ARTICLE

OPEN

Target control of complex networks

Jianxi Gao[1], Yang-Yu Liu[1,2,3], Raissa M. D'Souza[4,5] & Albert-László Barabási[1,2,3]

Controlling large natural and technological networks is an outstanding challenge. It is typically neither feasible nor necessary to control the entire network, prompting us to explore target control: the efficient control of a preselected subset of nodes. We show that the structural controllability approach used for full control overestimates the minimum number of driver nodes needed for target control. Here we develop an alternate 'k-walk' theory for directed tree networks, and we rigorously prove that one node can control a set of target nodes if the path length to each target node is unique. For more general cases, we develop a greedy algorithm to approximate the minimum set of driver nodes sufficient for target control. We find that degree heterogeneous networks are target controllable with higher efficiency than homogeneous networks and that the structure of many real-world networks are suitable for efficient target control.

[1] Center for Complex Network Research and Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA. [2] Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [3] Department of Medicine and Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [4] Complexity Sciences Center, Department of Mechanical and Aerospace Engineering, and Department of Computer Science, University of California, Davis, California 95616, USA. [5] Santa Fe Institute, Santa Fe, New Mexico 87501, USA. Correspondence and requests for materials should be addressed to A.-L.B. (email: alb@neu.edu).

1

In the past decade network science offered deep insights into the structure and dynamics of complex networked systems[1–7]. Yet, we continue to lack tools to efficiently control the dynamics of complex networks. According to control theory[8–10], a dynamic system is controllable if suitable external inputs can move the internal state of the system from any initial state to an arbitrary accessible final state in a finite time interval[11]. Recently, we introduced an analytical framework to study the controllability of complex networks via a combination of tools from network science, control theory and statistical physics[12]. By mapping the structural controllability problem[13] to a maximum matching problem[14], we were able to identify a minimum set of driver nodes of size N_D, sufficient to control the whole network[12,15]. For engineered systems, such as the auto-pilot system of an airplane, full control is essential. However, many biological, technological and social systems are massive in size and complexity, hence it is neither feasible nor necessary to control the full network. Instead, it is more realistic and sufficient to achieve target control, that is, to control a subset of target nodes (or a subsystem) that are essential for the system's mission pertaining to a selected task. While potential applications of target control have been developed in domain-specific areas in biology[16], chemical engineering[17], epidemics[18] and economic networks[19], a principled approach to identify a minimum set of driver nodes sufficient for the targeted control of an arbitrary complex network remains an open problem.

We study two distinct schemes for choosing the subset of nodes we wish to control. In the random scheme, a fraction f of nodes are chosen uniformly at random. In the local scheme, the chosen nodes form a connected component, capturing a well-defined local network neighbourhood. To develop an efficiency measure of target control, we compare P_D, which is the minimum number of driver nodes needed to control the fraction f of target nodes, with fN_D, which is the corresponding relative fraction of the driver nodes needed for full control. If P_D is less than fN_D then target control is more efficient than full control for a given f. We also establish the overall target control efficiency by considering the integrated efficiency across the entire range of $0 < f \leq 1$. For both scale-free (SF)[20] and Erdös–Rényi (ER) networks[21,22], we find that in general, local target control is more efficient than random target control. More surprisingly, we find that degree heterogeneous networks, such as SF networks, have higher specific and overall target control efficiency than degree homogeneous networks, for both random and local schemes, when the average degree of the network is large. In contrast, for full control, degree homogeneous networks, such as ER networks, are more efficient than the corresponding degree heterogeneous networks[12]. Finally, we apply our methods to real data, confirming that many real-world networks display high efficiency of target control.

Results

Model. The dynamics of most real systems is driven by nonlinear processes. However, the dynamical rules of real-world networks are so diverse that writing a general dynamical equation that captures them all is plainly impossible. Moreover, for many networks, especially biological networks, we do not even know the dynamical rules. Thus, before we explore the fully nonlinear dynamical problem, we have to understand the impact of the topological characteristics on linear control, which naturally serves as a prerequisite of the nonlinear controllability problem. Furthermore, the controllability of nonlinear systems is often structurally similar to and determined by the system's linearized dynamics[23]. Indeed, a basic starting point for exploring the controllability of any nonlinear system is the study of the

linearized version of the nonlinear dynamical system. Therefore, we start with the canonical linear time-invariant dynamics.

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

where $x \in \mathbb{R}^N$, $u \in \mathbb{R}^M$ and $y \in \mathbb{R}^S$ represent the system's state, input and output vector, respectively. $A \in \mathbb{R}^{N \times N}$, $B \in \mathbb{R}^{N \times M}$ and $C \in \mathbb{R}^{S \times N}$ denote the state, input and output matrices, respectively. A captures the wiring diagram of the system; B identifies the nodes that are controlled by an external controller and u is the time-dependent input applied to the nodes in B; C is the output matrix identifying the target nodes we want to control. For a network with node set $\mathcal{N} = \{1, 2, ..., N\}$, we are interested in controlling a target node set $\mathcal{C} = \{c_1, c_2, ..., c_S\}$ of size $S = |\mathcal{C}| = fN$. We set the output matrix $C = [I(c_1), I(c_2),... I(c_S)]$, where $I(i)$ denotes the ith row of an $N \times N$ identity matrix I. The system (A, B, C) is said to be target controllable with respect to a given target node set C if there exists a time-dependent input vector $u(t) = (u_1(t),... u_M(t))^T$ that can drive the state of the target nodes to any desired final state in finite time. Target controllability can be viewed as a special type of output controllability and the system (A, B, C) is target controllable if and only if the dimension of the output controllable subspace $d(A, B, C)$ satisfies

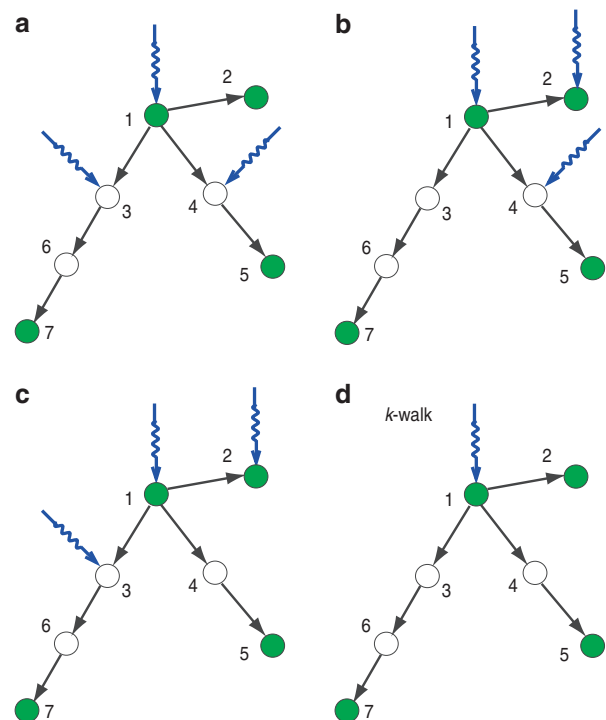$$d(A, B, C) \equiv \text{rank}[CB, CAB, CA^2B,..., CA^{N-1}B] = S, \quad (2)$$

Figure 1 | Structural control theory versus the k-walk theory for target control. To control the whole network, the structural control theory predicts that we need at least three driver nodes. The possible choices for the sets of driver nodes are shown in **a–c**. If instead we want to control a subset of nodes, for example, {1, 2, 5, 7} (green nodes) with a minimum set of nodes, we need to solve the target control problem. The upper bound obtained by structural control theory indicates that we need at least three driver nodes (the same sets are shown in **a,b** and **c**). (**d**) But, in reality, we only need one driver node (node 1), which can be obtained from both the k-walk theory and the greedy algorithm, because the length of the path from the node 1 to each target node {1, 2, 5, 7} is unique.
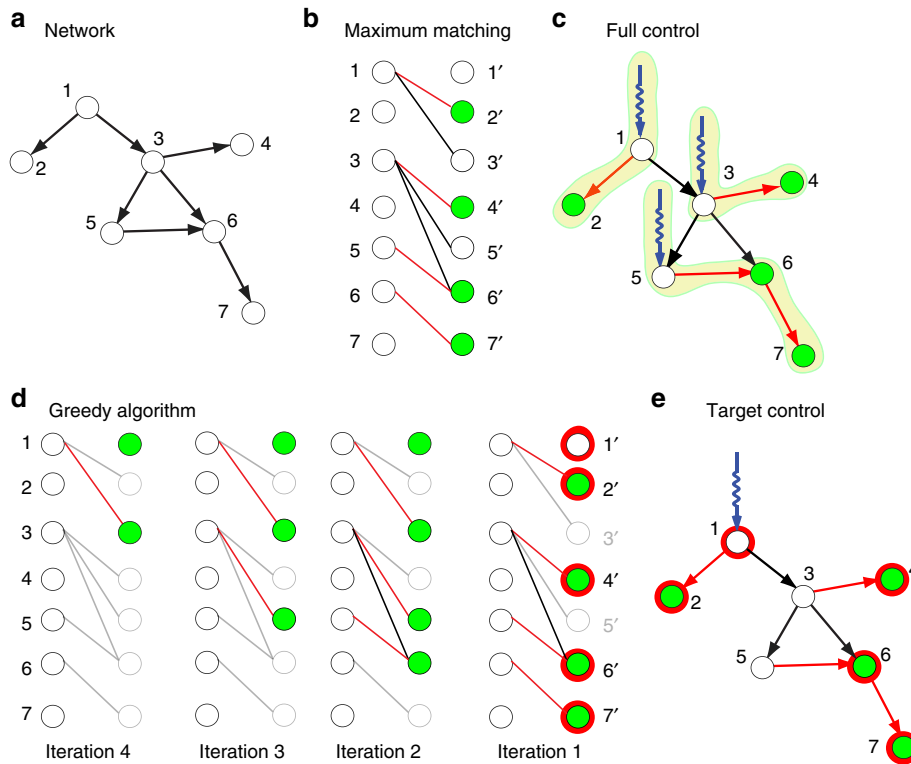
Figure 2 | Controlling a small network. (**a**) A directed network with seven nodes. (**b**) Maximum matching of the network in its bipartite representation. Matching edges are shown in red, matched nodes $\{x_2, x_4, x_6, x_7\}$ are green and unmatched nodes $\{x_1, x_3, x_5\}$ are white. (**c**) By controlling the three unmatched nodes (driver nodes) and ensuring that there are no inaccessible nodes, the system is guaranteed to be structurally controllable. The underlying control skeleton (or the cactus structure) is highlighted. (**d**) Greedy algorithm developed for target control. Here the target set is $\{x_1, x_2, x_4, x_6, x_7\}$ (highlighted in red). In the first iteration, we match all target nodes by solving a maximum matching problem on an induced bipartite graph. Target nodes $\{x_2, x_4, x_6, x_7\}$ (in green) are matched by $\{x_1, x_3, x_5, x_6\}$, which are the new target nodes considered in iteration 2. After four iterations, we obtain that node x_1 is the driver node for the target set $\{x_1, x_2, x_4, x_6, x_7\}$. (**e**) By controlling the unmatched node x_1 calculated from the GA, the target set $\{x_1, x_2, x_4, x_6, x_7\}$ is guaranteed to be controllable. The red lines are the matched links obtained in the first iteration of greedy algorithm. This can be proved by interpreting GA as a backtracking process on the dynamic graph representation of the system (Supplementary Note 3). The upper bound of P_D can be calculated by the minimal number of disjoint cacti that cover all target nodes, that is, $P_D^{UB} = 3$ for the target nodes $\{1, 2, 4, 6, 7\}$. The lower bound of P_D is just given by the first iteration of the greedy algorithm, that is, $P_D^{LB} = 1$ for the target nodes $\{1, 2, 4, 6, 7\}$, because in the first iteration only node 1 is unmatched.

representing the mathematical condition for target controllability[24]. Note that when $S = N$, C is the identity matrix and (2) reduces to the Kalman condition for full controllability[11]. If (2) is satisfied, for initial condition $x(0) = 0$, we can compute the optimal input vector $u(t)$ such that we can reach the desired final state $\bar{y}$ with minimum energy cost in time $\bar{t} > 0$ (that is, $y(\bar{t}) \equiv \bar{y}$) (ref. 24).

$$u(t) = B^T e^{A^T(\bar{t}-t)} C^T \left(CX(t)C^T \right)^{-1} \bar{y}, \quad (3)$$

where $X(t) = \int_0^t e^{A\delta} BB^T e^{A^T\delta} \mathrm{d}\delta$ (Supplementary Note 1).

In some ways the target control problem is more difficult than the full control problem. Full control has a graphical condition, which can be easily checked by exactly mapping the controllability problem to the maximum matching problem. Target control lacks such an exact mapping. Therefore, to solve the target control problem in the single-input case, we develop a new approach that we call the 'k-walk theory' (Fig. 1; Supplementary Fig. 1; Supplementary Note 2). This theory is based on the principle that a node can control a target set of nodes provided that the length of the path from the control node to each target node in the set is unique. Using the k-walk theory, we can identify all sets of nodes that can be controlled by the given node. We show that the k-walk theory is more powerful than the standard approach as it can find controllable subsystems that the

structural control theory misses (Fig. 1; Supplementary Note 2). We rigorously prove that the k-walk theory is correct for directed tree-like networks, that is, networks with no loops (Supplementary Note 2).

Although powerful, the k-walk theory is only applicable for the single-input case. For networks that require >1 control input, we formulate the target control problem in graph theoretic terms, allowing us to develop a greedy algorithm (GA) that offers a good approximation to the minimum set of inputs sufficient for target control (Fig. 2). We rigorously prove that the input set selected by our algorithm can indeed control all target nodes (Fig. 3; Supplementary Note 3). Both the k-walk theory and the GA are based on the structural control theory, that is, the system parameters are either fixed at zero or are independent free parameters. This approach has lead to several recent advances on network control[19,25–34]. Structural control theory, in general, relies on the canonical linear time-invariant dynamics of the system as discussed in the Methods.

Previous studies based on the structural control theory shed some light on how to determine the upper bound of the minimum number of driver nodes for target controllability[35,36]. We can use the maximum matching method to find the minimum set of driver nodes (N_D nodes) for full control of the network (Fig. 2a–c). This results in N_D disjoint 'cacti' (Supplementary Note 3), each denoting the control region of a
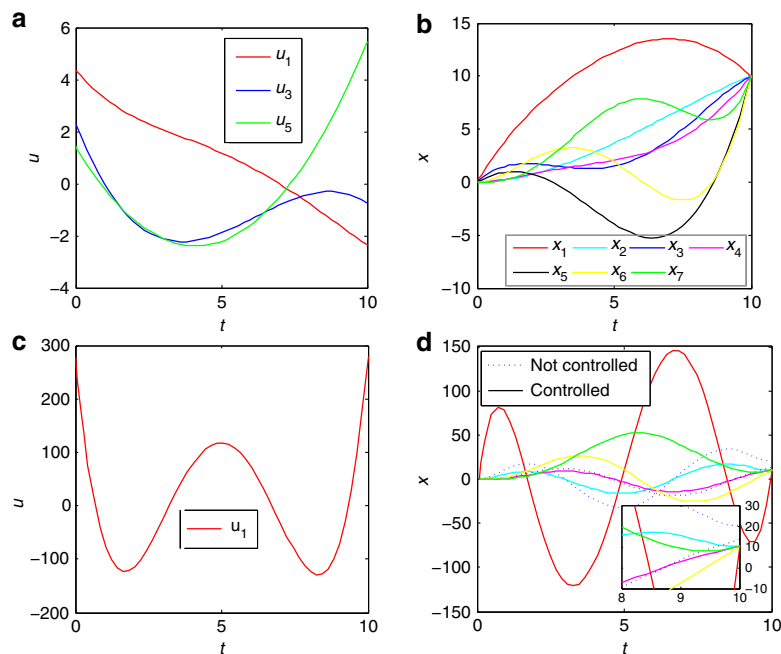
Figure 3 | Full versus target control of the simple network shown in Fig. 1. (**a**,**b**) Full control. (**a**) The time-dependent input signals $u_1(t)$, $u_3(t)$ and $u_5(t)$ predicted by equation (3) that can drive the state of the network from the initial state ($x_1 = \ldots = x_7 = 0$) to the final state ($x_1 = \ldots = x_7 = 10$). (**b**) The state trajectories of all nodes from initial state to the desired final state. (**c**,**d**) Target control. (**c**) The input required for the driver node 1 to control the state of the target nodes {1, 2, 4, 6, 7} from the initial state ($x_1 = \ldots = x_7 = 0$) to the final state ($x_1 = x_2 = x_4 = x_6 = x_7 = 10$). (**d**) The state trajectories of all nodes for $t \in [0, 10]$, where the target nodes go from 0 to the desired state 10 (solid curves) and the uncontrolled nodes go from 0 to other values (dotted curves) when $t = 10$. The inset in **d** shows the detail of the final state of all seven nodes, which clearly shows that all the target nodes are controlled to be 10 and other uncontrolled nodes are not 10 and with finite values.
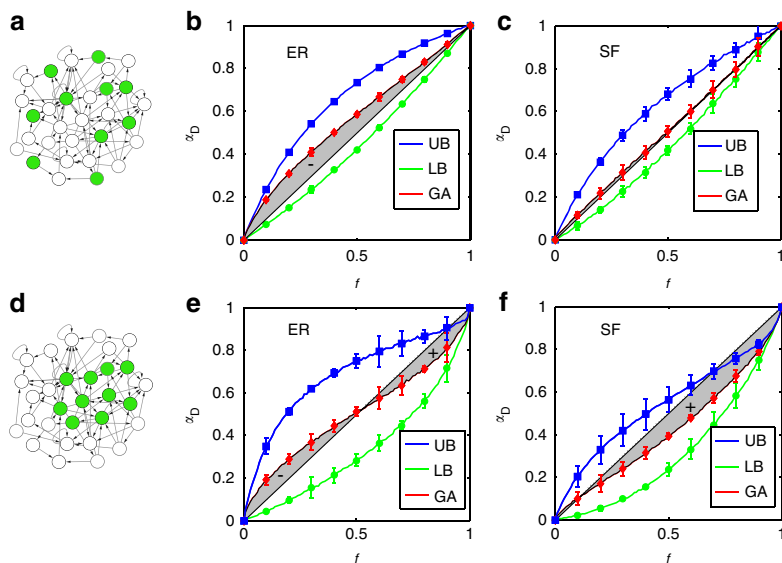

Figure 4 | Random versus local target controllability of two canonical network models. (**a**–**c**) Random selection scheme. (**a**) Illustrating random selection of $f = 1/3$ target nodes. (**b**) For ER networks with average degree $\langle k \rangle = 5.6$, we show the normalized fraction of driver nodes (α_D) in function of the target node fraction f for the random selection scheme. (**c**) For SF networks with $\langle k \rangle = 5.6$ and exponent $\gamma = 2.4$, we show the normalized fraction of driver nodes (α_D) as a function of the target node fraction f for the random scheme. (**d**–**f**) Local selection scheme. (**d**) The selection of an $f = 1/3$ target node with the local scheme. (**e**) For ER networks with the same average degree as in **b**, we show the normalized fraction of driver nodes (α_D) in function of the target node fraction f for the local scheme. (**f**) For SF networks with the same $\langle k \rangle$ and γ as in **c**, we show the normalized fraction of driver nodes (α_D) as a function of the target node fraction f for the local scheme. The black line describes the neutral case, that is, $\alpha_D = f$ (or $P_D = f N_D$) when the target control has the same efficiency as the full control. In each figure, we denote UB as the upper bound, GA as the greedy algorithm and LB as the lower bound of the minimum number of driver nodes to control an f fraction of target nodes. For ER networks, random target control is always less efficient than the full control, and for local target control efficiency depends on f. For SF networks random target control is as efficient as full control, but significant gains in efficiency are observed for local target control. Furthermore, for ER and SF networks local target control is more efficient than random target control. All results are averaged over 200 independent realizations of the networks with 10^4 nodes.
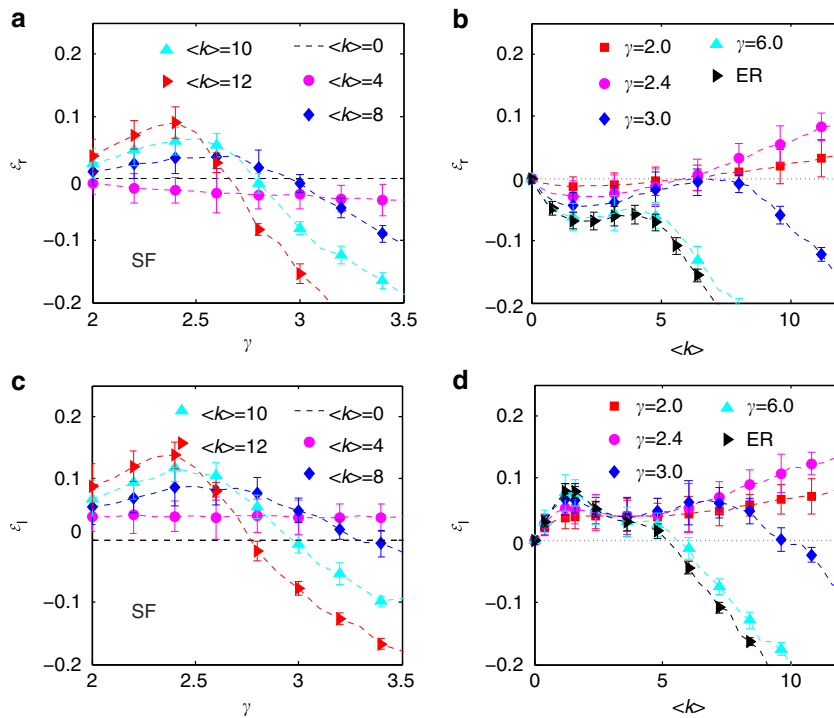
Figure 5 | Overall target control efficiency of model networks. (**a,b**) Random selection scheme. (**a**) The overall target control efficiency $\mathcal{E}$ (equation (4)) for SF networks in function of degree exponent γ for different values of average degree $\langle k \rangle$ in the random scheme. (**b**) The overall target control efficiency $\mathcal{E}$ (equation (4)) in function of the average degree $\langle k \rangle$ for different values of the degree exponent γ for SF networks and ER networks for random scheme. (**c,d**) Local selection scheme. (**c**) The overall target control efficiency $\mathcal{E}$ (equation (4)) for SF networks in function of degree exponent γ for different values of average degree $\langle k \rangle$ for local scheme. (**d**) The overall target control efficiency $\mathcal{E}$ (equation (4)) in function of the average degree $\langle k \rangle$ for different values of degree exponent γ of SF networks and ER networks for local scheme. From **a** and **b** we can see that for large $\langle k \rangle$ when γ is small the system has positive target control efficiency, but when γ is greater than a certain value, the efficiency decreases monotonically with γ becoming negative. Comparing **a,b** with (**c,d**), we find that the local control can enhance the target control efficiency. The results are obtained by averaging over 200 realizations of networks with 10^4 nodes.

particular driver node under one specific maximum matching. We can then count the minimal number of disjoint cacti that cover all target nodes, which yields an upper bound of P_D for target control (denoted as P_D^{UB}) (see Fig. 2; Supplementary Note 3 for details). The lower bound of P_D, denoted with P_D^{LB}, can also be obtained using maximum matching. We consider the maximum matching in a bipartite graph connecting on one side (1) the target nodes ($\mathcal{C}$ nodes), and on the other (2) the nodes that can reach the target nodes via (3) the edges among them. If a target node is unmatched, we must drive it directly, and the number of unmatched target nodes is the lower bound of P_D. Taking Fig. 2 for example, the lower bound of P_D is given by the first iteration of the GA, that is, $P_D^{LB} = 1$ for the target nodes {1, 2, 4, 6, 7}, because in the first iteration only node 1 is unmatched.

Target controllability. To quantify the efficiency of target control for a specific fraction f, we define the target controllability parameter $\alpha_D \equiv P_D/N_D$ (Methods section) and investigate it for both random and local control schemes, as shown in Fig. 4a,d, respectively. Note that for $f = 1$, we have $\alpha_D = 1$ as target control reduces to full control. In Fig. 4 the black lines denote the neutral condition ($\alpha_D = f$) and serve as a benchmark because they mean that controlling an f fraction of target nodes requires an f fraction of the driver nodes needed for full control, that is, $P_D = fN_D$. In each figure, UB denotes the upper bound of the minimum number of driver nodes needed to control an f fraction of target nodes, predicted by the structural control theory (Supplementary Note 3). Likewise, GA denotes the results of the greedy algorithm

and LB (Supplementary Note 3) denotes the lower bound, which is obtained by the first step of the GA (Methods section). Figure 4b shows the results of random target control of ER networks. Here the GA curve is above the neutral line ($\alpha_D > f$), which indicates that the target control is less efficient than the neutral expectation. Figure 4c shows the results of random target control on SF networks and the GA curve is almost at the neutral line ($\alpha_D \approx f$), which indicates that target control is as efficient as the neutral expectation. If we apply the local target control scheme to an ER network, we observe reduced efficiency for $f < 0.5$ (minus symbol in Fig. 4e) and enhanced efficiency for $f > 0.5$ (plus symbol in Fig. 4e). Figure 4f shows the case of local target control on SF networks, charactering that the target control is more efficient (that is, $\alpha_D < f$).

In general, we find that the local target control scheme requires fewer driver nodes when compared with the random target control scheme. This can be explained as follows. The GA helps us obtain an approximately minimal set of driver nodes sufficient to control the target nodes. Yet, the size of the controllable subsystem (or equivalently, the dimension of the controllable subspace) can be larger than the size of the target node set. In other words, we may actually be able to control a larger subsystem than necessary.

In Supplementary Fig. 3, we show that the random target control scheme has larger controllable subsystems than the local target control scheme. This is true for any fraction f of target nodes. Hence the local control scheme is more efficient than the random control scheme. Furthermore, we find that α_D is robust to changes in network size as shown in Supplementary Fig. 6.
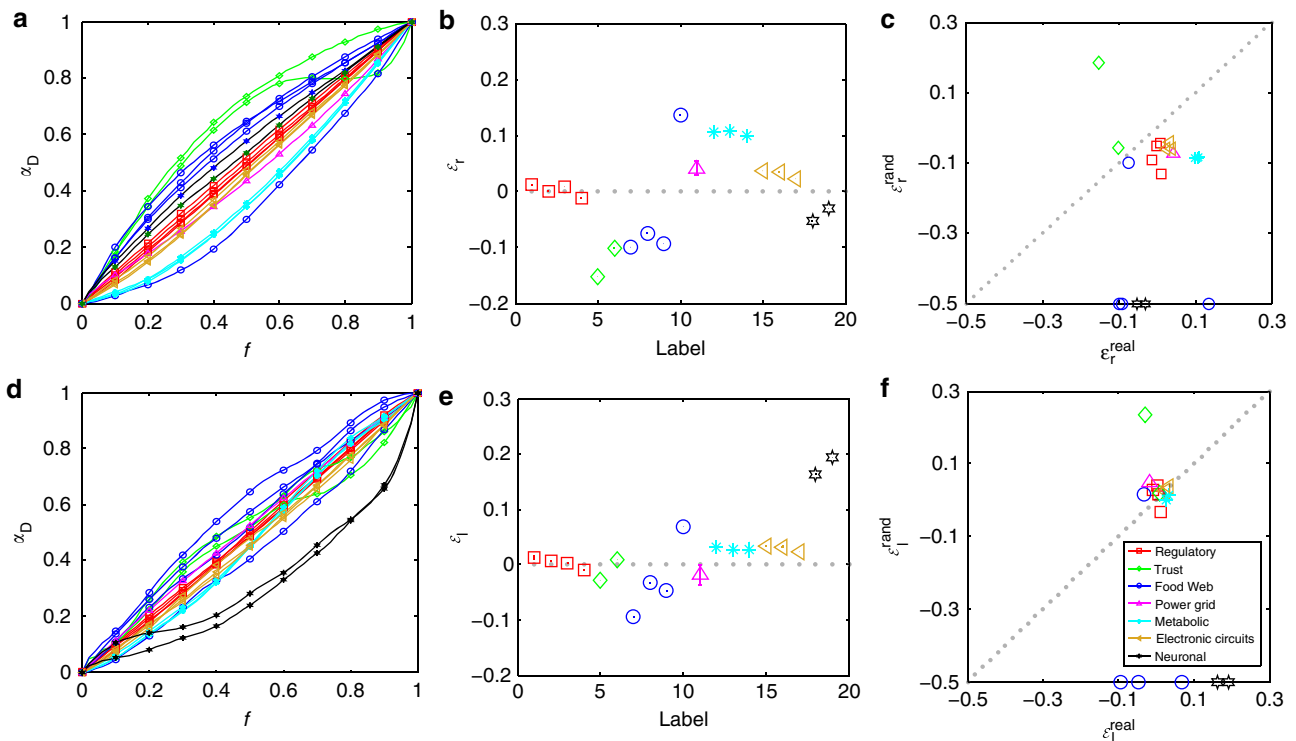
Figure 6 | Target control of real-world networks. (a–c) show random selection scheme. **(a)** The normalized fraction of driver nodes (α_D) in function of the target node fraction f for the random scheme for 19 different real-world networks. Network labels are in Table 1. **(b)** The overall target control efficiency $\mathcal{E}_r$ (equation (4)) of the real-world networks for the random scheme. **(c)** Comparing the target control efficiency of real networks with their randomized counterparts for the random scheme after, we randomly rewire each network using degree-preserving randomization. **(d–f)** Local selection scheme. **(d)** The normalized fraction of driver nodes (α_D) in function of the target node fraction f for the local scheme for different real-world networks. **(e)** The overall target control efficiency $\mathcal{E}_1$ (equation (4)) of real-world networks for the local scheme. **(f)** Comparing the target control efficiency of real networks with their randomized counterparts for the local scheme after, we randomly rewire each network using degree-preserving randomization. The results are averaged over 500 realizations. In general, real networks and more efficient than their randomized counterparts with respect to random target control.

Finally, since the hubs play an important role in complex networks, we study the case of controlling an f fraction of the highest in- and out-degree nodes as target nodes (Supplementary Fig. 5). In general, we find that controlling the high-degree nodes reduces the minimum number of driver nodes.

Target control efficiency. The above observations raise a fundamental question: what kind of network topology is more efficient for target control and which control scheme (random or local) facilitates target control for a general f fraction of target nodes? To address this question, we define the overall target control efficiency in the Methods. The overall target control efficiency of two model networks (ER and SF) are shown in Fig. 5 as provided by simulations. When $\langle k \rangle = 0$ (a network with N isolated nodes), we have $\mathcal{E} = 0$. That is, if we wish to control an f fraction of nodes we need to drive the same fraction of driver nodes as in the case of full control. We observe a peak of target control efficiency $\mathcal{E}$ for SF networks with degree exponent $\gamma \in [2, 3]$, which is the range of γ relevant for many real-world networks[37,38] (Fig. 5a,c).

This optimal target control efficiency is likely due to the presence of the hubs. Indeed, starting from an ER network, we can fix the number of links and rewire the network to increase the degree of a preselected node, turning it into a hub. We find that the target control efficiency increases as we develop a hub (Supplementary Fig. 4; Supplementary Note 4), explaining the increase of efficiency as we lower γ from $\gamma > 3$ to $\gamma \approx 2.5$.

Overall, Fig. 5a suggests that many real-world networks may have been optimized for efficient target controllability. Interestingly, we find that networks, which are easier to control if we wish to achieve full control, that is, those with large average degree $\langle k \rangle$ or large degree exponent γ[12], do not show high efficiency for target control. Moreover, SF networks with $2 \leq \gamma \leq 3$, which are harder to fully control than ER networks, have high random target control efficiency (Fig. 5c). Compared with ER networks, a SF network has lower local target control efficiency when the average degree $\langle k \rangle$ is small but higher local target control efficiency when the average degree $\langle k \rangle$ is large (Fig. 5d). Figure 5d also indicates the existence of a critical $\langle k \rangle_c$ so that when $\langle k \rangle > \langle k \rangle_c$ SF networks are more efficient than ER networks for target control. In general, sparse and homogeneous networks display high efficiency for target control, especially in the local scheme.

Target control of real networks. Next we apply the tools developed above to several real networks divided into seven categories, which are chosen for their diversity in applications and scope (Fig. 6; Table 1). We find that networks from the same category display similar target control efficiency (Fig. 6a,d). Interestingly, only networks with large average degree ($\langle k \rangle > \langle k \rangle_c 2.4$) are more efficient than their randomized counterparts with respect to the local target control scheme, that is, $\mathcal{E}_1^{rand} < \mathcal{E}_1^{real}$ (with values indicated in italics in Table 1, column $\langle k \rangle$ and in Fig. 6f). This is in agreement with simulations (Fig. 5d). We observe that local target control markedly increases

Table 1 | The properties of the real network analysed in this paper.

Label	Name	N	L	$\langle k \rangle$	n_D	$\mathcal{E}_r$	$\mathcal{E}_1$	$\mathcal{E}_r^{rand}$	$\mathcal{E}_1^{rand}$
Regulatory									
1	TRN-Yeast-1	4,441	12,873	2.9	0.965	0.0116	0.0127	− 0.1322	− 0.0334
2	TRN-Yeast-2	688	1,079	1.67	0.821	− 0.0002	0.0077	− 0.0523	0.0168
3	TRN-EC-1	1,550	3,340	2.15	0.891	0.0089	0.0021	− 0.045	0.0393
4	TRN-EC-2	418	519	1.24	0.751	− 0.0126	− 0.0087	− 0.0911	0.0289
Trust									
5	College student	32	96	3	0.188	− 0.1527	− 0.0281	0.1848	0.2352
6	Prison inmate	67	182	2.72	0.134	− 0.1015	0.0088	− 0.057	0.0178
Food web									
7	Ythan	135	601	4.45	0.511	− 0.0925	− 0.0469	− 0.5	− 0.5
8	Little rock	183	2,494	13.63	0.541	0.1364	0.0672	− 0.5	− 0.5
9	Grassland	88	137	1.56	0.523	− 0.0747	− 0.0312	− 0.0997	0.0162
10	Seagrass	49	226	4.61	0.265	− 0.0994	− 0.0937	− 0.5	− 0.5
Power grid									
11	TexasPowerGrid	4,889	5,855	1.2	0.325	0.0417	− 0.0189	− 0.0739	0.0501
Metabolic									
12	*E. coli*	2,275	5,763	2.53	0.382	0.0984	0.0277	− 0.0865	− 0.0005
13	*S. cerevisiae*	1,511	3,833	2.54	0.329	0.1079	0.0266	− 0.0837	− 0.0011
14	*C. elegans*	1,173	2,864	2.44	0.302	0.1059	0.0318	− 0.0877	0.0116
Electronic circuits									
15	s838	512	819	1.6	0.232	0.0372	0.0340	− 0.0593	0.0362
16	s420	252	399	1.58	0.234	0.0340	0.0325	− 0.0411	0.0358
17	s208	122	189	1.55	0.238	0.0227	0.0238	− 0.0604	0.0256
Neuronal									
18	*C. elegans*.Op	297	2,345	7.9	0.165	− 0.0516	0.1623	− 0.5	− 0.5
19	*C. elegans*.Ne	297	2,345	7.9	0.165	− 0.0295	0.1927	− 0.5	− 0.5

For each network, we show its type, name, number of nodes (N), edges (L), average degree $\langle k \rangle$, fraction of driver nodes for full control (n_D), random and local target control efficiency $\mathcal{E}_r$ and $\mathcal{E}_1$ of the original network and the corresponding randomized $\mathcal{E}_r^{rand}$ and $\mathcal{E}_1^{rand}$. For some networks the randomized target control efficiency is − 0.5, because these networks are small and dense, hence after randomization one node can control the whole network. Networks with large average degree ($\langle k \rangle > 2.4$) are more efficient than their randomized counterparts with respect to the local target control scheme, with values indicated in italics, column $\langle k \rangle$.

the control efficiency of neuronal networks (Fig. 6e). The opposite result is obtained for the power grid, finding that random target control is more efficient than the local control (Fig. 6b,e). This implies that the structure of the power grid is not optimized to facilitate local target control. For all gene networks target control efficiency is close to 0, indicating that they have evolved towards a topology that has comparable control efficiency within any sub-segment. Also, in the random target control scheme, the comparison between real networks and randomized networks support our earlier conclusion: real networks are optimized to have high random target control efficiency, because $\mathcal{E}_r > \mathcal{E}_r^{rand}$ for almost all of them. The two counter-examples are the trust networks, which are extremely small, however, with only 32 and 67 nodes. The target control efficiency of real networks is summarized in Table 1.

Discussion
In this work, we studied the target controllability of complex networks. We developed a new theoretical approach, the *k*-walk theory, to identify the controllable sub-graph that one node can control, and a GA to identify an approximately minimum set of driver nodes to control a specified target set of nodes. We studied both random and local target control schemes and analysed how the network topology impacts the target control efficiency. The GA proposed in this work is efficient in finding the driver nodes for target control when the network structure is completely known. In reality, we have very incomplete maps of many real-world networks and the full controllability of networks with

missing links has been recently addressed in ref. 39. Note that if we add the missing links back, they will only enhance the target controllability of the system, unless the link weights satisfy some particular algebraic constraints, which is a typical zero-measure event.

Our results raise several open questions: what higher-order network characteristics (such as communities, degree correlations) determine target controllability and target control efficiency? How to design the control inputs to steer the target nodes towards a desired final state? How to apply the GA to the study the target control of link dynamics?

Finally, as discussed in the Methods section, our formulation considers the canonical linear time-invariant dynamics of a system. Extending these findings to networks with fully nonlinear dynamics may require connecting our current work with complimentary approaches that consider the basins of attraction that describe the steady-state properties of nonlinear dynamics[10,40]. Understanding these questions can significantly improve our understanding of the control principles of complex systems.

Methods
Greedy algorithm. The GA we introduce here is based, in part, on iterating the procedure for the lower-bound calculation to ultimately approximate the minimum set of driver nodes (P_D^{GA}) for target control (Fig. 2d,e). The lower-bound of P_D can be calculated as follows: (1) Construct a bipartite graph $\mathcal{B}$, where the right side $\mathcal{R}$ consists of all the target nodes, and the left side $\mathcal{L}$ consists of all the nodes that can reach the target nodes. There is a link between node $u \in \mathcal{L}$ and $v \in \mathcal{R}$ if there is a link $u \rightarrow v$ in the original directed network $\mathcal{G}$. (2) Find a maximum matching in $\mathcal{B}$. The number of unmatched nodes in $\mathcal{R}$ yields the lower-bound of P_D. The greedy algorithm works as follows. (1) Initialize the set P_D^{GA} to be the set of the unmatched nodes found from the lower-bound calculation. (2) Identify the set of nodes in $\mathcal{L}$

that match the nodes in $\mathcal{R}$, and let this node set to be the new $\mathcal{R}$ set and get a new bipartite graph. (3) Calculate a maximum matching in the updated bipartite graph and add unmatched nodes in $\mathcal{R}$ to the set P_D^{GA}. (4) Repeat (2) and (3) until all nodes have been matched or are in the set P_D^{GA}. In Fig. 2d,e, we offer a specific example and the details of the GA. The proof of its sufficiency for target controllability is provided in Supplementary Fig. 2 and Supplementary Note 2. Figure 2e can also be obtained by the k-walk theory, because the length of the path from the node 1 to each target node {1, 2, 4, 6, 7} is unique. Note that if the GA converges after only one iteration, we obtain the exact number for the minimum number of driver nodes for target control.

A direct illustration of our algorithm's utility is illustrated in Fig. 3, where we consider control of the network shown in Fig. 2. To control the entire network, we need at least three driver nodes. Indeed, as we show in Fig. 3a,b using these three nodes we can move the state of all nodes to the desired final state $x_i = 10$. But if we just want to control a subsystem {1, 2, 4, 6, 7} (highlighted in red in Fig. 2e), the GA predicts that we need a single driver node, node 1. Indeed, as we show in Fig. 3c,d we can now move the state of the target nodes to the desired final state through an input to node 1 only. The nodes outside of the target list take arbitrary values as we do not control them.

Target control efficiency. We define the overall target control efficiency of an arbitrary network as

$$\mathcal{E} = 0.5 - \int_0^1 \alpha_D(f) \mathrm{d}f, \qquad (4)$$

denoting the efficiency of random and local target control scheme by $\mathcal{E}_r$ and $\mathcal{E}_1$, respectively. For example, for $\mathcal{E} = 0$ the overall network efficiency is neutral, that is, to control an f fraction of target nodes we need fN_D driver nodes. When $\mathcal{E} < 0 (\mathcal{E} > 0)$, the network is less (or more) efficient than neutral expectation. Furthermore, $\mathcal{E} \in [-0.5, 0.5]$, so $\mathcal{E} = 0.5$ corresponds to the most efficient case and $\mathcal{E} = -0.5$ shows the least efficient case. One example of the least efficient case is when only one driver node is needed to control the whole network. If we control any fraction of nodes, we still need one driver node, thus $\mathcal{E} = -0.5$ (observed for Food Web and Neuronal networks in Table 1). Note that target controllability for a specific fraction, $\alpha_D(f)$, depends on the fraction f of driver nodes, but overall target control efficiency, $\mathcal{E}$, is a property of the whole network, independent of the fraction of target nodes.

References

1. Dorogovtsev, S. N. & Mendes, J. F. *Evolution of Networks: From Biological Nets to the Internet and WWW (Physics)* (Oxford University Press, 2003).
2. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
3. Cohen, R. & Havlin, S. *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, 2010).
4. Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, 2010).
5. Song, C., Havlin, S. & Makse, H. A. Self-similarity of complex networks. *Nature* **433**, 392–395 (2005).
6. Barrat, A., Barthelemy, M. & Vespignani, A. *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).
7. Gallos, L. K., Song, C., Havlin, S. & Makse, H. A. Scaling theory of transport in complex biological networks. *Proc. Natl Acad. Sci. USA* **104**, 7746–7751 (2007).
8. Dorf, R. C. *Modern Control Systems* (Addison-Wesley Longman Publishing Co. Inc., 1991).
9. Boyd, S. P., El Ghaoui, L., Feron, E. & Balakrishnan, V. *Linear Matrix Inequalities in System and Control Theory* 15 (Siam, 1994).
10. Cornelius, S. P., Kath, W. L. & Motter, A. E. Realistic control of network dynamics. *Nat. Commun.* **4**, 1942 (2013).
11. Kalman, R. E. Mathematical description of linear dynamical systems. *J. Soc. Indus. Appl. Math. Ser. A* **1**, 152 (1963).
12. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
13. Lin, C.-T. Structural controllability. *IEEE T. Automat. Contr.* **19**, 201–208 (1974).
14. Hopcroft, J. E. & Karp, R. M. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput.* **2**, 225–231 (1973).
15. Nepusz, T. & Vicsek, T. Controlling edge dynamics in complex networks. *Nat. Phys.* **8**, 568–573 (2012).
16. Kuchtey, J., Fulton, S. A., Reba, S. M., Harding, C. V. & Boom, W. H. Interferon-αβ mediates partial control of early pulmonary mycobacterium bovis bacillus calmette-guérin infection. *Immunology* **118**, 39–49 (2006).
17. Baldea, M. & Daoutidis, P. Model reduction and control of reactor-heat exchanger networks. *J Process Contr.* **16**, 265–274 (2006).
18. Cohen, R., Havlin, S. & ben Avraham, D. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.* **91**, 247901 (2003).
19. Galbiati, M., Delpini, D. & Battiston, S. The power to control. *Nat. Phys.* **9**, 126–128 (2013).
20. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
21. Erdős, P. R. A. On random graphs. *I. Publ. Math.* **6**, 290–297 (1959).
22. Erdős, P. & Rényi, A. On the evolution of random graphs. *Inst. Hung. Acad. Sci.* **5**, 17–61 (1960).
23. Slotine, J.-J. & Li, W. *Applied Nonlinear Control* (Prentice-Hall, 1991).
24. Murota, K. & Poljak, S. Note on a graph-theoretic criterion for structural output controllability. *IEEE T. Automat. Contr.* **35**, 939–942 (1990).
25. Ruths, J. & Ruths, D. Control profiles of complex networks. *Science* **343**, 1373–1376 (2014).
26. Wu, F.-X., Wu, L., Wang, J., Liu, J. & Chen, L. Transittability of complex networks and its applications to regulatory biomolecular networks. *Sci. Rep.* **4**, 4819 (2014).
27. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Observability of complex systems. *Proc. Nat.l Acad. Sci.* **110**, 2460–2465 (2013).
28. Wang, W.-X., Ni, X., Lai, Y.-C. & Grebogi, C. Optimizing controllability of complex networks by minimum structural perturbations. *Phys. Rev. E* **85**, 026115 (2012).
29. Yan, G., Ren, J., Lai, Y.-C., Lai, C.-H. & Li, B. Controlling complex networks: how much energy is needed? *Phys. Rev. Lett.* **108**, 218703 (2012).
30. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Control centrality and hierarchical structure in complex networks. *Plos ONE* **7**, e44459 (2012).
31. Pósfai, M., Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Effect of correlations on network controllability. *Sci. Rep.* **3**, 1067 (2013).
32. Sun, J. & Motter, A. E. Controllability transition and nonlocality in network control. *Phys. Rev. Lett.* **110**, 208701 (2013).
33. Yuan, Z., Zhao, C., Di, Z., Wang, W.-X. & Lai, Y.-C. Exact controllability of complex networks. *Nat. Commun.* **4**, 2447 (2013).
34. Jia, T. *et al.* Emergence of bimodality in controlling complex networks. *Nat. Commun.* **4**, 2002 (2013).
35. Chan, B. Y. & Shachter, R. D. In *Proceedings of the Eighth International Conference on Uncertainty in Artificial Intelligence* 25–32 (Morgan Kaufmann Publishers Inc., 1992).
36. Blackhall, L. & Hill, D. J. On the structural controllability of networks of linear systems (2nd IFAC Workshop on Distributed Estimation and Control in Networked Systems) 245–250 (2010).
37. Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
38. Caldarelli, G. *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford University Press, 2007).
39. Slotine, J.-J. & Liu, Y.-Y. Complex networks: the missing link. *Nat. Phys.* **8**, 512–513 (2012).
40. Lai, Y.-C. *National Science Review* (Oxford University Press, 2014).

Author contributions

All authors designed and did the research. J.G. analysed the empirical data and did the analytical and numerical calculations. A.-L.B., R.M.D'S. and Y.-Y.L. were the lead writers of the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Gao, J. *et al.* Target control of complex networks. *Nat. Commun.* 5:5415 doi: 10.1038/ncomms6415 (2014).