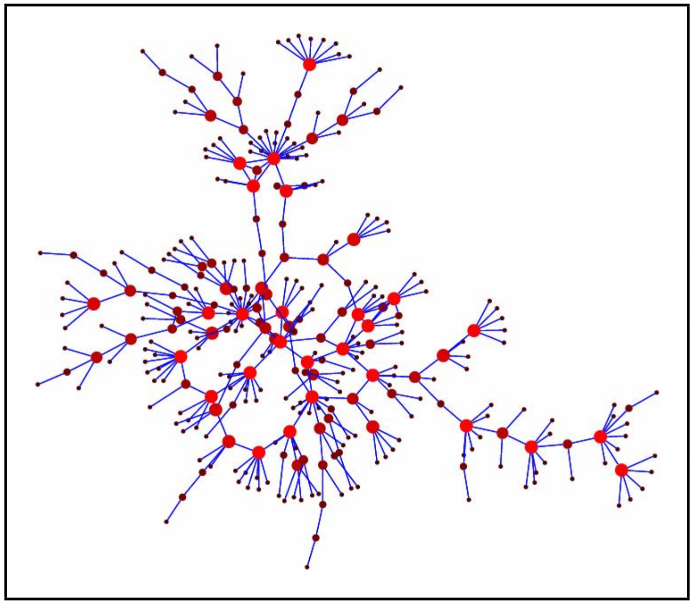
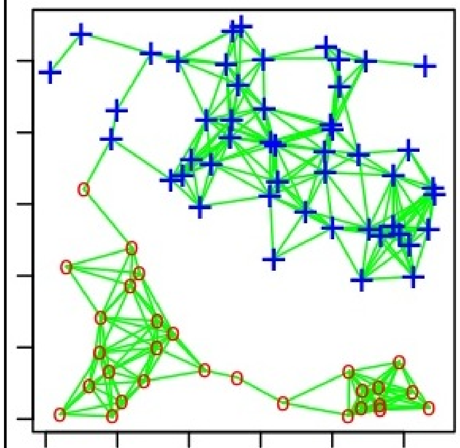
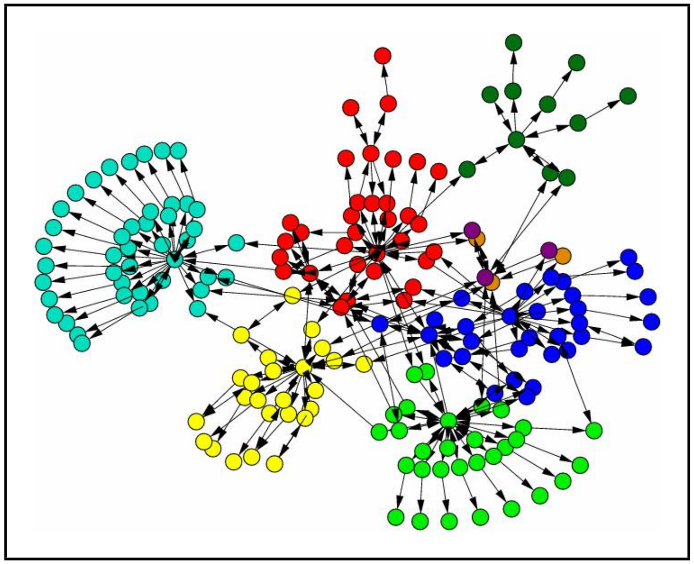
Overview of Network Theory, I







ECS 253 / MAE 253, Spring 2016, Lecture 1

Prof. Raissa D'Souza
University of California, Davis

Raissa's background:

- 1999, PhD, Physics, Massachusetts Inst of Tech (MIT):
  – Joint appointment: Statistical Physics and Lab for Computer Science

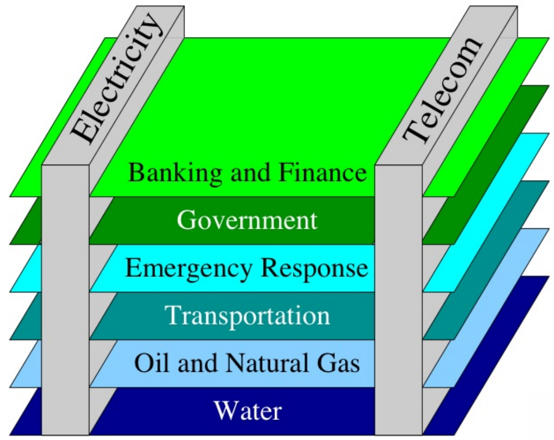
- 2000-2002, Postdoctoral Research Fellow, Bell Laboratories:
  – Joint appointment: Fundamental Mathematics and Theoretical Physics Research Groups.

- 2002-2005, Postdoctoral Research Fellow, Microsoft Research:
  – "Theory Group" (Physics and Theoretical Computer Science)

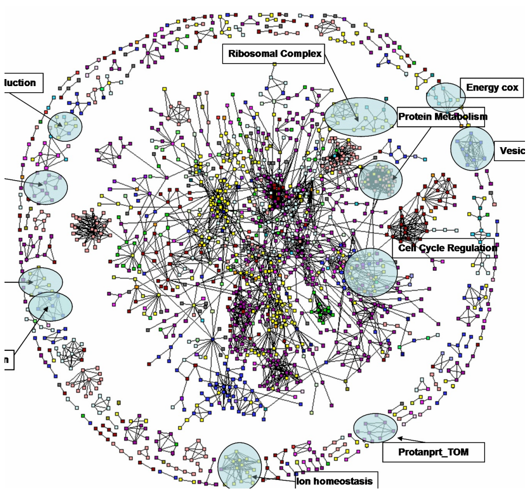
- Fall 2005-present, UC Davis:
  – Dept of Computer Science, Dept of Mechanical and Aeronautical Eng., Complexity Sciences Center, Grad Group Applied Math.

- 2007-present, External Faculty Member, Santa Fe Institute

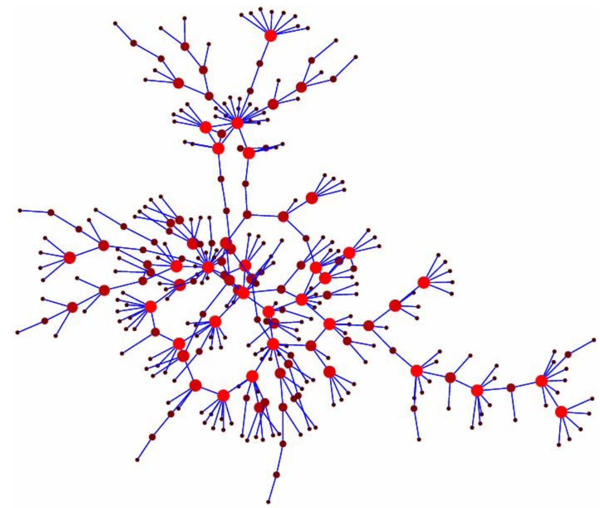
Complex networks are ubiquitous:



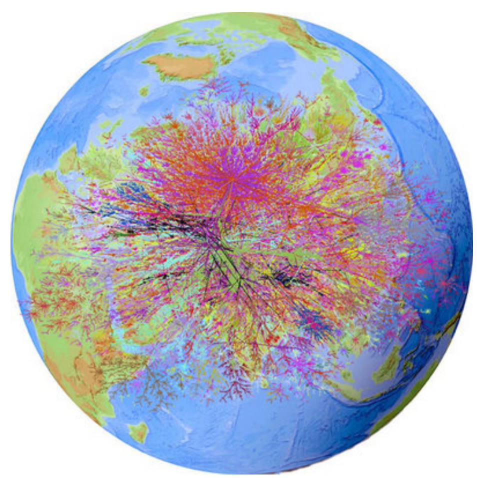
Critical Infrastructure



Biological & Ecological networks



Social networks: Economics & Epidemics



Information and Communication technology

What is a Network?

- Topology (i.e., structure: nodes/vertices and edges/links)
  Measures of topology

- Activity (i.e., function, processes on networks, dynamics of nodes and edges)

Modeling networks

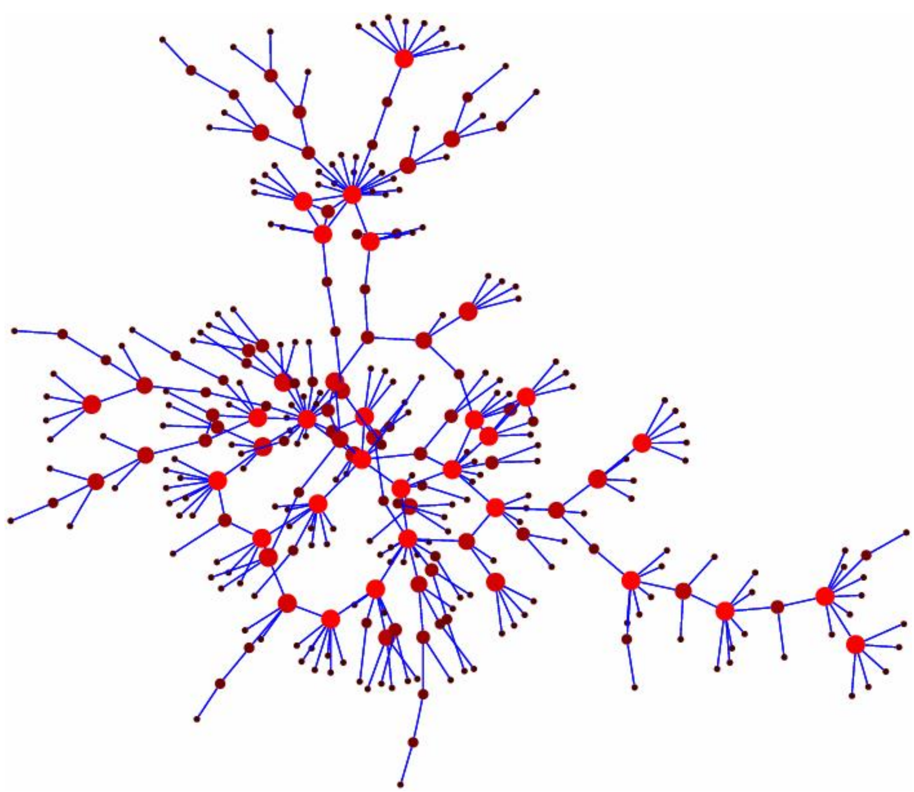
- Network growth

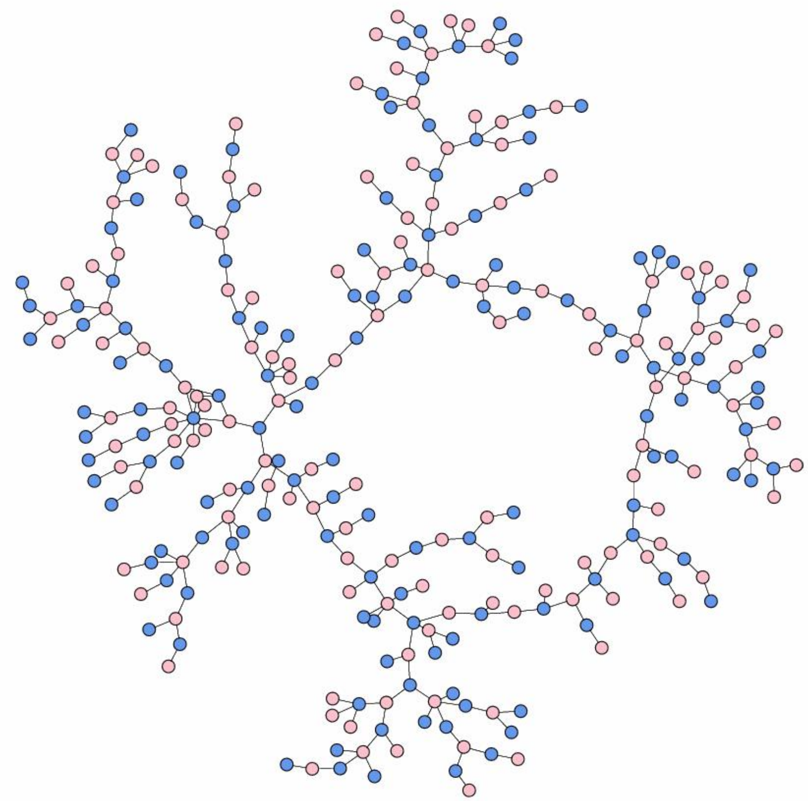
- Phase transitions

- Algorithms: analysis, growth/formation, searching and spreading

- Processes on networks

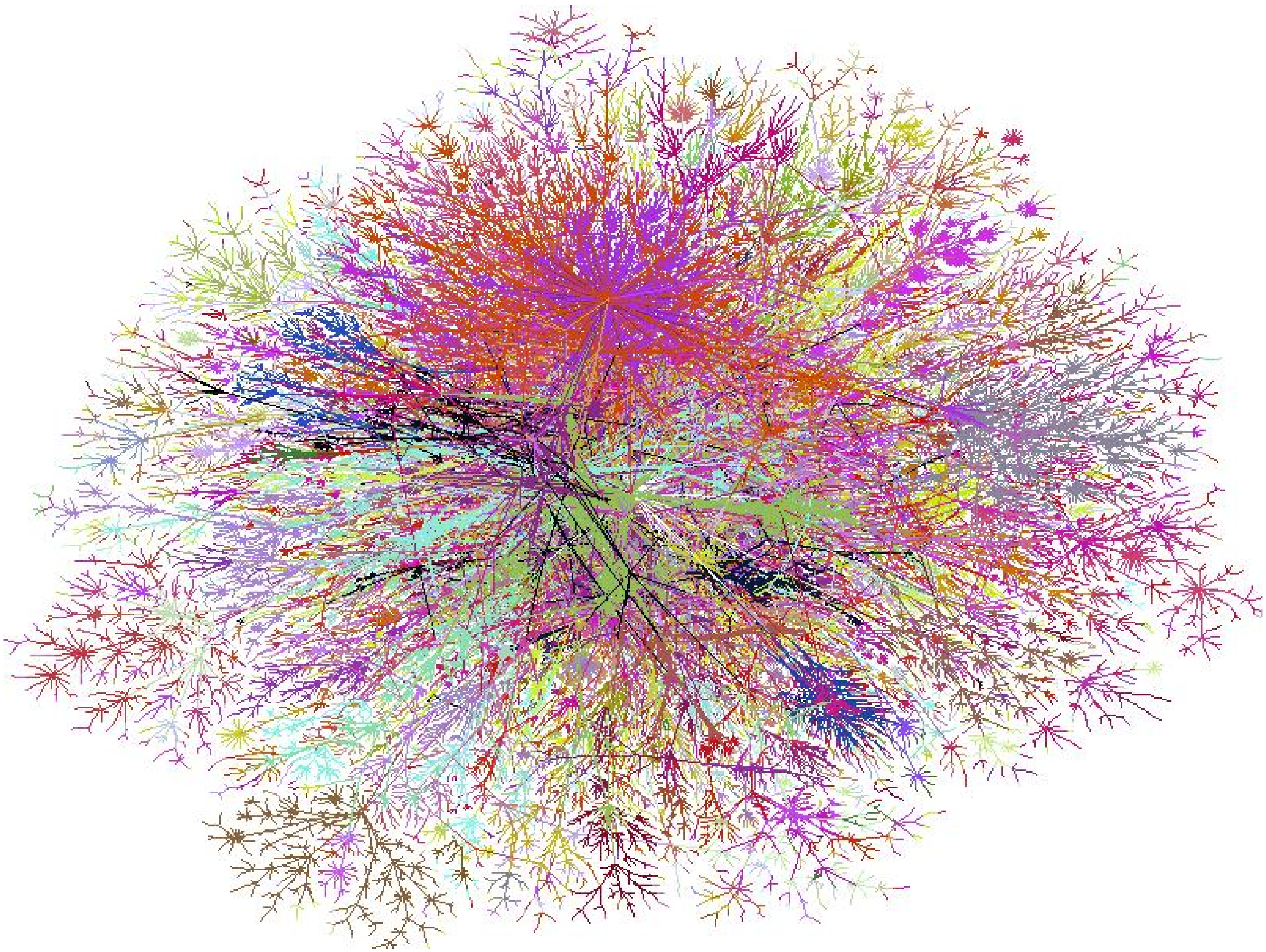
Example social networks
(Immunology; viral marketing; aliances/policy)





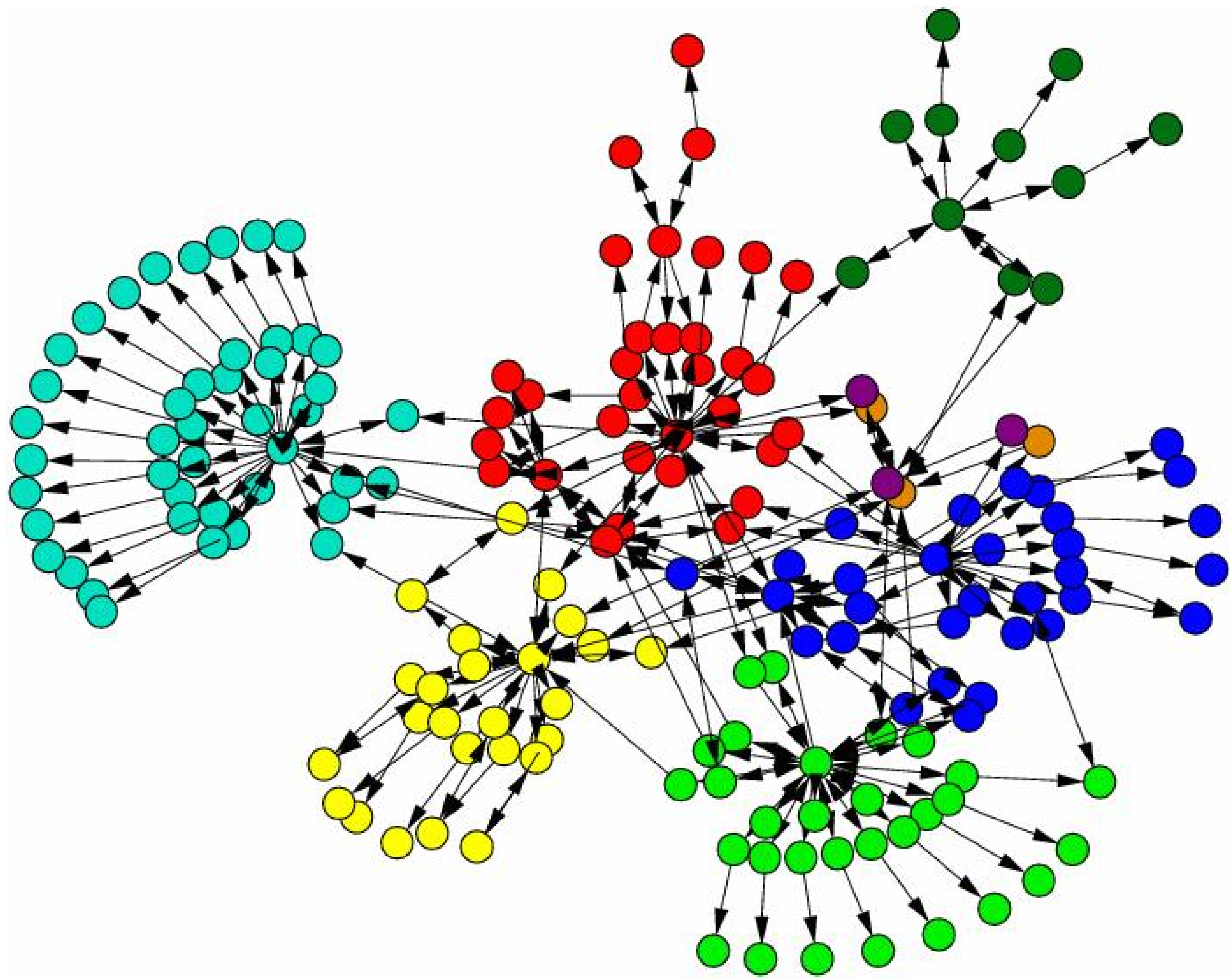
M. E. J. Newman

The Internet
(Robustness to failure; optimizing future growth; testing protocols on sample topologies)



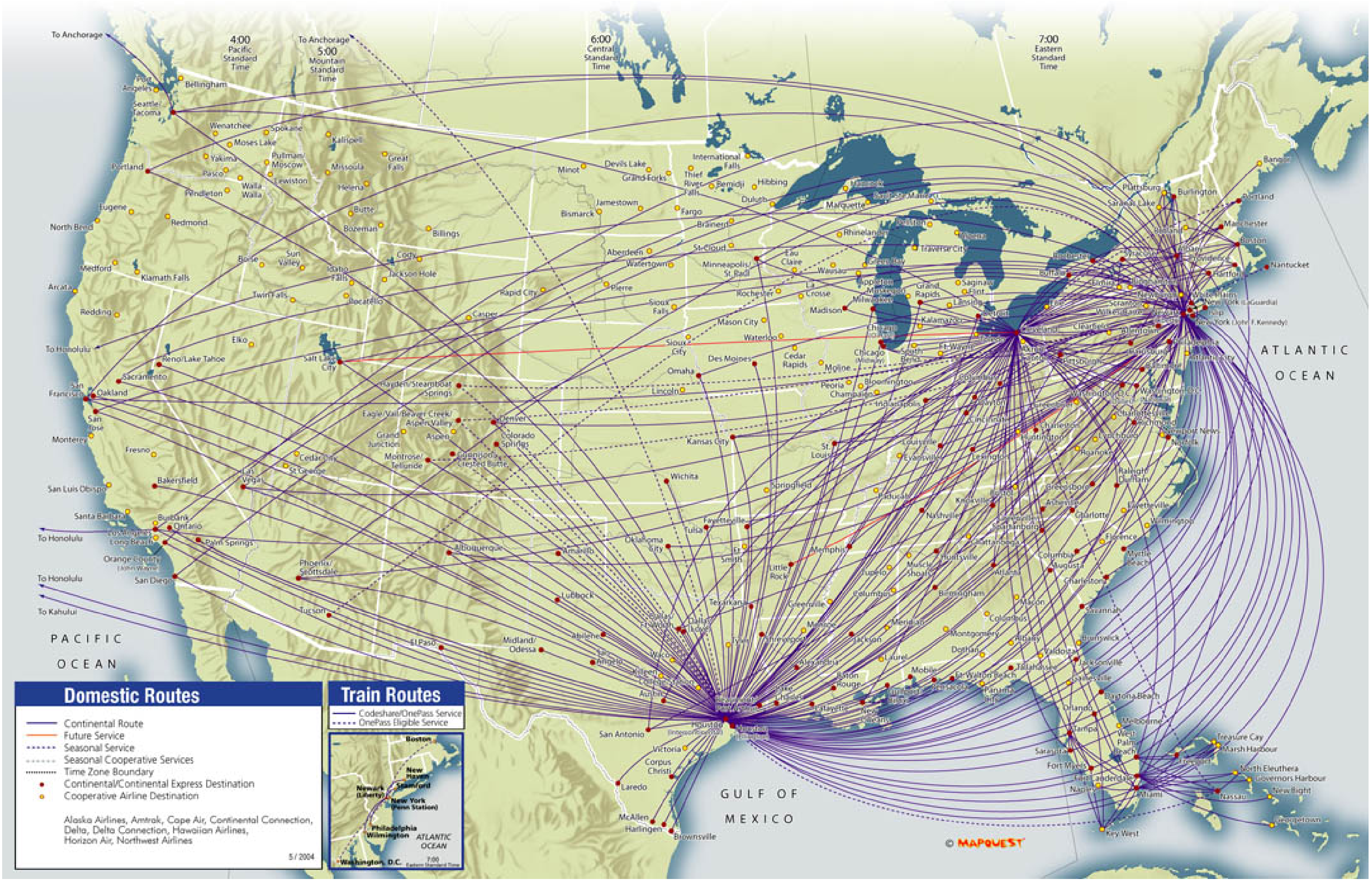
H. Burch and B. Cheswick

A typical web domain
(Web search/organization and growth centralized vs. decentralized protocols)



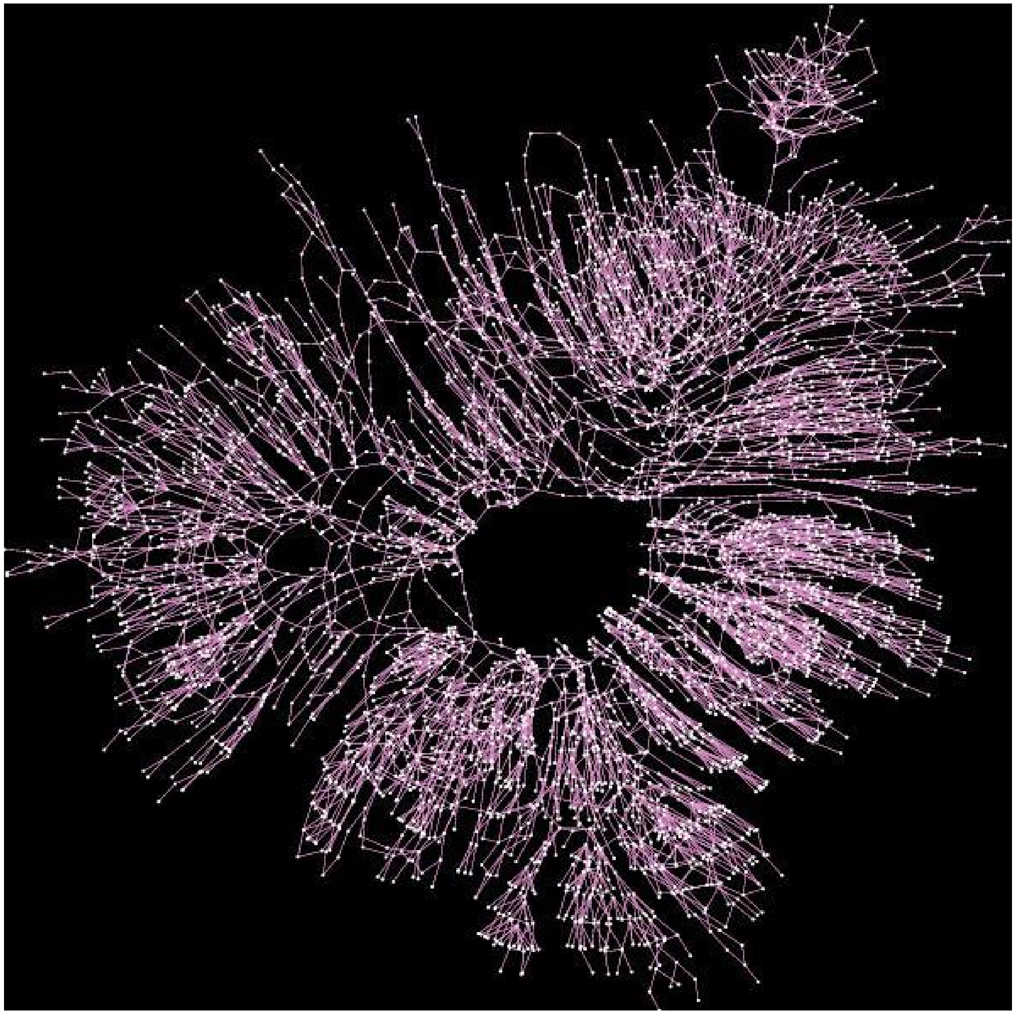
M. E. J. Newman

The airline network
(Optimization; dynamic external demands)



Continental Airlines

The power grid
(Mitigating failure; Distributed sources)

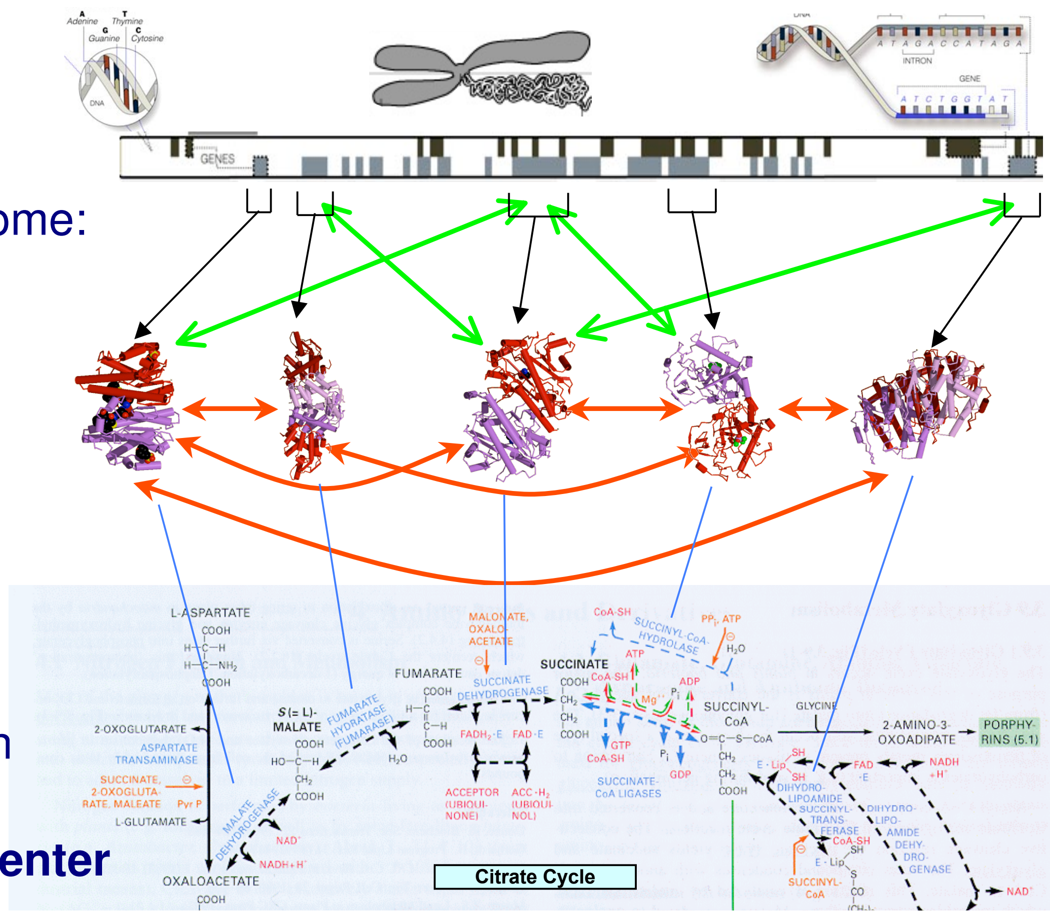


M. E. J. Newman

Biology: Networks at many levels

Control mechanisms / drug design/ gene therapy / biomarkers of disease

- Genome, Proteome: **Dandekar Lab**
- Metabolome: **Fiehn Lab**
- Data intergration BIOshare **Lin, Genome Center**
- Network structure / search for biomarkers: **D'Souza**



GENOME

protein-gene interactions

PROTEOME

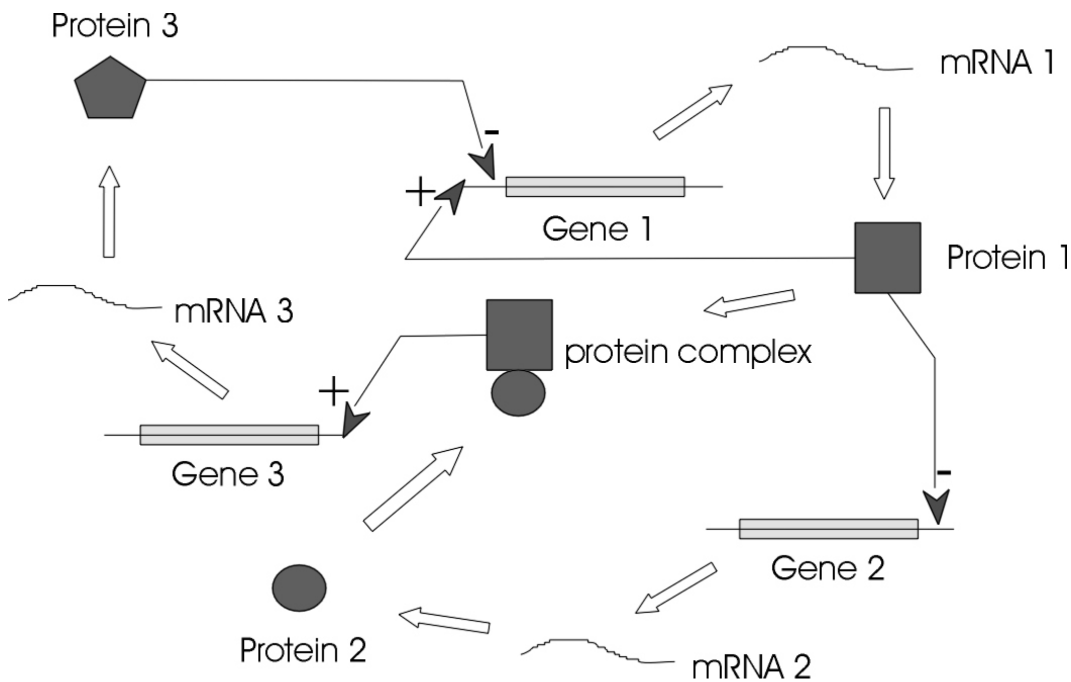
protein-protein interactions

METABOLISM

Bio-chemical reactions

research in biological networks

- gene regulatory networks
  - humans have only 30,000 genes, 98% shared with chimps
  - the complexity is in the interaction of genes
  - can we predict what result of the inhibition of one gene will be?



Slide from Adamic's course

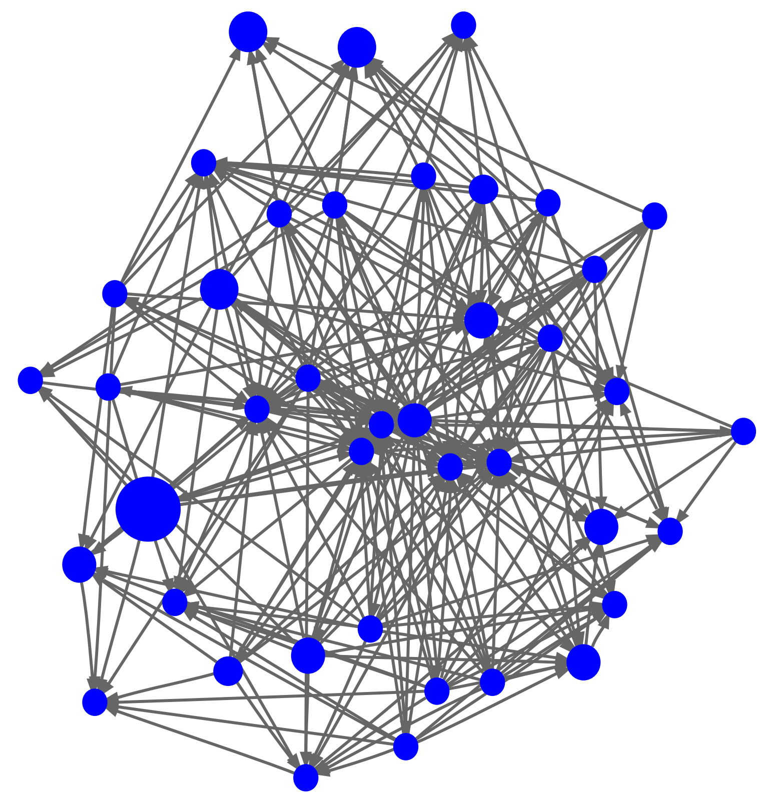
Software systems

(Highly evolveable, modular, robust to mutation, exhibit punctuated eqm)

Open-source software as a “systems” paradigm.

Networks:

- Function calls
- Email communication
- Socio-Technical congruence



Bird, Devanbu, D’Souza, Filkov, Saul, Wen

The past decade, a "Science of Networks":

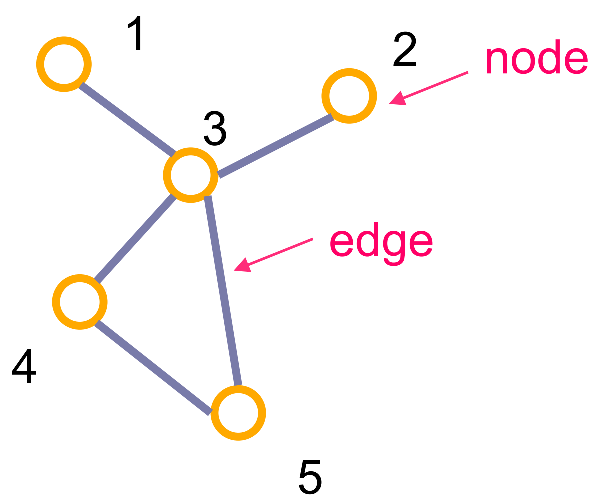
(Physical, Biological, Social)

- **Geometric** versus **virtual** (Internet versus WWW).
- **Natural** /spontaneously arising versus **engineered** /built.
- **Directed** versus **undirected** edges.
- Each network **optimizes** something unique.
- Identifying **similarities** and fundamental **differences**
- Interplay of **topology** and **function** ?
- Unifying features: – **Broad heterogeneity in node degree.**
  – **Small Worlds** (Diameter $\sim \log(N)$).

What are networks?

- Networks are collections of points joined by lines.

"Network" ≡ "Graph"



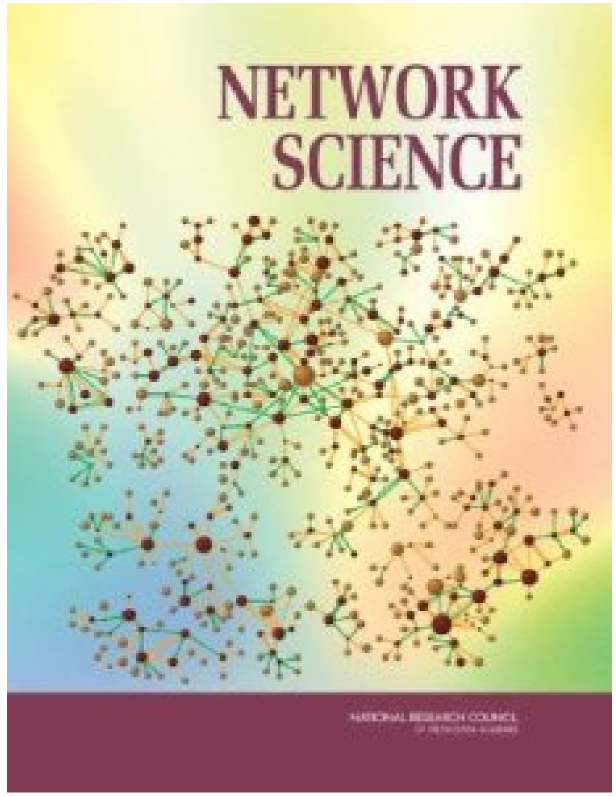
points	lines	
vertices	edges, arcs	math
nodes	links	computer science
sites	bonds	physics
actors	ties, relations	sociology

Slide from Adamic's course

Explosion of work and tools

- R, Graphviz, Pajek, igraph, Network Workbench, NetworkX, Netdraw, UCInet, Bioconductor, Ubigraph....

Natn Acam Sciences/Natn Research Council Study (2005)

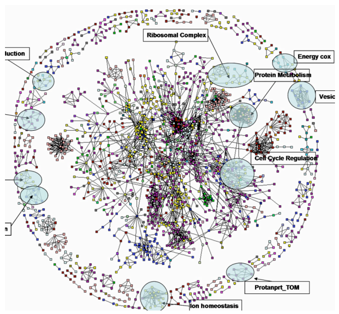


“all our modern critical infrastructure relies on networks... too much emphasis on specific applications/jargon/disciplinary stovepipes... need a cross-cutting science of networks... Research for the 21st century”

In reality a collection of interacting networks:

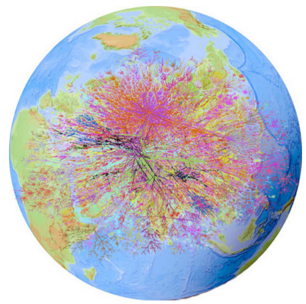
Networks:

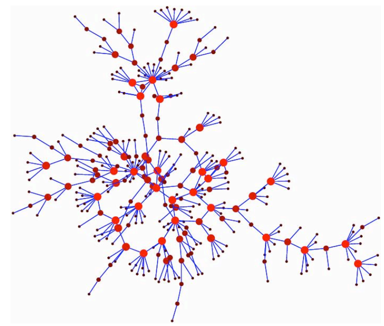
Transportation Networks/ Power grid
(distribution/ collection networks)



Biological networks
- protein interaction
- genetic regulation
- drug design

Computer networks





Social networks
- Immunology
- Information
- Commerce

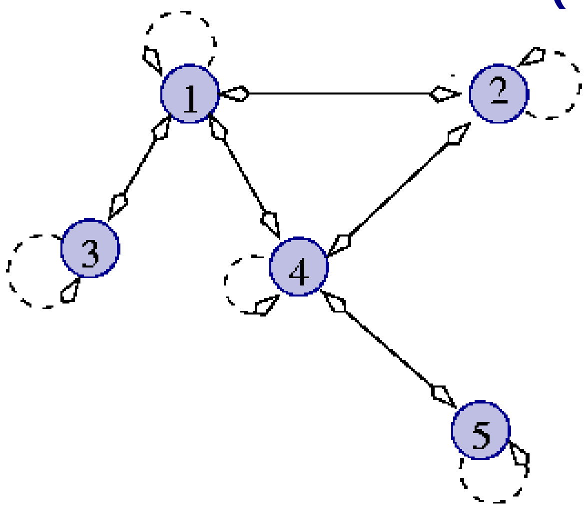
- E-commerce $\rightarrow$ WWW $\rightarrow$ Internet $\rightarrow$ Power grid $\rightarrow$ River networks.
- Biological virus $\rightarrow$ Social contact network $\rightarrow$ Transportation networks $\rightarrow$ Communication networks $\rightarrow$ Power grid $\rightarrow$ River networks.
  (Historical progression: Spatial waves (Black plague)   Regional outbreaks (ships)   Global pandemics (airplanes))

How do we represent a simple individual network as a mathematical object?

NETWORK TOPOLOGY

Connectivity matrix, M:

$$M_{ij} = \begin{cases} 1 \text{ if edge exists between } \; i \text{ and } j \\ 0 \text{ otherwise.} \end{cases}$$

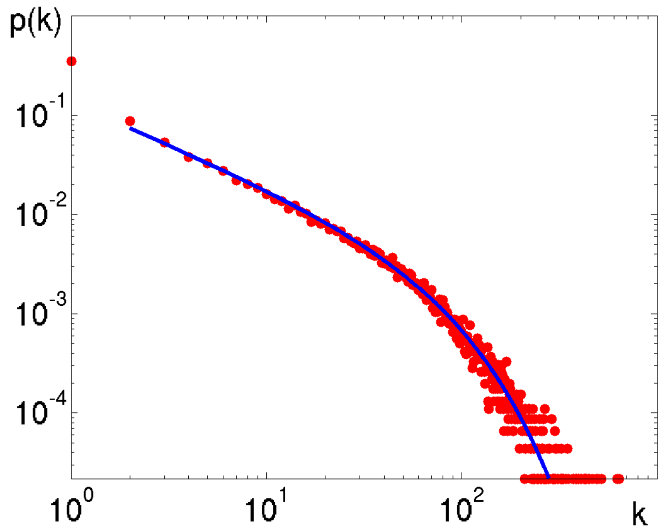


$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} = M$$

Node **degree** is number of links.

Typical measures of network topology

- **Degree distribution** (fraction of nodes with degree k, for all k)



- **Clustering coefficient** (fraction of triangles in the graph/transitivity: Are my friends friends with each other?)

  Also a local measure, for each node c_i is number of connections existing between neighbors/total number of possible connections.

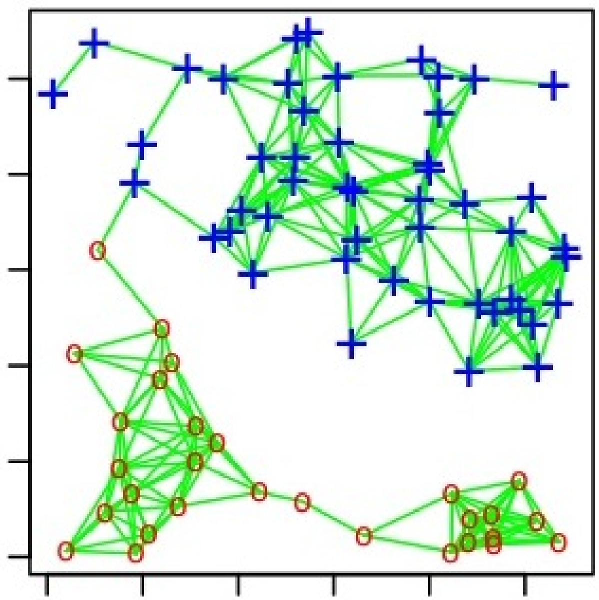
Typical measures of network topology, cont

- **Diameter** (Greatest distance between any two connected nodes)

  "Small world" if $d \sim \log N$ and strong clustering.

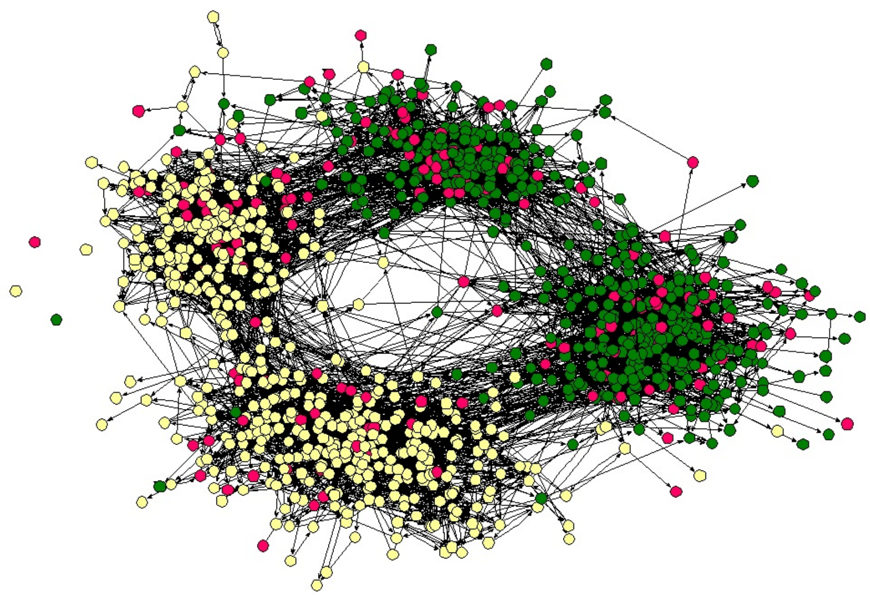
  (Watts Stogatz, *Nature* **393**, 1998.)

- **Betweenness centrality** (Fraction of shortest paths passing through a node, i.e., is a node a bottleneck for flow?)



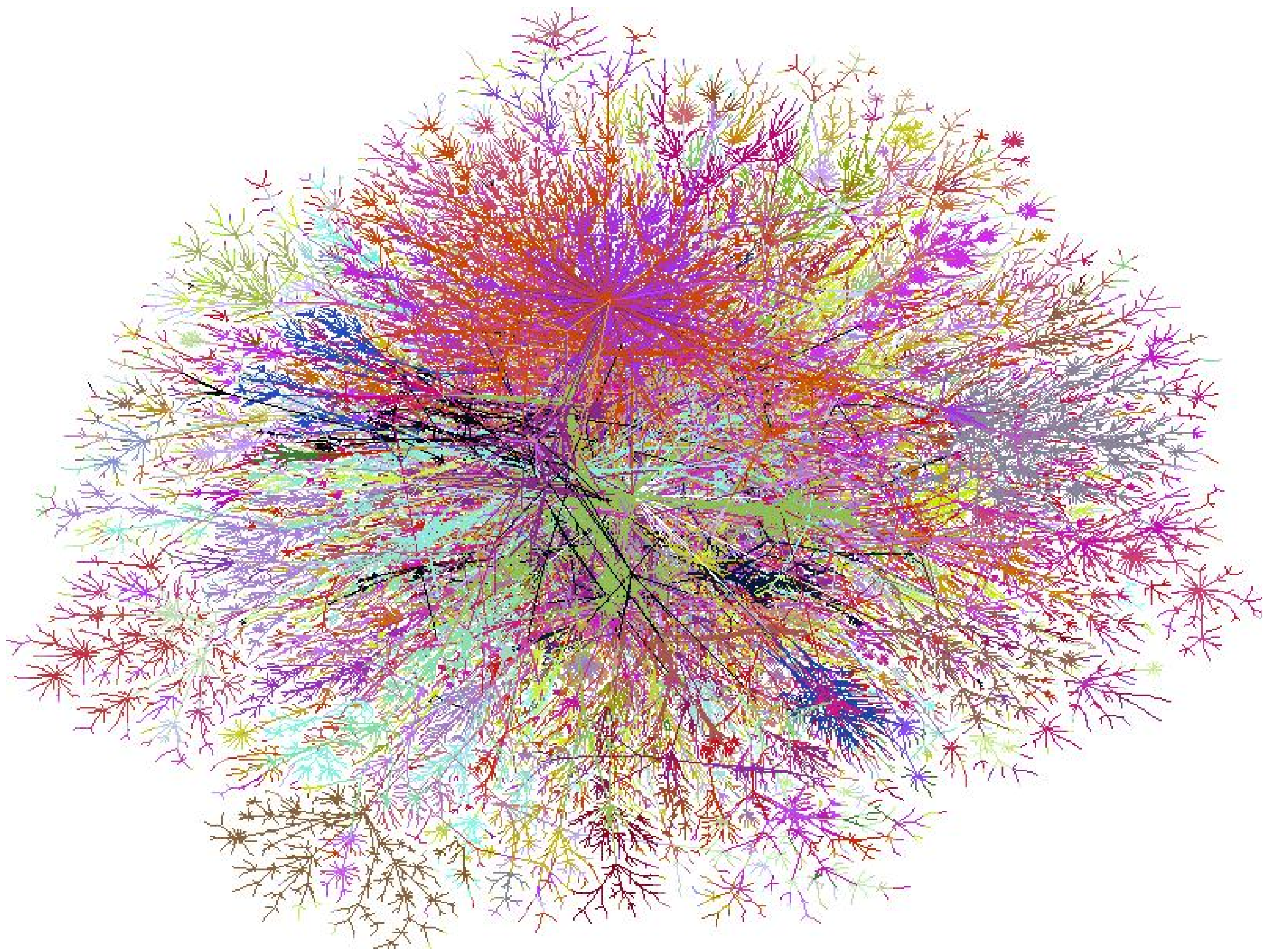
Typical measures of network topology, cont

- **Assortative/dissortative mixing** (Are nodes with similar attributes more or less likely to link to each other? Mixing by node degree common. Also, in social networks mixing by gender and race.)



(Example of assortative mixing by race. Friendship network of HS students: White, African American and Other.)

Degree distribution of "real-world" networks



Extremely broad range of node degree observed: from biological, to technological, to social.

Typical distribution in node degree

The "Internet"

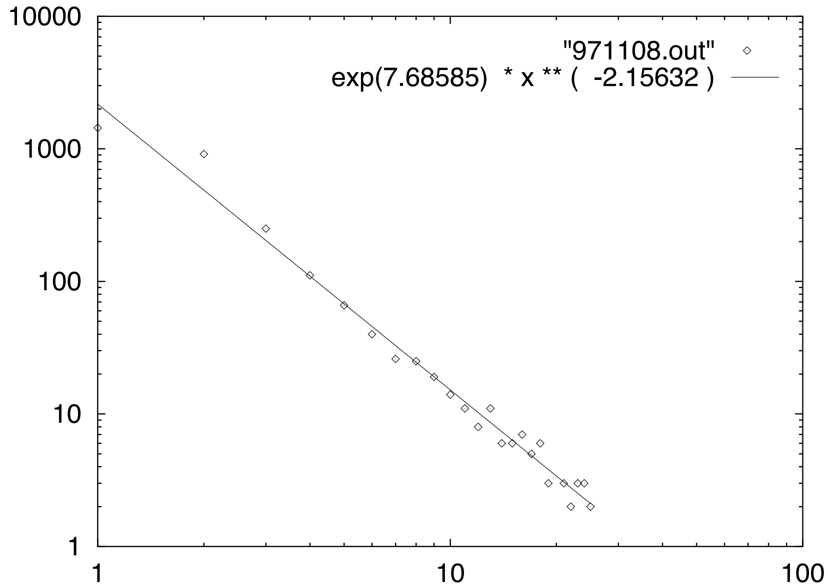
Faloutsos[3], *SIGCOMM* 1999

$$p(k) \sim k^{-2.16}$$

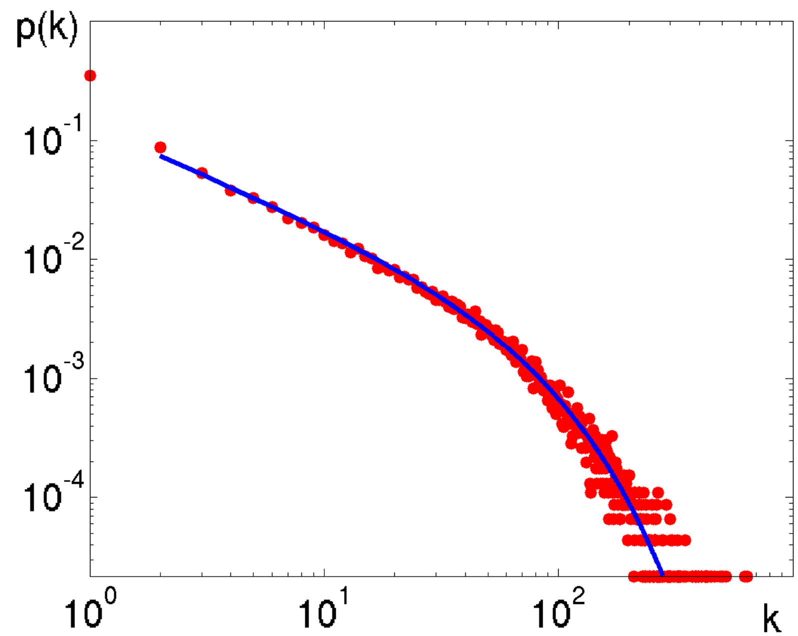
"Who-is-Who" network

Szendröi and Csányi

$$p(k) = ck^{-\gamma}e^{-\alpha k}$$



(a) Int-11-97



- Small data sets, power laws vs other similar distributions?
  - What is the "Internet"/ what level? (e.g., router vs AS)

Power law with exponential tail

Ubiquitous empirical measurements:

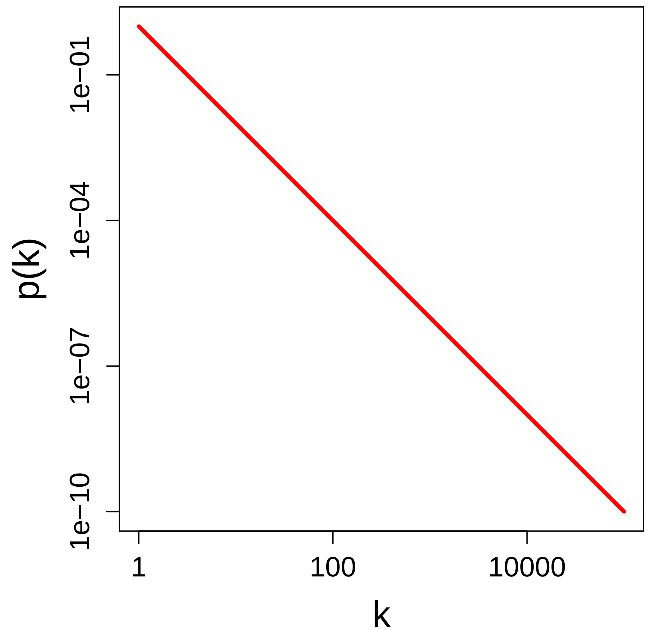
System with: $p(x) \sim x^{-B} \exp(-x/C)$	B	C
Full protein-interaction map of *Drosophila*	1.20	0.038
High-confidence protein-interaction map of *Drosophila*	1.26	0.27
Gene-flow/hydridization network of plants as function of spatial distance	0.75	10^5 m
Earthquake magnitude	1.35 - 1.7	$\sim 10^{21}$ Nm
Avalanche size of ferromagnetic materials	1.2 - 1.4	$L^{1.4}$
ArXiv co-author network	1.3	53
MEDLINE co-author network	2.1	~ 5800
PNAS paper citation network	0.49	4.21

What is a power law?

(Also called a “Pareto Distribution” in statistics).

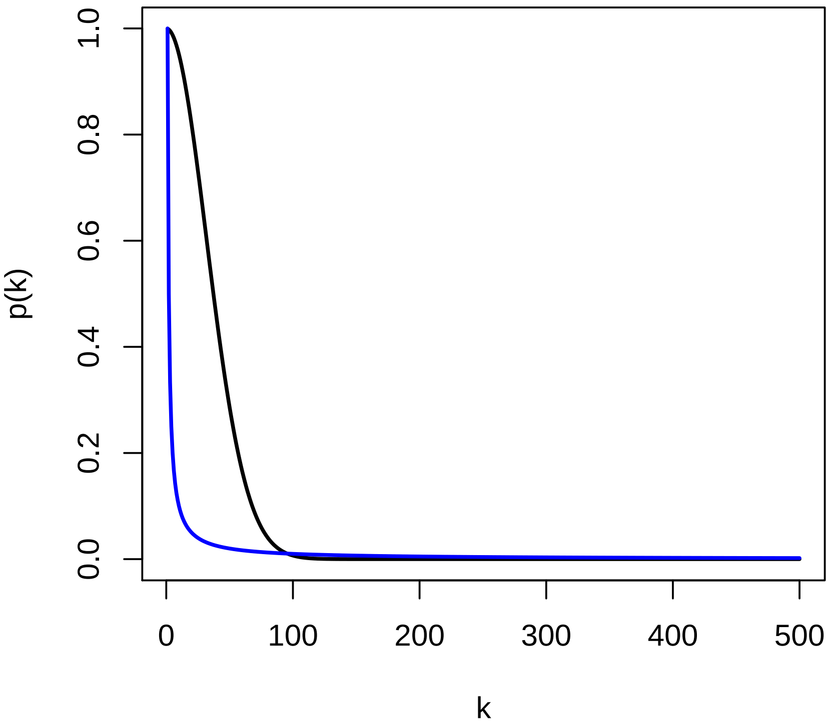
$$p_k \sim k^{-\gamma}$$

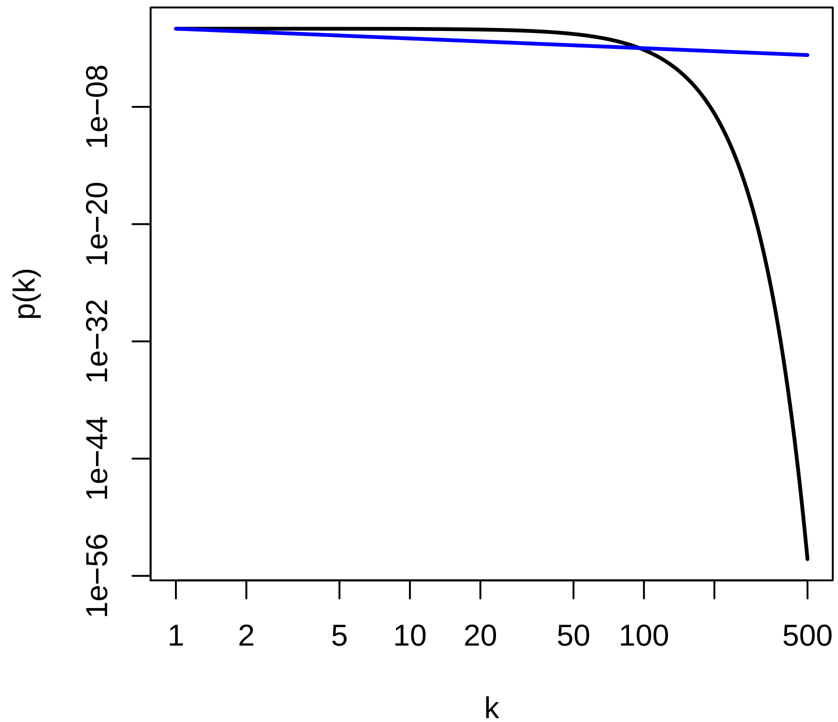
$$\ln p_k \sim -\gamma \ln k$$



Power Laws versus Bell Curves: “Heavy tails”

- Power law distribution: $p_k \sim k^{-\gamma}$.
- Gaussian distribution: $p_k \sim \exp(-k^2/2\sigma^2)$.





If $1 < \gamma < 2$, mean and variance $\to \infty$.

If $2 < \gamma < 3$ mean is finite, but variance $\to \infty$.

Many network growth models produce power law degree distribution

- Preferential attachment
- Copying models (WWW, biological networks, ...)
- Optimization models

Some outstanding challenges

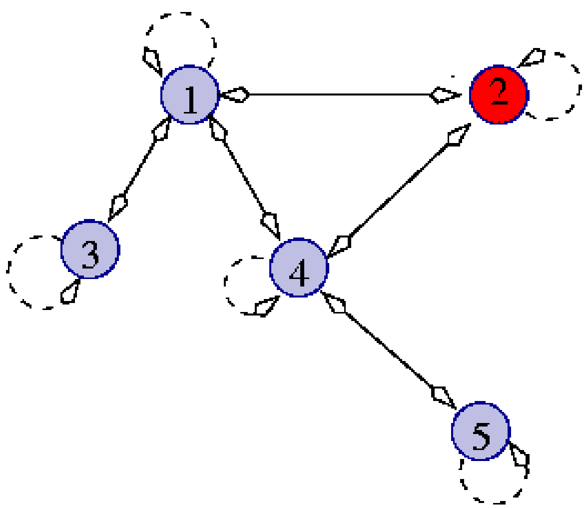
- Incorporating additional attributes beyond degree
- Validation

Network Activity: FLOWS on NETWORKS

(Spread of disease, routing data, materials transport/flow, gossip spread/marketing)

FLOWS on NETWORKS : Random walks

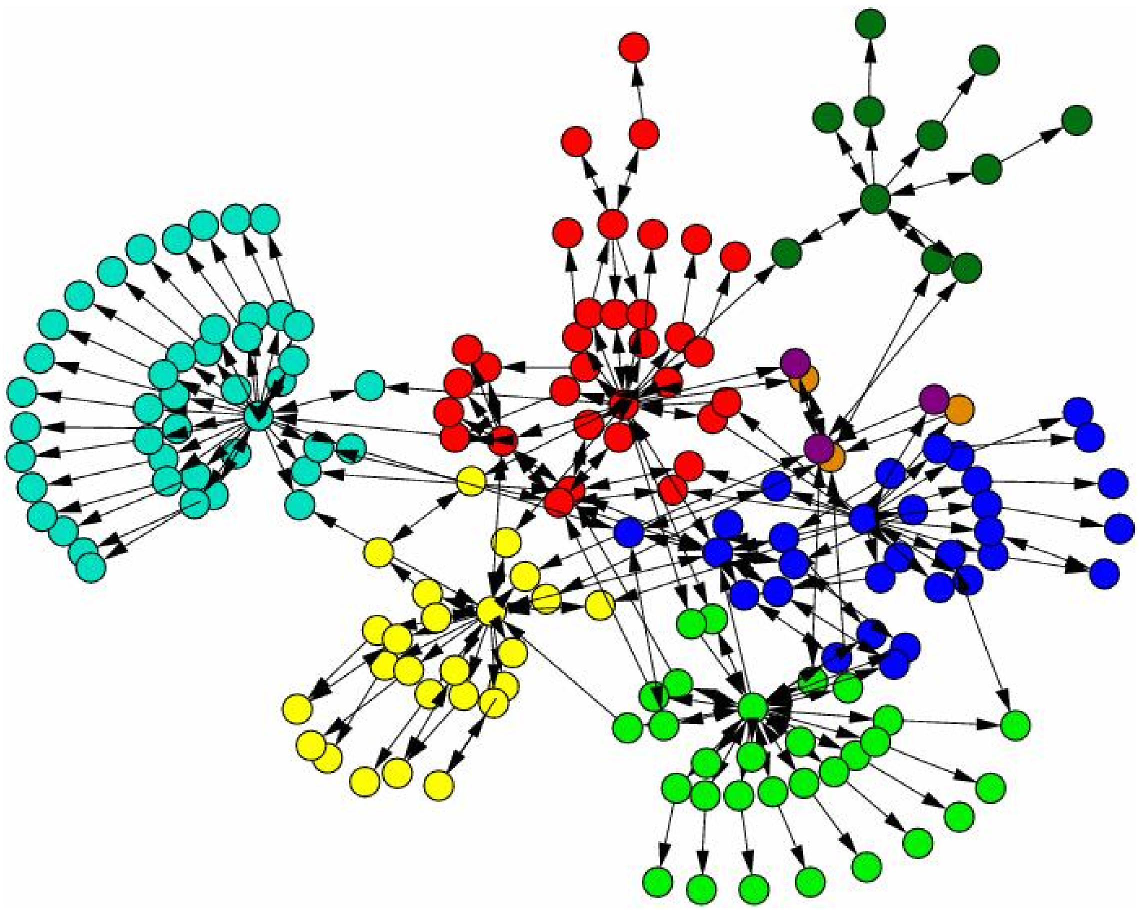
Random walk on the network has state transition matrix, P:



$$\begin{pmatrix} 1/4 & 1/3 & 1/2 & 1/4 & 0 \\ 1/4 & 1/3 & 0 & 1/4 & 0 \\ 1/4 & 0 & 1/2 & 0 & 0 \\ 1/4 & 1/3 & 0 & 1/4 & 1/2 \\ 0 & 0 & 0 & 1/4 & 1/2 \end{pmatrix} = P$$

The eigenvalues and eigenvectors convey much information.
Markov Chains, Spectral Gap.

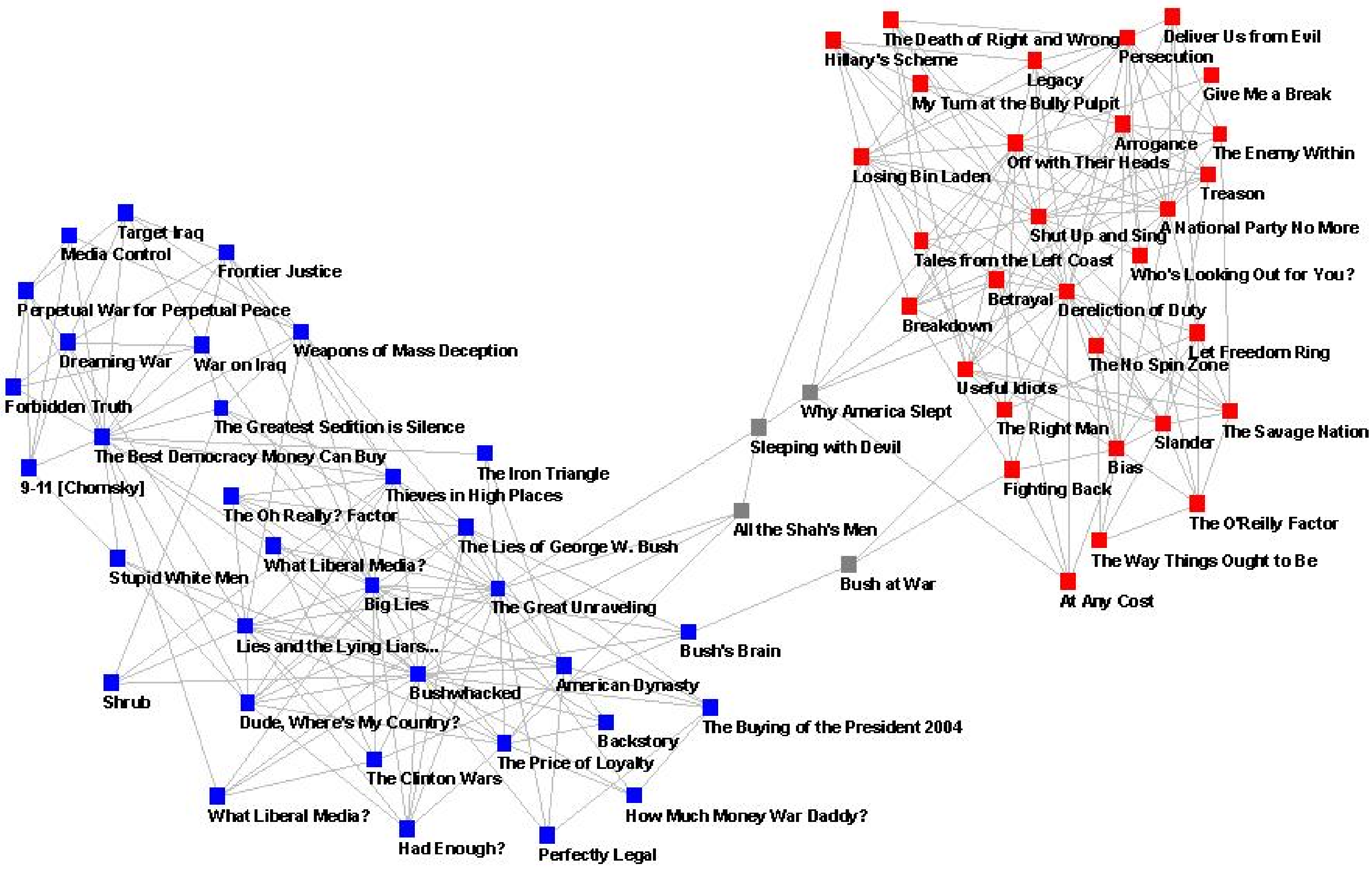
Random walk on the WWW is the "Page Rank"



Page Rank of a node is the steady-state random walk occupancy probabilty.

(We will discuss building a search engine in detail later.)

Example Eigen-technique: Community structure (Political Books 2004)



M. Girvan and M. E. J. Newman

Concepts covered today

- Social, physical and biological networks
- Simple network metrics (recapped next page)
- Random walks on networks
- Random graphs
- Phase transitions in connectivity
- Next time: Preferential attachment and network growth, Robustness, basic Internet structure, optimization.

Outstanding challenges

- How do we connect network **structure** to **function?**
  - Degree
  - Clustering Coefficient
  - Motifs
  - Betweeness Centrality
  - Assortativity
  - Flow and transport
  - Growth/evolution mechanisms.
- **Interacting networks**
- **Strategic interactions** / Game theory on networks

Sketchy outline of course

- Today: intro to different types of networks (physical, social, biological)
- Models of network topology:
  - random graphs
  - growth mechanisms
  - robustness and resilience
- Measures of network topology
- Processes on networks
  - Percolation
  - Epidemic spreading
  - Synchronization
  - Web search
- Optimization
  - User optimal versus system optimal
  - Braess's paradox
- Domain specifics and applications: CS, traffic, biology, social nets