Control Centrality and Hierarchical Structure in Complex Networks

Yang-Yu Liu[1,2], Jean-Jacques Slotine[3,4], Albert-László Barabási[1,2,5]*

1 Center for Complex Network Research and Department of Physics,Northeastern University, Boston, Massachusetts, United States of America, **2** Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America, **3** Nonlinear Systems Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **4** Department of Mechanical Engineering and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **5** Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America

Abstract

We introduce the concept of control centrality to quantify the ability of a single node to control a directed weighted network. We calculate the distribution of control centrality for several real networks and find that it is mainly determined by the network's degree distribution. We show that in a directed network without loops the control centrality of a node is uniquely determined by its layer index or topological position in the underlying hierarchical structure of the network. Inspired by the deep relation between control centrality and hierarchical structure in a general directed network, we design an efficient attack strategy against the controllability of malicious networks.

Introduction

Complex networks have been at the forefront of statistical mechanics for more than a decade [1–4]. Studies of them impact our understanding and control of a wide range of systems, from Internet and the power-grid to cellular and ecological networks. Despite the diversity of complex networks, several basic universal principles have been uncovered that govern their topology and evolution [3,4]. While these principles have significantly enriched our understanding of many networks that affect our lives, our ultimate goal is to develop the capability to control them [5–17].

According to control theory, a dynamical system is controllable if, with a suitable choice of inputs, it can be driven from any initial state to any desired final state in finite time [18–20]. By combining tools from control theory and network science, we proposed an efficient methodology to identify the minimum sets of driver nodes, whose time-dependent control can guide the whole network to any desired final state [12]. Yet, this minimum driver set (MDS) is usually not unique, but one can often achieve multiple potential control configurations with the same number of driver nodes. Given that some nodes may appear in some MDSs but not in other, a crucial question remains unanswered: what is the role of each individual node in controlling a complex system? Therefore the question that we address in this paper pertains to the importance of a given node in maintaining a system's controllability.

Historically, various types of centrality measures of a node in a network have been introduced to determine the relative importance of the node within the network in appropriate circumstances. For example, the degree centrality, closeness centrality [21], betweenness centrality [22], eigenvector centrality [23,24], PageRank [25], hub centrality and authority centrality [26], routing centrality [27], and so on. Here, we introduce control centrality to quantify the ability of a single node in controlling the whole network. Mathematically, control centrality of node i captures the dimension of the controllable subspace or the size of the controllable subsystem when we control node i *only*. This agrees well with our intuitive notion about the "power" of a node in controlling the whole network. We notice that control centrality is fundamentally different from the concept of control range, which quantifies the "duty" or "responsibility" of a node i in controlling a network *together with other driver nodes* [28].

Results

Control Centrality

Consider a complex system described by a directed weighted network of N nodes whose time evolution follows the linear time-invariant dynamics.

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \qquad (1)$$

where $\mathbf{x}(t) = (x_1(t), x_2(t), \cdots, x_N(t))^{\mathrm{T}} \in \mathbb{R}^N$ captures the state of each node at time t. $\mathbf{A} \in \mathbb{R}^{N \times N}$ is an $N \times N$ matrix describing the weighted wiring diagram of the network. The matrix element $a_{ij} \in \mathbb{R}$ gives the strength or weight that node j can affect node i. Positive (or negative) value of a_{ij} means the link $(j \rightarrow i)$ is excitatory (or inhibitory).

$\mathbf{B} \in \mathbb{R}^{N \times M}$ is an $N \times M$ input matrix ($M \leq N$) identifying the nodes that are controlled by the time dependent input vector $\mathbf{u}(t) = (u_1(t), u_2(t), \cdots, u_M(t))^{\mathrm{T}} \in \mathbb{R}^M$ with M independent signals imposed by an outside controller. The matrix element $b_{ij} \in \mathbb{R}$ represents the coupling strength between the input signal $u_j(t)$ and node i. The system (1), also denoted as $(\mathbf{A}, \mathbf{B})$, is controllable if and only if its controllability matrix $\mathbf{C} = (\mathbf{B}, \mathbf{AB}, \cdots, \mathbf{A}^{N-1}\mathbf{B}) \in \mathbb{R}^{N \times NM}$ has full rank, a criteria often called Kalman's controllability rank condition [18]. The rank of the controllability matrix $\mathbf{C}$, denoted by $\mathrm{rank}(\mathbf{C})$, provides the dimension of the controllable subspace of the system $(\mathbf{A}, \mathbf{B})$ [18,19]. When we control node i only, $\mathbf{B}$ reduces to the vector $\mathbf{b}^{(i)}$ with a single non-zero entry, and we denote $\mathbf{C}$ with $\mathbf{C}^{(i)}$. We can therefore use $\mathrm{rank}(\mathbf{C}^{(i)})$ as a natural measure of node i's ability to control the system: if $\mathrm{rank}(\mathbf{C}^{(i)}) = N$, then node i alone can control the whole system, i.e. it can drive the system between any points in the N-dimensional state space in finite time. Any value of $\mathrm{rank}(\mathbf{C}^{(i)})$ less than N provides the dimension of the subspace i can control. In particular if $\mathrm{rank}(\mathbf{C}^{(i)}) = 1$, then node i can only control itself.

The precise value of $\mathrm{rank}(\mathbf{C})$ is difficult to determine because in reality the system parameters, i.e. the elements of $\mathbf{A}$ and $\mathbf{B}$, are often not known precisely except the zeros that mark the absence of connections between components of the system [29]. Hence $\mathbf{A}$ and $\mathbf{B}$ are often considered to be structured matrices, i.e. their elements are either fixed zeros or independent free parameters [29]. Apparently, $\mathrm{rank}(\mathbf{C})$ varies as a function of the free parameters of $\mathbf{A}$ and $\mathbf{B}$. However, it achieves the maximal value for all but an exceptional set of values of the free parameters which forms a proper variety with Lebesgue measure zero in the parameter space [30,31]. This maximal value is called the *generic rank* of the controllability matrix $\mathbf{C}$, denoted as $\mathrm{rank}_g(\mathbf{C})$, which also represents the generic dimension of the controllable subspace. When $\mathrm{rank}_g(\mathbf{C}) = N$, the system $(\mathbf{A}, \mathbf{B})$ is *structurally controllable*, i.e. controllable for almost all sets of values of the free parameters of $\mathbf{A}$ and $\mathbf{B}$ except an exceptional set of values with zero measure [29,30,32,33]. For a single node i, $\mathrm{rank}_g(\mathbf{C}^{(i)})$ captures the "power" of i in controlling the whole network, allowing us to define the *control centrality* of node i as

$$C_c(i) \equiv \mathrm{rank}_g(\mathbf{C}^{(i)}). \tag{2}$$

The calculation of $\mathrm{rank}_g(\mathbf{C})$ can be mapped into a combinatorial optimization problem on a directed graph $G(\mathbf{A}, \mathbf{B})$ constructed as follows [31]. Connect the M input nodes $\{u_1, \cdots, u_M\}$ to the N state nodes $\{x_1, \cdots, x_N\}$ in the original network according to the input matrix $\mathbf{B}$, i.e. connect u_j to x_i if $b_{ij} \neq 0$, obtaining a directed graph $G(\mathbf{A}, \mathbf{B})$ with $N + M$ nodes (see Fig. 1a and b). A state node j is called *accessible* if there is at least one directed path reaching from one of the input nodes to node j. In Fig. 1b, all state nodes $\{x_1, \cdots, x_7\}$ are accessible from the input node u_1. A *stem* is a directed path starting from an input node, so that no nodes appear more than once in it, e.g. $u_1 \rightarrow x_1 \rightarrow x_5 \rightarrow x_7$ in Fig. 1b. Denote with G_s the *stem-cycle disjoint* subgraph of $G(\mathbf{A}, \mathbf{B})$, such that G_s consists of stems and cycles only, and the stems and cycles have no node in common (highlighted in Fig. 1b). According to Hosoe's theorem [31], the generic dimension of the controllable subspace is given by

$$\mathrm{rank}_g(\mathbf{C}) = \max_{G_s \in G} |E(G_s)| \tag{3}$$

with G the set of all stem-cycle disjoint subgraphs of the accessible part of $G(\mathbf{A}, \mathbf{B})$ and $|E(G_s)|$ the number of edges in the subgraph G_s. For example, the subgraph highlighted in Fig. 1b, denoted as $G_s^{\max}$, contains the largest number of edges among all possible stem-cycle disjoint subgraphs. Thus, $C_c(1) = \mathrm{rank}_g(\mathbf{C}^{(1)}) = 6$, which is the number of red links in Fig. 1b. Note that $\mathrm{rank}_g(\mathbf{C}^{(1)}) = 6 < N = 7$, the whole system is therefore not structurally controllable by controlling x_1 only. Yet, the nodes covered by the $G_s^{\max}$ highlighted in Fig. 1b, e.g. $\{x_1, x_2, x_3, x_4, x_5, x_7\}$, constitute a structurally controllable subsystem [33]. In other words, by controlling node x_1 with a time dependent signal $u_1(t)$ we can drive the subsystem $\{x_1, x_2, x_3, x_4, x_5, x_7\}$ from any initial state to any final state in finite time, for almost all sets of values of the free parameters of $\mathbf{A}$ and $\mathbf{B}$ except an exceptional set of values with zero measure. In general $G_s^{\max}$ is not unique. For example, in Fig. 1b we can get the same cycle $x_2 \rightarrow x_3 \rightarrow x_4 \rightarrow x_2$ together with a different stem $u_1 \rightarrow x_1 \rightarrow x_5 \rightarrow x_6$, which yield a different $G_s^{\max}$ and thus a different structurally controllable subsystem $\{x_1, x_2, x_3, x_4, x_5, x_6\}$. Both subsystems are of size six, which is exactly the generic dimension of the controllable subspace. Note that we can fully control each subsystem individually, yet we cannot fully control the whole system.

The advantage of Eq.(3) is that $\max_{G_s \in G} |E(G)|$ can be calculated via linear programming [34], providing us an efficient numerical tool to determine the control centrality and the structurally controllable subsystem of any node in an arbitrary complex network (see Fig. S1).

Distribution of Control Centrality

We first consider the distribution of control centrality. Shown in Fig. 2 is the distribution of the normalized control centrality ($c_c(i) \equiv C_c(i)/N$) for several real networks. We find that for the intra-organization network, $P(c_c)$ has a sharp peak at $c_c = 1$, suggesting that a high fraction of nodes can individually exert full control over the whole system (Fig. 2a). In contrast, for company-ownership network, $P(c_c)$ follows an approximately exponential distribution or a very short power-law distribution (Fig. 2d), indicating that most nodes display low control centrality. Even the most powerful node, with $c_c \sim 0.01$, can control only one percent of the total dimension of the system's full state space. For other networks $P(c_c)$ displays a mixed behavior, indicating the coexistence of a few powerful nodes with a large number of nodes that have little control over the system's dynamics (Fig. 2b,c). Note that under full randomization, turning a network into a directed Erdös-Rényi (ER) random network [35,36] with number of nodes (N) and number of edges (L) unchanged, the c_c distribution changes dramatically. In contrast, under degree-preserving randomization [37,38], which keeps the in-degree (k_{in}) and out-degree (k_{out}) of each node unchanged, the c_c distribution does not change significantly. This result suggests that $P(c_c)$ is mainly determined by the underlying network's degree distribution $P(k_{in}, k_{out})$. (Note that similar results were also observed for the minimum number of driver nodes [12] and the distribution of control range [28].) This result is very useful in the following sense: $P(k_{in}, k_{out})$ is easy to calculate for any complex network, while the calculation of $P(c_c)$ requires much more computational efforts (both CPU time and memory space). Studying $P(c_c)$ for model networks of prescribed $P(k_{in}, k_{out})$ will give us qualitative understanding of how $P(c_c)$ changes as we vary network parameters, e.g. mean degree $\langle k \rangle$. See Fig. S7 for more details.
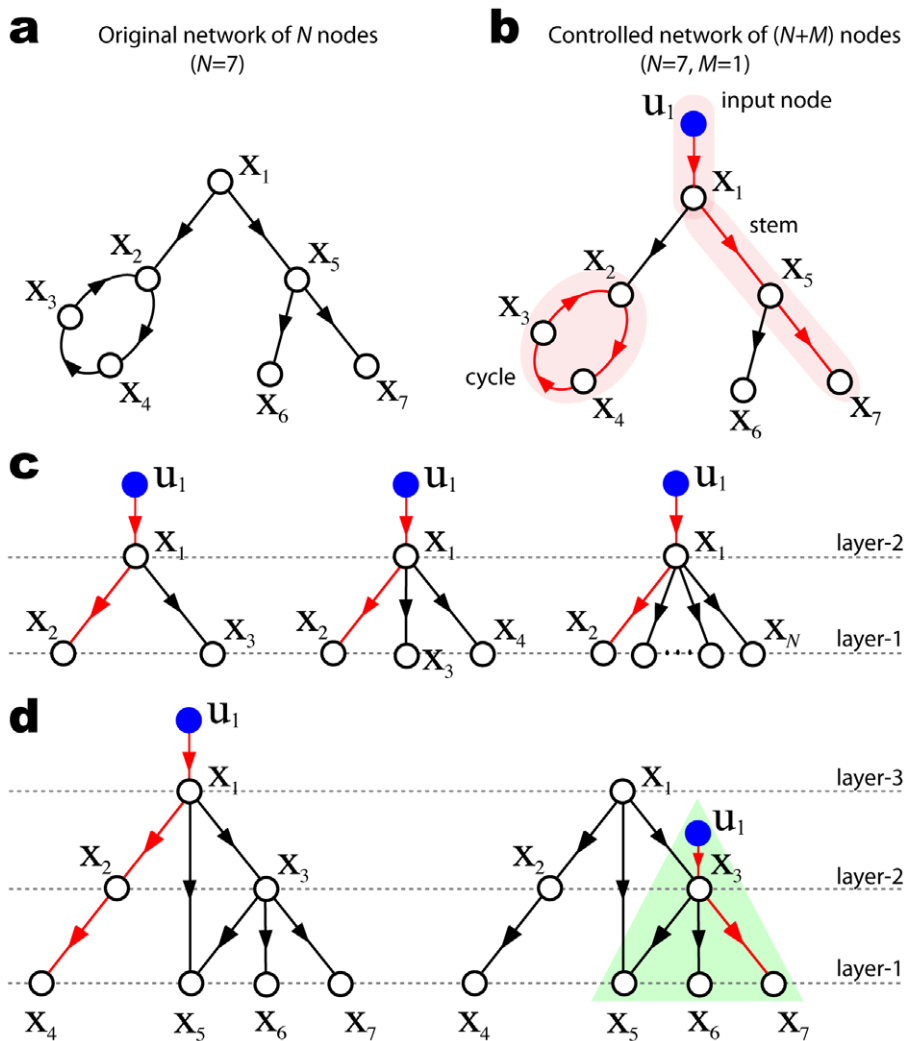
Figure 1. Control centrality. (a) A simple network of $N=7$ nodes. (b) The controlled network is represented by a directed graph $G(\mathbf{A},\mathbf{B})$ with an input node u_1 connecting to a state node x_1. The stem-cycle disjoint subgraph G_s (shown in red) contains six edges, which is the largest number of edges among all possible stem-cycle disjoint subgraphs of the directed graph $G(\mathbf{A},\mathbf{B})$ and corresponds to the generic dimension of controllable subspace by controlling node x_1. The control centrality of node 1 is thus $C_c(1)=6$. (c) The control centrality of the central hub in a directed star is always 2 for any network size $N \geq 2$. (d) The control centrality of a node in a directed acyclic graph (DAG) equals its layer index. In applying Hosoe's theorem, if not all state nodes are accessible, we just need to consider the accessible part (highlighted in green) of the input node(s).
doi:10.1371/journal.pone.0044459.g001

Control Centrality and Topological Features

To understand which topological features determine the control centrality itself, we compared the control centrality for each node in the real networks and their randomized counterparts (denoted as rand-ER and rand-Degree). The lack of correlations indicates that both randomization procedures eliminate the topological feature that determines the control centrality of a given node (see Fig. S2). Since accessibility plays an important role in maintaining structural controllability [29], we conjecture that the control centrality of node i is correlated with the number of nodes $N_r(i)$ that can be reached from it. To test this conjecture, we calculated $N_r(i)$ and $C_c(i)$ for the real networks shown in Fig. 2, observing only a weak correlation between the two quantities (see Fig. S3). This lack of correlation between $N_r(i)$ and $C_c(i)$ is obvious in a directed star, in which a central hub (x_1) points to $N-1$ leaf nodes ($x_2, \cdots, x_N$) (Fig. 1c). As the central hub can reach all nodes, $N_r(1)=N$, suggesting that it should have high control centrality. Yet, one can easily check that the central hub has control centrality $C_c(1)=2$ for any $N \geq 2$ and there are $N-1$ structurally controllable subsystems, i.e. $\{x_1, x_2\}, \cdots, \{x_1, x_{N-1}\}$. In other words, by controlling the central hub we can fully control each leaf node individually, but we cannot control them collectively.

Note that in a directed star each node can be labeled with a unique *layer index*: the leaf nodes are in the first layer (bottom layer) and the central hub is in the second layer (top layer). In this case the control centrality of the central hub equals its layer index (see Fig. 1c). This is not by coincidence: we can prove that for a directed network containing no cycles, often called a directed acyclic graph (DAG), the control centrality of any node equals its *layer index*.

$$C_c(i)=l_i. \tag{4}$$

Indeed, lacking cycles, a DAG has a unique *hierarchical structure*, which means that each node can be labeled with a unique layer
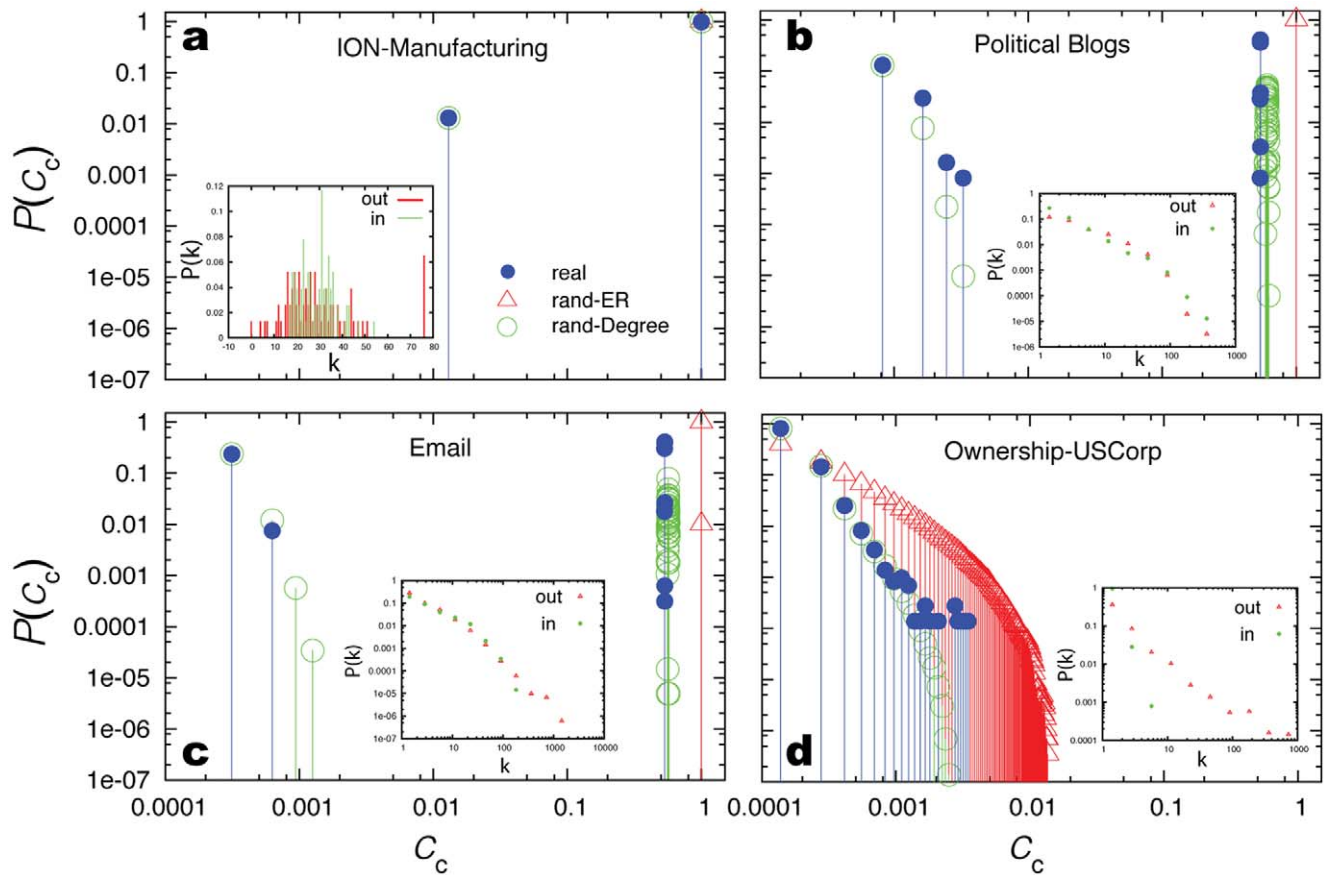
Figure 2. Distribution of normalized control centrality of several real-world networks (blue) and their randomized counterparts: rand-ER (red), rand-Degree (green), plotted in log-log scale. (a) Intra-organizational network of a manufacturing company [49]. (b) Hyperlinks between weblogs on US politics [50]. (c) Email network in a university [51]. (d) Ownership network of US corporations [52]. In- and out-degree distributions for each network are shown in the insets. See Table 1 for other network characteristics.
doi:10.1371/journal.pone.0044459.g002

index (l_i), calculated using a recursive labeling algorithm [39]: (1) Nodes that have no outgoing links ($k_{out}=0$) are labeled with layer index 1 (bottom layer). (2) Remove all nodes in layer 1. For the remaining graph identify again all nodes with $k_{out}=0$ and label them with layer index 2. (3) Repeat step (2) until all nodes are labeled. As the DAG lacks cycles, each subgraph in the set G of the directed graph $G(\mathbf{A}, \mathbf{b}^{(i)})$ consists of a stem only, which starts from the input node pointing to the state node i and ends at a state node in the bottom layer, e.g. $u_1 \rightarrow x_1 \rightarrow x_2 \rightarrow x_4$ in Fig. 1d. The number of edges in this stem is equal to the layer index of node i, so $\text{rank}_g(\mathbf{C}^{(i)}) = C_c(i) = l_i$. Therefore in DAG the higher a node is in

Table 1. Real networks analyzed in the paper.

name	N	L	$\langle k \rangle$	r	c
ION-Manufacturing [49]	77	2,228	57.9	−0.017	0.244
Political blogs [50]	1,224	19,025	31.1	−0.196	0.174
Email network [51]	3,188	39,256	24.6	−0.240	0.128
Ownership-USCorp [52]	7,253	6,726	1.9	−0.181	0.004

For each network, we show its name and reference; number of nodes (N) and edges (L); mean degree ($\langle k \rangle$); degree correlation (r) [4]; and clustering coefficient (c) [53], respectively.
doi:10.1371/journal.pone.0044459.t001

the hierarchy, the higher is its ability to control the system. Though this result agrees with our intuition to some extent, it is surprising at the first glance because it indicates that in a DAG the control centrality of node i is only determined by its topological position in the hierarchical structure, rather than any other importance measures, e.g. degree or betweenness centrality. This result also partially explains why driver nodes tend to avoid hubs [12]. (Note that similar phenomena about have been observed in other problems, e.g. networked transportation [40], synchronization [41] and epidemic spreading [42]).

Despite the simplicity of Eq. (4), we cannot apply it directly to real networks, because most of them are not DAGs. Yet, we note that any directed network has a underlying DAG structure based on the strongly connected component (SCC) decomposition (see Fig. S4). A subgraph of a directed network is *strongly connected* if there is a directed path from each node in the subgraph to every other node. The SCCs of a directed network G are its maximal strongly connected subgraphs. If we contract each SCC to a single supernode, the resulting graph , called the *condensation* of G, is a DAG [43]. Since a DAG has a unique hierarchical structure, a directed network can then be assigned an underlying hierarchical structure. The layer index of node i can be defined to be the layer index of the corresponding supernode (i.e. the SCC that node i belongs to) in . With this definition of l_i, it is easy to show that $C_c(i) \geq l_i$ for general directed networks (see Fig. S6 for more details). Furthermore, for an edge ($i \rightarrow j$) in a general directed
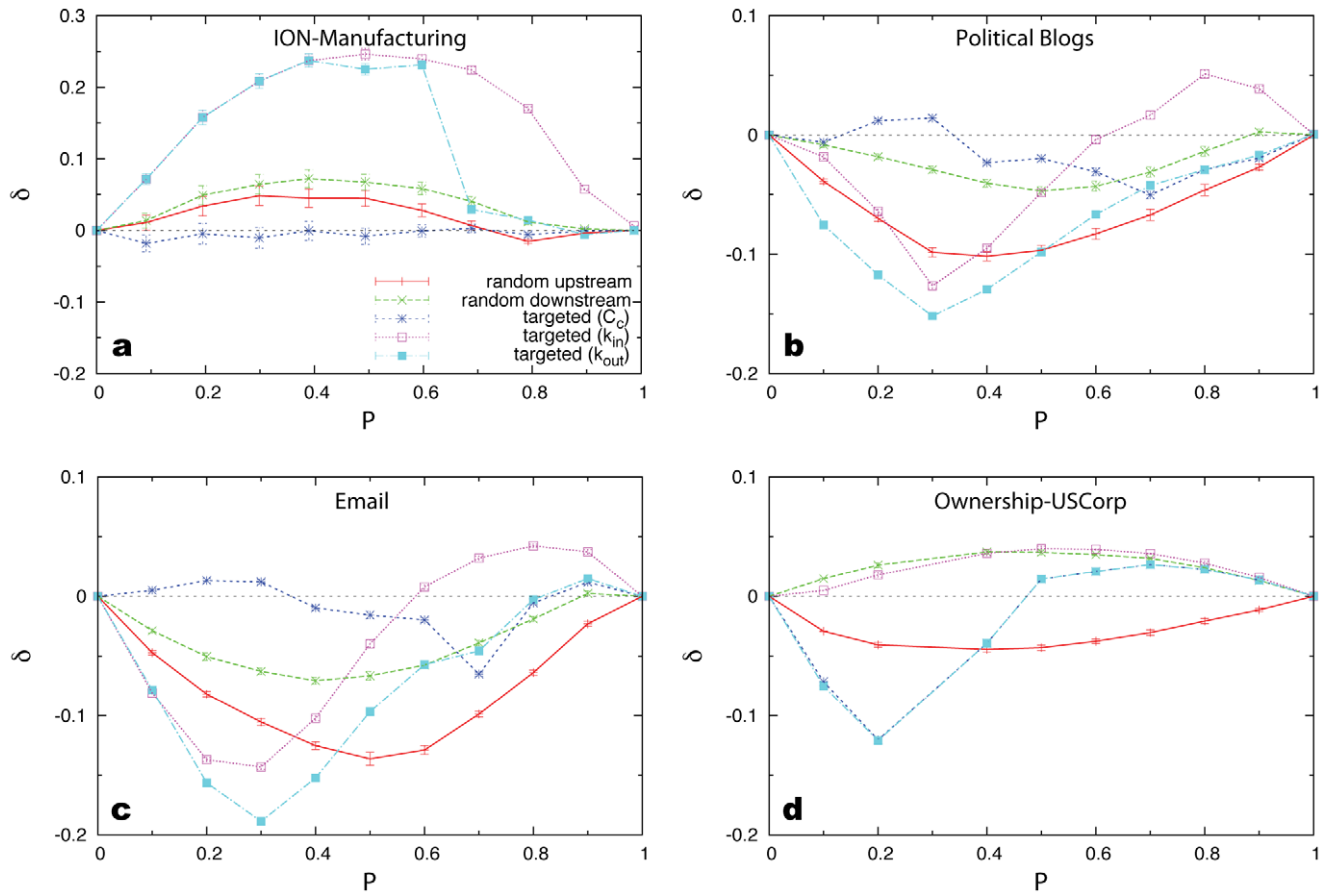
Figure 3. The impact of different attack strategies on network controllability with respective to the random attack.
$\delta \equiv [\text{rank}_g^{\text{Strategy}-j}(\mathbf{C}') - \text{rank}_g^{\text{Random}}(\mathbf{C}')]/N$ with $\text{rank}_g^{\text{Strategy}-j}(\mathbf{C}')$ represents the generic dimension of controllable subspace after removing a P fraction of nodes using strategy-j. The nodes are removed according to six different strategies. (Strategy-0) Random attack: randomly remove P fraction of nodes. (Strategy-1 or 2) Random upstream (or downstream) attack: randomly choose P fraction of nodes, randomly remove one of their upstream neighbors (or downstream neighbors). The results are averaged over 10 random choices of P fraction of nodes with error bars defined as s.e.m. Lines are only a guide to the eye. (Strategy-3,4, or 5) Targeted attacks: remove the top P fraction of nodes according to their control centralities (or in-degrees or out-degrees).
doi:10.1371/journal.pone.0044459.g003

network, if node i is topologically "higher" than node j (i.e. $l_i > l_j$), then $C_c(i) > C_c(j)$. Since $C_c(i)$ has to be calculated via linear programming which is computationally more challenging than the calculation of l_i, the above results suggest an efficient way to calculate the lower bound of $C_c(i)$ and to compare the control centralities of two neighboring nodes. Note that if $l_i > l_j$ and there is no directed edge $(i \rightarrow j)$ in the network, then in general one cannot conclude that $C_c(i) > C_c(j)$ (see Fig. S5 for more details).

Attack Strategy

Our finding on the relation between control centrality and hierarchical structure inspires us to design an efficient attack strategy against malicious networks, aiming to affect their controllability. The most efficient way to damage the controllability of a network is to remove all input nodes $\{u_1, u_2, \cdots, u_M\}$, rendering the system completely uncontrollable. But this requires a detailed knowledge of the control configuration, i.e. the wiring diagram of $G(\mathbf{A},\mathbf{B})$, which we often lack. If the network structure ($\mathbf{A}$) is known, one can attempt a *targeted attack*, i.e. rank the nodes according to some centrality measure, like degree or control centrality, and remove the nodes with highest centralities [44,45]. Though we still lack systematic studies on the effect of a targeted

attack on a network's controllability, one naively expects that this should be the most efficient strategy. But we often lack the knowledge of the network structure, which makes this approach unfeasible anyway. In this case a simple strategy would be *random attack*, i.e. remove a randomly chosen P fraction of nodes, which naturally serves as a benchmark for any other strategy. Here we propose instead a *random upstream attack* strategy: randomly choose a P fraction of nodes, and for each node remove one of its incoming or upstream neighbors if it has one, otherwise remove the node itself. A *random downstream attack* can be defined similarly, removing the node to which the chosen node points to. In undirected networks, a similar strategy has been proposed for efficient immunization [45] and the early detection of contagious outbreaks [46], relying on the statistical trend that randomly selected neighbors have more links than the node itself [47,48]. In directed networks we can prove that randomly selected upstream (or downstream) neighbors have more outgoing (or incoming) links than the node itself. Thus a random upstream (or downstream) attack will remove more hubs and more links than the random attack does. But the real reason why we expect a random upstream attack to be efficient in a directed network is because $C_c(i) \geq C_c(j)$ for most edges $(i \rightarrow j)$, i.e. the control centrality of the starting node

is usually no less than the ending node of a directed edge (see Fig. S8). In DAGs, for any edge $(i{\to}j)$, we have strictly $C_c(i) > C_c(j)$. Thus, the upstream neighbor of a node is expected to play a more important or equal role in control than the node itself, a result deeply rooted in the nature of the control problem, rather than the hub status of the upstream nodes.

To show the efficiency of the random upstream attack we compare its impact on fully controlled networks with several other strategies. We start from a network that is fully controlled ($\mathrm{rank}_g(\mathbf{C}) = N$) via a minimum set of N_D driver nodes. After the attack a P faction of nodes are removed, denoting with $\mathrm{rank}_g(\mathbf{C}')$ the dimension of the controllable subspace of the damaged network. We calculate $\mathrm{rank}_g(\mathbf{C}')$ as a function of P, with P tuned from 0 up to 1. Since the random attack serves as a natural benchmark, we calculate the difference of $\mathrm{rank}_g(\mathbf{C}')$ between a given strategy and the random attack, denoted as $\delta = [\mathrm{rank}_g^{\mathrm{Strategy}-j}(\mathbf{C}') - \mathrm{rank}_g^{\mathrm{Random}}(\mathbf{C}')]/N$. Obviously, the more negative is δ, the more efficient is the strategy compared to a fully random attack. We find that for most networks random upstream attack results in $\delta < 0$ for $0 < P < 1$, i.e. it causes more damage to the network's controllability than random attack (see Fig. 3b,c,d). Moreover, random upstream attack typically is more efficient than random downstream attack, even though in both cases we remove more hubs and more links than in the random attack. This is due to the fact that the upstream (or downstream) neighbors are usually more (or less) "powerful" than the node itself.

The efficiency of the random upstream attack is even comparable to targeted attacks (see Fig. 3). Since the former requires only the knowledge of the network's local structure rather than any knowledge of the nodes' centrality measures or any other global information (i.e. the structure of the $\mathbf{A}$ matrix) while the latter rely heavily on them, this finding indicates the advantage of the random upstream attack. The fact that those targeted attacks do not always show significant superiority over the random attacks is intriguing and would be explored in future work. Notice that for the intra-organization network all attack strategies fail in the sense that δ is either positive or very close to zero (Fig. 3a). This is due to the fact this network is so dense (with mean degree $\langle k \rangle \approx 58$) that we have $C_c(i) = C_c(j) = N$ for almost all the edges $(i{\to}j)$. Consequently, both random upstream and downstream attacks are not efficient and the C_c-targeted attack shows almost the same impact as the random attack. This result suggests that when the network becomes very dense its controllability becomes extremely robust against all kinds of attacks, consistent with our previous result on the core percolation and the control robustness against link removal [12]. We also tested those attack strategies on model networks (see Fig. S9, S10 and S11). The results are qualitatively consistent with what we observed in real networks.

Discussion

In sum, we study the control centrality of single node in complex networks and find that it is related to the underlying hierarchical structure of networks. The presented results help us better understand the controllability of complex networks and design an efficient attack strategy against network control. Due to the duality of controllability and observability [18,19], a similar centrality measure can be defined to quantify the ability of a single node in observing the whole system, i.e. inferring the state of the whole system.

Supporting Information

Figure S1 Calculation of control centrality (or the generic dimension of the controllable subspace). (a) The original controlled system is represented by a digraph $G(A,B)$. (b) The modified digraph $G'(A,B)$ used in solving the linear programming. Dotted and solid lines are assigned with weight $w_{ij} = 0$ and 1, respectively. The maximum-weight cycle partition is shown in red, which has weight 3, corresponding to the generic dimension of controllable subspace by controlling node x_1 or equivalently the control centrality of node x_1.
(TIF)

Figure S2 Control centrality of nodes in several real-world networks and their randomized counterparts: rand-ER (red), rand-Degree (green). (a) Intra-organizational network of a manufacturing company. (b) Hyperlinks between weblogs on US politics. (c) Email network in a university. (d) Ownership network of US corporations.
(TIF)

Figure S3 Control centrality vs. the number of reachable nodes. The real networks are the same as used in Fig. S2.
(TIF)

Figure S4 Any directed network has a underlying hierarchical structure. (a) A directed network of 50 nodes. There are seven SCCs highlighted in different colors. The nodes are colored according to their control centrality. The edge $(i{\to}j)$ is colored in green, red, or blue if $C_c(i)$ is larger than, smaller than, or equal to $C_c(j)$, respectively. For all edges with $l_i > l_j$, we have $C_c(i) > C_c(j)$. But this is not true for general node pairs $\{i, j\}$. (b) The condensation of the network in (a) is a DAG with three layers. Each node in the DAG represents a SCC in the original network.
(TIF)

Figure S5 Even if a lower node is accessible from a higher node, it is still possible that the control centrality of the higher node is smaller than or equal to the lower one.
(TIF)

Figure S6 Control centrality as a function of layer index in several real-world networks. The real networks are the same as used in Fig. S2. Symbol ('+') represents the average value of C_c with error bar defined as the C_c range, i.e. $[C_c^{\min}, C_c^{\max}]$, for all the nodes in the same layer of the largest connected component of the network. Dotted lines represents $C_c(i) = l_i$.
(TIF)

Figure S7 Variation of the hierarchical structure and its impact on the distribution of control centrality. (a) Number of layers (N_L). (b) Size of the giant SCC. Both ER and SF networks are generated from the Chung-Lu model with $N = 10^4$ and the results are averaged over 100 realizations with error bars defined as s.e.m. Dotted lines are only a guide to the eye. (c,d,e) Distribution of control centrality for ER networks at different $\langle k \rangle$ values ($\langle k \rangle = 1,2,8$).
(TIF)

Figure S8 Fraction of edges $(i{\to}j)$ which satisfy $C_c(i) > C_c(j)$. Fractions of edges $(i{\to}j)$ with $C_c(i) > C_c(j)$, $C_c(i) < C_c(j)$, and $C_c(i) = C_c(j)$, are denoted as $f_>$, $f_<$, and $f_=$, respectively. Both ER and SF networks are generated from the Chung-Lu model with $N = 10^3$ and the results are averaged over 100 realizations with error bars defined as s.e.m. Dotted lines are only a guide to the eye. (a) ER network. (b) SF network with $\gamma = 2.5$. (c) SF network with $\gamma = 2.1$.
(TIF)

Figure S9 Impact of different attack strategies on network controllability. $\mathrm{rank}_g^{\mathrm{Strategy}-j}(C')$ represents the

generic dimension of controllable subspace after removing a P fraction of nodes using strategy-j. The nodes are removed according to 10 different strategies (see text). Both ER and SF networks are generated from the Chung-Lu model with $N = 10^3$ and the results are averaged over 10 random choices of P fraction of nodes with error bars defined as s.e.m. Lines are only a guide to the eye.
(TIF)

Figure S10 Impact of different attack strategies on network controllability with respect to random attack. $\delta \equiv [\text{rank}_g^{\text{Strategy}-j}(C') - \text{rank}_g^{\text{Strategy}-0}(C')]/N$ denotes the generic dimension difference of the controllable subspace after removing a P fraction of nodes using strategy-j and random attack. The more negative is δ, the more efficient is the strategy compared to a random attack. Symbols are the same as used in Fig. S9.
(TIF)

Figure S11 Impact of different attack strategies on network connectivity. n_{lc} represents the normalized size of the largest connected component of the network after removing a P fraction of nodes. The nodes are removed according to 10 different strategies (see text). Both ER and SF networks are generated from the Chung-Lu model with $N = 10^3$ and the results are averaged over 10 random choices of P fraction of nodes with error bars defined as s.e.m. Lines are only a guide to the eye.
(TIF)

Author Contributions

Conceived and designed the experiments: Y-YL J-JS A-LB. Wrote the paper: Y-YL A-LB. Analyzed the empirical data and did the analytical and numerical calculations: Y-YL. Edited the manuscript: J-JS.

References

1. Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74: 47–97.
2. Newman M, Barabási AL, Watts DJ (2006) The Structure and Dynamics of Networks. Princeton: Princeton University Press.
3. Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286: 509.
4. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393: 440–442.
5. Wang XF, Chen G (2002) Pinning control of scale-free dynamical networks. Physica A 310: 521–531.
6. Tanner HG (2004) On the controllability of nearest neighbor interconnections. Decision and Control, 2004 CDC 43rd IEEE Conference on 3: 2467–2472.
7. Sorrentino F, di Bernardo M, Garofalo F, Chen G (2007) Controllability of complex networks via pinning. Phys Rev E 75: 046103.
8. Yu W, Chen G, Lü J (2009) On pinning synchronization of complex dynamical networks. Automatica 45: 429–435.
9. Lombardi A, Hörnquist M (2007) Controllability analysis of networks. Phys Rev E 75: 056110.
10. Rahmani A, Ji M, Mesbahi M, Egerstedt M (2009) Controllability of multi-agent systems from a graph-theoretic perspective. SIAM J Control Optim 48: 162–186.
11. Mesbahi M, Egerstedt M (2010) Graph Theoretic Methods in Multiagent Networks. Princeton: Princeton University Press.
12. Liu YY, Slotine JJ, Barabási AL (2011) Controllability of complex networks. Nature 473: 167–173.
13. Liu YY, Slotine JJ, Barabási AL (2011) Few inputs reprogram biological networks (reply). Nature 478: E4–E5.
14. Egerstedt M (2011) Complex networks: Degrees of control. Nature 473: 158–159.
15. Nepusz T, Vicsek T (2012) Controlling edge dynamics in complex networks. Nature Physics 8: 568–573.
16. Cowan NJ, Chastain EJ, Vilhena DA, Freudenberg JS, Bergstrom CT (2011) Nodal dynamics determine the controllability of complex networks. arXiv: 11062573v3.
17. Wang WX, Ni X, Lai YC, Grebogi C (2012) Optimizing controllability of complex networks by minimum structural perturbations. Phys Rev E 85: 1–5.
18. Kalman RE (1963) Mathematical description of linear dynamical systems. J Soc Indus and Appl Math Ser A 1: 152.
19. Luenberger DG (1979) Introduction to Dynamic Systems: Theory, Models, & Applications. New York: John Wiley & Sons.
20. Slotine JJ, Li W (1991) Applied Nonlinear Control. Prentice-Hall.
21. Sabidussi G (1966) The centrality index of a graph. Psychometrika 31.
22. Freeman L (1977) A set of measures of centrality based upon betweenness. Sociometry 40.
23. Bonacich P (1987) Power and centrality: A family of measures. American Journal of Sociology 92: 1170–1182.
24. Bonacich P, Lloyd P (2001) Eigenvector-like measures of centrality for asymmetric relations. Social Networks 23: 191–201.
25. Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. In: Seventh International World-Wide Web Conference (WWW 1998).
26. Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. J ACM 46: 604–632.
27. Dolev S, Elovici Y, Puzis R (2010) Routing betweenness centrality. J ACM 57: 25: 1–25: 27.
28. Wang B, Gao L, Gao Y (2012) Control range: a controllability-based index for node significance in directed networks. Journal of Statistical Mechanics: Theory and Experiment 2012: P04011.
29. Lin CT (1974) Structural controllability. IEEE Trans Auto Contr 19: 201.
30. Shields RW, Pearson JB (1976) Structural controllability of multi-input linear systems. IEEE Trans Auto Contr 21: 203.
31. Hosoe S (1980) Determination of generic dimensions of controllable subspaces and its application. IEEE Trans Auto Contr 25: 1192.
32. Dion JM, Commault C, van der Woude J (2003) Generic properties and control of linear structured systems: a survey. Automatica 39: 1125–1144.
33. Blackhall L, Hill DJ (2010) On the structural controllability of networks of linear systems. In: 2$^{\text{nd}}$ IFAC Workshop on Distributed Estimation and Control in Networked Systems. 245–250.
34. Poljak S (1990) On the generic dimension of controllable subspaces. IEEE Trans Auto Contr 35: 367.
35. Erdős P, Rényi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5: 17–60.
36. Bollobás B (2001) Random Graphs. Cambridge: Cambridge University Press.
37. Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. Science 296: 910–913.
38. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: Simple building blocks of complex networks. Science 298: 824–827.
39. Yan KK, Fang G, Bhardwaj N, Alexander RP, Gerstein M (2010) Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. Proc Natl Acad Sci USA.
40. Yan G, Zhou T, Hu B, Fu ZQ, Wang BH (2006) Efficient routing on complex networks. Physical Review E 73.
41. Motter AE, Zhou C, Kurths J (2005) Network synchronization, diffusion, and the paradox of heterogeneity. Phys Rev E 71: 016116.
42. Yang R, Zhou T, Xie YB, Lai YC, Wang BH (2008) Optimal contact process on complex networks. Phys Rev E 78: 066109.
43. Harary F (1994) Graph Theory. Westview Press.
44. Albert R, Jeong H, Barabási AL (2000) Error and attack tolerance of complex networks. Nature 406: 378–382.
45. Cohen R, Havlin S, ben Avraham D (2003) Efficient immunization strategies for computer networks and populations. Phys Rev Lett 91: 247901.
46. Christakis NA, Fowler JH (2010) Social network sensors for early detection of contagious outbreaks. PLoS ONE 5: e12948.
47. Feld SL (1991) Why your friends have more friends than you do. Am J Soc 96: 1464.
48. Newman MEJ (2003) Ego-centered networks and the ripple effect. Soc Netw 25: 83.
49. Cross R, Parker A (2004) The Hidden Power of Social Networks. Boston, MA: Harvard Business School Press.
50. Adamic LA, Glance N (2005) The political blogosphere and the 2004 us election. Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem.
51. Eckmann JP, Moses E, Sergi D (2004) Entropy of dialogues creates coherent structures in e-mail traffic. Proc Natl Acad Sci USA 101: 14333.
52. Norlen K, Lucas G, Gebbie M, Chuang J (2002) Eva: Extraction, visualization and analysis of the telecommunications and media ownership network. Proceedings of International Telecommunica- tions Society 14th Biennial Conference,Seoul Korea.
53. Newman MEJ (2002) Assortative mixing in networks. Phys Rev Lett 89: 208701.