Systems biology

Intervention in context-sensitive probabilistic Boolean networks

Ranadip Pal[1], Aniruddha Datta[1], Michael L. Bittner[2] and
Edward R. Dougherty[1,3,*]

[1]Department of Electrical Engineering, Texas A&M University, College Station, TX, 77843, USA,
[2]Translational Genomics Research Institute, 400 North Fifth Street, Suite 1600, Phoenix, AZ 85004,
USA and [3]University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA

ABSTRACT

Motivation: Intervention in a gene regulatory network is used to help it avoid undesirable states, such as those associated with a disease. Several types of intervention have been studied in the framework of a probabilistic Boolean network (PBN), which is essentially a finite collection of Boolean networks in which at any discrete time point the gene state vector transitions according to the rules of one of the constituent networks. For an instantaneously random PBN, the governing Boolean network is randomly chosen at each time point. For a context-sensitive PBN, the governing Boolean network remains fixed for an interval of time until a binary random variable determines a switch. The theory of automatic control has been previously applied to find optimal strategies for manipulating external (control) variables that affect the transition probabilities of an instantaneously random PBN to desirably affect its dynamic evolution over a finite time horizon. This paper extends the methods of external control to context-sensitive PBNs.

Results: This paper treats intervention via external control variables in context-sensitive PBNs by extending the results for instantaneously random PBNs in several directions. First, and most importantly, whereas an instantaneously random PBN yields a Markov chain whose state space is composed of gene vectors, each state of the Markov chain corresponding to a context-sensitive PBN is composed of a pair, the current gene vector occupied by the network and the current constituent Boolean network. Second, the analysis is applied to PBNs with perturbation, meaning that random gene perturbation is permitted at each instant with some probability. Third, the (mathematical) influence of genes within the network is used to choose the particular gene with which to intervene. Lastly, PBNs are designed from data using a recently proposed inference procedure that takes steady-state considerations into account. The results are applied to a context-sensitive PBN derived from gene-expression data collected in a study of metastatic melanoma, the intent being to devise a control strategy that reduces the WNT5A gene's action in affecting biological regulation, since the available data suggest that disruption of this influence could reduce the chance of a melanoma metastasizing.

Contact: edward@ee.tamu.edu

*To whom correspondence should be addressed.

1 INTRODUCTION

A probabilistic Boolean network (PBN) is essentially a discrete collection of Boolean networks in which at any discrete time point, the gene state vector transitions according to the rules of one of the Boolean networks (Shmulevich *et al.*, 2002a,b). As originally introduced, the governing Boolean network is randomly chosen at each time point, which means that the rule for updating each gene is randomly chosen at each time step from among several possible rules in accordance with a fixed probability distribution (Shmulevich *et al.*, 2002a). Such PBNs are referred to as *instantaneously random* PBNs. The intent is to generalize the Boolean model (Kauffman, 1969; Glass and Kauffman, 1973; Kauffman, 1993; Somogyi and Sniegoski, 1996; Huang, 1999) so as to incorporate uncertainty in the functions governing network transitions, whether uncertainty arises from inherent biological considerations or inference from data. In this way, PBNs form a subclass of Markovian genetic regulatory networks. A subsequent modification to the PBN structure allows each gene to randomly change value at each instant with a small perturbation probability p. This insures that all the possible states communicate, thereby resulting in an ergodic Markov chain possessing a steady-state distribution (Shmulevich *et al.*, 2002c).

Switching the constituent Boolean network of a PBN at every instant corresponds to switching the wiring diagram of the network from one time step to the next. While this can help incorporate uncertainty into the original Boolean rule-based paradigm, a model that is more appropriate to the stability of biological systems is achieved by limiting the switching between constituent networks. In such a PBN, the Boolean functions (and their variable sets, known as their *predictor sets*) remain fixed for an interval of time until the occurrence of some random event, perhaps corresponding to an external stimulus. This random switching is governed by a binary random variable with a typically small probability q of forcing a switch in constituent Boolean network. Taking the perspective that a switch corresponds to a change in context for the cell, these more general PBNs are referred to as *context-sensitive* PBNs.

To date, intervention in the context of PBNs has been approached in three ways: (i) resetting the state of an instantaneously random PBN with perturbation to a more desirable initial state and letting the network evolve from there (Shmulevich *et al.*, 2002c); (ii) changing the steady-state (long-run) probability distribution of an instantaneously random PBN with perturbation by minimally

altering its rule structure (Shmulevich *et al.*, 2002d); and (iii) manipulating external (control) variables that affect the transition probabilities of an instantaneously random PBN (without perturbation) to desirably affect its dynamic evolution over a finite time horizon (Datta *et al.*, 2003, 2004). All of these results have been obtained by exploiting the fact that the dynamic behavior of the PBN can be modeled by a Markov chain, thereby making the PBN amenable to the theory of Markov chains and Markov decision processes. A key difference between instantaneously random and context-sensitive PBNs that affects intervention analysis is that, whereas an instantaneously random PBN forms a Markov chain whose state space is composed of gene vectors, the Markov chain corresponding to a context-sensitive PBN is composed of pairs $(x, \mathbf{f})$, where x is the current gene vector occupied by the network and $\mathbf{f}$ is the vector of functions defining the current constituent Boolean network.

In this paper we consider intervention via external control variables in context-sensitive PBNs. This involves significant extension of the results in Datta *et al.* (2003) in several directions. First, constituent network switching only occurs with probability q at each time instant. Second, there is random gene perturbation at each instant with probability p. Third, we use the concept of gene *influence*, introduced in Shmulevich *et al.* (2002a), to choose the particular gene with which to intervene and demonstrate that intervening with a higher influence gene results in better performance. Finally, when applying the control theory, the designed PBN is derived from steady-state considerations, which makes the intervention strategy more biologically appealing.

Regarding the final point, we note that most microarray-based gene-expression studies do not involve controlled time series experimental data; rather, it is assumed that data result from sampling from the steady state. Under this assumption, a key criterion for checking the validity of a designed network is that much of its steady-state mass lies in the states observed in the sample. In the Boolean-network framework, this signifies a close resemblance between the observed data points and the attractors of the designed Boolean network. Here we use the Bayesian connectivity-based approach (Zhou *et al.*, 2004) to construct Boolean networks with the expectation of generating networks having a few very strong attractors (thereby reflecting biological stability) that are highly similar to the data. This approach yields a number of highly probable Boolean networks and Bayesian scores for each of them. These networks are combined with probabilities proportional to their scores to form a PBN. In this paper, we derive expressions for the state transition probabilities of a PBN formed from a number of such Boolean networks and devise a strategy to shift the state vector from undesirable states towards more desirable ones using genomic control.

2 DEFINITIONS

A *Boolean network* (BN) $B = (V, F)$ on n genes is defined by a set of nodes/genes $V = \{x_1, \ldots, x_n\}$, $x_i \in \{0, 1\}$, $i = 1, \ldots, n$, and a list $F = (f_1, \ldots, f_n)$ of Boolean functions, $f_i : \{0, 1\}^n \to \{0, 1\}$, $i = 1, \ldots, n$. Each node x_i represents the state/expression of the gene x_i, where $x_i = 0$ means that gene i is OFF and $x_i = 1$ means that gene i is ON. The function f_i is called the *predictor function* for gene i. Updating the states of all genes in B is done synchronously at every time step according to their predictor functions. A *probabilistic Boolean network* (*PBN*) consists of a set of

nodes/genes $V = \{x_1, \ldots, x_n\}$, $x_i \in \{0, 1\}$, $i = 1, \ldots, n$, and a set of vector-valued network functions, $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k$, governing the state transitions of the genes, each network function being of the form $\mathbf{f}_j = (f_{j1}, f_{j2}, \ldots, f_{jn})$, where $f_{ji} : \{0, 1\}^n \to \{0, 1\}$, $i = 1, \ldots, n$. The choice of which network function $\mathbf{f}_j$ to apply is governed by a selection procedure. Specifically, at each time point a random decision is made as to whether to switch the network function for the next transition, with the probability q of a switch being a system parameter. If a decision is made to switch the network function, then a new function is chosen from among $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k$, with the probability of choosing $\mathbf{f}_j$ being the selection probability c_j. In other words, each network function $\mathbf{f}_j$ determines a BN and the PBN behaves as a fixed BN until a random decision (with probability q) is made to change the network function according to the probabilities $c_1, c_2, \ldots, c_k$ from among $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k$. The PBN just described is called a *context-sensitive PBN*. In the special case when $q = 1$, the network function is switched at every time point and the PBN is called an *instantaneously random PBN*. We consider context-sensitive PBNs with perturbation, meaning that at each time point there is a probability p of any gene changing its value uniformly randomly. Since there are n genes, the probability of there being a random perturbation at any time point is $1 - (1 - p)^n$. The state space S of the network together with the set of network functions, in conjunction with transitions between the states and network functions, determines a Markov chain. The random perturbation makes the Markov chain ergodic, meaning that it has the possibility of reaching any state from any other state and that it possesses a steady-state distribution.

The state vector $x(t)$ at any time step t is essentially an n-digit binary number $[x_1 x_2 \ldots x_n]$ whose decimal equivalent is given by

$$z(t) = \sum_{j=1}^{n} 2^{n-j} x_j(t). \tag{1}$$

As $x(t)$ ranges from $000 \ldots 0$ to $111 \ldots 1$, $z(t)$ takes on all values from 0 to $2^n - 1$. For a context-sensitive PBN, the state $z(t)$ at time t could be originating from any one of the k possible networks. To keep track of the network emitting a particular state, let us redefine the states by incorporating the network number inside the state label. Since we have k different BNs forming the PBN, the total number of states becomes $2^n k$ and we label these states as $S_0, S_1, \ldots, S_{2^n k-1}$, where for each $r = 1, 2, \ldots, k$, states $S_{2^n(r-1)}, S_{2^n(r-1)+1}, \ldots, S_{2^n r-1}$ belong to network r. Equivalently $S_{2^n(r-1)+i}$ corresponds to z_{r_i}, where z_{r_i} is the decimal representation of the ith state in the network r. Let the redefined state at time t be denoted by $w(t)$.

3 TRANSITION PROBABILITIES

In this section, we derive expressions for the transition probabilities in a context-sensitive PBN subject to perturbations by recognizing that the following mutually exclusive events can occur at any time point t:

(1) The current network function is applied, the PBN transitions accordingly and the network function remains the same for the next transition.

(2) The current network function is applied, the PBN transitions accordingly and a new network function is selected for the next transition.

(3) There is a random perturbation and the network function remains the same for the next transition.

(4) There is a random perturbation and a new network function is selected for the next transition.

Assuming that the individual genes perturb independently, and letting $\mathrm{mod}(v, w)$ denote the remainder left over when v is divided by w, we consider two cases for determining the transition probability of going from state a to state b:

Case 1. $[a/2^n] = [b/2^n]$, meaning $2^n(r - 1) \le a, b \le 2^n r - 1$ for the same r. This corresponds to the events (1) and (3) above and the transition probabilities are given by

$$
\begin{aligned}
Pr(w(t + 1) = b|w(t) = a) \\
= (1 - q)(1 - p)^n f_{r,a,b} \\
+ (1 - q)(1 - p)^{n-h} p^h s(h)
\end{aligned}
\tag{2}
$$

where h is the Hamming Distance between $\mathrm{mod}(a, 2^n)$ and $\mathrm{mod}(b, 2^n)$, i.e. the number of genes which differ between the two states,

$$
f_{r,a,b} = \begin{cases} 1, & \text{if } a \text{ transitions to } b \text{ in a single step in} \\ & \text{network } r, \\ 0, & \text{otherwise,} \end{cases}
$$

and

$$
s(h) = \begin{cases} 0, & \text{if } h = 0, \\ 1, & \text{otherwise.} \end{cases}
$$

The first term in Equation (2) corresponds to event (1) above, where $1 - q$ is the probability that the network selection does not change, $(1 - p)^n$ is the probability that none of the n genes undergo a perturbation, we assume that network selection and random gene perturbation are independent events, and $f_{r,a,b} = 1$ if that particular transition is possible in the rth Boolean network. The second term corresponds to event (3), where h genes have to be perturbed to go from state a to state b.

Case 2. $2^n(r_1 - 1) \le a \le 2^n r_1 - 1$ and $2^n(r_2 - 1) \le b \le 2^n r_2 - 1$, where $r_1 \ne r_2$. This corresponds to events (2) and (4) above and the transition probabilities are given by

$$
\begin{aligned}
Pr(w(t + 1) = b|w(t) = a) \\
= q \frac{c_{r_2}}{\sum_{i=1, i \ne r_1}^k c_i} (1 - p)^n f_{r_1,a,b} \\
+ q \frac{c_{r_2}}{\sum_{i=1, i \ne r_1}^k c_i} (1 - p)^{n-h} p^h s(h).
\end{aligned}
\tag{3}
$$

If we define

$$
g(a, b) = \begin{cases} 1, & \text{if } [a/2^n] - [b/2^n] = 0, \\ 0, & \text{otherwise,} \end{cases}
$$

then a unified transition probability expression encompassing the two cases is given by

$$
\begin{aligned}
& Pr(w(t + 1) = b|w(t) = a) \\
&= [(1 - q)(1 - p)^n f_{r,a,b} \\
& \quad + (1 - q)(1 - p)^{n-h} p^h s(h)] g(a, b) \\
& \quad \times \left[q \frac{c_{r_2}}{\sum_{i=1, i \ne r_1}^k c_i} (1 - p)^n f_{r_1,a,b} \right. \\
& \quad \left. + q \frac{c_{r_2}}{\sum_{i=1, i \ne r_1}^k c_i} (1 - p)^{n-h} p^h s(h) \right] \\
& \quad \times [1 - g(a, b)].
\end{aligned}
\tag{4}
$$

By letting a and b range over all integers from 0 to $2^n k - 1$ and using Equation (4), we can determine all the entries of the $2^n k \times 2^n k$ matrix of transition probabilities.

In practice, it will likely be impossible to detect the Boolean network from which the current gene activity profile is being emitted. In most cases, we will only have knowledge of the states of the genes. To handle such situations, we can derive an expression for the transition probability from state s_2 to state s_1, where these states run from 0 to $2^n - 1$ and reflect only the expression status of the n-gene state vector:

$$
\begin{aligned}
& Pr[z(t + 1) = s_1|z(t) = s_2] \\
&= \sum_{i=1}^k Pr[z(t + 1) = s_1, s_2 \text{ belongs to network } i|z(t) = s_2] \\
&= \sum_{i=1}^k Pr[z(t + 1) = s_1|z(t) = s_2, s_2 \text{ belongs to} \\
& \quad \text{network } i] \cdot Pr[s_2 \text{ belongs to network } i] \\
&= \sum_{i=1}^k Pr[z(t + 1) = s_1|w(t) = s_2 + 2^n(i - 1)] . c_i \\
&= \sum_{i=1}^k \sum_{j=1}^k c_i \cdot Pr[w(t + 1) = s_1 + 2^n(j - 1)|w(t) \\
& \qquad\qquad\qquad = s_2 + 2^n(i - 1)]
\end{aligned}
\tag{5}
$$

where s_1 and s_2 run from 0 to $2^n - 1$. Note that here state s_1 is equivalent to the distinct states $s_1, s_1 + 2^n, \ldots, s_1 + (k - 1)2^n$ in the previous $2^n k$ formulation. Similarly s_2 here is equivalent to $s_2, s_2 + 2^n, \ldots, s_2 + (k - 1)2^n$ in the earlier formulation. By letting s_1 and s_2 range from 0 to $2^n - 1$ and using Equation (5), we can derive the $2^n \times 2^n$ transition probability matrix A corresponding to the context-sensitive PBN.

4 CONTROL IN CONTEXT-SENSITIVE PBNS

In this section, we consider the problem of external control in a context-sensitive PBN. Towards this end, suppose that a PBN with n genes has m control inputs, $u_1, u_2, \ldots, u_m$, each of which can take on only the binary values 0 or 1. Then at any time t, the row vector $u(t) \stackrel{\Delta}{=} [u_1(t), u_2(t), \ldots, u_m(t)]$ describes the complete

status of all the control inputs. $u(t)$ can take on all binary values from $[0, 0, \ldots, 0]$ to $[1, 1, \ldots, 1]$. One can equivalently represent the control input status using the decimal number

$$v(t) = \sum_{i=1}^{m} 2^{m-i} u_i(t). \tag{6}$$

As $u(t)$ takes on binary values from $[0, 0, \ldots, 0]$ to $[1, 1, \ldots, 1]$, the variable $v(t)$ ranges from 0 to $2^m - 1$. We can equivalently use $v(t)$ as an indicator of the complete control input status of the PBN at time t.

If a control action is applied, then the transition probability expressions will change. Suppose that our control action consists of forcibly altering the value of a single gene, g, from 0 to 1 or from 1 to 0. Thus, $m = 1$ here. Then the new transition probabilities with control, denoted by $Prc1$, are given by

$$Prc1(w(t + 1) = b | w(t) = a)$$
$$= Pr(w(t + 1) = b | w(t) = a + 2^{n-g}) func(a)$$
$$+ Pr(w(t + 1) = b | w(t) = a - 2^{n-g})(1 - func(a)), \tag{7}$$

where

$$func(a) = \begin{cases} 1, & \text{if state of gene } g \text{ is 0 for } a, \\ 0, & \text{if state of gene } g \text{ is 1 for } a, \end{cases}$$

and the transition probabilities, Pr, without control are given by Equation (4).

Here, a and b range over 0 through $2^n k - 1$. As before we can reduce the dimension of the state space by replacing the ws in Equation (7) by zs and using Equation (5) to determine the transition probabilities without the control action:

$$Prc1(z(t + 1) = b | z(t) = a)$$
$$= Pr(z(t + 1) = b | z(t) = a + 2^{n-g}) func(a)$$
$$+ Pr(z(t + 1) = b | z(t) = a - 2^{n-g})(1 - func(a)). \tag{8}$$

By letting a and b vary over 0 to $2^n - 1$ and making use of Equation (8), we can determine the $2^n \times 2^n$ matrix $A(v(t))$ of control-dependent transition probabilities.

In the rest of this section, we formulate and solve the control problem assuming 2^n states and the availability of full state information. The same development can be carried out for the $2^n k$ state formulation if we simultaneously have the gene state information and the network labels. As shown in Datta *et al.* (2003), the one-step evolution of the probability distribution vector in the case of a PBN containing 2^n states with control inputs takes place according to the equation:

$$pd(t + 1) = pd(t) A(v(t)) \tag{9}$$

where $pd(t)$ is the 2^n dimensional state probability distribution vector and $A(v(t))$ is the $2^n \times 2^n$ matrix of control-dependent transition probabilities determined by Equation (8). Since the transition probability matrix is a function of the control input $v(t)$, the evolution of the probability distribution vector of the PBN with control now depends not only on the initial distribution vector but also on the values of the control input at different time steps. Furthermore, intuitively it appears possible to make the states of the network evolve

in a desirable fashion by appropriately choosing the control input at each time step.

These ideas have been formalized in Datta *et al.* (2003) to arrive at the following finite horizon optimization problem. Given an initial state z_0,

$$\min_{\mu_0, \mu_1, \ldots, \mu_{M-1}} E\left[\sum_{t=0}^{M-1} C_t(z_t, \mu_t(z_t)) + C_M(z_M) \right] \tag{10}$$

subject to $Pr(z(t + 1) = j | z(t) = i, v(t))$, given by Equation (8), where

- M represents the treatment/intervention window;
- $\mu_t : [0, 1, 2, \ldots, 2^n - 1] \rightarrow [0, 1, 2, \ldots, 2^m - 1]$, $t = 0, 1, 2, \ldots, M - 1$ are functions mapping the state space into the control space;
- $C_t(z_t, v_t)$ is the one-step cost of applying the control v_t at state z_t;
- $C_M(z_M)$ is the terminal cost associated with the state z_M.

As discussed in Datta *et al.* (2003), the consideration of such an optimization problem can be naturally motivated in the context of cancer treatment applications where one must choose between a number of alternative treatments to be applied over a finite horizon of time. Once input from biologists/clinicians has been used to select an appropriate cost function and an appropriate treatment window, the control problem is essentially reduced to that of controlling a Markov chain over a finite horizon.

The dynamic programming solution to Equation (10) is given by (Bertsekas, 1976; Datta *et al.*, 2003):

$$J_M(z_M) = C_M(z_M), \tag{11}$$

$$J_t(z_t)$$
$$= \min_{v_t \in \{0, 1, \ldots, 2^m - 1\}} \left[C_t(z_t, v_t) + \sum_{j=0}^{2^n - 1} Pr(z_t | j, v_t) \cdot J_{t+1}(j) \right],$$
$$t = 0, 1, \ldots, M - 1. \tag{12}$$

If $v_t^* = \mu_t^*(z_t)$ minimizes the right-hand side of Equation (12) for each z_t and t, then the control law $\pi^* = \{\mu_0^*, \mu_1^*, \ldots, \mu_{N-1}^*\}$ is optimal.

The optimal control problem [Equation (10)] and its solution [Equations (11) and (12)] are from a very general setting; however, in our case, the class of allowable controls is severely constrained since our control action consists of forcibly altering the expression status of only a *single* gene. This limited control objective is dictated primarily by limitations on the kind of interventions that appear to be within the realm of biological possibility.

5 SELECTING THE CONTROL GENE

Given a particular target gene, there may be several genes that are good predictors for it. Among a set of predictors for a particular gene, some of them may have more impact on the value of the target gene than others. For instance, in cancer studies it has been shown that p53 has a more profound effect on the cell cycle regulator gene WAF1/p21 than other predictors of WAF1, such as AP2 or BRCA1 (Gartel and

Tyner, 1999). In view of this, one can define the *influence* of the variable x_j on the Boolean function f (Shmulevich *et al.*, 2002a). To do so, let D be the probability mass distribution over the states of a Boolean network and let $\partial f(x)/\partial x_j$ be the partial derivative of the Boolean function f with respect to the argument x_j. Then the influence of x_j on f is defined by

$$I_j(f) = E_D\left[\frac{\partial f(x)}{\partial x_j}\right] = Pr\left\{\frac{\partial f(x)}{\partial x_j} = 1\right\}$$
$$= Pr\{f(x) \neq f(x^{(j)})\}, \tag{13}$$

where $x^{(j)}$ is the same as x except that the jth component is toggled. In this paper, we will assume that the distribution D is uniform.

The main idea behind the influence definition is to quantify the amount by which the gene x_j affects the value of the function f. If the value of the function f changes on toggling the value of gene x_j for most gene activity profiles x, then the influence of the jth gene on f is high. For the case of PBNs, let F_i be the set of predictors for gene x_i with corresponding probabilities $c_1^{(i)}, \ldots, c_{l(i)}^{(i)}$. Let $I_k(f_j^{(i)})$ be the influence of variable x_k on the predictor $f_j^{(i)}$. Then the influence of gene x_k on gene x_i is given by (Shmulevich *et al.*, 2002a):

$$I_k(x_i) = \sum_{j=1}^{l(i)} c_j^{(i)} I_k(f_j^{(i)}). \tag{14}$$

We can use the *influence* to select the control gene. For example, suppose we have treatments $d_1, d_2, \ldots, d_r$ that can affect genes $g_1, g_2, \ldots, g_r$, respectively. Biological or economic considerations may constrain us to use only one treatment at a time. Then we can use the gene that has the highest influence on the target gene g_t. The influence can be directly calculated from the PBN as given by the previous formula or it can be approximated from the observed gene activity profiles. The hope is that by selecting a gene with high influence as the control gene, we will be able to carry out a more cost-effective intervention. The simulation results presented in the next section show that such an expectation is met.

6 MELANOMA APPLICATION

In this section, we apply the results of this paper to a context-sensitive PBN derived from gene expression data collected in a study of metastatic melanoma (Bittner *et al.*, 2000). In this study, the abundance of mRNA for the gene WNT5A was found to be highly discriminating between cells with properties typically associated with high versus low metastatic competence. These findings were validated and expanded in a second study in which experimentally increasing the levels of the Wnt5a protein secreted by a melanoma cell line via genetic engineering methods directly altered the metastatic competence of that cell as measured by the standard *in vitro* assays for metastasis (Weeraratna *et al.*, 2002). Furthermore, it was found that an intervention that blocked the Wnt5a protein from activating its receptor, the use of an antibody that binds the Wnt5a protein, could substantially reduce Wnt5a's ability to induce a metastatic phenotype. This suggests that a reasonable control strategy would be to use an intervention that reduces the WNT5A gene's action in affecting biological regulation, since the available data suggests that disruption of this influence could reduce the chance of a melanoma metastasizing, a desirable outcome. Instantaneously random
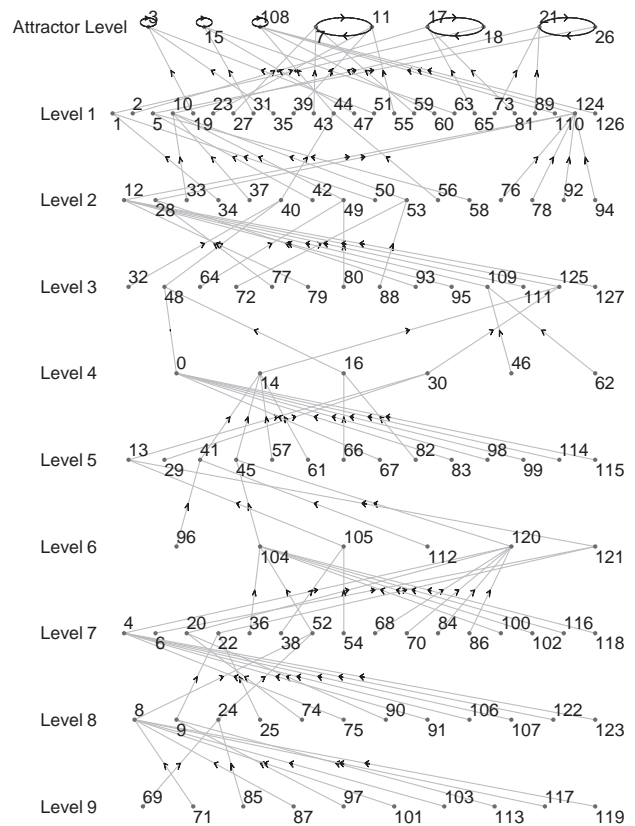

Fig. 1. Network 1.

PBNs derived from the same expression data have been used in Datta *et al.* (2003, 2004) for demonstrating the earlier intervention strategies.

Here, we consider a 7-gene network containing the genes WNT5A, pirin, S100P, RET1, MART1, HADHB and STC2. To obtain the PBN, we have used the Bayesian connectivity-based approach of Zhou *et al.* (2004) to construct four highly probable Boolean networks that are used as the constituent Boolean networks in the PBN, with their selection probabilities based on their Bayesian scores. The four generated Boolean networks are shown in Figures 1–4, where the states are labeled from 0 to $127 = 2^7 - 1$. Each constituent network is assumed to be derived from steady-state gene-expression data, and the attractor states and the level sets are shown in the figures. Observe that in each of these networks, the state enters an attractor cycle in a small number of steps (at most nine), which is consistent with what is expected in real networks (Zhou *et al.*, 2004).

The control strategy of the previous section has been applied to the designed PBN with pirin chosen as the control gene and $p = q = 0.01$. Figure 5 shows the expected cost for a finite horizon problem of length 5 originating from each of the 128 states. In these simulations, the problem formulation for 2^n states has been used. The cost of control is assumed to be 0.5 and the states are assigned a terminal penalty of 5 if WNT5A is 1 and 0 if WNT5A is 0. The control objective is to down-regulate the WNT5A gene. From Figure 5, it is clear that the expected cost with control is much lower than that without control, which agrees with our objective. If the length of the control horizon is increased, then Figure 6 shows that all the
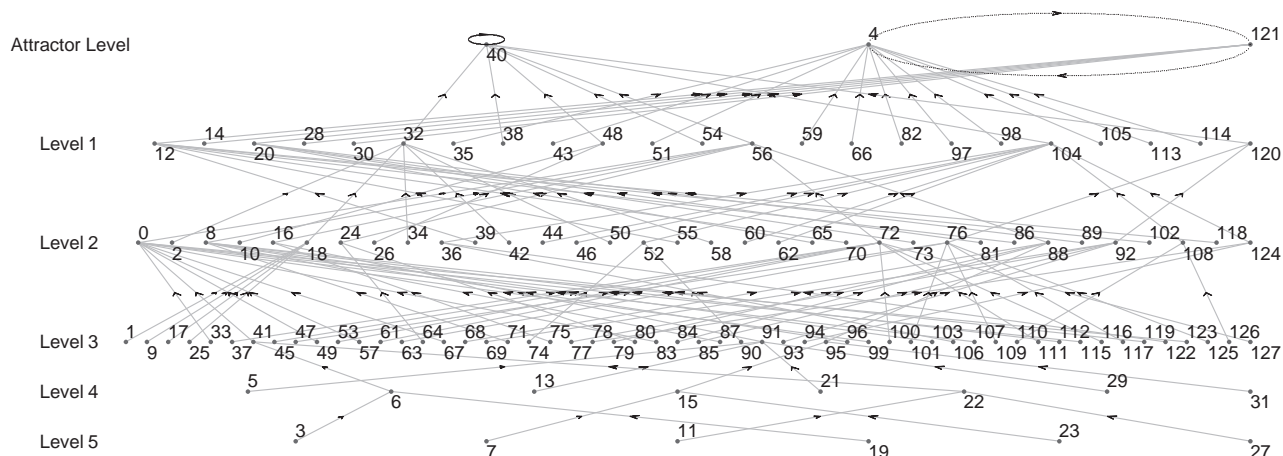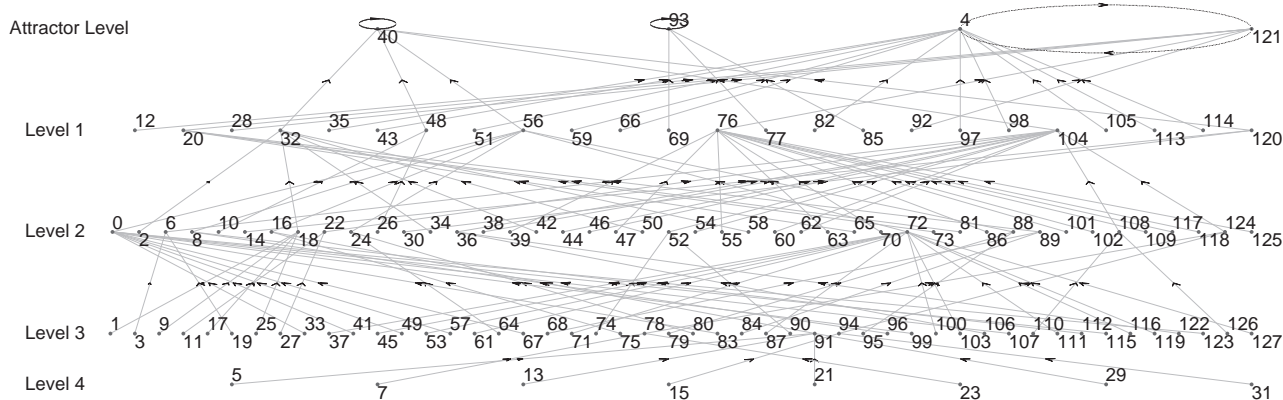
Fig. 2. Network 2.
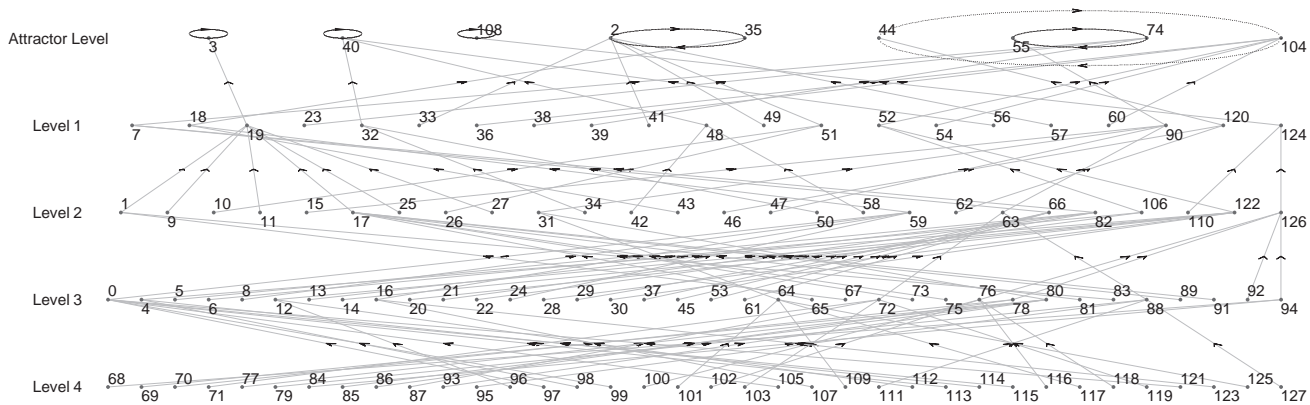

Fig. 3. Network 3 .

Fig. 4. Network 4.

initial states start yielding almost the same expected cost. This may be due to the fact that the maximum level of the constituent networks is 9 and the Markov chain is ergodic. If, on the other hand, the $2^n k$ formulation is used, then the expected costs for different initial states become almost equal after a larger number of time steps (data not shown). This is possibly due to the fact that no averaging is used in that formulation.

Next we consider the relationship between the influence of a control gene and its effectiveness in carrying out the intervention. The influences of the other six genes on WNT5a are as follows: pirin = 1,
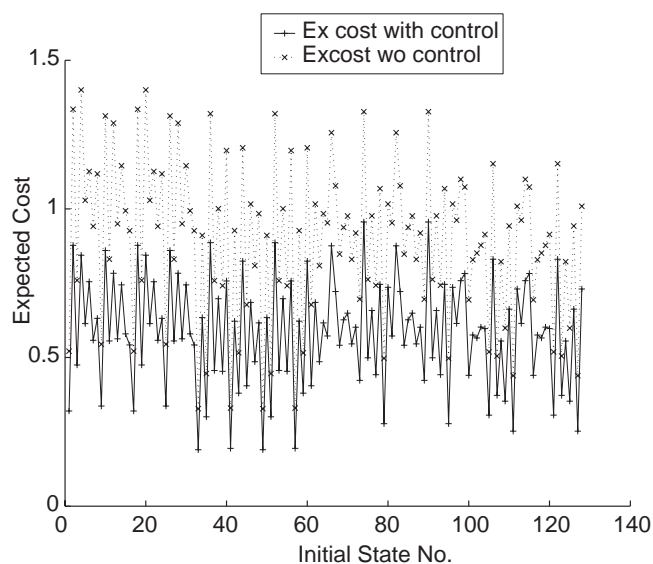
Fig. 5. Expected cost for a finite horizon problem of length 5 originating from the different initial states.
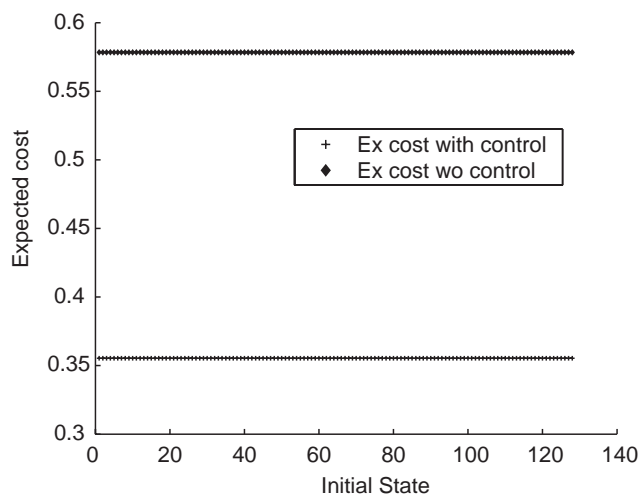
Fig. 6. Expected cost for a finite horizon problem of length 30 originating from the different initial states.

$S100P = 0.75$, $RET1 = 0$, $MART1 = 0$, $HADHB = 1$ and $STC2 = 1$. The influence has been calculated from the influences of the genes in the four constituent Boolean networks, assuming equal probabilities for each network. These influence values (GI) are tabulated alongside the control genes (CG) in Table 1. The perturbation probability p is not taken into account for the influence calculations because it has a very low value. If the starting gene activity profile is pirin = 0, S100P = 0, RET1 = 0, MART1 = 0, HADHB = 1, STC2 = 0 and WNT5A = 1, then the expected costs for finite horizon control problems of lengths (Ln) 5 and 30 are shown in Table 1. Here, $Ec1$ represents the expected cost when the 2^n state formulation is used, $Ec2$ represents the expected cost when the $2^n k$ state formulation is used, the suffix wc denotes with control, and the suffix woc denotes without control. The table shows that the expected

Table 1. Expected cost table

CG	GI	Ln	$Ec1wc$	$Ec1woc$	$Ec2wc$	$Ec2woc$
pirin	1	30	.355352	.5784	.566017	.949586
mart1	0	30	.568611	.5784	.743938	.949586
hadhb	1	30	.398291	.5784	.300602	.949586
stc2	1	30	.413105	.5784	.569817	.949586
pirin	1	5	.652455	.974544	.396288	.61994
mart1	0	5	.963684	.974544	.53374	.61994
hadhb	1	5	.762097	.974544	.304567	.61994
stc2	1	5	.830185	.974544	.398155	.61994

cost is much lower (0.35 and 0.39) when the high-influence genes pirin and HADHB are used, as compared to the expected cost (0.56) obtained when the low-influence gene MART1 is used to control the network.

7 CONCLUSION

This paper extends earlier results on intervention in instantaneously random PBNs without perturbation to context-sensitive PBNs with perturbation. The extension is significant because the latter class more closely models small biological subnetworks whose logical behavior is affected by conditions outside the genes represented in the model network. The results show that the expected cost with control is much lower than without control. In addition, the results indicate that we can achieve a much better control outcome if a gene with high influence is selected as the control gene.

ACKNOWLEDGEMENTS

REFERENCES

Bertsekas,D.P. (1976). *Dynamic Programming and Stochastic Control*, Academic Press.

Bittner,M., Meltzer,P., Chen,Y., Jiang,Y., Seftor,E., Hendrix,M., Radmacher,M., Simon,R., Yakhini,Z., Ben-Dor,A. *et al.* (2000). Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature*, **406**, 536–540.

Datta,A., Choudhary,A., Bittner,M.L. and Dougherty,E.R.(2003) External control in markovian genetic regulatory networks. *Machine learning*, **52**, 169–191.

Datta,A., Choudhary,A., Bittner,M.L. and Dougherty,E.R. (2004) External control in markovian genetic regulatory networks: the imperfect information case. *Bioinformatics*, **20**, 924–930.

Gartel,A.L. and Tyner,A.L. (1999) Transcriptional regulation of p21 (WAF1/CIp1) gene. *Exp. Cell Res.*, **246**, 280–289.

Glass,L. and Kauffman,S. (1973) The logical analysis of continuous non-linear biochemical control networks. *J. Theor. Biol.*, **39**, 103–129.

Huang,S. (1999) Gene expression profiling, genetic networks, and cellular states: An integrating concept for tumorigenesis and drug discovery. *J. Mol. Med.*, **77**, 469–480.

Kauffman,S.A. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *Theoret. Biol.*, **22**, 437–467.

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

Shmulevich,I., Dougherty,E.R., Kim,S. and Zhang,W. (2002a) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, **18**, 261–274.

Shmulevich,I., Dougherty,E.R. and Zhang,W. (2002b) From Boolean to probabilistic Boolean networks as models of genetic regulatory networks. *Proc. IEEE*, **90**, 1778–1792.

Shmulevich,I., Dougherty,E.R. and Zhang,W. (2002c) Gene perturbation and intervention in probabilistic boolean networks. *Bioinformatics*, **18**, 1319–1331.

Shmulevich,I., Dougherty,E.R. and Zhang,W. (2002d) Control of stationary behavior in probabilistic Boolean networks by means of structural intervention. *Biol. Syst.*, **10**, 431–446.

Somogyi,R. and Sniegoski,C.A. (1996) Modeling the complexity of genetic networks: understanding multigenic and pleiotropic regulation. *Complexity*, **1**, 45–63.

Weeraratna,A.T., Jiang,Y., Hostetter,G., Rosenblatt,K., Duray,P., Bittner,M. and Trent,J.M. (2002) Wnt5a signaling directly affects cell motility and invasion of metastatic melanoma. *Cancer Cell*, **1**, 279–288.

Zhou,X., Wang,X. Pal,R. Ivanov,I. Bittner,M. and Dougherty,E. (2004) A Bayesian connectivity-based approach to constructing probabilistic gene regulatory networks. *Bioinformatics*, **20**, 2918–2927.