\mathcal{B}| = (10C_2^{n-1})^n$, where $C_2^{n-1} := (n-1)(n-2)/2$. Define $A_{i,j} = \{BN \in \mathcal{B} : x_i \in W_j\}$ and $B_{i,j} = A_{i,j} \cap A_{j,i}$. The number of different $B_{i,j}$ is C_2^n . Then, if one considers the uniform discrete probability measure over \mathcal{B} , the probability of selecting a BN $\in \mathcal{B}$ that has at least two genes having a bidirectional relationship equals

$$P\left(\bigcup_{i,j} B_{i,j}\right) = \sum_{i,j} P(B_{i,j}) - \sum_{i \leq k, j \leq l} P(B_{i,j} \cap B_{k,l}) \\ + \sum_{i \leq k \leq p, j \leq l \leq q} P(B_{i,j} \cap B_{k,l} \cap B_{p,q}) - \dots \pm P(\bigcap_{i,j} B_{i,j})$$

Many of the terms in the above formula are equal to 0 because the corresponding intersections of the sets $B_{i,j}$ are empty. This happens because for all of the intersections of more than n sets $B_{i,j}$, at least one of the indices has to appear at least three times, that is, the corresponding gene has more than two predictor genes, which contradicts the condition $|W_i| = 2$. $|B_{i,j}|$ is the same for all of the pairs i,j , and can be computed by noticing that one of the inputs is fixed for

each possible predictor function of the genes x_i and x_j . Thus, $|B_{i,j}| = 10(n-2)10(n-2)((10C_2^{n-1})^{n-2})$ which gives, for all i, j , $P(B_{i,j}) = (n-2)^2(C_2^{n-1})^{-2}$.

The computation of the probabilities in the subsequent terms in the above formula is more involved. To compute $P(B_{i,j} \cap B_{k,l})$, one has to consider two different cases depending on how many indices are the same or different. If there are four mutually different indexes, then $|B_{i,j} \cap B_{k,l}| = (10(n-2))^4((10C_2^{n-1})^{n-4})$, and $P(B_{i,j} \cap B_{k,l}) = (n-2)^4(C_2^{n-1})^{-4}$. If there are exactly two identical indices, then the gene corresponding to that index has a fixed pair of predictor genes, and thus $P(B_{i,j} \cap B_{k,l}) = (n-2)^2(C_2^{n-1})^{-3}$. Note that the number of different $B_{i,j} \cap B_{k,l}$ intersections is $C_2^{C_2^n}$.

In the supplementary material, we compute $P(\cup_{i,j} B_{i,j})$ for the cases $n=4$ and $n=5$ genes. For $n=4$, $P(\cup_{i,j} B_{i,j}) = 1$. For $n=5$, $P(\cup_{i,j} B_{i,j}) = 0.9969$.

When the number n of genes makes the exact computation of the probability $P(\cup_{i,j} B_{i,j})$ difficult, one can use the estimate

$$P(\cup_{i,j} B_{i,j}) \geq \sum_{i,j} P(B_{i,j}) - \sum_{i \leq k, j \leq l} P(B_{i,j} \cap B_{k,l})$$

2.5 Lower bound of the probability of forming non-singleton attractor cycles

Again consider the class \mathcal{B} of BNs on n genes such that, for $i = 1, 2, \dots, n$, $|W_i| = 2$, $x_i \in W_i$, and both variables in W_i are essential ones for the corresponding predictor function. The next proposition derives a lower bound estimate for the probability of a BN with such relationships having at least one non-singleton attractor cycle.

Proposition 4: Consider the class $\mathcal{A} \subset \mathcal{B}$ for which (1) $x_1 \in W_2$ and $x_2 \in W_1$, and (2) N_1 genes are predicted by gene x_1 and N_2 genes are predicted by gene x_2 . Then the probability of forming a $\text{BN} \in \mathcal{A}$ with non-singleton attractor cycles conditioned on the existence of at least one singleton attractor cycle in that BN is greater than $(0.18) \times (0.4)^{N_1+N_2}$.

The above lower bound is very conservative because it considers only one pair of genes that have a bidirectional relationship of connectivity 2. The previous proposition helps explain the phenomenon observed in our simulations: for randomly generated BNs with at least one bidirectional relationship of connectivity 2, singleton attractors are accompanied most of the time by attractor cycles of length 2 or more, and such that at least one of the states in those cycles is 1 Hamming distance away from one of the singleton attractors.

Fig. 1 shows simulations on random BNs. In this simulation, 10^6 random BNs from the class \mathcal{A} on n genes

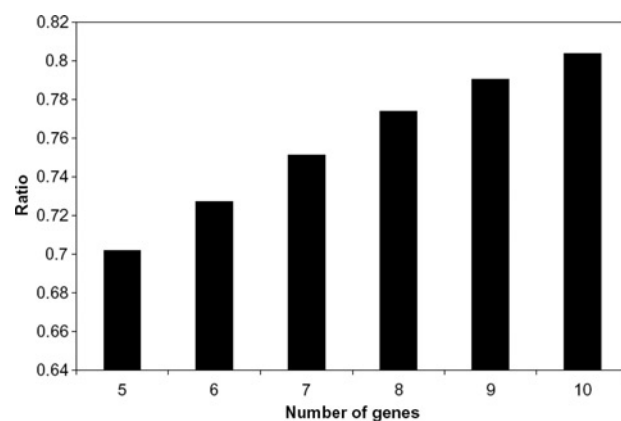


Figure 1 The ratio of randomly generated BNs $\in \mathcal{A}$ with at least one singleton attractor and one non-singleton attractor to all tried BNs $\in \mathcal{A}$ with at least one singleton attractor

are generated for each value of $n = 5, \dots, 10$. All the generated BNs in \mathcal{A} have at least one singleton attractor state (compare to Proposition 3). The number of random BNs, which have at least one non-singleton attractor cycle are counted. The simulations confirm the observation that the lower bound estimate is very conservative.

3 Algorithm

3.1 CoD-based inference of BNs

The CoD is a general non-linear statistical method to select a set of predictors for a given gene. It measures the degree to which the transcriptional levels of an observed (predictor) gene set can be used to improve the prediction of the transcriptional level of a target gene relative to the best prediction in the absence of observations. If x_i , W_i , and f_i are the target gene, the predictor set, and the predictor function for the target gene, respectively, then the CoD for the target gene x_i is given by

$$\theta^i = \frac{\varepsilon_0 - \varepsilon(x_i, f_i(W_i))}{\varepsilon_0}$$

where ε_0 is the error of the best estimate of x_i in the absence of any conditional variables and $\varepsilon(x_i, f_i(W_i))$ is the prediction error of the target gene according to the observations of the predictor set W_i [11]. For minimum mean-square error estimation, ε_0 is the error of the prediction of x_i with its mean.

The previous propositions explain why very often the CoD-inferred BNs possess spurious non-singleton attractors. We propose an algorithm to overcome the aforementioned shortcoming. We make the typical assumption that the data come from the steady state, and we apply the constraint that each data point is a singleton attractor.

Since the predictor function of each target gene is estimated from the steady-state data, not time-series data, each gene is a perfect estimator of itself (CoD equal to 1). To eliminate this trivial case, no gene can be a member of its own predictor set. Therefore given a target gene, for each target gene, there are $\sum_{m=1}^{k_i} C_m^{n-1}$ possible combinations for W_i , where k_i is the maximum cardinality of W_i , where n represents the number of genes in the network.

We employ a method called full-logic to estimate the predictor functions and consequently the CoDs for all possible combinations of predictor sets of each target gene. The CoD estimates a predictor function from the highest occurrence frequency of the target gene based on the values of all of the possible sets of predictor genes in the data set. More details regarding the full-logic method can be found in [21]. Note that there might be more than one high-CoD predictor set for a target gene.

3.2 Singleton attractor CoD (SA-CoD) inference algorithm

On the basis of our analysis of bidirectional relationships, in particular, their effect on the attractor structure of a BN, we have formulated an algorithm that limits the number of such bidirectional relationships when predictor sets are chosen using the CoD method.

The algorithm's input is the binary gene expression data. The outcome of the algorithm is a BN with no non-singleton attractors. The following parameters are set in advance: (1) a threshold, T_{CoD} , for the CoD ($T_{\text{CoD}} = 0.7$ in our study); (2) M_{BR} , the maximum number of permitted bidirectional relationships (keep in mind that, as we have shown, there is a substantial probability of there being at least two genes with bidirectional relationships in an arbitrary BN, $M_{\text{BR}} = 3$ in our study); and (3) m_{A} , the minimum number of points in the sample that appear as singleton attractors in the inferred BN ($m_{\text{A}} = 3$ in our study). Any predictor function that exceeds T_{CoD} is called a high-CoD predictor function. We now describe the singleton attractor CoD (SA-CoD) algorithm.

SA-CoD inference algorithm

1. Estimate the CoD and f_i for all the combinations of predictor sets W_i , for $i = 1, 2, \dots, n$.
2. Save all W_i and f_i with CoD exceeding T_{CoD} . For each target gene, save the high-CoD predictor sets and their associated predictor functions into two columns. The length, C_i , $i = 1, \dots, n$, of both columns, depends on T_{CoD} .
3. For each gene, randomly select W_i from the high-CoD predictor sets obtained in Step 2. Form a BNs from W_i and the corresponding f_i . If the total number of bidirectional

relationships in the BN does not exceed M_{BR} , go to Step 4. Otherwise, repeat Step 3.

4. If there is a non-singleton attractor in the BN, then go to Step 3; otherwise, continue.
5. If the number of data points appearing as singleton attractors in the BN is less than m_{A} , then go to Step 3; otherwise, STOP.

The steps of the algorithm accomplish certain goals. Step 3 constrains bidirectional relationships, thereby limiting spurious attractor cycles resulting from bidirectional relationships. The algorithm never allows bidirectional relationships for connectivity 1 since this case guarantees the formation of non-singleton attractor cycles. It should be noted that for each gene, there could be more than one possibility for the selection of W_i and f_i among high-CoD predictor sets. Step 3 selects W_i for each target gene randomly among high-CoD predictor sets such that the total number of bidirectional relationships in the generated BN does not exceed M_{BR} . Step 4 checks to see if any non-singleton attractor cycles have 'slipped through'. If the generated BN possesses at least one non-singleton attractor, the algorithm goes back to Step 3 and randomly selects another set of W_i and f_i among high-CoD predictor sets for all genes. The goal of this iteration is to generate a BN with only singleton attractors with at least m_{A} data points as singleton attractors. Step 5 insures that some minimal number of data points appear as singleton attractors in the inferred BN. The algorithm does not guarantee that the inferred BN will not contain singleton attractors that are not data points, but it does guarantee that there will be no non-singleton attractors. It is spurious non-singleton attractors that are ubiquitous in unconstrained CoD design. The algorithm does not guarantee that all data points will be singleton attractors, although it guarantees a minimum number, m_{A} , of these.

We select the parameters in the SA-CoD algorithm in the following manner. The selection of T_{CoD} depends on the gene expression data. We consider the CoDs of the best predictors of all genes in a network. When the connectivity is K , that is, maximum connectivity of genes, given a target gene, we compute the CoDs for all the combinations of predictors with length $1, \dots, K$. We choose T_{CoD} to be the lowest CoD in this set. As a result, for each target gene, the algorithm is allowed to select predictor sets in which their corresponding CoD is greater than T_{CoD} . The two parameters M_{BR} and m_{A} constrain the search space. The values of M_{BR} and m_{A} should be selected such that the constrained problem is feasible. For very large values of m_{A} and very small values of M_{BR} , the constrained search space may be empty. On the other hand, very small values of m_{A} may introduce BNs with large numbers of singleton attractors, which are not data points in the gene expression data. Similarly, very large values of M_{BR} may lead to BNs that possess at least one non-singleton attractors. The

selection of these two parameters also depends on the data. Given the gene expression data, by trial and error, one can choose M_{BR} and m_A such that the constrained search space is non-empty.

The algorithm can be run a number of times to produce a number of BNs, with each data point appearing in one or more BNs as a singleton attractor. This is somewhat similar to the design of PBNs under the requirement of contextual data consistency [14], where every data point must appear as a singleton attractor in at least one constituent BN of the PBN. There are, however, two key differences. First the method of [14] does not involve the CoD, but instead involves a constrained optimisation relative to the data distribution in the sample, and second, the number of BNs is determined by the data and it is theoretically certain that each data point will appear in at least one of the BNs as a singleton attractor. Nonetheless, the analogy is useful because the PBN design method first proposed in [12] applied CoD inference without constraint and then took combinations of high-CoD predictor functions to construct the BNs forming the PBN, with the threshold ultimately determining the number of constituent BNs.

Regarding algorithm complexity, the total number of BNs that can be generated from high-CoD predictor functions is $N = \prod_{i=1}^n C_i$, where n is the total number of genes and C_i is the number of high-CoD predictor functions. Thus, the search space has N members. In the worst-case scenario, Step 3 will be repeated N times.

4 Results and discussion

4.1 Comparison of SA-CoD algorithm with unconstrained CoD design

We have applied the preceding BN design procedure using gene-expression profiles from a study of 31 malignant melanoma samples in which messenger RNA was isolated directly from melanoma biopsies, and fluorescent cDNA from the message was prepared and hybridised to a microarray containing probes for 8150 cDNAs (representing 6971 unique genes) [22]. The seven genes used for the model are pirin, WNT5A, S100P, RET1, MART1, HADHB and STC2 (this being their order in the state space) were chosen from a set of 587 genes from the data set that have been subjected to an analysis of their ability to cross predict each other's state in a multivariate setting [23]. The gene-expression profiles were binarised to arrive at 31 binary vectors with seven columns corresponding to the selected seven genes.

The binary data have been taken from a previous paper that considered the external control of gene regulatory networks [24]. In that paper, the raw gene-expression data are quantised to ternary values according to the hypothesis testing procedure in [25]. The mapping from ternary to

binary data was done gene by gene, depending on the ternary values: RET1, MART1, and HADHB had only two ternary values, so the larger one was set to 1 and the smaller one to 0. Pirin and S100P each had large numbers of +1 and 0 ternary values with no more than a few -1 ternary values, so +1 was mapped to 1 and both 0 and -1 were mapped to 0. STC2 had large numbers of 0 and -1 ternary values and only a few +1 ternary values, so 0 and +1 were mapped to 1 and -1 was mapped to 0. Finally, WNT5A had twenty-two -1 values and nine values almost even split between 0 and +1. For this gene, 0 and +1 were mapped to 1 and -1 was mapped to 0.

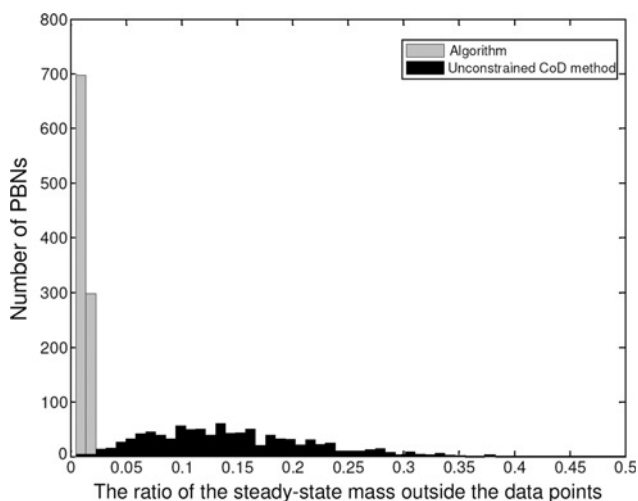
Table 1 gives the seven-gene profiles for the 18 distinct data points and their corresponding frequencies. The assumption is that the data points correspond to the steady state of the underlying gene regulatory system.

The SA-CoD algorithm is applied 500 times to the gene expression data to generate 500 BNs. On the basis of the specifications of the algorithm, the BNs possess no non-singleton attractors and there are at least three data points as singleton attractors in each of them. Figures of some of these BNs are shown in the supplementary material. We randomly choose ten BNs from the pool of 500 BNs. Setting the perturbation and switching probabilities equal to 0.01, we generate a PBN from these ten BNs. The BNs are selected randomly with equal probability. The PBN is run sufficiently long so that its steady-state distribution can be estimated. The steady-state probability corresponding to each attractor represents the probability of being in that attractor in the long-run. We consider the distribution of the gene expression data as our benchmark. The proportion of the steady-state probability mass of the generated PBN lying outside the data states is computed. This procedure is repeated 1000 times to generate 1000 PBNs, in each case the proportion of the steady-state probability mass outside the data states being computed. These 1000 proportions are used to form the light histogram in Fig. 2. The mass of this histogram is concentrated very close to 0.

To compare the performance of the SA-CoD algorithm with the unconstrained CoD method, we repeat the same experiment with the predictor sets and predictor functions with high-CoD chosen without the constraint of the SA-CoD algorithm. Figs. of some of these BNs are also shown in the supplementary material. Proceeding without constraint, 500 BNs are generated and 1000 PBNs composed of ten BNs randomly chosen from the 500 BNs are generated and run into their steady states. The dark histogram in Fig. 2 is formed from the proportions of mass of the 1000 steady-state distributions lying outside the data states. These are well dispersed between 0 and 0.35. By eliminating spurious attractors, the SA-CoD algorithm puts a much higher concentration of the steady-state probability mass on the data points.

Table 1 Expression profiles for melanoma

Profile#	Gene							Count
	pirin	WNT5A	S100P	RET1	MART1	HADHB	STC2	
1	1	0	0	1	1	1	1	2
2	1	1	0	1	0	0	0	1
3	1	0	1	0	1	1	1	1
4	1	0	0	1	1	1	0	2
5	0	1	0	1	0	0	1	1
6	1	0	1	1	1	1	1	1
7	0	1	0	1	1	1	1	1
8	0	1	0	0	0	0	1	4
9	0	0	0	1	0	0	1	1
10	0	1	1	0	0	0	1	1
11	1	0	1	0	1	0	0	1
12	1	0	1	1	1	1	0	2
13	1	0	1	0	1	1	0	8
14	0	0	1	0	0	0	0	1
15	0	0	1	0	1	1	0	1
16	0	1	0	1	0	0	0	1
17	0	0	1	1	1	0	0	1
18	0	0	1	0	1	0	0	1

**Figure 2** Proportion of the steady-state probability mass outside the data states in 1000 PBNs

Light histogram: when the SA-CoD algorithm is used

Dark histogram: when the unconstrained CoD method is used

A key issue for PBN design is to compose a PBN with enough BNs so that each data state appears as an attractor in the PBN (that is, appears as an attractor in one of the constituent BNs) but not to include so many BNs that there is a large number of spurious attractors. To compare the SA-CoD algorithm with unconstrained design in this regard, in the next experiment we compare the number of data points appearing as attractors with the number of attractors that are not data points in a collection of n BNs generated by either the SA-CoD algorithm or unconstrained CoD design. Let D be the number of distinct points in the data, N be the number of data points appearing as attractors in the generated BNs, and M be the number of non-data-point attractors appearing in the generated BNs. A reasonable measure of performance for the desired comparison is

$$R = a(D - N) + (1 - a)M$$

where $0 \leq a \leq 1$, a being chosen depending on what we want to emphasise. Smaller R means better performance.

Since N and M are functions of the number n of BNs, R is a function of n . We compute $R(a, n)$ for $n = 1, 2, \dots, 80$ and $0 \leq a \leq 1$ by taking $R(a, n)$ to be the average of 1000

trials of computation of R , each trial involving randomly choosing n BNs from a pool of 500 designed BNs. Fig. 3a shows the surface graph of $R(a, n)$ when using the SA-CoD algorithm. The dots on the surface indicate the minimum value of $R(a, n)$ for a given value of a , the value of n for the minimum being the optimal number of BNs relative to the measure R . For small a , the emphasis is on avoiding spurious attractors and hence the optimum n is smaller. For large a , the emphasis is on recovering data points as attractors and hence the optimum n is larger.

Fig. 3b shows the surface graph of $R(a, n)$ using unconstrained CoD design. The difference between Fig. 3a and Fig. 3b demonstrate the benefits of the SA-CoD algorithm. First, notice the different scales of the graphs. The values of R for unconstrained CoD design tend to greatly exceed those for the SA-CoD algorithm. Second, in Fig. 3b, the optimal number of BNs is 1 for all but very large values of a , which validates the point that, if we are

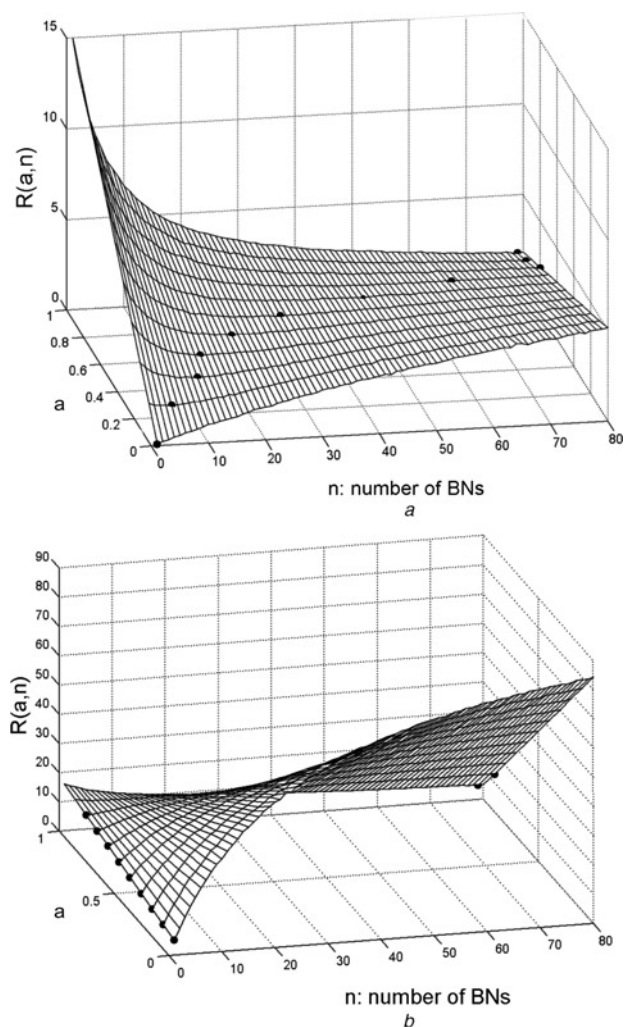


Figure 3 Surface graph of $R(a, n)$

a Value of $R(a, n)$ for a from 0 to 1 and n from 1 to 80 when the BNs are generated by the SA-CoD algorithm

b Value of $R(a, n)$ for a from 0 to 1 and n from 1 to 80 when the BNs are generated by the unconstrained CoD method

concerned about spurious attractors, then unconstrained CoD design performs poorly.

4.2 Concluding remarks

The creation of bidirectional relationships is an important shortcoming when applying CoD inference procedures to non-temporal microarray data. The usual assumption regarding such data is that it comes from the steady state of the underlying genomic regulatory system. For BN modelling, this translates into the assumption that the data profiles correspond to attractors in the network. Since there is no dynamic information in the sample points, it is prudent to produce a model in which they are singleton attractors. Given the tendency of bidirectional gene relationships produced by CoD inference to result in non-singleton attractors, we have proposed a novel algorithm to infer BNs from non-temporal data. The algorithm avoids non-singleton attractors and, as demonstrated by simulation studies, yields few attractors that are not data points while at the same time capturing data points as attractors in the designed networks.

5 References

- [1] DE JONG H.: 'Modeling and simulation of genetic regulatory systems: a literature review', *J. Comput. Biol.*, 2002, **9**, (1), pp. 67–103
- [2] IVANOV I., DOUGHERTY E.R.: 'Modeling genetic regulatory networks: continuous or discrete?', *J. Biol. Syst.*, 2006, **14**, (2), pp. 219–229
- [3] VAN SOMEREN E.P., WESSELS L.F.A., BACKER E., REINDERS M.J.T.: 'Multi-criterion optimization for genetic network modeling', *Signal Process.*, 2003, **83**, (4), pp. 763–775
- [4] LIANG S., FUHRMAN S., SOMOGYI R.: 'REVEAL, a general reverse engineering algorithm for inference of genetic network architectures'. *Pac. Symp. Biocomput.*, 1998, vol. 3, pp. 18–29
- [5] AKUTSU T.: 'Identification of genetic networks from a small number of gene expression patterns under the Boolean network model'. *Pac. Symp. Biocomput.*, 1999, pp. 17–28
- [6] SHMULEVICH I., SAARINEN A., YLI-HARJA O., ASTOLA J.: 'Inference of genetic regulatory networks under the best-fit extension paradigm'. *Computational and Statistical Approaches To Genomics*, 2002
- [7] LAHDESMÄKI H., SHMULEVICH I., YLI-HARJA O.: 'On learning gene regulatory networks under the boolean network model', *Mach. Learn.*, 2003, **52**, (1, 2), pp. 147–167

- [8] YU J., SMITH V.A., WANG P.P., HARTEMINK A.J., JARVIS E.D.: 'Advances to bayesian network inference for generating causal networks from observational biological data', *Bioinformatics*, 2004, **20**, (18), pp. 3594–3603
- [9] KAUFFMAN S.A.: 'Metabolic stability and epigenesis in randomly constructed genetic nets', *J. Theor. Biol.*, 1969, **22**, (3), pp. 437–467
- [10] KAUFFMAN S.: 'The origins of order: self-organization and selection in evolution' (Oxford University Press, New York, 1993)
- [11] DOUGHERTY E.R., KIM S., CHEN Y.D.: 'Coefficient of determination in nonlinear signal processing', *Signal Process.*, 2000, **80**, (10), pp. 2219–2235
- [12] SHMULEVICH I., DOUGHERTY E.R., KIM S., ZHANG W.: 'Probabilistic boolean networks: a rule-based uncertainty model for gene regulatory networks', *Bioinformatics*, 2002, **18**, (2), pp. 261–274
- [13] PAL R., IVANOV I., DATTA A., BITTNER M.L., DOUGHERTY E.R.: 'Generating boolean networks with a prescribed attractor structure', *Bioinformatics*, 2005, **21**, (21), pp. 4021–4025
- [14] DOUGHERTY E.R., XIAO Y.: 'Design of probabilistic boolean networks under the requirement of contextual data consistency', *IEEE Trans. Signal Process.*, 2006, **54**, (9), pp. 3603–3613
- [15] ZHAO W., SERPEDIN E., DOUGHERTY E.R.: 'Inferring connectivity of genetic networks using information theoretic criteria', *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 2008, **5**, (2), pp. 262–274
- [16] DOUGHERTY E.R.: 'Validation of inference procedures for gene regulatory networks', *Curr. Genom.*, 2008, **8**, (6), pp. 351–359
- [17] KAUFMAN V., MIHALJEV T., DROSSEL B.: 'Scaling in critical random Boolean networks', *Phys. Rev. E*, 2005, **72**, (4), part 2, article no. 046124, 1–9
- [18] KAUFMAN V., DROSSEL B.: 'On the properties of cycles of simple Boolean networks', *Eur. Phys. J. B*, 2005, **43**, (1), pp. 115–124
- [19] B. Drossel: 'Random boolean networks', <http://arxiv.org/abs/0706.3351>, 2008
- [20] KAUFMAN V., DROSSEL B.: 'Relevant components in critical random Boolean networks', *New J. Phys.*, 2006, **8**, article no. 228, 1–16
- [21] KIM S.C., DOUGHERTY E.R., CHEN Y.D., ET AL.: 'Multivariate measurement of gene expression relationships', *Genomics*, 2000, **67**, (2), pp. 201–209
- [22] BITTNER M., MEITZER P., CHEN Y., ET AL.: 'Molecular classification of cutaneous malignant melanoma by gene expression profiling', *Nature*, 2000, **406**, (6795), pp. 536–540
- [23] KIM S., LI H., DOUGHERTY E.R., ET AL.: 'Can markov chain models mimic biological regulation?', *J. Biol. Syst.*, 2002, **10**, (4), pp. 337–357
- [24] DATTA A., CHOUDHARY A., BITTNER M.L., DOUGHERTY E.: 'External control in markovian genetic regulatory networks: the imperfect information case', *Bioinformatics*, 2004, **20**, pp. 924–930
- [25] CHEN Y., KAMAT V., DOUGHERTY E.R., BITTNER M.L., MELTZER P., TRENT J.: 'Ratio statistics of gene expression levels and applications to microarray data analysis', *Bioinformatics*, 2002, **18**, (9), pp. 207–215
- [26] PAUL U., KAUFMAN V., DROSSEL B.: 'Properties of attractors of canalizing random Boolean networks', *Phys. Rev. E*, 2006, **73**, (2), part 2, article no. 026118, 1–9

6 Appendix

Proposition 1: If there are two genes in a BN having a bidirectional relationship of connectivity 1, then the BN has at least one non-singleton attractor cycle.

Proof: Without loss of generality assume the two genes are x_1 and x_2 . There are four possible transition pairs of predictor functions for these genes: (1) $f_1 \equiv x_2$ and $f_2 \equiv x_1$; (2) $f_1 \equiv x_2$ and $f_2 \equiv \bar{x}_1$; (3) $f_1 \equiv \bar{x}_2$ and $f_2 \equiv x_1$; and (4) $f_1 \equiv \bar{x}_2$ and $f_2 \equiv \bar{x}_1$, where the overbar denotes negation.

Consider the first possible pair: $f_1 \equiv x_2$ and $f_2 \equiv x_1$. If the transitions start from the point $01\mathbf{y}$, then after finitely many transitions, the BN will enter an attractor $01\mathbf{x}_0$ or $10\mathbf{y}_0$, where \mathbf{y} , \mathbf{y}_0 and \mathbf{x}_0 denote vectors of the remaining gene values. Assume that the first visited attractor state is $01\mathbf{x}_0$ (the other possibility $10\mathbf{y}_0$ can be considered in the same way). Because x_1 and x_2 depend only on each other and $01\mathbf{x}_0$ is an attractor state, from this point on the network must follow a transition sequence of the form $01\mathbf{x}_0, 10\mathbf{x}_1, 01\mathbf{x}_2, \dots, 01\mathbf{x}_k$, where $\mathbf{x}_k = \mathbf{x}_0$ and $\mathbf{x}_r \neq \mathbf{x}_0$ for $1 \leq r < k$. Thus, the sequence forms an attractor cycle of length $k = 2m > 1$, where m is a natural number. It is straightforward to show that similar cycles are formed when $f_1 \equiv \bar{x}_2$ and $f_2 \equiv \bar{x}_1$.

Next, we consider the predictor functions pair: $f_1 \equiv x_2$ and $f_2 \equiv \bar{x}_1$. If the transitions start from any point of the form $x_1x_2\mathbf{y}$, then after finitely many transitions the BN will enter an attractor state that is of one of the following forms: $00\mathbf{x}_0, 01\mathbf{y}_0, 10\mathbf{z}_0$ or $11\mathbf{u}_0$. Here we consider the case when the first visited attractor state is $00\mathbf{x}_0$ (the other possibilities can be considered similarly). Because x_1 and x_2 depend only on each other and $00\mathbf{x}_0$ is an attractor state,

from this point on the network must follow a transition sequence of the form, $00x_0, 01x_1, 11x_2, 10x_3, \dots, 00x_k$, where $x_k = x_0$ and $x_r \neq x_0$ for $1 \leq r < k$. Thus, the sequence forms an attractor cycle of length $k = 4m > 1$. It is straightforward to show that similar cycles are formed when $f_1 \equiv \bar{x}_2$ and $f_2 \equiv x_1$. In short, one of the following four possible cases happens and guarantees the formation of non-singleton attractors

$$\begin{matrix} 01x_0 & 10x_1 & \dots & 01x_k \\ 00x_0 & 11x_1 & \dots & 00x_k \\ 00x_0 & 01x_1 & 11x_1 & 10x_1 & \dots & 00x_k \\ 10x_0 & 11x_1 & 01x_1 & 00x_1 & \dots & 10x_k \end{matrix} \quad \square$$

Proposition 2: If a BN possesses a pair of genes that have a bidirectional relationship of connectivity 2, then at least 1/8 of the states in its state space cannot be singleton attractors of the network.

Proof: Suppose x_1 and x_2 have a bidirectional relationship of order 2 with $W_1 = \{x_2, x_4\}$ and $W_2 = \{x_1, x_3\}$. Because all predictor variables are essential, the following conditions cannot occur (refer to the truth tables for f_1 and f_2):

1. $(a_1 = c_1 \text{ and } b_1 = d_1) \text{ or } (a_2 = c_2 \text{ and } b_2 = d_2)$
2. $(a_1 = b_1 \text{ and } c_1 = d_1) \text{ or } (a_2 = b_2 \text{ and } c_2 = d_2)$

These conditions exclude cases where at least one gene is non-essential. In the first condition, $(a_1 = c_1 \text{ and } b_1 = d_1)$ excludes three cases (1) f_1 being constant (2) f_1 being equal to x_4 (3) f_1 being equal to \bar{x}_4 . Under (1) both genes are non-essential. Under (2) and (3), x_2 is non-essential. Furthermore, $(a_2 = c_2 \text{ and } b_2 = d_2)$ excludes three cases (1) f_2 being constant (2) f_2 being equal to x_3 (3) f_2 being equal to \bar{x}_3 . Under (1) both genes are non-essential. Under (2) and (3), x_1 is non-essential. Likewise, condition 2 excludes cases where x_3, x_4 , or both are non-essential.

Moreover, any combination of f_1 and f_2 belongs to at least one of the following (not mutually exclusive) classes:

$$F_1: a_1 = \bar{c}_1 \text{ and } a_2 = \bar{c}_2$$

$$F_2: a_1 = \bar{c}_1 \text{ and } b_2 = \bar{d}_2$$

$$F_3: b_1 = \bar{d}_1 \text{ and } a_2 = \bar{c}_2$$

$$F_4: b_1 = \bar{d}_1 \text{ and } b_2 = \bar{d}_2$$

Without loss of generality, suppose x_1 and x_2 have a bidirectional relationship of order 2 with $W_1 = \{x_2, x_4\}$ and $W_2 = \{x_1, x_3\}$. To prove the proposition we consider the following four cases: (a) $(f_1, f_2) \in F_1$ and $x_3x_4 = 00$; (b) $(f_1, f_2) \in F_2$ and $x_3x_4 = 01$; (c) $(f_1, f_2) \in F_3$ and $x_3x_4 = 10$; and (d) $(f_1, f_2) \in F_4$ and the states of the BN were $x_3x_4 = 11$.

For case (a), first consider $(f_1, f_2) \in F_1$ such that $a_1 = \bar{c}_1 = 0$ and $a_2 = \bar{c}_2 = 0$. Examination of the truth tables of f_1 and f_2 , Table 2, where $a_1 = \bar{c}_1 = 0$ and $a_2 = \bar{c}_2 = 0$, together with the assumed constant values of x_3 and x_4 , shows that any state with $x_3x_4 = 00$ and $x_1 = \bar{x}_2$ cannot be a singleton attractor. A simple counting argument shows that the states where $x_3x_4 = 00$ and $x_1 = \bar{x}_2$ account for exactly one-eighth of all of the states in state space. Reasoning in the same way, one can check that when $(f_1, f_2) \in F_1$ with $a_1 = \bar{c}_1 = 1$ and $a_2 = \bar{c}_2 = 1$, the states with $x_3x_4 = 00$ and $x_1 = x_2$ cannot be singleton attractors, and that there are exactly one-eighth such states in the state space. To complete the analysis of case (a), consider the situation where $(f_1, f_2) \in F_1$ with $a_1 = \bar{c}_1, a_2 = \bar{c}_2$, and $a_1 = \bar{a}_2$. In this case, examination of the the truth tables of f_1 and f_2 shows that all of the states where $x_3x_4 = 00$ cannot be singleton attractors. It is straightforward to count that there are exactly one-fourth of the states in the state space that are of this type.

Using similar arguments and symmetry considerations, one can show that the proposition holds for cases (b), (c) and (d). \square

Proposition 3: In the class of BNs possessing a pair of genes with bidirectional relationship of connectivity 2, there exists a BN with at least one singleton attractor cycle.

Proof: Without a loss of generality one can assume that $x_1 = f_1(x_2, x_3)$ and $x_2 = f_2(x_1, x_4)$. Table 2 shows the truth-tables for f_1 and f_2 . Next, consider a restricted class of BNs with n nodes for which the predictor functions for the genes $x_i, i = 3, 4, \dots, n$, are constants, that is $f_i \equiv x_i$. Suppose $x_3x_4 = 00$ (the other possible cases can be considered in a similar fashion). From the truth table for f_1 we have $x_1 = a_1$ or $x_1 = c_1$. It is straightforward to check that when $a_1 = c_1$, the state $\bar{x} = a_1a_200\bar{y}$ or the state $\bar{x} = a_1c_200\bar{y}, \bar{y} = x_5x_6 \dots x_n$, is a singleton attractor state for the BN in consideration, depending on the value of a_1 . Next, if one has a pair of predictor functions (f_1, f_2) with $a_1 \neq c_1$, we show how to find a singleton attractor state \bar{x} in the BN for the case when $a_1 = 0$ (the other case, $a_1 = 1$, can be considered in the same way). If $a_1 = 0$ and $a_2 = 0$, then the state $0000\bar{y}$ is a singleton attractor for the BN. If $a_1 = 0$ and $a_2 = 1$, then the state $1100\bar{y}$ is a singleton attractor for the BN, provided that $c_2 = 1$. If $c_2 = 0$, we consider truth tables in Table 3 for the predictor

Table 2 Truth tables for f_1 and f_2

x_2	x_4	f_1	x_1	x_3	f_2
0	0	a_1	0	0	a_2
0	1	b_1	0	1	b_2
1	0	c_1	1	0	c_2
1	1	d_1	1	1	d_2

Table 3 Truth tables for f_1 and f_2

x_2	x_4	f_1	x_1	x_3	f_2
0	0	0	0	0	1
0	1	b_1	0	1	b_2
1	0	1	1	0	0
1	1	d_1	1	1	d_2

functions f_1 and f_2 . In such a case, states \bar{x} where $x_3 = x_4 = 0$ cannot be singleton attractors for the network, and we consider states where $x_3 = x_4 = 1$. If $b_1 = d_1$, then the state $\bar{x} = b_1 b_2 11 \bar{y}$ or the state $\bar{x} = b_1 d_2 11 \bar{y}$ is a singleton attractor state for the BN in consideration, depending on the value of b_1 . If $b_1 \neq d_1$, then states where $x_3 = x_4 = 1$ cannot be singleton attractors for the network; however, owing to the conditions stated at the outset of this section, $b_1 \neq 0$. Thus, $b_1 = 1$, which implies $d_1 = 0$. Therefore in this case either ($a_1 = d_1 = 0, b_1 = c_1 = 1, a_2 = 1$ and $b_2 = c_2 = d_2 = 0$) or ($a_1 = d_1 = 0, b_1 = c_1 = 1, a_1 = b_2 = d_2 = 1$ and $c_2 = 2$), which implies that either the state $\bar{x} = 0001 \bar{y}$ or the state $\bar{x} = 0110 \bar{y}$ is a singleton attractor state for the network. \square

Proposition 4: Consider the class $\mathcal{A} \subset \mathcal{B}$ for which (1) $x_1 \in W_2$ and $x_2 \in W_1$, and (2) N_1 genes are predicted by gene x_1 and N_2 genes are predicted by gene x_2 . Then the probability of forming a $BN \in \mathcal{A}$ with non-singleton attractor cycles conditioned on the existence of at least one singleton attractor cycle in that BN is greater than $(0.18 \times 0.4)^{N_1+N_2}$.

Proof: Without loss of generality we assume that $W_1 = (x_2, x_4)$ and $W_2 = (x_1, x_3)$. Let us define a new variable called C_t . This variable shows the current state of the BN at time t . Now consider a $BN \in \mathcal{A}$ with a singleton attractor of the form $S = x_1^s x_2^s x_3^s x_4^s \bar{x}^s$, where \bar{x}^s denotes a vector of the remaining gene values. Since S is a singleton attractor state, if $C_t = S$, then $C_{t+1} = S$ for all $t \geq 0$. This implies in $C_{t+1}: x_1 = f_1(W_1^s) = x_1^s, x_2 = f_2(W_2^s) = x_2^s, x_3 = f_3(W_3^s) = x_3^s, x_4 = f_4(W_4^s) = x_4^s$, and $x_i = f_i(W_i^s) = x_i^s$ for any gene i in the vector \bar{x}^s .

Now consider the following state, which is only one Hamming distance away from the singleton attractor S : $A = x_1^s x_2^s x_3^s x_4^s \bar{x}^s$. We focus on the following transitions $A \rightarrow B \rightarrow C$. When the current state of the BN (C_t) is equal to A , the next transition (C_{t+1}) is called B and the transition after that (C_{t+2}) is called C . State B is of the form $B = x_1^b x_2^b x_3^b x_4^b \bar{x}_0^b$, where $x_1^b = f_1(W_1^a), x_2^b = f_2(W_2^a), x_3^b = f_3(W_3^a), x_4^b = f_4(W_4^a)$, and $x_i^b = f_i(W_i^a)$ for any gene i in the vector \bar{x}_0^b . For any gene m in the window of

$x_3^b x_4^b \bar{x}_0^b$ that is not predicted by gene 1 (i.e. $x_1 \notin W_m$), one can conclude that $x_m^b = x_m^s$ since $W_m^a = W_m^s$ (recalling that state A is one Hamming distance away from

state S and S is a singleton attractor state). However, for any gene k in the window of $x_3^b x_4^b \bar{x}_0^b$, if $W_k^a = (x_1^a, x_j^a)$, then x_k^b may or may not be equal to x_k^s . A function is canalising if one input can fix the output of a node, irrespective of the value of the other inputs [26]. If function f_k is a canalising function with the canalising variable x_j (canalising value x_j^c), then $W_k^a = W_k^s$. When this is true for any gene k where $x_1 \in W_k$, the window

$$x_3^b x_4^b \bar{x}_0^b \text{ is equal to the window } x_3^s x_4^s \bar{x}^s$$

As noted in the previous proof, there are ten possible predictor functions for each gene k . Assume gene k is one of the genes in the window $x_3^b x_4^b \bar{x}_0^b$ where it is predicted by gene 1 and gene j . Table 4 shows all ten possible functions.

When gene j is equal to 0, there are four canalising functions, $f_k^1, f_k^3, f_k^8, f_k^{10}$, with x_j being the canalising variable. Hence, for these functions, $f_k(x_1 = 0, x_j = 0) = f_k(x_1 = 1, x_j = 0)$. Similarly, when gene j is equal to 1, there are four canalising functions, $f_k^2, f_k^5, f_k^6, f_k^9$, with x_j being the canalising variable. Since for any $BN \in \mathcal{A}$, x_j^c is equally likely to be 0 or 1, the probability of having a canalising function for f_k with canalising variable x_j^c is equal to $4/10$ (0.4).

If the number of genes that are predicted by gene 1 is equal to N_1 , then the probability of the window $x_3^b x_4^b \bar{x}_0^b$ to be equal to the window $x_3^s x_4^s \bar{x}^s$ is equal to $(0.4)^{N_1}$.

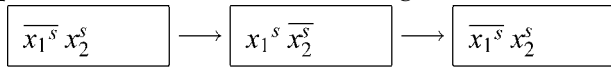
Now let us look at the state of gene 1 and gene 2 in the transition A to B . Conditional on x_3^b being equal to x_3^s and x_4^b being equal to x_4^s (we found its probability in the previous paragraph), we can find the number of functions of f_1 and f_2 that generate a non-singleton attractor state.

Table 2 shows predictors f_1 and f_2 . Assume $x_3^s x_4^s = 00$. If $a_1 = a_2 = 0$ and $c_1 = c_2 = 1$, then it is straightforward to check that a non-singleton attractor cycle is formed. When a_1, a_2, c_1 , and c_2 are fixed, there are three possibilities for (b_1, d_1) in f_1 and three possibilities for (b_2, d_2) in f_2 . Similarly, if $a_1 = a_2 = 1$ and $c_1 = c_2 = 0$, then 9 other

Table 4 All ten possible f_k for a $BN \in \mathcal{A}$ when gene k is predicted by gene 1 and gene j

x_1	x_j	f_k^1	f_k^2	f_k^3	f_k^4	f_k^5	f_k^6	f_k^7	f_k^8	f_k^9	f_k^{10}
0	0	0	0	0	0	0	1	1	1	1	1
0	1	0	0	1	1	1	0	0	0	1	1
1	0	0	1	0	1	1	0	0	1	0	1
1	1	1	0	0	0	1	0	1	1	1	0

possibilities of f_1 and f_2 form a non-singleton attractor cycle. Since it is equally likely to have $x_3^s x_4^s = 00, 01, 10, 11$, one can conclude that among 100 possible functions of f_1 and f_2 , 18 combinations allow the following transitions:



State A State B State C

Since gene 2 in B is different from gene 2 in C , the probability

of having the window $\boxed{x_3^c x_4^c x_0^c}$ being equal to the window

$\boxed{x_3^b x_4^b x_0^b}$ is equal to $(0.4)^{N_2}$ where N_2 is the number of genes predicted by gene 2.

Since the probability of $\boxed{x_3^b x_4^b x_0^b}$ being equal to the window $\boxed{x_3^s x_4^s \overline{x_0^s}}$ is equal to $(0.4)^{N_1}$, the probability of

$\boxed{x_3^c x_4^c x_0^c}$ being equal to the window $\boxed{x_3^s x_4^s \overline{x_0^s}}$ is equal to $(0.4)^{N_1+N_2}$.

Consequently, state C is equal to state A and a non-singleton attractor cycle state with length 2 is formed with probability of $(0.18) \times (0.4)^{N_1+N_2}$.

Hence under the proposition assumptions the probability of forming a non-singleton attractor cycle in a $\text{BN} \in \mathcal{A}$ conditional on the existence of a singleton attractor is greater than $(0.18) \times (0.4)^{N_1+N_2}$. \square