Journal of Theoretical Biology 380 (2015) 9–15

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Minimum network constraint on reverse engineering to develop biological regulatory networks

Bin Shao [a], Jiayi Wu [b], Binghui Tian [a], Qi Ouyang [a,b,*]

[a] Center for Quantitative Biology and Peking-Tsinghua Center for Life Sciences at Peking University, Beijing 100871, China
[b] School of Physics and the State Key Laboratory for Artificial Microstructures and Mesoscopic Physics, Peking University, Beijing 100871, China

HIGHLIGHTS

- We propose a novel constraint, i.e., minimal network constraint, to facilitate the reconstruction of biological networks.
- Two scenarios were considered in the network reconstruction process: generating data using different initial conditions; and generating data from knock out and over-expression experiments.
- Feasibility of our constraint for uncovering biological networks may offer new clues of the design principle of biological regulatory networks.

ARTICLE INFO

Article history:
Received 18 September 2014
Received in revised form
23 April 2015
Accepted 4 May 2015
Available online 14 May 2015

Keywords:
Reverse engineering
Biological network
Minimum network
Boolean mathematics

ABSTRACT

Reconstructing the topological structure of biological regulatory networks from microarray expression data or data of protein expression profiles is one of major tasks in systems biology. In recent years, various mathematical methods have been developed to meet this task. Here, based on our previously reported reverse engineering method, we propose a new constraint, i.e., the minimum network constraint, to facilitate the reconstruction of biological networks. Three well studied regulatory networks (the budding yeast cell cycle network, the fission yeast cell cycle network, and the SOS network of *Escherichia coli*) were used as the test sets to verify the performance of this method. Numerical results show that the biological networks prefer to use the minimal networks to fulfill their functional tasks, making it possible to apply minimal network criteria in the network reconstruction process. Two scenarios were considered in the reconstruction process: generating data using different initial conditions; and generating data from knock out and over-expression experiments. In both cases, network structures are revealed faithfully in a few steps using our approach.

1. Introduction

Technological advances facilitate accumulation of various types of experimental data in biological processes. For example, the development of microarray technique enables researchers to get genome wide gene expression data of cells status under different conditions. The explosive growth of biological data put forward a challenge to uncover underlying regulatory networks, which belongs to the study of biological network reverse engineering (Bansal et al., 2007; D'haeseleer et al., 2000). Comparing to old manual way of network construction (Kaizu et al., 2010), development of the reverse engineering techniques enables researchers to systematically reconstruct biological regulatory networks in a rather high speed. Many algorithms with different computational complexity have been employed, including simple correlation-based method (Maucher et al., 2011), Boolean Network Model (Zhang et al., 2013; Martin et al., 2007; Haider and Pal, 2012) and Bayesian networks (Sachs ;et al., 2005; Friedman et al., 2000). Their performances were tested in different situations (Marbach et al., 2010, 2012).

On the basis of the reverse engineering, it is possible to rationally design a sequence of experiments to maximally extract the information of underlying networks (Zhang et al., 2013). In general, existing knowledge from literatures and former experiments may imply a large set of possible networks, so that new discrimination experiments are needed to generate additional information for reducing the number of possible networks. However, the information content in different experiments may vary a lot, how to design new experiments to get maximal information on network structure becomes a focus of systems biology.

* Corresponding author at: The State Key Laboratory for Artificial Microstructures and Mesoscopic Physics, School of Physics, Peking University, 100871 Beijing, China. Tel./fax: +861062756943.
E-mail address: qi@pku.edu.cn (Q. Ouyang).

http://dx.doi.org/10.1016/j.jtbi.2015.05.005
0022-5193/

Recently, under the assumption that all the possible networks have equal probability to be selected in nature, we have proposed several methods to optimize experiment design, i.e., using a small number of experiments to decipher network structure through selecting of optimal initial condition (Zhang et al., 2013). We demonstrated that reduction of number of possible networks is largely facilitated if we apply such rational design procedures.

Here we propose a novel constraint to optimize the reverse engineering process. Our new assumption is that in the whole set of possible networks that can provide a specific biological function, those with the minimal number of edges tends to be the biological network. This is equivalent to adding a large weight on networks with the minimum number of edges. It is similar to some linear regression algorithm which induces penalty terms of coefficient, like Lasso method. We test our assumption with three well-known biological networks, including the budding yeast cell cycle network, the fission yeast cell cycle network and the bacteria *Escherichia coli* SOS network. Two different types of data are examined, one is generated by using optimal initial conditions, the other is data from expression perturbations like knockout and over-expression. Numerical simulation results show that edges shared by the minimal networks are contained in the biological networks. After limited number of experiments, the only minimal network turns to be the biological network.

Although our study was conducted in Boolean Network Model, our approach may provide a general framework to limit possible network number and can further incorporate other dynamic constraints, such as attractor basin size (Li et al., 2004), regulation entropy (Wu et al., 2009) in biological regulatory network construction.

2. Model

2.1. Biological networks

We test our method on three networks, the SOS network in *E. coli*, the cell cycle networks of budding yeast (Li et al., 2004) and fission yeast (Davidich and Bornholdt, 2008), as demonstrated in Fig. 1. These networks are well established, they can be viewed as representatives of biological networks. The interactions in these regulatory networks govern various cellular functions, which we termed as functional tasks. For cell cycle network of budding yeast, Cln3 activates a pair of transcription factor groups, SBF and MBF, which stimulates the transcription of Cln1/2 and Clb5/6. Activation of Clb5 triggers the entry of the S phase, after which Clb2 is activated and drives the cell into the M phase. After the M phase, Sic1 and Cdh1 are activated and the cell comes back to the stationary G1 phase. For cell cycle network of fission yeast, the Start Kinase (SK) inactivates Ste9 and Rum1 and allows accumulation of phosphorylated (inactive) form of Cdc2/Cdc13. During the G2/M transition, Cdc2/Cdc13 becomes activated due to the activation of Cdc25 and inhibition of Wee1. In the M phase, PP is activated and drives the cell back to the G1 stationary phase. For SOS response pathway of *E. coli*., RecA is recruited to the single stranded regions of DNA and becomes activated. Activation of RecA releases the inhibition of SOS genes which are responsible for the repair of DNA damage. When the DNA repair is completed, LexA is activated and the expression of SOS genes is down-regulated. In general, we refer functional task as a sequence of events which should be tightly regulated for biological systems to exert corresponding functions, which are presented in Tables 1–3.

2.2. Boolean network model and reverse engineering

We use Boolean Network model to generate dynamic behavior (function) of above described systems. In Boolean model, a biological species, such as DNA, mRNA and protein, are treated as a node. Its associated Boolean states$S_i \in \{0,1\}$ correspond to the on (active) and off (inactive) states of the node. Regulation from node j to node i is represented by coefficienta_{ji}. Weight of regulation is reflected by the value of a_{ji}. We choose dominant inhibition form of regulation (Zhang et al., 2013; Wang et al., 2010, 2012), with $a_{ji}=1$ for activation, $a_{ji}=-\infty$ for inhibition and $a_{ji}=0$ for no regulation. This captures the combinational rule of transcription, in which presence of an inhibitor may block the process of transcription. States of nodes in the network is updated via the following rule:

$$\begin{cases} S_i(t+1)=\theta(\sum_j S_j(t)a_{ji}), & \sum_j S_j(t)a_{ji}\neq 0 \\ S_i(t+1)=S_i(t), & \sum_j S_j(t)a_{ji}=0 \end{cases} \quad (1)$$

where θ is Heaviside step function: $\theta(x)=1$ when $x>0$ and $\theta(x)=0$ when $x<0$. For a given initial state, the system is updated synchronously according to Eq. (1) until it reaches a steady state known as attractor. The sequence of states is termed as a trajectory of the system. Different initial states lead to different trajectories, contributing to different responses and functions of the system. In our work, we use these trajectories as output data of different kinds of experiments instead of using realistic data. New experiments are conducted by changing the initial condition or states of nodes. For biological networks, normal trajectories can be generated using the nature initial conditions, these trajectories are termed as biological trajectories. The biological trajectories of the three networks are presented in Tables 1–3.

In the reverse engineering process, the question is to derive network structures that can produce experimental data. For the Boolean Network model that discussed above, it is to derive the network connection matrix from specific trajectories. Trajectory constraints on network topology can be written as logic expressions. Let logic variable r_{ji} denotes inhibition regulation from node j to node i, with $r_{ji}=1$ (True) for such a regulation exist and $r_{ji}=0$ (False) for the opposite case. Similarly, let g_{ji} represents positive regulation from node j to node i, and $r_{ij}\times g_{ji}=0$. Auto- activation and inhibition of node i is represented by g_{ii} and r_{ii}, respectively. For a give trajectory, one can get logic constrain of node i,

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

Eq. (2) of different time steps are combined to get the final constraint of a trajectory. For multiple trajectories, one can get all possible regulation pattern of each node according to the combination of logical expressions, thus get all the possible networks. It is shown that our reverse engineering method is applicable to *in silico* data (Zhang et al., 2013). On the basis of this approach, we have proposed the max-distance method, the sampling method, and the entropy method to optimize experiment design (Zhang et al., 2013). They diverge in performance on deduction of possible network number and computational complexity. Here we choose the sampling method as the benchmark for further discussion because of its high efficiency. It mainly works as follows: for given trajectories of former experiments, networks are sampled from whole possible space of topology. The initial condition that can maximally separate the sampled networks is chosen as the initial condition of next experiment to further confine network structure. This process is repeated until the number of possible networks converges. For more details see Zhang et al. (2013).

2.3. Minimal network constraint

Among all the possible networks, minimal networks are the subset with the fewest edges, contributing to the backbone motif