Complex Networks

Albert-László Barabási

Center for Complex Networks Research and Department of Physics

Northeastern University, Boston

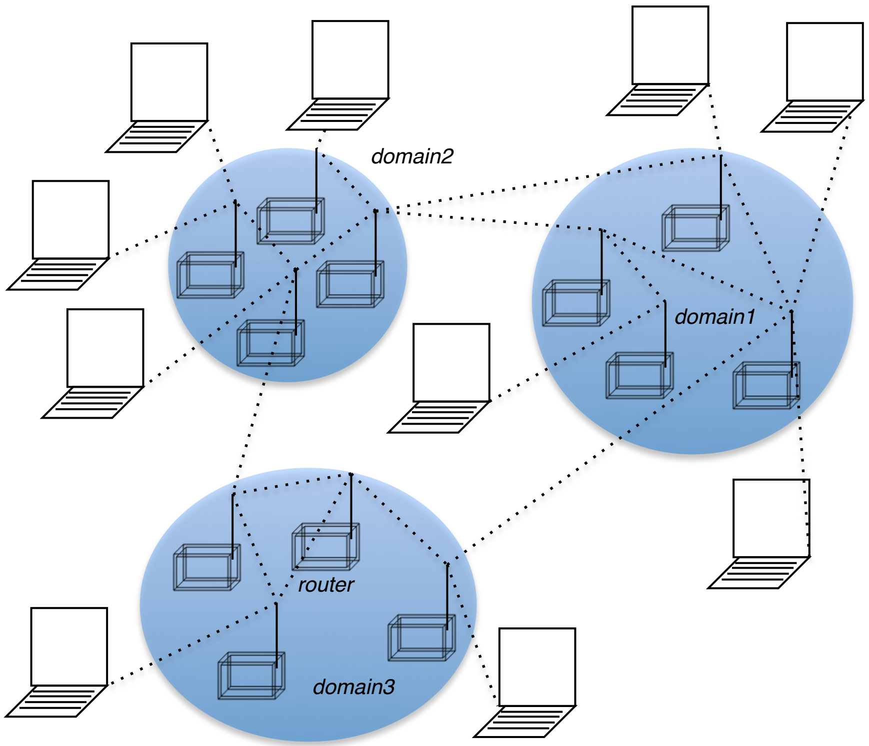
Central European University, Budapest

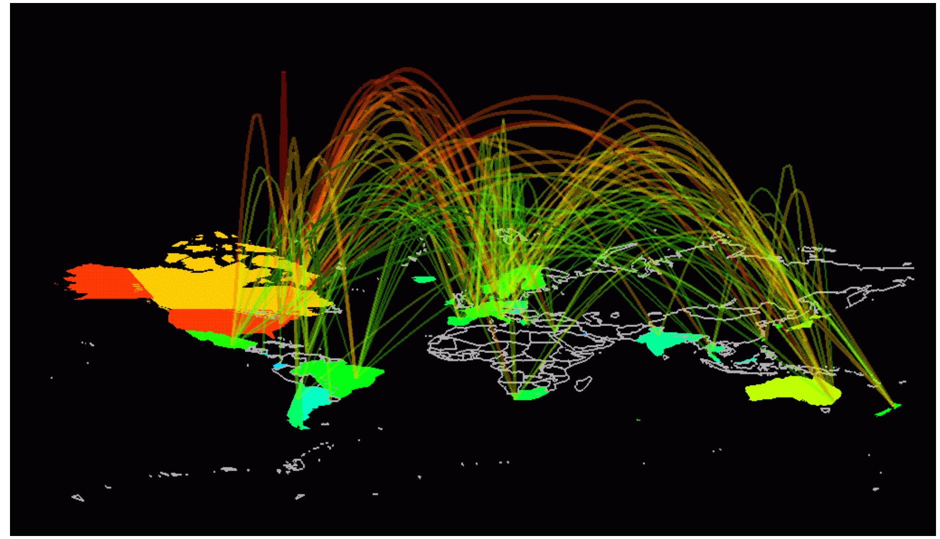
Division of Network Medicine

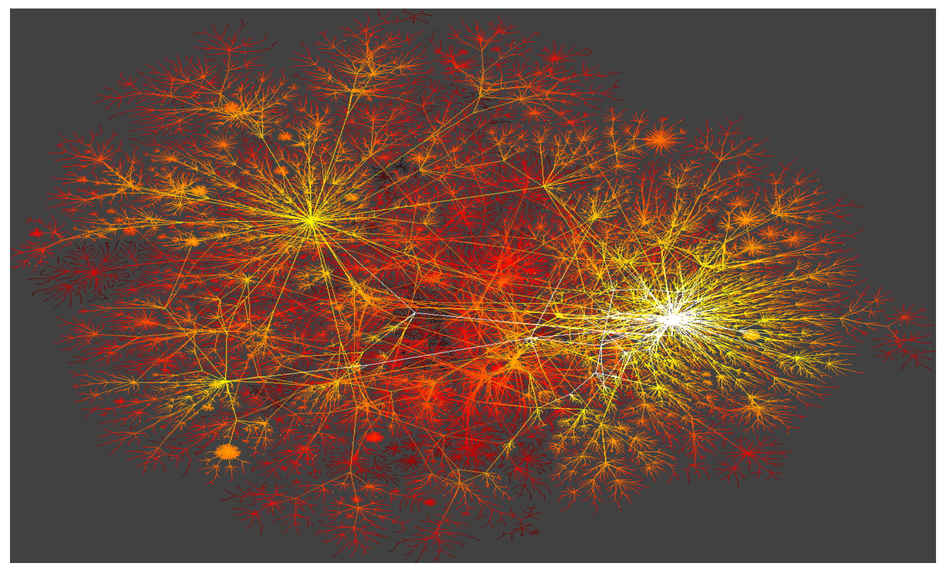
Harvard Medical School

www.BarabasiLab.com

INTERNET

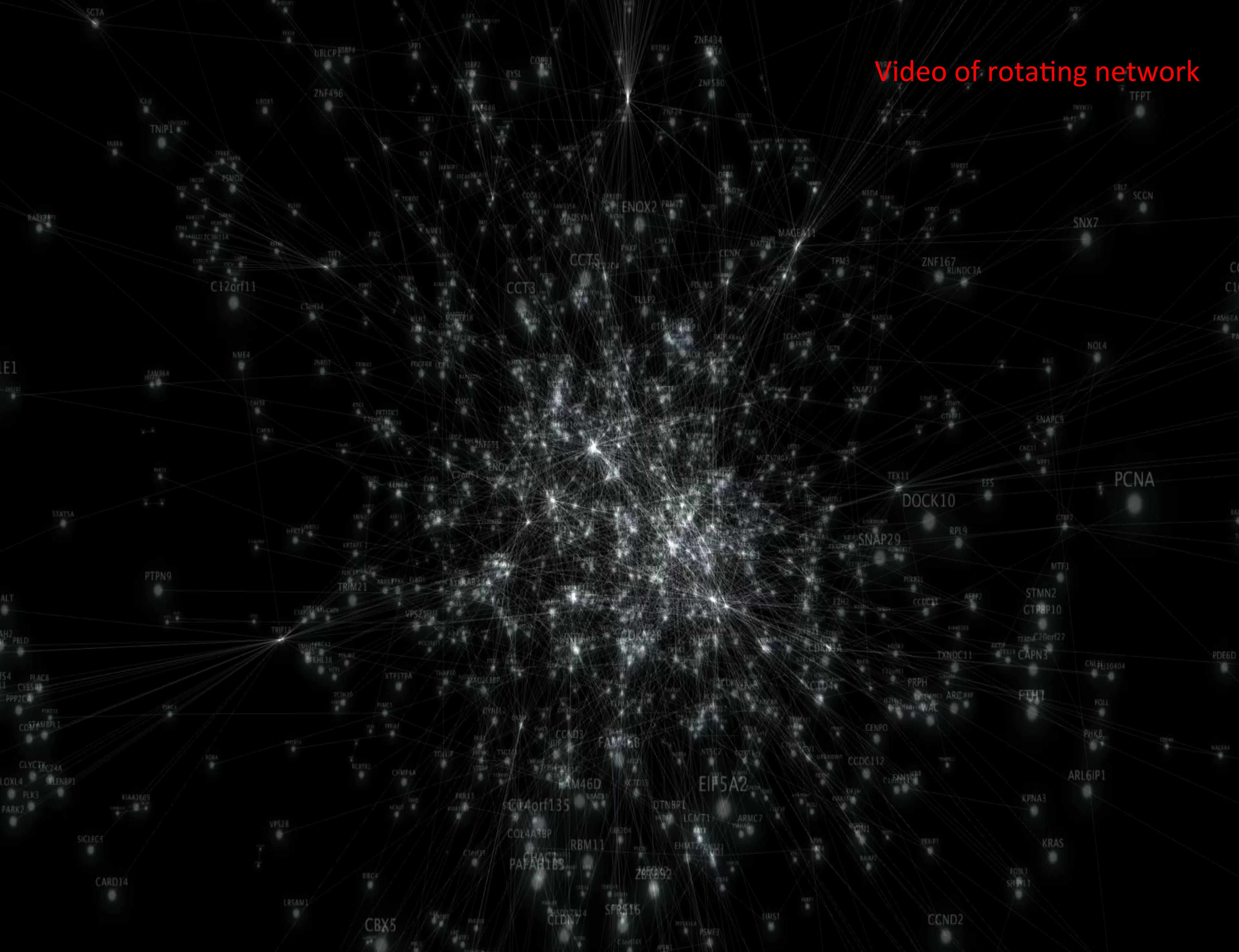






Facebook: The Social Graph

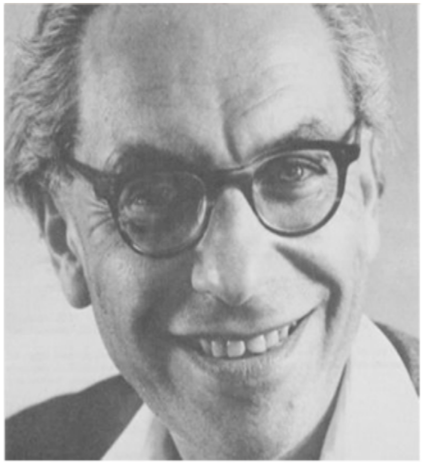
Keith Shepherd's "Sunday Best". http://baseballart.com/2010/07/shades-of-greatness-a-story-that-needed-to-be-told/



Video of rotating network
SCTA
ZNF434
UBLCP1
ZNF496
TNIP1
ZNF580
TFPT
ENOX2
MAGEA11
SCGN
SNX7
C12orf11
CCT5
CCT3
TULP2
ZNF167
RUNDC3A
NOL4
SNAPC5
PCNA
DOCK10
TEX11
SNAP29
RPL9
MTF1
STMN2
GTPBP10
C20orf27
CAPN3
TXNDC11
PDE6D
PTPN9
TRIM21
TRIP13
STAT5A
CDKN2B
CDKN1A
PRPH
FTH1
ARC
WAC
PHKB
ARL6IP1
KPNA3
KRAS
CENPO
CCDC112
EIF5A2
FAM46B
FAM46D
CCND3
C14orf135
COL4A3BP
DTNBP1
LGMT1
ARMC7
RBM11
PAFAH1B3
ZBTB32
SFRS16
CLDN7
CBX5
CCND2
CARD14
SIGLEC5
VPS28
LRSAM1
PARK2
PLK3
LOXL4
GLYCTK
STAMBPL1
PLAC8
PBLD
NME4

RANDOM NETWORK MODEL

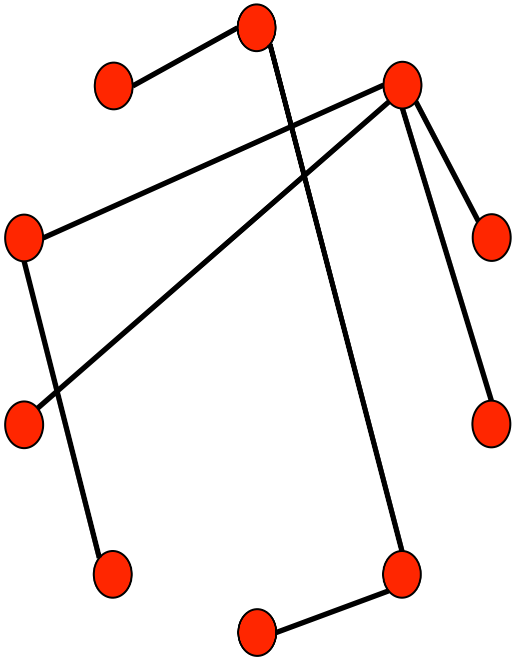
Pál Erdös
(1913-1996)



Alfréd Rényi
(1921-1970)

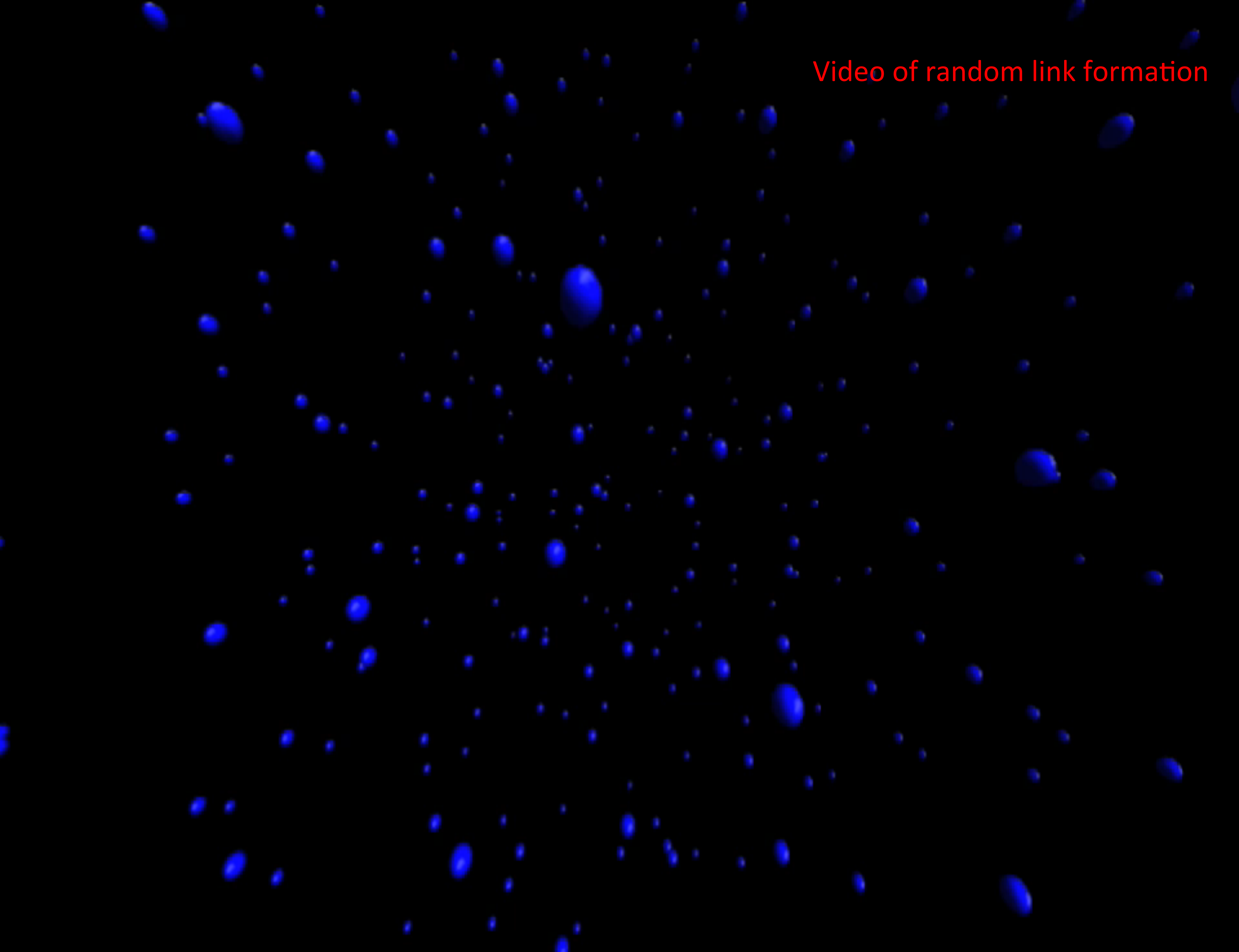
Erdös-Rényi model (1960)

Connect with probability p



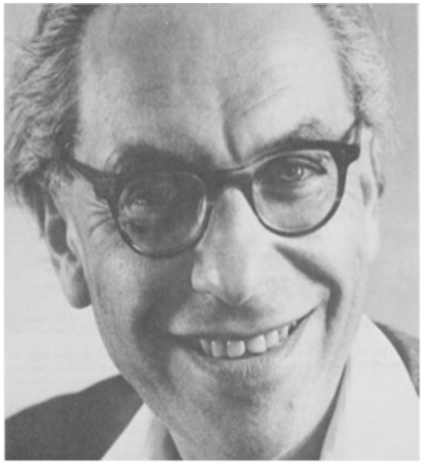
p=1/6 N=10

$\langle k \rangle \sim 1.5$


Video of random link formation

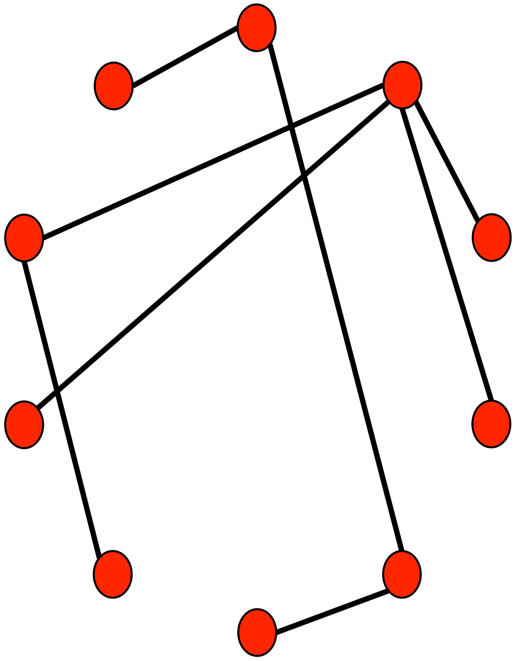
RANDOM NETWORK MODEL

Pál Erdös
(1913-1996)



Alfréd Rényi
(1921-1970)

Erdös-Rényi model (1960)

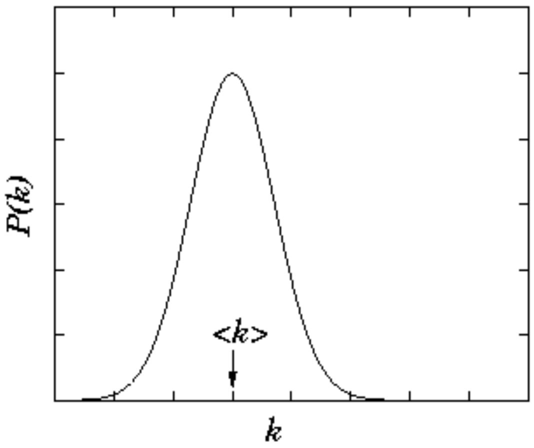


Connect with probability p

p=1/6 N=10

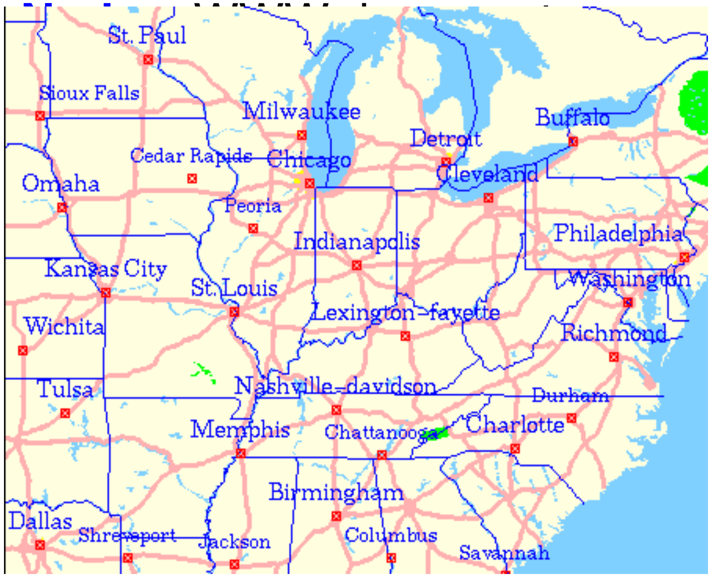
$\langle k \rangle \sim 1.5$

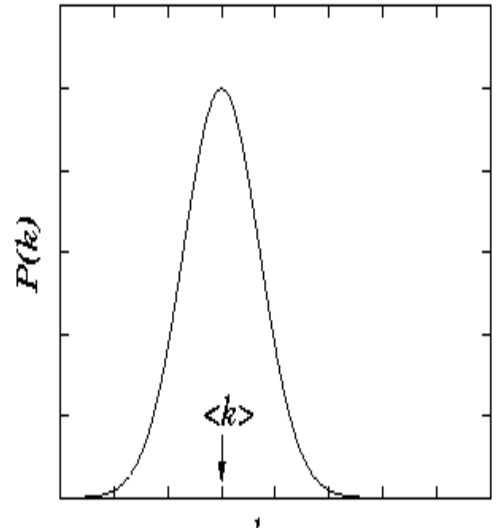
Degree distribution



WORLD WIDE WEB

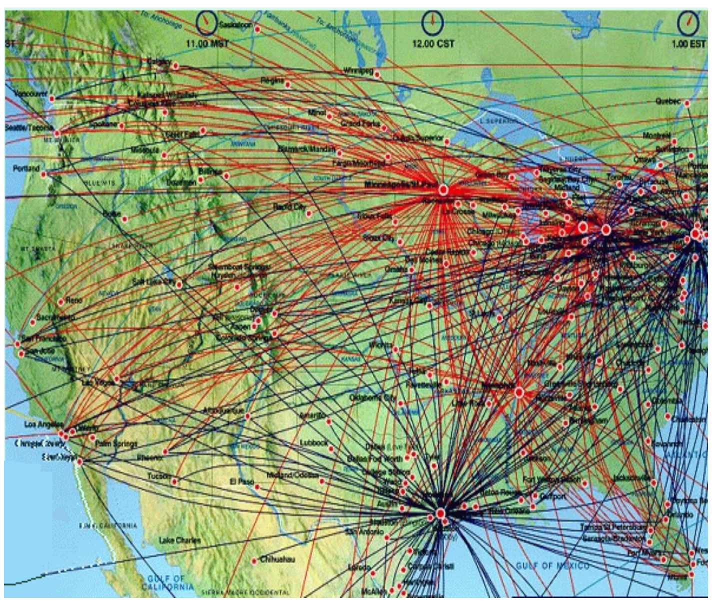
Random Network

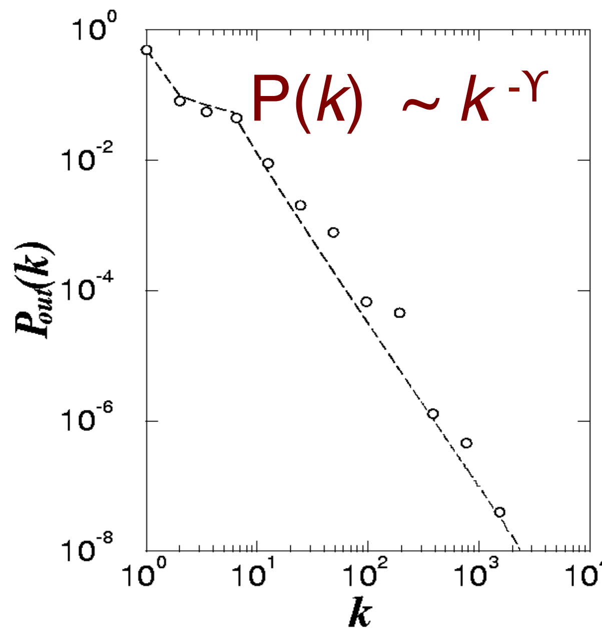




Expected

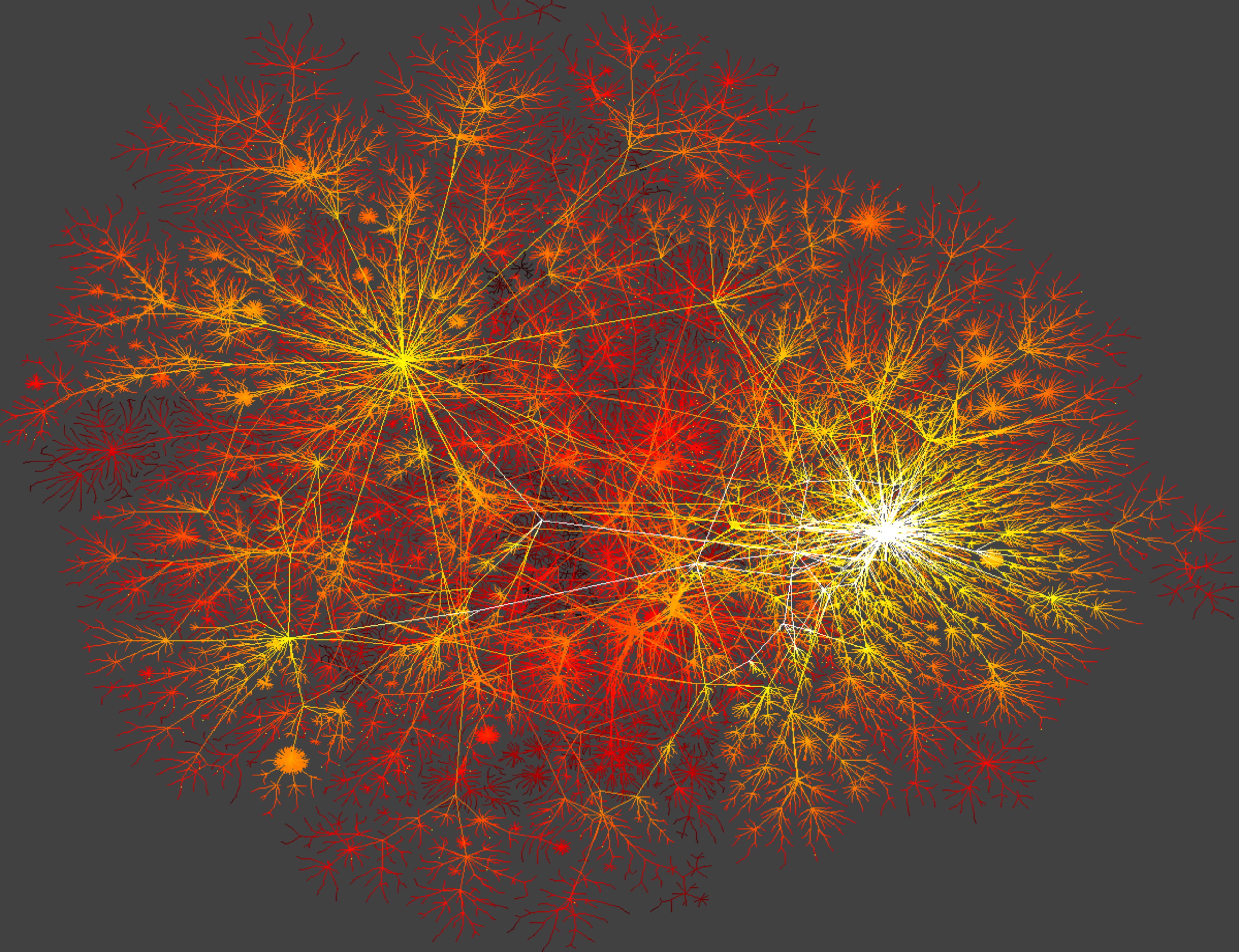
Scale-free Network





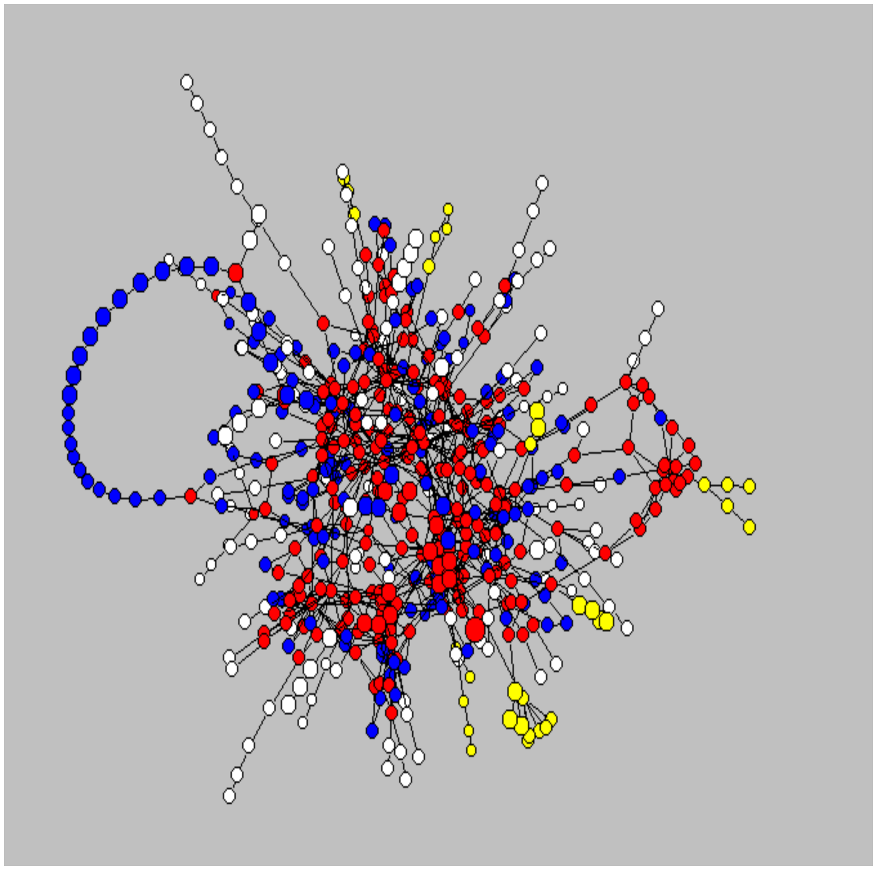
Found

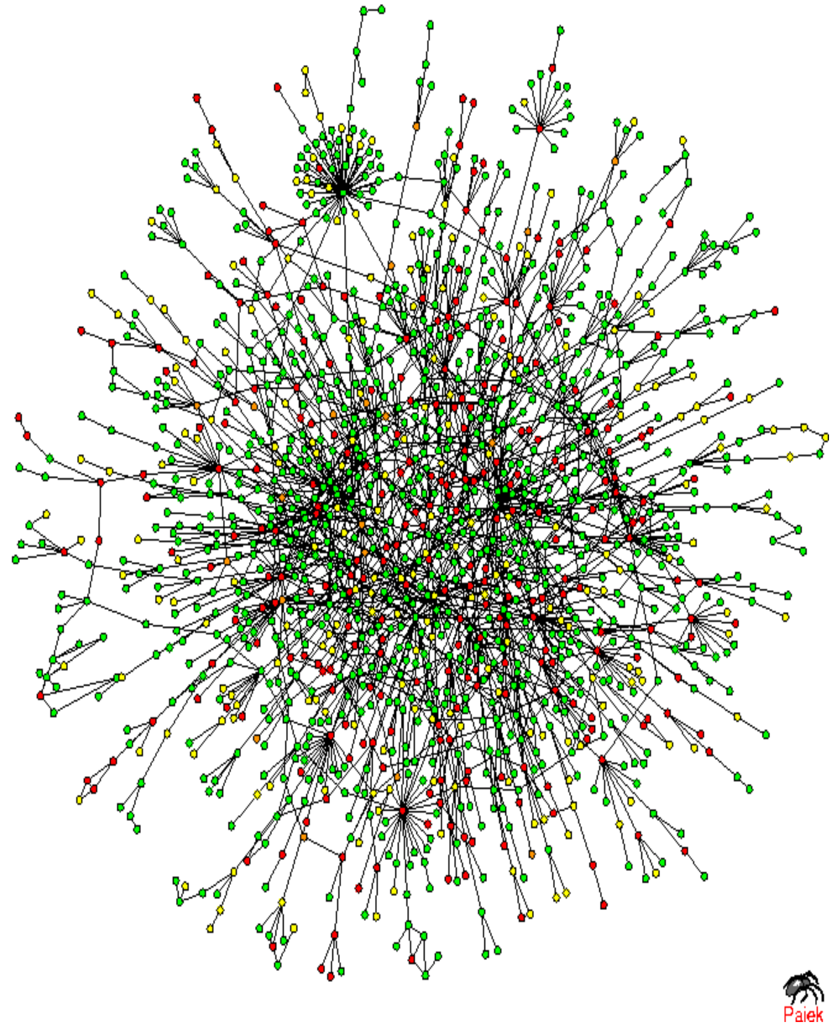
R. Albert, H. Jeong, A-L Barabási, *Nature*, 401 130 (1999).



METABOLIC NETWORK

PROTEIN INTERACTIONS





Jeong, Albert, Oltvai, and Barabási, *Nature* 407: 651 (2000); *Nature* 411: 41 (2001).

MANY REAL WORLD NETWORKS HAVE A SIMILAR ARCHITECTURE:

Scale-free networks

WWW, Internet (routers and domains), electronic circuits, computer software, movie actors, coauthorship networks, sexual web, instant messaging, email web, citations, phone calls, metabolic, protein interaction, protein domains, brain function web, linguistic networks, comic book characters, international trade, bank system, encryption trust net, energy landscapes, earthquakes, astrophysical network…

ORIGIN OF SF NETWORKS Growth and preferential attachment

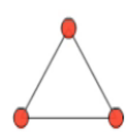
(1) Networks continuously expand by the addition of new nodes

WWW : addition of new documents

(2) New nodes prefer to link to highly connected nodes.

WWW : linking to well known sites

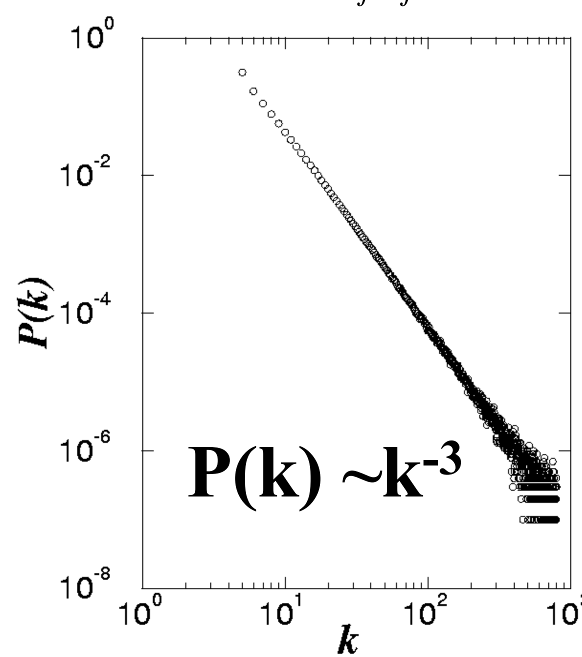
Video of preferential attachment



GROWTH:
add a new node with m links

PREFERENTIAL ATTACHMENT:
the probability that a node connects to a node with *k* links is proportional to *k*.

$$\Pi(k_i) = \frac{k_i}{\Sigma_j k_j}$$

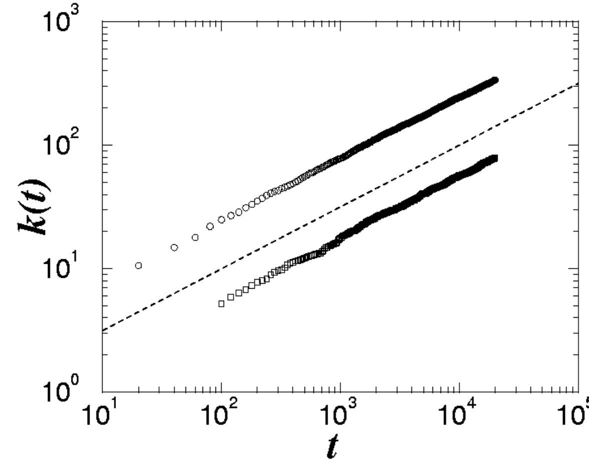


Barabási & Albert, *Science* **286,** 509 (1999)

CONTINUUM THEORY

$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i) = A\frac{k_i}{\sum_j k_j} = \frac{k_i}{2t}$$, with initial condition $$k_i(t_i) = m$$

$$k_i(t) = m\sqrt{\frac{t}{t_i}}$$



$$P(k_i(t) < k) = P_t(t_i > \frac{m^2 t}{k^2}) = 1 - P_t(t_i \le \frac{m^2 t}{k^2}) = 1 - \frac{m^2 t}{k^2(m_0 + t)}$$

$$\therefore P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^2 t}{m_o + t}\frac{1}{k^3} \sim k^{-3}$$

$$\gamma = 3$$

Barabási, Albert & Jeong, *Physica* A 272, 173 (1999).
Bollobás *et al. Rand. Structures & Algorithms* (2001).

DEGREE DISTRIBUTION

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}$$

$$P(k) \sim k^{-3}$$ for large *k*

$$\gamma = 3$$

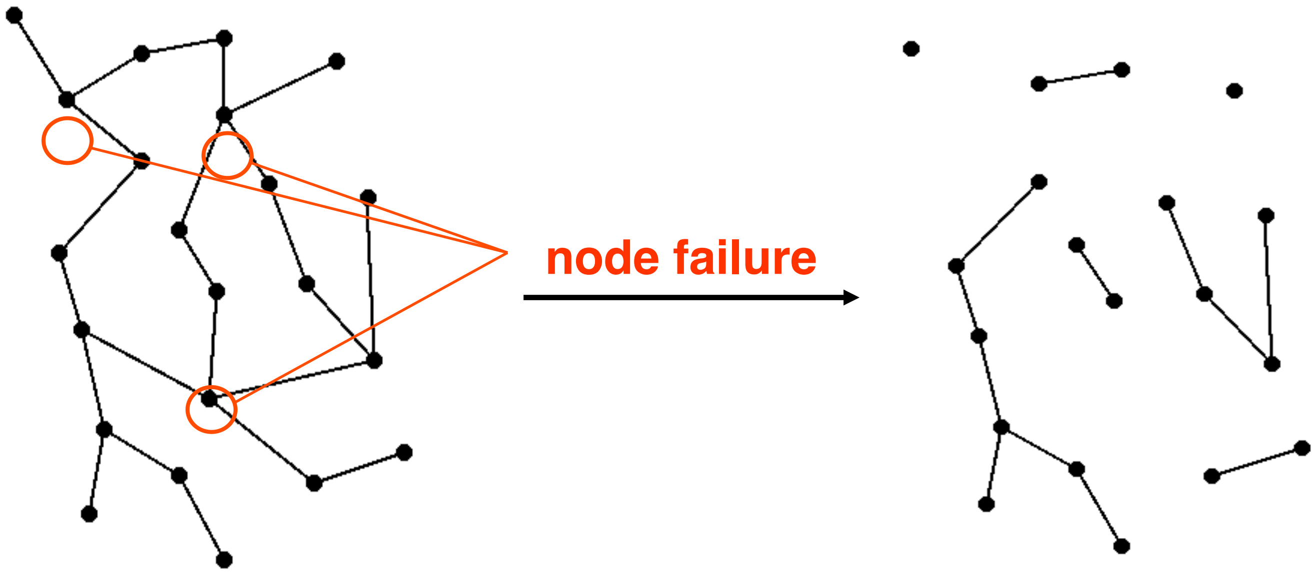
(i) The degree exponent is independent of *m*.

(ii) The network reaches a stationary scale-free state.

(iii) The coefficient of the power-law distribution is proportional to m^2.

Krapivsky, Redner, Leyvraz, *Phys. Rev. Lett*. 2000; Dorogovtsev, Mendes, Samukhin, *Phys. Rev. Lett.* 2000; Bollobas et al. *Random Struc. Alg*. 2001

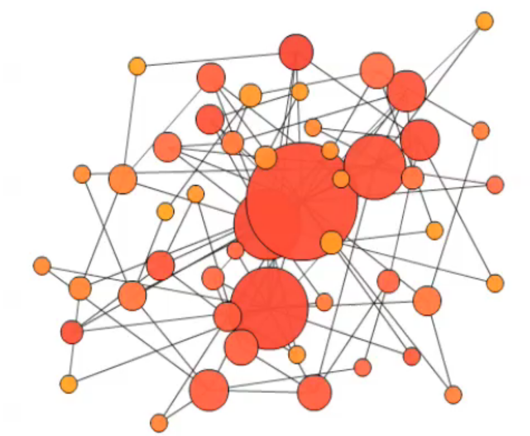
ROBUSTNESS OF SCALE-FREE NETWORKS

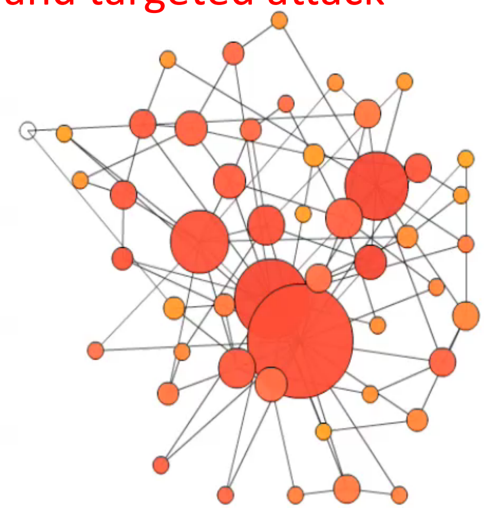


Failures

Two videos of networks under random and targeted attack

Attacks





Albert, Jeong, Barabási, *Nature* **406** 378 (2000)

ROBUSTNESS OF SCALE-FREE NETWORKS

$$f_c = 1 - \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$$

$$\frac{<k^2>}{<k>} = \left|\frac{2-\gamma}{3-\gamma}\right| K_{\min} \begin{cases} 1 & \gamma > 3 \\ N^{\frac{3-\gamma}{\gamma-1}} & 3 > \gamma > 2 \\ N^{\frac{1}{\gamma-1}} & 2 > \gamma > 1 \end{cases}$$

γ>3: $<k^2>$ is finite; the network will break apart at a finite f_c

γ<3: $<k^2>$ diverges in the $N \rightarrow \infty$ limit, so $f_c \rightarrow 1$
we need to remove all the nodes to break the system.

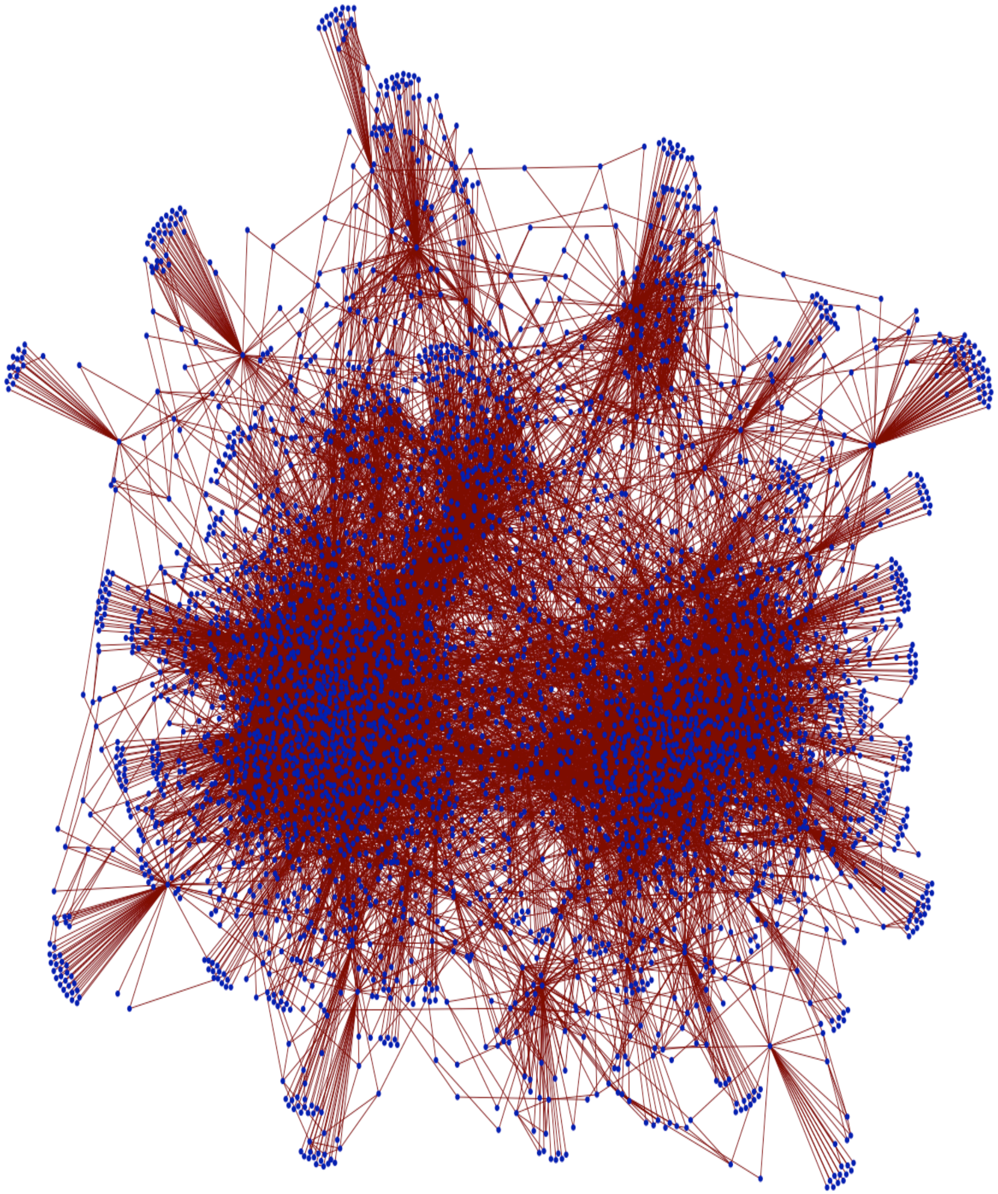
Finite systems:

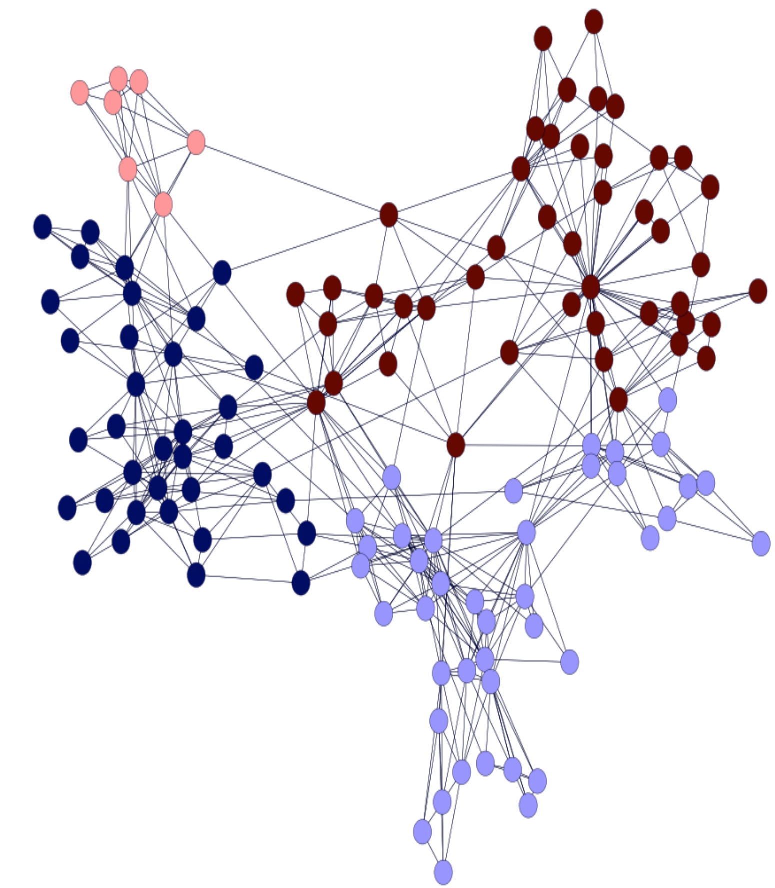
$$f_c \cong 1 - CN^{-\frac{3-\gamma}{\gamma-1}}$$

Internet: Router level map, N=228,263; γ=2.1±0.1; *κ=28* → $f_c=0.962$

Cohen & Havlin, *Physical Review Letters* (2000) (2001)

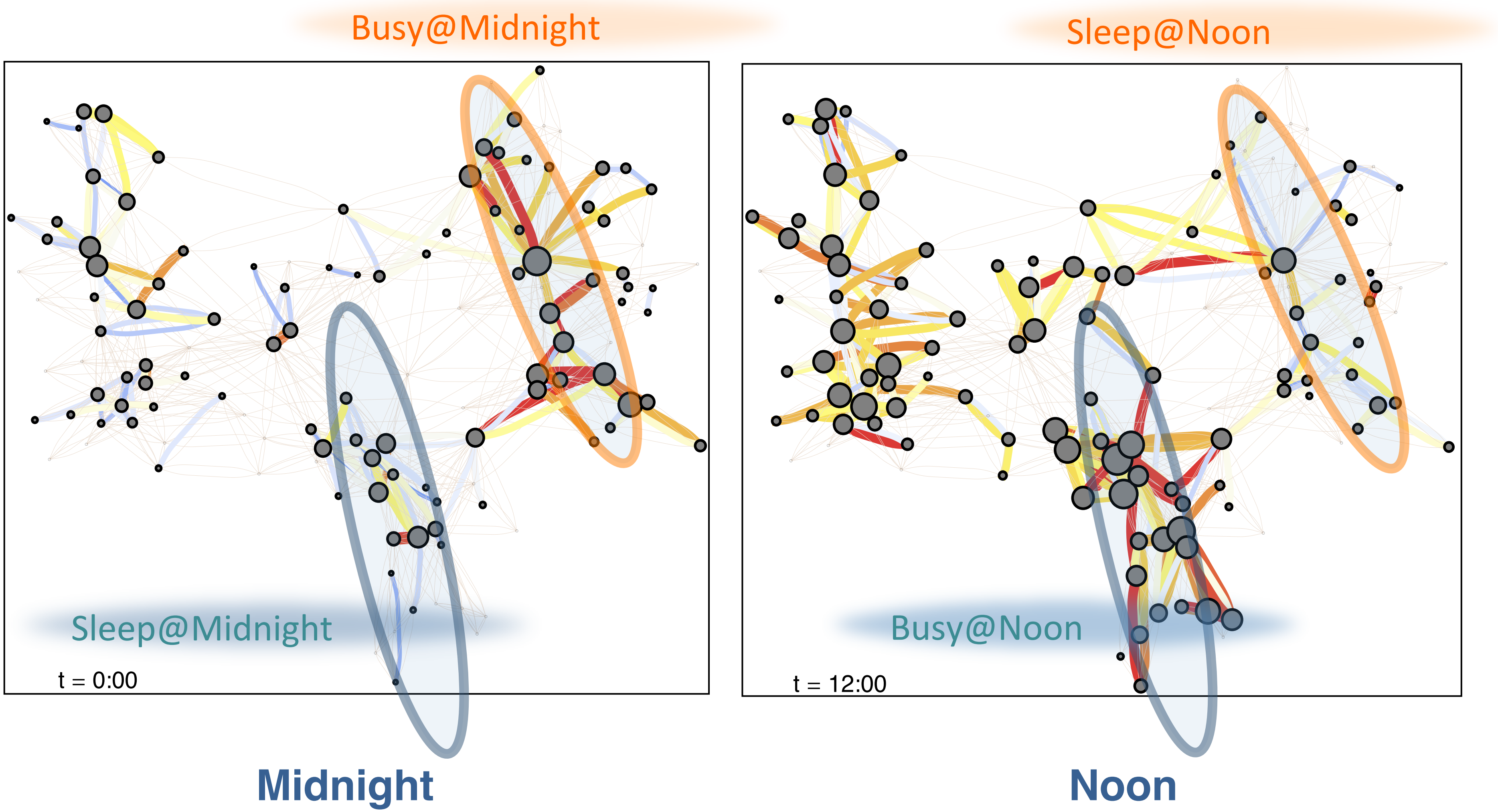
COMMUNITY STRUCTURE





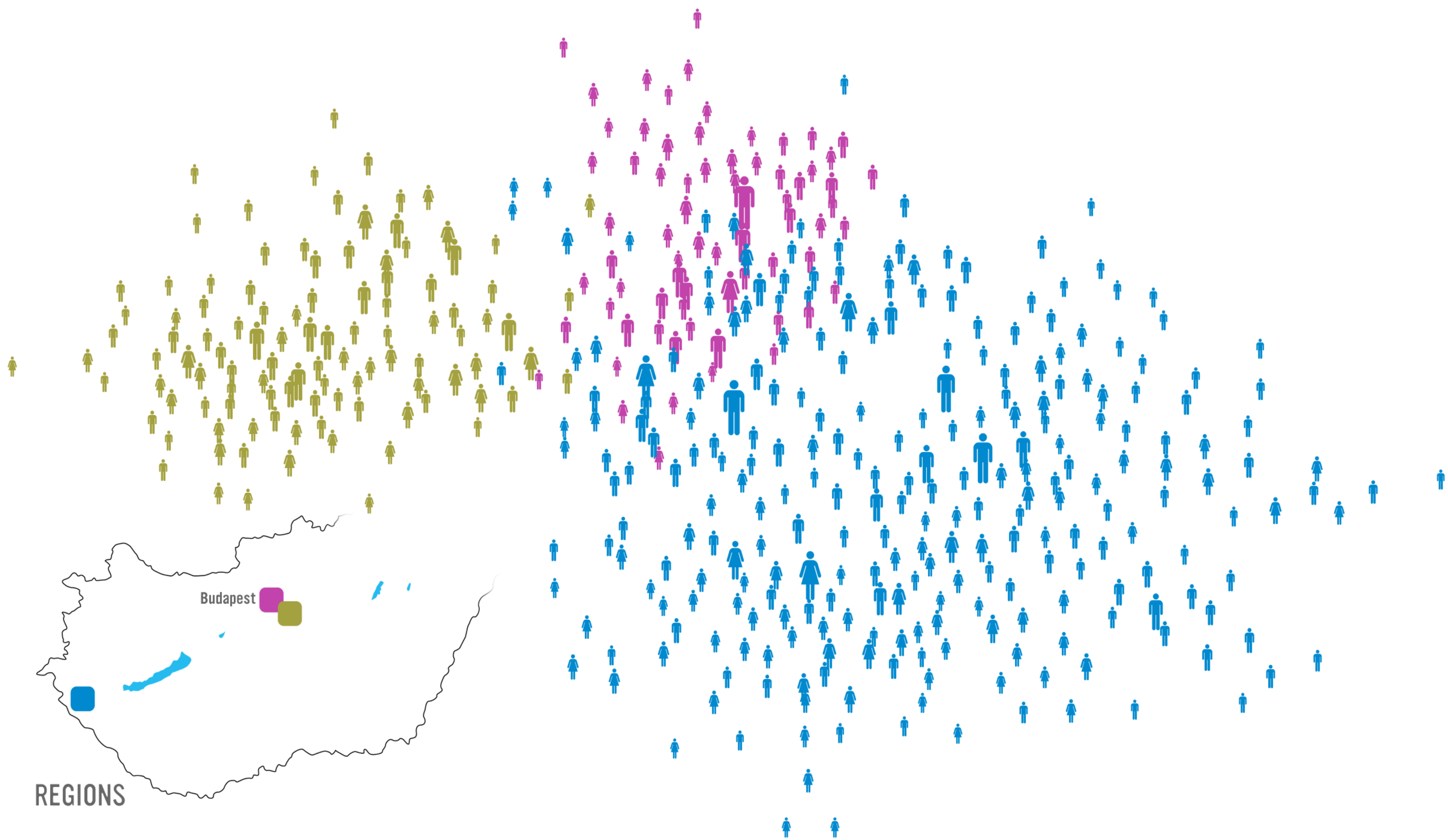
Y.-Y. Ahn, J. P. Bagrow, S. Lehmann, *Nature* (2010)

COMMUNITY STUCTURE Nodes in the same community are alike



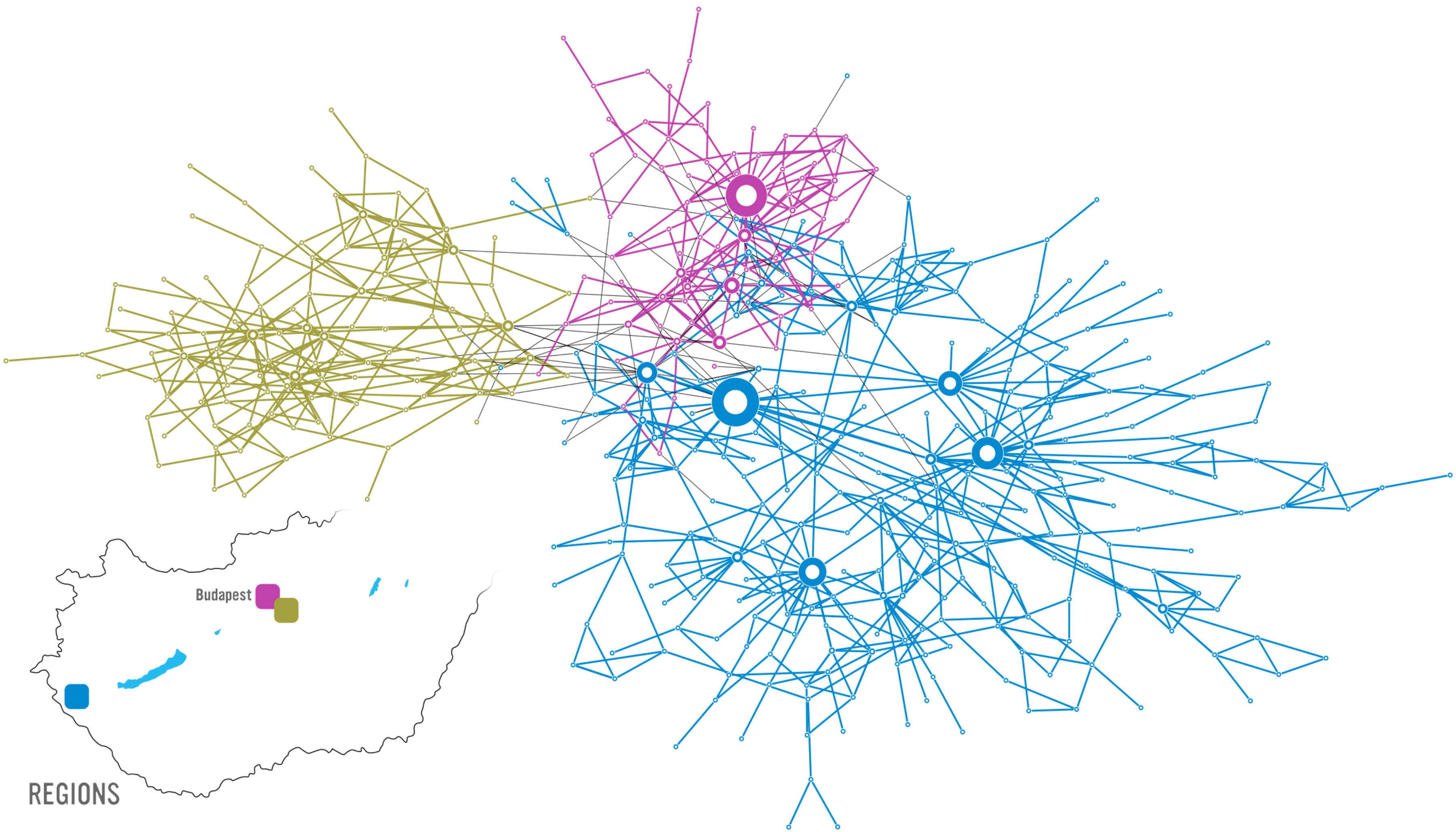
Ahn, Bagrow, Lehmann, *Nature* (2010)

THE POWER OF MAPS



maven7 connecting knowledge

Barabasi Lab



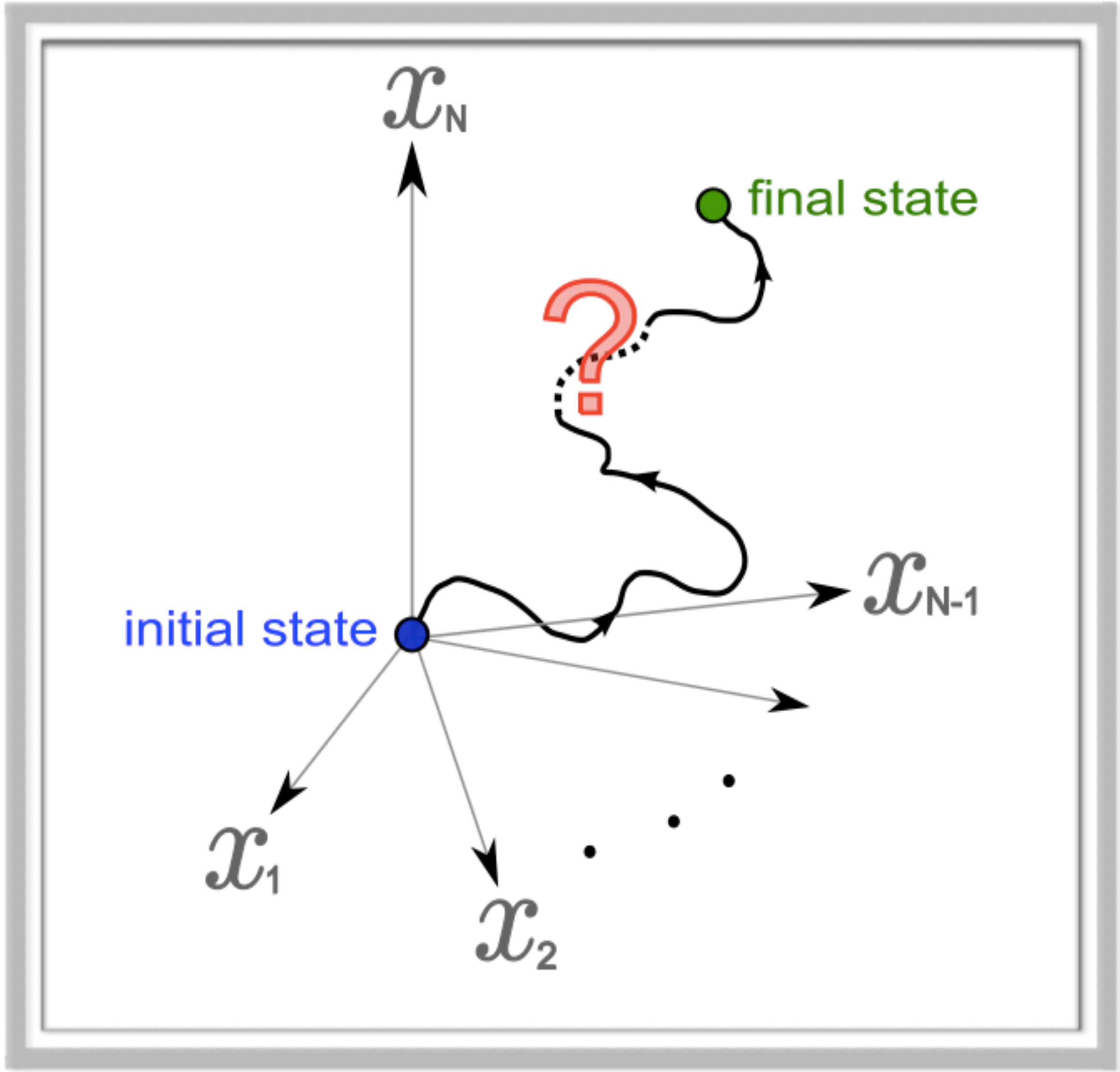
maven7
connecting knowledge

Barabasi Lab

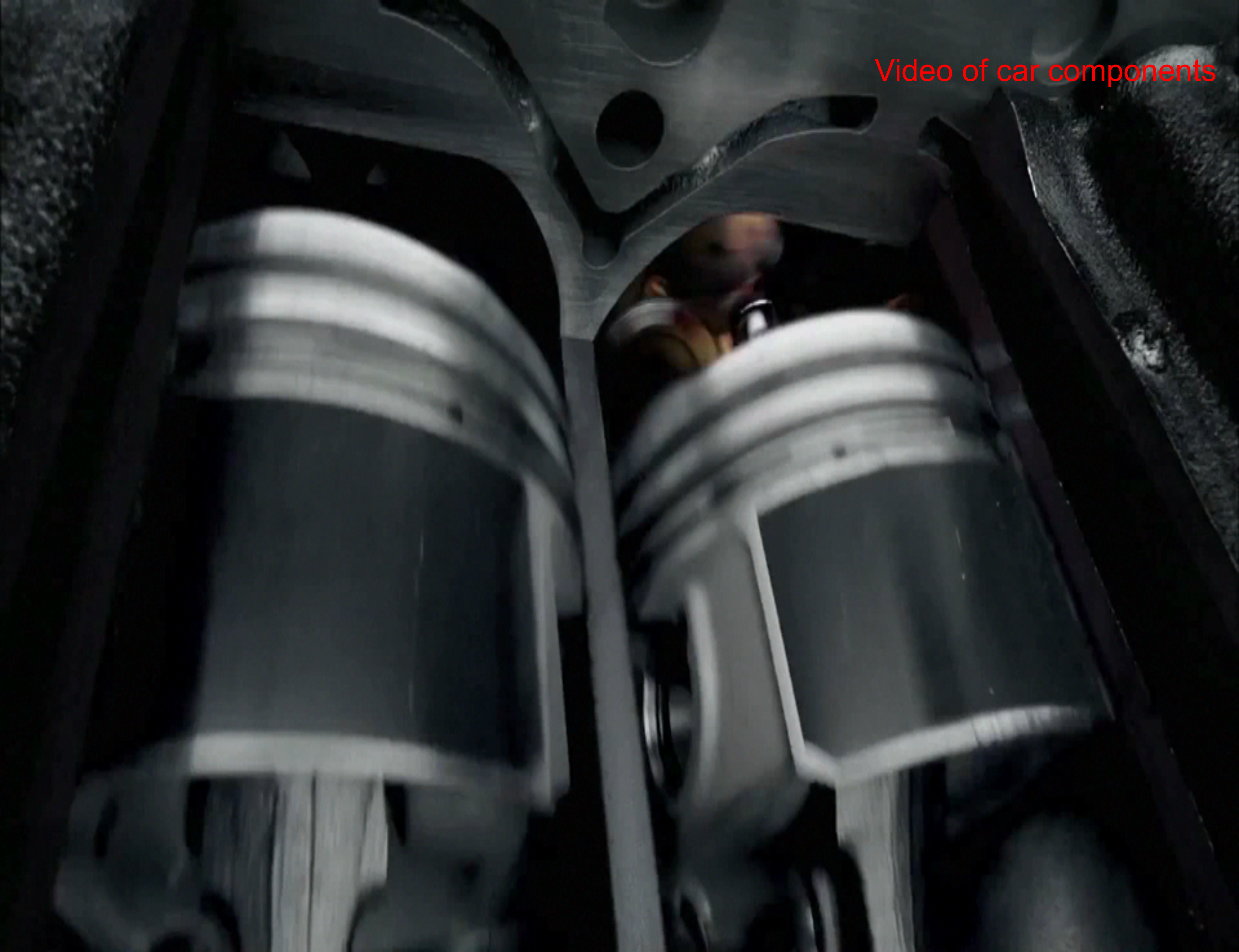

Directors/CEO
Top Managers
Managers
Group Leaders
Associates

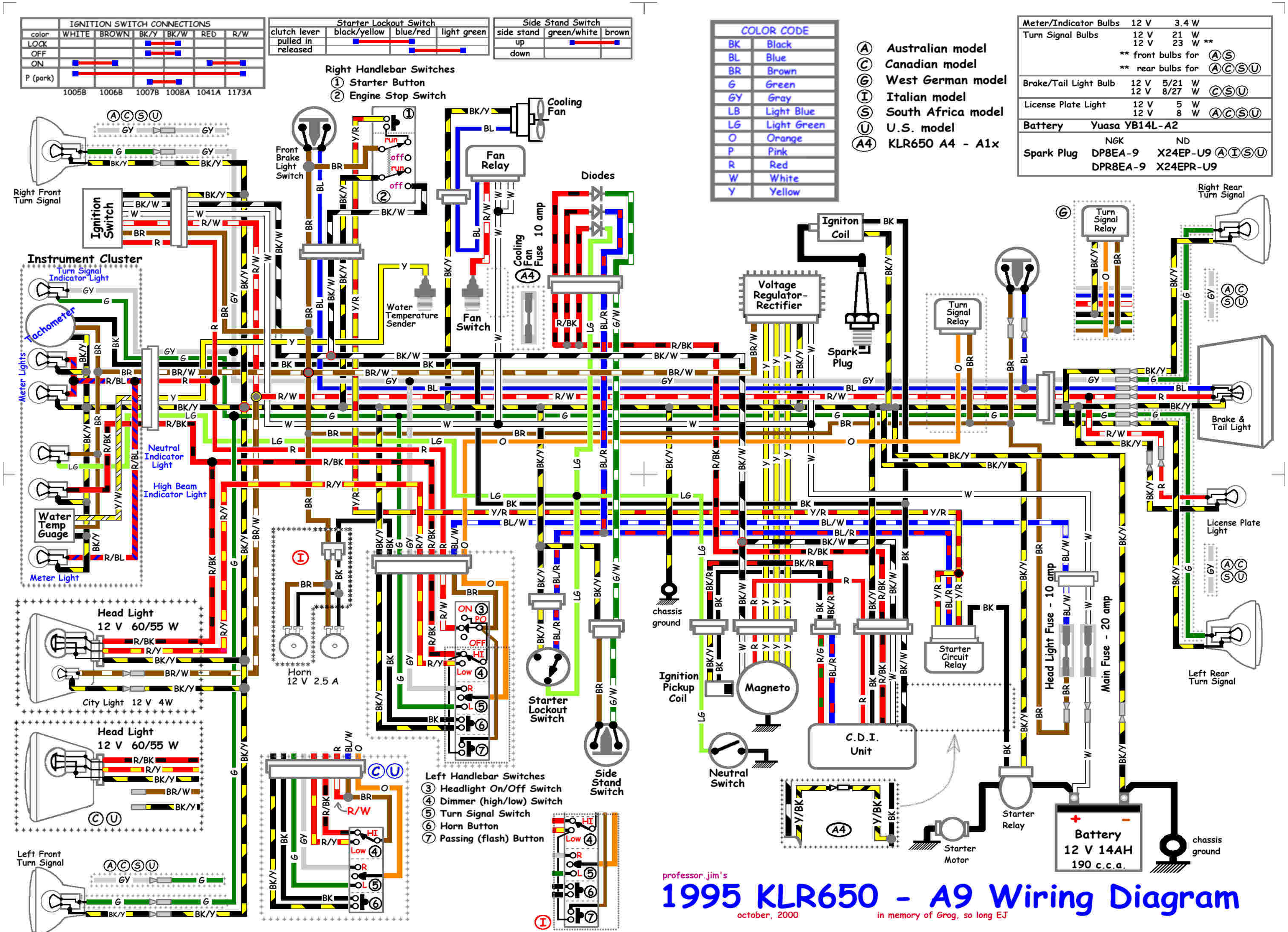
HIERARCHY


1
2
22 LINKS



A system is controllable if it can be driven from any *initial state* to any desired *final state* in finite time.


Video of car components

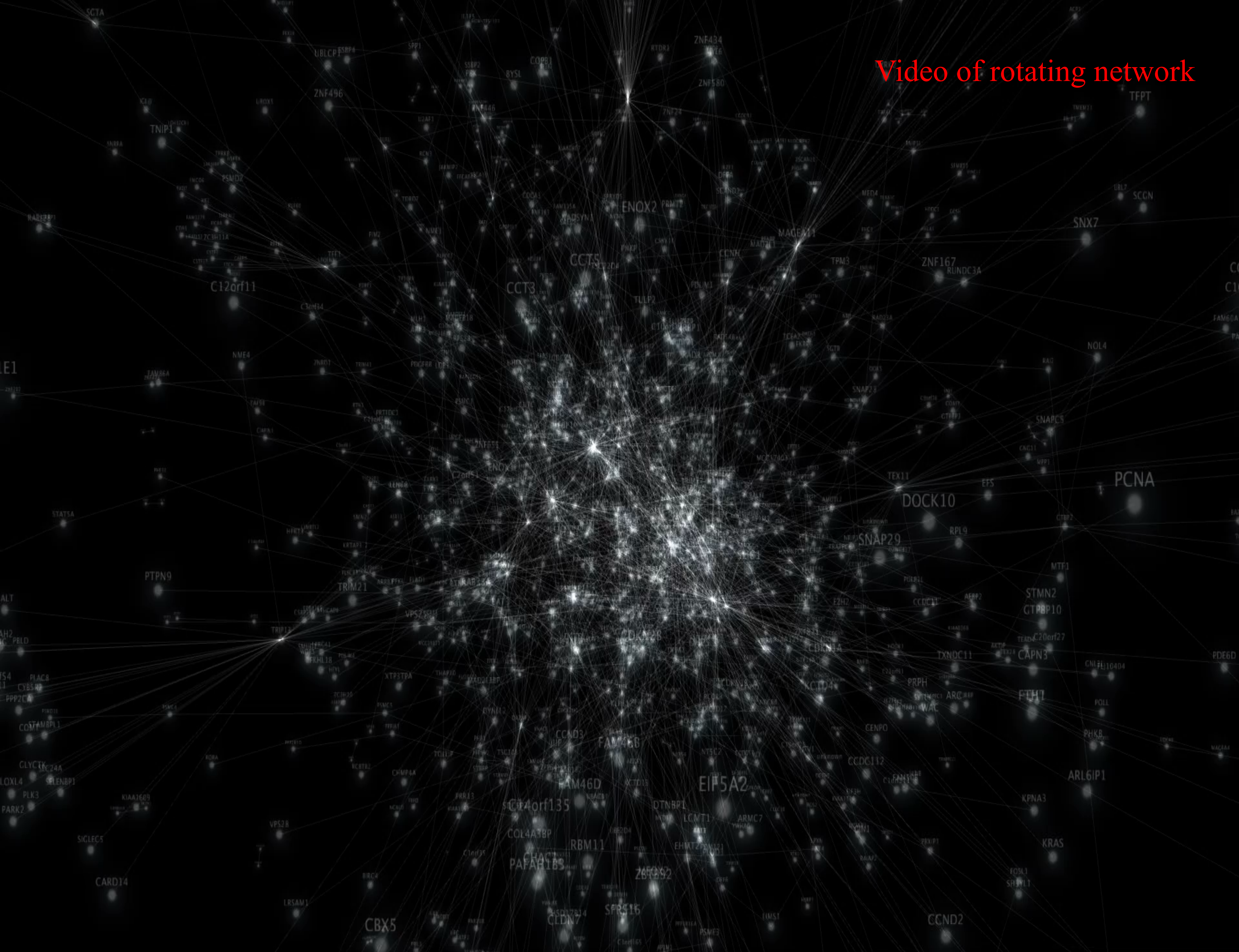


professor.jim's

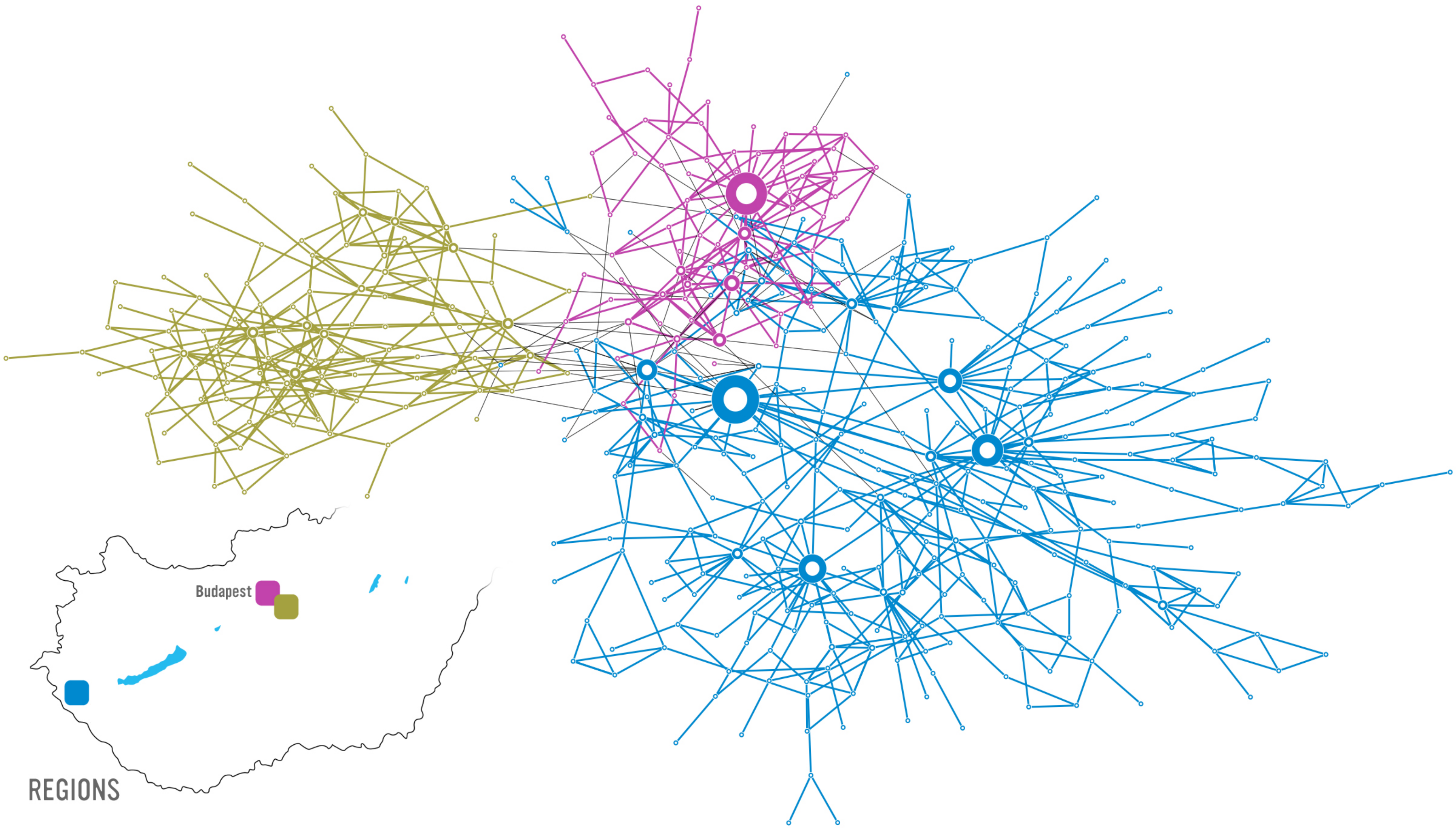
1995 KLR650 - A9 Wiring Diagram

october, 2000 in memory of Grog, so long EJ

http://static.ddmcdn.com/gif/mars1.jpg



Video of rotating network



maven7
connecting knowledge

ARTICLE

doi:10.1038/nature10011

Controllability of complex networks

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

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

LINEAR SYSTEMS   Kalman's Controllability Condition

- **Linear Time-Invariant Dynamics**

$$\frac{d\mathbf{X}}{dt} = \mathbf{A}\cdot\mathbf{X}(t) + \mathbf{B}\cdot\mathbf{u}(t)$$

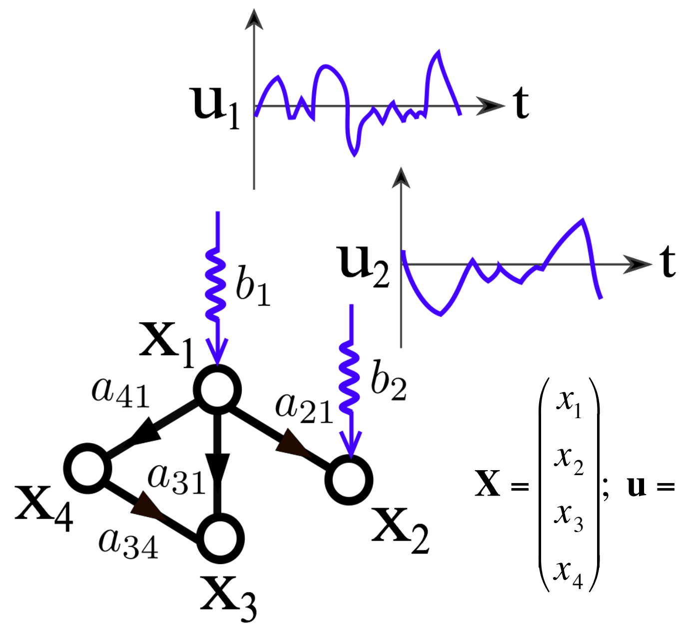
$\mathbf{A} \in R^{N\times N}$: weighted wiring diagram

$\mathbf{X}(t) \in R^{N\times 1}$: state vector.

$\mathbf{u}(t) \in R^{M\times 1}$: input vector ($M \le N$).

$\mathbf{B} \in R^{N\times M}$: input matrix

($\Rightarrow$ control configuration).



$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix};\ \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.$$

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ a_{21} & 0 & 0 & 0 \\ a_{31} & 0 & 0 & a_{34} \\ a_{41} & 0 & 0 & 0 \end{pmatrix};\ \mathbf{B} = \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

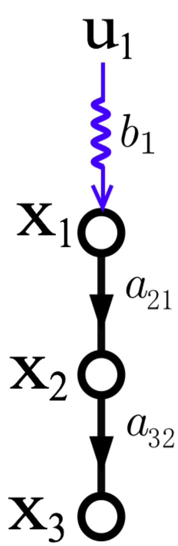
- **Kalman's Rank Condition:**
  A system is controllable iff its controllability matrix has full rank.

$$\text{rank}\ \mathbf{C} = N$$

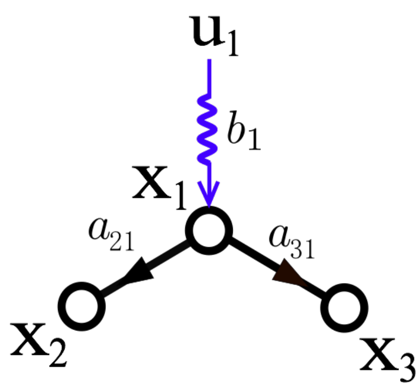
$$\mathbf{C} = [\mathbf{B}, \mathbf{A}\cdot\mathbf{B}, \mathbf{A}^2\cdot\mathbf{B}, \cdots, \mathbf{A}^{N-1}\cdot\mathbf{B}]$$

R. E. Kalman, *J.S.I.A.M. Control* (1963)

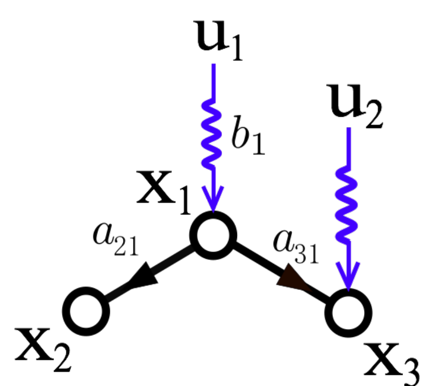
EXAMPLES: Controllable or not controllable?



Yes



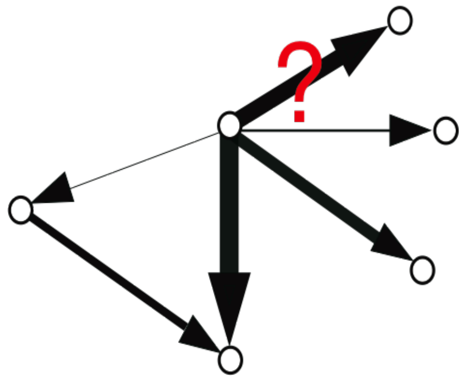
No



Yes

DIFFICULTIES

1. Parameters (link weights): usually unknown.
e.g. gene regulatory network, Internet, etc.



2. If brute-force search: (2^N-1) **combinations.**

$$\binom{N}{1}+\binom{N}{2}+\cdots+\binom{N}{N}=2^N-1$$

3. Kalman's rank condition is hard to check for large system.

$$\text{rank}\ \mathbf{C}=N$$

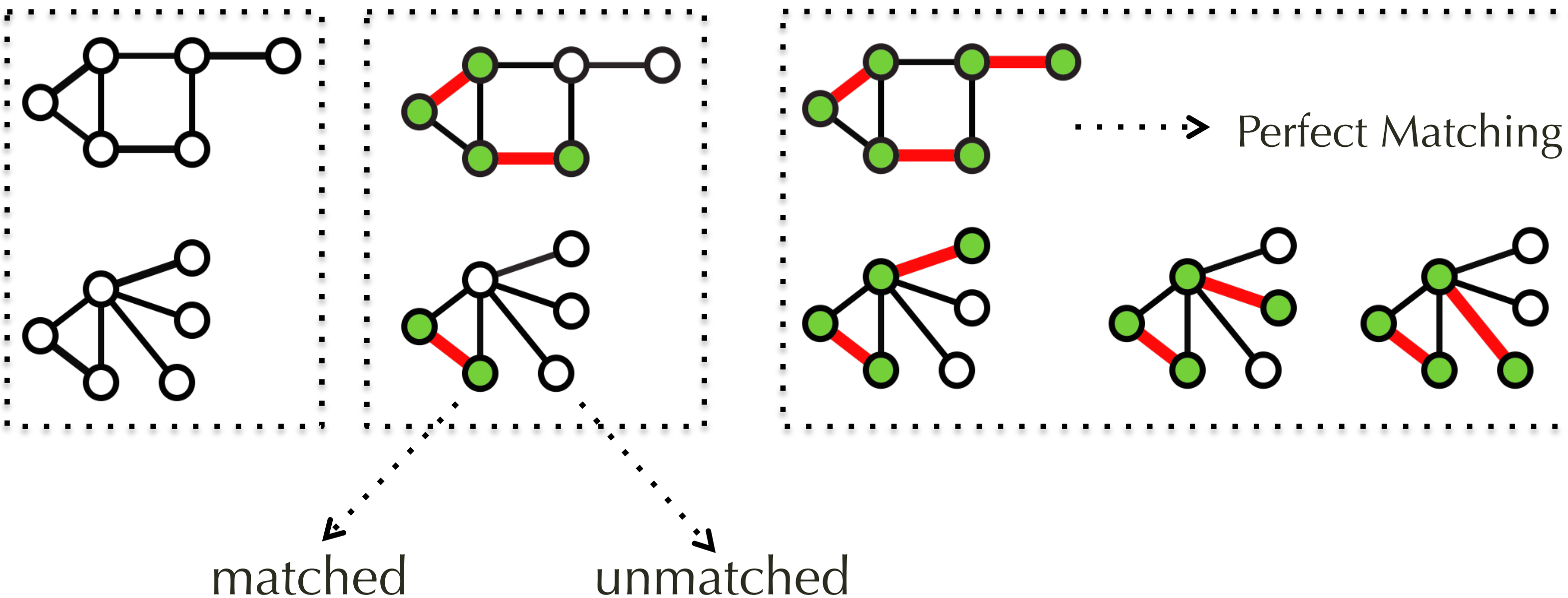
$$\mathbf{C}=[\mathbf{B},\ \mathbf{A}\cdot\mathbf{B},\ \mathbf{A}^2\cdot\mathbf{B},\cdots,\ \mathbf{A}^{N-1}\cdot\mathbf{B}]\ \text{ has dimension } N\times NM.$$

Matching

Network

Matching :
a set of edges without common vertices.

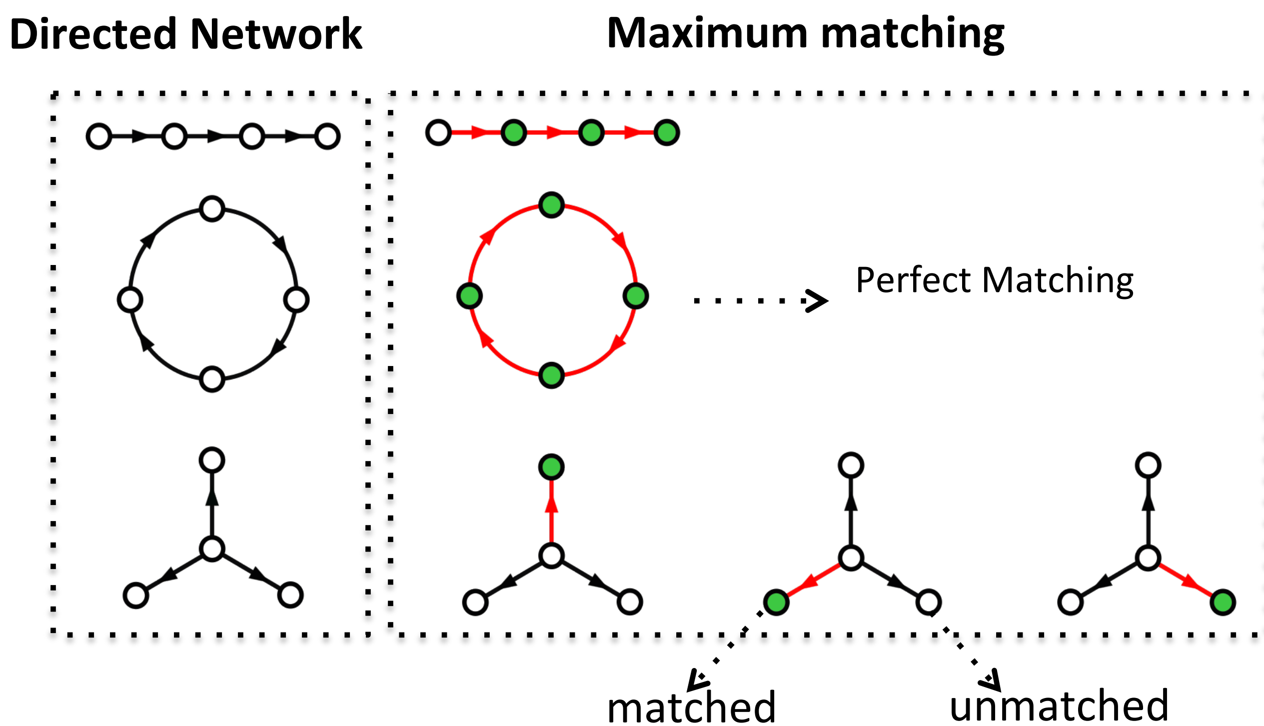
Maximum matching :
a matching of the largest size.



Lovász, L. & Plummer, M.D., *Matching Theory*

MATCHING IN DIRECTED NETWORKS:

Matching : a set of edges without common heads or tails.

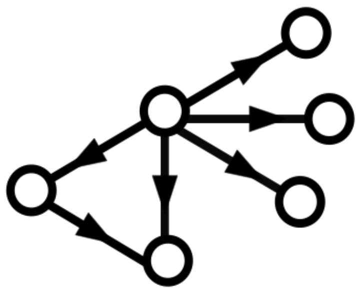


Minimum Input Theorem: Driver nodes = Unmatched nodes

Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási, *Nature* (2011)

EXAMPLES: Identifying the driver nodes

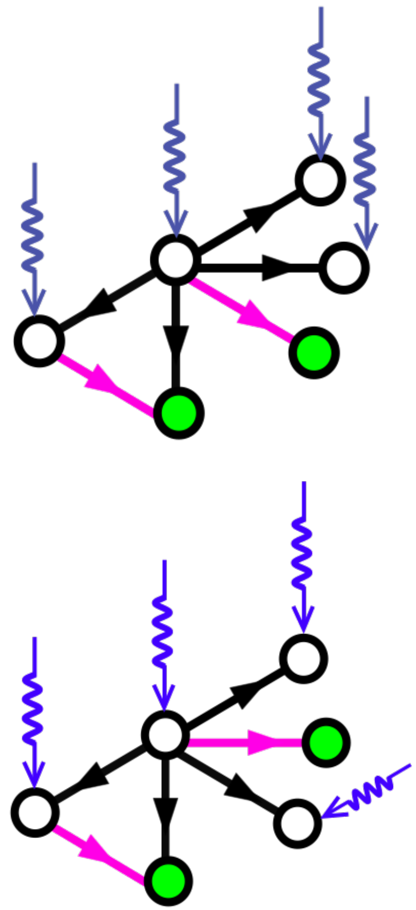
network



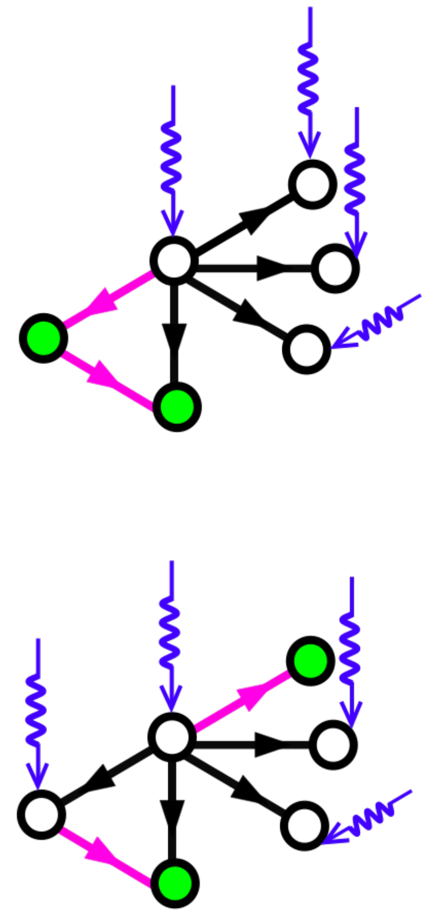
Brute-force search
$O(2^N)$~10^{30} for N=100.
Hopeless!

Hopcroft-Karp Algorithm
$O(N^{1/2}L)$ Polynomial!
Fast even for N~10^6.

Maximum matching

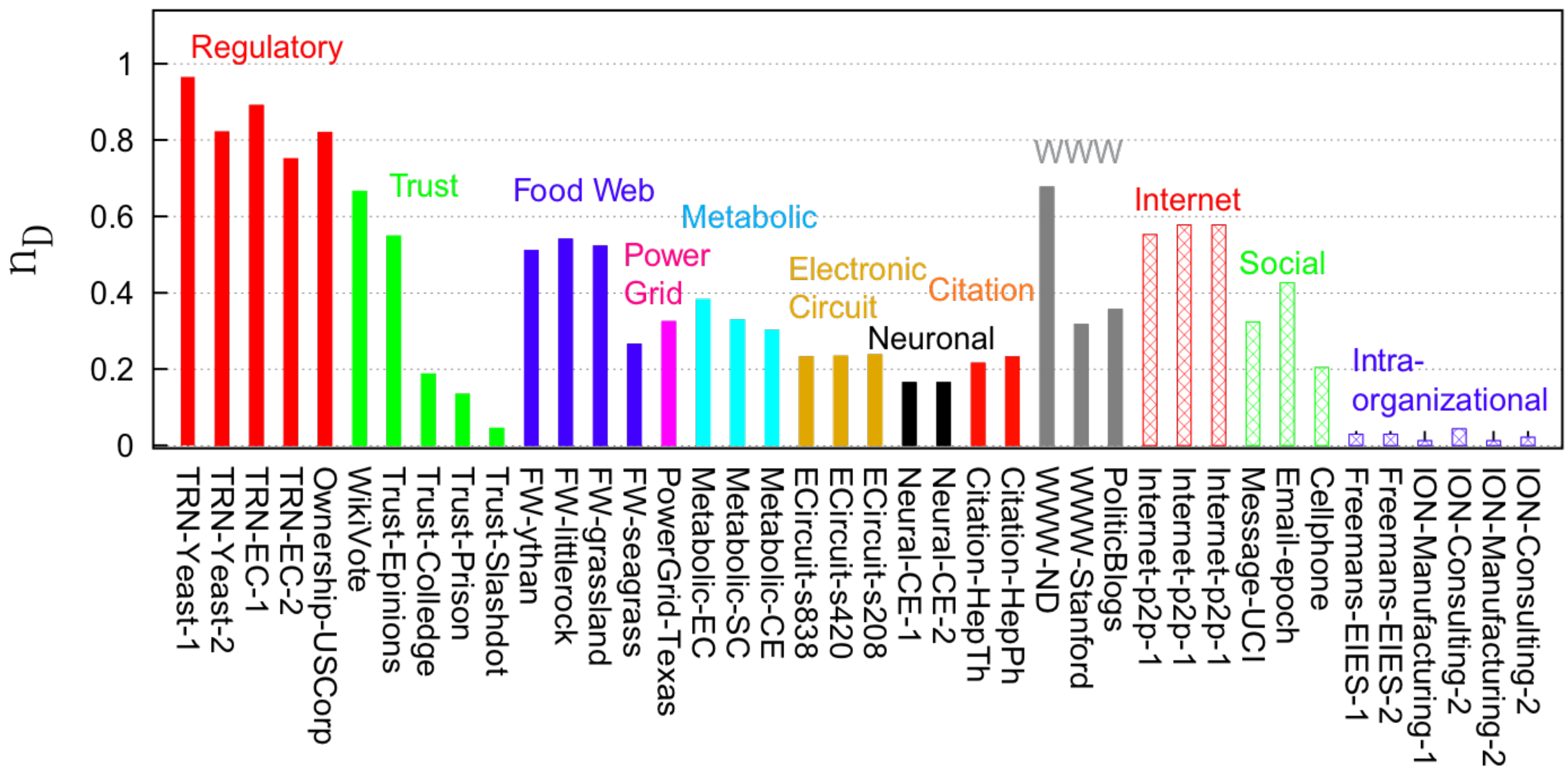


controlled network



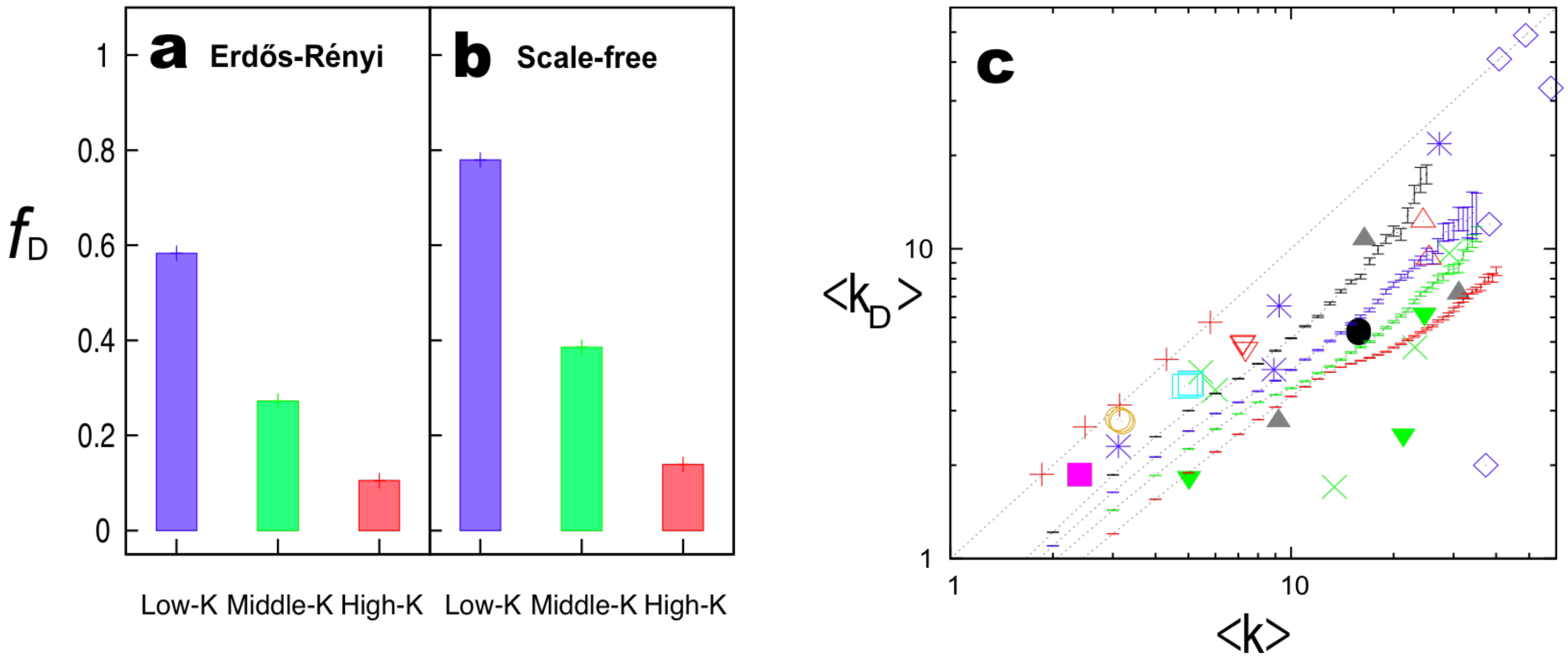
Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási, *Nature* (2011)

N_D of real networks



1. Overall we see no obvious trend in n_D (or N_D) across these networks.
2. As a group, regulatory networks display very high $n_D \approx 0.8$.
3. A few social networks display the smallest observed n_D values.

Role of hubs



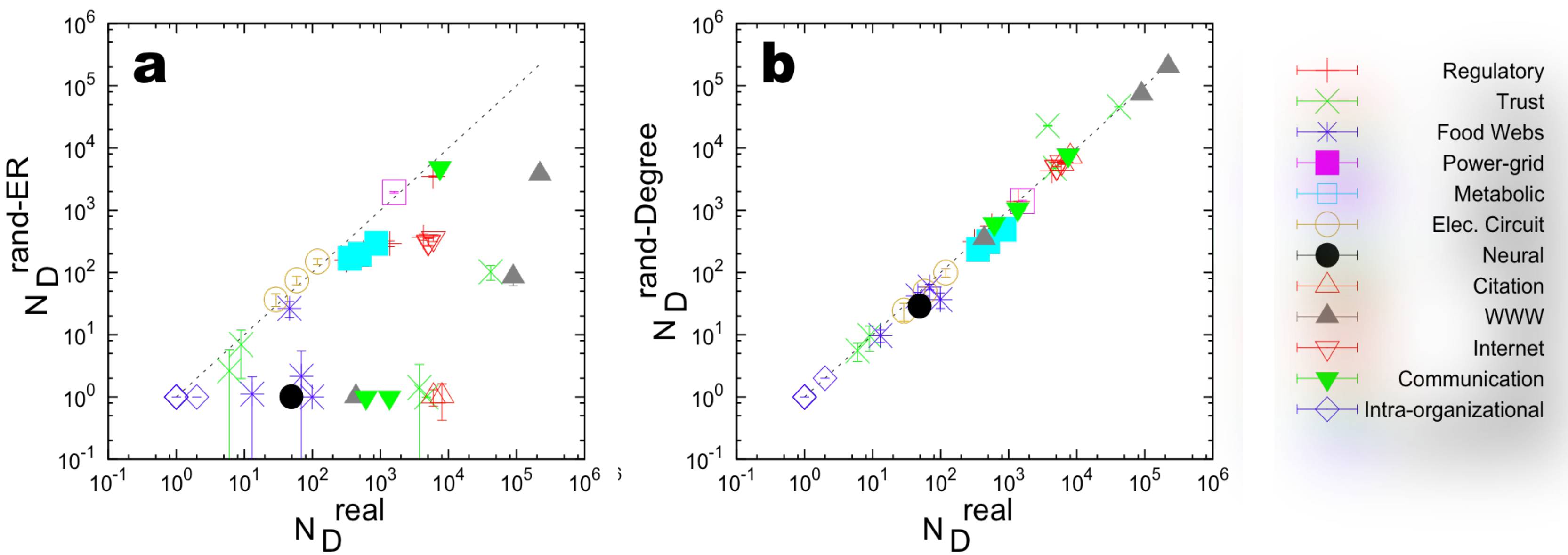
1. The fraction of driver nodes is significantly higher among low degree nodes than among the hubs.
2. Mean degree of driver nodes $<k_D>$ is either significantly smaller or comparable to $<k>$.

Driver nodes tend to avoid the hubs.

N_D^{real} vs. N_D^{rand}

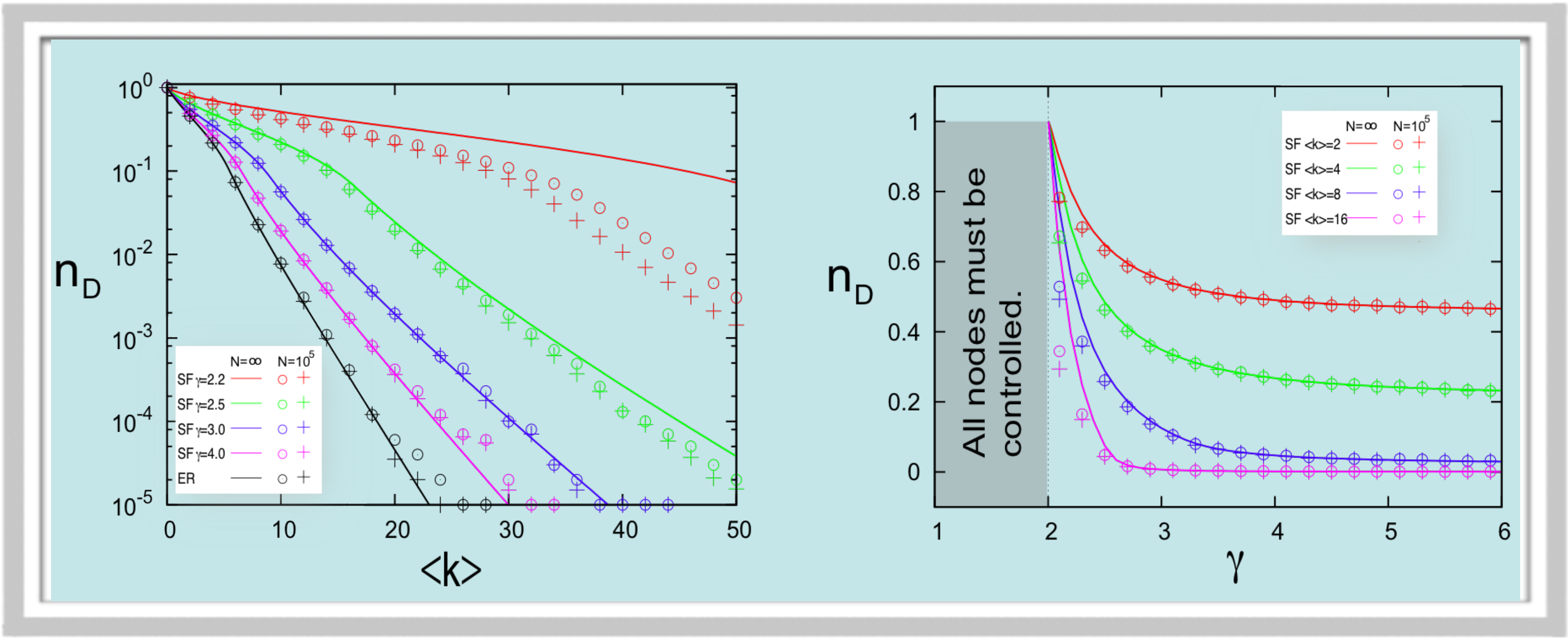
Complete randomization

Degree-preserving randomization



N_D is mainly determined by degree distribution.

Degree Dependence



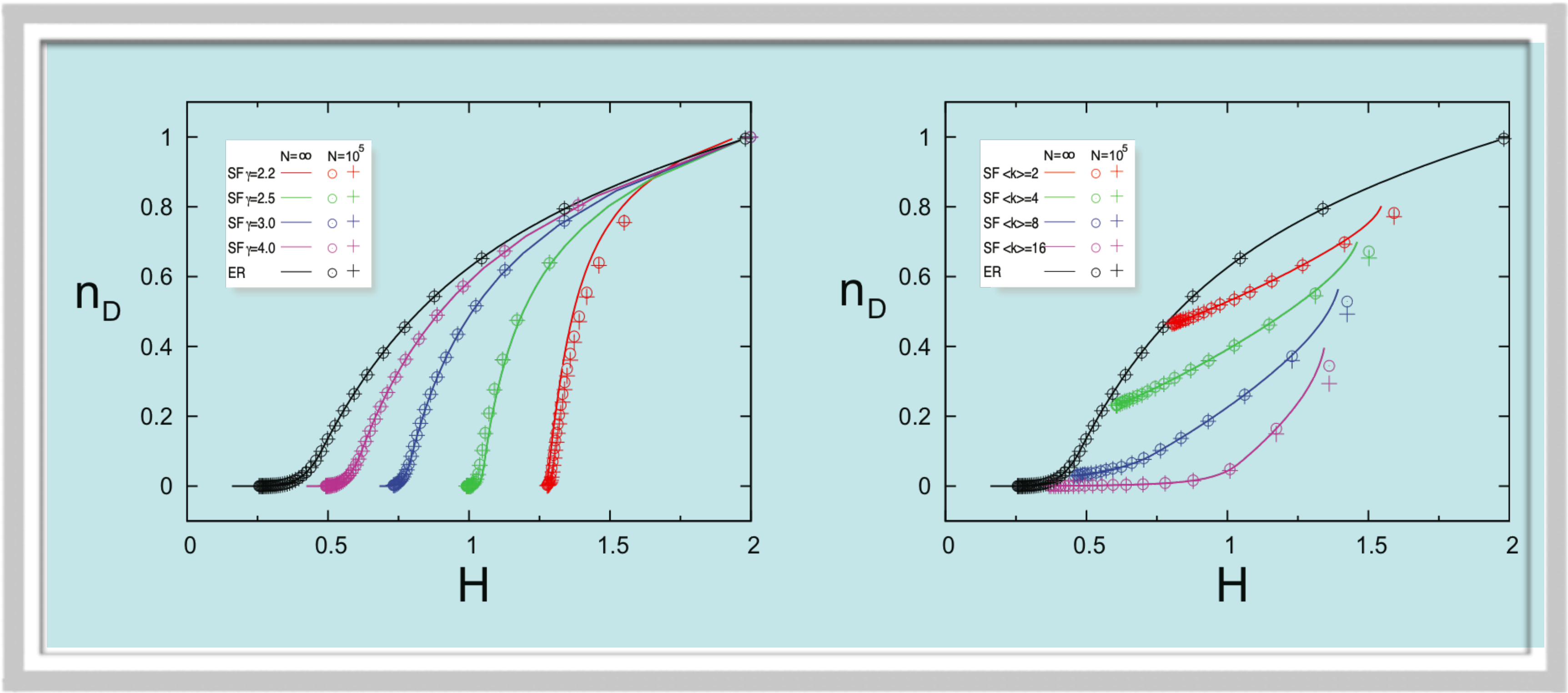
Construct ER and SF networks using the static model (Goh *et al*. PRL 2001)

1. ER : $n_D(\langle k \rangle) \propto e^{-\langle k \rangle /2}$ as $\langle k \rangle >> 1$.

2. SF : $n_D(\langle k \rangle, \gamma) \propto e^{-\left(1-\frac{1}{\gamma-1}\right)\langle k \rangle /2}$ as $\langle k \rangle >> 1$

(consistent with $\gamma_c = 2$ SF : $n_D(\gamma) \to 1$ as $\gamma \to \gamma_c = 2$.).

Degree Heterogeneity

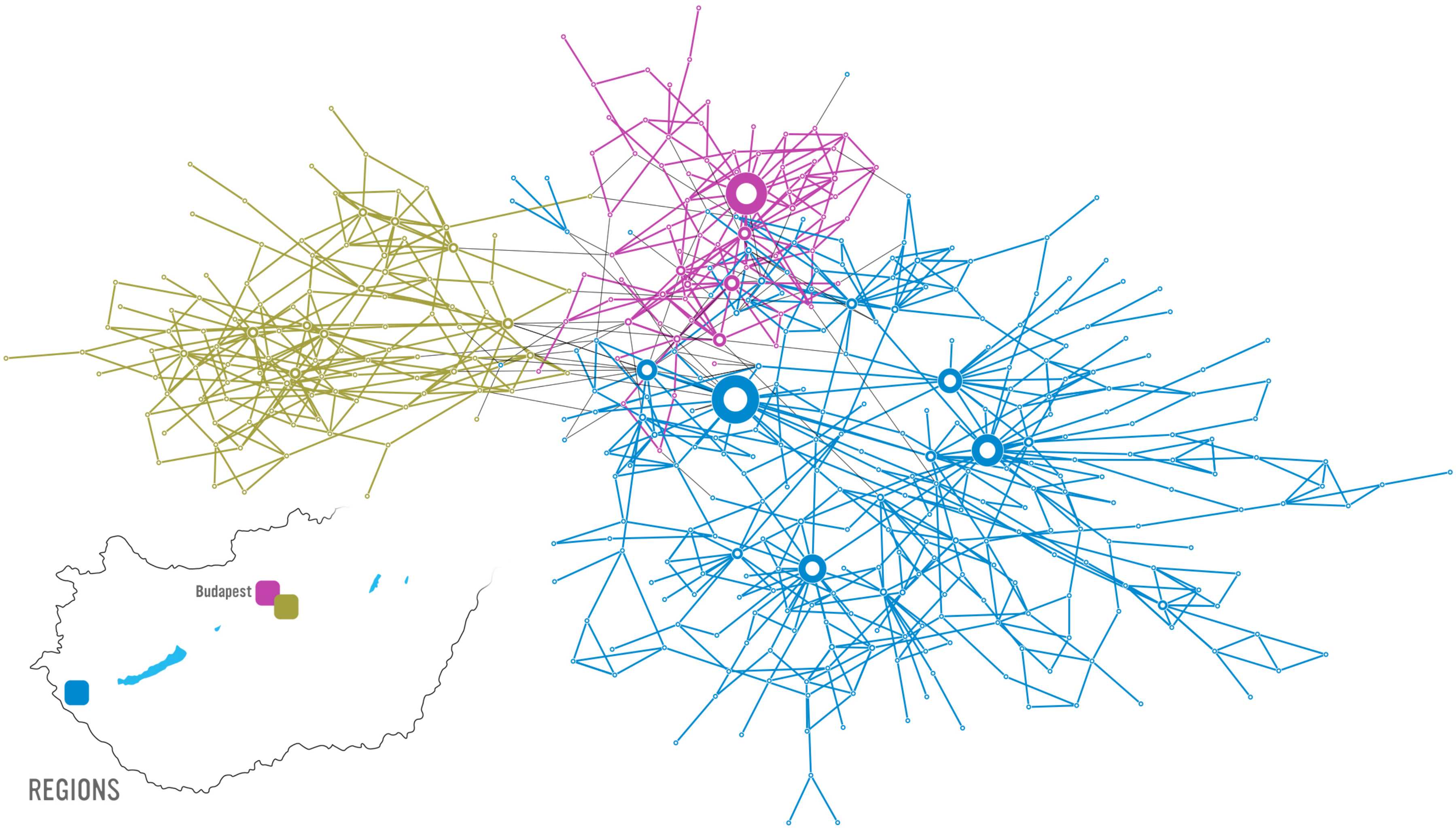


Degree heterogeneity $H = 2 \times$ Gini coefficient

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

Results

- **Mean degree** $\langle k \rangle$ and **degree heterogeneity** H are the two main factors that determine N_D.
- **Sparse** and **heterogeneous** networks are harder to control than **dense** and **homogeneous** networks.

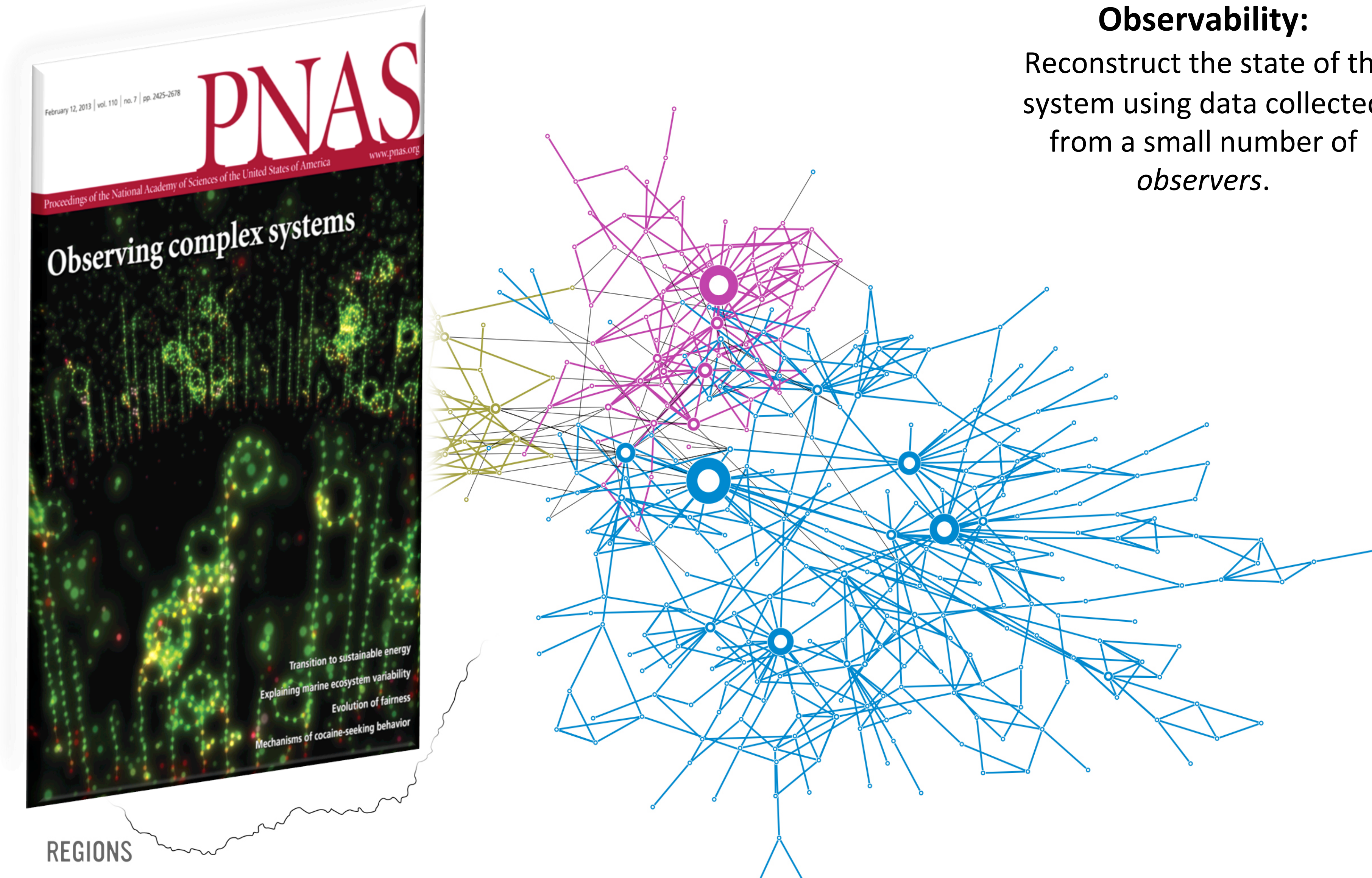


connecting knowledge

Video of individuals that are NOT control nodes

Video of individuals that ARE control nodes

OBSERVABILITY: Reconstruct the state of a complex system



Observability:
Reconstruct the state of the system using data collected from a small number of *observers*.

Liu, Slotine, Barabási *PNAS* (2013)

Network Science

an interactive textbook

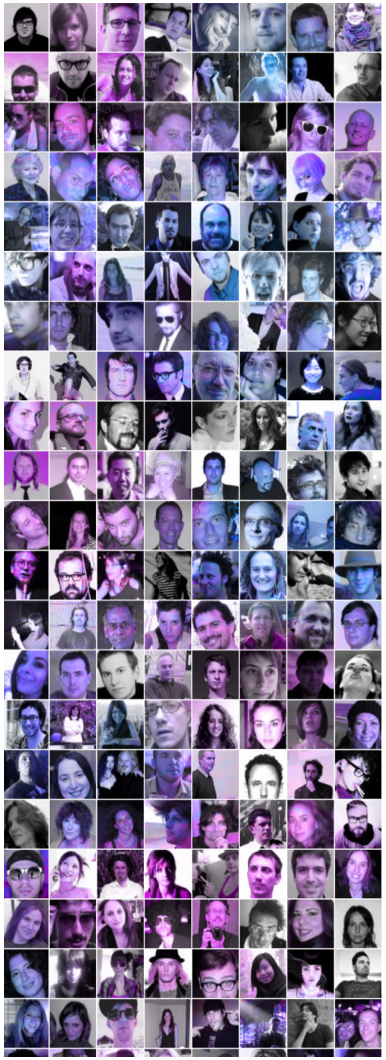
barabasilab.com/**NetworkScienceBook/**

facebook.com/**NetworkScienceBook**

Download PDF

Download ibooks

Download SLIDES



CHAPTER 1

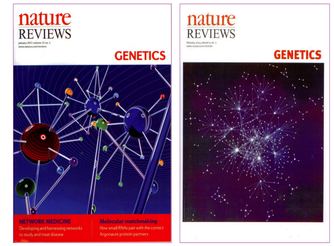


Image 1.7a, 1.7b

Networks in biology and medicine.

a) The cover of two issues of *Nature Reviews Genetics*, the top review journal in genetics. The cover from 2004, focuses on network biology [11], the cover from 2011 discuses network medicine [12].

b) The prominent role networks play in both cell biology and medical research is illustrated by the fact that the 2004 article on network biology is the second most cited article in the history of Nature Reviews Genetics.

genes and other cellular components interact with each other. Most cellular processes, from the processing of food by our cells to sensing changes in the environment, rely on molecular networks. The breakdown of these networks is responsible for most human diseases. This has led to the emergence of network biology, a new subfield of biology that aims to understand the behavior of cellular networks. A parallel movement within medicine, called network medicine, aims to uncover the role of networks in human disease (Image 1.7a/b). Networks are particularly important in drug development. The ultimate goal of network pharmacology is to develop drugs that can cure diseases without significant side effects. This goal is pursued at many levels, from millions of dollars invested to map out cellular networks to the development of tools and databases to store, curate, and analyze patient and genetic data. Several new companies take advantage of these opportunities, from *GeneGo* that aims to collect accurate maps of cellular interactions from scientific literature to *Genomatica* that uses the predictive power behind metabolic networks to identify drug targets in bacteria and humans. Recently most major pharmaceutical companies have made signifi-

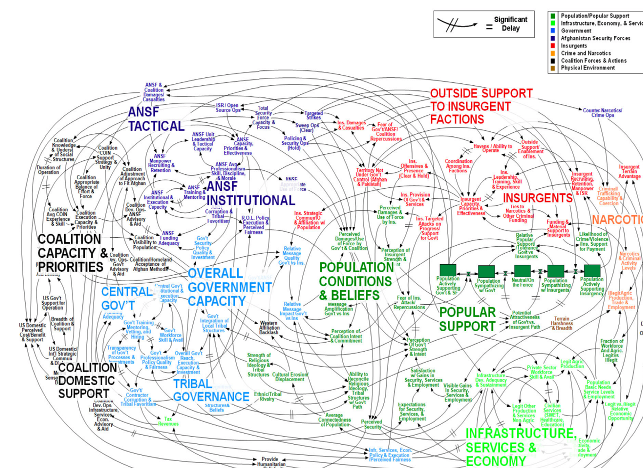


Image 1.8

The network behind a military engagement.

This diagram was designed during the Afghan war to portray the American strategy in Afghanistan. While it has been occasionally ridiculed in the press, it portrays well the complexities and the interconnected nature of a military's engagement. *(Image from New York Times)*

barabasilab.com/**NetworkScienceBook/**

facebook.com/**NetworkScienceBook**

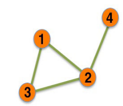
CHAPTER 2

Image 2.16

Graphology.

In network science we encounter many networks distinguished by some elementary property of the underlying graph. Here we summarize the most commonly encountered elementary network types, together with their basic properties, and an illustrative list of real systems that share the particular property. Note that in many real network we need to combine several of these elementary network characteristics. For example the WWW is a directed multi-graph with self-interactions. The mobile call network is directed and weighted, without self-loops.

Undirected



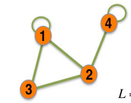
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{2L}{N}$$

UNDIRECTED NETWORK: a network whose links do not have a predefined direction. Examples: Internet, power grid, science collaboration networks, protein interactions.

Self-interactions



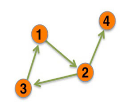
$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$A_{ii} \neq 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1,i\neq j}^{N} A_{ij} + \sum_{i=1}^{N} A_{ii} \qquad ?$$

SELF-INTERACTIONS: in many networks nodes do not interact with themselves, so the diagonal elements of adjacency matrix are zero, $A_{ii} = 0$, $i = 1,...,N$. In some systems self-interactions are allowed; in such networks, representing the fact that node *i* has a self-interaction. Examples: WWW, protein interactions.

Directed



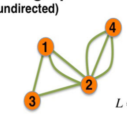
$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} \neq A_{ji}$$

$$L = \sum_{i,j=1}^{N} A_{ij} \qquad <k> = \frac{L}{N}$$

DIRECTED NETWORK: a network whose links have selected directions. Examples: WWW, mobile phone calls, citation network.

Multigraph (undirected)



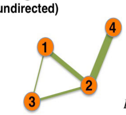
$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad <k> = \frac{2L}{N}$$

MULTIGRAPH: in a multigraph nodes are permitted to have multiple links (or parallel links) between them. Hence A_{ij} can have any positive integer.

Weighted (undirected)



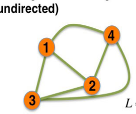
$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2}\sum^{N} nonzero(A_{ij}) \qquad <k> = \frac{2L}{N}$$

WEIGHTED NETWORK: a network whose links have a predefined weight, strength or fow parameter. The elements of the adjacency matrix are $A_{ij} = 0$ if *i* and *j* are not connected, or $A_{ij} = w_{ij}$ if there is a link with weight wij between them. For unweighted (binary) networks, the adjacency matrix only indicates the presence ($A_{ij} = 1$) or the absence ($A_{ij} = 0$) of a link between two nodes. Examples: Mobile phone calls, email network.

Complete Graph (undirected)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \qquad A_{i\neq j} = 1$$

$$L = L_{max} = \frac{N(N-1)}{2} \qquad <k> = N-1$$

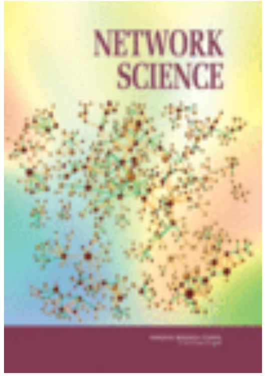
COMPLETE GRAPH: in a complete graph all nodes are connected to each other; no self-connections.

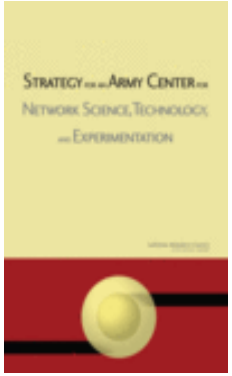
barabasilab.com/**NetworkScienceBook/** facebook.com/**NetworkScienceBook**

WHAT IS “NETWORK SCIENCE”?

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

NRC Report on “Network Science”





An attempt to understand networks emerging in nature, technology and society using a unified set of tools and principles.

What is new here?

Despite the apparent differences, many networks emerge and evolve driven by a fundamental set of laws and mechanism.

www.BarabasiLab.com