DReSS: A difference measurement based on reachability between state spaces of Boolean networks

Ziqiao Yin^{1,2,3}, Binghui Guo^{1,2†}*, Shuangge Steven Ma^{3†}, Yifan Sun⁴, Zhilong Mi^{1,2}, Zhiming Zheng^{1,2}

¹ Shenyuan Honors College and School of Mathematics and Systems Science, Beihang University, Beijing, 100191, China

² BDBC, LMIB and Peng Cheng Laboratory, Shenzhen, 518055, Guangdong, China

³ School of Public Health, Yale University, New Haven, 06520, US

⁴ School of Statistics, Renmin University of China, Beijing, 100872, China

* Corresponding author

E-mail: guobinghui@buaa.edu.cn

† These authors contributed equally to this work.

Abstract

Researches on dynamical features of biological systems are mostly based on fixed network structure. However, both biological factors and data factors can cause structural perturbations to biological regulatory networks. There are researches focus on the influence of such structural perturbations to the systems' dynamical features. Reachability is one of the most important dynamical features, which describe whether a state can automatically evolve into another state. However, there is still no method can quantitatively describe the reachability differences of two state spaces caused by structural perturbations. DReSS, Difference based on Reachability between State Spaces, is proposed in this research to solve this problem. First, basic properties of DReSS such as non-negativity, symmetry and subadditivity are proved based on the definition. And two more indexes, diagDReSS and iDReSS are proposed based on the definition of DReSS. Second, typical examples like $DReSS = 0$ or 1 are shown to explain the meaning of DReSS family, and the differences between DReSS and traditional graph distance are shown based on the calculation steps of DReSS. Finally, differences of DReSS distribution between real biological regulatory network and random networks are compared. Multiple interaction positions in real biological regulatory network show significant different DReSS value with those in random networks while none of them show significant different diagDReSS value, which illustrates that the structural perturbations tend to affect reachability inside and between attractor basins rather than to affect attractor set itself.

Author summary

Boolean network is a kind of networks which is widely used to model biological regulatory systems. There are structural perturbations in biological systems based on both biological factors and data-related factors. We propose a measurement called DReSS to describe the difference between state spaces of Boolean networks, which can be used to evaluate the influence of specific structural perturbations of a network to its state space quantitatively. We can use DReSS to detect the sensitive interactions in a regulatory network, where structural perturbations can influence its state space significantly. We proved properties of DReSS including non-negativity, symmetry and subadditivity, and gave examples to explain the meaning of some special DReSS values. Finally, we present an example of using DReSS to detect sensitive vertexes in yeast cell cycle regulatory network. DReSS can provide a new perspective on how different interactions affect the state space of a specific regulatory network differently.

Introduction

Network model is a widely used method to analysis biological systems. There are many kinds of different biological network models are proposed, including protein-protein interaction network, gene co-expression network, gene regulatory network, miRNA network, metabolic network, etc. Researches based on biological network method are mainly focused on two things, structural feature and dynamical feature [31]. The goal of studies which focused on structural features is to explore the key vertexes and society structure of biological systems. This can give us knowledges about the different role of each factor in a system and help us do classification and clustering. The goal of studies which focused on dynamical features is to find the relation between the behavior of biological system and its corresponding outcome. The most representative example for this kind of research is to model the biological regulatory system and predict its outcome. This outcome mostly represents biological phenotype which can be observed.

There is a widely accepted result that cellular regulatory networks are robustly designed [1]. Since then, robustness is one of the most mentioned properties for biological systems [2,3]. Here, the definition of robustness is that the invariance or low variation of a given phenotype when faced with a given incoming variation [4]. Therefore, most calculation methods of robustness are focused on the relation between the perturbation of incoming and the variation of outcoming. Mostly, to biological models based on Boolean networks, the incoming is the initial state. Hence, the perturbation to incoming is the perturbation to the initial state. However, perturbations not only exist in initial state, but in network structure as well. There are both biological and data-related factors can influence the system structure. First, biological factors can affect the biological regulatory interactions. Such as the growth factor can activate the expression of Myc and CyclinD1 [6], and temperature can influence the functional coupling of STIM1 and Orai1 [7] etc. Moreover, there is research pointed out both a passive and active role of DNA methylation to regulatory interactions, where active means changes in methylation can drive changes in regulatory interactions and passive

in contrary [8]. Second, networks constructed based on noisy and incomplete biological data are inexact [9]. That is, there may be missing interactions and incorrect interactions in our object networks. There are already methods to assess exploration stage of gene interaction networks [10], which can be used to estimate the possibility of missing and incorrect interactions in our object networks. These researches just focus on the structure itself, but the influence of structural perturbation on systems' dynamical features are not considered.

Structural perturbations do affect the systems' dynamical features. To show the affection of structural perturbation to biological system's stable state, we take fission yeast cell cycle regulatory network as an example. Because fission yeast cell cycle has been studied for years, and there are many widely accepted models that can model the cell cycle process very well [1,12,18]. Here, we use the model from Davidich et al [12] as an example. Boolean network is proposed to model fission yeast cell cycle, which includes 9 vertexes: Start kinase (SK), Cdc2/Cdc13 complex, Ste9, Rum1, Slp1, Cdc2/Cdc13*, Wee1/Mik1, Cdc25 and phosphatase PP. There are two similar vertexes, Cdc2/Cdc13 and Cdc2/Cdc13*. Cdc2/Cdc13 is responsible for the intermediate activity of the complex when Try-15 is in inactive form, while Cdc2/Cdc13* is an indicator of high activity of Cdc2/Cdc13 when Try-15 is unphosphorylated [12]. Each vertex in this network has a state with binary value, 0 or 1. 0 denotes the vertexes is inactive, while 1 denotes the vertexes is active. The states of vertexes are updated in discrete time step under the following function:

$$S_i(t+1) = \begin{cases} 0, & \text{if } \sum_j a_{ji} S_j(t) < \theta_i \\ 1, & \text{if } \sum_j a_{ji} S_j(t) > \theta_i \\ S_i(t), & \text{if } \sum_j a_{ji} S_j(t) = \theta_i \end{cases}$$

Where a_{ij} is the elements in network's adjacency Matrix A , $a_{ji} = 1$ means there is an activating interaction from vertexes j to vertexes i , $a_{ji} = -1$ means there is an inhibiting interaction from vertexes j to vertexes i , and $a_{ji} = 0$ means there is no interaction from vertexes j to vertexes i . $S_i(t)$ is the state of vertexes i at time step t . θ_i is the threshold of vertexes i , which is set to 0 for all vertexes except Cdc2/Cdc13. Because Cdc2/Cdc13 is positively regulated, so we need to add self-activation to it, which means its θ_i need to be -1. The network is shown in Fig 1.

Fig 1 Yeast Cell cycle network. There are 9 vertices in yeast cell cycle network. Activation interactions are shown in black solid line, and inhibition interactions are shown in red dotted line. Different colors in the state space of yeast cell cycle network represent different connected components, i.e. attractor basins. There are three main attractors in the state space, two fixed points and one limit cycle. Details of these three main attractors are shown in the Fig.

This system has Three main attractors, a limit cycle and two fixed point. Obviously, the limit cycle corresponds to the cell cycle and two fixed point corresponds to G1 state and the quiescent G0 cells. Because there are 9 vertexes with binary value in the network, there are 2^9 states in the state space. Among them, 490 of 512 states are leads to the three main attractors. The state space of this network is show in Fig 1.

To show structural perturbation can affect network's state space, three kinds of perturbations are used: delete an exist interaction, add an activation interaction between two vertexes without interaction, add an inhibition interaction between two vertexes without interaction. Without a quantitate index to describe the affection of structural perturbation to the networks state space, we can see its qualitative affection to system attractors. One of the most significant feature of cell cycle system is that there is a special attractor, a limit cycle. However, there are structural perturbations can affect this limit cycle significantly. The influences of structural perturbations to limit cycle are

shown in Fig 2. The ones which can change the length of the limit cycle from a 3-membered ring into 2,4,6,8-membered ring are shown in orange, while the ones which can make the limit cycle disappear are shown in green. All boxes in table of Fig 2 are in one of three colors, red, black and grey. Red ones denote to there is a inhibition interaction from the vertex on left to the vertex on top, black ones denote to there is a activation interaction from the vertex on left to the vertex on top, grey ones denote to there is not a known interaction between the vertex on left and the vertex on top. All grey boxes are divided into two part, the upper one denotes to add an activation interaction, and the lower one denotes to add an inhibition interaction. There are 137 kind of one step perturbation to this network. There are 46 of 137 position can make the limit cycle disappear, and 9 of 137 position can change the length of limit cycle.

Fig 2 The influence of the structural perturbations of the regulatory network on the limit cycle. There are three colors in the table, red represents there is an inhibition interaction, black represents there is an activation interaction, and grey represents there is no interaction. Perturbations on red and black positions are to remove the original interaction, while perturbations on grey positions are to add a new activation/inhibition interaction. The upper position in a grey box shows the result of adding an activation interaction, while the lower position shows the result of adding an inhibition interaction. The results can be divided into three kinds, limit cycle disappear, change of the length of limit cycle, and no significant change. The results are shown in the box, while the blank positions represent no significant change.

Moreover, these significant changes are not equally distributed to every position. For example, all possible perturbations to interactions from vertex SK cannot make any change to the limit cycle. However, almost every possible perturbation to interaction from vertex Cdc25 can make the limit cycle disappear. That is Cdc25 related interactions are more sensitive to structural perturbation than SK related interactions. Besides, although there are 82 of 137 perturbations that remain the 3-membered ring limit cycle, they still change the state space. Fig 2 gives three examples of each kind of result, disappearance of limit cycle, change of the length of limit cycle, and no significant change. The example perturbations are deleting SK-SK self-inhibition, add a CDC2/CDC13 self-activation and add a CDC25 self-inhibition. Compared with the original state space shown in Fig 1, the original limit cycle disappears in two example state spaces. Moreover,

Although there are no significant change in attractor set when SK-SK self-inhibition is deleted, there are still differences between the new state space and the original one. The purple component become larger, while the green component become smaller. That is, part of the states that lead to the purple attractor have changed their destination because of the structural perturbation.

Part of researches which focus on system dynamical features are based on given fixed network structures and do not consider the structural perturbation [11,12]. However, there are researches focused on the influence of perturbations on network structure to the system attractor. As mentioned above, part of the structural perturbation come from the incomplete and noisy data. Therefore, there are researchers trying to recovery the original systems from these incomplete and noisy data. Lü et al proposed a method to predict the position of missing links in a specific network based on the known part [13,14]. Because of the absence of the kinetic parameters, Barabási et al try to predict perturbation patterns from the topology of biological networks [15]. Chao et al tried to evaluate the influence of deleting, adding and switching a specific interaction in a cellular regulatory network [1]. They focused on the influence of structural perturbations to the size change of the largest attract basin, which is an important part of systems' dynamical feature. Whereas, there is still not a method can evaluate the influence of a specific structure perturbation to the system's state space quantitatively.

Structure perturbation can affect not only the attractor of the system, but the basin of attractor as well. That is, Structure perturbation can change the system's state space from one to another. A system's state space is a network, whose vertexes are possible states of the original system. Two vertexes are linked with a directed edge, if one state can transfer into another. In other word, to evaluate the influence of a specific structure perturbation to the system's state space, is actually to evaluate the difference between two directed networks. There are widely accepted methods to calculate the distance between two networks, such as methods based on maximal common subgraph [16] and based on Laplacian matrix [17]. They consider different feature of networks: similarity, degree distribution, connectivity, etc. However, to state space networks, the most important feature is its reachability, i.e. the ability to get from one state to another within the network. Reachability is one of the most important feature of state space, since it can describe the destination of a specific initial state. Not all stable states are common in biological systems, there are also morbid stable states [32]. Changes of reachability can cause part of the initial states that originally lead to common attractor finally turn into the ones that lead to morbid attractor, which is important in many disease researches. Thus, to evaluate the influence of a specific structure perturbation to the system's state space, the difference between two directed networks based on their reachability need to be evaluated.

Therefore, DReSS, Difference based on Reachability of State Spaces, is proposed. First, it can quantitatively describe the changes of reachability of networks' state spaces. Second, DReSS obey basic properties including non-negativity, symmetry and subadditivity. Third, a family of index can be defined follow the definition of DReSS to help screen the important interactions (edges) and factors (vertexes) in a system and describe the difference of attractor sets before and after being structural perturbed. In details, diagDReSS is proposed to describe the influence of structural perturbation to systems' attractor set and iDReSS is proposed to screen the sensitive vertex to structural perturbation. Finally, DReSS can show researchers where they should pay extra attention to in a system to avoid the significant bias on dynamical analysis causing by the potential structural

perturbations.

Result

In this article, basic related concepts are introduced at first. Then, the definition of DReSS is given. Based on the definition of DReSS, two more related index, diagDReSS and iDReSS, are introduced. With such definition, basic properties of DReSS index family including non-negativity, symmetry and subadditivity, etc. are proved in the following part. Next, examples are shown to explain the meaning of DReSS value and its differences against the original network distances. Finally, DReSS index family is applied to real biological network and random networks. Related conclusions are discussed based on the comparison of DReSS distribution on real networks and random networks.

Definition of DReSS - Difference based on Reachability between State Spaces

Many biological systems are modeled by Boolean networks. A Boolean network $G = (V, E)$ is a network with Boolean function (details see Methods) as its updating function, where V is the set of vertexes and E is the set of edges in network. Therefore, each vertex in a Boolean network have a 0-1 state, where 0 represents inhibited and 1 represents activated. In biological systems, vertexes can be DNAs, RNAs, proteins, genes, SNPs, biological factors, etc. And edges are interaction relations, like activation, inhibition, binding, etc, between these vertexes.

The states of vertexes in G are continuously updated by updating function until they are stable. At each time step, states of all vertexes form a state array. This state array is the state of G at that time step. And all the possible states of G and their transfer relations form the state space S_G of Boolean network G .

State space S_G is also a network, where all possible states of G are its vertexes and the transfer relations are its edges. If *state i* can updated into *state j* in one time step, there is an edge from *state i* to *state j* in S_G . State space can describe the transfer relations between states of the original network G , which is an important feature of a system. With transfer relations, the final stable states and all intermediate states of any specific initial states can be told. Stable states are always related with stable cellular phenotypes [30,31]. Because only stable states can stable exist and be observed, all unstable states will evolve into stable ones at last. But still, there are lots of perturbations can transfer a stable state into an unstable one, like environmental factors, mutations and external interventions, etc. The unstable state transferred from one stable state not always can return the original stable states but may possibly evolve into another stable states. Therefore, to study the transfer relations between states in state space is very important. However, a state space corresponding one original system. The structural perturbation to the original system would also change the state space as well. Thus, these changes may cause initial states possibly evolve into different stable states before and after the structural perturbation. Here, an index is proposed to describe how much of the initial states changed their destination after a specific structural perturbation. That is, to give an index to describe the reachability differences between the original

state space and the new state space.

First of all, A well-defined index to describe the reachability differences between state spaces, especially biological systems' state spaces, should have its explanation meaning. Second, the index should have basic properties such as commutative, non-negative and obey the triangle inequality etc. Therefore, the definition of DReSS (Difference based on Reachability between State Spaces) is proposed as followed.

To define DReSS, the definition of modified reachability matrix has to be introduced first. Reachability matrix $R = (r_{ij})_{N \times N}$ is a matrix to describe the reachability of networks. $r_{ij} = 1$ if there is a directed path from state i to state j in S_G , else $r_{ij} = 0$. Particularly, $r_{ii} = 1$ for all i in S_G . A state space S_G 's Reachability matrix R can be calculated by its adjacency Matrix $A = (a_{ij})_{N \times N}$. $a_{ij} = 1$ if there is a directed edge from state i to state j in S_G , else $a_{ij} = 0$. And $R = I + A + A^2 + \dots + A^N$, where N is the number of all possible states in state space S_G , and operators in the equation are Boolean operators (details see Methods). That is, Reachability matrix R is a Boolean matrix, consist of 0 and 1. However, reachability matrix R cannot describe the attractor set, which is one of the most important feature of state spaces.

Therefore, the definition of modified reachability matrix $\tilde{R}$ is proposed as followed.

Definition 1 – Modified reachability matrix Let $G = (V, E)$ be a Boolean network with n vertexes. S_G is the state space of G , and there are $N = 2^n$ possible states in S_G . $\tilde{R} = (\tilde{r}_{ij})_{N \times N}$ is a $N \times N$ matrix. $\tilde{R}$ is the modified reachability matrix of S_G , if $\tilde{R}$ satisfies

$$\tilde{r}_{ij} = \begin{cases} 0, & \text{if there is not a directed path from state } i \text{ to state } j \text{ in } S_G \\ 1, & \text{if there is a directed path from state } i \text{ to state } j \text{ in } S_G \end{cases}$$

Particularly, $\tilde{r}_{ii} = 1$ if and only if there is a path from state i and can return to state i at last including one-step self-loop, otherwise, $\tilde{r}_{ii} = 0$.

The only difference between modified reachability matrix $\tilde{R}$ and reachability matrix R is the diagonal of the matrix. In the tradition reachability matrix, the diagonal consists of 1 due to $R = I + A + A^2 + \dots + A^N$. But in state spaces, not all states can reach themselves. Only those states which are attractors or part of attractors can have a path eventually return themselves. Therefore, to state space networks, the original reachability matrix R need to be modified into $\tilde{R}$, where $\tilde{R} = A + A^2 + \dots + A^N$. With such a modification, the diagonal of $\tilde{R}$ are not all 1 anymore. Only those states in attractors, no matter fixed points or limit cycles, their corresponding position in the diagonal of $\tilde{R}$ are 1, which can describe the feature of state spaces better than the original reachability matrix R . Besides, all states in S_G are n dimension Boolean vector. These Boolean vectors, states, can be sorted by its corresponding binary number. That is, there is one and only one unique modified reachability $\tilde{R}_{S_G}$ to each state space S_G .

With the definition of modified reachability matrix, the Difference based on Reachability between States Spaces can be defined as followed.

Definition 2 – DReSS Let $G_1 = (V, E_1)$, $G_2 = (V, E_2)$ be two Boolean networks with same vertex set V . S_{G_1} and S_{G_2} are the state spaces of G_1 and G_2 . Let n be the number of vertexes in V , there are $N = 2^n$ possible states in S_{G_1} and S_{G_2} . Then, the difference based on Reachability of state spaces (DReSS) between S_{G_1} and S_{G_2} can be defined as

$$DReSS(S_{G_1}, S_{G_2}) = \frac{Hamming(\tilde{R}_{S_{G_1}}, \tilde{R}_{S_{G_2}})}{N^2}$$

Where $\tilde{R}_{S_{G_1}}$ and $\tilde{R}_{S_{G_2}}$ are the modified reachability matrixes of S_{G_1} and S_{G_2} , and $Hamming(\tilde{R}_{S_{G_1}}, \tilde{R}_{S_{G_2}})$ is the Hamming distance of Boolean matrix $\tilde{R}_{S_{G_1}}$ and $\tilde{R}_{S_{G_2}}$.

As $N = 2^n$, N^2 is always a large number in most situations. Mostly, $N^2 \gg N$. Meanwhile, the number of attractors must less than N , and mostly far less than N . Therefore, the difference between two attractor sets is a very small fraction of the difference between two whole state space. However, the difference between attractor sets is an important part of the difference of two state spaces. Thus, diagDReSS is introduced to describe the difference between two attractor sets.

Definition 3 – diagDReSS Let $G_1 = (V, E_1)$, $G_2 = (V, E_2)$ be two Boolean networks with same vertex set V . S_{G_1} and S_{G_2} are the state spaces of G_1 and G_2 . Let n be the number of vertexes in V , there are $N = 2^n$ possible states in S_{G_1} and S_{G_2} . Then, the difference based on Reachability of state spaces (DReSS) between S_{G_1} and S_{G_2} can be defined as

$$diagDReSS(S_{G_1}, S_{G_2}) = \frac{Hamming(diag(\tilde{R}_{S_{G_1}}), diag(\tilde{R}_{S_{G_2}}))}{N}$$

Where $diag(\tilde{R}_{S_{G_1}})$ and $diag(\tilde{R}_{S_{G_2}})$ are the diagonal of $\tilde{R}_{S_{G_1}}$ and $\tilde{R}_{S_{G_2}}$.

Although as the definition, DReSS is a difference measurement of state spaces, it can describe the state space difference caused by structural perturbation as well. Therefore, it can be used as an edge centrality at mean time. Because of the heterogeneity of different vertexes and edges in networks, DReSS values are not equally distributed at every possible position of edges. That is, edges with high DReSS values are more likely to be concentrated at specific positions, while the most of other edges are with low DReSS value. Therefore, iDReSS, a vertex centrality based on its surrounding edges' DReSS value is proposed as followed to describe where structural sensitive perturbations may be clustered at.

Definition 4 – iDReSS Let $\mathbb{D}$ be the set of DReSS of all possible structural perturbation positions in network $G = (V, E)$. Then, the iDReSS of vertex i in V can be defined as followed,

$$iDReSS(i) = \frac{\sum_{j \in \mathbb{P}_i} (1 - p_j(i)) D_j(i)}{\sum_{j \in \mathbb{P}_i} (1 - p_j(i))}$$

Where $\mathbb{P}$ is the position set of all possible perturbations related with vertexes i , $D_j(i)$ is the value of DReSS of perturbation j , and $p_j(i)$ is the p-value of perturbation j .

Since iDReSS is a weighted average value of DReSS values of all possible perturbations related with vertexes i , iDReSS is also a value between 0 and 1. The average value of $D_j(i)$ can reflect the structural perturbation sensitivity of vertexes i already. But here a weight coefficient $(1 - p_j(i))$ is added to it. Perturbations in random networks can also cause affection to the state space. If the significant different structural perturbation sensitive vertexes with biological meanings need to be screened out, those perturbation with significant different DReSS values than those in random networks should be considered at first. Thus, the weight coefficient $(1 - p_j(i))$ can help us do this.

Due to the definition of modified reachability matrix $\tilde{R}$, $\tilde{r}_{ii} = 1$ if and only if there is a path from state i and can return to state i at last including one-step self-loop, otherwise, $\tilde{r}_{ii} = 0$. That is, to all states inside the attractor set, their corresponding position in the diagonal of $\tilde{R}$ is 1, and all states outside the attractor set, their corresponding position in the diagonal of $\tilde{R}$ is 0. Therefore, when only the difference between diagonal of $\tilde{R}_{S_G}$ and $\tilde{R}_{S_{G_2}}$ is considered, the index can describe the difference between two attractor sets between S_{G_1} and S_{G_2} . Because the length of the diagonal of $\tilde{R}$ is N rather than N^2 , so the N^2 is changed into N from DReSS to diagDReSS. Besides, as a difference measurement, DReSS obey basic properties below.

Property 1 - DReSS conform to distance function properties. To any network G , $\mathcal{S}_G$ is the set of all possible state space of Boolean networks that have the same number of vertexes with G . $DReSS: \mathcal{S}_G \times \mathcal{S}_G \rightarrow \mathbb{R}$, to $\forall S_1, S_2, S_3 \in \mathcal{S}_G$, conform to all properties below when network is with deterministic update function and no limit cycle attractor:

- (a) $DReSS(S_1, S_2) \geq 0$
- (b)* $DReSS(S_1, S_2) = 0 \Leftrightarrow S_1 = S_2$
- (c) $DReSS(S_1, S_2) = D(S_2, S_1)$
- (d) $DReSS(S_1, S_3) \leq D(S_1, S_2) + D(S_2, S_3)$

And when network is with nondeterministic update function or there is at least one limit cycle attractor, $DReSS(S_G, \tilde{S}_G)$ conform to all distance function properties above except (b). Here, deterministic update function means with same input, there will always the same output through the update function, and nondeterministic update function means with same input, there will multiple possible output through the update function.

- (a) As $\tilde{R}_{S_G}$ and $\tilde{R}_{\tilde{S}_G}$ are all N^2 dimension matrix and $Hamming(S_G, \tilde{S}_G)$ is the total number of

difference positions of Boolean matrix $\tilde{R}_{S_G}$ and $\tilde{R}_{\tilde{S}_G}$, $Hamming(S_G, \tilde{S}_G)$ is number between 0 and N^2 . That is, $DReSS(S_G, \tilde{S}_G)$ is a number between 0 and 1. Besides, $DReSS(S_G, \tilde{S}_G)$ is the percentage of vertexes pairs with different reachability in two networks.

- (b) When network is with deterministic update function, there will be only one 1 in the $\{A_{S_1}(i, 1), A_{S_1}(i, 2), \dots, A_{S_1}(i, N)\}$. ($\Leftrightarrow$) if $S_1 = S_2$, $A_{S_1} = A_{S_2}$. According to the definition, $\tilde{R}_{S_1} =$

$$A_{S_1} + A_{S_1}^2 + \dots + A_{S_1}^n = A_{S_2} + A_{S_2}^2 + \dots + A_{S_2}^n = \tilde{R}_{S_2}. \text{ Therefore, } \text{Hamming}(\tilde{R}_{S_1}, \tilde{R}_{S_2}) = 0.$$

That is, $DReSS(S_1, S_2) = 0$. ($\Rightarrow$) Here, contradiction is introduced to proof $DReSS(S_1, S_2) = 0 \Rightarrow S_1 = S_2$. Assume that $DReSS(S_1, S_2) = 0$ and $S_1 \neq S_2$. As $DReSS(S_1, S_2) = 0$, $\text{Hamming}(\tilde{R}_{S_1}, \tilde{R}_{S_2}) = 0$. That is, $\tilde{R}_{S_1} = \tilde{R}_{S_2}$. Since $S_1 \neq S_2$, there will be at least in elements

(i, j) that $A_{S_1}(i, j) \neq A_{S_2}(i, j)$. Let's make $A_{S_1}(i, j) = 1$, then $A_{S_2}(i, j) = 0$. And $\tilde{R}_{S_1}(i, j) = A_{S_1}$

$(i, j) + A_{S_1}(i, j)^2 + \dots + A_{S_1}(i, j)^n = 1$, $\tilde{R}_{S_1} = \tilde{R}_{S_2}$, so $\tilde{R}_{S_2}(i, j) = 1$. And because $A_{S_2}(i, j) = 0$, so there is at least one k , so that $\tilde{R}_{S_2}(i, k) = 1$ and $\tilde{R}_{S_2}(k, j) = 1$. And $\tilde{R}_{S_1} = \tilde{R}_{S_2}$, so $\tilde{R}_{S_1}(i, k) = \tilde{R}_{S_2}(i, k) = 1$ and $\tilde{R}_{S_1}(k, j) = \tilde{R}_{S_2}(k, j) = 1$. That is, there is a pathway between (i, k) and a pathway between (k, j) in S_1 . So, state k is at downstream of state i , and upstream of state j . And because $A_{S_1}(i, j) = 1$, that is state j is the next upstate state of state i . Therefore, i, j, k formed a limit cycle, or there is more than one upstate state, which contradicts the null hypothesis.

(c) and (d) can be easily proved by the definition of Hamming Distance.

Fig 3 Examples for situations with $DReSS(S_1, S_2) = 0$ and $DReSS(S_1, S_2) = 1$.

Each situation is shown with two parts, a. example state spaces for regulatory networks with nondeterministic update function, b. example state spaces for regulatory networks with limit cycle attractor(s).

Fig 3 shows two examples that $DReSS(S_1, S_2) = 0$, but $S_1 \neq S_2$. Although they are different state space with same $\tilde{R}$, these two state spaces are actually very similar. According to the definition of $\tilde{R}$, if two state space have same $\tilde{R}$, they must have same attractors. If not so, there will be at least one different element on the diagonal of $\tilde{R}$. Besides, their reachability must be the same to all intermediate state. These two properties make sure that, from any specific initial state, although the state space may have differences, which makes differences of intermediate states, they always lead to the same attractor (stable state).

With such properties, DReSS strengthens its meaning. With $DReSS(S_1, S_2) = DReSS(S_2, S_1)$, DReSS can ensure there is only one difference value between two specific state spaces, no matter which is the original state space and which is the one after perturbation. With triangle inequality,

$DReSS(S_1, S_3) \leq DReSS(S_1, S_2) + DReSS(S_2, S_3)$, any two state spaces perturbed from the original state space, their differences is smaller than the sum of their differences to the original one.

Although the property of $DReSS(S_1, S_2) = 0 \Leftrightarrow S_1 = S_2$ is valid with condition, when it is without the condition, it can still tell a lot of information. First, $S_1 = S_2 \Rightarrow DReSS(S_1, S_2) = 0$ is always true without any condition. That is, if perturbation cause no change to the state space, the DReSS of the new state space and the original one would be 0. Second, with $DReSS(S_1, S_2) = 0$, S_1, S_2 have same attractor and reachability between states without any condition. Because $DReSS(S_1, S_2) = 0 \Leftrightarrow \tilde{R}_{S_1} = \tilde{R}_{S_2}$, and the diagonal of $\tilde{R}$ can tell the attractor and the reachability of its corresponding state space. That is, with same $\tilde{R}$, S_1 and S_2 must have same attractor and reachability. Therefore, according to the definition of attractor and reachability, if $DReSS(S_1, S_2) = 0$, from same specific initial state in both state spaces, they can always reach the same attractors.

Examples when DReSS are exactly 0 and 1

$DReSS(S_1, S_2)$ can be exactly 0 and 1, and examples are shown in Fig 3. Obviously, when there is no change in state space, $S_1 = S_2$, the difference $DReSS(S_1, S_2)$ will be exactly 0 according to the definition. However, because the elements in $\tilde{R}_{S_1}$ are not independent, that is, $\tilde{R}_{S_1}(i, j) = 1$ and $\tilde{R}_{S_1}(j, k) = 1 \rightarrow \tilde{R}_{S_1}(i, k) = 1$, the example for $DReSS(S_1, S_2) = 1$ cannot be constructed without limit. Fig 3 shows an example of $DReSS(S_1, S_2) = 1$. To make the state space ‘totally different’, the path cannot be simply reversed. Modified reachability matrix $\tilde{R}_{S_1}$ includes two kinds of information, what is the reachability between states and which states are attractors. Therefore, to make $DReSS(S_1, S_2) = 1$, both these two things need to be reversed from S_1 to S_2 .

To reverse the reachability between states, there are multiple situations. First, if state a is at downstream of state b in S_1 , state a must be at upstream of state b in S_2 , vice versa. This is the most basic situation. Second, if state a cannot reach state b and state b cannot reach state a in S_1 , state a and state b must be in a ring in S_2 , vice versa.

To reverse the attractors, the network must with nondeterministic update function when there is more than one attractor in S_1 . To make $DReSS(S_1, S_2) = 1$, all states which are not attractors in S_1 must become attractor in S_2 , vice versa. When there is more than one attractor in S_1 and the network is with deterministic update function, there is no path between any of two different attractor basins. To reverse this, as mentioned above, any states from one attractor in S_1 must be in a ring with all states in another in S_2 . If S_2 is also with deterministic update function, all the states are in a large ring, which against the rule of states in one attractor basin in S_1 cannot reach each other in S_2 . And when there is only one attractor, the nondeterministic update function is not a necessary condition. Fig 3 gives an example of $DReSS(S_1, S_2) = 1$ and S_1, S_2 are both with deterministic update function when there are just one attractor in S_1 and S_2 .

Differences between DReSS and traditional network difference measurements

Having proved such properties of DReSS above, the difference between DReSS and traditional network difference measurements should be discussed next. Mostly, graph difference measurements

are based on an assumption that all the vertexes in the graph are equal, which makes their order is commutative. However, to biological networks, vertexes are mostly unequal. Different DNAs, RNAs, proteins and other biological complexes, always have their unique function and cannot be replaced by others. And each vertex in state space represents a unique expression pattern of the given biological system. Therefore, two states cannot be considered as the same or similar just because they have same or similar statistical features in state space.

Therefore, different from the former graph difference measurements [16,17], vertexes are considered with order in DReSS. Fig 4 gives an example to show the difference between the former graph difference measurements and DReSS. There are two similar regulatory networks G_1 and G_2 . The only difference between them is the position of self-activation interaction. The differences in the original network make their state space different. S_1 and S_2 have same attractors (0,1) and (1,0). In S_1 the rest states are attracted to (1,0), while in S_2 they are attracted to (0,1).

Fig 4 Example for DReSS applied to evaluate the influence of structural perturbation to network's state space. From the top to bottom, all steps are the original regulatory network, the update function, the state space the modified reachability matrix and the DReSS & diagDReSS value.

There are examples for this kind of mutual inhibition regulatory in biological systems, such as GCR and c-Jun [19] or PTEN and PIP3 [20]. In network science, G_1 and G_2 are exactly the same graph. They have totally the same structure. But vertexes are non-commutative, the conclusion could be totally different. As an example, PTEN and PIP3 have mutual inhibition regulatory relations. PIP3 functions to activate the downstream signaling component including protein kinase AKT, whose isoforms are overexpressed in a variety of human tumors. PTEN, as a phosphatase to dephosphorylate PIP3, always acts the role of tumor suppressor gene. If the mutual inhibition regulatory relations between them is maintained, but give one of them a self-activation, the result would be totally different. Assuming that gene 1 in Fig 4 is PTEN and gene 2 is PIP3, most states in S_1 are attracted to states (1,0), which is the normal states. But in S_2 , most states are attracted to states (0,1), which is the cancerous states. Although S_1 and S_2 have the same structural in network science meaning, they lead to totally different results in biological meaning.

However, under the former graph difference measurements, as S_1 and S_2 have totally the same structure, their difference is 0. Although these two state spaces have same attractors and similar state transfer relations, there are still differences between their reachability, that is, $\tilde{R}_{S_1}$ and $\tilde{R}_{S_2}$ matrix. Thus, their DReSS $DReSS(S_1, S_2) = 1/8$ rather than 0. Because half of the states in state space changed their destination attractor, these two state spaces are not totally the same. This makes the systems more likely to be stable at (0,1) rather than (1,0). Therefore, these two state spaces are same, which shows that DReSS can describe this kind of changes in state space better.

Data analysis of DReSS in real regulatory networks and random networks

To compare DReSS distribution in the real biological regulatory network and random networks, yeast cell cycle regulatory network is chosen as an example. The structure of real regulatory network is referred from two former related researches [1,12]. The DReSS distribution of all possible structural perturbation to the fission yeast cell cycle regulatory network is calculated. There are three kinds of structural perturbation: delete an existing edge, add an activation edge and add an inhibition edge. To compare the result, the distributions of DReSS under these three kinds of perturbation on three kind of random networks are also calculate: ER random network, BA scale-free network and WS small world network. In random networks, most value of DReSS are clustered around 0, which means structural perturbation has less affection to state space.

However, in yeast cell cycle network, the peak position of DReSS distribution moves toward 1 obviously, and the distribution is more dispersed. There are multiple samples with significant larger DReSS than most samples in random networks. Specifically, there are 2 samples in real regulatory network that with extremely large DReSS value, which are even larger than the maximum value of all 43098 DReSS value among 300 random networks. These two samples are both perturbations toward vertex SK. This result shows that those corresponding structural

perturbations make huge difference to the original state space.

On the other hand, there are no sample in real regulatory network with significant smaller DReSS value than the random results except 8 samples with DReSS of exactly 0. All this 8 samples are also structural perturbation towards vertex SK. SK is a special vertex in the original network. It has no input from other vertex, and it only has one self-inhibition as its input. That is, it must be inhibited in any stable state. Thus, it is very sensitive to any input perturbation. SK, Start Kinase, is a simplified vertex as a beginning signal to the system. Thus, it is reasonable that there is rare similar situation in random networks. Meanwhile, there is rare similar situation in real regulatory networks as well. However, this result still reminds us that structural perturbations toward vertex with much more output than input can influence the state space significantly.

Concluded from the results above, DReSS of real biological networks may be more dispersed than the random ones. That is, there are significant sensitive positions and also more stable ones. This result shows that biological network has a clearer role assignment than random network when against structural perturbation. When facing structural perturbations, part of positions in the regulatory network are more sensitive to response, while the rest positions are more stable to keep the system robust, which also shows the importance of DReSS as an index to describe this kind of difference quantitatively.

Fig 5 Results analysis on yeast cell cycle regulatory network. a. Distributions of DReSS and diagDReSS in yeast cell cycle network and random networks. The red line is the 5% significant level threshold, and yellow line is the significant level threshold with Bonferroni

correction. b. Distribution of DReSS value of all possible structural perturbation around 9 different vertexes in yeast cell cycle network. c. The structural perturbations with significant difference compared to those in random networks are highlighted in the Fig. Solid lines are activation interactions, while dotted lines are inhibition interactions. The ones with solid color are the perturbations of delete an existed edge, and the ones with translucent are the perturbations of add a new edge.

When it comes to diagDReSS, there is no any position with significant different in real biological network compared to random ones. Fig 5.a shows the distribution of diagDReSS of real biological network and random networks. Because there is always only a small part of states can become attractor among of states in state spaces, the diagDReSS distribution is stepped compared to the DReSS distribution. Furthermore, there is no position with significant large diagDReSS value in real biological network compared to random networks, which shows that in both real and random networks, structural perturbation can hardly affect networks' attractors significantly.

T-test with Bonferroni correction method (details see Methods) has been introduced to test whether there are samples with significant difference in the yeast cycle network to the ones in random networks. The result shows that there are indeed 22 of 137 possible structural perturbations with significant difference to the ones in random networks, which means perturbation on these positions are more structural sensitive than nearly all the possible ones in random networks. Besides, none of these 137 possible structural perturbations have significant different value compared to the random networks. The positions in real regulatory network with significant different DReSS value from random networks has been light up in Fig 5.c. The structural perturbation positions with significant different DReSS are mostly related with specific vertexes.

Fig 5.b shows DReSS distribution of surrounding possible perturbations of all 9 vertexes in yeast cell cycle regulatory network. The most different vertex is SK whose DReSS clustered around both maximum value and minimum value of the interval. This has already been discussed above. For vertexes like STE9, RUM1 and CDC2/CDC13*, their DReSS is mainly clustered around the lower side of the interval and have no DReSS at top ranks. WEE1/MIK1 and CDC25 are two vertexes whose DReSS are equally distributed in whole interval. That is, when the perturbation is between them and another sensitive vertex, DReSS would be high, and when perturbation is between them and another stable vertex, DReSS would be low. Finally, for CDC2/CDC13, SLP1 and PP, most of their DReSS is clustered at medium and high part of the whole interval, which means perturbations around them may be sensitive. Therefore, iDReSS is calculated for these vertexes in yeast cell cycle regulatory network.

Table 1 iDReSS value of all 9 vertexes in yeast cell cycle regulatory network.

Vertex Symbol	SK	CDC2/ CDC13	STE9	RUM1	SLP1	CDC2/ CDC13*	WEE1/ MIK1	CDC25	PP
iDReSS × 10	4.16	4.85	3.12	3.12	4.76	3.65	4.01	4.48	4.85

Table 2 shows the iDReSS value for all vertexes in fission yeast cell cycle regulatory network.

Three kinds of vertexes which has been discussed above can be separated under iDReSS. Although SK is the vertex with most high rank DReSS, but it is still not the vertex with largest iDReSS value. This result shows that iDReSS focus on those vertexes whose surrounding positions are mostly sensitive rather than those which have extremely abnormal ones.

All three selected vertexes by iDReSS have their proved biological influence on cell cycle process. CDC genes, Cell Division Cycle genes, are known as key regulatory genes in cell cycle process for years [24]. CDK1 is the protein product of CDC2, whose phosphorylation can lead to the begin of cell cycle. CDC13 can regulates telomere replication, which affects the proliferative capacity of the cell. SLP1 is an effector of the Mad2-dependent spindle checkpoint, which can form a complex with Mad2 essential for the onset of anaphase [25]. PP, phosphatase, can drive protein phosphorylation, which can achieve the widespread reorganization of cellular architecture in mitosis [21].

Moreover, the interactions between these three top vertexes are important as well. CDK1 is the protein product of CDC2. When form a complex with its partner, CDK1 phosphorylation leads to cell cycle progression, which is regulated by PP. Meanwhile, Grallert et al described a PP1–PP2A phosphatase relay which controls the mitotic progression [21]. However, the pathway is blocked in early mitosis due to the inhibition of CDC2 towards PP, and declining CDC2 levels later in mitosis permit PP1 to self-activation again [26,27]. SLP1 is an activator of APC/C (anaphase promoting complex/ cyclosome), which can destroy cyclin B and therefore inhibit Cdk1 (CDC2). These biological results show the importance of interactions between these selected vertexes by iDReSS.

Discussion

DReSS brings a new quantitatively view to difference between state spaces, especially based on structural perturbation. Reachability is one of the most important features for state space, especially for systems with deterministic update function. To systems with deterministic update function, there no path between two attractor basins. That is, the differences of reachability between two state spaces may cause a same initial state to evolve into different stable states in two state spaces. DReSS can show how much of the vertex pairs have changed their reachability between two state spaces. Moreover, besides reachability, attractor set is also an important feature of a state space. diagDReSS can describe the differences of attractor sets between two state spaces. Due to the definition of attractor set of Boolean networks, a state is an attractor of a part of an attractor if and only if it has a path from itself towards itself. Therefore, the differences between the diagonals of two modified reachability matrix $\tilde{R}$ can be used to describe the differences between two attractor sets of two state spaces. Besides, DReSS provides a new method to evaluate the centrality of edges in a graph. There are plenty of researches focus of finding key vertexes of a network. There are many different centrality indexes for vertex, such as degree, betweenness, closeness, eigenvector, k-core, Katz, PageRank, etc. However, there are no so much centrality index for edges compared to for vertexes. Moreover, the edge centralities are mostly based the network structure itself, such as paths, edge betweenness, random walk, information, etc. However, there are rare centrality connecting structural features with dynamical features. DReSS is a measurement that can connect structural

features with dynamical features. In this research, only structural perturbation of one edge has been considered. However, DReSS can not only describe the difference between two state spaces caused by only one structural perturbation, it can describe the difference between two state spaces caused by the combination effect of multiple structural perturbation. In biological regulatory networks, changes of environmental factors may cause multiple change to our regulatory system at one time. Meanwhile, the missing and the noisy interactions may not be only one in each system, that is, there may be multiple missing and noisy interactions in a biological system. It is one of DReSS's advantage that it is with same computational complexity to calculate one or multiple perturbation at one time. Different from the existing edge centralities, DReSS can not only evaluate the centrality of existing edges of a network but can evaluate the potential ones as well. DReSS is actually an index to evaluate the difference between two state spaces. Meanwhile, any structural perturbation can correspond to changes of state space differences. Therefore, DReSS can be used to evaluate the influence of a specific position in a network, no matter there is a known edge at that position or not.

However, the calculation may become a main problem when calculating DReSS of a large system. The computational complexity of $\tilde{R}$ is $O(N^2)$ when the system is with deterministic update function, and $O(N^3)$ when it is with nondeterministic update function. Meanwhile, here N is 2^n where n is the number of vertexes in the original system. Thus, large systems' $\tilde{R}$ cannot be calculated without simplification such as network module partitions. The average time of calculating a single structural perturbation of yeast cell cycle regulatory network is about 5.82 second and the time of calculating all possible perturbation of yeast cell cycle regulatory network is 797.03 second. When system become larger, the state space will be larger at same time. However, the edges in state spaces are mostly sparse. Furthermore, when the system is with deterministic update function, there will be exactly N edges in state space of size N^2 because there is one and only one out-edge for each vertex. That is, there are always much more 0 than 1 in large systems' $\tilde{R}$. Therefore, these 0 positions may more likely be same in two state spaces, which makes the DReSS of large systems would be much smaller than those of small systems. This can be also explained as when system become larger, the influence of a single structural perturbation to the state space become smaller. Therefore, mostly, the calculation of DReSS value of a single structural perturbation in a large system is not necessary. When the computational complexity exceeds acceptable range, alternative calculating methods such as using random seeds sampling the reachability of state space to get the approximation of DReSS value.

The results above show that there are indeed differences between positions in a network on their sensitivity to structural perturbations. Moreover, these differences are even more significant in real biological networks than those in random networks. This may because the roles of different vertexes in biological systems are clearer and more different than roles of different vertexes in a random network. Based on the results above, part of the interactions are designed to response the changes to the system, while the rest of the interactions are designed to keep the system robust. However, there is not any structural perturbation with significant diagDReSS value in real biological network compared to random networks, which shows the structural perturbations tend to affect reachability inside and between attractor basins rather than to affect attractor set itself. Finally, from the analysis of iDReSS value of real biological network, vertexes can be divided into three groups: the ones aim to keep the robustness of system, the ones aim to make sensitive response to

perturbations, and the ones to link the former two group together.

The future research can focus on the theoretical distribution of the DReSS in different situations. With theoretical distribution, the position with significant difference DReSS value can be selected easily. Moreover, the theoretical distribution can help us make a better understanding of the functions of interactions in regulatory systems. Also, the knowledge about the different roles of interactions in a system may help us to find the optimal control method of the system. That is, when DReSS is combined with optimization theory, it can provide new views to find optimal control methods of a system from any initial state evolve into the any stable state with minimum cost. The code can be downloaded at <https://github.com/yinziqiao/DReSS-package>.

Methods

In this part, details of basic concepts mentioned in the article are shown to help make a better understanding of the research. The contents are listed in the order in which it appears in the article.

Boolean function and Boolean network

Boolean function is a kind of function whose arguments assume values from a binary set (usually $\{0,1\}$). A Boolean function is a mapping from $\mathbb{B}^k$ to $\mathbb{B}$, where $\mathbb{B} = \{0,1\}$. Operation ' + ' mentioned above represents logical operation OR, and ' $\times$ ' represents logical operation AND. And the power operations for matrix used above are still based matrix multiplication, whose inside operations are replaced by logical operations. Boolean networks are networks with Boolean function as their update functions. Mostly, the Boolean networks which are used to model biological regulatory networks are directed networks. That is, each edge in the network is with a direction.

State space

A state space is a set of all possible states of a system. Specifically, to a Boolean network with k vertexes, the state space would be $\mathbb{B}^k$, where $\mathbb{B} = \{0,1\}$. That is, to a Boolean network with k vertexes, there are 2^k states in its corresponding state space. The states can be ordered by their corresponding binary value, i.e. $\{(0,0,\dots,0,0),(0,0,\dots,0,1),\dots,(1,1,\dots,1,1)\}$. Besides, as a discrete system, a Boolean network's state space can be described by a directed graph. States are vertexes in the graph, and there is a directed edge from s_1 to s_2 , if and only if $f(s_1) = s_2$ where f is the update function of the system. When the system is with nondeterministic update function, the condition of $f(s_1) = s_2$ turn into $T(s_1, s_2) \neq 0$, where T is the transition matrix of the system.

Attractor and attractor basin

Attractor is a set of numerical values that a system tends to evolve towards. Specifically, to a Boolean network, attractor is a state or a group states, when the system update into such state or such group of states, the system would keep in this state or this group of states unless there are perturbations from outside of the system. The attractor which is a single state called fixed point. And the attractor which is formed with a group of states called limit cycle. The set of all the attractors is the attractor set. Attractors set is the minimal set of states which can only update into states inside

the set.

An attractor's basin of attraction, or attractor basin in short, is set of all the states which will eventually evolve into that attractor. When the system is with deterministic update function, there is no common subset between any two different attractor basins.

Hamming distance

Hamming distance is a concept from information theory to describe the difference between two string with same length. Its value is the number of positions at which the corresponding symbols are different in two strings. Here, the concept of Hamming distance is borrowed to describe the difference between two matrixes with same size. Its value is the number of positions at which the corresponding elements are different in two matrixes. Thus, its value is between 0 to the size of matrix.

Modified reachability matrix $\tilde{R}$

$\tilde{R}_{S_G}$ can be simplified into quasi-diagonal matrix $s\tilde{R}_{S_G}$ by matrix elementary transformation when network G is with deterministic update function, the matrix can be described as:

$$s\tilde{R}_{S_G} = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_m \end{pmatrix}$$

Where m is the number of attractors in S_G . And when attractor i is a fixed point,

$$R_i = \begin{pmatrix} 1 & 0 & 0 & & 0 & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & & 0 & & 0 & 0 & 0 \\ & \vdots & & \ddots & & \vdots & \\ 1 & & & & 0 & 0 & 0 \\ 1 & & & \dots & 0 & 0 & \\ 1 & & & & & 0 & 0 \\ & & & & & & 0 \end{pmatrix}$$

When attractor i is a limit cycle,

$$R_i = \begin{pmatrix} 1 & 1 & 1 & & 0 & 0 & 0 \\ 1 & \ddots & 1 & \dots & 0 & 0 & 0 \\ 1 & 1 & 1 & & 0 & 0 & 0 \\ & \vdots & & \ddots & & \vdots & \\ 1 & 1 & 1 & & 0 & 0 & 0 \\ 1 & 1 & 1 & \dots & 0 & 0 & \\ 1 & 1 & 1 & & & 0 & 0 \\ & & & & & & 0 \end{pmatrix}$$

With deterministic update function, one state can have one and only one direct downstream state. Let the basin size of attractor i is L , then from any state in basin i , it takes less than L step to reach the attractor i . What is more, in a attractor's basin, there is no ring except the limit cycle. All states in attractor i 's basin can reach attractor i , and all the states except attractor i formed a tree. Therefore, the R_i can be simplified into the form above. And according to the definition, there is no path between attractors. Thus, $\tilde{R}_{S_G}$ can be simplified into quasi-diagonal matrix $s\tilde{R}_{S_G}$ by matrix elementary transformation when network G is with deterministic update function. But what need to be noted is that this simplified quasi-diagonal matrix $s\tilde{R}_{S_G}$ cannot be used for the calculation of DReSS, since the vertexes are ordered in the calculation of DReSS.

Random networks

In this research, 3 kinds of random networks are used as comparison: ER random network, BA scale-free network and WS small world network. The random networks have to be with same size, same average degree and same activation-inhibition rate with fission yeast cell cycle network. Therefore, random networks are generated of same size and same average degree with yeast cell cycle network by Python package networkx. Then, a probability p whose value is the percentage of inhibition interactions in real cell cycle network is set. Finally, edges are turned into inhibition interaction by probability p , and keep the rest edges as activation interactions.

Bonferroni correction

Bonferroni correction is a method to solve the multiple comparison analysis problems. Let $H_1, H_2, \dots, H_m$ be a family of hypotheses, and $p_1, p_2, \dots, p_m$ are their corresponding p-value. Bonferroni correction rejects all the null hypotheses if their corresponding $p \leq \alpha/m$ to control $FWER \leq \alpha$. Specifically, in this research, m is the number of all possible position for perturbation in yeast cell cycle networks. The null hypotheses is the DReSS value of perturbation in yeast cell cycle networks obey the same distribution of DReSS value of perturbation in random networks.

References

- [1] Li F, Long T, Lu Y, et al. The yeast cell-cycle network is robustly designed[J]. Proceedings of the National Academy of Sciences, 2004, 101(14): 4781-4786.
- [2] Wollman R. Robustness, accuracy, and cell state heterogeneity in biological systems[J]. Current opinion in systems biology, 2018, 8: 46-50.
- [3] Bruner A, Sharan R. A Robustness Analysis of Dynamic Boolean Models of Cellular Circuits[J]. Journal of Computational Biology, 2019.
- [4] Félix M A, Barkoulas M. Pervasive robustness in biological systems[J]. Nature Reviews Genetics, 2015, 16(8): 483-496.
- [5] Willadsen K, Triesch J, Wiles J. Understanding robustness in Random Boolean Networks[C]//ALIFE. 2008: 694-701.
- [6] Deritei D, Aird W C, Ercsey-Ravasz M, et al. Principles of dynamical modularity in biological regulatory networks[J]. Scientific reports, 2016, 6: 21957.
- [7] Xiao B, Coste B, Mathur J, et al. Temperature-dependent STIM1 activation induces Ca^{2+} influx and modulates gene expression[J]. Nature chemical biology, 2011, 7(6): 351.
- [8] Gutierrez-Arcelus M, Lappalainen T, Montgomery S B, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation[J]. elife, 2013, 2: e00523.
- [9] Barabási A L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease[J]. Nature reviews genetics, 2011, 12(1): 56-68.
- [10] Yin Z, Guo B, Mi Z, et al. Gene saturation: an approach to assess exploration stage of gene interaction networks[J]. Scientific reports, 2019, 9(1): 1-17.
- [11] Deritei D, Aird W C, Ercsey-Ravasz M, et al. Principles of dynamical modularity in biological regulatory networks[J]. Scientific reports, 2016, 6: 21957.

- [12] Davidich M I, Bornholdt S. Boolean network model predicts cell cycle sequence of fission yeast[J]. PloS one, 2008, 3(2).
- [13] Lü L, Pan L, Zhou T, et al. Toward link predictability of complex networks[J]. Proceedings of the National Academy of Sciences, 2015, 112(8): 2325-2330.
- [14] Pan L, Zhou T, Lü L, et al. Predicting missing links and identifying spurious links via likelihood analysis[J]. Scientific reports, 2016, 6(1): 1-10.
- [15] Santolini M, Barabási A L. Predicting perturbation patterns from the topology of biological networks[J]. Proceedings of the National Academy of Sciences, 2018, 115(27): E6375-E6383.
- [16] Bunke H, Shearer K. A graph distance metric based on the maximal common subgraph[J]. Pattern recognition letters, 1998, 19(3-4): 255-259.
- [17] Shimada Y, Hirata Y, Ikeguchi T, et al. Graph distance for complex networks[J]. Scientific reports, 2016, 6: 34944.
- [18] Novak B, Pataki Z, Ciliberto A, et al. Mathematical model of the cell division cycle of fission yeast[J]. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2001, 11(1): 277-286.
- [19] Yang-Yen H F, Chambard J C, Sun Y L, et al. Transcriptional interference between c-Jun and the glucocorticoid receptor: mutual inhibition of DNA binding due to direct protein-protein interaction[J]. Cell, 1990, 62(6): 1205-1215.
- [20] Matsuoka S, Ueda M. Mutual inhibition between PTEN and PIP3 generates bistability for polarity in motile cells[J]. Nature communications, 2018, 9(1): 1-15.
- [21] Grallert A, Boke E, Hagting A, et al. A PP1-PP2A phosphatase relay controls mitotic progression[J]. Nature, 2015, 517(7532): 94-98.
- [22] Izawa D, Goto M, Yamashita A, et al. Fission yeast Mes1p ensures the onset of meiosis II by blocking degradation of cyclin Cdc13p[J]. Nature, 2005, 434(7032): 529-533.
- [23] Chikashige Y, Yamane M, Okamasa K, et al. Fission yeast APC/C activators Slp1 and Fzr1 sequentially trigger two consecutive nuclear divisions during meiosis[J]. FEBS letters, 2017, 591(7): 1029-1040.
- [24] Morgan D O. The cell cycle: principles of control[M]. New Science Press, 2007.
- [25] Kim S H, Lin D P, Matsumoto S, et al. Fission yeast Slp1: an effector of the Mad2-dependent spindle checkpoint[J]. Science, 1998, 279(5353): 1045-1047.
- [26] Wu J Q, Guo J Y, Tang W, et al. PP1-mediated dephosphorylation of phosphoproteins at mitotic exit is controlled by inhibitor-1 and PP1 phosphorylation[J]. Nature cell biology, 2009, 11(5): 644-651.
- [27] Kwon Y G, Lee S Y, Choi Y, et al. Cell cycle-dependent phosphorylation of mammalian protein phosphatase 1 by cdc2 kinase[J]. Proceedings of the National Academy of Sciences, 1997, 94(6): 2168-2173.
- [28] Chandra A, Hughes T R, Nugent C I, et al. Cdc13 both positively and negatively regulates telomere replication[J]. Genes & Development, 2001, 15(4): 404-414.
- [29] Meadows J C, Millar J B A. Sharpening the anaphase switch[J]. 2015.
- [30] Choi M, Shi J, Jung S H, et al. Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage[J]. Sci. Signal., 2012, 5(251): ra83-ra83.
- [31] Carter H, Hofree M, Ideker T. Genotype to phenotype via network analysis[J]. Current opinion in genetics & development, 2013, 23(6): 611-621.
- [32] Choi M, Shi J, Zhu Y, et al. Network dynamics-based cancer panel stratification for systemic prediction of anticancer drug response[J]. Nature communications, 2017, 8(1): 1-12.