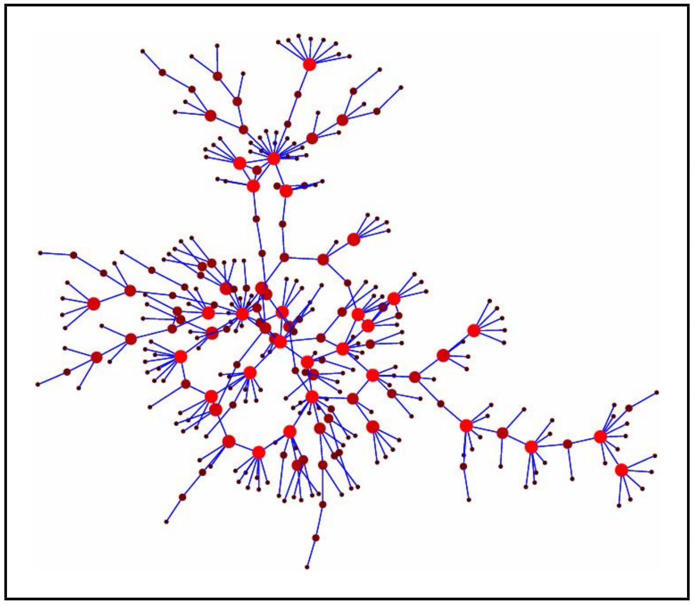
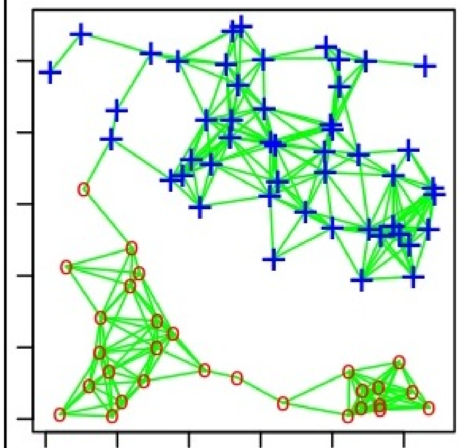
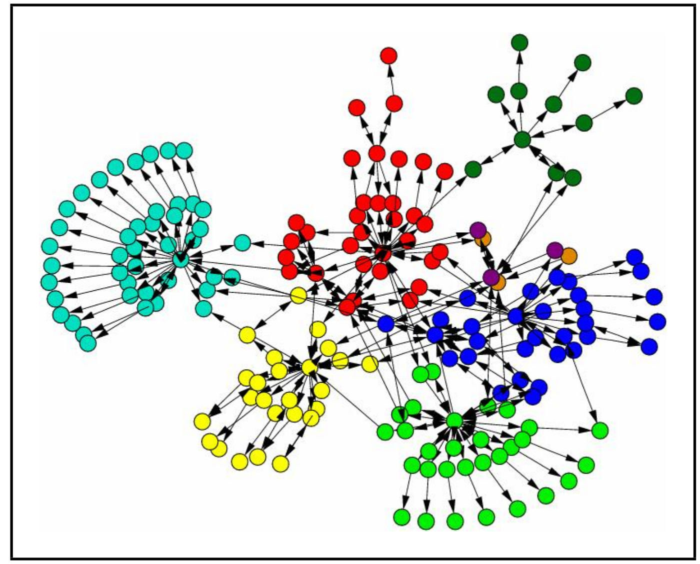
Overview of Network Theory, I







ECS 253 / MAE 253, Spring 2023, Lecture 1

Prof. Raissa D'Souza
University of California, Davis

Class enrollment

- 70 requested; 28 on waitlst – We cannot change class rooms

Many different graduate programs represented:

- Computer Science
- MAE
- Physics
- Applied Math
- Statistics
- Animal behavior
- Political Science

ECS/MAE 253 Network theory and applications

Course description:

Develops the mathematical theory underlying growth, structure and function of networks with applications to physical, social, biological and engineered systems. Topics include network growth, resilience, epidemiology, phase transitions, software and algorithms, routing and search control, cascading failures.

(This is not a computer networking course. Instead see ECE 273 Networking Architecture and Resource Management.)

Raissa's background:

- 1999, PhD, Physics, Massachusetts Inst of Tech:
  – 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)

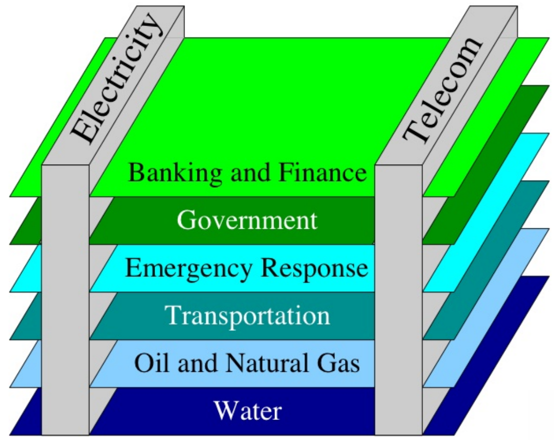
- Fall 2005-present, UC Davis:
  – Dept of Computer Science, Dept of Mech and Aerospace Eng., Grad Group Applied Math, Physics Grad Group, Biosystems Eng Grad Group.

- 2007-present, External Faculty Member, Santa Fe Institute.
  2018-onwards, Science Board Member.

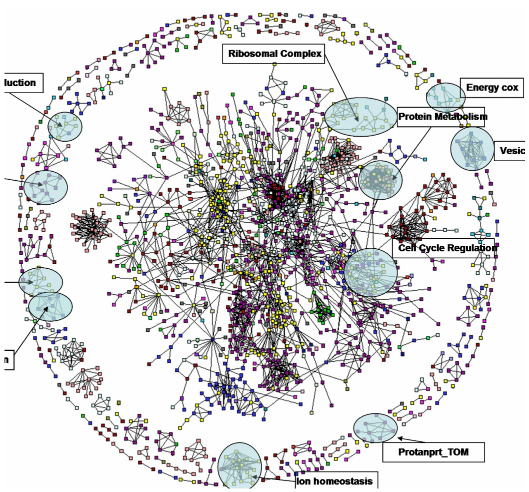
- July 2022-present, UC Davis:
  – Associate Dean of Research, COE

(email me at profdsouza@gmail.com)

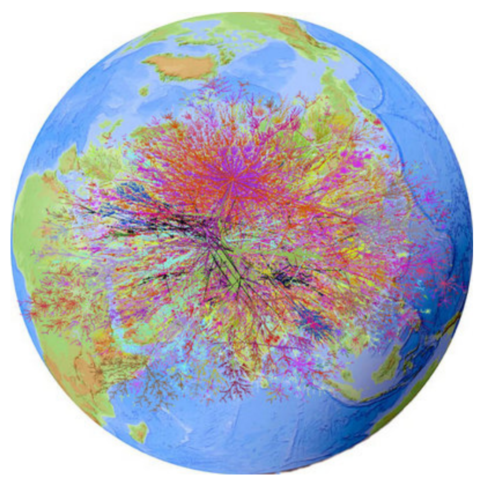
Complex networks are ubiquitous:



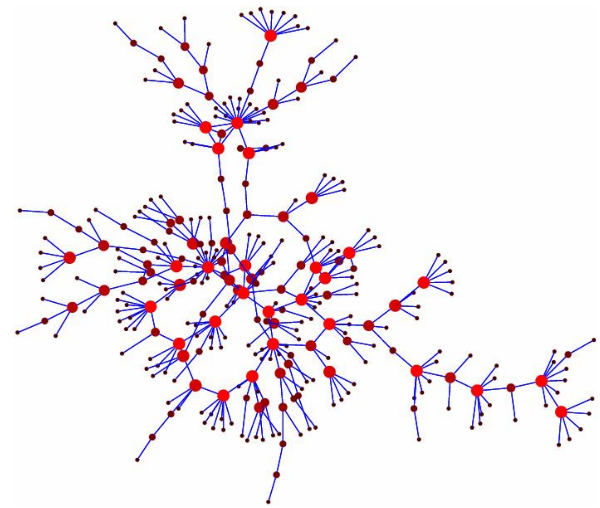
Critical Infrastructure



Biological & Ecological networks



Information and Communication technology



Social networks: Economics & Epidemics

What is a Network?

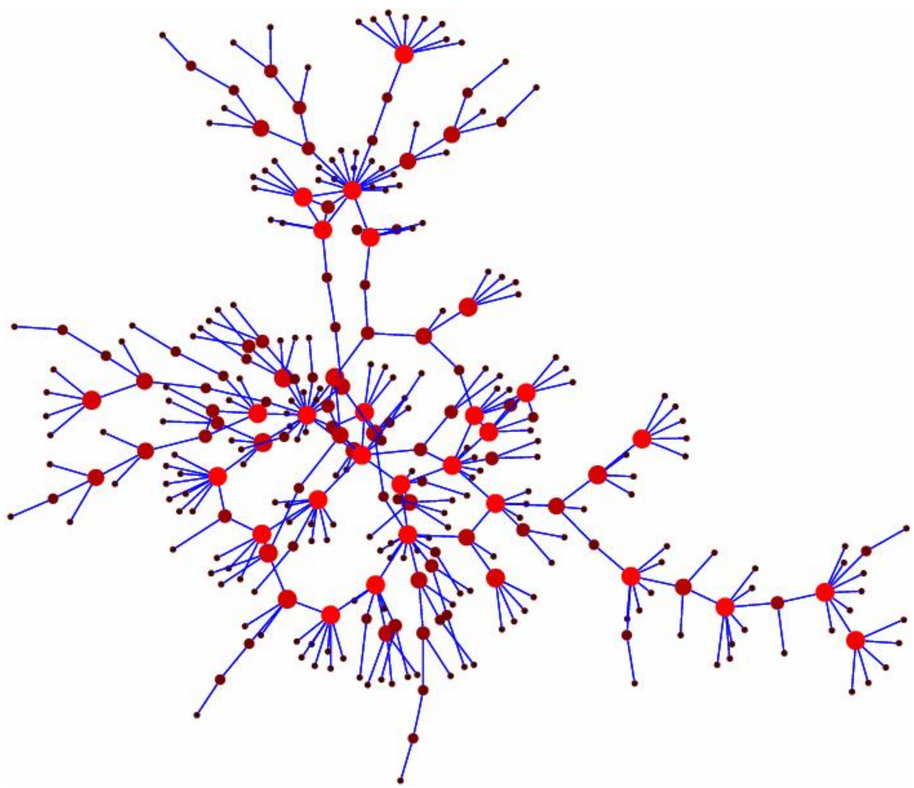
- Topology (i.e., structure: nodes/vertices and edges/links)
  Measures of topology

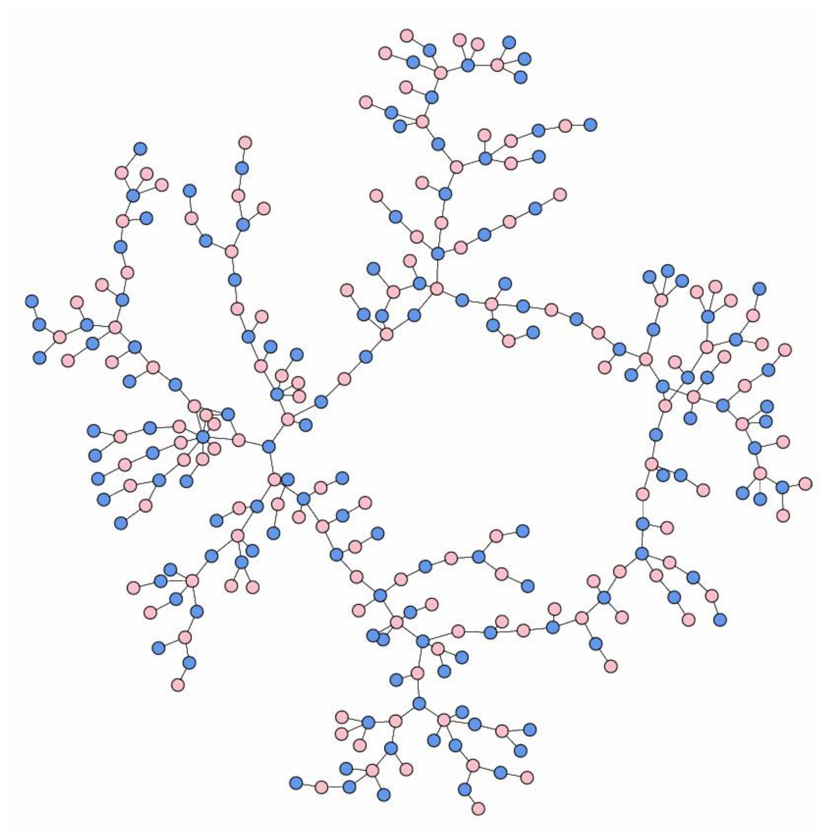
- Activity (i.e., function, processes on networks, dynamics of nodes and edges)

Modeling networks

- Network growth
- Phase transitions
- Algorithms: analysis, growth/formation, searching and spreading
- Processes on networks

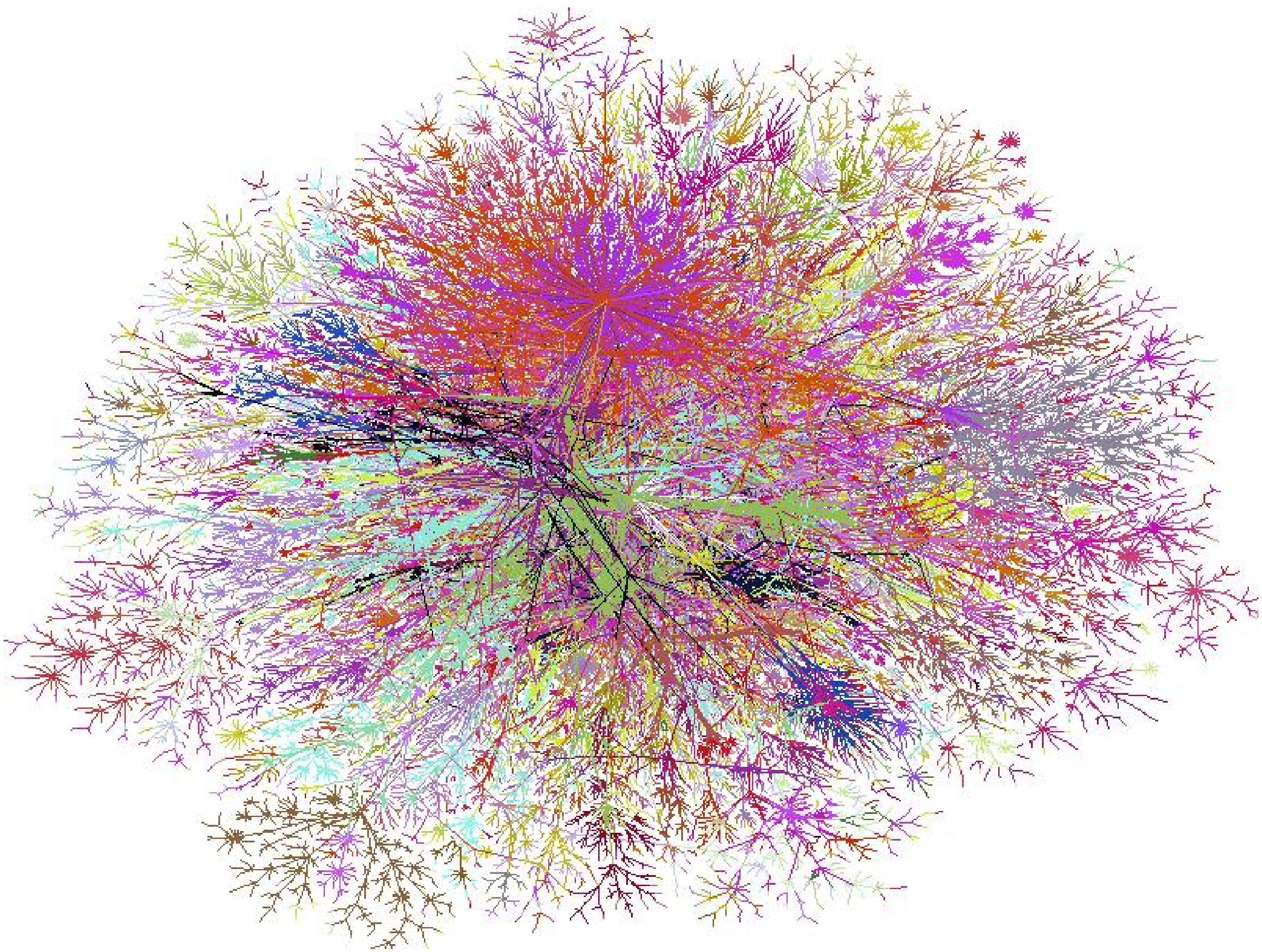
Example social networks
(Immunology; viral marketing; aliances/policy)





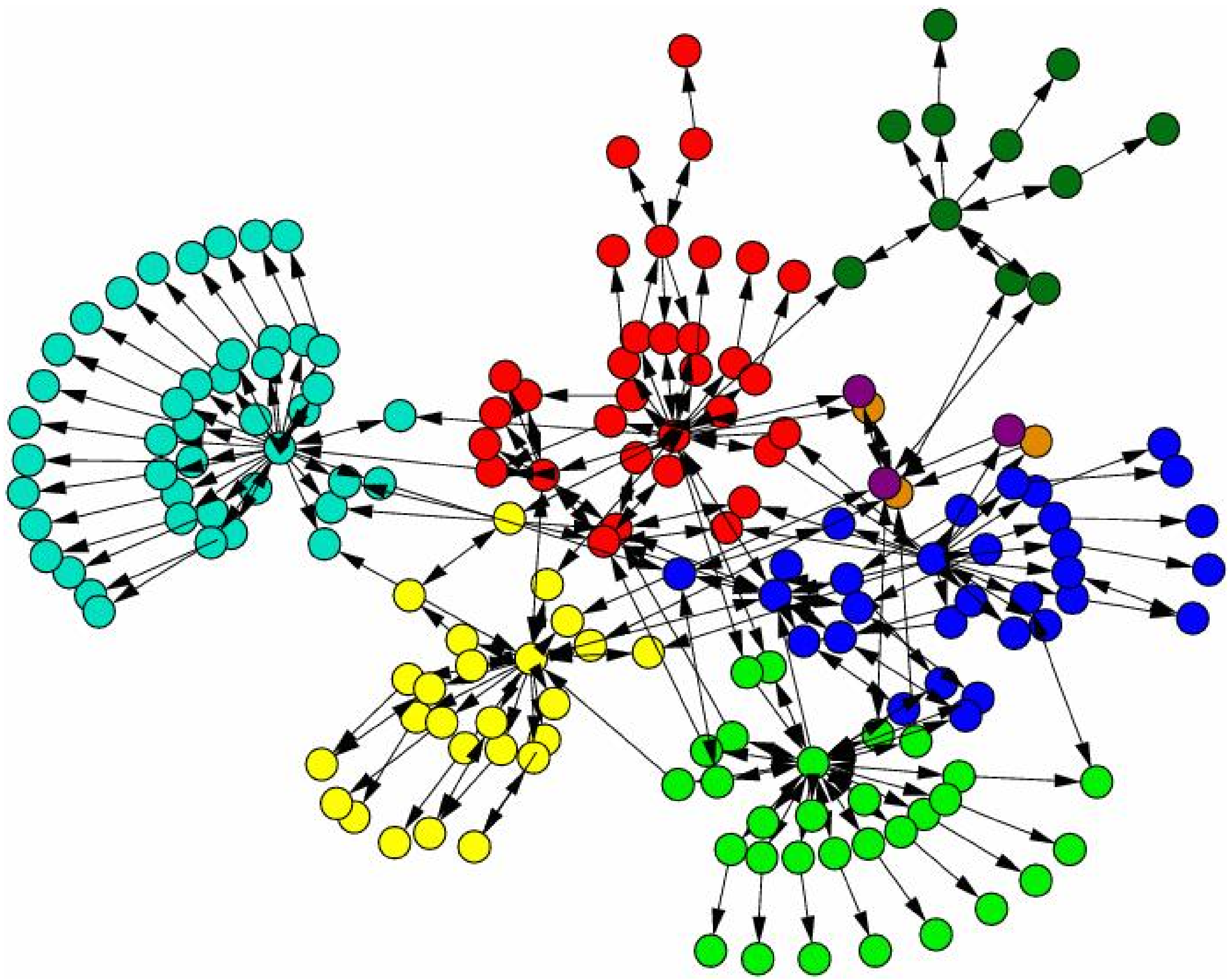
M. E. J. Newman

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



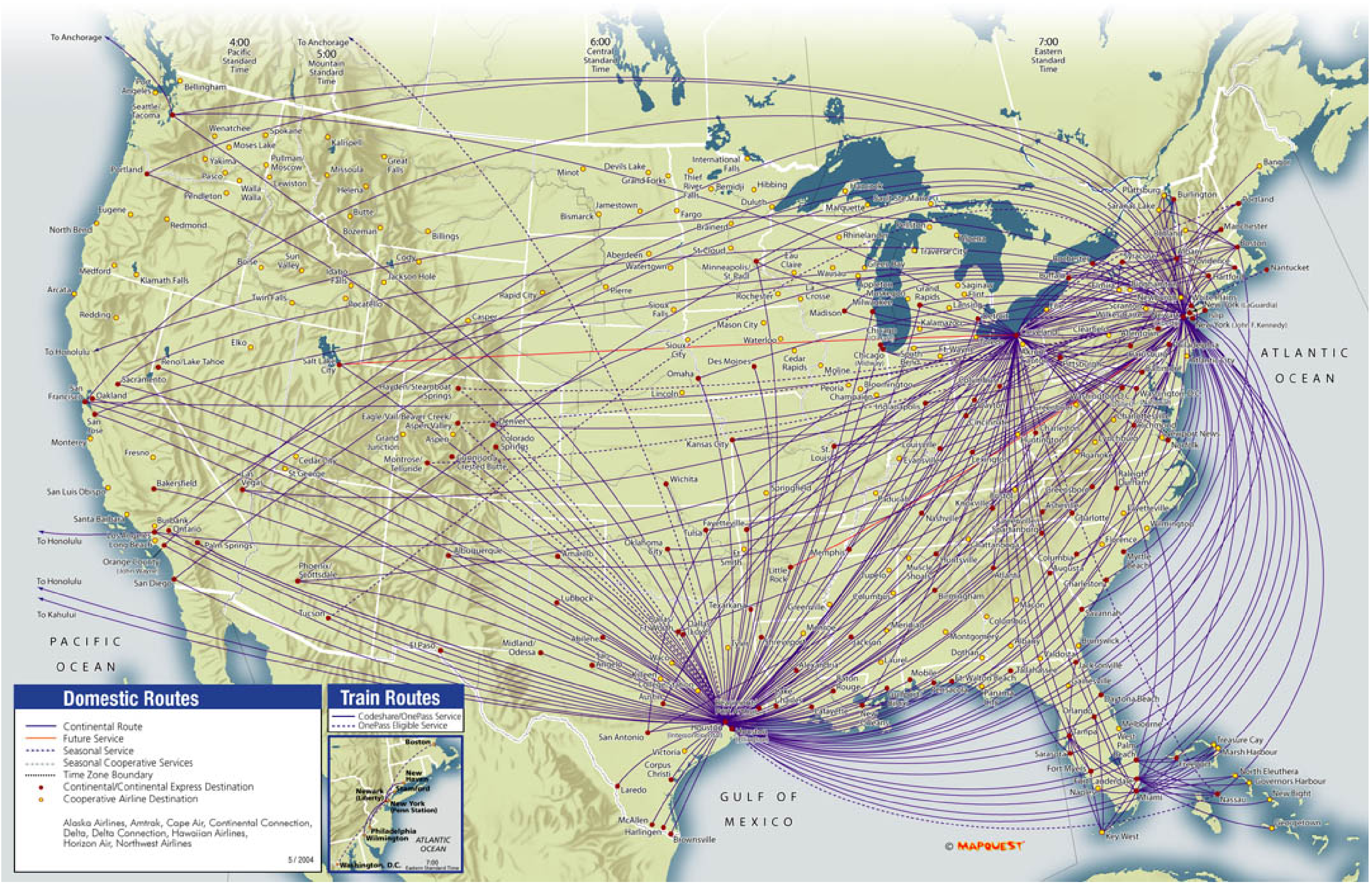
H. Burch and B. Cheswick

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



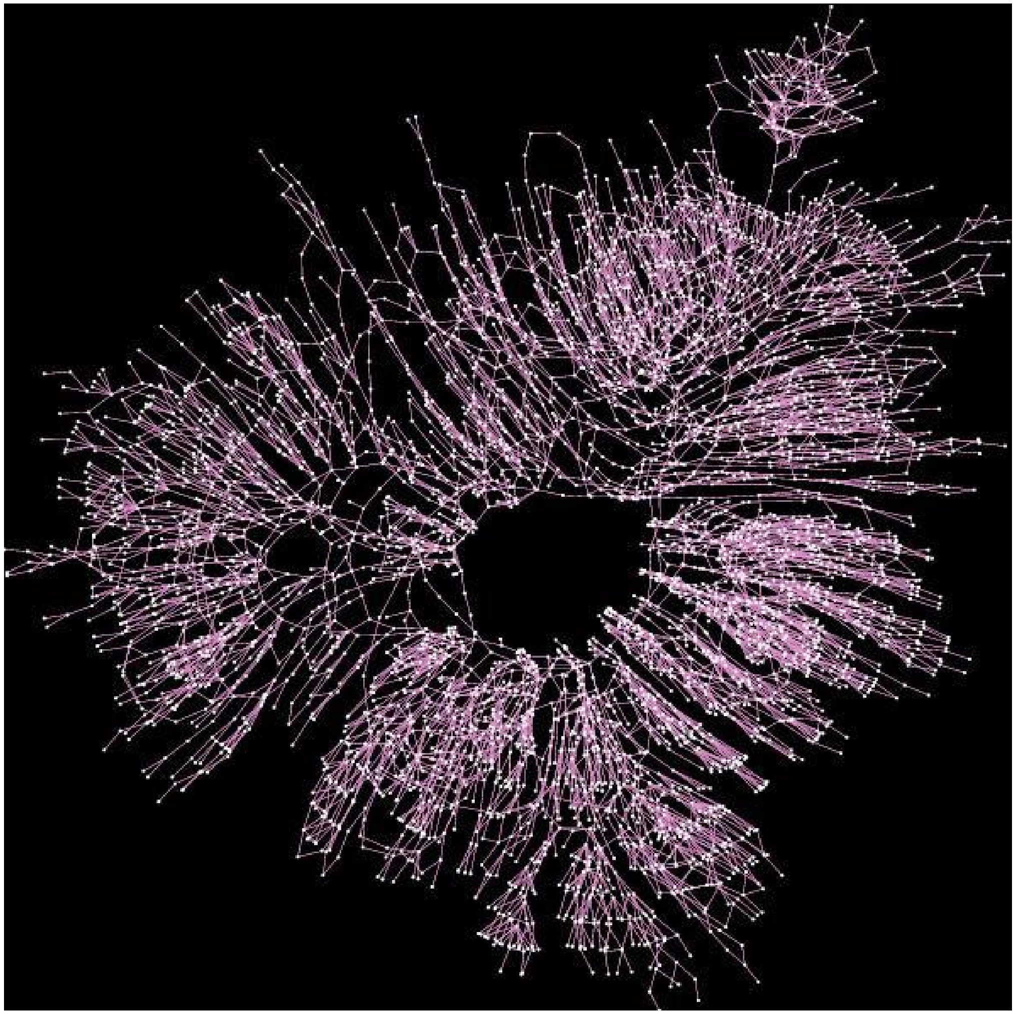
M. E. J. Newman

The airline network
(Optimization; dynamic external demands)



Continental Airlines

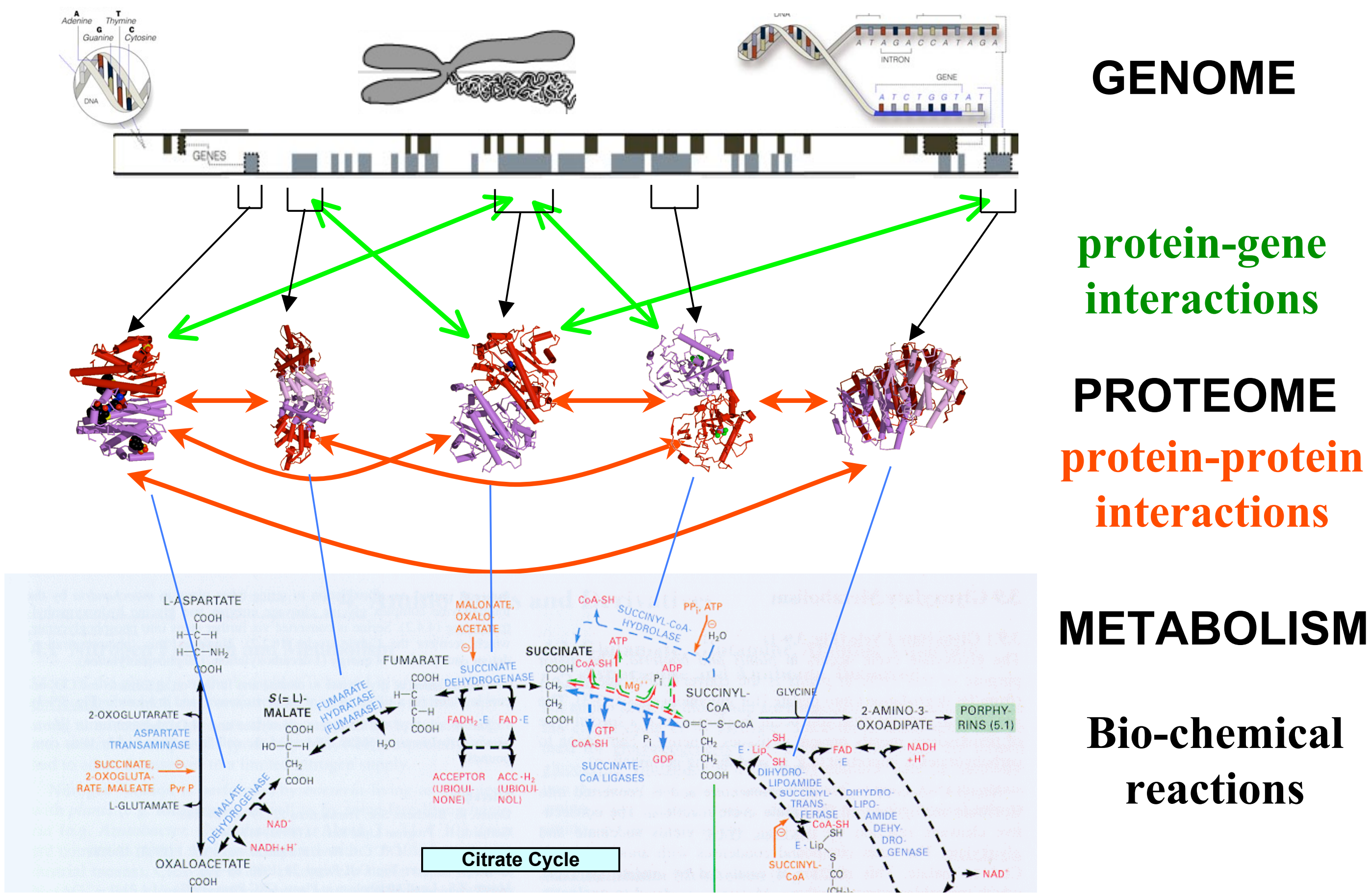
The power grid
(Mitigating failure; Distributed sources)



M. E. J. Newman

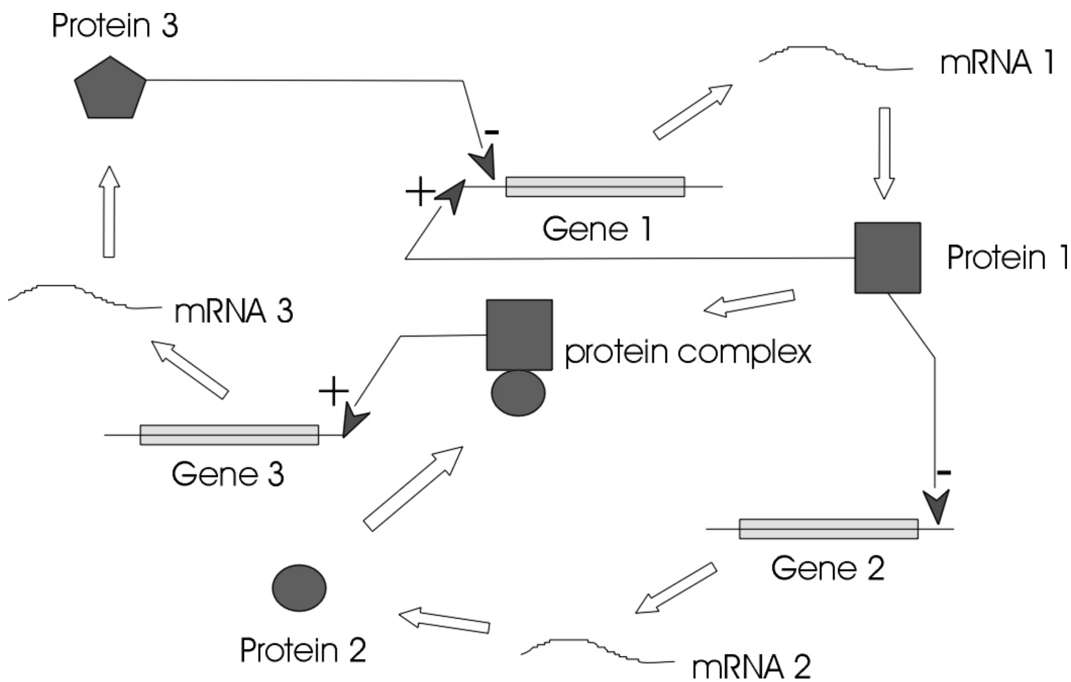
Biology: Networks at many levels

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



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

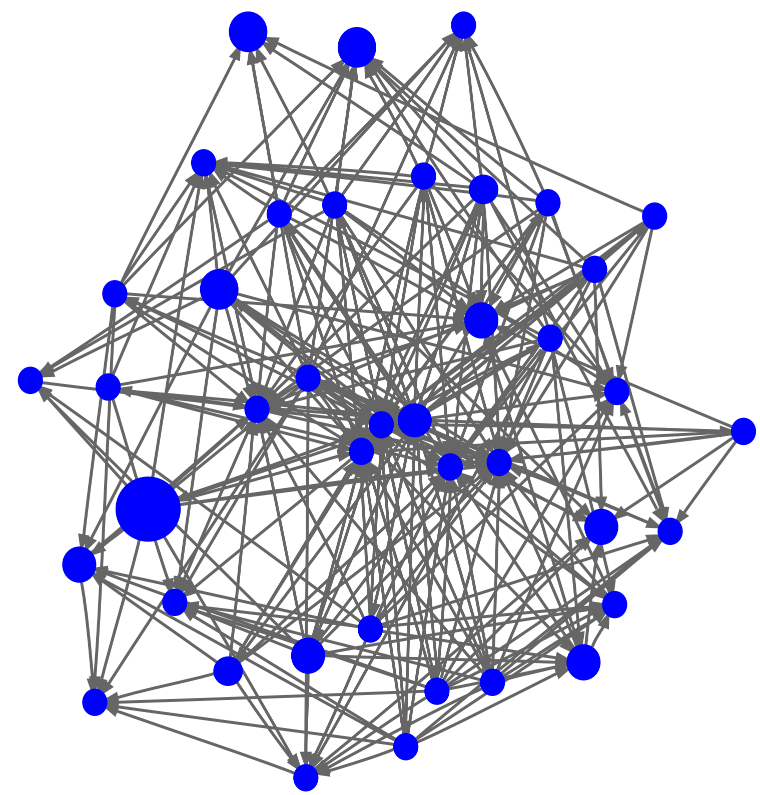
Software systems

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

Open-source software as a "systems" paradigm.

Networks:

- Function calls
- Email communication
- Socio-Technical congruence



The past decade, a "Science of Networks":

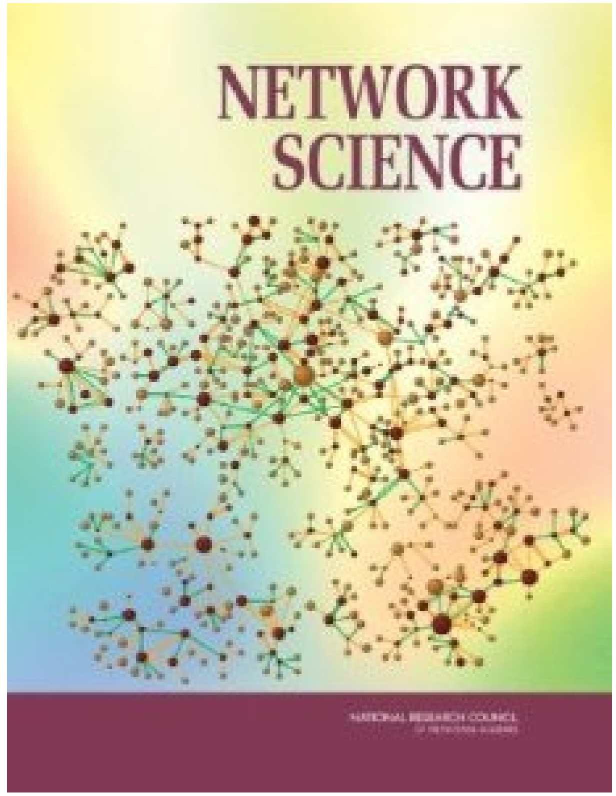
(Physical, Biological, Social)

- **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 $\sim \log(N)$).

Explosion of work and tools

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

Natn Acam Sciences/Natn Research Council Study (2005)

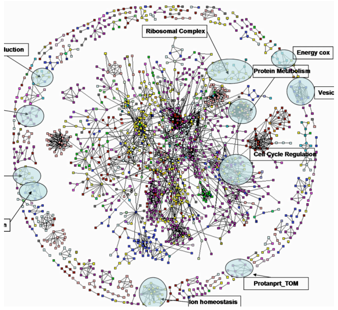


"all our modern critical infrastructure relies on networks... too much emphasis on specific applications/jargon/disciplinary stovepipes... need a cross-cutting science of networks... Research for the 21st century"

In reality a collection of interacting networks:

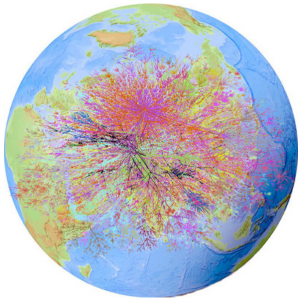
Networks:

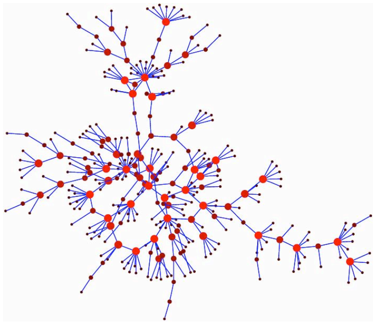
Transportation Networks/ Power grid
(distribution/ collection networks)



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))

THE HISTORY OF NETWORK ANALYSIS

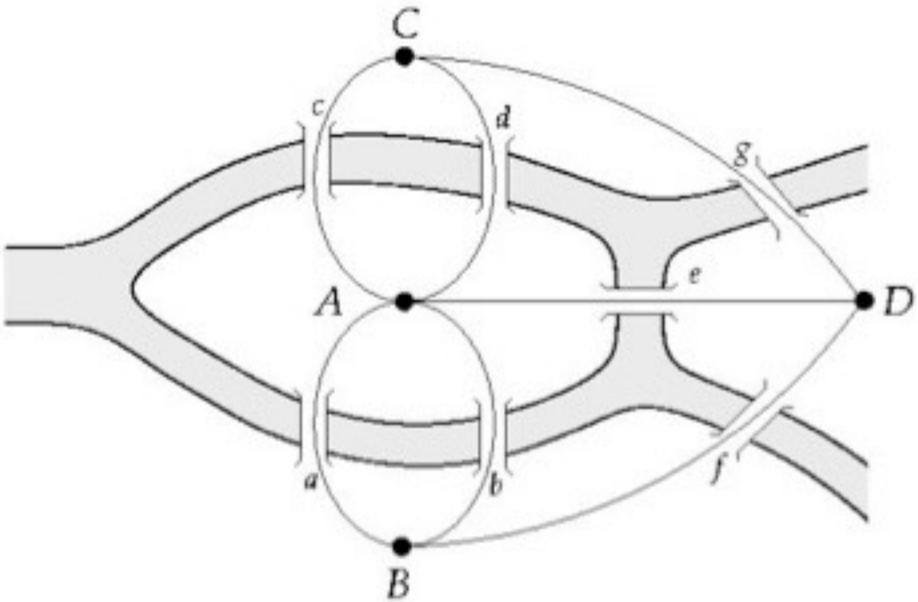
Graph theory: 1735, Euler

Social Network Research: 1930s, Moreno

Communication networks/internet: 1960s

Ecological Networks: May, 1979.

THE BRIDGES OF KONIGSBERG



CAN ONE WALK ACROSS THE SEVEN BRIDGES AND NEVER CROSS THE SAME BRIDGE TWICE?

1735: EULER'S THEOREM:

(A) IF A GRAPH HAS MORE THAN TWO NODES OF ODD DEGREE, THERE IS NO PATH.

(B) IF A GRAPH IS CONNECTED AND HAS NO ODD DEGREE NODES, IT HAS AT LEAST ONE PATH.

http://www.numericana.com/answer/graphs.htm

The founding of “network science”:

• 1998: Watts-Strogatz on **small world networ**ks is the most cited **Nature** publication from 1998; highlighted by ISI as one of the ten most cited papers in physics in the decade after its publication.

•1999: Faloutsos, Faloutsos, & Faloutsos, **ACM SIGCOMM** “**On power law relationships of the Internet topology**”.

•1999: Barabasi and Albert paper on **scale-free networ**ks is the most cited **Science** paper in 1999;highlighted by ISI as one of the ten most cited papers in physics in the decade after its publication.

•2001: Pastor -Satorras and Vespignani on **epidemic spreading** is one of the two most cited papers among the papers published in 2001 by **Physical Review Letters**.

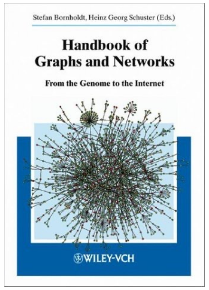
•2002: Girvan-Newman on **community structur**e is the most cited paper in 2002 **Proceedings of the National Academy of Sciences**.

Complex systems and networks.

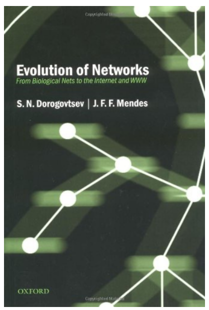
•**Science**:

Special Issue for the 10 year anniversary of Barabasi & Albert 1999 paper.

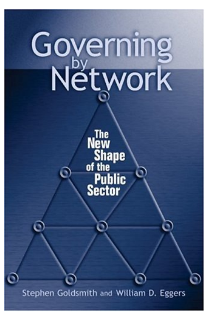
BOOKS



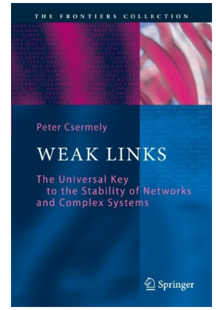
Handbook of Graphs and Networks: From the Genome to the Internet (Wiley-VCH, 2003).



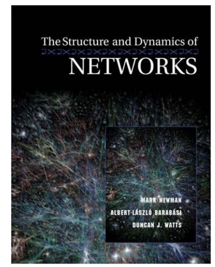
S. N. Dorogovtsev and J. F. F. Mendes, Evolution of Networks: From Biological Nets to the Internet and WWW (Oxford University Press, 2003).



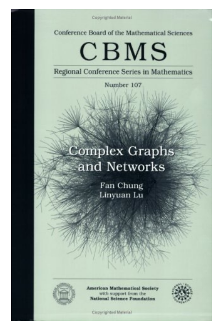
S. Goldsmith, W. D. Eggers, Governing by Network: The New Shape of the Public Sector (Brookings Institution Press, 2004).



P. Csermely, Weak Links: The Universal Key to the Stability of Networks and Complex Systems (The Frontiers Collection) (Springer, 2006), rst edn.

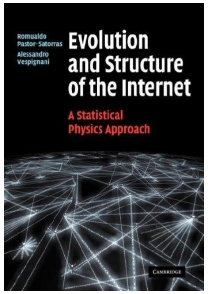


M. Newman, A.-L. Barabasi, D. J. Watts, The Structure and Dynamics of Networks: (Princeton Studies in Complexity) (Princeton University Press, 2006), rst edn.

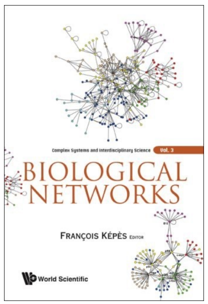


L. L. F. Chung, Complex Graphs and Networks (CBMS Regional Conference Series in Mathematics) (American Mathematical Society, 2006).

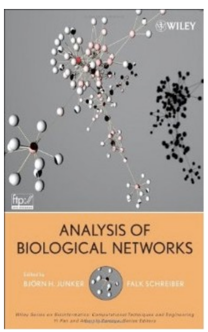
BOOKS



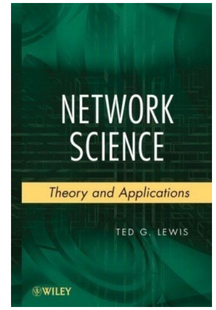
R. Pastor-Satorras, A. Vespignani, Evolution and Structure of the Internet: A Statistical Physics Approach (Cambridge University Press, 2007), rst edn.



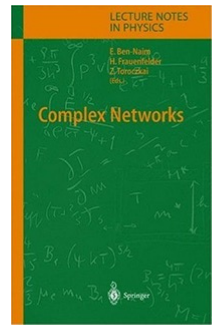
F. Kopos, Biological Networks (Complex Systems and Interdisciplinary Science) (World Scientic Publishing Company, 2007), rst edn.



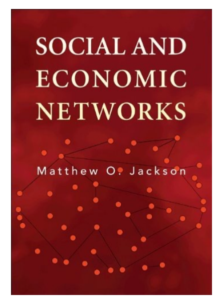
B. H. Junker, F. Schreiber, Analysis of Biological Networks (Wiley Series in Bioinformatics) (Wiley-Interscience, 2008).



T. G. Lewis, Network Science: Theory and Applications (Wiley, 2009).

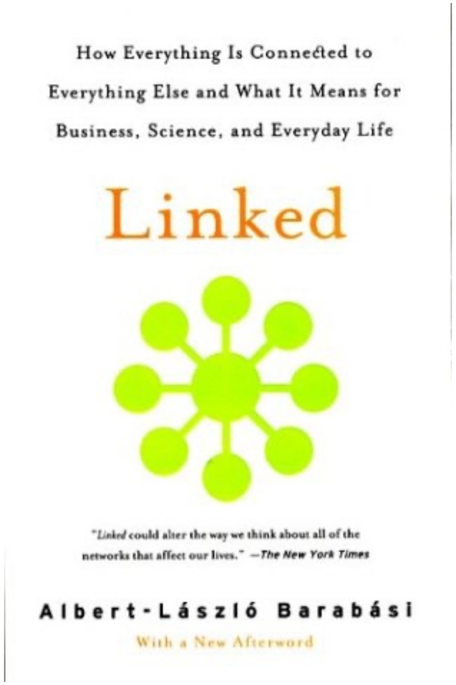


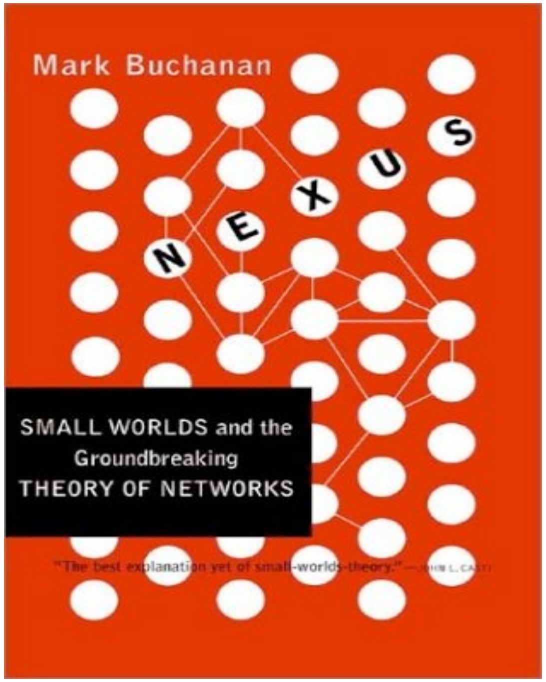
E. Ben Naim, H. Frauenfelder, Z.Torotzai, Complex Networks (Lecture Notes in Physics) (Springer, 2010), rst edn.

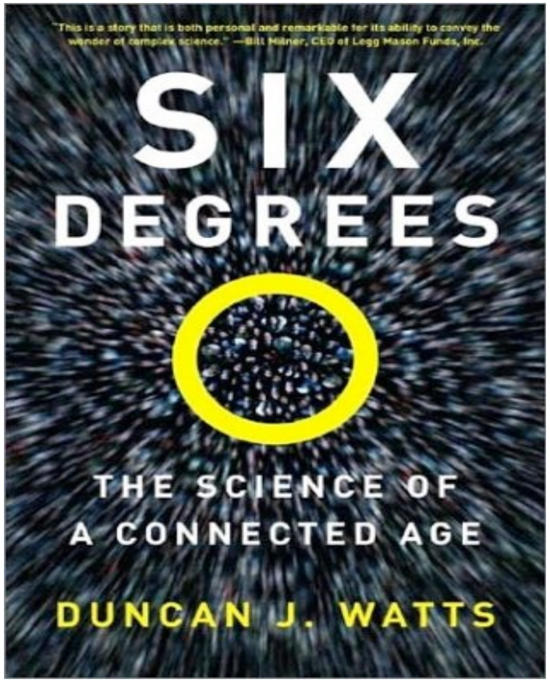


M. O. Jackson, Social and Economic Networks (Princeton University Press, 2010).

GENERAL AUDIENCE

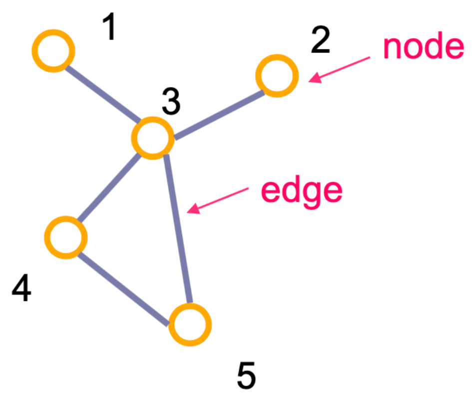






What are networks?

- 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

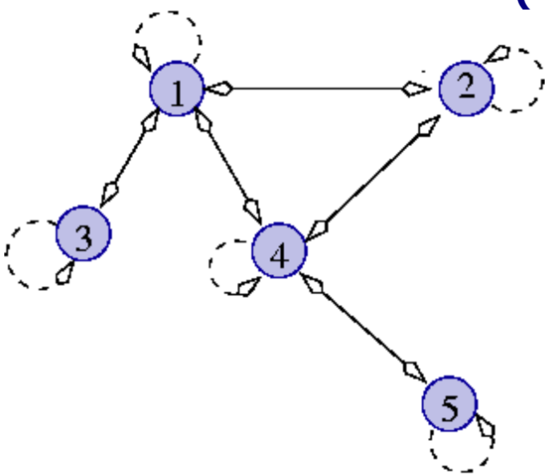
Slide from Adamic's course

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

NETWORK TOPOLOGY

Connectivity matrix, M:

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



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

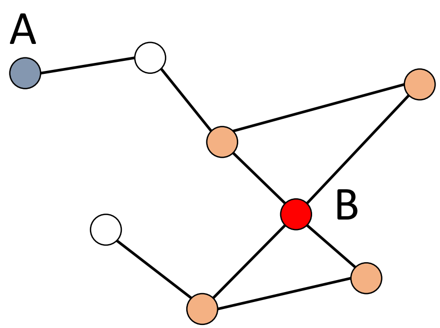
Node **degree** is number of links.

Typical measures of network topology

1) Node degree
2) Degree distribution
3) Connectivity
4) Clustering coefficient
5) Diameter
6) Betweenness centrality
7) Assortative/dissortative mixing

1) NODE DEGREES

Undirected

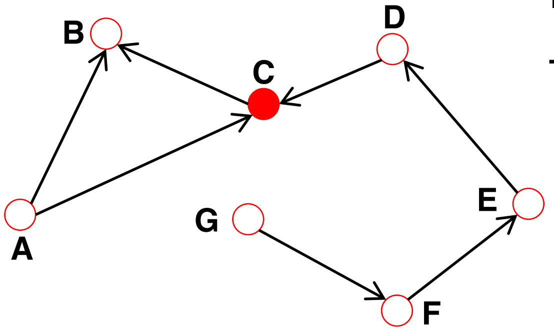


Node degree: the number of links connected to the node.

$$k_A = 1 \qquad k_B = 4$$

Undirected links :
coauthorship links
Actor network
protein interactions

Directed



In *directed networks* we can define an in-degree and out-degree.

The (total) degree is the sum of in- and out-degree.

$$k_C^{in} = 2 \qquad k_C^{out} = 1 \qquad k_C = 3$$

Source: a node with $k^{in} = 0$; Sink: a node with $k^{out} = 0$.

Directed links :
URLs on the www
phone calls
metabolic reactions

A BIT OF STATISTICS

BRIEF STATISTICS REVIEW

Four key quantities characterize a sample of N values $x_1, \dots, x_N$:

Average (mean):

$$\langle x \rangle = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

The n^{th} moment:

$$\langle x^n \rangle = \frac{x_1^n + x_2^n + \dots + x_N^n}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i^n$$

Standard deviation:

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \langle x \rangle\right)^2} = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$$

Distribution of x:

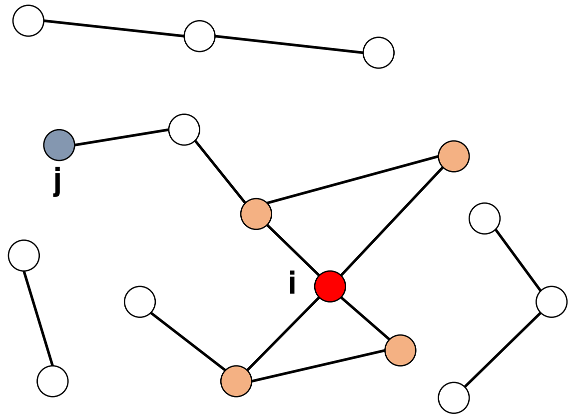
$$p_x = \frac{1}{N}\sum_i \delta_{x,x_i}$$

where p_x follows

$$\sum_i p_x = 1 \left(\int p_x \, dx = 1 \right)$$

AVERAGE DEGREE

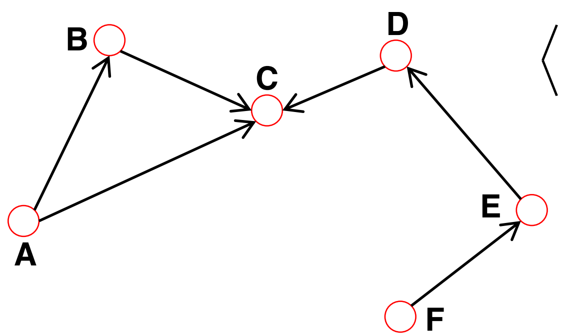
Undirected



$$\langle k \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} k_i \qquad \langle k \rangle \equiv \frac{2L}{N}$$

N – the number of nodes in the graph

Directed



$$\langle k^{in} \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N} \sum_{i=1}^{N} k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

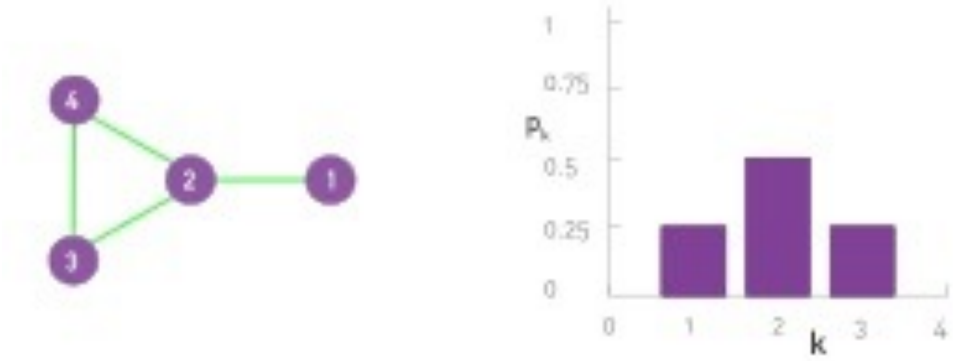
$$\langle k \rangle \equiv \frac{L}{N}$$

Average Degree

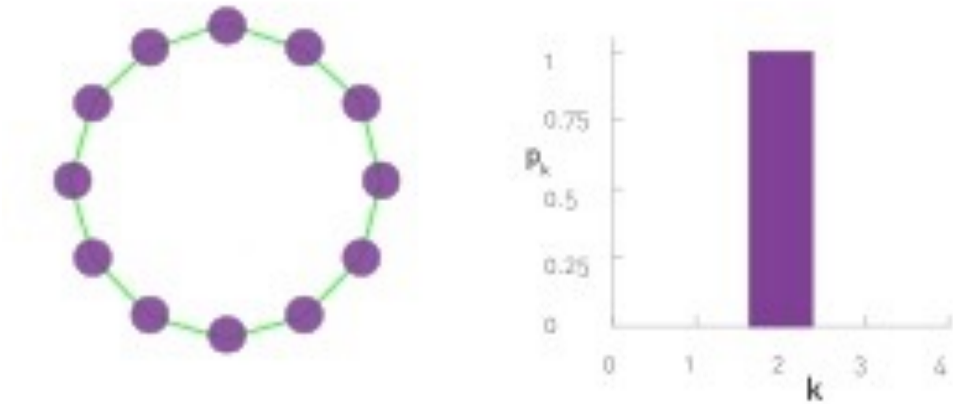
NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

Degree distribution

P(k): probability that a randomly chosen node has degree *k*

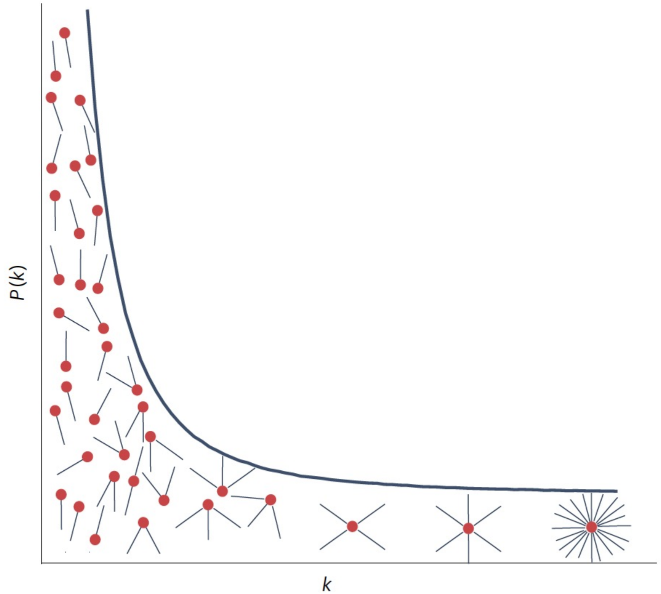


N_k = # nodes with degree k

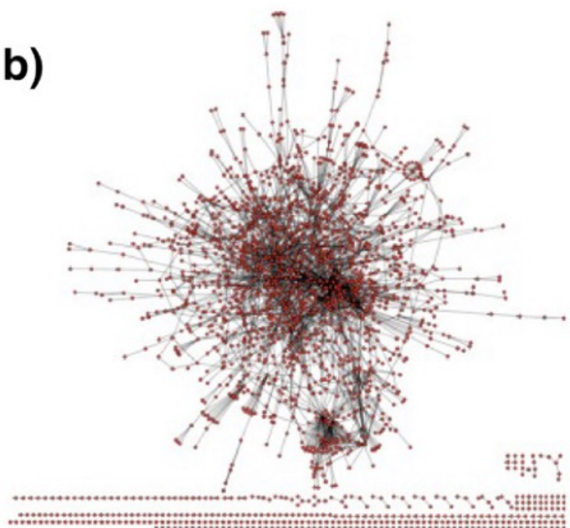


$P(k) = N_k / N$ ➔ plot

DEGREE DISTRIBUTION



k



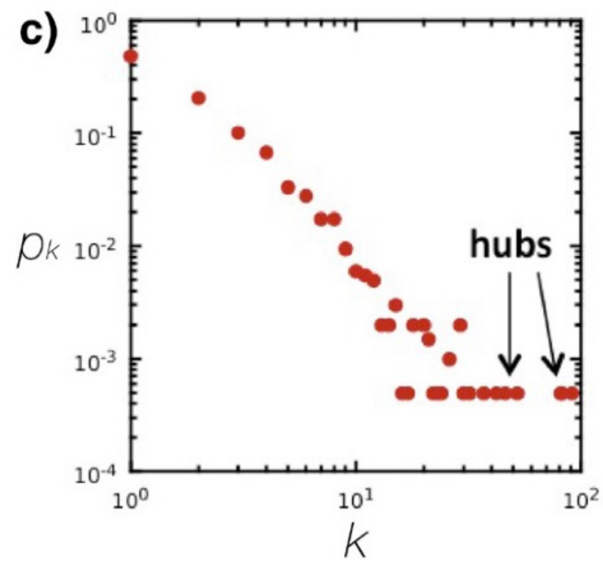


Image 2.4b

DEGREE DISTRIBUTION

Discrete Representation: p_k is the probability that a node has degree **k**.

Continuum Description: **p(k)** is the pdf of the degrees, where

$$\int_{k_1}^{k_2} p(k)dk$$

represents the probability that a node's degree is between k_1 and k_2.

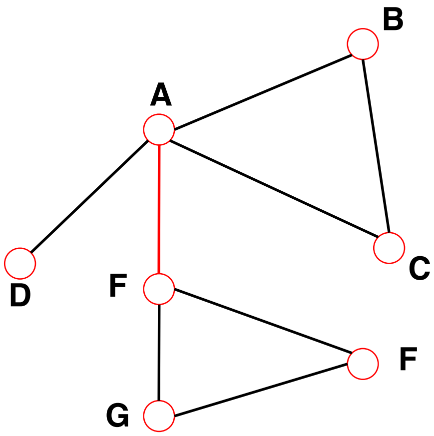
Normalization condition:

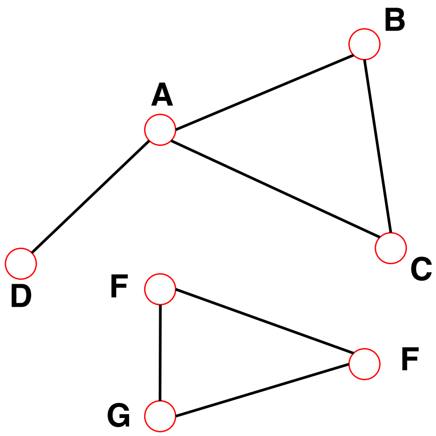
$$\sum_{0}^{\infty} p_k = 1 \qquad \int_{K_{\min}}^{\infty} p(k)dk = 1$$

where K_{min} is the minimal degree in the network.

3) CONNECTIVITY OF UNDIRECTED GRAPHS

Connected (undirected) graph: any two vertices can be joined by a path.
A disconnected graph is made up by two or more connected components.





Largest Component:
Giant Component

The rest: **Isolates**

Bridge: if we erase it, the graph becomes disconnected.

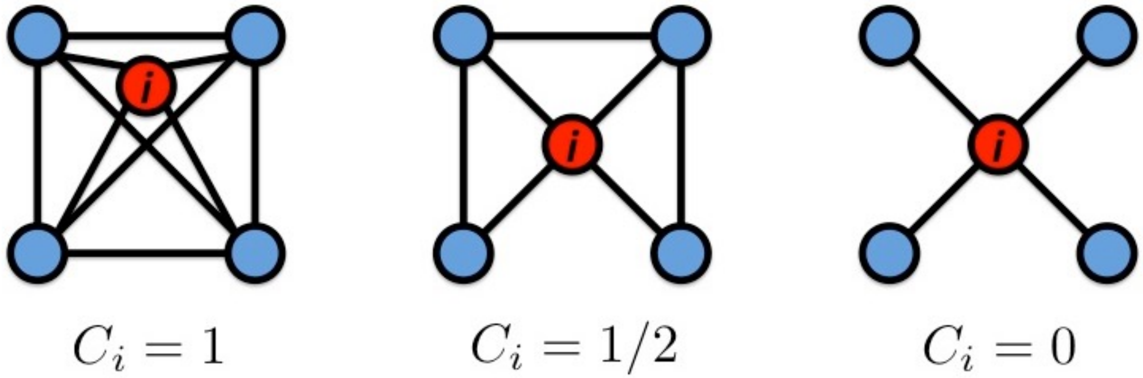
*Clustering coefficient:

what fraction of your neighbors are connected?

* Node i with degree k_i

* C_i in [0,1]

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$



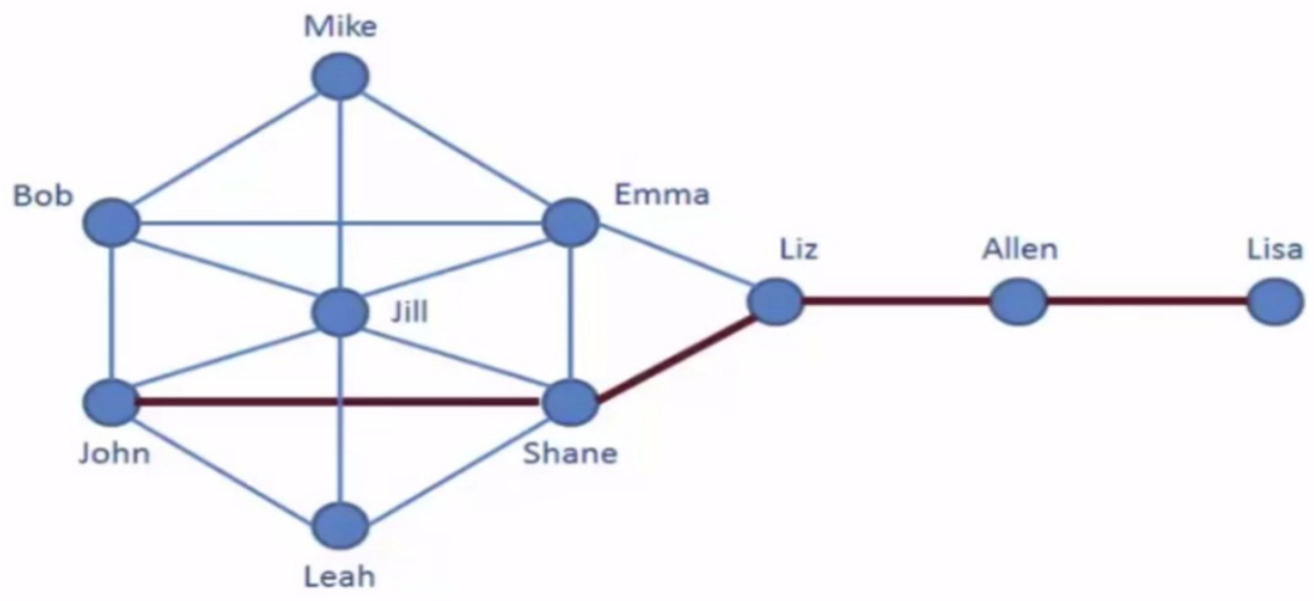
Watts & Strogatz, Nature 1998.

5) DIAMETER

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

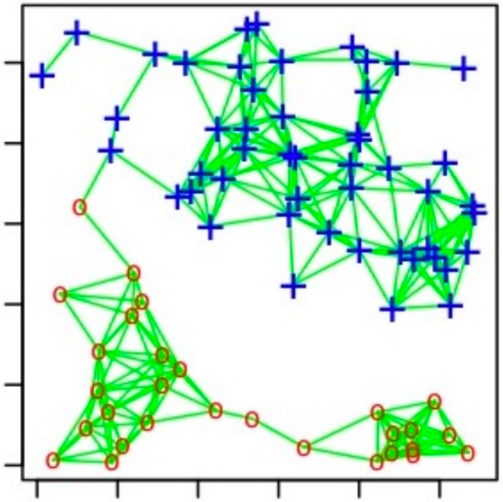
  "Small world" if $d \sim \log N$ and strong clustering.

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



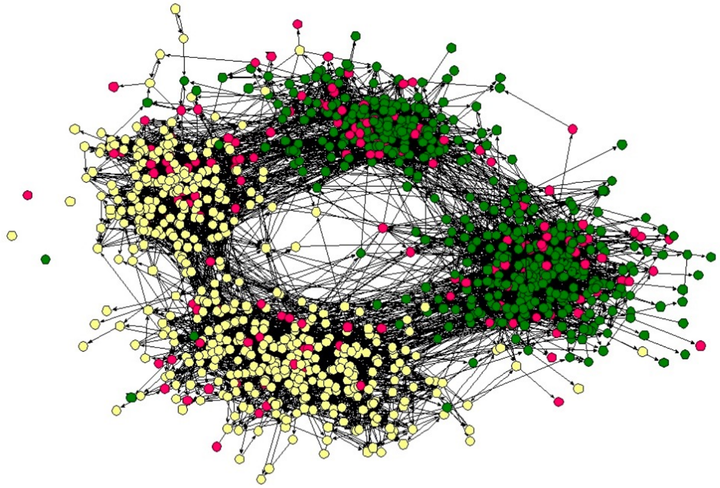
6) BETWEENNESS CENTRALITY

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



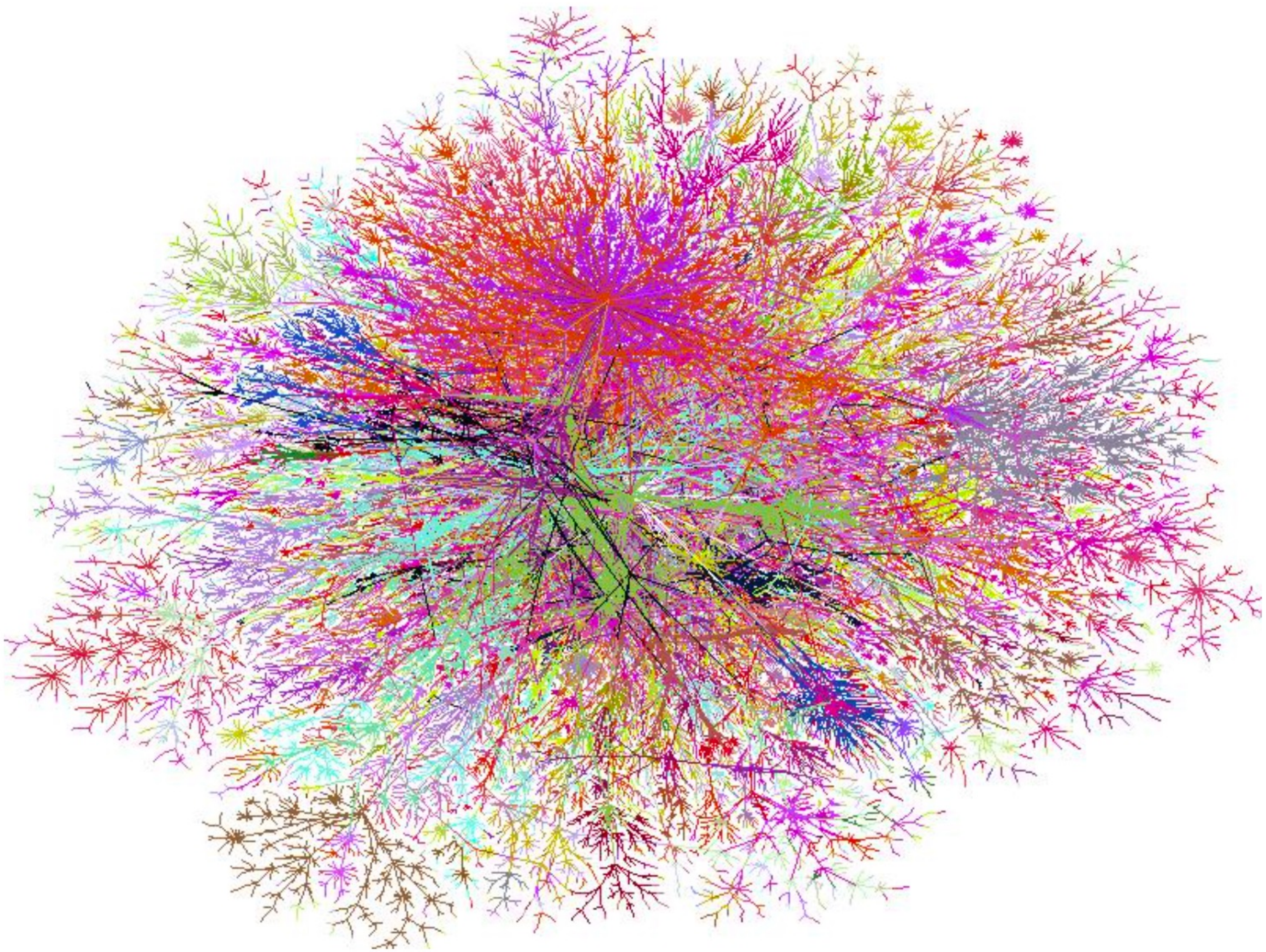
7) ASSORTATIVE/DISSORTATIVE MIXING

- **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



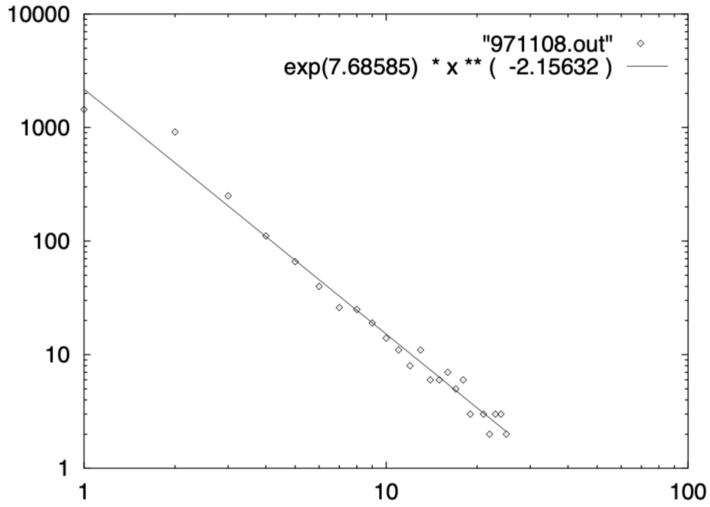
Extremely broad range of node degree observed: from biological, to technological, to social.

Typical distribution in node degree

The "Internet"

Faloutsos^3, *SIGCOMM* 1999

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

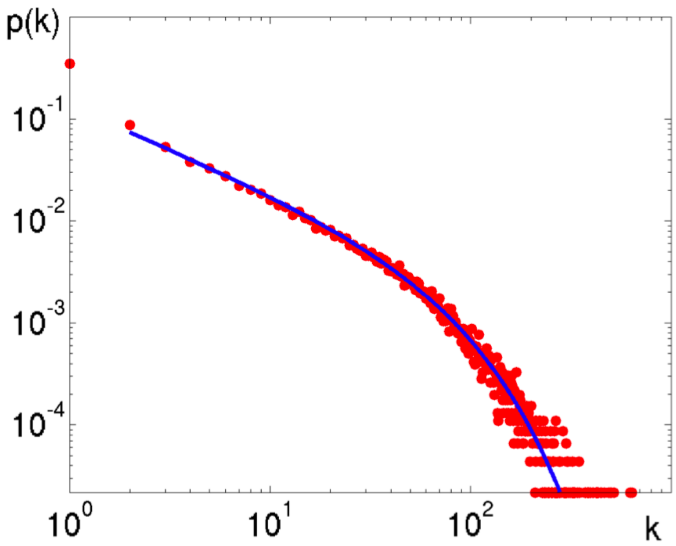


(a) Int-11-97

"Who-is-Who" network

Szendröi and Csányi

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



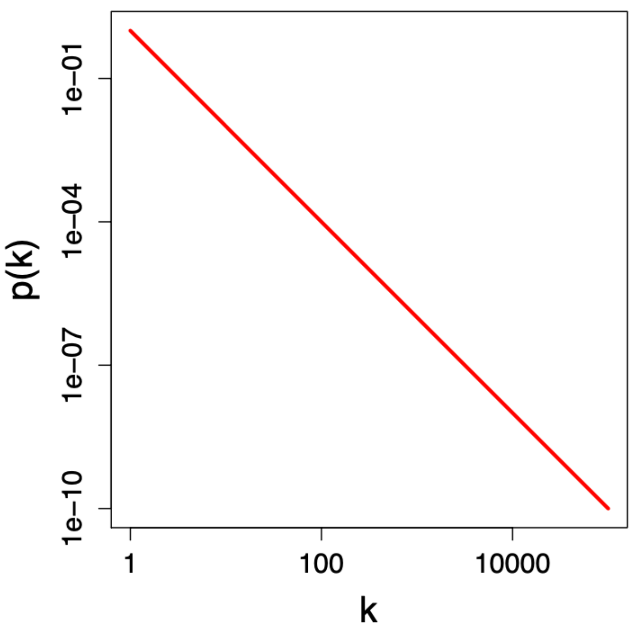
- Small data sets, power laws vs other similar distributions?
  - What is the "Internet"/ what level? (e.g., router vs AS)

What is a power law?

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

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

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



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

Sketch 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