DECSAI
Departamento de Ciencias de la Computación e I.A.
Universidad de Granada

# Networks

© Fernando Berzal berzal@acm.org

# Networks

- Network Analysis
  - Applications
  - Network Properties
- Network Models
  - Random-Graph Models
  - Growing Random Models
  - Strategic Network Formation
- Network Structure & Dynamics
  - Network Centrality
  - Community Detection
  - Diffusion through Networks
  - Search on Networks
- Bibliography

# Network Analysis

Networks permeate our lives.

Networks play a central role in determining

- the transmission of information about job opportunities,
- how diseases spread,
- which products we buy,
- our likelihood of succeeding professionally,
- …



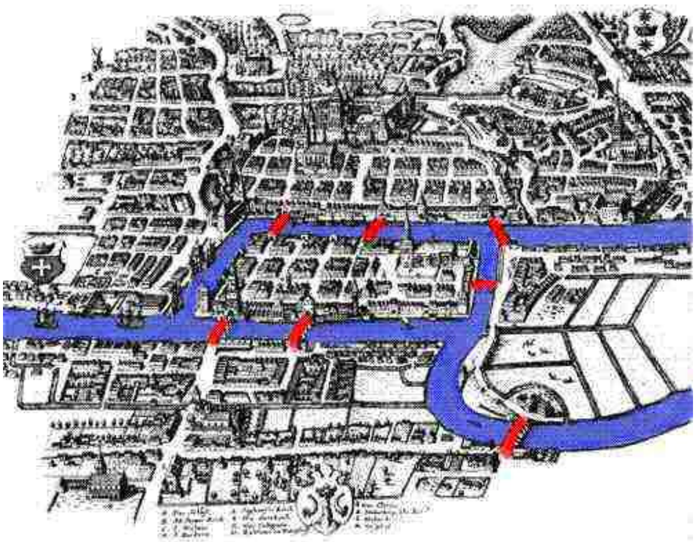
# Network Analysis

As a field of study…

- How relationships between parts give rise to the collective behaviors of a system and how the system interacts and forms relationships with its environment (complex systems).
- Common principles, algorithms and tools that govern network behavior (network science).

# Network Analysis

**Origins: Graph Theory**



The Seven Bridges of Könisberg
(Leonhard Euler, 1736)



# Network Analysis

**Networks as graphs "on steroids"…**

- **Objects**: Graph vertices.
  - Objects can be of different kinds.
  - Objects can be labeled.
  - Objects can have attributes
- **Links** between objects: Graph edges.
  - Links can be of different kinds.
  - Links can be directed (arcs) or undirected (edges).
  - Links can have attributes.

# Network Analysis

**A formal definition of network**
[Ted G. Lewis: "Network Science," 2009]

$$G(t) = \{ N(t), L(t), f(t) : J(t) \}$$

where
t = time (simulated or real)
N = nodes (a.k.a. vertices or "actors")
L = links (a.k.a. edges)
f = topology (connections through links)
J = behavior of nodes and links (algorithm)



# Network Analysis

**An interdisciplinary field: Complex systems**
("networks of heterogeneous components that interact")

- Physics: **Nonlinear dynamics & chaos.**
  Dynamical systems that are highly sensitive to initial conditions (a.k.a. butterfly effect).
- Economics: **Markets.**
  Spontaneous (or emergent) order as the result of human action, but not the execution of any human design [Austrian perspective].
- Information theory: **Complex adaptive systems.**
  (focus on the ability to change and learn from experience).

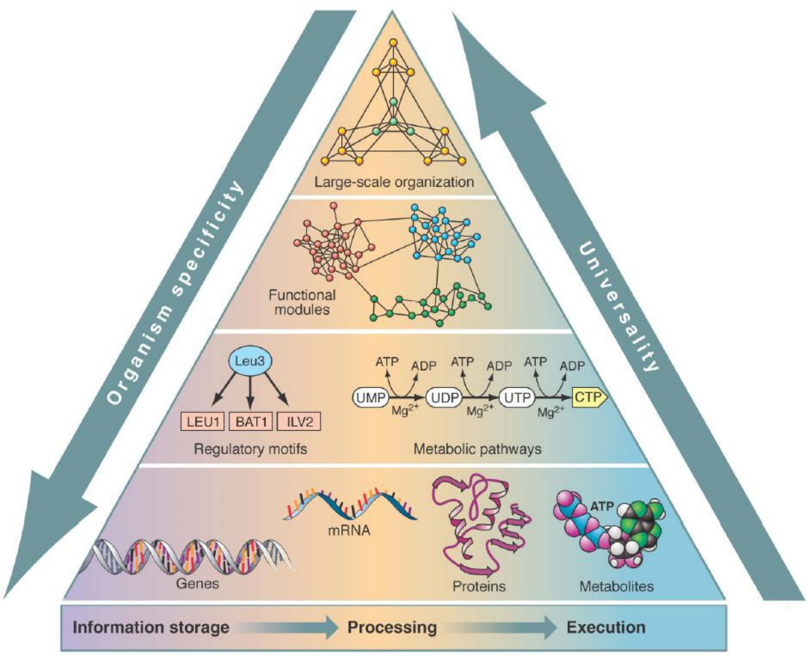
# Applications

- “Cheminformatics”: Chemical compounds.
- “Bioinformatics”: Protein networks & bio-pathways
- Software Engineering: Program analysis…
- Network flow analysis (transport, workflows…)
- Semi-structured databases, e.g. XML
- Knowledge management: Ontologies & semantic nets
- Computer-aided design (CAD): IC design…
- Geographic information systems (GIS) & cartography
- Social networks, e.g. Web
- Economic networks, e.g. markets



# Applications

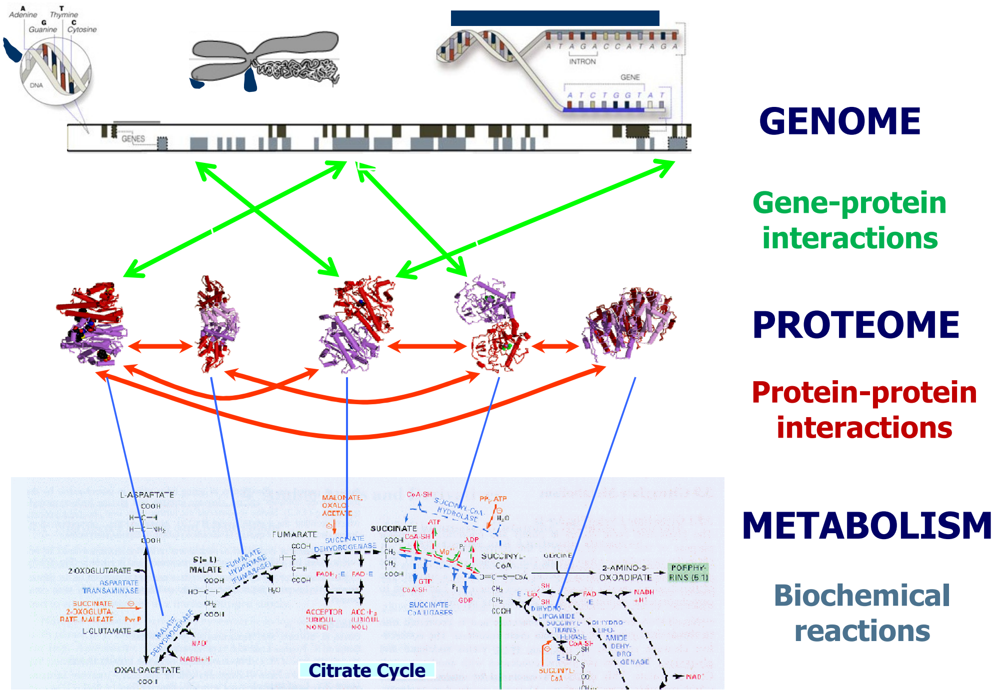
“Life complexity pyramid”




from Z.N. Oltvai and A.-L. Barabasi. Science, 2002

# Applications

## Biological networks





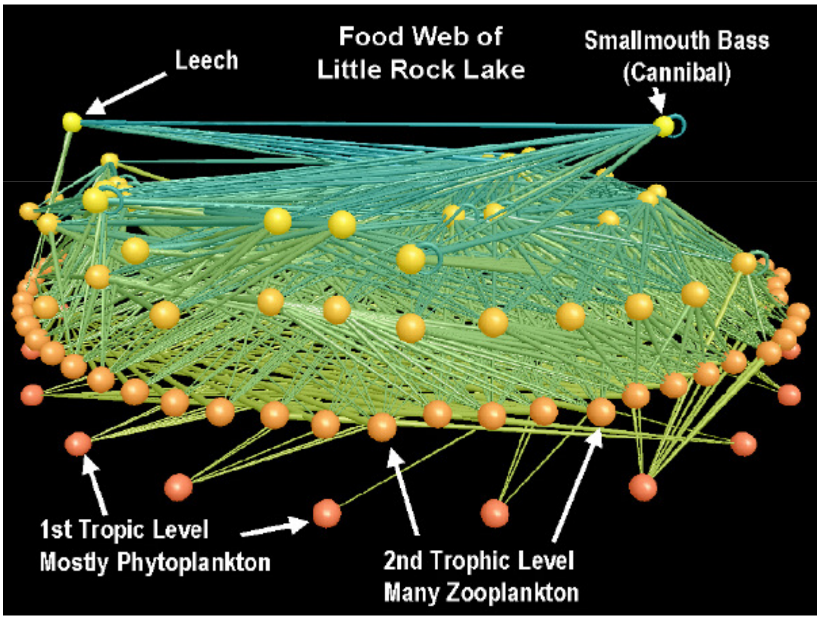

# Applications

Yeast protein interaction network

from H. Jeong et al Nature 411, 41 (2001)

# Applications

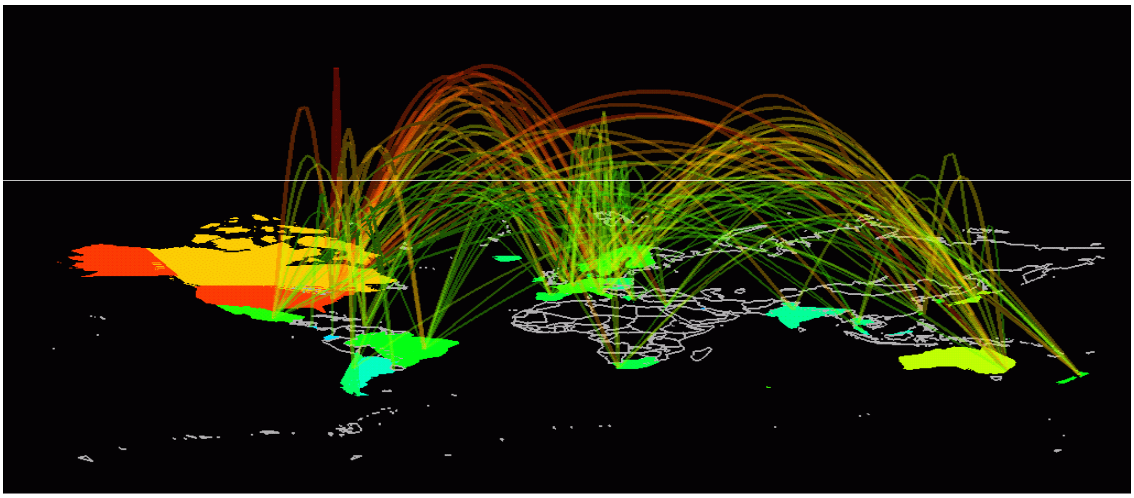
**Ecological network**: Trophic relationships in a food web.





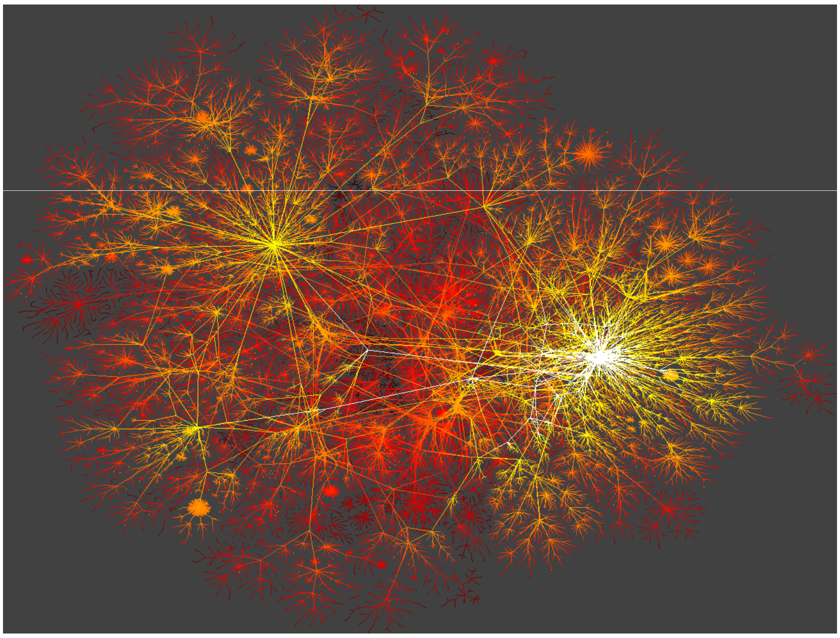

# Applications

Telecommunication network



# Applications

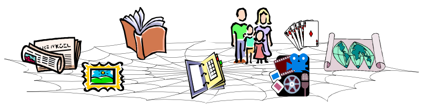
Internet

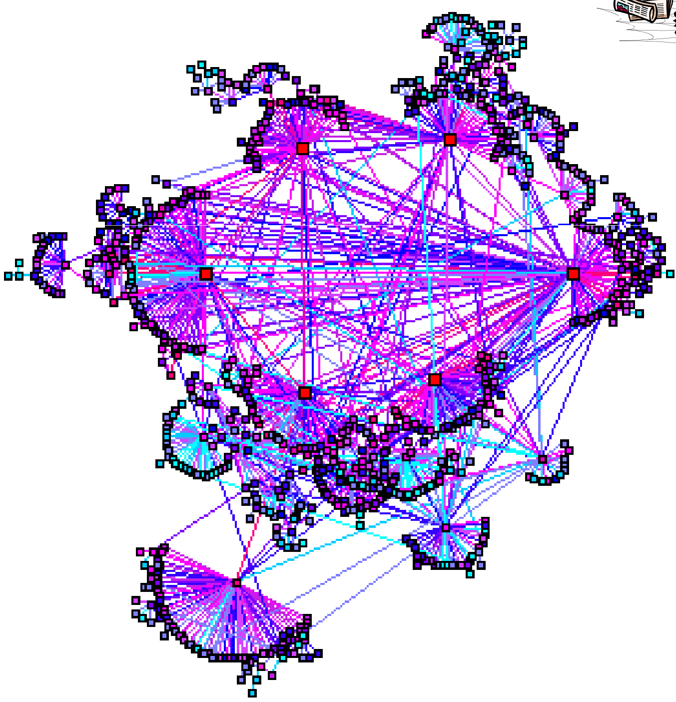




# Applications

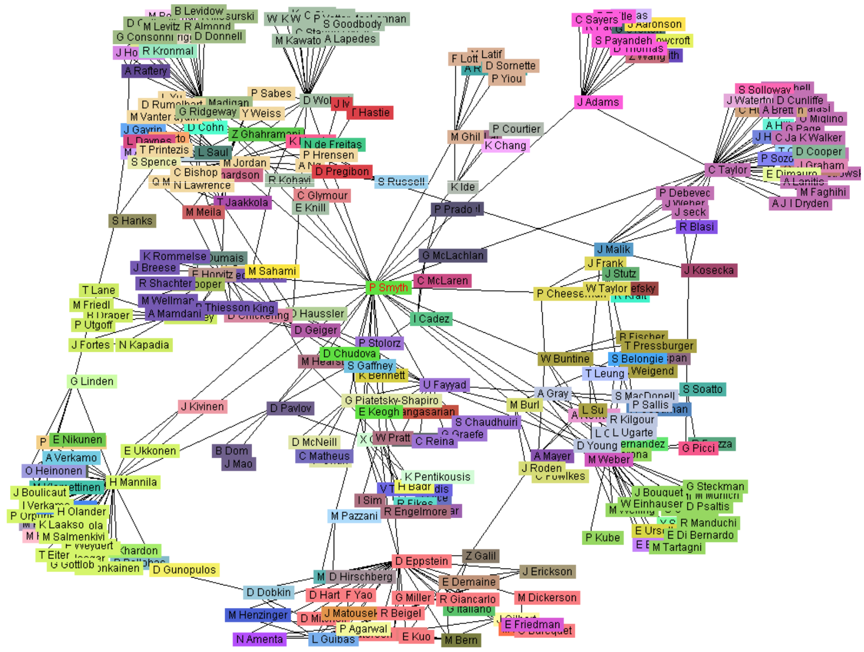
World Wide Web





# Applications

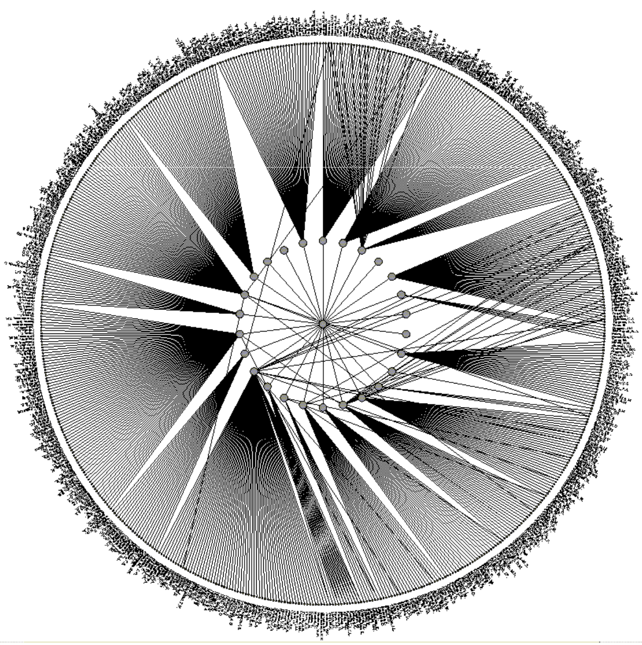
**Social network**: Bibliographic network (coauthors)





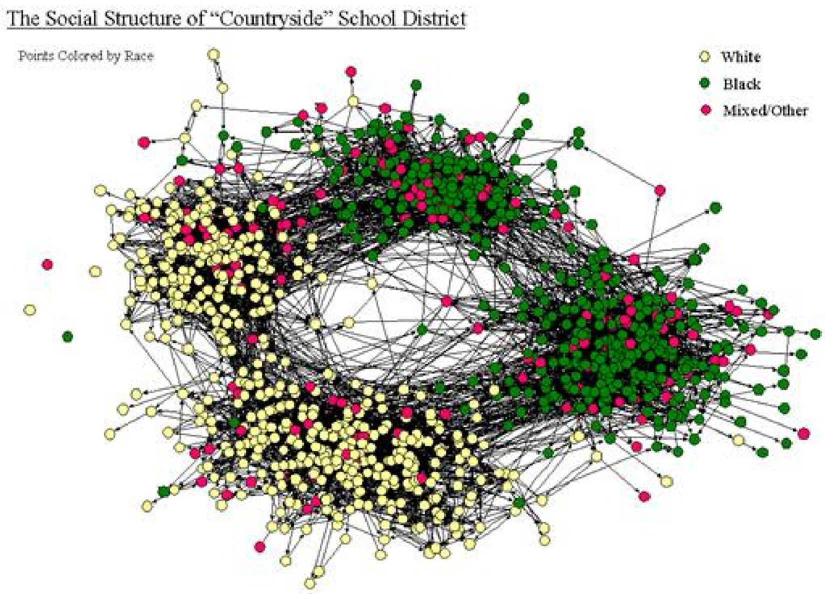

# Applications

**Social network**: Bibliographic network (coauthors)



# Applications

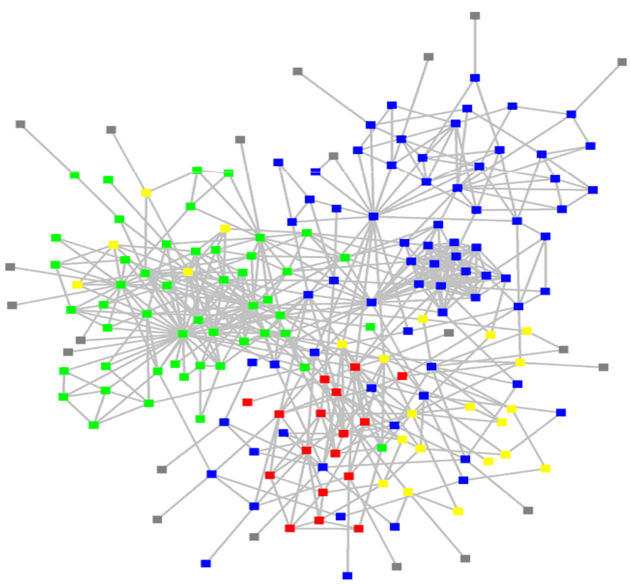
**Social network**: FOAF ("friend of a friend")





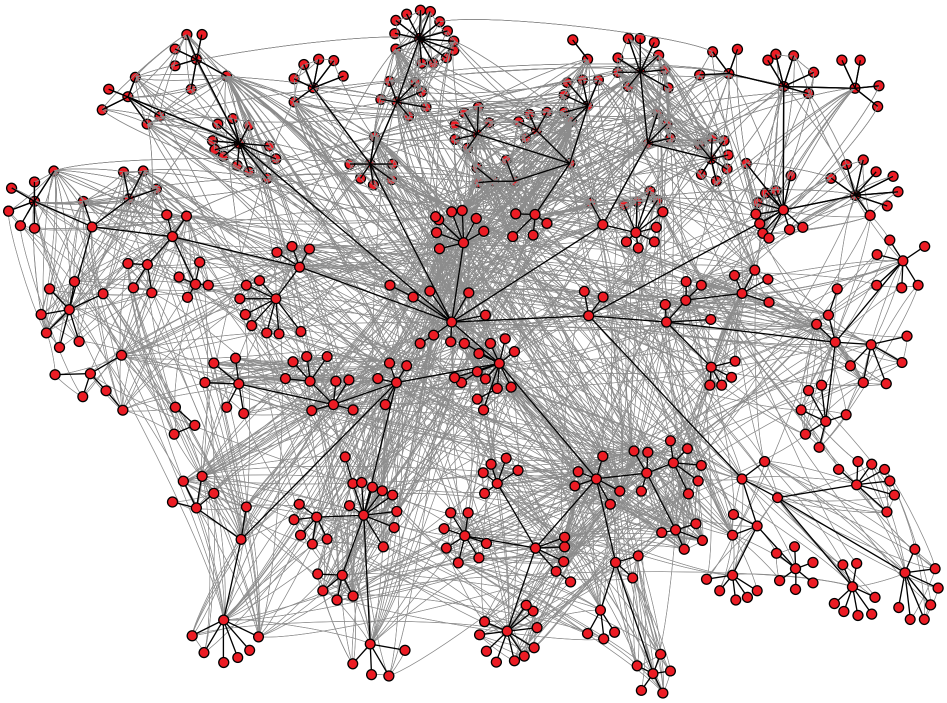

# Applications

**Social network**: Organization



# Applications

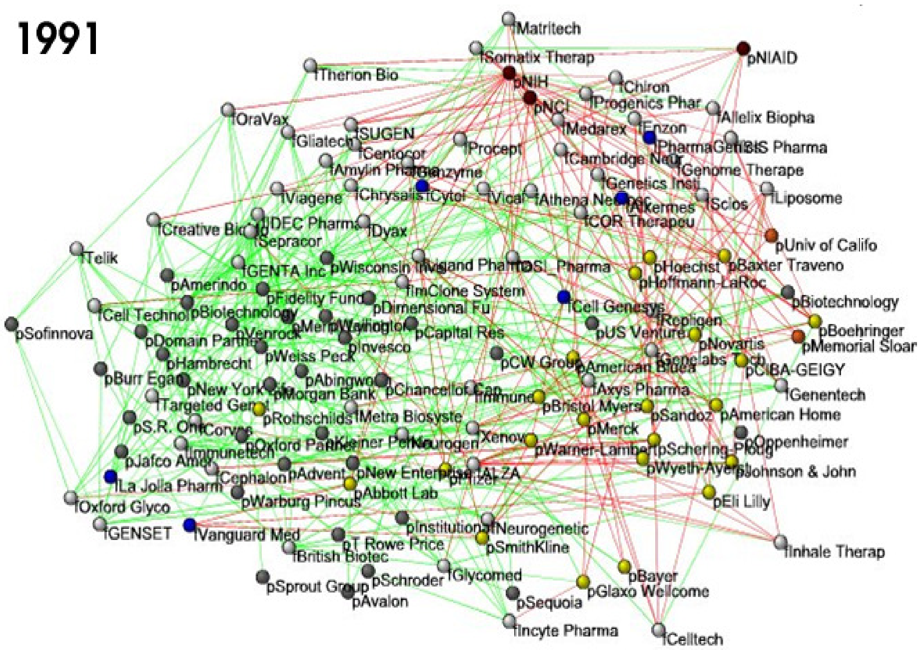
**Social network**: E-mail spectroscopy





# Applications

**Social network**: US Biotech Industry



# Network Properties

**Common network features:**

- Large scale.
- Continuous evolution.
- Distribution (nodes decide their connections).
- Interactions only through existing links.



# Network Properties

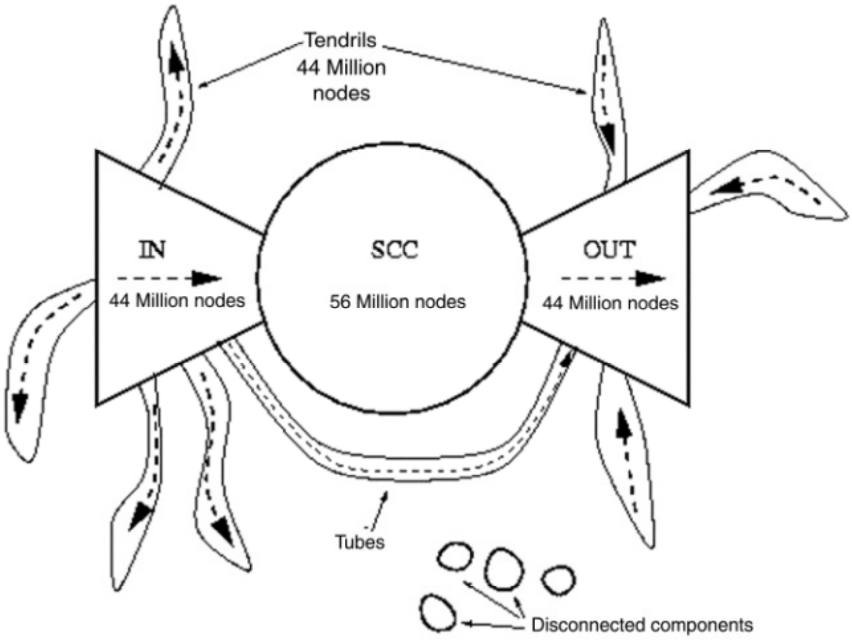
**Some interesting structural properties:**

- Connected components: How many? Of what size?.
- Network diameter: Average distance, worst case…
- Node degree distribution
  & existence of "hubs" (heavily-connected nodes).
- Groupings (balance between local and large-distance connections, as well as their roles).

# Network Properties

**Network Connectivity**

WWW

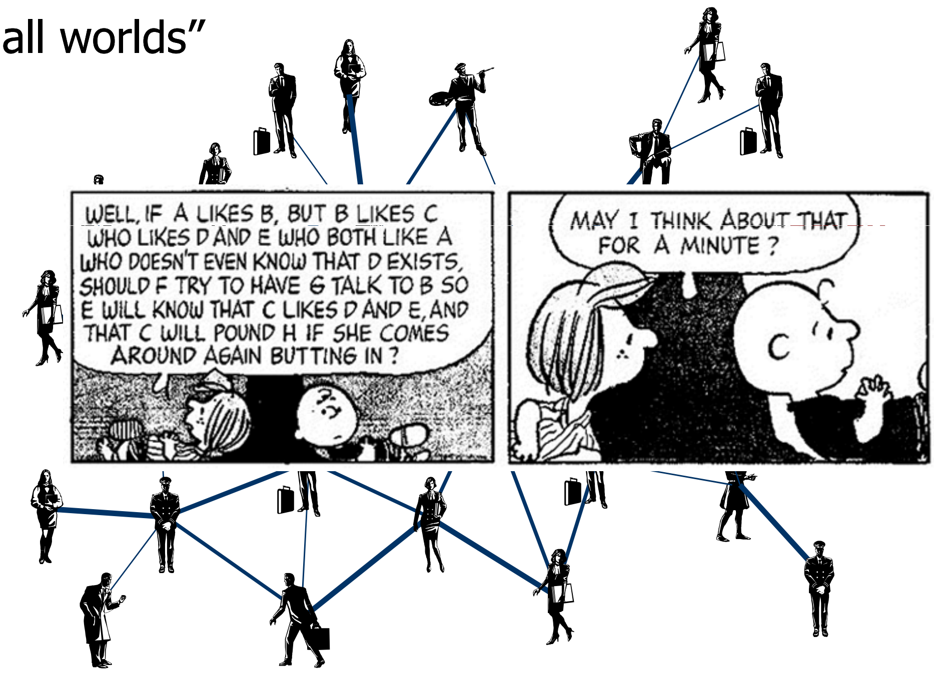





# Network Properties

**Network Diameter**

"small worlds"



# Network Properties

**Clustering coefficient**

| | |
|---|---|
| nbr(u) | Neighbors of the node u in the network. |
| k | Number of neighbors of u, i.e. \|nbr(u)\|. |
| max(u) | Maximum number of links among the neighbors of u, e.g. k*(k-1)/2. |

Clustering coefficient for the node u:
c(u) = (#links among neighbors of u) / max(u)

Clustering coefficient for the graph G:
C = average of c(u) for every node in G

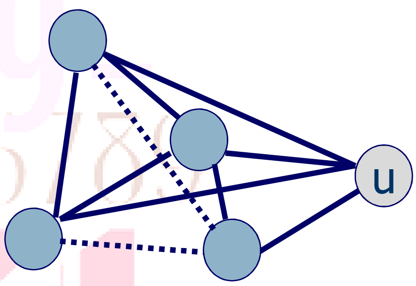


# Network Properties

**Clustering coefficient**

k = 4
m = 6

c(u) = 4/6 = 0.66




**0 <= c(u) <= 1**

Similarity of u neighbors to a clique (complete graph).

Informal interpretation:
"My friends tend to be friends among them."

# Network Properties

**Clustering coefficient** for some real networks

| Network | N | C | $C_{rand}$ | L |
|---|---|---|---|---|
| WWW | 153127 | 0.1078 | 0.00023 | 3.1 |
| Internet | 3015-6209 | 0.18-0.30 | 0.001 | 3.7-3.76 |
| Actor | 225226 | 0.79 | 0.00027 | 3.65 |
| Coauthorship | 52909 | 0.43 | 0.00018 | 5.9 |
| Metabolic | 282 | 0.32 | 0.026 | 2.9 |
| Foodweb | 134 | 0.22 | 0.06 | 2.43 |
| C. elegance | 282 | 0.28 | 0.05 | 2.65 |

Clustering coefficient (C): $C>C_{rand}$

Path length (L): $L<L_{rand}$

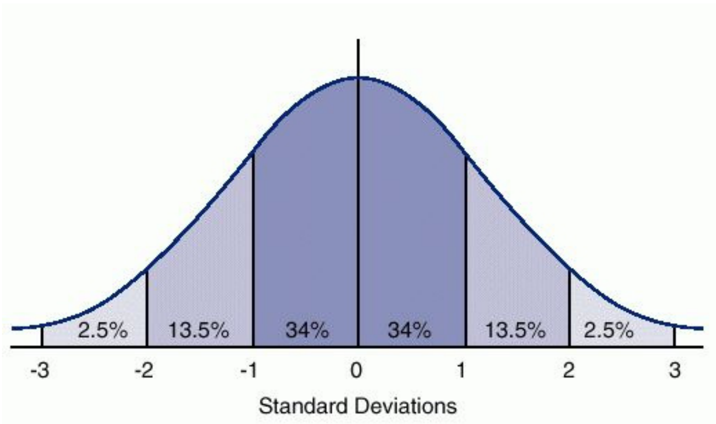


# Network Properties

**Node degree distribution**

Normal distribution

Parameters: Average & deviation

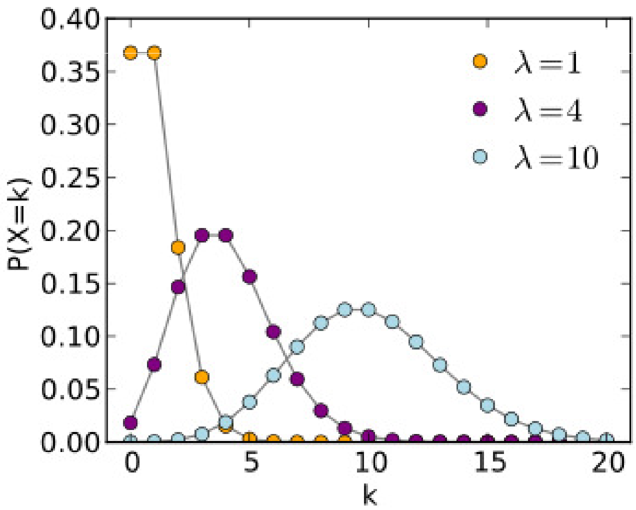


## Network Properties

**Node degree distribution**

Poisson distribution

Single parameter: λ (mean & deviation)



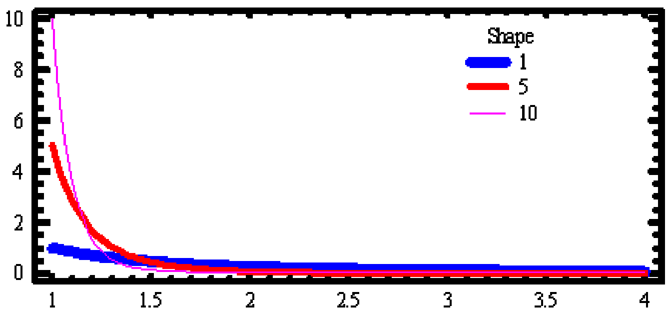



## Network Properties

**Node degree distribution**

Pareto distribution (a.k.a. "power law")

Single parameter: α




**$P(x) \sim x^{-\alpha}$**

**The Pareto principle (the "80-20 rule"):**

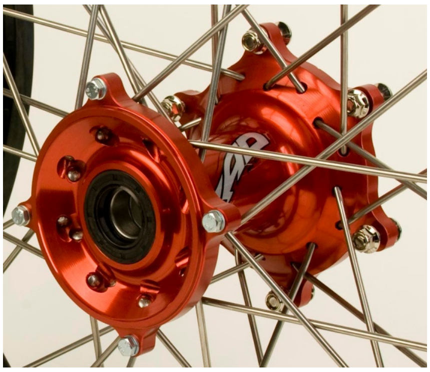
20% of the population controls 80% of the wealth.

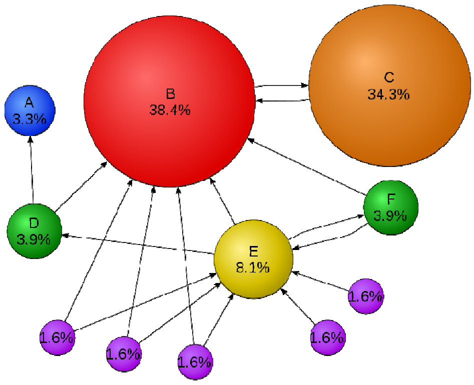
## Network Properties

**Node degree distribution**

Hubs

Small number of nodes with a very high degree.






- Hubs appear with power laws (**P(x)** ~ $x^{-\alpha}$), but not with normal/binomial/Poisson distributions.



## Network Properties

**Node degree distribution**

Log-log plot

- **Pareto distribution**
  - $\log(\Pr[X = x]) = \log(1/x^{\alpha}) = -\alpha \log(x)$
  - Linear, $-\alpha$ slope.
- **Normal distribution**
  - $\log(\Pr[X = x]) = \log(a \exp(-x^2/b)) = \log(a) - x^2/b$
  - Nonlinear, concave around the average.
- **Poisson distribution**
  - $\log(\Pr[X = x]) = \log(\exp(-\lambda)\ \lambda^x/x!)$
  - Nonlinear.

# Network Properties

**Node degree distribution**

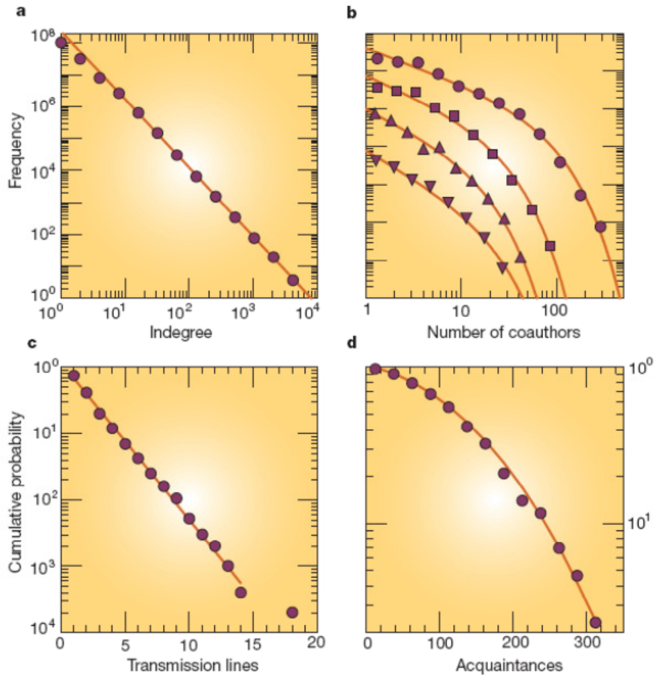
Log-log plot

**a** **WWW**
power law

**b** **Coauthorship networks**
power law with exponential cutoff

**c** **Power grid**
exponential

**d** **Social network**
Gaussian




# Network Models

"Natural" networks tend to have…

- One (or a few) connected components.
  - Independent of network size.
- A small diameter ("six degrees of separation").
  - Constant, logarithmically increasing, or even decreasing with network size.
- High clustering ("communities").
  - Much larger than expected from a random network (and, even so, with a small diameter!).
- A mixture of connections.
  - Local vs. "long-distance" connections

Do they share some "universal" features?

# Network Models

- Random networks.
- Random-biased networks.
- Small-world networks.
- Scale-free networks.
- Hierarchical & modular networks.
- Affiliation networks.

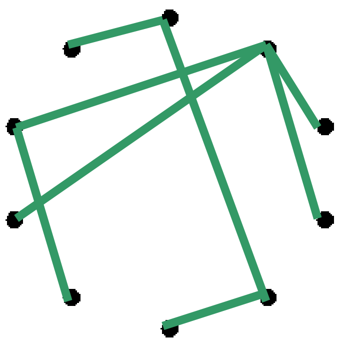


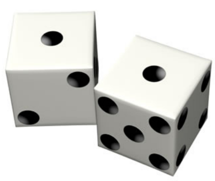
# Network Models

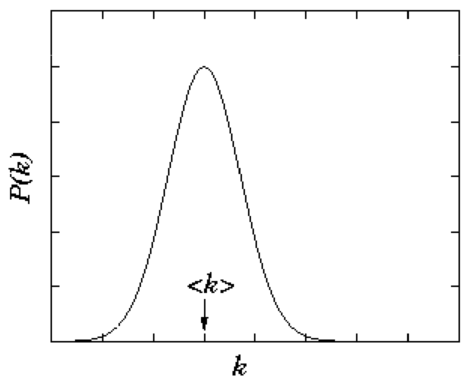
**Random Networks**

Erdös-Rényi model

- Small number of connected components (typically one).
- Low clustering coefficient.
- Poisson distribution.



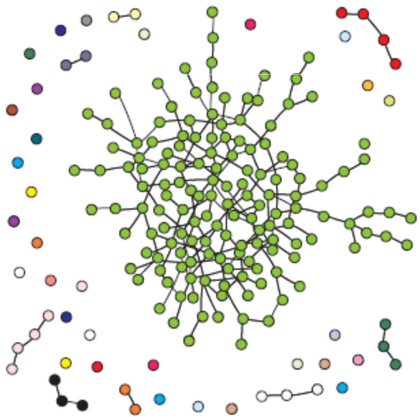


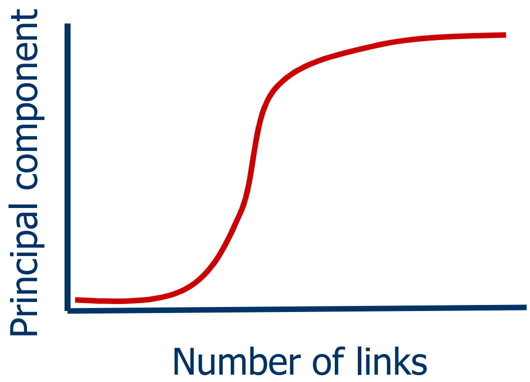


# Network Models

**Random Networks**

Erdös-Renyi model



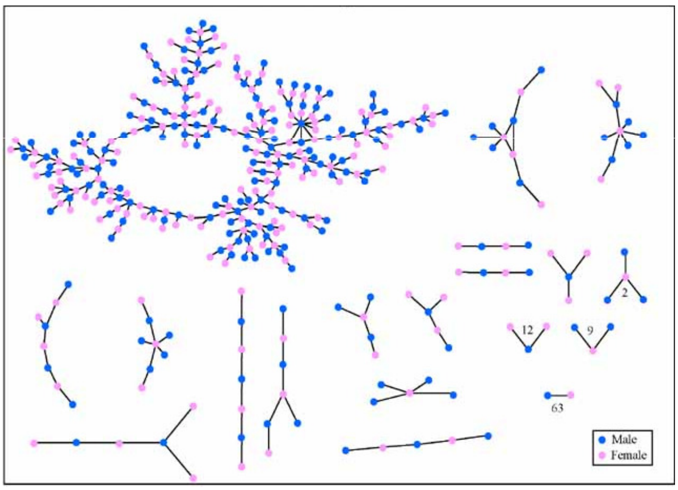


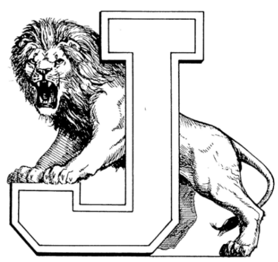



# Network Models

**Random Networks**

Example: Romantic relationships in the Add Health data set.



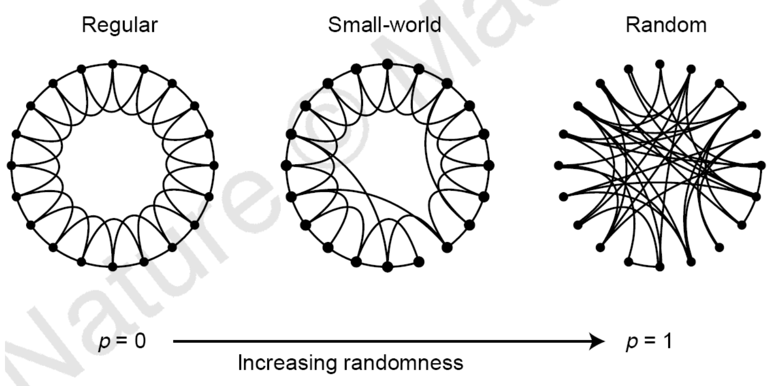
Peter S. Bearman, James Moody & Katherine Stovel:
"Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks"
American Journal of Sociology, 110(1):44–91, July 2004

# Network Models

## Small-World Networks

Watts & Strogatz model

- Small number of connected components (typically one).
- **Small diameter.**
- Poisson distribution.
- **High clustering coefficient.**

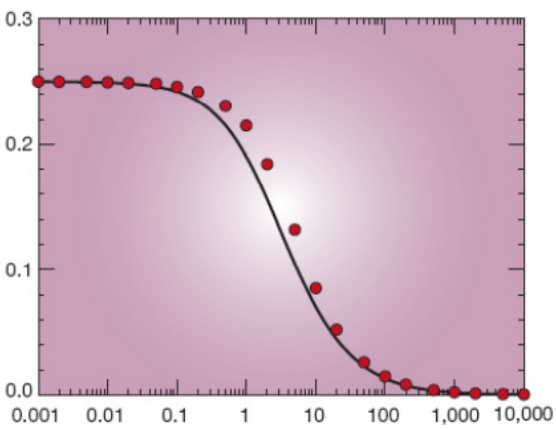





# Network Models

## Small-World Networks

Watts & Strogatz model



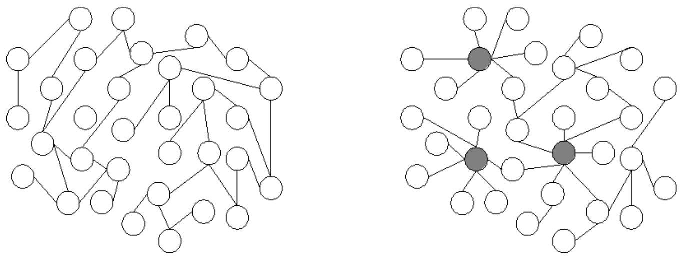

Average path length, normalized by system size, plotted as a function of the average number of shortcuts.

# Network Models

**Scale-Free Networks**

Barabási & Albert model

- Small number of connected components (typically one).
- Small diameter.
- **Pareto distribution.**
- **Small clustering coefficient.**
- **Hubs.**



(a) Random network (b) Scale-free network



# Network Models

**Scale-Free Networks**

Barabási & Albert model

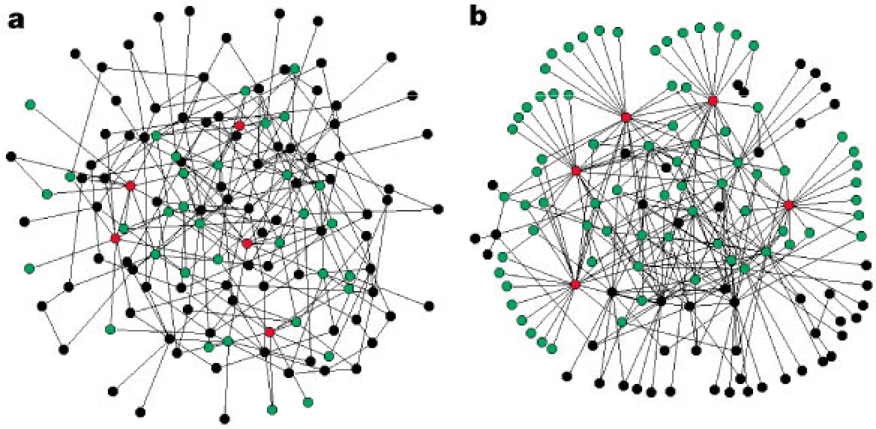
“Natural” interpretation of the model:

- Variable number of nodes:
  Network grows as new nodes are added.
- **Preferential attachment**:
  The more connected a node is,
  the more likely it is to receive new links
  (“rich get richer” or Matthew effect).

# Network Models

**Scale-Free Networks**

Barabási & Albert model




Exponential model…
… without hubs.

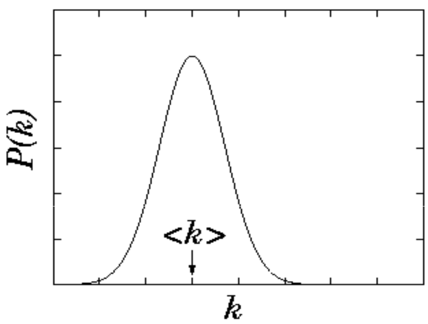
Scale-free model…
… with hubs.

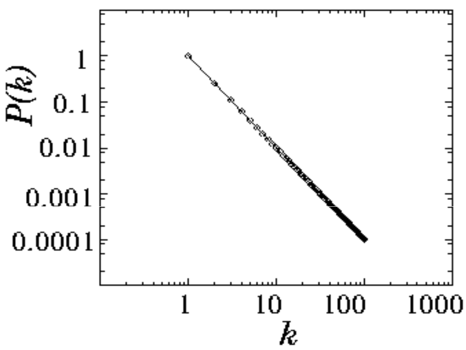


# Network Models

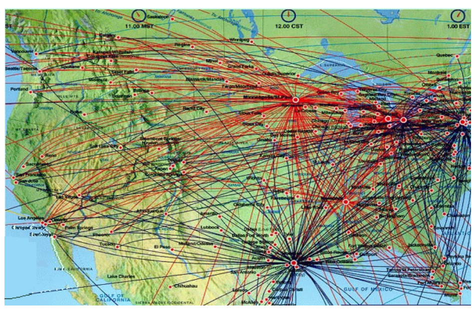
Poisson

Pareto (power law)







# Network Models

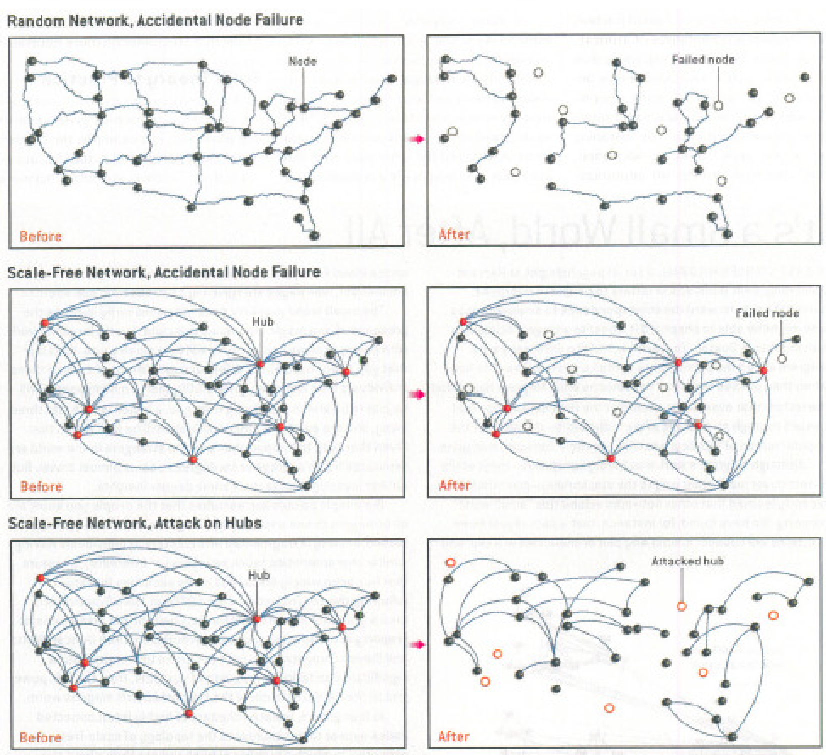
## Scale-Free Networks

Features

- Self-organization traits: Links are not random (a feature found in many complex systems).
- Tolerance to random attacks, which easily disrupt random networks but not scale-free networks.
- Vulnerability to targeted attacks: "Hubs" are essential to maintain connectedness.



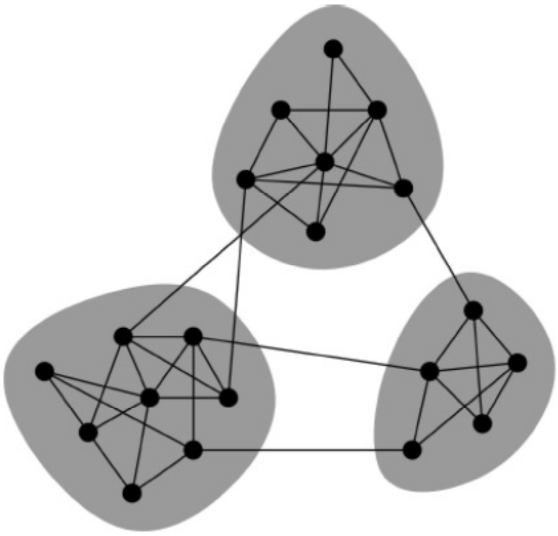
# Network Models



# Network Models

**Hierarchical/Modular Networks**

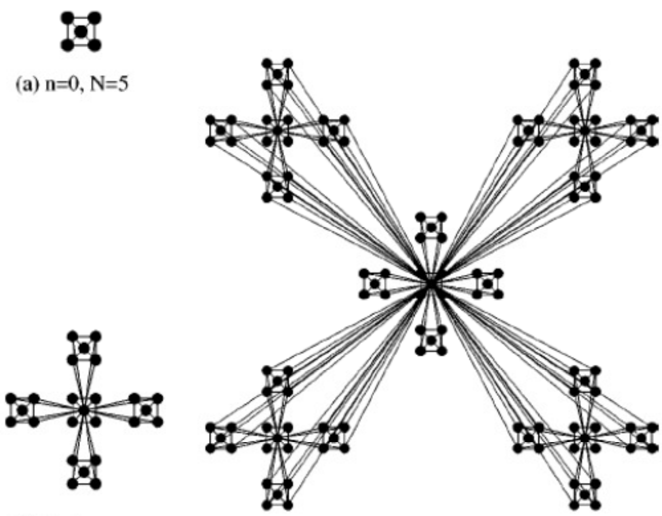
- Hierarchical organization.
- Hubs.
- **Cliques**.





# Network Models

**Hierarchical/Modular Networks**



(a) n=0, N=5

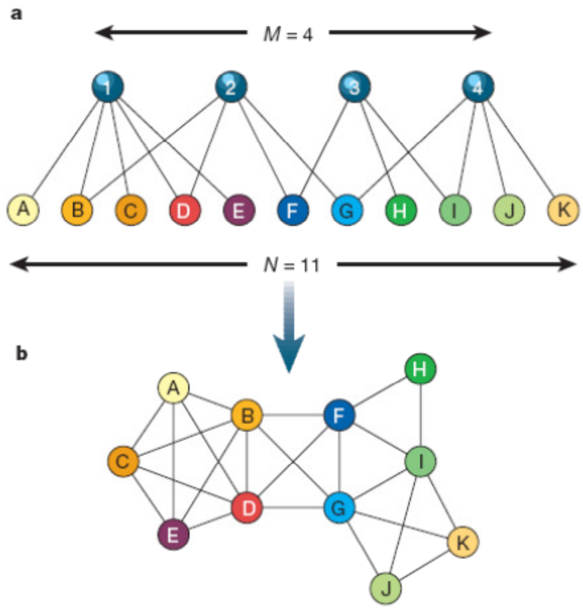
(b) n=1, N=25

(c) n=2, N=125

# Network Models

**Affiliation Networks**

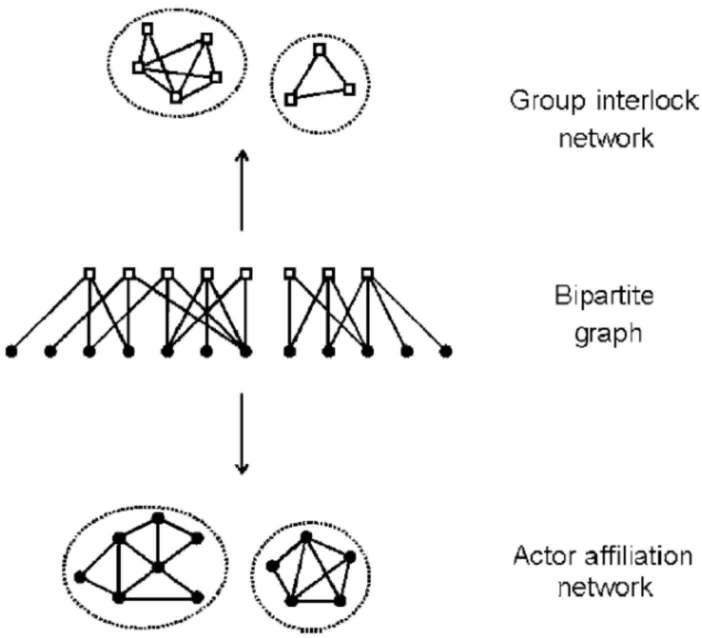
Bipartite graph to model social interactions:



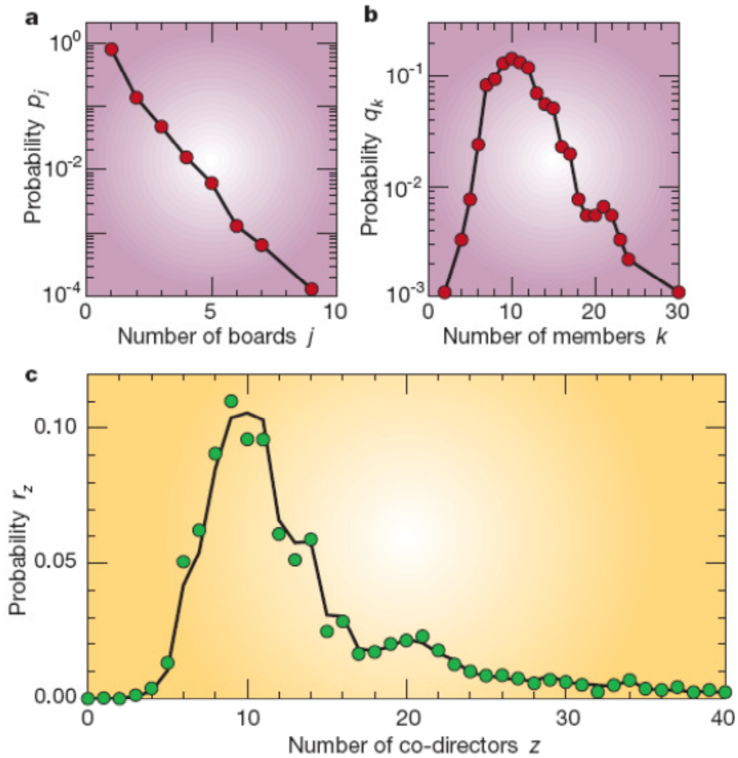



# Network Models

**Affiliation Networks**



# Network Models






# Network Structure & Dynamics

The countless ways in which network structures affect our lives make it critical to understand:

1. How network structure affect behavior.
2. Which network structure is likely to emerge.

# Network Structure & Dynamics

A complex system is a system composed of interconnected parts that, as a whole, exhibit one or more properties (behavior) not obvious from the properties of the individual parts (i.e. emergence).



# Network Structure & Dynamics

**Research problems**

- Search on networks
  (with partial local information)
- Diffusion problems:
  epidemics, social contagion (ideas, fads, products…)
- Analysis of network properties
  e.g. robustness/vulnerability

# Network Structure & Dynamics

## From an algorithmic point of view…

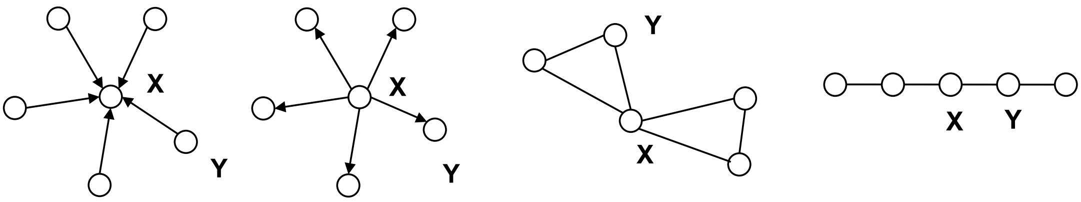
- Objects:
  - Ranking (HITS, PageRank…).
  - Classification & anomaly detection.
  - Clustering & community detection.
  - Object identification (e.g. "entity resolution").
- Links:
  - Link prediction.
- Graphs:
  - Subgraph detection.
  - Graph classification.
  - Graph generation models.
  - …



# Network Structure: Centrality

## Different notions of centrality

In each of the following networks, X has higher centrality than Y according to a particular measure



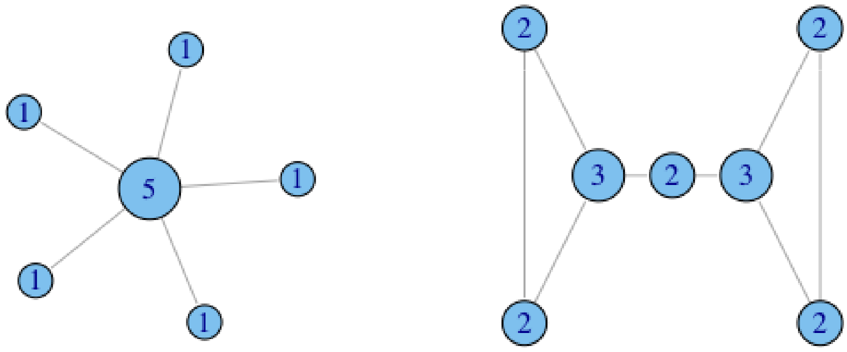

in-degree out-degree betwenness closeness

[Lada Adamic, "Social Network Analysis", https://www.coursera.org/course/sna]

# Network Structure: Centrality

**Degree**


Nodes with more connections are more central…



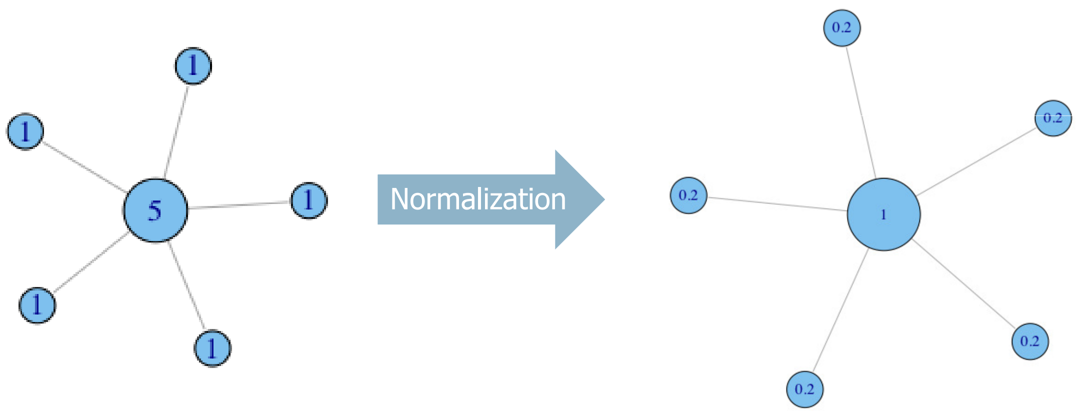

[Lada Adamic, "Social Network Analysis", https://www.coursera.org/course/sna]



# Network Structure: Centrality

**Degree**

Nodes with more connections are more central…

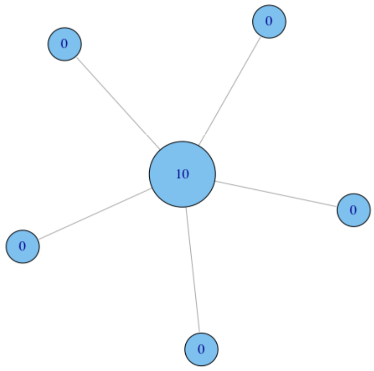


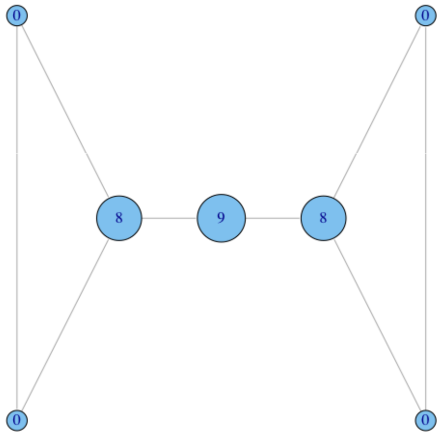

[Lada Adamic, "Social Network Analysis", https://www.coursera.org/course/sna]

# Network Structure: Centrality

**Betweenness**


Brokerage (not captured by degree)



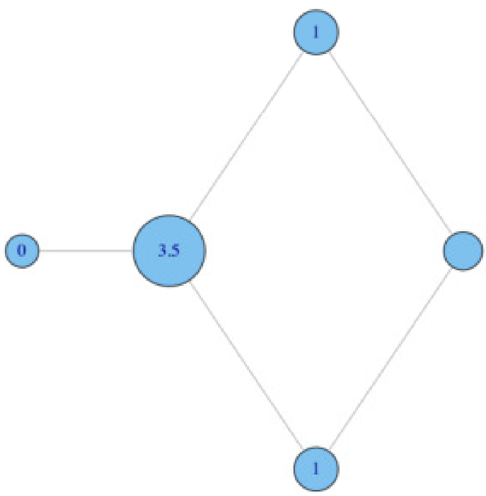

IDEA: How many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?



# Network Structure: Centrality

**Betweenness**

Brokerage (not captured by degree)

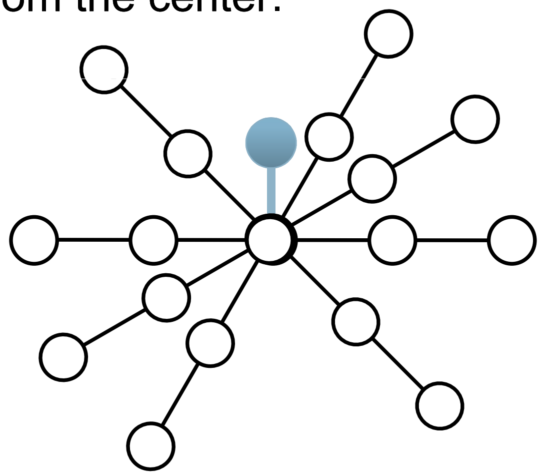



Partial credit for lying in one of several shortest paths…

# Network Structure: Centrality

**Closeness**

When it is not so important to have many connections, nor be between others, but be in the middle of things... not too far from the center.



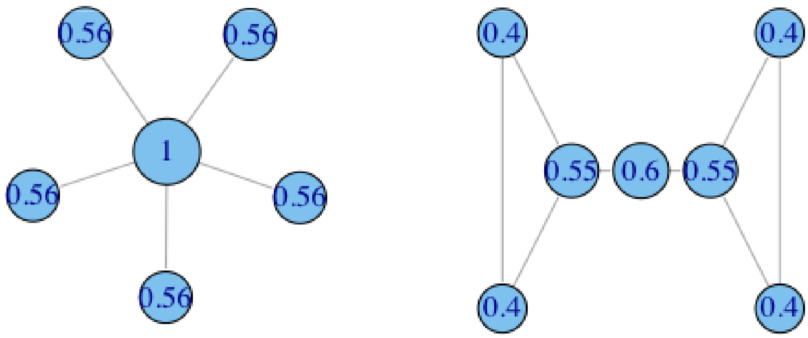
[Lada Adamic, “Social Network Analysis”, https://www.coursera.org/course/sna]



# Network Structure: Centrality

**Closeness**


based on the length of the average shortest path between a node and all other nodes in the network.

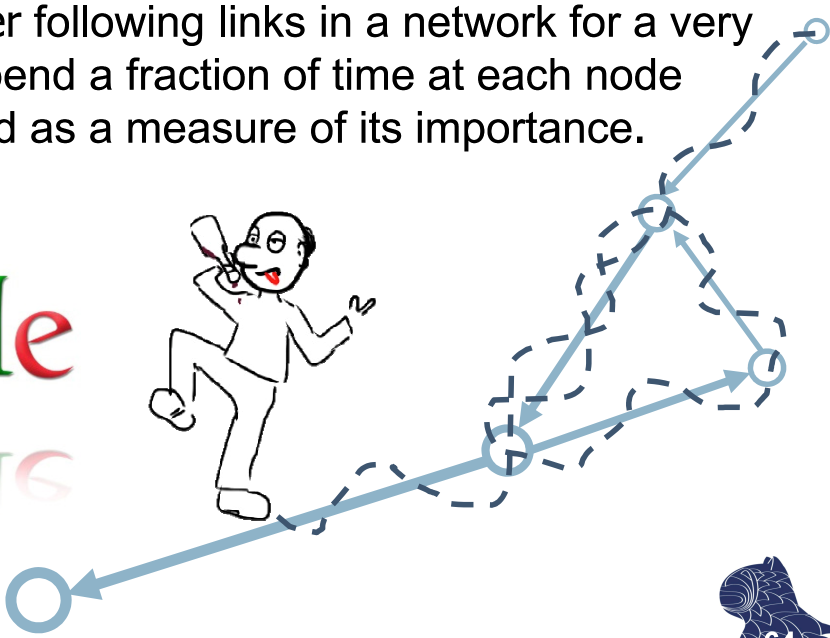


# Network Structure: Centrality

## PageRank

A random walker following links in a network for a very long time will spend a fraction of time at each node that can be used as a measure of its importance.

Google

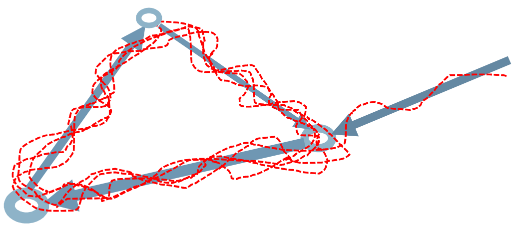


[Lada Adamic, “Social Network Analysis”, https://www.coursera.org/course/sna]



# Network Structure: Centrality

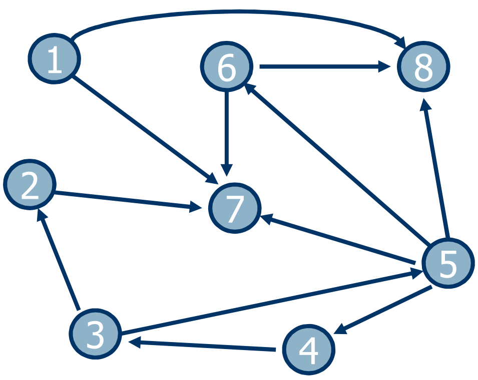
## PageRank

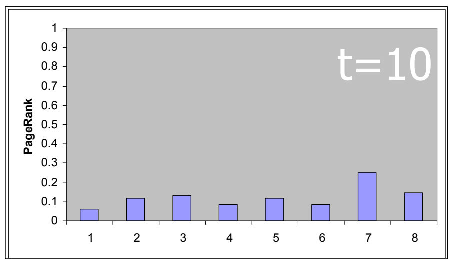


Problem: Stuck in the network

Solution: Teleportation

A random jump to anywhere else with a given probability.

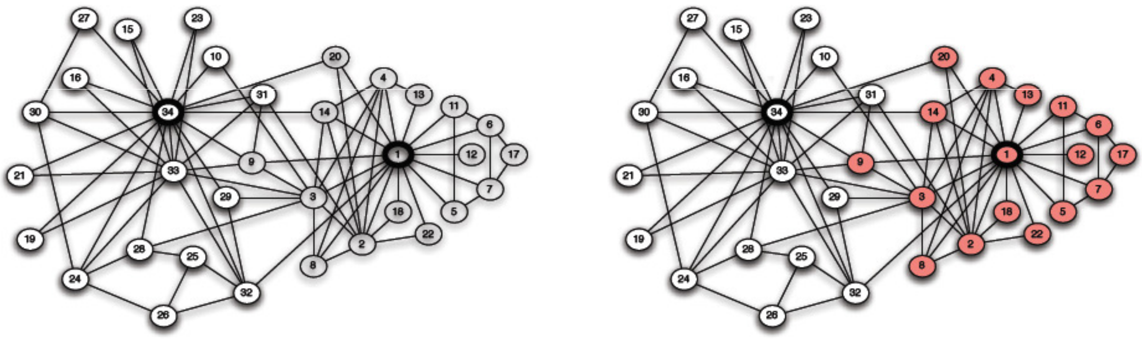


# Network Structure: Communities

**Community detection (i.e. clustering)**

Identification of groups of nodes within a network…



(a) *Karate club network* (b) *After a split into two clubs*

David Easley & Jon Kleinberg:
"Networks, Crowds & Markets: Reasoning About a Highly Connected World",
http://www.cs.cornell.edu/home/kleinber/networks-book/



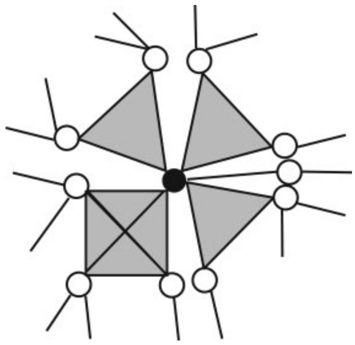
# Network Structure: Communities

**Heuristics**

- Mutual ties
- Frequency of ties within a community (cliques & k-cores)
- Closeness/reachability of community members (n-cliques)
- Relative frequency of ties within a community
  (ties among members compared to ties to non-members)

**Cliques & k-cores**

- Cliques (complete subgraphs)
  - A single missing links disqualifies the clique
  - Overlapping cliques



- K-cores
  (every node connected to at least k other nodes)





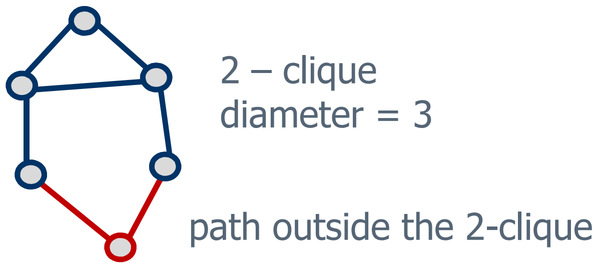
**n-cliques**

Maximal distance between any two nodes is n

IDEA: Information flow throw intermediaries.

Problems:

- Diameter > n
- Disconnected n-cliques




Solution: **n-clubs** (maximal subgraphs of diameter n)

# Network Structure: Communities

## Example: Political blogs

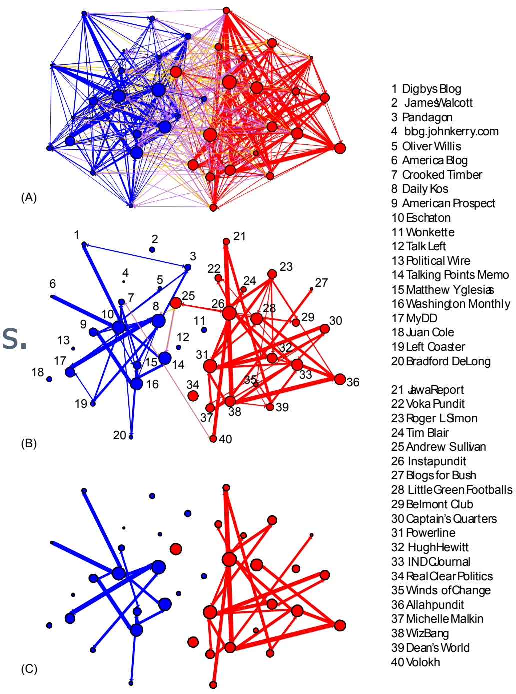
A) All citations between blogs.

B) Blogs with at least 5 citations in both directions.

C) Edges further limited to those exceeding 25 combined citations.

*only 15% of the citations bridge communities*

[Adamic & Glance, LinkKDD2005]

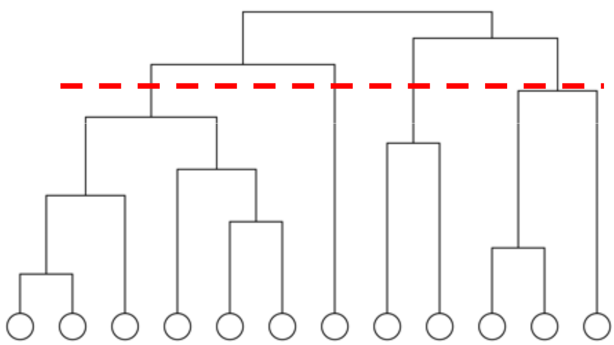





# Network Structure: Communities

## Community detection algorithms

Hierarchical clustering

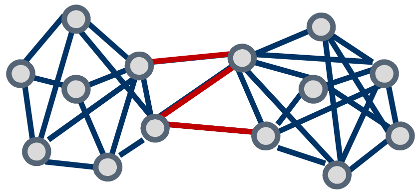


Michelle Girvan & Mark E.J. Newman:
"Community structure in social and biological networks"
PNAS **99**(12):7821–7826, 2002
doi:10.1073/pnas.122653799

# Network Structure: Communities

**Betweenness clustering**

Hierarchical clustering using edge betweenness



compute the betweenness of all edges
while (betweenness of any edge > threshold)
    remove edge with highest betweenness
    recalculate betweenness

- Betweenness clustering is inefficient due to the need to recompute edge betweenness in every iteration.



# Network Structure: Communities

**Modularity clustering**

- Consider links that fall within a community (vs. links between a community and the rest of the network)

Modularity Q

if vertices are in the same community

$$Q = \frac{1}{2m} \sum_{vw} \left[ A_{vw} - \frac{k_v k_w}{2m} \right] \delta(c_v, c_w)$$

adjacency matrix

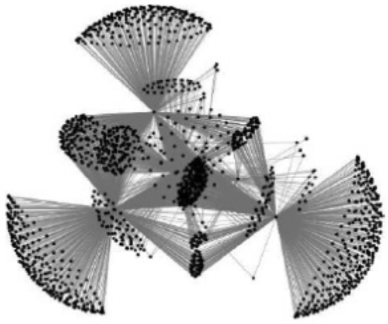
probability of an edge between two vertices is proportional to their degrees

NOTE: For a random network, Q=0

# Network Structure: Communities

**Modularity clustering**

Algorithm



start with all vertices as isolates
do
    join clusters with the greatest increase in modularity ($\Delta$Q)
while ($\Delta$Q > 0)

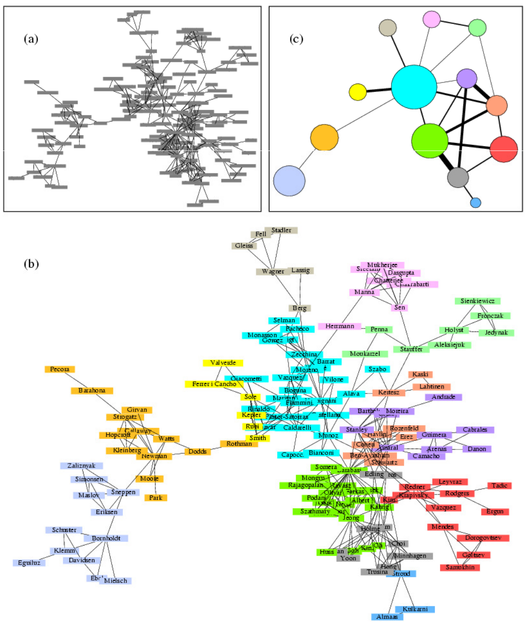
Aaron Clauset, Mark E. J. Newman, Cristopher Moore:
"Finding community structure in very large networks"
Physical Review E 70(6):066111, 2004
doi:10.1103/physreve.70.066111



# Network Structure: Communities

**Modularity clustering**

An application: Visualization of large networks (Gephi)



# Network Structure: Communities

## Limitations of current community detection methods

- Scalability: Identification of large communities.
- Existence of overlapping communities in large networks.
- Unrealistic models (algorithms make oversimplified assumptions over the networks or community structures, but perform poorly against real world data sets).
- Heuristics without performance guarantees (for those heuristic algorithms that work well in practice, there is no performance guarantee over the quality of their output).



# Bibliography

Networks: Origins & Applications (social networks, Web…)


- Stanley Milgram: **The small world problem**. Psychology Today, 2:60-67 (1967)
- Phillip W. Anderson: **More is different**. Science, 177:393-396 (1972)
- Mark S. Granovetter: **The strength of weak ties**. American Journal of Sociology, 78:1360-1380 (1973)
- Stanley Wasserman & Katherine Faust: **Social Network Analysis: Methods and Applications**. Cambridge University Press, 1994
- John P. Scott: **Social Network Analysis**, 2nd edition. Sage Publications Ltd., 2000.
- Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins & Janet Wiener: **Graph structure in the Web**. Computer Networks 33:309–320 (2000)
- Steven H. Strogatz: **Exploring Complex Networks**. Nature, 410:268-275 (2001)
- Albert-Laszlo Barabasi: **Linked: How Everything Is Connected to Everything Else and What It Means**. Plume, 2003. ISBN 0452284392
- Duncan J. Watts: **Six Degrees: The Science of a Connected Age**. W. W. Norton & Company, 2004. ISBN 0393325423
- Jure Leskovec, Jon M. Kleinberg & Christos Faloutsos: **Graphs over time: densification laws, shrinking diameters and possible explanations**. KDD'2005

# Bibliography

## Network Models


- Paul Erdös & Alfred Rényi: **On the evolution of random graphs**. Mathematical Institute of the Hungarian Academy of Sciences, 5:17-61 (1960) reprinted in Duncan, Barabasi & Watts (eds.): "The Structure and Dynamics of Networks"
- Ray Solomonoff & Anatol Rapoport: **Connectivity of random nets**. Bulletin of Mathematical Biophysics, 13:107-117 (1951) reprinted in Duncan, Barabasi & Watts (eds.): "The Structure and Dynamics of Networks"
- Duncan J. Watts & Steven H. Strogatz: **Collective dynamics of 'small-world' networks**. Nature, 393:440-442 (1998)
- Albert-László Barabási & Réka Albert: **Emergence of scaling in random networks**. Science, 286:509-512 (1999)
- Réka Albert, Hawoong Jeong & Albert-László Barabási: **Error and attack tolerance of complex networks**. Nature 406:378-382 (2000)
- M.E.J. Newman, S.H. Strogatz & D.J. Watts: **Random graphs with arbitrary degree distributions and their applications**. Physical Review E, 64:026118 (2001)
- M.E.J. Newman, S.H. Strogatz & D.J. Watts: **Random graphs models of social networks**. PNAS 99:2566-2572 (2002)
- Erzsébet Ravasz & Albert-László Barabási: **Hierarchical organization in complex networks**. Physical Review E, 67:026112 (2003)
- Mark Newman: **The structure and function of complex networks**. SIAM Review 45:167-256 (2003)




# Bibliography

## Community Detection


- Michelle Girvan & Mark E.J. Newman: **Community structure in social and biological networks**, PNAS 99(12):7821–7826 (2002)
- Aaron Clauset, Mark E. J. Newman & Cristopher Moore: **Finding community structure in very large networks**, Physical Review E 70(6):066111 (2004)
- Gergely Palla, Imre Derényi, Illés Farkas & Tamás Vicsek: **Uncovering the overlapping community structure of complex networks in nature and society**, Nature 435:814-818 (2005)
- Jure Leskovec, Kevin J. Lang, Anirban Dasgupta & Michael W. Mahoney: **Statistical Properties of Community Structure in Large Social and Information Networks**, International World Wide Web Conference, WWW'08 (2008). Extended version: **Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters**, arxiv:0810.1355 (2008)
- Martin Rosvall & Carl T. Bergstrom: **Maps of random walks on complex networks reveal community structure**, PNAS 105(4):1118-1123 (2008)
- Jure Leskovec, Kevin J. Lang & Michael W. Mahoney: **Empirical Comparison of Algorithms for Network Community Detection,** WWW 2010 (2010)
- Santo Fortunato: **Community detection in graphs**. Physics Reports, 486(3-5):75-174, (2010).
- S. Arora, R. Ge, S. Sachdeva & G. Schoenebeck: **Finding overlapping communities in social networks: toward a rigorous approach**. Proceedings of the 13th ACM Conference on Electronic Commerce, EC '12, pp. 37-54 (2012)

# Bibliography

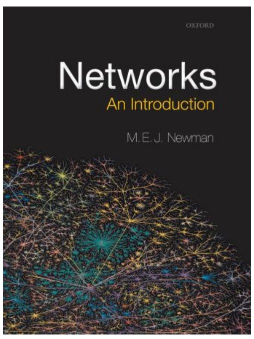
## Search on Networks

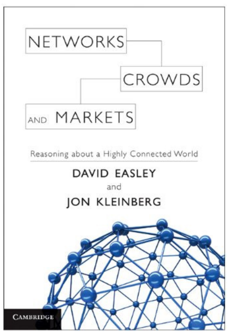

- Sergey Brin & Lawrence Page: **The anatomy of a large-scale hypertextual Web search engine**. Computer Networks and ISDN Systems, April 1998
- David Gibson, Jon M. Kleinberg & Prabhakar Raghavan: **Inferring Web Communities from Link Topology**. ACM Conference on Hypertext and Hypermedia, June 1998
- Jon M. Kleinberg: **Authoritative sources in a hyperlinked environment**. Journal of the ACM, September 1999
- Toby Walsh: **Search in a Small World**. IJCAI'1999
- Jon M. Kleinberg. **Navigation in a Small World**. Nature, August 2000.
- Jon M. Kleinberg: **The small-world phenomenon: An algorithm perspective**. STOC'2000
- Scott White & Padhraic Smyth: **Algorithms for Estimating Relative Importance in Networks**. KDD'2003
- Hanghang Tong & Christos Faloutsos: **Center-Piece Subgraphs: Problem Definition and Fast Solutions**. KDD'2006
- Alekh Agarwal, Soumen Chakrabarti & Sunny Aggarwal: **Learning to Rank Networked Entities**. KDD'2006
- Jeffrey Davitz, Jiye Yu, Sugato Basu, David Gutelius & Alexandra Harris: **iLink: Search and Routing in Social Networks**. KDD'2007.

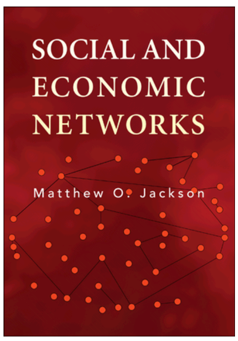



# Bibliography


- David Easley & Jon Kleinberg: **Networks, Crowds, and Markets: Reasoning About a Highly Connected World**. Cambridge University Press, 2010. ISBN 0521195330 http://www.cs.cornell.edu/home/kleinber/networks-book/
- Mark Newman: **Networks: An Introduction**. Oxford University Press, 2010. ISBN 0-19-920665-1
- Matthew O. Jackson: **Social and Economic Networks**, Princeton University Press, 2008. ISBN 0-691-13440-5

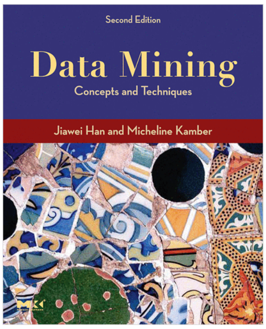



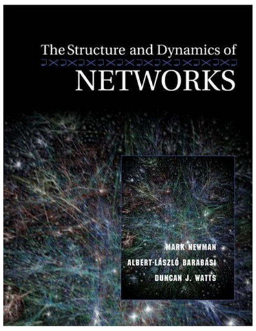



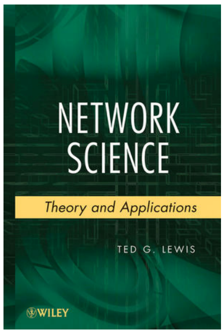



# Bibliography


- Jiawei Han & Micheline Kamber:
  **Data Mining: Concepts and Techniques** [2nd edition], section 9.2. Addison-Wesley, 2006. ISBN 1-55860-901-3
- Mark Newman, Albert-Laszlo Barabasi & Duncan J. Watts (editors): **The Structure and Dynamics of Networks**. Princeton University Press, 2006. ISBN 0-691-11357-2
- Ted G. Lewis: **Network Science: Theory and Applications**. Wiley, 2009. ISBN 0-470-33188-7




# Bibliography


- Albert-Laszlo Barabási: **Linked: How Everything Is Connected to Everything Else and What It Means**. Plume, 2003. ISBN 0452284392
- Duncan J. Watts: **Six Degrees: The Science of a Connected Age**. W. W. Norton & Company, 2004. ISBN 0393325423
- Albert-Laszlo Barabási: **Bursts: The Hidden Pattern Behind Everything We Do**. Dutton, 2010. ISBN 0525951601