2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Approximation Method to Topological Structure and Stability Analysis of Large-Scale Boolean Networks *

Haitao Li,[†] Xinrong Yang and Wenrong Li[‡]

Abstract

This paper investigates the approximation problem of large-scale Boolean networks (BNs). The whole network is partitioned into several blocks and the input-output representation of each block is obtained via the observed data. By analyzing the simplified network composed of the input-output representation of each block, the properties of the original network can be obtained. The relations of topological structure and finite-time stability between the original system and the approximated system are discussed. Several illustrative examples are given to demonstrate the obtained results about the approximation of large-scale BNs.

Keywords: Large-scale Boolean network, Input-output representation, Approximation, Topological structure, Stability.

1. INTRODUCTION

Boolean networks (BNs) are a special class of finite dynamical systems to model gene regulatory networks, and the variables of BNs take values from the finite set $\{0,1\}$. Starting from modeling molecular networks [1] and then developing into popular logical models [2], BNs have been widely used in systems biology and computational biology [3,4]. By designing the appropriate control strategies of BNs, one can carry out the intervention and treatment of some diseases [5,6]. The updating of each node in a BN is determined by the Boolean function and node interactions. However, it is worth noting that the node interactions will be more complex when the number of nodes increasing rapidly. Therefore, a fundamental topic of studying large-scale BNs is how to reduce the computational load [7].

Recently, the semi-tensor product (STP) of matrices has been proposed for the investigation of BNs [8,9]. By using STP, the algebraic state space representation of BNs can be obtained. Then the classical control theory can be applied to the analysis and control of BNs [10,11]. Under this framework, there exist lots of excellent results for the exploration of BNs, including controllability and observability [12,13], stability and stabilization [14,15], optimal control [16,17], and so on. However, it is worth pointing out that most of the existing results are hardly applied to the large-scale BNs due to the high computational complexity. For large-scale BNs, there exist several useful methods to reduce the computational load, including network aggregation [18,19], logical matrix factorization [20], distributed pinning control [21,22] and so on [23,24].

In [23], the approximation of large-scale BNs was investigated via STP method. Using the approximation method, one can simplify a large-scale network into a relatively small network through the observed data. However, in the process of estimating input-output time-varying matrices, not all the observed data comes into play. Only the observed data which is corresponding to the state transitions with the highest frequency is adopted. In order to make full use of the observed data and to reduce errors in the process of estimation as much as possible, we will use the model of probabilistic Boolean networks (PBNs) to estimate time-varying BNs in this paper.

The main contributions of this paper are listed as follows: (i) The model of PBNs is used to simplify large-scale BNs by describing the input-output representation of each subnetwork. Compared with [23], the observed data can be better applied for the approximation; (ii) The relations of topological structure and stability between the original large-scale BNs and the simplified network are proposed.

The remainder of this paper is organized as follows. Section 2 formulates the approximation problem of large-scale BNs. Section 3 investigates the approximation of large-scale BNs via PBNs. The topological

*The research was supported by the National Natural Science Foundation of China under grant 62073202, and the Taishan Scholar Project of Young Experts under grant tsqn201909076.

†Corresponding author.

‡Haitao Li, Xinrong Yang and Wenrong Li are with the School of Mathematics and Statistics, Shandong Normal University, Jinan 250014, P.R. China (e-mail: haitaoli09@gmail.com; xinrongyang2019@163.com; liwenrong0503@163.com).

979-8-3503-0123-6/23/$31.00 ©2023 IEEE

structure of large-scale BNs is analyzed in Section 4, which is followed by a brief conclusion in Section 5.

For statement convenience, the notations used in this paper are listed here. $\mathscr{D} := \{0,1\}$ and $\mathscr{D}^k := \underbrace{\mathscr{D} \times \cdots \times \mathscr{D}}_{k}$. $\Delta_k := \{\delta_k^i : i = 1, \cdots, k\}$, where δ_k^i is the i-th column of identity matrix I_k. $P = [\delta_k^{i_1} \ \cdots \ \delta_k^{i_s}]$ is called a logical matrix, and it can be simply denoted by $P = \delta_k[i_1 \ \cdots \ i_s]$. The set of $k \times s$ logical matrices is denoted by $\mathscr{L}_{k\times s}$. $|\mathscr{M}|$ denotes the cardinality of set $\mathscr{M}$. $\varphi(R,S)$ is the natural projection from $R = \{\ltimes_{i=1}^n \alpha_i : \alpha_i \in \Delta_2, i = 1, \cdots, n\} \subseteq \Delta_{2^n}$ to $S = \{\ltimes_{j=1}^m \alpha_{i_j} : \alpha_{i_j} \in \Delta_2, i_j \in \{1, \cdots, n\}, j = 1, \cdots, m\} \subseteq \Delta_{2^m}$, $1 \le m < n$.

2. PROBLEM FORMULATION

Consider the following BN:

$$\begin{cases} x_1(t+1) = f_1(X(t)), \\ \quad \vdots \\ x_n(t+1) = f_n(X(t)), \end{cases} \tag{1}$$

where $X(t) = (x_1(t), \cdots, x_n(t))^\top$ is the state vector, and $f_i : \mathscr{D}^n \to \mathscr{D}$, $i = 1, \cdots, n$ are Boolean functions.

Denote the state node set as $\mathscr{X} = \{x_1, \cdots, x_n\}$. In general, the number of state nodes is large for a real gene regulatory network. This motivates us to develop an appropriate approximation method for the study of large-scale BNs.

First of all, by resorting to the network partition [18], we split the whole state node set into s blocks as follows:

$$\mathscr{X} = \mathscr{X}_1 \cup \cdots \cup \mathscr{X}_s, \tag{2}$$

where $\mathscr{X}_i = \{x_{i,1}, \cdots, x_{i,n_i}\}$ is a nonempty proper subset of $\mathscr{X}$, $\mathscr{X}_i \cap \mathscr{X}_j = \emptyset$, $i \neq j$, $i, j = 1, \cdots, s$. For the i-th block, there may exist input nodes from the other blocks and output nodes to the other blocks, denoted by $\mathscr{Z}_i = \{z_{i,1}, \cdots, z_{i,q_i}\}$ and $\mathscr{Y}_i = \{y_{i,1}, \cdots, y_{i,p_i}\}$, respectively. It is obvious that

$$\bigcup_{i=1}^{s} \mathscr{Z}_i = \bigcup_{j=1}^{s} \mathscr{Y}_j := \mathscr{C} \subseteq \mathscr{X}. \tag{3}$$

Without loss of generality, denote the i-th block as subnetwork Σ_i, $i = 1, \cdots, s$ and suppose $|\mathscr{C}| = \beta$. Since gene regulatory networks are generally sparse [25], it is evident that $\beta \ll n$.

Based on the above notations, the dynamics of Σ_i is

$$\begin{cases} x_{i,1}(t+1) = f_{i,1}(X_i(t), Z_i(t)), \\ \quad \vdots \\ x_{i,n_i}(t+1) = f_{i,n_i}(X_i(t), Z_i(t)), \\ y_{i,l}(t) = h_{i,l}(X_i(t)),\ l = 1, \cdots, p_i, \end{cases} \tag{4}$$

where $X_i(t) = (x_{i,1}(t), \cdots, x_{i,n_i}(t))^\top$ and $Z_i(t) = (z_{i,1}(t), \cdots, z_{i,q_i}(t))^\top$ are the state vector and input vector of Σ_i, respectively, and $f_{i,j} : \mathscr{D}^{n_i+q_i} \to \mathscr{D}$, $h_{i,l} : \mathscr{D}^{n_i} \to \mathscr{D}$, $j = 1, \cdots, n_i$, $l = 1, \cdots, p_i$, $i = 1, \cdots, s$ are Boolean functions.

By using the STP method, the algebraic form of subnetwork (4) is

$$\begin{cases} \theta_i(t+1) = F_i \theta_i(t) \rho_i(t), \\ \eta_i(t) = H_i \theta_i(t), \end{cases} \tag{5}$$

where $\theta_i(t) = \ltimes_{j=1}^{n_i} x_{i,j}(t)$, $\rho_i(t) = \ltimes_{j=1}^{q_i} z_{i,j}(t)$, $\eta_i(t) = \ltimes_{j=1}^{p_i} y_{i,j}(t)$, $F_i \in \mathscr{L}_{2^{n_i} \times 2^{n_i+q_i}}$, and $H_i \in \mathscr{L}_{2^{p_i} \times 2^{n_i}}$, $i = 1, \cdots, s$.

For the sake of avoiding the appearance of isolated blocks, we need the following assumption, which is also necessary for deriving the input-output description of each subnetwork.

Assumption 1 *The network graph of system (1) is weakly connected. Furthermore, for a network partition, each block is weakly connected.*

Remark 2 *For convenience, we assume that the components in $\mathscr{Y}_i$ keep the order in $\mathscr{X}$. Moreover, we also assume that the component indices in $\mathscr{Y}_i$ are smaller than that in $\mathscr{Y}_j$, $i < j$, $i, j = 1, \cdots, s$. For instance, for nodes $y_{i,k}$ and $y_{j,r}$, it holds that indices $(i,k) < (j,r)$, where $k \in \{1, \cdots, p_i\}$, $r \in \{1, \cdots, p_j\}$, $i < j$, $i, j = 1, \cdots, s$.*

Furthermore, based on (5), the input-output representation of system (4) is obtained as

$$\eta_i(t+1) = F_i(t) \rho_i(t), \tag{6}$$

where $F_i(t) = H_i F_i \theta_i(t) \in \mathscr{L}_{2^{p_i} \times 2^{q_i}}$ is called the input-output transition matrix of Σ_i, which is a time-varying logical matrix. Since p_i and q_i are generally much smaller than n, $i = 1, \cdots, s$, the size of $F_i(t)$ is much smaller than that in the whole system (1). Therefore, if we can obtain a time-invariant matrix to properly approximate $F_i(t)$, $i = 1, \cdots, s$, then a simplified network can be derived for the original large-scale network.

3. APPROXIMATION OF LARGE-SCALE BNS

In this section, we firstly investigate the approximation of time-varying BNs via PBNs, based on which, we present the approximation of large-scale BNs.

3.1. Approximation of Time-Varying BNs based on PBNs

Consider the following time-varying BN:

$$w(t+1) = g(t, w_1(t), \cdots, w_n(t)), \tag{7}$$

where $w(t)=(w_1(t),\cdots,w_n(t))\in\mathscr{D}^n$ is the state variable, and $g:\mathbb{N}\times\mathscr{D}^n\to\mathscr{D}^n$ is the time-varying Boolean function. System (7) is a time-varying system, which is difficult to handle. We aim to find a proper time-invariant system to approximate system (7) via the following observed data:

$$\mathscr{O}=\{(d_k,e_k):k=0,1,\cdots,N-1\}, \tag{8}$$

where $N\in\mathbb{Z}_+$ is the number of observed pairs of states, $d_k\in\mathscr{D}^n$ and $e_k\in\mathscr{D}^n$ depict the states of system (7) at two adjacent moments, respectively.

In order to make full use of the observed data, we develop a PBN model to approximate system (7). Suppose that the approximating PBN is

$$v(t+1)=\hat{g}(v_1(t),\cdots,v_n(t)), \tag{9}$$

where $v(t)=(v_1(t),\cdots,v_n(t))\in\mathscr{D}^n$ is the state variable, and $\hat{g}:\mathscr{D}^n\to\mathscr{D}^n$ is a time-invariant Boolean function, which is chosen from $\{\hat{g}_1,\cdots,\hat{g}_\tau\}$ with probability $p_j,\ j=1,\cdots,\tau$ at every time step. Obviously, $\sum\limits_{j=1}^{\tau}p_j=1$, and τ is the number of modes in system (9).

By using the STP method, the algebraic form of system (9) is

$$V(t+1)=LV(t), \tag{10}$$

where $V(t)=\ltimes_{i=1}^{n}v_i(t)$, $L\in\{L_1,\cdots,L_\tau\}$, and $\mathbb{P}\{L=L_j\}=p_j,\ j=1,\cdots,\tau$. The derivation of approximating PBN becomes the determination of L_j and p_j, $j=1,\cdots,\tau$ via the observed data.

For the observed data $(d_k,e_k)\in\mathscr{O}$, using the canonical vector form, assume that

$$d_k=\ltimes_{i=1}^{n}w_i(k)\in\Delta_{2^n},\ e_k=\ltimes_{i=1}^{n}w_i(k+1)\in\Delta_{2^n},$$

where $k\in\{0,1,\cdots,N-1\}$.

Split $\mathscr{O}$ into 2^n disjoint groups as

$$\mathscr{O}_i=\{(d_k,e_k)\in\mathscr{O}:d_k=\delta_{2^n}^{i}\},\ i=1,\cdots,2^n.$$

Then split $\mathscr{O}_i$ into 2^n disjoint groups as

$$\mathscr{O}_i^j=\{(d_k,e_k)\in\mathscr{O}_i:e_k=\delta_{2^n}^{j}\},\ i,j=1,\cdots,2^n. \tag{11}$$

For a given threshold $\varepsilon\in\mathbb{Z}_+$, if $|\mathscr{O}_i^j|\geq\varepsilon$, then the corresponding observed data is effective; otherwise, it is not sufficient to determine the state transition from $\delta_{2^n}^{i}$ to $\delta_{2^n}^{j}$ of system (7), and thus we omit it.

Let $J_i=\{j_i:|\mathscr{O}_i^{j_i}|\geq\varepsilon\}:=\{j_i^1,\cdots,j_i^{\tau_i}\}$, where $j_i^1<\cdots<j_i^{\tau_i}$, and $\tau_i=|J_i|$, $i=1,\cdots,2^n$. We use the lexicographical order to arrange elements in $J_1\times\cdots\times J_{2^n}$, and each combination is given as follows:

$$\begin{array}{l}
j_1^1,\ j_2^1,\ \cdots,\ j_{2^n-1}^1,\ j_{2^n}^1,\\
j_1^1,\ j_2^1,\ \cdots,\ j_{2^n-1}^1,\ j_{2^n}^2,\\
\cdots\\
j_1^1,\ j_2^1,\ \cdots,\ j_{2^n-1}^1,\ j_{2^n}^{\tau_{2^n}},\\
\cdots\\
j_1^{\tau_1},\ j_2^{\tau_2},\ \cdots,\ j_{2^n-1}^{\tau_{2^n-1}},\ j_{2^n}^1,\\
j_1^{\tau_1},\ j_2^{\tau_2},\ \cdots,j_{2^n-1}^{\tau_{2^n-1}},\ j_{2^n}^2,\\
\cdots\\
j_1^{\tau_1},\ j_2^{\tau_2},\ \cdots,\ j_{2^n-1}^{\tau_{2^n-1}},\ j_{2^n}^{\tau_{2^n}}.
\end{array} \tag{12}$$

Each line of (12) determines a mode in the approximating PBN. For instance, the second line of (12) is used to describe mode $L_2=\delta_{2^n}[j_1^1\ j_2^1\ \cdots\ j_{2^n-1}^1\ j_{2^n}^2]$. Based on the operation, $\{L_1,\cdots,L_\tau\}$ of the approximating PBN can be obtained, where $\tau=\prod\limits_{i=1}^{2^n}\tau_i$. Moreover,

$$\mathbb{P}\{j_i^*=j_i^k\}=\frac{|\mathscr{O}_i^{j_i^k}|}{\sum\limits_{k=1}^{\tau_i}|\mathscr{O}_i^{j_i^k}|}:=p_{i,k}, \tag{13}$$

where $k=1,\cdots,\tau_i$, $i=1,\cdots,2^n$. Therefore, for

$$L_\zeta=\delta_{2^n}\left[j_1^{k_1}\ j_2^{k_2}\ \cdots\ j_{2^n-1}^{k_{2^n-1}}\ j_{2^n}^{k_{2^n}}\right], \tag{14}$$

it holds that

$$\mathbb{P}\{L=L_\zeta\}=\prod_{i=1}^{2^n}p_{i,k_i}, \tag{15}$$

where $\zeta=1,\cdots,\tau$.

Based on (12), (14) and (15), we establish a time-invariant PBN in the form of (10) to approximate the time-varying BN (7). Please see Algorithm 1.

Remark 3 *The approximation method proposed in [23] just considered $\mathscr{O}_i^j$ with the maximum cardinality and thus obtained a time-invariant BN to approximate the time-varying BNs. It is worth pointing out that the time-invariant PBN model makes full use of the observed data and can reduce the approximating error.*

Remark 4 *The specific value of threshold $\varepsilon\in\mathbb{Z}_+$ is determined according to the amount of observed data under the particular network.*

3.2. Approximation of Large-Scale BNs via PBNs

In this part, based on the results about the approximation of time-varying BNs, we give the procedure for the approximation of large-scale BNs.

Algorithm 1 Calculation of approximating PBN via observed data

Input: $\mathscr{O}$, ε.

Output: $L \in \{L_1, \cdots, L_\tau\}$, and $\mathbb{P}\{L = L_\zeta\}$, $\zeta = 1, \cdots, \tau$.

Split $\mathscr{O}$ into 2^n groups as $\mathscr{O}_i = \{(d_k, e_k) \in \mathscr{O} : d_k = \delta_{2^n}^i\}$, $i = 1, \cdots, 2^n$.
if $\mathscr{O}_i = \emptyset$ **then**
    the observed data is not enough, stop
else
    go to the next step
Split $\mathscr{O}_i$ into 2^n groups as $\mathscr{O}_i^j = \{(d_k, e_k) \in \mathscr{O}_i : e_k = \delta_{2^n}^j\}$, $i, j = 1, \cdots, 2^n$.
Let $J_i = \{j_i : |\mathscr{O}_i^{j_i}| \geq \varepsilon\} := \{j_i^1, \cdots, j_i^{\tau_i}\}$, $i = 1, \cdots, 2^n$.
$L \in \{L_1, \cdots, L_\tau\}$, where $L_1 = \delta_{2^n}[j_1^1\ j_2^1\ \cdots\ j_{2^n}^1]$, $\cdots$, $L_\tau = \delta_{2^n}[j_1^{\tau_1}\ j_2^{\tau_2}\ \cdots\ j_{2^n}^{\tau_{2^n}}]$ and $\tau = \prod_{i=1}^{2^n} \tau_i$.
$\mathbb{P}\{j_i^* = j_i^k\} = \dfrac{|\mathscr{O}_i^{j_i^k}|}{\sum_{k=1}^{\tau_i} |\mathscr{O}_i^{j_i^k}|} = p_{i,k}$.
$\mathbb{P}\{L = L_\zeta\} = \prod_{i=1}^{2^n} p_{i,k_i}$ is satisfied for $L_\zeta = \delta_{2^n}\big[j_1^{k_1}\ j_2^{k_2}\ \cdots\ j_{2^n-1}^{k_{2^n-1}}\ j_{2^n}^{k_{2^n}}\big]$.

According to (2), system (1) is split into s blocks, and the input-output representation of each subnetwork is obtained as (6). Based on (3) and properties of STP, there exists unique matrix $W_i \in \mathscr{L}_{2^{q_i} \times 2^\beta}$ such that $\rho_i(t) = W_i \eta(t)$, where $\eta(t) = \ltimes_{i=1}^s \eta_i(t)$. Then, system (6) can be further described as

$$\eta_i(t+1) = \tilde{F}_i(t)\eta(t), \tag{16}$$

where $\tilde{F}_i(t) = F_i(t)W_i \in \mathscr{L}_{2^{p_i} \times 2^\beta}$. Then, we obtain the following time-varying BN:

$$\eta(t+1) = \tilde{F}(t)\eta(t), \tag{17}$$

where $\tilde{F}(t) = \tilde{F}_1(t) * \cdots * \tilde{F}_s(t) \in \mathscr{L}_{2^\beta \times 2^\beta}$.

Since $\beta \ll n$, there exist sparse connections between each block in (2). Therefore, it is feasible to directly observe the state nodes in the set $\mathscr{C}$. Given a set of observed data, denoted by $\tilde{\mathscr{O}} = \{(\tilde{d}_k, \tilde{e}_k) : k = 0, 1, \cdots, N-1\}$, where $\tilde{d}_k = \ltimes_{i=1}^s \eta_i(k) \in \Delta_{2^\beta}$ and $\tilde{e}_k = \ltimes_{i=1}^s \eta_i(k+1) \in \Delta_{2^\beta}$. For any $i, j = 1, \cdots, 2^\beta$, similar to (11), one can define $\tilde{\mathscr{O}}_i^j$. Fix a threshold $\varepsilon \in \mathbb{Z}_+$. Then, by resorting to Algorithm 1, one can obtain the following PBN to approximate system (17):

$$\eta(t+1) = \tilde{F}\eta(t), \tag{18}$$

where $\tilde{F} \in \{\tilde{F}_1, \cdots, \tilde{F}_\tau\}$, $\tilde{F}_\zeta \in \mathscr{L}_{2^\beta \times 2^\beta}$, and $\mathbb{P}\{\tilde{F} = \tilde{F}_\zeta\} = p_\zeta$, $\zeta = 1, \cdots, \tau$. We call system (18) the approximating PBN of large-scale BN (1).

4. TOPOLOGICAL STRUCTURE AND STABILITY OF LARGE-SCALE BNs

Based on the approximating PBN (18) of large-scale BN (1), this section explores the relation of topological structure between systems (1) and (18), and then analyzes the stability.

First of all, as for the topological structure between systems (1) and (18), we have the following results.

Theorem 5 *If $x_e = \delta_{2^n}^\theta$ is a fixed point of system (1), then $\eta_e = \ltimes_{i=1}^s \varphi(x_e, \mathscr{Y}_i) := \delta_{2^\beta}^\vartheta$ is a positive-probability fixed point of system (18).*

Proof: Since $x_e = \delta_{2^n}^\theta$ is a fixed point of system (1), for some $(\tilde{d}_k, \tilde{e}_k) \in \tilde{\mathscr{O}}$, $\tilde{d}_k = \eta_e$ implies $\tilde{e}_k = \eta_e$. Then, $\tilde{\mathscr{O}}_\vartheta^\vartheta \neq \emptyset$. Hence, for the enough observed data, one has $\tilde{J}_\vartheta := \{j_\vartheta : |\tilde{\mathscr{O}}_\vartheta^{j_\vartheta}| \geq \varepsilon\} \supseteq \{\vartheta\}$. Correspondingly, there exists $\zeta \in \{1, \cdots, \tau\}$ such that $(\tilde{F}_\zeta)_{\vartheta,\vartheta} = 1$.

A straightforward calculation gives $\mathbb{P}\{\eta(t+1) = \delta_{2^\beta}^\vartheta \mid \eta(t) = \delta_{2^\beta}^\vartheta\} = \sum_{i=1}^\tau p_i(\tilde{F}_i)_{\vartheta,\vartheta} \geq p_\zeta > 0$, which shows that $\eta_e = \delta_{2^\beta}^\vartheta$ is a positive-probability fixed point of system (18). □

Suppose that system (1) has a limit cycle $\{\delta_{2^n}^{\gamma_1} \to \cdots \to \delta_{2^n}^{\gamma_l}\}$. Using the natural projection, denote $\delta_{2^\beta}^{\vartheta_j} = \ltimes_{i=1}^s \varphi(\delta_{2^n}^{\gamma_j}, \mathscr{Y}_i)$, $j = 1, \cdots, l$.

Theorem 6 *If $\{\delta_{2^n}^{\gamma_1} \to \cdots \to \delta_{2^n}^{\gamma_l}\}$ is a limit cycle of system (1) and $\vartheta_i \neq \vartheta_j$, $\forall\ i \neq j$, $i, j = 1, \cdots, l$, then $\{\delta_{2^\beta}^{\vartheta_1} \to \cdots \to \delta_{2^\beta}^{\vartheta_l}\}$ is a positive-probability basic cycle of system (18).*

Proof: Since $\{\delta_{2^n}^{\gamma_1} \to \cdots \to \delta_{2^n}^{\gamma_l}\}$ is a limit cycle of system (1), then for some $(\tilde{d}_k, \tilde{e}_k) \in \tilde{\mathscr{O}}$, $\tilde{d}_k = \delta_{2^\beta}^{\vartheta_i}$ implies $\tilde{e}_k = \delta_{2^\beta}^{\vartheta_{i+1}}$, $i = 1, \cdots, l$, where $l + 1 := 1$. Then, $\tilde{\mathscr{O}}_{\vartheta_i}^{\vartheta_{i+1}} \neq \emptyset$. Hence, for the enough observed data, $\tilde{J}_{\vartheta_i} := \{j_{\vartheta_i} : |\tilde{\mathscr{O}}_{\vartheta_i}^{j_{\vartheta_i}}| \geq \varepsilon\} \supseteq \{\vartheta_{i+1}\}$.

Correspondingly, for any $i = 1, \cdots, l$, there exists $\zeta_i \in \{1, \cdots, \tau\}$ such that $(\tilde{F}_{\zeta_i})_{\vartheta_{i+1},\vartheta_i} = 1$. Then,

$$\mathbb{P}\{\eta(t+1) = \delta_{2^\beta}^{\vartheta_{i+1}} \mid \eta(t) = \delta_{2^\beta}^{\vartheta_i}\} \tag{19}$$
$$= \sum_{\zeta=1}^\tau p_\zeta(\tilde{F}_\zeta)_{\vartheta_{i+1},\vartheta_i} \geq p_{\zeta_i} > 0,\ \forall\, i = 1, \cdots, l,$$

which shows that $\{\delta_{2^\beta}^{\vartheta_1} \to \cdots \to \delta_{2^\beta}^{\vartheta_l}\}$ is a positive-probability basic cycle of system (18). □

Remark 7 *Suppose that system (1) has a limit cycle $\{\delta_{2^n}^{\gamma_1} \to \cdots \to \delta_{2^n}^{\gamma_l}\}$. If there exist two distinct positive* *integers $i, j \in \{1,\cdots,l\}$ such that $\vartheta_i = \vartheta_j$, then system (18) has at least two positive-probability basic cycles with length being less than l.*

Next, for the stability analysis between systems (1) and (18), we have the following results.

Theorem 8 *If system (1) is globally stable at $x_e = \delta_{2^n}^{\theta}$, then system (18) is stable at $\eta_e = \ltimes_{i=1}^{s}\varphi(x_e,\mathscr{Y}_i)$ with positive probability.*

Proof: Assume that system (1) is globally stable at x_e. On one hand, for any initial state $\delta_{2^n}^{\gamma_0} \in \Delta_{2^n}$, there exist $l \in \mathbb{Z}_+$ and $\gamma_1,\cdots,\gamma_l \in \{1,\cdots,2^n\}$ satisfying $x(1,\delta_{2^n}^{\gamma_i}) = \delta_{2^n}^{\gamma_{i+1}}$, $\forall\ i = 0,\cdots,l$, where $\gamma_{l+1} = \theta$. Accordingly, for some $(\tilde{d}_k,\tilde{e}_k) \in \tilde{\mathscr{O}}$, $\tilde{d}_k = \delta_{2^\beta}^{\vartheta_i}$ implies $\tilde{e}_k = \delta_{2^\beta}^{\vartheta_{i+1}}$, $i = 0,\cdots,l$, where $\delta_{2^\beta}^{\vartheta_{l+1}} = \eta_e$. Then, similar to the derivation of (19), for any $i = 0,\cdots,l$, there exists $\zeta_i \in \{1,\cdots,\tau\}$ such that $\mathbb{P}\{\eta(l+1) = \eta_e \mid \eta(0) = \delta_{2^\beta}^{\vartheta_0}\} \geq (\sum_{\zeta=1}^{\tau} p_\zeta(\tilde{F}_\zeta))_{\vartheta_{l+1},\vartheta_l}\cdots(\sum_{\zeta=1}^{\tau} p_\zeta(\tilde{F}_\zeta))_{\vartheta_1,\vartheta_0} \geq p_{\zeta_l}\cdots p_{\zeta_0} > 0$. On the other hand, since x_e is a fixed point of system (1), we obtain from Theorem 5 that η_e is a positive-probability fixed point of system (18).

To sum up, system (18) is stable at η_e with positive probability. □

In order to demonstrate the effectiveness of the approximation method in simplifying large-scale BNs, we give an example of BN with 400 nodes.

Example 9 *Consider the following BN with 400 nodes:*

$$\begin{cases} x_i(t+1) = x_{i+1}(t) \vee x_{i+3}(t),\ i = 1,\cdots,50, \\ x_i(t+1) = \neg x_{i+1}(t) \wedge x_{i+2}(t), \\ \qquad\qquad\qquad i = 51,\cdots,100, \\ x_i(t+1) = x_{i+2}(t),\ i = 101,\cdots,150, \\ x_i(t+1) = \neg(x_{i+1}(t) \wedge x_{i+2}(t)), \\ \qquad\qquad\qquad i = 151,\cdots,200, \\ x_i(t+1) = x_{i+1}(t) \vee x_{i+3}(t), \\ \qquad\qquad\qquad i = 201,\cdots,250, \\ x_i(t+1) = \neg(x_{i+1}(t) \wedge x_{i+3}(t)), \\ \qquad\qquad\qquad i = 251,\cdots,350, \\ x_i(t+1) = \neg x_{i+2}(t),\ i = 351,\cdots,398, \\ x_{399}(t+1) = \neg x_1(t),\ x_{400}(t+1) = \neg x_2(t), \end{cases} \tag{20}$$

where $x_i \in \mathscr{D}$, $i = 1,\cdots,400$ are state variables.

Assume that the network graph of system (20) is partitioned into 7 blocks as

$$\mathscr{X}_1 = \{x_1,\cdots,x_{50}\},\ \mathscr{Y}_1 = \mathscr{Z}_7 = \{x_1,x_2\},$$
$$\mathscr{X}_2 = \{x_{51},\cdots,x_{100}\},\ \mathscr{Y}_2 = \mathscr{Z}_1 = \{x_{51},x_{52},x_{53}\},$$
$$\mathscr{X}_3 = \{x_{101},\cdots,x_{150}\},\ \mathscr{Y}_3 = \mathscr{Z}_2 = \{x_{101},x_{102}\},$$
$$\mathscr{X}_4 = \{x_{151},\cdots,x_{200}\},\ \mathscr{Y}_4 = \mathscr{Z}_3 = \{x_{151},x_{152}\},$$
$$\mathscr{X}_5 = \{x_{201},\cdots,x_{250}\},\ \mathscr{Y}_5 = \mathscr{Z}_4 = \{x_{201},x_{202}\},$$
$$\mathscr{X}_6 = \{x_{251},\cdots,x_{350}\},\ \mathscr{Y}_6 = \mathscr{Z}_5 = \{x_{251},x_{252},x_{253}\},$$
$$\mathscr{X}_7 = \{x_{351},\cdots,x_{400}\},\ \mathscr{Y}_7 = \mathscr{Z}_6 = \{x_{351},x_{352},x_{353}\}.$$

By setting $\eta_1 = \rho_7 = x_1 \ltimes x_2$, $\eta_2 = \rho_1 = x_{51} \ltimes x_{52} \ltimes x_{53}$, $\eta_3 = \rho_2 = x_{101} \ltimes x_{102}$, $\eta_4 = \rho_3 = x_{151} \ltimes x_{152}$, $\eta_5 = \rho_4 = x_{201} \ltimes x_{202}$, $\eta_6 = \rho_5 = x_{251} \ltimes x_{252} \ltimes x_{253}$, and $\eta_7 = \rho_6 = x_{351} \ltimes x_{352} \ltimes x_{353}$, we firstly aim to obtain the following simplified system:

$$\eta(t+1) = F\eta(t), \tag{21}$$

where $\eta(t) = \ltimes_{i=1}^{7}\eta_i(t)$, $F \in \mathscr{L}_{2^{17}\times 2^{17}}$ is the state transition matrix.

Considering that the model of system (20) is known, then the observed data can be obtained by choosing the initial states arbitrarily. Generate data randomly using Python. In this example, we choose the initial states randomly for 100 times, and we observe 1000 data at each time. According to Algorithm 1, the structural matrices F of system (21) can be estimated as $F = F_1 * \cdots * F_7$, where

$$\begin{aligned} & F_1 = (\mathbf{1}_4^\top \otimes I_8 \otimes \mathbf{1}_{4096}^\top)H_1, \\ & H_1 = \delta_4[j_{1,1}\ 3\ j_{1,3}\ 2\ 2\ j_{1,6}\ 4\ 3], \\ & F_2 = (\mathbf{1}_{32}^\top \otimes I_4 \otimes \mathbf{1}_{1024}^\top)H_2,\ H_2 = \delta_8[3\ 1\ 6\ j_{2,4}], \\ & F_3 = (\mathbf{1}_{128}^\top \otimes I_4 \otimes \mathbf{1}_{256}^\top)H_3,\ H_3 = \delta_4[j_{3,1}\ 2\ 4\ 4], \\ & F_4 = (\mathbf{1}_{512}^\top \otimes I_4 \otimes \mathbf{1}_{64}^\top)H_4,\ H_4 = \delta_4[2\ 4\ 4\ 2], \\ & F_5 = (\mathbf{1}_{2048}^\top \otimes I_8 \otimes \mathbf{1}_{8}^\top)H_5, \\ & H_5 = \delta_4[4\ 3\ 3\ 1\ 3\ 3\ 2\ 4], \\ & F_6 = (\mathbf{1}_{2^{14}}^\top \otimes I_8)H_6,\ H_6 = \delta_8[4\ 4\ 6\ 3\ 3\ 8\ 5\ 7], \\ & F_7 = (I_4 \otimes \mathbf{1}_{2^{15}}^\top)H_7,\ H_7 = \delta_8[8\ j_{7,2}\ 1\ 2], \end{aligned}$$

$j_{1,1} \in \{2,3\}$, $j_{1,3} \in \{1,4\}$, $j_{1,6} \in \{2,4\}$, $j_{2,4} \in \{1,3,7\}$, $j_{3,1} \in \{1,4\}$, $j_{7,2} \in \{5,7\}$, and $j_{1,1}, j_{1,3}, j_{1,6}, j_{2,4}, j_{3,1}, j_{7,2}$ take their values from the corresponding set with equal probability, respectively.

After obtaining the structural matrices of system (21), a simplified network can be derived for system (20). By using the dummy operator, one can obtain the algebraic representation of system (21) in the form of (18). Then we can analyze the topological structure of system (20) based on the approximating system (21). □

Remark 10 *We can see from Example 9 that the approximation method can obtain a small-scale system to estimate the original large-scale system. If we directly deal with system (20), we should handle matrix of size $2^{400} \times 2^{400}$. On the contrary, for the simplified system, the matrix we should handle is of size $2^{17} \times 2^{17}$, which is far less than the former. Though the simplified system is an approximation of original network, the computational load can be reduced greatly.*

5. CONCLUSION

In this paper, we have studied the approximation problem of large-scale BNs. The time-varying input-output matrix of each block has been approximated after partitioning the whole network into several blocks. Furthermore, the approximating PBN for the original system has been derived, and the relations of topological structure and stability between the original system and the approximating PBN have been discussed. Future work will devote to establishing several necessary and sufficient conditions for the stability of original BNs via the approximation method.

References

[1] S. Kauffman, Homeostasis and differentiation in random genetic control networks, *Nature*, vol. 224, no. 5215, 1969, pp 177-178.

[2] R. Thomas and R. D'Ari, *Biological Feedback*, CRC press, 1990.

[3] S. Kauffman, Metabolic stability and epigenesis in randomly constructed genetic nets, *Journal of Theoretical Biology*, vol. 22, no. 3, 1969, pp 437-467.

[4] L. Raeymaekers, Dynamics of Boolean networks controlled by biologically meaningful functions, *Journal of Theoretical Biology*, vol. 218, no. 3, 2002, pp 331-341.

[5] S. M. Choo and K. H. Cho, An efficient algorithm for identifying primary phenotype attractors of a large-scale Boolean network, *BMC Systems Biology*, vol. 10, no. 1, 2016, pp 95.

[6] Z. Ma, Z. Wang and M. J. Mckeown, Probabilistic Boolean network analysis of brain connectivity in parkinson's disease, *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 6, 2008, pp 975-985.

[7] T. Akutsu, M. Hayashida, W. Ching and M. Ng, Control of Boolean networks: Hardness results and algorithms for tree structured networks, *Journal of Theoretical Biology*, vol. 244, no. 4, 2007, pp 670-679.

[8] D. Cheng, H. Qi and Z. Li, *Analysis and Control of Boolean Networks: A Semi-Tensor Product Approach*, London: Springer, 2011.

[9] D. Cheng and H. Qi, A linear representation of dynamics of Boolean networks, *IEEE Transactions on Automatic Control*, vol. 55, no. 10, 2010, pp 2251-2258.

[10] J. Lu, H. Li, Y. Liu and F. Li, Survey on semi-tensor product method with its applications in logical networks and other finite-valued systems, *IET Control Theory & Applications*, vol. 11, no. 13, 2017, pp 2040-2047.

[11] H. Li, G. Zhao, M. Meng and J. Feng, A survey on applications of semi-tensor product method in engineering, *Science China Information Sciences*, vol. 61, 2018, pp 010202.

[12] D. Cheng, C. Li and F. He, Observability of Boolean networks via set controllability approach, *Systems & Control Letters*, vol. 115, 2018, pp 22-25.

[13] Y. Liu, H. Chen and B. Wu, Controllability of Boolean control networks with impulsive effects and forbidden states, *Mathematical Methods in the Applied Sciences*, vol. 37, no. 1, 2014, pp 1-9.

[14] H. Li, X. Yang and S. Wang, Robustness for stability and stabilization of Boolean networks with stochastic function perturbations, *IEEE Transactions on Automatic Control*, vol. 66, no. 3, 2021, pp 1231-1237.

[15] L. Lin, J. Cao, M. Abdel-Aty and U. A. Al-Juboori, Topological structure, reachability, and stabilization of constrained Boolean control networks via event-triggered control, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, 2021, pp 3565-3574.

[16] E. Fornasini and M. E. Valcher, Optimal control of Boolean control networks, *IEEE Transactions on Automatic Control*, vol. 59, no. 5, 2014, pp 1258-1270.

[17] Y. Wu, X. Sun, X. Zhao and T. Shen, Optimal control of boolean control networks with average cost: A policy iteration approach, *Automatica*, vol. 100, 2019, pp 378-387.

[18] Y. Zhao, J. Kim and M. Filippone, Aggregation algorithm towards large-scale Boolean network analysis, *IEEE Transactions on Automatic Control*, vol. 58, no. 8, 2013, pp 1976-1985.

[19] Y. Zhao, B. Ghosh and D. Cheng, Control of large-scale Boolean networks via network aggregation, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 7, 2016, pp 1527-1536.

[20] H. Li and Y. Wang, Logical matrix factorization with application to topological structure analysis of Boolean network, *IEEE Transactions on Automatic Control*, vol. 60, no. 5, 2015, pp 1380-1385.

[21] S. Zhu, J. Lu, S. Azuma and W. Zheng, Strong structural controllability of Boolean networks: Polynomial-time criteria, minimal node control, and distributed pinning strategies, *IEEE Transactions on Automatic Control*, DOI: 10.1109/TAC.2022.3226701.

[22] S. Zhu, J. Cao, L. Lin, J. Lam and S. Azuma, Towards stabilizable large-scale Boolean networks by controlling the minimal set of nodes, *IEEE Transactions on Automatic Control*, DOI: 10.1109/TAC.2023.3269321.

[23] D. Cheng, Y. Zhao, J. Kim and Y. Zhao, "Approximation of Boolean networks", *Proceedings of the 10th World Congress on Intelligent Control and Automation*, Beijing, 2012, pp. 2280-2285.

[24] J. Zhong, D. W. C. Ho and J. Lu, A new approach to pinning control of Boolean networks, *IEEE Transactions on Control of Network Systems*, vol. 9, no. 1, 2022, pp 415-426.

[25] M. Andrecut and S. Kauffman, On the sparse reconstruction of gene networks, *Journal of Computational Biology*, vol. 15, no. 1, 2008, pp 21-30.

[26] K. Kobayashi, A graph-theoretic approach to design of probabilistic Boolean networks, *IFAC PapersOnLine*, vol. 51, no. 33, 2018, pp 217-222.