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 β_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,

such as brain networks (Gu, Pasqualetti, Cieslak, Grafton, & Bassett, 2014, p. 14065197), protein interaction networks (Wuchty, 2014), immune networks (Matsuoka et al., 2013), gene regulatory networks (Qian & Dougherty, 2013), and metabolic networks (Liu & Pan, 2014). Biological network controllability analysis provides a system level paradigm for potentially identifying intervention targets in complex biological systems, which offers a new strategy for drug discovery and development (Csermely, Korcsmáros, Kiss, London, & Nussinov, 2013; Kariya, Honma, & Suzuki, 2013; Kim, Park, & Cho, 2013; Poland et al., 2013; Pujol, Mosca, Farrés, & Aloy, 2010; Ruths & Ruths, 2014).

In traditional control theories, system controllability is determined by the Kalman's rank condition. However, it is difficult to apply this criterion to large scale biological systems, because it is computationally expensive, numerically unstable and it is also hard to design an appropriate input matrix B for a given system matrix A to achieve the goal of the system (Klamka, 2013; Nepusz & Vicsek, 2012). Recently, Liu, Slotine, and Barabasi (2011) adopted the structural controllability (SC) and maximum matching theory, and successfully resolved these problems. The system (1) is structurally controllable if there exists a set of parameters for the non-zero components in A and B such that model (1) is controllable (Chen, 2013; Liu et al., 2011). Under the SC framework, the maximum matching algorithm is used to determine the number of driver nodes, which is defined as the minimum number of inputs to maintain full control of the network, as the minimum set of unmatched nodes (Chen, 2013; Liu et al., 2011). Liu et al (Liu et al., 2011). provided a new insight into the controllability framework of nodal dynamics (Liu, 2014), which is based on the assumption that driver nodes can be directly controlled by external input signals in the whole network (Asgari, Salehzadeh-Yazdi, Schreiber, & Masoudi-Nejad, 2013; Cho, 2011). However, one drawback of this method is that they did not consider the strength of the links in the complex networks. To resolve this problem, Yuan, Zhao, Di, Wang, and Lai (2013) developed an exact controllability (EC) framework for analysing the controllability of nodal dynamics with link weights of complex networks. Besides these two methods, Nepusz and Vicsek (2012) took a novel perspective to handle the controllability by using the edge dynamics (ED) controllability framework (Nacher & Akutsu, 2013). In this method, each node acts as a switchboard that can accept information using its internal link edge and also spread the signals to neighbouring nodes through external link edges.

Stability is another important property of a dynamic system. It characterizes whether the solutions (trajectories) of a dynamic system is stable under small perturbations of the initial conditions. Over the past decades, many researchers have studied the stability of various kinds of dynamic systems, such as linear systems (Lawrence, 1991), nonlinear systems (Khalil & Grizzle, 2002; Sastry, 1999), time-invariant systems (Doan, Kalauch, & Siegmund, 2009; Pötzsche, Siegmund, & Wirth, 2003), time-varying systems (DaCunha, 2005; Hinrichsen, Ilchmann, & Pritchard, 1989) and systems with time-delay (Gu, Chen, & Kharitonov, 2003; Park, 2007). For a dynamic system with deterministic parameters, the stability is dichotomous and deterministic, *i.e.*, the dynamic system can be classified as stable or unstable. In reality, the structure and parameters of a dynamic system cannot be measured precisely due to errors pertain to the experimental measurements or the approximation techniques used in parameter estimation. Hence, it is necessary to study the stability in the presence of randomness of system matrix. In this case, the stability should be quantified probabilistically. This concept of "probability of stability" was first proposed by Ashby (1960), i.e., for a real-valued random system matrix generated from a given ensemble, the probability of stability is the probability that all eigenvalues of this matrix have negative real parts. Simulation studies based on some low-dimensional systems showed that the probability of stability decreases rapidly to zero as the system dimension increases (Ashby, 1960). Porter (1972); Porter and Crossley (1972) derived an analytical formula for the probability of stability for a linear system in which the elements of the Jordan canonical form of the system matrix are independently and identically distributed between interval $[-a, a]$ with a uniform distribution. Using the eigenvalue linearization, Lim and Junkins (1987) reduced the expression of probability of stability into a simple form for random matrices in which the elements follow a joint Gaussian distribution. However, for a high-dimensional system, linearizing eigenvalues with respect to the differentials of parameters is too complex and impractical. Dankovic, Vidojkovic, & Vidojkovic, 2007; Jovanovic, Dankovic, & Antic, 2005; Jovanović & Danković, 2004; Zlatkovic & Samardzic, 2012 presented a method for determining the probability of stability of one-dimensional, nth-order discrete-time systems. In their work, the probability of stability is defined as the multiple integral of the joint distribution function of all random parameters over the stable region of the parameter space. Although their systems can be transformed to certain high-dimensional, first-order ODE systems, their results are not applicable for such systems in general. Raghavan and Barmish (2006) investigated the probability of stability for dynamic systems with random parameters in the case that the mean of the system matrix is stable. These investigators derived an upper bound for the variation of parameter perturbation under which the stability of system is assured with probability one. This bound is inversely proportional to the square root of the network dimension, so it converges to zero as the dimension increases. However, there is no estimation of the probability of stability when the variance of the perturbation is larger than this bound and the system is not almost surely stable. Moreover, their results are not applicable if the mean system matrix is unstable. To the best of our knowledge, so far there are no methods that can estimate the probability of stability for high-dimensional dynamic systems with a general random system matrix.

The time course gene expression and viral load data for 17 healthy human volunteers who received intranasal inoculation of influenza H3N2/Wisconsin were first described in (Huang et al., 2011) and reanalysed in (Linel, Wu, Deng, & Wu, 2014) using systems biology approaches. Among these 17 subjects, 9 developed symptoms in response to influenza infection (symptomatic subjects) and the other 8 did not (asymptomatic subjects). For each subject, temporally differentially expressed genes were identified and clustered into co-expressed modules based on the similarity of gene expression patterns. Module-based ODE network was constructed for each subject and the network structure, and nonzero components in the system matrix A of model (1), were identified and estimated using the pipeline proposed in Wu et al. (2014). In this work, we will

perform controllability analysis and stability analysis after parameter refinement for the reconstructed ODE networks in Linel et al. (2014). Our goal of controllability analysis is to understand how to control the dynamic gene response networks in human subjects under influenza infection with minimum number of nodes and how the network topology affects their controllability. Stability analysis will also be performed to evaluate how stable the dynamic gene response network is under influenza infection.

2. Methods

2.1. Pipeline for identifying the network structure through module-based ODE model

In Wu et al. (2014), a novel pipeline was proposed, based on the work in Lu et al. (2011), to reconstruct genome-wide dynamic regulatory networks through time course gene expression data using high-dimensional ordinary differential equation (ODE) models. This pipeline incorporated a series of statistical methods to perform the following tasks. First, the temporally differentially expressed genes, i.e., genes that exhibit significant expression changes during the experimental period, were identified using the FPCA (functional principal component analysis)-based significant test (Wu & Wu, 2013). In this method, the expression profile of each gene was estimated through a data-driven eigen-representation using FPCA, and a modified F-statistic was adopted to quantify the signal-to-noise ratio of each gene. The significant genes were determined by a permutation test coupled with the multiple testing adjustment method proposed by Benjamini and Hochberg (1995).

Next, the time course expressions of the temporally differentially expressed genes were standardized and clustered into a number of co-expressed modules. These modules were considered as the nodes of the dynamic gene regulatory network, so the dimension of the network was significantly reduced. The module-based linear ODE model for the GRN can be written as model (1), where coefficients $A = \{a_{ki}\}_{k,i=1,\ldots,K}$ quantify the regulation effects of other modules, including self-regulation on the rate of expression change of the k-th module. The identification of network structures of the GRN is then transformed to an equivalent statistical problem of identifying the nonzero coefficients $S = \{1 \le k, i \le K : a_{ki} \neq 0\}$ in the ODE model (1).

To this end, the two-stage estimation method for ODE models was employed to decouple model (1) into K pseudo-regression models so that the problem of identifying the ODE model structure can be transformed into variable selection problems of these K regression models. Then the smoothly clipped absolute deviation (SCAD) (Fan & Li, 2001) method was applied to select nonzero a_{ki}'s (Lu et al., 2011). The benefits of this approach are that it avoids numerically solving the differential equations and allows independent model selection and parameter estimation for one equation at a time, which significantly reduce the computational cost (Lu et al., 2011). More details of the pipeline can be found in (Wu et al., 2014).

2.2. Refining parameter estimates

The parameter estimates obtained from the two-stage method are not efficient in terms of estimation accuracy, because of the approximation errors brought in by the estimates of the mean expression curves $X_k(t)$ of the modules and their derivatives $X'_k(t)$. When the data are measured at a sparse grid or with large noise signals, such errors could be quite large and may devastate the parameter estimates. So it is necessary to refine the parameter estimates for the selected ODE model using the more rigorous maximum likelihood method (Lu et al., 2011) or nonlinear least squares (NLS) method (Xue, Miao, & Wu, 2010). The SCAD estimators from the two-stage method can be used as the initial estimates in these nonlinear optimization procedures. Here we propose using a Trust-Region-Reflective (TRR) algorithm (Branch et al., 1999; Byrd, Schnabel, Shultz, 1988; Moré & Sorensen, 1983; Coleman and Li, 1994, 1996) to solve the NLS equation for the ODE model (1). While the mixed-effect modelling approach adopted in (Lu et al., 2011) is more efficient in estimating gene-specific parameters, the TRR algorithm solves the ODEs directly and provides better parameter estimation at the module level, which better fits the objective of this work. In addition, it has been proven that the trust region approach possesses the global convergence property, robustness and large scale affinity (Fletcher, Gould, Leyffer, Toint, & Wächter, 2002; Wright & Nocedal, 1999; Yuan, 2000). Some works have suggested combining the global optimization method, such as the global genetic algorithm with a local optimization approach, which ultimately will find the "true" global solution to the NLS equation (Xue et al., 2010). However, the improvement in the residual sum of squares is minimum compared to the significant difference in the computational time cost between the global optimization and the trust region approach, especially for high-dimensional ODE models. Therefore, we believe that the trust region approach is the best in terms of both solution quality and computational cost.

This TRR algorithm allows the box constraint optimization, where the upper and lower bounds can be derived from the data. Some parameters can be deduced from prior knowledge. For example, the regulatory relationships between some genes are known to be positive or negative during the biological process of interest, so the lower or upper bounds can be determined correspondingly. In addition, this algorithm does not require forming and solving the normal equation for the NLS problem directly, but rather optimizes the trust-region subspace through a least-square type of function that return the square difference of each component, i.e.,

$$\min_\theta f(\theta) = \min_\theta (f_1(\theta) + f_2(\theta) + \ldots + f_K(\theta)) \tag{2}$$

where $f_k(\theta) = \sum_{i=1}^{n_k} \|Y_{k_i}(t) - X_k(\theta,t)\|_2^2$, with $Y_{k_i}(t)$ being the time course expression data for the i-th gene of the k-th module, n_k being the size of the k-th module, $X_k(\theta,t)$ being the state variable in model (1) and $\|\cdot\|_2$ being the L^2 norm. We use an implementation algorithm that takes advantage of the sparsity of the system matrix A along with a cost-less Jacobian evaluation. The Jacobian matrix is not evaluated by finite-difference but through the sensitivity equation of the linear ODE model (1). This implementation algorithm makes our NLS estimates more efficient for the sparse linear ODE estimation (Linel et al., 2014).

2.3. Controllability analysis

To introduce the controllability framework for the complex, high-dimensional ODE networks, we consider the linear time-invariant dynamics as equation (1). We can make a linear transform of the input variable $V(t)$ as following:

$$\overline{V}(t) = V(t) + B^{-1}\beta_0 \tag{3}$$

This linear transform does not change the property of the system. So system (1) becomes:

$$X'(t) = \beta_0 + AX(t) + BV(t) = AX(t) + B\left(\overline{V}(t) - B^{-1}\beta_0\right) + \beta_0 = AX(t) + B\overline{V}(t) \tag{4}$$

In this work, we apply and compare the SC unweight (Liu et al., 2011), EC weight (Yuan et al., 2013) and ED methods Nepusz & Vicsek, 2012. In the SC unweight method (Liu et al., 2011), the driver nodes are calculated by:

$$N_D = \max\{N - |M^*|, 1\} \tag{5}$$

where $|M^*|$ is the number of maximum matching and N is the number of vertices. Similarly, the edge nodes are defined in the ED method and the number of edge nodes can be computed by:

$$N_D = \sum_{ver=1}^{n} \max\left(dout_i^+ - din_i^-, 0\right) + \sum_{i=1}^{cn} \beta eta_i \tag{6}$$

where, $dout_i^+$ is the out-degree; din_i^- is the in-degree; ver is the vertex; cn is the number of connected components; βeta_i is 1 if the i-th component is balanced and 0 otherwise (Yuan et al., 2013). In the EC weighted method (Yuan et al., 2013), the driver nodes for link weight directed network can be either the sparse case or the dense case. The number of driver nodes under the sparse case can be computed as

$$N_D = \max\{1, N - \mathrm{rank}(A)\}, \tag{7}$$

and under the dense case, it is defined as

$$N_D = \max\{1, N - \mathrm{rank}(\omega I + A)\}, \tag{8}$$

where I is the identity matrix.

2.4. Stability analysis

In model (1), the system matrix A is a deterministic matrix, and the homogenous system associated with model (1) is as follows:

$$\begin{aligned} X'(t) &= AX(t) \\ X(t = t_0) &= X_0 \end{aligned} \tag{9}$$

The exponential stability of system (9) can be determined by the following criterion (Doan et al., 2009): Every solution is exponentially stable if and only if all the eigenvalues of A have negative real parts. The solution of system (1) is of the following form (Lawrence, 1991):

$$X(t) = e^{(t-t_0)A}M_0 + \int_{t_0}^{t} e^{(t-s)A}(\beta_0 + BV(s))ds \tag{10}$$

where the first term on the right hand side of (10) is the solution to system (9); and the second term is an integral of function of A, B, $V(t)$ and β_0. Thus, the condition of system (1) being exponentially stable is the same as that of system (9). For a given matrix A, it can be rewritten into the form of PJP^{-1} by conducting the Jordan decomposition over the field of real numbers, in

which P is a nonsingular matrix with columns being generalized eigenvectors of A, and J is a block-diagonal matrix with the diagonal elements being the real parts of the generalized eigenvalues of A. Taking into consideration the randomness of the system matrix A due to estimation error, the dynamic system with randomness associated with system (1) can be written as:

$$\begin{aligned} X'(t) &= \beta_0 + W_K X(t) + BV(t) \\ A &= \widehat{A} + G_K \end{aligned} \tag{11}$$

where the estimated system matrix $\widehat{A}$ is considered as a deterministic matrix here; β_0, B and $V(t)$ are the same as those in (1); G_K is a random matrix representing the estimation error. In this work, G_K is chosen to be the real Ginibre ensemble in which all elements are independently and identically distributed Gaussian random variables with mean zero (Rider & Sinclair, 2014). Thus, the true system matrix A is a random matrix whose elements have different means but the same variance. Since the stability of a dynamic system is determined by its corresponding system matrix, based on the definition of probability of stability in (Lim & Junkins, 1987), the probability of stability of system (11) can be computed as follows:

$$\int_{\max(\mu(A))<0} f(\max(\mu(A)))d(\max(\mu(A))) \tag{12}$$

where $max(\mu(A))$ is the largest real part of eigenvalues of A and $f(\max(\mu(A)))$ is the probability density function (pdf) of $max(\mu(A))$.

Here the estimated system matrix $\widehat{A}$ is considered as a deterministic matrix, and $A = \widehat{A} + G_K$ is a random matrix, since we assume that G_K is the real Ginibre ensemble with a standard deviation cr_A, where $0 < c < 1$ is a coefficient that quantifies the magnitude of randomness relative to the estimated $\widehat{A}$ and $r_A = |\widehat{A}|_F/K$ with $|\widehat{A}|_F$ being the Frobenius norm of $\widehat{A}$. We conducted Monte Carlo simulation studies as follows. We first generate $N = 1{,}000{,}000$ random matrices G_{Ks} ($s = 1,\ldots,N$) with standard deviation cr_A for $c = 0.05$, 0.1 and 0.5 respectively; then, for each $A = \widehat{A} + G_{Ks}$, we compute the value of $max(\mu(A))$, the largest real part of the eigenvalues of A, by conducting the Jordan-decomposition over the field of real numbers; finally, we obtain the empirical probability $\Pr(max(\mu(A))<0) =$ number of $max(\mu(A))<0/N$.

3. Results

Using the parameter refinement method described in the Methods section, we obtain the refined parameter estimates for the gene response ODE networks identified in Linel et al (Linel et al., 2014). for the 17 subjects. Based on these parameter estimates, we then conduct the following controllability analysis and stability analysis.

3.1. Controllability analysis

We performed the controllability analysis using the SC unweighted method (Liu et al., 2011), EC weighted method (Yuan et al., 2013) and ED method Nepusz & Vicsek, 2012, respectively. We compared the number/proportion of driver nodes and different edge types computed by these three methods, and explored the control paths of the dynamic networks. We also compared the observed proportion of driver nodes with those from null random network models.

Fig. 1 shows the proportions of driver nodes determined by the SC unweighted, EC weighted and ED methods for asymptomatic and symptomatic subjects, respectively. The corresponding numeric values can be found in ATable 1. From Fig. 1, we can easily see that the proportions of driver nodes identified by the SC unweighted and EC weighted methods are the same for most subjects; for subjects 3,11 and 15, the proportions of driver nodes identified using the EC weighted method are larger than those identified using the SC unweighted method. The similarity of these two methods in the results is due to the fact that both these two methods are based on the linear nodal dynamics and both consider the structure of the estimated system matrix $\widehat{A}$. If one assigns each directed link a random parameter, linear nodal dynamics with a given exact link weight can reproduce the results by the linear nodal dynamics without the link weight. This means that the EC weighted method produces the same proportion of driver nodes as the SC unweighted method if the true system matrix A is given. However, in practice, there are inevitable numerical errors in the estimated system matrix $\widehat{A}$. The EC weighted method is under an exact controllability framework based on the eigenvalues and the rank of $\widehat{A}$, while the SC unweighted method is under a structural controllability framework based on a maximum matching algorithm. So, when $\widehat{A}$ has numerical errors, these two methods may produce different results, as seen in subjects 3, 11 and 15. Since these two methods produce similar results for most subjects and the SC unweighted method is not subject to the numeric errors in the estimated matrix $\widehat{A}$, we prefer using the SC unweighted method to the EC weighted method.

We can also see that the proportions of driver nodes based on the ED method are greater than or equal to those based on the SC unweighted method, with the only exception of subject 7. It is not surprising to see that these two methods produce quite different results, as the SC unweighted method is based on nodal dynamics while the ED method is based on edge dynamics. Here we prefer the SC unweighted method over the ED method, because the number of driver nodes computed by the former method is no more than that calculated by the later. Our choice is also supported by the fact that GRNs are known to be homogeneous or homogeneous-mixing networks (Bouyioukos & Kim, 2011) and the SC unweighted method is more

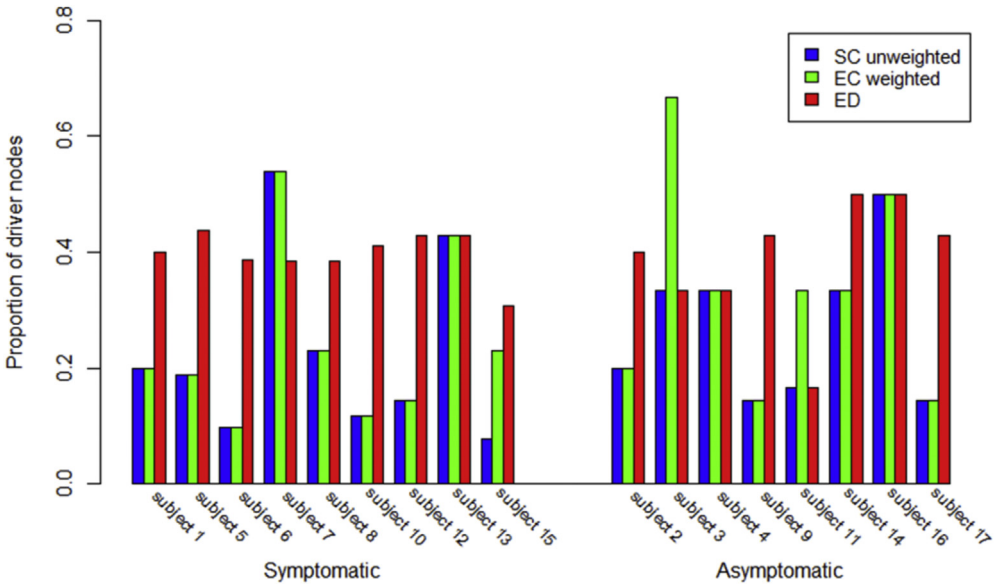


Fig. 1. The proportions of driver nodes determined by the SC unweighted, EC weighted and ED methods for symptomatic and asymptomatic subjects.

suitable for the homogeneous networks than the ED method (Asgari et al., 2013; Liu, 2014; Tang, Qian, Gao, & Kurths, 2014). In the following, we only show the results using the SC unweighted method.

One commonly used criterion for evaluating the controllability of complex networks is the minimal number of driver nodes (Yan-Dong, Song-Yang, Lv-Lin, & Liang, 2014). The minimal number of driver nodes is directly associated with how difficult it is to control the network; the larger the number of driver nodes is, the more difficult it is to control the network. From Fig. 2, we can see that the minimal numbers of driver nodes for asymptomatic subjects are smaller than those for symptomatic subjects. This indicates that the GRNs of asymptomatic subjects are easier to be controlled back to the health state from disease state after viral infection compared to those of symptomatic subjects. However, it may be suboptimal to compare only the absolute values of the minimal number of driver nodes across subjects, since the total number of nodes is different for different subjects. In regard to this problem, we also compared the fraction of minimum driver nodes, i.e., the proportion of the minimum number of driver nodes with respect to the total number of nodes in the network. A higher value of the fraction of minimum driver nodes means that it is more difficult to control the network. The proportions of minimum driver nodes for asymptomatic and symptomatic subjects are displayed in Fig. 2 and there is no significant difference between these two groups.

There are three types of links in a network: redundant, critical and ordinary links (Liu et al., 2011). If the number of driver nodes remains unchanged when an edge is removed, this edge is a redundant link; if the number of driver nodes increases with the absence of an edge, then this edge is a critical link; if an edge is neither redundant link nor critical link, then it is an ordinary link. The numbers of redundant links and critical links are related to the robustness and stability of a complex network; a larger number of redundant links corresponds to a more robust and stable dynamic network, while a larger number of critical links indicates that the network is more difficult to control. The numbers and proportions of the redundant, ordinary and critical link edges for asymptomatic and symptomatic subjects using the SC unweight method are shown in Fig. 3. We could see that most asymptomatic subjects had larger proportions of redundant links than symptomatic subjects. This implies that the gene response networks of asymptomatic subjects are more robust and stable than those of symptomatic subjects in response to influenza infection. On the other hand, asymptomatic subjects generally had smaller proportions of critical links than symptomatic subjects, except for subject 2 and subject 4, indicating that most asymptomatic subjects are easier to be controlled than symptomatic subjects. These results are also consistent with our findings in Fig. 2.

Next, we test whether the proportions of driver nodes for the GRNs of the 17 subjects are significantly different from those in null random network models. Complex networks are described by many property-related statistics such as the node degree, clustering coefficient, degree distribution and so on. The absolute values of these statistics are not enough to fully describe the macro characters of complex networks in quantitative precision and accuracy because there are so many different kinds of structure topologies and sizes in complex networks. So in the field of complex networks, the significance of observed network statistics is often tested against statistics computed based on null-model networks that preserve basic properties of the original network, such as the degree distribution, density, edges, nodes and so on. There are two most commonly used null-model networks, the Erdos-Renyi (ER) random network and the configuration model (CM). The ER random network is a random graph that models the edge between each of its nodes with degrees following a Poisson distribution. The configuration model has the same degree distribution as the original network and it is a directed random multi-

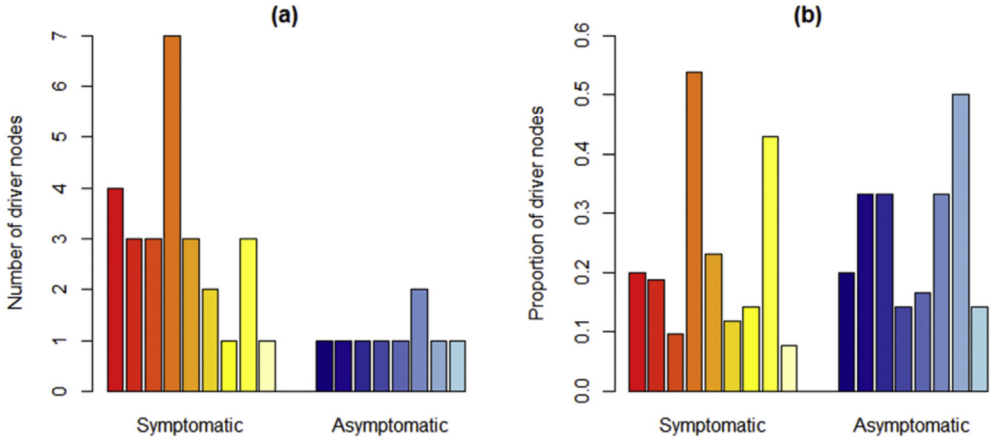


Fig. 2. (a) The minimum numbers and (b) proportions of driver nodes for symptomatic and asymptomatic subjects by the SC unweighted method.

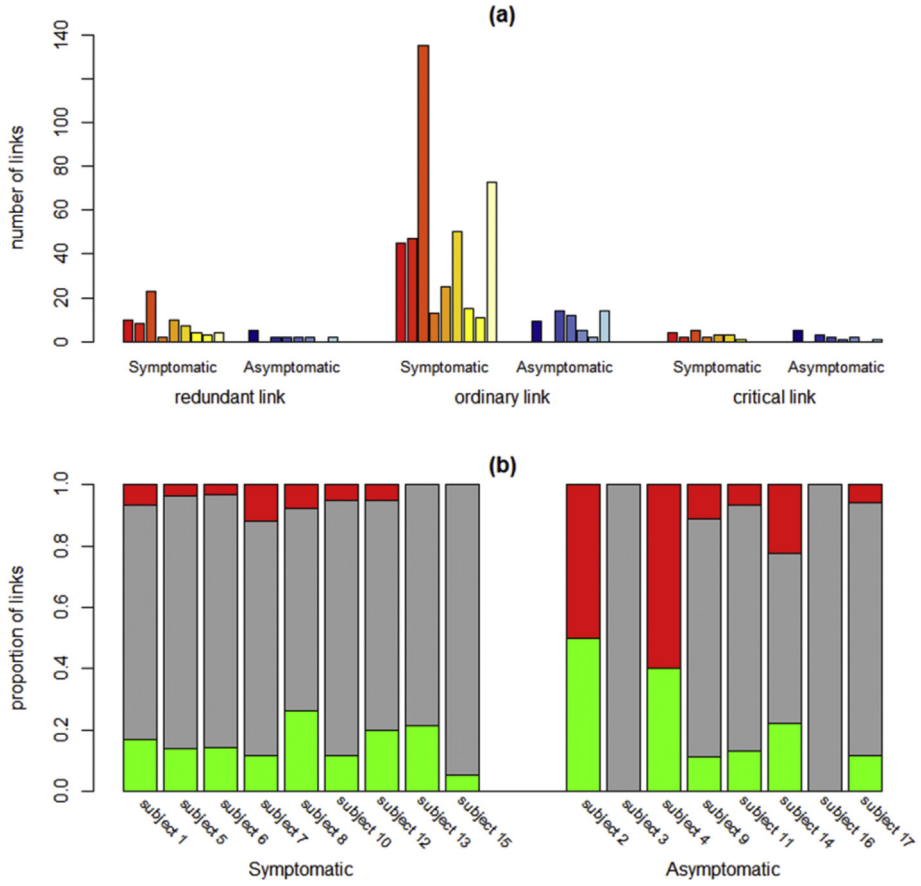


Fig. 3. (a) Bar plot of the numbers of the redundant, ordinary and critical links for symptomatic and asymptomatic subjects, respectively. (b) The proportions of the redundant (green), ordinary (grey) and critical (red) links for symptomatic and asymptomatic subjects, respectively.

graph with a given degree sequence. The edges in the configuration model are randomly paired with every pair uniformly, allowing self-loops and multi-edges. Intuitively, the joint degree distribution is the probability that a uniformly and randomly selected edge is between two node degrees, which is estimated by the degree-degree correlation between two nodes connected by the edge. Considering that many biological networks have very different joint degree distributions, we also construct a null-model with a fixed degree distribution using the joint degree distribution corresponding to the real-world

networks, which is rooted in the configuration model. To evaluate if the proportions of driver nodes are significantly over- or under-represented in the identified GRNs, null-model networks are generated with the randomness, degree sequence and joint degree distribution preserved.

Here we choose the ER random network and configuration model with a joint degree distribution as the null models Nepusz & Vicsek, 2012. For each null model, we use the ED method to generate 100 random samples and then compute the proportion of driver nodes for all randomized instances. The proportions of driver nodes for each null model and the real network are visualized in Fig. 4. For most asymptomatic subjects, the proportion of driver nodes for the ER random network is almost the same as that of the identified GRNs, except for subject 14; while for symptomatic subjects, the proportion of driver nodes for the ER random network is much smaller than that of the identified GRNs. It implies that the ER network is easier to control than the GRNs of symptomatic subjects and is similarly controllable as the GRNs of asymptomatic subjects. The proportions of driver nodes for the configuration model with and without a joint degree distribution are both closer to that of the identified GRNs than the ER random network, no matter for asymptomatic subjects or symptomatic subjects. Comparing the proportion of driver nodes between the identified GRNs and the configuration model, the differences are within ± 0.0254 for most subjects, except that subjects 13 and 14 had ± 0.13 differences. The fact that the proportion of driver nodes for the identified GRNs is almost equal to that for the configuration model indicates that the number of driver nodes to control a network only depends on the degree distribution of homogeneous networks. These findings are consistent with those in Liu et al. (2011).

3.2. Stability analysis

The ODE dynamic systems (networks) for temporal gene response to influenza infection in human subjects were established using time course gene expression data. It is challenging to study the stability of the dynamic system with the estimated system matrix $\widehat{A}$. The concept of probability stability was introduced to quantify the stability uncertainty for a dynamic system with a random system matrix (Ashby, 1960; Dankovic et al., 2007; Jovanovic et al., 2005; Jovanović & Danković, 2004; Lim & Junkins, 1987; Porter, 1972; Porter & Crossley, 1972; Raghavan & Barmish, 2006; Zlatkovic & Samardzic, 2012). The random matrix theory of the distribution of eigenvalues can be explored in order to investigate the stability of a dynamic system with an uncertain system matrix (Chiani, 2014; Rider & Sinclair, 2014; Silverstein, 1994; Tao & Vu, 2008). However, all previous results in random matrix theory are concerned with cases in which either the mean of the random matrix is a zero matrix or a low-rank perturbation. In our study, most of the system matrices $\widehat{A}$ estimated in (Linel et al., 2014) are not of low-rank so these theories are not directly applicable. Here we employ a Monte Carlo simulation approach to study the probability of stability for the established dynamic gene response networks. The basic idea is to use the Monte Carlo approach to obtain the empirical distribution of the largest (real part) eigenvalue of the estimated system matrix $\widehat{A}$ under certain realistic assumptions, so that we could evaluate the probability of stability of the dynamic system in practice. The results of the probability stability analysis for the dynamic gene response networks for the 17 subjects infected by influenza virus (Linel et al., 2014) are summarized as follows.

We display $max(\mu(\widehat{A}))$, the largest real part of eigenvalues of the estimated system matrix $\widehat{A}$ in column 5 of Table 1 for all 17 subjects. From this table, we can see that the estimated system matrices $\widehat{A}$ are unstable for all 17 subjects because $max(\mu(\widehat{A}))$ are all non-negative. We have also found from Table 1 that the network dimensions (K) of the asymptomatic subjects are

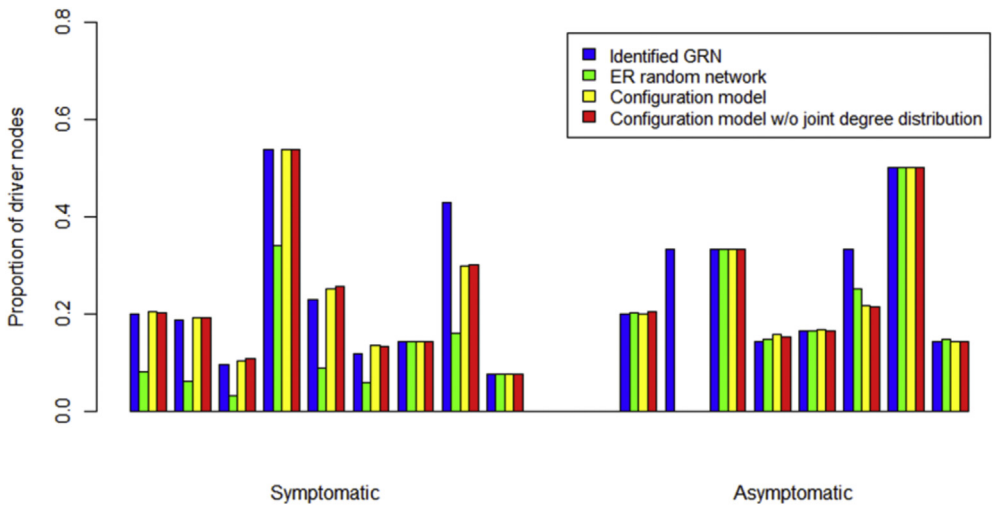


Fig. 4. The proportions of driver nodes for symptomatic and asymptomatic subjects in the identified GRNs, ER random network, configuration model and configuration model without joint degree distribution.

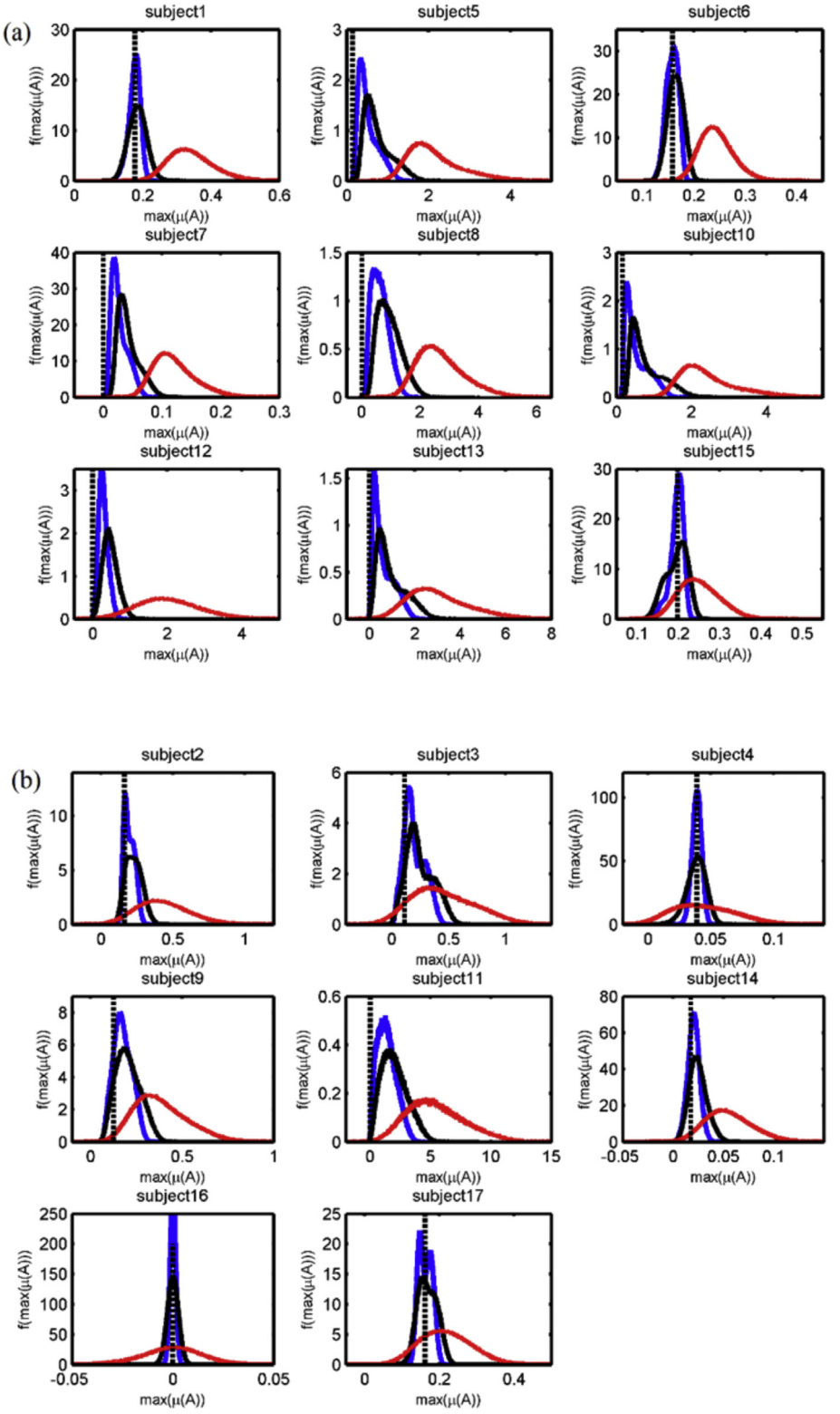


Fig. 5. The pdfs of the largest real part of eigenvalues $max(\mu(A'))$ for (a) symtomatic subjects and (b) asymptomatic subjects, where $A = \widehat{\mathbf{A}} + G_K$, and $\widehat{\mathbf{A}}$ is the estimated system matrix, and G_K is the real Ginibre ensemble with a standard deviation cr_A, in which $r_A = |\widehat{\mathbf{A}}|_F/K$ with $|\widehat{\mathbf{A}}|_F$ being the Frobenius norm of $\widehat{\mathbf{A}}$. The blue, black and red solid lines denote the cases of coefficient $c = 0.05$, 0.1 and 0.5, respectively; the black dash line represents the value of $max(\mu(\widehat{\mathbf{A}}))$.

smaller (no more than 7) than those of the symptomatic subjects (no less than 7). Although the system with the estimated system matrix $\widehat{A}$ is not stable for all 17 subjects, we further explored whether there is a probability that the system is stable if the true system matrix $A = \widehat{A} + G_K$. Here G_K is a random matrix due to estimation error with a standard deviation cr_A, where $r_A = \left|\widehat{A}\right|_F / K$ with $\left|\widehat{A}\right|_F$ being the Frobenius norm of $\widehat{A}$, and $0 < c < 1$ is a coefficient that quantifies the magnitude of randomness relative to the norm of estimated $\widehat{A}$. We employed the Monte Carlo approach (see the Methods section below) to evaluate the probability that the system with the true system matrix A is stable, i.e., the probability that the largest real part of eigenvalues of A, $max(\mu(A))$, is negative.

Fig. 5a and b display the empirical probability density functions (pdfs) of $max(\mu(A))$ for symptomatic subjects and asymptomatic subjects, respectively. The pdfs with different colors indicate cases of different magnitudes of randomness added to $\widehat{A}$, i.e., small, medium and large standard deviation of G_K or $c = 0.05$, 0.1 and 0.5 respectively. In columns 6–8 of Table 1, we list the empirical probability of the dynamic system being stable for all 17 subjects, *i.e.*, $\Pr(max(\mu(A))<0)$. We have found from both Fig. 5a and Table 1 that for all degrees of randomness in the random matrix G_K (i.e., $c = 0.05$, 0.1 and 0.5), 7 out of the 9 symptomatic subjects had practically zero probability of stability and 2 symptomatic subjects (subjects 12 and 13) had very small probability of stability (smaller than 0.004). Moreover, we see from the third column in Table 1 that the network dimensions of the 7 symptomatic subjects with zero probability of stability are much larger than 7, while the network dimensions of subjects 12 and 13 are 7. On the other hand, we have seen from Fig. 5b and Table 1 that for $c = 0.5$, none of the 8 asymptomatic subjects had zero probability of stability, and in particular subjects 3, 4 and 16 had much larger probability of stability than any of the symptomatic subjects. We have also noticed that the network dimensions of subjects 3, 4 and 16 are much smaller than 7, while the network dimensions of the other asymptomatic subjects are 7 or close to 7. All of these results have indicated that the symptomatic subjects and asymptomatic subjects are quite different in stability property presumably due to the difference in system dimensions.

Since we suspect that a high-dimensional stable system can be easily identified as an unstable system if its system matrix is measured or estimated with error, we also performed Monte Carlo simulations to demonstrate that, for a stable system, it could have a high probability to be unstable if its dimension is high and its system matrix is measured with a small random error (see Appendix 1). For example, we simulated an artificial system which is stable but with similar properties to that of Subject 5 with 16 dimensions, our simulation results show that the simulated system has a high probability (from 0.60 to 0.76) to be unstable if a small random error is added to the stable system matrix (see AFig. 1). As a matter of fact, this observation is conceptually consistent with the random matrix theory in that the mean of the largest real part of eigenvalues of a random matrix with zero-mean elements is positively related to the square root of the dimension of the random matrix (Chiani, 2014; Rider & Sinclair, 2014; Tao & Vu, 2008).

4. Conclusions and discussions

Network controllability analysis provides a system-level understanding of how biological systems respond to diseases and also a novel paradigm for discovery of potential intervention targets. In this paper, we have compared three methods of controllability analysis, the SC unweighted method, the EC weighted method and the ED method, to study the GRNs of human subjects challenged by influenza virus. The SC unweighted method is preferred over the EC weighted method, because the former only relies on the structure of the estimated system matrix $\widehat{A}$ and has fewer sources of errors compared to the latter. In

Table 1
The network dimension K (column 3), the average value of elements in the estimated system matrix $\widehat{\mathbf{A}}$ (column 4), the largest real part of generalized eigenvalues of $\widehat{\mathbf{A}}$ (column 5) and the empirical probability $\Pr(max(\mu(A)) < 0)$ for different values of coefficient c in case 1 (columns 6–8) for the GRN of each subject.

Case	Subject	K	r_A	$max(\mu(\widehat{A}))$	$\Pr(max(\mu(A))<0\|c)$		
					$c = 0.05$	$c = 0.1$	$c = 0.5$
Symptomatic	subject1	20	0.106	0.1781	0	0	0
	subject 5	16	0.8731	0.144	0	0	0
	subject 6	31	0.0542	0.1585	0	0	0
	subject 7	13	0.054	0	0	0	0
	subject 8	13	1.2191	0.0267	0	0	0
	subject 10	17	0.9763	0.1669	0	0	0
	subject 12	7	2.0127	0.003	0	0.002366	0.002893
	subject 13	7	2.1202	0.0014	0.003158	0.001837	0.000725
	subject 15	13	0.0807	0.1987	0	0	0
Asymptomatic	subject 2	5	0.3169	0.162	0	0	0.001946
	subject 3	3	0.4525	0.1156	0.000027	0.001272	0.033702
	subject 4	3	0.0414	0.0391	0	0	0.020151
	subject 9	7	0.2167	0.1258	0	0	0.000044
	subject 11	6	3.9621	0.0185	0.001638	0.00158	0.001779
	subject 14	6	0.0412	0.0177	0.000048	0.000226	0.00239
	subject 16	2	0.0264	0	0.521634	0.521558	0.504336
	subject 17	7	0.1037	0.1621	0	0	0.00001

addition, the SC unweighted method is more suitable for the homogeneous networks than the ED method, and the GRNs are known to be homogeneous and homogeneous-mixing networks (Asgari et al., 2013; Bouyioukos & Kim, 2011; Liu, 2014; Tang et al., 2014).

Comparing the minimal numbers of driver nodes between the asymptomatic and symptomatic subjects using the SC unweighted method, we find that the GRNs of asymptomatic subjects are easier to be controlled than those of symptomatic subjects. Similar results have also been observed when comparing the redundant, ordinary and critical links of the GRNs between asymptomatic and symptomatic subjects. As seen from Fig. 3, the GRNs of most asymptomatic subjects have larger proportions of redundant links, smaller proportions of critical links and smaller numbers of driver nodes than those of symptomatic subjects, indicating that the GRNs of asymptomatic subjects are more robust and stable, and are easier to be controlled than those of symptomatic subjects. This may be one of the reasons that the asymptomatic subjects did not develop symptoms after influenza challenge. It may also suggest that these asymptomatic subjects are easier to recover from disease status after viral infection than those symptomatic subjects.

The minimum number of driver nodes required to control a dynamic GRN depends on the degree of homogeneity of the network. The less heterogeneous the degree distribution is, the less the proportion of minimum driver nodes is required. This is demonstrated by the observations that the proportions of driver nodes required to control the GRNs identified for the 17 subjects are no less than that for an ER random network and are almost the same as that of the configuration model. Thus, the degree distribution, rather than the structure of network, plays an important role in controlling the GRN in response to influenza infection. In addition, there exist significant differences in the proportions of driver nodes between the GRNs and the ER random network for asymptomatic and symptomatic subjects. GRNs of symptomatic subjects are harder to be controlled compared to the ER random network, while GRNs of asymptomatic subjects are almost the same to be controlled as the ER random network.

Considering the empirical probability of stability of the 17 GRNs, we first treated the estimated system matrix $\widehat{A}$ for each subject as a deterministic matrix and added different degrees of randomness to this matrix. Differences have been observed between the symptomatic and asymptomatic subjects. Most symptomatic subjects had zero probability of stability when adding different degrees of randomness in the estimated matrix $\widehat{A}$, while all the asymptomatic subjects had nonzero probability of stability when the randomness is high. We have also found that higher dimensional systems (dimension $K \geq 5$) were of zero probability to be stable, while the probabilities of stability of smaller systems (dimension $K < 5$) were much higher, especially with high randomness. Thus, the dimension of the GRN plays an important role in determining the probability of stability of the system.

We also used the Monte Carlo simulation approach to explore whether a stable system can be identified as an unstable system if its dimension is high and its system matrix is measured or estimated with a small random error. Our simulation results show that it is very easy for a high-dimensional exponentially stable system to become an unstable one after adding randomness into the system matrix. Since estimation errors are inevitable in the estimated system matrix $\widehat{A}$ for the dynamic gene response network systems for the 17 subjects in (Linel et al., 2014), it is possible that the unobserved true system matrix A for some of these 17 subjects (if not all) are exponentially stable. Thus, the fact that the system based on the estimated matrix $\widehat{A}$ is unstable is presumably due to the random estimation error in $\widehat{A}$. Therefore, the network stability is very sensitive to randomness in the system matrix for high dimensional systems, we conclude that a true stable system is most likely to be identified as an unstable system if its system matrix is high-dimensional and needs to be estimated with noisy data.

Conflicts of interest

The authors declare that they have no competing interests.

Authors' contributions

XS performed the controllability analysis; FH and XQ performed the stability analysis; and PL performed the parameter refinement analysis. XS, FH, SW and PL drafted the manuscript. HW conceived the idea, directed the research and revised the manuscript.

Acknowledgements

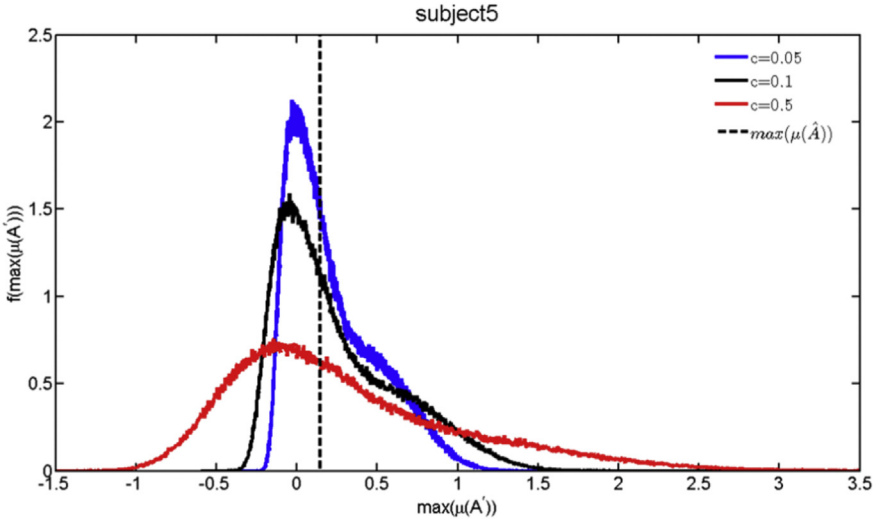
This research was partially supported by the NIH grants HHSN272201000055C, P30AI078498, R01 AI087135, HHSN27220201200005C, and HHSN266200700008C. The authors would like to thank Yang-Yu Liu, Tamás Nepusz, Wen-Xu Wang and Zhesi Shen for assistance in code preparation and helpful discussions about the controllability analysis.

Appendix 1. Additional Stability Analyses

We consider the estimated system matrix $\widehat{A}$ as an unobserved true system matrix A^{true} imposed by measurement errorsG_K, $\widehat{A} = A^{true} + G_K$ where G_K is a Ginibre ensemble. For each estimated matrix $\widehat{A}$, we create a deterministic matrix A^{new} such that, after adding the same real Ginibre ensemble G_K as in Eq. (11) (denote $A^{new} + G_K$ by A'), the median of the empirical pdf of

$max(\mu(A'))$ equals to $max(\mu(\widehat{A}))$. By construction, the probability of A' being more stable than $\widehat{A}$, i.e., $\Pr(max(\mu(A'))< max(\mu(\widehat{A})))$ is 0.5. It suggests that the stability of unobserved true system matrix A^{true} is comparable to that of matrix A^{new}. We constructed the deterministic matrix A^{new} as follows. Firstly, for each estimated system matrix $\widehat{A}$, we calculate its corresponding block diagonal matrix J by conducting Jordan decomposition. For Jordan blocks that have positive diagonal elements and no nilpotent components, we set the diagonal elements to be zero and keep the rest elements unchanged. For Jordan blocks that have positive diagonal elements and nilpotent components, we set this Jordan block to be a zero matrix. For the rest of Jordan blocks, we keep them unchanged. The obtained new matrix is denoted by J', and we define a deterministic matrix $R_K = PJ'P^{-1}$. Then, by using the same real Ginibre ensemble G_K and the simulation method as stated in the Methods Section, we can obtain the empirical pdf of $max(\mu(R_K + G_K))$, and computed its median, denoted by δ. Finally, we construct the required deterministic matrix $A^{new} = R_K + (max(\mu(\widehat{A}))\text{-}\delta)E_K$, where E_K is a K-dimensional unit matrix. The empirical pdf of $max(\mu(A'))$ is the pdf of $max(\mu(R_K + G_K))$ shifted by a distance of $max(\mu(\widehat{A}))\text{-}\delta$.

For simplicity of representation, we only show the results for the symptomatic subject 5. The results for other subjects are available upon request. The values of $max(\mu(A^{new}))$ are -0.275, -0.498 and -1.875 for $c = 0.05$, 0.1 and 0.5, respectively, meaning that the constructed matrix A^{new} is exponentially stable. Since A^{new} is constructed to match the stability property of the unobserved true system matrix A^{true}, the fact that A^{new} is exponentially stable suggests that it is possible that the true system matrix A^{true} is exponentially stable, although the estimated system matrix $\widehat{A}$ is unstable. If we impose the real Ginibre ensemble G_K on A^{new}, the empirical pdfs of $max(\mu(A'))$ for different values of the coefficient c for subject 5 are displayed in AFig. 1 and the representations of the colorful solid lines and black dash line are the same as those in.Fig. 5. We can obtained from AFig. 1 that the empirical probability of stability, *i.e.*, $\Pr(max(\mu(A'))<0)$ are 0.2389, 0.3086 and 0.4037 for $c = 0.05$, 0.1 and 0.5, respectively. These results show that an exponentially stable system can be more prone to be unstable if the parameters are estimated with errors. Thus, the apparent instability of the estimated system matrix $\widehat{A}$ is probably an artefact due to randomness brought by errors in measurements and estimation.



AFigure 1. The pdfs of the largest real part of eigenvalues $max(\mu(A'))$, for subject 5, where $A' = A^{new} + G_K$, and A^{new} is the constructed stable deterministic matrix from the estimated matrix $\widehat{\mathbf{A}}$, and G_K is the real Ginibre ensemble with a standard deviation cr_A, in which $r_A = |\widehat{\mathbf{A}}|_F/K$ with $|\widehat{\mathbf{A}}|_F$ being the Frobenius norm of $\widehat{\mathbf{A}}$. The three colourful solid lines denote the cases for different coefficient c indicating the magnitude of randomness, and the black dash line represents the value of $max(\mu(\widehat{\mathbf{A}}))$.

ATable 1
The proportions of driver nodes determined by the SC unweighted, EC weighted and ED methods for symptomatic and asymptomatic subjects.

Case	Subject	# Of nodes	# Of edges	SC unweighted	EC weighted	ED
Symptomatic	subject1	20	59	0.2	0.2	0.4
	subject5	16	57	0.1875	0.1875	0.4375
	subject6	31	163	0.0967742	0.0967741935484	0.387097
	subject7	13	17	0.538462	0.538461538462	0.384615
	subject8	13	38	0.230769	0.230769230769	0.384615
	subject10	17	60	0.117647	0.117647058824	0.411765
	subject12	7	20	0.142857	0.142857142857	0.428571
	subject13	7	14	0.428571	0.428571428571	0.428571
	subject15	13	77	0.0769231	0.230769230769	0.307692

ATable 1 (*continued*)

Case	Subject	# Of nodes	# Of edges	SC unweighted	EC weighted	ED
Asymptomatic	subject2	5	10	0.2	0.2	0.4
	subject3	3	9	0.333333	0.666666666667	0.333333
	subject4	3	5	0.333333	0.333333333333	0.333333
	subject9	7	18	0.142857	0.142857142857	0.428571
	subject11	6	15	0.166667	0.333333333333	0.166667
	subject14	6	9	0.333333	0.333333333333	0.5
	subject16	2	2	0.5	0.5	0.5
	subject17	7	17	0.142857	0.142857142857	0.428571

ATable 2
The number of driver nodes (DN), redundant links (RL), ordinary links (OL) and critical links (CL) for symptomatic and asymptomatic subjects by the SC unweighted method.

Symptomatic					Asymptomatic				
Subject	DN	RL	OL	CL	Subject	DN	RL	OL	CL
subject1	4	10	45	4	subject2	1	5	0	5
subject5	3	8	47	2	subject3	1	0	9	0
subject6	3	23	135	5	subject4	1	2	0	3
subject7	7	2	13	2	subject9	1	2	14	2
subject8	3	10	25	3	subject11	1	2	12	1
subject10	2	7	50	3	subject14	2	2	5	2
subject12	1	4	15	1	subject16	1	0	2	0
subject13	3	3	11	0	subject17	1	2	14	1
subject15	1	4	73	0					

ATable 3
The proportions of redundant links (RL), ordinary links (OL) and critical links (CL) for symptomatic and asymptomatic subjects by the SC unweighted method.

Case	Subject	RL	OL	CL
Symptomatic	subject1	0.169492	0.762712	0.0677966
	subject5	0.140351	0.824561	0.0350877
	subject6	0.141104	0.828221	0.0306748
	subject7	0.117647	0.764706	0.117647
	subject8	0.263158	0.657895	0.0789474
	subject10	0.116667	0.833333	0.05
	subject12	0.2	0.75	0.05
	subject13	0.214286	0.785714	0
	subject15	0.0519481	0.948052	0
Asymptomatic	subject2	0.5	0	0.5
	subject3	0	1	0
	subject4	0.4	0	0.6
	subject9	0.111111	0.777778	0.111111
	subject11	0.133333	0.8	0.0666667
	subject14	0.222222	0.555556	0.222222
	subject16	0	1	0
	subject17	0.117647	0.823529	0.0588235

ATable 4
The proportions of driver nodes in comparison to null models by the SC unweighted method.

Case	Subject	Identified GRN	ER random network	Configuration model	Configuration model w/o joint
Symptomatic	subject1	0.2	0.0825	0.2055	0.2035
	subject5	0.1875	0.063125	0.193125	0.191875
	subject6	0.096774	0.0322581	0.104839	0.109032
	subject7	0.538462	0.34	0.538462	0.538462
	subject8	0.230769	0.09	0.253077	0.256154
	subject10	0.117647	0.0594118	0.136471	0.132941
	subject12	0.142857	0.142857	0.142857	0.142857
	subject13	0.428571	0.161429	0.298571	0.301429
	subject15	0.076923	0.0769231	0.0769231	0.0769231
Asymptomatic	subject2	0.2	0.202	0.2	0.204
	subject3	0.333333	–	–	–
	subject4	0.333333	0.333333	0.333333	0.333333
	subject9	0.142857	0.148571	0.157143	0.154286
	subject11	0.166667	0.166667	0.168333	0.166667
	subject14	0.333333	0.251667	0.218333	0.215
	subject16	0.5	0.5	0.5	0.5
	subject17	0.142857	0.148571	0.144286	0.144286

ATable 5
The complete list of driver nodes for each subject by the SC unweighted method.

subject1	19	4	12	17			
subject2	1						
subject3	1						
subject4	1						
subject5	12	3	11				
subject6	4	7	15				
subject7	3	8	10	13	6	9	12
subject8	5	11	13				
subject9	7						
subject10	15	16					
subject11	5						
subject12	1						
subject13	5	6	7				
subject14	6	2					
subject15	7						
subject16	2						
subject17	6						

ATable 6
The complete list of the redundant links (RL) and critical links (CL) for each subject by the SC unweighted method.

Subject	Link type	Source	Target
subject1	CL	15	17
		4	0
		12	10
		8	6
	RL	9	10
		14	17
		10	6
		10	17
		17	0
		18	0
		4	17
		2	10
		12	0
		5	0
subject2	CL	0	1
		2	2
		3	4
		1	3
		4	0
	RL	3	2
		3	1
		1	0
		1	2
		1	1
subject3	CL		
	RL	9	10
subject4	CL	0	0
		1	2
		2	1
	RL	0	1
		1	1
subject5	CL	7	9
		12	13
	RL	5	9
		1	13
		10	13
		2	9
		2	13
		11	9
		11	13
		13	13
subject6	CL	4	15
		1	13
		21	25
		27	29
		17	11

ATable 6 (*continued*)

Subject	Link type	Source	Target
	RL	3	11
		3	13
		3	15
		19	25
		19	15
		7	11
		20	25
		20	15
		10	25
		10	29
		28	11
		28	15
		2	15
		12	13
		13	13
		22	25
		25	25
		14	13
		15	13
		15	25
		29	11
		29	15
		30	13
subject7	CL	6	7
		5	8
	RL	6	8
		11	7
subject8	CL	0	0
		1	2
		2	9
	RL	3	0
		7	0
		7	2
		7	9
		8	2
		6	0
		6	2
		6	9
		11	9
		12	2
subject9	CL	0	2
		4	5
	RL	0	0
		0	1
subject10	CL	0	1
		2	3
		1	15
	RL	7	15
		7	3
		8	15
		9	15
		9	3
		3	3
		12	15
subject11	CL	3	5
	RL	4	5
		5	5
subject12	CL	6	2
	RL	2	2
		2	1
		6	1
		1	2
subject13	CL		
	RL	1	1
		1	2
		2	1
subject14	CL	1	5
		5	0
	RL	1	0
		5	4
subject15	CL		

(*continued on next page*)

ATable 6 (*continued*)

Subject	Link type	Source	Target
	RL	11	12
		9	12
		3	12
		10	12
subject16	CL		
	RL		
subject17	CL	1	0
	RL	6	0
		3	0

ATable 7
The complete list of control paths for each subject by the SC unweighted method.

subject1	stem	8	6	1								
		9	13									
		11										
		12	10	15	17	19						
	bud	0	2	7	14	18	16	3	4	(assigned to stem 12 10 15 17 19)		
		5	(assigned to stem 9 13)									
subject2	bud	0	1	3	4							
		2										
subject3	bud	0										
		1										
		2										
subject4	bud	0										
	bud	1	2									
subject5	stem	2	7	9								
		4										
		14										
	bud	0	1	12	13	10	3	5	11	8	15	6	assigned to stem 2 7 9
subject6	stem	5	22	20	28	0	1	13	27	29	14	17	11
							19	23	30	26	4	15	24
		6	18										
		21	25	2	10	8							
	bud	3	7	(assigned to stem 5 22 20 28 0 1 13 27 29 14 17 11 19 23 30 26 4 15 24)									
		12	(assigned to stem 5 22 20 28 0 1 13 27 29 14 17 11 19 23 30 26 4 15 24)										
subject7	stem	2										
		3	0	1								
		4										
		5	8									
		6	7									
		11	9	10								
		12										
subject8	stem	7	4									
		11	5	6	1	2	9					
		12	10									
	bud	0	(assigned to stem 11 5 6 1 2 9)									
		3	(assigned to stem 11 5 6 1 2 9)									
		8	(assigned to stem 11 5 6 1 2 9)									
subject9	stem	6										
	bud	0	2	3	4	5	(assigned to stem 6)					
		1										
subject10	stem	11	14	10	8	6	13					
		16	12	7	0	1	15					
	bud	2	3	4	(assigned to stem 11 14 10 8 6 13)							
		5	(assigned to stem 11 14 10 8 6 13)									
		9	(assigned to stem 11 14 10 8 6 13)									
subject11	stem	3	5	4	2							
	bud	0	1	(assigned to stem: 3 5 4 2)								
subject12	stem	0	1	6	2	5	4	3				
subject13	stem	4										
		5										
		6										
	bud	0	1	(assigned to stem 6)								
		2	3									
subject14	stem	2	4									
		3										
	bud	0	1	5								
subject15	stem	11	9	2	3	4	12	10	6	8	7	

	bud	0	1	(assigned to stem: 11 9 2 3 4 12 10 6 8 7)
		5	(assigned to stem: 11 9 2 3 4 12 10 6 8 7)	
subject16	stem	1		
	bud	0		
subject17	stem	3	2	6 4 5
	bud	0	1	(assigned to stem 3 2 6 4 5)

References

Arbeitman, M. N., Furlong, E. E., Imam, F., Johnson, E., Null, B. H., Baker, B. S., et al. (2002). Gene expression during the life cycle of Drosophila melanogaster. *Science, 297*(5590), 2270–2275.

Asgari, Y., Salehzadeh-Yazdi, A., Schreiber, F., & Masoudi-Nejad, A. (2013). Controllability in cancer metabolic networks according to drug targets as driver nodes. *PLoS ONE, 8*(11), e79397.

Ashby, W. R. (1960). *Design for a brain*. Springer Science & Business Media.

Bahri, P. A., Bandoni, J. A., & Romagnoli, J. A. (1997). Integrated flexibility and controllability analysis in design of chemical processes. *AIChE Journal, 43*(4), 997–1015.

Baranzini, S. E., Mousavi, P., Rio, J., Caillier, S. J., Stillman, A., Villoslada, H., et al. (2004). Transcription-based prediction of response to IFNβ using supervised computational methods. *PLoS Biology, 3*(1), e2.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B Methodology, 57*(1), 289–300.

Bouyioukos, C., & Kim, J. (2011). Gene regulatory network properties linked to gene expression dynamics in spatially extended systems. In G. Kampis, I. Karsai, & E. Szathmáry (Eds.), *Advances in artificial life Darwin meets von Neumann* (vol. 5777, pp. 321–328). Springer Berlin Heidelberg.

Branch, M. A., Coleman, T. F., & Li, Y. (1999). A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM Journal on Scientific Computing, 21*(1), 1–23.

Byrd, R., Schnabel, R., & Shultz, G. (1988). Approximate solution of the trust region problem by minimization over two-dimensional subspaces. *Mathematical Programming, 40*(1–3), 247–263.

Charbonnier, C., Chiquet, J., & Ambroise, C. (2010). Weighted-LASSO for structured network inference from time course data. *Statistical Applications in Genetics and Molecular Biology, 9*. Article 15.

Chen, G.-R. (2013). Problems and challenges in control theory under complex dynamical network environments. *Acta Automatica Sinica, 39*(4), 312–321.

Chen, K. C., Wang, T. Y., Tseng, H. H., Huang, C. Y., & Kao, C. Y. (2005). A stochastic differential equation model for quantifying transcriptional regulatory network in Saccharomyces cerevisiae. *Bioinformatics, 21*(12), 2883–2890.

Chiani, M. (2014). Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy–Widom distribution. *Journal of Multivariate Analysis, 129*, 69–81.

Cho, A. (2011). Scientific link-up yields 'control panel'for networks. *Science, 332*(6031), 777–777.

Coleman, T., & Li, Y. (1994). On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds. *Mathematical Programming, 67*(1–3), 189–224.

Coleman, T. F., & Li, Y. (1996). An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization, 6*(2), 418–445.

Csermely, P., Korcsmáros, T., Kiss, H. J., London, G., & Nussinov, R. (2013). Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. *Pharmacology & Therapeutics, 138*(3), 333–408.

DaCunha, J. J. (2005). Stability for time varying linear dynamic systems on time scales. *Journal of Computational and Applied Mathematics, 176*(2), 381–410.

Dankovic, B., Vidojkovic, B. M., & Vidojkovic, B. (2007). The probability stability estimation of discrete-time systems with random parameters. *Control and Intelligent Systems, 35*(2), 134–139.

Doan, T. S., Kalauch, A., & Siegmund, S. (2009). Exponential stability of linear time-invariant systems on time scales. *Nonlinear Dynamics and Systems and Theory, 9*, 37–50.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association, 96*, 1348–1360.

Fletcher, R., Gould, N. I. M., Leyffer, S., Toint, P. L., & Wächter, A. (2002). Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming. *SIAM Journal on Optimization, 13*(3), 635–659.

Golightly, A., & Wilkinson, D. J. (2006). Bayesian sequential inference for stochastic kinetic biochemical network models. *Journal of Computational Biology : a Journal of Computational Molecular Cell Biology, 13*(3), 838–851.

Gu, K., Chen, J., & Kharitonov, V. L. (2003). *Stability of time-delay systems*. Springer.

Gu, S., Pasqualetti, F., Cieslak, M., Grafton, S. T., & Bassett, D. S. (2014). *Controllability of brain networks*. arXiv preprint arXiv.

Hinrichsen, D., Ilchmann, A., & Pritchard, A. J. (1989). Robustness of stability of time-varying linear systems. *Journal of Differential Equations, 82*(2), 219–250.

Hirose, O., Yoshida, R., Imoto, S., Yamaguchi, R., Higuchi, T., Charnock-Jones, D. S., et al. (2008). Statistical inference of transcriptional module-based gene networks from time course gene expression profiles by using state space models. *Bioinformatics, 24*(7), 932–942.

Holter, N. S., Maritan, A., Cieplak, M., Fedoroff, N. V., & Banavar, J. R. (2001). Dynamic modeling of gene expression data. *Proceedings of the National Academy of Sciences of the United States of America, 98*(4), 1693–1698.

Huang, Y., Zaas, A. K., Rao, A., Dobigeon, N., Woolf, P. J., Veldman, T., et al. (2011). Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. *PLoS genetics, 7*(8), e1002234.

Husmeier, D. (2003). Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics, 19*(17), 2271–2282.

de Jong, H. (2002). Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology : a Journal of Computational Molecular Cell Biology, 9*(1), 67–103.

Jovanović, Z., & Danković, B. M. (2004). On the probability stability of discrete-time control systems. *Facta Universitatis-series: Electronics and Energetics, 17*(1), 11–20.

Jovanovic, Z., Dankovic, B., & Antic, D. (2005). Reliability estimation of the discrete-time control systems with random parameters. *Acta Electrotechnica et Informatica, 5*(1), 4.

Kaczorek, T. (1992). *Linear control systems: Analysis of multivariable systems*. John Wiley & Sons, Inc.

Kariya, Y., Honma, M., & Suzuki, H. (2013). Systems-based understanding of pharmacological responses with combinations of multidisciplinary methodologies. *Biopharmaceutics & Drug Disposition, 34*(9), 489–507.

Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology, 22*(3), 437–467.

Khalil, H. K., & Grizzle, J. W. (2002). *Nonlinear systems, vol. 3. Upper Saddle river*. Prentice hall.

Kim, J., Park, S.-M., & Cho, K.-H. (2013). Discovery of a kernel for controlling biomolecular regulatory networks. *Scientific Reports, 3*.

Klamka, J. (1991). *Controllability of dynamical systems*. The Netherlands: Kluwer Academic Publishers Dordrecht.

Klamka, J. (2013). Controllability of dynamical systems. A survey. *Bulletin of the Polish Academy of Sciences: Technical Sciences, 61*(2), 335–342.

Kojima, K., Yamaguchi, R., Imoto, S., Yamauchi, M., Nagasaki, M., Yoshida, R., et al. (2010). A state space representation of VAR models with sparse learning for dynamic gene networks. *Genome Informatics International Conference on Genome Informatics, 22*, 56–68.

Lawrence, P. (1991). *Differential equations and dynamical systems*. New York: Springer-Verlag.
Lim, K. B., & Junkins, J. L. (1987). Probability of stability: New measures of stability robustness for linear dynamical systems. *Journal of the Astronautical Sciences, 35*(4), 383–397.
Linel, P., Wu, S., Deng, N., & Wu, H. (2014). Dynamic transcriptional signatures and network responses for clinical symptoms in influenza-infected human subjects using systems biology approaches. *Journal of Pharmacokinetics and Pharmacodynamics, 41*, 509–521.
Liu, Y.-Y. (2014). Theoretical progress and practical challenges in controlling complex networks. *National Science Review, 1*(3), 341–343.
Liu, X., & Pan, L. (2014). Detection of driver metabolites in the human liver metabolic network using structural controllability analysis. *BMC Systems Biology, 8*(1), 51.
Liu, Y. Y., Slotine, J. J., & Barabasi, A. L. (2011). Controllability of complex networks. *Nature, 473*(7346), 167–173.
Lu, T., Liang, H., Li, H. Z., & Wu, H. (2011). High-dimensional ODEs coupled with mixed-effects modeling techniques for dynamic gene regulatory network identification. *Journal of the American Statistical Association, 106*(496), 1242–1258.
Matsuoka, Y., Matsumae, H., Katoh, M., Eisfeld, A. J., Neumann, G., Hase, T., et al. (2013). A comprehensive map of the influenza A virus replication cycle. *BMC Systems Biology, 7*(1), 97.
Moré, J. J., & Sorensen, D. C. (1983). Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing, 4*(3), 553–572.
Murphy, K., & Mian, S. (1999). In *Modelling gene expression data using dynamic Bayesian networks*. Berkeley: University of California.
Nacher, J. C., & Akutsu, T. (2013). Structural controllability of unidirectional bipartite networks. *Scientific Reports, 3*.
Nepusz, T., & Vicsek, T. (2012). Controlling edge dynamics in complex networks. *Nature Physics, 8*(7), 568–573.
Opgen-Rhein, R., & Strimmer, K. (2007). Learning causal networks from systems biology time course data: An effective model selection procedure for the vector autoregressive process. *BMC Bioinformatics, 8*(2), S3.
Park, P. (2007). Wan Ko J: Stability and robust stability for systems with a time-varying delay. *Automatica, 43*(10), 1855–1858.
Perrin, B.-E., Ralaivola, L., Mazurie, A., Bottani, S., Mallet, J., & d'Alché–Buc, F. (2003). Gene networks inference using dynamic Bayesian networks. *Bioinformatics, 19*(suppl 2), ii138–ii148.
Poland, G. A., Kennedy, R. B., McKinney, B. A., Ovsyannikova, I. G., Lambert, N. D., Jacobson, R. M., et al. (2013). Vaccinomics, adversomics, and the immune response network theory: Individualized vaccinology in the 21st century. In *Seminars in immunology* (pp. 89–103). Elsevier.
Porter, B. (1972). Probability of stability of a class of linear dynamical systems. *International Journal of Systems Science, 3*(1), 113–116.
Porter, B., & Crossley, T. R. (1972). Probability of stability of linear dynamical systems. *Kybernetes, 1*(4), 207–209.
Pötzsche, C., Siegmund, S., & Wirth, F. (2003). A spectral characterization of exponential stability for linear time-invariant systems on time scales. *Discrete Continuous Dynamic Systems, 9*(5), 1223–1241.
Pujol, A., Mosca, R., Farrés, J., & Aloy, P. (2010). Unveiling the role of network and systems biology in drug discovery. *Trends in Pharmacological Sciences, 31*(3), 115–123.
Qian, X., & Dougherty, E. R. (2013). Validation of gene regulatory network inference based on controllability. *Frontiers in Genetics, 4*.
Raghavan, V., & Barmish, B. R. (2006). Stability of systems with random parameters. In *In decision and control, 2006 45th IEEE Conference on* (San Diego, CA, USA).
Rajapakse, I., Groudine, M., & Mesbahi, M. (2011). Dynamics and control of state-dependent networks for probing genomic organization. *Proceedings of the National Academy of Sciences, 108*(42), 17257–17262.
Rangel, C., Angus, J., Ghahramani, Z., Lioumi, M., Sotheran, E., Gaiba, A., et al. (2004). Modeling T-cell activation using gene expression profiling and state-space models. *Bioinformatics, 20*(9), 1361–1372.
Reehorst, A., Chung, J., Potapczuk, M., & Choo, Y. (2000). Study of icing effects on performance and controllability of an accident aircraft. *Journal of Aircraft, 37*(2), 253–259.
Rider, B., & Sinclair, C. D. (2014). Extremal laws for the real Ginibre ensemble. *The Annals of Applied Probability, 24*(4), 1621–1651.
Ruths, J., & Ruths, D. (2014). Control profiles of complex networks. *Science, 343*(6177), 1373–1376.
Sastry, S. (1999). *Nonlinear systems: Analysis, stability, and control*. New York: Springer. vol. 10.
Schafer, J., & Strimmer, K. (2005). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics, 21*(6), 754–764.
Shimamura, T., Imoto, S., Yamaguchi, R., Fujita, A., Nagasaki, M., & Miyano, S. (2009). Recursive regularization for inferring gene networks from time-course gene expression profiles. *BMC Systems Biology, 3*, 41.
Shmulevich, I., Dougherty, E. R., Kim, S., & Zhang, W. (2002). Probabilistic boolean networks: A rule-based uncertainty model for gene regulatory networks. *Bioinformatics, 18*(2), 261–274.
Shojaie, A., & Michailidis, G. (2010). Discovering graphical Granger causality using the truncating lasso penalty. *Bioinformatics, 26*(18), i517–i523.
Silverstein, J. W. (1994). The spectral radii and norms of large dimensional non-central random atrices matrices. *Stochastic Models, 10*(3), 525–532.
Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., et al. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Molecular Biology of the Cell, 9*(12), 3273–3297.
Steuer, R., Kurths, J., Daub, C. O., Weise, J., & Selbig, J. (2002). The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics, 18*(2), S231–S240.
Stuart, J. M., Segal, E., Koller, D., & Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science, 302*(5643), 249–255.
Tang, Y., Qian, F., Gao, H., & Kurths, J. (2014). Synchronization in complex networks and its application – A survey of recent advances and challenges. *Annual Reviews in Control, 38*(2), 184–198.
Tao, T., & Vu, V. (2008). Random matrices: The circular law. *Communications in Contemporary Mathematics, 10*(2), 261–307.
Wang, W.-X., Ni, X., Lai, Y.-C., & Grebogi, C. (2012). Optimizing controllability of complex networks by minimum structural perturbations. *Physical Review E, 85*(2), 026115.
Wright, S., & Nocedal, J. (1999). *Numerical optimization*. New York: Springer.
Wuchty, S. (2014). Controllability in protein interaction networks. *Proceedings of the National Academy of Sciences, 111*(19), 7156–7160.
Wu, S., Liu, Z.-P., Qiu, X., & Wu, H. (2014). Modeling genome-wide dynamic regulatory network in mouse lungs with influenza infection using high-dimensional ordinary differential equations. *PLOS ONE, 9*(5), e95276.
Wu, S., & Wu, H. (2013). More powerful significant testing for time course gene expression data using functional principal component analysis approaches. *BMC Bioinformatics, 14*(1), 6.
Xue, H., Miao, H., & Wu, H. (2010). Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Annals of Statistics, 38*(4), 2351–2387.
Yan-Dong, X., Song-Yang, L., Lv-Lin, H., & Liang, B. (2014). Optimization of robustness of network controllability against malicious attacks. *Chinese Physics B, 23*(11), 118902.
Yeung, M. K. S., Tegnér, J., & Collins, J. J. (2002). Reverse engineering gene networks using singular value decomposition and robust regression. *Proceedings of the National Academy of Sciences, 99*(9), 6163–6168.
Yuan, Y. (2000). A review of trust region algorithms for optimization. In *ICM99: Proceeding of the fourth international Congress on Industrial and applied mathematics* (pp. 271–282). Oxford University Press.
Yuan, Z., Zhao, C., Di, Z., Wang, W.-X., & Lai, Y.-C. (2013). Exact controllability of complex networks. *Nature Communications, 4*.
Zlatkovic, B. M., & Samardzic, B. (2012). One way for the probability of stability estimation of discrete systems with randomly chosen parameters. *IMA Journal of Mathematical Control and Information, 29*(3), 329–341.
Zou, M., & Conzen, S. D. (2005). A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics, 21*(1), 71–79.