Journal of Theoretical Biology 353 (2014) 170–178

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Refining network reconstruction based on functional reliability

Yunjun Zhang [a,*], Qi Ouyang [b], Zhi Geng [c]

[a] Faculty of Foundation Education, Peking University Health Science Center, 100191 Beijing, China
[b] Center for Theoretical Biology, School of Physics, Peking University
[c] School of Mathematical Sciences, Peking University, 100871, Beijing, China

ARTICLE INFO

Article history:
Received 31 January 2013
Received in revised form
25 February 2014
Accepted 28 February 2014
Available online 11 March 2014

Keywords:
Boolean model
Robustness
Reverse engineering

ABSTRACT

Reliable functioning is crucial for the survival and development of the genetic regulatory networks in living cells and organisms. This functional reliability is an important feature of the networks and reflects the structural features that have been embedded in the regulatory networks by evolution. In this paper, we integrate this reliability into network reconstruction. We introduce the concept of *dependency probability* to measure the dependency of functional reliability on network edges. We also propose a method to estimate the *dependency probability* and select edges with high contributions to functional reliability. We use two real examples, the regulatory network of the cell cycle of the budding yeast and that of the fission yeast, to demonstrate that the proposed method improves network reconstruction. In addition, the *dependency probability* is robust in calculation and can be easily implemented in practice.

1. Introduction

Genetic regulatory networks (GRNs) in living organism consist of a set of interacting biomolecules ('Nodes') and interactions among them ('Edges'). These networks are central to numerous cellular processes (Karlebach and Shamir, 2008). In system biology, the reconstruction of GRNs (or the reverse engineering of GRNs) has attracted great interest because it is important to understand the development, functioning, and pathology of various organisms.

Recent advances in genomics, especially high-throughput profiling of gene expression patterns with DNA microarrays, have provided large amounts of measurement data about the expression states of GRNs in regulatory processes (Li et al., 2004; Bornholdt, 2005; Davidich and Bornholdt, 2008; Wang et al., 2010). These expression data have made it possible to develop sophisticated reconstruction algorithms. Most expression-based reconstruction algorithms are based on statistical inference. These algorithms employ various forms of models to describe the biological mechanisms contributing to the regulatory relationships in the GRNs. For example, the method based on mutual-information describes the mechanism with the entropy of information (Irrthum et al., 2010), the differential equation method models the mechanism by differential equations (Wang et al., 2006), and the Bayesian network approach models the mechanism with the Bayesian rule (Friedman et al., 2000). There are a vast body of literature about the expression-based reconstruction methods, and we refer the reader to some latest review papers for more details (De Smet and Marchal, 2010; Emmert-Streib et al., 2012; Huang et al., 2009).

In expression-based reconstruction methods, a huge attention has been placed on transforming the given expression information into a set of constraint equations and solving these equations to obtain structural solutions, while the functional reliability of these solutions is largely overlooked. The functional reliability is defined as the ability to evolve reliably along a given sequence of states regardless of the intrinsic noises in biomolecular interactions. This ability is crucial for real GRNs because they must perform their functions reliably to ensure the survival and development of living cells and organisms (Klemm and Bornholdt, 2005). On the other hand, the structural solutions which can fulfill the same expression process may have a different reliability. And the true GRNs prefer the reliable solutions. For example, in Klemm and Bornholdt (2005), the authors compared the functional reliability of small motifs (3 nodes) fulfilling the same expression sequence. They found that reliable motifs are more likely to be employed by real networks. Moreover, the cell-cycle control network of budding yeast has been shown strong reliability (Braunewell and Bornholdt, 2007). Additionally in Braunewell and Bornholdt (2009), the authors simulated the evolution of network structure under the guidance of functional reliability.

In this paper, we integrate functional reliability into the network reconstruction of GRNs to refine the outcome of reconstruction algorithm based on expression data. Our study is based on the framework of a generalized Boolean network model with continuous timing and noise as in Braunewell and Bornholdt (2009). For a given biological function (a sequence of Boolean data), the previous reconstruction

* Corresponding author.
E-mail addresses: yjzhang@bjmu.edu.cn, yunjun.zhang2008@gmail.com (Y. Zhang).

http://dx.doi.org/10.1016/j.jtbi.2014.02.041
0022-5193

algorithm (Wang et al., 2010) can infer some network edges and also generates a huge number of similar edges (called interchangeable edges) which have equal contributions to the function. We introduce the concept of *dependency probability* to measure the dependency of functional reliability on these interchangeable edges. Based on this concept, we perform edge selection within the set of interchangeable edges. In two real examples of regulatory networks (the cell-cycle regulatory networks of budding yeast and fission yeast), we can identify 4 and 7 true edges (the proportions are 25 percent and 50 percent) with false discovery rates of 0.33 and 0, respectively.

The implications of this study are important for the following reasons: first of all, the *dependency probability* helps us to understand the structural origin of functional reliability. The *dependency probability* is also very robust in calculation and easy to implement. With the help of functional reliability, we can reveal more edges without further experimental observation. This will reduce experimental expense in practice. Finally, although the method of utilizing functional reliability is developed in the context of the Boolean model, the idea can be easily extended to other reverse engineering algorithms based on more sophisticated models.

2. Methods

2.1. Review of synchronous Boolean model and backbone motif

The synchronous Boolean model (or discrete Boolean model) was first introduced in Kauffman (1969). Suppose that there is a genetic regulatory network with N kinds of interacting molecules. The Boolean model assigns to every node (e.g., node i) a binary variable ($s_i(t) \in \{0, 1\}$) to denote whether the node is 'on' (active or highly expressed) or 'off' (inactive). The state of the network at time t is represented by a Boolean vector $\vec{S}(t) = (s_1(t), s_2(t)), \ldots, s_N(t))$. And the trajectory of network state is the collection of network states at a series of time points, denoted as $(\vec{S}(1), \vec{S}(2), \ldots, \vec{S}(T))$

The synchronous Boolean model assumes that the states of all nodes are updated simultaneously in discrete time steps according to the Boolean logic rule:

$$s_i(t+1) = \begin{cases} 1, & \sum_j a_{ji} s_j(t) > 0 \\ 0, & \sum_j a_{ji} s_j(t) < 0 \\ s_i(t), & \sum_j a_{ji} s_j(t) = 0 \end{cases} \quad (1)$$

where (a_{ji}) is an $N \times N$ matrix encoding the network structure. There are four kinds of regulatory effects (edges) between nodes: self-inhibition ($a_{ii} = -0.5$), self-activation ($a_{ii} = 1$), inhibition ($a_{ji} = -\infty$) and activation ($a_{ji} = 1$). Furthermore, under the Boolean logic assumptions (addition represents the Boolean operator OR, multiplication represents AND, the bar on a variable represents NOT), the four kinds of edges can be represented by four indicators: r_{ii}, g_{ii}, r_{ji} and g_{ji} respectively. These indicators are unknown structure parameters. We have the following Boolean equation:

$$s_i(t+1) = \left(\sum_{j \neq i} (s_j(t) g_{ji}) + s_i(t) \bar{r}_{ii} + \overline{s_i(t)} g_{ii} \right) \prod_{j \neq i} (\overline{s_j(t) r_{ji}}). \quad (2)$$

Given the state updating rule (1) or (2), it is possible to calculate the trajectory of network state starting from an initial state. For example, the biological process summarized in Table 1 can be deduced from this rule. On the other hand, given a process of network state, it is possible to infer the network structure by solving Eq. (2). In fact, node states at T time instants can be represented as $T-1$ Boolean equations of the structure parameters. It is possible to enumerate all the solutions of these equations for node i. This method is referred to as the Process-based network decomposition method by Wang et al. (2010).

As to the sixth node (node Swi5) in the budding yeast, for example, the changing process of its state is (0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0) and the corresponding solutions can be any of the following three equivalent sets: (r_{66}, g_{76}, $r_{10,6}$), (r_{66}, g_{76}, $r_{4,6}$) and (r_{66}, g_{76}, $r_{8,6}$). We refer to the edge r_{66} and g_{76} as the *rigid edge*. They appear in all solutions. On the other hand, the edges of $r_{10,6}$, $r_{4,6}$ and $r_{8,6}$ are defined as *interchangeable edges*. These *interchangeable edges* are necessary for all structural solutions but they can be substituted by each other. In reality, the edge $r_{10,6}$ does exist in the regulatory network, so $r_{10,6}$ is referred to as a real interchangeable edge.

For a given biological process, a backbone motif is composed of the structure solutions of all nodes, that is, a backbone motif contains all the rigid edges and one set of interchangeable edges for each node. So the total number of backbone motifs is the product of the number of interchangeable edge sets of all nodes. Therefore, there are 108,864 backbone motifs for the budding yeast and 49,152 backbone motifs for the fission yeast. All the backbone motifs (including the rigid and interchangeable edges) of the budding yeast is summarized in Table 2. All the edges in bold type are the real interchangeable edges, and the sets of real interchangeable edges in the first row (in bold type) form the real backbone motif (please see the Fig. 1 for the full network and the real backbone motif). For the example of the cell-cycle regulatory network of fission yeast, the biological process and the list of backbone motifs are presented in supplementary data.

2.2. Review of asynchronous Boolean model

To study the functional reliability of a GRN, Braunewell and Bornholdt (2007) introduced an asynchronous Boolean model. In this paper, we follow the same line to study functional reliability. An obvious benefit of this model is that it does not require the assumption of synchronous updating of all nodes. It allows for the continuous concentration variation of each node

Table 1
Budding yeast: *biological process*. The time course of the 11 nodes of the regulatory network, a representation of its biological function.

Time t	Cln3 s_1	MBF s_2	SBF s_3	Cln1,2 s_4	Cdh1 s_5	Swi5 s_6	Cdc20/14 s_7	Clb5,6 s_8	Sic1 s_9	Clb1,2 s_{10}	Mcm1/SFF s_{11}	Phase
0	1	0	0	0	1	0	0	0	1	0	0	START
1	0	1	1	0	1	0	0	0	1	0	0	G1
2	0	1	1	1	1	0	0	0	1	0	0	G1
3	0	1	1	1	0	0	0	0	0	0	0	G1
4	0	1	1	1	0	0	0	1	0	0	0	S
5	0	1	1	1	0	0	0	1	0	1	1	G2
6	0	0	0	1	0	0	1	1	0	1	1	M
7	0	0	0	0	0	0	1	0	0	0	1	M
8	0	0	0	0	1	1	1	0	1	0	0	M
9	0	0	0	0	1	1	0	0	1	0	0	G1
10	0	0	0	0	1	0	0	0	1	0	0	G1