

SCIENTIFIC REPORTS



OPEN

Minimum energy control for complex networks

Gustav Lindmark & Claudio Altafini

The aim of this paper is to shed light on the problem of controlling a complex network with minimal control energy. We show first that the control energy depends on the time constant of the modes of the network, and that the closer the eigenvalues are to the imaginary axis of the complex plane, the less energy is required for complete controllability. In the limit case of networks having all purely imaginary eigenvalues (e.g. networks of coupled harmonic oscillators), several constructive algorithms for minimum control energy driver node selection are developed. A general heuristic principle valid for any directed network is also proposed: the overall cost of controlling a network is reduced when the controls are concentrated on the nodes with highest ratio of weighted outdegree vs indegree.

Understanding the basic principles that allow to control a complex network is a key prerequisite in order to move from a passive observation of its functioning to the active enforcement of a desired behavior. Such an understanding has grown considerably in recent years. For instance, classical control-theoretical notions like structural controllability have been used to determine a minimal number of driver nodes (i.e., nodes of the network which must be endowed with control authority) that guarantee controllability of a network¹. Several works have explored the topological properties underlying such notions of controllability^{2–7}, or have suggested to use other alternative controllability conditions^{8–10}. Several of these approaches are constructive, in the sense that they provide receipts on how to identify a subset of driver nodes that guarantees controllability. However, controllability is intrinsically a yes/no concept that does not take into account the effort needed to control a network¹¹. A consequence is that even if a network is controllable with a certain set of driver nodes, the control energy that those nodes require may result unrealistically large. Achieving “controllability in practice” i.e., with a limited control energy, is a more difficult task, little understood in terms of the underlying system dynamics of a network. In addition, in spite of the numerous attempts^{9,11–20}, no clear strategy has yet emerged for the related problem of selecting the driver nodes so as to minimize the control energy.

The aim of this paper is to tackle exactly these two issues, namely: i) to shed light on what are the dynamical properties of a network that determine the input energy needed to control a network; and ii) to develop driver node placement strategies requiring minimum control energy.

We show in the paper that for linear dynamics the natural time constants of the modes of the system are key factors in understanding how much energy a control must use. Since the time constants of a linear system are inversely proportional to the real part of its eigenvalues, systems that have eigenvalues near the imaginary axis (i.e., slow dynamics and nearly oscillatory behavior) are easier to control than systems having eigenvalues with large real parts (i.e., fast dynamics, stable or unstable), regardless of the specific metric used to quantify the control energy (in the paper we use Gramian-based measures^{17,21}).

The main driver node selection strategy we propose in the paper is based on ranking nodes according to the ratio between weighted outdegree and weighted indegree. This strategy, which is inspired by topological considerations but is still lacking a complete theoretical justification, is very different from choosing a minimum number of driver nodes that guarantee structural controllability. We show that whenever controllability is not an issue (e.g. when the network is strongly connected), such method systematically outperforms a random driver node assignment even by orders of magnitude, in all the considered control metrics and in all considered networks (typically of size $\sim 10^3$ in this paper).

Methods like the one proposed in this paper are based on computing eigenvalues, and hence are necessarily quantitative, i.e., require to assign weights to the edges of a network. However, we show in the paper that our quantitative analysis sheds light on the intrinsic weakness of purely qualitative methods like those based on structural controllability: there are cases in which the Gramian associated to structural controllability is closer to singularity than the one associated with our ranking, regardless of the latter passing or less the structural

Division of Automatic Control, Dept. of Electrical Engineering, Linköping University, SE-58183, Linköping, Sweden. Correspondence and requests for materials should be addressed to C.A. (email: claudio.altafini@liu.se)

controllability test. In terms of energy this translates into the following paradox: the control energy of a structurally controllable system can be higher than that of a “structurally uncontrollable” one.

We also show that for the special case of networks of coupled harmonic oscillators, which have purely imaginary spectra, it is possible to obtain explicit criteria for minimum energy driver node placement. Unlike for our main driver node selection strategy, based on weighted outdegree/indegree ratio, which is supported mainly by heuristic topological considerations, the strategies for the special case of purely imaginary spectra rely instead on arguments such as diagonal dominance of the Gramian, and are supported by a fully theoretical analysis. In this case, essentially every Gramian-based controllability measure leads to a tractable minimum energy driver node placement strategy. One of these strategies in particular is exact, and turns out to be very similar to the node ranking mentioned above.

Methods

Complete Controllability: merging Reachability and Controllability to 0. A linear system

$$\dot{x} = Ax + Bu \quad (1)$$

is controllable if there exists an input $u(t)$ that transfers the n -dimensional state vector $x(t)$ from any point x_0 to any other point x_f in \mathbb{R}^n . The Kalman rank condition for controllability, $\text{rank}([BAB^2B^3 \dots A^k B]) = n$ for k sufficiently large, only provides a yes/no answer but does not quantify what is the cost, in term of input effort, of such state transfer. Once controllability is guaranteed, a possible approach to investigate “controllability in practice” consists in quantifying the least energy that a control requires to accomplish the state transfer, i.e., in minimizing $\mathcal{E}(t_f) = \int_0^{t_f} \|u(\tau)\|^2 d\tau$ for $u(t)$ that maps x_0 in x_f in a certain time t_f . Although in this work we focus on the infinite horizon case (i.e., $t_f \rightarrow \infty$), for pedagogical reasons it is instructive to start the analysis from the finite horizon case. In fact, for linear systems like (1), a closed form solution to this problem exists when t_f is finite, and the resulting cost is

$$\mathcal{E}(t_f) = (x_f - e^{At_f}x_0)^T W_r^{-1}(t_f) (x_f - e^{At_f}x_0), \quad (2)$$

where the matrix $W_r(t_f) = \int_0^{t_f} e^{A\tau} B B^T e^{A^T \tau} d\tau$ is called the *reachability Gramian*²². The control that achieves the state transfer $x_0 \rightarrow x_f$ with minimal cost can be computed explicitly:

$$u(t) = B^T e^{A^T(t_f-t)} W_r^{-1}(t_f) (x_f - e^{At_f}x_0), \quad t \in [0, t_f]. \quad (3)$$

Various metrics have been proposed to quantify the difficulty of the state transfer based on the Gramian²¹, like its minimum eigenvalue $\lambda_{\min}(W_r)$, its trace $\text{tr}(W_r)$, or the trace of its inverse $\text{tr}(W_r^{-1})$, see SI for a more detailed description.

We would like now to describe how (2) depends on the eigenvalues of A . In order to do that, one must observe that (2) can be decomposed into contributions originating from two distinct problems: (1): *controllability-from-0* (or *reachability*, as it is normally called in control theory²²) and (2): *controllability-to-0*. The first problem consists in choosing $x_0 = 0$, in which case (2) reduces to $\mathcal{E}_r(t_f) = x_f^T W_r^{-1}(t_f) x_f$, while in the second $x_f = 0$ leads to $\mathcal{E}_c(t_f) = x_0^T W_c^{-1}(t_f) x_0$ where $W_c(t_f) = e^{-A^T t_f} W_r(t_f) e^{-A t_f}$ is a second Gramian, called the *controllability Gramian*. The two problems mentioned above, reachability and controllability-to-0, are characterized by different types of difficulties when doing a state transfer, all related to the stability of the eigenvalues of A . For instance the reachability problem is difficult along the stable eigendirections of A because the control has to win the natural decay of the unforced system to 0, while the unstable eigenvalues help the system escaping from 0 by amplifying any small input on the unstable eigenspaces, see Fig. 1 for a graphical explanation. The surfaces of $\mathcal{E}_r(t_f)$ shown in Fig. 1(a) reflect these qualitative differences. On the contrary, the influence of the eigenvalues of A is the opposite for the controllability-to-0 problem shown in Fig. 1(b). Hence if we want to evaluate the worst-case cost of a transfer between any x_0 and any x_f (a problem sometimes referred to as *complete controllability*²³), we have to combine the difficult cases of the two situations just described.

Measuring control energy through the infinite-time horizon mixed Gramian. In the complete controllability problem, all the cases leading to a high control energy can be taken into account by combining the two Gramians W_r and W_c into a “mixed” Gramian W_m obtained splitting A into its stable and antistable parts and forming a reachability subGramian for the former and a controllability subGramian for the latter^{24–26}. Such Gramian can be computed in closed form only when the time of the transfer tends to infinity. In the infinite time horizon, in fact, both W_r and W_c diverge, but their inverses are well-posed and depend only on the stable modes the former and the unstable modes the latter. These are the parts composing the inverse of W_m , see Fig. 1(c). Computing W_m and its inverse requires solving jointly two Lyapunov equations. The procedure is explained in detail in the SI. Assume that the spectrum of A contains k , $0 \leq k \leq n$, eigenvalues with negative real part, and $n - k$ eigenvalues with positive real part (and no purely imaginary eigenvalues). Then there exist a change of basis V bringing A into the form:

$$\begin{bmatrix} \bar{A}_1 & 0 \\ 0 & \bar{A}_2 \end{bmatrix} = V A V^{-1} \quad (4)$$

and, correspondingly,

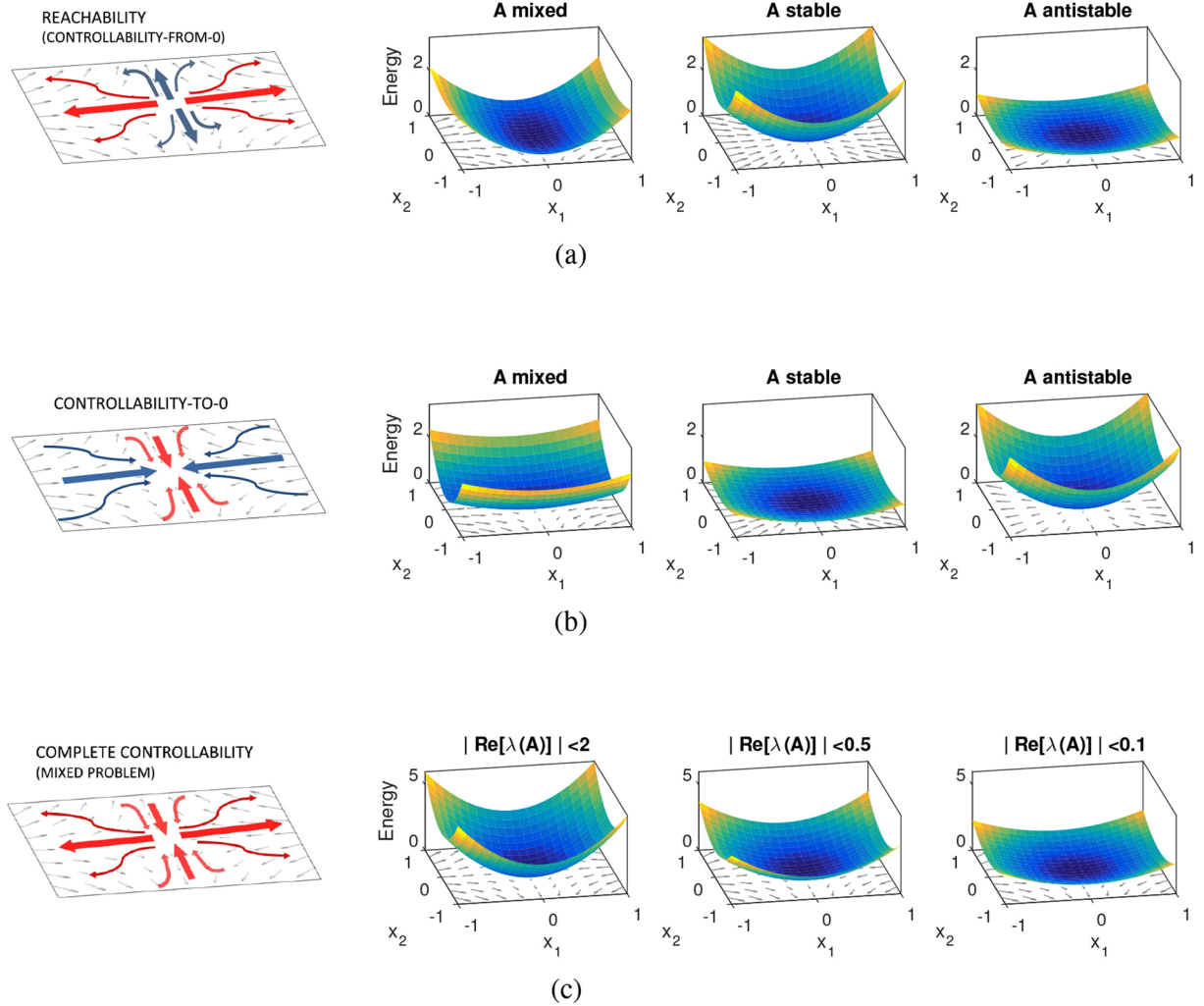


Figure 1. Reachability, Controllability-to-0 and Complete Controllability problems. (a) The reachability (or controllability-from-0) problem is difficult along the stable eigendirections of A (red curves in the leftmost panel) and easy along the unstable ones (blue). This is reflected in the surfaces of $\mathcal{E}_r(t_f) = x_f^T W_r^{-1}(t_f) x_f$ shown in the 3 rightmost panels. In particular, the reachability problem requires limited control energy when A is antistable (rightmost panel). (b) The controllability-to-0 problem is difficult along the unstable eigendirections of A (red) and easy along the stable ones (blue). The input energy surfaces, $\mathcal{E}_c(t_f) = x_0^T W_c^{-1}(t_f) x_0$, reflect these properties. The case of A stable requires the least control energy. (c) The problem studied in this paper, complete controllability, is a mixture of the two cases, collecting the worst-case situations of both. When the real part of the eigenvalues of A is squeezed towards the imaginary axis as in the right panels of Fig. 2(a), the input energy reduces accordingly.

$$\begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \end{bmatrix} = V B$$

with $\text{Re}[\lambda(\bar{A}_1)] < 0$ and $\text{Re}[\lambda(\bar{A}_2)] > 0$. In the new basis, the following two Lyapunov equations hold (see the SI for details on how they are obtained)

$$\begin{bmatrix} \bar{A}_1 & 0 \\ 0 & -\bar{A}_2 \end{bmatrix} \bar{W} + \bar{W} \begin{bmatrix} \bar{A}_1 & 0 \\ 0 & -\bar{A}_2 \end{bmatrix}^T + \begin{bmatrix} \bar{B}_1 \bar{B}_1^T & 0 \\ 0 & \bar{B}_2 \bar{B}_2^T \end{bmatrix} = 0 \tag{5}$$

with

$$\bar{W}_m = \begin{bmatrix} \bar{W}_{1,r} & 0 \\ 0 & \bar{W}_{2,c} \end{bmatrix}$$

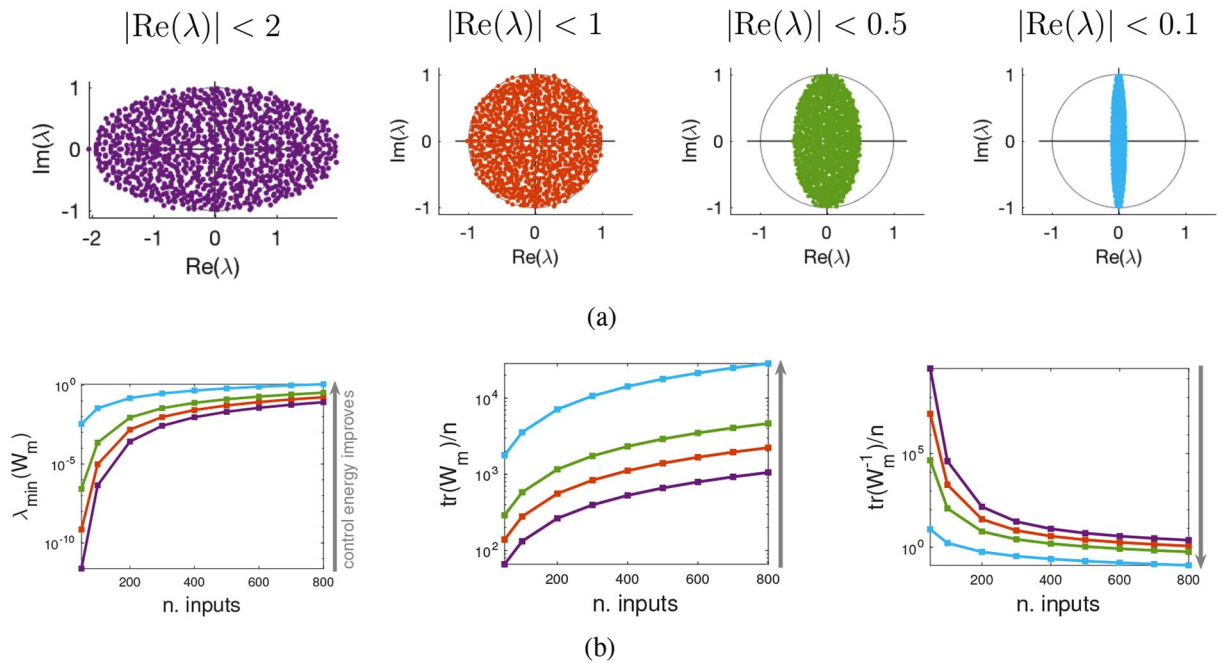


Figure 2. Real part of the eigenvalues and control energy. **(a)** Eigenvalues of a random matrix. The circular/elliptic laws allow to obtain state matrices A with eigenvalues in prescribed locations, for instance with predetermined real part. **(b)** Control energy for various metrics when the number of (randomly chosen) inputs grows and $\text{Re}[\lambda(A)]$ changes. The data show a mean over 100 realizations of dimension $n = 1000$ (for each realization 100 different edge weights assignments are considered). The color code is as in **(a)**. Of the three metrics used to measure the control energy, $\lambda_{\min}(W_m)$ (left) and $\text{tr}(W_m)$ (middle) should both be maximized, while $\text{tr}(W_m^{-1})$ (right) should be minimized in order to improve the control energy (see direction of the arrow on the left of each panel). For all three metrics, the performances are strictly a function of the position of the eigenvalues of A . The minimum of the control energy is achieved when the eigenvalues have very small real part (cyan) and worsen with growing real part, following the order: cyan, green, red, violet.

the mixed Gramian (the subGramians $W_{1,r}$ and $W_{2,c}$ are derived in detail in the SI). Following²⁴, the expression of the mixed Gramian in the original basis is $W_m = V^{-1} \bar{W}_m V^{-T}$. By construction, the mixed Gramian matrix W_m always exists when A has no purely imaginary eigenvalues, and it summarizes the infinite-horizon contribution of the stable eigenvalues to the reachability problem and of the unstable eigenvalues to the controllability to 0 problem (i.e., all cases leading to a high control energy). The three figures of merit $\lambda_{\min}(W_m)$, $\text{tr}(W_m)$ and $\text{tr}(W_m^{-1})$ are the measures that are mostly used in the paper to quantify the control energy. Although a finite-horizon version of W_m (and W_m^{-1}) can be deduced from the infinite horizon ones (see SI), in this paper we deal exclusively with the infinite horizon case, as often done in the literature^{17,22,24}.

Constructing networks with preassigned spectra. The circular and elliptic laws²⁷ are used in the paper to construct random networks with eigenvalues in desired regions. The circular law states that a matrix A of entries a_{ij}/\sqrt{n} where a_{ij} are i.i.d. random variables with zero-mean and unit variance has spectral distribution which converges to the uniform distribution on the unit disk as $n \rightarrow \infty$, regardless of the probability distribution from which the a_{ij} are drawn. A numerical example is shown in Fig. 2(a) (second panel from left). A random matrix is typically a full matrix, meaning that the underlying graph of interactions is fully connected. The circular law is however valid also for sparse matrices, for instance for Erdős-Rényi (ER) topologies. If p is the edge probability, then $A = (a_{ij})/\sqrt{p \cdot n}$ still has eigenvalues distributed uniformly on the unit disk, see Fig. S1(a).

A generalization of the circular law is the *elliptic law*, in which the unit disk containing the eigenvalues of A is squeezed in one of the two axes. To do so, the pairs of entries $\{a_{ij}, a_{ji}\}$ of A have to be drawn from a bivariate distribution with zero marginal means and covariance matrix expressing the compression of one of the two axes. Various examples of elliptic laws are shown in the panels of Fig. 2(a). Also elliptic laws generalize to sparse matrices, see Fig. S1(a).

Results

Control energy as a function of the real part of the eigenvalues of A . In a driver node placement problem, the inputs affect a single node, hence the columns of B are always elementary vectors, i.e., vectors having one entry equal to 1 and the rest equal to 0. When A is a random matrix, the underlying graph is generically fully connected, hence issues like selection of the minimal number of driver nodes based on the connectivity of the graph become irrelevant. In this work we always consider cases in which the driver nodes we have available are enough to guarantee controllability. Having disentangled the problem from topological aspects, the dependence

of the control energy from other factors, like the spectrum of A , becomes more evident and easier to investigate. If for instance we place driver nodes at random and use the infinite-horizon mixed Gramian W_m to form the various energy measures mentioned above for quantifying the control energy, then we have the results shown in Fig. 2(b). As expected, all indicators improve with the number of inputs, because the control effort gets spread over more nodes. What is more interesting is that when we repeat the computation for the various spectral distributions of Fig. 2(a), the result is that the control energy decreases when the (absolute value of the) real part of the eigenvalues of A decreases. All measures give an unanimous answer on this dependence, regardless of the number of inputs considered. In particular, when A has eigenvalues which are very close to the imaginary axis (rightmost panel of Fig. 2(a) and cyan curves in Fig. 2(b)) then $\lambda_{\min}(W_m)$ is bigger (i.e., the worst-case controllability direction is easiest to control), but also $\text{tr}(W_m)$ increases and $\text{tr}(W_m^{-1})$ decreases (meaning, in both cases, that the average energy needed for controllability on all directions decreases).

An identical result is valid also for sparse matrices. In particular, for ER graphs with edge probability $p = 0.05$ and coefficients from a bivariate normal distribution (yielding elliptic laws as in Fig. S1(b)), the various norms used to quantify the control energy are shown in Fig. S1(b). Their pattern is identical to the full graph case of Fig. 2(b).

The computations shown in Fig. 2(b) are performed with the infinite-horizon mixed Gramian W_m described in the SI, because such W_m can be easily computed in closed form. A finite-horizon $W_m(t_f)$ can be approximately obtained from it, but the arbitrariness of t_f makes it hard to set up an unbiased comparison of the various spectral distributions of A of Fig. 2(a), which are characterized by widely different time constants (inversely correlated to the amplitude of the real part of $\lambda(A)$). Observe in Fig. S2 how the various measures of controllability computed with a finite-time $W_m(t_f)$ tend all to the infinite-time W_m but with different speeds.

Driver node placement based on weighted connectivity. In the analysis carried out so far the driver nodes were chosen randomly. A topic that has raised a remarkable interest in recent times (and which is still open in our knowledge) is devising driver node placement strategies that are efficient in terms of input energy^{12–17,19,20}. If we consider as weighted indegree and outdegree at node i the sum of the weights in absolute value of all incoming or outgoing edges, i.e., $w_{\text{in}}(i) = \sum_{j=1}^n |a_{ji}|$ and $w_{\text{out}}(i) = \sum_{j=1}^n |a_{ij}|$ (a normalization factor such as $\sqrt{p \cdot n}$ can be neglected), see Fig. 3(a), then a strategy that systematically beats random input assignment consists in ranking the nodes according to the ratio $r_w(i) = w_{\text{out}}(i)/w_{\text{in}}(i)$ and placing inputs on the nodes with highest r_w . In Fig. 3(c) the $\lambda_{\min}(W_m)$ of this driver node placement strategy is compared with a random driver node selection. If for all graphs the improvement is minimal, as the graphs become sparser it increases, and for ER networks with $p = 0.01$ the $\lambda_{\min}(W_m)$ obtained by controlling nodes with high r_w is more than twice that of the random choice of controls, see Fig. 3(b). As can be seen in Fig. S3, all measures of input energy show a qualitatively similar improvement. The topology of the network can be used to render the values of r_w more extreme, for instance choosing direct scale-free (SF) graphs with indegree exponent bigger than outdegree exponent²⁸ (and at the same time guaranteeing strong connectivity of the ensuing graphs, in order to avoid problems with minimal controllability), see Fig. 3(a,b). For these skewed SF networks, the improvement in choosing driver nodes with high r_w becomes much more substantial, even of orders of magnitude bigger than a random selection, see Figs 3(c), S4 and S5 for more details.

Application to real-world networks. In order to verify whether the proposed driver node placement strategy is effective also on more realistic data, we tested our algorithm on several networks collected from the literature, and representing complex systems of biological (transcriptional, metabolic and signaling), ecological (food-web), social, transportation and trade type, see Table 1. As for most real datasets, the underlying graphs are not strongly connected, hence controllability is not automatically verified. In order to guarantee it, it is necessary to assign a certain number m_c of driver nodes, for instance chosen according to structural controllability and computed using a maximum matching algorithm¹, see Table 1. If edge weights are assigned (here drawn from a uniform distribution), then in correspondence of such minimum number of driver nodes it is possible to estimate the control energy via the three measures $\lambda_{\min}(W_m)$, $\text{tr}(W_m)$, and $\text{tr}(W_m^{-1})$, see Figs 4, S7 and S8. When instead the driver nodes are chosen according to our ranking strategy r_w , then there is no guarantee of achieving controllability in a structural sense. In theory, lack of structural controllability corresponds to a singular Gramian, but in practice it corresponds to a poorly conditioned Gramian, with $\lambda_{\min}(W_m)$ small and $\text{tr}(W_m^{-1})$ big. One expects then a higher control energy than when structural controllability is guaranteed. This in fact happens in some of the networks we are considering. For instance in the *E. coli-metab* network of Fig. 4(a), indeed when m_e nodes are chosen according to r_w , then if $m_e = m_c$ two of the three control energy measures ($\lambda_{\min}(W_m)$ and $\text{tr}(W_m^{-1})$) are worst than in the structurally controllable case. However, for other networks, such as the *US Airports* network of Fig. 4(b), when $m_e = m_c$ the control energy for driver nodes chosen according to r_w is already better than for structural controllability, and it improves further when $m_e > m_c$. A similar behavior is observed in several of the networks analyzed, see Figs S7 and S8. In all networks, the control energy improves as m_e increases (as expected), and it becomes better than that of the m_c driver nodes required by structural controllability before the r_w -ranked nodes begin to fulfill the structural controllability test. Notice how in nearly all networks $\text{tr}(W_m)$ (which is less sensitive to singularity of W_m) improves when we choose r_w -ranked driver nodes.

An alternative test that can be carried out to evaluate the effect of our ranking strategy on the control energy consists in choosing the m_c driven nodes prescribed by structural controllability, plus extra m_f nodes according to the ranking r_w . Here we have decided to select a number of extra driver nodes equal to $m_f = (n - m_c)/2$ (i.e., half of the remaining nodes), in two different ways: i) according to the ratio r_w , and ii) randomly. The resulting numbers for our networks are given in Fig. 5(a) and in Table 1. The results in terms of control energy are illustrated in Fig. 5(b). For all networks, all three metrics $\lambda_{\min}(W_m)$, $\text{tr}(W_m)$, and $\text{tr}(W_m^{-1})$ are essentially always confirming that

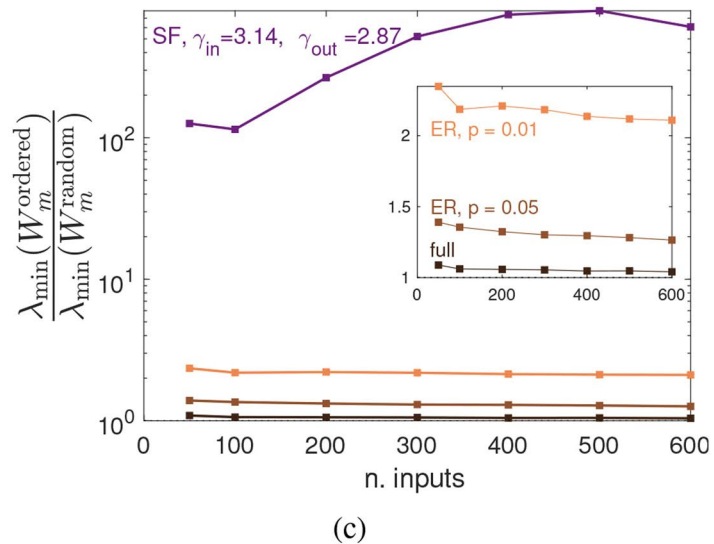
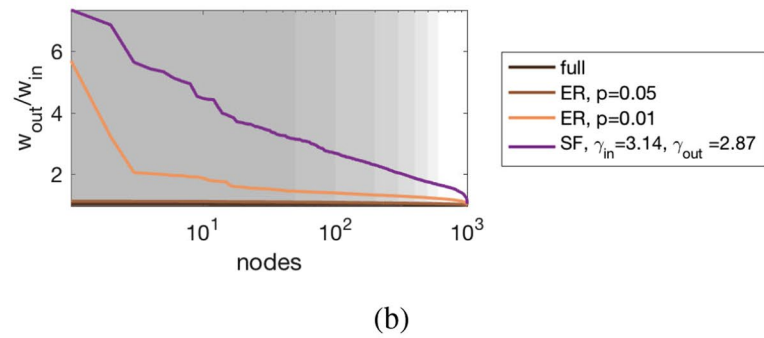
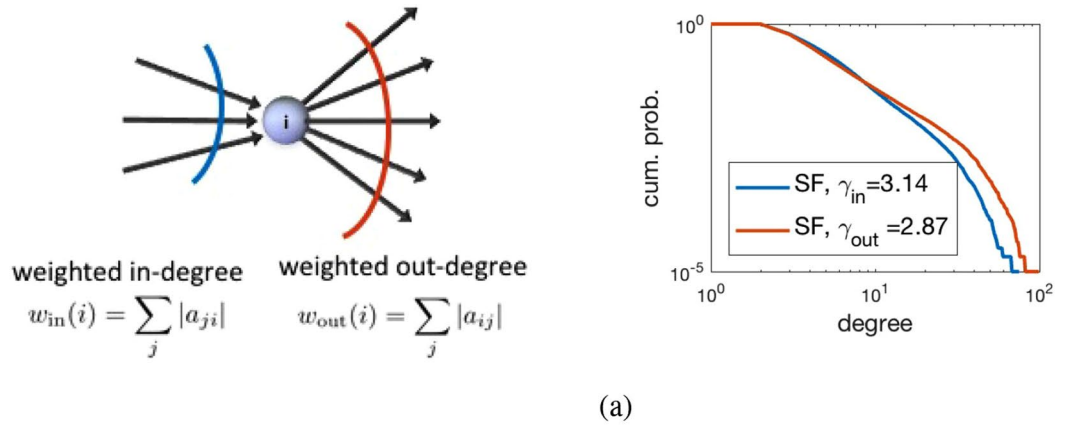


Figure 3. Driver node placement strategy: ranking according to $r_w = w_{out}/w_{in}$. **(a)** Interpretation of the ranking strategy (left panel). The selection of driver nodes starts from those having the highest ratio r_w , meaning those that are dominated by (weighted) out-degree. Networks that are SF directed graphs with indegree exponent bigger than out-degree exponent (in the right panel $\gamma_{in} = 3.14$ and $\gamma_{out} = 2.87$) have a large fraction of nodes with this characteristic. **(b)** The ratio r_w of ranked nodes is shown for ER networks of different densities and for the SF network shown in **(a)**. Indeed for SF the fraction of nodes having high r_w is much bigger than that of ER networks. The shaded areas represent the values of m used in our computations of control energy. **(c)** Comparison between control energy (here measured according to $\lambda_{min}(W_m)$) for the r_w -ranking (denoted $\lambda_{min}(W_m^{ordered})$) and random driver node selection (denoted $\lambda_{min}(W_m^{random})$). The ratio between the two quantities is shown for the ER and SF networks mentioned in **(a,b)**. For ER networks, the improvement in control energy increases with the sparsity of the graph (inset: zoomed comparison in linear scale). For the SF networks, the improvement is remarkably more significant (two orders of magnitude, violet curve, see also Fig. S5 for more details). Measures are means over 100 realizations of size $n = 1000$; for each realization 100 edge weight assignments are tested.

Network	nodes	edges	m_c	m_f	source (ref. n.)
Biological, transcriptional					
<i>E. coli</i>	1623	3515	1472	75	38
Yeast	664	1066	550	57	39
Biological, metabolic					
<i>E. coli</i>	757	6116	102	327	40
Yeast	780	4421	142	319	41
Biological, signalling					
Macrophage	660	1549	179	240	42
Toll-like	672	2194	144	264	43
Food-web					
Florida	128	2106	30	49	44
Mangdry	97	1491	22	37	44
Social					
Moreno highschool	70	366	3	33	45
Advogato	5042	47067	1194	1924	46
Transport and Trade					
US Airport	1572	28235	582	495	47
Wheat	166	1782	35	65	48

Table 1. Various networks used in this study. m_c represents the number of driver nodes needed to guarantee structural controllability. m_f the number of extra driver nodes (in addition the previous m_c) selected to reduce the control energy.

the driver node selection strategy based on r_w is outperforming significantly a random node assignment. In particular, for the two nonlinear metrics ($\lambda_{\min}(W_m)$ and $\text{tr}(W_m^{-1})$), the improvement is often of several orders of magnitude, i.e., even better than in the skewed SF networks discussed in Fig. 3.

Systems with purely imaginary eigenvalues: the case of coupled harmonic oscillators. From what we have seen above, the control energy is least when the real part of the eigenvalues tends to vanish. In the extreme case of purely imaginary eigenvalues, it is possible to obtain explicit driver node placement criteria that minimize the control energy. A special case of linear system with purely imaginary eigenvalues is a network of n coupled harmonic oscillators, represented by a system of second order ODEs

$$M\ddot{q} + Kq = Bu \quad (6)$$

where $M = M^T > 0$ is the inertia matrix, $K = K^T \geq 0$ is the stiffness matrix, typically of the form $K = K_d + L$, with $K_d \geq 0$ diagonal and L a Laplacian matrix representing the couplings. In (6) the controls are forces, and the columns of the input matrix B are elementary vectors in correspondence of the driver nodes. The state space representation of (6) is

$$\dot{x} = A_o x + B_o u \quad (7)$$

with

$$x = \begin{bmatrix} M\dot{q} \\ Mq \end{bmatrix} \in \mathbb{R}^{2n}, \quad A_o = \begin{bmatrix} 0 & I \\ -K & M^{-1} \end{bmatrix}, \quad \text{and} \quad B_o = \begin{bmatrix} 0 \\ B \end{bmatrix}.$$

The system (7) has purely imaginary eigenvalues equal to $\pm i\omega_j$, $j = 1, \dots, n$, where ω_j are the natural frequencies of the oscillators. If $\Omega^2 = \text{diag}(\omega_1^2, \dots, \omega_n^2)$ and $\Psi = [\psi^1 \dots \psi^n]$ is the matrix of corresponding eigenvectors, then in the so-called modal basis the oscillators are decoupled and one gets the state space representation

$$\begin{aligned} \dot{z} &= A_1 z + B_1 u \\ &= \begin{bmatrix} 0 & I \\ -\Omega^2 & 0 \end{bmatrix} z + \begin{bmatrix} 0 \\ \Psi^T M^{-1} B \end{bmatrix} u \end{aligned} \quad (8)$$

where $z = \begin{bmatrix} \Psi^{-1} & 0 \\ 0 & \Psi^{-1} \end{bmatrix} x$. See SI for the details. When a system has purely imaginary eigenvalues, the finite time reachability Gramian diverges as $t_f \rightarrow \infty$. However, in the modal basis (8) the Gramian (here denoted W_z) is diagonally dominant and linear in t_f , hence as t_f grows it can be approximated by a diagonal matrix which can be computed explicitly²⁹.

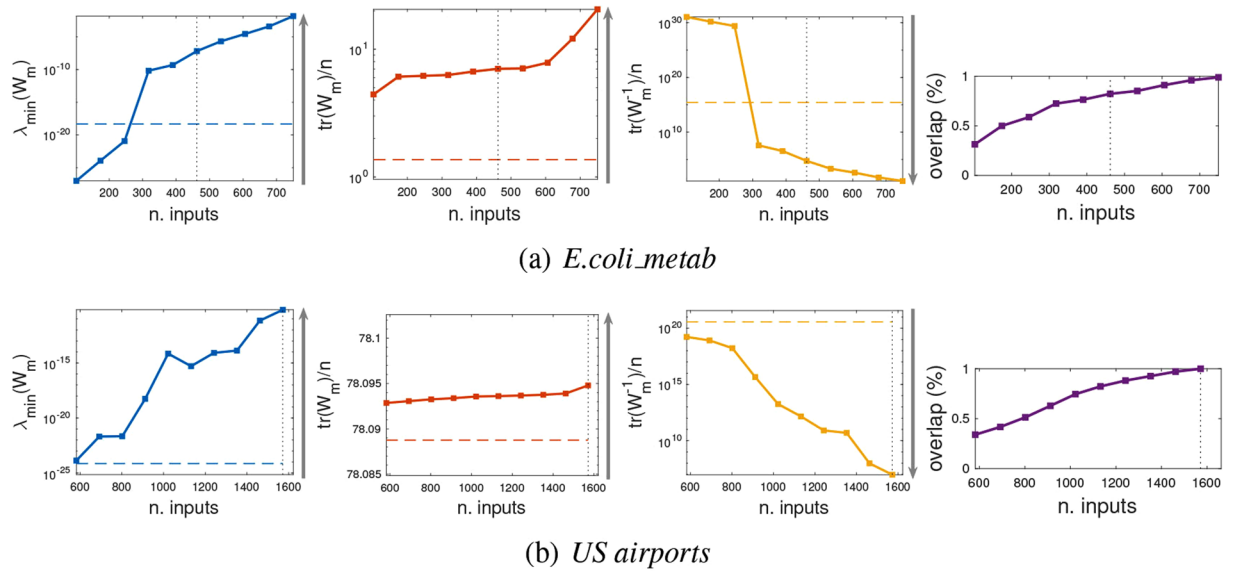


Figure 4. Driver node placement according to $r_w = w_{out}/w_{in}$ for two real-world networks. **(a) *E.coli-metabol*.** **(b) *US Airports*.** In both cases the control energy measures $\lambda_{min}(W_m)$, $tr(W_m)$, and $tr(W_z^{-1})$ are shown in the 3 leftmost panels (solid lines), when the number m_c of driver nodes (chosen according to r_w) grows. The horizontal dashed line is the control energy in correspondence of the m_c driver nodes required by structural controllability. The vertical dotted line is the value of m_c at which the r_w -ranked nodes achieve structural controllability (it is $m_c = n$ in *US Airports*). The rightmost panel shows how many of the m_c r_w -ranked nodes overlap with the m_c nodes of structural controllability.

$$W_z(t_f) \approx \sum_{j=1}^n \frac{\beta_j}{2M_{jj}^2} \begin{pmatrix} \frac{(\psi_j^1)^2}{\omega_1^2} \\ \vdots \\ \frac{(\psi_j^n)^2}{\omega_n^2} \\ \vdots \\ (\psi_j^1)^2 \\ \vdots \\ (\psi_j^n)^2 \end{pmatrix} t_f, \tag{9}$$

where $\beta_j = 1$ if the j -th input is present and 0 otherwise, and M_{jj} is the j -th diagonal entry of M , see SI for the details. For purely imaginary spectra (of which a network of coupled harmonic oscillators is a special case), using (9), the three measures of control energy adopted in this paper give rise to simple strategies for minimum energy driver nodes placement, which in some cases can be computed exactly for any n (for instance for the metric $tr(W_z)$, see SI). The three algorithms are explained in detail in the SI. Figure 6 shows that such strategies are always beating a random driver node placement, often by orders of magnitude.

Also w_{out}/w_{in} is still a good heuristic for driver node placement strategy. This can be understood by observing that the model (6) is symmetric hence for it in- and out-degrees are identical. However, since A_o has rows rescaled by M^{-1} , w_{out} is affected directly: when the inertia M_{ii} is big, the corresponding $w_{out}(i) = \sum_{j=1}^n K_{ji}/M_{ii}$ is small and viceversa. No specific effect is instead induced on w_{in} . In the representation (7), selecting nodes according to w_{out}/w_{in} means placing control inputs on the lighter masses, see Fig. 6(d). When the harmonic oscillators are decoupled ($L=0$) then $m < n$ means controllability is lost, but nevertheless selecting the nodes with least inertia as driver nodes maximizes the impact of a limited control authority ($tr(W_z)$ is maximized). A weak (and sparse) coupling allows to recover controllability, while the least inertia as optimal driver node strategy becomes suboptimal. When the coupling becomes stronger (for instance when the coupling graph is more connected) then the inertia of an oscillator is less significant as a criterion for selection of driver nodes: the modes of the system are now spread throughout the network and no longer localized on the individual nodes. As shown in Fig. S9, in a fully connected network of harmonic oscillators, driver node strategies based on w_{out}/w_{in} and on $tr(W_z)$ tend to perform considerably worse for the other measures ($\lambda_{min}(W_z)$ and $tr(W_z^{-1})$), while for a sparse graph (here ER graphs with $p=0.05$), of the three explicit optimal driver node placement strategies available in this case, $tr(W_z)$ has a high overlap with w_{out}/w_{in} , see Fig. 6(b), while the other two tend to rank controls in somewhat different ways. Given that in this case we have three strategies that are (near) optimal for the chosen measure of control energy, the dissimilarity of the node rankings of these three strategies means that the driver node placement problem is heavily dependent on the way control energy is quantified.

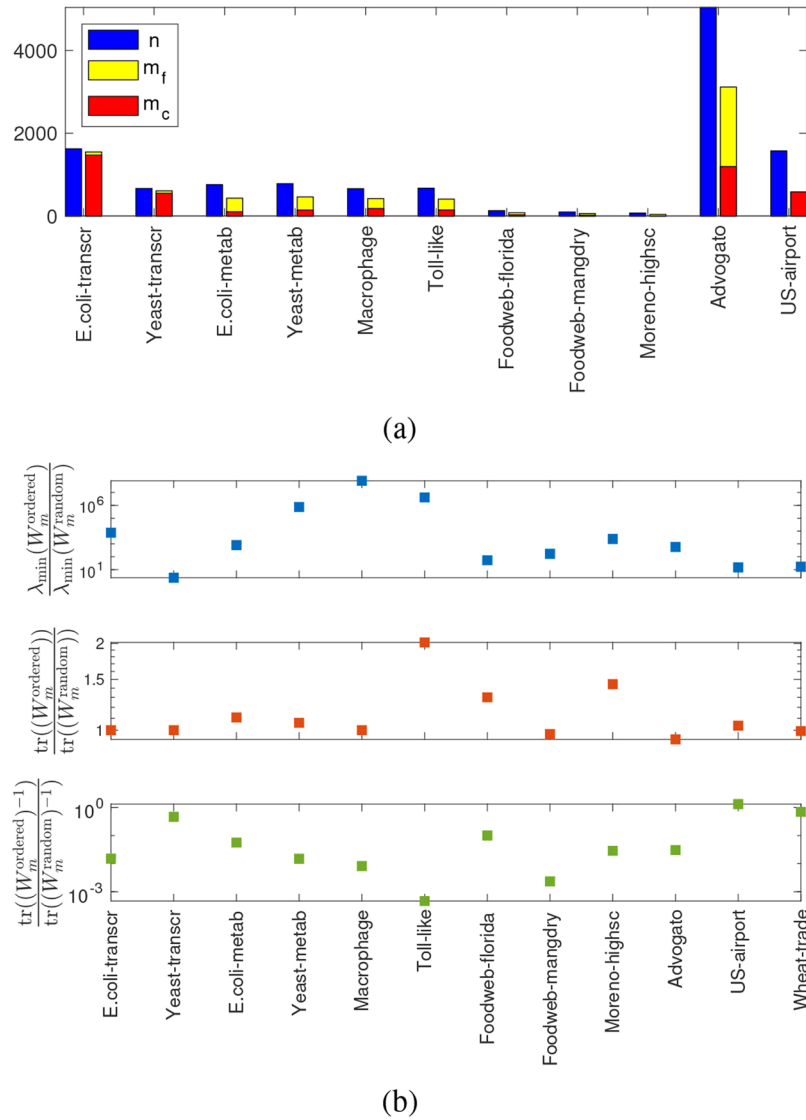


Figure 5. Control energy of driver node placement according to $r_w = w_{\text{out}}/w_{\text{in}}$ for the real-world networks of Table 1. (a) The real-world networks used. The number of nodes (n) is shown in blue, the minimal number of driver nodes needed to achieve controllability (m_c) is shown in red, and the number of additional driver nodes selected (m_f) in yellow. See also Table 1. For some networks, like for the two transcriptional networks, m_c amounts to a large fraction of n . (b) Comparison of control energies between a choice of m_f nodes according to r_w and a random choice. The two choices give rise to two different mixed Gramians, denoted respectively W_m^{ordered} and W_m^{random} . These Gramians are used to form the three measures of control energy $\lambda_{\min}(W_m)$, $\text{tr}(W_m)$, and $\text{tr}(W_m^{-1})$. The ratios between the two resulting values for these three metrics are shown in the three panels for the real networks considered in this study. In basically all cases the control energy improves (i.e., $\frac{\lambda_{\min}(W_m^{\text{ordered}})}{\lambda_{\min}(W_m^{\text{random}})} \gg 1$, $\frac{\text{tr}(W_m^{\text{ordered}})}{\text{tr}(W_m^{\text{random}})} > 1$, and $\frac{\text{tr}((W_m^{\text{ordered}})^{-1})}{\text{tr}((W_m^{\text{random}})^{-1})} \ll 1$), often by several orders of magnitude, even when m_f is small.

Application to minimum energy control of power grids. In the linear regime, power grids can be modeled as networks of weakly damped coupled harmonic oscillators³⁰. The so-called swing equation corresponds in fact to the following network of damped and coupled harmonic oscillators

$$M\ddot{q} + D\dot{q} + Kq = Bu, \tag{10}$$

where D is the matrix of dampings which we assume to be proportional, that is, that in the modal basis $D_1 = \Psi^T M^{-1} D M^{-1} \Psi$ is diagonal. In the state space representation (7), one gets then

$$\dot{x} = \begin{bmatrix} 0 & I \\ -K M^{-1} & -D M^{-1} \end{bmatrix} x + \begin{bmatrix} 0 \\ B \end{bmatrix} u, \tag{11}$$

while in the modal basis

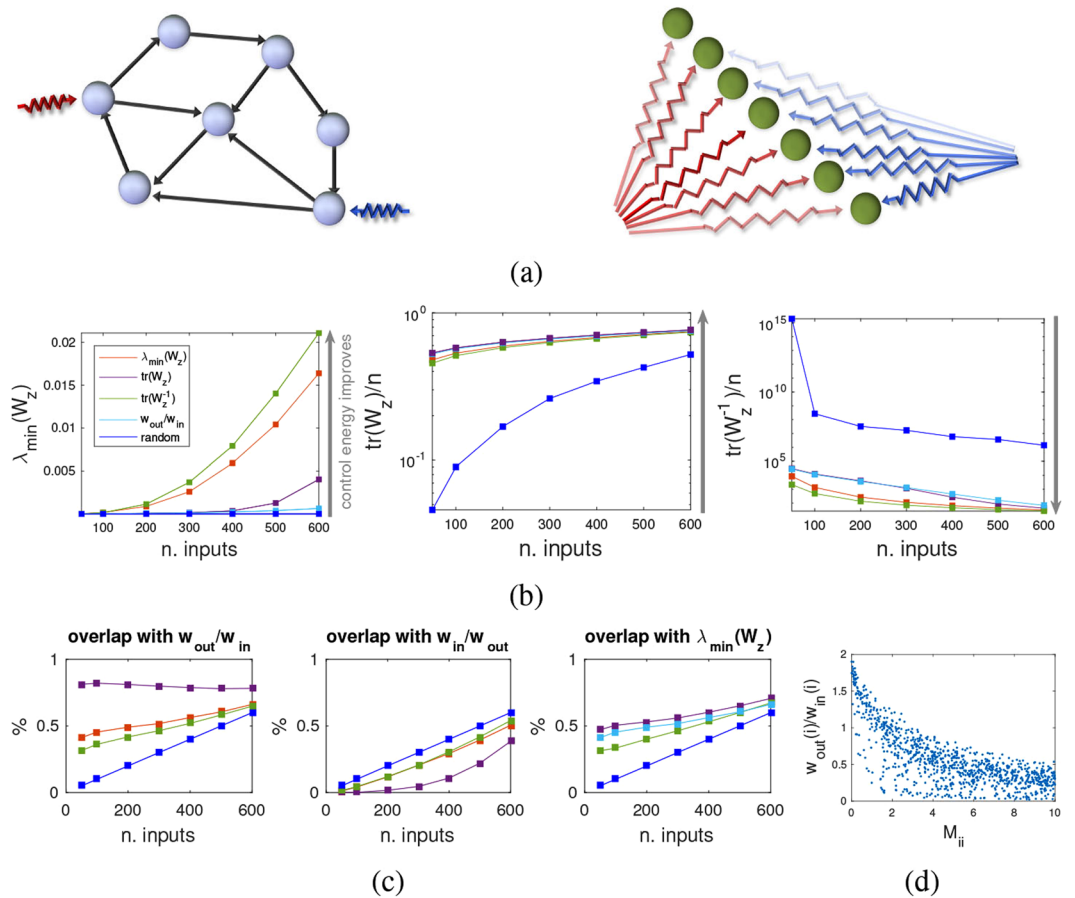


Figure 6. Driver node placement strategies for a network of coupled harmonic oscillators. (a) The concept of driver node is basis dependent: when the basis changes in state space (for instance we pass from (7) to (8)), the control inputs no longer target a single node, but become spread across the entire state space (now decoupled into non-interacting modes). (b) Comparison of different driver node placement strategies for $n = 1000$ coupled harmonic oscillators. Shown are means over 100 realizations (with 100 edge weights samples taken for each realization). The diagonal Gramian W_z of (9) is used to quantify the control energy. Red: driver node placement based on $\lambda_{\min}(W_z)$. Violet: placement based on $\text{tr}(W_z)$. Green: placement based on $\text{tr}(W_z^{-1})$. Cyan: placement based on $w_{\text{out}}/w_{\text{in}}$. Blue: random input assignment. All driver node placement strategies always beat a random assignment, often by orders of magnitude. The green and red curves give similar performances and so do the cyan and violet. Notice that for $\text{tr}(W_z)$ the violet curve gives the exact optimum. (c) Overlap in the node ranking of the different driver node placement strategies. Color code is the same as in (b). The only highly significant overlap is between $w_{\text{out}}/w_{\text{in}}$ and $\text{tr}(W_z)$, while $\lambda_{\min}(W_z)$ and $\text{tr}(W_z^{-1})$ correspond to different node ranking patterns. Notice that none of the strategies orders nodes according to $w_{\text{in}}/w_{\text{out}}$ (mid panel). (d) Inverse correlation between the inertia at node i , M_{ii} , and $w_{\text{out}}/w_{\text{in}}$ (correlation coefficient around -0.75 in average).

$$\dot{z} = \begin{bmatrix} 0 & I \\ -\Omega^2 & -D_1 \end{bmatrix} z + \begin{bmatrix} 0 \\ \Psi^T M^{-1} B \end{bmatrix} u. \tag{12}$$

The presence of damping makes the state update matrix stable, hence $W_m = W_r$. For weak damping, the driver node selection strategies illustrated above can be applied to the model (11) and so can the method based on $w_{\text{out}}/w_{\text{in}}$. We have investigated the minimum energy control of several power grids listed in Table S2, varying the dampings across several orders of magnitude, see Fig. 7(a). As expected, for all of them the energy required to achieve controllability increases as the real part of the eigenvalues moves away from the imaginary axis, see Figs 7(b) and S10–S13. All strategies still beat a random driver node placement, even those based on the Gramian (9), formally valid only for undamped dynamics.

Discussion

This paper contains two distinct contributions to the problem of understanding what determines the control energy required to completely control a complex network. The first is to show that the time constants of the free evolution of the system play a key role. Recall that in a linear unforced dynamical system the real part of the eigenvalues of A determines how fast/slow a system converges to the origin (stable eigenvalues, when real part of $\lambda(A)$

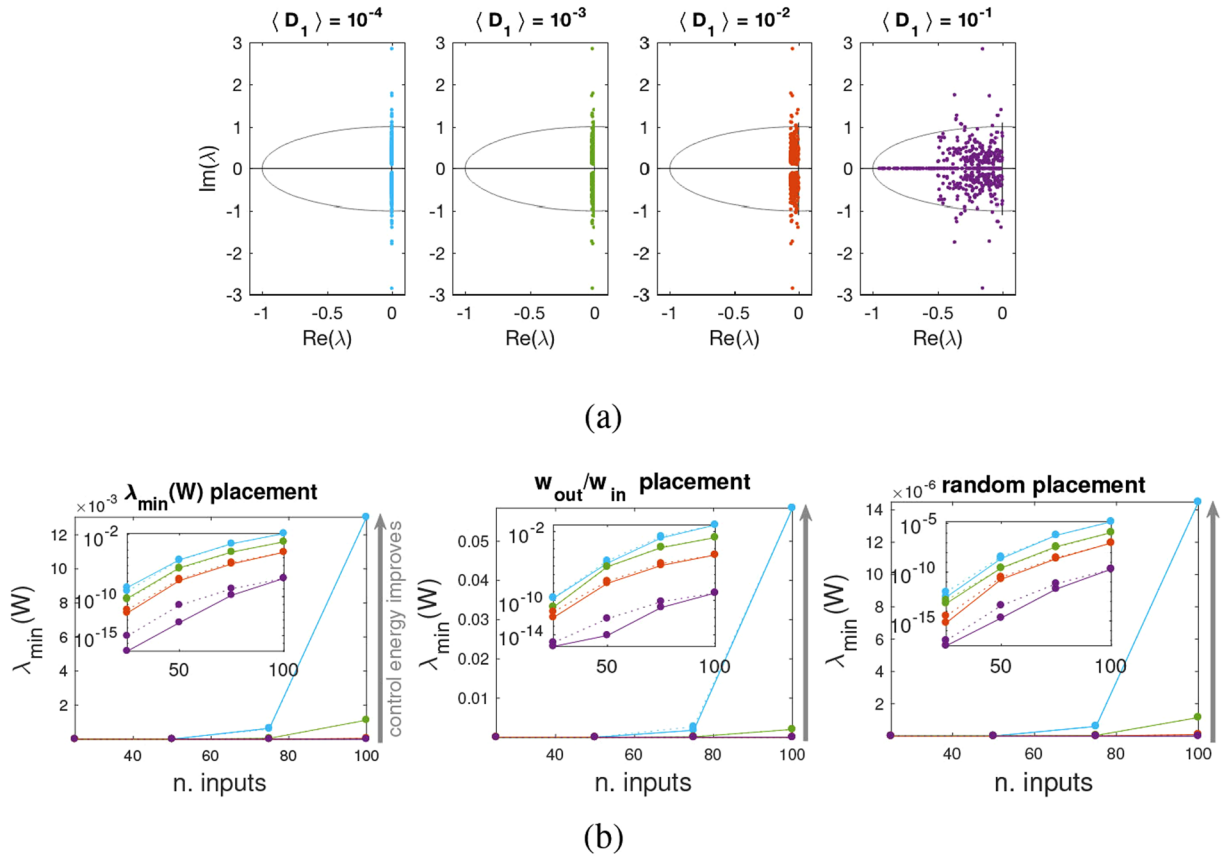


Figure 7. Minimum energy control of power grids for varying damping coefficients. (a) The eigenvalues of the state space system (11) for the North EU power grid³⁷ with uniformly distributed masses ($\langle M_{ii} \rangle = 10$) and damping coefficients that vary across 4 orders of magnitude. Since the system is always stable, $W_m = W_r$. (b) Control energy for the metric $\lambda_{\min}(W_r)$ when the driver nodes are placed according to $\lambda_{\min}(W_r)$ (left panel), $w_{\text{out}}/w_{\text{in}}$ (mid panel), or randomly (right panel). The values of $\lambda_{\min}(W_r)$ corresponding to the 4 choices of damping made in (a) are shown in solid lines (same color code as in (a)), while in dotted lines the values of $\lambda_{\min}(W_z)$ are shown (suitably normalized to eliminate the explicit dependence from t_p , see (9)). The insets show the same quantities in log scale. Values are all averages over 100 realizations. For all three driver node placement strategies, the performances worsen as the damping is increased. Comparing the three panels, $w_{\text{out}}/w_{\text{in}}$ performs similarly to $\lambda_{\min}(W_r)$, and both outperform a random placement by orders of magnitude.

is negative) or diverges to ∞ (unstable eigenvalues, when real part of $\lambda(A)$ is positive). Such convergence/divergence speed grows with the absolute value of the real part of $\lambda(A)$. In the complete controllability problem, both stable and unstable modes of A are gathered together, and all have to be “dominated” by the controls to achieve controllability. This result is valid regardless of the topology of the network, and of the number of controls used (provided, of course, that controllability is guaranteed). When the modes of the system are all slow, like when they are very close to the imaginary axis, then the energy needed to dominate them all is lower than when some of them are fast (i.e., the eigenvalues have large real part).

In the limit case of all eigenvalues of A being purely imaginary, we can also recover a known result from control theory affirming that controllability from any x_o to any x_f can be achieved in finite time by means of control inputs of bounded amplitude. As a matter of fact, an alternative approach used in control theory to take into account the control energy of a state transfer is to impose that the amplitude of the input stays bounded for all times (rather than the total energy), and to seek for conditions that guarantee controllability with such bounded controls^{31–33}. Assume $u \in \Omega$, with Ω a compact set containing the origin, for instance $\Omega = [-1, 1]^m$, where m is the number of control inputs. The constraint $u \in \Omega$ guarantees that we are using at all times a control which has an energy compatible with the physical constraints of our network. The consequence is, however, that reaching any point in \mathbb{R}^n may require a longer time, or become unfeasible. In particular a necessary and sufficient condition for any point to be reachable from 0 in finite time when $u \in \Omega$ is that no eigenvalue of A has a negative real part, see SI. This is clearly connected with our previous considerations on the reachability problem without bounds on u : when all modes of A are unstable then the input energy required to reach any state from 0 is low (Fig. 1(a)) and becomes negligible for sufficiently long time horizons. On the contrary, transferring any state to 0 in finite time with $u \in \Omega$ is possible if and only if no eigenvalue of A has a positive real part. Also in this case the extra constraints on the input amplitude reflects the qualitative reasoning stated above and shown in Fig. 1(b). Also in the

bounded control case, considering a generic transfer from any state x_o to any other state x_f means combining the two scenarios just described: formally a system is completely controllable from any x_o to any x_f in finite time and with bounded control amplitude $u \in \Omega$ if and only if all eigenvalues of A have zero real part, see SI for the details. The findings discussed above for u unbounded are completely coherent with this alternative approach to “practical controllability”.

The second contribution of this paper is a general strategy for driver node placement which reduces the control energy. Once the technical issues associated with minimal controllability can be neglected (in the sense that we are choosing sufficiently many inputs so as to guarantee controllability), a general criterion for controlling a network with a limited input energy is to drive the nodes having the maximal disbalance between their weighted outdegree and indegree. This driver node selection strategy has a direct topological interpretation, in terms of outdegree “dominance” (in a sense, it resembles the idea of dominating sets discussed in the literature⁹ although for directed edges only, as for undirected networks $r_w(i) = 1 \forall i$). Our node ranking strategy has also another indirect topological interpretation. For random edge weights, nodes that have high r_w tend to be outdegree dominated, and hence to be associated to (topological) dilations, see Fig. 3(a). A set \mathcal{S} of nodes forms a dilation if $|\mathcal{S}| > |T(\mathcal{S})|$, where $T(\mathcal{S})$ are the parent nodes of the nodes in \mathcal{S} and $|\mathcal{S}|$ (and $|T(\mathcal{S})|$) denotes set cardinality^{1,34}. When seeking minimal (structural) controllability, dilations cannot appear, meaning that controls must be added to the children nodes in excess³⁴. However, when minimal controllability is not the main concern and a number of driver nodes achieving controllability has already been selected, the strategy we are suggesting for reducing the control energy tends to favor nodes that are parent nodes of a dilation, rather than children. The rationale behind it is that in a dilation controlling parent nodes impacts several eigenmodes of the system (and hence reduces the control energy) while controlling children nodes enlarges the reachable subspace and for instance allows to discriminate among children of the same parents (and hence improves the structural controllability properties)³⁵. We think that this trade-off could be responsible for the limited success in finding effective minimum control energy driver node strategies in the literature.

It has already been pointed out¹¹ that guaranteeing controllability in a structural sense does not correspond to guaranteeing it in practice, as the required control energy can be exceedingly high. Our analysis of real-world networks confirms and strengthens this observation. Quite remarkably, we show that there are cases in which structural controllability fails, but nevertheless the control energy (which is theoretically infinite) turns out to be numerically lower than the one obtained in structurally controllable cases, even with equal numbers of controls. Such paradox follows from the fact that in both situations the Gramian is near-singular, which impacts measures like $\lambda_{\min}(W_m)$ and $\text{tr}(W_m^{-1})$. On the contrary, $\text{tr}(W_m)$ does not require W_m to be invertible, and in fact in nearly all networks we find that even with equal numbers of controls our r_w -ranked driver node strategy outperforms the driver node selection based on structural controllability. This confirms the effectiveness of our node ranking strategy.

Obviously our analysis requires to assign sets of weights to the edges, not to rely exclusively on structural properties. The consistency in the results we find across all investigated real-world networks makes us confident that our observations are not spurious numerical artifacts. The conclusion one can draw from our quantitative analysis is that “binary” (yes/no) responses to the controllability question have limited value and should be taken with extreme care.

As can be seen comparing Fig. 5(a) (and Table 1) with Fig. S6(a), the topology drastically influences the minimal number of controls needed to achieve structural controllability. It also influences the minimum energy problem, and makes it difficult to disentangle the two aspects analytically. For instance, the transcriptional networks we consider (*E. coli-transcr* and *Yeast-transcr*) are essentially directed acyclic graphs, and as such require an extremely high number of driver nodes to achieve minimal controllability¹. Nonetheless, by adding a few more driver nodes according to our r_w ranking, at least in *E. coli-transcr* the control energy can still be reduced by several orders of magnitude, meaning that even in these “extreme” topologies there is a lot to gain in terms of practical control by carefully selecting extra driver nodes. By comparing Figs S6 and 5(b), it is possible to realize that, rather than the indegree and outdegree distribution *per se*, it is the fraction of nodes having a high r_w (after the m_c nodes needed to guarantee controllability have been disregarded, violet curves in Fig. S6(b)) that determines the efficacy of our strategy. Although it is difficult to obtain a sharp classification, indicatively the networks that show a limited improvement with our driver node selection procedure (for instance *Yeast-transcr*, *US-airport* and *Wheat-trade*) correspond to those having a fairly limited number of nodes with high r_w in the m_f extra nodes being selected. Clearly a future step of our research is to extend our driver node strategy so as both minimal number of controls and minimum control energy are achieved at the same time.

Notice that our computation of weighted out/indegrees considers the total sum of weights in absolute value. When signs are taken into account in computing w_{in} and w_{out} , then no significant improvement over random input placement is noticeable. This is connected to the quadratic nature of the Gramian.

Finally, it is worth emphasizing that for a dynamical system the concept of driver node is not intrinsic, but basis-dependent. In fact, just like the idea of adjacency matrix of a graph is not invariant to a change of basis in state space, so inputs associated to single nodes (i.e., to single state variables) in the original basis become scattered to all variables in another representation of the same system, see Fig. 6(a). Our calculations of driver nodes placement are always in the original basis, in which the adjacency matrix is expressed (for instance the representation (7) for the coupled harmonic oscillator, and (11) for the power grid). If we take a special basis in which the modes are decoupled (for instance the Jordan basis), then the contribution of the nodes to the modes (i.e., the eigenvectors of A) provide useful information for the investigation of minimum input energy problems. The topic is closely related to the so-called participation factors analysis in power networks³⁶. Also quantities like w_{in} and w_{out} are basis-dependent and become nearly equal for instance if in (1) we pass to a Jordan basis. On the contrary, the eigenvalues of A are invariant to a change of basis. Hence as a general rule, the control energy considerations

that are consequence of the time constants of the system (like the dependence on the real part of the eigenvalues illustrated in Fig. 2) are “more intrinsic” than those that follow from the particular basis representation we have available for a network.

References

- Liu, Y.-Y., Slotine, J.-J. & Barabasi, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).
- Commault, C., Dion, J.-M. & van der Woude, J. W. Characterization of generic properties of linear structured systems for efficient computations. *Kybernetika* **38**, 503–520 (2002).
- Gao, J., Liu, Y.-Y., D’Souza, R. M. & Barabási, A.-L. Target control of complex networks. *Nature communications* **5** (2014).
- Liu, Y.-Y. & Barabási, A.-L. Control Principles of Complex Networks. *ArXiv e-prints* (2015).
- Nepusz, T. & Vicsek, T. Controlling edge dynamics in complex networks. *Nature Physics* (2012).
- Olshevsky, A. Minimal controllability problems. *Control of Network Systems, IEEE Transactions on* **1**, 249–258 (2014).
- Ruths, J. & Ruths, D. Control profiles of complex networks. *Science* **343**, 1373–1376 (2014).
- Ding, J., Lu, Y.-Z. & Chu, J. Studies on controllability of directed networks with extremal optimization. *Physica A: Statistical Mechanics and its Applications* **392**, 6603–6615 (2013).
- Nacher, J. C. & Akutsu, T. Analysis of critical and redundant nodes in controlling directed and undirected complex networks using dominating sets. *Journal of Complex Networks* **2**, 394–412 (2014).
- Yuan, Z., Zhao, C., Di, Z., Wang, W.-X. & Lai, Y.-C. Exact controllability of complex networks. *Nat Commun* **4** (2013).
- Yan, G., Ren, J., Lai, Y.-C., Lai, C.-H. & Li, B. Controlling complex networks: How much energy is needed? *Phys. Rev. Lett.* **108**, 218703 (2012).
- Bof, N., Baggio, G. & Zampieri, S. On the Role of Network Centrality in the Controllability of Complex Networks. *ArXiv e-prints* (2015).
- Chen, Y.-Z., Wang, L.-Z., Wang, W.-X. & Lai, Y.-C. Energy scaling and reduction in controlling complex networks. *Royal Society Open Science* **3** (2016).
- Li, G. *et al.* Minimum-cost control of complex networks. *New Journal of Physics* **18**, 013012 (2016).
- Olshevsky, A. Eigenvalue Clustering, Control Energy, and Logarithmic Capacity. *ArXiv e-prints* (2015).
- Pasqualetti, F., Zampieri, S. & Bullo, F. Controllability metrics, limitations and algorithms for complex networks. *IEEE Transactions on Control of Network Systems* **1**, 40–52 (2014).
- Summers, T. H., Cortesi, F. L. & Lygeros, J. On submodularity and controllability in complex dynamical networks. *IEEE Transactions on Control of Network Systems* **3**, 91–101 (2016).
- Sun, J. & Motter, A. E. Controllability transition and nonlocality in network control. *Phys. Rev. Lett.* **110**, 208701 (2013).
- Tzoumas, V., Rahimian, M. A., Pappas, G. J. & Jadbabaie, A. Minimal actuator placement with bounds on control effort. *IEEE Transactions on Control of Network Systems* **3**, 67–78 (2016).
- Yan, G. *et al.* Spectrum of controlling and observing complex networks. *Nat Phys* **11**, 779–786 (2015).
- Müller, P. & Weber, H. Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems. *Automatica* **8**, 237–246 (1972).
- Antsaklis, P. & Michel, A. *Linear Systems* (Birkhäuser Boston, 2005).
- Sontag, E. D. *Mathematical Control Theory: Deterministic Finite Dimensional Systems* (2Nd Ed.) (Springer-Verlag New York, Inc., New York, NY, USA, 1998).
- Zhou, K., Salomon, G. & Wu, E. Balanced realization and model reduction for unstable systems. *International Journal of Robust and Nonlinear Control* **9**, 183–198 (1999).
- Shaker, H. R. & Tahavori, M. Optimal sensor and actuator location for unstable systems. *Journal of Vibration and Control* (2012).
- Lee, H. & Park, Y.-J. Degree of disturbance rejection capability for linear anti-stable systems. In *Proceedings of the Int. Conf. on Control, Autom., and Syst. (ICCAS)* (2014).
- Allesina, S. & Tang, S. The stability–complexity relationship at age 40: a random matrix perspective. *Population Ecology* **57**, 63–75 (2015).
- Bollobás, B., Borgs, C., Chayes, J. & Riordan, O. Directed scale-free graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 03*, 132–139 (Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2003).
- Arbel, A. Controllability measures and actuator placement in oscillatory systems. *International Journal of Control* **33**, 565–574 (1981).
- Grigsby, L. *Power System Stability and Control*. The Electric Power Engineering Hbk, Second Edition (CRC Press, 2007).
- Brammer, R. F. Controllability in linear autonomous systems with positive controllers. *SIAM J of Control* **10**, 339–353 (1972).
- Jacobson, D. *Extensions of Linear-Quadratic Control, Optimization and Matrix Theory*, vol. 133 of Mathematics in Science and Engineering (Academic Press, London, 1977).
- Lee, E. & Markus, L. *Foundations of Optimal Control Theory* (R.E. Krieger Publishing Company, 1986).
- Lin, C. T. Structural controllability. *Automatic Control, IEEE Transactions on* **19**, 201–208 (1974).
- Lindmark, G. & Altafini, C. Controllability of complex networks with unilateral inputs. *Scientific Reports* **7**, 1824 (2017).
- Perez-arriaga, I. J., Verghese, G. C. & Schweppe, F. C. Selective modal analysis with applications to electric power systems, Part I: Heuristic introduction. *IEEE Transactions on Power Apparatus and Systems* **PAS-101**, 3117–3125 (1982).
- Menck, P. J., Heitzig, J., Kurths, J. & Joachim Schellnhuber, H. How dead ends undermine power grid stability. *Nat Commun* **5** (2014).
- Gama-Castro, S. *et al.* Regulondb version 7.0: transcriptional regulation of *Escherichia coli* k-12 integrated within genetic sensory response units (sensor units). *Nucleic Acids Research* (2010).
- Milo, R. *et al.* Network motifs: Simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Reed, J. L., Vo, T. D., Schilling, C. H. & Palsson, B. Ø. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* **4**, R54 (2003).
- Förster, J., Famili, I., Fu, P., Palsson, B. Ø. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244–253 (2003).
- Oda, K. *et al.* Molecular interaction map of a macrophage. *AfCS Research Reports* **2** (2004).
- Oda, K. & Kitano, H. A comprehensive map of the toll-like receptor signaling network. *Mol. Syst. Biol.* **2**, 2006.0015 (2006).
- Batagelj, V. & Mrvar, A. Pajek data sets. <http://pajek.imfm.si/doku.php?id=data:index> (Accessed: 09-21-2016).
- Coleman, J. *Introduction to mathematical sociology* (Free Press of Glencoe, 1964).
- Massa, P., Salvetti, M. & Tomasoni, D. Bowling alone and trust decline in social network sites. In *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, 658–663 (2009).
- Opsahl, T. The network of airports in the United States. <https://toreopsahl.com/datasets/#usairports>.
- De Domenico, M., Nicosia, V., Arenas, A. & Latora, V. Structural reducibility of multilayer networks. *Nat Commun* **6** (2015).

Acknowledgements

Work supported in part by a grant from the Swedish Research Council (grant no. 2015-04390 to C.A.).

Author Contributions

C.A. did the theoretical analysis, C.A. and G.L. developed the methods, implemented the algorithms, and analyzed the results. C.A. wrote the paper. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-21398-7>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018