East Asian Journal on Applied Mathematics
doi: 10.4208/eajam.191012.221112a

Vol. **2**, No. 4, pp. 353-372
November 2012

On Generating Optimal Sparse Probabilistic Boolean Networks with Maximum Entropy from a Positive Stationary Distribution

Hao Jiang, Xi Chen, Yushan Qiu* and Wai-Ki Ching

Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong.

Received 19 October 2012; Accepted (in revised version) 22 November 2012

Available online 29 November 2012

Abstract. To understand a genetic regulatory network, two popular mathematical models, Boolean Networks (BNs) and its extension Probabilistic Boolean Networks (PBNs) have been proposed. Here we address the problem of constructing a sparse Probabilistic Boolean Network (PBN) from a prescribed positive stationary distribution. A sparse matrix is more preferable, as it is easier to study and identify the major components and extract the crucial information hidden in a biological network. The captured network construction problem is both ill-posed and computationally challenging. We present a novel method to construct a sparse transition probability matrix from a given stationary distribution. A series of sparse transition probability matrices can be determined once the stationary distribution is given. By controlling the number of nonzero entries in each column of the transition probability matrix, a desirable sparse transition probability matrix in the sense of maximum entropy can be uniquely constructed as a linear combination of the selected sparse transition probability matrices (a set of sparse irreducible matrices). Numerical examples are given to demonstrate both the efficiency and effectiveness of the proposed method.

AMS subject classifications: 65C20, 92B05

Key words: Boolean Networks (BNs), Entropy, Probabilistic Boolean Networks (PBNs), genetic regulatory networks, sparsity, stationary probability distribution, transition probability matrix.

1. Introduction

In the post-genome era, rapidly evolving genomic technologies have paved the way for massive amounts of genomic data. This enhances the fast development in systems biology, a field of study focusing on the interactions among the components of biological systems. Through the study, one can better understand the functions and behavior of a

*Corresponding author. *Email addresses:* haohao@hkusuc.hku.hk (H. Jiang), dlkcissy@hku.hk (X. Chen), yushanqiu2526374@163.com (Y. Qiu), wching@hku.hk (W.-K. Ching)

 ©2012 Global-Science Press

biological system in a holistic manner. Building genetic and related biological networks have been enhanced by the advancement of computational and statistical techniques. A tremendous amount of mathematical and computing approaches have been used to glean the understanding of biological processes over the past few decades. Directed graphs can be viewed as the most straightforward way to model a Genetic Regulatory Network (GRN). Bower and Bolouri [14] introduced some classic models of genetic networks. Other models like multivariate Markov chain models [9] and regression models [38] can also be found in the literature.

A Bayesian network [13] depicts the genetic regulatory process from a probability perspective. The dynamic Bayesian network, an extension of Bayesian network can describe statistical temporal dependencies among genes. However, it does not explicitly describe temporal relations among genes in a functional form. In the perspective of dynamical systems, differential equations have been employed to describe the change rate of expression levels. Discrete Dynamical System (DDS) Model [23], a discrete version of ODEs, assists one to understand the interactions among variables systematically. It has gained a solid foot in quantitative modeling of GRNs. Other stochastic models for studying the dynamical properties of GRNs can be found in [26]. Mathematical models based on genetic programming and fuzzy logic have been studied in [22]. Reviews on other mathematical formalisms can be found in [12,33].

Two popular mathematical models, Boolean Networks (BNs) and its extension Probabilistic Boolean Networks (PBNs) have been proposed in the literature. From a logical standpoint, the expression of a gene in the network is quantized to be two states: "ON" and "OFF", see for instance [18,19]. This helps us in understanding the key dynamic properties of a regulatory process. BNs belong to a class of discrete dynamical systems in that genes interact with each other precisely determined by molecular interactions over a set of Boolean variables [20]. BN models have been applied in various aspects for its simplicity and deterministic property. The uncertainty in genetic regulation process and errors of microarray data caused by experimental noise require more realistic models other than deterministic model like BN model. PBN model [29–32] is an extension of BN model that incorporates the stochastic characteristics of GRNs. Each gene is regulated through a set of Boolean functions with corresponding selection probabilities. The model combines deterministic functional aspects and the inherent probabilistic characteristics of complex systems. A PBN can be regarded as a Markov chain process [6] and therefore it can be studied using the well established Markov chain theory [5]. Given a PBN, its stationary distribution characterizes the network behavior. Efficient numerical methods [21,36], approximation methods [6] and perturbation methods [35] for computing transition probability matrix and the resulting stationary distribution have been developed. These methods are important for one to understand the structure of a genetic regulatory network and it also facilitates the study and the design of optimal control policies for gene intervention [8,10,28].

Network inference from steady-state data is essential in that most microarray data sets are presumed to be obtained from sampling the steady-state. Two algorithms have been proposed in [25] to find attractors composing a BN. Here we consider an inverse problem of constructing a PBN based on the prescribed positive stationary probability distribution.

This problem was first formulated as the inverse problem in [37] in the form of a constrained least squares problem. The basic idea is to impose a suitable criteria for selecting PBNs, the criteria is to maximize the entropy rate of a Markov chain [17]. For more details about information theory and entropy, we refer readers to [11]. A heuristic method based on Conjugate Gradient (CG) algorithm, an iterative method, was then proposed to solve the resulting least squares problem. The computational cost of this problem is huge [25, 37], we therefore seek for further reduction and approximation. The idea is to introduce sparsity into the construction. On one hand, it can reduce the complexity of the problem and on the other hand it agrees with the gene network property in certain extend [1]. The problem can be split into two sub-problems as follow: (i) Construct a sparse transition probability matrix from a given positive stationary distribution (Problem 1); and (ii) Construct a PBN from the obtained sparse transition probability matrix (Problem 2). We note that a sparse matrix is preferable, as it is easier to identify the major component of a network. For Problems 1 and 2, a favorable result has been obtained through adding an α-norm term to the objective function [3]. Here we shall focus more on the following problem: given a positive stationary distribution, we try to obtain a series of sparse transition probability matrices and then construct a PBN from them.

Fixing the number of nonzero entries in the column of transition probability matrix, one can obtain a unique solution among all the linear combinations of the sparsest transition probability matrices in the context of maximum entropy [27]. Details will be elucidated in the following sections with proof. The construction of PBN is then tackled efficiently once knowing the PBN structures for the sparsest transition probability matrices respectively. In this framework, the inverse problem can be efficiently solved.

The remainder of the paper is structured as follows. In Section 2, a review on BNs and PBNs will be given. We will also briefly review our previous works on Problems 1 and 2. In Section 3, we present the mathematical formulation of the inverse problem. Mathematical properties of the model will also be discussed. Numerical experiments are given to demonstrate the effectiveness of our proposed method in Section 4. Finally, concluding remarks are given in the last section.

2. A Brief Review on BNs and PBNs

In this section, we first give an brief introduction to BNs and PBNs. We then review some previous works related to the construction of PBNs (Problems 1 and 2). Since the transition matrix A is sparse, for the rest of this section, we assume that the transition probability matrix is A and its size is $2^n \times 2^n$, where n is number of genes (vertices) in the network and each column of A has at most m non-zero entries.

A Boolean Network (BN) $G(V,F)$ consists of a set of vertices $V = \{v_1, v_2, \cdots, v_n\}$ and a list of Boolean functions $F = \{f_1, f_2, \cdots, f_n\}$ where $f_i : \{0,1\}^n \to \{0,1\}$. Define $v_i(t)$ to be the state (0 or 1) of the vertex v_i at time t. The rules of the regulatory interactions among the genes are then represented by

$$v_i(t+1) = f_i(\mathbf{v}(t)), \quad i = 1, 2, \cdots, n$$

Table 1: The Truth Table.

State	$v_1(t)$	$v_2(t)$	$f^{(1)}$	$f^{(2)}$
1	0	0	0	0
2	0	1	1	0
3	1	0	0	1
4	1	1	1	0

where $\mathbf{v}(t) = [v_1(t), v_2(t), \cdots, v_n(t)]^T$ is called the Gene Activity Profile (GAP). The GAP can take any possible forms (states) from the set

$$S = \{(v_1, v_2, \cdots, v_n)^T : v_i \in \{0, 1\}\} \tag{2.1}$$

and totally there are 2^n possible states in the network. It is known that eventually a BN will enter into a cycle (attractor cycle) and stay there forever. The cycles actually have biological significance such as cell proliferation, differentiation and apoptosis [15, 16].

The following is an example of a two-gene BN (taken from [7]) with its truth table being given in Table 1. From the table, one can see that if the current network state is 1 then it will make a transition to itself in one step. The next transition step is state 3 if the current state is either 2 or 4. Finally if the current state is 3, the state in the next step will be 2. The transition probability matrix (Boolean Network matrix) of the 2-gene BN is then given by

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{2.2}$$

Since the network is deterministic, each column in B has only one non-zero element and the column sum is one. We remark that there is an one-to-one relation between a BN and its corresponding BN matrix.

Since there are random errors in data measurement, stochastic model is more preferable. In view of this, to overcome the deterministic rigidity of a BN model, it is natural to consider a stochastic extension of a BN model. To extend the BN model to a stochastic model, for each vertex v_i in a PBN, instead of having only one Boolean function as in the case of a BN, there are a number of Boolean functions (predictor functions) $f_j^{(i)}(j = 1, 2, \cdots, l(i))$ to be chosen for determining the state of gene v_i. Here $l(i) \leq 2^{2^n}$ and $l(i)$ is the total number of possible Boolean functions of gene i available. Then there are

$$N = \prod_{i=1}^{n} l(i)$$

different possible realizations of BNs. The probability of choosing $f_j^{(i)}$ as the predictor

Table 2: The Truth Table.

State	$v_1(t)$	$v_2(t)$	$f_1^{(1)}$	$f_2^{(1)}$	$f_1^{(2)}$	$f_2^{(2)}$
1	0	0	0	1	0	1
2	0	1	1	0	0	1
3	1	0	0	1	1	0
4	1	1	1	0	0	1

function is $c_j^{(i)}$ where

$$0 \le c_j^{(i)} \le 1 \quad \text{and} \quad \sum_{j=1}^{l(i)} c_j^{(i)} = 1, \quad i = 1, 2, \cdots, n.$$

Since there are N possible realizations of BNs and they are characterized by N vector functions $f_1, f_2, \cdots, f_N$ ordered lexicographically. Here $f_1 = (f_1^{(1)}, f_1^{(2)}, \cdots, f_1^{(n)})$ is the first vector function for the first BN and $f_N = (f_{l(1)}^{(1)}, f_{l(2)}^{(2)}, \cdots, f_{l(n)}^{(n)})$ is the last vector function for the Nth BN. Then in an independent PBN (the selection of the Boolean function for each gene is assumed to be independent), the probability of choosing the kth BN having the vector function $(f_{k_1}^{(1)}, f_{k_2}^{(2)}, \cdots, f_{k_n}^{(n)})$ is given by

$$q_k = \prod_{i=1}^{n} c_{k_i}^{(i)}, \quad k = 1, 2, \cdots, N. \tag{2.3}$$

We note that the transition process among the states of the set S in (2.1) is a Markov chain process. Let $\mathbf{a}$ and $\mathbf{b}$ be any two column vectors in the set S. Then the transition probability of the Markov chain is given by

$$\begin{aligned} &\text{Prob}\left\{\mathbf{v}(t+1) = \mathbf{a} \mid \mathbf{v}(t) = \mathbf{b}\right\} \\ = &\sum_{j=1}^{N} \text{Prob}\left\{\mathbf{v}(t+1) = \mathbf{a} \mid \mathbf{v}(t) = \mathbf{b}, \text{the } k\text{th network is selected}\right\} \cdot q_k. \end{aligned}$$

The transition probability matrix A of the captured PBN (Markov chain) can then be obtained by computing the above probabilities for all the possible states in the set S. It is straightforward to show that the transition probability matrix A can be written as the sum of the BN matrices A_k (see for instance [6]):

$$A = \sum_{k=1}^{N} q_k A_k. \tag{2.4}$$

Here q_k is the probability of choosing the BN having the BN matrix A_k.

Let us consider a 2-gene example, a PBN. The truth table is given in Table 2. In this example, each gene has two Boolean functions and therefore we have $l(i) = 2$ $(i = 1,2)$ and there are $N = \prod_{i=1}^{2} l(i) = 4$ possible realizations and they are

$$f_1 = (f_1^{(1)}, f_1^{(2)}), \quad f_2 = (f_1^{(1)}, f_2^{(2)}), \quad f_3 = (f_2^{(1)}, f_1^{(2)}), \quad f_4 = (f_2^{(1)}, f_2^{(2)}).$$

Moreover, if

$$c_1^{(1)} = 0.6, \qquad c_2^{(1)} = 0.4, \qquad c_1^{(2)} = 0.5, \qquad c_2^{(2)} = 0.5,$$

then we can get the selection probabilities of the corresponding BNs as follow:

$$\begin{cases} q_1 = c_1^{(1)} \cdot c_1^{(2)} = 0.3, & q_2 = c_1^{(1)} \cdot c_2^{(2)} = 0.3, \\ q_3 = c_2^{(1)} \cdot c_1^{(2)} = 0.2, & q_4 = c_2^{(1)} \cdot c_2^{(2)} = 0.2. \end{cases}$$

The transition probability matrices of corresponding BNs can be given as follow:

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad A_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

We note that transition probability matrix can be written as follows:

$$A = 0.3 \cdot A_1 + 0.3 \cdot A_2 + 0.2 \cdot A_3 + 0.2 \cdot A_4.$$

2.1. A maximum entropy approach for construction of a transition probability matrix

It is known that both Problems 1 and 2 are difficult inverse problems. In [7], some heuristic algorithms are proposed to construct a PBN from a given transition probability matrix (Problem 2). It has been shown that entropy can be used as a measurement for finding good candidates. In this subsection, we briefly review the maximum entropy method proposed in [2] for Problem 2. As we mentioned before, because the entries in $\mathbf{q}$ and A are non-negative, assuming that each column of A has m non-zero entries, there are at most m^{2^n} BNs constituting this PBN. We label the transition probability matrices of these BNs by $A_1, A_2, \cdots, A_{m^{2^n}}$ systematically.

To construct a PBN, the parameters $\mathbf{q}$ (a distribution) has to satisfy the following constraints:

$$\sum_{i=1}^{m^{2^n}} q_i A_i = A \tag{2.5}$$

and

$$0 \le q_i \le 1 \quad \text{and} \quad \sum_{i=1}^{m^{2^n}} q_i = 1.$$

Usually, there are too many feasible solutions and we then adopt entropy as the measurement to narrow down the solution set or even get an unique (optimal) solution. In fact, entropy is a measure of the uncertainty associated with a random variable. It measures, in the sense of an expected value, the information contained in a message [27]. Entropy can also be regarded as a measure of the multiplicity associated with the state of the objects [2,37].

To simplify the discussion, we define (see for instance [2])

$$M\left(\begin{bmatrix} a_{11} & \cdots & a_{1l} \\ \vdots & \vdots & \vdots \\ a_{l1} & \cdots & a_{ll} \end{bmatrix}\right) = [a_{11}, \cdots, a_{l1}, a_{12}, \cdots, a_{l2}, \cdots, \cdots, a_{1l}, \cdots a_{ll}]^T \tag{2.6}$$

and we let

$$U = [M(A_1), M(A_2), \cdots, M(A_{m^{2^n}})] \quad \text{and} \quad \mathbf{p} = M(A). \tag{2.7}$$

To ensure the condition

$$\sum_{i=1}^{m^{2^n}} q_i = 1$$

we add a row vector of $[1,1,\cdots,1]$ to the bottom of the matrix U and form a new matrix $\bar{U}$. Meanwhile, we add an entry 1 at the end of the vector $\mathbf{p}$ to get a new vector $\bar{\mathbf{p}}$. The maximum entropy algorithm can be formulated as follows:

$$\max_{\mathbf{q}} \sum_{i=1}^{m^{2^n}} (-q_i \log q_i) \tag{2.8}$$

subject to

$$\bar{U}\mathbf{q} = \bar{\mathbf{p}} \quad \text{and} \quad 0 \le q_i, \quad i = 1,2,\cdots,m^{2^n}.$$

We remark that the constraint $q_i \le 1$ can be discarded as we require that

$$\sum_{i=1}^{m^{2^n}} q_i = 1, \quad 0 \le q_i, \quad i = 1,2,\cdots,m^{2^n}.$$

Newton's method in conjunction with CG method can be applied to solve the above problem [2,3].

For demonstration purpose, we consider a PBN with $n = 2$ and $m = 2$. Suppose the observed transition probability matrix of the PBN is given as follows:

$$A_{2,2} = \begin{bmatrix} 0.1 & 0.3 & 0.5 & 0.6 \\ 0.0 & 0.7 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 \\ 0.9 & 0.0 & 0.0 & 0.4 \end{bmatrix}.$$

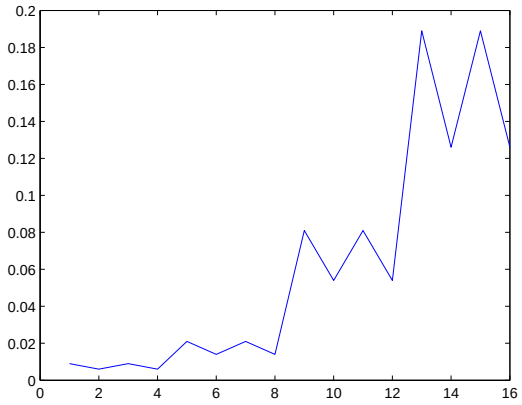

Figure 1: The distribution $\mathbf{q}$ obtained by the maximum entropy approach (Taken from [2]).

Using this maximum entropy approach, one can obtain the solution as shown in Fig. 1. The x-axis represents the kth BN and the y-axis is the probability distribution q choosing the BN. It is been shown that the re-constructed PBN is supposed to be dominated by the 9th, the 11st and the last 4 BNs. Here we see that this method can be used to identify the major components of the BNs constituting the PBN.

A modified entropy approach has been proposed in [3]. By adding an α-norm to the objective function (2.8), we can generate a PBN with sparser $\mathbf{q}$ as suggested in [4,34]. The mathematical formulation for Problem 2 becomes

$$\max_{\mathbf{q}} \left\{ -\sum_{i=1}^{m^{2^n}} q_i \log q_i - \beta \sum_{i=1}^{m^{2^n}} q_i^{\alpha} \right\} \tag{2.9}$$

subject to

$$\bar{U}\mathbf{q} = \bar{\mathbf{p}} \quad \text{and} \quad 0 \le q_i, \quad i = 1, 2, \cdots, m^{2^n},$$

where $0 < \alpha < 1$ and $\beta \ge 0$. The first term is the entropy as in (2.8) and the second term is the α-norm part which helps in getting a sparse solution $\mathbf{q}$. Here α and β are two parameters. In practice, a grid search method was adopted to find optimal values of α and β.

We use the same example given in (2.9) to demonstrate this method and compare with the maximum entropy method as well. Using this modified entropy approach, we obtain the solution as shown in Fig. 2. The optimal solution is reached when $\alpha = 0.63$ and $\beta = 1.40$. Compared it with the result in Fig. 1, one can see that the modified entropy approach can get a sparser solution.

Now the approach can also be applied to solving Problem 1. For Problem 1, suppose the

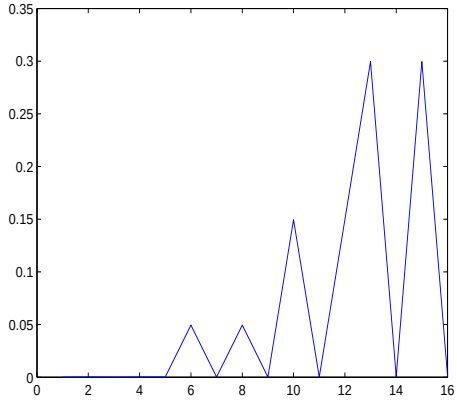

Figure 2: The distribution $\mathbf{q}$ obtained by the modified entropy approach (Taken from [3]).

observed stationary distribution is $\boldsymbol{\pi} = [\pi_1, \cdots, \pi_{2^n}]^T$, we have the following algorithm:

$$\max_{p_{ij}} \left\{ \sum_{j=1}^{2^n} \pi_j \left(-\sum_{i=1}^{2^n} p_{ij} \log p_{ij} \right) - \sum_{j=1}^{2^n} \left(\beta \sum_{i=1}^{2^n} p_{ij}^{\alpha} \right) \right\} \tag{2.10}$$

subject to

$$\begin{cases} \sum_{i=1}^{2^n} p_{ij} = 1, \ j = 1, 2, \cdots, 2^n, \\ P\boldsymbol{\pi} = \boldsymbol{\pi}, \\ p_{ij} \geq 0, \ i, j = 1, 2, \cdots, 2^n, \end{cases} \tag{2.11}$$

where the first term in (2.10) is the entropy rate of the Markov chain having transition probability matrix $P = (p_{ij})_{2^n \times 2^n}$. Here $0 < \alpha < 1$ and $\beta \geq 0$ are weightings that one can also use grid search method to find their optimal values. We remark that for Problem 1, suppose the observed steady-state distribution is

$$\boldsymbol{\pi} = [0.1, 0.2, 0.3, 0.4]^T.$$

Using the modified entropy method, we get the optimal transition probability matrix as follows:

$$P = \begin{bmatrix} 0.0000 & 0.0830 & 0.1126 & 0.1240 \\ 0.0000 & 0.2097 & 0.2234 & 0.2276 \\ 0.1902 & 0.3250 & 0.3115 & 0.3063 \\ 0.8098 & 0.3824 & 0.3525 & 0.3420 \end{bmatrix}.$$

The optimal solution is reached when $\alpha = 0.94$ and $\beta = 1.6$ as suggested in [3].

One of the major drawbacks of the proposed methods is that the computational cost is very expensive and is infeasible for a large network. To reduce the computational cost, we consider constructing PBNs from a class of sparse matrices. This idea is motivated by the fact that biological networks are robust in general and more than one network can be inferred from the data [24].

3. The New Approach

In this section, we would like to provide the mathematical formulation of the inverse problem. Since it can be divided into two sub-problems : (i) constructing a sparse transition probability matrix from the stationary distribution (Problem 1) and (ii) approximating the sparse transition probability matrix with PBNs, we will study them one by one (Problem 2).

3.1. Sparse transition probability matrix construction

We assume the number of genes to be n, $\pi = [\pi_1, \pi_2, \cdots, \pi_N]^T$ is a positive stationary probability distribution where $N = 2^n$. Once knowing the stationary distribution π, we can obtain $N-1$ sparse transition probability matrices as follow:

$$T_1 = \begin{bmatrix} 1-t_1 & 0 & \cdots & 0 & 0 & t_N \\ t_1 & 1-t_2 & \cdots & 0 & 0 & 0 \\ 0 & t_2 & \ddots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & t_{N-2} & 1-t_{N-1} & 0 \\ 0 & 0 & \cdots & 0 & t_{N-1} & 1-t_N \end{bmatrix}, \tag{3.1}$$

$$T_2 = \begin{bmatrix} 1-t_1 & 0 & \cdots & 0 & t_{N-1} & 0 \\ 0 & 1-t_2 & \cdots & 0 & 0 & t_N \\ t_1 & 0 & 1-t_3 & 0 & 0 & 0 \\ 0 & t_2 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 1-t_{N-1} & 0 \\ 0 & 0 & \cdots & t_{N-2} & 0 & 1-t_N \end{bmatrix}, \tag{3.2}$$

$$\vdots$$

$$T_{N-1} = \begin{bmatrix} 1-t_1 & t_2 & \cdots & 0 & 0 & 0 \\ 0 & 1-t_2 & t_3 & 0 & 0 & 0 \\ 0 & 0 & 1-t_3 & t_4 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 & 1-t_{N-1} & t_N \\ t_1 & 0 & \cdots & 0 & 0 & 1-t_N \end{bmatrix}. \tag{3.3}$$

Here

$$t_i = \frac{\dfrac{1}{\pi_i}}{\sum\limits_{i=1}^{i=N} \dfrac{1}{\pi_i}}, \quad i = 1, 2, \cdots, N.$$

We note that these matrices are the sparsest irreducible matrices.

Next, we consider the linear combinations of the $N-1$ sparsest transition probability matrices. Suppose we fix the number of nonzero entries in each column of the transition probability matrix to be k, we are interested in getting a desirable (optimal) matrix in the sense of maximizing the entropy. Mathematically speaking, we are to solve the following problem:

$$\max_{p_{ij}} \left\{ -\sum_{j=1}^{N} \pi_j \sum_{i=1}^{N} p_{ij} \log p_{ij} \right\}$$

subject to

$$\begin{cases} \sum\limits_{i=1}^{N} p_{ij} = 1, \ j = 1, 2, \cdots, N, \\ P\pi = \pi, \ P = [p_{ij}], \\ p_{ij} \geq 0, \ i, j = 1, 2, \cdots, N, \\ P = \sum\limits_{j=1}^{k-1} q_{i_j} T_{i_j}, \\ \sum\limits_{j=1}^{k-1} q_{i_j} = 1, \\ q_{i_j} > 0, \\ \{i_1, i_2, \cdots, i_{k-1}\} \subset \{1, 2, \cdots, N\}. \end{cases} \tag{3.4}$$

Theorem 3.1. *If the number of nonzero entries in each column of the transition probability matrix is fixed to be k, then one can have C_{N-1}^{k-1} optimal solutions.*

Proof. There are $N-1$ sparsest transition probability matrices. We also note that for all the $N-1$ matrices, the number of nonzero entries in each column is 2. Together with the particular characteristics embedded in these matrices we need exactly $k-1$ matrices to make a transition probability matrix with k nonzero entries in each column. We prove that the uniformly (probability $1/(k-1)$) linear combination of any $k-1$ matrices satisfy the above optimization problem.

We assume that

$$P = \sum_{j=1}^{k-1} q_{i_j} T_{i_j}$$

then the entropy can be expressed as follows:

$$\begin{aligned}
&-\sum_{j=1}^{N}\pi_j\sum_{i=1}^{N}p_{ij}\log p_{ij}\\
&=-\sum_{j=1}^{N}\pi_j\left[(1-t_j)\log(1-t_j)+\sum_{l=1}^{k-1}q_{i_l}t_j\log(q_{i_l}t_j)\right]\\
&=-\sum_{j=1}^{N}\pi_j(1-t_j)\log(1-t_j)-\sum_{j=1}^{N}\pi_j\sum_{l=1}^{k-1}\left[q_{i_l}t_j(\log(q_{i_j})+\log(t_j)))\right]\\
&=-\sum_{j=1}^{N}\pi_j(1-t_j)\log(1-t_j)-\sum_{j=1}^{N}\pi_jt_j\sum_{l=1}^{k-1}\left[q_{i_l}\log(q_{i_l})\right]-\sum_{j=1}^{N}\pi_jt_j\log(t_j)\\
&=-\sum_{j=1}^{N}\pi_j\left[(1-t_j)\log(1-t_j)+t_j\log(t_j)\right]-\sum_{j=1}^{N}\pi_jt_j\sum_{l=1}^{k-1}\left[q_{i_l}\log(q_{i_l})\right].
\end{aligned}$$

Since π_i, t_i, $i=1,2,\cdots,N$ are already known, they can be viewed as constants. Apart from that, because of the convexity of the function $x\log(x)$, one can easily achieve the optimal solution when

$$q_{i_c}=q_{i_d}=\frac{1}{k-1},\quad c,\ d\in\{1,2,\cdots,k-1\}.$$

This clearly states the irrelevance in selection of the $k-1$ matrices. Therefore, we can have C_{N-1}^{k-1} kinds of combinations as long as all the combination coefficients are $1/(k-1)$. Therefore this completes the proof. □

From the above theorem, we know that for any fixed number k, there are C_{N-1}^{k-1} optimal solutions among all the $k-1$ linear combinations of the sparsest transition probability matrices with the same maximum entropy. By varying the parameter k, one can improve the entropy and get a better solution. The following theorem addresses this issue.

Theorem 3.2. *The optimal entropy of the transition probability matrix is an increasing function in k (number of nonzero entries in each column).*

Proof. From Theorem 3.1, we define the optimal entropy function $E(k)$ as follows:

$$\begin{aligned}
E(k)&=-\sum_{j=1}^{N}\pi_j\left[(1-t_j)\log(1-t_j)+t_j\log(t_j)\right]-\sum_{j=1}^{N}\pi_jt_j\sum_{l=1}^{l=k-1}\left[\frac{1}{k-1}\log\left(\frac{1}{k-1}\right)\right]\\
&=-\sum_{j=1}^{N}\pi_j\left[(1-t_j)\log(1-t_j)+t_j\log(t_j)\right]+\left[\sum_{j=1}^{N}\pi_jt_j\right]\log(k-1)
\end{aligned}$$

and

$$E(k+1)-E(k)=\left[\sum_{j=1}^{N}\pi_j t_j\right]\log\left(1+\frac{1}{k-1}\right)>0.$$

This completes the proof. □

Since for a fixed number k in transition probability matrix construction, we can have C_{N-1}^{k-1} optimal solutions within all the possible $k-1$ linear combinations of the sparsest transition probability matrices. Among the C_{N-1}^{k-1} possibilities, we would like to choose the most preferable one. We have the following theorem.

Theorem 3.3. *The solution to the following optimization problem:*

$$\max_{p_{ij}}\left\{-\sum_{j=1}^{N}\pi_j\sum_{i=1}^{N}p_{ij}\log p_{ij}\right\}$$

subject to

$$\begin{cases}\displaystyle\sum_{i=1}^{N}p_{ij}=1,\ j=1,2,\cdots,N,\\ P\boldsymbol{\pi}=\boldsymbol{\pi},\ P=[p_{ij}],\\ p_{ij}\geq 0,\ i,j=1,2,\cdots,N,\end{cases}\tag{3.5}$$

is achieved when $p_{ij}=\pi_i,\ i,j=1,2,\cdots,N$.

Proof. The optimization problem is equivalent to the following minimization problem:

$$\min_{p_{ij}}\left\{\sum_{j=1}^{N}\pi_j\sum_{i=1}^{N}p_{ij}\log p_{ij}\right\}$$

with the same constraints. We then apply the method of Lagrange multiplier and we rewrite

$$\begin{cases}f=\displaystyle\sum_{j=1}^{N}\pi_j\sum_{i=1}^{N}p_{ij}\log p_{ij},\\ k_i=\displaystyle\sum_{j=1}^{N}p_{ij}\pi_j-\pi_i,\quad i=1,2,\cdots,N,\\ h_j=\displaystyle\sum_{i=1}^{N}p_{ij}-1,\quad j=1,2,\cdots,N,\\ g_{ij}=-p_{ij},\quad i,j=1,2,\cdots,N.\end{cases}$$

Then we have for $\boldsymbol{\lambda} \in R^N$, $\mathbf{r} \in R^N$, $\boldsymbol{\mu} \in R^{N^2}$, such that

$$\begin{cases} \nabla f + \nabla \mathbf{h}\boldsymbol{\lambda} + \nabla \mathbf{k}\mathbf{r} + \nabla \mathbf{g}\boldsymbol{\mu} = \mathbf{0}, \\ \mathbf{g}\boldsymbol{\mu} = \mathbf{0}, \\ \boldsymbol{\mu} \geq \mathbf{0}, \end{cases}$$

where

$$\begin{cases} \mathbf{g} = [g_{11}, g_{12}, \cdots, g_{1N}, g_{21}, g_{22}, \cdots, g_{2N}, \cdots, g_{NN}], \\ \mathbf{h} = [h_1, h_2, \cdots, h_N], \\ \mathbf{k} = [k_1, k_2, \cdots, k_N]. \end{cases}$$

We note that $p_{ij} > 0$, so the non-negativity constraints are indeed inactive. Then we can have the following equations:

$$\begin{cases} \pi_j(1 + \log(p_{ij})) + \lambda_j + r_i \pi_j = 0, \quad i, j = 1, 2, \cdots, N, \\ \sum_{i=1}^{i=N} p_{ij} = 1, \quad j = 1, 2, \cdots, N, \\ \sum_{j=1}^{N} p_{ij} \pi_j = \pi_i, \quad i = 1, 2, \cdots, N. \end{cases}$$

We therefore have

$$p_{ij} = e^{-1 - \frac{\lambda_j}{\pi_j} - r_i}, \quad i, j = 1, 2, \cdots, N.$$

Using the condition that

$$\sum_{i=1}^{N} p_{ij} = 1, \quad j = 1, 2, \cdots, N$$

one can obtain

$$e^{-1 - \frac{\lambda_j}{\pi_j}} \sum_{i=1}^{N} e^{-r_i} = 1, \quad j = 1, 2, \cdots, N. \tag{3.6}$$

On the other hand, with

$$\sum_{j=1}^{N} p_{ij} \pi_j = \pi_i, \quad i = 1, 2, \cdots, N$$

we get

$$e^{-r_i} \sum_{j=1}^{N} \pi_j e^{-1 - \frac{\lambda_j}{\pi_j}} = \pi_i, \quad i = 1, 2, \cdots, N. \tag{3.7}$$

Using Equation (3.6), we know that $e^{-1-(\lambda_j/\pi_j)}$ is a constant. If we define

$$e^{-1 - \frac{\lambda_j}{\pi_j}} = C$$

then from Equation (3.7) we have

$$e^{-r_i}C = \pi_i, \quad i = 1, 2, \cdots, N.$$

This implies that

$$p_{ij} = e^{-r_i - 1 - \frac{\lambda_j}{\pi_j}} = e^{-r_i}C = \pi_i, \quad i, j = 1, 2, \cdots, N.$$

We note that

$$\nabla^2 f + \sum_{i=1}^{N} \lambda_i \nabla^2 h_i + \sum_{j=1}^{N} r_j \nabla^2 k_j + \sum_{l=1}^{N}\sum_{m=1}^{N} \mu_{lm} \nabla^2 g_{ij} = \mathrm{Diag}(\tilde{p})$$

where

$$\tilde{p} = \left[\frac{\pi_1}{p_{11}}, \frac{\pi_2}{p_{12}}, \cdots, \frac{\pi_N}{p_{1N}}, \frac{\pi_1}{p_{21}}, \cdots, \frac{\pi_N}{p_{NN}}\right]^T.$$

It is clear that the Hessian matrix is positive definite, hence $p_{ij} = \pi_i$, $i, j = 1, 2, \cdots, N$ is the global optimal minimum point of the function f subject to the constraints. Hence the proof is complete. □

Knowing that $T_{std} = [p_{ij}]$ where $p_{ij} = \pi_i$, $(i, j = 1, 2, \cdots, N)$ is the transition probability matrix with maximum entropy, we can differentiate (select the best) the C_{N-1}^{k-1} possible transition probability matrices by using the Euclidean distances between the matrices and the matrix T_{std}. Fixing the number k, for matrix $TestM$ in C_{N-1}^{k-1} transition probability matrices, the smallest distance between $TestM$ and T_{std} indicates that the corresponding matrix contains the most abundant information. We therefore choose the matrix as our preferable transition probability matrix.

3.2. Construction of PBNs from the selected transition probability matrix

In the process of constructing PBNs from the transition probability matrix, a favorable result has been obtained through the technique of adding α-norm to the objective function. We thus can utilize the algorithm to get the desired PBNs.

For T_i, $i = 1, 2, \cdots, N$, as there are only two nonzero entries in each column in these matrices, it is very fast to get the desired PBNs for all the $N - 1$ matrices. Without loss of generality, we hypothesize that

$$T_j = \sum_{l=1}^{M} coef_l^j B_l^j, \quad j = 1, 2, \cdots, N.$$

For a fixed number k, if we assume the optimal transition probability matrix is in the following expression:

$$OP_k = \frac{1}{k-1}\sum_{j=1}^{k-1} T_{i_j}, \quad \{i_j, j = 1, 2, \cdots, k-1\} \subset \{1, 2, \cdots, N\}.$$

Then we can directly get the desired PBNs for OP_k without further computation.

$$OP_k = \frac{1}{k-1}\sum_{j=1}^{k-1}\sum_{l=1}^{M} coef_l^{i_j} B_l^{i_j}.$$

This would reduce much time for the construction of PBNs, in particular when k is large, as the major computational complexity was efficiently reduced from $O(k^3 2^{3n})$ to $O(k2^{3n})$ [2], thereby offering a new perspective in PBN construction.

4. Numerical Experiments

In this section, we present numerical experiments to illustrate the effectiveness of our proposed method. Suppose the number of genes is $n = 2$, then the number of states $N = 2^n = 4$. Further assume the stationary distribution is

$$\pi = [0.1, 0.2, 0.3, 0.4]^T.$$

Once we know the information, it's straightforward to construct the set of sparsest probability transition matrices T_1, T_2, T_3.

$$T_1 = \begin{bmatrix} 0.52 & 0.00 & 0.00 & 0.12 \\ 0.48 & 0.76 & 0.00 & 0.00 \\ 0.00 & 0.24 & 0.84 & 0.00 \\ 0.00 & 0.00 & 0.16 & 0.88 \end{bmatrix}, \qquad T_2 = \begin{bmatrix} 0.52 & 0.00 & 0.16 & 0.00 \\ 0.00 & 0.76 & 0.00 & 0.12 \\ 0.48 & 0.00 & 0.84 & 0.00 \\ 0.00 & 0.24 & 0.00 & 0.88 \end{bmatrix}, \quad (4.1)$$

$$T_3 = \begin{bmatrix} 0.52 & 0.24 & 0.00 & 0.00 \\ 0.00 & 0.76 & 0.16 & 0.00 \\ 0.00 & 0.00 & 0.84 & 0.12 \\ 0.48 & 0.00 & 0.00 & 0.88 \end{bmatrix} \quad \text{and} \quad T_{std} = \begin{bmatrix} 0.1 & 0.1 & 0.1 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.2 \\ 0.3 & 0.3 & 0.3 & 0.3 \\ 0.4 & 0.4 & 0.4 & 0.4 \end{bmatrix}. \quad (4.2)$$

Applying the α-norm algorithm proposed in [3] to construct the PBNs, we get the major components of BNs and their corresponding coefficients as follows:

$$T_1 \approx 0.12\times \begin{bmatrix} 1&0&0&1\\0&1&0&0\\0&0&1&0\\0&0&0&0 \end{bmatrix} + 0.16\times \begin{bmatrix} 1&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&1&1 \end{bmatrix} + 0.24\times \begin{bmatrix} 1&0&0&0\\0&0&0&0\\0&1&1&0\\0&0&0&1 \end{bmatrix} + 0.48\times \begin{bmatrix} 0&0&0&0\\1&1&0&0\\0&0&1&0\\0&0&0&1 \end{bmatrix},$$

$$T_2 \approx 0.12\times \begin{bmatrix} 1&0&0&0\\0&1&0&1\\0&0&1&0\\0&0&0&0 \end{bmatrix} + 0.16\times \begin{bmatrix} 1&0&1&0\\0&1&0&0\\0&0&0&0\\0&0&0&1 \end{bmatrix} + 0.24\times \begin{bmatrix} 1&0&0&0\\0&0&0&0\\0&0&1&0\\0&1&0&1 \end{bmatrix} + 0.48\times \begin{bmatrix} 0&0&0&0\\0&1&0&0\\1&0&1&0\\0&0&0&1 \end{bmatrix},$$

$$T_3 \approx 0.12\times \begin{bmatrix} 1&0&0&0\\0&1&0&0\\0&0&1&1\\0&0&0&0 \end{bmatrix} + 0.16\times \begin{bmatrix} 1&0&0&0\\0&1&1&0\\0&0&0&0\\0&0&0&1 \end{bmatrix} + 0.24\times \begin{bmatrix} 1&1&0&0\\0&0&0&0\\0&0&1&0\\0&0&0&1 \end{bmatrix} + 0.48\times \begin{bmatrix} 0&0&0&0\\0&1&0&0\\0&0&1&0\\1&0&0&1 \end{bmatrix}.$$

Table 3: 2-norm of the C_{N-1}^{k-1} matrices with T_{std} for different k.

$k=2$	T_1	T_2	T_3
2-norm with T_{std}	0.9763	1.0554	0.9208
$k=3$	$(T_1+T_2)/2$	$(T_1+T_3)/2$	$(T_2+T_3)/2$
2-norm with T_{std}	0.8834	0.8704	0.8482
$k=4$	$(T_1+T_2+T_3)/3$		
2-norm with T_{std}	0.8374		

Table 4: Construction of optimal transition probability matrices for different k.

	PBNs
$k=2$	$0.12\times T_{13}+0.16\times T_{23}+0.24\times T_{33}+0.48\times T_{43}$
$k=3$	$\frac{1}{2}\Big[0.12\times T_{12}+0.16\times T_{22}+0.24\times T_{32}+0.48\times T_{42}+0.12\times T_{13}$ $+0.16\times T_{23}+0.24\times T_{33}+0.48\times T_{43}\Big]$
$k=4$	$\frac{1}{3}\Big[0.12\times T_{11}+0.16\times T_{21}+0.24\times T_{31}+0.48\times T_{41}+0.12\times T_{12}+0.16\times T_{22}$ $+0.24\times T_{32}+0.48\times T_{42}+0.12\times T_{13}+0.16\times T_{23}+0.24\times T_{33}+0.48\times T_{43}\Big]$

Now we are ready to construct PBNs for certain given conditions. Let $k=2$, with the support of Theorem 3.3, we can find the desirable optimal transition probability matrix: T_3 having smallest distance with T_{std}. Let $k=3$, we can find the desired optimal transition probability matrix: $(T_2+T_3)/2$. When $k=4$, in a similar manner, the desired optimal transition probability matrix is $(T_1+T_2+T_3)/3$. For better illustration, see Table 3.

Since we have already got the decomposition for transition probability matrices T_1, T_2, T_3, the corresponding PBNs for the optimal transition probability matrices with different k can then be efficiently constructed. If we rewrite the PBNs for all the probability transition matrices T_1, T_2, T_3 in the following way:

$$\begin{cases} T_1 \approx 0.12\times T_{11}+0.16\times T_{21}+0.24\times T_{31}+0.48\times T_{41}, \\ T_2 \approx 0.12\times T_{12}+0.16\times T_{22}+0.24\times T_{32}+0.48\times T_{42}, \\ T_3 \approx 0.12\times T_{13}+0.16\times T_{23}+0.24\times T_{33}+0.48\times T_{43}. \end{cases}$$

Then, for different values of k, the optimal transition probability matrices are given in Table 4.

5. Conclusions

In this paper we have proposed a novel perspective to tackle the problem of PBNs construction from a prescribed positive stationary distribution. Here we present a novel

method to construct a sparse transition probability matrix based on a given stationary distribution. A series of sparsest transition probability matrices can be determined once giving the stationary distribution. Then by fixing the number of nonzero entries in each column of the transition probability matrix, the desired sparse transition probability matrix in the sense of maximum entropy can be uniquely constructed as a linear combination of the selected sparsest transition probability matrices. Compelling support in theory and efficiency in realization constitute a powerful demonstration for our developed model.

Acknowledgments

The authors would like to thank the anonymous referees for their helpful comments and suggestions. Research support in part by HK GRF Grant, HKU CERG Grants and National Natural Science Foundation of China Grant Nos. 10971075 and 11271144.

References

[1] M. Arnone and M. Davidson, *The Hardwiring of Development: Organization and Function of Genomic Regulatory Systems*, Development, 124 (1997), pp. 1851–1864.

[2] X. Chen, W. Ching, X.S. Chen, Y. Cong and N. Tsing, *Construction of Probabilistic Boolean Networks from a Prescribed Transition Probability Matrix: A Maximum Entropy Rate Approach*, East Asian Journal of Applied Mathematics, 1 (2011), pp. 132–154.

[3] X. Chen, H. Jiang and W. Ching, *On Construction of Sparse Probabilistic Boolean Networks*, East Asian Journal of Applied Mathematics, 2 (2012), pp. 1–18.

[4] E. Candes and T. Tao, *Near-optimal Signal Recovery from Random Projections: Universal Encoding Strategies*, IEEE Transactions on Information Theory, 52 (2006), pp. 5406–5425.

[5] W. Ching and M. Ng, *Markov Chains: Models, Algorithms and Applications*, International Series on Operations Research and Management Science, Springer, New York, (2006).

[6] W. Ching, S. Zhang, M. Ng and T. Akutsu, *An Approximation Method for Solving the Steady-state Probability Distribution of Probabilistic Boolean Networks*, Bioinformatics, 23 (2007), pp. 1511–1518.

[7] W. Ching, X. Chen and N. Tsing, *Generating Probabilistic Boolean Networks from a Prescribed Transition Probability Matrix*, IET on Systems Biology, 6 (2009), pp. 453–464.

[8] W. Ching, S. Zhang, Y. Jiao, T. Akutsu, N. Tsing and A. Wong, *Optimal Control Policy for Probabilistic Boolean Networks with Hard Constraints*, IET on Systems Biology, 3 (2009), pp. 90–99.

[9] W. Ching, E. Fung, M. Ng and T. Akutsu, *On Construction of Stochastic Genetic Networks Based on Gene Expression Sequences*, International Journal of Neural Systems, 15 (2005), pp. 297–310.

[10] Y. Cong, W. Ching, N. Tsing and H. Leung, *On Finite-Horizon Control of Genetic Regulatory Networks with Multiple Hard-Constraints*, BMC Systems Biology, 4(Suppl 2) (2010), S14, doi:10.1186/1752-0509-4-S2-S14.

[11] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley and Sons, Inc.,(1991).

[12] H. de Jong, *Modeling and Simulation of Genetic Regulatory Systems: a Literature Review*, Journal of Computational Biology, 9 (2002), pp. 69–103.

[13] N. Friedman, M. Linial, I. Nachman and D. Pe'er, *Using Bayesian Networks to Analyze Expression Data*, Journal of Computational Biology, 7 (2000), pp. 601–620.

[14] M. Gibson and E. Mjolsness *Modeling the Activity of Single Genes*, in Computational Modeling of Genetic and Biochemical Networks, J. Bower and H. Bolouri (Eds.), Cambridge MA: MIT Press, (2001), Chapter 1.

[15] S. Huang, *Gene Expression Profiling, Genetic Networks, and Cellular States: An Integrating Concept for Tumorigenesis and Drug Discovery*, Journal of Molecular Medicine, 77 (1999), pp. 469–480.

[16] S. Huang and D. Ingber, *Shape-dependent Control of Cell Growth, Differentiation, and Apoptosis: Switching Between Attractors in Cell Regulatory Networks*, Experimental Cell Research, 261 (2000), pp. 91–103.

[17] E. Jaynes *Information Theory and Statistical Mechanics*, The Physical Review, 106 (1957), pp. 620–630.

[18] S. Kauffman, *Metabolic Stability and Epigenesis in Randomly Constructed Gene Nets*, Journal of Theoretical Biology, 22 (1969), pp. 437–467.

[19] S. Kauffman, *Homeostasis and Differentiation in Random Genetic Control Networks*, Nature, 224 (1969), pp. 177–178.

[20] S. Kauffman, *The Origins of Order: Self-organization and Selection in Evolution*, New York: Oxford Univ. Press, (1993).

[21] W. Li, L. Cui and M. Ng, *On Computation of the Steady-State Probability Distribution of Probabilistic Boolean Networks with Gene Perturbation*, Journal of Computational and Applied Mathematics, 236 (2012), pp. 4067-4081.

[22] R. Linden and A. Bhaya, *Evolving Fuzzy Rules to Model Gene Expression*, Biosystems, 88 (2007), pp. 76–91.

[23] M. Song, Z. Ouyang and Z. Liu, *Discrete Dynamical System Modeling for Gene Regulatory Networks of 5-hxymethylfurfural Tolerance for Ethanologenic Yeast*, IET on Systems Biology, 3 (2009), pp. 203–218.

[24] S. Martin, A. Zhang, A. Martino and J. Faulon, *Boolean Dynamics of Genetic Regulatory Networks inferred from Microarray time Series Data*, Bioinformatics, 23 (2007), pp. 886–874.

[25] R. Pal, I. Ivanov, A. Datta, M. Bittner and E. Dougherty, *Generating Boolean Networks with a Prescribed Attractor Structure*, Bioinformatics, 21 (2005), pp. 4021–4025.

[26] A. Ribeiro, R, Zhu, and S. Kauffman, A Genetic Modeling Strategy for Gene Regulatory Networks with Stochastic Dynamics, Journal of Computational Biology, 13 (2006), pp. 1630–1639.

[27] C. Shannon, *A Mathematical Theory of Communication*, Bell System Technical Journal, 27 (1948), pp. 379–423.

[28] Y. Shin and L. Bleris, *Linear Control Theory for Gene Network Modeling*, PLoS ONE 5(9): e12785. doi:10.1371/journal.pone.0012785.

[29] I. Shmulevich, E. Dougherty, S. Kim and W. Zhang, *Probabilistic Boolean Networks: A Rule-based Uncertainty Model for Gene Regulatory Networks*, Bioinformatics, 18 (2002), pp. 261–274.

[30] I. Shmulevich, E. Dougherty, S. Kim and W. Zhang, *Control of Stationary Behavior in Probabilistic Boolean Networks by Means of Structural Intervention*, Journal of Biological Systems, 10 (2002), pp. 431–445.

[31] I. Shmulevich, E. Dougherty, S. Kim and W. Zhang *From Boolean to Probabilistic Boolean Networks as Models of Genetic Regulatory Networks*, Proceedings of the IEEE, 90 (2002), pp. 1778–1792.

[32] I. Shmulevich and E. Dougherty, *Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks*, Philadelphia: Society for Industrial and Applied Mathematics, (2010).

[33] P. Smolen, D. Baxter and J. Byrne, *Mathematical Modeling of Gene Networks*, Neuron, 26 (2000), pp. 567–580.

[34] Z. Xu, H. Zhang, Y. Wang and X. Chang, $L_{1/2}$ *Regularization*, Science in China Series F - Information Sciences, 40 (2010), pp. 1–11.

[35] W. Xu, W. Ching, S. Zhang, W. Li and X. Chen, *A Matrix Perturbation Method for Computing the Steady-state Probability Distributions of Probabilistic Boolean Networks with Gene Perturbations*, Journal of Computational and Applied Mathematics, 235 (2011), pp. 2242–2251.

[36] S. Zhang, W. Ching, M. Ng and T. Akutsu, *Simulation Study in Probabilistic Boolean Network Models for Genetic Regulatory Networks*, Journal of Data Mining and Bioinformatics, 1 (2007), pp. 217–240.

[37] S. Zhang, W. Ching, X. Chen and N. Tsing, *Generating Probabilistic Boolean Networks from a Prescribed Stationary Distribution*, Information Sciences, 180 (2010), pp. 2560–2570.

[38] S. Zhang, W. Ching, N. Tsing, H. Leung and D. Guo, *A New Multiple Regression Approach for the Construction of Genetic Regulatory Networks*, Journal of Artificial Intelligence in Medicine, 48 (2010), pp. 153–160.