Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control

New Journal of Physics

Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control

Jose C Nacher[1,2] **and Tatsuya Akutsu**[3]

[1] Department of Information Science, Faculty of Science, Toho University, Miyama 2-2-1, Funabashi, Chiba 274-8510, Japan
[2] Department of Complex and Intelligent Systems, Future University Hakodate, Hokkaido 041-8655, Japan
[3] Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji 611-0011, Japan
E-mail: nacher@is.sci.toho-u.ac.jp and takutsu@kuicr.kyoto-u.ac.jp

Abstract. The possibility of controlling and directing a complex system's behavior at will is rooted in its interconnectivity and can lead to significant advances in disparate fields, ranging from nationwide energy saving to therapies that involve multiple targets. In this work, we address complex network controllability from the perspective of the minimum dominating set (MDS). Our theoretical calculations, simulations using artificially generated networks as well as real-world network analyses show that the more heterogeneous a network degree distribution is, the easier it is to control the entire system. We demonstrate that relatively few nodes are needed to control the entire network if the power-law degree exponent is smaller than 2, whereas many nodes are required if it is larger than 2.

IOP Institute of Physics   Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Contents

1. Introduction

Momentous advances in information technology are rapidly changing the world and immensely expanding human interconnectedness on a global scale [1]. The components of complex systems, from cells and organs such as the brain to the Internet, WiFi communications and economic systems, are wired together in networks [2, 3]. Empirical data and analytical models have shown that connection patterns in many real networks converge to similar architectures, exhibiting a heterogeneous degree distribution characterized by a power law with a handful of highly connected nodes [4–7]. Specific network connectivity has a considerable effect not only on system behavior [8] but also on how the system's entire dynamics can be directed at will.

    Both techno-social and natural systems must be controlled or regulated to achieve optimal performance. However, because of the formidable size of real systems, a detailed complexity analysis of control dynamics in networks is impractical in most cases or limited to canonical linear time-invariant approximations at best [9]. Although recent works have addressed network controllability [9–16], general answers to the key questions regarding the factors that have a high impact on the controllability of techno-social and natural systems are still lacking. For example, how does the underlying network structure in socio-technical systems influence the minimum number of infected computers, smart phones or individuals necessary to damage the entire system? How does the topology of gene regulatory networks or human protein interactions influence treatment strategies necessary to determine the minimum number of possible drugs to regulate the whole network? To address these questions, here we focus on the problem of controlling a network by considering a model of reduced complexity, where a minimum set of

possible nodes dominates the whole system. For the purposes of control design, we investigate the dependence of the size of the minimum dominating set (MDS) of nodes on topological features of real networks, such as a variable scaling exponent that correlates to a different decay steepness of the degree distribution and the connectivity of high-degree nodes. The concept of MDS has been applied to the design and/or control of various kinds of discrete systems, which include mobile *ad hoc* networks (MANET) [17–20], transportation routing [21] and computer communication networks [21, 22].

In this work, we compute the MDS in real networks and in computer-generated networks with a variety of topologies. Our extensive analysis of real data and computer simulations together with mathematically developed tools provides new insights into control and network regulation as well as the development of new control techniques. Firstly, our theoretical findings suggest that scale-free networks with small scaling exponent values ($\gamma < 2$), where high-degree nodes are present, require relatively few nodes to be controlled. Conversely, networks with large-degree exponent values ($\gamma > 2$) or faster exponential decay, where hubs are weakly connected or almost absent, require more nodes to be fully regulated. The existence of the critical scaling exponent, predicted by our theoretical analysis, indicates a structural boundary below which the system could be more easily controlled by a significantly reduced set of nodes. Secondly, the MDS size also depends on the average degree of all nodes in the network. This fact also permeates into other network features, including clustering degree, diameter, centrality and shortest path length, uncovering multiple new ways to significantly increase or decrease controllability, based exclusively on topological design. In sharp contrast to a recent work on controllability that uses maximum matching [9], our analysis demonstrates that the MDS tends to target highly connected nodes, whereas the previous study suggested that driver nodes tend to avoid high-degree nodes. Although Müller and Schuppert [10] recently suggested that iPS cells can be controlled by a few driver nodes, they did not show general results or consider structural properties of networks. In contrast, our conclusions are supported by database analysis, computer simulations and analytical predictions. Taken together, these findings provide new leads in complex network control design, ranging from better energy-distribution networks planning and optimal mobile phone tower deployment to controlling a complex disease by simultaneously disrupting multiple targets in distant network pathways.

2. Simulation analysis

2.1. Computation of the minimum dominating set (MDS)

A set $S \subseteq V$ of nodes in a graph $G = (V, E)$ is a dominating set if every node $v \in V$ is either an element of S or adjacent to an element of S [21] (see figure 1). Since the domination problem is a classical NP-complete decision problem in computational complexity theory [21], it is not plausible that there exists a polynomial time algorithm that finds a smallest dominating set for arbitrary graphs. Here, we reduce this problem to the computation of a binary integer programming problem. Binary integer programming is the problem of finding a binary vector x that minimizes a linear function $f(x)$ subject to linear constraints. Each node i is assigned to a binary integer variable x_i that takes the value 1 or 0. A node i that belongs to the dominating set will take the value $x_i = 1$. To compute the MDS, we consider the minimization of the following
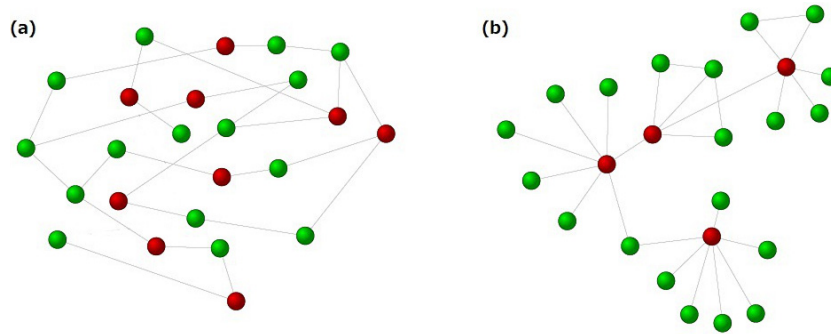
Figure 1. Two network topologies consisting of the same number of nodes $N = 24$ and edges $E = 26$. (a) A random network with average degree $\langle k \rangle = 2.16$, average clustering degree $C = 0$, diameter $d = 12$, network centrality $NC = 0.040$ and average path length $l = 4.33$. (b) A scale-free network with $\langle k \rangle = 2.16$, $C = 0.206$, $d = 6$, $NC = 0.229$ and $l = 3.36$. In each network, each green node is adjacent to at least one red node. This situation indicates that the green node is dominated or controlled by the red node. The impact of the network structure on the size of the MDS is illustrated by the small number of red nodes displayed in the scale-free network, less than half of those from the random network.

linear function:

$$f(x) = x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i \tag{1}$$

subject to the constraints

$$x_i + \sum_{j \in \partial i} x_j \geqslant 1, \tag{2}$$

where ∂i indicates the neighbors of node i and n is the total number of nodes in the network.

To solve this integer linear programming (ILP) problem, we have used an algorithm that uses a LP-based classic branch and bound method to determine the optimal solution (BintProg), allowing us to find solutions for moderate sizes of graphs with up to a few tens of thousands of nodes. The algorithm searches for an optimal solution to the binary integer programming problem by solving a series of LP-relaxation problems, in which the binary integer constraints on the variables are substituted by the weaker constraint $0 \leqslant x \leqslant 1$ [23, 24]. The BintProg algorithm is available in the computer libraries of the Optimization ToolBox of MatLab version R2010a.

2.2. Computational experiments on the MDS size in scale-free networks

To analyze in detail the dependence of the scaling exponent and mean degree on the MDS size, we generated a variety of scale-free networks using computer simulations. The configuration model [7, 25, 26] has often been used for network simulations; however, this model generates graphs that are neither simple nor connected. We then used an algorithm that generates random simple connected graphs with a prescribed degree sequence [27]. For each configuration of a
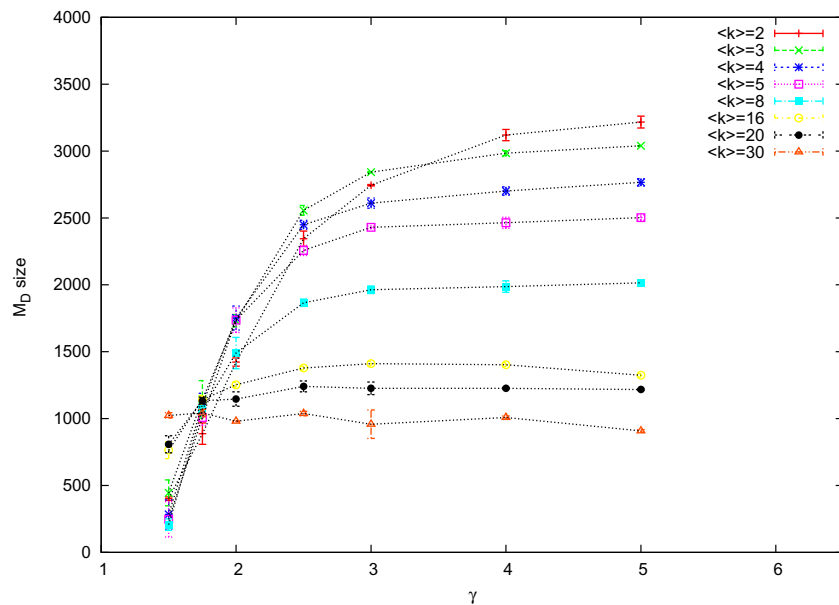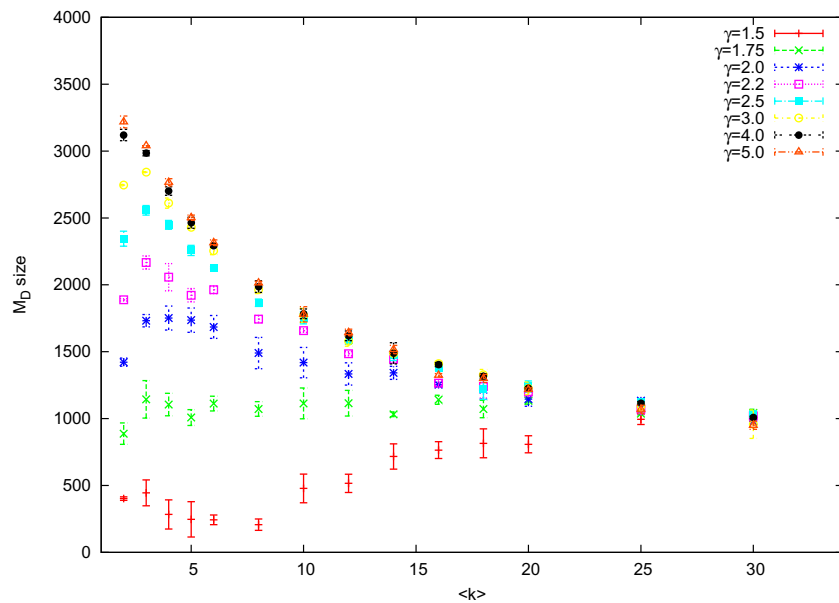
Figure 2. Computer simulations of the MDS size as a function of the scaling exponent in simulated networks with different average degrees $\langle k \rangle$. Scale-free networks with $N = 10\,000$ nodes are generated as shown in section 2.2. The error bars (s.e.m.) are shown in the figure.

given γ and average degree $\langle k \rangle$, ten networks of $10\,000$ nodes were constructed. The results of the MDS were averaged over all realizations. Then, the s.e.m. were computed and are shown in the figures. We first plot the MDS size as a function of the degree exponent, keeping fixed the average degree $\langle k \rangle$. Figure 2 shows that, for a given average degree value, the MDS size increases as γ increases. However, we also observe that the MDS decreases as the average degree increases for values of γ above 2. Conversely, for $\gamma < 2$, the MDS increases as the average degree increases. It is worth noting that $\gamma = 1.7$ is a crossing point of curves for various $\langle k \rangle$. However, this crossing point does not have any meaning. The results shown in figure 2 along with the mathematical analysis provided in section 3 suggest that a few individuals could in principle control a network characterized by both a small scaling exponent and average degree, providing a simple control design principle for large networks.

These results are complemented by figure 3, where we plot the fraction of nodes in the MDS as a function of the average degree, keeping the scaling exponent fixed. The MDS generally decreases as the average degree increases. For values of γ above 2, the MDS decreases as the average degree increases. Conversely, when $\gamma < 2$, the MDS slightly increases or becomes constant as the average degree increases. Moreover, the jump size in the number of dominating nodes between $\gamma = 1.5$ and $\gamma = 2$ is considerably large, suggesting a drastic change in the network's dominance phenomena when the scaling exponent takes values near $\gamma = 2$, which is in agreement with the tendency observed in figure 2.

2.3. Phase diagram highlighting optimal dominating configurations

The previous results suggest that the network structure significantly limits the number of individuals that could control the entire system, thus requiring a phase diagram to uniquely

Figure 3. Computer simulations of the MDS size as a function of the average degree of all nodes in simulated networks with different scaling exponents γ. Scale-free networks are generated as shown in section 2.2 with $N = 10\,000$ nodes. Error bars (s.e.m.) are shown in the figure.

characterize the optimal configurations available to the system. To clearly visualize the impact of the network structure on the number of dominating nodes, we plot a phase diagram to help illustrate the properties of scale-free networks by showing various values of γ and average degree (figure 4). The MDS size is shown using a gradient of colors as indicated in the colorbar legend. The plot highlights optimal configurations for scale-free networks that minimize the number of individuals necessary for dominating the whole system (dark blue regions for $\gamma < 2$). Conversely, networks with higher γ values (or showing decays faster than a power law) and small average degrees are more difficult to dominate, thus requiring more nodes.

3. Theoretical analysis

3.1. Mathematical analysis of the MDS size in scale-free networks

We analyze the size of the MDSs for scale-free graphs. Although rather precise results are known for the Barabási–Albert networks with exponent $\gamma = 3$ [28], scale-free trees [29] and general graphs [30], the dependence of the size of the MDSs on the power-law scaling exponent γ remains unknown.

Let $G(V, E)$ be an undirected random graph with node set V and edge set E such that the degree distribution follows a power-law $k^{-\gamma}$. We let $n = |V|$. In this section, we show the following:

- if $\gamma > 2$, the size of the MDS is $\Theta(n)$,
- if $\gamma < 2$, the expected size of the MDS is $o(n)$.
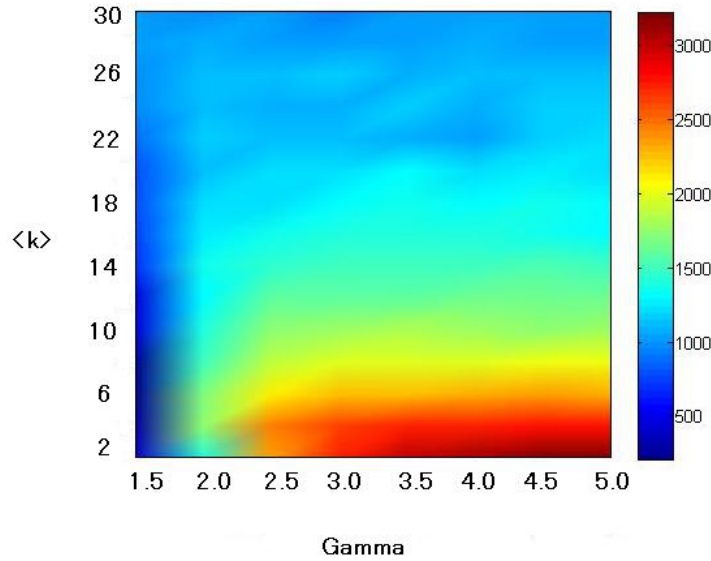
Figure 4. Phase diagram showing optimal dominating configurations. The MDS size is indicated using a gradient of colors as shown in the colorbar legend. The plot highlights optimal configurations for scale-free networks that minimize the number of individuals necessary to dominate the whole system (see dark blue regions below $\gamma = 2$).

This result means that a phase transition occurs at $\gamma = 2$. Here we note that $f(n)$ is $\Theta(g(n))$ if it is both $O(g(n))$ and $\Omega(g(n))$, that is, $f(n)$ is $\Theta(g(n))$ if $k_1 g(n) \leqslant f(n) \leqslant k_2 g(n)$ holds for sufficiently large n where k_1 and k_2 are some positive constants. Moreover, $f(n)$ is $o(g(n))$ if, for every $\epsilon > 0$, $|f(n)| \leqslant |g(n)| \epsilon$ holds for sufficiently large n.

3.1.1. The case of $\gamma > 2$. We assume that the degree distribution follows $\alpha k^{-\gamma}$ with a cutoff at $k = n$, where $\gamma > 2$. From

$$\alpha n \int_1^n k^{-\gamma} \, \mathrm{d}k = \frac{\alpha n}{\gamma - 1} \left(1 - n^{-\gamma+1}\right) = n, \tag{3}$$

we have $\alpha \approx \gamma - 1$.

For $S \subseteq V$, $\mathcal{C}(S)$ denotes the set of edges between S and $V - S$ (i.e. $\mathcal{C}(S) = \{\{u, v\} \mid u \in S$ and $v \in V - S\}$). The following property is well known [28, 30]:

if $|S| + |\mathcal{C}(S)| < n$, S is not a dominating set.

Let S be the set of nodes whose degree is greater than or equal to K. Here we assume without loss of generality that $|S| < n/2$ holds, because otherwise S is already large (i.e. $|S| \geqslant n/2$). It is to be noted that S is chosen so that the total degree (i.e. the number of edges incident on S) is maximized among the sets with the same cardinality.

We estimate the size of $\mathcal{C}(S)$ as follows:

$$|\mathcal{C}(S)| < \alpha n \int_K^n k \cdot k^{-\gamma} \, \mathrm{d}k \approx n(\gamma - 1) \int_K^n k^{-\gamma+1} \, \mathrm{d}k$$

$$= n \left(\frac{\gamma - 1}{\gamma - 2}\right) \left(\frac{1}{K^{\gamma-2}} - \frac{1}{n^{\gamma-2}}\right) < n \left(\frac{\gamma - 1}{\gamma - 2}\right) \frac{1}{K^{\gamma-2}}. \tag{4}$$

If S is a dominating set, the last term should be greater than $n - |S| > n/2$. Therefore, the following inequality should be satisfied:

$$n \left(\frac{\gamma - 1}{\gamma - 2}\right) \frac{1}{K^{\gamma - 2}} > n/2. \tag{5}$$

By solving this inequality, we have

$$K < \left[2 \left(\frac{\gamma - 1}{\gamma - 2}\right)\right]^{1/(\gamma - 2)}. \tag{6}$$

Then, the size of S is estimated as

$$|S| \approx \alpha n \int_K^n k^{-\gamma} \, \mathrm{d}k \approx n \left(\frac{1}{K^{\gamma - 1}} - \frac{1}{n^{\gamma - 1}}\right) \approx n \frac{1}{K^{\gamma - 1}}$$

$$> \left[2 \left(\frac{\gamma - 1}{\gamma - 2}\right)\right]^{-\frac{\gamma - 1}{\gamma - 2}} n. \tag{7}$$

From this inequality and the fact that V is a trivial dominating set, we can see that the size of the MDS is $\Theta(n)$ (for fixed γ) and the coefficient increases as γ increases.

3.1.2. The case of $\gamma < 2$. We only consider the case of $\gamma > 1$ because there exist very few networks with $\gamma \leqslant 1$. We assume as above that the degree distribution follows equation (3), where $1 < \gamma < 2$.

We use a well-known result on the *coupon collector* problem [31], which states that we can collect all n kinds of coupons with probability greater than $1 - (1/n)$ if we randomly draw $2n \ln n$ coupons. We associate coupons with nodes in $V - S$. Since we assume that edges in $G(V, E)$ are randomly selected under a power-law degree distribution, we can see that if $|\mathcal{C}(S)| > 2n \ln n$ holds, S is expected to be a dominating set.

As in the case of $\gamma > 2$, we let S be the set of nodes whose degree is greater than or equal to K. Furthermore, we let $K = \beta n$. Then, to be shown later, the size S is $o(n)$. We can estimate $|\mathcal{C}(S)|$ as follows:

$$|\mathcal{C}(S)| \approx n(\gamma - 1) \int_K^n k \, k^{-\gamma} \, \mathrm{d}k = n \left(\frac{\gamma - 1}{2 - \gamma}\right) \left[n^{2-\gamma} - K^{2-\gamma}\right]$$

$$= n \left(\frac{\gamma - 1}{2 - \gamma}\right) \left[(1 - \beta^{2-\gamma}) n^{2-\gamma}\right]$$

$$= \left(\frac{\gamma - 1}{2 - \gamma}\right) \left(1 - \beta^{2-\gamma}\right) n^{3-\gamma}. \tag{8}$$

Since we assumed $\gamma < 2$, the last term is greater than $cn \ln n$ for any constant c if n is sufficiently large. Therefore, S is expected to be a dominating set.

The size of S is estimated by

$$|S| \approx \alpha n \int_K^n k^{-\gamma} \, \mathrm{d}k \approx n(\gamma - 1) \int_K^n k^{-\gamma} \, \mathrm{d}k$$

$$= n \left(K^{1-\gamma} - n^{1-\gamma}\right) = \left(\left(\frac{1}{\beta}\right)^{\gamma - 1} - 1\right) n^{2-\gamma}. \tag{9}$$

Since we assumed $1 < \gamma < 2$, the last term is $o(n)$, which means that the size of S is very small. It should be noted that this bound is not tight because $n^{2-\gamma}$ increases as γ decreases, whereas, as suggested by simulation results shown in section 2.2, the size of the MDS should decrease as γ decreases.

3.1.3. Dependence on the average degree. We have not considered the average degree $\langle k \rangle$ so far. Here, we consider the dependence of the size of MDS on $\langle k \rangle$ for the case of $\gamma > 2$, in a similar way as in section 3.1.1. Since $\langle k \rangle$ cannot be a constant (i.e. $\langle k \rangle$ depends on n) if $\gamma < 2$, we do not consider the case of $\gamma < 2$ here.

Suppose that the frequency of nodes of degree βk follows a power law $P(k) \propto k^{-\gamma}$.[4] As in equation (3), we have

$$\alpha n \int_1^{(n/\beta)} k^{-\gamma} \, \mathrm{d}k = \frac{\alpha n}{\gamma - 1} \left(1 - (n/\beta)^{-\gamma+1} \right) = n \tag{10}$$

and $\alpha \approx \gamma - 1$.

Since the average degree is $c \equiv \langle k \rangle$, we have

$$\alpha n \int_1^{n/\beta} \beta k \cdot k^{-\gamma} \, \mathrm{d}k \approx \frac{\gamma - 1}{\gamma - 2} \beta n = cn, \tag{11}$$

from which $\beta \approx \frac{\gamma-2}{\gamma-1} c$ follows. As in equation (4), we have

$$|\mathcal{C}(S)| < \alpha n \int_K^{n/\beta} \beta k \cdot k^{-\gamma} \, \mathrm{d}k \approx cn(\gamma - 2) \int_K^{n/\beta} k^{-\gamma+1} \, \mathrm{d}k$$

$$= cn \left(\frac{1}{K^{\gamma-2}} - \frac{1}{(n/\beta)^{\gamma-2}} \right) < cn \frac{1}{K^{\gamma-2}}. \tag{12}$$

Since the last term should be greater than $n/2$, $K < (2c)^{\frac{1}{\gamma-2}}$ should hold. Therefore, the lower bound of the size of the MDS is estimated as

$$\alpha n \int_K^{n/\beta} k^{-\gamma} \, \mathrm{d}k \approx (2\langle k \rangle)^{-\frac{\gamma-1}{\gamma-2}} n. \tag{13}$$

This result suggests that the size of the MDS tends to decrease as the average degree increases. It is also seen that the constant factor ($(2\langle k \rangle)^{-\frac{\gamma-1}{\gamma-2}}$) rapidly decreases as γ approaches 2. This analytical prediction is in agreement with the tendency observed in computer simulations and empirical data analysis to be shown later (figures 3 and 9, respectively).

3.2. The relation between the MDS and structural controllability

Hereafter we show a link between the MDS and structural controllability. We begin with the following theorem:

Theorem 1. *Suppose that every edge in a network is bi-directional and every node in the MDS can control all of its outgoing links separately. Then, the network is structurally controllable by selecting the nodes in the MDS as the driver nodes.*

---

[4] This β is different from that in section 3.1.2.
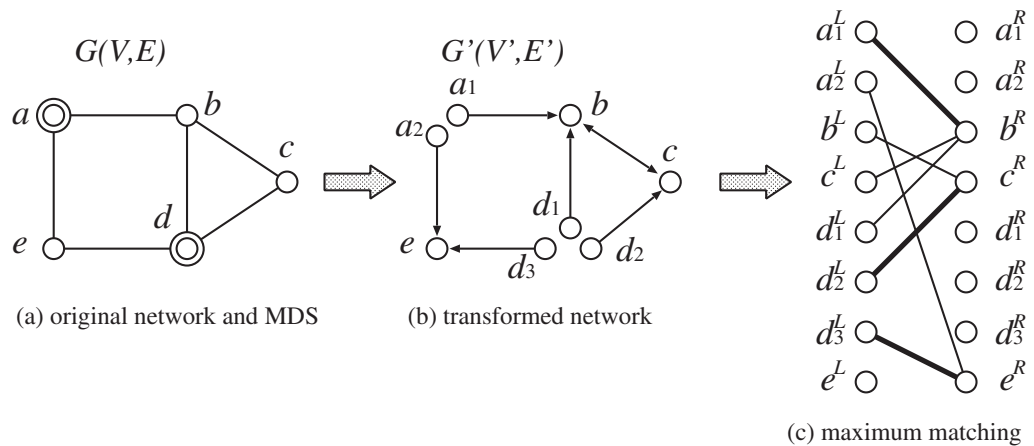
Figure 5. Illustration of the proof of theorem 1. (a) The original network and MDS, where the MDS is denoted by double circles. (b) Transformed network. (c) Constructed bipartite network where bold lines correspond to maximum matching.

Proof. From a given network $G(V, E)$ and MDS S, we first construct a directed network as follows (see figure 5). For each node $v \in S$, we split this node into k nodes $v_1, v_2, \ldots, v_k$ where k is the degree of v and create a directed edge from each v_i to a node adjacent to v so that the out-degree of each v_i is 1. For each pair of nodes (u, v) such that $u, v \notin S$ and $\{u, v\} \in E$, we create directed edges (u, v) and (v, u). Let $G'(V', E')$ denote the resulting network (see also figure 5(b)).

We next construct a bipartite graph from $G'(V', E')$ as follows (see also figure 5(c)). For each node $v \in V'$, we create two nodes v^L and v^R. For each edge $(u, v) \in E'$, we create an edge between u^L and v^R.

Since S is a dominating set, there exists a matching M satisfying:

- for each node $v \notin S$, v^R appears in M,
- for each node $v \in S$, none of v_i^R appear in M.

Furthermore, we can see that M is maximum because no v_i^R with $v \in S$ has an edge in the bipartite graph. Therefore, by selecting $\{v_i | v \in S\}$ as the set of driver nodes, the theorem follows from theorem 2 in [9]. □

This result not only links our approach with that from Liu *et al* but also shows that our findings can be applied to directed networks. Although we have considered undirected networks, the theoretical results shown in this work also hold for directed networks if networks exhibit a power-law for the out-degree distribution, because we are only considering outgoing edges in the theoretical analysis.

4. Results on real-world networks

4.1. Datasets

We have collected 16 networks to investigate the size of the MDS. In particular, we focus on how MDS size correlates with topological features, such as the scaling exponent in

IOP Institute of Physics **Φ** DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Table 1. The main topological features of the analyzed networks. From left to right, GCC is the giant connected component size, m_D fraction of dominating nodes, $\langle k \rangle$ average degree, l average shortest path, d diameter, C average clustering degree and NC the network centrality.

Name	Nodes	GCC	m_D	$\langle k \rangle$	l	d	C	NC
PPI *C. elegans*	2 651	2 386	0.182	3.20	4.80	14	0.022	0.077
PPI *D. melanogaster*	7 498	7 351	0.199	6.14	4.40	12	0.012	0.023
PPI *E. coli*	1 865	1 447	0.229	8.12	3.81	12	0.109	0.109
PPI *H. sapiens*	1 607	805	0.239	2.92	6.53	19	0.107	0.042
PPI *M. musculus*	599	50	0.220	2.20	4.42	9	0.060	0.208
PPI *S. cerevisiae*	4 963	4 902	0.179	7.03	4.14	11	0.097	0.056
TRN *S. cerevisiae*	688	662	0.126	3.20	5.20	15	0.049	0.103
TRN *E. coli*	418	328	0.176	2.78	4.83	13	0.110	0.213
US airports	500	500	0.102	11.92	2.99	7	0.617	0.268
Word adjacency (*Japanese*)	2 704	2 698	0.109	5.92	3.07	8	0.220	0.267
Word adjacency (*Spanish*)	12 642	11 558	0.067	7.44	2.91	10	0.376	0.258
Collaboration (*ca-HepTh*)	9 877	8 638	0.205	5.74	5.94	18	0.482	0.007
Collaboration (*ca-GrQc*)	5 242	4 158	0.186	6.45	6.04	17	0.557	0.018
Wiki-Vote	7 115	7 066	0.154	28.5	3.24	7	0.141	0.140
Electronic circuit S420	252	252	0.260	3.167	5.806	13	0.056	0.044
Electronic circuit S208	122	122	0.250	3.098	4.928	11	0.059	0.058

power-law degree distributions and their average degree. The networks are representative of datasets used in complex network analysis corresponding to the major fields of biological and socio-technical systems. The list of the networks as well as the main topological features are shown in table 1. The network centrality NC (or degree centralization) measures the extent to which certain nodes are far more central than others. Networks whose topologies resemble a perfect star have a centralization close to 1, whereas decentralized networks are characterized by a centralization close to 0 [32].

4.1.1. Socio-technical networks. The collected networks include eight datasets from disparate fields, ranging from US airline routes, word adjacency in texts, Wikipedia administrator vote to collaboration networks in two fields of expertise and two types of electronic circuit designs. The US airline network corresponds to the 500 busiest commercial airports in the United States. A link is placed between two airports if a flight was scheduled between them. The data correspond to the US commercial flights operated during 2002 [33]. The word adjacency networks of a text in Spanish and Japanese languages and the networks for two designs of electronic circuits S208 and S420 were downloaded from Uri Alon's website[5] [34]. The collaboration networks in two fields of expertise and the Wikipedia network were downloaded from the Stanford large network dataset collection website[6]. The HEP-TH (High Energy Physics-Theory) collaboration network was built using the e-print arXiv and contains scientific collaboration between authors. For example, a paper co-authored by three researchers will lead

to a complete graph of degree $k = 3$. Similarly, a collaboration network corresponding to the GR-QC (General Relativity and Quantum Cosmology) was considered. Both datasets cover papers in the period from January 1993 to April 2003 (124 months). These two networks were first reported in [35]. The Wikipedia vote represents an example of an online social network. An online vote process decides who are the users to be promoted to Wikipedia administrators. This network has the highest density among the analyzed networks with 103 689 edges, which leads to a high average degree (see table 1). We used an undirected representation of the network where two users i and j are connected by an edge if user i voted on user j or vice versa (see footnote 7) [36].

4.1.2. Biological networks. We have analyzed eight datasets corresponding to biological networks. First, we have used the database of interacting proteins[7] to construct the protein–protein interacting networks for six model organisms, including *H. sapiens*, *C. elegans*, *S. cerevisiae*, *D. melanogaster*, *E. coli* and *M. musculus* organisms. Protein interactions are essential for many biological functions. In these networks, the nodes represent proteins and two nodes are connected if an interaction exists between them. Interactions often take place when two proteins bind together in order to perform a specific cellular function. On the other hand, we have also considered the transcription interactions in the bacteria *E. coli* and in the yeast *S. cerevisiae*. Transcriptional regulatory networks play a crucial role in controlling the gene expression programs of thousands of genes in a cell. The datasets were obtained from [34] (see footnote 5).

4.1.3. Fit of power-law distributions using real-world data. The above set of networks was fitted to power-law distributions to determine the degree exponent γ. First, we computed the GCC of each network. The number of nodes in each GCC is shown in table 1. We then used each GCC network to fit the power-law distribution. By following [37, 38], let x represent a sequence of observations of some variable whose distribution we wish to fit as a power law. There must be some lower bound to the power-law behavior. This value is denoted as $x_{\min}$. Therefore, given a set containing n observations $x_i \geqslant x_{\min}$ and provided that $\gamma > 1$, it can be shown that the continuous distribution with the corresponding normalizing constant is

$$P(x) = \frac{\gamma - 1}{x_{\min}} \left[\frac{x}{x_{\min}} \right]^{-\gamma}. \tag{14}$$

Then, the probability that the data are drawn from a distribution that follows the power law for $x \geqslant x_{\min}$ reads as

$$p(x|\gamma) = \prod_{i=1}^{n} \frac{\gamma - 1}{x_{\min}} \left(\frac{x_i}{x_{\min}} \right)^{-\gamma}. \tag{15}$$

This expression is also called the likelihood of the data given the model. Then, the maximum likelihood estimate (after maximization of the likelihood) for the scaling exponent reads as

$$\gamma = 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}} \right]^{-1}. \tag{16}$$

---

7 http://dip.doe-mbi.ucla.edu/dip

However, a more accurate expression for the computation of the scaling exponent when using discrete distributions (such as degree distributions) can be considered. We write the power-law distribution over an integer variable as

$$P(x) = \frac{x^{-\gamma}}{\zeta(\gamma, x_{\min})}, \tag{17}$$

where $\zeta(\gamma, x_{\min})$ is the Hurwitz zeta function. Following a similar derivation, the likelihood function can be written and the scaling exponent solved numerically. For moderate values of $x_{\min} \gtrsim 6$, however, we can still write the scaling exponent as

$$\gamma = 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min} - \frac{1}{2}} \right]^{-1}. \tag{18}$$

The fitting procedure is as follows [38][8]. The maximum likelihood, as described above, estimates the scaling parameter γ for each possible value of $x_{\min}$. Then, the Kolmogorov–Smirnov goodness-of-fit statistic KS is computed. This is done by computing the maximum distance between the cumulative distribution function (CDF) of the data and the fitted model:

$$KS = \max_{x \geqslant x_{\min}} |S(x) - P_{\text{cum}}(x)|, \tag{19}$$

where $S(x)$ is the CDF of the data for the observations with value at least $x_{\min}$, and P_{cum} is the CDF for the power-law distribution that best fits the data. The estimate of $x_{\min}$ is determined as the value that gives the minimum value KS over all values of $x_{\min}$. The results for the scaling exponent γ are shown in figure 8, together with the standard errors on γ, which are derived from the width of the likelihood maximum [38].

4.2. The MDS targets high-degree nodes in real networks

In figure 6, we plot the average degree of MDS nodes $\langle k_D \rangle$ as a function of the average degree $\langle k \rangle$ of each network in table 1. The results show that $\langle k_D \rangle$ is significantly higher than the average degree of the network $\langle k \rangle$, suggesting that MDS tends to hit high-degree nodes in real systems. This finding is also supported by analyzing the top 5% highly connected nodes, shown in figure 7. In all cases except two, the fraction of the top 5% high-degree nodes included in the dominating set is above about 50% and exceeds 80% in several cases. The lowest fractions correspond to a type of electronic circuit network whose degree distribution decays faster than a power law and the Wiki-Vote network. Moreover, by classifying the networks into two types, the results indicate that cellular networks tend to target a much larger fraction of hubs than socio-technical networks.

Furthermore, we have also investigated whether the tendency observed in figure 7 is also present in computer simulated networks. The fraction of the top 5% high-degree nodes within the MDS was analyzed in synthetic scale-free networks with different size, from $N = 100$, 500, 1000, 5000 and 10 000 nodes. The degree exponent was set as $\gamma = 2.5$ and with a fixed average degree $\langle k \rangle = 3$. The results show that for small network sizes of $N = 100$, 500 and 1000, all the top 5% high-degree nodes are included in the MDS. For larger networks ($N = 5000$ and 10 000) the fraction of 5% nodes within the MDS is 97.6 and 97.8%, respectively. For real biological

---

[8] http://tuvalu.santafe.edu/aaronc/powerlaws
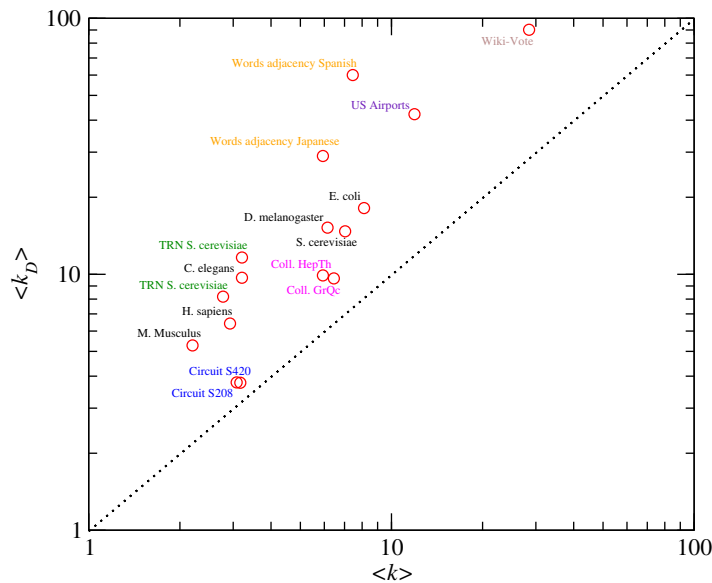
Figure 6. Average degree of the nodes in MDS $\langle k_{\mathrm{D}} \rangle$ compared with the average degree $\langle k \rangle$ of all nodes in each real network shown in table 1.

networks this fraction represents 80%. Although the MDS in large synthetic networks tends to target more high-degree nodes than real-world networks of similar size, we also observe that some of the small real networks, such as electronic circuit S208 and PPI *M. musculus*, also show larger fractions of high-degree nodes within the MDS. A reason for the small discrepancy between the synthetic and real networks may be that the real-world networks have a richer local structure, which includes clustering and degree correlations. These kinds of local patterns may have an influence on the selection of MDS.

4.3. The dependence of the scaling exponent and average degree on the MDS in real networks

To investigate in more detail the topology behind the MDS selection of high-degree nodes in real-world networks, we plot the MDS size as a function of each network's scaling exponent γ. A small degree exponent correlates with hubs with higher degrees. In contrast, high-degree exponents exhibit weaker connected hubs, which tend to vanish when the decay is exponential or faster than a power law. The results in figure 8 indicate that MDS size tends to decrease as γ decreases and only reaches higher values as γ increases. This highlights the role of the scaling exponent in scale-free systems, suggesting that networks with a small scaling exponent could be more easily controlled by a smaller set of nodes. These results agree well with the predictions of the theoretical analysis shown in section 3. Network dominance in the average degree is also investigated by plotting MDS size as a function of the average degree observed in real networks. Figure 9 shows that networks with higher average degrees tend to have smaller MDSs. In table 1, we show the main network features of the analyzed networks. There, we can see that the Wiki network has a extraordinarily high average degree, much higher than the values observed for networks of a similar size. We consider this network as an exceptional case with values $(\langle k \rangle, m_{\mathrm{D}}(\%)) = (28.3, 15.4)$. We also remark here that a strict comparison of the results shown in figures 8 and 9 with those from computer simulations shown in figures 2 and 3
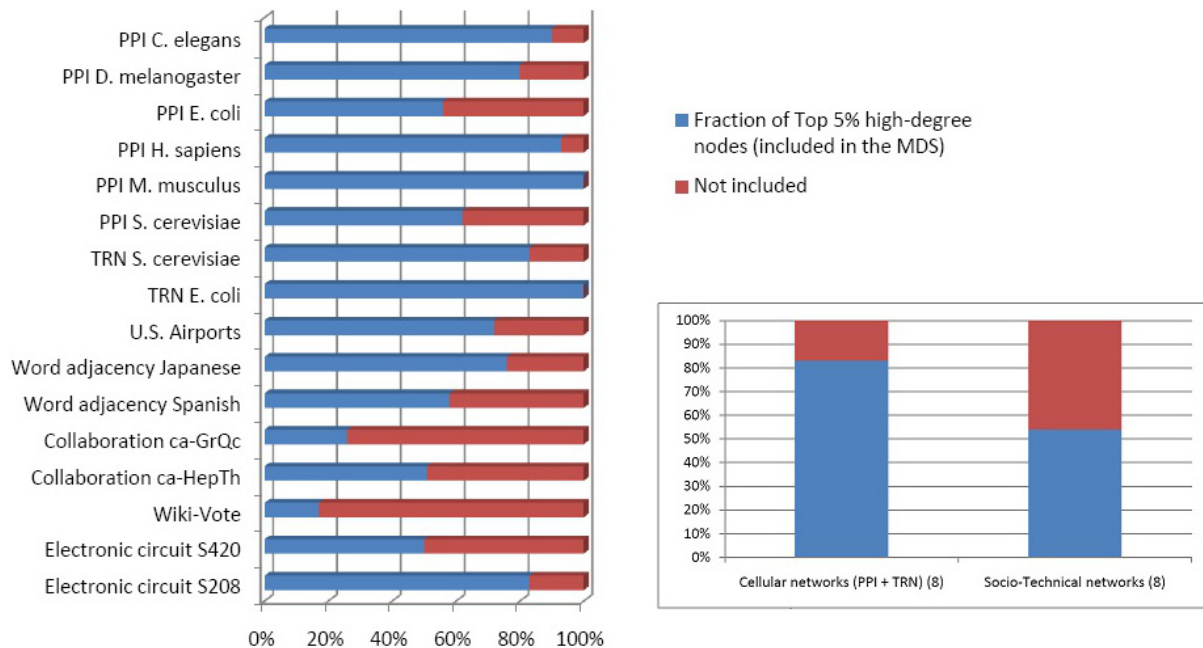
Figure 7. The bars in blue show the fraction of the top 5% high-degree nodes included in the dominating set. Red bars show the fraction of the top 5% high-degree nodes not included in the dominating set. In all cases except two, the fraction of the top 5% high-degree nodes included in the dominating set is above about 50% and exceeds 80% in several cases. The classification of the real networks into two types indicates that the MDS in cellular networks tends to target on average a larger fraction of hubs than socio-technical networks. We note that in the socio-technical networks we have included electronic circuits with decay faster than a power law.

should be made with great care. As shown in figure 4, MDS strongly depends on both γ and $\langle k \rangle$ parameters. Figure 3 also shows that the dependence on γ is strong. In real networks, we do not have enough statistics to classify networks with specific values of average degree and scaling exponent. Therefore, the results shown in figures 8 and 9 contain networks with various values of $\langle k \rangle$ and γ, which explains the dispersion of the data when compared with figures 2 and 3, respectively.

4.4. Dependence on more complex network metrics

In figure 10, we plot MDS size as a function of the average shortest path, diameter, average clustering degree and network centrality. The result shows a positive correlation between the MDS size and the average shortest path and the diameter. In contrast, the MDS size is anti-correlated with the average clustering degree and network centrality. It seems that, although both correlations are not strong because as discussed before the networks also do not have the same $\langle k \rangle$ or γ values, there is some tendency to have a smaller dominating set when both the path length and the diameter of the network decrease. Moreover, when the network centrality and clustering degree increase, the size of the dominating set also decreases.
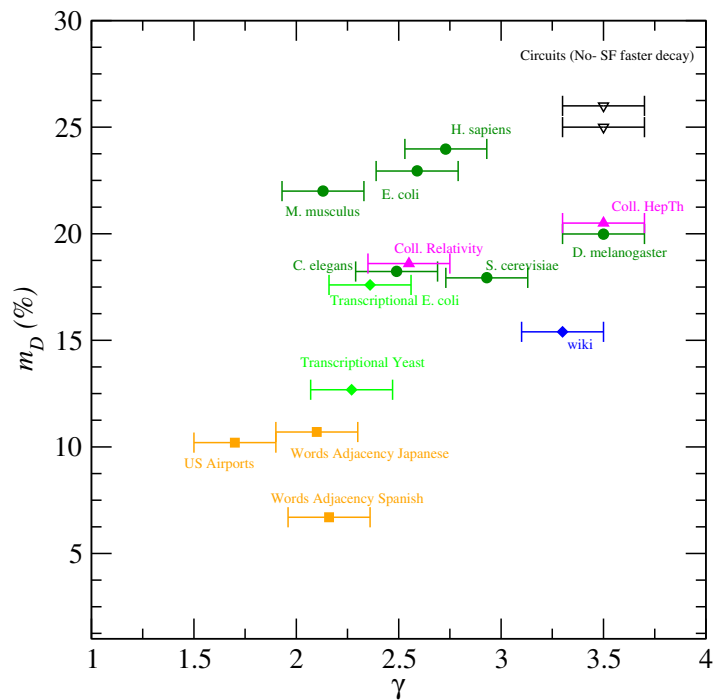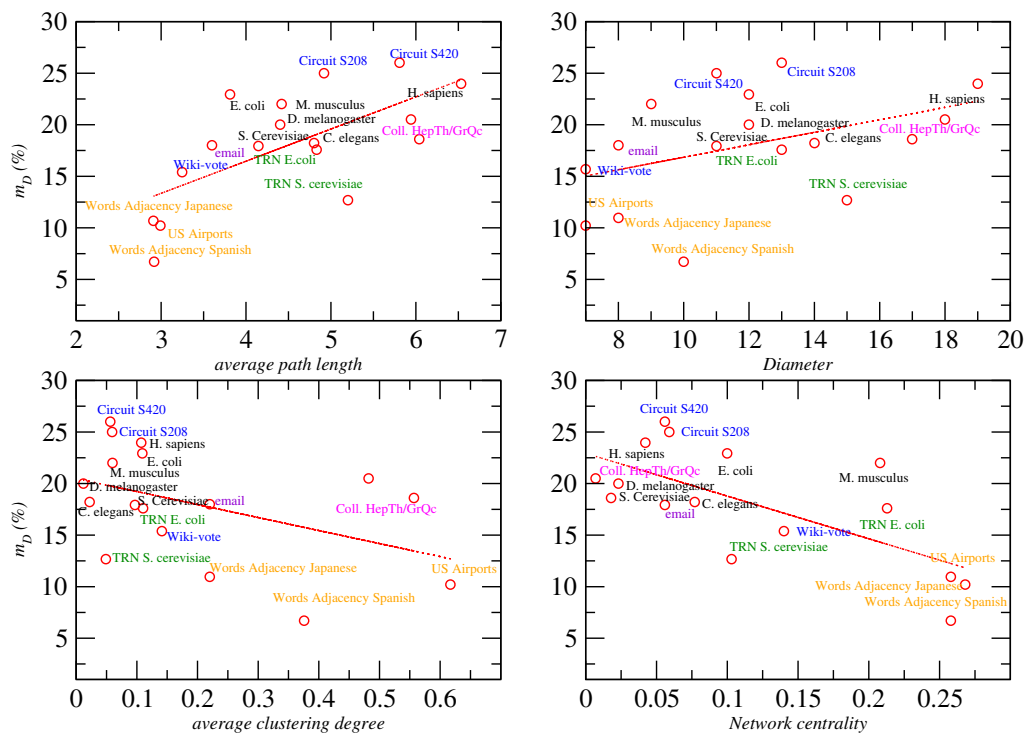
Figure 8. MDS size as a function of the scaling exponent γ of each real network. Electronic circuits exhibit a decay faster than a power-law. Error bars are derived from the width of the likelihood maximum [38].
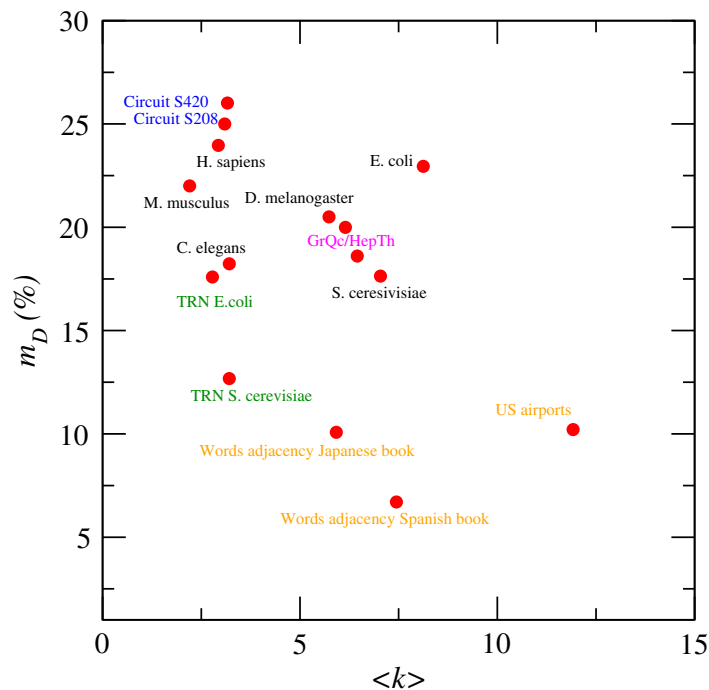
Figure 9. MDS size as a function of the average degree of all nodes in real networks.

Figure 10. MDS size as a function of the average shortest path (a) ($r = 0.661$), diameter (b) ($r = 0.415$), average clustering degree (c) ($r = -0.449$) and network centrality (d) ($r = -0.686$). The correlation coefficient is indicated in parentheses.

4.5. Degree distribution of nodes within the MDS

As shown in section 4.2, the MDS tends to target high-degree nodes in real networks. In particular, the fraction of the top 5% hubs within MDS is relatively high. However, this observation does not mean that all the nodes within a dominating set are highly connected nodes. Note that we are just focusing on the top 5% hubs in the previous analysis. In order to shed light on this issue, we investigate what the degree distribution of nodes within MDS looks like. The analysis shown in figures 11 and 12 indicates that the MDS is not only the upper part of the high-degree nodes. A broad distribution of heterogeneous degree nodes is visible in most of the real-world networks. It is worth noting that in some networks, the distribution also resembles a power-law functional form.

4.6. Relationship between assortativity and the MDS

In figure 7, we showed that cellular networks tend to target a larger number of hubs than social networks. On the other hand, degree correlations between nodes of similar degree have been identified in real-world networks. In social networks, highly connected nodes tend to be connected with other high-degree nodes. This tendency is referred to as assortative mixing, or assortativity. In contrast, technological and biological networks usually show disassortative mixing, or dissortativity, as high-degree nodes tend to connect to low-degree nodes. We have computed the assortativity of the analyzed networks. In figures 13 and 14, we show that
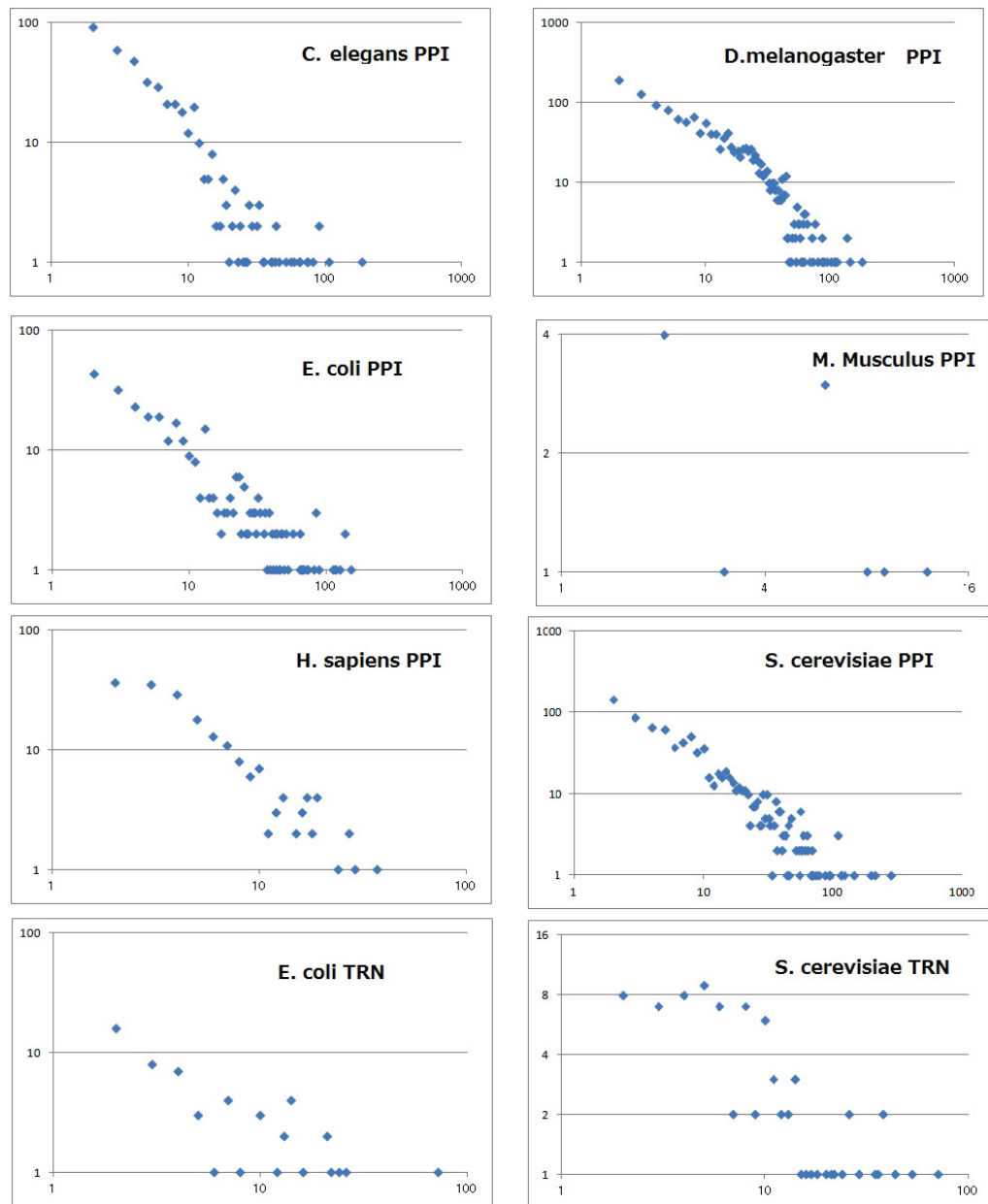
Figure 11. Degree distribution of nodes in biological networks within the MDS. Vertical axis: frequency of nodes with degree k. Horizontal axis: node with degree k.

our biological networks are dissortative and social networks, such as research collaborations, display clear assortative behavior. Based on figure 7, we could then conclude at first glance that assortative networks tend to have a lower fraction of high-degree nodes within the MDS. However, we should be careful since there are exceptions. The Wiki-Vote network, for example, tends to target a small number of hubs and its assortativity behavior is not clear. Electronic circuits do not have a definite degree correlation pattern. The words adjacency network could be classified as a social network; however, the analysis shows clear dissortative behavior. To
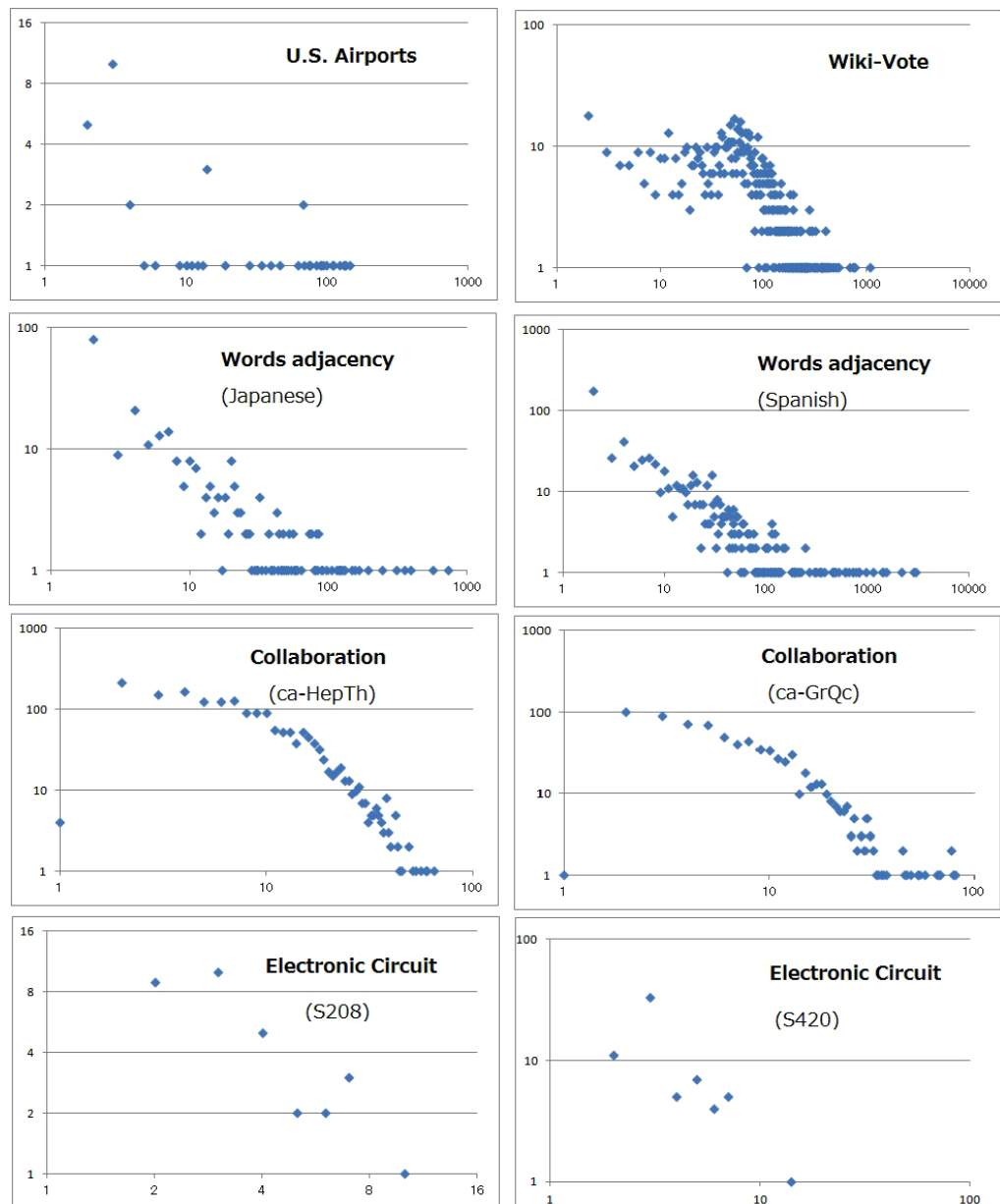
Figure 12. Degree distribution of nodes in socio-technical networks within the MDS. Vertical axis: frequency of nodes with degree k. Horizontal axis: node with degree k.

sum up, although our study does not show strong evidence for a correlation between MDS and dissortative/assortative behavior, further specific studies on this issue are encouraged.

4.7. Dependence of the MDS size on the system size

Here we investigate whether there is a characteristic scaling between the size of the MDS and the system size. From a theoretical viewpoint, for $\gamma > 2.0$ the MDS size should scale as $O(n)$. Figure 15 shows a similar dependence for computer-simulated networks with $\gamma = 2.5$ and
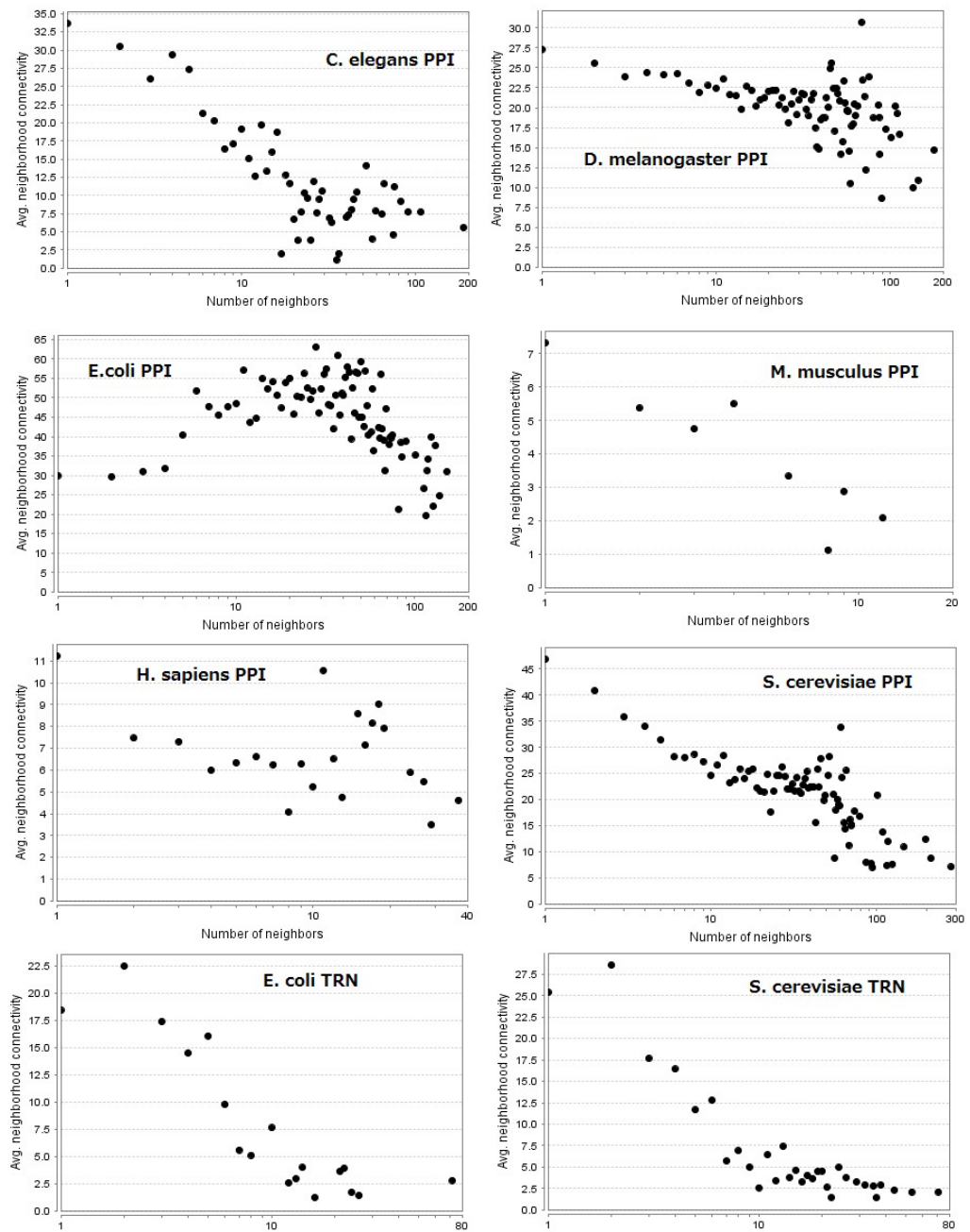
Figure 13. Average degree of neighbors of a node with degree k for biological networks.

$\gamma = 3.5$. For $\gamma < 2$, as n increases, the MDS size should decrease. Although the simulated data show this tendency to some extent, it is not completely clear from figure 15 for the cases $\gamma = 1.5$ and 1.75. Note that we have only provided a non-tight upper bound of the size of MDS for the case of $\gamma < 2$. A more detailed theoretical analysis of the case $\gamma < 2$ is left for future work.
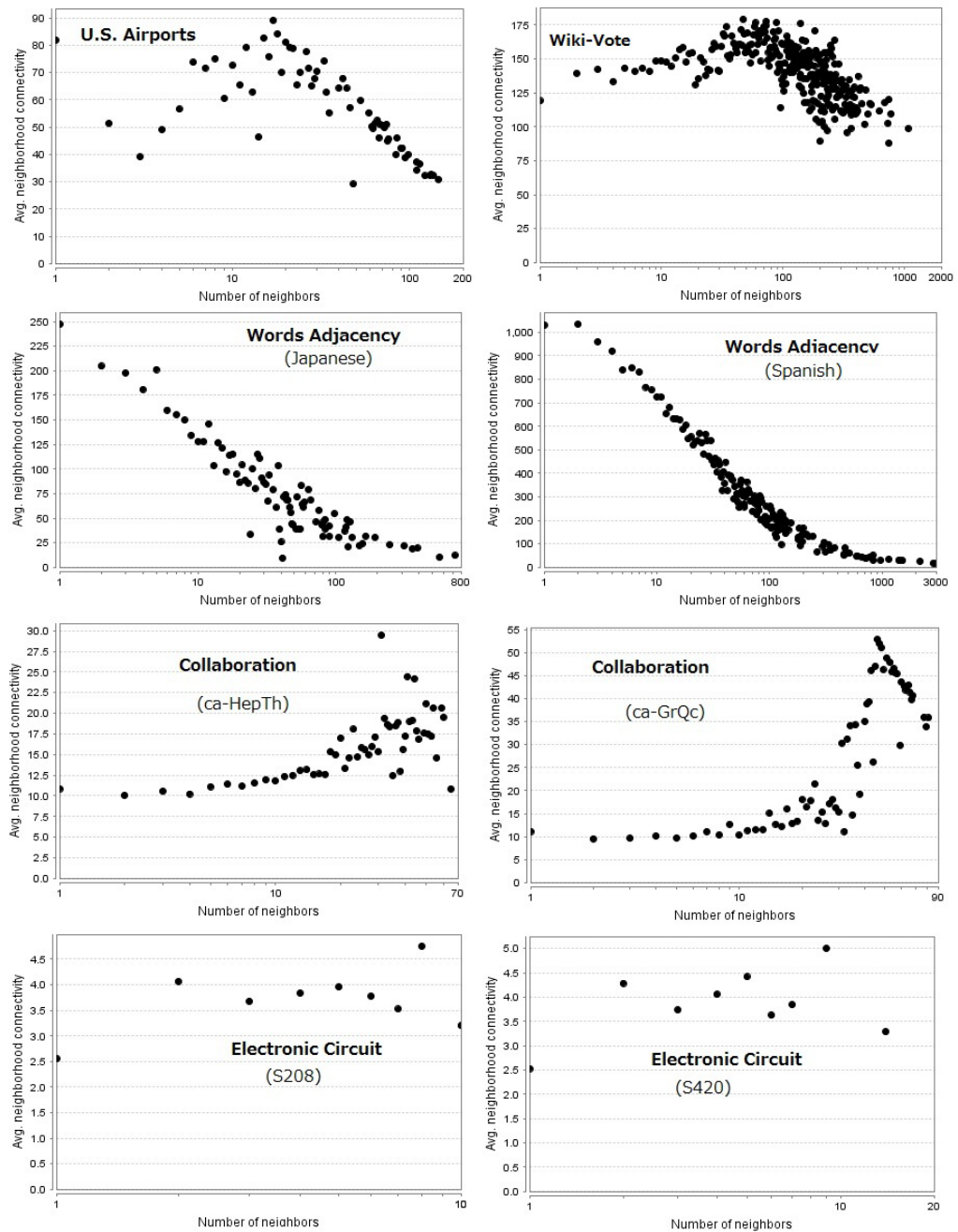
Figure 14. Average degree of neighbors of a node with degree k for socio-technical networks.

5. Discussion and conclusion

Although topological features have been extensively studied over the last decade, the dynamic and control aspects of complex networks have not followed the same pace of development. We have provided a set of tools that facilitate designing networks with increased controllability. We find that not all the real networks with fat-tailed degree distributions lead to similar
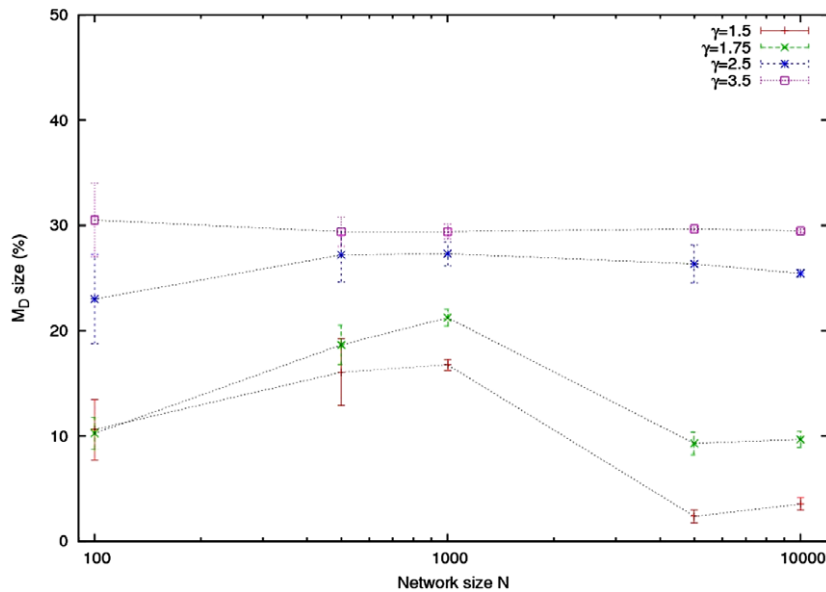
Figure 15. Computer simulations of the MDS size as a function of the network size. Each line indicates a synthetic network constructed with a specific degree exponent γ as shown in the legend and with average degree $\langle k \rangle = 3$.

(a) Model by Liu et al.   (b) MDS model
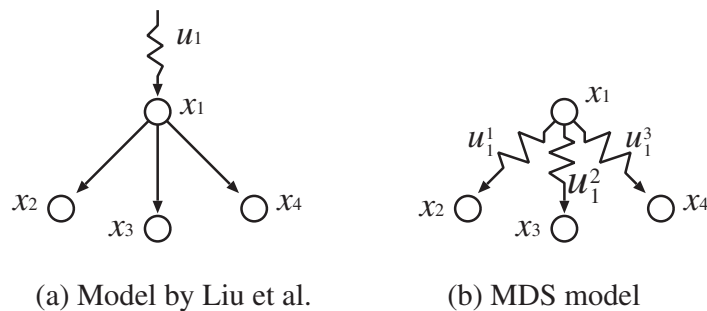
Figure 16. Comparison between the model proposed by Liu *et al* [9] and our approach. The network is structurally controllable by selecting an MDS because each dominated node has its own control signal.

MDS sizes. Our findings highlight a specific configuration of scale-free networks based on a scaling exponent $\gamma < 2$ that leads to a significant reduction of MDS size. Therefore, a few key individuals or devices could in principle dominate a very large network. In sharp contrast to a previous study on network control based on the maximum matching problem [9], we find that the degree of heterogeneity and denseness facilitates network control. However, our results do not contradict the recent results of Liu *et al* [9]. As in figure 16, they assumed that only driver node values can be directly controlled through external signals. Conversely, the MDS approach assumes that each driver node can control its links individually. Therefore, a node with degree k is treated as if it were a set of k nodes. This assumption is reasonable for artificial networks, such as the Internet and power grids, where each node is sufficiently smart to control individual links separately, although this assumption may remain unclear for biological networks. Therefore, our findings complement the results of Liu *et al*.

IOP Institute of Physics ⊕ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

A recent study by Müller and Schuppert [10] has suggested that a few inputs could reprogram biological networks, giving iPS cells as an example. Although their work shows that network control techniques could also make an impact on synthetic molecular biology, their study is entirely empirical and does not present general results. Moreover, the analysis does not include network features. In contrast, we have provided here an extensive data analysis and computer simulations that demonstrate correlations between network structures and MDS size, in agreement with the theoretical predictions.

Our approach has demonstrated that the more homogeneous a network, the larger the fraction of individuals required for dominating the entire system. The theoretical analysis and computational experiments unveiled a critical value of the scaling exponent $\gamma = 2$ below which scale-free networks show drastically different behavior for control purposes. More complex network metrics, such as clustering degree, network centrality, average shortest path and diameter, also correlating with MDS size, provide new tools for addressing network control design.

In summary, our results offer a new view of the large-scale network control problem from the MDS perspective and provide simple control design tools and principles to address controllability for complex networks. We envision that, in an increasingly smart e-society, mobile and MANET networks [17], computer and cluster communication networks and sensor and camera vigilance systems could provide optimal coverage and monitoring performance, or control an entire network, using only a small fraction of devices as long as the network connectivity is highly heterogeneous with $\gamma < 2$, dense, clustered and central, and exhibits both a small average shortest path and diameter. Although the deployment of such smart infrastructures is at an early stage, the possibility of predicting optimal device connectivity in large-scale networks using inexpensive computer simulations and network science tools suggests a promising start.

Acknowledgments

References

[1] Vespignani A 2009 Predicting the behavior of techno-social systems *Science* **325** 425–8
[2] Schweitzer F, Fagiolo G, Sornette D, Vega-Redondo F, Vespignani A and White D R 2009 Economic networks: the new challenges *Science* **325** 422–5
[3] Barabási A-L, Gulbahce N and Loscalzo J 2011 Network medicine: a network-based approach to human disease *Nature Rev. Genet.* **12** 56–68
[4] Albert R and Barabási A-L 2002 Statistical mechanics of complex networks *Rev. Mod. Phys.* **74** 47–97
[5] Dorogovtsev S N and Mendes J F F 2003 *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford: Oxford University Press)
[6] Kim J S, Goh K-I, Kahng B and Kim D 2007 Fractality and self-similarity in scale-free networks *New J. Phys.* **9** 177
[7] Newman M E J 2010 *Networks: An Introduction* (New York: Oxford University Press)
[8] Oh E, Lee D-S, Kahng B and Kim D 2007 Synchronization transition of heterogeneously coupled oscillators on scale-free networks *Phys. Rev.* E **75** 011104
[9] Liu Y-Y, Slotine J-J and Barabási A-L 2011 Controllability of complex networks *Nature* **473** 167–73
[10] Müller F-J and Schuppert A 2011 Few inputs can reprogram biological networks *Nature* **478** E4

[11] Meyn S 2008 *Control Techniques for Complex Networks* (New York: Cambridge University Press)

[12] Akutsu T, Hayashida M, Ching W-K and Ng M K 2007 Control of Boolean networks: hardness results and algorithms for tree structured networks *J. Theor. Biol.* **244** 670–9

[13] Lombardi A and Hornquist M 2007 Controllability analysis of networks *Phys. Rev.* E **75** 56110

[14] Liu B, Chu T, Wang L and Xie G 2008 Controllability of a leader–follower dynamic network with switching topology *IEEE Trans. Autom. Control* **53** 1009–13

[15] Rahmani A, Ji M, Mesbahi M and Egerstedt M 2009 Controllability of multi-agent systems from a graph-theoretic perspective *SIAM J. Control Optim.* **48** 162–89

[16] Kim D-H and Motter A E 2009 Slave nodes and the controllability of metabolic networks *New J. Phys.* **11** 113047

[17] Blum J, Ding M, Thaeler A and Cheng X 2004 Applications of connected dominating sets in wireless networks *Handbook of Combinatorial Optimization* ed D-Z Du and P Pardalos (Dordrecht: Kluwer) pp 329–69

[18] Alzoubi K, Wan P J and Frieder O 2002 Message-optimal connected dominating sets in mobile *ad hoc* networks *Proc. 3rd Int. Symp. on Mobile ad hoc Networks and Computing* (New York: ACM) pp 157–64

[19] Stojmenovic I, Seddigh M and Zunic J 2002 Dominating sets and neighbor elimination-based broadcasting algorithms in wireless networks *IEEE Trans. Parallel Distrib. Syst.* **13** 14–25

[20] Cokuslu D, Erciyes K and Dagdeviren O 2006 A dominating set-based clustering algorithm for mobile *ad hoc* networks *Proc. Int. Conf. on Computational Science* (*Lecture Notes in Computer Science* vol 3991) (Berlin: Springer) pp 571–8

[21] Haynes T W, Hedetniemi S T and Slater P J 1998 Fundamentals of domination in graphs *Pure Applied Mathematics* (New York: Chapman and Hall/CRC)

[22] Sampath A and Thampi S M 2011 An ACO algorithm for effective cluster head selection *J. Adv. Inf. Technol.* **2** 50–6

[23] Wolsey L A 1998 *Integer Programming* (New York: Wiley)

[24] Nemhauser G L and Wolsey L A 1998 *Integer and Combinatorial Optimization* (New York: Wiley)

[25] Molloy M and Reed B 1995 A critical point for random graphs with a given degree sequence *Random Struct. Algorithms* **6** 161–79

[26] Newman M E J 2003 The structure and function of complex networks *SIAM Rev.* **45** 167–203

[27] Viger F and Latapy M 2005 Efficient and simple generation of random simple connected graphs with prescribed degree sequence *Proc. COCOON'05 (Computing and Combinatorics Conf.)* pp 1–15

[28] Cooper C, Klasing R and Zito M 2005 Lower bounds and algorithms for dominating sets in web graphs *Internet Math.* **2** 275–300

[29] Cooper C and Zito M 2009 An analysis of the size of the minimum dominating sets in random recursive trees, using the Cockayne–Goodman–Hedetniemi algorithm *Discrete Appl. Math.* **157** 2010–14

[30] Artmann S, Göring F, Harant J, Rautenbach D and Schniermeyer I 2010 Random procedures for dominating sets in graphs *Electron. J. Comb.* **17** R102

[31] Mitzenmacher M and Upfal E 2005 *Probability and Computation* (Cambridge: Cambridge University Press)

[32] Dong J and Horvath S 2007 Understanding network concepts in modules *BMC Syst. Biol.* **1** 24

[33] Colizza V, Pastor-Satorras R and Vespignani A 2007 Reaction–diffusion processes and metapopulation models in heterogeneous networks *Nature Phys.* **3** 276–82

[34] Milo R *et al* 2004 Superfamilies of evolved designed networks *Science* **303** 1538–42

[35] Leskovec J, Kleinberg J and Faloutsos C 2007 Graph evolution: densification and shrinking diameters *ACM Trans. Knowl. Discovery Data (ACM TKDD)* **1** 2

[36] Leskovec J, Huttenlocher D and Kleinberg J 2010 Signed networks in social media *Proc. 28th ACM Conf. of Human Factors and Computing Systems (CHI) (Atlanta, GA, USA)*

[37] Newman M E J 2005 Power laws, Pareto distributions and Zipf's law *Contemp. Phys.* **46** 323–51

[38] Clauset A, Shalizi C R and Newman M E J 2009 Power-law distributions in empirical data *SIAM Rev.* **51** 661–703