ARTICLE

doi:10.1038/nature10011

Controllability of complex networks

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

The ultimate proof of our understanding of natural or technological systems is reflected in our ability to control them. Although control theory offers mathematical tools for steering engineered and natural systems towards a desired state, a framework to control complex self-organized systems is lacking. Here we develop analytical tools to study the controllability of an arbitrary complex directed network, identifying the set of driver nodes with time-dependent control that can guide the system's entire dynamics. We apply these tools to several real networks, finding that the number of driver nodes is determined mainly by the network's degree distribution. We show that sparse inhomogeneous networks, which emerge in many real complex systems, are the most difficult to control, but that dense and homogeneous networks can be controlled using a few driver nodes. Counterintuitively, we find that in both model and real systems the driver nodes tend to avoid the high-degree nodes.

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 within finite time[1–3]. This definition agrees with our intuitive notion of control, capturing an ability to guide a system's behaviour towards a desired state through the appropriate manipulation of a few input variables, like a driver prompting a car to move with the desired speed and in the desired direction by manipulating the pedals and the steering wheel. Although control theory is a mathematically highly developed branch of engineering with applications to electric circuits, manufacturing processes, communication systems[4–6], aircraft, spacecraft and robots[2,3], fundamental questions pertaining to the controllability of complex systems emerging in nature and engineering have resisted advances. The difficulty is rooted in the fact that two independent factors contribute to controllability, each with its own layer of unknown: (1) the system's architecture, represented by the network encapsulating which components interact with each other; and (2) the dynamical rules that capture the time-dependent interactions between the components. Thus, progress has been possible only in systems where both layers are well mapped, such as the control of synchronized networks[7–10], small biological circuits[11] and rate control for communication networks[4–6]. Recent advances towards quantifying the topological characteristics of complex networks[12–16] have shed light on factor (1), prompting us to wonder whether some networks are easier to control than others and how network topology affects a system's controllability. Despite some pioneering conceptual work[17–23] (Supplementary Information, section II), we continue to lack general answers to these questions for large weighted and directed networks, which most commonly emerge in complex systems.

Network controllability

Most real systems are driven by nonlinear processes, but the controllability of nonlinear systems is in many aspects structurally similar to that of linear systems[3], prompting us to start our study using the canonical linear, time-invariant dynamics

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

where the vector $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))^{\mathrm{T}}$ captures the state of a system of N nodes at time t. For example, $x_i(t)$ can denote the amount of traffic that passes through a node i in a communication network[24] or transcription factor concentration in a gene regulatory network[25]. The $N \times N$ matrix A describes the system's wiring diagram and the interaction strength between the components, for example the traffic on individual communication links or the strength of a regulatory interaction. Finally, B is the $N \times M$ input matrix ($M \leq N$) that identifies the nodes controlled by an outside controller. The system is controlled using the time-dependent input vector $\mathbf{u}(t) = (u_1(t), \ldots, u_M(t))^{\mathrm{T}}$ imposed by the controller (Fig. 1a), where in general the same signal $u_i(t)$ can drive multiple nodes. If we wish to control a system, we first need to identify the set of nodes that, if driven by different signals, can offer full control over the network. We will call these 'driver nodes'. We are particularly interested in identifying the minimum number of driver nodes, denoted by N_{D}, whose control is sufficient to fully control the system's dynamics.

The system described by equation (1) is said to be controllable if it can be driven from any initial state to any desired final state in finite time, which is possible if and only if the $N \times NM$ controllability matrix

$$C = (B, AB, A^2B, \ldots, A^{N-1}B) \qquad (2)$$

has full rank, that is

$$\mathrm{rank}(C) = N \qquad (3)$$

This represents the mathematical condition for controllability, and is called Kalman's controllability rank condition[1,2] (Fig. 1a). In practical terms, controllability can be also posed as follows. Identify the minimum number of driver nodes such that equation (3) is satisfied. For example, equation (3) predicts that controlling node x_1 in Fig. 1b with the input signal u_1 offers full control over the system, as the states of nodes x_1, x_2, x_3 and x_4 are uniquely determined by the signal $u_1(t)$ (Fig. 1c). In contrast, controlling the top node in Fig. 1e is not sufficient for full control, as the difference $a_{31}x_2(t) - a_{21}x_3(t)$ (where a_{ij} are the elements of A) is not uniquely determined by $u_1(t)$ (see Fig. 1f and Supplementary Information section III.A). To gain full control, we must simultaneously control node x_1 and any two nodes among $\{x_2, x_3, x_4\}$ (see Fig. 1h, i for a more complex example).

To apply equations (2) and (3) to an arbitrary network, we need to know the weight of each link (that is, the a_{ij}), which for most real

[1]Center for Complex Network Research and Departments of Physics, Computer Science and Biology, Northeastern University, Boston, Massachusetts 02115, USA. [2]Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [3]Nonlinear Systems Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Department of Mechanical Engineering and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

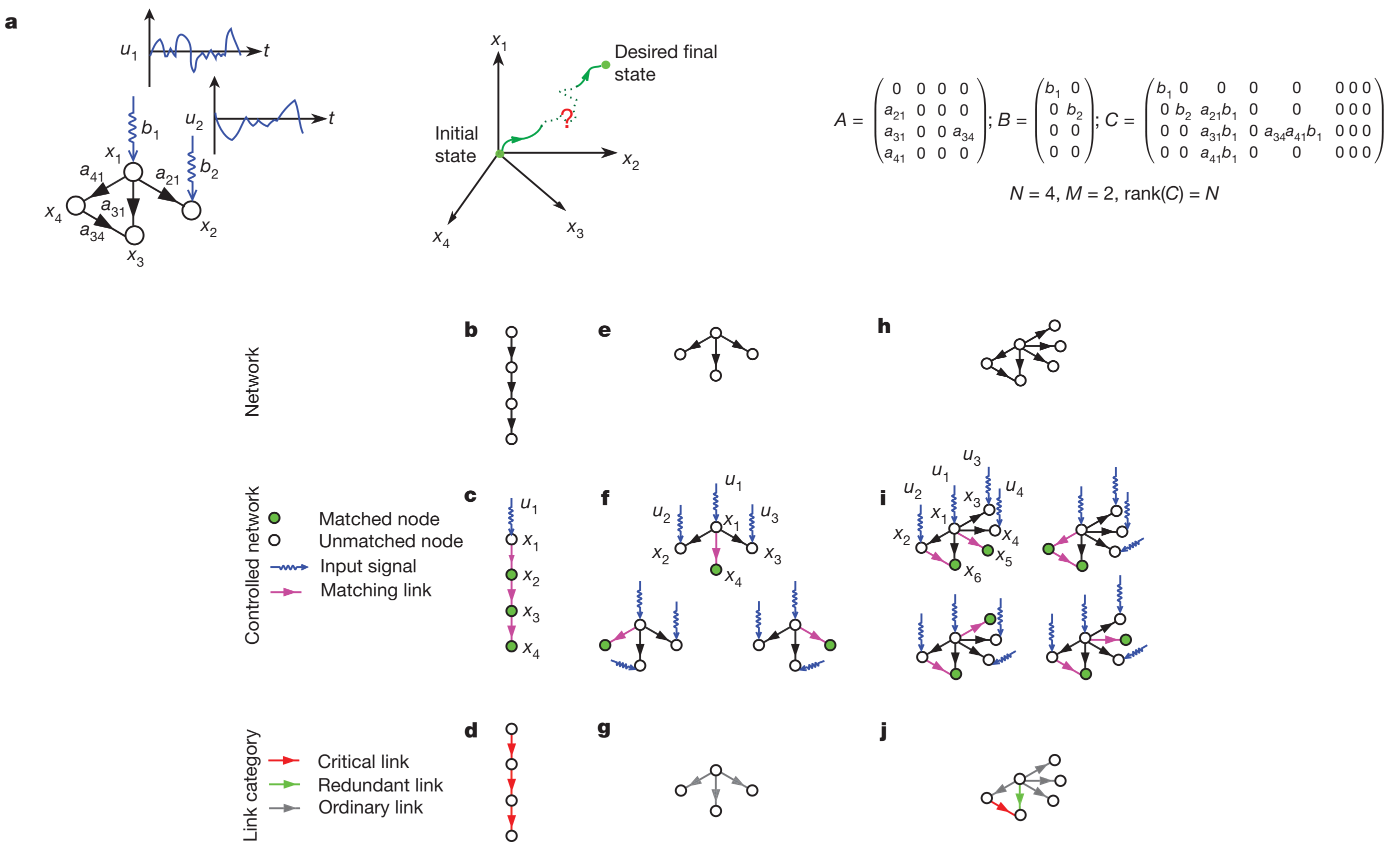

Figure 1 | Controlling a simple network. **a**, The small network can be controlled by an input vector $\mathbf{u} = (u_1(t), u_2(t))^{\mathrm{T}}$ (left), allowing us to move it from its initial state to some desired final state in the state space (right). Equation (2) provides the controllability matrix (C), which in this case has full rank, indicating that the system is controllable. **b**, Simple model network: a directed path. **c**, Maximum matching of the directed path. Matching edges are shown in purple, matched nodes are green and unmatched nodes are white. The unique maximum matching includes all links, as none of them share a common starting or ending node. Only the top node is unmatched, so controlling it yields full control of the directed path ($N_{\mathrm{D}} = 1$). **d**, In the directed path shown in **b**, all links are critical, that is, their removal eliminates our ability to control the network. **e**, Small model network: the directed star. **f**, Maximum matchings of the directed star. Only one link can be part of the maximum matching, which yields three unmatched nodes ($N_{\mathrm{D}} = 3$). The three different maximum matchings indicate that three distinct node configurations can exert full control. **g**, In a directed star, all links are ordinary, that is, their removal can eliminate some control configurations but the network could be controlled in their absence with the same number of driver nodes N_{D}. **h**, Small example network. **i**, Only two links can be part of a maximum matching for the network in **h**, yielding four unmatched nodes ($N_{\mathrm{D}} = 4$). There are all together four different maximum matchings for this network. **j**, The network has one critical link, one redundant link (which can be removed without affecting any control configuration) and four ordinary links.

networks are either unknown (for example regulatory networks) or are known only approximately and are time dependent (for example Internet traffic). Even if all weights are known, a brute-force search requires us to compute the rank of C for $2^N - 1$ distinct combinations, which is a computationally prohibitive task for large networks. To bypass the need to measure the link weights, we note that the system (A, B) is 'structurally controllable'[26] if it is possible to choose the non-zero weights in A and B such that the system satisfies equation (3). A structurally controllable system can be shown to be controllable for almost all weight combinations, except for some pathological cases with zero measure that occur when the system parameters satisfy certain accidental constraints[26,27]. Thus, structural controllability helps us to overcome our inherently incomplete knowledge of the link weights in A. Furthermore, because structural controllability implies controllability of a continuum of linearized systems[28], our results can also provide a sufficient condition for controllability for most non-linear systems[3] (Supplementary Information, section III.A).

To avoid the brute-force search for driver nodes, we proved that the minimum number of inputs or driver nodes needed to maintain full control of the network is determined by the 'maximum matching' in the network, that is, the maximum set of links that do not share start or end nodes (Fig. 1c, f, i). A node is said to be matched if a link in the maximum matching points at it; otherwise it is unmatched. As we show in the Supplementary Information, the structural controllability problem maps into an equivalent geometrical problem on a network: we can gain full control over a directed network if and only if we directly control each unmatched node and there are directed paths from the input signals to all matched nodes[29]. The possibility of determining N_{D}, using this mapping, is our first main result. As the maximum matching in directed networks can be identified numerically in at most $O(N^{1/2}L)$ steps[30], where L denotes the number of links, the mapping offers an efficient method to determine the driver nodes for an arbitrary directed network.

Controllability of real networks

We used the tools developed above to explore the controllability of several real networks. The networks were chosen for their diversity: for example, the purpose of the gene regulatory network is to control the dynamics of cellular processes, so it is expected to evolve towards a structure that is efficient from a control perspective, potentially implying a small number of driver nodes (that is, small $n_{\mathrm{D}} \equiv N_{\mathrm{D}}/N$). In contrast, for the World Wide Web or citation networks controllability has no known role, making it difficult even to guess n_{D}. Finally, it might be argued that social networks, given their perceived neutrality (or even resistance) to control, should have a high n_{D}, as it is necessary to control most individuals separately to control the whole system. We used the mapping into maximum matching to determine the minimum set of driver nodes (N_{D}) for the networks in Table 1, the obtained trend defying our expectations: as a group, gene regulatory networks display high n_D ($\sim$0.8), indicating that it is necessary to independently control about 80% of nodes to control them fully. In contrast, several social networks are characterized by some of the smallest n_D values, suggesting that a few individuals could in principle control the whole system.

Given the important role hubs (nodes with high degree) have in maintaining the structural integrity of networks against failures and attacks[31,32], in spreading phenomena[32,33] and in synchronization[8,34], it is natural to expect that control of the hubs is essential to control a network. To test the validity of this hypothesis, we divided the nodes into three groups of equal size according to their degree, k (low, medium and high). As Fig. 2a, b shows for two canonical network models (Erdős–Rényi[35,36] and scale-free[15,37–39]), the fraction of driver nodes is significantly higher among low-k nodes than among the hubs. In Fig. 2c, we plot the mean degree of the driver nodes, $\langle k_D \rangle$, as a function of the mean degree, $\langle k \rangle$, of each network in Table 1 and several network models. In all cases, $\langle k_D \rangle$ is either significantly smaller than or comparable to $\langle k \rangle$, indicating that in both real and model systems the driver nodes tend to avoid the hubs.

To identify the topological features that determine network controllability, we randomized each real network using a full randomization procedure (rand-ER) that turns the network into a directed Erdős–Rényi random network with N and L unchanged. For several networks there is no correlation between the N_D of the original network and the N_D of its randomized counterpart (Fig. 2d), indicating that full randomization eliminates the topological characteristics that influence controllability. We also applied a degree-preserving randomization[40,41] (rand-Degree), which keeps the in-degree, k_{in}, and out-degree, k_{out}, of each node unchanged but selects randomly the nodes that link to each other. We find that this procedure does not alter N_D significantly, despite the observed differences in N_D of six orders of magnitude (Fig. 2e). Thus, a system's controllability is to a great extent encoded by the underlying network's degree distribution, $P(k_{in}, k_{out})$, which is our second and most important finding. It indicates that N_D is determined mainly by the number of incoming and outgoing links each node has and is independent of where those links point.

An analytical approach to controllability

The importance of the degree distribution allows us to determine N_D analytically for a network with an arbitrary $P(k_{in}, k_{out})$. Using the cavity method[42–44], we derived a set of self-consistent equations (Supplementary Information, section IV) whose input is the degree distribution and whose solution is the average n_D (or N_D) over all network realizations compatible with $P(k_{in}, k_{out})$, which is our third key result. As Fig. 2f shows, the analytically predicted N_D agrees perfectly with $N_D^{\text{rand-Degree}}$ (and hence is in good agreement with the exact value, N_D^{real}), offering an effective analytical tool to study

Table 1 | The characteristics of the real networks analysed in the paper

Type	Name	N	L	n_D^{real}	$n_D^{\text{rand-Degree}}$	$n_D^{\text{rand-ER}}$
Regulatory	TRN-Yeast-1	4,441	12,873	0.965	0.965	0.083
	TRN-Yeast-2	688	1,079	0.821	0.811	0.303
	TRN-EC-1	1,550	3,340	0.891	0.891	0.188
	TRN-EC-2	418	519	0.751	0.752	0.380
	Ownership-USCorp	7,253	6,726	0.820	0.815	0.480
Trust	College student	32	96	0.188	0.173	0.082
	Prison inmate	67	182	0.134	0.144	0.103
	Slashdot	82,168	948,464	0.045	0.278	1.7×10^{-5}
	WikiVote	7,115	103,689	0.666	0.666	1.4×10^{-4}
	Epinions	75,888	508,837	0.549	0.606	0.001
Food web	Ythan	135	601	0.511	0.433	0.016
	Little Rock	183	2,494	0.541	0.200	0.005
	Grassland	88	137	0.523	0.477	0.301
	Seagrass	49	226	0.265	0.199	0.203
Power grid	Texas	4,889	5,855	0.325	0.287	0.396
Metabolic	*Escherichia coli*	2,275	5,763	0.382	0.218	0.129
	Saccharomyces cerevisiae	1,511	3,833	0.329	0.207	0.130
	Caenorhabditis elegans	1,173	2,864	0.302	0.201	0.144
Electronic circuits	s838	512	819	0.232	0.194	0.293
	s420	252	399	0.234	0.195	0.298
	s208	122	189	0.238	0.199	0.301
Neuronal	*Caenorhabditis elegans*	297	2,345	0.165	0.098	0.003
Citation	ArXiv-HepTh	27,770	352,807	0.216	0.199	3.6×10^{-5}
	ArXiv-HepPh	34,546	421,578	0.232	0.208	3.0×10^{-5}
World Wide Web	nd.edu	325,729	1,497,134	0.677	0.622	0.012
	stanford.edu	281,903	2,312,497	0.317	0.258	3.0×10^{-4}
	Political blogs	1,224	19,025	0.356	0.285	8.0×10^{-4}
Internet	p2p-1	10,876	39,994	0.552	0.551	0.001
	p2p-2	8,846	31,839	0.578	0.569	0.002
	p2p-3	8,717	31,525	0.577	0.574	0.002
Social communication	UClonline	1,899	20,296	0.323	0.322	0.706
	Email-epoch	3,188	39,256	0.426	0.332	3.0×10^{-4}
	Cellphone	36,595	91,826	0.204	0.212	0.133
Intra-organizational	Freemans-2	34	830	0.029	0.029	0.029
	Freemans-1	34	695	0.029	0.029	0.029
	Manufacturing	77	2,228	0.013	0.013	0.013
	Consulting	46	879	0.043	0.043	0.022

For each network, we show its type and name; number of nodes (N) and edges (L); and density of driver nodes calculated in the real network (n_D^{real}), after degree-preserved randomization ($n_D^{\text{rand-Degree}}$) and after full randomization ($n_D^{\text{rand-ER}}$). For data sources and references, see Supplementary Information, section VI.

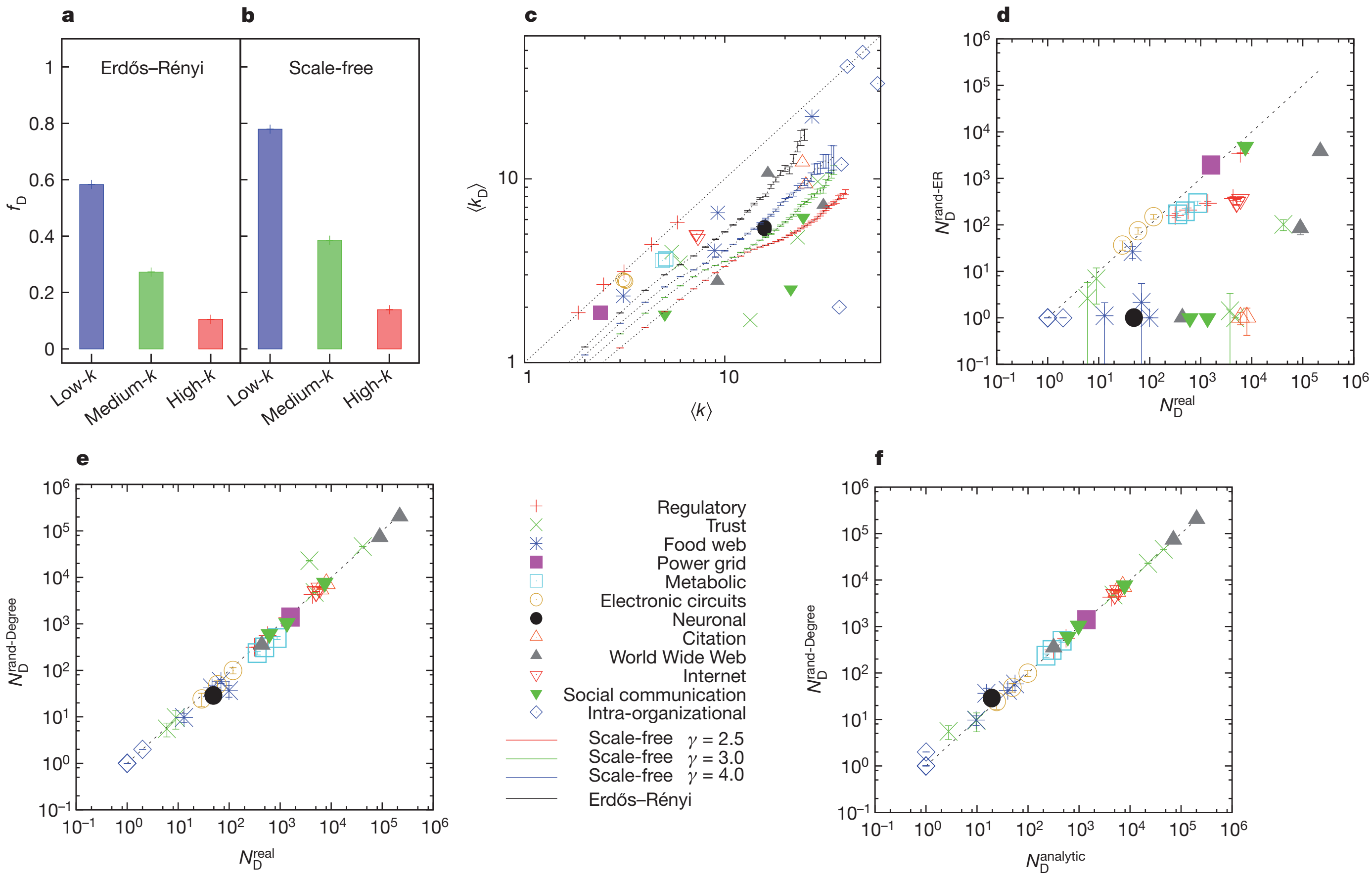

Figure 2 | Characterizing and predicting the driver nodes (N_D). **a**, **b**, Role of the hubs in model networks. The bars show the fractions of driver nodes, f_D, among the low-, medium- and high-degree nodes in two network models, Erdős–Rényi (**a**) and scale-free (**b**), with $N = 10^4$ and $\langle k \rangle = 3$ ($\gamma = 3$), indicating that the driver nodes tend to avoid the hubs. Both the Erdős–Rényi and the scale-free networks are generated from the static model[38] and the results are averaged over 100 realizations. The error bars (s.e.m.), shown in the figure, are smaller than the symbols. **c**, Mean degree of driver nodes compared with the mean degree of all nodes in real and model networks, indicating that in real systems the hubs are avoided by the driver nodes. **d**, Number of driver nodes, $N_D^{\text{rand-ER}}$, obtained for the fully randomized version of the networks listed in Table 1, compared with the exact value, N_D^{real}. **e**, Number of driver nodes, $N_D^{\text{rand-Degree}}$, obtained for the degree-preserving randomized version of the networks shown in Table 1, compared with N_D^{real}. **f**, The analytically predicted N_D^{analytic} calculated using the cavity method, compared with $N_D^{\text{rand-Degree}}$. In **d**–**f**, data points and error bars (s.e.m.) were determined from 1,000 realizations of the randomized networks.

the impact of various network parameters on N_D. Although the cavity method does not offer a closed-form solution, we can derive the dependence of n_D on key network parameters in the thermodynamic limit ($N \to \infty$). We find, for example, that for a directed Erdős–Rényi network n_D decays as

$$n_D \approx e^{-\langle k \rangle /2} \tag{4}$$

for large $\langle k \rangle$. For scale-free networks with degree exponent $\gamma_{in} = \gamma_{out} = \gamma$ in the large-$\langle k \rangle$ limit[38], we have

$$n_D \approx \exp\left[-\frac{1}{2}\left(1-\frac{1}{\gamma-1}\right)\langle k \rangle\right] \tag{5}$$

which has the same $\langle k \rangle$ dependence as equation (4) in the $\gamma \to \infty$ limit. Equation (5) predicts that $\gamma_c = 2$ is a critical exponent for the controllability of an infinite scale-free network, as only for $\gamma > \gamma_c$ can we control the full system through a finite subset of nodes (that is, $n_D < 1$). For $\gamma \le \gamma_c$ in the thermodynamic limit, all nodes must be individually controlled (that is, $n_D = 1$). We note that γ_c is different from $\gamma = 3$, which is the critical exponent for a number of network phenomena driven by the divergence of $\langle k^2 \rangle$, from network robustness to epidemic spreading[31–33,45]. To check the validity of the analytical predictions, we determined the $\langle k \rangle$ dependence of n_D numerically for both Erdős–Rényi and scale-free networks, confirming the asymptotic exponential dependence of n_D on $\langle k \rangle$, as predicted by equations (4) and (5). Furthermore, the predicted n_D value is in excellent agreement with the numerical results for $\gamma > 3$ (Fig. 3d, e). Near $\gamma = 2$, however, n_D as predicted by the cavity method deviates from the exact n_D value owing to degree correlations that are prominent at $\gamma_c = 2$ and can be eliminated by imposing a degree cut-off in constructing the scale-free networks[39,46] (Supplementary Information, section IV.B).

Equation (5) also shows that n_D decreases as γ increases (for fixed $\langle k \rangle$), indicating that n_D is affected by degree heterogeneity, representing the spread between the less connected and the more connected nodes. We defined the degree heterogeneity as $H = \Delta / \langle k \rangle$, where $\Delta = \sum_i \sum_j |k_i - k_j| P(k_i) P(k_j)$ is the average absolute degree difference of all pairs of nodes (i and j) drawn from the degree distribution $P(k)$. The degree heterogeneity is zero ($H = 0$) for networks in which all nodes have the same degree, such as the random regular digraph (Fig. 3a), in which the in- and out-degrees of the nodes are fixed to $\langle k \rangle /2$ but the nodes are connected randomly. For $\langle k \rangle \ge 2$, this graph always has a perfect matching[47], which means that a single driver node can control the whole system (Supplementary Information, section IV.B1). The degree heterogeneity increases as we move from the random regular digraph to an Erdős–Rényi network (Fig. 3b) and eventually to scale-free networks with decreasing γ (Fig. 3c). Overall, the fraction of driver nodes, n_D, increases monotonically with H, whether we keep γ (Fig. 3f) or $\langle k \rangle$ (Fig. 3g) constant.

Taking these results together, we find that the denser a network is, the fewer driver nodes are needed to control it, and that small changes in the average degree induce orders-of-magnitude variations in n_D.

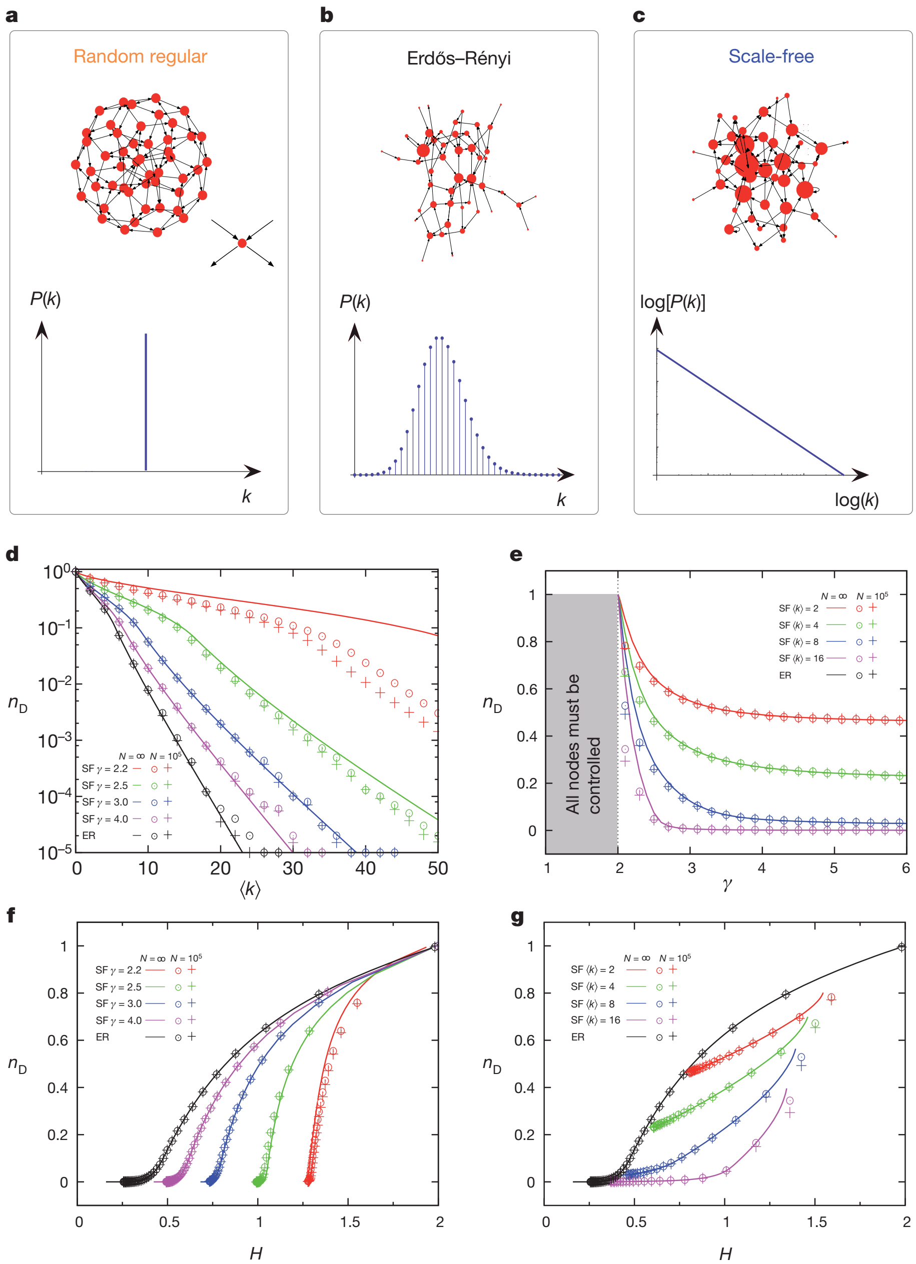

Figure 3 | The impact of network structure on the number of driver nodes. **a**–**c**, Characteristics of the explored model networks. A random-regular digraph (**a**), shown here for $\langle k \rangle = 4$, is the most degree-homogeneous network as $k_{in} = k_{out} = \langle k \rangle/2$ for all nodes. Erdős–Rényi networks (**b**) have Poisson degree distributions and their degree heterogeneities are determined by $\langle k \rangle$. Scale-free networks (**c**) have power-law degree distributions, yielding large degree heterogeneities. **d**, Driver node density, n_D, as a function of $\langle k \rangle$ for Erdős–Rényi (ER) and scale-free (SF) networks with different values of γ. Both the Erdős–Rényi and the scale-free networks are generated from the static model[38] with $N = 10^5$. Lines are analytical results calculated by the cavity method using the expected degree distribution in the $N \rightarrow \infty$ limit. Symbols are calculated for the constructed discrete network: open circles indicate exact results calculated from the maximum matching algorithm, and plus symbols indicate the analytical results of the cavity method using the exact degree sequence of the constructed network. For large $\langle k \rangle$, n_D approaches its lower bound, N^{-1}, that is, a single driver node ($N_D = 1$) in a network of size N. **e**, n_D as a function of γ for scale-free networks with fixed $\langle k \rangle$. For infinite scale-free networks, $n_D \rightarrow 1$ as $\gamma \rightarrow \gamma_c = 2$, that is, it is necessary to control almost all nodes to control the network fully. For finite scale-free networks, n_D reaches its maximum as γ approaches γ_c (Supplementary Information). **f**, n_D as a function of degree heterogeneity, H, for Erdős–Rényi and scale-free networks with fixed γ and variable $\langle k \rangle$. **g**, n_D as a function of H for Erdős–Rényi and scale-free networks for fixed $\langle k \rangle$ and variable γ. As γ increases, the curves converge to the Erdős–Rényi result (black) at the corresponding $\langle k \rangle$ value.

Furthermore, the larger are the differences between node degrees, the more driver nodes are needed to control the system. Overall, networks that are sparse and heterogeneous, which are precisely the characteristics often seen in complex systems like the cell or the Internet[13–16], require the most driver nodes, underscoring that such systems are difficult to control.

Robustness of control

To see how robust is our ability to control a network under unavoidable link failure, we classify each link into one of the following three categories (Fig. 1d, g, j): 'critical' if in its absence we need to increase the number of driver nodes to maintain full control; 'redundant' if it can be removed without affecting the current set of driver nodes; or 'ordinary' if it is neither critical nor redundant. Figure 4 shows the densities of critical ($l_c = L_c/L$), redundant ($l_r = L_r/L$) and ordinary ($l_o = L_o/L$) links for each real network, and indicates that most networks have few or no critical links. Most links are ordinary, meaning that they have a role in some control configurations but that the network can still be controlled in their absence.

To understand the factors that determine l_c, l_r and l_o, in Fig. 5a, c, f we show their $\langle k \rangle$ dependence for model systems. The behaviour of l_c is the easiest to understand: for small $\langle k \rangle$, all links are essential for control ($l_c \approx 1$). As $\langle k \rangle$ increases, the network's redundancy increases, decreasing l_c. The increasing redundancy suggests that the density of redundant links, l_r, should always increase with $\langle k \rangle$, but it does not: it reaches a maximum at a critical value of $\langle k \rangle$, $\langle k \rangle_c$, after which it decays. This non-monotonic behaviour results from the competition of two topologically distinct regions of a network, the core and leaves[43]. The core represents a compact cluster of nodes left in the network after applying a greedy leaf removal procedure[48], and leaves are nodes with $k_{in} = 1$ or $k_{out} = 1$ before or during leaf removal. The core emerges through a percolation transition (Fig. 5b, d): for $k < \langle k \rangle_c$, $n_{core} = N_{core}/N = 0$, so the system consists of leaves only (Fig. 5e). At $\langle k \rangle = \langle k \rangle_c$, a small core emerges, decreasing the number of leaves. For Erdős–Rényi networks, we predict that $\langle k \rangle_c = 2e \approx 5.436564$ in agreement with the numerical result (Fig. 5a, b), a value that coincides with $\langle k \rangle$ where l_r reaches its maximum. Indeed, l_r starts decaying at $\langle k \rangle_c$ because for $\langle k \rangle > \langle k \rangle_c$ the number of distinct maximum

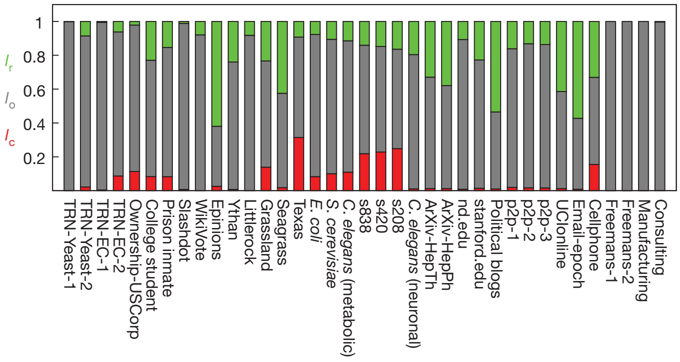

Figure 4 | Link categories for robust control. The fractions of critical (red, l_c), redundant (green, l_r) and ordinary (grey, l_o) links for the real networks named in Table 1. To make controllability robust to link failures, it is sufficient to double only the critical links, formally making each of these links redundant and therefore ensuring that there are no critical links in the system.

matchings increases exponentially (Supplementary Information, section IV.C) and, as a result, the chance that a link does not participate in any control configuration decreases. For scale-free networks, we observe the same behaviour, with the caveat that $\langle k \rangle_c$ decreases with γ (Fig. 5c, d).

Discussion and conclusions

Control is a central issue in most complex systems, but because a general theory to explore it in a quantitative fashion has been lacking, little is known about how we can control a weighted, directed network—the configuration most often encountered in real systems. Indeed, applying Kalman's controllability rank condition (equation (3)) to large networks is computationally prohibitive, limiting previous work to a few dozen nodes at most[17–19]. Here we have developed the tools to address controllability for arbitrary network topologies and sizes. Our key finding, that N_D is determined mainly by the degree

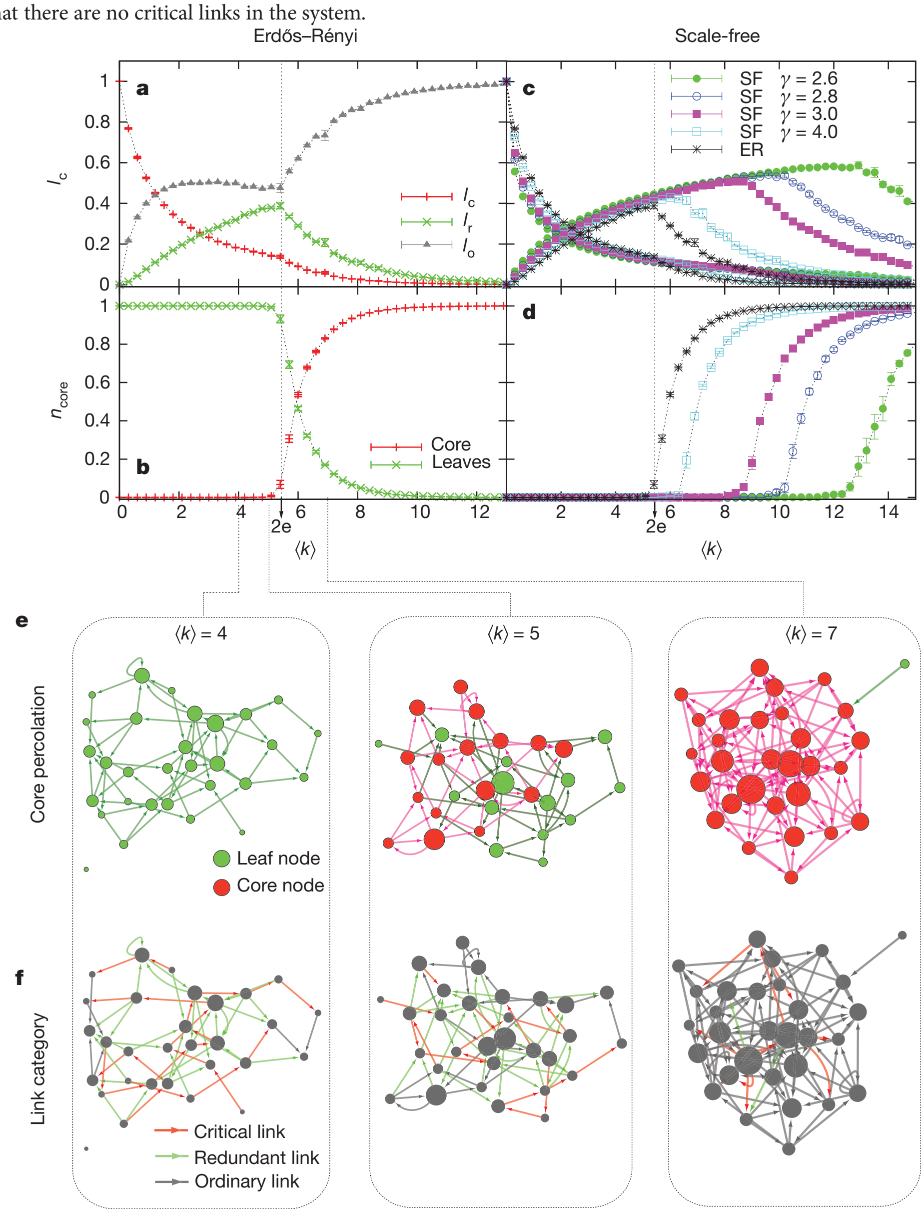

Figure 5 | Control robustness. a, Dependence on $\langle k \rangle$ of the fraction of critical (red, l_c), redundant (green, l_r) and ordinary (grey, l_o) links for an Erdős–Rényi network: l_r peaks at $\langle k \rangle = \langle k \rangle_c = 2e$ and the derivative of l_c is discontinuous at $\langle k \rangle = \langle k \rangle_c$. **b**, Core percolation for Erdős–Rényi network occurs at $k = \langle k \rangle_c = 2e$, which explains the l_r peak. **c**, **d**, Same as in **a** and **b** but for scale-free networks. The Erdős–Rényi and scale-free networks[38] have $N = 10^4$ and the results are averaged over ten realizations with error bars defined as s.e.m. Dotted lines are only a guide to the eye. **e**, The core (red) and leaves (green) for small Erdős–Rényi networks ($N = 30$) at different $\langle k \rangle$ values ($\langle k \rangle = 4, 5, 7$). Node sizes are proportional to node degrees. **f**, The critical (red), redundant (green) and ordinary (grey) links for the above Erdős–Rényi networks at the corresponding $\langle k \rangle$ values.

distribution, allows us to use the tools of statistical physics to predict N_D from $P(k_{in}, k_{out})$ analytically, offering a general formalism with which to explore the impact of network topology on controllability.

The framework presented here raises a number of questions, answers to which could further deepen our understanding of control in complex environments. For example, although our analytical work focused on uncorrelated networks, the algorithmic method we developed can identify N_D for arbitrary networks, providing a framework in which to address the role of correlations systematically[40,49,50]. Taken together, our results indicate that many aspects of controllability can be explored exactly and analytically for arbitrary networks if we combine the tools of network science and control theory, opening new avenues to deepening our understanding of complex systems.

Received 18 November 2010; accepted 16 March 2011.

1. Kalman, R. E. Mathematical description of linear dynamical systems. *J. Soc. Indus. Appl. Math. Ser. A* **1,** 152–192 (1963).
2. Luenberger, D. G. *Introduction to Dynamic Systems: Theory, Models, & Applications* (Wiley, 1979).
3. Slotine, J.-J. & Li, W. *Applied Nonlinear Control* (Prentice-Hall, 1991).
4. Kelly, F. P., Maulloo, A. K. & Tan, D. K. H. Rate control for communication networks: shadow prices, proportional fairness and stability. *J. Oper. Res. Soc.* **49,** 237–252 (1998).
5. Srikant, R. *The Mathematics of Internet Congestion Control* (Birkhäuser, 2004).
6. Chiang, M., Low, S. H., Calderbank, A. R. & Doyle, J. C. Layering as optimization decomposition: a mathematical theory of network architectures. *Proc. IEEE* **95,** 255–312 (2007).
7. Wang, X. F. & Chen, G. Pinning control of scale-free dynamical networks. *Physica A* **310,** 521–531 (2002).
8. Wang, W. & Slotine, J.-J. E. On partial contraction analysis for coupled nonlinear oscillators. *Biol. Cybern.* **92,** 38–53 (2005).
9. Sorrentino, F., di Bernardo, M., Garofalo, F. & Chen, G. Controllability of complex networks via pinning. *Phys. Rev. E* **75,** 046103 (2007).
10. Yu, W., Chen, G. & Lü, J. On pinning synchronization of complex dynamical networks. *Automatica* **45,** 429–435 (2009).
11. Marucci, L. *et al.* How to turn a genetic circuit into a synthetic tunable oscillator, or a bistable switch. *PLoS ONE* **4,** e8083 (2009).
12. Strogatz, S. H. Exploring complex networks. *Nature* **410,** 268–276 (2001).
13. Dorogovtsev, S. N. & Mendes, J. F. F. *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Univ. Press, 2003).
14. Newman, M., Barabási, A.-L. & Watts, D. J. *The Structure and Dynamics of Networks* (Princeton Univ. Press, 2006).
15. Caldarelli, G. *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford Univ. Press, 2007).
16. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74,** 47–97 (2002).
17. Tanner, H. G. in *Proc. 43rd IEEE Conf. Decision Contr.* Vol. 3, 2467–2472 (2004).
18. Lombardi, A. & Hörnquist, M. Controllability analysis of networks. *Phys. Rev. E* **75,** 56110 (2007).
19. Liu, B., Chu, T., Wang, L. & Xie, G. Controllability of a leader–follower dynamic network with switching topology. *IEEE Trans. Automat. Contr.* **53,** 1009–1013 (2008).
20. Rahmani, A., Ji, M., Mesbahi, M. & Egerstedt, M. Controllability of multi-agent systems from a graph-theoretic perspective. *SIAM J. Contr. Optim.* **48,** 162–186 (2009).
21. Kim, D.-H. & Motter, A. E. Slave nodes and the controllability of metabolic networks. *N. J. Phys.* **11,** 113047 (2009).
22. Mesbahi, M. & Egerstedt, M. *Graph Theoretic Methods in Multiagent Networks* (Princeton Univ. Press, 2010).
23. Motter, A. E., Gulbahce, N., Almaas, E. & Barabási, A.-L. Predicting synthetic rescues in metabolic networks. *Mol. Syst. Biol.* **4,** 168 (2008).
24. Pastor-Satorras, R. & Vespignani, A. *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge Univ. Press, 2004).
25. Lezon, T. R., Banavar, J. R., Cieplak, M., Maritan, A. & Fedoroff, N. V. Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc. Natl Acad. Sci. USA* **103,** 19033–19038 (2006).
26. Lin, C.-T. Structural controllability. *IEEE Trans. Automat. Contr.* **19,** 201–208 (1974).
27. Shields, R. W. & Pearson, J. B. Structural controllability of multi-input linear systems. *IEEE Trans. Automat. Contr.* **21,** 203–212 (1976).
28. Lohmiller, W. & Slotine, J.-J. E. On contraction analysis for nonlinear systems. *Automatica* **34,** 683–696 (1998).
29. Yu, W., Chen, G., Cao, M. & Kurths, J. Second-order consensus for multiagent systems with directed topologies and nonlinear dynamics. *IEEE Trans. Syst. Man Cybern. B* **40,** 881–891 (2010).
30. Hopcroft, J. E. & Karp, R. M. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput.* **2,** 225–231 (1973).
31. Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406,** 378–382 (2000).
32. Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* **85,** 4626–4628 (2000).
33. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86,** 3200–3203 (2001).
34. Nishikawa, T., Motter, A. E., Lai, Y.-C. & Hoppensteadt, F. C. Heterogeneity in oscillator networks: are smaller worlds easier to synchronize? *Phys. Rev. Lett.* **91,** 014101 (2003).
35. Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5,** 17–60 (1960).
36. Bollobás, B. *Random Graphs* (Cambridge Univ. Press, 2001).
37. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286,** 509–512 (1999).
38. Goh, K.-I., Kahng, B. & Kim, D. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.* **87,** 278701 (2001).
39. Chung, F. & Lu, L. Connected component in random graphs with given expected degree sequences. *Ann. Combin.* **6,** 125–145 (2002).
40. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296,** 910–913 (2002).
41. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298,** 824–827 (2002).
42. Mézard, M. & Parisi, G. The Bethe lattice spin glass revisited. *Eur. Phys. J. B* **20,** 217–233 (2001).
43. Zdeborová, L. & Mézard, M. The number of matchings in random graphs. *J. Stat. Mech.* **05,** 05003 (2006).
44. Zhou, H. & Ou-Yang, Z.-c. Maximum matching on random graphs. Preprint at ⟨http://arxiv.org/abs/cond-mat/0309348⟩ (2003).
45. Callaway, D. S., Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Network robustness and fragility: percolation on random graphs. *Phys. Rev. Lett.* **85,** 5468–5471 (2000).
46. Boguñá, M., Pastor-Satorras, R. & Vespignani, A. Cut-offs and finite size effects in scale-free networks. *Eur. Phys. J. B* **38,** 205–209 (2004).
47. Lovász, L. & Plummer, M. D. *Matching Theory* (American Mathematical Society, 2009).
48. Bauer, M. & Golinelli, O. Core percolation in random graphs: a critical phenomena analysis. *Eur. Phys. J. B* **24,** 339–352 (2001).
49. Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89,** 208701 (2002).
50. Pastor-Satorras, R., Vázquez, A. & Vespignani, A. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.* **87,** 258701 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Song, G. Bianconi, H. Zhou, L. Vepstas, N. Gulbahce, H. Jeong, Y.-Y. Ahn, B. Barzel, N. Blumm, D. Wang, Z. Qu and Y. Li 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 Office of Naval Research under Agreement Number N000141010968; the Defense Threat Reduction Agency awards WMD BRBAA07-J-2-0035 and BRBAA08-Per4-C-2-0033; and the James S. McDonnell Foundation 21st Century Initiative in Studying Complex Systems.

Author Contributions All authors designed and did the research. Y.-Y.L. analysed the empirical data and did the analytical and numerical calculations. A.-L.B. was the lead writer of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.-L.B. (alb@neu.edu).