A novel structure-based control method for analyzing nonlinear dynamics in biological networks

Wei-Feng Guo[1], Shao-Wu Zhang[1]*, Tao Zeng[2], Yan Li[1], Jianxi Gao[4]*, Luonan Chen[2,3]*

[1] Key Laboratory of Information Fusion Technology of Ministry of Education, School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China

[2] Key Laboratory of Systems Biology, Shanghai Institutes for Biological Science, Chinese Academy Science, Shanghai 200031, China

[3] Collaborative Research Center for Innovative Mathematical Modeling, Institute of Industrial Science, University of Tokyo, Tokyo 153-8505, Japan.

[4] Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA.

* Corresponding author

Email: zhangsw@nwpu.edu.cn, gaoj8@rpi.edu, lnchen@sibs.ac.cn

Abstract

Exploring complex biological systems requires adequate knowledge of the system's underlying wiring diagram but not its specific functional forms. Thus, exploration actually requires the concepts and approaches delivered by structure-based network control, which investigates the controllability of complex networks through a minimum set of input nodes. Traditional structure-based control methods focus on the structure of complex systems with linear dynamics and may not match the meaning of control well in some biological systems. Here we took into consideration the nonlinear dynamics of some biological networks and formalized the nonlinear control problem of undirected dynamical networks (NCU). Then, we designed and implemented a novel and general graphic-theoretic algorithm (NCUA) from the perspective of the feedback vertex set to discover the possible minimum sets of the input nodes in controlling the network state. We applied our NCUA to both synthetic networks and real-world networks to investigate how the network parameters, such as the scaling exponent and the degree heterogeneity, affect the control characteristics of networks with nonlinear dynamics. The NCUA was applied to analyze the patient-specific molecular networks corresponding to patients across multiple datasets from The Cancer Genome Atlas (TCGA), which demonstrates the advantages of the

nonlinear control method to characterize and quantify the patient-state change over the other state-of-the-art linear control methods. Thus, our model opens a new way to control the undesired transition of cancer states and provides a powerful tool for theoretical research on network control, especially in biological fields.

Keywords: Controllability; Nonlinear dynamics; Biological networks; Graphic-theoretic algorithm; Single patient system

Author summary

Complex biological systems usually have nonlinear dynamics, such as the biological gene (protein) interaction network and gene co-expression networks. However, most of the structure-based network control methods focus on the structure of complex systems with linear dynamics. Thus, the ultimate purpose to control biological networks is still too complicated to be directly solved by such network control methods. We currently lack a framework to control the biological networks with nonlinear and undirected dynamics theoretically and computationally. Here, we discuss the concept of the nonlinear control problem of undirected dynamical networks (NCU) and present the novel graphic-theoretic algorithm from the perspective of a feedback vertex set for identifying the possible sets with minimum input nodes in controlling the networks. The NCUA searches the minimum set of input nodes to drive the network from the undesired attractor to the desired attractor, which is different from conventional linear network control, such as that found in the Maximum Matching Sets (MMS) and Minimum Dominating Sets (MDS) algorithms. In this work, we evaluated the NCUA on multiple synthetic scale-free networks and real complex networks with nonlinear dynamics and found the novel control characteristics of the undirected scale-free networks. We used the NCUA to thoroughly investigate the sample-specific networks and their nonlinear controllability corresponding to cancer samples from TCGA which are enriched with known driver genes and known drug target as controls of pathologic phenotype transitions. We found that our NCUA control method has a better predicted performance for

indicating and quantifying the patient biological system changes than that of the state-of-the-art linear control methods. Our approach provides a powerful tool for theoretical research on network control, especially in a range of biological fields.

Introduction

Numerous biological systems can be represented as networks, and several approaches have been developed to construct reliable biological networks[1,2]. Since the control process is dominated by the intrinsic structure and dynamic propagation within the system, the concepts and approaches of structure-based network control are emergently required to investigate the controllability of complex networks through a minimum set of input nodes[3-13]. The analysis of biological systems from the structure-based control viewpoint provides a deeper understanding of the dynamics of complex large-scale biological systems[14-16]. So far, the studies exploiting the structure-based control of complex networks can be mainly divided into two categories according to the styles of the networks, that is, the approaches focusing on directed networks[3-6,10-13,17] and the methods focusing on undirected networks[7-9] For directed networks, many researchers have developed linear structural control tools to identify the minimum number of input nodes that need to be controlled by external signals for the system to achieve the desired control objectives[5,6,13]. Although those linear control tools have many applications to biomolecular systems, such as in the detection of driver metabolites in the human liver metabolic network [14] and driver gene discovery in pan-cancer datasets [15], those tools may only give an incomplete view of the network control properties of a system with nonlinear dynamics[17]. Recently, an analytical tool called a feedback vertex set control (FC) has been shown to study the control of large directed networks in a reliable and nonlinear manner, where the network structure is prior-known and the functional form of the governing equations is not specified, but must satisfy some properties [12,18]. This formalism identifies the set of feedback vertex nodes (FVS) in networks, uniquely determining the long-term dynamics of the entire network. With such a scheme, the source nodes can converge to a unique state (or trajectory) without

independent control [12,18]. Zañudo et al. showed that both the state of the source nodes and FVS can change the dynamic attractors available to the network; they identified the source nodes and FVS as the input nodes to control the direct networks with nonlinear dynamics [17]. The above approaches only focus on the linear or nonlinear dynamics of directed networks. There are few approaches to investigate the linear dynamics on undirected networks. For example, an exact controllability framework [7], an analytical framework, offers a tool to treat the structural controllability of undirected networks; the Minimum Dominating Sets (MDS) [8] is an alternative way to investigate the controllability of undirected linear networks, since it works with a strong assumption that the controllers can control its outgoing links independently. Therefore, there is still a need for efficient tools to analyze the structural controllability of the undirected networks with nonlinear dynamics.

In this paper, we first formalize the nonlinear control problem of undirected networks (NCU), that is, how to choose the proper input nodes to drive the network from one attractor to a desired attractor in the networks with nonlinear and undirected dynamics. We developed a novel graphic-theoretic algorithm (NCUA) to measure the controllability of undirected networks based on the feedback vertex sets. Specifically, (i) we assume that each edge in a network is bidirectional; (ii) we construct a bipartite graph from the original undirected network, in which the nodes of the top side are the nodes of the original graph and the nodes of the bottom side are the edges of the original graph (Figure 1 (b)); (iii) we adopt an equivalent optimization procedure for determining the MDS of the top side nodes to cover the bottom side nodes in the bipartite graph that can control the whole network using mathematical terms; and (iv) we apply random Markov chain sampling to obtain the distribution of the input nodes set and uncover the possible sets of the input nodes to control the undirected network.

Since most real world networks have a statistically significant power-law distribution, we generally have defined the control characteristics as the fraction of identified minimum input nodes and applied NCUA for multiple synthetic scale-free (SF) networks and real-world networks, and obtained several counterintuitive findings: i) the fraction of input nodes in the network increases when the degree exponent value

increases for fixed average degree, indicating that control characteristics is affected by degree heterogeneity; ii) new degree heterogeneity is defined and the fraction of input nodes decreases monotonically when degree heterogeneity becomes larger for fixed average degree. Furthermore, the degree heterogeneity and the average degree determine the minimum number of control input nodes; iii) the set of input nodes tends to be highly target-connected nodes, whereas the previous linear control study suggested that driver nodes tend to avoid high-degree nodes [9-12].

We also investigated the network transition between the disease state and normal state identifiable with the stable network states (dynamical attractors) in personalized patient networks. For each sample of each cancer patient from 10 kinds of cancer sites in TCGA, we constructed a personalized differential network between the normal state and disease state, and applied the NCUA for finding their key control genes on pathologic phenotype transitions. We found that (i) although most of the cancer samples have a similar nonlinear controllability, the determining control genes still differ for different cancer samples; (ii) we identified the controllability of the reconstructed individual networks for single samples across 10 cancer datasets, and we found the high confidence cancer-specific key genes have significant enrichments in the cancer genes census (CGC) set and the FDA-approved drug target genes (DTG) set. Compared with the traditional control model of linear networks (Exact control and Liu's linear control) [7,8], our results imply that a single-patient system in cancer may be more controllable than predicted on linear dynamical networks due to the ubiquity of the nonlinear features in biological networks. In contrast to another model on the network control of undirected networks called MDS[8], our NCUA also showed a higher performance in identifying the key genes in the CGC and DTG, which were underestimated by the MDS. In conclusion, our model provides a new powerful tool for theoretical and empirical study of network controllability, especially in biological and biomedical fields.

Methods

Formulation of the NCU

Network dynamics are commonly nonlinear, especially at the level of nodes or small groups of nodes in the network [19]. In past decades, the focus of network control research has shifted from linear dynamics to nonlinear dynamics [12,20-24]. One of these methods, namely the feedback vertex set control (FC) [12,18], can be reliably applied to large complex networks in which the structure is well known and the functional form of the governing equations is not specified but must satisfy some properties. Although Zañudo et al. applied the FC to study dynamic models of direct networks to predict nodes for the control of various technical, social, and biological networks [17], we still lack a framework to solve the nonlinear control problem on undirected networks [8]. Here, we focus on the nonlinear control problem of undirected networks. Given an undirected network G (V, E), we generally consider the following broader class of the model [23] to be the following:

$$\begin{aligned} dx_i / dt &= f_i(x_i) - \sum_{j=1}^{N} l_{ij} \Gamma x_j + B_i u \\ &\equiv F_i(x_i, x_{I_i}) + B_i u, i = 1, \ldots, N \end{aligned} \qquad (1)$$

where $x_i \in R^d$ denotes the d-dimensional state variable of the i-th node at time t. The set I_i is a set of neighborhood nodes of node i; $L=(l_{ij})$ is the Laplacian matrix determined by the network topological structure, and Γ describes the internal coupling configuration at each node. $\mathbf{B}_i \in R^{N*N_C}$ characterizes the driving by the N_C controllers with the network. $F_i(x_i, x_{I_i})$ shows the enhancement of the activity of node i, satisfying that (i) is the continuous differentiability of $F_i(x_i, x_{I_i})$, that is, $F_i(x_i, x_{I_i}) \in C^1$, and (ii) dissipativity, that is, for any initial condition $x(0)$ and for a finite time $t \geq 0$, the dynamical state $x(t)$ is bounded by a positive constant C:

$\|x_n(t)\| \le C$; and iii) the decay condition is $\partial_1 F_i(x_i, x_{i_j}) < 0$ (more details of the formulation are shown in Supplementary Note 1). Then, we formalize the concept of the nonlinear control of the undirected networks, which is how we chose the set of input nodes that are injected by input signal u with the minimum cost to control the above equation (1) from an initial attractor to a desired attractor. In Figure 1, we give a diagram illustration of our NCU with a simple example.

Algorithm for the Nonlinear Control of an Undirected Network (NCUA)

In many complex biological systems, there is adequate knowledge of the underlying wiring diagram, but not of the specific functional forms [17]. Analyzing such complicated systems requires concepts and approaches of structure-based control, which investigates the controllability of complex networks through a minimum set of input nodes. The traditional structure-based control methods, such as the Structural Controllability-based Control (SC) and MDS-based Control and Feedback Control (FC) [12,18] focus on the structure of complex systems with linear dynamics, and may not always match the meaning of control in real systems, where control tends to involve only naturally-occurring system states [25]. To drive the state of a network to any one of its naturally occurring end states (i.e., dynamical attractors), FC needs to manipulate a set of nodes (i.e., the FVS) that intersects every feedback loop in the network. Mochizuki et al. have mathematically proved that, for a direct network with the nonlinear dynamics of Equation (1), the control action of forcing (overriding) the state variables of the FVS ensures that the network will asymptotically approach the desired dynamical attractor, regardless of the specific form of the functions. For this FC, it requires only a few conditions (e.g., continuous, dissipative, and decaying) on the nonlinear functions that are very mild and satisfied by a wide range of nonlinear dynamical systems [12,18]. Recently, Zañudo et al. have applied the FVS to study dynamic models of direct networks, illustrating that the controllability is determined by the cycle structure and the source nodes of the network [17]. However, they focused on the structure control of direct networks with nonlinear dynamics. We still lack an analytical framework for the feedback control of undirected networks.

Therefore, to solve the above proposed NCU, we developed a novel algorithm, the NCUA, which is based on the assumption that the edges of the undirected networks are modeled as the bi-directed edges.

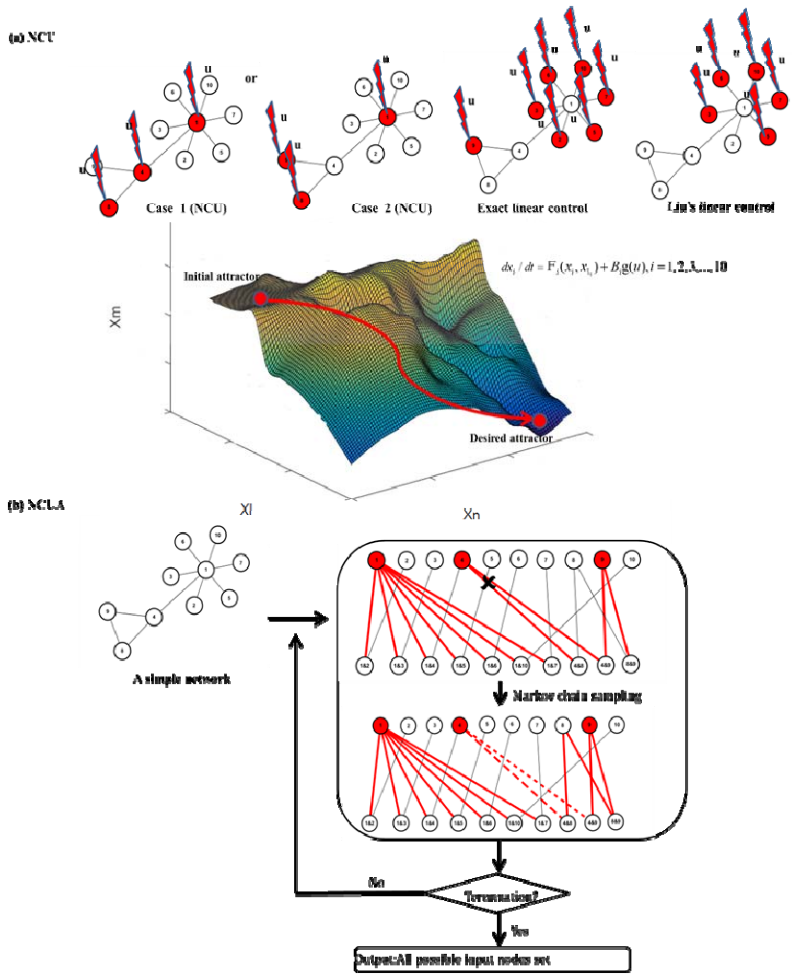

Figure 1. A schematic diagram illustrating nonlinear control problem of undirected network (NCU) and our NCU algorithm (NCUA). (a) NCU: dynamical attractors have been found to be identifiable with stable network states in networks with nonlinear dynamic, we chose the set of input nodes with the minimum cost to control the system which is represented by an undirected network from the initial attractor to the desired attractor. By controlling the three minimum feedback vertex nodes $\{v_1, v_4, v_9\}$ and ensuring that the removal of the three nodes leaves the graph without cycles, the system is guaranteed to be controllable from initial attractor to desired attractor. By using Liu's linear control and exact control method, we identified $\{v_1, v_3, v_5, v_6, v_7, v_{10}\}$ and $\{v_1, v_2, v_3, v_5, v_6, v_7, v_9, v_{10}\}$ respectively to control the network from initial state to desired state. (b) NCUA: We firstly assume that each edge in a network is bi-directional and construct a bipartite graph from the original undirected network, in which the nodes of top side are the nodes of original graph and the nodes of the bottom side are the edges of the original graph. Then, we adapt an equivalent optimization procedure for obtaining the initial input nodes $M_1 = \{v_1, v_4, v_9\}$ as a initial Markov chain within the top side nodes to cover the bottom side nodes in the bipartite graph that are sufficient to control the whole network with nonlinear dynamics in mathematical term. Finally, we generate a new Markov chain $M_2 = \{v_1, v_8, v_9\}$ by replacing node $v4$ with node $v8$ in the Markov chain M_1 which can also cover edge and generated

the new Markov chain M_2 and repeat this process until the absolute percentage error $MAPE = \sum_i |\phi_t(i) - \phi(i)| / n$ is less than the tolerance error $\tau = 0.01$ where $\phi_t(i)$ is the frequency of node i at time t and $\phi(i)$ is the average frequency of node i in the process of sampling and n is number of nodes with $0 < \phi < 1$. The diagram of EC control scheme and Liu's control scheme and MDS control scheme and our NCUA are shown in Figure S2.

The definition of FVS is a subset of nodes in the graph, such that the removal of the set leaves the graph without cycles. Therefore our NCUA searches the minimum set whose removal leaves the graph without feedback loops. This algorithm consists of three main steps: (i) constructing a bipartite graph from the original undirected network, in which the nodes of the top side are the nodes of the original graph and the nodes of the bottom side are the edges of the original graph, shown in Figure 1 (b), (ii) determining the MDS of the top side nodes to cover the bottom side nodes in the bipartite graph by using Integer Linear Programming (ILP), and (iii) designing random Markov chain sampling to obtain different input node sets. In Figure 1 (b), we give a diagram to illustrate the process of our NCUA for discovering the possible input nodes. The details of the NCUA are introduced below.

I. Constructing a bipartite graph from the original undirected network

For a given undirected network G (V, E), we assume that each edge is bi-direct and convert G (V, E) into a bipartite graph $G(V_{\mathrm{T}}, V_{\perp}, E_1)$, where $V_{\mathrm{T}} \equiv V$ and $V_{\perp} \equiv E$. If $v_i \in V_{\mathrm{T}}$ is 1 of the nodes for $v_j \in V_{\perp}$, we add 1 edge e_1 connecting $v_i \in V_{\mathrm{T}}$ and $v_j \in V_{\perp}$ into the set E_1.

II. Obtaining the cover set with minimum cost by using Integer Linear Programming (ILP)

After we obtain the bipartite graph $G(V_{\mathrm{T}}, V_{\perp}, E_1)$, we adopt a modified version of the dominating set, in which the dominating set S must be selected from V_{T} and is also sufficient to dominate all of the nodes in $V_{\perp}$. We use a minimum dominating set cover

problem for determining the nodes to control the whole network, that is, how to select a proper node set *S*, in which for each node $v_j \in V_\perp$, there exists a node $s_i \in S \subseteq V_T$ such that $(v_j, s_i) \in E_1$. This problem can be solved by solving the following ILP model,

$$\min \sum_{v \in V_\perp} x_v \qquad (2)$$
$$s.t. \sum_{\{v,u\} \in E} x_v \geq 1 (\text{every } u \in V_T), x_v \in \{0,1\}$$

where it will take the value x_i=1 when node *i* belongs to the cover set; the object is to obtain the minimum number of nodes to cover set $V_\perp$. Although it is an NP-hard problem[8], the optional solution is obtained efficiently for moderate sizes of graphs with up to a few tens of thousands of nodes by utilizing an algorithm that uses the LP-based classic branch and bound method [26,27] to determine the optimal solution.

III. Obtaining different input nodes by using random Markov chain sampling.

Here, we define the minimum dominating nodes in the "bipartite graph"as a Markov chain. The state space ***G*** of the Markov Chain (MC) is the set of all the possible minimum dominating nodes of the "bipartite graph". The different MCs need to be sampled from the state space ***G*** so that a random MC method is used. The MC approach samples different sets of minimum dominating nodes randomly. Thus, the MC method can give different sets of input nodes to control the undirected network with nonlinear dynamics. The basic idea of the MC method is to build a Markov Chain whose states are collections of the minimum dominating nodes in the top nodes covering the bottom nodes in the bipartite graph $G(V_T, V_\perp, E_1)$, as shown in Figure 1b. The MC method is described as follows:

Initialization: By using ILP, obtain the initial Markov Chain M_0.

Iteration: For *t*=1, 2,…, obtain M_{t+1} from M_t as follows:

- Choose a node w uniformly at random in M_t. Then, delete node w and add a new node $v \in V_T - M_t$ which can cover the edges connected by node w in the bipartite graph $G(V_T, V_\perp, E_1)$. A new Markov Chain $M_{t+1} = M_t - \{w\} + \{v\}$ has been obtained.
- Accept the new Markov Chain M_{t+1} randomly.

We terminate the procedure of the MC sampling when the absolute percentage error, $MAPE = \sum_i |\phi_t(i) - \phi(i)| / n$, is less than the tolerance error $\tau = 0.01$, where $\phi_t(i)$ is the frequency of node i at time t, $\phi(i)$ is the average frequency of node i in the process of sampling, and n is number of nodes, satisfying $0 < \phi_i < 1$. Otherwise, the search process is terminated if the iteration time exceeds the fixed default value $N_{\max}$ = 10,000.

Complexity analysis of the NCUA

The NCUA consist of three steps: the first step is to convert the undirected network G (V, E) into a bipartite graph $G(V_T, V_\perp, E_1)$, where the top nodes V_T and the bottom nodes $V_\perp$ are the nodes and the edges (feedback loop) in the undirected network, respectively; the second step is to obtain the minimum set among the top nodes to cover the bottom nodes in the bipartite graph $G(V_T, V_\perp, E_1)$, where $V_T \equiv V$ and $V_\perp \equiv E$; the third step is to obtain different input nodes by using random Markov Chain sampling. The computational complexity of our NCUA method stems from the second part and the third part. In the phase of the minimum set cover problem, we utilize a branch-and-bound algorithm on the bipartite graph, an automatic method with the computational complexity $O(\log\|V+E\|)$ [28]; in the phase of the random Markov Chain sampling with N_m times, we employ a greedy algorithm on the bipartite graph with the computational complexity, where $N_T \equiv \|V\|$ is the network size, and N_D is the

number of the input nodes. In fact, the computational complexity $O\left(N_m*(N-N_D)\right)$ can be approximately considered as $O\left(N_m*N\right)$. Therefore, the overall complexity of our NCUA approach is $O\left(\log\|V+E\|\right)+O\left(N_m*N\right)$.

Results

Controllability of the SF network revealed by the NCU in synthetic networks

In order to evaluate the control characteristics of the NCU, we applied our NCUA to the synthetic SF networks generated by the static model [13,29] (more details are listed in Supplementary Note 1 of Additional File 1). We assumed the degree distribution of the undirected network *G* (*V*, *E*) follows $P(k)\propto k^{-\gamma}$. We first defined the fraction of the input nodes $n_d=\frac{\|S\|}{\|V\|}$, where $\|S\|$ denotes the set of input nodes to control the whole network and $\|V\|$ denotes the number of connected nodes in the network. Then, we applied our NCUA to estimate the minimum number of input nodes to control the networks with nonlinear dynamics. For a given γ and average degree <k>, 100 networks of 10,000 nodes were constructed. The results of the NCUA were averaged over all realizations. We list the numerical results of our NCUA for the synthetic networks in Figure 2.

In fact, we plotted the NCU size as a function of the degree exponents and the average degrees and list the results in Figure 2 (a-c). In Figure 2(a), we show that for γ < 2, the number of input nodes increases as γ increases, while the number of input nodes does not depend on the average degree <k>. However, if the value of γ is above 2, the number of input nodes is governed by both γ and <*k*>. Furthermore, SF networks with a large value of γ above 2 or large value of <*k*> are hard to control, as shown in Figure 2(a, c). These results are complemented by Figure 2 (b-c), where it

shows that, compared with the Erdös-Rényi random networks (ER), only a few nodes are needed to control the entire network if the power law degree exponent γ is smaller than or around 2, whereas it is more difficult to controlled with a value of γ above 2. This result gives insight into which SF network will be easier to control with the minimum number of input nodes. To more clearly visualize the impact of the network structure on the number of input nodes, we plot the NCU size as a function of the network degree heterogeneity for fixed <*k*> in Figure 2(d). Here, we define the degree heterogeneity [30], $H = \langle k^2 \rangle - \langle k \rangle = \sum (k_i)^2 / N - \sum k_i / N$, where k_i is the degree of node v_i in the undirected network and N is the number of nodes in the network. We observe that for the fixed average degree <k>, the NCU size decreases as the degree heterogeneity increases. These results illustrate that heterogeneous networks are not difficult to control, which is opposite to the conclusions of the EC linear control scheme [7] and Liu's linear control scheme[13]; however, these results are in agreement with the results of the MDS control scheme [8]. Note that the diagram of EC control, Liu's linear control, MDS control, and our NCUA are shown in Figure S2.

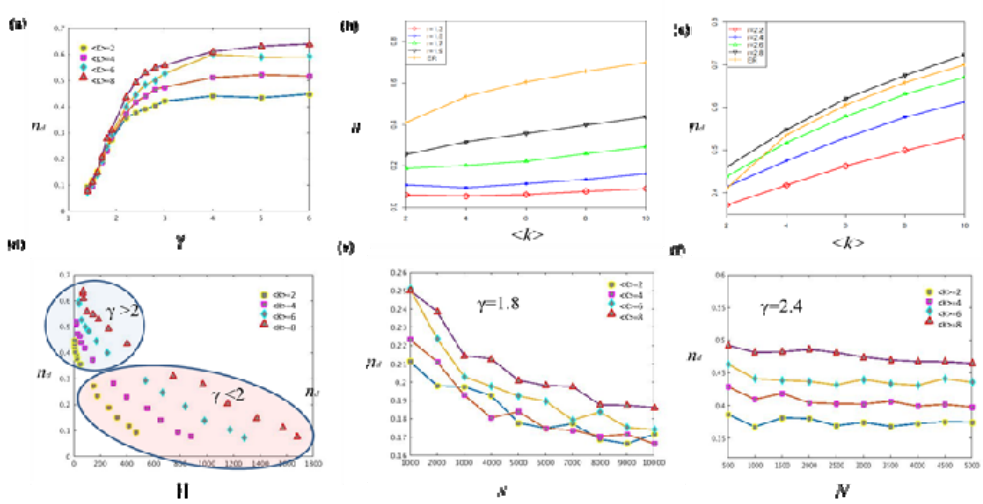

Figure 2. Nonlinear Controllability of synthetic scale free undirected networks with 10,000 nodes. (a) The number of input nodes for NCU control cost n_d in function of the average degree $\langle k \rangle$ and the degree exponent γ for SF networks. All results are averaged over 10 independent realizations of the networks with 1000 nodes. **(b-c)** The number of input nodes for NCU

control cost n_d in function of the average degree $<k>$ compared with the ER networks, for the degree exponent γ larger than 2 and less than 2 respectively. It shows that compared with the ER networks, few nodes are needed to control the entire network if the power law degree exponent γ is smaller than around 2, whereas more difficultly it is to be controlled, with larger value of γ above 2. **(d)** The NCU size as a function of the network degree heterogeneity at fixed <k>.(e-f) The number of input nodes in function of network size for different average degree exponent with γ=1.8 and γ=2.4 respectively

In Figures 2(e) and (f), S3, and S4, we list the number of input nodes in the function of the network size for a fixed degree exponent with γ = 1.4, 1.6, 1.8, and γ = 2.4, 2.6, 3.4, and 3.6. We find that the number of input nodes decreases with the increasing network size for $\gamma < 2$, while for $\gamma > 2$, the number of input nodes are not affected by the network size.

Counterintuitive findings of the controllability from the NCU on real-world networks

We collected 17 networks with 11 categories, which were chosen for their diversity in applications and scopes (Additional File 2). By calculating the P-value of the Kolmogorov-Smirnov goodness-of-fit statistic [31], whose results are listed in Table S3, we found that the above networks are significantly subject to the power-law distribution; the detailed results are shown in Supplementary Note 5 of Additional File 1. First, we focused on how the number of input nodes correlates with the topological features, such as the scaling exponent in the power-law degree distributions and their average degree. The results of the application of the NCU on real-world networks are shown in Figure 3. In Figure 3(a), we show the scatter plot of the number of input nodes as a function of the degree exponents of SF networks and their average degrees. As shown in Figure 3 (a), we show that the number of input nodes has a tendency to increase as the exponent and the average degree increase. Furthermore, in Figure 3(a), we can evaluate the value of scaling exponent approximately by fitting its control characteristic to that on the synthetic networks.

We observed that the degree heterogeneity becomes larger as the number of

network nodes increases (Figure S3 and Figure S4). To clearly visualize the impact of the degree heterogeneity and the average degree on the number of input nodes, we first eliminated the effect of the network number on the degree heterogeneity by using the following formula,

$$H_N = H_0 - H_{n_0} + H'_N$$

where H_0 is value of the degree heterogeneity in the given network and H_{n_0} is the mean value of the degree heterogeneity in the synthetic network with same value of the scaling exponent whose values are evaluated by matching the controllability on the synthetic scale free networks and average degree and number of network nodes compared with the given network. H'_N is the mean value of the degree heterogeneity in the synthetic network with the number of network nodes N=10,000, but with the same values of the scaling exponent and average degree compared with the given network.

We list the results of the number of input nodes in the function of the average degree and the new converted degree heterogeneity measure in Figure 3(b). As shown in Figure 3 (b), we find that networks with a lower average degree and higher degree heterogeneity are easier to control than those with a large average degree. The control characteristics of networks can be fully discriminated by the new converted degree heterogeneity and the average degree. We also find that the set of input nodes tends to highly target connected nodes, whereas the previous linear control study suggested that driver nodes tend to avoid high degree nodes, as shown in Figure 3(c)[7].

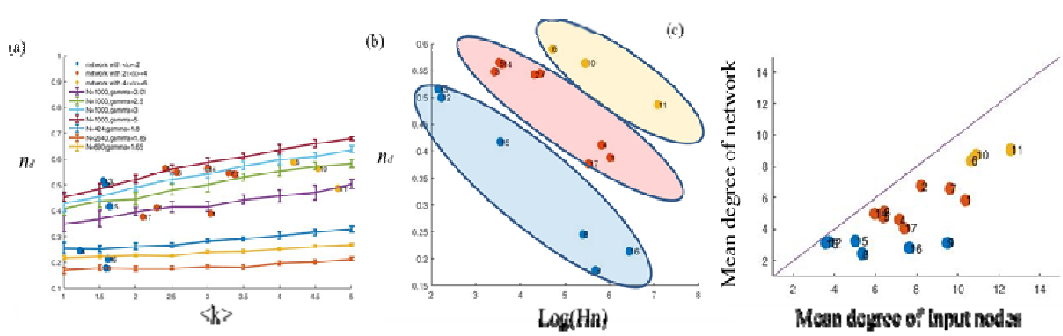

Figure 3. Results of NCU on real-world networks. (a) For each real network in Additional file 2, the scatter plot of the number of input nodes for NCU control cost n_d in function of the average degree $<k>$. The error bar plot denotes the results of synthetic networks whose control characteristics match that of real networks. (b) Scatter plot for each real network with the fraction of input nodes in NCU vs the new degree heterogeneity with eliminating the effect of network number (n=10,000). Note that for the three networks with $\gamma<2$, we estimate the value of the degree exponent by mapping the controllability on the synthetic networks with the same number of given real-world network, while for other networks with $\gamma>2$, we estimate the value of the degree exponent by mapping the controllability on the synthetic networks with N=1000. (c) Scatter plot with the mean degree of the whole networks vs mean degree of the input nodes for each real network by using our NCUA. The bold line denotes the positions in the plot where the mean degree of the whole networks is equal to that of the input nodes;

We observe that most types of biological networks (e.g., gene regulatory, PPI, and genetic networks) require the control of a smaller fraction of nodes than social networks (trust and social communication networks); the fraction of input nodes is between 10% and 30% in biological networks vs. more than 40% in social networks. These predictions match well with recent experimental results in cellular reprogramming and large scale social network experiments[32,33]. Note that this prediction stands in contrast with those of linear control[7] on the same type of networks, and to some extent, can address the initial arguments on network controllability[34,35].

To ensure that our NCU is physically significant, we then focused on the required control energy and the control time to achieve control for networks with nonlinear dynamics. We applied a 3-dimensional stable nonlinear Lorenz oscillator system[36,37] on the real-world network to control the networked system to the desired attractor $\{8.484, 8.484, 27\}$. The time cost of the network is defined as $T=\max\{T_i\}, \|T_i - T_{i-1}\| \le \varepsilon$, where ε is the tolerance error. The energy cost of the network is defined as $E=\int_0^{T_i}\sum_{i=1}^{N}\|u_i(t)\|dt$, where T_i is the convergence time for the node i.

The feedback controllers for node i are defined as $u_i=-k(x_i-\bar{x})$, where k is the control strength and $\bar{x}$ is the desired state. The closed loop controllers for node i are defined as

$$u_i = \begin{cases} -k(x_i - \bar{x}), \|x_i - \bar{x}\| > 1 \\ -ksig(x_i - \bar{x})^{\alpha}, \|x_i - \bar{x}\| < 1 \end{cases}$$

where $sig(x_i - \bar{x})^{\alpha} = [\text{sign}(x_{i1} - \bar{x})\|x_{i1} - \bar{x}\|^{\alpha}, ..., \text{sign}(x_{id} - \bar{x})\|x_{id} - \bar{x}\|^{\alpha}]$, sign(.) is a sign function, and $\alpha \in (0,1)$. Figure 4 demonstrates that the networks with a smaller number of input nodes would demand a greater energy cost for the pinning control to the real-world networks with the feedback controllers and 1 type of closed loop controllers ($\alpha = 0.5$). Note that in Figure 4, the energy cost and time cost of a given network are the average energy and time cost of different input nodes, respectively.

Finally, we evaluated the differences between closed-loop controller and linear feedback controller on the nonlinear network control. Here, we adopt the local feedback controllers[36,37] and closed-loop controllers [23] on the real-world networks. Figure 4 shows that closed-loop controllers demand a greater number of determining nodes, but require less control time and control energy than the traditional linear feedback controllers.

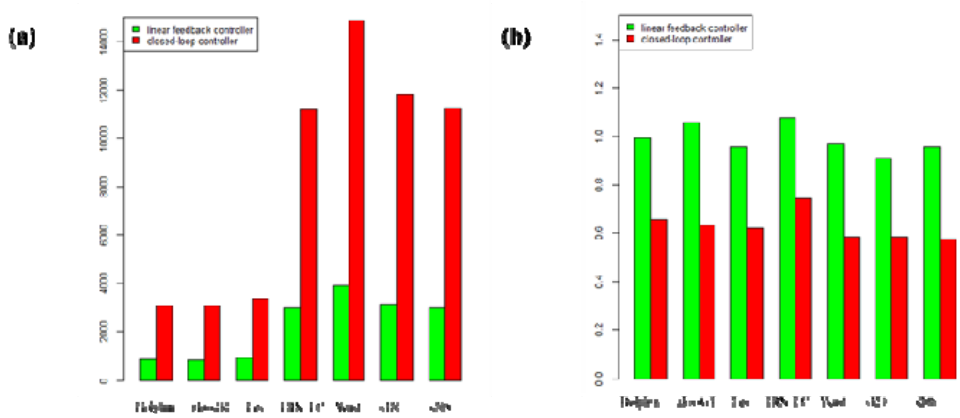

Figure 4. The compared results in terms of control energy and control time for linear feedback controller and the closed-loop controller. All the energy cost and the time cost is averaged in the process of sampling different input nodes to control the networks. **(a)** Result with the average energy cost of different input nodes set by using NCUA in **linear feedback controller** and **the closed-loop controller** respectively in eleven real-world networks **(b)** Result with the average time cost of different input nodes set by using NCUA in **linear feedback controller** and **the closed-loop controller** respectively in eleven real-world networks.

Advanced discovery of individual phenotype-transition genes in cancer samples using the NCUA

Cancer is a complex disease that generally results from a dysfunction of the relevant system or network with nonlinear dynamics, which dynamically changes with time and conditions [38-40]. To further evaluate the merits of our NCU, we applied the NCUA to estimate the nonlinear controllability of a network corresponding to a single patient with cancer and to find the key control genes related with the phenotype transition from the normal state to the tumor state. Ten cancer datasets were analyzed, including the datasets for breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), kidney chromophobe (KICH) and kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), stomach adenocarcinoma (STAD), and uterine corpus endometrial carcinoma (UCEC). According to the requirements of the NCUA, only normal-disease paired samples were obtained from TCGA data portal[39,40] (112 paired samples for BRCA, 50 paired samples for COAD, 23 paired samples for KICH, 72 paired samples for KIRC, 31 paired samples for KIRP, 50 paired samples for LIHC, 57 paired samples for LUAD, 49 paired samples for LUSC, 32 paired samples for STAD, and 23 paired samples for UCEC cancer). Each individual had paired samples (a control sample and a tumor sample). More detailed information is given in Supplementary Note 2 in Additional File 1.

We constructed the patient-manner network by integrating the tumor and normal expression data for each patient and gene interaction network data[41]. First, we chose all the normal data as the reference data and constructed the tumor network and normal network, respectively, based on the reference data with SSN method[38]. Next, we constructed the differential expression network where the edge between gene i and gene j will exist if the p-value of the edge is less than (greater than) 0.05 in the tumor network, but greater than (less than) 0.05 in the normal network. We calculated the differential expression genes with the +/-1 log2 fold change and defined the patient-manner network as the sub-network, which consists of the differential

expressed genes and where the edge exists in the both gene-gene interaction network and the differential expression network for each patient (Figure S1 in Additional File 1).

We first used KS statistics to find that these networks have scale free properties. We then computed the linear regression coefficient between the frequency of the network degree $d(\log10(f(d)))$ and the log 10 transformed degree d ($\log10(d)$). We found that the scale free exponents of the single sample networks in different types of cancer are less than 2 (Additional File 3). We also found that the NCU controllability for the single sample network in different types of cancer, that is the ability to alter the normal state and tumor state, will be much easier to control than the controllability of the network linear dynamics, including the EC control scheme and Liu's control scheme (Figures 5 (a) and S5 of Additional File 1). This result reveals that for the cancer patient, we only need a small fraction of genes to change the network state between the stable states, which is not applicable for controlling the biological network from initial states to any states in linear dynamics. This observation is in agreement with previous biological conclusions[34,35].

The key control genes in the patient-manner network were further investigated using the NCUA method. In fact, the NCUA method provides a ranking of the nodes as the input control nodes according to the value of the frequencies of the nodes, in which the input control nodes are ordered by decreasing the sampling frequency in the random Markov chain sampling. We first defined the personalized key control genes as the genes that appear as the key control nodes with a high frequency (f>0.6) in the patient-manner network. Then, we calculated the frequency of the personalized key control genes for different cancer datasets. We defined the high confidence cancer-specific determining genes (f>0.6), middle confidence cancer-specific key control genes (0.3<f<=0.6), and low confidence cancer-specific key control genes (f<=0.3). The computational results of 10 cancer datasets are listed in Figure 5 (b1). Figure 5 (b1) shows that the low cancer-specific key control genes account for the majority, demonstrating that the determining control genes vary for most of the cancer

patients. The high confidence cancer-specific key control genes in each cancer dataset are in listed in Additional File 4.

Finally, we computed the p-value of the high-confidence key control genes enriching the cancer genes census set [42] or FDA-approved drug targeted genes set [43] by using the hyper-geometric test [44]. If the calculated p-value was less than 0.05, then we regarded that this cancer gene set is significantly enriched in the Cancer Genes Census set and FDA-approved DTG set. Figure 5 (b2) shows that the high-confidence determining control genes for different cancer datasets have a good enrichment in the cancer genes census set and the FDA-approved DTG set. Furthermore, we find that the set of input nodes tends to target highly connected nodes, as shown in Figure S6 of Additional File 1. These results are in agreement with previous biological observations[45,46].

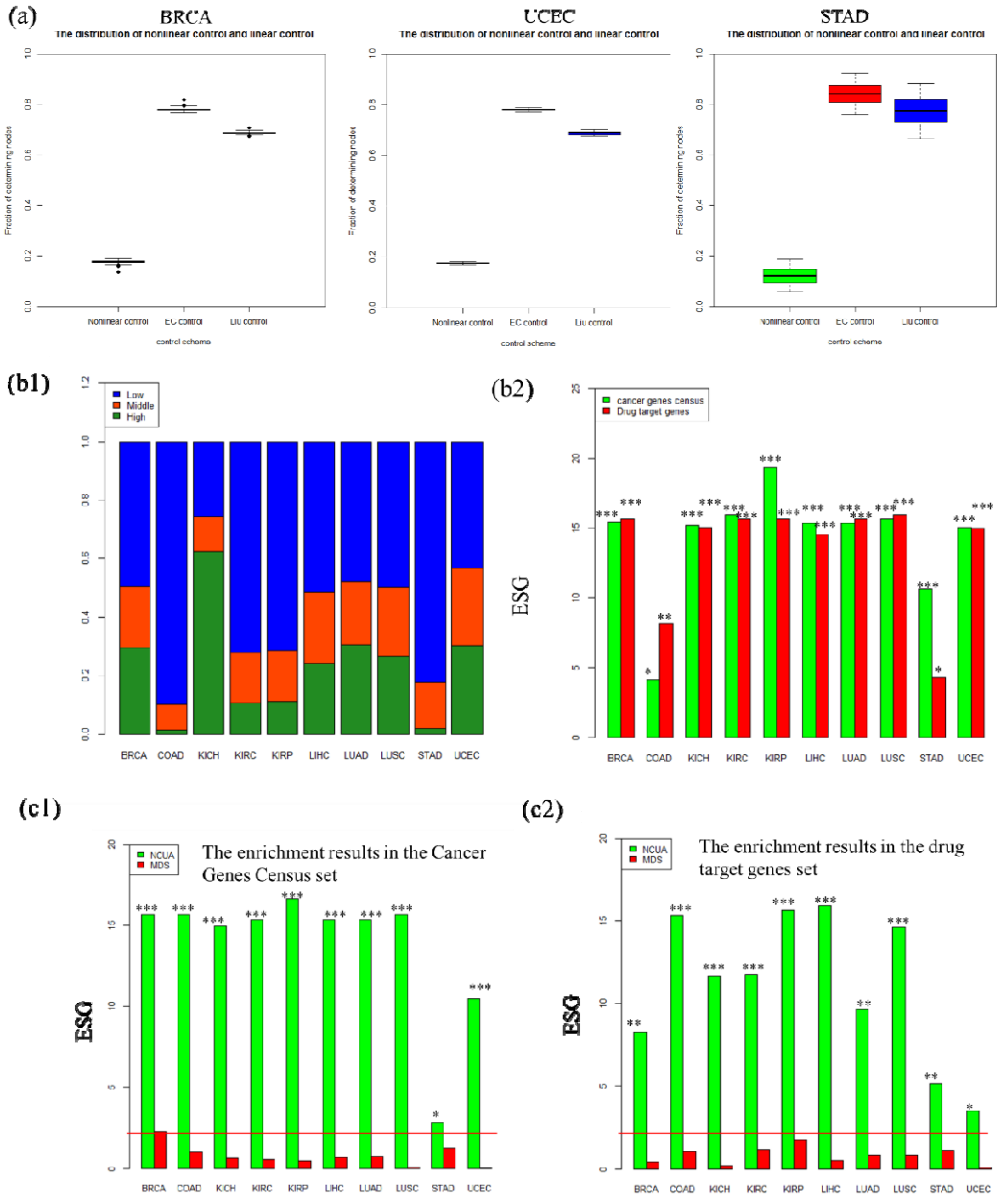

Figure 5. Applying the NCUA to discover the individual phenotype-transition genes in cancer. **(a)** Box plot with the distribution of the fraction of control nodes in the NCU and linear control (EC control and Liu's control scheme) methods, respectively, for each single sample network for the KICH, BRCA, and LUAD cancer data. **(b1)** The frequency distribution of the determining control genes in all samples of different cancers. (b2) The p-value of the high-confidence determining control genes enriching the cancer genes census set or FDA-approved drug targeted genes set. (c1) The significant enrichment of the input nodes of the NCU control model, which are not in the minimum dominating nodes in the Cancer Genes Census set and FDA-approved drug targeted genes set. (c2) The significant enrichment of the input nodes of the MDS control model, which are not in the NCU nodes in the Cancer Genes Census set and FDA-approved drug targeted genes set. Note: *scores of the ESG are larger than 5, but less than 10; **scores of the ESG are larger than 10, but less than 15; ***scores of the ESG are larger than 15.

Discussions

Nacher and Akutsu introduced the MDS to study the controllability of undirected networks by assuming that each node in the MDS can control all of its outgoing edges separately [8]. However, the MDS-based model assumes that more powerful control is possible (because each driver node can control its outgoing links independently), which has the possible drawback of requiring higher costs and may not be possible in many kinds of networks. Even if such powerful controllers exist, the MDS-based model still suffers from the underestimated nonlinear control of complex systems (networks). Despite its success and widespread application in searching for the important genes in the protein interaction network [16,47-50], the MDS-based model may give an incomplete view of the undirected network control properties. In the case of a network with nonlinear dynamics, the definition of control (full control; from any initial to any final state) for the MDS-based model does not always match the meaning of control in biological, technological, and social systems, where control tends to involve only naturally occurring system states.

In this work, our control model NCU drives the whole-networked system from the initial state toward its desired dynamical attractors (e.g., the steady states and limit state cycles) by steering the input nodes to the desired dynamic attractors. Our NCU algorithm (NCUA) predicts the input nodes whose override (by an external controller or drive signals) can steer a network's dynamics toward its desired long-term dynamic behaviors (its desired dynamical attractors). Furthermore, we used the NCU control model on biological, technological, and social networks, and we identified the topological characteristics underlying the predicted node overrides. We also identified that the networks with a low average degree are easier to control than those with a large average degree, which is opposite to the previous observation from the MDS theory, as shown in Figure S7. We summarize the difference between the MDS-based method and the NCU method in Table S2 of Additional File 1.

The NCU and MDS methods are very different methods, so one should be careful about extending their predictions beyond their realm of applicability. In fact, in the case of network with MDS's assumption, the key nodes identified using our

NCU control model can provide sufficient conditions to control the system from any initial state to any desired final steady state. For example, in Figure 1, the key nodes $\{v_1, v_4, v_9\}$ identified using our NCU control model dominate the nodes of the whole network for the MDS model, but the key nodes $\{v_1, v_4\}$ identified using the MDS model cannot cover the edges of the whole network for our NCU control model. To further emphasize the advantage of the NCU method over the MDS method, we provide the enrichment results from the CGC set and DTG set of the input nodes, which are nodes of the NCU, but not of the MDS for individual (paired) samples in the 10 cancer datasets. Figure 5(c) shows that the NCUA can identify the key genes in the CGC set and the FDA-approved DTG set, which are missed using the MDS method. The NCU model provides us with a more complete insight into the control of network-based systems.

Conclusions

Generally, complex biological networks whose data are limited can be diagrammed less accurately than networks, such as power grid networks. Recently, several control principles have been developed to control complex networks, but controlling complex biological networks is still hindered by network data [24]. In biological networks, we usually utilize undirected networks to model the protein interactions. Controlling the network dynamics by regulating some key nodes in the undirected networks to achieve optimal performance is still a big challenge. Two conventional control frameworks for undirected network dynamics, that is, the exact controllability and MDS-based model, focus on the linear dynamics in undirected networks. A theoretical control framework is urgently required to solve the nonlinear control problem in the undirected networks. Instead of focusing on how to obtain the state transitions of the undirected network with linear dynamics, a new concept, the nonlinear control of undirected networks (NCU), is introduced to understand how we can estimate the ability of the proper set of input nodes to achieve the control from the initial attractor to the desired attractor in undirected networks. To solve this problem, an NCUA based

on feedback vertex sets was designed and implemented. The NCUA has been evaluated on multiple synthetic SF networks and real complex networks, and it has exhibited the novel control characteristics of the undirected SF networks with nonlinear dynamics. The NCUA has also been applied to investigate the networks and their nonlinear controllability of cancer samples from TCGA by screening known driver genes and known drug targets as controls of their phenotype transitions, as well as to provide meaningful predictions with biological significance. Interestingly, we find that the control performance of our nonlinear control method in the single-patient system in cancer is much better than that of the traditional linear control methods, which are limited to a canonical linear time-invariant approximation. The key control genes for the individual cancer samples have significant enrichments both in the CGC set and the FDA-approved DTG set. Furthermore, it is worth exploring how to solve the NCU model with more constrained conditions (such as the target control and constrained target control [5,6,13]) and how to extend our method to the edge dynamics [10] to create new avenues to tackle complex systems. Note that although the NCUA is applied to the analysis of undirected networks, we believe that, in the future, it can be extended to the analysis of directed or semi-directed networks after the implementation of a module processing technique on directed or semi-directed networks with a network community detection algorithm from the microcosmic perspective [47,51].

Acknowledgments

This paper was supported by the National Natural Science Foundation of China (61873202, 61473232, 91430111, 91439103, 91529303, 31771476, 81471047 and 11871456) and National Key R&D Program (2017YFA0505500), Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB13040700), National Key R&D Program (Special Project on Precision Medicine) (2016YFC0903400), and Natural Science Foundation of Shanghai (17ZR1446100). We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

References

1. Boguñá M (2006) The Structure and Dynamics of Networks: Princeton University Press. 419-421 p.
2. Barabási AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nature Reviews Genetics 12: 56.
3. Lin CT (1974) Structural controllability. IEEE Transactions on Automatic Control 19: 201-208.
4. Lombardi A, Hörnquist M (2007) Controllability analysis of networks. Physreve 75: 056110.
5. Gao J, Liu YY, D'Souza RM (2014) Target control of complex networks. Nature Communications 5: 5415.
6. Guo WF, Zhang SW, Wei ZG, Zeng T, Liu F, et al. (2017) Constrained target controllability of complex networks. Journal of Statistical Mechanics Theory & Experiment 2017: 063402.
7. Yuan Z, Zhao C, Di Z, Wang WX, Lai YC (2013) Exact controllability of complex networks. Nature Communications 4: 2447.
8. Nacher JC, Akutsu T (2012) Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control. New Journal of Physics 14: 73005-73028(73024).
9. Nacher JC, Akutsu T (2013) Structural controllability of unidirectional bipartite networks. Scientific Reports 3: 1647.
10. Nepusz T, Vicsek T (2012) Controlling edge dynamics in complex networks. Nature Physics 8: 568-573.
11. Wu FX, Wu L, Wang J, Liu J, Chen L (2014) Transittability of complex networks and its applications to regulatory biomolecular networks. Sci Rep 4: 4819.
12. Fiedler B, Kurosawa G, Saito D (2013) Dynamics and Control at Feedback Vertex Sets. I: Informative and Determining Nodes in Regulatory Networks. Journal of Dynamics & Differential Equations 25: 563-604.
13. Liu, Y.-Y., Slotine, J.-J. and Barabási, A.-L. (2011) Controllability of complex networks. Nature, 473, 167-173.
14. Liu X, Pan L (2014) Detection of driver metabolites in the human liver metabolic network using structural controllability analysis. BMC Systems Biology,8: 51.
15. Guo WF, Zhang SW, Liu LL, Liu F, Shi QQ, et al. (2018) Discovering personalized driver mutation profiles of single samples in cancer by network control strategy. Bioinformatics.
16. Wuchty S (2014) Controllability in protein interaction networks. Proceedings of the National Academy of Sciences of the United States of America 111: 7156.
17. Jgt Z, Yang G, Albert R (2017) Structure-based control of complex networks with nonlinear dynamics. Proceedings of the National Academy of Sciences of the United States of America: 7234-7239.
18. Mochizuki A, Fiedler B, Kurosawa G, Saito D (2013) Dynamics and control at feedback vertex sets. II: a faithful monitor to determine the diversity of molecular activities in regulatory networks. Journal of Theoretical Biology 335: 130-146.
19. Lai YC (2014) Controlling complex, non-linear dynamical networks. National Science Review 1: 339.
20. Schiff S, Whalen A, Brennan S, Sauer T (2015) Observability and Controllability of Nonlinear

Networks: The Role of Symmetry. Physical Review X 5.
21. Wang LZ, Su RQ, Huang ZG, Wang X, Wang WX, et al. (2016) A geometrical approach to control and controllability of nonlinear dynamical networks. Nature Communications 7: 11323.
22. Karl S, Dandekar T (2015) Convergence behaviour and Control in Non-Linear Biological Networks. Scientific Reports 5: 9746.
23. Sun YZ, Leng SY, Lai YC, Grebogi C, Lin W (2017) Closed-Loop Control of Complex Networks: A Trade-Off between Time and Energy. Physrevlett.
24. Li M, Gao H, Wang J, Wu F-X (2018) Control principles for complex biological networks. Briefings in Bioinformatics: bby088-bby088.
25. Müller FJ, Schuppert A (2011) Few inputs can reprogram biological networks. Nature 478: E4; discussion E4.
26. Lenstra HW (1983) Integer Programming with a Fixed Number of Variables. Mathematics of Operations Research 8: 538-548.
27. Williams HP (1990) Integer and Combinatorial Optimization. Journal of the Operational Research Society 41: 177-178.
28. Lund C, Yannakakis M (1994) On the hardness of approximating minimization problems. Journal of the Acm 41: 960-981.
29. Goh KI, Kahng B, Kim D (2001) Universal behavior of load distribution in scale-free networks. Physical Review Letters 87: 278701.
30. Gao J, Barzel B, Barabási AL (2016) Universal resilience patterns in complex networks. Nature 530: 307.
31. Clauset A, Shalizi CR, Newman MEJ (2009) Power-Law Distributions in Empirical Data. Siam Review 51: 661-703.
32. FJ M, A S (2011) Few inputs can reprogram biological networks. Nature 478: 4-5.
33. Kramer AD, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences of the United States of America 111: 8788.
34. Müller FJ, Schuppert A (2011) Few inputs can reprogram biological networks. Nature 478: 4-5.
35. Zañudo JG, Albert R (2014) Cell fate reprogramming by control of intracellular network dynamics. Plos Computational Biology 11.
36. Huang W, Sun H, He W (2010) Pinning control of complex networks with general topology. 51: 360-364.
37. Wang XF, Chen G (2002) Pinning control of scale-free dynamical networks. Physica A Statistical Mechanics & Its Applications 310: 521-531.
38. Liu X, Wang Y, Ji H, Aihara K, Chen L (2016) Personalized characterization of diseases using sample-specific networks. Nucleic Acids Research 44: e164-e164.
39. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, et al. (2015) TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Research 44: e71-e71.
40. Yu X, Zhang J, Sun S, Zhou X, Zeng T, et al. (2017) Individual-specific edge-network analysis for disease prediction. Nucleic acids research 45: e170.
41. Bertrand D, Chng KR, Sherbaf FG, Kiesel A, Chia BK, et al. (2015) Patient-specific driver gene prediction and risk assessment through integrated network analysis of cancer omics profiles. Nucleic Acids Research 43: e44-e44.

42. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, et al. (2004) A census of human cancer genes. Nature Reviews Cancer 4: 177-183.
43. David, Craig, Knox, An, Chi, et al. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Research 36: 901-906.
44. Rivals I, Personnaz L, Taing L, Potier MC (2007) Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics 23: 401-407.
45. Jeong H, Mason SP, Barabási AL, Oltvai ZN (2001) Lethality and centrality in protein networks. Nature 411: 41.
46. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, et al. (2008) High-quality binary protein interaction map of the yeast interactome network. Science 322: 104-110.
47. Sun PG, Ma X (2017) Understanding the controllability of complex networks from the microcosmic to the macrocosmic. New Journal of Physics 19.
48. Sun PG (2015) Controllability and modularity of complex networks. Information Sciences 325: 20-32.
49. Ishitsuka M, Akutsu T, Nacher JC (2016) Critical controllability in proteome-wide protein interaction network integrating transcriptome. Scientific Reports 6: 23541.
50. Wuchty S, Boltz T, Küçükmcginty H (2017) Links between critical proteins drive the controllability of protein interaction networks. Proteomics 17.
51. Guo WF, Zhang SW (2016) A general method of community detection by identifying community centers with affinity propagation. Physica A Statistical Mechanics & Its Applications 447: 508-519.