

# On the Minimum Number of Genes Required for Discriminating Steady States under a Boolean Model

XIAOQING CHENG<sup>1</sup> TAKEYUKI TAMURA<sup>2</sup> WAI-KI CHING<sup>3</sup> TATSUYA AKUTSU<sup>2,a)</sup>

**Abstract:** Identifying the type of each cell by observing gene expression patterns of a few genes is an important research topic. Since attractors are often regarded to correspond to cell types, we study the problem of finding a minimum set of nodes for discriminating given attractors, under the Boolean network model. We present combinatorial algorithms and analyze their time complexities. In particular, we prove that one node is not necessarily enough but two nodes are always enough to discriminate two periodic attractors.

## 1. Introduction

Inferring internal states of biological systems from observation on a small set of genes is an important topic in bioinformatics, systems biology, and complex networks. However, it is reported that observation on many genes/nodes is required to identify internal states of complex biological networks under both linear models [2] and nonlinear Boolean models [3].

In medical diagnosis, it is not necessarily required to infer the whole states of cells. Instead, it is often required to identify the type of each cell by observing expression patterns of a few genes (e.g., biomarkers or marker genes). Therefore, extensive studies have been done to find a small number of marker genes so that disease types or cell types can be discriminated by observing expression of these genes [1]. However, in these studies, the target types of diseases or cells are mostly limited to those related with specific diseases. Furthermore, practical aspects are focused on in most of these studies and thus it is not guaranteed that the minimum set of marker genes is always found.

Since attractors are often regarded to correspond to cell types, we study the problem of finding a minimum set of nodes that is required to discriminate given attractors, under the Boolean network model. In this short report, we give a summary of the results. Details on the algorithms, theoretical analyses, and computational experiments are given in [4].

## 2. Discrimination of Singleton Attractors

In this section, we consider discrimination of singleton attractors. Let  $B$  be an  $m \times n$  binary matrix, where each row corresponds to a singleton attractor and each column corresponds to a node in

a BN. Therefore,  $m$  and  $n$  correspond to the number of cell types and the number of genes.  $B[i, j]$  denotes the element at  $i$ -th row and  $j$ -th column. Here  $B[i, -]$  and  $B[-, j]$  denote the  $i$ -th row and  $j$ -th column of  $B$ , respectively. Let  $J = \{j_1, \dots, j_k\}$  be a set of column indices. Then,  $B_J$  denotes the submatrix of  $B$  consisting of the  $j_1, j_2, \dots, j_k$ -th columns.

### Definition 1 [MinDiscSatt]

**Input:** A set of singleton attractors represented as an  $m \times n$  binary matrix  $B$ ,

**Output:** A minimum cardinality set  $J$  of columns (i.e., nodes) such that  $B_J[i_1, -] \neq B_J[i_2, -]$  holds for all  $i_1, i_2$  with  $1 \leq i_1 \neq i_2 \leq m$ .

For example, consider the following matrix.

$$B = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}.$$

Then,  $J = \{2, 3, 5, 6\}$  is a solution of **MinDiscSatt** for this  $B$  because any pair of rows are different in

$$B_{\{2,3,5,6\}} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

By using a dynamic programming algorithm, we have the following.

**Theorem 2** **MinDiscSatt** can be solved in  $O((m/1.146)^m \text{Poly}(m, n))$  time.

## 3. Discrimination of Attractors

In Section 2, we have considered discrimination of singleton

<sup>1</sup> School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China

<sup>2</sup> Bioinformatics Center, Institute for Chemical Research, Kyoto University, Kyoto 611-0011, Japan

<sup>3</sup> Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong, China

<sup>a)</sup> takutsu@kuicr.kyoto-u.ac.jp

attractors. Here, we consider the main problem: discrimination of periodic attractors, where singleton attractors can also be included. If we handle the discrimination problem as in **MinDiscSatt**, we may not be able to get a correct solution. For example, suppose that two periodic attractors  $A_1 = [01, 10]$  and  $A_2 = [00, 11]$  are given, where 01 and 00 denote states in  $t = 0, 2, 4, \dots$  and 10 and 11 denote states in  $t = 1, 3, 5, \dots$  in  $A_1$  and  $A_2$ , respectively. If we focus on 10 and 11, we can discriminate these states by looking at the second node. However, the corresponding infinite time series data are  $[1, 0, 1, 0, \dots]$  and  $[0, 1, 0, 1, \dots]$  for  $A_1$  and  $A_2$ , respectively. These two time series are identical except the starting states. Since we cannot know which is the starting state in practice, we cannot discriminate  $A_1$  from  $A_2$  by looking at time series data for the second node. The situation is the same for the first node. However, if we look at both nodes, the time series are clearly different,  $[01, 10, 01, 10, \dots]$  versus  $[00, 11, 00, 11, \dots]$ , and thus we can discriminate  $A_1$  from  $A_2$ .

For a set  $U \subseteq V$  and an  $n$ -dimensional 0-1 vector  $\mathbf{v}$ ,  $\mathbf{v}_U$  denotes the  $|U|$ -dimensional vector consisting of elements of  $\mathbf{v}$  that correspond to  $U$ . For  $v_i \in V$ ,  $\mathbf{v}_i$  denotes the  $i$ -th coordinate value of  $\mathbf{v}$ . For example, if  $n = 5$ ,  $\mathbf{v} = (1, 0, 1, 1, 0)$ , and  $U = \{v_2, v_4, v_5\}$ , then  $\mathbf{v}_U = (0, 1, 0)$  and  $\mathbf{v}_5 = 0$ . Let

$$A_h = [\mathbf{v}(0), \mathbf{v}(1), \dots, \mathbf{v}(p(A_h) - 1)]$$

be an attractor of period  $p(A_h)$  (i.e.,  $\mathbf{v}(0) = \mathbf{v}(p(A_h))$ ). Then,  $Ser(A_h, U, t)$  denotes an infinite sequence of vectors defined by

$$Ser(A_h, U, t) = [\mathbf{v}_U(t), \mathbf{v}_U(t + 1), \mathbf{v}_U(t + 2), \dots].$$

Note that two periodic attractors  $A_h$  and  $A_k$  are identical if and only if  $Ser(A_h, V, 0) = Ser(A_k, V, 0)$  holds for some  $t \geq 0$ .

**Definition 3 [MinDiscAtt]**

**Input:** A set of attractors  $A = \{A_1, A_2, \dots, A_m\}$ , where each  $A_h$  consists of  $p(A_h)$  states represented as a  $p(A_h) \times n$  binary matrix, and  $p(A_h)$  denotes the period of  $A_h$ .

**Output:** A minimum cardinality set  $U$  of nodes such that  $Ser(A_h, U, 0) \neq Ser(A_k, U, 0)$  for any  $t \geq 0$  when  $h \neq k$ .

The following lemma states that any two periodic attractors can be discriminated by observing time series data of two genes, which is proven by using the Chinese Remainder Theorem.

**Lemma 4** For any two distinct (i.e., non-identical) attractors  $A_h = [\mathbf{v}(0), \dots, \mathbf{v}(p_h - 1)]$  and  $A_k = [\mathbf{w}(0), \dots, \mathbf{w}(p_k - 1)]$ , there exists  $U \subseteq V$  of  $|U| = 2$  for which  $Ser(A_h, U, 0) \neq Ser(A_k, U, 0)$  holds for any  $t \geq 0$ .

By using this lemma, we have the following.

**Theorem 5 MinDiscAtt** can be solved in  $O(n^{|U_{\min}}| Poly(m, n, p))$  time, where  $U_{\min}$  is an optimal solution of **MinDiscAtt** and  $p$  is the longest period of input attractors. Furthermore,  $|U_{\min}| \leq m(m - 1)$  holds.

We applied the algorithms corresponding to Theorems 2 and 5 to identification of a minimum set of nodes discriminating artificially generated attractors. The results suggest that if the number of attractors is small (e.g., less than 10), an exact solution can be found in short time. We also applied these algorithms to four

biological processes containing 10-90 nodes. In all cases, the minimum sets of nodes were found, which included genes having regulatory roles.

**References**

[1] Whitfield, M., George, L.K., Grant, G.D., and Perou, C.M.: Common markers of proliferation. *Nature Reviews Cancer*, Vol. 6, 99-106 (2006).  
 [2] Liu, Y., Slotine, J., and Barabási, A.; Observability of complex systems, *Proceedings of National Academy of Sciences, USA*, Vol. 110, 2460-2465 (2013).  
 [3] Li, R., Yang, M., and Chu, T.: Controllability and observability of Boolean networks arising from biology. *Chaos*, Vol. 24, 023104 (2015).  
 [4] Cheng, X., Tamura, T., Ching, W-K., and Akutsu, T.: Discrimination of singleton and periodic attractors in Boolean networks. *Automatica*, Vol. 84, 205-214 (2017).