Infectious Disease Modelling 1 (2016) 52–70

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

# Controllability and stability analysis of large transcriptomic dynamic systems for host response to influenza infection in human

Xiaodian Sun [a, 1], Fang Hu [b, 1], Shuang Wu [c, 1], Xing Qiu [b], Patrice Linel [c], Hulin Wu [d, *]

[a] Biostatistics and Bioinformatics Core, Sylvester Comprehensive Cancer Center, University of Miami, Miami, USA
[b] Department of Biostatistics and Computational Biology, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA
[c] Genus PLC, ABS Global, Deforest, WI, USA
[d] Department of Biostatistics, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA

ARTICLE INFO

*Article history:*
Received 22 April 2016
Accepted 8 July 2016
Available online 13 September 2016

ABSTRACT

*Background:* Gene regulatory networks are complex dynamic systems and the reverse-engineering of such networks from high-dimensional time course transcriptomic data have attracted researchers from various fields. It is also interesting and important to study the behavior of the reconstructed networks on the basis of dynamic models and the biological mechanisms. We focus on the gene regulatory networks reconstructed using the ordinary differential equation (ODE) modelling approach and investigate the properties of these networks.

*Results:* Controllability and stability analyses are conducted for the reconstructed gene response networks of 17 influenza infected subjects based on ODE models. Symptomatic subjects tend to have larger numbers of driver nodes, higher proportions of critical links and lower proportions of redundant links than asymptomatic subjects. We also show that the degree distribution, rather than the structure of networks, plays an important role in controlling the network in response to influenza infection. In addition, we find that the stability of high-dimensional networks is very sensitive to randomness in the reconstructed systems brought by errors in measurements and parameter estimation.

*Conclusions:* The gene response networks of asymptomatic subjects are easier to be controlled than those of symptomatic subjects. This may indicate that the regulatory systems of asymptomatic subjects are easier to recover from disease stimulations, so these subjects are less likely to develop symptoms. Our results also suggest that stability constraint should be considered in the modelling of high-dimensional networks and the estimation of network parameters.



* Corresponding author.
*E-mail addresses:* Xiaodian.Sun@med.miami.edu (X. Sun), hufang87@gmail.com (F. Hu), swu115@gmail.com (S. Wu), Xing_Qiu@urmc.rochester.edu (X. Qiu), patrice.linel@genusplc.com (P. Linel), Hulin.Wu@uth.tmc.edu (H. Wu).

Peer review under responsibility of KeAi Communications Co., Ltd.

[1] Contribute equally.

http://dx.doi.org/10.1016/j.idm.2016.07.002

## 1. Background

Genes, proteins and metabolites in a cell interact with each other and also with the environment, forming large complex networks. When a cell responds to different stimulus, different sets of genes are expressed and these genes and their products may form distinct regulatory networks (Arbeitman et al., 2002; Spellman et al., 1998). In recent years, various analytical tools and methods have been developed to reconstruct the gene regulatory networks (GRNs) from experimental data and many of these methods are based on high-dimensional time course gene expression data. Examples include Boolean networks (Baranzini et al., 2004; Kauffman, 1969; Shmulevich, Dougherty, Kim, & Zhang, 2002), information theory models (Steuer, Kurths, Daub, Weise, & Selbig, 2002; Stuart, Segal, Koller, & Kim, 2003), graphical Gaussian models (Schafer & Strimmer, 2005), graphical Granger causality (Shojaie & Michailidis, 2010), dynamic Bayesian networks (DBN) (Husmeier, 2003; Murphy & Mian, 1999; Perrin, Ralaivola, Mazurie, Bottani, & Mallet, 2003; Zou & Conzen, 2005), vector autoregressive (VAR) models (Charbonnier, Chiquet, & Ambroise, 2010; Opgen-Rhein & Strimmer, 2007; Shimamura et al., 2009), state space models (Hirose et al., 2008; Kojima et al., 2010; Rangel et al., 2004) and differential equation models (Chen, Wang, Tseng, Huang, & Kao, 2005; Golightly & Wilkinson, 2006; Holter, Maritan, Cieplak, Fedoroff, & Banavar, 2001; Lu, Liang, Li, & Wu, 2011; Yeung, Tegnér, & Collins, 2002; de Jong, 2002).

Among the aforementioned network modelling approaches, we focus on the ordinary differential equation (ODE) model. An ODE model is formed by taking the derivative of a gene expression as a function of expression levels of all related genes, which results in a direct graph model and the dynamic feature of GRNs is automatically captured and quantified. Recently, Lu et al. (2011). and Wu, Liu, Qiu, & Wu (2014). proposed a novel pipeline to reverse engineer genome-wide dynamic GRNs from time course gene expression data using high-dimensional ODE models. To deal with the high-dimensionality of the genome-wide GRN, genes are clustered with similar expression patterns into co-expressed modules and module-based dynamic ODE networks are constructed. The advanced parameter estimation method for ODE models is coupled with statistical variable selection techniques to identify the sparse structure of the network. A series of cutting-edge statistical techniques are combined to efficiently reduce the dimension of the network reconstruction and account for the correlations of time series expression measurements from the same gene. In addition, it also allows us to perform model selection and parameter estimation of the ODE model for one equation at a time, which is highly efficient in reconstructing large-scale networks from a computational perspective.

In this paper, we focus on the analysis of the networks identified using the data-driven pipeline developed in Lu et al. (2011) and Wu et al. (2014). Specifically, we are interested in the parameter estimation, the controllability analysis and stability analysis of the reconstructed ODE networks. Consider a $K$ dimensional dynamic network by the following ODE model

$$\begin{aligned} \mathbf{X}'(t) &= \beta_0 + A\mathbf{X}(t) + B\mathbf{V}(t), \quad \mathbf{X}(t=t_0) = \mathbf{X}_0, \\ \mathbf{Y}(t) &= \mathbf{C}\mathbf{X}(t) + \mathbf{W}(t), \end{aligned} \tag{1}$$

where $\mathbf{X}(t) = (X_1(t), \ldots, X_K(t))^T$ is the vector of state variables; $\mathbf{X}'(t)$ indicates the derivative of $\mathbf{X}(t)$; $\mathbf{V}(t) = (V_1(t), \ldots, V_m(t))^T$ is the vector of input variables such as the stimulation variables to the gene regulatory network (e.g., the viral load in our influenza infection example); $A = (a_{ki})_{k,i=1,\ldots,K}$ is the system matrix that quantifies the regulatory effects between network components; $B = (b_{kj})_{k=1,\ldots,K;j=1,\ldots,m}$ is the input matrix that represents the effect of the input variables; parameters $\beta_0$ and $\mathbf{X}_0$ are the intercept and initial values, respectively; $\mathbf{Y}(t) = (Y_1(t), \ldots, Y_K(t))^T$ is the vector of observation variables and the observation matrix $\mathbf{C}$ is an identity matrix for our gene regulatory network examples (the expression levels for all genes are measured by microarray or RNA-Seq techniques); and $\mathbf{W}(t)$ represents the measurement error, which is usually assumed to follow a Gaussian distribution with mean zero. For simplicity of presentation, we consider model (1) with only one input variable, i.e., $m$=1. But the methods presented in this paper can be easily extended to cases where $m$>1.

In biological systems, most nodes are only directly connected to a small number of other nodes, so it is reasonable to assume that the system matrix $A$ is a sparse matrix. In this work, we further assume that the structure and nonzero components of $A$ have already been identified and estimated. Since the network parameters estimated during the network structure identification step are based on each differential equation separately, they may not be accurate and need further refinement (Lu et al., 2011). This refined estimation needs to be done very carefully, as the estimated parameters will be utilized in the following controllability and stability analysis. We propose to use a trust-region-reflective algorithm with box constraint (Branch, Coleman, & Li, 1999; Byrd, Schnabel, Shultz, 1988; Moré & Sorensen, 1983; Coleman and Li, 1994, 1996) in the parameter estimation (refinement) step. Moreover, we take advantage of the matrix sparsity of $A$ along with a cost-less Jacobian evaluation so that the algorithm is most efficient for sparse linear ODE parameter estimation.

The ability of controlling complex biological networks is of paramount importance since it enables us to obtain a deeper understanding of the networks. A dynamic system is controllable, if and only if an arbitrary initial state is steerable to an arbitrary desired final state within a finite time interval using the external inputs (Klamka, 2013). Controllability is a generic characteristic of the dynamic system and is also strongly related to stability and stabilizability (Klamka, 1991). The main goal of controllability analysis is to assess whether a dynamic system is controllable and how to control the system state to the desired state effectively (Kaczorek, 1992). Controllability analysis has also been proven to be especially useful for studying dynamic systems in many different fields, including biology (Rajapakse, Groudine, & Mesbahi, 2011), engineer (Reehorst, Chung, Potapczuk, & Choo, 2000), chemical processes (Bahri, Bandoni, & Romagnoli, 1997) and physics (Wang, Ni, Lai, & Grebogi, 2012) and so on. In recent years, studies of controllability have been applied in a variety of biological systems,