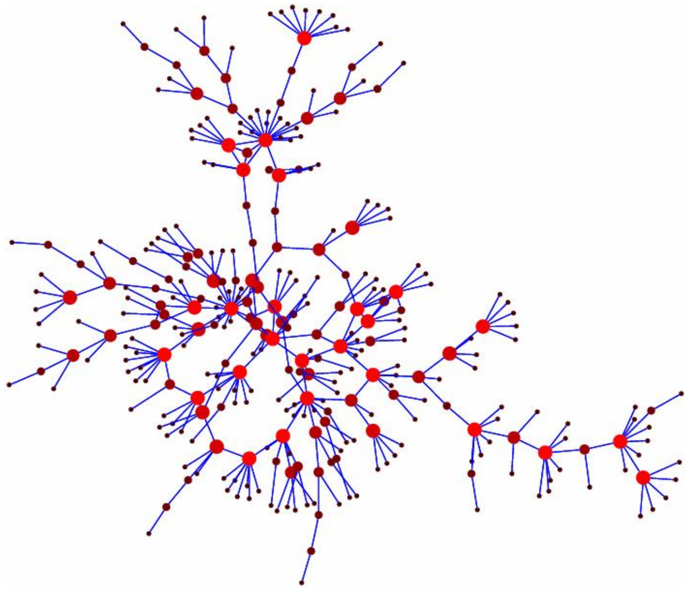
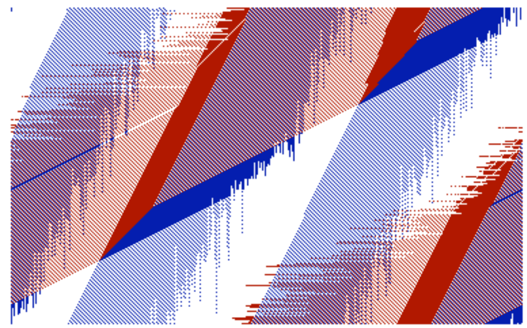
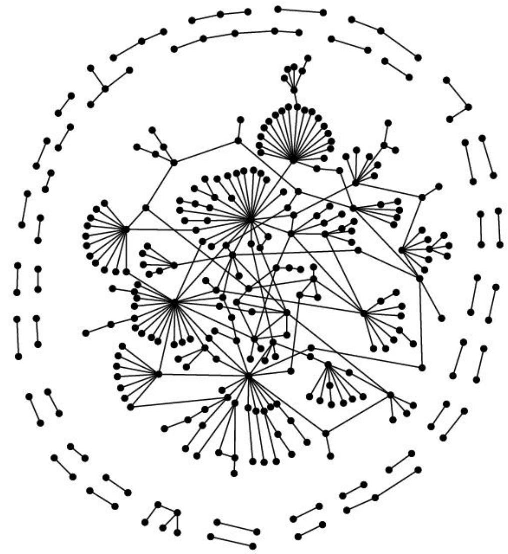
"What is a network? Topology and Activity"







MAE 298, Lecture 1

Jan 7, 2008

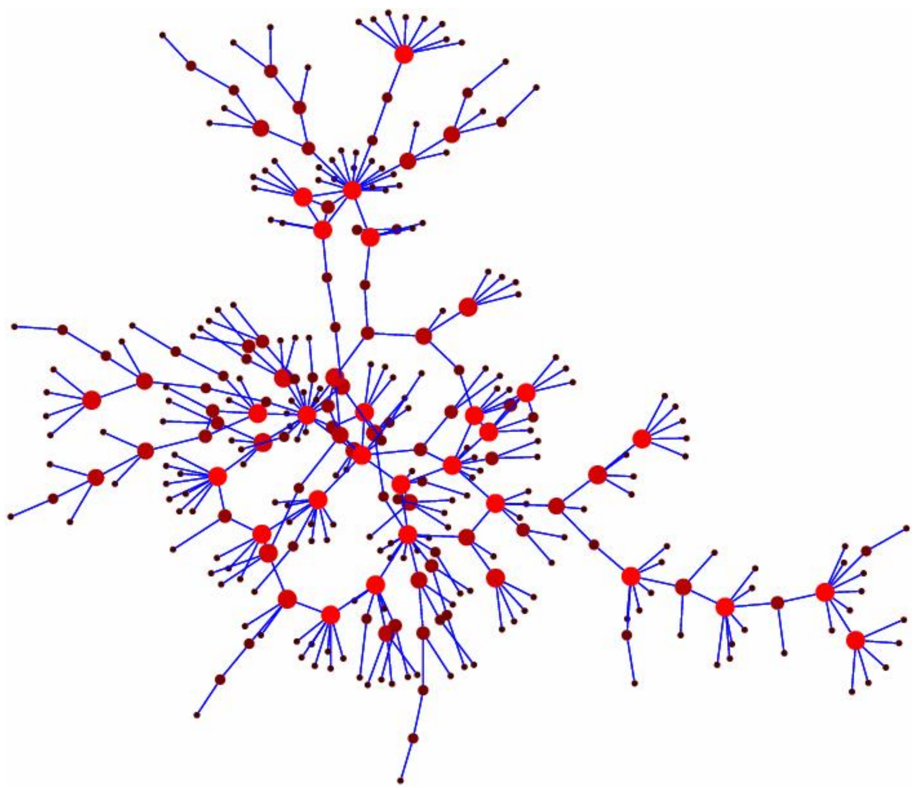
Raissa's Professional history: i.e., (How did I get here?)

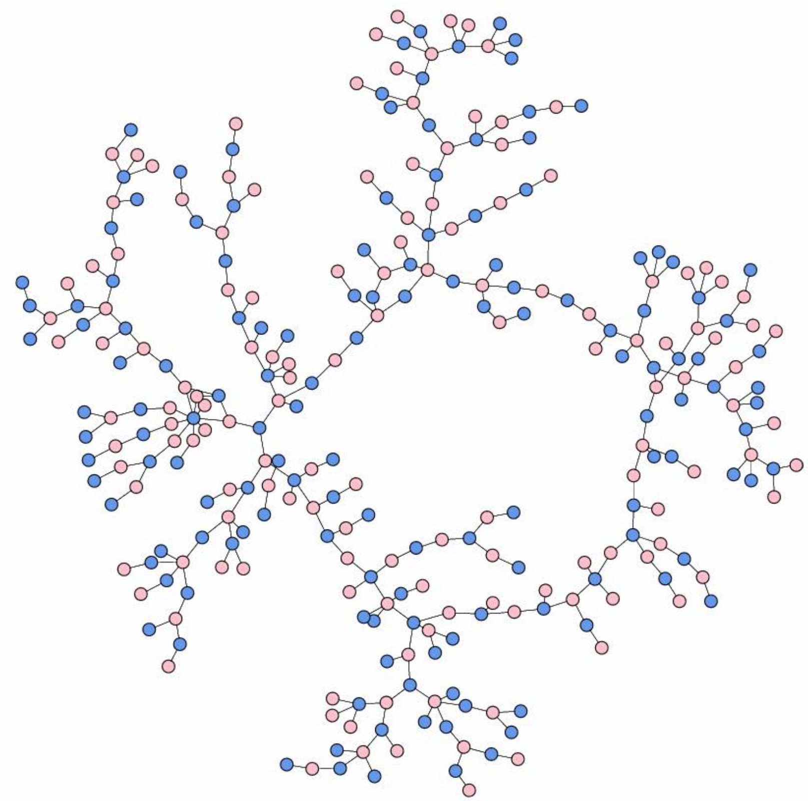
- 1999, PhD, Physics, Massachusetts Inst of Tech (MIT):
  - Joint appointment: Statistical Physics and Lab for Computer Science
- 2000-2002, Postdoctoral Research Fellow, Bell Laboratories:
  - Joint appointment: Fundamental Mathematics and Theoretical Physics Research Groups.
- 2002-2005, Postdoctoral Research Fellow, Microsoft Research:
  - "Theory Group" (Interdisciplinary group in Physics and Theoretical Computer Science)
- 2005-present, Assistant Professor, UC Davis:
  - Dept of Mechanical and Aeronautical Eng., and Center for Computational Science and Eng.
- 2007-present, External Faculty Member, Santa Fe Institute

What is a Network?

- Topology (i.e., structure)
- Activity (i.e., function)
- Phase transitions

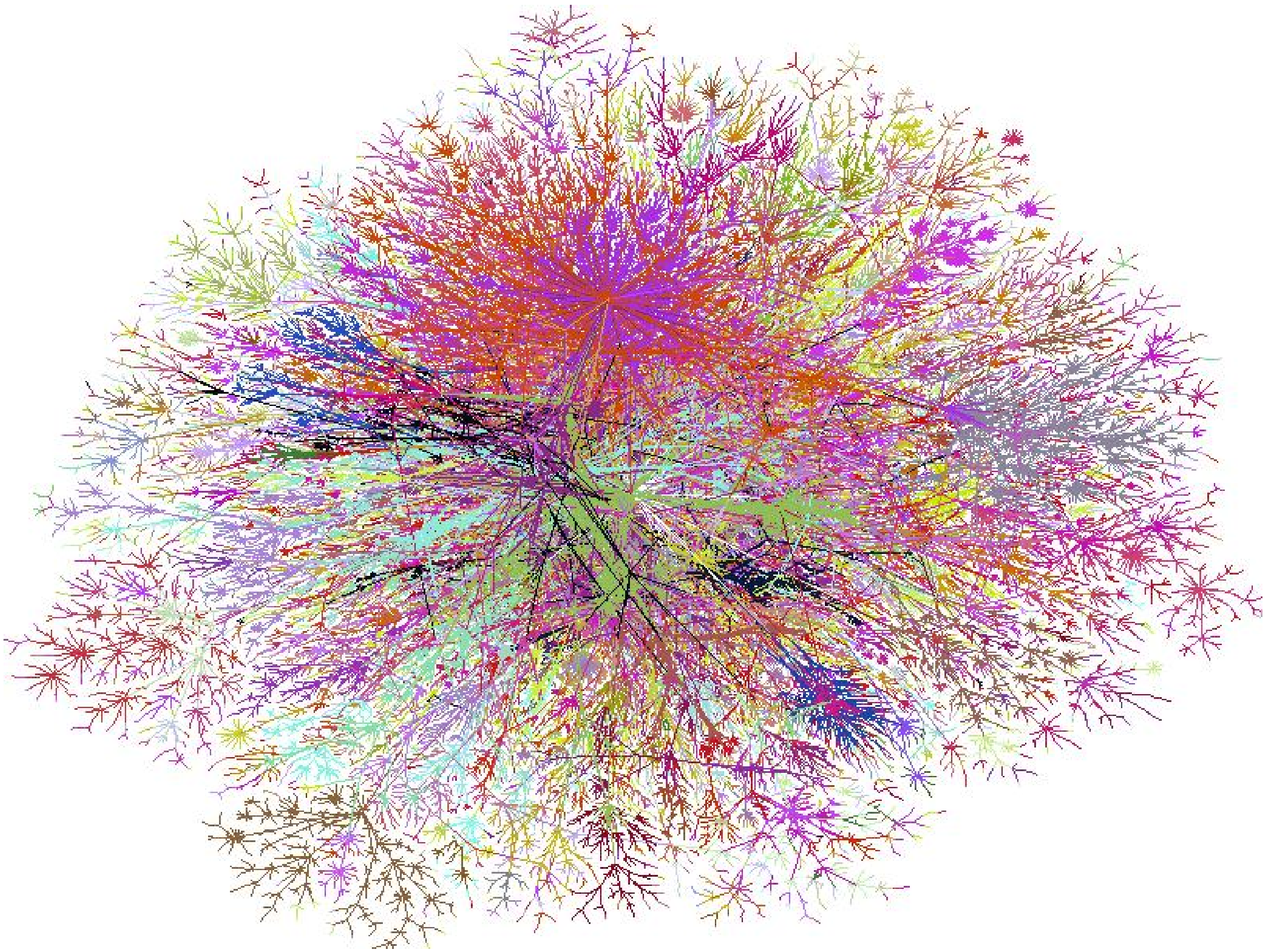
Example social networks
(Immunology; viral marketing)





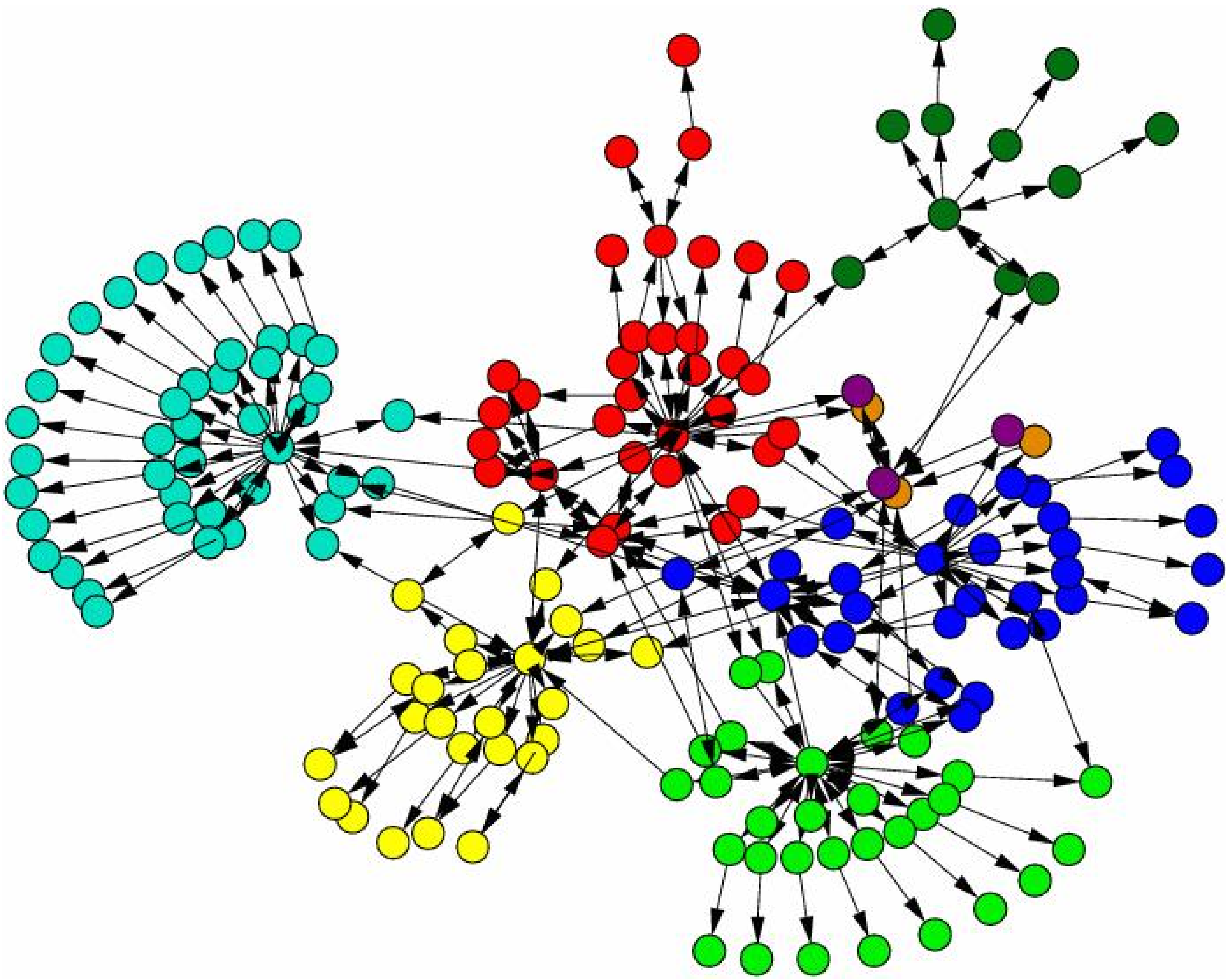
M. E. J. Newman

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



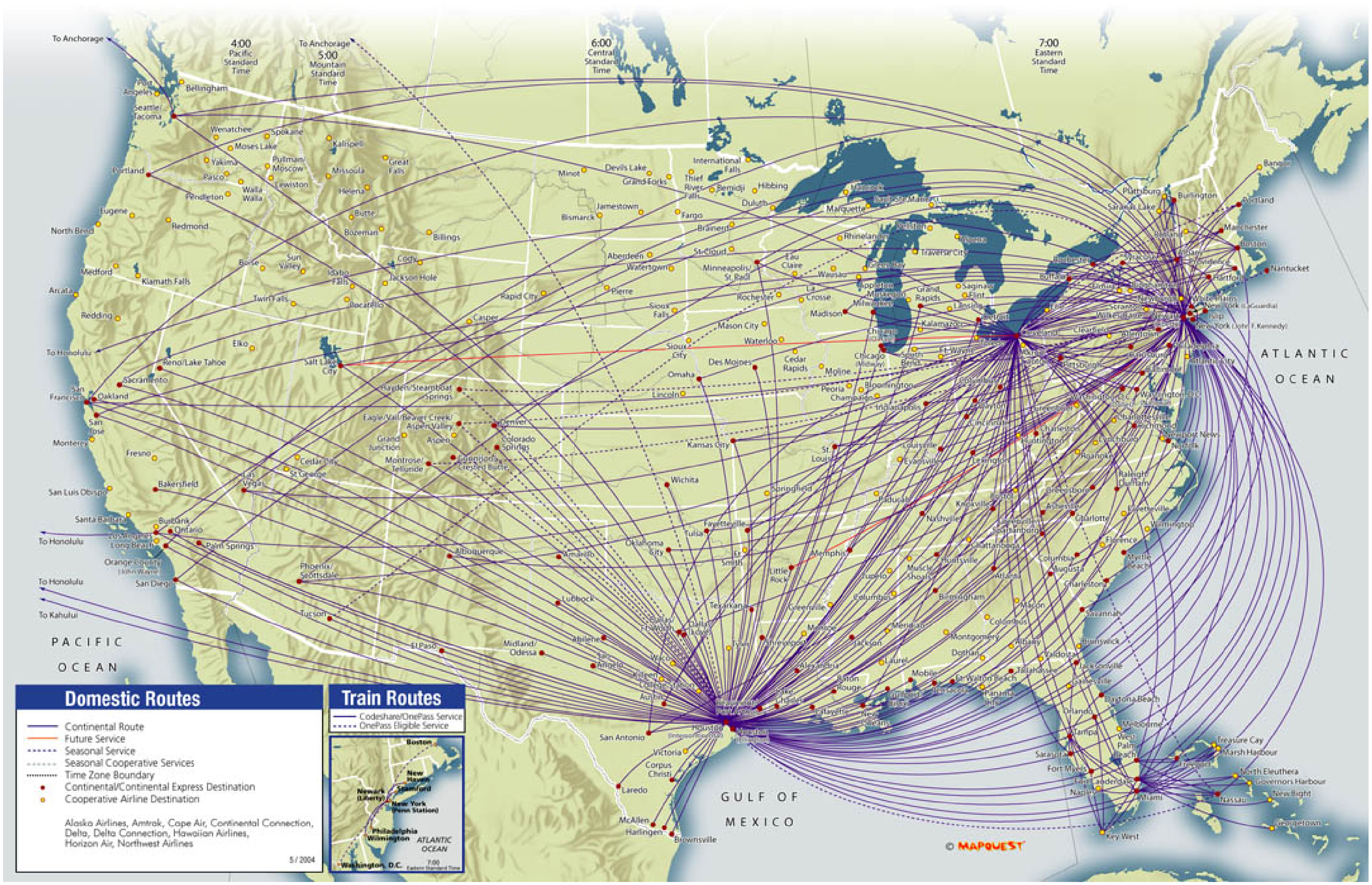
H. Burch and B. Cheswick

A typical web domain
(Web search/organization and growth)



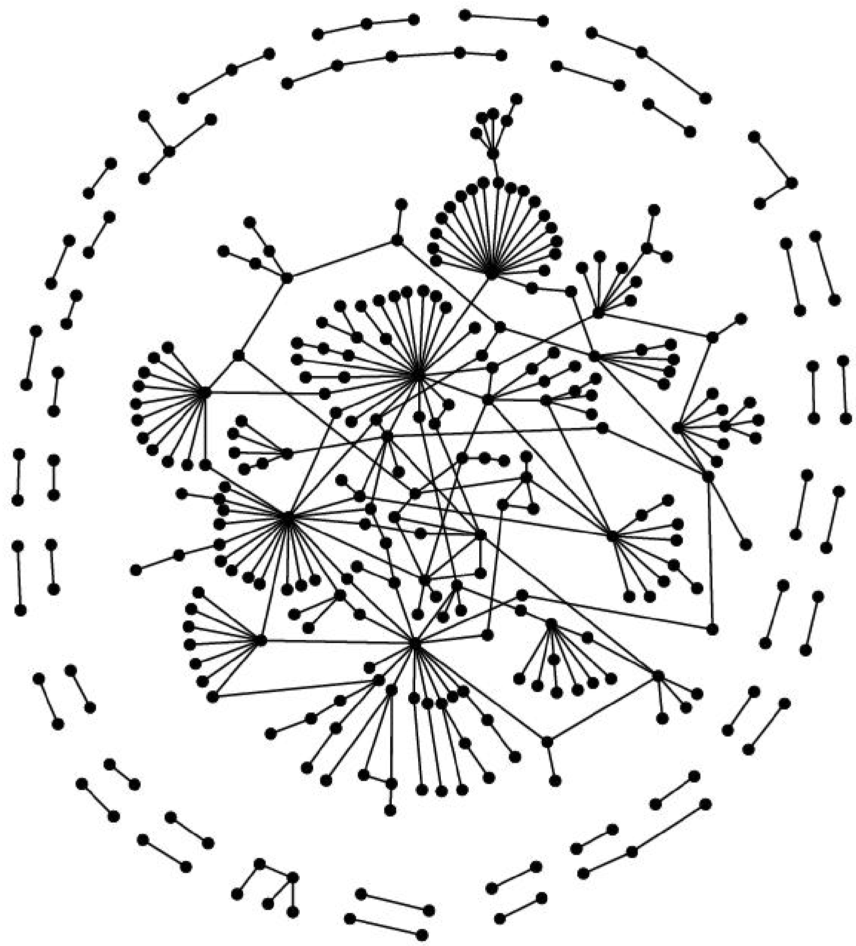
M. E. J. Newman

The airline network
(Optimization; dynamic external demands)



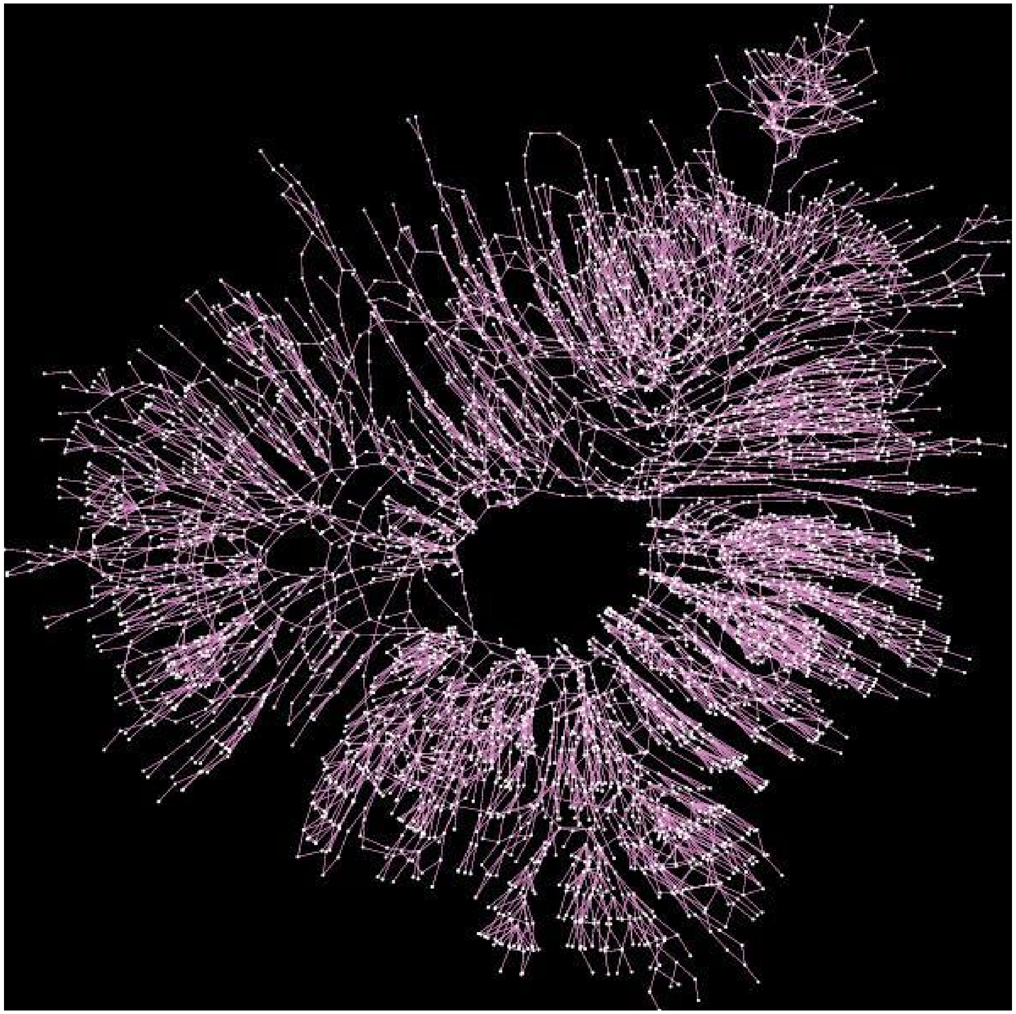
Continental Airlines

Yeast protein signaling network
(Control mechanisms in biology)



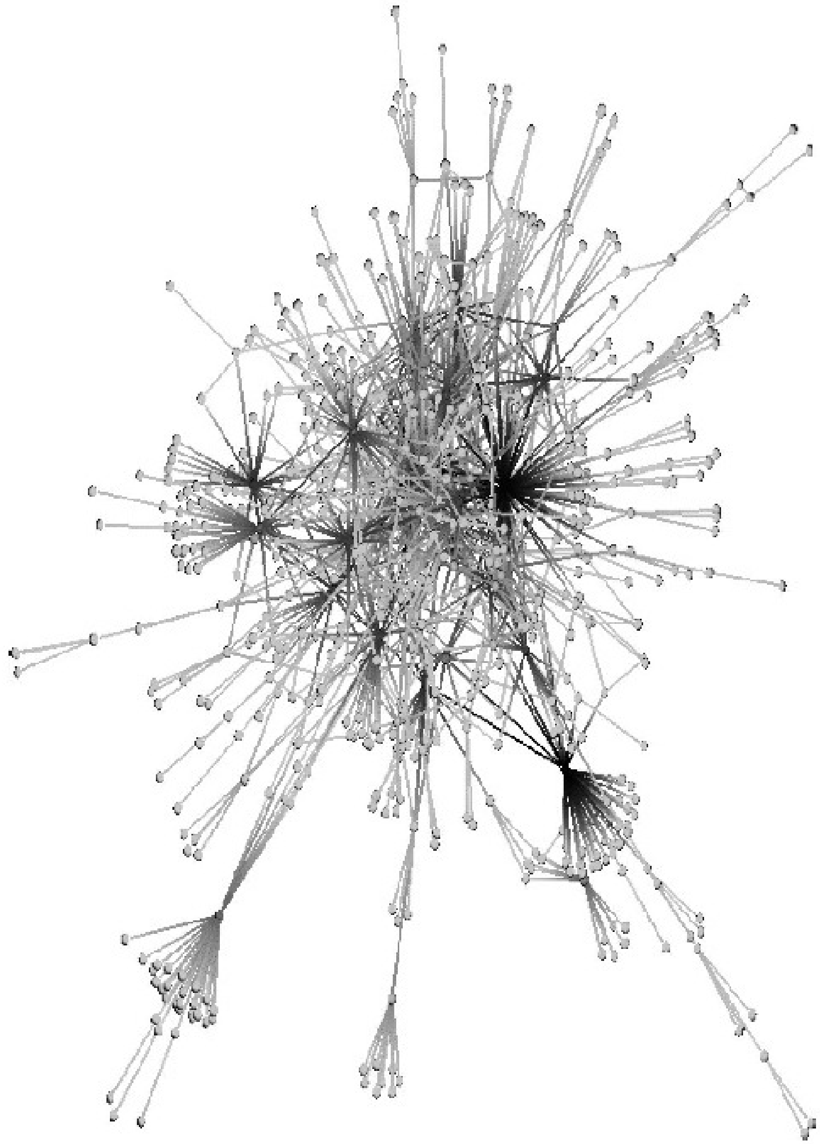
S. Masloc and K. Sneppen; M. E. J. Newman

The power grid
(Mitigating failure; Distributed sources)



M. E. J. Newman

Software call graph
(Uncovering design principles/robustness to mutation)



Chris Myers

Networks: basic properties

- Network made of nodes connected together by edges.
- Edges can be directed or undirected (i.e., one-way or two-way connections).

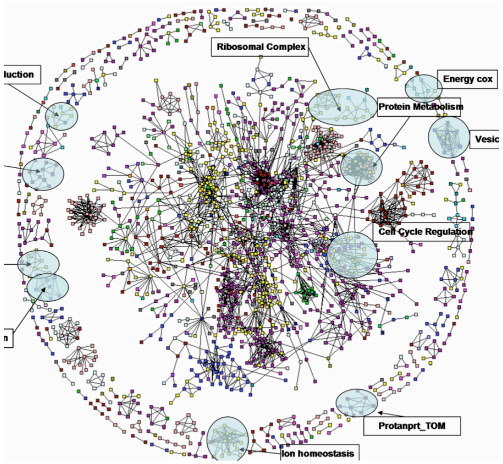
  Example one-way: Web pages.

  Example two-way: Family tree (relatives).

  Example hybrid: Road networks with some one-way and some two-way links (city of Boston prime example!).
- Geometric versus geometry free (e.g., Internet vs WWW)
- STRUCTURE (topology) and FUNCTION (information flow/ dynamics on the network)

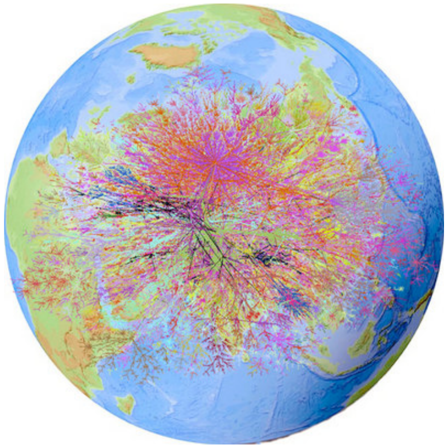
Networks:

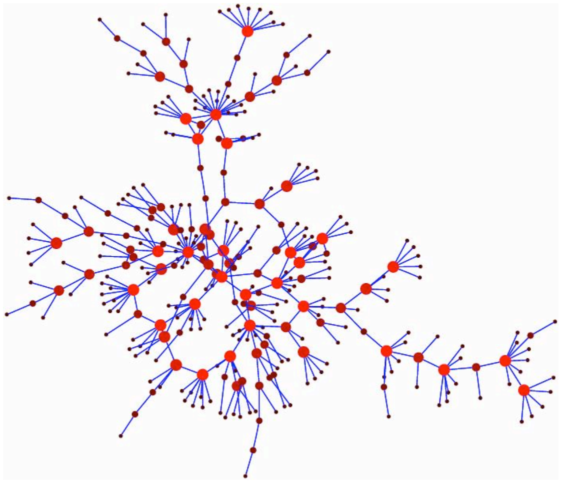
Transportation Networks/ Power grid
(distribution/ collection networks)



Biological networks
- protein interaction
- genetic regulation
- drug design

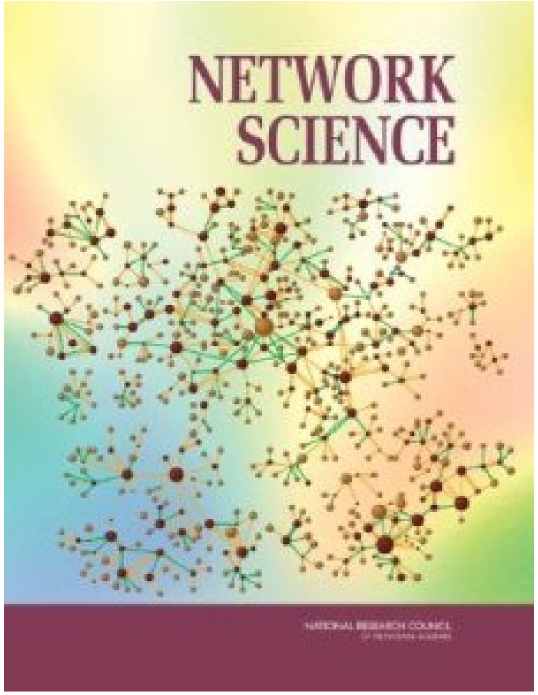
Computer networks





Social networks
- Immunology
- Information
- Commerce

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”

Why do networks exist?
Physical, Biological, Social, Engineered

- More efficient control, esp through hierarchy?
- Robustness to noise and fluctuations?
- Can we learn function from structure?
- Can we apply these lessons to engineered systems?

  $\longrightarrow$ Would a modern power grid look like the one we have?

Lessons from existing networks?

All networks, all quantitatively different, each optimizes something different.

What are the key parameters that distinguish them?

Which kinds work best for which application?

Example question

Is network connectivity a good thing?

General Considerations/Tradeoffs

- For what purpose are we building the network?

- CONNECTIVITY?
  - Preserve at all costs (Internet),
  - Or break at all costs (Immunology)?

- ROBUSTNESS to which failure modes?
  - Random failure (biology),
  - Or targeted attacks (technological).

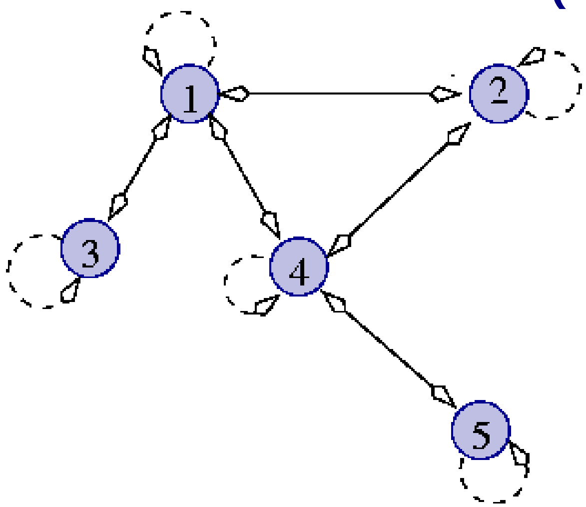
- Fully decentralized or some centralized control?

How do we represent a network as a mathematical object?

Matrix representation of a network: TOPOLOGY

Connectivity matrix, M:

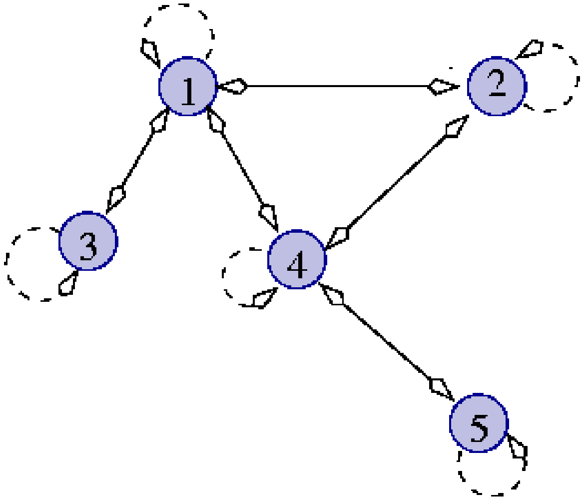
$$M_{ij} = \begin{cases} 1 \text{ if edge exists between } \; i \text{ and } j \\ 0 \text{ otherwise.} \end{cases}$$



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

The *degree* of a node, is how many links it has.

TOPOLOGY: Common measures of fine structure

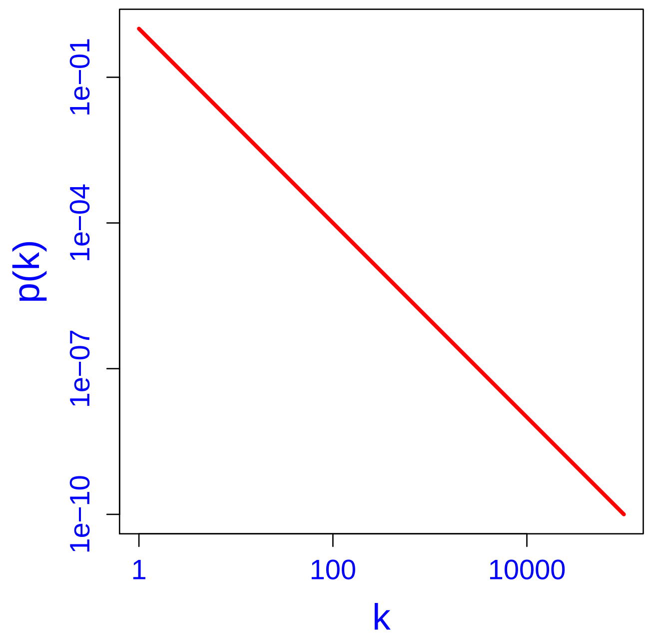


$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} = M$$

- Degree distribution (fraction of nodes with degree k)
- Clustering coefficient
- Diameter
- Betweenness centrality
- Assortative/dissortative mixing

Many different types of networks exhibit

Power Law Degree Distributions (i.e. “scale free”)

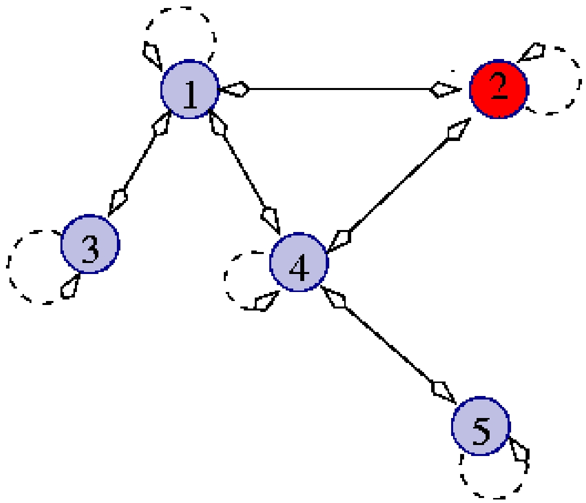


$$
\begin{aligned}
p(k) &\sim k^{-\gamma} \\
\ln(p(k)) &= -\gamma \ln(k) + \text{const.}
\end{aligned}
$$

Matrix representation of a network: ACTIVITY

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

Consider a random walk on the network. The state transition matrix, P:



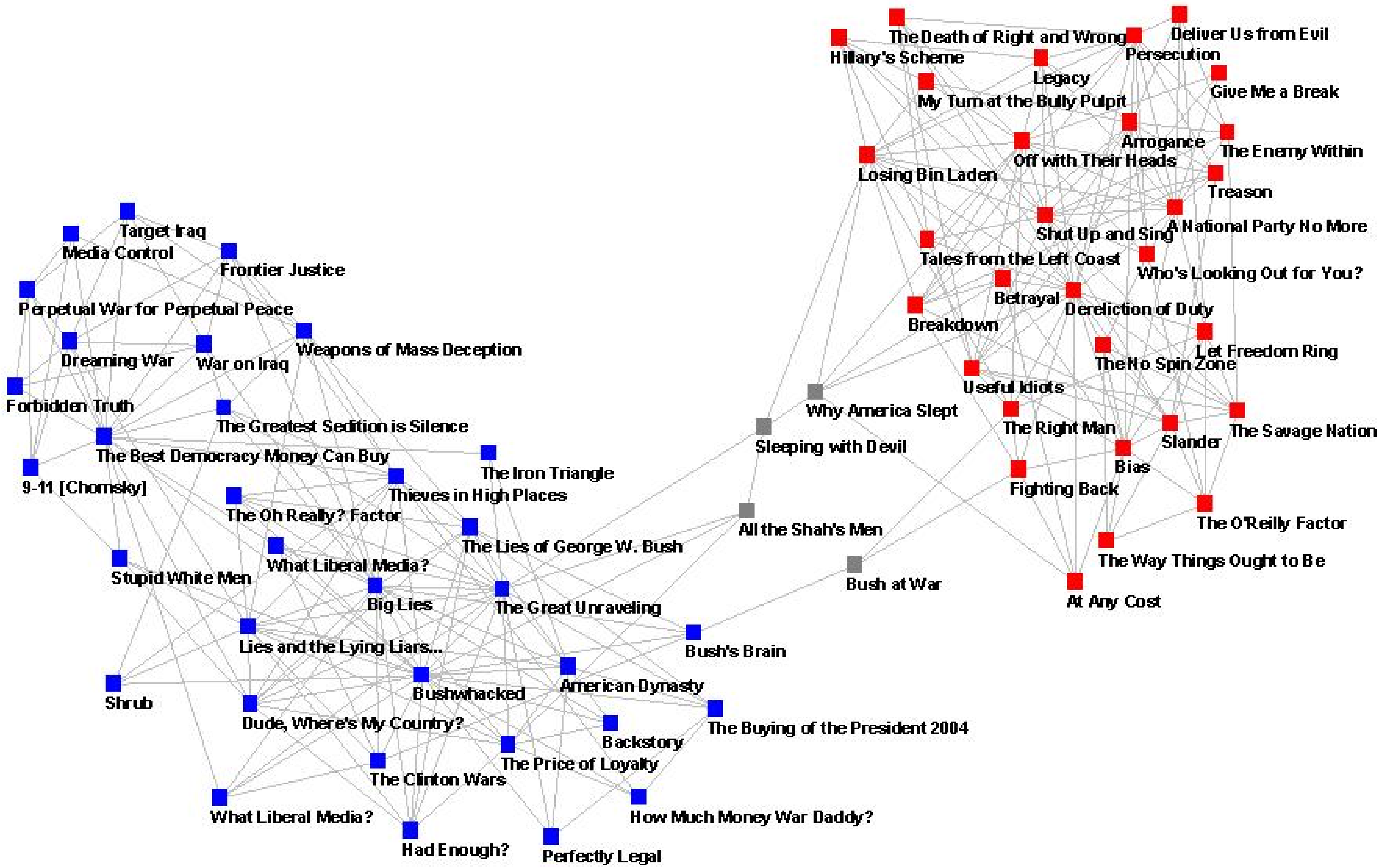
$$\begin{pmatrix} 1/4 & 1/3 & 1/2 & 1/4 & 0 \\ 1/4 & 1/3 & 0 & 1/4 & 0 \\ 1/4 & 0 & 1/2 & 0 & 0 \\ 1/4 & 1/3 & 0 & 1/4 & 1/2 \\ 0 & 0 & 0 & 1/4 & 1/2 \end{pmatrix} = P$$

The eigenvalues and eigenvectors convey much information.

Eigenvalues and eigenvectors of the state transition matrix

- The stationary distribution of a random walk on the graph:
  - “cover time” of a random walker
  - mixing time
  - occupancy probabilities
- Partitioning graphs into subgraphs/communities

Partitioning networks: Community structure: Political Books USA, 2004



M. E. J. Newman

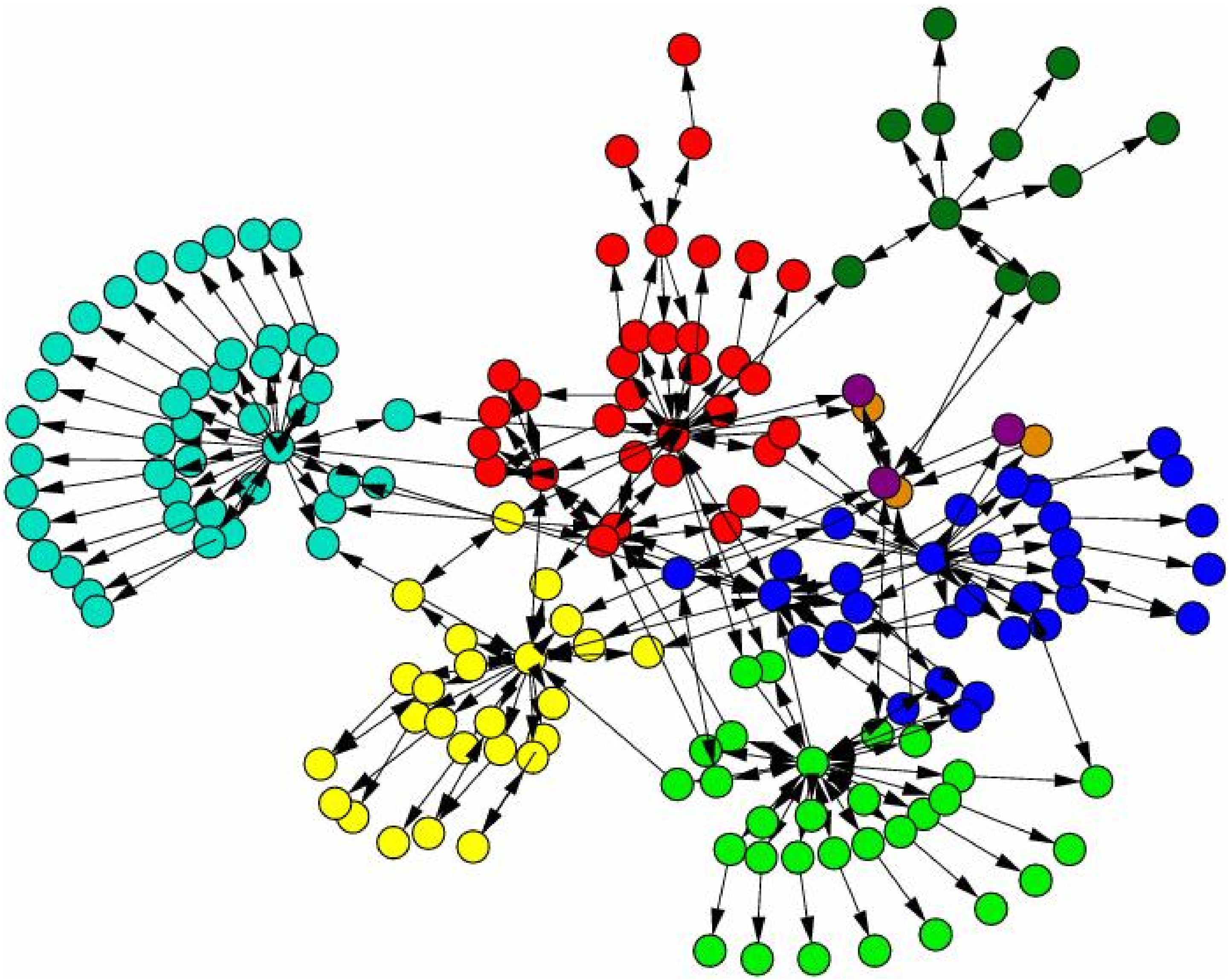
Dynamics on networks

- Markov chains (i.e., state transition matrix) current approach.
- Tracks one field (the random walker), and has only one timescale (the rate of the walker).
- Partial differential equations track multiple fields with multiple scales simultaneously.
- How do we develop techniques for PDEs on a network?

A Sample Application

Web Search

A typical web domain



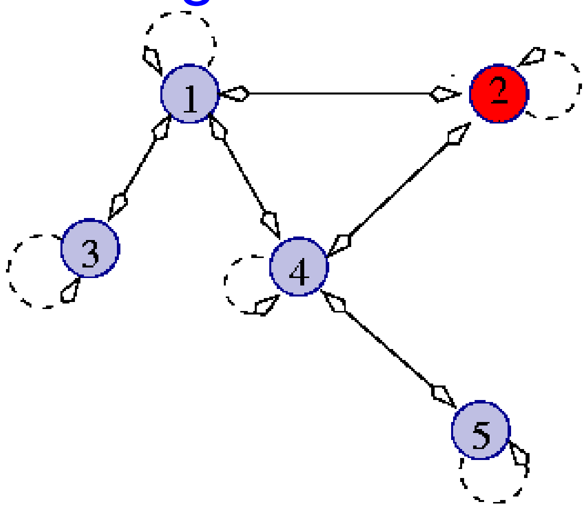
M. E. J. Newman

Pieces of a WWW search engine

- First thing needed is a web map!
  - -Nodes: Web page (unique url, semantic content, in-links, out-links).
  - -Edges: Connectivity between pages.
- Use a crawler (also called spider) to crawl the WWW.
- Over 10^9 web pages!! How long will this take? Will it ever be complete?
- Selectively choose which out-links to follow during the crawl (avoid spam-holes, link-farms, etc.)
  - -What heuristics characterize quality pages?
  - -How often should recrawl occur, and in what order?

Page Rank

Brings us back to random walks on a graph (Markov chains):



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

The stationary probabilities are essentially the Page Rank!
(The celebrated algorithm which was the foundation for Google.)

Summary: Terms introduced today

- Graph/network (also nodes and edges)
- Connectivity matrix (M)
- State transition matrix (random walk on M)
- Degree
- Degree distribution
- Graph diameter

Networks

- Network structures are pervasive – physical, biological, social, engineered
- Networks are made of discrete nodes – hard to envision continuum description
- Nodes can live in geometric, or geometry free space (Internet vs WWW)
- Need to consider:
  - Topology of network
  - Activity on network
- Can we learn lessons from existing networks?