

Research Article

Several Indicators of Critical Transitions for Complex Diseases Based on Stochastic Analysis

Gang Wang,¹ Yuanyuan Li,¹ and Xiufen Zou^{1,2}

¹*School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China*

²*Computational Science Hubei Key Laboratory, Wuhan University, Wuhan 430072, China*

Correspondence should be addressed to Xiufen Zou; xfzou@whu.edu.cn

Received 18 April 2017; Revised 19 June 2017; Accepted 19 June 2017; Published 1 August 2017

Academic Editor: Luca Faes

Copyright © 2017 Gang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Many complex diseases (chronic disease onset, development and differentiation, self-assembly, etc.) are reminiscent of phase transitions in a dynamical system: quantitative changes accumulate largely unnoticed until a critical threshold is reached, which causes abrupt qualitative changes of the system. Understanding such nonlinear behaviors is critical to dissect the multiple genetic/environmental factors that together shape the genetic and physiological landscape underlying basic biological functions and to identify the key driving molecules. Based on stochastic differential equation (SDE) model, we theoretically derive three statistical indicators, that is, coefficient of variation (CV), transformed Pearson's correlation coefficient (TPC), and transformed probability distribution (TPD), to identify critical transitions and detect the early-warning signals of the phase transition in complex diseases. To verify the effectiveness of these early-warning indexes, we use high-throughput data for three complex diseases, including influenza caused by either H3N2 or H1N1 and acute lung injury, to extract the dynamical network biomarkers (DNBs) responsible for catastrophic transition into the disease state from predisease state. The numerical results indicate that the derived indicators provide a data-based quantitative analysis for early-warning signals for critical transitions in complex diseases or other dynamical systems.

1. Introduction

A sudden change of a system is a recurrent phenomenon in many complex diseases, which often occurs at a critical threshold, or the so-called “tipping point,” and can be interpreted as the fact that the system shifts abruptly from one asymptotically stable equilibrium to another one [1]. During the progression of many complex diseases, for example, in chronic diseases such as cancer, the deterioration is abrupt [2, 3]. In other words, there exists a sudden catastrophic shift during the progress of gradual health deterioration that results in a drastic transition to a disease state. Intuitively, between the healthy state and the disease state, there is a critical state which is the so-called “critical transition” [4, 5]. Therefore, in recent years, there has been a growing interest in developing quantitative and qualitative approaches for detecting early-warning signals to avoid such undesirable transitions. The model-based and network-based approaches, respectively, have been used to extract warning signals from

observed time series. Therefore, a variety of empirical studies based on analysis of time-series data have suggested that some statistical signatures, such as variance, Pearson's correlation coefficient (PCC), autocorrelation, and coefficient of variation (CV), may be used to predict the critical transition [6–14]. To extend and generalize the proposed indicators and find new statistical characteristics, in this study, we use a generic stochastic differential equation (SDE) model to represent the complex system based on time-series data with noise and try to understand the critical transition from a mathematical viewpoint. Combining the qualitative theory of fast-slow dynamical systems, probability theory, and statistical analysis, we theoretically prove that three statistical indicators, that is, coefficient of variation (CV), transformed Pearson's correlation coefficient (TPC), and transformed probability distribution (TPD), can distinguish the early-warning signals of the critical transition in complex systems. Finally, we use real high-throughput data for three complex diseases, including influenza caused by either H3N2

virus [15] or H1N1 virus [16] and acute lung injury induced by carbonyl chloride inhalation exposure [17], to demonstrate the applicability of these early-warning indexes for extracting the dynamical network biomarkers (DNBs) responsible for catastrophic transition into the disease state from predisease state.

2. Methods

2.1. Preliminaries for Dynamical Systems and Stochastic Differential Equations. The dynamics for the progression of a complex system can be expressed by the following nonlinear continuous-time equations (ODEs):

$$\frac{dZ(t)}{dt} = f(Z(t); P), \quad (1)$$

where $Z(t) = (z_1(t), z_2(t), \dots, z_n(t))^T \in \mathbb{R}$ is an n -dimensional state vector at time instant t with $t \in \mathbb{R}_+$ and $P = (p_1, p_2, \dots, p_s)^T \in \mathbb{R}^s$ is a s -dimensional parameter vector or driving factors. $f : \mathbb{R}^n \times \mathbb{R}^s \rightarrow \mathbb{R}^n$ is an n -dimensional nonlinear function vector. Suppose that the bifurcation of the nonlinear dynamical system occurs at the bifurcation point (Z^*, P^*) , where $Z^* = (z_1^*, z_2^*, \dots, z_n^*)^T$ and $P^* = (p_1^*, p_2^*, \dots, p_s^*)^T$. For system (1) near Z^* , before P reaches P^* , the system is supposed to stay at an asymptotically stable equilibrium Z^* and therefore all the real parts of the eigenvalues are negative [18, 19].

Generally, there exist the noises in the real system [20, 21]. Thus, we add the white Gaussian noises to the original ordinary differential system (1) and obtain the following stochastic differential equations:

$$\frac{dZ(t)}{dt} = f(Z(t); P) + H \frac{dB(t)}{dt}, \quad (2)$$

where $B(t) = (B_1(t), B_2(t), \dots, B_m(t))^T$ are m -dimensional standard Brownian motion.

Consider the linearized equations of (1) with the perturbations of white Gaussian noises near Z^* . Namely, introducing new variables $Y(t) = (y_1(t), y_2(t), \dots, y_n(t))^T$ and a transformation matrix S , that is, $Y(t) = S^{-1}(Z(t) - Z^*)$ (see the detailed process in Supplementary Information A1 in the Supplementary Material available online at <https://doi.org/10.1155/2017/7560758>), we have

$$\frac{dY(t)}{dt} = \Lambda(P)Y(t) + G \frac{dB(t)}{dt}, \quad (3)$$

where $\Lambda(P)$ is a diagonal matrix which is the Jordan normal form for the Jacobian matrix $(\partial f(Z; P)/\partial Z)|_{Z=Z^*}$ and

$$G = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \vdots & \vdots & & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nm} \end{pmatrix} \in \mathbb{R}^{n \times m}, \quad (4)$$

with

$$\sum_{l=1}^m g_{il}g_{jl} \neq 0, \quad i, j = 1, 2, \dots, n. \quad (5)$$

Actually, there are typically three cases arising in diagonal matrix $\Lambda(P)$. For the sake of simplicity, here we only illustrate case 1, that is, $\Lambda(P) = \text{diag}(\lambda_1(P), \lambda_2(P), \dots, \lambda_n(P))$ with each λ_i negative. The other two cases are described in Supplementary Information A1: a generic model in abstract phase space and some preliminaries.

Among the eigenvalues of Λ in case 1, the largest one, say λ_1 , first approaches 0 when parameter $P \rightarrow P^*$. This eigenvalue characterizes the rate of the system around the steady state approximately and is called the dominant eigenvalue. We can calculate the solution of the stochastic differential equation (3):

$$y_i(t) = e^{\lambda_i(t-t_0)} y_i(t_0) + \sum_{k=1}^m g_{ik} \int_{t_0}^t e^{\lambda_i(t-s)} dB_k(s), \quad (6)$$

$$i = 1, 2, \dots, n.$$

It is easy to obtain that

$$E(y_i(t)) = e^{\lambda_i(t-t_0)} y_i(t_0),$$

$$\text{var}(y_i(t)) = \left(\sum_{k=1}^m g_{ik}^2 \right) \frac{1 - e^{2\lambda_i(t-t_0)}}{-2\lambda_i}, \quad (7)$$

$$\text{cov}(y_i(t), y_j(t)) = \left(\sum_{k=1}^m g_{ik}g_{jk} \right) \frac{1 - e^{(\lambda_i+\lambda_j)(t-t_0)}}{-(\lambda_i + \lambda_j)},$$

$$i, j = 1, 2, \dots, n.$$

Here when the dominant eigenvalue $\lambda_1 \rightarrow 0^-$ because of the change in the parameter values, $\text{var}(y_1) \rightarrow +\infty$, but $\text{var}(y_i)$ is bounded for $i \neq 1$ because $\lambda_i < \lambda_1 < 0$, ($i = 2, \dots, n$). For any $i \neq j$, $\text{cov}(y_i, y_j)$ tends to have a positive bounded value.

Returning to the original variables Z whose elements can be expressed by $z_i = \sum_{j=1}^n s_{ij}y_j + z_i^*$, where s_{ij} is the element of the linear transformation S , we have (Supplementary Information A2)

$$E(z_i) = \sum_{j=1}^n s_{ij}E(y_j) + z_i^*,$$

$$\text{cov}(z_i, z_j) = s_{i1}s_{j1} \text{var}(y_1) + \cdots + s_{in}s_{jn} \text{var}(y_n)$$

$$+ \sum_{k=1}^n \sum_{m=1, m \neq k}^n s_{ik}s_{jm} \text{cov}(y_k, y_m), \quad (8)$$

$$\text{PCC}(z_i, z_j) = \frac{\text{cov}(z_i, z_j)}{\sqrt{\text{var}(z_i) \text{var}(z_j)}}.$$

Notice that variable y_1 is related to the dominant eigenvalue λ_1 . From the above equations, it is obvious that as $\lambda_1 \rightarrow$

0^- , the variance, that is, $\text{var}(z_i) = \text{cov}(z_i, z_j)$, increases greatly, or $\text{var}(z_i) \rightarrow +\infty$ if s_{i1} is not vanishing, and $|\text{PCC}(z_i, z_j)|$ approaches 1 drastically if both s_{i1} and s_{j1} are nonzero. In this case, variables z_i and z_j are directly affected by the dominant eigenvalue. A group composed of such variables is called the dominant group in the network. On the other hand, as $\lambda_1 \rightarrow 0^-$, $|\text{PCC}(z_i, z_j)|$ between the dominant group (e.g., including z_i) and others (e.g., including z_j , which does not belong to the dominant group) reduces to zero if $s_{i1} \neq 0$ but $s_{j1} = 0$.

In addition, from dynamical viewpoint, a system that leads to critical transition is due to the change of the dominant eigenvalue. This viewpoint can be understood from the process of the theoretical derivation in Supplementary Information A1. In other words, the fact that the eigenvalue approaches zero when the critical transition occurs is very important. Simultaneously, a very necessary question arises: how can we describe the concept “related to the dominant eigenvalue”? In order to achieve our goal, using the theory of the Taylor expansion and the Jordan normal form of matrix, the key point is that the model can be projected onto the center manifold under the linear transformation. Then by identifying the coefficient with respect to the dominant eigenvalue being zero or not, we can determine which group of variables can lead to critical transition. Therefore, all the variables z_i with $s_{i1} \neq 0$ form a group, which indicates that a transition from a steady state to another steady state will occur. It is easy to see that this group characterizes the statistical features of the underlying system. Especially in complex diseases, the molecules in this group are strongly and dynamically correlated in the predisease state (see Section 2.3), which can be viewed as DNBs for early-warning signals.

2.2. Method for Identifying Dynamical Network Biomarkers (DNBs) for Complex Diseases Based on Real Data. The evolutionary process of the complex diseases can be generally categorized into three states: that is, a normal state, a predisease state, and a disease state [20–25]. The normal state is a steady state that represents a relatively healthy stage during which the disease is under control. The predisease state corresponds to the critical state before transition to the disease state. In this stage, it may be reversible to the healthy stage under appropriate and therapeutic interventions [26]. Therefore, it is of great importance to identify DNBs representing an early-warning signals of disease progression, or as a leading biomarker that drives the system into the disease state. In this paper, the identification of DNBs includes four steps. For a given real dataset, the first step is to calculate the indicators CV and TPD of each protein in the symptomatic group and the asymptomatic group, respectively. Then, in second step, the t -test method is used to select the significantly different proteins of the indicators CV and TPD between symptomatic and asymptomatic groups as candidates. In third step, we choose the intersection of candidate of indicators CV and TPD. In final step, the proposed TPC is calculated for above candidates and the distinguished DNBs are selected.

2.3. Data Collection. Three gene expressing profiling datasets were downloaded from the NCBI GEO database (GSE30550, GSE52428, and GSE2565). Probe sets from these datasets lacking the corresponding gene symbols were ignored in our analysis. The expression values of the probe sets that mapped to the same gene were averaged. The diseases from the first two datasets were two influenza strains, H3N2 and H1N1, whereas the other one dataset was for acute lung injury.

The biological data GSE3055024 contained 17 healthy subjects who received intranasal inoculations of influenza H3N2/Wisconsin [15]. Nine of these 17 subjects developed severe infection symptoms, and the other 8 subjects remained in good health. Gene expression profiles were measured in whole peripheral blood drawn from all subjects approximately every 8 hours postinoculation (hpi) through 108 hpi. In total, 268 gene expression profiles were obtained for all subjects at 16 time points, including baseline (–24 hpi). The gene expression profiles of subject 8 at 21 hpi, subject 13 at baseline and 36 hpi, and subject 17 at 36 hpi were missing.

The biological dataset GSE5242825 contained 24 healthy subjects who received intranasal inoculations of influenza H1N1/Brisbane [16]. 12 of these 24 subjects developed severe infection symptoms and 11 subjects remained in good health. One subject was excluded from all analyses because the symptoms began late and were thought to be related to infection acquired in the facility rather than from the primary infection related to inoculation. Gene expression profiles were measured as described for the biological dataset GSE30550.

The biological dataset GSE256526 contained 6 control samples (control group) and 6 case samples (case group) [17]. CD-1 male mice were divided into two groups that were exposed to air or phosgene. Lung tissues were collected from air- or phosgene-exposed mice at 0.5, 1, 4, 8, 12, 24, 48, and 72 hours (h) after exposure.

3. Results

3.1. Statistical Characteristics near Critical Phase Transition. In this paper, we focus on the following three statistical indexes.

For the convenience of the following statements, we will introduce some symbols and equations mentioned in the “Methods”; that is,

$$\frac{dY(t)}{dt} = \Lambda(P)Y(t) + G\frac{dB(t)}{dt}, \quad (9)$$

where $Y(t) = (y_1(t), y_2(t), \dots, y_n(t))^n \in \mathbb{R}^n$, $\Lambda(P) \in \mathbb{R}^{n \times n}$, $G \in \mathbb{R}^{n \times m}$, and $B(t) = (B_1(t), B_2(t), \dots, B_m(t))^T$ are m -dimensional standard Brownian motion.

In addition,

$$z_i(t) = \sum_{j=1}^n s_{ij}y_j(t) + z_i^*, \quad i = 1, 2, \dots, n, \quad (10)$$

where

$$\begin{aligned} Z(t) &= (z_1(t), z_2(t), \dots, z_n(t))^T \in \mathbb{R}^n, \\ Z^* &= (z_1^*, z_2^*, \dots, z_n^*)^T \in \mathbb{R}^n, \\ P &= (p_1, p_2, \dots, p_s)^T \in \mathbb{R}^s, \\ P^* &= (p_1^*, p_2^*, \dots, p_s^*)^T \in \mathbb{R}^s, \\ S &= (s_{ij})_{n \times n} \in \mathbb{R}^{n \times n}. \end{aligned} \quad (11)$$

(1) *The Coefficient of Variation CV*. For a given variable X , we have the following equation for CV:

$$\text{CV}(X) = \frac{\sqrt{\text{var}(X)}}{E(X)}, \quad (12)$$

where $\text{var}(X)$ and $E(X)$ are the variance and mean value for X , respectively.

(2) *Transformed Pearson's Coefficient (TPC)*. As we know, $-1 \leq \text{PCC}(X, Y) \leq 1$, where $\text{PCC}(X, Y)$ is Pearson's correlation coefficient (PCC) between the random variables X and Y . To make it more effective in practice, we can define an indicator with respect to PCC as follows:

$$\text{TPC}(X, Y) = -\ln(1 - |\text{PCC}(X, Y)|). \quad (13)$$

(3) *Transformed Probability Distribution (TPD)*. Obviously,

$$z_i(t) \sim N(\mu_i(t), (\sigma_i(t))^2), \quad i = 1, 2, \dots, n, \quad (14)$$

where

$$\begin{aligned} \mu_i(t) &= E(z_i(t)), \\ \sigma_i(t) &= \sqrt{\text{var}(z_i(t))}. \end{aligned} \quad (15)$$

Thus, the probability density of the random variable $z_i(t)$ is

$$p_i(x) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(x-\mu_i)/2\sigma_i^2}. \quad (16)$$

Then, we can define the indicator TPD as

$$\text{TPD}_i([a, b]) = -\ln\left(\int_a^b p_i(s) ds\right), \quad (17)$$

where $0 < a < b$.

The theoretical results for these indicators are presented in parts below.

Coefficient of Variation (CV). According to the definition of CV, we can consider case 1 described in Supplementary Information A1 for which Λ is diagonal.

Theorem 1. *Suppose that Λ in (9) is diagonal; then*

$$\lim_{t \rightarrow +\infty} \lim_{P \rightarrow P^*} \text{CV}(z_i(t)) = \begin{cases} +\infty & s_{i1} \neq 0, \\ \text{A finite number} & s_{i1} = 0. \end{cases} \quad (18)$$

Proof. A detailed proof is given in Supplementary Information A2.1. \square

We recall that $z_i(t) = \sum_{j=1}^n s_{ij} y_j(t) + z_i^*$, $i = 1, 2, \dots, n$. $s_{i1} \neq 0$ implies that $z_i = \sum_{l=1}^n s_{il} y_l(t) + z_i^*$ is related to the dominant eigenvalue. $s_{i1} \neq 0$ and $s_{i2} = 0$ mean that $z_{i_1} = \sum_{l=1}^n s_{i_1 l} y_l(t) + z_{i_1}^*$ may lead to the critical transition and $z_{i_2} = \sum_{l=2}^n s_{i_2 l} y_l(t) + z_{i_2}^*$ may not lead to the critical transition. For the convenience to discuss in the following sections, we denote $z_{i_1}(t)$ and $z_{i_2}(t)$ as a biomarker and a nonbiomarker, respectively. Then, we can prove the following corollary.

Corollary 2. *As the parameter P approaches the bifurcation value P^* , that is, the system is close to the critical transition,*

- (1) *there are no drastic changes for the coefficient of variation CV of nonbiomarkers;*
- (2) *the coefficient of variation CV for biomarkers is much larger than the coefficient of variation CV for nonbiomarkers;*
- (3) *the coefficient of variation CV for biomarkers drastically increases.*

Proof. A detailed proof is given in Supplementary Information A2.1. \square

Transformed Pearson's Correlation Coefficient (TPC). According to the definition of TPC, we can derive the following theorem.

Theorem 3. *Suppose that Λ in (9) is diagonal; then*

$$\lim_{t \rightarrow +\infty} \lim_{P \rightarrow P^*} \text{TPC}(z_i(t), z_j(t)) = \begin{cases} +\infty & s_{i1} \neq 0, s_{j1} = 0, \\ 0 & s_{i1} = 0, s_{j1} = 0, \\ \text{A finite number} & s_{i1} = 0, s_{j1} = 0. \end{cases} \quad (19)$$

Proof. A detailed proof is given in Supplementary Information A2.2. \square

Based on Theorem 3, we can prove the following corollary.

Corollary 4. *As the parameter P approaches the bifurcation value P^* ,*

- (1) *there are no drastic changes for the indicator TPC between nonbiomarkers;*
- (2) *the indicator TPC between biomarkers is much larger than TPC between nonbiomarkers;*
- (3) *the indicator TPC between a biomarker and a nonbiomarker is much smaller than TPC between nonbiomarkers;*
- (4) *the indicator TPC between biomarkers drastically increases.*

TABLE 1: DNBs for H3N2, H1N1, and lung with respect to the indicators CV, TPC, and TPD.

Data for diseases	Number for CV	Number for TPD	DNBs
H3N2	264	116	22 APOL6, CASP1, CNP, CXCL10, DHX58, DRAP1, DYNLT1, FAM46A, GEMIN4, GORASP1, IFITM1, IRF7, PLSCR1, RTP4, SAMD9, SAMHD1, SIGLEC1, TAF1C, TLR7, TNFAIP6, TREX1, ZBP1
H1N1	32	6	22 ACP6, BTG1, CCNA1, DDX18, DKC1, H2AFV, HMGNI, IRF7, ITK, LAX1, NPAT, NR2C1, NUCBI, PFN2, POLR1C, RBM4B, RPS2, SEPT4, SLBP, SPI10, STMN1, VPRBP
Lung	34	104	60 Abcd3, Actn1, Adcy8, Adss, Anxa1, Aplp2, Aqp1, Atp6v1d, Capn1, Clstn1, Csf1r, Dapk1, Ddx39, Ensmusg00000050347, Ereg, Faf1, Fzd2, G6pd2, Gimap4, Glrx, Gnb1, Gp49a, Grem2, Gtf2i, H1f0, Hhip, Hist2h2bb, Hnrnpd, Hpirt1, Htra1, Kcnq1, Klhl13, Lox, Lrg1, Macf1, Mcee, Mmp19, Nagk, Ncl, Nr2f6, Nrp1, Phlpp, Pla2g15, Prelp, Prpf40a, Psmal, Psm7, Psmb5, Psmc3, Psmd13, Psmd4, Psme4, Rad17, Rad23b, Sin3b, Stxbp1, Thbs3, Tjp2, Ulk2, Wbp1

Proof. A detailed proof is given in Supplementary Information A2.2. \square

Transformed Probability Distribution (TPD). According to the definition of TPD, we have Theorem 5.

Theorem 5. Suppose that Λ in (9) is diagonal; then

$$\lim_{t \rightarrow +\infty} \lim_{P \rightarrow P^*} \text{TPD}_i([a, b]) = \begin{cases} +\infty & s_{i1} \neq 0, \\ \text{A finite number} & s_{i1} = 0, \end{cases} \quad (20)$$

where $0 < a < b$.

Proof. A detailed proof is given in Supplementary Information A2.3. \square

Similarly, we can easily obtain the corresponding corollary.

Corollary 6. As the parameter P approaches the bifurcation value P^* ,

- (1) there are no drastic changes for the indicator TPD of nonbiomarkers;
- (2) the indicator TPD for biomarkers is much larger than TPD for nonbiomarkers;
- (3) the indicator TPD for biomarkers drastically increases.

3.2. Results for Predicting DNBs in Real Datasets for Complex Diseases. Three real datasets for complex diseases, that is, H3N2, H1N1, and acute lung injury, respectively, are used to illustrate the above theoretical results. According to the steps proposed in the above part “Method for Identifying Dynamical Network Biomarkers (DNBs) for Complex Diseases Based on Real Data,” we calculate three indicators CV, TPD, and TPC for these datasets. The obtained DNBs are listed in Table 1. The second column *number for CV* and third column *number for TPD* represent the numbers of proteins chosen by CV and TPD, respectively. The last column denotes the final selected DNBs by combining CV, TPC, and TPD. In the following, we only present the application on the dataset of the complex disease H3N2 as an example. The detailed descriptions of the results on the other two datasets H1N1 and lung are given in Figures S1–S9 and Figures S10–S17 in Supplementary Information B, respectively.

To further present the results for the H3N2, the change curves of two indicators CV and TPD for the chosen 22 proteins are depicted in Figures 1 and 2, respectively. From these two figures, we can observe that as the critical transition occurs, that is, the time evolves towards $t = 45$ h (see the vertical black line), the CV and TPD for 22 proteins above are significantly increasing in the symptomatic group. But they have no obvious change in the asymptomatic group. Therefore, these 22 proteins can be viewed as DNBs for early-warning signals.

To explore the biological link between arbitrary two proteins, the analysis for the indicator TPC is implemented on these identified DNBs according to the coefficient of variation CV using the gene expression datasets, respectively.

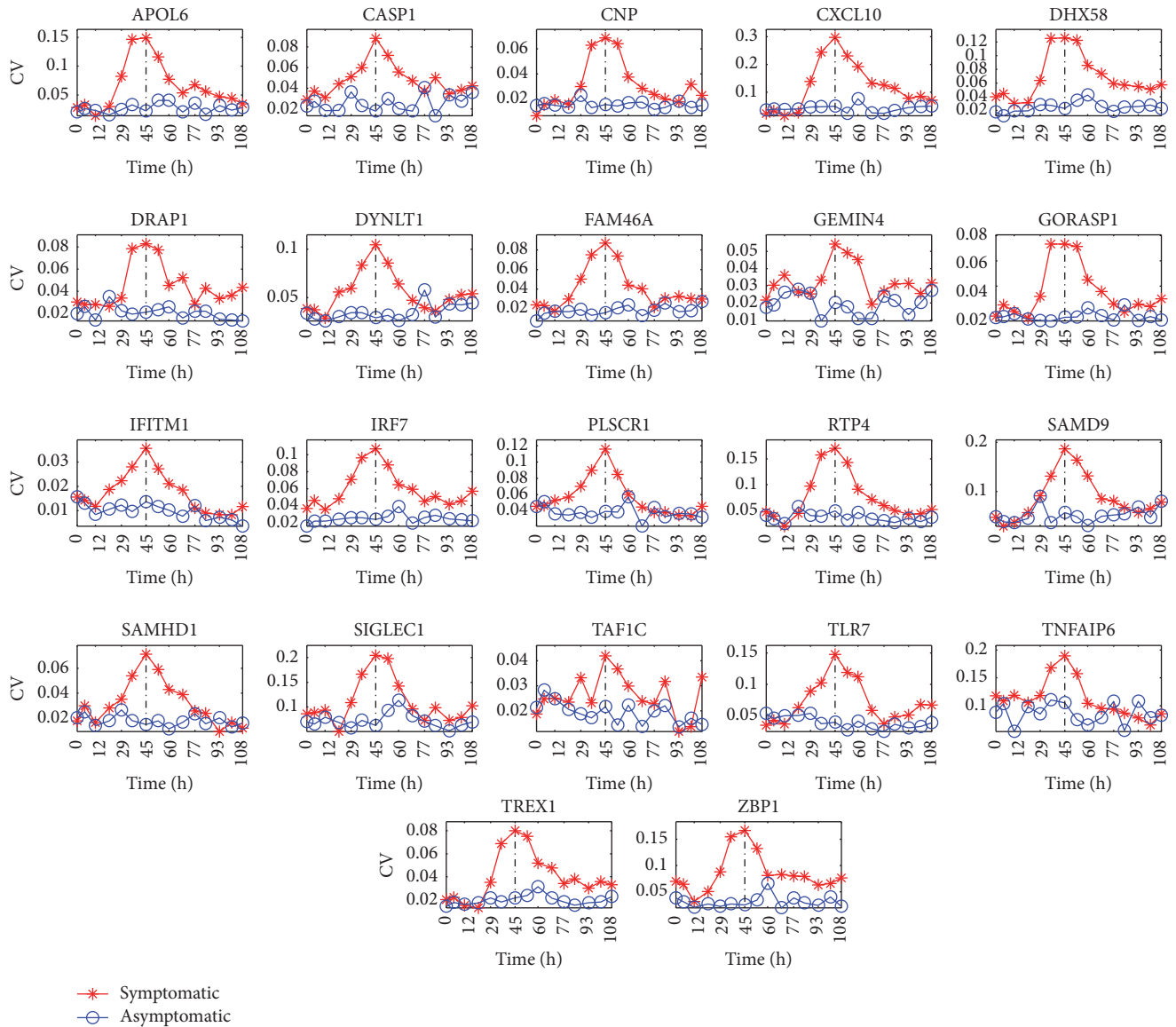


FIGURE 1: The change of the coefficient of variation CV for 22 proteins: APOL6, CASP1, CNP, CXCL10, DHX58, DRAP1, DYNLT1, FAM46A, GEMIN4, GORASP1, IFITM1, IRF7, PLSCR1, RTP4, SAMD9, SAMHD1, SIGLEC1, TAF1C, TLR7, TNFAIP6, TREX1, and ZBP1. The x -axis denotes the time (unit: h). It indicates that as the critical transition occurs, that is, the time evolves towards $t = 45$ h (see the vertical black line), the coefficient of variation CV for 22 proteins above significantly increases in the symptomatic group and has no obvious change in the asymptomatic group. Therefore, these 22 proteins can be viewed as DNBs for early-warning signals.

Figure 3 shows the change of the indicator TPC of a small part of the protein pairs from the DNBs from Figure 1 for the disease H3N2. The TPC of other protein pairs for H3N2 is presented in Supplementary Information B. It is obvious that the identified DNBs form a strongly correlated subnetwork to provide the significant warning signal near the critical state (45 h).

From Figures 1–3, it is easy to find that during the predisease state, for DNBs, there are significant differences between the members of the symptomatic group and those of the asymptomatic group which behave in a considerably different manner in terms of three indicators CV, TPD, and TPC that are theoretically derived. However, after the system

is driven into the disease state, interestingly the members of both the symptomatic group and the asymptomatic group appear to behave in a manner similar to each other (see the curve after the time $t = 45$ h). Therefore, we can conclude that during the normal state and the disease state, for each protein, the stochastic behaviors for the symptomatic group and the asymptomatic group display similarity. Then, the significant differences for the stochastic behaviors for the symptomatic and asymptomatic groups of DNBs only occur during the predisease state. From this statement, we know that the stochastic signals, such as the indicators CV, TPD, and TPC, make sense for predicting the impending critical transition.

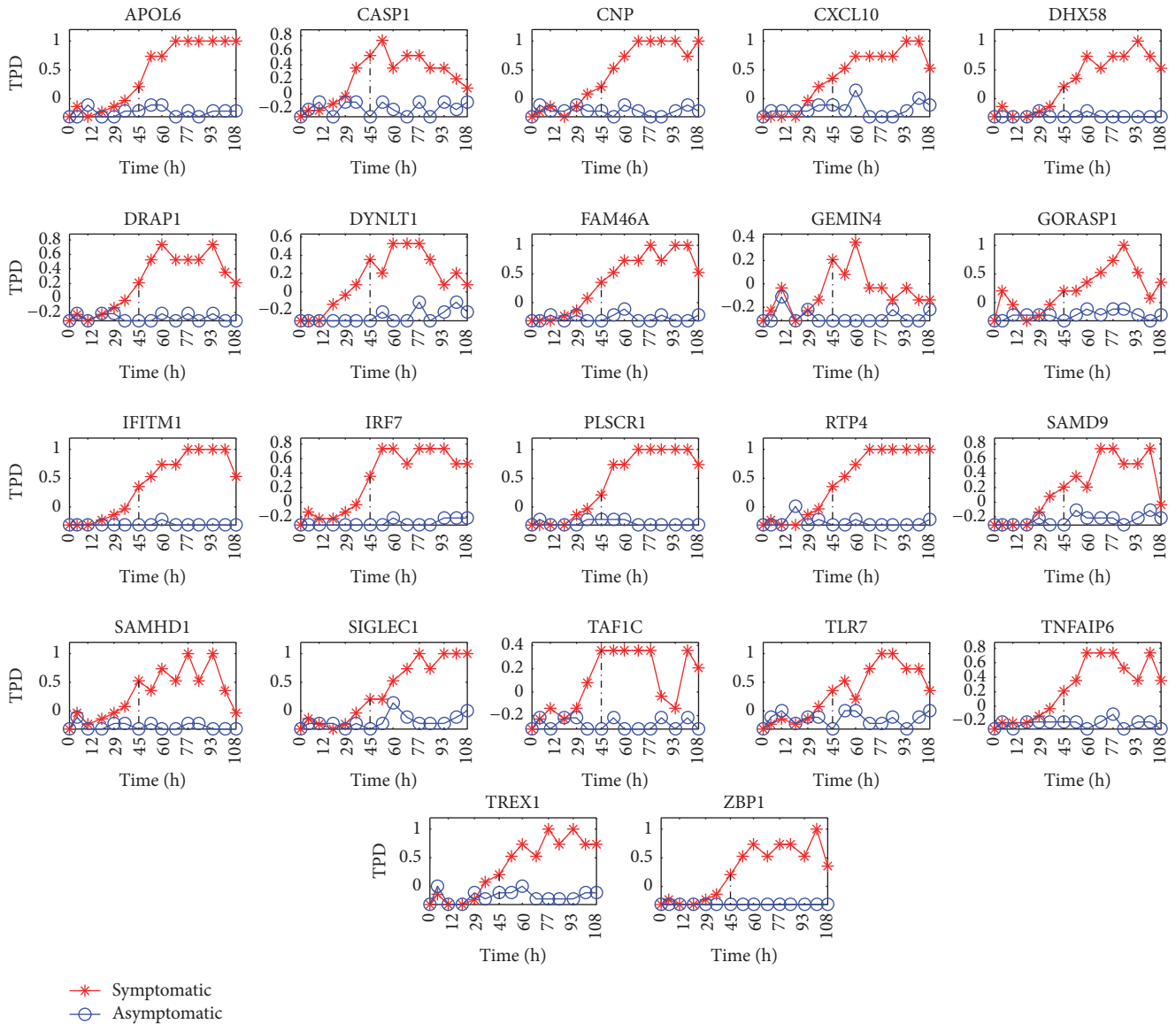


FIGURE 2: The change of the indicator TPD for 22 proteins: APOL6, CASP1, CNP, CXCL10, DHX58, DRAP1, DYNLT1, FAM46A, GEMIN4, GORASP1, IFITM1, IRF7, PLSCR1, RTP4, SAMD9, SAMHD1, SIGLEC1, TAF1C, TLR7, TNFAIP6, TREX1, and ZBP1. The x -axis denotes the time (unit: h). It indicates that as the critical transition occurs, that is, the time evolves towards $t = 45$ h (see the vertical black line), the indicator TPD for 22 proteins above significantly increases. Therefore, these 22 proteins can be viewed as DNBs for early-warning signals.

3.3. *The Criteria to Predict DNBs in Real Datasets for Complex Diseases.* Through theoretical analysis presented in Corollaries 2, 4, and 6 in Section 3.1, we obtain three statistical indicators CV, TPC, and TPD that can distinguish the biomarkers from nonbiomarkers. To validate the theoretical results, we calculate these three indicators for three real datasets. Figures 1–3 demonstrate that these three indicators can clearly distinguish the biomarkers and nonbiomarkers when they reach the predisease state. These results can help us to detect the early-warning signals for complex diseases. From Figures 1–3, we can also observe that three indicators exhibit different behaviors after critical transition. That is, the values of the indicators CV and TPC only display a peak at the critical transition point (see $t = 45$ h) while the value of the

indicator TPD keeps a high value after the critical transition compared to the value before the critical transition (see $t = 45$ h). We can give their explicit explanations intuitively from the perspectives of the qualitative theory of ordinary differential equations and statistics. First, far before $t = 45$ h, the system is in the normal state or in an asymptotically stable state from the perspective of the ordinary differential equation. That is, there exist only stable manifolds, or the state has strong attraction or strong robustness. In statistics, this corresponds to the phenomenon that the system fluctuates weakly. Thus, in this state, the fluctuation is weak. Just far after $t = 45$ h, the system is in the disease state or in another asymptotically stable state. Due to the same reason, in this state, the fluctuation is weak. However, at the bifurcation

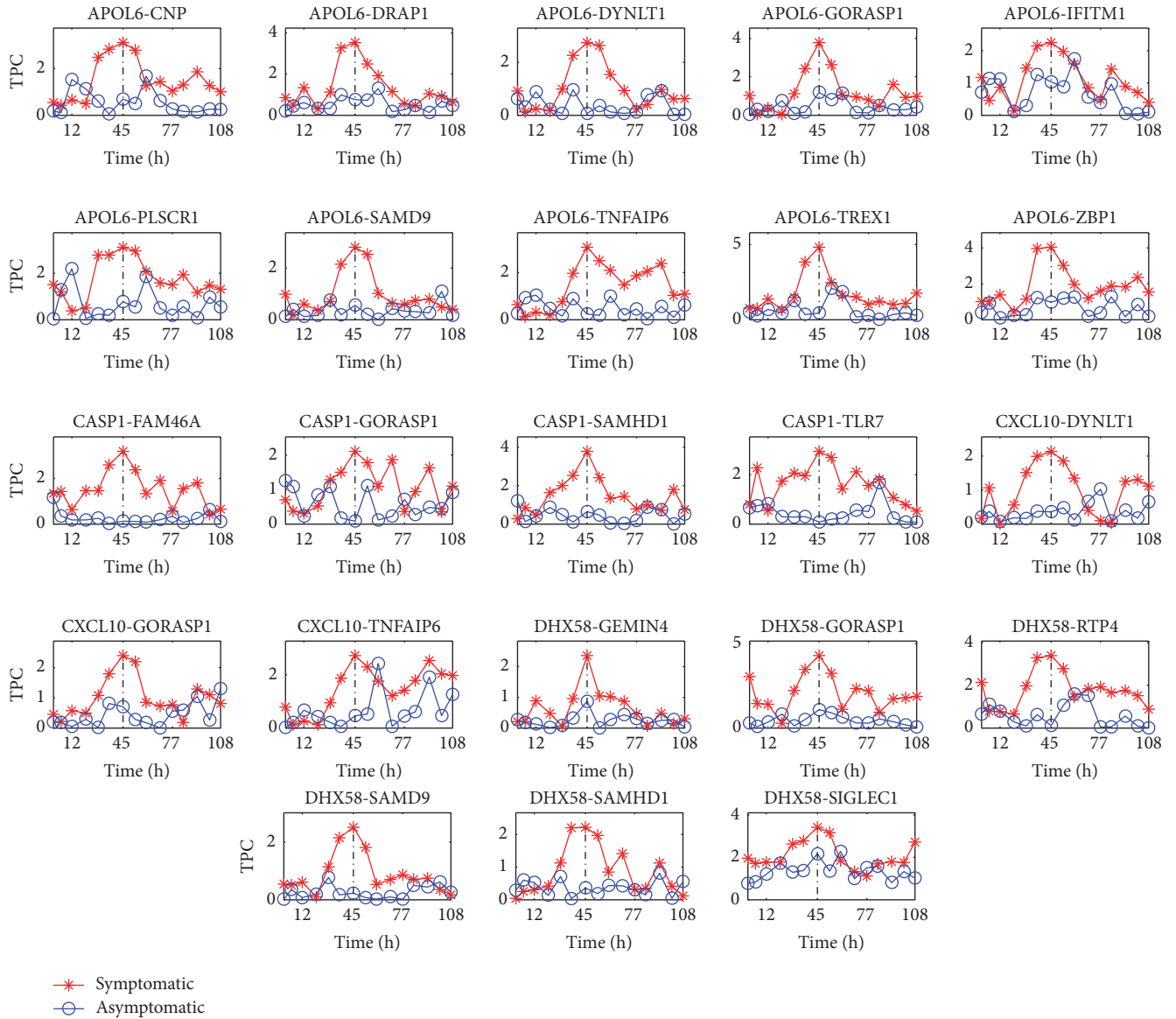


FIGURE 3: The change of the indicator TPC for 23 protein pairs: APOL6-CNP, APOL6-DRAP1, APOL6-DYNLT1, APOL6-GORASP1, APOL6-IFITM1, APOL6-PLSCR1, APOL6-SAMD9, APOL6-TNFAIP6, APOL6-TREX1, APOL6-ZBP1, CASP1-FAM46A, CASP1-GORASP1, CASP1-SAMHD1, CASP1-TLR7, CXCL10-DYNLT1, CXCL10-GORASP1, CXCL10-TNFAIP6, DHX58-GEMIN4, DHX58-GORASP1, DHX58-RTP4, DHX58-SAMD9, DHX58-SAMHD1, and DHX58-SIGLEC1. The x -axis denotes the time (time: h).

point, the center manifold occurs. Thus, near $t = 45$ h or in the predisease state, compared with the normal state or the disease state, the system is in a less stable state and has weaker attraction? Thus, in this state, the fluctuation becomes stronger (before $t = 45$ h) or becomes weaker (after $t = 45$ h). In addition, this kind of phenomenon is verified by many real experiments.

Second, in fact, the indicator $TPC(X, Y) = -(1 - |PCC(X, Y)|)$ describes the linear correlation relationship between two random variables X and Y or the similarity relationship between two random variables. Far before $t = 45$ h, the system is in the normal state or in an asymptotically stable state. In this state, each element of the state variable has different recovery rate, and thus the similarity

of their behaviors is weak. Correspondingly, in statistics, the linear correlation relationship is weak. Similarly, we can understand the case in the disease state or after $t = 45$ h. However, near $t = 45$ h or in the predisease state, the dominant eigenvalue is the largest eigenvalue for the Jacobian matrix and often is unique. As we know, the biomarkers are related to the dominant eigenvalue, and their recovery rate can be described by the dominant eigenvalue. Thus, the behaviors of the biomarkers are similar. Thus, in this state, the linear correlation relationship becomes stronger and stronger (before $t = 45$ h) or becomes weaker (after $t = 45$ h). In addition, this kind of phenomenon is verified by many real experiments.

Third, as mentioned above, we know that

$$\text{TPD}_i([a, b]) = -\ln\left(\int_a^b p_i(s) ds\right), \quad (21)$$

where $p_i(\cdot)$ is the probability density of the random variable $z_i(t)$ and a, b can be taken as the maximum and minimum values of the real data at the initial time instant. We know that the critical transition can be understood as the fact that the system state abruptly switches from one asymptotically stable equilibrium to another one at a critical threshold as the complex disease evolves. In general, we can think that the asymptotically stable equilibrium corresponds to a probability distribution in statistics. Far before $t = 45$ h, the system state is stable keeping the value between the maximum value and the minimum value due to the weak fluctuations, and thus the indicator displays little change. As the critical transition occurs (before $t = 45$ h), the values of some of the proteins gradually evolve beyond the interval $[a, b]$ due to the stronger and stronger fluctuations, and then the indicator TPD becomes larger and larger. After $t = 45$ h or in the disease state, the values of the biomarkers are all not in the interval $[a, b]$. That is, in this state, the system state is another asymptotically stable state. Because this attractor is asymptotically stable, the values of biomarkers cannot return to the interval $[a, b]$. Thus, after $[a, b]$, the indicator TPD displays little change and always keeps a high value.

3.4. Functional Analysis of DNBs. To verify the biological significance of the DNBs identified by our method, DAVID was used to carry out functional enrichment analysis, respectively, for the three diseases [27]. The results of function annotation for three datasets showed that the DNBs were significantly enriched in diseases-related biological processes. In H3N2 infection, some DNBs are involved in defense response (e.g., SIGLECI, TNFAIP6, IRF7, SAMHD1, TLR7, DHX58, and CXCL10) and inflammatory response (e.g., SIGLECI, TNFAIP6, IRF7, TLR7, and CXCL10). In H1N1 infection, 4 DNBs (SEPT4, NPAT, STMN1, and CCNA1) participate in cell cycle. In the lung injury study, we found that 2 DNBs (Erge and Aplp2) are associated with the regulation of epidermal growth factor receptor activity and epidermal growth factor receptor signaling pathway. At the pathway enrichment level, some pathways were also highly related to complex diseases. For example, for H3N2 influenza, 5 of 22 DNBs (IRF7, TREX1, CASP1, ZBP1, and CXCL10) are observed in the cytosolic DNA-sensing pathway, which plays an important role in initiating innate immunity and adaptive immunity [28]. In addition, another two significantly enriched pathways include RIG-I-like receptor signaling pathway with 3 DNBs (IRF7, DHX58, and CXCL10) and Toll-like receptor signaling pathway with 3 DNBs (IRF7, TLR7, and CXCL10). For H1N1 influenza, 2 DNBs (IRF7, POLRIC) are also observed in the cytosolic DNA-sensing pathway. For acute lung injury, 7 DNBs (Psmb5, Psmal1, Psmid13, Psmc3, Psmid4, Psme4, and Psmal7) are observed in the pathways of regulating by proteasome mediated degradation, which played a key role in regulating many processes of cellular biology [29, 30].

4. Discussion and Conclusions

To detect the predisease state of complex diseases or identify DNBs, we analyzed a series of statistic indicators, such as CV, TPC, and TPD. Although the previous works relevant to DNBs have proposed how to detect the predisease state of complex diseases by using the statistical methods [8, 22, 31], the rigorous mathematical derivation is not given and the networks based on real data are needed to be constructed. In our viewpoints, we improved the mathematical approaches by using a continuous-time dynamical system. Based on a general dynamical model, we theoretically derive the statistical indicators to detect specific early-warning signals for the predisease state to complex diseases. Furthermore, we have also conducted numerical experiments to identify the DNBs based on high-throughput data for three complex diseases. The function analysis verified the meaningfulness of the detected DNBs. In our paper, some statistical indicators, such as skewness and kurtosis, are not theoretically analyzed [10, 14]. It does not mean that the skewness and kurtosis are not statistical characteristics which are useful for us to detect early-warning signals. The analysis whether they can be used as indicators would be important topic in the future.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors thank Professor Xicheng Zhang in School of Mathematics and Statistics in Wuhan University, China, for his stimulating and helpful discussions on the mathematical proofs. This work was supported by the Major Research Plan of the National Natural Science Foundation of China (no. 91530320), the National Basic Research Program of China (973 Program) (no. 2013CB911100), and the Chinese National Natural Science Foundation (no. 61672388).

References

- [1] J. Gao, B. Barzel, and A.-L. Barabási, "Universal resilience patterns in complex networks," *Nature*, vol. 530, no. 7590, pp. 307–312, 2016.
- [2] S. Huang, I. Ernberg, and S. Kauffman, "Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective," *Seminars in Cell and Developmental Biology*, vol. 20, no. 7, pp. 869–876, 2009.
- [3] S. Jin, F.-X. Wu, and X. Zou, "Domain control of nonlinear networked systems and applications to complex disease networks," *Discrete and Continuous Dynamical Systems. Series B. A Journal Bridging Mathematics and Sciences*, vol. 22, no. 6, pp. 2169–2206, 2017.
- [4] C. Kuehn, "A mathematical framework for critical transitions: bifurcations, fast-slow systems and stochastic dynamics," *Physica D: Nonlinear Phenomena*, vol. 240, no. 12, pp. 1020–1035, 2011.
- [5] C. Kuehn, "A mathematical framework for critical transitions: normal forms, variance and applications," *Journal of Nonlinear Science*, vol. 23, no. 3, pp. 457–510, 2013.

- [6] P. Chen, R. Liu, Y. Li, and L. Chen, "Detecting critical state before phase transition of complex biological systems by hidden Markov model," *Bioinformatics*, vol. 32, no. 14, pp. 2143–2150, 2016.
- [7] S. Jin, Y. Li, R. Pan, and X. Zou, "Characterizing and controlling the inflammatory network during influenza A virus infection," *Scientific Reports*, vol. 4, no. 3799, pp. 1–14, 2014.
- [8] Y. Li, S. Jin, L. Lei, Z. Pan, and X. Zou, "Deciphering deterioration mechanisms of complex diseases based on the construction of dynamic networks and systems analysis," *Scientific Reports*, vol. 5, article 9283, 2015.
- [9] L. Dai, K. S. Korolev, and J. Gore, "Relation between stability and resilience determines the performance of early warning signals under different environmental drivers," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 32, pp. 10056–10061, 2015.
- [10] V. Dakos, S. R. Carpenter, W. A. Brock et al., "Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data," *PLoS ONE*, vol. 7, no. 7, Article ID e41010, 2012.
- [11] S. J. Lade and T. Gross, "Early warning signals for critical transitions: a generalized modeling approach," *PLoS Computational Biology*, vol. 8, no. 2, Article ID e1002360, 2012.
- [12] M. Scheffer, J. Bascompte, W. A. Brock et al., "Early-warning signals for critical transitions," *Nature*, vol. 461, no. 7260, pp. 53–59, 2009.
- [13] L. Dai, K. S. Korolev, and J. Gore, "Slower recovery in space before collapse of connected populations," *Nature*, vol. 496, no. 7445, pp. 355–358, 2013.
- [14] M. Scheffer, S. R. Carpenter, V. Dakos, and E. H. Van Nes, "Generic indicators of ecological resilience: inferring the chance of a critical transition," *Annual Review of Ecology, Evolution, and Systematics*, vol. 46, pp. 145–167, 2015.
- [15] Y. Huang, A. K. Zaas, A. Rao et al., "Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection," *PLoS Genetics*, vol. 7, no. 8, Article ID e1002234, 2011.
- [16] C. W. Woods, M. T. McClain, M. Chen et al., "A host transcriptional signature for presymptomatic detection of infection in humans exposed to influenza H1N1 or H3N2," *PLoS ONE*, vol. 8, no. 1, Article ID e52198, 2013.
- [17] A. M. Sciuto, C. S. Phillips, L. D. Orzolek, A. I. Hege, T. S. Moran, and J. F. Dillman III, "Genomic analysis of murine pulmonary tissue following carbonyl chloride inhalation," *Chemical Research in Toxicology*, vol. 18, no. 11, pp. 1654–1660, 2005.
- [18] Y. Hart, Y. E. Antebi, A. E. Mayo, N. Friedman, and U. Alona, "Design principles of cell circuits with paradoxical components," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 21, pp. 8346–8351, 2012.
- [19] Y. A. Kuznetsov, *Elements of Applied Bifurcation Theory*, Springer Science & Business Media, 2013.
- [20] Y. Li, M. Yi, and X. Zou, "The linear interplay of intrinsic and extrinsic noises ensures a high accuracy of cell fate selection in budding yeast," *Scientific Reports*, vol. 4, article 5764, 2014.
- [21] W. Zhang, T. Tian, and X. Zou, "Negative feedback contributes to the stochastic expression of the interferon- γ gene in virus-triggered type I interferon signaling pathways," *Mathematical Biosciences*, vol. 265, pp. 12–27, 2015.
- [22] L. Chen, R. Liu, Z.-P. Liu, M. Li, and K. Aihara, "Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers," *Scientific Reports*, vol. 2, article 342, 2012.
- [23] S. Jin, L. Niu, G. Wang, and X. Zou, "Mathematical modeling and nonlinear dynamical analysis of cell growth in response to antibiotics," *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, vol. 25, no. 7, Article ID 1540007, 1540007, 12 pages, 2015.
- [24] M. A. Dahlem, S. Rode, A. May et al., "Towards dynamical network biomarkers in neuromodulation of episodic migraine," *Translational Neuroscience*, vol. 4, no. 3, pp. 282–294, 2013.
- [25] D. Wang, H. Wang, and X. Zou, "Identifying key nodes in multilayer networks based on tensor decomposition," *Chaos*, vol. 27, no. 6, 063108, 12 pages, 2017.
- [26] Y. Liu and X. Zou, "Mathematical modeling of HIV-like particle assembly in vitro," *Mathematical Biosciences*, vol. 288, pp. 46–51, 2017.
- [27] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.
- [28] F. Civril, T. Deimling, C. C. De Oliveira Mann et al., "Structural mechanism of cytosolic DNA sensing by cGAS," *Nature*, vol. 498, no. 7454, pp. 332–337, 2013.
- [29] F. Vafaei, "Using multi-objective optimization to identify dynamical network biomarkers as early-warning signals of complex diseases," *Scientific Reports*, vol. 6, Article ID 22023, 2016.
- [30] A. Ciechanover, "Proteolysis: from the lysosome to ubiquitin and the proteasome," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 1, pp. 79–87, 2005.
- [31] R. Liu, X. Yu, X. Liu, D. Xu, K. Aihara, and L. Chen, "Identifying critical transitions of complex diseases based on a single sample," *Bioinformatics*, vol. 30, no. 11, pp. 1579–1586, 2014.