ARTICLES
PUBLISHED ONLINE: 27 MAY 2012 | DOI: 10.1038/NPHYS2327

nature physics

Controlling edge dynamics in complex networks

Tamás Nepusz[1,2] and Tamás Vicsek[1,2]*

The interaction of distinct units in physical, social, biological and technological systems naturally gives rise to complex network structures. Networks have constantly been in the focus of research for the past decade, with considerable advances in the description of their structural and dynamical properties. However, much less effort has been devoted to studying the controllability of the dynamics taking place on them. Here we introduce and evaluate a dynamical process defined on the edges of a network, and demonstrate that the controllability properties of this process significantly differ from simple nodal dynamics. Evaluation of real-world networks indicates that most of them are more controllable than their randomized counterparts. We also find that transcriptional regulatory networks are particularly easy to control. Analytic calculations show that networks with scale-free degree distributions have better controllability properties than uncorrelated networks, and positively correlated in- and out-degrees enhance the controllability of the proposed dynamics.

The past decade has witnessed an explosive growth of interest in the descriptive analysis of the complex natural and technological systems that permeate many aspects of everyday life[1–3]. Research in network science has mostly been focused on measuring[4–6], modelling[7,8] and decomposing[9–11] network representations of existing natural phenomena to deepen our understanding of the underlying systems. Considerably less attention has been dedicated to the various types of network dynamics[12–15] and even less to the problem of controllability[16–18], that is, determining the conditions under which the dynamics of a network can be driven from any initial state to any desired final state within finite time[19–22].

Structural controllability[23] has been proposed recently as a framework for studying the controllability properties of directed complex networks[22]. In this framework, a linear time-invariant nodal dynamics is assumed on the network, governed by the following equation:

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

where $\mathbf{A}$ is the transpose of the (weighted) adjacency matrix of the network, $\mathbf{x}(t)$ is a time-dependent vector of the state variables of the nodes, $\mathbf{u}(t)$ is the vector of input signals, and $\mathbf{B}$ is the so-called input matrix, which defines how the input signals are connected to the nodes of the network. The dynamics is structurally controllable if there exists a pair of matrices $(\mathbf{A}^*,\mathbf{B}^*)$ with the same structure as $(\mathbf{A},\mathbf{B})$ such that the network can be driven from any initial state to any final state by appropriately choosing the input signals $\mathbf{u}(t)$ (ref. 23). The structural equivalence of $\mathbf{A}$ and $\mathbf{A}^*$ means that $\mathbf{A}^*$ is not allowed to contain a non-zero entry when the corresponding entry in $\mathbf{A}$ is zero. Structural controllability is a generic property in the sense that almost all weight combinations of a given network are controllable if the network is structurally controllable for a given $\mathbf{B}$ (refs 23,24). The minimum number of input signals is then determined by finding a maximum matching in the network, that is, a maximum subset of edges such that each node has at most one inbound and at most one outbound edge from the matching. The number of nodes without inbound edges from the matching is then equal to the number of input signals required for structural controllability[22].

Perhaps the most striking feature of the structural-controllability approach to linear nodal dynamics is that input signals tend to control the hubs of the network only indirectly. In addition, real-world networks that seem to have evolved to control an underlying process (such as transcriptional regulatory networks) need many input signals[22], as driven nodes (that is, those which receive an input signal directly) are not able to control their subordinates independently of each other. Here, we describe a dynamics that takes place on the edges of the network, and show that this dynamics leads to significantly different controllability properties for the same real-world networks.

Switchboard dynamics in complex networks

We study a dynamical process on the edges of a directed complex network $G(V,E)$ as follows. Let $\mathbf{x}=[x_j]$ denote the state vector of the process, where one state variable corresponds to each edge of the network. Let $\mathbf{y}_i^-$ and $\mathbf{y}_i^+$ be vectors consisting of those x_j values that correspond to the inbound and outbound edges of vertex i, respectively, and let $\mathbf{M}_i$ denote a matrix with the number of rows being equal to the out-degree and the number of columns being equal to the in-degree of vertex i. Furthermore, we assume that the dynamics can be influenced from the environment by adding an offset vector $\mathbf{u}_i$ to the state vector of the outbound edges of any node i. The equations governing the dynamics of the network are then as follows:

$$\dot{\mathbf{y}}_i^+(t)=\mathbf{M}_i\mathbf{y}_i^-(t)-\boldsymbol{\tau}_i\otimes\mathbf{y}_i^+(t)+\sigma_i\mathbf{u}_i(t) \qquad (1)$$

where $\boldsymbol{\tau}_i$ is a vector of damping terms corresponding to the edges in $\mathbf{y}_i^+(t)$, σ_i is 1 if vertex i is a so-called driver node and zero otherwise, and $\otimes$ denotes the entry-wise product of two vectors of the same size.

We call the above the switchboard dynamics (SBD) because each vertex i acts as a small switchboard-like device mapping the signals of the inbound edges to the outbound edges using a linear operator $\mathbf{M}_i$, which is called the mixing or switching matrix from now on. To simplify the equations, state variables and signals such as $\mathbf{y}_i^+$, $\mathbf{y}_i^-$ and $\mathbf{u}_i$ are implicitly considered as time-dependent, even if the time variable t is omitted. Furthermore, note that for an edge $v \to w$, exactly one of the coordinates of $\mathbf{u}_v$ affects the state of this edge, therefore we can simply introduce a unified input vector $\mathbf{u}$ where

[1]Department of Biological Physics, Eötvös Loránd University, Pázmány Péter sétány 1/a, 1117 Budapest, Hungary, [2]Statistical and Biological Physics Research Group of the Hungarian Academy of Sciences, Pázmány Péter sétány 1/a, 1117 Budapest, Hungary. *e-mail: vicsek@hal.elte.hu.

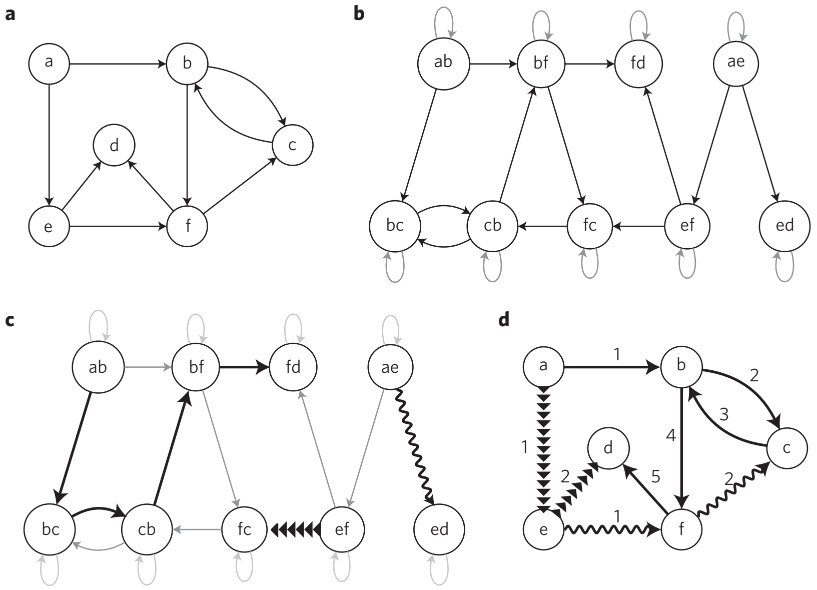

Figure 1 | SBD on a simple network and its mapping to the line graph. a, An example network *G* with six vertices and nine edges. The SBD takes place on the edges of the network. **b**, The line graph *L*(*G*) corresponding to *G*. A linear time-invariant dynamics on the vertices of this network is equivalent to the SBD on *G*. Node labels refer to the endpoints of the edges in *G* to which they correspond. **c**, Applying the minimum input theorem to *L*(*G*) yields disjoint control paths, denoted by different line styles. **d**, The control paths in *G*, mapped back from *L*(*G*). Note how each path in *L*(*G*) became an edge-disjoint walk in *G*. Numbers represent the order in which the edges have to be traversed in the walks. The two driver nodes are a and e because each walk starts from either a or e.

the *j*th element u_j is simply the component of the offset vectors that affects edge *j* directly.

In some sense, the SBD provides a simplified representation of the underlying dynamic processes of many real-world networks. For instance, in social communication networks, a node (in this case, a person) is constantly processing the information received via its inbound edges and makes decisions which are then communicated to other nodes via the outbound edges. The inbound and outbound signals are then represented by the state variables x_j, while the decision process is modelled by the mixing matrices $\mathbf{M}_i$.

We must also explain the motivation for introducing the offset vectors as a means of controlling the system instead of assuming external input signals. In most networks, one usually cannot take control over a single edge as the connections do not always have a physical realization. Therefore, to control an edge in a network, one has to take control over the vertex from which the edge originates, and adjust the output vector of the vertex appropriately. This adjustment is represented by the term $\sigma_i\mathbf{u}_i$ for each vertex *i*, although we note that the actual adjustment can also be achieved by simply replacing the original node with an appropriate control unit. Throughout this paper, we will be interested in determining an optimal control configuration for the SBD of a given network, where optimality is measured by the number of driver nodes $\sigma = \sum_i \sigma_i$—in other words, we assume that the cost of controlling a node does not depend on the number of controlled outputs (see Supplementary Information for a more detailed discussion).

First, we make a connection between the SBD and a standard linear dynamical system by re-writing the equations of the SBD (equation (1)) in terms of x_i. Note that the derivative of the state of an arbitrary edge *j* originating in some vertex *r* and terminating in vertex *s* depends only on itself and on the states of edges whose head is *r*. Let us denote this latter set by Γ_j^-, simplifying our dynamical equation to

$$\dot{x}_j = \sum_{k\in\Gamma_j^-} w_{kj}x_k - \tau_j x_j + \sigma_s u_j$$

where w_{kj} is the element in the mixing matrix $\mathbf{M}_r$ of vertex *r* that corresponds to edge *k* (as inbound edge) and edge *j* (as outbound edge), τ_j is the damping term related to edge *j*, and u_j is equal to the value of the input signal affecting the state variable of edge *j*. Defining $w_{kj} = 0$ for all $k \notin \Gamma_j^-$ yields

$$\dot{\mathbf{x}} = (\mathbf{W} - \mathbf{T})\mathbf{x} + \mathbf{H}\mathbf{u} \qquad (2)$$

where the unknown variables are as follows: $\mathbf{W} = [w_{kj}]$ is a matrix where w_{kj} may be nonzero if and only if the head of edge *k* is the tail of edge *j*. $\mathbf{T}$ is a diagonal matrix with the damping terms of each edge in the main diagonal. $\mathbf{H}$ is a diagonal matrix where the *j*th diagonal element is σ_s if vertex *s* is the tail of edge *j*.

Equation (2) essentially describes a simple linear time-invariant dynamical system of the form $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$ with the substitution $\mathbf{A} = \mathbf{W} - \mathbf{T}$ and $\mathbf{B} = \mathbf{H}$. It is also easy to see that $\mathbf{W}$ is the transpose of the adjacency matrix of the line digraph *L*(*G*) of the original digraph *G* by definition. The nodes of *L*(*G*) thus correspond to the edges of the original network *G*, and each edge of *L*(*G*) represents a length-two directed path of *G*. An example network *G* is shown in Fig. 1a, and its corresponding line digraph on Fig. 1b. The loop edges on Fig. 1b,c arise from the damping term $-\mathbf{T}$ in equation (1).

Structural controllability of the SBD

To apply the minimum input theorem of Liu *et al.*[22] to *L*(*G*), we must first show that the matrices $\mathbf{W} - \mathbf{T}$ and $\mathbf{H}$ that occur in equation (2) are indeed structured matrices. $\mathbf{W} - \mathbf{T}$ is a structured matrix by definition because its non-zero coefficients w_{kj} and τ_j are independent of each other. $\mathbf{H}$ may seem not to be structural at first glance, as edges with the same tail vertex correspond to the same σ_s value in the matrix. However, note that we can multiply any of the elements in $\mathbf{H}$ by an arbitrary factor c_j (where *j* is the index of an edge) if we divide the corresponding u_j input signal by c_j to keep the dynamical equations the same. Therefore, we can treat both $\mathbf{W} - \mathbf{T}$ and $\mathbf{H}$ as structured matrices.

Applying the minimum input theorem of Liu *et al.*[22] to *L*(*G*) (Fig. 1b) gives us a set of control paths and driven nodes in the line

Table 1 | Controllability properties of the real networks analysed in this paper.

Type	No.	Name	Nodes	Edges	n_D^{SBD}	m_D^{SBD}	n_D^{Liu}	n_D^{ER}	n_D^{Degree}
Regulatory	1	Ownership-USCorp	7,253	6,726	0.160	0.924	0.820	0.339	0.085
	2	TRN-EC-2	418	519	0.222	0.923	0.751	0.366	0.148
	3	TRN-Yeast-1	4,441	12,873	0.034	0.965	0.965	0.415	0.033
	4	TRN-Yeast-2	688	1,079	0.177	0.952	0.821	0.381	0.137
Trust	5	College*	32	96	0.344	0.188	0.188	0.418	0.315
	6	Epinions*	75,888	508,837	0.336	0.375	0.549	0.445	0.448
	7	Prison*	67	182	0.403	0.319	0.134	0.411	0.451
	8	Slashdot*	82,168	948,464	0.323	0.117	0.045	0.458	0.392
	9	WikiVote*	7,115	103,689	0.281	0.653	0.666	0.463	0.620
Food web	10	Grassland	88	137	0.318	0.606	0.523	0.381	0.297
	11	Little Rock	183	2,494	0.639	0.603	0.541	0.463	0.649
	12	Seagrass	49	226	0.449	0.518	0.265	0.436	0.433
	13	Ythan	135	601	0.304	0.597	0.511	0.432	0.337
Metabolic	14	*C. elegans*	1,173	2,864	0.182	0.116	0.302	0.409	0.309
	15	*E. coli*	2,275	5,763	0.182	0.121	0.382	0.409	0.309
	16	*S. cerevisiae*	1,511	3,833	0.185	0.116	0.329	0.409	0.313
Electronic circuits	17	s208a	122	189	0.451	0.344	0.238	0.381	0.431
	18	s420a	252	399	0.456	0.348	0.234	0.385	0.440
	19	s838a	512	819	0.459	0.350	0.232	0.381	0.442
Neuronal and brain	20	*C. elegans*	297	2,359	0.549	0.377	0.165	0.449	0.499
	21	Macaque	45	463	0.333	0.067	0.022	0.446	0.457
Citation	22	arXiv-HepPh*	34,546	421,578	0.356	0.586	0.232	0.459	0.577
	23	arXiv-HepTh*	27,770	352,807	0.359	0.561	0.216	0.460	0.569
World Wide Web	24	Google	15,763	171,206	0.670	0.578	0.337	0.457	0.612
	25	Polblogs	1,490	19,090	0.509	0.526	0.471	0.460	0.501
	26	nd.edu	325,729	1,497,134	0.271	0.376	0.677	0.433	0.301
	27	stanford.edu	281,904	2,312,497	0.665	0.575	0.317	0.450	0.653
Internet	28	p2p-1	10,876	39,994	0.334	0.591	0.552	0.425	0.344
	29	p2p-2	8,846	31,839	0.344	0.628	0.578	0.423	0.344
	30	p2p-3	8,717	31,525	0.343	0.625	0.577	0.424	0.344
Social communication	31	Twitter*†	41.7×10^6	1.47×10^9	0.402	0.403	–	0.476	0.434
	32	UCIOnline	1,899	20,296	0.216	0.236	0.323	0.456	0.375
	33	WikiTalk	2,394,385	5,021,410	0.022	0.744	0.968	0.399	0.026
Organizational	34	Consulting*	46	879	0.522	0.150	0.043	0.458	0.460
	35	Freemans-1*	34	645	0.412	0.113	0.088	0.441	0.476
	36	Freemans-2*	34	830	0.588	0.071	0.029	0.439	0.465
	37	Manufacturing*	77	2,228	0.597	0.135	0.013	0.468	0.424
	38	University*	81	817	0.519	0.245	0.012	0.451	0.532

Notations are as follows: fraction of driver nodes (n_D^{SBD}) and driven edges (m_D^{SBD}) under the SBD; fraction of driver nodes under the simple nodal dynamics[22] (n_D^{Liu}); fraction of driver nodes under the SBD in randomized networks using the Erdős–Rényi model (n_D^{ER}) and the degree-preserving configuration model (n_D^{Degree}). Note that this latter model does not preserve the joint degree distribution. Results for null models are averaged from 100 randomizations. *Networks where the edges were reversed compared to the original publication (see Supplementary Section S3.1). †Results calculated directly from the degree distribution (that is, not taking into account balanced components).

digraph (Fig. 1c), or equivalently, a set of driven edges in the original graph G. As edges can be controlled only via the offset vectors, the set of driver nodes is given by collecting those vertices that have at least one outbound driven edge. However, note that the minimum input theorem guarantees only that the number of driven nodes in $L(G)$ will be minimal, and this does not imply that the obtained set of driver nodes in G is also minimal. In particular, selecting all the loop edges in the line digraph (Fig. 1b) would make every edge in G a driven edge, requiring us to take control of all the nodes except the ones without outgoing edges (node d on Fig. 1a). This is clearly not optimal; a better solution is shown in Fig. 1c, which requires only two driver nodes: a and e.

Let us now compare the control paths obtained from the maximum matching in the line graph $L(G)$ in Fig. 1c with the corresponding control paths in the original graph G in Fig. 1d. It can be seen that the maximum matching consists of vertex-disjoint open and closed paths (also called stems and buds) in $L(G)$, and mapping these paths back to G yields edge-disjoint open and closed walks in G. The walks together form a complete cover of the edges of G. As the first vertex of each stem has to be driven in $L(G)$, the driver nodes in G are those from which the corresponding open edge-disjoint walks originate. Our goal is thus to find a cover that minimizes the number of nodes from which open walks originate in G.

Let us denote the out-degree of vertex v with d_v^+ and the in-degree of vertex v with d_v^-. Vertex v is called divergent if $d_v^+ > d_v^-$, convergent if $d_v^+ < d_v^-$, and balanced if $d_v^+ = d_v^-$. Let us define a balanced component as a connected component consisting solely of balanced vertices and at least one edge. Our key result is that the minimum set of driver nodes required to control the

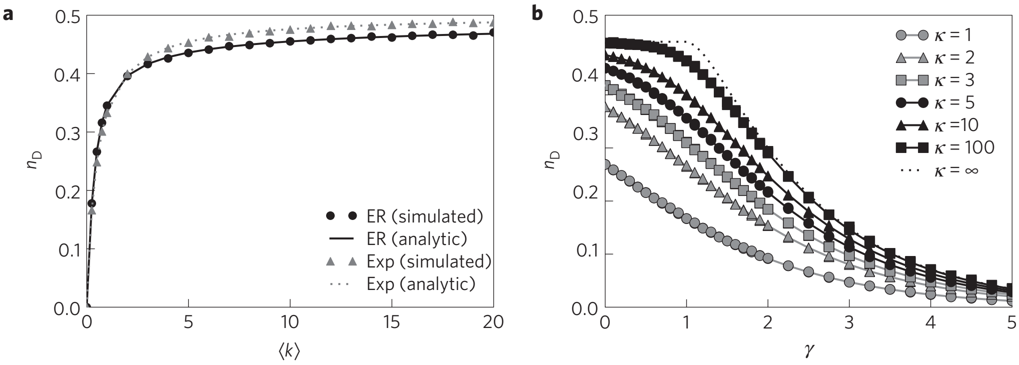

Figure 2 | Expected fraction of driver nodes n_D in various model networks. a, n_D in Erdős–Rényi (ER) and exponential (Exp) networks as a function of the average degree ⟨k⟩. **b**, n_D in scale-free networks with exponential cutoff as a function of the exponent γ of the degree distribution, for different cutoff values κ. On both panels, symbols denote the results of simulations on networks with 10^5 nodes, solid lines correspond to the analytical results.

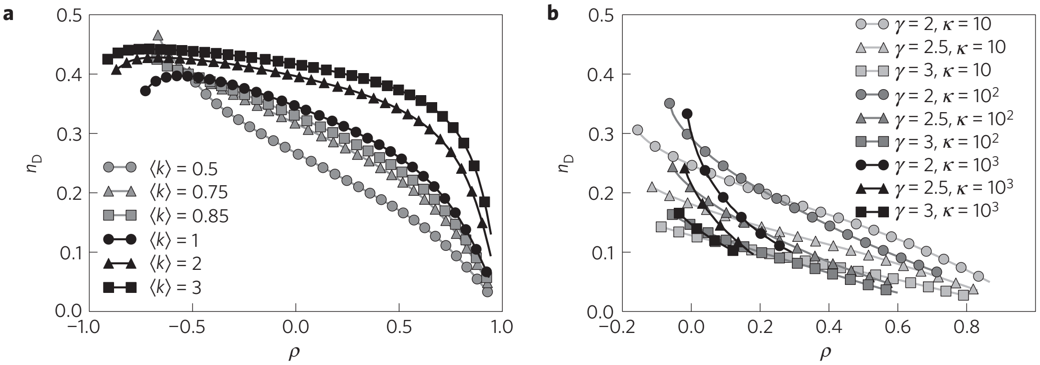

Figure 3 | The dependence of the fraction of driver nodes n_D on the one-point degree correlation ρ. a, n_D in Erdős–Rényi networks with different average degree ⟨k⟩ as a function of ρ. **b**, Fraction of driver nodes n_D in scale-free networks with different exponents γ as a function of ρ. On both panels, every fifth data point is marked by a symbol. Each data point was obtained by averaging at least 20 different realizations of the network model; error bars were omitted as they were smaller than the symbols. Note that it is hard to introduce negative degree correlations in the case of scale-free networks and none of our test runs managed to decrease the correlation below −0.2.

SBD on a network $G(V, E)$ can be determined by selecting the divergent vertices of G and one arbitrary vertex from each balanced component. We also showed that each divergent vertex must control $d_v^+ - d_v^-$ of its outbound edges, and the selected vertices in each balanced component must control only one of its outbound edges (see Supplementary Information for the formal proofs). This implies that the number of driver nodes and the number of driven edges (that is, the number of independent input signals) required to control the SBD is almost completely determined by the joint degree distribution of the network. This is in concordance with the results of Liu *et al.*[22] for the linear time-invariant nodal dynamics.

Controllability of real networks

We have determined the set of driver nodes under the SBD for 38 real networks, classified into 11 categories, and compared the fraction of driver nodes n_D and the fraction of driven edges m_D with the model of Liu *et al.*[22] and with the expected values of n_D after different types of randomizations (Table 1; randomization results for m_D are given in the Supplementary Information). A striking difference between the SBD and the model of Liu *et al.*[22] can be seen for two classes of networks. Regulatory networks such as the transcriptional regulatory network of *Echerichia coli* (TRN-EC-2 (ref. 25)) and *Saccharomyces cerevisiae* (TRN-Yeast-1 (ref. 26), TRN-Yeast-2 (ref. 25)) and the ownership network of US telecommunications and media corporations (Ownership-USCorp (ref. 27)) turned out to be well-controllable under the SBD using a few driver nodes only, but they are hard to control in the linear nodal dynamics. This can be explained by the fundamental difference between the two models. In the linear nodal dynamics, a driven node may not influence its subordinates independently of each other, thus the presence of out-hubs in a network degrades its controllability significantly. In the SBD, out-hubs behave the opposite way, allowing one to control many state variables with a single out-hub. It follows that driver nodes prefer out-hubs in the SBD, whereas they are shown to avoid hubs in the linear nodal dynamics. Therefore, hubs have an important role not only in maintaining the connectivity of a network in the case of random failures[28–30] and containing epidemic spreading[31,32], but they also make it possible to control the network efficiently with a smaller number of driver nodes. However, note that this comes at a cost: although one has to control a few nodes only, each such node may have to emit multiple independent input signals along different outbound edges, as reflected by the markedly higher m_D values for regulatory networks compared to their low n_D values. (See Supplementary Information for a more detailed discussion.)

The other class of networks with the largest difference between the two models is the case of intra-organizational networks[33–35]. In the model used by Liu *et al.*, all these networks can be controlled by at most three nodes. On closer examination, it turns out that 75–80% of the connections in each of these networks is reciprocal, that is, an edge exists between vertices A and B in both directions. A reciprocal edge pair can easily form a bud in a maximum matching, requiring no driver node on its own; therefore high reciprocity in a network always implies a low fraction of driver nodes in the linear nodal dynamics, whereas this is not necessarily true for the SBD. Dense networks may also show a high fraction of driver nodes for the SBD and a low fraction of driver nodes for the linear nodal dynamics: as SBD assumes a state variable for each edge, such networks are harder to control under the SBD than under the linear nodal dynamics because of the large number of edges and the high probability of reciprocal edge pairs.

The difference between the nodal and edge dynamics is probably the most prominent in highly hierarchical, tree-like networks, where the presence of central out-hubs rapidly increase the required number of driver nodes for the linear nodal dynamics of Liu *et al.*, although the same out-hubs allow efficient control of many subordinate nodes and thus decrease the required number of driver nodes in the SBD. Such hierarchies are ubiquitous in nature and society[36], from scales as small as gene regulatory networks[25,26], through leader–follower relationships of flocking pigeons[37], to the large-scale organization of some man-made social structures such as the Wikipedia talk network[38] or the ownership network of US media and telecommunications corporations[27]. The presence or absence of hierarchy thus seems to be an important contributing factor to the controllability properties of large dynamical systems.

Comparing the fraction of driver nodes for the SBD with the randomized variants reveals that, in most cases, the fraction of driver nodes required to control a random Erdős–Rényi network[39,40] of the same size is larger than the fraction of driver nodes for the real-world network, suggesting that the structure of these networks is at least partially optimized for controllability. Notable exceptions are electronic circuits[25], the neural network of *Caenorhabditis elegans*[4,41], most World Wide Web networks[8,42–44], and intra-organizational networks[33–35]. Preserving the in- and out-degree distributions (but not the joint distribution) brings the fraction of driver nodes closer to the observed one after randomization, and keeping the joint degree distribution makes the fraction of driver nodes practically the same up to a difference of ±0.002 in the networks we have studied, confirming that the effect of balanced components on the fraction of driver nodes is indeed negligible for large real-world networks. Edge deletion experiments (see Supplementary Information) also indicate that the optimal control configurations in the studied networks are robust to single link failures, as the networks mostly remain controllable with the same number of driver nodes after the removal of a single edge.

Analytical results for model networks

The dependence of n_{D} on the joint degree distribution allows us to derive analytical formulae for the expected fraction of driver nodes for a wide variety of model networks (see Supplementary Information for the exact derivations). For Erdős–Rényi digraphs[39,40] with n vertices and an edge probability of p, n_{D} is given as follows:

$$n_{\mathrm{D}}^{\mathrm{ER}} = \frac{1}{2} - \frac{\mathrm{e}^{-2\langle k\rangle}}{2} I_0(2\langle k\rangle)$$

where $\langle k\rangle = np$ is the average in- and out-degree and $I_0(x)$ is the modified zero-order Bessel function of the first kind. The function converges rapidly to 0.5 as $\langle k\rangle$ increases. Similar results are obtained for graphs with independent exponential in- and out-degree distributions $C\mathrm{e}^{-k/\kappa}$, where $\kappa = 1/\log((1+\langle k\rangle)/\langle k\rangle)$:

$$n_{\mathrm{D}}^{\mathrm{exp}} = \frac{\langle k\rangle}{2\langle k\rangle + 1}$$

which also approaches 0.5 rapidly as $\langle k\rangle \to \infty$ (Fig. 2a). For power-law distributed digraphs[5,45] with $\mathbf{P}(d_v^+ = k) = \mathbf{P}(d_v^- = k) = Ck^{-\gamma}\mathrm{e}^{-k/\kappa}$, n_{D} is given by

$$n_{\mathrm{D}}^{\mathrm{power}} = \frac{1}{2} - \frac{\mathrm{Li}_{2\gamma}(\mathrm{e}^{-2/\kappa})}{2\mathrm{Li}_{\gamma}(\mathrm{e}^{-1/\kappa})^2}$$

where $\mathrm{Li}_s(z)$ is the base s polylogarithm function. As $\kappa \to \infty$, this converges to

$$n_{\mathrm{D}}^{\mathrm{power}} = \frac{1}{2} - \frac{\zeta(2\gamma)}{2\zeta(\gamma)^2}$$

(where $\zeta(x)$ is the Riemann zeta function) in the absence of any exponential cutoff (Fig. 2b). The Supplementary Information also contains the analytical treatment of k-regular networks.

It is worthwhile to compare these analytical results with those of Liu *et al.*[22], who have found that the fraction of driver nodes n_{D} decreases for both Erdős–Rényi and scale-free networks as $\langle k\rangle \to \infty$, whereas these networks behave the opposite way under the SBD. For $\langle k\rangle \to \infty$, the fraction of driver nodes tends to 1/2 for Erdős–Rényi networks and to $1/2 - \zeta(2\gamma)/(2\zeta(\gamma)^2)$ for scale-free networks. The consequence is that denser networks are harder to control (as expected by our intuition), and that scale-free networks with a given $\langle k\rangle$ are easier to control than an Erdős–Rényi network with the same average degree. This can partly be attributed to the higher frequency of short loops[46] in scale-free networks: these loops can be covered by closed walks and do not require extra driver nodes.

The effect of degree correlations

Our analytical results assumed that the in-degree and out-degree of a node are uncorrelated, which was true for all of the model networks we have studied. However, one-point degree correlations in real networks are significantly different from zero[47–51]. Positive one-point degree correlations were recently shown to facilitate information spreading[51], thus it is not unreasonable to assume that such correlations may affect the controllability properties of a given network. To answer this question, we have performed simulations on Erdős–Rényi networks and scale-free networks with an exponential cutoff and varied the one-point degree correlation (Fig. 3a,b and Methods).

The plots in Fig. 3 clearly show a general trend: increasing the correlation between the in- and out-degrees decreases the fraction of driver nodes. Negative one-point correlations yield a higher fraction of driver nodes because these networks are unlikely to contain balanced nodes: a vertex either has a high in-degree and a low out-degree or a high out-degree and a low in-degree. In other words, negative correlations indicate a clear separation of responsibilities between the nodes of the network: divergent nodes are strongly divergent with a large difference between the out-degree and the in-degree, whereas convergent nodes are strongly convergent. Positive correlations indicate that nodes often represent complex decision processes which map a high-dimensional input space into a similarly high-dimensional output space. Strong positive correlations also yield networks with a higher number of short loops[50], which can then be covered by closed walks that do not require driver nodes on their own.

As happens so often in scientific research, the framework we have presented raises more questions than answers. For instance, it is yet unknown how the SBD would behave in the presence of noise or nonlinearity, or in cases when it is enough to control only a subset of the state variables (output controllability) and only ensure that the uncontrollable ones have stable dynamics (stabilizability). However, as we have shown, even the first steps along our approach could be used to deepen our understanding of the origins of controllability of real-world networks.

Methods

Analytical solutions for n_{D} in model networks. Analytical formulae for the expected fraction of driver nodes in the model networks we have studied are based on two observations. First, we note that the expected fraction of driver nodes depends almost completely on the joint degree distribution of the network, therefore it is enough to calculate the expected density of divergent nodes in the model networks. Second, we also note that the probability of observing a node with in-degree k and out-degree l is equal to that of a node with in-degree l and out-degree k. This means that the density of divergent nodes will be equal to the density of convergent nodes, and either of them can be obtained by subtracting the density of balanced nodes from 1 and multiplying the result by 1/2. The exact derivations for the model networks are presented in the Supplementary Information.

Generating networks with correlated in- and out-degrees. Correlated in- and out-degrees in Erdős–Rényi and scale-free networks were achieved as follows. First, we generated an instance of the network model with $n = 10^5$ nodes and calculated the in- and out-degree sequences. These instances were obviously uncorrelated as neither the Erdős–Rényi model nor the configuration model (which we have used to generate scale-free networks) introduces correlations between the in- and out-degree of the same node. Next, while keeping the in-degree sequence intact, we started swapping elements in the out-degree sequence randomly such that only those swaps were performed that increased the correlation. The process was continued until we were not able to increase the correlation any more in the last t steps (where $t = 10^4$ in our simulations). A similar greedy algorithm was executed from the original degree sequences in the opposite direction, performing swaps only if it decreased the correlation, terminating when it was not possible to decrease the correlation any more in the last t steps. The fraction of driver nodes n_D was then calculated in the original configuration and whenever the absolute difference of the calculated in- and out-degree correlation between the last examined state and the current state became larger than 0.01.

Received 16 November 2011; accepted 24 April 2012; published online 27 May 2012

References

1. Newman, M. The structure and function of complex networks. *SIAM Rev.* **45,** 167–256 (2003).
2. Albert, R. & Barabási, A-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74,** 47–97 (2002).
3. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Huang, D-U. Complex networks: Structure and dynamics. *Phys. Rep.* **424,** 175–308 (2006).
4. Watts, D. & Strogatz, S. Collective dynamics of ‘small-world’ networks. *Nature* **393,** 440–442 (1998).
5. Barabási, A-L. & Albert, R. Emergence of scaling in random networks. *Science* **286,** 509–512 (1999).
6. Amaral, L., Scala, A., Barthelémy, M. & Stanley, H. Classes of small-world networks. *Proc. Natl Acad. Sci. USA* **97,** 11149–11152 (2000).
7. Newman, M., Watts, D. & Strogatz, S. Random graph models of social networks. *Proc. Natl Acad. Sci. USA* **99** (suppl 1), 2566–2572 (2002).
8. Leskovec, J. & Faloutsos, C. *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining* (KDD, 2005).
9. Fortunato, S. Community detection in graphs. *Phys. Rep.* **486,** 75–174 (2010).
10. Newman, M. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69,** 026113 (2004).
11. Palla, G., Derényi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435,** 814–818 (2005).
12. Ebel, H., Davidsen, J. & Bornholdt, S. Dynamics of social networks. *Complexity* **8,** 24–27 (2002).
13. Luscombe, N., Madan Babu, M., Yu, H., Snyder, M., Teichmann, A. & Gerstein, M. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431,** 308–312 (2004).
14. Prill, R., Iglesias, P. & Levchenko, A. Dynamic properties of network motifs contribute to biological network organization. *PLoS Biol.* **3,** 1881–1892 (2005).
15. Palla, G., Barabási, A-L. & Vicsek, T. Quantifying social group evolution. *Nature* **446,** 664–667 (2007).
16. Yu, W., Chen, G. & Lü, J. On pinning synchronization of complex dynamical networks. *Automatica* **45,** 429–435 (2009).
17. Lombardi, A. & Hörnquist, M. Controllability analysis of networks. *Phys. Rev. E* **75,** 056110 (2007).
18. 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).
19. Kalman, R. Mathematical description of linear dynamical systems. *J. Soc. Indus. Appl. Math. Ser. A* **1,** 152–192 (1963).
20. Sontag, E. *Mathematical Control Theory* (Springer, 1998).
21. Slotine, J-J. & Li, W. *Applied Nonlinear Control* (Prentice-Hall, 1991).
22. Liu, Y., Slotine, J. & Barabási, A. Controllability of complex networks. *Nature* **473,** 167–173 (2011).
23. Lin, C. Structural controllability. *IEEE Trans. Automat. Contr.* **19,** 201–208 (1974).
24. Shields, R. & Pearson, J. Structural controllability of multi-input linear systems. *IEEE Trans. Automat. Contr.* **21,** 203–212 (1976).
25. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. Network motifs: Simple building blocks of complex networks. *Science* **298,** 824–827 (2002).
26. Balaji, S., Babu, M., Iyer, L., Luscombe, N. & Aravind, L. Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J. Mol. Biol.* **360,** 213–227 (2006).
27. Norlen, K., Lucas, G., Gebbie, M. & Chuang, J. *Proc. Int. Telecommunications Society 14th Biennial Conference (ITS2002)* (2002).
28. Albert, R., Jeong, H. & Barabási, A-L. Error and attack tolerance of complete networks. *Nature* **406,** 378–382 (2000).
29. Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* **85,** 4626–4628 (2000).
30. Jeong, H., Mason, S., Barabási, A-L. & Oltvai, Z. Lethality and centrality in protein networks. *Nature* **411,** 41–42 (2001).
31. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86,** 3200–3203 (2001).
32. Pastor-Satorras, R. & Vespignani, A. Immunization of complex networks. *Phys. Rev. E* **65,** 036104 (2002).
33. Cross, R. & Parker, A. *The Hidden Power of Social Networks* (Harvard Business School Press, 2004).
34. Freeman, S. & Freeman, L. *Social Science Research Reports 46.* Technical report (Univ. California, 1979).
35. Nepusz, T., Petróczi, A., Négyessy, L. & Bazsó, F. Fuzzy communities and the concept of bridgeness in complex networks. *Phys. Rev. E* **77,** 016107 (2008).
36. Mones, E., Vicsek, L. & Vicsek, T. Hierarchy measure for complex networks. *PLoS ONE* **7,** e33799 (2012).
37. Nagy, M., Ákos, Z., Biro, D. & Vicsek, T. Hierarchical group dynamics in pigeon flocks. *Nature* **464,** 890–893 (2010).
38. Leskovec, J., Huttenlocher, D. & Kleinberg, J. *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (CHI)* 1361–1370 (ACM, 2010).
39. Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5,** 17–60 (1960).
40. Bollobás, B. *Cambridge Studies in Advanced Mathematics* 2nd edn (Cambridge Univ. Press, 2001).
41. Achacoso, T. & Yamamoto, W. *AY’s Neuroanatomy of* C. elegans *for Computation* 1st edn (CRC Press, 1992).
42. Palla, G., Farkas, I., Pollner, P., Derényi, I. & Vicsek, T. Directed network modules. *New J. Phys.* **9,** 186 (2007).
43. Adamic, L. & Glance, N. *Proc. WWW-2005 Workshop on the Weblogging Ecosystem* (ACM, 2005).
44. Albert, R., Jeong, H. & Barabási, A-L. Diameter of the World Wide Web. *Nature* **401,** 130–131 (1999).
45. Caldarelli, G. *Scale-Free Networks: Complex Web in Nature and Technology* (Oxford Univ. Press, 2007).
46. Bianconi, G. & Marsili, M. Loops of any size and hamilton cycles in random scale-free networks. *J. Stat. Mech.* P06005 (2005).
47. Schwartz, N., Cohen, R., ben Avraham, D., Barabási, A-L. & Havlin, S. Percolation in directed scale-free networks. *Phys. Rev. E* **66,** 015104 (2002).
48. Nanavati, A. A., Gurumurthy, S., Das, G., Chakraborty, D., Dasgupta, K., Mukherjea, S. & Joshi, A. *Proc. 2006 ACM CIKM Int. Conf. Information and Knowledge Management* 435–444 (ACM, 2006).
49. Serrano, M., Maguitman, A., Boguñá, M., Fortunato, S. & Vespignani, A. Decoding the structure of the WWW: Comparative analysis of web crawls. *ACM Trans. Web* **1,** 10 (2007).
50. Bianconi, G., Gulbahce, N. & Motter, A. Local structure of directed networks. *Phys. Rev. Lett.* **100,** 118701 (2008).
51. Peruani, F. & Tabourier, L. Directedness of information flow in mobile phone communication networks. *PLoS ONE* **6,** e28860 (2011).

Acknowledgements

This research was supported by the European Union, European Research Council COLLMOT project and the European Social Fund No: TAMOP 4.2.1/B-09/1/KMR-2010-0003. We are grateful to E. Mones for useful discussions.

Author contributions

T.N. devised the SBD model and performed analytical calculations and simulations. T.V. initiated the research and supervised the project. T.N. and T.V. wrote the paper.

Additional information

The authors declare no competing financial interests. Supplementary information accompanies this paper on www.nature.com/naturephysics. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.V.