

## 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), dlkciissy@hku.hk (X. Chen), yushanqiu2526374@163.com (Y. Qiu), wching@hku.hk (W.-K. Ching)

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.