frontiers in Physiology

METHODS
published: 08 June 2018
doi: 10.3389/fphys.2018.00695

Identification of Boolean Network Models From Time Series Data Incorporating Prior Knowledge

*Thomas Leifeld, Zhihua Zhang and Ping Zhang**

Institute of Automatic Control, Technische Universität Kaiserslautern, Kaiserslautern, Germany

Motivation: Mathematical models take an important place in science and engineering. A model can help scientists to explain dynamic behavior of a system and to understand the functionality of system components. Since length of a time series and number of replicates is limited by the cost of experiments, Boolean networks as a structurally simple and parameter-free logical model for gene regulatory networks have attracted interests of many scientists. In order to fit into the biological contexts and to lower the data requirements, biological prior knowledge is taken into consideration during the inference procedure. In the literature, the existing identification approaches can only deal with a subset of possible types of prior knowledge.

Results: We propose a new approach to identify Boolean networks from time series data incorporating prior knowledge, such as partial network structure, canalizing property, positive and negative unateness. Using vector form of Boolean variables and applying a generalized matrix multiplication called the semi-tensor product (STP), each Boolean function can be equivalently converted into a matrix expression. Based on this, the identification problem is reformulated as an integer linear programming problem to reveal the system matrix of Boolean model in a computationally efficient way, whose dynamics are consistent with the important dynamics captured in the data. By using prior knowledge the number of candidate functions can be reduced during the inference. Hence, identification incorporating prior knowledge is especially suitable for the case of small size time series data and data without sufficient stimuli. The proposed approach is illustrated with the help of a biological model of the network of oxidative stress response.

Conclusions: The combination of efficient reformulation of the identification problem with the possibility to incorporate various types of prior knowledge enables the application of computational model inference to systems with limited amount of time series data. The general applicability of this methodological approach makes it suitable for a variety of biological systems and of general interest for biological and medical research.

Keywords: Boolean networks, identification, prior knowledge, time series data, network inference

OPEN ACCESS

Edited by:
Tomáš Helikar,
University of Nebraska-Lincoln,
United States

Reviewed by:
Aurélien Naldi,
École Normale Supérieure, France
Ruisheng Wang,
Department of Medicine, Brigham and Women's Hospital, United States

****Correspondence:***
Ping Zhang
pzhang@eit.uni-kl.de

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Physiology

Received: *01 February 2018*
Accepted: *18 May 2018*
Published: *08 June 2018*

Citation:
Leifeld T, Zhang Z and Zhang P (2018) Identification of Boolean Network Models From Time Series Data Incorporating Prior Knowledge. Front. Physiol. 9:695. doi: 10.3389/fphys.2018.00695

1. INTRODUCTION

Boolean networks (BNs) are discrete-time systems, whose variables can take only two possible values (i.e., 0 and 1). Since Stuart Kaufman firstly introduced BNs in Kauffman (1969) for qualitative description of gene regulatory interactions, BNs have attracted great attention from many scientists and several results have been proposed, for instance, analysis (Albert and Barabási, 2000) and control (Fauré et al., 2006). An overview can be found in Wang et al. (2012) and a database for established models and compatible tools has been introduced (Naldi et al., 2015).

Mathematical models are important to explain dynamic behavior of a system and to understand the functionality of system components (Grieb et al., 2015) and can help scientists to design model-based targeted therapy and diagnosis (Fumia and Martins, 2013). Hence, the inference of models capturing the relevant behavior of the system is an important topic. The inference can be based on the connection of known biochemical reactions, like BN model for the yeast cell cycle in Davidich and Bornholdt (2008), or on experimental data, if the latter is the case it is also called the identification problem. One of the first approaches to identify a BN was REVEAL which is based on mutual information (Liang et al., 1998). In Akutsu et al. (1999) a similar but less complex approach is presented. Both cannot handle errors in the dataset which was solved in Lähdesmäki et al. (2003). The modeled quantities are not Boolean in the experimental data and need to be binarized first. For the binarization several approaches can be found in the literature ranging from mixture model based clustering (Zhou et al., 2003) to more complex methods where the significance of a jump in the time series is estimated in Hopfensitz et al. (2012). A comparison of some identification and binarization approaches and their combinations can be found in Berestovsky and Nakhleh (2013). Most identification approaches are based on previously binarized data, but there also exist approaches directly based on continuous data (e.g., Karlebach and Shamir, 2012). In Higa et al. (2011) the data is considered as given constraint and the set of systems fulfilling the constraints is searched. This approach was then further improved by reducing the sensitivity to noise in Ouyang et al. (2014). An example of recent research is the identification of Boolean models for transient dynamics after perturbations from time course data with answer set programming (Ostrowski et al., 2016). A BN can simply be extended to a Boolean control network (BCN) by considering manipulated external stimuli as control signal of the network. Recently, a powerful tool called semi-tensor product (STP) of matrices has been proposed in Cheng (2001), which can convert the dynamics of BCNs into a model where all information of the dynamics and the structure of the BCN is contained in two matrices (Cheng et al., 2011a). Using the STP based matrix description of BCN several approaches for identifying BCN have been proposed (Cheng and Zhao, 2011; Fornasini and Valcher, 2014; Zhang et al., 2017a).

However, in general, in order to identify the dynamical model of a BCN from its input and output data, a huge number of data is required (Cheng and Zhao, 2011; Cheng et al., 2011b). Though, in practice, data size is limited by the cost of experiments (Geier et al., 2007). In order to reduce the search space and improve the accuracy of the model, the benefit of biological prior knowledge should be taken into consideration. Cheng and Zhao (2011) pointed out that, if the network graph is known, then the data required can be reduced considerably. In the literature there are several approaches to include different types of prior knowledge. For example the known network structure and known steady state activity is considered in Videla et al. (2015). Moreover, two common properties of the Boolean function, canalizing and unateness, can be further utilized according to Breindl et al. (2013) and Faisal et al. (2010). A Boolean function is canalizing, if a variable takes on a certain "canalizing" value, then the output of the boolean function is always the same (Waddington, 1942). Different from canalizing function, an unate function has monotonic properties, which in biology indicates that a gene acts exclusively as an inducer or as an inhibitor for the expression of another gene (Porreca et al., 2010). The prior knowledge is used in different ways either by introducing additional constraints in the optimization (Breindl et al., 2013), or reducing the number of parameters in the optimization (Cheng and Zhao, 2011). In Dorier et al. (2016) and Terfve et al. (2012) genetic algorithms are used to handle the complexity problem of large networks while satisfying prior knowledge network graphs as constraints. However, these approaches to handle prior knowledge are not compatible and the advantages of different types of prior knowledge can not be combined. In the approach proposed in this paper, all different types of prior knowledge can be utilized simultaneously and it can additionally handle hypotheses for interactions, which could be used for researcher bias free distinction between alternative hypotheses. Furthermore existing approaches can not handle the case that at some time instances some measurement values are missing, which cannot be avoided in practice due to the limitation of measuring techniques like mass spectrometry-based proteomics.

In this paper, we consider the identification problem of BCNs utilizing biological prior knowledge. A part of the results was presented at the 56th IEEE Conference on Decision and Control in Melbourne (Zhang et al., 2017b). However, the BCN model considered in Zhang et al. (2017b) contains a general output equation. By applying prediction error method (PEM), a high-dimensional BCN (i.e., $2^n \times 2^{n+m}$) cannot be avoided. Different from that, although the handling of unmeasurable processes is considered in this paper, the proposed approach leads to a low-dimensional matrix for PEM. Besides, more prior biological knowledge is considered in the paper, like potential interactions, known attractors and limit cycles. Moreover, it is discussed how to deal with alternative hypotheses for interactions and missing measurement points. The main contributions of this paper are as follows:

- A suitable way to handle the prior knowledge such as known network graph, hypotheses for interactions, canalizing and unateness properties or attractor is introduced. For this purpose the BCN is described by two matrices with unknown parameters as entries. If possible, some parameters are inferred directly. Otherwise, relationships between the parameters are set up.
- An approach to deal with the identification of BCNs, in particular, from noisy measurements and missing data points

is proposed. The identification problem of BCNs is formulated as a nonlinear pseudo-Boolean optimization, which can be equivalently transformed into a linear binary optimization problem and then solved efficiently.

The remainder of the paper is organized as follows. Section 2 introduces some fundamental definitions and notations. In Section 3, the identification problem of BCNs addressed in this paper will be formulated. Section 4 introduces a way to utilize prior knowledge in identification procedure. The formulation of identification problem of BCNs as an integer linear programming problem is derived and an example is given in Section 5 to illustrate the approach. Finally, a short discussion on the advantages and limitations of the proposed approach is given in Section 6.

2. PRELIMINARIES

In this part, we list some necessary notations, which will be used in the subsequent sections.

1. $\neg$, $\wedge$ and $\vee$ denote the logical negation (not), conjunction (and) and disjunction (or), respectively.
2. $\mathcal{D} := \{1, 0\}$ and $\mathcal{D}^n = \underbrace{\mathcal{D} \times \mathcal{D} \times \cdots \times \mathcal{D}}_{n}$.
3. $\Delta_n := \{\delta_n^k | 1 \leq k \leq n\}$, where δ_n^k denotes the k-th column of the identity matrix I_n.
4. For a vector $v \in \mathbb{R}^m$, its j-th entry is denoted by $[v]_j, j = 1, 2, \cdots, m$.
5. An $n \times t$ matrix L is called a logical matrix, if $L = [\delta_n^{i_1}\ \delta_n^{i_2}\ \cdots\ \delta_n^{i_t}]$, where $i_1, i_2, \cdots, i_t \in \{1, 2, \cdots, n\}$, and we express L briefly as $L = \delta_n[i_1\ i_2\ \cdots\ i_t]$. Denote the set of $n \times t$ logical matrices by $\mathcal{L}_{n\times t}$. $Col_i(M)$ denotes the i-th column of the matrix M.
6. $\mathbf{0}_n := [\underbrace{0\ 0\ \cdots\ 0}_{n}]^{\mathrm{T}}$, where the superscript T denotes the transpose.

The concept of the semi-tensor product of matrices (STP) has been introduced by Cheng et al. (2011a). The STP of two matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ is defined as

$$A \ltimes B = (A \otimes I_{l/n}) \cdot (B \otimes I_{l/p}) \tag{1}$$

where $\otimes$ is the Kronecker product and $l = lcm\{n, p\}$ is the least common multiple of n and p. The following property of the STP will be used in the subsequent sections.

Lemma 1. *Let $X \in \mathbb{R}^{m\times 1}$ and $Y \in \mathbb{R}^{n\times 1}$. Then $Y \ltimes X = W_{[m,n]} \ltimes X \ltimes Y$, where $W_{[m,n]}$ is the swap matrix (Cheng et al., 2011a).*

So the order of two vectors which are multiplied can be altered by multiplying a suitable matrix from the left, this is also called the pseudo-commutativity of the STP. In the following parts the symbol $\ltimes$ will be omitted.

3. PROBLEM FORMULATION

System identification is the determination of a model describing the dynamic behavior of a system based on measured data and known perturbations. In the context of Boolean modeling it is assumed that the transient behavior of the system can be qualitatively described by a finite number of Boolean states and that the interaction of these states can be described by Boolean functions. The perturbations are inputs to the system and cause transient behavior of the interacting states in the system. A measured time series of inputs and states form together the data basis for the identification. Depending on the system which is to be modeled, the states might represent the activity of genes or the abundance of proteins and the perturbations could be a stress like heat or oxygen or a chemical substance. In the following the identification process will be formulated as mathematical optimization problem. Therefore the mathematical model of a BCN needs to be defined first. A Boolean control network (BCN) can be described by the following equations (Cheng and Qi, 2010):

$$\begin{cases} X_1(t+1) &= f_1(X_1(t), \cdots, X_n(t), U_1(t), \cdots, U_m(t)) \\ &\vdots \\ X_n(t+1) &= f_n(X_1(t), \cdots, X_n(t), U_1(t), \cdots, U_m(t)) \end{cases} \tag{2}$$

where $X(t) = [X_1(t)\ X_2(t)\ \cdots\ X_n(t)]^{\mathrm{T}} \in \mathcal{D}^n$, $U(t) = [U_1(t)\ U_2(t)\ \cdots\ U_m(t)]^{\mathrm{T}} \in \mathcal{D}^m$ are, respectively, the state vector, input vector at time t, f_i are logic functions. At the discrete time instances t the state variables are updated synchronously according to the logic functions f_i. As shown in Cheng and Qi (2010), a vector form of Boolean variable $X_i, i = 1, 2, \cdots, n$ can be simply expressed as

$$x_i = \begin{bmatrix} X_i \\ \neg X_i \end{bmatrix}. \tag{3}$$

Let $x = \ltimes_{i=1}^n x_i \in \Delta_{2^n}, u = \ltimes_{i=1}^m u_i \in \Delta_{2^m}$. According to Cheng and Qi (2010), (2) can be equivalently represented in a vector form:

$$\begin{cases} x_1(t+1) = S_1 u(t) x(t) \\ \quad\vdots \\ x_n(t+1) = S_n u(t) x(t) \end{cases}, \tag{4}$$

where $S_i \in \mathcal{L}_{2\times 2^{n+m}}, i = 1, 2, \cdots, n$ are logical matrices. Multiplying all Equations in (4) together, there is

$$x(t+1) = Lu(t)x(t) \tag{5}$$

where $L \in \mathcal{L}_{2^n\times 2^{n+m}}$ is a logical matrix and $Col_i(L) = \ltimes_{j=1}^n Col_i(S_j), i = 1, 2, \cdots, 2^{n+m}$.

A polynomial $P_{ml} : \mathbb{R}^k \to \mathbb{R}$ with k variables $\{\theta_1, \theta_2, \cdots, \theta_k\}$ is called multi-linear polynomial, if its degree in each variable is at most 1 (Alon et al., 1991). So, a multi-linear polynomial can be generally expressed as

$$P_{ml}(\theta_1, \theta_2, \cdots, \theta_k) = c + \sum_{i=1}^{k} c_i\theta_i + \sum_{\alpha=1}^{q} c_{\mathcal{I}_\alpha} \prod_{j\in\mathcal{I}_\alpha} \theta_j \tag{6}$$

where $c, c_i, c_{\mathcal{I}_\alpha} \in \mathbb{R}$ for $\mathcal{I}_\alpha \subset V = \{1, 2, \cdots, k\}$ and the set $\mathcal{I}_\alpha$ has a cardinality of at least 2, i.e., $|\mathcal{I}_\alpha| \geq 2, \alpha = 1, 2, \cdots, q$.

Generally, the identification problem of BCNs can be described as reconstruction of Boolean functions $f_i, i = 1, 2, \cdots, n$ that explain the experimental data as well as possible. Because of equivalent representation of a Boolean function by a logical matrix, the identification problem is reformulated as searching for logical matrices $S_i \in \mathcal{L}_{2\times 2^{n+m}}, i = 1, 2, \cdots, n$ based on the input and measurement state data.

Note that any logical matrix in $\mathcal{L}_{2^a\times 2^b}$ can be expressed by multi-linear polynomials in a binary parameter vector $\boldsymbol{\theta}$ of dimension $a \cdot 2^b$. For example, any logical matrix in $\mathcal{L}_{4\times 8}$ can be expressed by a binary parameter vector $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \cdots \ \theta_{16}]^{\mathrm{T}}$ as

$$\begin{bmatrix} \theta_1\cdot\theta_2 & \theta_1\cdot(1-\theta_2) & (1-\theta_1)\cdot\theta_2 & (1-\theta_1)\cdot(1-\theta_2) \\ \theta_3\cdot\theta_4 & \theta_3\cdot(1-\theta_4) & (1-\theta_3)\cdot\theta_4 & (1-\theta_3)\cdot(1-\theta_4) \\ \theta_5\cdot\theta_6 & \theta_5\cdot(1-\theta_6) & (1-\theta_5)\cdot\theta_6 & (1-\theta_5)\cdot(1-\theta_6) \\ \theta_7\cdot\theta_8 & \theta_7\cdot(1-\theta_8) & (1-\theta_7)\cdot\theta_8 & (1-\theta_7)\cdot(1-\theta_8) \\ \theta_9\cdot\theta_{10} & \theta_9\cdot(1-\theta_{10}) & (1-\theta_9)\cdot\theta_{10} & (1-\theta_9)\cdot(1-\theta_{10}) \\ \theta_{11}\cdot\theta_{12} & \theta_{11}\cdot(1-\theta_{12}) & (1-\theta_{11})\cdot\theta_{12} & (1-\theta_{11})\cdot(1-\theta_{12}) \\ \theta_{13}\cdot\theta_{14} & \theta_{13}\cdot(1-\theta_{14}) & (1-\theta_{13})\cdot\theta_{14} & (1-\theta_{13})\cdot(1-\theta_{14}) \\ \theta_{15}\cdot\theta_{16} & \theta_{15}\cdot(1-\theta_{16}) & (1-\theta_{15})\cdot\theta_{16} & (1-\theta_{15})\cdot(1-\theta_{16}) \end{bmatrix}^{\mathrm{T}}$$

where the superscript T denotes the transpose. In this way, each realization of the binary parameter vector $\boldsymbol{\theta} \in \mathcal{D}^{a2^b}$ corresponds to a unique logical matrix. It is straightforward to equivalently convert this logical matrix into readable logical equations. Based on this, the objective of the paper is to find a binary parameter vector $\boldsymbol{\theta}$, such that dynamic behavior of the BCN (5) is consistent with the important dynamics captured in the observed input-state data.

4. INCORPORATION OF PRIOR KNOWLEDGE

In this section, we shall show how to utilize known network graph, potential interactions, canalizing and unateness properties and attractors in the identification procedure.

4.1. Known or Potential Interactions

Often some or all interaction partners are known in a biological system which is subject of identification. This knowledge can come from databases or can be constructed based knowledge about the underlying biochemical reactions. In some cases a known signaling network is to be complemented and different hypothesis for potential interactions shall be evaluated. If all interaction partners and the direction of the interactions are known, the underlying directed network graph of the BN is known.

In graph theory, a directed graph can be denoted by $G = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is a finite set of nodes and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is a finite set of edges (Bollobas, 2012). If $(v_i, v_j) \in \mathcal{E}$, then there is an edge from $v_i \to v_j$. According to Cheng et al. (2011a), a BCN can be represented by a directed graph, where each gene is considered as a node. If there is an edge from $X_i \to X_j$, then X_j is affected by X_i.

Assume that a directed graph for a BCN $G = \{\mathcal{V}, \mathcal{E}\}$ is known. Then we have the following result.

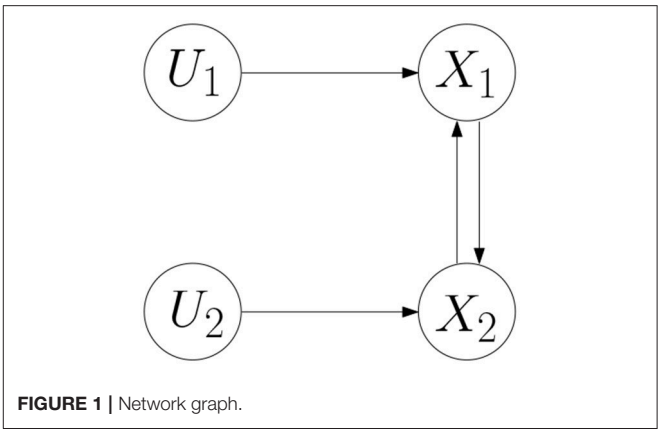

FIGURE 1 | Network graph.

Lemma 2. *If the node X_i is affected by w nodes, then 2^w binary parameters are enough to describe the corresponding logical matrix S_i.*

Proof: As the node x_i is affected by w nodes, then the Boolean function can be represented in a vector form as

$$x_i(t+1) = S_i x_{i_1}(t) x_{i_2}(t) \cdots x_{i_w}(t)$$

where the matrix S_i is a logical matrix of dimension 2×2^w. Recall that the logical matrix S_i is a matrix containing only columns belonging to Δ_2 (Cheng et al., 2011a). Hence, 2^w binary parameters are enough for the description of the logical matrix S_i.

An example is given below to express logical matrices of a BCN with a known network graph with the help of binary parameters.

Example 1. Consider a BCN as follows.

$$\begin{cases} X_1(t+1) &= f_1(X_2(t), U_1(t)) \\ X_2(t+1) &= f_2(X_1(t), U_2(t)) \end{cases} \tag{7}$$

where the network graph of the BCN is shown in **Figure 1** (Cheng and Zhao, 2011). According to Cheng and Qi (2010), the algebraic form of the BCN is obtained,

$$\begin{cases} x_1(t+1) = S_1 u_1(t) x_2(t) \\ x_2(t+1) = S_2 u_2(t) x_1(t) \end{cases} \tag{8}$$

where the logical matrices $S_1, S_2 \in \mathcal{L}_{2\times 4}$ can be expressed by the binary parameter vector $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \cdots \ \theta_8]^{\mathrm{T}}$ in the following form:

$$S_1 = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 \\ 1-\theta_1 & 1-\theta_2 & 1-\theta_3 & 1-\theta_4 \end{bmatrix},$$
$$S_2 = \begin{bmatrix} \theta_5 & \theta_6 & \theta_7 & \theta_8 \\ 1-\theta_5 & 1-\theta_6 & 1-\theta_7 & 1-\theta_8 \end{bmatrix}.$$

Potential interactions can be treated in the same way as known interactions as long as all of them could potentially be simultaneously true. If there are two alternative hypotheses and the question is which fits better to the data, then this can be done by introducing a constraint on the parameters θ.

Example 2. Assume that X_1 is influenced either by X_2 or by U_1, this could be ensured by imposing the constraint

$$\lambda(\theta_1-\theta_2)\cdot(\theta_3-\theta_4)+(1-\lambda)(\theta_1-\theta_3)\cdot(\theta_2-\theta_4)=0,\ \ \lambda\in\{0,1\}, \tag{9}$$

4.2. Canalizing Boolean Functions

The concept of "canalizing" values in Boolean functions was introduced in developmental biology in 1940s (Waddington, 1942). The idea is, that one input is dominant and if it takes a certain value it determines the output. After that, in order to explain the phenomenon that absence of repressor or high levels of allolactose assures the operator cannot bind repressor in *lac operon* of the bacterium *Escherichia coli*, Kauffman applied this concept to BN modeling of gene regulatory networks (Kauffman, 1974).

Canalizing functions are defined as follows.

Definition 1. A Boolean function $f:\mathcal{D}^n \xrightarrow{f} \mathcal{D}$ is canalizing if there exist a variable $X_i, i\in\{1,2,\cdots,n\}$ and a Boolean function $g(X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n)$ and $a,b\in\mathcal{D}$, such that

$$f(X_1,\cdots,X_n)=\begin{cases} b, & \text{if } X_i=a,\\ g\neq b, & \text{if } X_i\neq a\end{cases} \tag{10}$$

where a is called the canalizing value for the variable X_i and b is the canalizing output value (Kauffman, 1974).

Based on Definition 1, this prior knowledge can be translated into imposing a specified value in the corresponding logical matrix. Assume that the logical matrix for the canalizing function (10) is denoted as S and the canalizing value a and canalizing output b can, respectively, be expressed in a vector form as δ_2^{2-a} and δ_2^{2-b}. Then, we can get the following result.

Theorem 1. *Given a canalizing function (10). The corresponding logical matrix $S\in\mathcal{L}_{2\times 2^n}$ satisfies*

$$SW_{[2,2^{i-1}]}\delta_2^{2-a}=\delta_2[\underbrace{2-b\ \ 2-b\ \cdots\ 2-b}_{2^{n-1}}]. \tag{11}$$

where $W_{[2,2^{i-1}]}$ is the swap matrix.

Proof: According to Lemma 1, it is easy to obtain $Sx_1x_2\cdots x_n = SW_{[2,2^{i-1}]}x_ix_1x_2\cdots x_{i-1}x_{i+1}\cdots x_n$. Applying (11), we have

$$\begin{aligned}&SW_{[2,2^{i-1}]}\delta_2^{2-a}x_1x_2\cdots x_{i-1}x_{i+1}\cdots x_n\\ &=\delta_2[\underbrace{2-b\ \ 2-b\ \cdots\ 2-b}_{2^{n-1}}]x_1x_2\cdots x_{i-1}x_{i+1}\cdots x_n=\delta_2^{2-b}\end{aligned}$$

which corresponds to $f(X_1,\cdots,X_{i-1},a,X_{i+1},\cdots,X_n)=b$ for any $X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n\in\{0,1\}$.

Let's take an example to illustrate the result of Theorem 1.

Example 3. Consider the BCN (7). Assume that the Boolean function f_1 is a canalizing function in x_2 for a canalizing value δ_2^2 and the corresponding canalizing output is δ_2^1. Due to the canalizing property, the logical matrix S_1 can be reduced to

$$S_1W_{[2,2]}\delta_2^2=\begin{bmatrix}1&1\\0&0\end{bmatrix}\Rightarrow S_1=\begin{bmatrix}\theta_1&1&\theta_3&1\\1-\theta_1&0&1-\theta_3&0\end{bmatrix}.$$

It can be checked that $S_1u_1\delta_2^2=\delta_2^1$, no matter whether $u_1=\delta_2^1$ or $u_1=\delta_2^2$. Note that the logical matrix S_1 contains only two binary parameters (i.e., θ_1 and θ_3). It shows that using canalizing property can reduce the number of binary parameters.

As an important subclass of canalizing function, k-canalizing function is defined as follows.

Definition 2. Let σ be a permutation on the set $\{1,2,\cdots,n\}$. A Boolean function $f:\mathcal{D}^n \xrightarrow{f} \mathcal{D}$ is k-canalizing in the variable order $X_{\sigma(1)},X_{\sigma(2)},\cdots,X_{\sigma(k)}$ with canalizing input values $a_1,a_2,\cdots,a_k$ and canalizing output values $b_1,b_2,\cdots,b_k$, if it can be represented in the form (Kauffman et al., 2003).

$$f(X_1,\cdots,X_n)=\begin{cases} b_1, & \text{if } X_{\sigma(1)}=a_1,\\ b_2, & \text{if } X_{\sigma(1)}\neq a_1, X_{\sigma(2)}=a_2,\\ \vdots & \\ b_k, & \text{if } X_{\sigma(1)}\neq a_1, X_{\sigma(2)}\neq a_2\cdots\\ & \quad X_{\sigma(k)}=a_k,\\ g\neq b_k, & \text{if } X_{\sigma(1)}\neq a_1, X_{\sigma(2)}\neq a_2\cdots\\ & \quad X_{\sigma(k)}\neq a_k.\end{cases} \tag{12}$$

Note that if all variables have certain canalizing values, then the function is called *nested canalizing function* (Kauffman et al., 2003).

As a Boolean variable can only take two values, i.e., $\{0,1\}$, (12) can be equivalently expressed as $f(X_1,\cdots,X_n)=b_i$, if $X_{\sigma(1)}=1-a_1, X_{\sigma(2)}=1-a_2,\cdots,X_{\sigma(i)}=a_k, i=1,2,\cdots,k$. Using the Boolean variables $[X_{\sigma(1)}\,X_{\sigma(2)}\cdots X_{\sigma(i)}]^{\mathrm{T}}$ to represent a multi-valued logic variable, it is straightforward to recognize that a k-canalizing function can be equivalently formulated as a canalizing function in a multi-valued logic variable. Therefore, Theorem 1 can be applied to specify the logical matrix for k-canalizing or nested canalizing function (12).

It is necessary to point out that different from the approaches proposed in Breindl et al. (2013) and Faisal et al. (2010), some binary parameters can be directly inferred, no matter which canalizing value the canalizing variable takes.

Example 4. Consider the BCN (7). Assume that the Boolean function f_2 is nested canalizing function, which can be represented as

$$f_2(U_2,X_1)=\begin{cases}1, & \text{if } U_2=1,\\ 0, & \text{if } U_2\neq 1, X_1=1.\end{cases}$$

Because $f_2(1, X_1) = 1$ for $X_1 \in \{0, 1\}$, we have

$$S_2\delta_2^1 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow S_2 = \begin{bmatrix} 1 & 1 & \theta_7 & \theta_8 \\ 0 & 0 & 1-\theta_7 & 1-\theta_8 \end{bmatrix}.$$

Moreover, due to $f_2(0, 1) = 0$, there is

$$S_2\delta_2^2\delta_2^1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow S_2 = \begin{bmatrix} 1 & 1 & 0 & \theta_8 \\ 0 & 0 & 1 & 1-\theta_8 \end{bmatrix}.$$

Remark 1. *Theorem 1 implies that considering canalizing property of a Boolean function, the corresponding logical matrix can be expressed with fewer binary parameters. For instance, if a Boolean function $f(X_1, X_2, \cdots, X_n)$ is a k-canalizing function, then 2^{n-k} different binary parameters are enough to represent the corresponding logical matrix.*

4.3. Unate Boolean Functions

The behavior of some substances or genes are well studied and it is known that they act as suppressing or activating in all reactions they are involved. If they always act inhibiting they have the so called negative unateness property. For the case that a quantity exclusively induces the expression of another gene or substance it has the positive unateness property (Porreca et al., 2010).

For the mathematical modeling of the unatess properties let us consider another important type of Boolean functions, which is called the unate function (Breindl et al., 2013).

Definition 3. (Breindl et al., 2013) A Boolean function $f : \mathcal{D}^n \xrightarrow{f} \mathcal{D}$ is unate in x_i, if for any $[X_1\ X_2 \cdots X_{i-1}\ X_{i+1} \cdots X_n]^{\mathrm{T}} \in \mathcal{D}^{n-1}$ it holds for positive unateness that

$$f(\cdots, X_{i-1}, 0, X_{i+1}, \cdots) \leq f(\cdots, X_{i-1}, 1, X_{i+1}, \cdots) \tag{13}$$

or it always holds for negative unateness that

$$f(\cdots, X_{i-1}, 0, X_{i+1}, \cdots) \geq f(\cdots, X_{i-1}, 1, X_{i+1}, \cdots) \tag{14}$$

In the same way as Breindl et al. (2013), unateness can be equivalently represented as linear formulation. Afterwards, this linear formulation can be seen as additional inequality constraints in the optimization problem. As Boolean function can be rewritten as a vector form (4) and according to Lemma 1, there is

$$Sx_1x_2 \cdots x_{i-1}x_ix_{i+1} \cdots x_n = SW_{[2,2^{i-1}]}x_ix_1x_2 \cdots x_{i-1}x_{i+1} \cdots x_n \tag{15}$$

where S is the logical matrix corresponding to the Boolean function f. Hence, $f(\cdots, X_{i-1}, 0, X_{i+1}, \cdots)$ and $f(\cdots, X_{i-1}, 1, X_{i+1}, \cdots)$ can, respectively, be represented in a vector form as

$$SW_{[2,2^{i-1}]}\delta_2^2x_1x_2 \cdots x_{i-1}x_{i+1} \cdots x_n \tag{16}$$

and

$$SW_{[2,2^{i-1}]}\delta_2^1x_1x_2 \cdots x_{i-1}x_{i+1} \cdots x_n \tag{17}$$

Furthermore, based on the vector form of Boolean variable (3) and according to (13) or (14), for each $x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n \in \Delta_2$ an inequality can be set up. Putting all inequality constraints together, we can find a matrix A for the following expression.

$$A \cdot \boldsymbol{\theta} \leq \mathbf{0}_n \tag{18}$$

Example 5. Consider the Boolean function $x_1 = f_1(x_2)$, this function f_1 is defined by two unknown parameters θ_1 and θ_2 . Assume that the Boolean function f_1 is a unate function with respect to x_2, which satisfies (13). As the first step, the matrix $S_1\delta_2^1$ and $S_1\delta_2^2$ are calculated, which yields

$$S_1\delta_2^1 = \begin{bmatrix} \theta_1 \\ 1-\theta_1 \end{bmatrix},\ S_1\delta_2^2 = \begin{bmatrix} \theta_2 \\ 1-\theta_2 \end{bmatrix}.$$

Then, the inequality constraint is

$$\theta_2 \leq \theta_1 \Longleftrightarrow \begin{bmatrix} -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \leq 0.$$

4.4. Known Attractors or Limit Cycles

When the BCN is not perturbed for a sufficiently long time it reaches the steady state. The steady state of a BCN can be exactly one state (i.e., attractor) or a fix cycle of some states (i.e., limit cycle). Attractors or limit cycles are assumed to determine the phenotype in the cell differentiation (Huang and Ingber, 2000). The experimental setup to measure the steady state of a system is simpler and measurements are easier to reproduce compared with transient dynamics. As a result, the steady state of the BN is often already known when the perturbation experiments for identification of the transient behavior are carried out. This knowledge can be utilized as follows.

An attractor corresponds to a self loop in the reachability graph. For a given input combination this fixes one specific coulumn in the matrix L. For the constant input $u(t) = \delta_{2^m}^i$ and the constant state $x(t) = \delta_{2^n}^j$ the k-th column is known to be $Col_k(L) = \delta_{2^n}^j$ with $k = (i-1)2^n + j$. A limit cycle can be analyzed in a similar manner. For the given state sequence of the limit cycle of length T and the constant input $u(t) = \delta_{2^m}^i$ one can calculate T columns of L. For each time instant t of the cycle the actual state $x(t) = \delta_{2^n}^j$ and the next state $x(t+1) = \delta_{2^n}^w$ is known. The information of this known transition is used by setting the k-th column to $Col_k(L) = \delta_{2^n}^w$ with $k = (i-1)2^n + j$.

5. IDENTIFICATION APPROACH

In this part, the identification problem of BCNs will be studied. At first, it will be shown that the identification problem can be reformulated as a nonlinear pseudo-Boolean optimization problem by applying the idea of the prediction error method.

The pseudo-Boolean optimization can be transformed into an equivalent linear binary integer programming problem that can be solved more efficiently. Then, we give a way to deal with missing measurement values. Finally, we discuss how dependencies between measured substances can be handled.

5.1. Optimization Problem

The prediction error method (PEM) is one of the most widely used identification methods (Isermann and Münchhof, 2011). The basic idea behind this method is to choose parameters to make the difference between a prediction based on the model and the measured values as small as possible. As the PEM minimizes the prediction error in the identified system, errors in the data set due to noise need no special treatment. Obviously the more noise is expected in the data set the more data should be acquired for identification of a reliable model.

Before applying PEM, it is necessary to specify a measure of prediction error. In information theory, the Hamming distance $d(X, Y)$ between two vectors $X, Y \in \mathcal{D}^n$ is defined as the number of positions, in which the entries differ (Hamming, 1950).

$$d(X, Y) = |\{j \in \{1, 2, \cdots, n\} | \, [X]_j \neq [Y]_j\}| \tag{19}$$

As each entry in the vectors X and Y belongs to the Boolean domain $\{0, 1\}$, (19) can be equivalently written as

$$d(X, Y) = \sum_{i=1}^{n} \left| [X]_i - [Y]_i \right| \tag{20}$$

Furthermore, let x_i, y_i be, respectively, the vector form of $[X]_i$ and $[Y]_i$. Then, it is straightforward to get

$$\left| [X]_i - [Y]_i \right| = 1 - x_i^{\mathrm{T}} \cdot y_i \tag{21}$$

Based on this, the Hamming distance $d(X, Y)$ can be rewritten as

$$d(X, Y) = \sum_{i=1}^{n} \left(1 - x_i^{\mathrm{T}} \cdot y_i \right) \tag{22}$$

Assume that the observed input and state data is $\{(U(t), X(t)), t = 0, 1, \cdots, T\}$. The vector form of the input data $\{U_1(t), U_2(t), \cdots, U_m(t)\}$ and state data $\{X_1(t), X_2(t), \cdots, X_n(t)\}$ are denoted, respectively, as $u_1(t), u_2(t), \cdots, u_m(t)$ and $x_1(t), x_2(t), \cdots, x_n(t)$. Since the logical matrix S_i for the state variable X_i can be represented by the parameter vector $\boldsymbol{\theta}$, we simply denote them as $S_i(\boldsymbol{\theta})$. Suppose that the state variable X_i can be influenced by the variables $X_{j_1}, X_{j_2}, \cdots, X_{j_k}$. According to (5), it is easy to get expression of the prediction $\hat{x}_i(\boldsymbol{\theta}, t)$:

$$\hat{x}_i(\boldsymbol{\theta}, t) = S_i(\boldsymbol{\theta}) u(t-1) \ltimes_{i=1}^{k} x_{j_i}(t-1) \tag{23}$$

Recalling (21) and (22), the PEM method will estimate the binary parameters by minimizing the prediction error, i.e.,

$$\min_{\boldsymbol{\theta} \in \mathcal{D}^k} \sum_{t=0}^{T} d(X(t), \hat{X}(\boldsymbol{\theta}, t)) = \min_{\boldsymbol{\theta} \in \mathcal{D}^k} \sum_{t=0}^{T} \sum_{i=1}^{n} \left(1 - x_i^{\mathrm{T}}(t) \cdot \hat{x}_i(\boldsymbol{\theta}, t) \right) \tag{24}$$

Furthermore, the optimization problem (24) can be equivalently rewritten as

$$\min_{\boldsymbol{\theta} \in \mathcal{D}^k} \left(T \cdot n - \sum_{t=0}^{T} \sum_{i=1}^{n} x_i^{\mathrm{T}}(t) \cdot \hat{x}_i(\boldsymbol{\theta}, t) \right)$$

which is actually equivalent to

$$\max_{\boldsymbol{\theta} \in \mathcal{D}^k} \sum_{t=0}^{T} \sum_{i=1}^{n} x_i^{\mathrm{T}}(t) \cdot \hat{x}_i(\boldsymbol{\theta}, t) \tag{25}$$

Next, it will be shown that the optimization problem (25) can be formulated as a pseudo-Boolean optimization (i.e., optimization of pseudo-Boolean functions). A pseudo-Boolean function is a mapping from a finite number of Boolean variables to a real number and can be uniquely represented by a multi-linear polynomial (Boros and Hammer, 2002).

As mentioned before, any logical matrix can be expressed by a multi-linear polynomial. After calculation, the term $\sum_{t=0}^{T} \sum_{i=1}^{n} x_i^{\mathrm{T}}(t) \hat{x}_i(\boldsymbol{\theta}, t)$ can be represented by a multivariate polynomial.

$$P_{mv}(\boldsymbol{\theta}) = c + \sum_{\mathcal{Q}_\beta \subset V} c_{\mathcal{Q}_\beta} \prod_{j \in \mathcal{Q}_\beta} \theta_j^{r_{\mathcal{Q}_\beta j}} \tag{26}$$

where $c, c_{\mathcal{Q}_\beta} \in \mathbb{R}$ for $\mathcal{Q}_\beta \subset V = \{1, 2, \cdots, k\}$ and the factor $r_{\mathcal{Q}_\beta, j}, \forall \beta, j$ is a natural number. In addition, using the property of Boolean variables $\theta_i^r = \theta_i, \forall r \in \mathbb{Z}_+$, the multivariate polynomial (26) is easily transformed into a multi-linear polynomial. Consequently, the term $\sum_{t=0}^{T} \sum_{i=1}^{n} x_i^{\mathrm{T}}(t) \cdot \hat{x}_i(\boldsymbol{\theta}, t)$ can be described by a multi-linear polynomial (6) and the optimization problem (25) is transformed into a pseudo-Boolean optimization problem

$$\max_{\boldsymbol{\theta} \in \mathcal{D}^k} P_{ml}(\boldsymbol{\theta}) = \max_{\boldsymbol{\theta} \in \mathcal{D}^k} c + \sum_{i=1}^{k} c_i \theta_i + \sum_{\alpha=1}^{q} c_{\mathcal{I}_\alpha} \prod_{j \in \mathcal{I}_\alpha} \theta_j \tag{27}$$

So far, several different ways to handle the nonlinear pseudo-Boolean optimization problems (27) exist, such as reduction to an equivalent linear or quadratic binary programming problem, branch-and-bound method, linear approximations (Boros and Hammer, 2002; Crama and Rodrí-guez-Heck, 2017). As the linear programming relaxation of an integer linear program can be solved efficiently and based on the solution integer solutions can be found, in this paper we consider "linearization", so that nonlinear binary optimization can be reduced to integer linear program (Crama and Rodrí-guez-Heck, 2017). The key is to introduce auxiliary Boolean variables $\boldsymbol{z} = [z_1 \;\; z_2 \;\; \cdots]^{\mathrm{T}}$ to replace the nonlinear monomial $\prod_{j \in \mathcal{I}_\alpha} \theta_j$ in (6) by means of the AND-expression $z_\alpha = \prod_{j \in \mathcal{I}_\alpha} \theta_j$. Simultaneously to satisfy the AND-expression, linear inequalities as constraints are considered

to get feasible value of the nonlinear monomial $\prod_{j\in\mathcal{I}_\alpha}\theta_j$. Finally, an optimization problem equivalent to (27) is obtained as

$$\begin{aligned}\max_{\boldsymbol{\theta},\boldsymbol{z}} L_P(\boldsymbol{\theta},\boldsymbol{z}) &= \max_{\boldsymbol{\theta},\boldsymbol{z}} c + \sum_{i=1}^{k} c_i\theta_i + \sum_{\alpha} c_{\mathcal{I}_\alpha} z_\alpha \\ s.t. \quad & z_\alpha \leq \theta_j, \forall j \in \mathcal{I}_\alpha, \\ & z_\alpha \geq \sum_{j\in\mathcal{I}_\alpha}\theta_j - (|\mathcal{I}_\alpha| - 1), \\ & z_\alpha \in \mathcal{D},\ \boldsymbol{\theta} \in \mathcal{D}^k.\end{aligned} \tag{28}$$

The constraints in the optimization problem in (27) can be complemented by additional constraints representing the prior knowledge of alternative hypotheses or unateness as shown in Section 4.1 and Section 4.3, respectively.

Remark 2. *It is important to note that minimizing or maximizing a pseudo-Boolean function is known to be $\mathcal{NP}$-hard (Crama and Rodrí-guez-Heck, 2017). However, Breindl et al. (2013) shows that the optimization problem (28) can be solved using a relaxed problem, i.e., linear programming solver based on the simplex method, which requires less computational effort than mixed integer linear program. The relaxed problem delivers an integer as optimal solution, which is also an optimal solution of the optimization problem (28).*

5.2. Handling of Large Scale Networks

With modern measurement techniques it is possible to quantify a huge amount of substances simultaneously. A Boolean network which describes the observed interactions is then also of large scale. But the number of substances which are direct relevant for the regulation of certain substance is usually limited, in other words the connectivity inside the network is bounded. For instance, as pointed out by Arnone and Davidson (1997), the connectivity is bounded by 8. Without prior knowledge the complexity of the algorithm is $\mathcal{O} = 2^{n+m}$ as all state and input combinations have to be considered as potential regulators for all states, even though only some of them are relevant in the end. This would limit the applicability of the approach to rather small networks. If one has hypotheses about potential interaction partners and the number of potential regulators per state is limited by a set of k variables, then the complexity of the algorithm is $\mathcal{O} = 2^k$, as the regulative functions for each state can be inferred separately. The hypotheses for the interaction partners are not necessarily based on prior-knowledge, but could also be computed based on the data set. In Margolin et al. (2006) an approach is presented, which is based on the information theoretic concept of mutual information ranking and the restriction to pairwise interactions that leads to a very good scaling with big data sets.

5.3. Handling of Missing Measurement Values

Dependent on the measurement technique it is sometimes not possible to measure all states at all time instances and the missing values must be handled in the data analysis. There are approaches in the literature to compute an imputation e.g., for microarrays in Gan et al. (2006) and gel-based proteomics in Albrecht et al. (2010). These approaches are based on interpolation or heuristics. An alternative is to use a data analysis approach which can deal with incomplete data matrices.

A missing measurement value can be estimated during the identification by adding additional binary parameters in the identification process. Because of vector expression of states, all possible states belong to the set Δ_{2^n}. In this way, n binary parameters are enough for vector expression of a completely unknown state at time k. For example, if $n = 2$, then we can generally express the unknown state as

$$x(k) = \begin{bmatrix} \gamma_1 \cdot \gamma_2 \\ \gamma_1 \cdot (1-\gamma_2) \\ (1-\gamma_1) \cdot \gamma_2 \\ (1-\gamma_1) \cdot (1-\gamma_2) \end{bmatrix}. \tag{29}$$

Furthermore, as the states of the system are known partially, then the number of binary parameters can be reduced accordingly. So for each missing value one parameter is added to the optimization and the imputation for this value is calculated which fits best to the other dynamic behavior of the system.

5.4. Handling of Unmeasurable Processes

In some systems post transcriptional protein-protein interactions induce dependencies between the measured abundances similar to the transcriptional regulation. This leads to the situation that the transcriptional regulation can not be observed directly and the identification procedure needs to be adapted accordingly (Geier et al., 2007). The dependencies between the states and the measured outputs can be included in boolean models easily by adding Boolean functions mapping from the actual stats $X(t)$ to the measured outputs $Y(t)$:

$$Y_j(t) = h_j(X(t)), \qquad j = 1, 2, \ldots, p \tag{30}$$

where $[Y(t) = Y_1(t)\ Y_2(t)\ \ldots\ Y_p(t)]^T \in \mathcal{D}^p$ is the output vector at time t , h_i are logic functions. All structural information on the logic functions can be expressed with a logical matrix H

$$y(t) = Hx(t) \tag{31}$$

which can be derived analogous to Equations (2–5). All approaches presented in this paper can be extended for the BN model with output mapping. As additional logic functions are to be identified, additional unknown parameters are added and these parameters cannot be separately identified from the parameters of the regulative functions, which impacts the computational burden drastically (Zhang et al., 2017b) .

5.5. Influence of Noise

In real world experiments measurement noise is unavoidable. With a sophisticated binarization method the influence of additive noise can often be suppressed (Hopfensitz et al., 2012). But noise can still lead to wrong binarized values in some cases and consequently errors in the input to the identification method

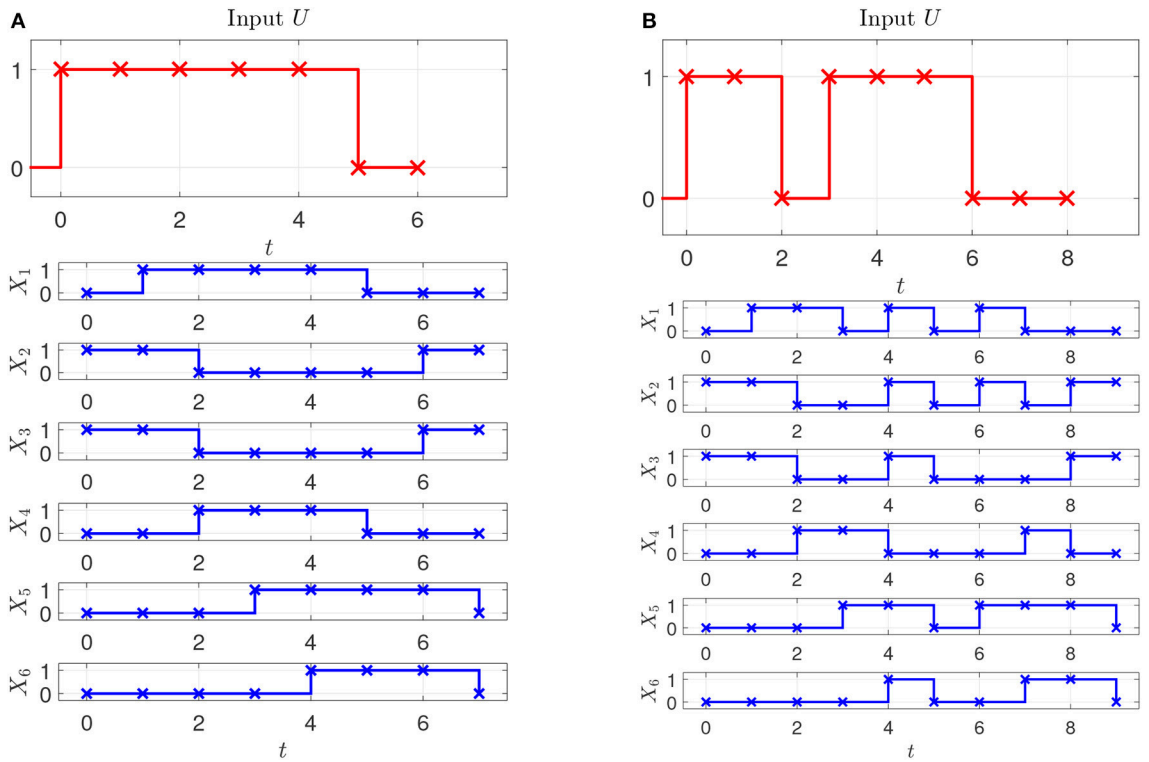

FIGURE 2 | Perturbation and state measurement. **(A)** First experiment. **(B)** Second experiment.

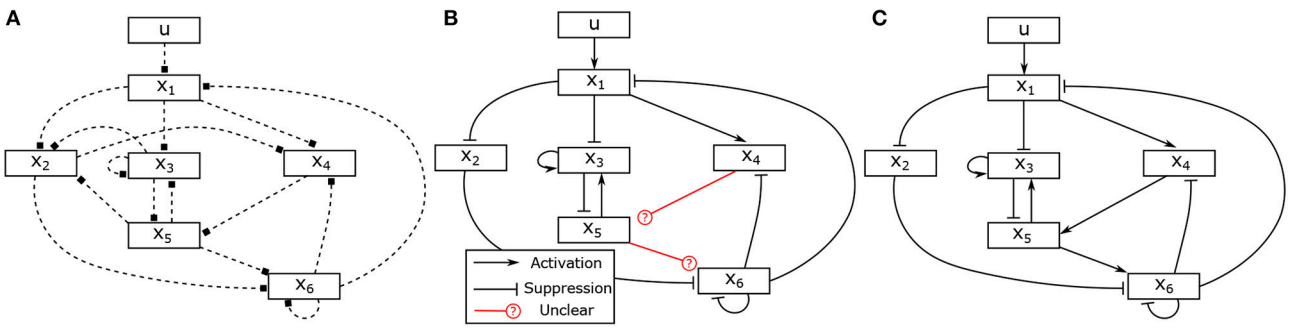

FIGURE 3 | Hypothesis, partially identified and fully identified network graph. **(A)** Hypotheses for regulative interactions. **(B)** Identified Boolean network without canalizing information. **(C)** Identified Boolean network with prior knowlege.

cannot be totally avoided. As the presented approach is based on an optimization, the network which optimally fits to the observed data is found. Inconsistent transitions caused by noise in the data set can be handled directly and lead to an identification result with a non-zero prediction error. If, due to noise, the observed transitions would lead to an identification result which is contradictory to prior knowledge, the identification approach ignores these transitions directly.

Example 6. Consider the BCN for oxidative stress response pathways with the PI3-Kinase-Akt pathway given in Sridharan et al. (2012).

$$\begin{cases} X_1(t+1) &= U(t) \wedge \neg X_6(t) \\ X_2(t+1) &= \neg X_1(t) \\ X_3(t+1) &= \neg X_1(t) \wedge (X_5(t) \vee X_3(t) \\ X_4(t+1) &= X_1(t) \wedge \neg X_6(t) \\ X_5(t+1) &= X_4(t) \vee \neg X_3(t) \\ X_6(t+1) &= X_5(t) \wedge (\neg X_6(t) \vee \neg X_2(t)) \end{cases} \tag{32}$$

In the model, X_1 represents stress reactive intermediaries, X_2 transcription factor A, X_3 key protein, X_4 protein kinase, X_5

transcription factor B, X_6 anti-stress response element, U stress signal. Using STP, (32) can be converted into the algebraic form (5) with $x(t) = \ltimes_{i=1}^{6} x_i(t) \in \Delta_{64}, u(t) \in \Delta_2$.

Assume that two experiments have been executed starting in steady state with two different stimuli, the corresponding input-state data is obtained as shown in **Figures 2A,B**. Assume further that as prior knowledge the candidates of regulative interactions (see the dashed lines in **Figure 3A**) and the attractor are given. The attractor of the BCN without stress is $X_1 = 0,\ X_2 = 1,\ X_3 = 1, X_4 = 0,\ X_5 = 0,\ X_6 = 0$.

Based on the candidates of regulative interactions, the number of unknown binary parameters $\boldsymbol{\theta}$ representing the logical matrices of the Boolean functions can be reduced from $6 \cdot 2^7 = 768$ to 40 as described in Section 4.1. For instance, since the variable X_2 is connected with the variables X_1, X_3 and X_5, it means that the Boolean function of the variable X_2 can be described by $f_2(X_1, X_5, X_5)$. Accordingly, 8 binary parameters are enough to represent the logical matrix S_2 of the Boolean function f_2, i.e.,

$$S_2 = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 & \theta_6 & \theta_7 & \theta_8 \\ 1-\theta_1 & 1-\theta_2 & 1-\theta_3 & 1-\theta_4 & 1-\theta_5 & 1-\theta_6 & 1-\theta_7 & 1-\theta_8 \end{bmatrix}. \tag{33}$$

The information about the steady state is used as described in Section 4.4 to determine one parameter in each matrix, which reduces the number of unknown variables to 34. In the next, we apply the proposed approach to identify the model of the BCN from the given input-state data. Solving the optimization problem (28), in total, 31 unknown binary parameters can be determined. The identification result is depicted in **Figure 3B** and the identified matrices are as follows,

$$\begin{aligned} S_1 &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix}, \qquad S_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \\ S_3 &= \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \\ S_4 &= \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}, \ S_5 = \begin{bmatrix} \theta_{29} & 0 & 1 & 1 \\ 1-\theta_{29} & 1 & 0 & 0 \end{bmatrix}, \\ S_6 &= \begin{bmatrix} 0 & 1 & \theta_{35} & 0 & 1 & 1 & \theta_{39} & 0 \\ 1 & 0 & 1-\theta_{35} & 1 & 0 & 0 & 1-\theta_{39} & 1 \end{bmatrix}. \end{aligned} \tag{34}$$

It can be seen that the logical matrices of the Boolean functions for X_5 and X_6 can not be uniquely determined. Combined with an additional information about activating or suppressing properties of the states, for instance, X_4 and X_5 are, respectively, activator to X_5 and X_6, the complete model can be uniquely reconstructed. The canalizing property of X_4 and X_5 can be utilized as described in Section 4.2. If this information is not available, one could conduct additional experiments with different stimuli and combine the data to have full reconstruction of the model as depicted in **Figure 3C**.

6. DISCUSSION

The proposed method facilitates the incorporation of various types of prior knowledge. The optimization problem can be solved by efficient linear programming solvers. By using the simplex method one can guarantee to find the network which optimally fits to the observed data. In comparison, the genetic algorithms based approaches may not guarantee the optimal solution. The proposed method is developed for synchronous Boolean networks. It can be applied to large scale networks, if the connectivity of the network to be identified is limited with aid of prior knowledge or application of information theory.

In future we plan to investigate data-based approaches to infer the connections in large networks and automated partitioning into smaller subsystems (e.g., with an adapted approach from discrete event systems like Saives et al., 2018). We also work on a new method for the binarization based on the idea that the qualitative system behavior before and after the binarization shall be the same.

AUTHOR CONTRIBUTIONS

TL, ZZ, and PZ conception and design of research. TL and ZZ performed simulation and analyzed data. TL, ZZ, and PZ interpreted simulation results. TL and ZZ prepared figures. TL, ZZ, and PZ drafted manuscript and approved final version of manuscript.

FUNDING

This work is supported by the Federal State of Rhineland-Palatinate, Germany in the framework of the project Complex Data Analysis in Life Sciences and Biotechnology (*BioComp*).

ACKNOWLEDGMENTS

The authors would like to thank Michael Schroda and Timo Mühlhaus for the discussion on the data characteristics in biological systems.

REFERENCES

Akutsu, T., Miyano, S., and Kuhara, S. (1999). "Identification of genetic networks from a small number of gene expression patterns under the boolean network model," in *Proceedings of the Pacific Symposium on Biocomputing* (Mauna Lani), 17–28.

Albert, R., and Barabási, A.-L. (2000). Dynamics of complex systems: Scaling laws for the period of boolean networks. *Phys. Rev. Lett.* 84, 5660–5663. doi: 10.1103/PhysRevLett.84.5660

Albrecht, D., Kniemeyer, O., Brakhage, A. A., and Guthke, R. (2010). Missing values in gel-based proteomics. *Proteomics* 10, 1202–1211. doi: 10.1002/pmic.200800576

Alon, N., Babai, L., and Suzuki, H. (1991). Multilinear polynomials and frankl-ray-chaudhuri-wilson type intersection theorems. *J. Combinat. Theory A* 58, 165–180. doi: 10.1016/0097-3165(91)90058-O

Arnone, M., and Davidson, E. (1997). The hardwiring of development: organization and function of genomic regulatory systems. *Development* 124, 1851–1864.

Berestovsky, N., and Nakhleh, L. (2013). An evaluation of methods for inferring boolean networks from time-series data. *PLoS ONE* 8:e66031. doi: 10.1371/journal.pone.0066031

Bollobas, B. (2012). *Graph Theory: An Introductory Course*, Vol. 63. New York, NY: Springer Science & Business Media.

Boros, E., and Hammer, P. L. (2002). Pseudo-boolean optimization. *Discrete Appl. Math.* 123, 155–225. doi: 10.1016/S0166-218X(01)00341-9

Breindl, C., Chaves, M., and Allgöwer, F. (2013). "A linear reformulation of boolean optimization problems and structure identification of gene regulation networks," in *Proceedings of the 52th IEEE Conference on Decision and Control* (Florence), 733–738.

Cheng, D. (2001). Semi-tensor product of matrices and its application to morgen's problem. *Sci. China Ser. Informat. Sci.* 2001, 195–212. doi: 10.1007/BF02714570

Cheng, D., and Qi, H. (2010). A linear representation of dynamics of boolean networks. *IEEE Trans. Automat. Cont.* 55, 2251–2258. doi: 10.1109/TAC.2010.2043294

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

Cheng, D., Qi, H., and Li, Z. (2011b). Model construction of boolean network via observed data. *IEEE Trans. Neural Netw.* 22, 525–536. doi: 10.1109/TNN.2011.2106512

Cheng, D., and Zhao, Y. (2011). Identification of boolean control networks. *Automatica* 47, 702–710. doi: 10.1016/j.automatica.2011.01.083

Crama, Y., and Rodrí-guez-Heck, E. (2017). A class of valid inequalities for multilinear 0-1 optimization problems. *Discrete Optimizat.* 25, 28–47. doi: 10.1016/j.disopt.2017.02.001

Davidich, M. I., and Bornholdt, S. (2008). Boolean network model predicts cell cycle sequence of fission yeast. *PLoS ONE* 3:e1672. doi: 10.1371/journal.pone.0001672

Dorier, J., Crespo, I., Niknejad, A., Liechti, R., Ebeling, M., and Xenarios, I. (2016). Boolean regulatory network reconstruction using literature based knowledge with a genetic algorithm optimization method. *BMC Bioinform.* 17:410. doi: 10.1186/s12859-016-1287-z

Faisal, S., Lichtenberg, G., Trump, S., and Attinger, S. (2010). Structural properties of continuous representations of boolean functions for gene network modelling. *Automatica* 46, 2047–2052. doi: 10.1016/j.automatica.2010.09.001

Fauré, A., Naldi, A., Chaouiya, C., and Thieffry, D. (2006). Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle. *Bioinformatics* 22, e124–e131. doi: 10.1093/bioinformatics/btl210

Fornasini, E., and Valcher, M. E. (2014). "Identification problems for boolean networks and boolean control networks," in *Proceedings of the 19th IFAC World Congress* (Cape Town), 5399–5404.

Fumia, H. F., and Martins, M. L. (2013). Boolean network model for cancer pathways: predicting carcinogenesis and targeted therapy outcomes. *PLoS ONE* 8:e69008. doi: 10.1371/journal.pone.0069008

Gan, X., Liew, A. W.-C., and Yan, H. (2006). Microarray missing data imputation based on a set theoretic framework and biological knowledge. *Nucleic Acids Res.* 34, 1608–1619. doi: 10.1093/nar/gkl047

Geier, F., Timmer, J., and Fleck, C. (2007). Reconstructing gene-regulatory networks from time series, knock-out data, and prior knowledge. *BMC Sys. Biol.* 1:11. doi: 10.1186/1752-0509-1-11

Grieb, M., Burkovski, A., Sträng, J. E., Kraus, J. M., Groß, A., Palm, G., et al. (2015). Predicting variabilities in cardiac gene expression with a boolean network incorporating uncertainty. *PLoS ONE* 10:e0131832. doi: 10.1371/journal.pone.0131832

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell Labs Techn. J.* 29, 147–160. doi: 10.1002/j.1538-7305.1950.tb00463.x

Higa, C. H., Louzada, V. H., Andrade, T. P., and Hashimoto, R. F. (2011). Constraint-based analysis of gene interactions using restricted boolean networks and time-series data. *BMC Proc.* 5:S5. doi: 10.1186/1753-6561-5-S2-S5

Hopfensitz, M., Müssel, C., Wawra, C., Maucher, M., Kühl, M., Neumann, H., and Kestler, H. A. (2012). Multiscale binarization of gene expression data for reconstructing boolean networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 9, 487–498. doi: 10.1109/TCBB.2011.62

Huang, S., and Ingber, D. E. (2000). Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Exp. Cell Res.* 261, 91–103. doi: 10.1006/excr.2000.5044

Isermann, R., and Münchhof, M. (2011). *Identification of Dynamic Systems: An Introduction With Applications*. Berlin;Heidelberg: Springer.

Karlebach, G., and Shamir, R. (2012). Constructing logical models of gene regulatory networks by integrating transcription factor-dna interactions with expression data: an entropy-based approach. *J. Comput. Biol.* 19, 30–41. doi: 10.1089/cmb.2011.0100

Kauffman, S. (1974). The large scale structure and dynamics of gene control circuits: an ensemble approach. *J. Theor. Biol.* 44, 167–190. doi: 10.1016/S0022-5193(74)80037-8

Kauffman, S., Peterson, C., Samuelsson, B., and Troein, C. (2003). Random boolean network models and the yeast transcriptional network. *Proc. Natl. Acad. Sci. U.S.A.* 100, 14796–14799. doi: 10.1073/pnas.2036429100

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 22, 437–467. doi: 10.1016/0022-5193(69)90015-0

Lähdesmäki, H., Shmulevich, I., and Yli-Harja, O. (2003). On learning gene regulatory networks under the boolean network model. *Mach. Learn.* 52, 147–167. doi: 10.1023/A:1023905711304

Liang, S., Fuhrman, S., and Somogyi, R. (1998). "Reveal: a general reverse engineering algorithm for inference of genetic network architectures," in *Proceedings of the Pacific Symposium on Biocomputing* (Hawaii), 18–29.

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., et al. (2006). Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformat.* 7:S7. doi: 10.1186/1471-2105-7-S1-S7

Naldi, A., Monteiro, P. T., Müssel, C., Consortium for Logical Models and Tools, Kestler, H. A., Thieffry, D., et al. (2015). Cooperative development of logical modelling standards and tools with colomoto. *Bioinformatics* 31, 1154–1159. doi: 10.1093/bioinformatics/btv013

Ostrowski, M., Paulevé, L., Schaub, T., Siegel, A., and Guziolowski, C. (2016). Boolean network identification from perturbation time series data combining dynamics abstraction and logic programming. *Biosystems* 149, 139–153. doi: 10.1016/j.biosystems.2016.07.009

Ouyang, H., Fang, J., Shen, L., Dougherty, E. R., and Liu, W. (2014). Learning restricted boolean network model by time-series data. *EURASIP J. Bioinform. Sys. Biol.* 2014:10. doi: 10.1186/s13637-014-0010-5

Porreca, R., Cinquemani, E., Lygeros, J., and Ferrari-Trecate, G. (2010). Identification of genetic network dynamics with unate structure. *Bioinformatics* 26, 1239–1245. doi: 10.1093/bioinformatics/btq120

Saives, J., Faraut, G., and Lesage, J. J. (2018). Automated partitioning of concurrent discrete-event systems for distributed behavioral identification. *IEEE Trans. Autom Sci. Eng.* 15, 832–841. doi: 10.1109/TASE.2017.2718244

Sridharan, S., Layek, R., Datta, A., and Venkatraj, J. (2012). Boolean modeling and fault diagnosis in oxidative stress response. *BMC Genomics* 13(Suppl. 6):S4. doi: 10.1186/1471-2164-13-S6-S4

Terfve, C., Cokelaer, T., Henriques, D., MacNamara, A., Goncalves, E., Morris, M. K., et al. (2012). Cellnoptr: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. *BMC Sys. Biol.* 6:133. doi: 10.1186/1752-0509-6-133

Videla, S., Guziolowski, C., Eduati, F., Thiele, S., Gebser, M., Nicolas, J., et al. (2015). Learning boolean logic models of signaling networks with asp. *Theor. Comput. Sci.* 599, 79–101. doi: 10.1016/j.tcs.2014.06.022

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565. doi: 10.1038/150563a0

Wang, R.-S., Saadatpour, A., and Albert, R. (2012). Boolean modeling in systems biology: an overview of methodology and applications. *Phys. Biol.* 9:055001. doi: 10.1088/1478-3975/9/5/055001

Zhang, X., Han, H., and Zhang, W. (2017a). Identification of boolean networks using premined network topology information. *IEEE Trans. Neural Netw. Learn. Sys.* 28, 464–469. doi: 10.1109/TNNLS.2016.2514841

Zhang, Z., Leifeld, T., and Zhang, P. (2017b). "Identification of boolean control networks incorporating prior knowledge," in *IEEE 56th Annual Conference on Decision and Control* (Melbourne, VIC), 5839–5844.

Zhou, X., Wang, X., and Dougherty, E. R. (2003). Binarization of microarray data on the basis of a mixture model. *Mol. Cancer Therapeut.* 2, 679–684.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.