Controllability of Complex Networks

Yang-Yu Liu, Jean-Jacques Slotine, Albert Laszlo Barabasi

Presented by Yuanyuan Bao

Controllability of Complex Networks

- Science-Reviews
  - Scientific Link-up Yields ‘Control Panel’ for Networks
- Guanrong Chen
- Network Science: Complex Networks
  - Control is Aim!

Authors

- **Yang-Yu Liu**

  Assistant Professor, Northeastern University

- **Jean-Jacques Slotine**

  Director, MIT Nonlinear Systems Laboratory

- **Albert Laszlo Barabasi**

  Director, Center for Complex Network Research, Northeastern University

Albert Laszlo Barabasi

This Man Could Rule the World!

POPULAR SCIENCE
THE FUTURE NOW

Nature:20

Science: 7

Albert Laszlo Barabasi

- Power-law distribution of world wide web--Science
- Scale-free Network----Scientific American
- Hubs
- Burst—93% accuracy
- Controllability of complex networks---Nature

Barabasi's research may soon allow us to not just understand and predict network behavior, but also to control it.

Complex Networks

- A complex network is a graph with non-trivial topological features—features that do not occur in simple networks such as lattices or random graphs but often occur in **real graphs**. The study of complex networks is a young and active area of scientific research inspired largely by the empirical study of real-world networks such as **computer networks and social networks**.
- Two well-known and much studied classes of complex networks are **scale-free networks** and **small-world networks**, whose discovery and definition are canonical case-studies in the field. Both are characterized by specific structural features—**power-law degree distributions** for the former and **short path lengths and high clustering** for the latter.

Controllability

- A dynamical system is controllable if, with a suitable choice of inputs, it can be driven from any initial state to any desired final state within finite time.
- Like a driver prompting a car to move with the desired speed and in the desired direction by manipulating the pedals and the steering wheel.
- Two independent factors contribute to controllability
    - ➢ The system's architecture, represented by the network encapsulating which components interact with each other;
    - ➢ The dynamical rules that capture the time-dependent interactions between the components.

Driver Nodes

- Driver nodes:
  - ➢ The set of nodes, if driven by different signals, can offer full control over the network.
- N_D:
  - ➢ The minimum number of driver nodes, whose control is sufficient to fully control the system's dynamics.

Motivation

- Whether some networks are easier to control than others?
- How network topology affect a system's controllability?

We continue to lack general answers to these questions for large weighted and directed networks, which most commonly emerge in complex systems.

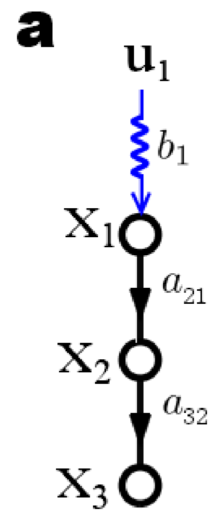
Questions

- What is the minimum number of driver nodes (N_D) of real-world networks?
- How to locate them efficiently?
- Which topological characteristics determine N_D?
- How robust of network controllability?

Outline

- **Network controllability**
  - **Minimum number of driver nodes (N_D)**
- Controllability of real networks
  - Network topology (degree distribution)
- An analytical approach to controllability
  - N_D compatible with $P(k_{in},k_{out})$
- Robustness of control

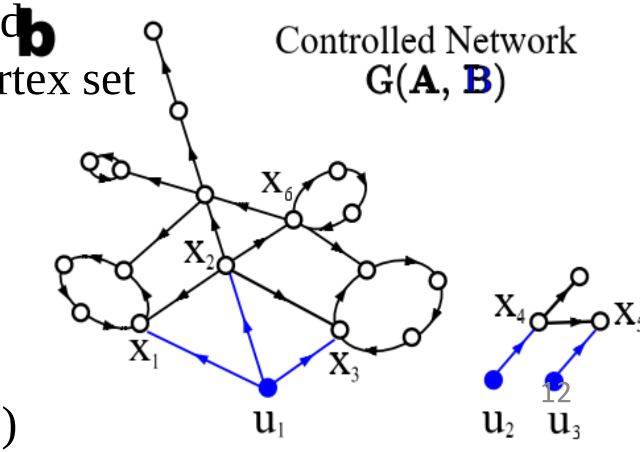
Network Controllability



- Linear time-invariant (LTI) System:
- $\frac{dx(t)}{dt} = Ax(t) + Bu(t)$
  - $x(t) \in R^{N\times 1}, state\ vector$
  - $A \in R^{N\times N}, state\ matrix$
  - $B \in R^{N\times M}, input\ matrix$
  - $u(t) \in R^{M\times 1}, input\ or\ control\ vector$
  - $A \coloneqq (a_{ij})_{N\times N}$, a_{ij} is 0 if (j→i) is not a link in G(A), i.e. node-j does not affect node-i.

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} b_1 \\ 0 \\ 0 \end{bmatrix} u(t)$$

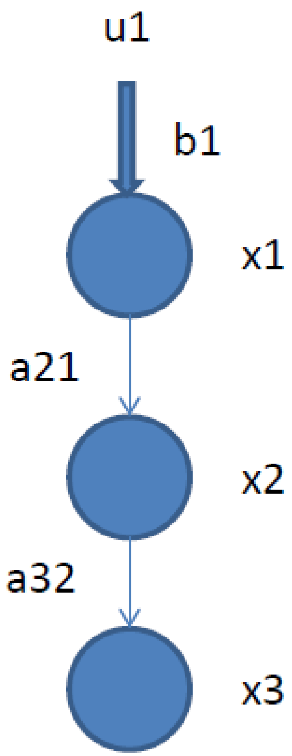
- The whole system, denoted as (A, B) can be represented by a diagraph G(A, B)=(V, E) with $V = V_A \cup V_B$ the vertex set and E $= E_A \cup E_B$ the edge set.
  - $V_A = \{v_1, \ldots, v_N\}$ is the set of state vertices
  - $V_B = \{u_1, \ldots, u_M\}$ is the set of input vertices
  - input vertices: u_1, u_2, u_3;
  - controlled nodes: x_1, x_2, x_3, x_4, x_5
  - driver nodes: x_2, x_4, x_5 (do not share input vertices)



LTI System and Kalman's controllability rank condition

- Linear time-invariant (LTI) System: $\frac{dx(t)}{dt} = Ax(t) + Bu(t)$ is controllability if the controllability matrix C has full rank.
  - $C=(B, AB, A^2B, \ldots A^{N-1}B)$
  - rank(C)=N

Network Controllability



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

$$\begin{bmatrix} x1\dot{}(t) \\ x2\dot{}(t) \\ x3\dot{}(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ a21 & 0 & 0 \\ 0 & a32 & 0 \end{bmatrix} \begin{bmatrix} x1(t) \\ x2(t) \\ x3(t) \end{bmatrix} + \begin{bmatrix} b1 \\ 0 \\ 0 \end{bmatrix} u1(t)$$

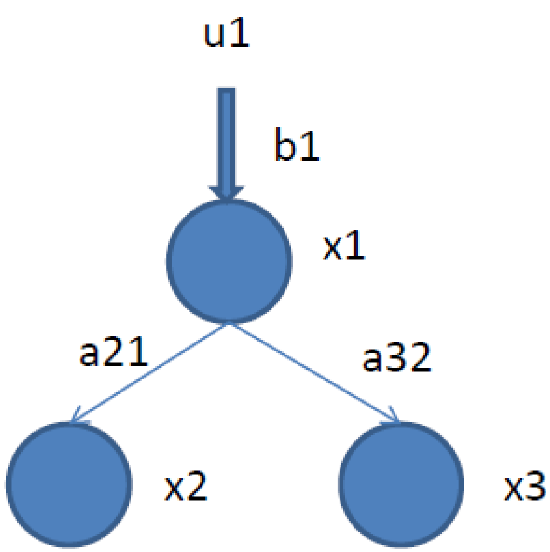
$$C = [B, AB, \cdots, A^{N-1}B]$$

$$= \begin{bmatrix} b1 & 0 & 0 \\ 0 & a21b1 & 0 \\ 0 & 0 & a32a21b1 \end{bmatrix}$$

rank(C)$= 3 = N$

This network is controllable.

Network Controllability



$$A = \begin{bmatrix} 0 & 0 & 0 \\ a21 & 0 & 0 \\ a31 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} b1 \\ 0 \\ 0 \end{bmatrix}$$

$$C = [B, AB, \cdots, A^{N-1}B] \\ = \begin{bmatrix} b1 & 0 & 0 \\ 0 & a21b1 & 0 \\ 0 & a31b1 & 0 \end{bmatrix}$$

rank(C)$= 2 \neq N$

This network is uncontrollable.

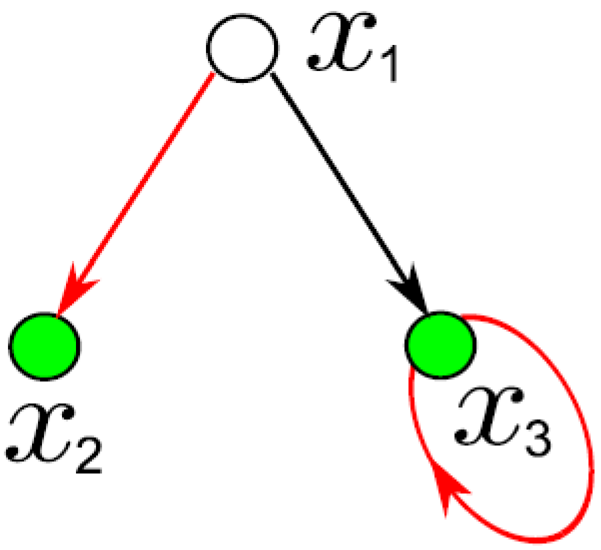
BUT

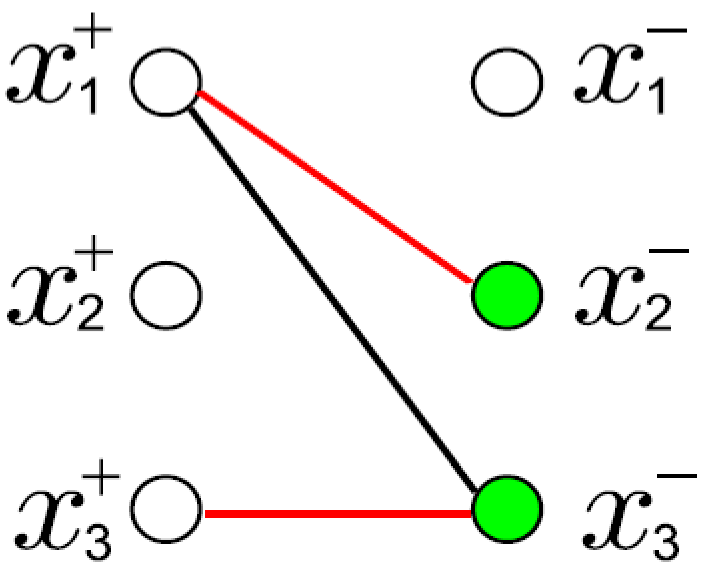
Network Controllability

- $\boldsymbol{A}$ is usually unknown or time-independent
  - Gene regulatory network, Internet
- Brute-force search
  - $C_N^1 + C_N^2 + \cdots + C_N^N = 2^N - 1$
- Kalman's rank condition is difficult to test for large-scale networks
  - $\boldsymbol{C} = [\boldsymbol{B}, \boldsymbol{AB}, \cdots, \boldsymbol{A}^{N-1}\boldsymbol{B}] \in R^{N \times NM}$

Matching

- Generalized matching M from undirected graph to diagraph:
  - Undirected: a set of edges without common nodes
  - Digraph: a set of edges without common starting or ending nodes
- A node is matched if it is an ending vertex of an edge in the matching. Otherwise, it is unmatched.





Minimum Input Theorem

- $N_I = N_D = \max\{N - |M^*|, 1\}$
  - Denote $|M^*|$ as the size of the maximum matching in the directed network G(A).
  - The minimum number of inputs (N_I) or equivalently the minimum number of driver nodes (N_D) need to fully control a network G(A) is one if there is a perfect matching in G(A). (In this case, any single node can be chosen as the driver node.) Otherwise, it equals the number of unmatched nodes with respective to any maximum matching. (In this case, the driver nodes are just the unmatched nodes.)
- Maximum matching M can be found efficiently using the well-known Hopcroft-Karp algorithm.
  - The algorithm runs in $O(N^{1/2}L)$, where L denotes the number of links.

Result 1

The possibility of determining N_D, using maximum matching, is the first main result of this paper.

Outline

- Network controllability
  - Minimum number of driver nodes (N_D)
- **Controllability of real networks**
  - **Network topology (degree distribution)**
- An analytical approach to controllability
  - N_D compatible with $P(k_{in},k_{out})$
- Robustness of control

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

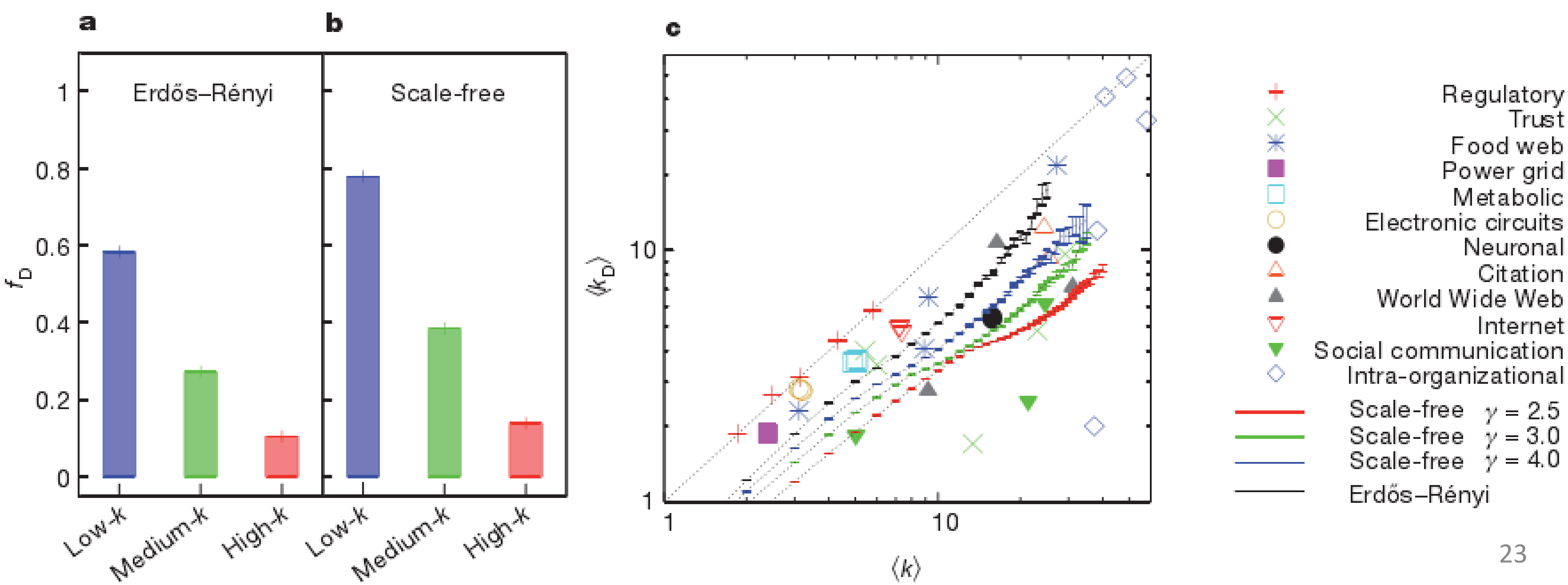
Type	Name	N	L	n_D^{real}	$n_D^{rand\text{-}Degree}$	$n_D^{rand\text{-}ER}$
Regulatory	TRN-Yeast-1	4,441	12,873	0.965	0.965	0.083
	TRN-Yeast-2	688	1,079	0.821	0.811	0.303
	TRN-EC-1	1,550	3,340	0.891	0.891	0.188
	TRN-EC-2	418	519	0.751	0.752	0.380
	Ownership-USCorp	7,253	6,726	0.820	0.815	0.480
Trust	College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

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

- Independently control about 80% of nodes to control them fully.
- Smallest N_D values, suggesting that a few individuals could in principle control the whole system.

HUBS

- Hypothesis: the control of hubs is essential to control a network
- Validation: divided the nodes into three groups of equal size according to their degree, k(low, medium and high)
- Conclusion: In both real and model systems the driver nodes tend to avoid the hubs.



Topology feature that determine network controllability

- Full randomization with N and L unchanged: $N_D^{\text{rand-ER}}$
- Degree-preserving randomization which keeps the in-degree, k_{in} and out-degree, k_{out}, of each node unchanged: $N_D^{\text{rand-Degree}}$
- Conclusion: N_D is determined mainly by the number of incoming and outgoing links each node has and is independent of where those links point.

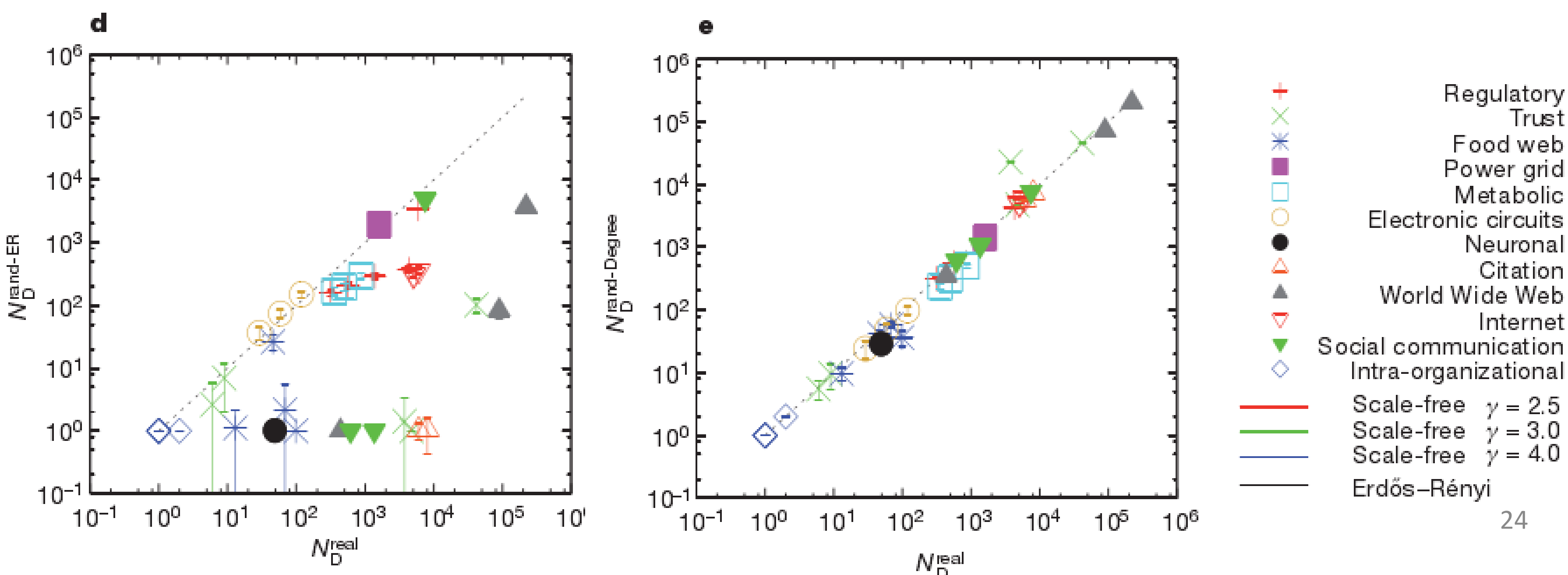


Table 1 | The characteristics of the real networks analysed in the paper

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

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

Result 2

A system's controllability is to a great extent encoded by the underlying network's degree distribution, $P(k_{in},k_{out})$, which is the second and most important finding.

Outline

- Network controllability
  - Minimum number of driver nodes (N_D)
- Controllability of real networks
  - Network topology (degree distribution)
- **An analytical approach to controllability**
  - **N_D compatible with $P(k_{in},k_{out})$**
- Robustness of control

Cavity Method: N_D & $P(k_{in}, k_{out})$

- The minimum density of unmatched nodes or equivalently the minimum density of driver nodes is given by

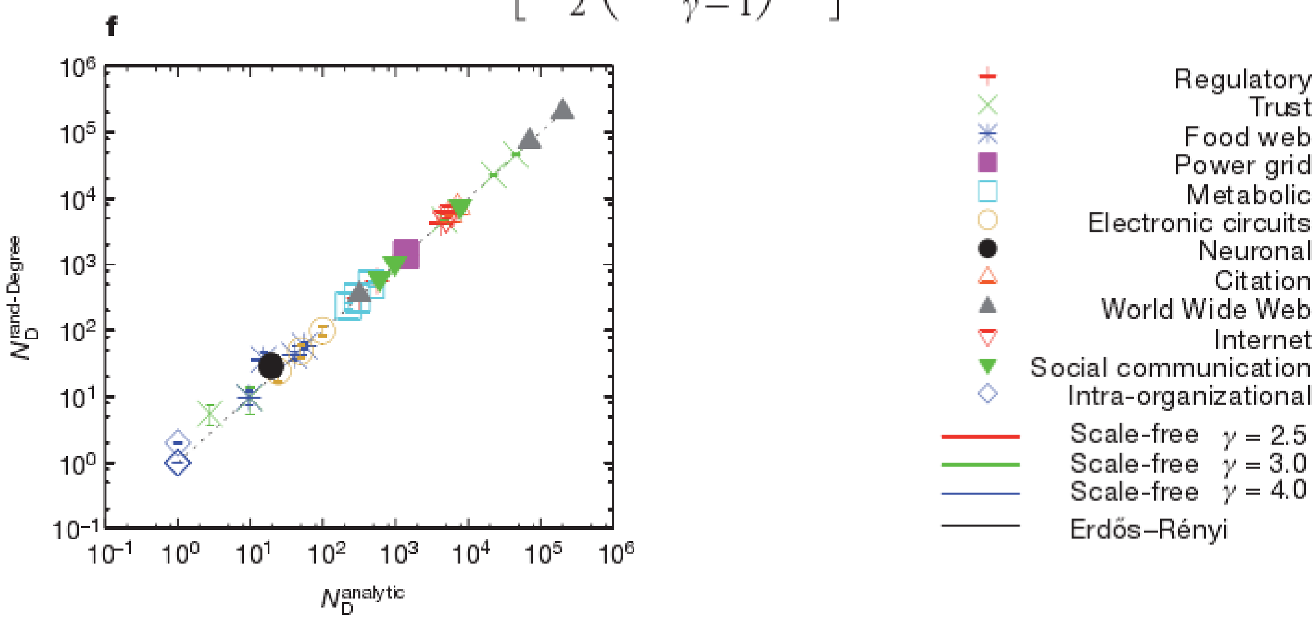
$$n_{\mathrm{D}} = \frac{1}{2}\left\{\left[G(\widehat{w}_2) + G(1-\widehat{w}_1) - 1\right] + \left[\widehat{G}(w_2) + \widehat{G}(1-w_1) - 1\right] + \frac{z}{2}\left[\widehat{w}_1(1-w_2) + w_1(1-\widehat{w}_2)\right]\right\}$$

- n_D of directed ER network:

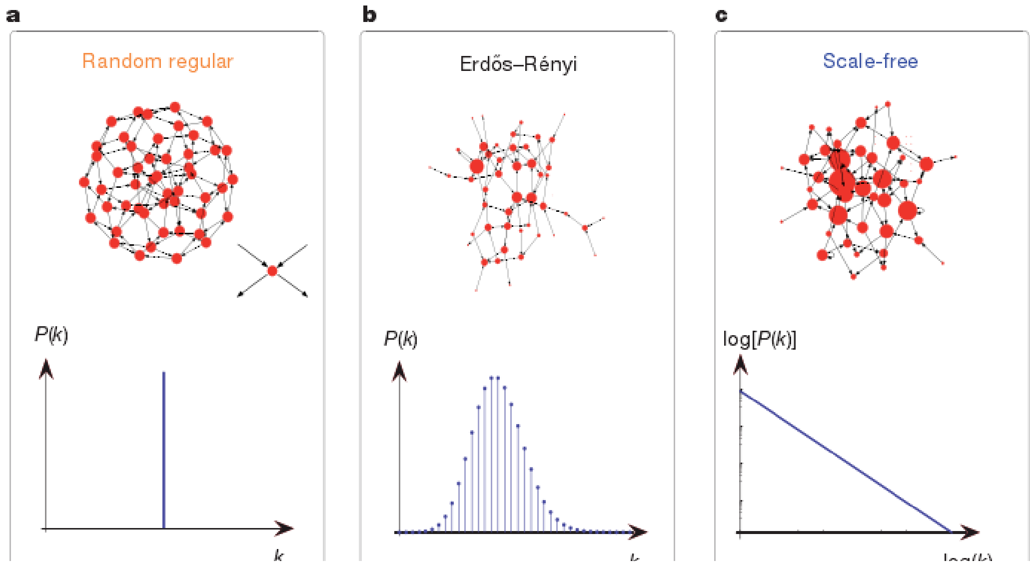
$$n_{\mathrm{D}} \approx \mathrm{e}^{-\langle k \rangle/2}$$

- n_D of scale-free network with degree exponent $Y_{in} = Y_{out} = Y$:

$$n_{\mathrm{D}} \approx \exp\left[-\frac{1}{2}\left(1-\frac{1}{\gamma-1}\right)\langle k \rangle\right]$$

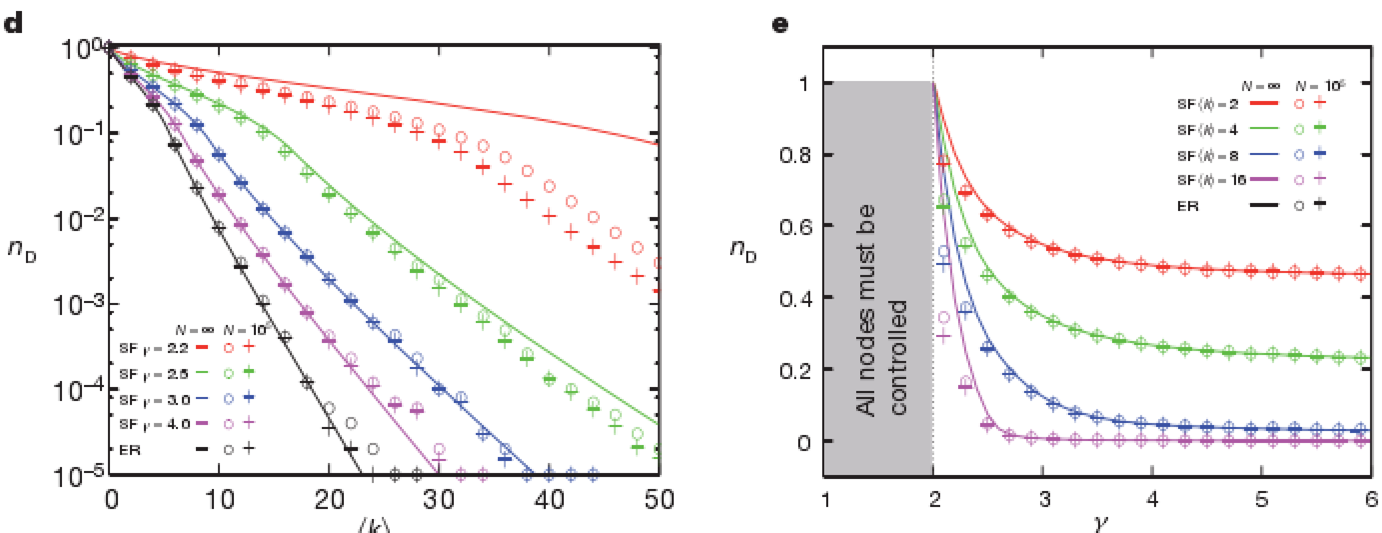


The impact of network structure on the number of driver nodes



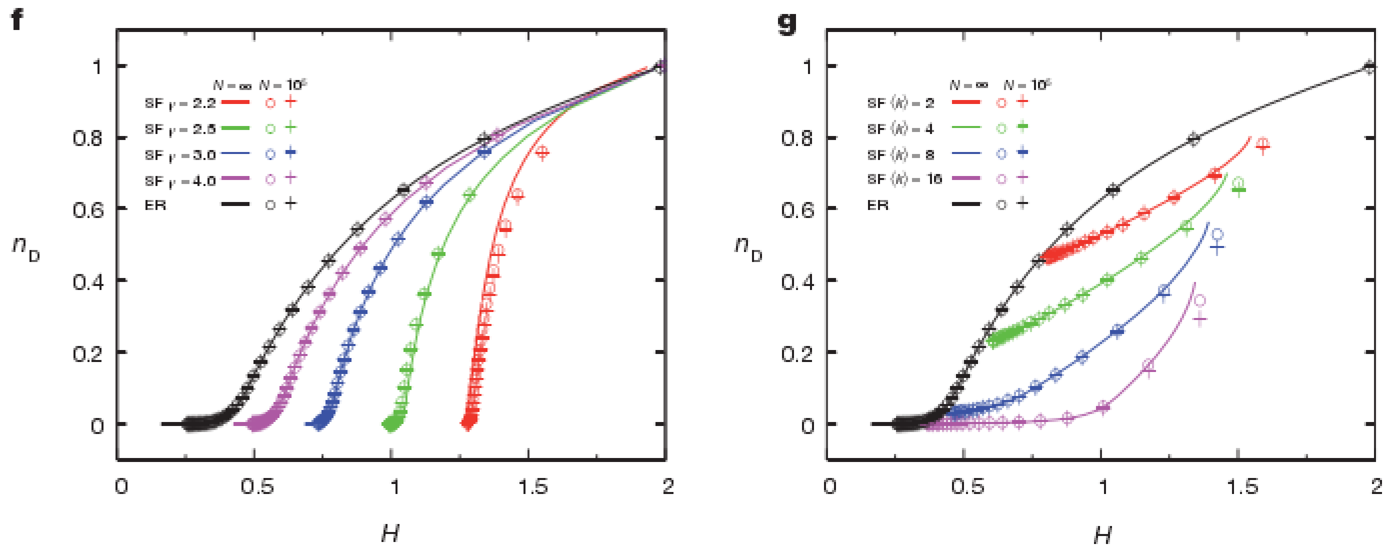
$$n_{\mathrm{D}} \approx e^{-\langle k \rangle/2}$$

$$n_{\mathrm{D}} \approx \exp\left[-\frac{1}{2}\left(1-\frac{1}{\gamma-1}\right)\langle k \rangle\right]$$



- <k> and n_D
- γ and n_D

The impact of network structure on the number of driver nodes



- Degree heterogeneity

$$H = \frac{\Delta}{\langle k \rangle} = \frac{\sum_i \sum_j |i-j| P(i) P(j)}{\langle k \rangle} = 2 \frac{\sum_i \sum_{j<i} (i-j) P(i) P(j)}{\langle k \rangle}$$

- H and n_D
  - f: fixed γ and variable <k>
  - g: fixed <k> and variable γ

Result 3

- Using the cavity method, we derived a set of self-consistent equations whose input is the degree distribution and whose solution is the average n_D(or N_D) over all network realizations compatible with $P(k_{in},k_{out})$, which is the third key result.
- The denser a network is, the fewer driver nodes are needed to control it, and that small changes in the average degree induce orders-of-magnitude variations in n_D.
- The larger are the differences between node degrees, the more driver nodes are needed to control the system.
- Overall, networks that are sparse and heterogeneous, which are precisely the characteristics often seen in complex systems like the cell or the Internet, require the most driver nodes, underscoring that such systems are difficult to control.

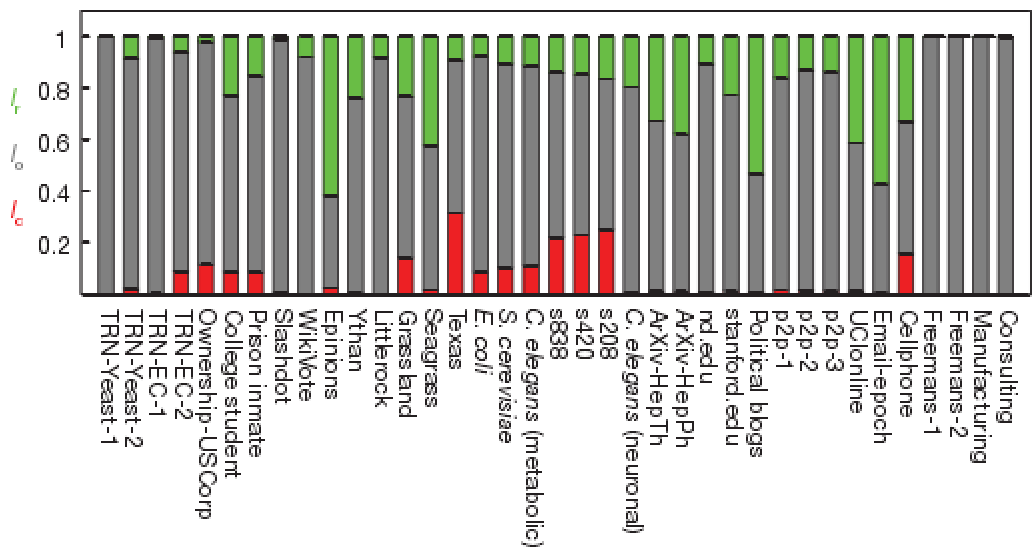
Outline

- Network controllability
  - Minimum number of driver nodes (N_D)
- Controllability of real networks
  - Network topology (degree distribution)
- An analytical approach to controllability
  - N_D compatible with $P(k_{in},k_{out})$
- **Robustness of control**

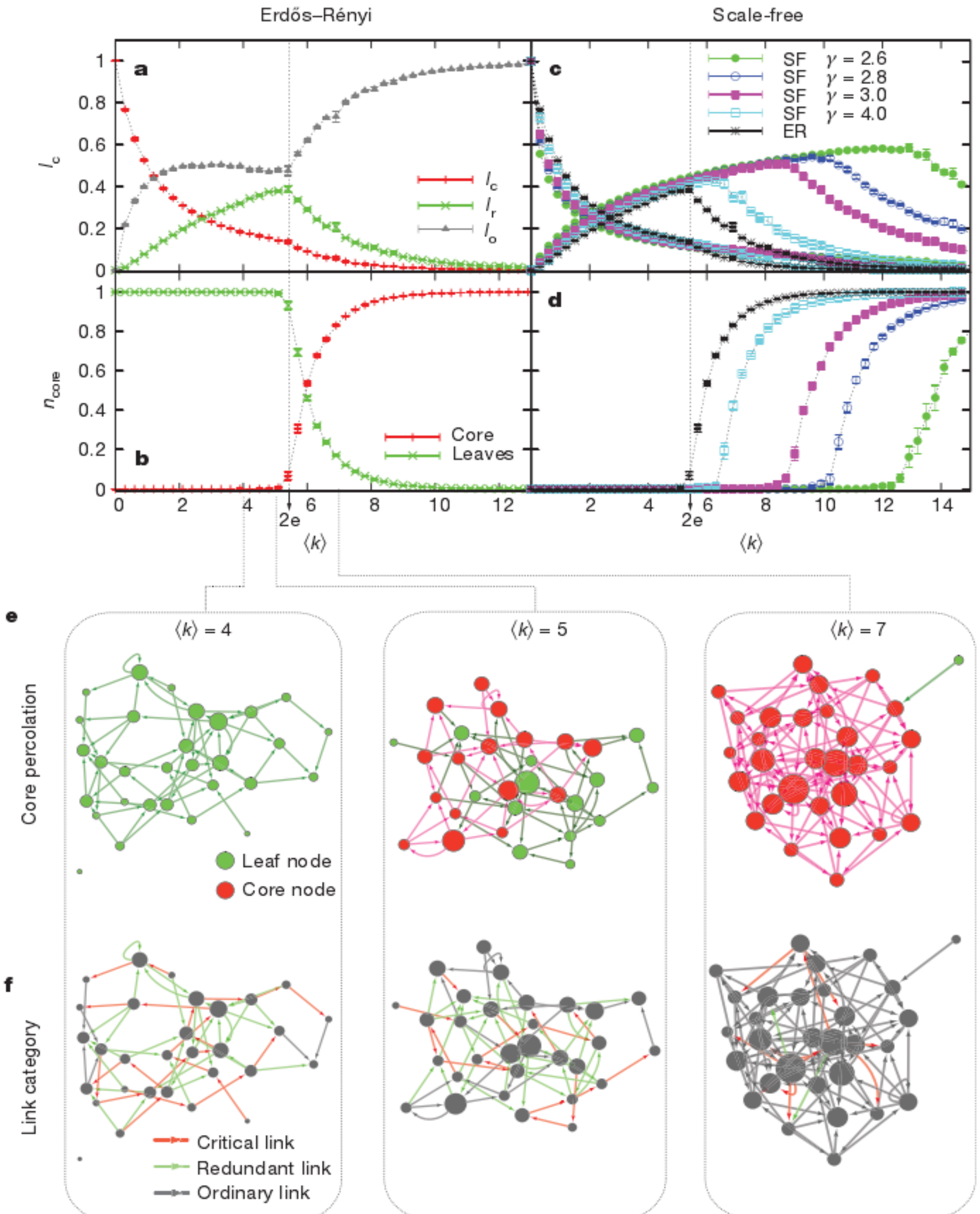
Robustness of control

- To see how robust is our ability to control a network under unavoidable link failure, we classify each link into one of the following three categories:
  - ‘critical’ if in its absence we need to increase the number of driver nodes to maintain full control; density of critical: lc=Lc/L
  - ‘redundant’ if it can be removed without affecting the current set of driver nodes; density of redundant: lr=Lr/L
  - ‘ordinary’ if it is neither critical nor redundant.

    density of ordinary: lo=Lo/L



Factors determine lc,lr and lo



- For small <k>, all links are essential for control (lc≈1). As <k> increases, the network's redundancy increases, decreasing lc.
- The increasing redundancy suggests that the density of redundant links, lr, reaches a maximum at a critical value of <k>, $<k>_c$, after which it decays.
- This non-monotonic behavior results from the competition of two topologically distinct regions of a network, the core and leaves.
- Indeed, lr starts decaying at $<k>_c$, because for <k>> $<k>_c$, the number of distinct maximum matchings increases exponentially and, as a result, the chance that a link does not participate in any control configuration decreases.

Result 4

- The relations between robustness of control and <k>

Questions

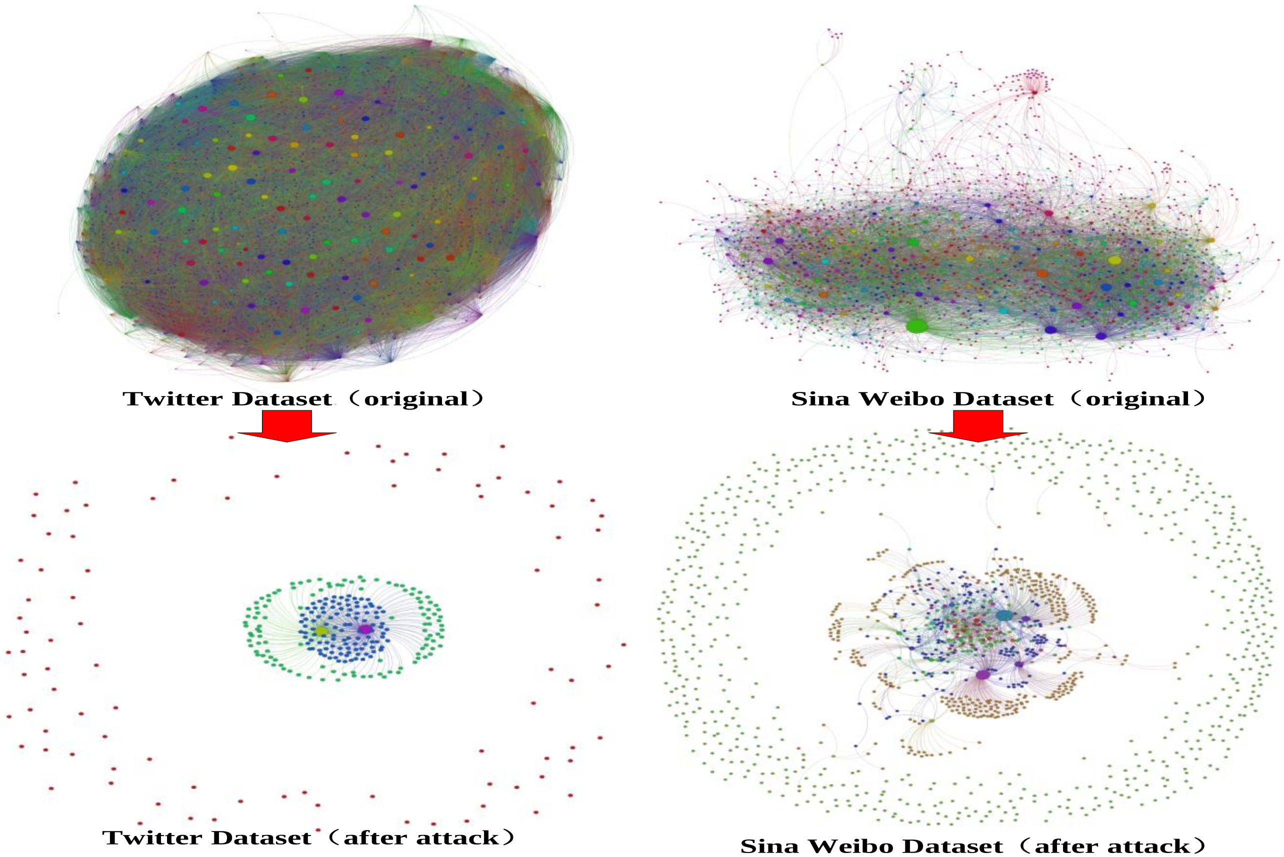
- What is the minimum number of driver nodes (N_D) of real-world networks? (maximum matching)
- How to locate them efficiently? (avoid hubs)
- Which topological characteristics determine N_D? (degree distribution, $P(k_{in},k_{out})$)
- How robust of network controllability?

Conclusion

- Developed the tools to address controllability for arbitrary network topologies and sizes.
- Our key finding, that N_D is determined mainly by the degree distribution.
- Use the tools of statistical physics to predict N_D from $P(k_{in},k_{out})$ analytically, offering a general formalism with which to explore the impact of network topology on controllability.

Comment

- Why driver nodes avoid hubs?
- Network topology affect N_D



Thank you

&

Q.A.

&

Discussion