



## OPEN

## Control Capacity and A Random Sampling Method in Exploring Controllability of Complex Networks

Tao Jia<sup>1,4,5</sup> & Albert-László Barabási<sup>1,2,3</sup>

<sup>1</sup>Center for Complex Network Research and Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA, <sup>2</sup>Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA, <sup>3</sup>Department of Medicine and Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA, <sup>4</sup>Social Cognitive Networks Academic Research Center, Rensselaer Polytechnic Institute, Troy, NY, 12180 USA, <sup>5</sup>Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY, 12180 USA.

SUBJECT AREAS:  
STATISTICAL PHYSICS  
COMPLEX NETWORKS  
APPLIED PHYSICS  
PHYSICS

Received  
14 May 2013

Accepted  
17 July 2013

Published  
5 August 2013

Correspondence and requests for materials should be addressed to A.-L.B. (barabasi@gmail.com)

Controlling complex systems is a fundamental challenge of network science. Recent advances indicate that control over the system can be achieved through a minimum driver node set (MDS). The existence of multiple MDS's suggests that nodes do not participate in control equally, prompting us to quantify their participations. Here we introduce control capacity quantifying the likelihood that a node is a driver node. To efficiently measure this quantity, we develop a random sampling algorithm. This algorithm not only provides a statistical estimate of the control capacity, but also bridges the gap between multiple microscopic control configurations and macroscopic properties of the network under control. We demonstrate that the possibility of being a driver node decreases with a node's in-degree and is independent of its out-degree. Given the inherent multiplicity of MDS's, our findings offer tools to explore control in various complex systems.

The need to control is ubiquitous in many complex systems. For example, a cellular system controls a series of chemical reactions during its division to guarantee sufficient genetic materials in each daughter cell<sup>1-3</sup>. A company controls the dynamics of information flow for efficient task execution or innovation<sup>4</sup>. In a supply chain, cost is reduced by controlling the commodity flow<sup>5</sup>. Therefore there is an increasing need to understand the control principles of complex systems.

Recent advances in applying control theory<sup>6-8</sup> to complex networks<sup>9,10</sup> shed new light on this problem<sup>11-28</sup>. According to control theory, a dynamical system is controllable if it can be driven from any initial state to any desired final state within finite time<sup>6,7</sup>. Obviously, when we influence every element in the system, we obtain full control. However, control in general can be achieved through the control of only a subset of nodes that we call *driver nodes*. In a linear time-invariant system, the *minimum driver node set* (MDS) can be efficiently identified, representing the minimum set of nodes through which we can yield control over the whole system<sup>12</sup>. It has been shown that the number of driver nodes necessary for control ( $N_D$ ) is fixed in a given network and primarily determined by the underlying degree distribution. Yet, there are often multiple control configurations with the same  $N_D$ <sup>24</sup>. For example, in the six-node network shown in Fig. 1a  $N_D = 4$ , but control can be achieved via five different MDS's: {1, 2, 4, 6}, {1, 2, 4, 5}, {1, 2, 3, 6}, {1, 2, 3, 5} and {1, 2, 3, 4}.

The existence of multiple MDS's indicates that not all nodes participate in control equiprobability, prompting us to quantify the role of each node in control. Here we introduce the concept of *control capacity*  $\phi(i)$ , defined as the fraction of MDS's in which node  $i$  is included. This quantity measures the participation of node  $i$  in MDS's, hence gives the likelihood that node  $i$  is a driver node when the network is under control via a random control configuration. For example, in Fig. 1a, the control capacity of each node is  $\phi(1) = \phi(2) = 1$ ,  $\phi(3) = \phi(4) = 0.6$  and  $\phi(5) = \phi(6) = 0.4$ . In connection with previous work that classifies nodes into three categories<sup>24</sup>, a node with  $\phi = 1$  is critical as it always acts as a driver node,  $\phi = 0$  is redundant as it never participates in MDS's whereas  $0 < \phi < 1$  is intermittent as it plays as a driver node in some control configurations but not all.

In spite of its direct relevance control capacity is difficult to measure, as only nodes with  $\phi = 0$  and  $\phi = 1$  can be identified in polynomial time<sup>24</sup>. Intuitively control capacity is readily obtained once all MDS's are known. However, enumeration of all MDS's in an arbitrary network is in the class of #P problem and computationally

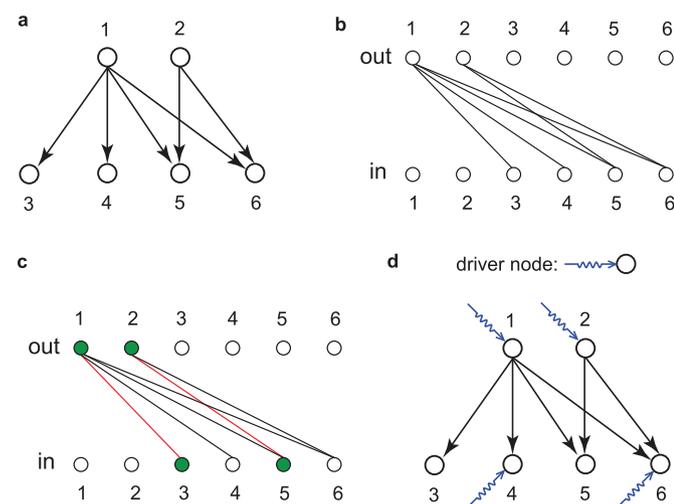


prohibitive for large networks. Indeed, the number of MDS's can grow exponentially with networks size, hence a network with only hundreds of nodes often leads to millions of MDS's<sup>29</sup>. To cope with this difficulty we propose a random sampling algorithm, allowing us to measure the control capacity within a limited number of MDS's, drawn randomly from all MDS's. In the following we show that our algorithm yields a random pick of MDS and provides reliable statistical estimation of control capacity of nodes in arbitrary networks.

## Results

**Random sampling algorithm.** We start by briefly reviewing the process in identifying  $N_D$  and MDS for an arbitrary directed network. First, a directed network is converted to a bipartite graph with two disjoint sets of nodes *out* and *in*. The *out* nodes can be considered as “superiors” that influence others internally. The *in* nodes are “subordinates” that need to be controlled. A directed link from node *i* to *j* corresponds to a connection between node *i* in the *out* set and node *j* in the *in* set in the bipartite graph (Fig. 1a, b). By performing the maximum matching in the bipartite graph<sup>12,30</sup>, the minimum driver nodes are unmatched nodes in the *in* set (Fig. 1c, d).

The method provides a direct connection between a *maximumly-matched set* (MMS) and a MDS, as the complementary set of a MMS yields a MDS. One can use different algorithms to find the maximum matching in a bipartite graph, such as Hopcroft-Karp algorithm<sup>30</sup>, FordFulkerson algorithm<sup>31</sup> and Hungarian algorithm<sup>32</sup>. All these algorithms aim to increase the matching size in each iteration via the augmenting path that starts at a matched node, end on a unmatched node and alternates between unmatched and matched links on the path<sup>17</sup>. Because there is no randomness in identifying an augmenting path, these algorithms will locate only one MMS for a given initial condition hence they are not appropriate for sampling purposes. Two simple modifications can be applied to bring randomness: one is to randomize the initial matching and the other is to randomly choose possible augmenting paths. However, sampling based on these methods are not guaranteed to be uniform among all MMS's and can be typically biased.



**Figure 1** | (a) An example of a directed network with six nodes. (b) The bipartite representation of the directed network in (a) where nodes are represented as two disjoint sets of nodes *out* and *in*. A directed link from node 1 to node 3 in (a) corresponds to a connection between node 1 in the *out* set and node 3 in the *in* set. (c) One maximum matching configuration of the bipartite graph (b) where one node can maximumly match another node through one link. Colored nodes and links are matched nodes and links respectively. (d) One choice of minimum driver node set (MDS) to control the network based on the maximum matching in (c), i.e. controlling nodes 1, 2, 4 and 6 to control the whole system.

Here we propose a novel algorithm that performs unbiased random sampling among all MMS's, which equivalently samples all MDS's and estimates the control capacity. The steps are as follows:

0. For simplicity, remove the always matched nodes in the *in* set and their links (the algorithm to identify always matched nodes is introduced in reference<sup>24</sup>). Denote by *G* the bipartite graph obtained after node removal.

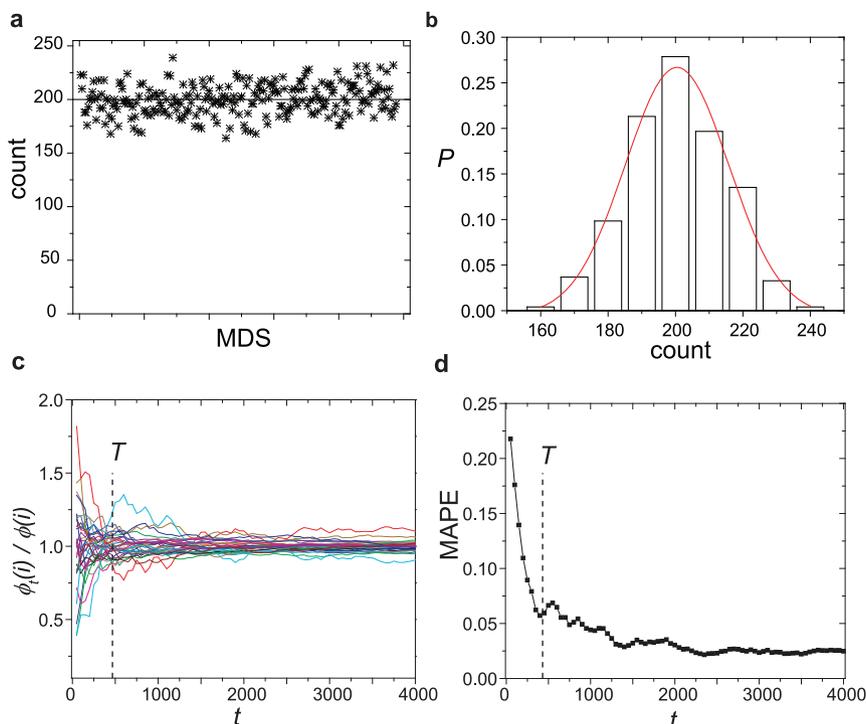
1. Obtain one MMS (denoted by *M*).
2. Randomly pick an element in *M* (denoted by node *i*).
3. Enumerate all alternative MMS's that include all other elements of *M* except node *i*. (see Methods)
4. Randomly pick one of these alternative MMS as the current MMS *M*.
5. Repeat step 2.

Now we prove that the above steps randomly samples MMS's. Considering each MMS as one state, our algorithm maps to a Markov chain characterized by a transition matrix *P* with the element  $p_{i,j}$  that equals the probability of transition from state *i* to *j*. Without losing generality, assume two MMS as sets of *m* nodes  $\{n_1, n_2, \dots, n_a, \dots, n_m\}$  and  $\{n_1, n_2, \dots, n_b, \dots, n_m\}$ , denoted by  $M_1$  and  $M_2$  respectively. Suppose that there are totally *z* other MMS's that include  $n_1, n_2, \dots, n_m$  except  $n_a$  or  $n_b$  and consider the MMS  $M_1$  and  $M_2$  as two state *i* and *j* in the Markov chain that our algorithm maps to. The transition from state *i* to *j* requires the pick of element  $n_a$  out of *m* elements with probability  $1/m$  and the pick of set  $M_2$  out of  $z + 1$  alternative sets with probability  $1/(z + 1)$ . Therefore  $p_{i,j} = \frac{1}{m(z + 1)}$ . Similarly, from state *j* to *i*, the probability to pick element  $n_b$  is  $1/m$ . As the number of alternative MMS's including  $n_1, n_2, \dots, n_N$  except  $n_b$  is also  $z + 1$ ,  $p_{j,i} = \frac{1}{m(z + 1)}$ . Hence  $p_{i,j} = p_{j,i}$  and the transition matrix *P* is symmetric. For Markov chain with symmetric transition matrix, the steady state distribution is with equal probabilities for all states<sup>33</sup>. This means that in the long run, each MMS is picked with the same probability as all others.

To verify the result, we construct a small network with 244 MDS's, perform our sampling algorithm 48,800 iterations and count the time that each MDS is picked. We find that each MDS is sampled approximately 200 times (Fig. 2a), which is the expected count a random sampling yields. The distribution of the counts follows a Gaussian distribution (Fig. 2b) centered at 200, implying the difference between actual and expected counts are due to random fluctuation.

One important attribute of a sampling method is the rate of convergence, capturing how fast the estimate converges to the actual value. Typically the rate of convergence is not known exactly unless an analytical solution can be found for the sampling process<sup>34</sup>. Via numerical tests, we find that the sampling results converges to the actual value after  $T = N \ln N$  iterations in a network with *N* nodes. The interpretation of *T* is intuitive: as our algorithm randomly draws the original elements in the MMS and replaces it with the new ones, the measure will not converge until we have the original MMS completely shuffled. Assuming that the size of the MMS is *m*, the expected number of iterations to obtain the first element replaced is 1, second element replaced is  $m/(m - 1)$  and the  $n^{\text{th}}$  element replaced is  $m/(m - n)$ . Therefore for *m* elements the expected iteration is  $\sum_{i=0}^{m-1} \frac{m}{m-i} \sim m \ln m$ . As *m* varies but is proportional to network size *N*, we replace *m* by *N* and hence have the characteristic time-step *T*.

The characteristic time-step provides the minimum number of iterations necessary for sampling, therefore capturing the complexity in quantifying control capacity. The time needed to find one maximum matching can be as small as  $\mathcal{O}(N^{0.5}L)$  with the Hopcroft-Karp algorithm in a network with *N* nodes and *L* links<sup>30</sup>. To find one alternative MMS with one node replaced, a breadth first search is needed with  $\mathcal{O}(L)$  time. As the number of alternative MMS is capped



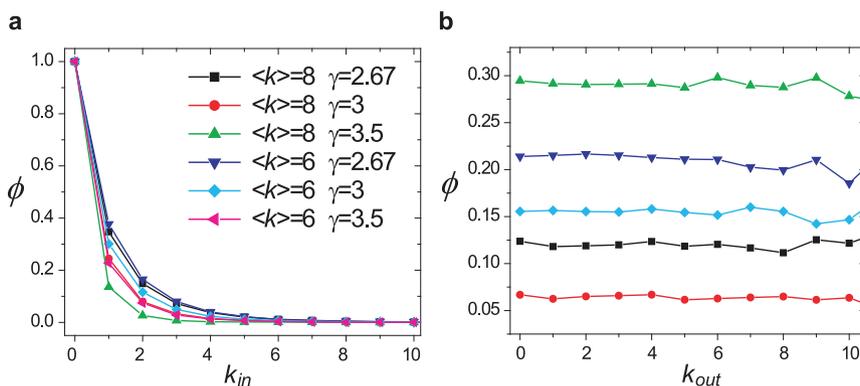
**Figure 2** | (a) The count on each of 244 MDS's is around the expected value 200 when taking 48800 samples. (b) The distribution of the counts that is centered at 200 and can be well fitted by a Gaussian distribution. (c) The time evolution of  $\phi_t(i)/\phi(i)$  of 33 nodes with  $0 < \phi < 1$ .  $\phi_t(i)$  is the control capacity of node  $i$  at time-step  $t$  based on the samples collected up to  $t$  and  $\phi(i)$  is the expected control capacity of node  $i$  that is explicitly measured through the enumeration of all 153,123 MDS's. Control capacity starts to converge at the characteristic time-step  $T$ . (d) The time evolution of the mean absolute percentage error  $\text{MAPE} = \sum_i |\phi_t(i) - \phi(i)|/n$ , where  $n$  is the number of nodes with  $0 < \phi < 1$ . MAPE drops quickly after  $T$  time-steps.

by  $N$ , the complexity of one sampling is  $\mathcal{O}(NL)$ . As the sampling time needed is proportional to  $T$ , the control capacity can be estimated in polynomial time as  $\mathcal{O}(N^2L \ln N)$ .

To test our proposition, we construct a network with 100 nodes. We explicitly enumerate all 153,123 MDS's (see Methods) and exactly measure the control capacity of all 33 nodes with  $0 < \phi < 1$ . Then we apply the random sampling and estimate the capacity  $\phi_t(i)$  of each node  $i$  at time-step  $t$  based on the samples collected up to  $t$ . It is observed that  $\phi_t(i)$  quickly converges to the expected value in  $T = N \ln N$  steps (Fig. 2c,d). It is also noteworthy that in 3000 iterations, which cover fewer than 2% of all MDS's, the mean absolute percentage error (MAPE) of the estimated control capacity is less than 3% (Fig. 2d), indicating the efficiency of utilizing random

sampling in estimating the control capacity that significantly reduces the computational complexity.

**Control capacity in model and real networks.** We check the relationship between a node's topological property and its role in control. On the one hand, we find that a node's out-degree does not affect its control capacity (Fig. 3b). This is because the outgoing links serve as means to control other nodes, which does not affect how this node itself would be controlled. On the other hand, control capacity does depend on in-degree. Particularly  $\phi = 1$  when  $k_{\text{in}} = 0$ , indicating that nodes without incoming links need to be always controlled, in line with our previous finding<sup>24</sup>. As in-degree increases,  $\phi$  decays rapidly (Fig. 3a), indicating that a node with more



**Figure 3** | **Dependency between control capacity  $\phi$  and nodes' in- and out-degree.** (a)  $\phi$  decays rapidly with in-degree  $k_{\text{in}}$ , suggesting that nodes with more incoming links are less likely to be a driver node.  $\phi = 0$  when  $k_{\text{in}} = 0$ , indicating that nodes without incoming links are always driver nodes. (b) For the same networks in (a),  $\phi$  does not vary with out-degree  $k_{\text{out}}$ , indicating control capacity is independent of out-degree. Networks analyzed are generated by static model (see Methods) with size  $N = 10,000$  where  $P(k_{\text{in}}) \sim k_{\text{in}}^{-\gamma_{\text{in}}}$ ,  $P(k_{\text{out}}) \sim k_{\text{out}}^{-\gamma_{\text{out}}}$ ,  $\gamma_{\text{in}} = \gamma_{\text{out}} = \gamma$  and  $\langle k_{\text{in}} \rangle = \langle k_{\text{out}} \rangle = \frac{1}{2} \langle k \rangle$ .



**Table 1 | Real networks Analyzed.** For each network, we show its type, name; number of nodes ( $N$ ) and links ( $L$ ); average degree ( $\langle k \rangle$ ); average degree of driver nodes ( $\langle k_D \rangle$ ) and average in- and out-degree of driver nodes ( $\langle k_{D,in} \rangle$  and  $\langle k_{D,out} \rangle$ ), respectively

Type	Name	$N$	$L$	$\langle k \rangle$	$\langle k_D \rangle$	$\langle k_{D,in} \rangle$	$\langle k_{D,out} \rangle$
Regulatory	TRN-Yeast-1 <sup>47</sup>	4,441	12,873	5.80	5.74	2.82	2.92
	TRN-Yeast-2 <sup>48</sup>	688	1,079	3.14	3.13	1.31	1.82
	TRN-EC-1 <sup>49</sup>	1,550	3,340	4.31	4.30	2.09	2.22
	TRN-EC-2 <sup>48</sup>	418	519	2.48	2.61	1.06	1.55
Trust	College student <sup>50,51</sup>	32	96	6	3.5	0.17	3.33
	Prison inmate <sup>50,51</sup>	67	182	5.43	3.98	1.43	2.55
Food Web	Ythan <sup>52</sup>	135	601	8.90	4.17	0.28	3.89
	Little Rock <sup>53</sup>	183	2,494	27.26	25.87	10.03	15.84
	Grassland <sup>52</sup>	88	137	3.11	2.30	0.29	2.00
	Seagrass <sup>54</sup>	49	226	9.22	6.05	0.55	5.51
Power Grid	TexasPowerGrid <sup>55</sup>	4,889	5,855	2.40	1.9	0.81	1.09
Metabolic	<i>E. coli</i> <sup>56</sup>	2,275	5,763	5.07	3.64	1.76	1.88
	<i>S. cerevisiae</i> <sup>56</sup>	1,511	3,833	5.07	3.71	1.79	1.92
	<i>C. elegans</i> <sup>56</sup>	1,173	2,864	4.88	3.60	1.72	1.88
	s838 <sup>48</sup>	512	819	3.20	2.40	0.87	1.54
Electronic Circuits	s420 <sup>48</sup>	252	399	3.17	2.43	0.84	1.59
	s208 <sup>48</sup>	122	189	3.10	2.49	0.79	1.70
	<i>C. elegans</i> <sup>57</sup>	297	2,345	15.79	8.09	2.69	5.40
Neuronal Internet	p2p-1 <sup>58</sup>	6301	20777	6.59	5.48	2.29	3.19
	p2p-2 <sup>58</sup>	8,846	31,839	7.20	6.27	2.73	3.53
	p2p-3 <sup>58</sup>	8,717	31,525	7.23	6.28	2.79	3.49
Social Communication	UClonline <sup>59</sup>	1,899	20,296	21.38	5.89	3.37	2.52
	Email-epoch <sup>60</sup>	3,188	39,256	24.63	7.13	2.70	4.43

incoming links are less likely to be a driver node as they are more likely to be influenced internally.

We extend the analysis to real systems and check the relationship between  $\langle k_D \rangle$  and  $\langle k \rangle$ , which are the average degree of the driver nodes and all nodes, respectively (Table 1). With control capacity,  $\langle k_D \rangle$  can be explicitly expressed as  $\langle k_D \rangle = \sum_i \phi(i)k(i)/N_D$  where  $k(i)$  is the degree of an arbitrary node  $i$ . One would expect  $\langle k_D \rangle < \langle k \rangle$  in real networks since  $\phi$  decreases with  $k_{in}$ . However, while the fact that the average in-degree of driver nodes is less than that of the network ( $\langle k_{D,in} \rangle < \langle k_{in} \rangle$ ) reflects the relationship between  $\phi$  and  $k_{in}$ , average out-degree of driver nodes ( $\langle k_{D,out} \rangle$ ) can be affected by the in- and out-degree correlation<sup>21,35–38</sup> featuring in real systems and the finite size effect<sup>39,40</sup>. Indeed several networks are found with  $\langle k_{D,out} \rangle > \langle k_{out} \rangle$  (e.g. Seagrass in food web and TRN-Yeast-2 in regulatory networks of Table 1) and in one network we even observe that  $\langle k_D \rangle$  is slightly higher than  $\langle k \rangle$  (TRN-EC-2 in regulatory networks of Table 1). But for majority of real networks the average degree of driver nodes are less than that of the whole network, leading to the conclusion that hubs are less likely to be driver nodes<sup>12</sup>.

Finally, we check how control capacity is distributed among nodes with  $0 < \phi < 1$  in Erdős-Rényi networks<sup>41</sup>, scale-free networks<sup>42</sup> (see Methods) and some real networks (Fig. 4). The distributions are found to depend on specific network configurations and there seems no simple universal function for the distribution. But as a common feature,  $\phi$  typically displays multi-modal distribution, implying that several clusters of nodes share about the same chance of being driver nodes. Recent work<sup>24</sup> discovered that dense networks with identical degree distribution can stay in one of the two control modes, centralized or distributed, depending on the fraction of nodes that can participate in MDS's. The distributions of control capacity corresponding to networks in the two control modes also show distinct features. For networks in distributed mode, a significant fraction of nodes are with control capacity close to zero (Fig. 4(b), (e)). This indicates that while many nodes can participate in MDS's, their participation is not frequent compared with the huge number of MDS's in distributed mode. For networks in centralized mode, the number of nodes with  $0 < \phi < 1$  is small, but the distribution of

capacity among these nodes is similar to that when  $\langle k \rangle$  is small (Fig. 4(c), (f)).

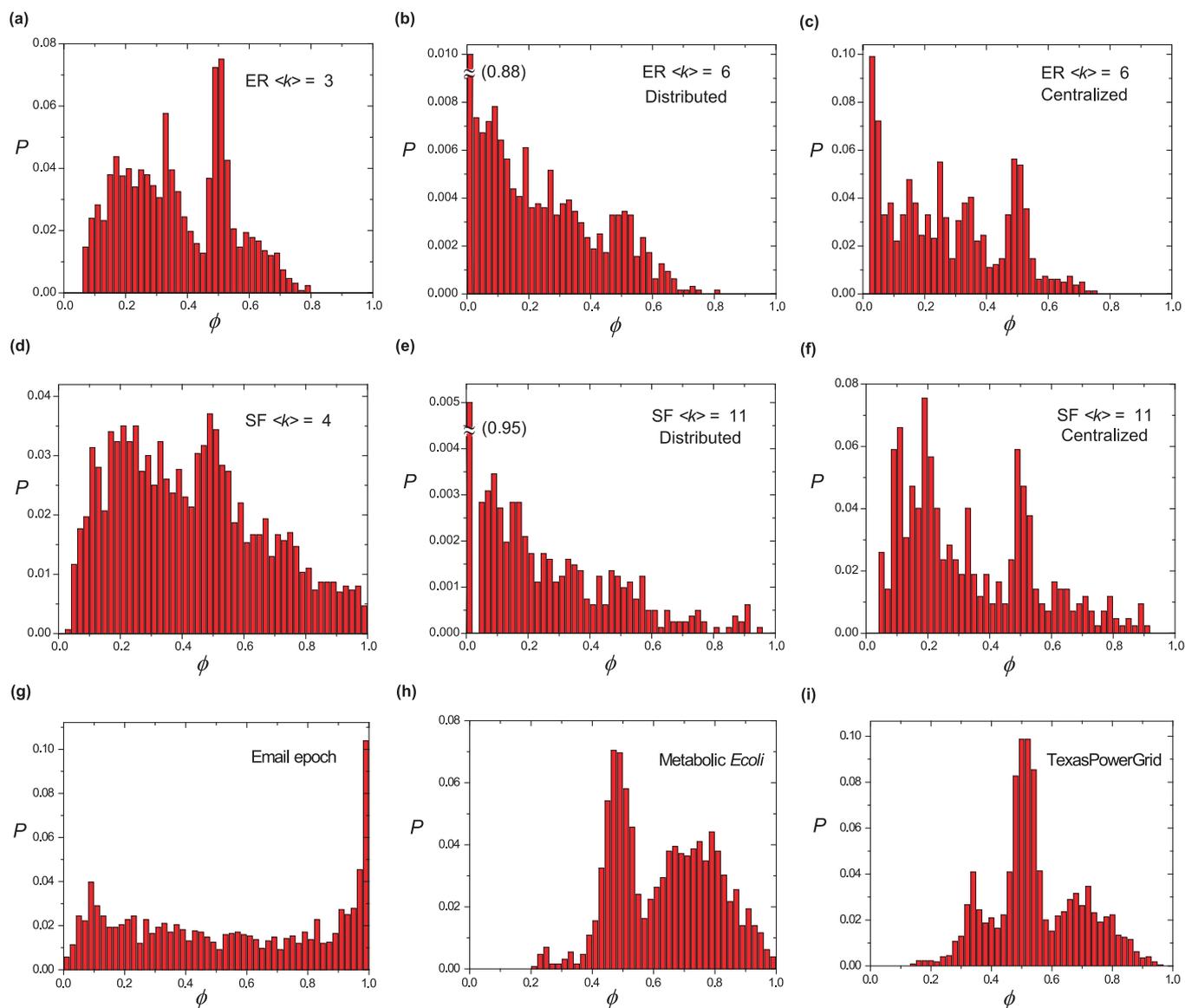
## Discussion

In summary, uncovering the role of individual nodes in controlling a network requires us to understand control capacity, a centrality measure quantifying a node's likelihood of being a driver node. While a network's control can be achieved via different MDS's and each may give rise to different outcomes, we lack a tool to average the effect of different MDS's or statistically analyze the consequences driven by different MDS's over the network. In this paper we propose a random sampling algorithm, allowing us to efficiently measure control capacity in arbitrary networks. The proposed algorithm bridges the gap between multiple microscopic control configurations and macroscopic properties of the network under control. One important example of its application is the study of  $\langle k_D \rangle$ , which can not be properly addressed without the random sampling method<sup>12</sup>.

The results presented have many potential applications in future works. For example, recent work on the controllability of bank systems investigated the time correlation of nodes' roles in control<sup>25,27</sup>. The measure of control capacity could be crucial in such tasks, especially in temporal networks where a node's role in control varies with time<sup>43–45</sup>. The relationship between control capacity and the efficiency or energy cost in control<sup>15,28</sup> are also important issues for further investigations. The random sampling method is useful in problems when an overall measure of a network is needed. As an example, when estimating the control robustness of a network, the random sampling algorithm has to be considered as different MDS's may facing different failure risks. Finally, links do not participate in control in an equal manner, allowing multiple link combinations to spread the control signal. Our approach can offer insights for future work exploring the participation of links in control. Given the inherent multiplicity feature in control, our findings offer fundamental tools to explore control in various complex systems.

## Methods

**Enumerating all alternative MMS's with one node replaced.** Suppose the maximum matching is obtained in a bipartite graph and denote  $M$  by the current maximum-matched set (MMS) of nodes in the  $in$  set. Assume node  $i$  is an element of the set  $M$ .



**Figure 4** | Distribution of control capacity  $\phi$  among nodes with  $0 < \phi < 1$  in Erdős-Rényi networks ((a)–(c)), scale-free networks ((d)–(f)) and real networks ((g)–(i)). Dense networks with identical degree distribution in centralized ((c) and (f)) and distributed ((b) and (e)) control modes are labeled. For Erdős-Rényi and scale-free networks, network size  $N = 10,000$ . For scale-free networks,  $\gamma_{in} = \gamma_{out} = 3$ . The information for the three real networks is listed in Table 1.

The following procedures can provide all MMS's that contain all other elements of  $M$  except node  $i$ . (0) Set node  $i$  as the removal node. (1) Identify the node in the  $out$  set that matches the removal node (denoted by node  $j$ ). (2) Keep the current matched nodes and links unchanged, remove the removal node with all its links. (3) Check if there is an augmenting path that starts from node  $j$ , ends at an unmatched node and alternates between unmatched and matched links on the path. (4) If so, we obtain a new MMS with node  $i$  replaced. Update the matched links and nodes correspondingly. Set the new matched node in the  $in$  set as the removal node and repeat step (1). (5) If not, there is no new MMS with node  $i$  replaced and all of them are enumerated already.

**Enumerating all MDS's.** For a given bipartite graph, we first remove all the always matched nodes in the  $in$  set and their links (algorithm discussed in reference<sup>24</sup>). Define  $S_i$  as a set of nodes that a  $out$  node  $i$  can reach. For example, in Fig. 1b there are  $S_1 = \{3, 4, 5, 6\}$  and  $S_2 = \{5, 6\}$ . Effectively  $S_i$  is the set of nodes that node  $i$  can match. In the bipartite graph with no always matched nodes in the  $in$  set, a MMS of  $in$  nodes is a set of nodes without duplication, each drawn from one set  $S$ . Therefore, we can repeatedly test all possible combinations to enumerate all MMS's that equivalently provides all MDS's. Note that nodes chosen from different  $S$ 's can sometimes give rise to the same MMS. For example, in Fig. 1b picking node 5 from  $S_1$  and node 6 from  $S_2$  yields the same MMS as picking node 6 from  $S_1$  and node 5 from  $S_2$ . All MMS's need to be recorded. Once a valid node combination is found, it needs to be checked with previously found MMS to avoid double count.

**Generating a scale free network.** The scale-free networks<sup>42</sup> analyzed are generated via the static model<sup>46</sup>. We start from  $N$  disconnected nodes indexed by integer number  $i$  ( $i = 1, \dots, N$ ). The weight  $w_i^{out,in} = i^{-\gamma_{out,in}}$  is assigned to each node in the  $out$  and the  $in$  set, with  $\gamma_{out,in}$  a real number in the range  $[0, 1)$ . Randomly selected two nodes  $i$  and  $j$  respectively from the  $out$  set and the  $in$  set, with probability proportional to  $w_i^{out}$  and  $w_j^{in}$ . Connect node  $i$  and  $j$  if there is no connection between them, corresponding to a directed link from node  $i$  to node  $j$  in the digraph. Otherwise randomly choose another pair. Repeated the procedure until  $\langle k_{in} \rangle = \langle k_{out} \rangle = \langle k \rangle / 2$  links are created. The degree distribution under this construction is  $P_{out,in}(k) \sim k^{-\left(1 + \frac{1}{\gamma_{out,in}}\right)} = k^{-\gamma_{out,in}}$  in the large  $k$  limit.

- Warren, G. Membrane partitioning during cell division. *Annual review of biochemistry* **62**, 323–348 (1993).
- Birky, C. W. Jr. *et al.* The partitioning of cytoplasmic organelles at cell division. *International review of cytology. Supplement* **15**, 49 (1983).
- Huh, D. & Paulsson, J. Random partitioning of molecules at cell division. *Proceedings of the National Academy of Sciences* **108**, 15004–15009 (2011).
- Georgakopoulos, D., Hornick, M. & Sheth, A. An overview of workflow management: from process modeling to workflow automation infrastructure. *Distributed and parallel Databases* **3**, 119–153 (1995).
- Beamon, B. M. Supply chain design and analysis: Models and methods. *International journal of production economics* **55**, 281–294 (1998).



6. Kalman, R. E. Mathematical description of linear dynamical systems. *J. Soc. Indust. and Appl. Math. Ser. A* **1**, 152–192 (1963).
7. Slotine, J.-J. & Li, W. *Applied Nonlinear Control* (Prentice-Hall, 1991).
8. Lin, C.-T. Structural controllability. *IEEE Trans. Auto. Contr.* **19**, 201–208 (1974).
9. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics* **74**, 47–97 (2002).
10. Newman, M. E. The structure and function of complex networks. *SIAM review* **45**, 167–256 (2003).
11. Wang, X. F. & Chen, G. Pinning control of scale-free dynamical networks. *Physica A: Statistical Mechanics and its Applications* **310**, 521–531 (2002).
12. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
13. Rajapakse, I., Groudine, M. & Mesbahi, M. Dynamics and control of state-dependent networks for probing genomic organization. *Proceedings of the National Academy of Sciences* **108**, 17257–17262 (2011).
14. Nepusz, T. & Vicsek, T. Controlling edge dynamics in complex networks. *Nature Physics* **8**, 568–573 (2012).
15. Yan, G., Ren, J., Lai, Y.-C., Lai, C.-H. & Li, B. Controlling complex networks: How much energy is needed? *Physical Review Letters* **108**, 218703 (2012).
16. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Control centrality and hierarchical structure in complex networks. *PLoS ONE* **7**, e44459 (2012).
17. Wang, W.-X., Ni, X., Lai, Y.-C. & Grebogi, C. Optimizing controllability of complex networks by minimum structural perturbations. *Physical Review E* **85**, 026115 (2012).
18. Tang, Y., Gao, H., Zou, W. & Kurths, J. Identifying controlling nodes in neuronal networks in different scales. *PLoS one* **7**, e41375 (2012).
19. Cowan, N. J., Chastain, E. J., Vilhena, D. A., Freudenberg, J. S. & Bergstrom, C. T. Nodal dynamics, not degree distributions, determine the structural controllability of complex networks. *PLoS one* **7**, e38398 (2012).
20. Gutiérrez, R., Sendiña-Nadal, I., Zanin, M., Papo, D. & Boccaletti, S. Targeting the dynamics of complex networks. *Scientific reports* **2**, 396 (2012).
21. Pósfai, M., Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Effect of correlations on network controllability. *Scientific Reports* **3**, 1067 (2013).
22. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Observability of complex systems. *Proceedings of the National Academy of Sciences* **110**, 2460–2465 (2013).
23. Nacher, J. C. & Akutsu, T. Structural controllability of unidirectional bipartite networks. *Scientific Reports* **3**, 1647 (2013).
24. Jia, T. *et al.* Emergence of bimodality in controlling complex networks. *Nature Communications* **4**, (2013).
25. Delpini, D. *et al.* Evolution of Controllability in Interbank Networks. *Scientific reports* **3**, (2013) doi: 10.1038/srep01626.
26. Caldarelli, G., Chessa, A., Pammolli, F., Gabrielli, A. & Puliga, M. Reconstructing a credit network. *Nature Physics* **9**, 125–126 (2013).
27. Galbiati, M., Delpini, D. & Battiston, S. The power to control. *Nature Physics* **9**, 126–128 (2013).
28. Sun, J. & Motter, A. E. Controllability transition and nonlocality in network control. *Physical Review Letters* **110**, 208701 (2013).
29. Zdeborová, L. & Mézard, M. The number of matchings in random graphs. *Journal of Statistical Mechanics: Theory and Experiment* **05**, P05003 (2006).
30. Hopcroft, J. E. & Karp, R. M. An  $n^{5/2}$  algorithm for maximum matchings in bipartite graphs. *SIAM Journal on computing* **2**, 225–231 (1973).
31. Ford, L. R. & Fulkerson, D. R. Maximal flow through a network. *Canadian Journal of Mathematics* **8**, 399–404 (1956).
32. Kuhn, H. W. The Hungarian method for the assignment problem. *Naval research logistics quarterly* **2**, 83–97 (1955).
33. Norris, J. R. *Markov chains*. 2008 (Cambridge university press, 1998).
34. Bishop, C. M. *et al.* *Pattern recognition and machine learning*, vol. 1, (Springer New York, 2006).
35. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A.-L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
36. Pastor-Satorras, R., Vázquez, A. & Vespignani, A. Dynamical and correlation properties of the internet. *Physical Review Letters* **87**, 258701 (2001).
37. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
38. Lee, J.-S., Goh, K.-I., Kahng, B. & Kim, D. Intrinsic degree-correlations in the static model of scale-free networks. *The European Physical Journal B-Condensed Matter and Complex Systems* **49**, 231–238 (2006).
39. Boguná, M., Pastor-Satorras, R. & Vespignani, A. Cut-offs and finite size effects in scale-free networks. *The European Physical Journal B-Condensed Matter and Complex Systems* **38**, 205–209 (2004).
40. Lü, L., Zhang, Z.-K. & Zhou, T. Zipf's law leads to Heaps' law: analyzing their relation in finite-size systems. *PLoS one* **5**, e14139 (2010).
41. Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–60 (1960).
42. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
43. Holme, P. & Saramäki, J. Temporal networks. *Physics reports* **519**, 97–125.
44. Perra, N., Gonçalves, B., Pastor-Satorras, R. & Vespignani, A. Activity driven modeling of time varying networks. *Scientific reports* **2** (2012) doi: 10.1038/srep00469.
45. Liu, S.-Y., Baronchelli, A. & Perra, N. Contagion dynamics in time-varying metapopulation networks. *Physical Review E* **87**, 032805 (2013).
46. Goh, K.-I., Kahng, B. & Kim, D. Universal behavior of load distribution in scale-free networks. *Physical Review Letters* **87**, 278701 (2001).
47. Balaji, S., Madan Babu, M., Iyer, L., Luscombe, N. & Aravind, L. Principles of combinatorial regulation in the transcriptional regulatory network of yeast. *Journal of molecular biology* **360**, 213–227 (2006).
48. Milo, R. *et al.* Network motifs: Simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
49. Gama-Castro, S. *et al.* RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Research* **36**, D120–D124 (2008).
50. Van Duijn, M. A. J., Huisman, M., Stokman, F. N., Wasseur, F. W. & Zeggelink, E. P. H. Evolution of sociology freshmen into a friendship network. *Journal of Mathematical Sociology* **27**, 153–191 (2003).
51. Milo, R. *et al.* Superfamilies of designed and evolved networks. *Science* **303**, 1538–1542 (2004).
52. Dunne, J. A., Williams, R. J. & Martinez, N. D. Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences* **99**, 12917–12922 (2002).
53. Martinez, N. Artifacts or attributes? effects of resolution on the little rock lake food web. *Ecological Monographs* **61**, 367–392 (1991).
54. Christian, R. R. & Luczkovich, J. J. Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. *Ecological Modelling* **117**, 99–124 (1999).
55. Bianconi, G., Gulbahce, N. & Motter, A. E. Local structure of directed networks. *Physical Review Letters* **100**, 118701 (2008).
56. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
57. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
58. Leskovec, J., Kleinberg, J. & Faloutsos, C. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **1**, 2 (2007).
59. Opsahl, T. & Panzarasa, P. Clustering in weighted networks. *Social Networks* **31**, 155 (2009).
60. Eckmann, J.-P., Moses, E. & Sergi, D. Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences* **101**, 14333–14337 (2004).

## Acknowledgements

We thank Y.-Y. Liu, M. Pósfai, E. Csóka, J.-J. Slotine, C. Song and D. Wang for discussions. This work was supported by the Network Science Collaborative Technology Alliance sponsored by the US Army Research Laboratory under Agreement Number W911NF-09-2-0053; the Defense Advanced Research Projects Agency under Agreement Number 11645021; the Defense Threat Reduction Agency award WMD BRBAA07-J-2-0035; and through the generous support of Lockheed Martin.

## Author contributions

T.J. and A.-L.B. designed the research. T.J. performed numerical simulation and analyzed the empirical data. T.J. and A.-L.B. prepared the manuscript.

## Additional information

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

**How to cite this article:** Jia, T. & Barabási, A.-L. Control Capacity and A Random Sampling Method in Exploring Controllability of Complex Networks. *Sci. Rep.* **3**, 2354; DOI:10.1038/srep02354 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0>