

# Online Observability of Boolean Control Networks\*

**Abstract**—Four types of observability of Boolean control networks (*BCNs*) have been proposed to solve different problems. However, all of them are offline observabilities that we can't adjust the input sequence by observing the output sequence in the process of determining the initial state of *BCNs*. In this paper, we firstly propose the online observability that it can determine the initial state of *BCNs* dynamically without presupposing its initial state. The online observability accomplishes this task by deciding input sequence and observing output sequence in every time step. The node values of *BCNs* update at discrete time, then we use the time step to represent the discrete time. Compared with offline observabilities, the online observability can determine the initial state of some biological systems which can be checked at most once.

**Index Terms**—Boolean control networks, online observability, directed graph, supertree

## I. INTRODUCTION

In 1960s, Nobel Prize winners Jacob and Monod found that “Any cell contains a number of ‘regulatory’ genes that act as switches and can turn one another on and off. If genes can turn one another on and off, then you can have genetic circuits.” [1], [2]. Inspired by these Boolean-type actions in genetic circuits, the Boolean networks (*BNs*) is firstly proposed by Kauffman [3] for modeling nonlinear and complex biological systems. Some general descriptions of the *BNs* and its applications to biological systems can be found in Kauffman. Since then research interests in *BNs* have been motivated by the large number of natural and artificial systems. These natural and artificial systems describing variables display only two distinct configurations, and hence these describing variables take only two values, i.e.,  $\{0, 1\}$  [4], [5], [6], [7], [8], [9].

When external regulation or perturbation is considered, *BNs* are naturally extended to Boolean control networks (*BCNs*) [10]. The study on control-theoretic problems of *BCNs* can date back to 2007 [11]. The work above also proves that the problem of determining the controllability of *BCNs* is *NP*-hard in the number of nodes. Furthermore, it points out that “One of the major goals of systems biology is to develop a control theory for complex biological systems.” Since then, the study on control-theoretic problems in the areas of *BNs* and *BCNs* has drawn great attention [2], [12], [13], [14], [9]. Besides controllability, observability is another basic control-theoretic problems and it also attract many attentions. Among the studies, semi-tensor product (STP) of matrices is one of useful tool to deal with both *BNs* and *BCNs* related problems [2]. Moreover, [2] gives equivalent conditions for controllability of *BCNs* and observability of controllable *BCNs*. To date, there are four types of observability have been proposed.

- 1) The first type of observability proposed in 2009 [2] means that every initial state can be determined by an input sequence.
- 2) The second observability proposed in 2010 [12] stands that for every two distinct initial states, there exists an input sequence which can distinguish them, and this observability is determined in [15].
- 3) The third observability proposed in 2011 [13] states that there is an input sequence that determines the initial state.
- 4) The fourth observability proposed in 2013 [9] is essentially the observability of linear control systems, i.e., every sufficient long input sequence can determine the initial state.

In above mentioned definitions an input is not the value of an input-node, but it represents the values of all input-nodes of the *BCN* on a time step. Therefore, an input is a vector of the values of all input-nodes of the *BCN* on a time step. An input sequence consists of several inputs in sequential time steps. A output is a vector of the values of all output-nodes of the *BCN* on a time step. In every step, there is a pair of input and output. A output sequence also consists of several outputs in sequential time steps. In the following, we will list the inform definition of four offline observabilities and they formal definition of four observabilities in the following pages.

The four types of observability provide many nice properties and they can be used into some useful applications. However, all of four types of observability of *BCNs* are offline observabilities which means that they can't adjust the input sequence by observing the output sequence in the process of determining the initial state of *BCNs*. Therefore, we propose the online observability that we can determine the initial state of *BCNs* dynamically. In other words, the online observability decides the input sequence in each time step by observes the out sequence. In the online observability, we infer the possible initial states set by observe the first  $k$  time steps output of *BCN*. Through the possible initial states, we can choose one of input to refine the possible initial states set in  $k + 1$  time step. Repeat above procedure until the cardinality of initial states set is one. We call this process is a dynamic process.

**Contribution:** Firstly, we give the formal definition of online observability of *BCNs*. Secondly, we provide two algorithms to determine the online observability for *BCNs*. Finally, we introduce some applications of the online observability of *BCNs* and advantages of online observability comparing with offline observabilities. No important/key points



The *STP* of matrices can be used to represent the algebraic forms of *BCNs* [2], the definition of *STP* is as follows.

**Definition 2 (STP):** [14] Let  $X \in \mathbb{R}_{m \times n}$ ,  $Y \in \mathbb{R}_{p \times q}$  and  $\alpha = \text{lcm}(n, p)$  be the least common multiple of  $n$  and  $p$ . The *STP* of  $X$  and  $Y$  is defined as

$$X \ltimes Y = (X \otimes I_{\alpha/n})(Y \otimes I_{\alpha/p}),$$

where  $\otimes$  denotes the Kronecker product.

After introduce the definition of *STP* of matrices, we introduce some related notations at first [16]:

- $\delta_n^i$ : the  $i$ -th column of the identity matrix  $I_n$ ;
- $\Delta_n$ : the set  $\{\delta_n^1, \dots, \delta_n^n\}$ ;
- $\delta_n[i_1, \dots, i_s]$ :  $[\delta_n^{i_1}, \dots, \delta_n^{i_s}]$  ( $i_1, \dots, i_s \in \{1, 2, \dots, n\}$ ) the logical matrix;
- $L_{n \times s}$ : the set of  $n \times s$  logical matrices.

Using *STP* of matrices, the formula (1) can be equivalently represented in the following algebraic form:

$$\begin{aligned} s(t+1) &= L \ltimes i(t) \ltimes s(t) \\ o(t) &= H \ltimes s(t) \end{aligned} \quad (2)$$

where  $s(t) \in \Delta_N$ ,  $i(t) \in \Delta_M$ , and  $o(t) \in \Delta_Q$  denote the states, inputs and outputs respectively the same as in (1), but they are with vector forms;  $L \in L_{N \times (NM)}$ ;  $H \in L_{Q \times N}$  denote the relation matrices; that  $N = 2^n$ ,  $M = 2^m$ , and  $Q = 2^q$ . Since *STP* keeps most properties of the conventional product [14], the associative law, the distributive law, etc., we usually omit the symbol “ $\ltimes$ ” hereinafter. For instance, the formula “ $s(t+1) = L \ltimes i(t) \ltimes s(t)$ ” can be written as “ $s(t+1) = Li(t)s(t)$ ”.

To construct algebraic form (2) we give a mapping  $\tau : \{0, 1\} \mapsto \{\delta_2^1, \delta_2^2\}$  where  $\tau(0) = \delta_2^2$ ,  $\tau(1) = \delta_2^1$ . Therefore, the logical variable  $A(t)$  takes value from these two vectors, i.e.,  $A(t) \in \{\delta_2^1, \delta_2^2\}$ . Using the *STP* of matrices, we have

$$\begin{aligned} i(t) &= i_1(t) \dots i_m(t); \\ s(t) &= s_1(t) \dots s_n(t); \\ o(t) &= o_1(t) \dots o_q(t). \end{aligned}$$

And according to [17], each logical function  $f_p$  of state-nodes can be found in the updating rules (1). The form of  $f_p$  as:

$$f_p(i_1(t), \dots, i_m(t), s_1(t), \dots, s_n(t))$$

and there exists a structure matrix  $L_p \in L_{2 \times NM}$  such that

$$\tau(f_p(i_1(t), \dots, i_m(t), s_1(t), \dots, s_n(t))) = L_p i(t) s(t) \quad (3)$$

For state-nodes  $s_1, \dots, s_n$ , we have  $n$  logical matrices  $L_1, \dots, L_n$  for them, respectively. If for each state-node  $s_p$  the logical matrix has form

$$L_p = [\delta_2^{p1}, \dots, \delta_2^{pNM}],$$

then we have that

$$L = [\delta_N^{R1}, \dots, \delta_N^{RNM}]$$

where

$$\delta_N^{R1} = \delta_2^{11} \dots \delta_2^{n1}; \dots; \delta_N^{RNM} = \delta_2^{1NM} \dots \delta_2^{nNM}.$$

By this relationship we can construct the  $L$  for the algebraic forms of *BCNs*. What's more we can also construct the logical matrix  $H$  in the similar way.

For instance, the *BCN* whose structure is depicted in Fig.1, and the updating rules of this *BCN* is described as truth table in Fig.2. We have that the updating rules of this *BCN* can be represented with the algebraic form:

$$\begin{aligned} s(t+1) &= \delta_{16}[\alpha] i(t) s(t) \\ o(t) &= \delta_4[\beta] s(t) \end{aligned} \quad (4)$$

where  $\alpha = \{10, 4, 11, 16, 9, 5, 1, 7, 15, 2, 3, 12, 7, 6, 8, 13, 8, 9, 15, 10, 14, 4, 3, 16, 1, 14, 12, 13, 5, 7, 2, 6, 7, 2, 3, 13, 13, 9, 5, 1, 16, 13, 6, 14, 11, 10, 4, 15, 1, 14, 7, 6, 9, 8, 11, 12, 5, 5, 13, 3, 10, 12, 16, 16\}$ ,  $\beta = \{1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4\}$ ,  $t \in \mathbb{N}$ ,  $s \in \Delta_{16}$ ,  $i \in \Delta_4$  and  $o \in \Delta_4$ .

### C. Four Existing Observability of BCNs

In this section we introduce four existing kinds of observability of *BCNs*. Let  $\Delta_N$ ,  $\Delta_M$ ,  $\Delta_Q$  be three alphabets, for all  $s_0 \in \Delta_N$  and all  $p \in \mathbb{Z}_+$ ;  $\infty$  is the infinite natural numbers. In order to introduce four existing kinds of observability of *BCNs*, we define the mappings [16]:

$$\begin{aligned} L_{s_0}^p : (\Delta_M)^p &\mapsto (\Delta_N)^p, u_0 \dots u_{p-1} \mapsto s_1 \dots s_p \\ L_{s_0}^\infty : (\Delta_M)^\infty &\mapsto (\Delta_N)^\infty, u_0 u_1 \dots \mapsto s_1 s_2 \dots \end{aligned} \quad (5)$$

$$\begin{aligned} (HL)_{s_0}^p : (\Delta_M)^p &\mapsto (\Delta_Q)^p, u_0 \dots u_{p-1} \mapsto o_1 \dots o_p \\ (HL)_{s_0}^\infty : (\Delta_M)^\infty &\mapsto (\Delta_Q)^\infty, u_0 u_1 \dots \mapsto o_1 o_2 \dots \end{aligned} \quad (6)$$

For all  $p \in \mathbb{Z}_+$ , all  $U = u_1 \dots u_p \in (\Delta_M)^p$ , and all  $1 \geq i \geq j \geq |U|$ , we use  $U[i, j]$  to denote the word  $u_i \dots u_j$  as a input sequence. Then four existing kinds of observability of *BCNs* can be define as:

**Definition 3:** The first kind of observability is that, *BCN* is called observable, if for every initial state  $s_0 \in \Delta_N$ , there exists an input sequence  $U \in (\Delta_M)^p$  for some  $p \in \mathbb{Z}_+$  such that for all states  $s_0 \neq s'_0 \in \Delta_N$ ,  $Hs_0 = Hs'_0$  implies  $(HL)_{s_0}^p(U) \neq (HL)_{s'_0}^p(U)$  [2].

Hence the first observability means that if a *BCN* is observable then every initial state of the *BCN* can be determined by an input sequence. But we can only use the corresponding input sequence of a state to check whether this state is the initial state of the *BCN* or not.

**Definition 4:** The second kind of observability is that a *BCN* is called observable if for any distinct states  $s_0, s'_0 \in \Delta_N$ , there exists an input sequence  $U \in (\Delta_M)^p$  for some  $p \in \mathbb{Z}_+$ , such that  $Hs_0 = Hs'_0$  implies  $(HL)_{s_0}^p(U) \neq (HL)_{s'_0}^p(U)$  [12].

The second observability means that a *BCN* is called observable if for every two distinct initial states of the *BCN*, there exists an input sequence which can distinguish them.

**Definition 5:** The third kind of observability is that, a *BCN* is called observable, if there exists an input sequence  $U \in (\Delta_M)^p$  for some  $p \in \mathbb{Z}_+$ , such that for any distinct states  $s_0, s'_0 \in \Delta_N$ ,  $Hs_0 = Hs'_0$  implies  $(HL)_{s_0}^p(U) \neq (HL)_{s'_0}^p(U)$  [13].

The third observability means that a *BCN* is called observable if there is an input sequence that determines the initial state of the *BCN*.

*Definition 6:* The fourth kind of observability is that, *BCN* is called observable, if for any distinct states  $s_0, s'_0 \in \Delta_N$ , for any input sequence  $U \in (\Delta_M)^\infty$ ,  $HS_0 = HS'_0$  implies  $(HL)_{s_0}^\infty(U) \neq (HL)_{s'_0}^\infty(U)$  [9].

The fourth observability means that a *BCN* is called observable if every sufficient long input sequence can determine the initial state of the *BCN*.

Then from the definitions of four existing kinds of observability, we know that [16]:

- the first one implies the second one;
- the third one implies the second one and first one;
- the fourth one implies the third one, second one and first one.

And we can't use the first one and second one to determine the initial state of *BCNs* which can be checked at most once. For example, in the first kind observability we need to assume the initial state of a *BCN*, and then check it by corresponding input sequence of this state. If the state we assume is correct, then we can determine the initial state. But if the assumption is not correct, we can't determine the initial state of the *BCN*. Therefore we need to check the *BCN* again and again until we can determine the initial state of it. However we can use the third existing observability and fourth existing observability to determine the initial state of *BCNs* in one time. And we need not to presuppose the initial state of *BCNs* when we use the third observability and fourth observability. But the requirements for *BCNs* are very harsh when we use the third observability and fourth observability. With these disadvantages of four existing observabilities, we propose the online observability of *BCNs* to solve this problem.

*Problem:* Finding the necessary and sufficient condition of determine the initial state of *BCNs* in one time and without presupposing the range of its initial state.

### III. THE ONLINE OBSERVABILITY OF *BCNs*

In this paper we propose the online observability, the informal definition of it is as follows.

*Definition 7:* A *BCN* is called online observable, if every initial state  $s_0 \in \Delta_N$  can be determined in one time by dynamically deciding input sequence and observing output sequence at every step without presupposing the initial state of *BCN*. And this process can be accomplished in finite steps.

In this section, firstly we present the definition of deduce function, secondly we present the definition of  $K$  steps deterministic. We take them as the preparations for defining online observability. Finally, we give the formal definition of online observability of *BCNs*.

#### A. Deduce Function

Different from four existing types, the observability we propose can determine the initial state online. Because in the process of determining the initial state every input of the input sequence is decided by the output we observe at

every time step. At the beginning, we can observe the output of *BCNs*, so that we can infer the possible values of state-nodes and treat them as possible states set  $S_0$ . Then as we can know the possible states set, we need to decide the input  $i_0$ . The input  $i_0$  will make sure any different possible states  $s_i, s_j \in S_0$  will not turn into the same state after affected by input  $LS_i i_0 \neq LS_j i_0$ . After decided input, we can observe the new output, and then we can infer the new possible states set. The cardinal number of possible states set after we input will not larger than the cardinal number of possible states set before we input. If the cardinal number of possible states set turn into be 1 then we can determine the state and the initial state of *BCN*. To simulate this deduction process, we give the definition of deduce function that.

*Definition 8 (Deduce Function):* The deduce function can be defined as  $D(S, I, O)$ . Based on deduction process, we have for any

$$s_i(t+1) \in D(S, I, O)$$

there exists the corresponding  $s_i(t)$  of  $s_i(t+1)$  in  $S$  that

$$s_i(t+1) = LIS_i(t)$$

and

$$O = HS_i(t+1).$$

where

- $S \in 2^{\Delta_N}$  is the possible states set;
- $I \in \Delta_M$  represents the input;
- $O \in \Delta_Q$  represents the output;
- $D(S, I, O) \in 2^{\Delta_N}$  is the possible states set after deduction.

From the definition of deduce, we have some equations for this function that

$$D(\emptyset, I_i, O_i) = \emptyset \quad (7)$$

Equation (7) represents that if the possible states set is an empty set  $\emptyset$ , no matter what we do we can only deduce the possible set is  $\emptyset$  too.

$$D(S_i, \varepsilon, \varepsilon) = S_i \quad (8)$$

If the possible states set is  $S_i$  and we neither input anything and nor observe the output. In this case we can only deduce that the possible states set is  $S_i$  shown in equation (8).

$$D(\Delta_N, \varepsilon, \delta_4^1) = \{\delta_{16}^1, \delta_{16}^2, \delta_{16}^3\} \quad (9)$$

Using the example mentioned before, when the possible states set  $S_i = \Delta_N$ , and we observe that the outputs of *BCN* is  $\delta_4^1$  before we decide input. In this case we can deduce that the possible states would be  $\delta_{16}^1, \delta_{16}^2$  or  $\delta_{16}^3$  shown in equation (9).

$$D(\{\delta_{16}^1, \delta_{16}^2, \delta_{16}^3\}, \delta_4^1, \varepsilon) = \{\delta_{16}^{10}, \delta_{16}^4, \delta_{16}^{11}\} \quad (10)$$

If the possible states set  $S_i = \{\delta_{16}^1, \delta_{16}^2, \delta_{16}^3\}$  we input  $\delta_4^1$ . Before we observe the output of *BCN* we can only deduce the

possible states would be  $\delta_{16}^{10}$ ,  $\delta_{16}^4$  or  $\delta_{16}^{11}$  shown in equation (10).

$$D(\{\delta_{16}^1, \delta_{16}^2, \delta_{16}^3\}, \delta_4^1, \delta_4^3) = \{\delta_{16}^{10}, \delta_{16}^{11}\} \quad (11)$$

But if we observe that the output of *BCN* is  $\delta_4^3$ , then we can deduce that the possible state can be  $\delta_{16}^{10}$  or  $\delta_{16}^{11}$  shown in equation (11);

$$D(\{\delta_{16}^4, \delta_{16}^5, \delta_{16}^6\}, \delta_4^3, \varepsilon) = \{\delta_{16}^9, \delta_{16}^{13}\} \quad (12)$$

Finally if the set of states is  $\{\delta_{16}^4, \delta_{16}^5, \delta_{16}^6\}$  and the inputs is  $\delta_4^3$ . Before we observe the output of *BCN* we can deduce that the possible state values can be  $\delta_{16}^9$  or  $\delta_{16}^{13}$  shown in equation (12), as the cardinality number of the possible states set decreased, we can't deduce the initial state any more.

### B. *K Steps Deterministic*

After we defined the deduce function, we can present the definition of *K* steps deterministic of the states set of *BCNs* and the range of *K* is the set of natural numbers. It may easier to define online observability by programming language. But we would like to define its mathematical form for preciseness of concepts. Therefore, before defining the online observability of *BCNs*, we need to define the *K* steps deterministic of the states set of *BCNs* at first.

*Definition 9 (K Steps Deterministic):* When  $K = 0$ , if for a set of states  $S'$  and  $|S'| = 1$ , then  $S'$  is *K* step deterministic. When  $K > 0$ , if for a set of states  $S'$  ( $|S'| > 1$ ), there exists  $I'$  in  $\Delta_M$  implies

$$|D(S', I', \varepsilon)| = |S'|,$$

and implies “For every  $O'$  in  $\Delta_Q$ ,

$$|D(S', I', O')| \neq 0$$

implies  $D(S', I', O')$  is  $K'$  ( $K' < K$ ) steps deterministic.’, then  $S'$  is *K* steps deterministic.

From the definition of *K* steps deterministic we know  $K = 0$  means that we can determine the state without any input and observing output. Because if we know the cardinality number of possible states set is 1, then we can know the state of *BCNs*. We can only discuss the case of  $K = 0$  when  $|S'| = 1$ . If  $K > 0$ , then the definition of *K* steps deterministic is defined recursively, and it need to use the definition of *K* ( $K = 0$ ) steps. When we talk that a states set of *BCNs* is *k* steps deterministic we default  $k \geq 0$ .

Furthermore, if  $S'$  is  $k_1$  steps deterministic and  $k_1 \leq k_2$ , then  $S'$  is  $k_2$  steps deterministic. But if  $S'$  is  $k_1$  steps deterministic and  $k_1 \geq k_2$ , we can not make sure whether  $S'$  is  $k_2$  steps deterministic or not. Therefore you can consider the “ $S'$  is  $k_i$  steps deterministic” as “We can determine the state of a *BCN* with possible states set  $S'$  in  $k_i$  steps. And we finish this process by deciding input sequence and observing out sequence at each time step”.

### C. Online Observability

After the previous preparation, we present the formal definition of the online observability. The formal definition of the online observability of *BCNs* is as follows.

*Definition 10 (Online Observability of BCNs):* If for every  $O'$  in  $\Delta_Q$  and  $|D(\Delta_N, \varepsilon, O')| \neq 0$ , there exists a  $k \geq 0$  implies  $D(\Delta_N, \varepsilon, O')$  is *k* steps deterministic, then this *BCN* is online observable. We even can define it simpler, if there exists  $k \geq 0$  implies  $\Delta_N$  is *k* steps deterministic, then this *BCN* is online observable.

The difference between the second definition and the first definition is that whether we observe the corresponding output of the initial state of *BCN* at first. For better performance, we use the first definition of online observability.

After defining online observability of *BCNs*, we discuss the comparison of online observability with the four existing observability. In the second existing kind of observability, we presuppose the initial state of *BCNs*, and then try to find the input sequence to distinguish it from other kinds of initial states. But the input sequence determined by the presupposed initial state may make other kinds of initial states turn into be the same state, so that other kinds of initial states can't be determined anymore. This problem has to be considered in the online observability of *BCNs*. Hence the online observability implies the first existing kind of observability, and then the online observability implies the second existing kind of observability. In the third existing kind of observability, there has to exist an input sequence that can distinguish any distinct states. However in online observability we can use different input sequences to distinguish any distinct states in different states sets. These different states sets are classified by their corresponding output. Therefore, we have the third existing kind of observability implies the online observability of *BCNs*, then the fourth existing kind of observability implies the online observability.

When I learn the existing four kinds of observability of *BCNs*, I find that if we want determine the initial state of a *BCNs* by first kind of observability, we need to guess the initial state of the *BCN* and then check it by its corresponding input sequence, if the initial state we guess is right, we can determine it, but if not, we need to guess again and input the corresponding input sequence untill we determine the initial state of the *BCN*. But if we can't repeat this process, we may can't determine the initial state of the *BCN* any more. Then I turn my gaze to the third observability, this kind of observability makes we can determine the initial state without presupposing the initial state. But I think if we can determine the possible states set of the *BCN* by observing the output at first, why can't we try to find corresponding input sequence for them? And then my teacher and I talk about this thinkness and expand it into the original idea of the online observability of *BCNs*.

From the informal definition and formal definition of online observability, we can know that the necessary and sufficient condition of determine the initial state of *BCNs* without presuppose the initial state is the online observability

of *BCNs*. By this definition we can build *BCNs* with least output-nodes when we want to determine the initial state of *BCNs*.

#### IV. DETERMINING THE ONLINE OBSERVABILITY OF ETERMINING THE ONLINE OBSERVABILITY OF BCNs

In this paper, we proposed two ways to determine the online observability of *BCNs*. The first way is by using supertree and the second way is by using directed graph. The construction process of supertree simulates deduction process mentioned before. We check the super tree based on the definition of online observability of *BCNs* depth first or breadth first. But when we used the super tree to determine the online observability of *BCNs*, we need to check the existence of loops. And many nodes in the tree are repeated, these nodes will take a lot of time overhead and space overhead. Therefore, we proposed the second way to determine the online observability by using directed graph. By this way we can avoid checking the existence of loop and avoid checking repeated nodes. There are also other advantages like determining observability earlier and select the input smarter when we use the second way. All of these advantages will reduce time and space overhead to determine the online observability.

##### A. Supertree

As we mentioned before, we can use the deduce function to determine the initial state of *BCNs*. According to the definition of online observability we will alternately observe the output and decide the input. When the cardinal number of the states set comes into be 1 we can determine the initial state, and stop deducing the initial state of *BCNs*. According to this process, we can define the supertree for *BCNs*. For convenience, we use the states set inside the node to represent the node, and output in the edge to represent the edge.

**Definition 11 (Super Tree):** The root node of the super tree is  $\Delta_N$ , the leaf nodes of the super tree are the nodes with cardinal number 1 ( $|S'| = 1$ ). In addition to the leaf nodes, if a node  $S'$  in the  $2k+1$  ( $k \in \mathbb{N}$ ) layer of the supertree and

$$|D(S', \varepsilon, O')| > 0,$$

then  $D(S', \varepsilon, O')$  is one of its son nodes, and  $O_i$  is the edge from  $S_i$  to  $D(S, \varepsilon, O_i)$ . If a node  $S'$  in the  $2k+2$  layer of the supertree and

$$|D(S', I', \varepsilon)| = |S'|,$$

then  $D(S', I', \varepsilon)$  is the son node of  $S'$  and  $I'$  is the edge from  $S'$  to  $D(S', I', \varepsilon)$ .

At the beginning we infer that the possible states set is  $\Delta_N$ , thus the root node of the super tree is  $\Delta_N$ . When the cardinal number of the possible states set turns into 1, we can determine the state of *BCN*. Therefore, the leaf nodes of the super tree are the nodes with cardinal number 1. We observe the output of *BCN* to infer the possible states set of *BCN* at first. After that, we decide the input and infer the new possible states set. We alternately observe the output

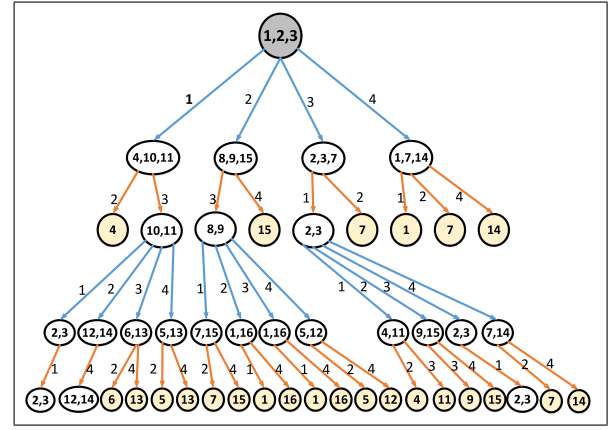


Fig. 3. Branch of the super tree which represents  $\{\delta_{16}^1, \delta_{16}^2, \delta_{16}^3\}$ . The blue edges and orange edges show the observing output processes and deciding input processes, respectively. The yellow nodes are leaf nodes.

and decide the input until we can determine the state of *BCN*. Therefore we use  $D(S', \varepsilon, O')$  in  $2k+1$  layer to find son nodes for  $S'$ , and use  $D(S', I', \varepsilon)$  in  $2k+2$  layer to find son nodes for  $S'$ . The formula  $|D(S', \varepsilon, O')| > 0$  ensures the node  $D(S', \varepsilon, O')$  has meaning. The formula  $|D(S', I', \varepsilon)| = |S'|$  guarantee we can determine the state of *BCN* in the end.

For example, the Fig.3 show branch of the tree which represents  $\{\delta_{16}^1, \delta_{16}^2, \delta_{16}^3\}$ . The nodes represent the states sets, the blue edges represent the observing output processes, and the orange edges represent the deciding input processes. This branch is not completed, because only the yellow nodes are the leaf nodes. If we want to find all of the ways to determine the initial state of *BCN*, we have to build the complete tree for *BCN*. This process takes many additional time and space overhead. Especially when there are loops in the tree, like the  $\{\delta_{16}^2, \delta_{16}^3\}$  in fourth layer and the  $\{\delta_{16}^2, \delta_{16}^3\}$  in fifth layer that will form a loop. In this case we can never build the complete tree, thus we need to check the existence of loops and omit it. There are also some nodes take the same states set which will also take additional overhead. For instance there are two nodes take the same states set  $\{\delta_{16}^1, \delta_{16}^{16}\}$  in the fifth layer. However, it would be a lot easier if we only need to find a way to determine the initial state. For instance, when we find the leaf nodes  $\delta_{16}^1$ ,  $\delta_{16}^7$  and  $\delta_{16}^{14}$  in third layer by breadth-first algorithm, we can make sure that the states set  $\{\delta_{16}^1, \delta_{16}^2, \delta_{16}^3\}$  is 1 step deterministic. After that, we use this conclusion to determine the initial state of *BCN*.

##### B. Directed Graph

To improve the shortcomings of the way by using supertree, we proposed the way by directed graph which takes less time and space overhead. The most difference between supertree and directed graph is that supertree is built from the root node to leaf nodes. However, the directed graph is built from smaller nodes (contain less states) to larger nodes (contain more states). There is not any repeated node in the directed graph because any node only appears once in the graph. And even there are some loops in the directed graph,



the loops would not influence us to build the directed graph.

The construction algorithm of derected graph is shown in the Algorithm.1. The algorithm to build nodes used in the Algorithm.1 is shown in the Algorithm.2.

---

**Algorithm 1** Algorithm to construct the directed graph of BCNs

---

**Input:** The algebraic forms of BCN

**Output:** The directed graph of BCN

```

1: Define a int variable  $k = 1$ , to represent the number of
   states in the nodes.
2: Define a boolean type variable  $Ob = 0$ , to represent the
   online observability of BCN.
3: buildnode(k)
4:  $k = k + 1$ 
5: while buildnode(k)==1 do
6:   if  $k == 2$  then
7:     We make  $\Delta_M$  as the suitable inputs set for every
     node with  $k$  states.
8:   else
9:     if  $k > 2$  then
10:      We use two nodes with  $(k - 1)$  states and a node
      with two states to find the suitable inputs set for
      every node with  $k$  states.
11:    end if
12:  end if
13:  for every node with  $k$  states do
14:    for every input in the suitable input set of the nodes
    do
15:      Check whether this input can make the nodes  $z$ 
      ( $z \geq 0$ ) steps deterministic, and build edges for
      this node.
16:    end for
17:  end for
18:  if there exist one node has not any edge connect it
  with other nodes then
19:     $Ob = 0$  return Ob
20:  end if
21:   $k = k + 1$ 
22: end while
23:  $Ob = 1$  return Ob

```

---

Some details in Algorithm.1 and Algorithm.2 are as follows:

- Sort the states in the nodes, and then sort the nodes: For example, the nodes  $\{\delta_{16}^1, \delta_{16}^2\}$ ,  $\{\delta_{16}^1, \delta_{16}^3\}$  and  $\{\delta_{16}^2, \delta_{16}^3\}$  shown in Fig.4.
- Using other nodes to find the suitable inputs set for nodes with  $k$  states: For example, we can search inputs sets which make  $\{\delta_{16}^4, \delta_{16}^5, \delta_{16}^6\}$ ,  $\{\delta_{16}^5, \delta_{16}^6, \delta_{16}^7\}$  and  $\{\delta_{16}^4, \delta_{16}^7\}$   $k$  steps deterministic at first. After that, take the intersection of these sets to be the suitable inputs set of  $\{\delta_{16}^4, \delta_{16}^5, \delta_{16}^6, \delta_{16}^7\}$ .
- Check whether an input ( $I'$ ) can make a node ( $S'$ )  $z$  steps deterministic: According to the order determined in previous steps, we check every node in order. If for one input  $I'$  which belongs to suitable inputs set of the

---

**Algorithm 2** Algorithm to build nodes with  $k$  states

---

**Input:** The number of states in the nodes  $p$

**Output:** The nodes with  $p$  states

```

1: int buildnode(int p)
2: {
3:   Try to build all the nodes with  $p$  states whose corre-
   sponding outputs are the same, and classify them by their
   corresponding outputs.
4:   if Failed to build then
5:     return 0;
6:   else
7:     Sort the states inside the nodes from small to large,
     and then sort the nodes based on the values inside the
     nodes.
8:     return 1;
9:   end if
10: }

```

---

states set  $S'$  implies  $|D(S', I', \varepsilon)| < |S'|$ , we can make sure the  $I'$  is a wrong input. Else if for all  $O' \in \Delta_Q$  and  $|D(S', I', O')| > 0$ , the  $D(S', I', O')$  is  $z$  steps deterministic then  $I'$  is a right input. Therefore, we can connect the node  $S'$  to all nodes  $D(S', I', O')$  with directed edges. The colour of directed edges represent its corresponding input. Else if there exist  $O' \in \Delta_Q$  and we can not make sure whether  $D(S', I', O')$  is  $z$  steps deterministic, we check it in the next round.

According the construction process, we have the definition of directed graph.

*Definition 12 (Directed Graph):* For every node  $S_j$  in the directed graph, the  $S_j$  is  $k$  steps deterministic. For every distinct two  $s_a, s_b \in S_j$  we have  $Hs_a = Hs_b$ . If  $|S_i| = 1$  there are not edge from it to other nodes, else if there are exist one edge  $I'$  from it to one nodes then there exist  $p$  ( $p \geq 1$ ) edges contain  $I'$  from it to nodes  $S_1, \dots, S_p$  that

$$|S_j| = |S_1| + \dots + |S_p|$$

and

$$D(S_j, I', \varepsilon) = S_1 \vee \dots \vee S_p.$$

When we trying to build the directed graph for a BCN, we check whether the nodes with less states are  $k$  steps deterministic first and then check whether the nodes with more states are  $k$  steps deterministic. As the the nodes with less states are not  $k$  steps deterministic, the nodes with more states would not be  $k$  steps deterministic. Therefore, once we can find a node is not  $k$  steps deterministic we can know that  $\Delta_N$  is not  $k$  steps deterministic and this BCN is not online observable. Moreover, we can use the nodes with less states that are  $k$  steps deterministic to help us check the nodes with more states. For instance, if the node  $S$  has two edges from it to two nodes  $S_1$  and  $S_2$ , and we have  $S_1$  and  $S_2$  are  $k$  steps deterministic. In this case, we can make sure that the node  $S$  is  $k$  steps deterministic.

Based on the definitions of existing four kinds of observability, we can also use the directed graph to determine the

existing second and fourth kinds of observability. When we trying to build bottom layer and penultimate layer of the directed graph, if there are exist some nodes in penultimate layer has no edges from it to other nodes. In this case this *BCN* is not satisfied existing second observability. When we trying to build edges for every layer, and if there exist one node whose right inputs set is not  $\Delta_M$ , then this *BCN* is not satisfied existing fourth observability.

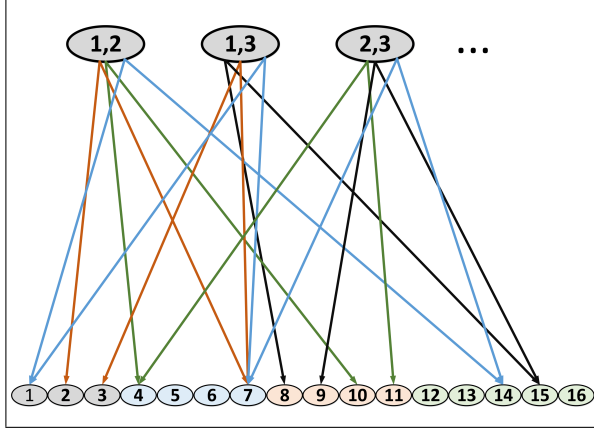


Fig. 4. Part of the directed graph which represents  $\{\delta_{16}^1, \delta_{16}^2\}$ ,  $\{\delta_{16}^1, \delta_{16}^3\}$  and  $\{\delta_{16}^2, \delta_{16}^3\}$ . The green, black, orange, blue edges show the inputs  $\delta_4^1$ ,  $\delta_4^2$ ,  $\delta_4^3$  and  $\delta_4^4$  respectively.

### C. Complexity Analysis

The way by the directed graph is better than by supertree, thus we analyze the complexity of it. We classify the states with their corresponding output. After that form the set of states set  $\{S_1, S_2, \dots, S_M\}$ , and every element in a states set has the same corresponding output. That is to say for each  $S_j \in \{S_1, S_2, \dots, S_M\}$  then for every  $s_k \in S_j$  we have  $Hs_k = \delta_M^j$ .

Firstly, we need to calculate the upper bound of the number of the states in the directed graph nodes  $k$  we have.

$$k_{upb} = \max(|S_1|, |S_2|, \dots, |S_M|) \quad (13)$$

The  $k_{upb}$  is the maximum value of  $|S_1|, |S_2|, \dots, |S_M|$ , because the states in the directed graph nodes should have the same corresponding output.

Secondly, we need to calculate the number of nodes with  $k$  states:

$$k_{non} = C_{|S_i|}^k + \dots + C_{|S_p|}^k \quad (14)$$

where  $S_i, \dots, S_p \in \{S_1, S_2, \dots, S_M\}$  and  $|S_i|, \dots, |S_p| \geq k$ .

Thirdly we need to calculate the cardinal number of suitable inputs set of each node. Finally we need to calculate the time used to check each input is a right input for a node.

After completing the previous steps, calculate the complexity by layer by layer. But the cardinal number of suitable inputs set of a node depends on the cardinal number of it and the other three nodes used to find the suitable inputs set for it. And the time used to check whether an input is a right input for a node also depends on the updating rules of *BCNs*.

Moreover, instead of taking  $\Delta_M$  as the suitable inputs set for every node in the directed graph. We would use the other three nodes like  $\{\delta_{16}^4, \delta_{16}^5, \delta_{16}^6\}$ ,  $\{\delta_{16}^5, \delta_{16}^6, \delta_{16}^7\}$  and  $\{\delta_{16}^4, \delta_{16}^7\}$  that are  $k$  steps deterministic to find the suitable inputs set for a node  $\{\delta_{16}^4, \delta_{16}^5, \delta_{16}^6, \delta_{16}^7\}$  with more than 2 states. By this way we can reduce the cardinal number of the suitable inputs set for every nodes with more than 2 states, and then reduce the time cost.

The reason why we can use this method is that only the input which make the subset of  $\{\delta_{16}^4, \delta_{16}^5, \delta_{16}^6, \delta_{16}^7\}$   $k$  steps deterministic will make the  $\{\delta_{16}^4, \delta_{16}^5, \delta_{16}^6, \delta_{16}^7\}$   $k$  steps deterministic. Furthermore, using these three nodes will be a good way to cover all the subset with cardinal number 2 of  $\{\delta_{16}^4, \delta_{16}^5, \delta_{16}^6, \delta_{16}^7\}$ . That is to say every subset  $s_k$  with cardinal number 2 belongs to  $\{\delta_{16}^4, \delta_{16}^5, \delta_{16}^6, \delta_{16}^7\}$  will belongs to  $\{\delta_{16}^4, \delta_{16}^5, \delta_{16}^6\}$ ,  $\{\delta_{16}^5, \delta_{16}^6, \delta_{16}^7\}$  or  $\{\delta_{16}^4, \delta_{16}^7\}$ . This conclusion can help us to select the nodes we need when we seek the suitable inputs set for a node. But it is hard to analyze the complexity of this method, and it makes the complexity analysis of the way by directed graph harder.

Therefore, it is hard to give a accurate complexity of the algorithm without the complete information of *BCNs*. We may finish this job in the future, and we will try to use real example to do complexity analysis.

## V. APPLICATIONS

If the *BCNs* we research is online observable, and we have built the directed graph for them, then we can use the online observability to determine the initial states of these *BCNs*. Using the the online observability, it will need less observation costs to determine the initial state of this *BCN*. Furthermore we can also use it to try to find the shortest path or avoid entering critical states when we determine the initial state of *BCNs*.

### A. Determine the Initial State

After we build the directed graph of the *BCN*, we can use it to determine the initial state of *BCNs* as Fig.5 shows.

- Firstly, we observe the output of the *BCN* mentioned before. If we observe the output is  $\delta_4^1$  then we can infer that the possible states set should be  $\{\delta_{16}^1, \delta_{16}^2, \delta_{16}^3\}$ , and we record them as initial states and current states in the table.
- After that we input  $\delta_4^1$  and observe the output is  $\delta_4^3$  then we can infer that the possible states set should be  $\{\delta_{16}^{10}, \delta_{16}^{11}\}$ , and we record them as current states set in the corresponding position.
- Repeat the second step until the cardinal number of the possible states set turns into 1. In that time we can determine the current state and the corresponding initial state of the *BCN*.

### B. Less observation costs

Some biological systems, such as the immune systems which can be depicted as the *BCN* T-cell receptor kinetics model [18]. There exist 3 input-nodes, and 37 state-nodes in this model, therefore the model has  $2^3$  inputs and  $2^{37}$  states.



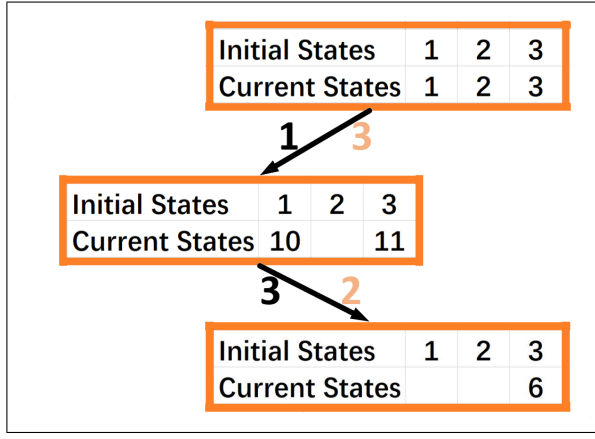


Fig. 5. The process of determining the initial state of BCNs, we change the values of current states by input and the output we observe.

For the purpose of obtain the initial state of the *BCN*, we must select some state-nodes to be observe at first.

If we use the online observability of *BCNs* to determine the initial state of the *BCN* T-cell receptor kinetics model, it needs less observation costs. Because compare with the existing third and fourth kind of observabilities the online observability need weakest preconditions to determine the initial state of *BCNs* without any presupposition. In other words, the *BCNs* needs less output-nodes to determine its initial state. As we can select less state-nodes to be observe, it will takes less ovservation costs in this example. In addition to this, there are also some other advantages, when we use the online observability of *BCNs*.

### C. Find Shortest Path

When we need to determine the initial state of a *BCN*, an important aspect that we will consider is to find the shortest path. In general, we can't find the shortest path definitely. Fortunately, we can use the directed graph to make the best decision. For the path, we introduce two functions  $\text{Pe}(S, I_z)$  and  $\text{Pv}(S, I_z)$  to describe its expected value and variance, respectively. For better explain our idea, we list some definitions in the following.

Some statements before definitions:

- $\{I_1, I_2, \dots, I_p\}$  : the right inputs set of  $S$ ;
- $\{S_z^1, S_z^2, \dots, S_z^k\}$  : the set of state sets, and its elements corresponding to the possible outputs  $\{O_1, O_2, \dots, O_k\}$  after input  $I_z$ , for every  $I_z$  in  $\{I_1, I_2, \dots, I_p\}$ .

**Definition 13** ( $\text{Spe}(S)$ ):

$$\text{Spe}(S) = \min(\text{Pe}(S, I_1), \text{Pe}(S, I_2), \dots, \text{Pe}(S, I_p)).$$

**Definition 14** ( $\text{Pe}(S, I_z)$ ): When the  $|S| = 1$ , we have that  $\text{Pe}(S, I_z) = 0$  for every  $I$  in  $\Delta_M$ . According **Definition 13**,  $\text{Spe}(S) = 0$  if  $|S| = 1$ . But when the  $|S| > 1$ , we have that

$$\text{Pe}(S, I_z) = 1 + \frac{\sum_{j=1}^k \text{Spe}(S_z^j) |S_z^j|}{|S|}$$

The function shortest path expected value  $\text{Spe}(S)$  is to find the  $I_z$  from  $\{I_1, I_2, \dots, I_p\}$  to calculat least  $\text{Pe}(S, I_z)$  for  $S$ . From the definition of  $\text{Pe}(S, I_z)$  we can know that

if  $|S| = 1$  then we can make sure the state of *BCNs*, and we need not input anymore to determine the state of *BCNs*. Therefore, for any input the path expected value  $\text{Pe}(S, I_z)$  would be 0 and so do the shortest path expected value  $\text{Spe}(S_i)$ . But if  $|S| > 1$  we still need more input and observe the output. Only by this way we can determine the state of *BCNs*. Therefore, we can recursively define the  $\text{Pe}(S, I_z)$  and  $\text{Spe}(S)$  for each input  $I_z$  in the right inputs set. The  $\text{Pv}(S, I_z)$  can be defined in the similar way. Hence we omit the details in this paper.

### D. Avoid Entering Critical States

In biological systems, some states of the genes may corresponding to unfavorable or dangerous situations [19]. So another important aspect that we will consider is to avoid entering critical states. We can also construct two functions  $\text{Ce}(S, I_z)$  and  $\text{Cv}(S, I_z)$  to describe expected value and variance of the times of entering critical states, the definition of  $\text{Ce}(S, I_z)$  is as follows:

**Definition 15** ( $\text{Lce}(S)$ ):

$$\text{Lce}(S) = \min(\text{Ce}(S, I_1), \text{Ce}(S, I_2), \dots, \text{Ce}(S, I_p))$$

**Definition 16** ( $\text{Ce}(S, I_z)$ ): When the  $|S| = 1$ , and for every  $I_z$  in  $\Delta_M$ , we have:

$$\text{Ce}(S, I_z) = |S \cap S_{cri}|$$

According **Definition 15**,

$$\text{Lce}(S) = \text{Ce}(S, I_z) = |S \cap S_{cri}|$$

But when the  $|S| > 1$  we have that

$$\text{Ce}(S, I_z) = |S \cap S_{cri}| + \frac{\sum_{j=1}^k \text{Lce}(S_z^j) |S_z^j|}{|S_z|}$$

Where  $S_{cri}$  is the critical states set of the *BCN* we research. The definition of  $\text{Ce}(S, I_z)$  has some difference with  $\text{Pe}(S, I_z)$ , because we need to use the critical states set  $S_{cri}$ . So that we can analyze the possibility of entering the critical states when we infer the possible states set of *BCNs*.

We can get the definitions of  $\text{Cv}(S, I_z)$  in similar ways, and use all of these four functions to make the best decision we like. When we choose an input  $I_z$  with least  $\text{Pe}(S, I_z)$ , we may find the shortes path to determine the initial state. But the output of *BCNs* after input  $I_i$  is uncertain, so selecting the least  $\text{Pe}(S, I_z)$  may leads to a very long path to determine the initial state of *BCNs*. For better performce, we the can also use the  $\text{Pv}(S, I_z)$  to avoiding risks. The uses of  $\text{Ce}(S, I_z)$  and  $\text{Cv}(S, I_z)$  are similar to  $\text{Pe}(S, I_z)$  and  $\text{Pv}(S, I_z)$ , when we trying to avoid entering critical states of *BCNs*.

In the four existing kinds of observability, we can not analyze the state of *BCNs* dynamically, so it would be hard to find the best way we like when we determine the initial state of *BCNs*. But this problem can be solved by the online observability of *BCNs*.

## VI. CONCLUSIONS

In this paper, firstly we proposed the online observability of *BCNs* and define its mathematical form. Secondly we use the super tree and directed graph to determine the online observability. After introduced the ways to determine the online observability we present some applications of the online observability of *BCNs* and talk about some advantages of it.

But even we use the directed graph, it is still hard to determine the the online observability of a *BCN* with a large number of nodes. Therefore, in the future we will try to separate the *BCNs*, and then determine their online observability respectively. Furthermore, we also want to try to use some knowledge about formal methods to earn scalability for *BCNs*. In addition to the theoretical aspect, the realistic application is also very important. Hence we will also try to find some realistic example which can be modeled by *BCNs*. So that we can research these realistic examples well and determine the online observability their models for better performance.

## REFERENCES

- [1] Waldrop and M. Mitchell, *Complexity : the emerging science at the edge of order and chaos*. Simon & Schuster, 1992.
- [2] D. Cheng and H. Qi, "Controllability and observability of boolean control networks," *Automatica*, vol. 45, no. 7, pp. 1659–1667, 2009.
- [3] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets [j]," *Journal of Theoretical Biology*, vol. 22, no. 3, pp. 437–467, 1968.
- [4] T. Akutsu, S. Miyano, and S. Kuhara, "Inferring qualitative relations in genetic networks and metabolic pathways," *Bioinformatics*, vol. 16, no. 8, pp. 727–734, 2000.
- [5] I. Shmulevich, E. R. Dougherty, and W. Zhang, "From boolean to probabilistic boolean networks as models of genetic regulatory networks," *Proceedings of the IEEE*, vol. 90, no. 11, pp. 1778–1792, 2002.
- [6] A. Faur, A. Naldi, C. Chaouiya, and D. Thieffry, "Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle," *Bioinformatics*, vol. 22, no. 14, pp. e124–e131, 2006.
- [7] D. G. Green, T. G. Leishman, and S. Sadedin, "The emergence of social consensus in boolean networks," in *Artificial Life, 2007. ALIFE '07. IEEE Symposium on*, 2007, pp. 402–408.
- [8] Y. Lou and Y. Hong, "Multi-agent decision in boolean networks with private information and switching interconnection," in *Chinese Control Conference*, 2010, pp. 4530–4535.
- [9] E. Fornasini and M. E. Valcher, "Observability, reconstructibility and state observers of boolean control networks," *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1390–1401, 2013.
- [10] T. Ideker, T. Galitski, and L. Hood, "A new approach to decoding life: systems biology," *Annu Rev Genomics Hum Genet*, vol. 2, no. 1, pp. 343–372, 2001.
- [11] T. Akutsu, M. Hayashida, W. K. Ching, and M. K. Ng, "Control of boolean networks: hardness results and algorithms for tree structured networks," *Journal of Theoretical Biology*, vol. 244, no. 4, p. 670, 2007.
- [12] Y. Zhao, H. Qi, and D. Cheng, "Input-state incidence matrix of boolean control networks and its applications," *Systems & Control Letters*, vol. 59, no. 12, pp. 767–774, 2010.
- [13] D. Cheng, H. Qi, and Z. Li, "Identification of boolean control networks," *Automatica*, vol. 47, no. 4, pp. 702–710, 2011.
- [14] —, *Analysis and Control of Boolean Networks*. Springer London, 2011.
- [15] R. Li, M. Yang, and T. Chu, "Controllability and observability of boolean networks arising from biology," *Chaos An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 2, pp. 1778–1792, 2015.
- [16] K. Zhang and L. Zhang, "Observability of boolean control networks: A unified approach based on finite automata," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2733–2738, 2016.
- [17] D. Cheng, "Semi-tensor product of matrices and its applications - a survey," *Methods & Applications of Analysis*, vol. 321, no. 2, pp. 61–79, 2003.
- [18] S. Klamt, J. Saez-Rodriguez, J. A. Lindquist, L. Simeoni, and E. D. Gilles, "A methodology for the structural and functional analysis of signaling and regulatory networks," *Bmc Bioinformatics*, vol. 7, no. 1, p. 56, 2006.
- [19] Z. Q. Li and J. L. Song, "Controllability of boolean control networks avoiding states set," *Science China(Information Sciences)*, vol. 57, no. 3, pp. 1–13, 2014.