Briefings in Bioinformatics, 20(6), 2019, 2253–2266

doi: 10.1093/bib/bby088
Advance Access Publication Date: 18 September 2018
Review Article

Control principles for complex biological networks

Min Li, Hao Gao, Jianxin Wang and Fang-Xiang Wu

Corresponding author: Min Li, School of Information Science and Engineering, Central South University, Changsha, Hunan 410083, China. E-mail: limin@mail.csu.edu.cn; Fang-Xiang Wu, School of Information Science and Engineering, Central South University, Changsha, Hunan 410083, China and Division of Biomedical Engineering and Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada. E-mail: faw341@mail.usask.ca

Abstract

Networks have been widely used to model the structure of various biological systems. Currently, a series of approaches have been developed to construct reliable biological networks. However, the ultimate understanding of a biological system is to steer its states to the desired ones by imposing signals. The control process is dominated by the intrinsic structure and the dynamic propagation. To understand the underlying mechanisms behind the life process, the control theory can be applied to biological networks with specific target requirements. In this article, we first introduce the structural controllability of complex networks and discuss its advantages and disadvantages. Then, we review the effective control to meet the specific requirements for complex biological networks. Moreover, we summarize the existing methods for finding the unique minimum set of driver nodes via the optimal control for complex networks. Finally, we discuss the relationships between biological networks and structural controllability, effective control and optimal control. Moreover, potential applications of general control principles are pointed out.

Key words: biological networks; control principles; structural controllability; optimal control

Min Li received the BS degree in communication engineering and the MS and PhD degrees in computer science from Central South University in Changsha, China, in 2001, 2004 and 2008, respectively. She is currently a professor at the School of Information Science and Engineering at Central South University. Her main research interests include bioinformatics and systems biology.

Hao Gao received the BS degree in surveys technology and instrument from Central South University in China in 2017. Currently, he is working toward the MS degree in computer science at the Central South University in Changsha, China. His current research interests include control principles for large biological networks and system biology.

Jianxin Wang received the BEng and MEng degrees in computer engineering from Central South University in Changsha, Hunan, China, in 1992 and 1996, respectively, and the PhD degree in computer science from Central South University in 2001. He is the chair of and a professor in the Department of Computer Science at Central South University. His current research interests include algorithm analysis and optimization, parameterized algorithm, bioinformatics and computer networks.

Fang-Xiang Wu received both the BSc and MSc degrees in Applied Mathematics from Dalian University of Technology in 1993 and 1990, respectively, and two PhD degrees, the first one in control theory and its applications from Northwestern Polytechnical University in 1998 and the second one in bioinformatics and computational biology at the University of Saskatchewan (U of S) in August 2004. From September 2004 to August 2005, he worked as a postdoctoral fellow with the Laval University Biomedical Research Center. He is currently a full professor in Department of Computer Science, Division of Biomedical Engineering, and Department of Mechanical Engineering at the U of S. His research mainly focuses on bioinformatics and computational systems biology, specifically in biomolecular network analytics, machine learning in bioinformatics, big biological/clinical data analytics, complex bionetwork controllability, brain image and brain networks and nonlinear biodynamic analytics. So far, he has published more than 300 refereed journal/conference papers. He has been invited to give research talks in about 70 international conferences or academic institutes worldwide. Dr Wu is serving as the editorial board member of six international journals and as the guest editor of numerous international journals and as the program committee chair or member of over 30 international conferences.

Submitted: 6 June 2018; **Received (in revised form):** 16 August 2018

Introduction

The last decade has witnessed paramount development of network science from several interdisciplinary areas, mainly in engineering, physics, mathematics, biology and social science [1, 2]. As the large data are assembling and the ability of computation is enhancing, more and more complex networks have been constructed for modeling and understanding the real complex systems. Most of complex networks have been found to show special properties including the small-world property [3] and the scale-free property [4]. With the development of network science, network structure and dynamics have been the key factors for network analysis [5]. The structural properties mainly contain nodal centrality for measuring the importance of a node [6, 7] and network communities for detecting the set of nodes with close relationships [8, 9], both of which try to identify the important nodes in complex networks. Jeong *et al.* [10] found that the proteins with high degree in protein–protein interaction (PPI) network play an important role in the survival of cells. Norton *et al.* [11] proposed a graph-based method called 3DNetMod for communities detection. On the other hand, the dynamics of complex networks has been used to analyze the mechanism of real systems [12], such as epidemic spreading [13]. Wu *et al.* [14] found that cell fate can be determined by initializing the state of gene regulatory networks (GRNs). Abel *et al.* [15] considered the average dynamics of the population as a single oscillator to control biological time. The study of controllability is to identify the set of steering nodes through the dynamics model. Generally, controlling nonlinear dynamics system is still so difficult that the linear system can be studied at the first step.

Although classical linear control theory is suitable for complex biological networks mathematically, there are still great challenges for controlling complex biological networks. The linear control theory is built on the state space represented by vectors and matrices. The purpose of control is to steer the states of complex systems to the optimal or desired states. Several approaches have been proposed to judge whether a linear system is controllable, including the Kalman's controllability condition [16] and Popov-Belevitch-Hautus (PBH) controllability condition [17], both of which can be algebraically verified in principle. However, the real complex biological networks usually have hundreds and thousands of nodes so that the algebraic approaches are prohibited. In addition, it is still impossible to accurately determine the parameters in complex biological systems although the technologies have been developed rapidly. Hence, the control principles of complex biological networks should be robust to large-scale models and parameter inaccuracy.

Recently, the concept of structural controllability gives us the ability to investigate the controllability of complex biological networks through a minimum set of steering nodes [18]. Structural controllability was first proposed by Lin [19] based on graph theory, which focuses on the structure of systems. In structural controllability, some special structures have been defined for testing network controllability. The graph-theoretic framework has two distinct advantages. First, structural controllability only has to know whether there is an edge or not, which is good for complex biological networks whose structures are known easier than their parameter values. Second, the graph-theoretic methods can be tested with efficient algorithms instead of computing complex matrix operations.

Structural controllability connects the structure with the dynamics of complex biological networks. Under this framework, the study of network structure can help understand the mathematical dynamics. Liu *et al.* [18] defined the classification of nodes as critical, redundant and ordinary nodes if its absence increases, decreases or equals to the driver nodes, which can be used to study the robust of network controllability [20, 21]. Wang *et al.* [22] showed that the networks with strong power-law degree distribution are easier to control, indicating a new way by adding edges to optimize structural controllability. Pósfai *et al.* [23] found that the degree correlation between in-degree and out-degree has a robust effect on network controllability while the clustering and modularity don't. Wang *et al.* [24] derived the control range to measure the size of subnetwork a node can control. Similarly, Liu *et al.* [25] proposed the control centrality to quantify the ability of a single node to control a directed complex network, which is closely related to hierarchical structure in complex networks. Jia *et al.* [26] redefined a classification of nodes as critical, intermittent and redundant nodes if it acts as a driver node in all, some or none of the control configurations and discovered two important control modes for complex systems: centralized and distributed control. Ruths *et al.* [27] defined the control profiles of complex networks from the properties of structure, which found that the control profiles in real-world networks are different from the random network models. Jia *et al.* [28] formulated the control capacity measure for quantifying the importance of a node, indicating that the possibility decreases by the in-degree while it is independent of the out-degree of the node. Menichetti *et al.* [29] showed that the density of nodes with in-degree and out-degree equal to one or two determines the number of driver nodes required to control complex networks. In addition, several control strategies have been proposed to control complex biological networks. Furthermore, Nepusz *et al.* [30] used switchboard dynamics model to control edge dynamics instead of nodal dynamics through signals imposed on the minimum set of nodes. Moreover, Yuan *et al.* [31] proposed a general control framework in which the complex weighted and undirected networks can also be controlled.

With the development of structural controllability, many applications to complex biological networks have verified that structural controllability can provide meaningful results, which can be divided into three aspects. The first aspect is based on Liu's classification [18], which has been applied to identify disease genes and drug targets [32], viral targeted proteins [33] in the directed human PPI (dPPI) network and find robust control structures in yeast stress response pathways [34]. Similarly, Jia's classification [26] is able to be applied to detect driver metabolites in the human liver metabolic network [35], driver proteins in human signaling network [36] and critical regulatory genes in cancer signaling network [37] as well. What's more, Wang *et al.* [38] studied both type classifications in a gene network for *Arabidopsis*, indicating that different sets of nodes may be preferentially related to specific biological function and progress by Gene Ontology enrichment analysis. The last aspect uses the control centrality [25] to identify dysregulated pathways in the tissue-specific GRN [39]. These applications can be found in Table 1.

Though the tools to analyze biological networks have been developed [40], the ultimate purpose to control complex biological systems is still too far to be directly solved by structural controllability. First, based on the fully structural controllability, the minimum set of driver nodes (MDSs) may contain a large proportion of nodes. However, the biological system is built to suit for different situations. The complex biological networks may have large non-functional nodes in special processes, indicating that full controllability should be shifted to partial controllability. On the other hand, the fully structural controllability cannot

Table 1. The list of applications to complex biological networks

Control principle	Biological network	Species	Notes	Year	Ref
Structural controllability	dPPI network	Human	Identify disease genes and drug targets	2016	[32]
	Liver metabolic network	Human	Detect driver metabolites	2014	[35]
	Signaling network	Human	Identify driver proteins	2015	[36]
	Stress response network	Yeast	Find robust control structures in yeast stress response pathways	2016	[34]
	Cancer signaling network	Human	Identify critical regulatory genes	2017	[37]
	Tissue-specific GRN	Human	Identify dysregulated pathways	2017	[41]
	Multitype networks	Human	Identify cancer-specific targets	2017	[42]
	GRN	*Arabidopsis thaliana*	Identify essential genes	2018	[38]
	HIV-1 infected network	Human	Identify viral targeted genes	2018	[33]
Control partial nodes	Neural network	*C. elegans*	Predict neuron function	2017	[43]
	dPPI network	Human	Identify essential proteins	2017	[44]
	Functional interactions network	Human	Discovering personalized driver profiles	2018	[45]
Control two-state transition	p53-mediated DNA damage response network	Human	Identify steering kernel	2014	[46]
	T helper differentiation cellular network	Human			
	Yeast cell cycle network	Human			
	EMT network	Human			
Optimal control	Brain network	Human	Identify controllable regions	2015	[47]
		Human	Cognitive control	2016	[48, 49]
		Human	Predicted dynamics in development	2017	[50]
		Human	Creativity and intelligence	2018	[51]

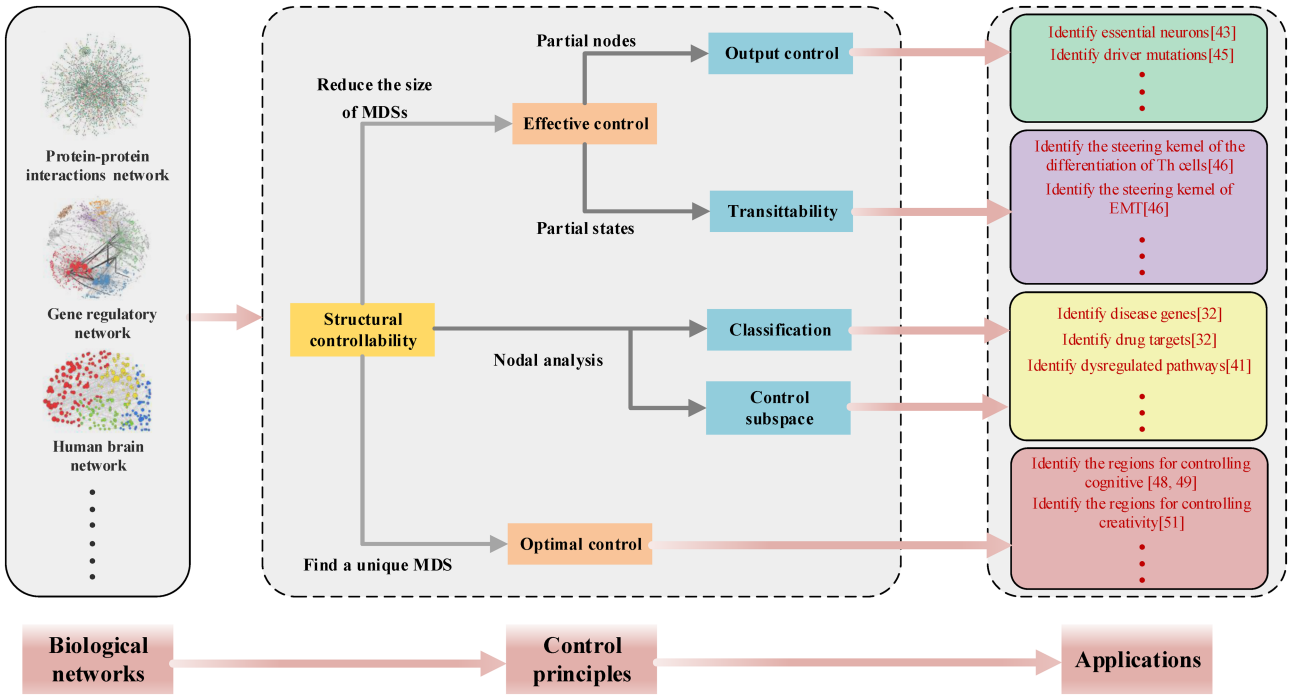

Figure 1. An outline of the control principles for biological networks. The biological networks serve as the complex dynamics system models to design powerful control principles so that different control tasks in biology can be understood distinctly.

give a reasonable set of driver nodes, indicating that the control strategy has not been well understood with structural controllability yet. From the aspect of biological function and systems biology, the structural controllability is non-efficient and non-optimal, the limitation of which hinders the applications of control theory.

In the rest of this article, as shown in Figure 1, we first introduce the structural controllability theory. Next, we discuss how to design effective control for complex biological networks. Then, we show how to address the controllability of complex biological networks based on the control energy. Finally, we emphasize the demands of effective control and optimal control in complex biological networks and discuss the challenges of general control principles which can be potentially applied to complex biological networks.

Controllability of complex networks

Generally, the classical linear control theory has its computational limitations to be applied to complex biological networks that may contain thousands of nodes, such as the *Saccharomyces cerevisiae* PPI network which has 1870 proteins [10]. Such challenges force us to develop feasible and effective approaches to determine whether the state of complex biological systems can be steered to its desired state. An important framework derived from the classical control condition is the structural controllability [18], which uses graph-theoretic algorithms to judge whether a complex network is controllable, even though the weight of edges may be unknown. So far, this approach has been widely applied to analyze the mechanism of biological networks.

Controllability of linear systems

The linear, time-invariant (LTI) dynamic system has been widely used as the model of systems, which can be described as follows [52]:

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \tag{1}$$

where A is a $N \times N$ state transition matrix representing interactions between state nodes, where $a_{ij} \neq 0$ if state node j affects state node i. B is a $N \times M$ input matrix associated to input control signals, where the strength of the j-th external signal injects into state node i denoted by b_{im}. C is a $S \times N$ output matrix describing the relationship between the measured signals and state nodes. $x(t) = \{x_1(t), x_2(t), \ldots, x_N(t)\}^T$ is the state vector representing the activity values of N state nodes at time t. $y(t) = \{y_1(t), y_2(t), \ldots, y_S(t)\}^T$ is the output vector representing the measured signals from the network. $u(t) = \{u_1(t), u_2(t), \ldots, u_M(t)\}^T$ is the control signals actuated on the driver nodes.

Controllability quantifies the ability to drive a system from an initial state to a desired final state in finite time. Kalman's rank condition [16] was first proposed to check whether an *LTI* system is fully controllable. Mathematically, a system (A, B) is fully controllable if and only if the controllability matrix

$$\mathcal{C}_C \equiv \left[B, AB, A^2B, \ldots, A^{N-1}B\right] \tag{2}$$

has the full rank, that is,

$$\text{rank}\,(\mathcal{C}_C) = N \tag{3}$$

The concept of controllability has been successfully applied to control complicated behavior of robot [53, 54] and optimize the process of industry [55, 56]. However, when N is large, the elements of A^N are very large (or small) when the absolute value of the eigenvalue of A is larger (smaller) than one. Hence, the controllability matrix $\mathcal{C}_C$ can become ill-defined and the rank $(\mathcal{C}_C)$

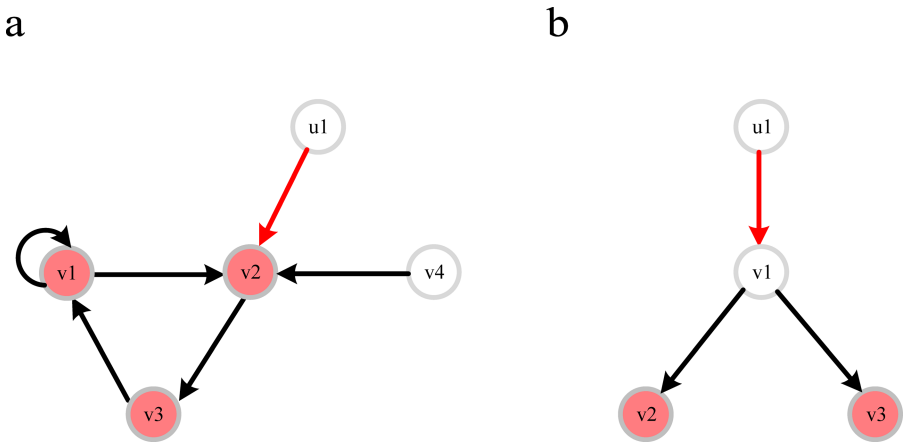

Figure 2. Inaccessibility and dilation. (a) There is no path from u_1 to v_4, and therefore, node v_4 is inaccessible. (b) Considering a set $S = \{v_2, v_3\}$, we have $T(S) = \{v_1\}$. Since $|T(S)| < |S|$, there exists a dilation.

cannot be accurately calculated. Moreover, the entries of A are impossible to be measured exactly in practice, directly resulting in worse computational results. These limitations in computing the rank of controllability matrix of large complex networks prompt us to develop efficient algorithms from the aspect of graph theory.

Structural controllability of complex networks

Structural controllability theorem

We introduce the Lin's structural controllability of *LTI* system [19], which has clear graph-theoretic interpretations. Usually, an *LTI* system can be represented by a digraph $G(A, B) = (V, E)$. The node set $V = V_A \cup V_U$ contains both state nodes V_A and input nodes V_U. The edge set is $E = E_A \cup E_U$ where E_A contains edges among state nodes in V_A while E_U contains edges between input nodes V_U and state nodes V_A. An *LTI* system is called a structural system (A, B) if the entries in A and B are either fixed zeros or independent free parameters. A structural system (A, B) is fully structurally controllable if it satisfies Kalman's condition [16] with some setting of the non-zero entries of A and B. We introduce two important definitions before the description of structural controllability theorem.

Consider a network represented by a digraph $G(A, B)$ (Figure 2),

Definition 1. *(Inaccessibility). A state node v_i is inaccessible in graph $G(A, B)$ if and only if there are no directed paths reaching v_i from any input nodes V_u.*

Definition 2. *(Dilation). A digraph $G(A, B)$ contains a dilation if and only if there is a subset of nodes $S \in V_A$ such that $|T(S)| < |S|$, where $T(S)$ is the neighborhood set of S containing the set of all nodes v_j which point to any node in S.*

Based on two definitions above, we can state the sufficient and necessary conditions for the structural controllability of a controlled network.

Theorem 1. *(Structural controllability theorem [19]). A structural system (A, B) is structurally controllable if and only if*

(i) *the digraph $G(A, B)$ contains no dilations,*
(ii) *no node in V_A is inaccessible.*

The structural controllability theorem has algebraic interpretations. In the view of algebra, an *LTI* system has inaccessible nodes if the structured matrix $[A; B]$ is reducible, i.e. there exists a permutation matrix P such that

$$A = P \begin{bmatrix} A_{11} & 0 \\ A_{12} & A_{22} \end{bmatrix} P^{-1} \text{ and } B = P \begin{bmatrix} 0 \\ B_2 \end{bmatrix} \tag{4}$$

where $A_{11} \in R^{(N-K)\times(N-K)}$, $A_{22} \in R^{K\times(N-K)}$ and $B_2 \in R^{K\times M}$ with $1 \le K \le N$. An *LTI* system has dilation if the structural matrix $[A; B]$ has generic rank less than N, i.e.

$$\text{rank}_g [A; B] < N \tag{5}$$

Hence, the system (A, B) is structural controllable if and only if both the structure matrix $[A; B]$ is irreducible and has generic rank equal to N.

Minimum input theorem

Generally, structural controllability is a yes/no concept, which cannot provide a feasible control strategy. Given a set of signals imposed on network, we can judge whether the system is controllable. However, if we impose signals on every node of a network, the network is controllable in any case, which is trivial. However, it is non-trivial to find out the minimum number of input signals required to steer the state of system to any desired state. Liu *et al.* [18] first formed the problem as an optimization problem named the MDSs and theoretically proved that the MDSs can be determined by the maximum matching algorithm proposed by Hopcroft and Karp [57].

To obtain the set of MDSs, we first introduce the concept of maximum matching.

Definition 3. *(Maximum matching). A maximum matching M is the maximum set of edges which do not share the same start node or end node. The end node of the matching edge is called matched; the others are called unmatched.*

From the definition of structural controllability, each node must have its unshared father node in order to make a system fully controllable. The inaccessible nodes cannot be influenced by signals from the others, indicating that the node doesn't have its unshared father node. The isolated node is a simple example. On the other hand, the structure of dilation cannot transfer independent signals from a node to more than one node, indicating that only one child node can have its unshared father node in the structure. The relationship between a father node and a child node forms a specific matching edge in structural

controllability. Obviously, a child node is the matched node, which can be controlled by its unshared father node independently. Moreover, an unmatched node does not have unshared father node, which should be controlled by an external signal. Based on such analysis, we can use maximum matching to find out the minimum set of unmatched nodes. In summary, the minimum inputs theorem can be described as:

Theorem 2. *(Minimum input theorem). If all nodes are matched, the matching is perfect and the number of driver node is 1; otherwise, the number of MDSs is* $N_D = N - M^*$*, where* M^* *is the number of matched nodes.*

Based on minimum input theorem, Liu *et al.* [18] found that the structure of complex networks may determine how to impose the controllers. The conclusion can be described from three important properties of structural controllability. First, the MDSs are determined by the degree distribution of network. Second, it is hard to control heterogeneous networks than homogeneous networks. Finally, the set of driver nodes tends to avoid the high-degree nodes in both real and model networks. However, MDSs from the minimum input theorem (Theorem 2) is a necessary condition for fully structural controllability. As it can only guarantee that condition (1) in structural controllability theorem, but not condition (2), is satisfied. For example, the source strongly connected component (sSCC), which is a cycle without input edges, is always a perfect matching, but all nodes in sSCC are inaccessible from the input signals [58].

Applications to biological networks

Identify disease genes and drug targets

The dPPI network can be inferred from original PPI data which maps the mutual regulations supporting the process of life in a cell [59]. The dPPI network contains 6339 proteins and 34 813 interactions, where the edge direction and the edge weight are related to the signal flow along the interaction proteins and the confidence of the predicted direction, respectively. A recent study applied the structural controllability theory to classify these proteins into one of the following three categories: indispensable, dispensable and neutral if its absence increases, decreases or equals to the number of driver nodes [32]. The result of classification showed that 21% of proteins are indispensable, 37% of proteins are dispensable and 42% of proteins are neutral. In the context of classification, the indispensable node proteins have been found its biological regulation effect (e.g. disease genes and drug targets) on altering a cell state transition between healthy and disease states. To further study the indispensable proteins, 56 genes had been discovered to be related to cancer through analyzing data from 1547 cancer patients, 46 of which are new potential disease genes of cancer.

Identify dysregulated pathway

Complex diseases usually are induced by a set of genes in dysregulated pathways, rather than a single gene [60]. GRNs primarily built from gene expression data have been used to predict dysregulated pathways [61]. Considering the tissue-specific GRN of Type 2 Diabetes (T2D), the set of genes predicted to control the tissue-specific GRN may be the reason inducing the progress of T2D. In the area of structural controllability, control centrality had been proposed, which quantifies the ability of a single node to control a directed complex network [25, 62]. Hence, a recent study found that those genes with high control centrality pathways might control other downstream genes resulting in disease manifestation [39]. The prediction of dysregulated pathways based on control centrality is superior to other centrality measures, including betweenness centrality, degree centrality, eigenvector centrality and closeness centrality. Moreover, the mechanistic connections of NFATC4 with downstream targets have been discovered from four important T2D pathways.

General limitations to biological networks

The results of structural controllability may be helpless to design efficient control strategy. The driver nodes to control biological networks are more than 80% of all nodes offering useless prediction of drug targets in biological experiment [63]. The phenomenon means that most of prediction nodes are useless and indicates that being fully controllable is not necessary. In practical applications, not all nodes can be altered by drugs. Hence, we may consider drug sensitivity, drug toxicity and so on to measure the ability of drugs. Furthermore, not all states need to be achieved. Hence, we are interested in transitting between disease states and health states so that the specific and personalized combination of drugs can be identified with known databases, which can be one of strategies for drug repurposing. On the other hand, the set of driver nodes usually is not unique for complex networks. Given a set of certain MDSs, the control trajectory may be non-local in phase space and the transition of states may diverge, which may result in control failure [64]. Because the control energy in some directions is very large so that it is only theoretically correct but practically infeasible. The practical explanation is that the dose of drugs must be limited. Hence, these principles should consider both control nodes and control energy so that they are feasible to biological networks.

The structural controllability of directed networks assumes that all edge weights are independent while the edge weights are dependent in the undirected networks as the system matrix A of an undirected biological network is a symmetry matrix, where the elements a_{ij} and a_{ji} in A must be equal. Therefore, the structural controllability theory [18, 19] of directed networks isn't suitable for undirected biological networks. Fortunately, if the system matrix A of undirected networks is diagonalizable, the MDSs has been proved to be determined by the maximum algebraic multiplicity of A's eigenvalues [31]. Under this condition, we can also control the undirected biological networks by a set of driver nodes and utilize the above constraints to find optimal control principles for undirected biological networks.

Effective control for complex biological networks

As full controllability of complex networks is hardly effective in real applications, especially biological applications, we consider more effective control cases in this section. First, we consider that only partial nodes can participate in control. On the other hand, we consider that only specific states with biological meaning need to be reached.

Control partial nodes

Output control of complex networks

A linear dynamic system (A, B, C) described by (1) is output controllable if and only if any output state of network (1) can be steered to any desired output states within finite time t_f. Mathematically, a dynamic system (A, B, C) is controllable output

if and only if the rank of output control matrix

$$C_O \equiv \left[CB, CAB, CA^2B, \ldots, CA^{N-1}B\right] \tag{6}$$

is equal to S, where S is the dimension of the output states [16]. For a structural system (A, B), the free parameters in matrices A and B determine the dimension of its controllable subspace. We define the dimension of the controllable subspace as the generic dimension of the controllable subspace of the structural system (A, B) which is denoted by generic dimension of the controllable subspace (*GDCS*) (A, B) [65]. Similarly, the dimension of the output controllable subspace of a system (A, B, C) is defined as the generic dimension of the controllable output subspace and denoted by *GDCOS* (A, B, C). Hence, a system (A, B, C) is output controllable if and only if *GDCOS* $(A, B, C) = S$.

Though there are exact graph-theoretic algorithms to calculate *GDCS* (A, B) [65] efficiently, there is no exact graph-theoretic algorithm to calculate *GDCOS* (A, B, C) efficiently. Hence, an approximate method is used to provide the lower bound of *GDCOS* (A, B, C) for determining the minimum driver nodes [66]. A weighted bipartite graph had been constructed corresponding to the network matrix A and the output matrix C [67], and Kuhn–Munkres algorithm [68, 69] can be used to find out the maximum weight complete matching. The method can efficiently predict some special drug targets in arachidonic acid metabolic network and human pathway networks. A similar study naming output controllability as target control used a greedy algorithm to find the minimum driver nodes approximately based on graph-theoretic algorithm as well [70].

The core difference between output controllability and structural controllability is to choose eligible output or target nodes. Note that if every node is considered as an output node, the output controllability is the full controllability. Generally, target nodes are chosen according to the importance of nodes [71] or special structure [72]. For example, phenotype genes can act as target nodes in GRN when one is interested in steering a cell from abnormal state to healthy state.

Application to neuronal network

The nervous system of *Caenorhabditis elegans* is the only organism which had been mapped with reasonable accuracy at the cellular level [73]. The locomotion patterns of *C. elegans* are only determined by the motor muscles controlled by the muscle neurons. Hence, we can choose the muscle neurons of *C. elegans* as target nodes in neuronal networks and the sensory neurons as the sensory inputs to control the locomotion of *C. elegans* [43]. In this condition, if the removal of a neuron affects the locomotion of *C. elegans*, the neuron plays an important role in controlling the corresponding locomotion. With this analysis framework, several neurons had been found, in which some neurons have been experimentally validated to be related to the locomotion of *C. elegans*. An example is the ablation of both DA and DB results in the loss of backward/forward locomotion of *C. elegans*. Moreover, the new neuron 'PDB' had been theoretically identified its impact on signal propagation and verified its relationship with the locomotion of deep body bends by new experiments [74].

Constrained target control of complex networks

Constrained target control (CTC) is the same kind problem of output control, where CTC should choose both the constrained control nodes and partial target nodes. For a linear dynamic system described by (1) containing a set of nodes $V = \{v_1, v_2, \ldots, v_N\}$, we assume that $O = \{v_{c_1}, v_{c_2}, \ldots, v_{c_{N_o}}\}$ and $U = \{v_{b_1}, v_{b_2}, \ldots, v_{b_{N_c}}\}$ represent the set of target nodes and the set of constrained control nodes, respectively, where both $\{c_1, c_2, \ldots, c_{N_o}\}$ and $\{b_1, b_2, \ldots, b_{N_c}\}$ are the subset of $\{1, 2, \ldots, N\}$. N_o and N_c are the number of target nodes and the number of constrained control nodes, respectively. A system is constrained target controllable if and only if

$$\max\left\{\text{rank}\left[CB, CAB, CA^2B, \ldots, CA^{N-1}B\right]\right\} = N_o \tag{7}$$

where A is a $N \times N$ system matrix; B is a $N \times M$ input matrix constrained by the set of nodes in O, where $M \leq N_c$; and C is a $N_o \times M$ output matrix determined by the set of nodes in U. As shown in Figure 3, we demonstrate how to control partial nodes. It compares different definitions between target control and CTC.

A system must be target controllable when all constrained control nodes in U act as the driver nodes. This condition guarantees the upper bound of driver nodes so that the lower bound of driver nodes can be found by removing the redundant nodes. Sometimes, it is impossible to satisfy such a condition. An extreme example is that target nodes are all of the source nodes while constrained control nodes are all of the sink nodes. Hence, whether the method is feasible is closely related to the strategy of choosing both constrained control nodes and target nodes. On the other hand, identifying the minimum driver nodes in constrained control nodes is to meet the requirement of practices. Guo *et al.* [75] proposed a novel graph-theoretic algorithm named constrained target control algorithm (CTCA) to find the minimum driver node of a given network with constraints. The approach can efficiently identify the MDSs, some of which are approved drug targets and new potentials in the case study of biological networks.

Applications to identify driver mutations

Recently, a single-sample controller strategy (SCS) had been proposed to discover personalized driver mutation profiles of single samples through a driver mutation network for each patient [45], which is derived from a large reference network containing 11 648 genes and 211 794 edges [76]. The applications assumed that the differentially expressed genes are controlled by gene mutation through other genes. Hence, the mutation genes are the set of constrained control nodes and the differentially expressed genes are target nodes. The results indicate that SCS can improve the precision of predicted driver genes compared with Dawnrank [76], OncoImpact [77] and so on. Moreover, SCS can efficiently identify personalized driver genes, some of which are rare driver genes.

Control two-state transition

Transittability of complex networks

Unexpected state transitions of complex networks are very harmful phenomenon in real biological processes, most of which are irreversible process. Compared to fully structural controllability that concerns the possibility to steer the complex networks from any unexpected state to any desired state [18], the transittablity concerns the possibility to steer the complex networks from a specific unexpected state to a specific desired state [46]. It has been mathematically proved that a system is transittable between two specific states if and only if

$$\text{rank}\,(\overline{C}) = \text{rank}\,(C) \tag{8}$$

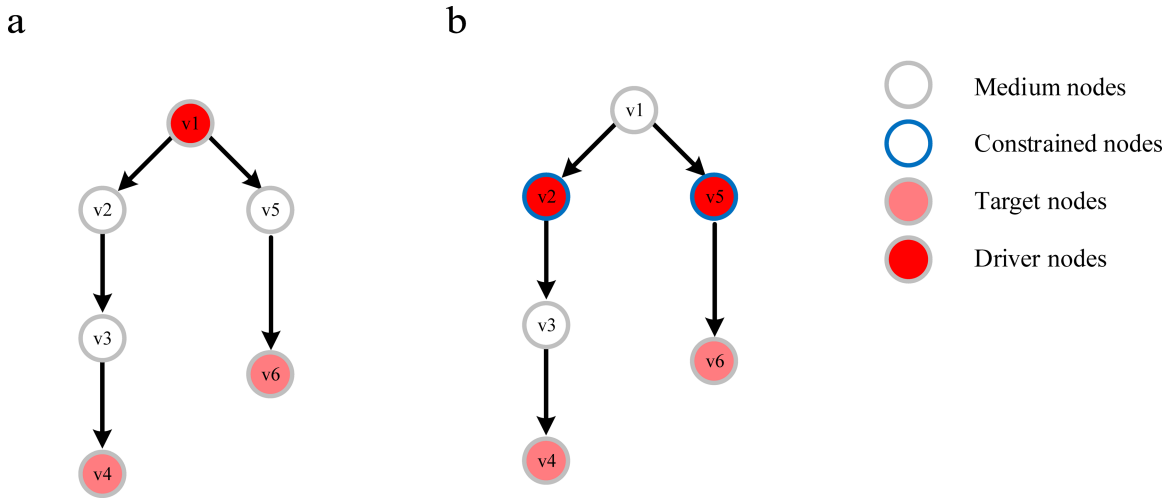

Figure 3. Target control and constrained target control. (a) In target control, the set of target nodes $\{v_4, v_6\}$ can be controlled by the set of a single driver node $\{v_1\}$. (b) In constrained target control, the set of target nodes $\{v_4, v_6\}$ can be controlled by the set of two driver nodes $\{v_2, v_5\}$.

where $\overline{C} = \left[\overline{B}, A\overline{B}, A^2\overline{B}, \ldots, A^{N-1}\overline{B}\right]$ and $\overline{B} = [x_0 - x_1, B]$. Generally, one of the states of transition must be stable. Given that x_1 be the stable state and $x_0 - x_1$ can be replaced by x_0 due to $x_1 = 0$ without loss of generality. Hence, a system is transittable from a specific state x_0 to the origin if and only if

$$\text{rank}\,(\mathcal{C}_0) = \text{rank}\,(\mathcal{C}) \tag{9}$$

where $\mathcal{C}_0 = \left[B_0, AB_0, A^2B_0, \ldots, A^{N-1}B_0\right]$ and $B_0 = [x_0, B]$. The form of equation (9) is very similar to Kalman's condition. Wu *et al.* [46] formulated the problem as an optimal assignment problem of a weighted bipartite graph to obtain the minimum steering nodes for transittability.

Transittability is more feasible than structural controllability with more practical constraints. First, transittability only focuses on effective transition. In biological network, states have specific functions in cell life, such as proliferation and apoptosis. These states are discrete, where any desired states defined in structural controllability are not suitable. Moreover, several state transitions are irreversible. When normal state has transferred to some disease state, we can usually cure the disease state to another healthy state, which may not be the original normal state. On the other hand, the disease states of complex networks are typically determined by a small set of nodes. In medical testing, we only need to detect a set of certain biomolecules to predict whether or not they are potentially sick. Furthermore, those unchanging nodes are redundant so that the number of minimum input signals in structural controllability can be largely reduced.

Application to T helper differentiation cellular network

T helper cells (Th cells) are a type of white blood cells that release T cell cytokines to stimulate the activity of other immune cells. Matured Th cells can be classified into one of the following three states: Th0 (precursor), Th1 and Th2 (effector) cells, according to the expressed surface proteins. To deeply understand the mechanism of differentiation, the helper differentiation cellular network containing 17 nodes and 27 edges has been constructed [78]. When applying the transittability to the T helper differential cellular network, Wu *et al.* [46] found three different types of steering strategies of transition, which are in agreement with existing results [79–81]. For example, the transition between Th0 and Th1 can be steered by nodes SOCS1 and T-bet, the transition between Th0 and Th2 can be steered by nodes IL-4 and GATA3 while the transition between Th1 and Th2 can be steered by T-bet and GATA2. These results indicate that the transittability is a feasible approach to study complex biological networks.

Application to epithelial–mesenchymal transition network

The invasion and metastasis of cancer cells is one of the critical hallmarks [82, 83]. Accumulating evidence shows that the progress of epithelial–mesenchymal transition (EMT) network, in which epithelial cells acquire the properties of mesenchymal cells, plays an important role in the initiation of the invasion of cancers beginning at the metastasis. Hence, the EMT networks have been constructed, which contains 6 nodes and 15 interactions [84]. To study the progress, Wu *et al.* [46] defined significantly differentially expressed nodes representing the phenotypes and found that only SNAI1 can steer the occurrence of transition, which has been verified in [84]. Moreover, except for SNAI1, any one of MIR203, MIR200, ZEB1 and ZEB2 can also steer the transitions between the two phenotypes. MIR203 and MIR200 have been verified to be able to steer the transitions [84, 85] while ZEB1 and ZEB2 are deserved the further investigations. These different selections provide the opportunities to investigate the mechanisms of the EMT from different aspects. Similarly, transittability can potentially be applied to others hallmarks of cancer, such as self-sufficiency in growth signals, evading programmed cell death and so on.

Optimal control for complex biological networks

The MDSs for structural controllability is usually not unique. Some preference approaches have been proposed to identify unique MDSs [86] through modifying the algorithm and adding more practical information, e.g. drug binding data [87]. To find the optimal MDSs in complex biological networks theoretically, control energy which should be minimized or limited to the acceptable scope has been investigated.

Control energy of complex networks

Mathematically, given an *LTI* system is controllable by a set of input signals, we define the energy required to control system to

a desired state as [88, 89]

$$E\left(t_f\right) = \int_{t_0}^{t_f} \|u(t)\|^2 dt \tag{10}$$

where t_0 and t_f are the initial time and the final time through the process of control, respectively. $u(t)$ is the input vector variable associated to time. We define the optimal problem to choose the optimal $u(t)$ as follows:

$$\min_{u(t)} J = \int_{t_0}^{t_f} \|u(t)\|^2 dt \tag{11}$$

subject to

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$x\left(t_0\right) = x_0; x\left(t_f\right) = x_{t_f}$$

The optimal problem can be solved by utilizing Potryagin's Maximum Principle. Specifically, the optimal input signals can be determined by $u(t) = B^T e^{A^T(t_f - t)} W^{-1}(x_{t_f} - e^{At_f} x_0)$, $t \in [t_0, t_f]$, where $W(t_f) \equiv \int_{t_0}^{t_f} e^{A\tau} BB^T e^{A^T\tau} d\tau$ is the positive-defined and symmetry Gramian matrix if system (A, B) is controllable. The minimum control energy can be determined by

$$E\left(t_f\right) = \int_{t_0}^{t_f} \|u(t)\|^2 dt = \left(x_{t_f} - e^{At_f} x_0\right)^T W^{-1} \left(x_{t_f} - e^{At_f} x_0\right) \tag{12}$$

Two useful special cases are as follows:

(i) $x_0 = 0$ (Reachability problem, which concerns the ability to steer the state of system from $x_0 = 0$ to any desired final state $x_{t_f} \neq 0$)

$$E\left(t_f\right) = x_{t_f}^T W_r^{-1} x_{t_f} \tag{13}$$

(ii) $x_{t_f} = 0$ (Null controllability problem, which concerns the ability to steer the state of system from any initial state $x_0 \neq 0$ to the desired final state $x_{t_f} = 0$)

$$E\left(t_f\right) = x_{t_0}^T W_c^{-1} x_{t_0} \tag{14}$$

where $W_r = W$ and $W_c = e^{-At_f} W e^{-A^T t_f}$ are the reachability Gramian matrix and null controllability Gramian matrix [88, 89], respectively.

Control metric

From (12), the minimum control energy is determined by the initial state x_0, final state x_{t_f} and the Gramian matrix W. Gramian matrix W is associated with system matrices A and B while initial state and final state are the inherent state of complex network. Hence, we can describe the control metrics associated with Gramian matrix [90].

$\lambda_{\min}(W)$: the smallest eigenvalue of the controllability Gramian is related to the worst-case of the maximum control energy, where the direction of state space is the hardest to control. Given $\|x_{t_f}\|^2 = 1$, we can normalize the control energy using the Rayleigh–Ritz theorem

$$\lambda_{\max}^{-1} = E_{\min} \leq E\left(t_f\right) \leq E_{\max} = \lambda_{\min}^{-1} \tag{15}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and the maximum eigenvalues of the Gramian matrix W, respectively. Especially, if x_{t_f} is the *ith* eigenvector of the eigenvalue λ_i, the minimum control energy is the λ_i^{-1}.

trace (W^{-1}): the trace of the inverse of the Gramian matrix is the measurement of the average control energy around on the state space. The average energy can be calculated as follows:

$$\frac{\int x^T W^{-1} x dx}{\int dx} = \frac{1}{n} trace(W^{-1}) \tag{16}$$

The trace of W^{-1} is ill-condition when the system is very large. Instead, we maximize the trace of W to minimize the average energy required to control system to any desired states.

det(W): the determinant of the controllability Gramian is proportional to the volume of ellipsoid in the state space consisting of all states reached by a unit-energy control input. The volume of ellipsoid can be calculated as follows:

$$V = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)} \sqrt[n]{det(W)} \tag{17}$$

where Γ is the Gamma function. Note that the ellipsoid volume is zero when the system is not fully controllable.

Applications to brain dynamic network

The human brain containing about 86 billion neurons [91] shows a very complex structure and various dynamics, some principles and mechanisms of which have been uncovered. Gu *et al.* [47] used the approaches from control theories and network analysis to provide deep understanding of how the state transitions of brains happen. A brain network is an undirected network and contains a number of nodes, each of which represents a specific region of a brain. Based on the linear control theory, the brain network is theoretical controllable because the smallest eigenvalue of Gramian matrix is consistently greater than zero, indicating that one node can control the whole system. However, the maximum energy increases as $E_{\max} \sim e^N$, indicating that controlling the system is energetically prohibitive in some directions of state spaces by one node [92]. On the other hands, we want to find the critical regions which facilitate the changes in brain state trajectories. Gu *et al.* [47] used the average energy matric *trace*(W) to identify those regions that can steer the system into different states with little energy cost. The results showed the default mode in human brain has large values, which is the densely connected areas in brain network. Moreover, the approach is robust to different scale brain networks and species.

Discussion and future work

This survey of control principles for complex biological networks has discussed structural controllability and analyzed the biological requirements for control theory. In this section, combining aforementioned control principles and applications in Table 1, we discuss more detail about the demands of effective control and optimal control in complex biological networks and challenges for general control principles when applying to complex biological networks.

Demands of effective control and optimal control

Structural controllability only needs to know the structure of complex networks and can provide the upper bound of the minimum number of driver nodes efficiently [18]. The mathematical form of classical control theory is not suitable for current complex networks, which may contain thousands of biological

nodes. Moreover, the weight of edges is hard to be precisely inferred from biological data. Under such practical conditions, structural controllability is good for complex biological networks when studying how to control cell fate, cure disease and so on.

However, the framework has two distinct disadvantages: (1) controlling biological networks needs too many driver nodes [63] and (2) the MDSs is usually not unique. Generally, the applications to complex biological networks are divided into one of three aspects: Liu' classification [18], Jia's classification [26] and control centrality [25], rather than identifying a set of driver nodes. Considering the aforementioned applications to identify disease genes and drug targets, Vinayagam *et al.* [32] don't take specific disease into consideration so that the results are only meaningful in the whole cell progress. Alternatively, Sharma *et al.* [42] only identified sets of critical targets of different phenotypes in osteosarcoma combining multi-type biological networks. Although these both studies identified the sets of critical targets based on the principles of network controllability, they did not focus on identifying the optimal sets of driver nodes for the transition between health states and disease states. Another application is to identify dysregulated pathway via the measurement of pathways [39]. Though the definition of control centrality takes the dynamics and directionality of signaling transduction into consideration, the measurement of control centrality lacks the ability to find a biologically meaningful set of driver nodes for biological networks. In practice, the meaningful set of driver nodes can enhance the designation of control strategy. Hence, these disadvantages should be carefully taken into consideration when one wants to propose a control strategy instead of identifying controllable related nodes for biological networks.

Therefore, we discuss effective control and optimal control to complement fully structural controllability. The key idea of effective control is to ignore the redundant nodes and states that aren't necessary to be controlled. For example, cancer is usually altered by a few driver mutations [93]. The cancer phenotypes are determined by part of differentially expressed genes [77]. The cancer state is just associated to a small part of whole state space [94]. Hence, effective control for complex biological networks can minimize the set of driver nodes to control. On the other hand, optimal control takes the cost of control strategy into consideration. At the first step, the steering nodes (drug targets) and corresponding drugs are selected for steering a cell from the cancer state to a healthy state. Then, whether these drugs can efficiently achieve the transitions should be considered in terms of the drug dosage, the transition time, the transition trajectories and so on. Usually, optimal control is to select a detailed control strategy through the balance between the practical constraints and the selection of driver nodes. In the view of current effective control and optimal control, we still face many challenges in controlling complex biological networks.

Effective control should also consider the side effect for controlling partial nodes or two-state transition. Output control and transittability study the partial nodes of networks representing special nodes or states. By actuating signals on the driver nodes, the task of partial control can be finished. However, when the effective nodes or states have been controlled, the side effect induced by the redundant signals or nodes may be observed. In medical practice, it's a common phenomenon that the drugs used to cure disease cause a side effect for allergic people. Generally, the side effect should be minimized so that the progress of control is suitable for most people. Hence, when developing a new algorithm, the controller placement should be sufficiently considered to minimize the side effect while making sure systems are controllable. For a specific person, partial control is constrained by the allergic drugs, indicating that some areas should be segregated from the external signals.

The objective function could take more constraints into consideration in full controllability. Though a set of driver nodes can be found based on the minimum energy control theory, there are more complex conditions for us to study. In real biological networks, the control energy has a bound distinguishing the feasible edges of control energy [95], indicating that the dose of drugs is feasible with a range. Moreover, given a set of drugs which has several targets, the form of control energy could be largely different. Hence, we can get more reasonable results by adding constraints corresponding with drug properties.

Both effective control and optimal control should compensate each other. The results of effective control are not unique as well. The algorithms to solve output control and two-state transition can only get approximate solutions that may have many combinations. The computation of optimal control is also time-consuming. The operation of a large matrix may consume a large amount of time to obtain a unique result. Recently, a cutting-edge approach is to optimize the output control of complex networks by solving Hamiltonian equation [96]. Studies indicated that the required energy for output control can be reduced substantially with the decrease of target nodes. On the other hand, the most importance in two-state transition is to design feasible trajectories with limited control energy [64, 97]. The results can provide fundamental insights into the mechanisms of state transitions between healthy state and disease state. Hence, the combination can be potentially applied to complex biological networks.

As outlined above, effective control and optimal control are important for complex biological networks. The development of effective control and optimal control can potentially reveal the mechanism of biological systems.

Challenges for general control principles

Recently, several control principles had been reported to control complex networks [98], but controlling complex biological networks is still hindered by network data. Generally, networks, such as power grid network, WWW and so on, can be diagramed more accurately than complex biological networks whose data are limited. In biological networks, we usually utilize directed biological networks to model the flow of signaling transduction. However, such a network only contains partial molecules which is far from complete. The typical example is the signaling pathway network. The low coverage of the genome in curated pathway databases including KEGG [99], Reactome [100] and Panther [101] enforces us to extend them to signaling pathway data with high coverage [102], roughly 50% of total SwissProt proteins. Similarly, we can also predict the direction of edges in human PPI network that shows high coverage [59]. On the basis of high coverage signaling pathway network, we can extract subnetworks possessing the properties including tissue specific, disease specific and state specific, which is crucial for us to use control principles to understand the mechanism of signaling transduction.

Though structural controllability is robust to the edges missing, controlling nonlinear dynamic in biological networks, which are constrained by precise data of biological networks, isn't robust. The advantage of nonlinear dynamics is able to deeply mine the dynamics properties of biological networks, while structural controllability guides us to find a set of driver nodes. Cornelius *et al.* [103] developed a physically admissible

compensatory perturbations on biological systems to steer the state to the attractor of desired state under the assumption that the cell state can evolve without drug stimulation. When applying it to T cell survival signaling networks, it can identify the drug targets to reprogram the state to its desired state, but it requires a complete data of the detailed model. Similarly, Wang *et al*. [104] proposed a control strategy of nonlinear dynamics by small parameter perturbation, which is constrained by the data of the detailed model as well.

Sometimes, these detailed models cannot be extracted from the biological data so that we can focus more on the structure of biological networks, by which controlling nonlinear dynamics may be more feasible. Mochizuki and Fiedler [105, 106] developed a structure-based control method through open-loop control applied to the feedback vertex set, which can completely steer the other nodes to the desired attractors by following desired trajectories. This method doesn't need external signals under the condition that the states of source nodes converge. Recently, Zanudo *et al*. [107] thought that the states of source nodes can affect the trajectories of attractors and brought up a new structure-based control for complex biological networks. Moreover, as the nonlinear dynamics could be properly approximated by several parts of linear dynamics in the finite time, each part can be considered as two-state transition problem. On the basis of transittability, the nonlinear dynamics can be controlled if a set of driver nodes to control all state transitions could be found. Hence, structure-based methods not only require a small amount of other data but also provide an efficient graph-theoretic algorithm.

On the other hand, the aforementioned control principles focus on controlling isolated complex networks, while realistic networks are coupled together [108, 109]. In biological networks, the PPI network changes with time forming a temporal network [110, 111]. The gene regulation and metabolic reaction interact with each other to form an interdependent network [112]. Obviously, control principles for such networks are more complicated than controlling isolated complex networks. However, some useful concepts in isolated networks can be extended to the networks interacting with each other.

In the area of temporal networks, Pósfai *et al*. [113] studied structural controllability of temporal networks to understand complex systems and found that the controllability is affected by the overall activity and the degree distribution. Moreover, Li *et al*. [114] found the fundamental advantages of temporal networks in controllability, including faster control, less control energy and distinct control trajectories. The advantages are attributed to the flexibility of edges in temporal networks that can enhance our power to control them. However, identifying a feasible set of driver nodes is still a challenge because the minimum set of feasible driver nodes and the combination of snapshots interact with each other. Actually, the snapshots of a temporal network are connected by time series, indicating that the orders of snapshots determine the direction of signals from a feasible set of control signals. Hence, how to balance the two objectives is a great challenge. Moreover, controlling specific targets in temporal biological networks needs to assign the orders of snapshots so that the desired stated can be achieved within the minimum time.

In the area of multilayer networks, the study of control focuses on two-layer networks including random duplex networks [115] and scale-free duplex network [116]. Pósfai *et al*. [117] studied the controllability of multilayer networks, each layer of which can operate at a different time scale, and utilized structural controllability to determine the minimum number of driver nodes via graph-theoretic methods so that the multilayer network is fully controlled. More important, it is found that controllability is enhanced when the faster layer is controlled. Observably, multilayer networks could be meaningful in complex biological networks with the accumulation of omics data. Each layer in multilayer networks represents a specific molecular level whose dynamics can be influenced by the inherent interactions, external interactions and external signals. Controlling multilayer biological networks should first define the signal layer and the phenotype layer. For example, genome editing can be used to regulate genes and change gene transcription levels. Furthermore, the more the layers are, the more meaningful controlling multilayer biological networks is. However, the development of control strategy is still a challenge with many layers because the coupled relationships are so complicated.

In summary, general control principles from traditional control theory can be potentially applied to complex biological networks if the data of biological networks are available. Moreover, the development of new control principles for complex biological networks should be on the basis of biological meaning.

Key points

- The advantages and disadvantages of structural controllability are discussed in Section *Controllability of complex networks*.
- Effective control methods are summarized in Section *Effective control for complex biological networks* to meet the requirement of complex biological networks.
- Optimal control based on control energy is introduced in Section *Optimal control for complex biological networks* to find a unique and useful set of driver nodes.

Funding

This work was supported in part by the National Natural Science Foundation of China [61622213, 61728211, 61772552].

References

1. Newman M. *Networks: An Introduction*. Oxford, United Kingdom: Oxford University Press, 2010.
2. Barabási A-L, Pósfai M. *Network Science*. Cambridge, United Kingdom: Cambridge University Press, 2016.
3. Watts DJ, Strogatz SH. Collective dynamics of small-world networks. *Nature* 1998;**393**:440.
4. Barabási A-L, Albert R. Emergence of scaling in random networks. *Science* 1999;**286**:509–12.
5. Boccaletti S, Latora V, Moreno Y, *et al*. Complex networks: structure and dynamics. *Phys Rep* 2006;**424**:175–308.
6. Lü L, Chen D, Ren X-L, *et al*. Vital nodes identification in complex networks. *Phys Rep* 2016;**650**:1–63.
7. Liao H, Mariani MS, Medo M, *et al*. Ranking in evolving complex networks. *Phys Rep* 2017;**689**:1–54.
8. Fortunato S, Hric D. Community detection in networks: a user guide. *Phys Rep* 2016;**659**:1–44.
9. Fortunato S. Community detection in graphs. *Phys Rep* 2010;**486**:75–174.
10. Jeong H, Mason SP, Barabási A-L, *et al*. Lethality and centrality in protein networks. *Nature* 2001;**411**:41.

11. Norton HK, Emerson DJ, Huang H, *et al*. Detecting hierarchical genome folding with network modularity. *Nat Methods* 2018;**15**:119.
12. Wang W-X, Lai Y-C, Grebogi C. Data based identification and prediction of nonlinear and complex dynamical systems. *Phys Rep* 2016;**644**:1–76.
13. Diekmann O, Heesterbeek JAP. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Hoboken, United States: John Wiley & Sons, 2000.
14. Wu M, Su R-Q, Li X, *et al*. Engineering of regulated stochastic cell fate determination. *Proc Natl Acad Sci U S A* 2013;**110**:10610–5.
15. Abel JH, Chakrabarty A, Doyle FJ. *Controlling Biological Time: Nonlinear Model Predictive Control for Populations of Circadian Oscillators*. New York, United States: Springer, 2018.
16. Kalman RE. Mathematical description of linear dynamical systems. *SIAM J Appl Math* 1963;**1**:152–92.
17. Hautus M. Controllability and observability conditions of linear autonomous systems. *Indag Math* 1969;**72**:443–8.
18. Liu Y-Y, Slotine J-J, Barabási A-L. Controllability of complex networks. *Nature* 2011;**473**:167.
19. Lin C-T. Structural controllability. *IEEE Trans Automat Contr* 1974;**19**:201–8.
20. Nie S, Wang X, Zhang H, *et al*. Robustness of controllability for networks based on edge-attack. *PLoS ONE* 2014;**9**: e89066.
21. Pu C-L, Pei W-J, Michaelson A. Robustness analysis of network controllability. *Physica A* 2012;**391**:4420–5.
22. Wang W-X, Ni X, Lai Y-C, *et al*. Optimizing controllability of complex networks by minimum structural perturbations. *Phys Rev E* 2012;**85**:026115.
23. Pósfai M, Liu Y-Y, Slotine J-J, *et al*. Effect of correlations on network controllability. *Sci Rep* 2013;**3**:1067.
24. Wang B, Gao L, Gao Y. Control range: a controllability-based index for node significance in directed networks. *J Stat Mech* 2012;**2012**:P04011.
25. Liu Y-Y, Slotine J-J, Barabási A-L. Control centrality and hierarchical structure in complex networks. *PLoS ONE* 2012;**7**:e44459.
26. Jia T, Liu Y-Y, Csóka E, *et al*. Emergence of bimodality in controlling complex networks. *Nat Commun* 2013;**4**:2002.
27. Ruths J, Ruths D. Control profiles of complex networks. *Science* 2014;**343**:1373–6.
28. Jia T, Barabási A-L. Control capacity and a random sampling method in exploring controllability of complex networks. *Sci Rep* 2013;**3**:2354.
29. Menichetti G, Dall'asta L, Bianconi G. Network controllability is determined by the density of low in-degree and out-degree nodes. *Phys Rev Lett* 2014;**113**:078701.
30. Nepusz T, Vicsek T. Controlling edge dynamics in complex networks. *Nat Phys* 2012;**8**:568.
31. Yuan Z, Zhao C, Di Z, *et al*. Exact controllability of complex networks. *Nat Commun* 2013;**4**:2447.
32. Vinayagam A, Gibson TE, Lee H-J, *et al*. Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets. *Proc Natl Acad Sci U S A* 2016;**113**:4976–81.
33. Ravindran V, Nacher JC, Akutsu T, *et al*. Network controllability: viruses are driver agents in dynamic molecular systems. bioRxiv 2018:311746.
34. Kawakami E, Singh VK, Matsubara K, *et al*. Network analyses based on comprehensive molecular interaction maps reveal robust control structures in yeast stress response pathways. *NPJ Syst Biol Appl* 2016;**2**:15018.
35. Liu X, Pan L. Detection of driver metabolites in the human liver metabolic network using structural controllability analysis. *BMC Syst Biol* 2014;**8**:51.
36. Liu X, Pan L. Identifying driver nodes in the human signaling network using structural controllability analysis. *IEEE/ACM Trans Comput Biol Bioinform* 2015;**12**:467–72.
37. Ravindran V, Sunitha V, Bagler G. Identification of critical regulatory genes in cancer signaling network using controllability analysis. *Physica A* 2017;**474**:134–43.
38. Wang P, Wang D, Lu J. Controllability analysis of a gene network for Arabidopsis thaliana reveals characteristics of functional gene families. *IEEE/ACM Trans Comput Biol Bioinform* 2018.
39. Sharma A, Halu A, Decano JL, *et al*. Controllability in an islet specific regulatory network identifies the transcriptional factor NFATC4, which regulates Type 2 Diabetes associated genes. *NPJ Syst Biol Appl* 2018;**4**:25.
40. Wu L, Min L, Wang J, *et al*. CytoCtrlAnalyser: a Cytoscape app for biomolecular network controllability analysis. *Bioinformatics* 2017;**1**:3.
41. Sharma A, Halu A, Decano JL, *et al*. Controllability in an islet specific regulatory network identifies the transcriptional factor NFATC4, which regulates Type 2 Diabetes associated genes. bioRxiv 2017:226456.
42. Sharma A, Cinti C, Capobianco E. Multitype network-guided target controllability in phenotypically characterized osteosarcoma: role of tumor microenvironment. *Front Immunol* 2017;**8**:918.
43. Yan G, Vértes PE, Towlson EK, *et al*. Network control principles predict neuron function in the Caenorhabditis elegans connectome. *Nature* 2017;**550**:519.
44. Kanhaiya K, Czeizler E, Gratie C, *et al*. Controlling directed protein interaction networks in cancer. *Sci Rep* 2017;**7**: 10327.
45. Guo W-F, Zhang S-W, Liu L-L, *et al*. Discovering personalized driver mutation profiles of single samples in cancer by network control strategy. *Bioinformatics* 2018;**34**:1893–1903.
46. Wu F-X, Wu L, Wang J, *et al*. Transittability of complex networks and its applications to regulatory biomolecular networks. *Sci Rep* 2014;**4**:4819.
47. Gu S, Pasqualetti F, Cieslak M, *et al*. Controllability of structural brain networks. *Nat Commun* 2015;**6**:8414.
48. Medaglia JD, Gu S, Pasqualetti F, *et al*. Cognitive control in the controllable connectome. arXiv preprint arXiv 2016:1606.09185.
49. Medaglia JD, Pasqualetti F, Hamilton RH, *et al*. Brain and cognitive reserve: translation via network control theory. *Neurosc Biobehav Rev* 2017;**75**:53–64.
50. Tang E, Giusti C, Baum GL, *et al*. Developmental increases in white matter network controllability support a growing diversity of brain dynamics. *Nat Commun* 2017;**8**:1252.
51. Kenett YN, Medaglia JD, Beaty RE, *et al*. Driving the brain towards creativity and intelligence: a network control theory analysis. *Neuropsychologia* 2018.
52. Kailath T. *Linear Systems*. Englewood Cliffs, NJ: Prentice Hall, 1980.
53. Raibert M, Blankespoor K, Nelson G, *et al*. Bigdog, the rough-terrain quadruped robot. *IFAC Proceedings Volumes* 2008;**41**:10822–5.
54. Semini C, Tsagarakis NG, Guglielmino E, *et al*. Design of HyQ–a hydraulically and electrically actuated quadruped robot. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering* 2011;**225**: 831–49.

55. Engell S. Feedback control for optimal process operation. *J Process Control* 2007;**17**:203–19.
56. Chai T, Qin SJ, Wang H. Optimal operational control for complex industrial processes. *Annu Rev Control* 2014;**38**: 81–92.
57. Hopcroft JE, Karp RM. An n^5/2 algorithm for maximum matchings in bipartite graphs. *SIAM J Comput* 1973;**2**: 225–31.
58. Wu L, Li M, Wang J, *et al*. Minimum steering node set of complex networks and its applications to biomolecular networks. *IET Syst Biol* 2016;**10**:116–23.
59. Vinayagam A, Stelzl U, Foulle R, *et al*. A directed protein interaction network for investigating intracellular signal transduction. *Sci Signal* 2011;**4**:rs8–8.
60. Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature* 2009;**461**:218.
61. Kim Y-A, Wuchty S, Przytycka TM. Identifying causal genes and dysregulated pathways in complex diseases. *PLoS Comput Biol* 2011;**7**:e1001095.
62. Poljak S. On the generic dimension of controllable subspaces. *IEEE Trans Automat Contr* 1990;**35**:367–9.
63. Müller F-J, Schuppert A. Few inputs can reprogram biological networks. *Nature* 2011;**478**:E4.
64. Sun J, Motter AE. Controllability transition and nonlocality in network control. *Phys Rev Lett* 2013;**110**:208701.
65. Hosoe S. Determination of generic dimensions of controllable subspaces and its application. *IEEE Trans Automat Contr* 1980;**25**:1192–6.
66. Murota K, Poljak S. Note on a graph-theoretic criterion for structural output controllability. *IEEE Trans Automat Contr* 1990;**35**:939–42.
67. Wu L, Shen Y, Li M, *et al*. Network output controllability-based method for drug target identification. *IEEE Trans Nanobioscience* 2015;**14**:184–91.
68. Munkres J. Algorithms for the assignment and transportation problems. *SIAM J Appl Math* 1957;**5**:32–8.
69. Kuhn HW. The Hungarian method for the assignment problem. *Nav Res Logist* 1955;**2**:83–97.
70. Gao J, Liu Y-Y, D'souza RM, *et al*. Target control of complex networks. *Nat Commun* 2014;**5**:5415.
71. Liu X, Pan L. Controllability of the better chosen partial networks. *Physica A* 2016;**456**:120–7.
72. Liu X, Pan L, Stanley HE, *et al*. Controllability of giant connected components in a directed network. *Phys Rev E* 2017;**95**:042318.
73. White JG, Southgate E, Thomson JN, *et al*. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci* 1986;**314**:1–340.
74. Chew YL, Walker DS, Towlson EK, *et al*. Recordings of Caenorhabditis elegans locomotor behaviour following targeted ablation of single motorneurons. *Sci Data* 2017;**4**:170156.
75. Guo W-F, Zhang S-W, Wei Z-G, *et al*. Constrained target controllability of complex networks. *J Stat Mech Theory Exp* 2017;**2017**:063402.
76. Hou JP, Ma J. DawnRank: discovering personalized driver genes in cancer. *Genome Med* 2014;**6**:56.
77. Bertrand D, Chng KR, Sherbaf FG, *et al*. Patient-specific driver gene prediction and risk assessment through integrated network analysis of cancer omics profiles. *Nucleic Acids Res* 2015;**43**:e44–4.
78. Mendoza L. A network model for the control of the differentiation process in Th cells. *Biosystems* 2006;**84**: 101–14.
79. Lee HJ, Takemoto N, Kurata H, *et al*. GATA-3 induces T helper cell type 2 (Th2) cytokine expression and chromatin remodeling in committed Th1 cells. *J Exp Med* 2000;**192**: 105–16.
80. Hwang ES, Szabo SJ, Schwartzberg PL, *et al*. T helper cell fate specified by kinase-mediated interaction of T-bet with GATA-3. *Science* 2005;**307**:430–3.
81. Szabo SJ, Kim ST, Costa GL, *et al*. A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell* 2000;**100**:655–69.
82. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;**100**:57–70.
83. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;**144**:646–74.
84. Moes M, Le Béchec A, Crespo I, *et al*. A novel network integrating a miRNA-203/SNAI1 feedback loop which regulates epithelial to mesenchymal transition. *PLoS ONE* 2012;**7**:e35440.
85. Park S-M, Gaur AB, Lengyel E, *et al*. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* 2008;**22**:894–907.
86. Wu L, Tang L, Li M, *et al*. The MSS of complex networks with centrality based preference and its application to biomolecular networks. In Bioinformatics and Biomedicine (BIBM), 2016 IEEE International Conference *on* 2016:229–34.
87. Wu L, Tang L, Li M, *et al*. Biomolecular network controllability with drug binding information. *IEEE Trans Nanobioscience* 2017;**16**:326–32.
88. Rugh WJ and Rugh WJ. Linear System Theory. Upper Saddle River, NJ:Prentice Hall, 1996.
89. Yan G, Ren J, Lai Y-C, *et al*. Controlling complex networks: How much energy is needed? *Phys Rev Lett* 2012; **108**:218703.
90. Pasqualetti F, Zampieri S, Bullo F. Controllability metrics, limitations and algorithms for complex networks. *IEEE Trans Cont Network Syst* 2014;**1**:40–52.
91. Bartheld CS, Bahney J, Herculano-Houzel S. The search for true numbers of neurons and glial cells in the human brain: a review of 150 years of cell counting. *J Comp Neurol* 2016;**524**:3865–95.
92. Yan G, Tsekenis G, Barzel B, *et al*. Spectrum of controlling and observing complex networks. *Nat Phys* 2015;**11**:779.
93. Haber DA, Settleman J. Cancer: drivers and passengers. *Nature* 2007;**446**:145.
94. Gupta PB, Fillmore CM, Jiang G, *et al*. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 2011;**146**:633–44.
95. Tzoumas V, Rahimian MA, Pappas GJ, *et al*. Minimal actuator placement with bounds on control effort. *IEEE Trans Cont Network Syst* 2016;**3**:67–78.
96. Klickstein I, Shirin A, Sorrentino F. Energy scaling of targeted optimal control of complex networks. *Nat Comm* 2017;**8**:15145.
97. Gu S, Betzel RF, Mattar MG, *et al*. Optimal trajectories of brain state transitions. *Neuroimage* 2017;**148**:305–17.
98. Liu Y-Y, Barabási A-L. Control principles of complex systems. *Rev Mod Phys* 2016;**88**:035006.
99. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28**:27–30.
100. Joshi-Tope G, Gillespie M, Vastrik I, *et al*. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* 2005;**33**:D428–32.

101. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* 2012;**41**:D377–86.
102. Wu G, Feng X, Stein L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol* 2010;**11**:R53.
103. Cornelius SP, Kath WL, Motter AE. Realistic control of network dynamics. *Nature Commun* 2013;**4**:1942.
104. Wang L-Z, Su R-Q, Huang Z-G, *et al*. A geometrical approach to control and controllability of nonlinear dynamical networks. *Nature Commun* 2016;**7**:11323.
105. Mochizuki A, Fiedler B, Kurosawa G, *et al*. Dynamics and control at feedback vertex sets. II: a faithful monitor to determine the diversity of molecular activities in regulatory networks. *J Theor Biol* 2013;**335**:130–46.
106. Fiedler B, Mochizuki A, Kurosawa G, *et al*. Dynamics and control at feedback vertex sets. I: informative and determining nodes in regulatory networks. *J Dyn Differ Equ* 2013;**25**:563–604.
107. Zañudo JGT, Yang G, Albert R. Structure-based control of complex networks with nonlinear dynamics. *Proc Natl Acad Sci U S A* 2017;**114**:7234–9.
108. Buldyrev SV, Parshani R, Paul G, *et al*. Catastrophic cascade of failures in interdependent networks. *Nature* 2010;**464**:1025.
109. Liu X, Stanley HE, Gao J. Breakdown of interdependent directed networks. *Proc Natl Acad Sci U S A* 2016;**113**:1138–43.
110. Wang J, Peng X, Li M, *et al*. Construction and application of dynamic protein interaction network based on time course gene expression data. *Proteomics* 2013;**13**:301–12.
111. Li M, Yang J, Wu F-X, *et al*. DyNetViewer: a Cytoscape app for dynamic network construction, analysis and visualization. *Bioinformatics* 2017;**34**:1597–1599.
112. Klosik DF, Grimbs A, Bornholdt S, *et al*. The interdependent network of gene regulation and metabolism is robust where it needs to be. *Nat Commun* 2017;**8**:534.
113. Pósfai M, Hövel P. Structural controllability of temporal networks. *New J Phy* 2014;**16**:123055.
114. Li A, Cornelius SP, Liu Y-Y, *et al*. The fundamental advantages of temporal networks. *Science* 2017;**358**:1042–6.
115. Menichetti G, Dall'asta L, Bianconi G. Control of multilayer networks. *Sci Rep* 2016;**6**:20706.
116. Zhang Y, Garas A, Schweitzer F. Value of peripheral nodes in controlling multilayer scale-free networks. *Phys Rev E* 2016;**93**:012309.
117. Pósfai M, Gao J, Cornelius SP, *et al*. Controllability of multiplex, multi-time-scale networks. *Phys Rev E* 2016;**94**:032316.