Discrimination of singleton and periodic attractors in Boolean networks☆

Xiaoqing Cheng [a], Takeyuki Tamura [b], Wai-Ki Ching [c,d,e], Tatsuya Akutsu [b]

[a] *School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China*
[b] *Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan*
[c] *Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong*
[d] *Hughes Hall, Wollaston Road, Cambridge, UK*
[e] *School of Economics and Management, Beijing University of Chemical Technology, North Third Ring Road, Beijing, China*

ARTICLE INFO

ABSTRACT

Determining the minimum number of sensor nodes to observe the internal state of the whole system is important in analysis of complex networks. However, existing studies suggest that a large number of sensor nodes are needed to know the whole internal state. In this paper, we focus on identification of a small set of sensor nodes to discriminate statically and periodically steady states using the Boolean network model where steady states are often considered to correspond to cell types. In other words, we seek a minimum set of nodes to discriminate singleton and periodic attractors. We prove that one node is not necessarily enough but two nodes are always enough to discriminate two periodic attractors by using the Chinese remainder theorem. Based on this, we present an algorithm to determine the minimum number of nodes to discriminate all given attractors. We also present a much more efficient algorithm to discriminate singleton attractors. The results of computational experiments suggest that attractors in realistic Boolean networks can be discriminated by observing the states of only a small number of nodes.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Knowing internal states of complex systems is important for diagnosing various kinds of artificial, social, and biological systems. It is particularly important to identify a small set of variables so that we can reconstruct the system's complete internal state at any given time step from time-series data of these variables. In such a case, the system is called *observable*. Recent studies on complex networks revealed relationships between network properties and the number of sensor nodes corresponding to those variables (Liu, Slotine, & Barabàsi, 2013; Yan, Tsekenis, Barzel, Liu, Slotine, & Barabàsi, 2015). However, existing studies focus on linear systems and certain types of nonlinear systems. Since biological systems contain highly nonlinear switch-like components, observability studies should also be done on systems with discrete components.

The Boolean network (BN) is known as a discrete mathematical model of gene regulatory networks (Kauffman, 1993) and has been applied to modeling of various biological systems (Albert & Thakar, 2014). In a BN, each node corresponds to a gene and takes one of the two values 0 and 1, where 0 (resp., 1) means that the corresponding gene is inactive (resp., active). The value of a node at a given time step is determined according to a regulation rule represented by a set of Boolean functions. Although there exist several variants, in a widely used model, the values of network nodes are updated synchronously by using the Boolean functions and the (global) *state* of a network at a given time step is the vector of its node values. Beginning with any initial state, the system eventually falls into an *attractor*, which is classified into two types: a *singleton attractor* corresponding to a stable state and a *periodic attractor* corresponding to a sequence of states that repeats periodically. Attractors are often considered as cell types: different attractors correspond to different cell types (Kauffman, 1993). Based on this interpretation, extensive studies have been done on the distribution and length of attractors in BNs (Drossel, Mihaljev, & Greil, 2005; Kauffman, Peterson, Samuelsson, & Troein, 2004; Samuelsson & Troein, 2003).

Although attractors in synchronous BNs are either singleton or periodic, attractors in asynchronous BNs are more complex. An attractor in an asynchronous BN is a strongly connected component without outgoing edges in a state transition diagram (Saadatpour, Albert, & Albert, 2010), which is also called a loose attractor (Harvey & Bossomaier, 1997), a complex loop (Garg, Di Cara, Xenarios, Mendoza, & De Micheli, 2008), and a terminal strongly connected

component (Fauré, Naldi, Chaouiya, & Thieffry, 2006). If we consider non-linear systems obtained by extending BNs, steady states may include intermediate states (Mochizuki, 2005). Since such complex attractors or complex steady states cannot be represented as sequences of 0–1 vectors and thus are very difficult to handle, we focus on synchronous BNs in this paper.

Recently, observability of BNs has also been studied (Cheng, Qi, & Li, 2011; Laschov, Margaliot, & Even, 2013; Li, Yang, & Chu, 2015). However, due to its high nonlinearity, it is impossible in most cases to observe complete internal states of BNs from a small number of nodes. For example, Li et al. showed that more than half of the nodes are required to guarantee the observability of internal states of all attractors in the Drosophila segment polarity network (Li et al., 2015). Therefore, we need to consider another approach for distinguishing the internal states of BNs.

We note here that it is important for medical diagnosis 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 (Whitfield, George, Grant, & Perou, 2006). Furthermore, both gene expression data and network structure data are utilized to find more reliable marker genes (Hayashida & Akutsu, 2016). Dynamical gene expression data are also combined with network structure data in order to find pre-disease states (Wu, Chen, & Wang, 2014). However, in these studies, the target types of cells or diseases are mostly limited to those related with specific diseases. Furthermore, most of these studies focus on practical aspects and thus it is not guaranteed to find the minimum set of marker genes. Therefore, in this paper, we consider the problem of identifying attractors by observing activities (0 or 1) of a small number of nodes using the BN model. In particular, we focus on finding the minimum number of nodes, by which all given attractors can be discriminated. We call this problem *discrimination of attractors*. Although it is computationally intractable (#P-hard) to enumerate all singleton attractors (Akutsu, Kuhara, Maruyama, & Miyano, 1998), some algorithms have been developed to enumerate all singleton attractors for up to moderate size networks by using network reduction (Veliz-Cuba, Aguilar, Hinkelmann, & Laubenbacher, 2014) and to enumerate all singleton and periodic attractors by using network reduction and stable motifs (Zañudo & Albert, 2013). Therefore, it is reasonable to assume that a set of attractors is given. Furthermore, we can even assume that this set is given independently of a BN because gene expression data for each cell type can be experimentally obtained without knowing the structure of the underlying genetic network.

In this paper, we begin with the discrimination problem for singleton attractors, and present an algorithm that works in polynomial time of the number of genes (n) and the exponential factor only depends on the number of attractors (m). Next, we present a key result, which states that any pair of (singleton and periodic) attractors can be discriminated by observing time-series data of two nodes, by making use of the Chinese Remainder Theorem. This result gives an upper bound of the number of sensor nodes to discriminate attractors. It also leads to development of an efficient algorithm whose polynomial degree depends only on the minimum number of sensor nodes. Then, we perform computational experiments using artificially generated BNs and BN models of real biological systems. Finally, we conclude with future work.

Note that Qiu et al. studied the same discrimination problem (Qiu, Cheng, Ching, Jiang, & Akutsu, 2015). However, in their work, a very restricted model was considered: periodic attractors were not considered and discrimination nodes had to be selected from consecutive nodes. The latter restriction is too strong and is not appropriate from a biological viewpoint. Furthermore, such a restricted problem can be trivially solved in polynomial time because it is enough to examine all possible $O(n^2)$ intervals (although
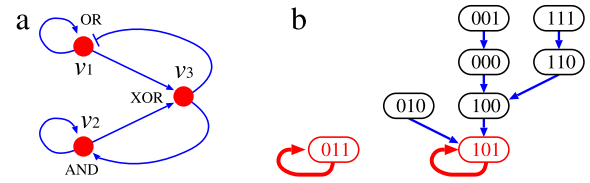
some improvement was done in their work). In this paper, we do not adopt such a non-realistic assumption. Cheng et al. also studied the same discrimination problem (Cheng, Qiu, Hou, & Ching, 2017). They developed an integer programming-based method for discrimination of singleton attractors. Although they also assumed that discrimination nodes had to be selected from consecutive nodes, this restriction can be easily removed and then the method might be practically useful. However, they did not perform any theoretical analysis and their method cannot handle periodic attractors. The main contributions of our work are discovery of a novel and useful property on periodic attractors, development of combinatorial algorithms for singleton and periodic attractors, and theoretical analyses of the time complexities of these algorithms. None of these was studied in the above-mentioned works.

2. Discrimination of singleton attractors

Before defining the problem, we briefly review *Boolean networks* (BNs). A BN $N(V, F)$ consists of a set of n nodes $V = \{v_1, \ldots, v_n\}$ and a corresponding set of Boolean functions $F = \{f^{(i)} | v_i \in V\}$. Let $v_i(t) \in \{0, 1\}$ represent the value of a node v_i at time t, and denote by $\mathbf{v}(t) = (v_1(t), \ldots, v_n(t))$ the *state of the network* at time t. The values of all nodes are updated simultaneously according to the corresponding Boolean functions, $v_i(t + 1) = f^{(i)}(\mathbf{v}(t))$, where $f^{(i)}$ may depend only on a few nodes in V. A directed graph can be associated with the network in which there exists a directed edge $(v_j, v_i) \in E$ if and only if $f^{(i)}$ depends on v_j. Dynamics of a BN is well represented by a *state transition diagram* in which nodes correspond to network states and there exists a directed edge from $\mathbf{u}$ to $\mathbf{v}$ if and only if network state $\mathbf{u}$ at time t transits to network state $\mathbf{v}$ at time $t + 1$. A sequence of states $\mathbf{v}(0), \mathbf{v}(1), \ldots,$ is called an *attractor with period p* if $\mathbf{v}(0) = \mathbf{v}(p)$ and $\mathbf{v}(i) \neq \mathbf{v}(j)$ for all i, j with $0 \leq i \neq j < p$. An attractor with period $p = 1$ is called a *singleton attractor*. An attractor with $p > 1$ is called a *periodic attractor* and is represented as $[\mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(p-1)]$. Note that if $[\mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(p-1)]$ is a periodic attractor, $\mathbf{v}(i) = \mathbf{v}(i+kp)$ holds for all i, k with $i \geq 0$ and $k > 0$. Therefore, $[\mathbf{v}(i), \mathbf{v}(i+1), \ldots, \mathbf{v}(i+p-1)]$ and $[\mathbf{v}(j), \mathbf{v}(j+1), \ldots, \mathbf{v}(j+p-1)]$ represent the same attractor for any i, j with $i \neq j \geq 0$. We say that two periodic attractors $[\mathbf{v}(0), \mathbf{v}(1), \ldots, \mathbf{v}(p-1)]$ and $[\mathbf{u}(0), \mathbf{u}(1), \ldots, \mathbf{u}(p-1)]$ are *identical* if there exists an integer k with $0 \leq k < p$ such that $\mathbf{u}((i+k) \bmod p) = \mathbf{v}(i)$ holds for all $i = 0, 1, \ldots, p-1$. Fig. 1 shows an example of a BN. In this BN, transition rules are given by

$$v_1(t + 1) = v_1(t) \vee \overline{v_3(t)},$$
$$v_2(t + 1) = v_2(t) \wedge v_3(t),$$
$$v_3(t + 1) = v_1(t) \oplus v_2(t),$$

where $x \wedge y$, $x \vee y$, $x \oplus y$, and $\bar{x}$ denote logical AND, OR, XOR, and NOT, respectively. There exist two singleton attractors, $(0, 1, 1)$ and $(1, 0, 1)$, and no periodic attractors.

Here we introduce the problem of determining the minimum discriminators for singleton attractors. Let B be an $m \times n$ binary matrix, where each row corresponds to a singleton attractor and

Fig. 1. Example of (a) BN and (b) its state transition diagram. T-type arrow in (a) means that the input is negated. In this BN, there exist two singleton attractors $(0, 1, 1)$ and $(1, 0, 1)$ which are shown in bold curves in (b).