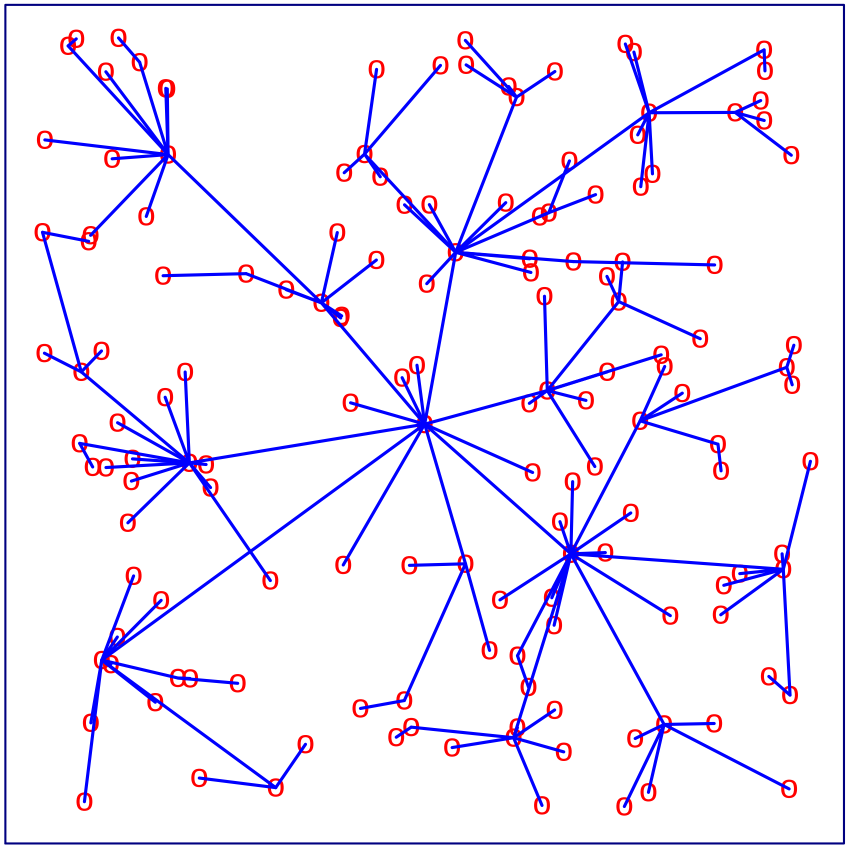
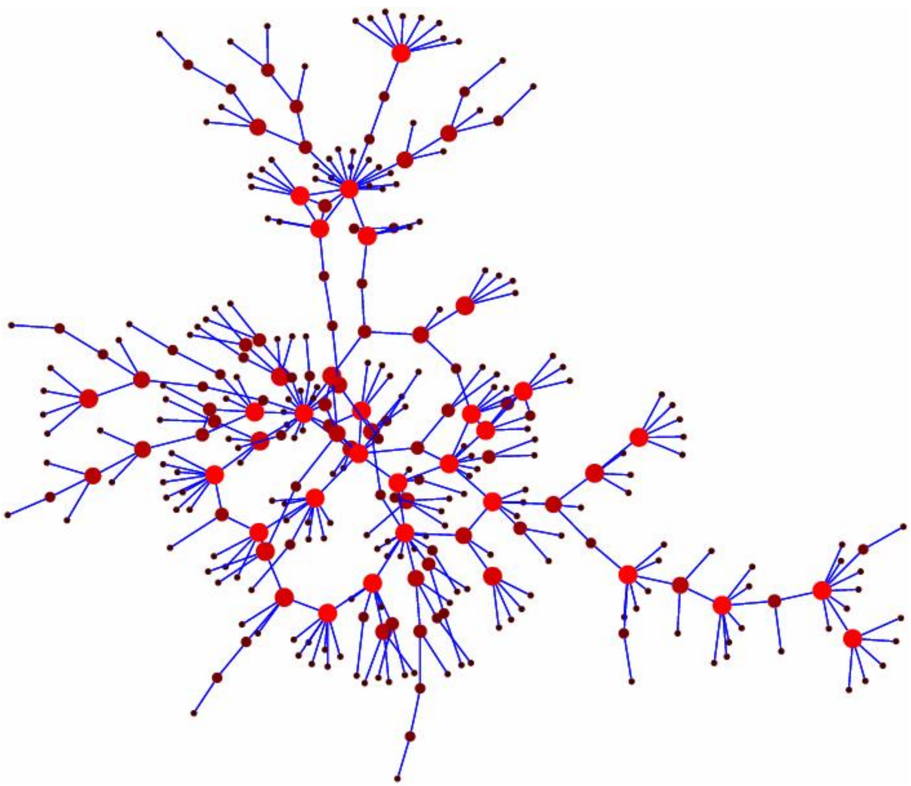
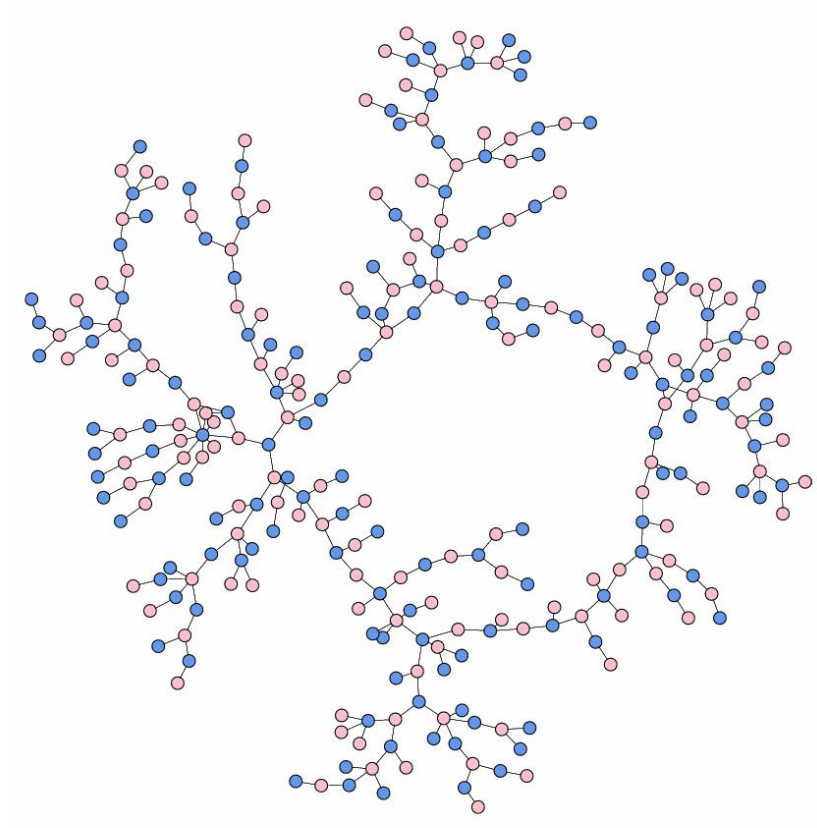
# “What is a network?”



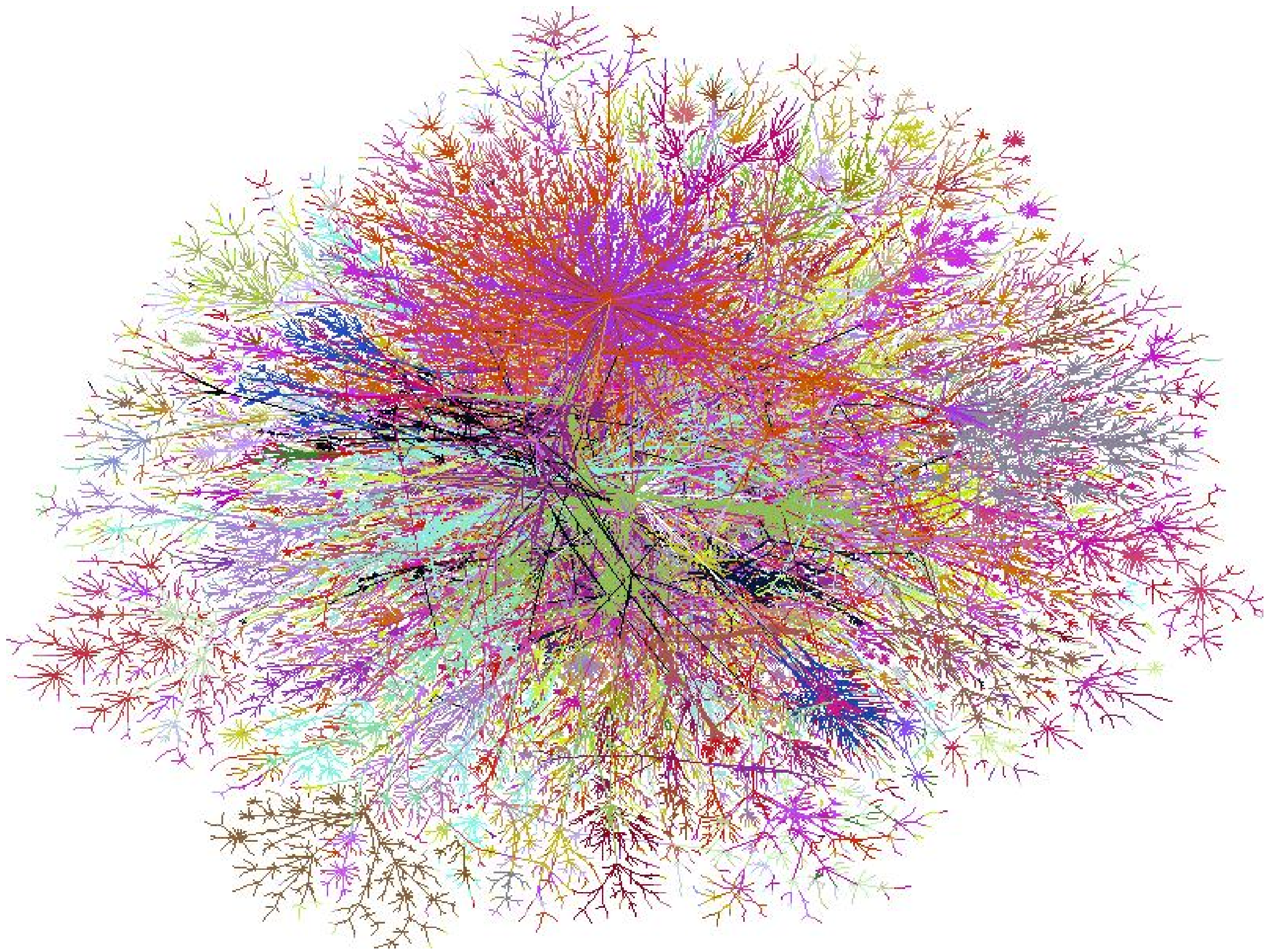
# Example social networks





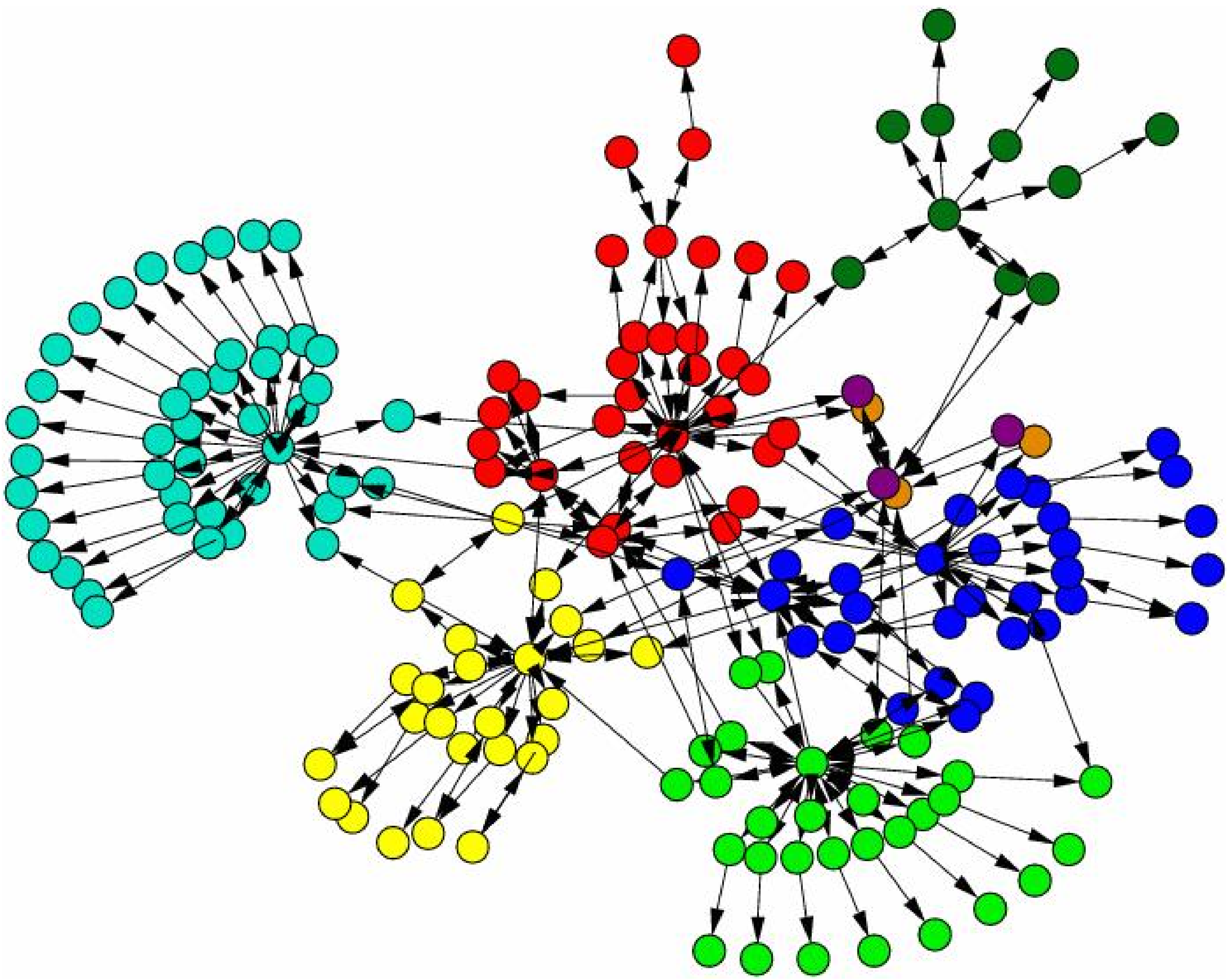
M. E. J. Newman

# The Internet



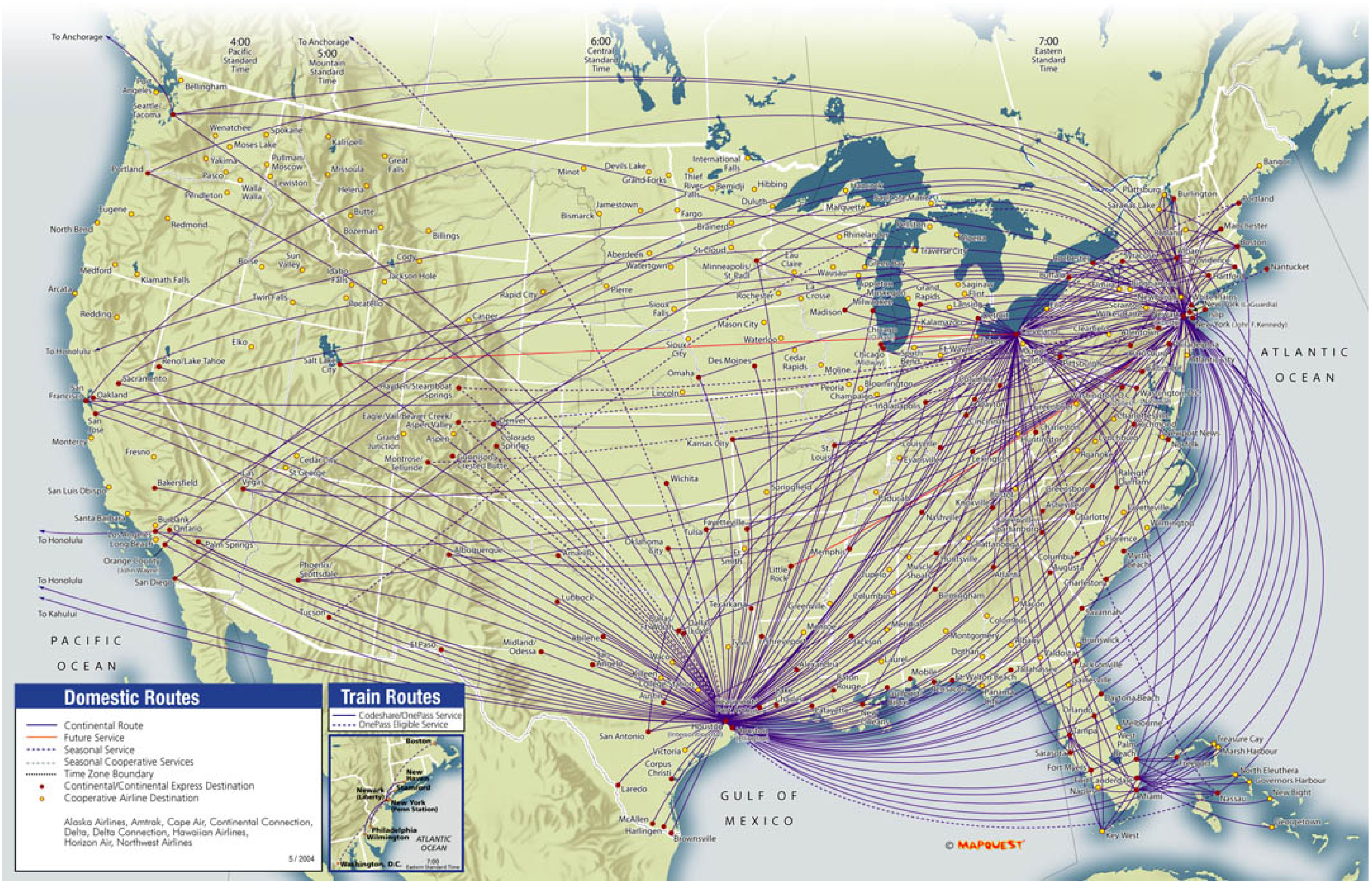
H. Burch and B. Cheswick

# A typical web domain



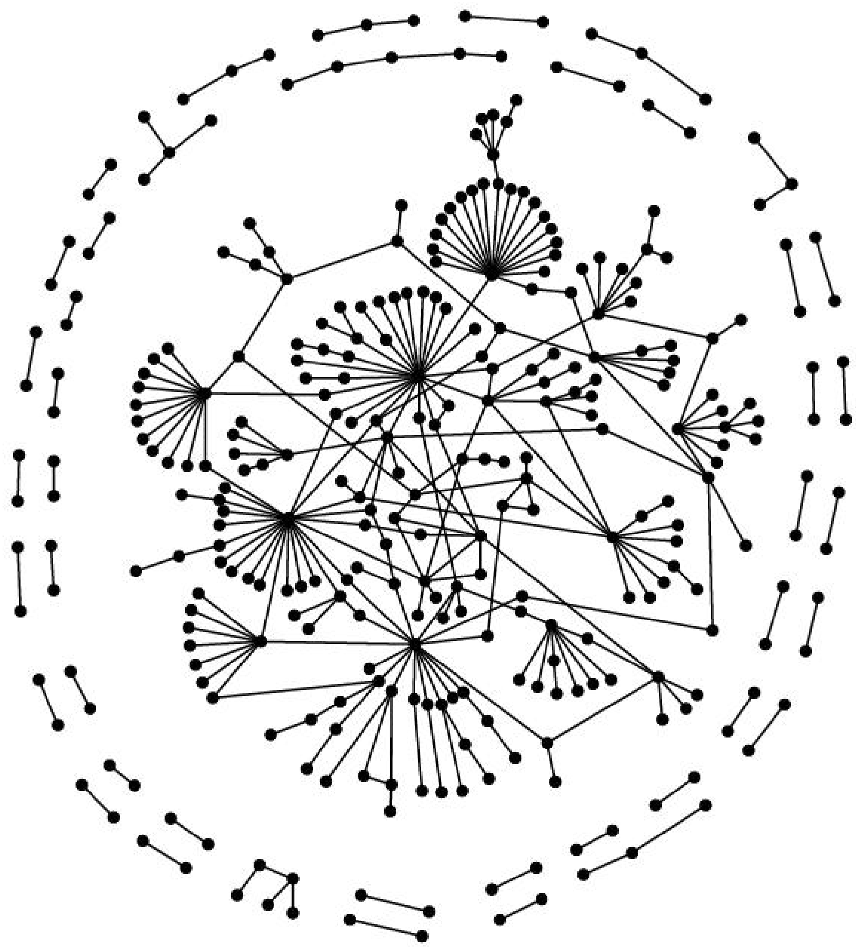
M. E. J. Newman

# The airline network



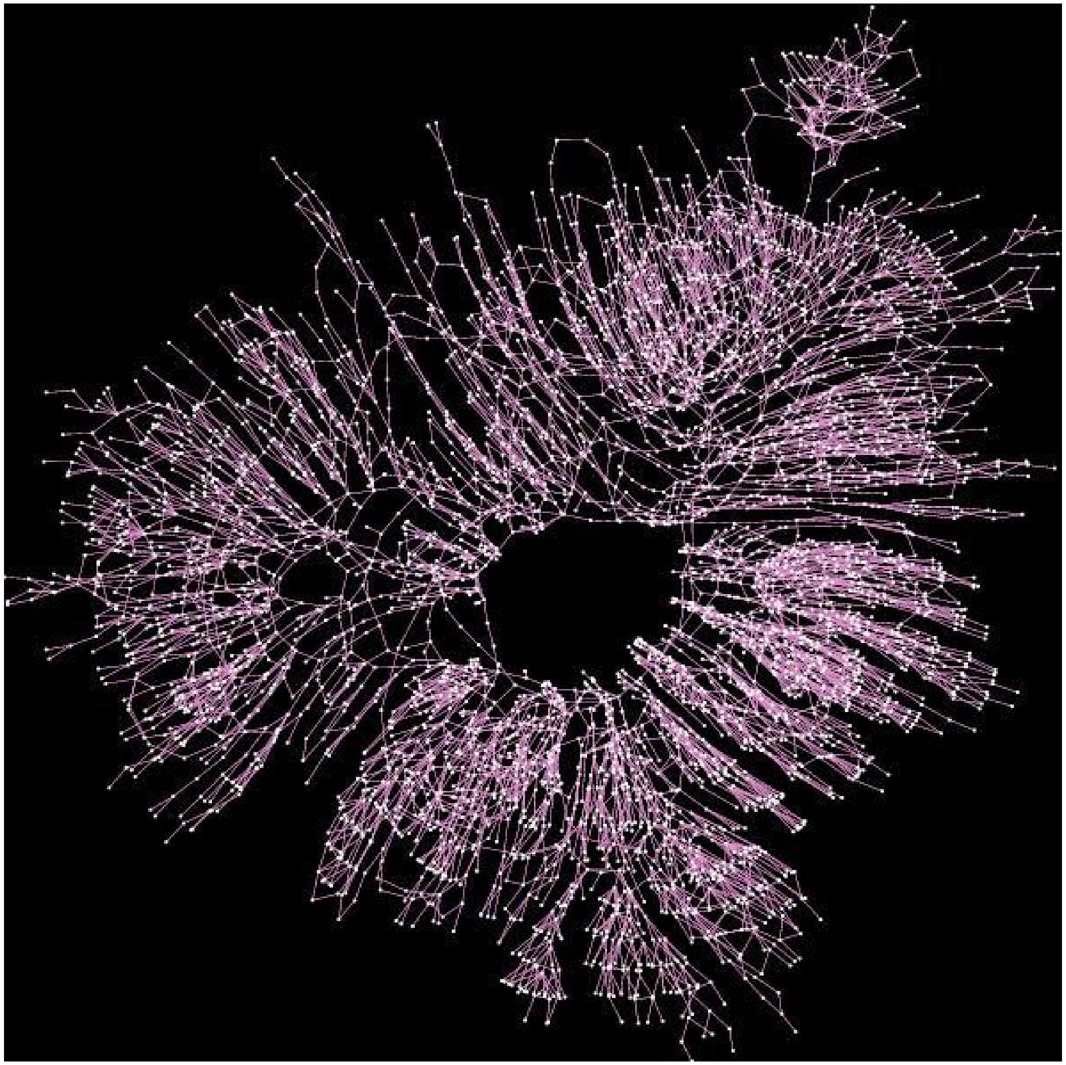

Continental Airlines

# Yeast protein signaling network



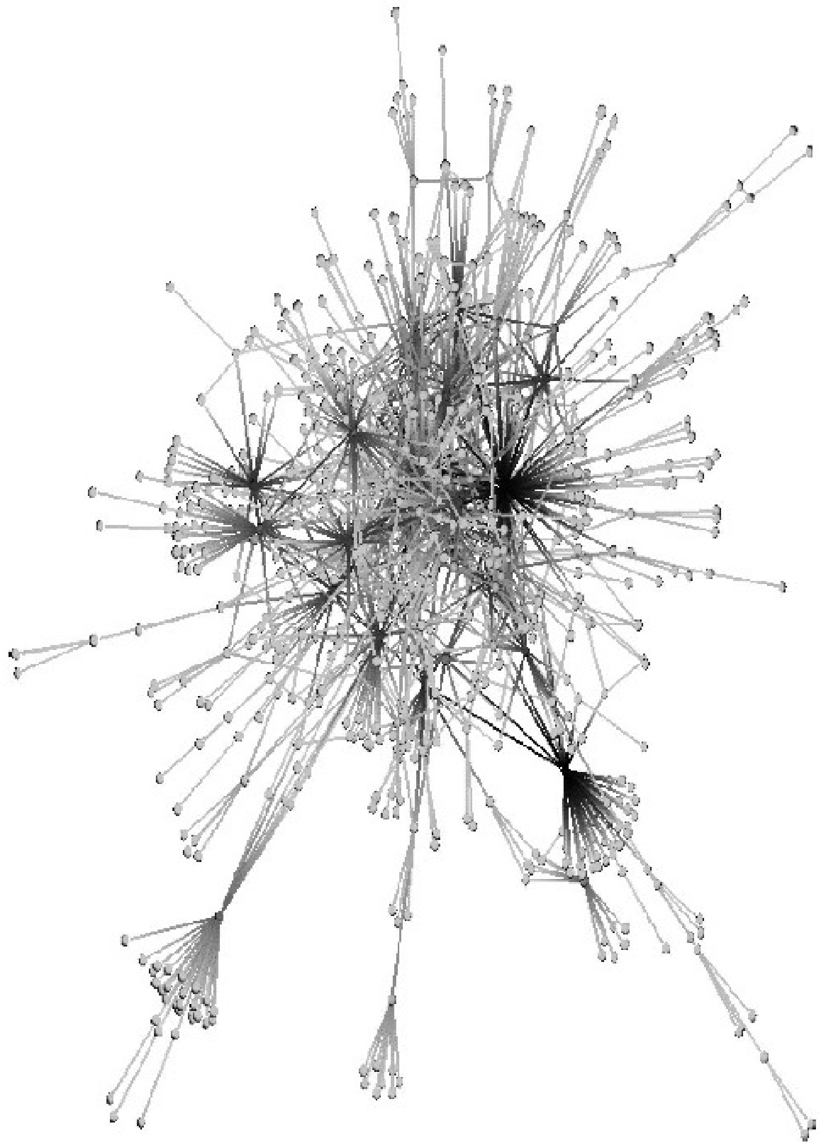
S. Masloc and K. Sneppen; M. E. J. Newman

# The power grid



M. E. J. Newman

# Software call graph



Chris Myers

# Networks: basic properties

- Network mode 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: Links connecting routers in the Internet.

  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 and FUNCTION (information flow/dynamics on the network)

## Networks exist:
## Physical, Biological, Social, Engineered

- They are made of discrete entities (nodes) and discrete links (edges). Hard to envision any continuum approximation.

- They have complicated interconnections; not the typical Euclidean world or regular lattice.

- Furthermore the connections can be dynamic!
  (This is uncharted territory).

- Statistical properties of the connections characterize topology.

- Markov chains are the main tool for dynamics on a network.

# Why do networks exist?

- 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?

  –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?

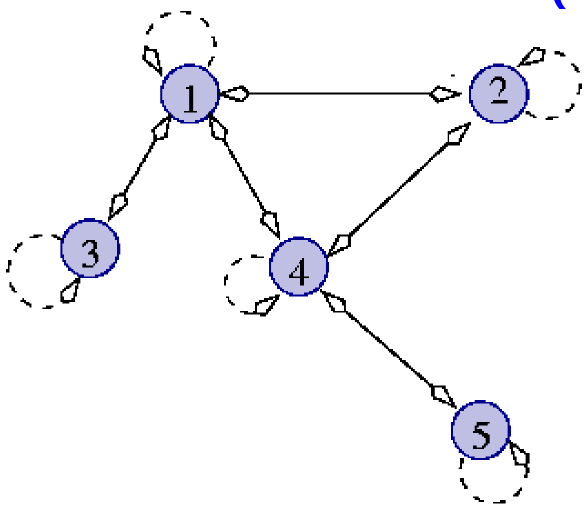
## General Considerations/Tradeoffs

- For what purpose are we building the network?
- CONNECTIVITY?
- Robustness to which failure modes?
  (e.g., random (biology) vs targeted (technological) attacks).
  (Furthermore, robustness to what? Connectivity? Functionality?)
- Smart-network/dumb-terminal vs Dumb-network/smart-terminal
- Fully decentralized or some centralized control?

# Matrix representation of a network: TOPOLOGY

Connectivity matrix, $M$:

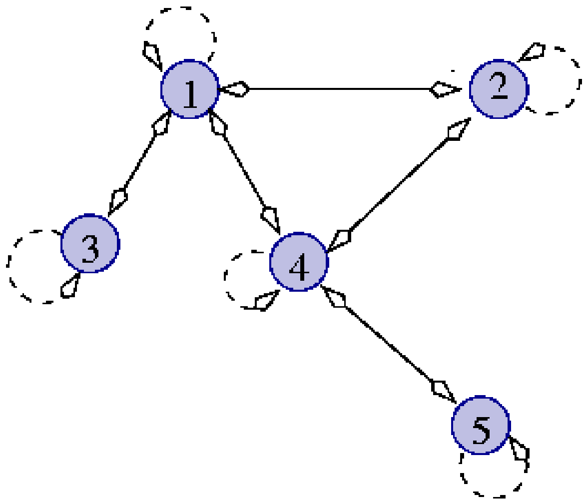
$$M_{ij} = \begin{cases} 1 \text{ if edge exists between } i \text{ and } j \\ 0 \text{ otherwise.} \end{cases}$$




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

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

# TOPOLOGY: Common measures of fine structure



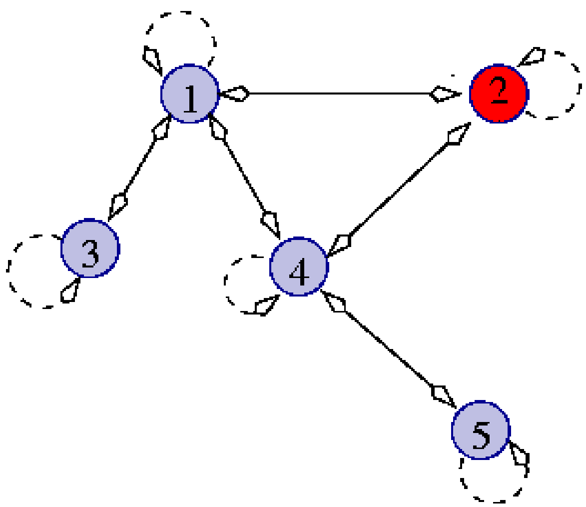

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} = M$$

- Degree distribution (fraction of nodes with degree $k$)
- Clustering coefficient
- Community structure
- Assortative/dissortative mixing
- Self-similarity even when pruned

# Matrix representation of a network: ACTIVITY

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

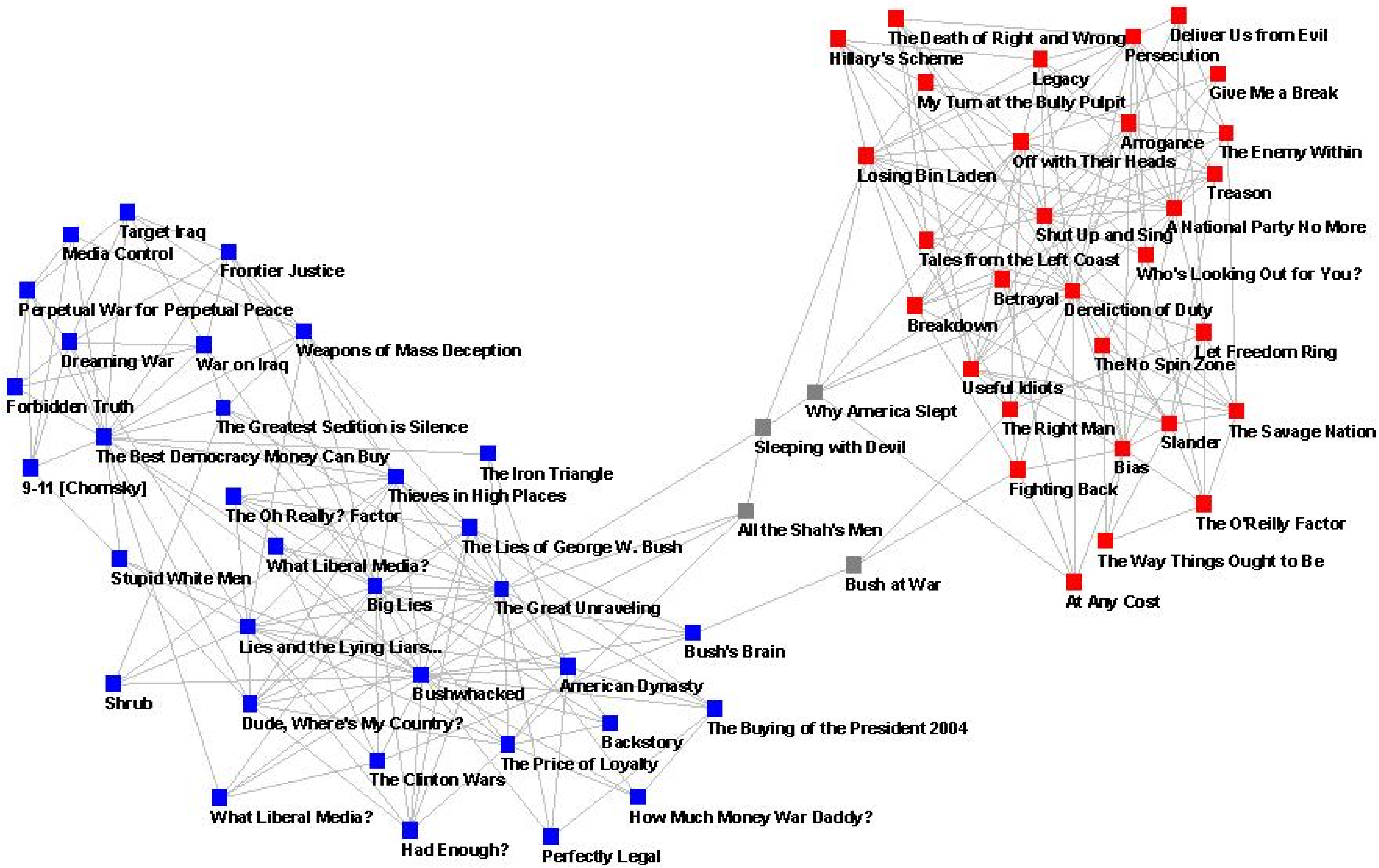
Consider a random walk on the network. The state transition matrix, $P$:




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

The eigenvalues and eigenvectors convey much information.

# Community structure: Political Books 2004



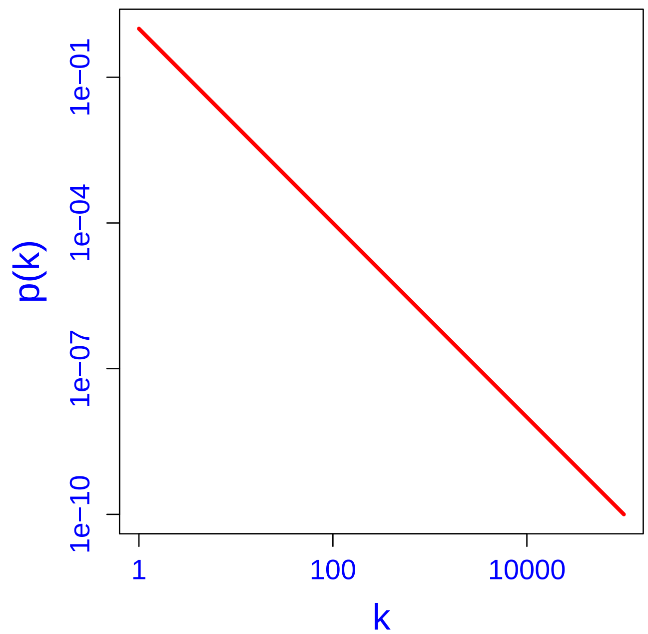

M. E. J. Newman

# 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?

# Many different types of networks exhibit

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



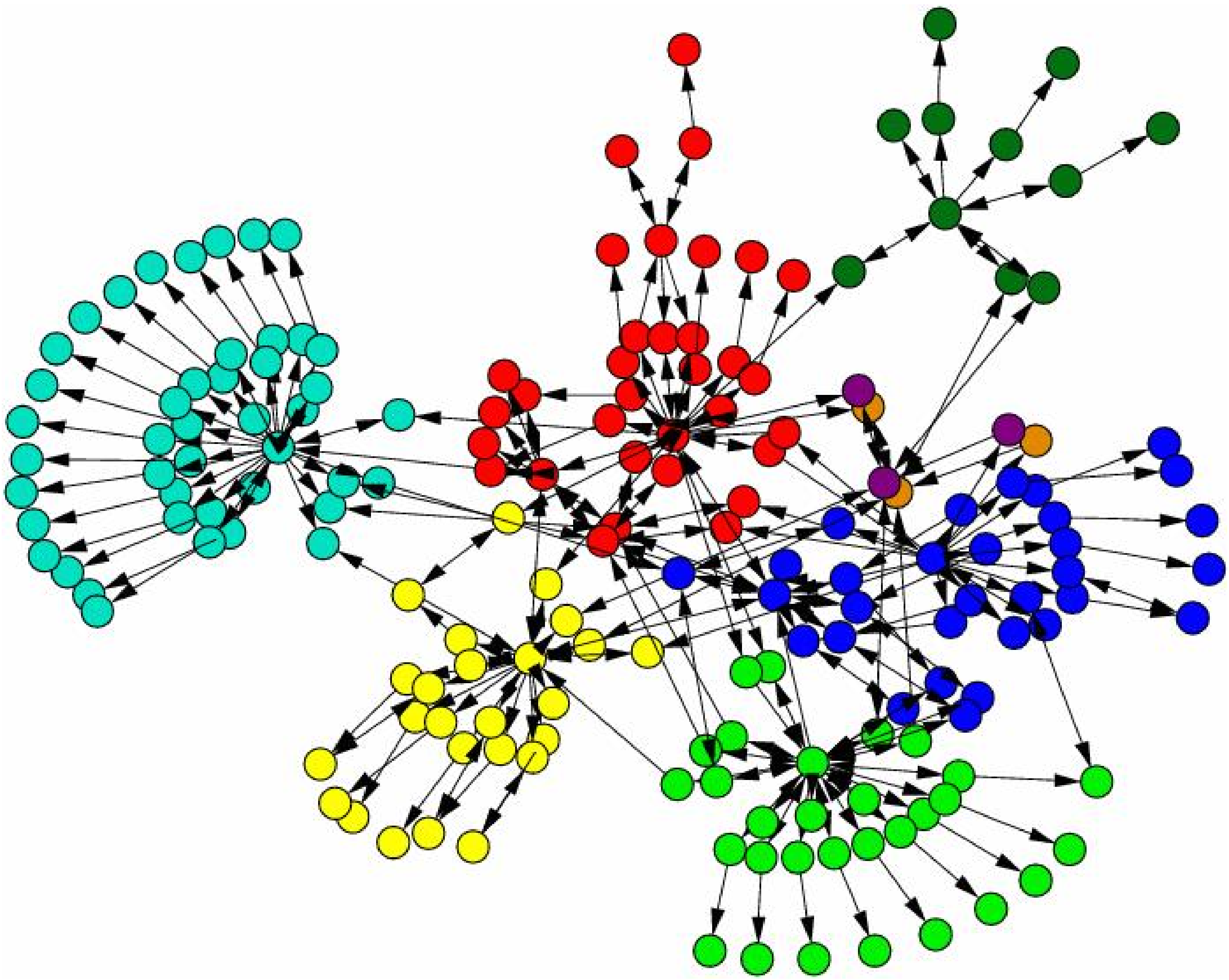

$$p(k) \sim k^{-\gamma}$$

$$\ln(p(k)) = -\gamma \ln(k) + \text{const.}$$

# A Sample Application

# Web Search

# A typical web domain



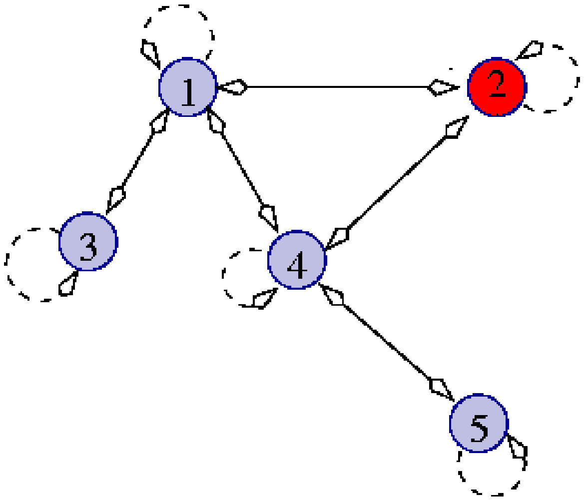
M. E. J. Newman

# 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.)

# 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?