# Overview of Complex Networks

## Complex Networks, Course 303A, Spring, 2009

Prof. Peter Dodds

Department of Mathematics & Statistics
University of Vermont







# Outline

# Class Admin

- Office hours:
  - Tuesday 2:30 pm–3:30 pm
  - Thursday 11:30 am–12:30 pm
- Course outline
- Projects

# Basic definitions

Complex: (Latin = with + fold/weave (com + plex))

Adjective

- Made up of multiple parts; intricate or detailed.
- Not simple or straightforward.

# Basic definitions

## Complex System—Basic ingredients:

- Relationships are nonlinear
- Relationships contain feedback loops
- Complex systems are open (out of equilibrium)
- Memory
- Modular (nested)/multiscale structure
- Opaque boundaries
- May result in emergent phenomena

# Basic definitions

Network: (net + work, 1500's)

Noun:

1. Any interconnected group or system
2. Multiple computers and other devices connected together to share information

Verb:

1. To interact socially for the purpose of getting connections or personal advancement
2. To connect two or more computers or other computerized devices

# Observation

- Many complex systems can be regarded as complex networks of physical or abstract interactions
- Opens door to mathematical and numerical analysis
- Dominant approach of last decade of a theoretical-physics/stat-mechish flavor.

# Basic definitions

Nodes = A collection of entities which have properties that are somehow related to each other

- e.g., people, forks in rivers, proteins, webpages, organisms,...

# Basic definitions

## Links = Connections between nodes

- links
  - may be real and fixed (rivers),
  - real and dynamic (airline routes),
  - abstract with physical impact (hyperlinks),
  - or purely abstract (semantic connections between concepts).
- Links may be directed or undirected.
- Links may be binary or weighted.

# Basic definitions

Node degree = Number of links per node

- Notation: Node $i$'s degree = $k_i$.
- $k_i = 0,1,2,\ldots$.
- Notation: the average degree of a network = $\langle k \rangle$ (and sometimes as $z$)

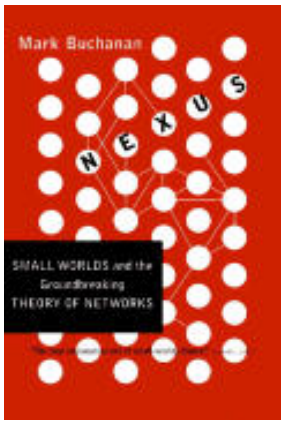
# Basic definitions

## Adjacency matrix:

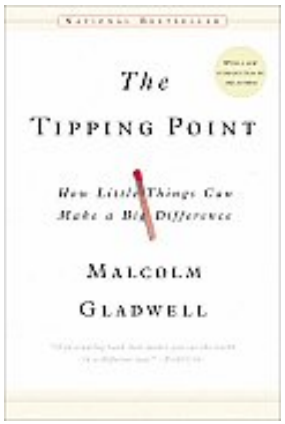
- We represent a graph or network by a matrix $A$ with link weight $a_{ij}$ for nodes $i$ and $j$ in entry $(i, j)$.
- e.g.,

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

# Books

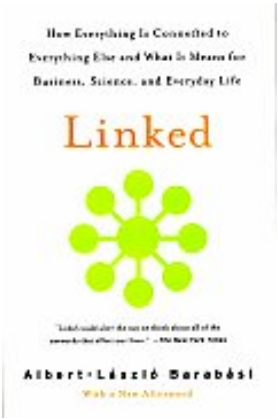



Nexus: Small Worlds and the Groundbreaking Science of Networks—Mark Buchanan

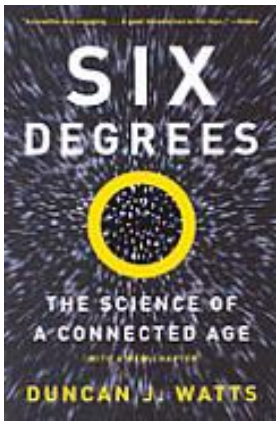



The Tipping Point: How Little Things can make a Big Difference—Malcolm Gladwell [7]

# Books

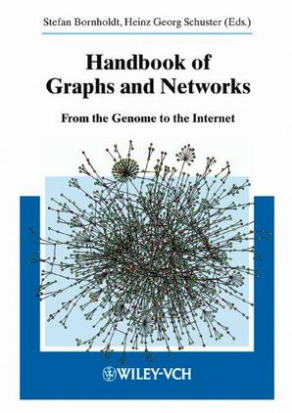



Linked: How Everything Is Connected to Everything Else and What It Means—Albert-Laszlo Barabási

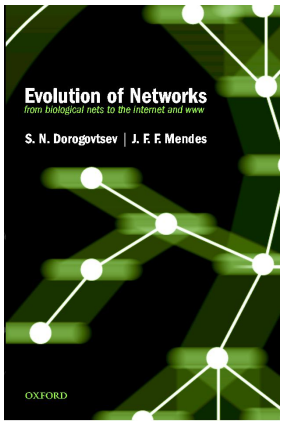



Six Degrees: The Science of a Connected Age—Duncan Watts [17]

# Books




Handbook of Graphs and Networks—editors: Stefan Bornholdt and H. G. Schuster [3]




Evolution of Networks—S. N. Dorogovtsev and J. F. F. Mendes [6]

# Books


Social Network Analysis—Stanley Wasserman and Kathleen Faust [16]

# Books

## Numerous others:

- Complex Social Networks—F. Vega-Redondo [15]
- Fractal River Basins: Chance and Self-Organization—I. Rodríguez-Iturbe and A. Rinaldo [12]
- Random Graph Dynamics—R. Durette
- Scale-Free Networks—Guido Caldarelli
- Evolution and Structure of the Internet: A Statistical Physics Approach—Romu Pastor-Satorras and Alessandro Vespignani
- Complex Graphs and Networks—Fan Chung

# Examples

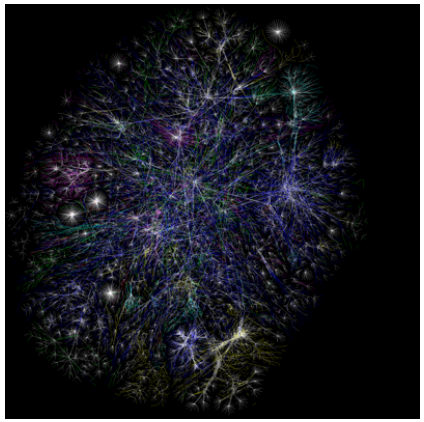
## What passes for a complex network?

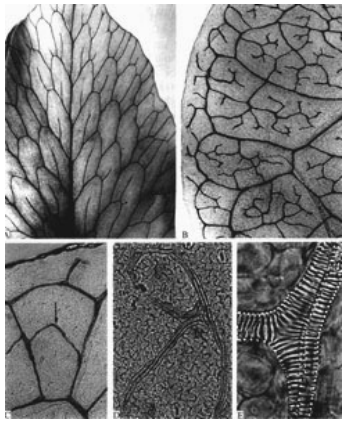
- Complex networks are large (in node number)
- Complex networks are sparse (low edge to node ratio)
- Complex networks are usually dynamic and evolving
- Complex networks can be social, economic, natural, informational, abstract, ...
- Isn't this graph theory?: Yes, but emphasis is on data and mechanistic explanations...

# Examples

## Physical networks

- River networks
- Neural networks
- Trees and leaves
- Blood networks
- The Internet
- Road networks
- Power grids





- Distribution (branching) versus redistribution (cyclical)

# Examples

## Interaction networks

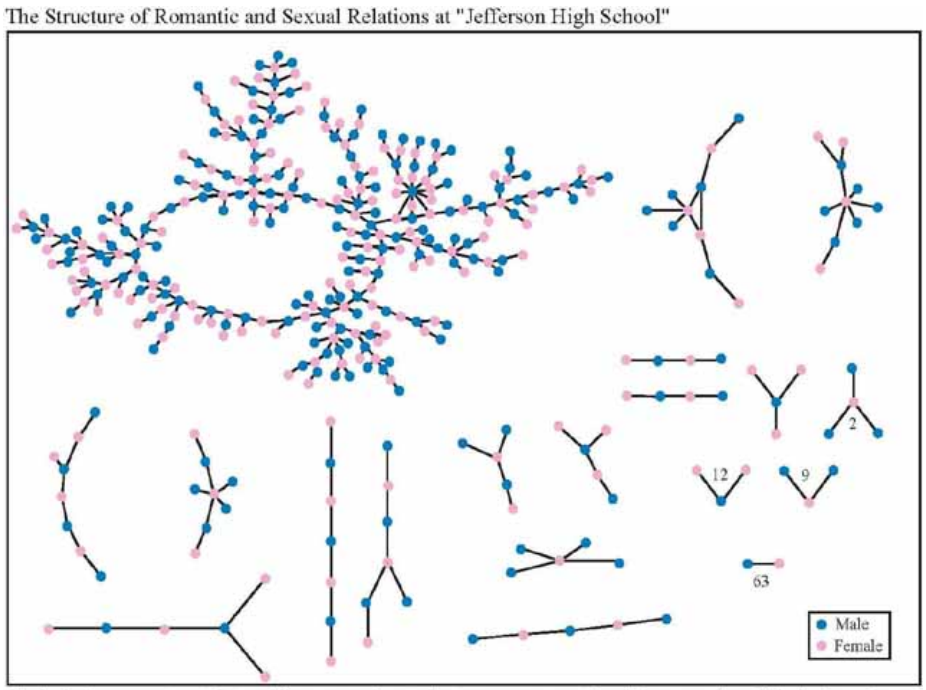
- The Blogosphere
- Biochemical networks
- Gene-protein networks
- Food webs: who eats whom
- The World Wide Web (?)
- Airline networks
- Call networks (AT&T)
- The Media

datamining.typepad.com (⊞)

# Examples

## Interaction networks: social networks

- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- myspace.com (⊞), facebook.com (⊞)

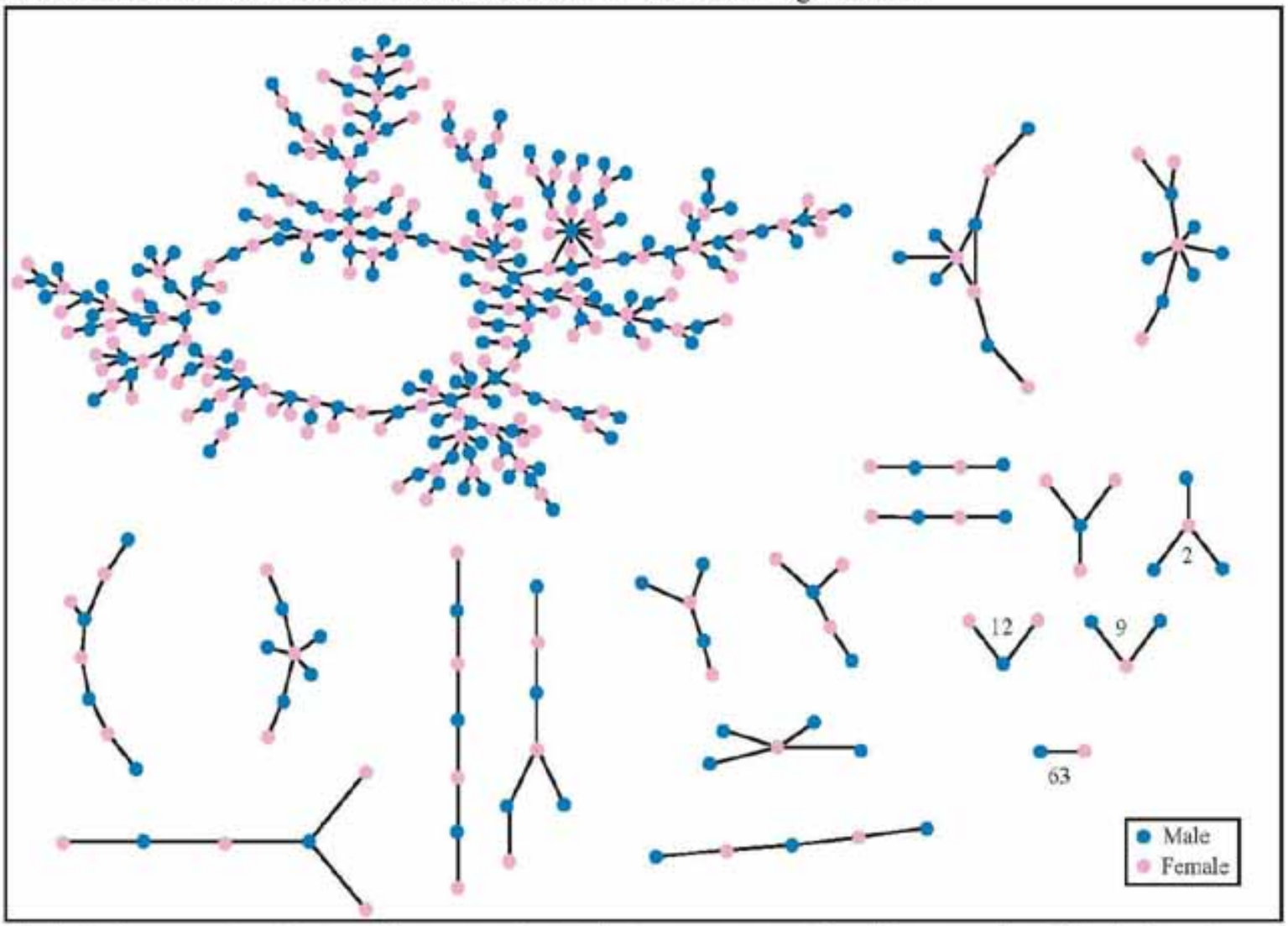



(Bearman *et al.*, 2004)

- 'Remotely sensed' by: email activity, instant messaging, phone logs (*cough*).

# Examples

The Structure of Romantic and Sexual Relations at "Jefferson High School"




Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

# Examples

## Relational networks

- Consumer purchases
  (Wal-Mart: $\approx$ 1 petabyte $= 10^{15}$ bytes)
- Thesauri: Networks of words generated by meanings
- Knowledge/Databases/Ideas
- Metadata—Tagging: del.icio.us (⊞), flickr (⊞)

**common tags** cloud | list

community daily dictionary education **encyclopedia**
english free imported info information internet knowledge
learning news **reference** research resource
resources search tools useful web web2.0 **wiki**
**wikipedia**

# Observations

## A notable feature of large-scale networks:

- Graphical renderings of complex networks are often just a big mess.

# Properties

## Some key aspects of real complex networks:

- degree distribution
- assortativity
- homophily
- clustering
- motifs
- modularity
- concurrency
- hierarchical scaling
- network distances
- centrality
- efficiency
- robustness
- + Coevolution of network structure and processes on networks.

# Properties

## 1. degree distribution $P_k$

- $P_k$ is the probability that a randomly selected node has degree $k$
- $k$ = node degree = number of connections
- ex 1: Erdős-Rényi random networks:

$$P_k = e^{-\langle k \rangle} \langle k \rangle^k / k!$$

- Distribution is Poisson

# Properties

## 1. degree distribution $P_k$

- ex 2: “Scale-free” networks: $P_k \propto k^{-\gamma} \Rightarrow$ ‘hubs’
- link cost controls skew
- hubs may facilitate or impede contagion

# Properties

## Note:

- Erdős-Rényi random networks are a *mathematical construct*.
- 'Scale-free' networks are growing networks that form according to a plausible mechanism.
- Randomness is out there, just not to the degree of a completely random network.

# Properties

## 2. assortativity/3. homophily:

- Social networks: Homophily (⊞) = birds of a feather
- e.g., degree is standard property for sorting: measure degree-degree correlations.
- Assortative network:[10] similar degree nodes connecting to each other.
  *Often social: company directors, coauthors, actors.*
- Disassortative network: high degree nodes connecting to low degree nodes.
  *Often techological or biological: Internet, WWW, protein interactions, neural networks, food webs.*

# Clustering

## 4. clustering:

- Your friends tend to know each other.
- Two measures:
  1. Watts & Strogatz [19]

$$C_1 = \left\langle \frac{\sum_{j_1 j_2 \in \mathcal{N}_i} a_{j_1 j_2}}{k_i(k_i - 1)/2} \right\rangle_i$$

  2. Newman [11]

$$C_2 = \frac{3 \times \#\text{triangles}}{\#\text{triples}}$$

# Properties

## First clustering measure:

- $C_1$ is the average fraction of pairs of neighbors who are connected.
- Fraction of pairs of neighbors who are connected is

$$\frac{\sum_{j_1 j_2 \in \mathcal{N}_i} a_{j_1 j_2}}{k_i(k_i-1)/2}$$

where $k_i$ is node $i$'s degree, and $\mathcal{N}_i$ is the set of $i$'s neighbors.

- Averaging over all nodes, we have

$$C_1 = \frac{1}{n}\sum_{i=1}^{n} \frac{\sum_{j_1 j_2 \in \mathcal{N}_i} a_{j_1 j_2}}{k_i(k_i-1)/2} = \left\langle \frac{\sum_{j_1 j_2 \in \mathcal{N}_i} a_{j_1 j_2}}{k_i(k_i-1)/2} \right\rangle_i$$

# Properties

- For sparse networks, $C_1$ tends to discount highly connected nodes.
- $C_2$ is a useful variant
- In general, $C_1 \neq C_2$.

# Properties

## Triples and triangles

- Nodes $i_1$, $i_2$, and $i_3$ form a triple around $i_1$ if $i_1$ is connected to $i_2$ and $i_3$.
- Nodes $i_1$, $i_2$, and $i_3$ form a triangle if each pair of nodes is connected
- The definition

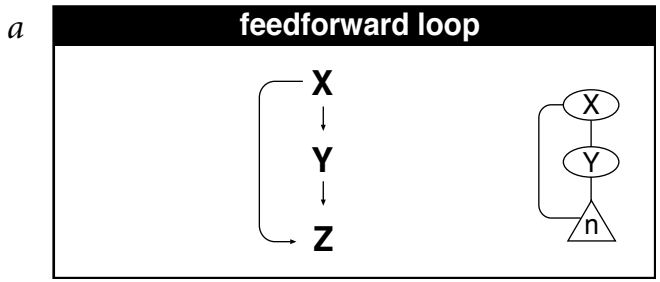
$$C_2 = \frac{3 \times \#\text{triangles}}{\#\text{triples}}$$

  measures the fraction of closed triples
- Social Network Analysis (SNA): fraction of transitive triples.
- The '3' appears because for each triangle, we have 3 closed triples.

# Properties

## 5. motifs:

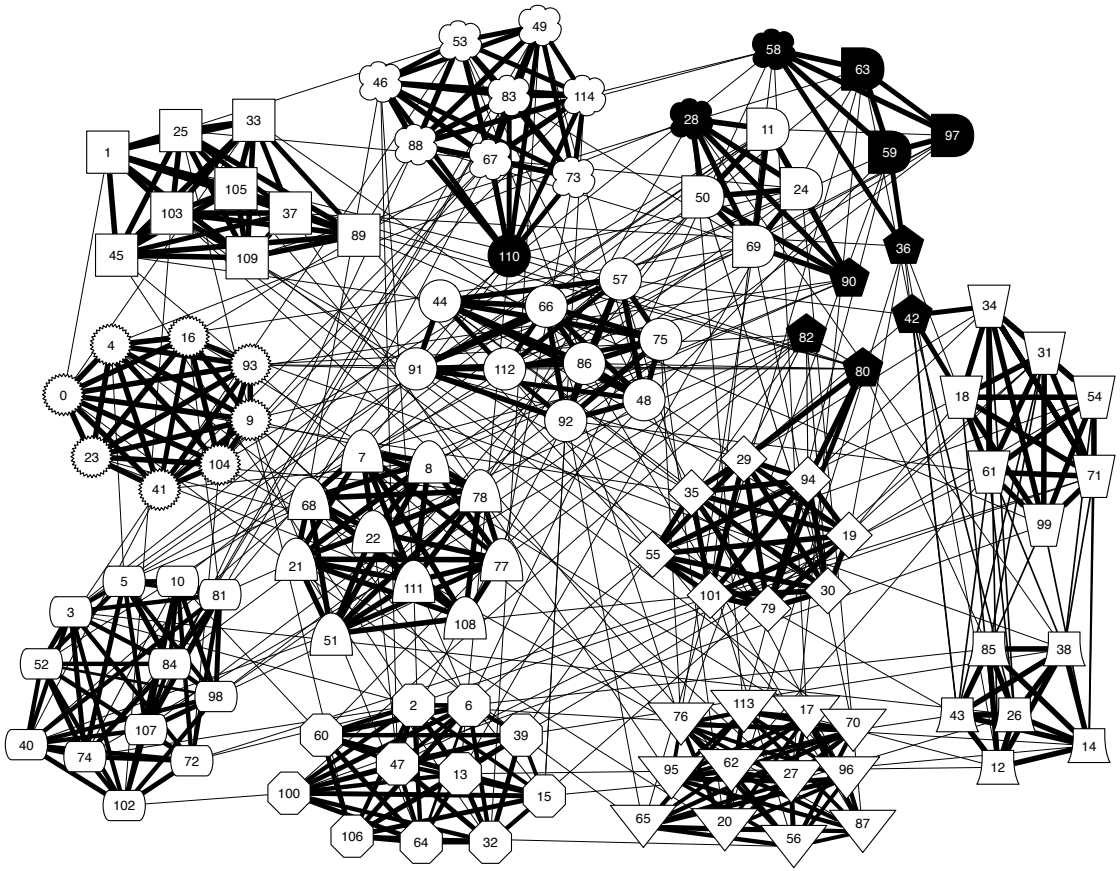
- small, recurring functional subnetworks
- e.g., Feed Forward Loop:




Shen-Orr, Uri Alon, *et al.* [13]

# Properties

## 6. modularity and structure/community detection:



Clauset *et al.*, 2006 [5]: NCAA football

# Properties

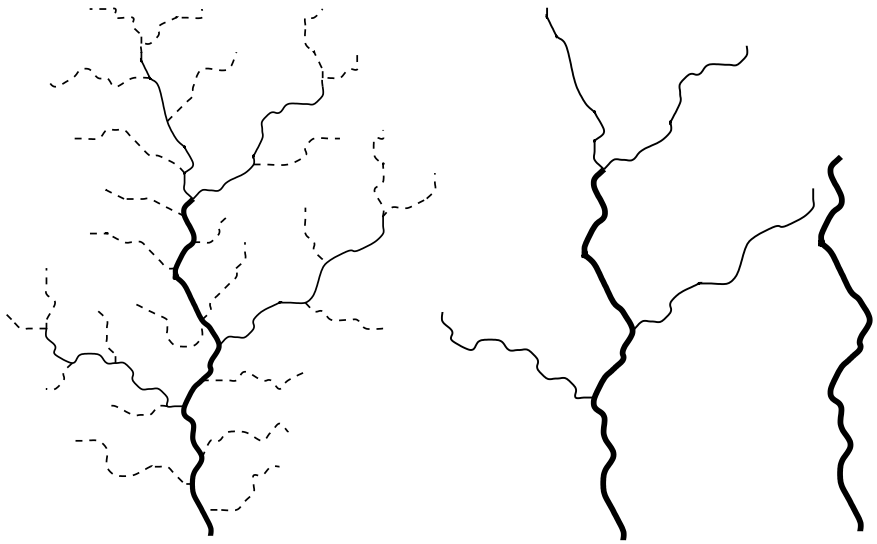
## 7. concurrency:

- transmission of a contagious element only occurs during contact
- rather obvious but easily missed in a simple model
- dynamic property—static networks are not enough
- knowledge of previous contacts crucial
- beware cumulated network data
- Kretzschmar and Morris, 1996 [9]

# Properties

## 8. Horton-Strahler ratios:

- Metrics for branching networks:
  - Method for ordering streams hierarchically
  - Number: $R_n = N_\omega / N_{\omega+1}$
  - Segment length: $R_l = \langle l_{\omega+1} \rangle / \langle l_\omega \rangle$
  - Area/Volume: $R_a = \langle a_{\omega+1} \rangle / \langle a_\omega \rangle$



# Properties

## 9. network distances:

### (a) shortest path length $d_{ij}$:

- Fewest number of steps between nodes $i$ and $j$.
- (Also called the chemical distance between $i$ and $j$.)

### (b) average path length $\langle d_{ij} \rangle$:

- Average shortest path length in whole network.
- Good algorithms exist for calculation.
- Weighted links can be accommodated.

# Properties

## 9. network distances:

- network diameter $d_{\max}$:
  Maximum shortest path length between any two nodes.
- closeness $d_{\mathrm{cl}} = [\sum_{ij} {d_{ij}}^{-1} / \binom{n}{2}]^{-1}$:
  Average 'distance' between any two nodes.
- Closeness handles disconnected networks ($d_{ij} = \infty$)
- $d_{\mathrm{cl}} = \infty$ only when all nodes are isolated.
- Closeness perhaps compresses too much into one number

# Properties

## 10. centrality:

- Many such measures of a node's 'importance.'
- ex 1: Degree centrality: $k_i$.
- ex 2: Node $i$'s betweenness
  = fraction of shortest paths that pass through $i$.
- ex 3: Recursive centrality: Hubs and Authorities (Kleinberg [8])

# Models

## Some important models:

1. generalized random networks (touched on in 300)
2. scale-free networks (⊞) (covered in 300)
3. small-world networks (⊞) (covered in 300)
4. statistical generative models ($p^*$)
5. generalized affiliation networks (partly covered in 300)

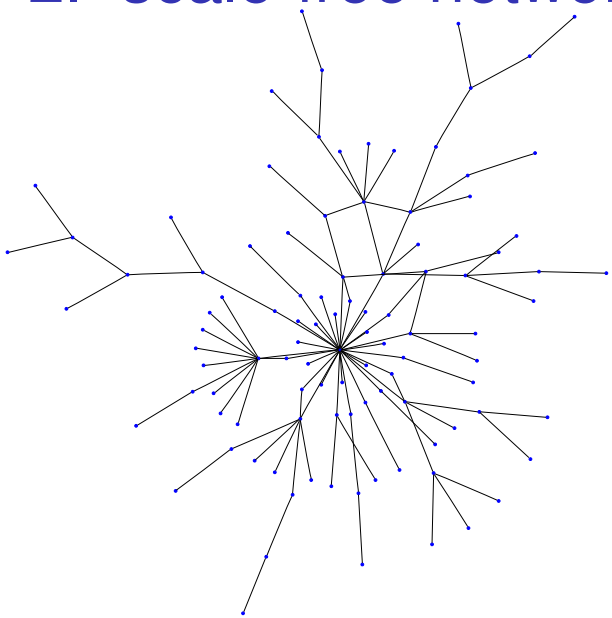
# Models

## 1. generalized random networks:

- Arbitrary degree distribution $P_k$.
- Wire nodes together randomly.
- Create ensemble to test deviations from randomness.
- Interesting, applicable, rich mathematically.
- We will have fun with these guys...

# Models

## 2. ‘scale-free networks’:



$\gamma = 2.5$
$\langle k \rangle = 1.8$
$N = 150$

- Introduced by Barabasi and Albert [1]
- Generative model
- Preferential attachment model with growth:
- $P$[attachment to node i] $\propto \mathrm{k_i^\alpha}$.
- Produces $P_k \sim k^{-\gamma}$ when $\alpha = 1$.
- Trickiness: other models generate skewed degree distributions.

# Models

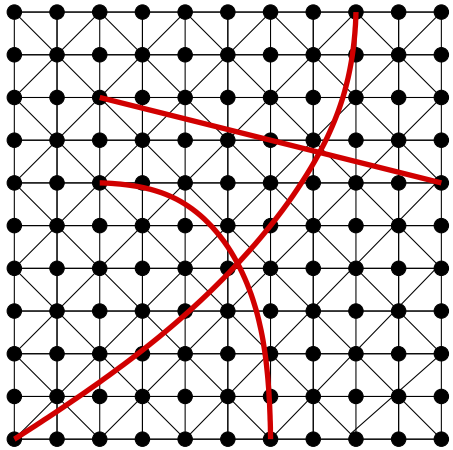
## 3. small-world networks

- Introduced by Watts and Strogatz [19]

Two scales:

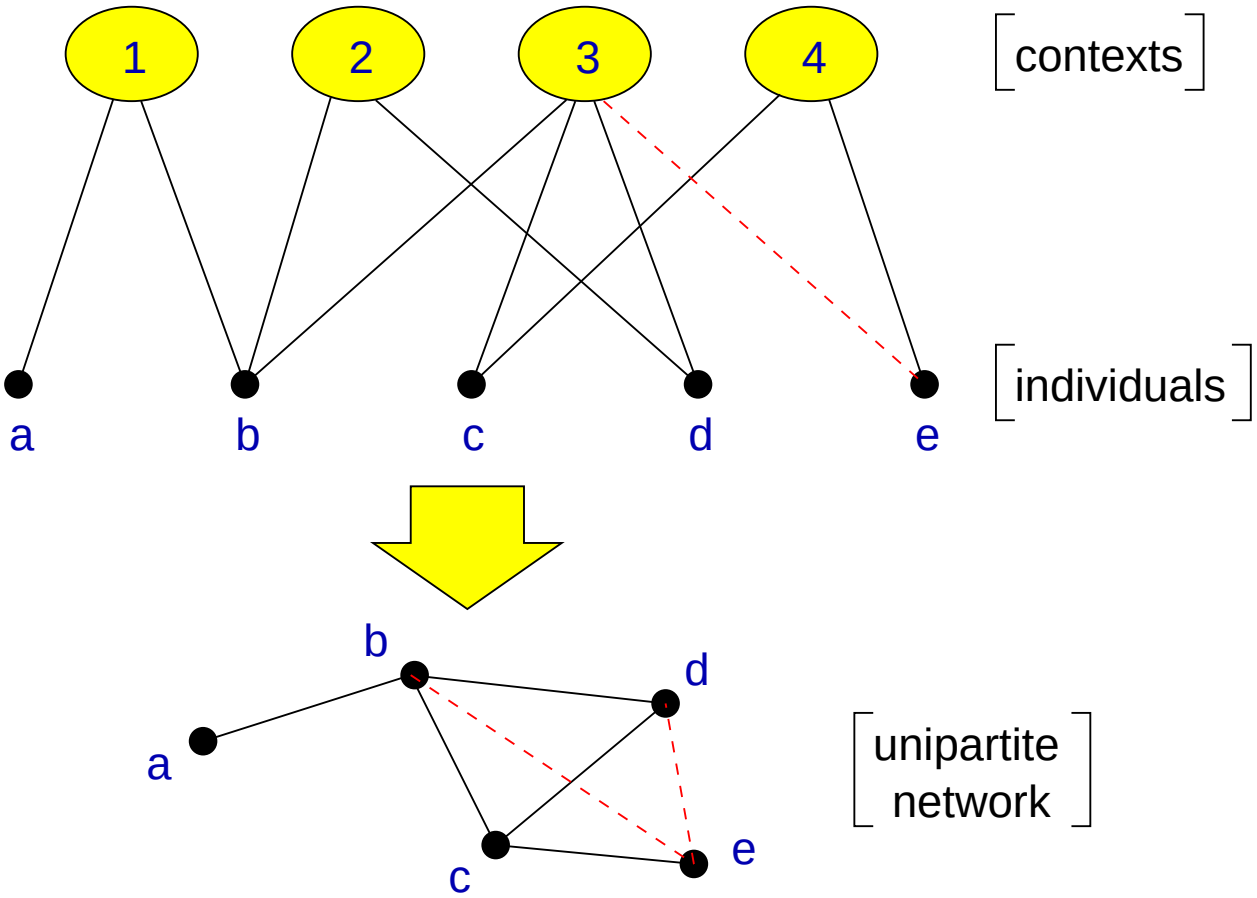
- local regularity (an individual's friends know each other)
- global randomness (shortcuts).

- Shortcuts allow disease to jump
- Number of infectives increases exponentially in time
- Facilitates synchronization



# Models

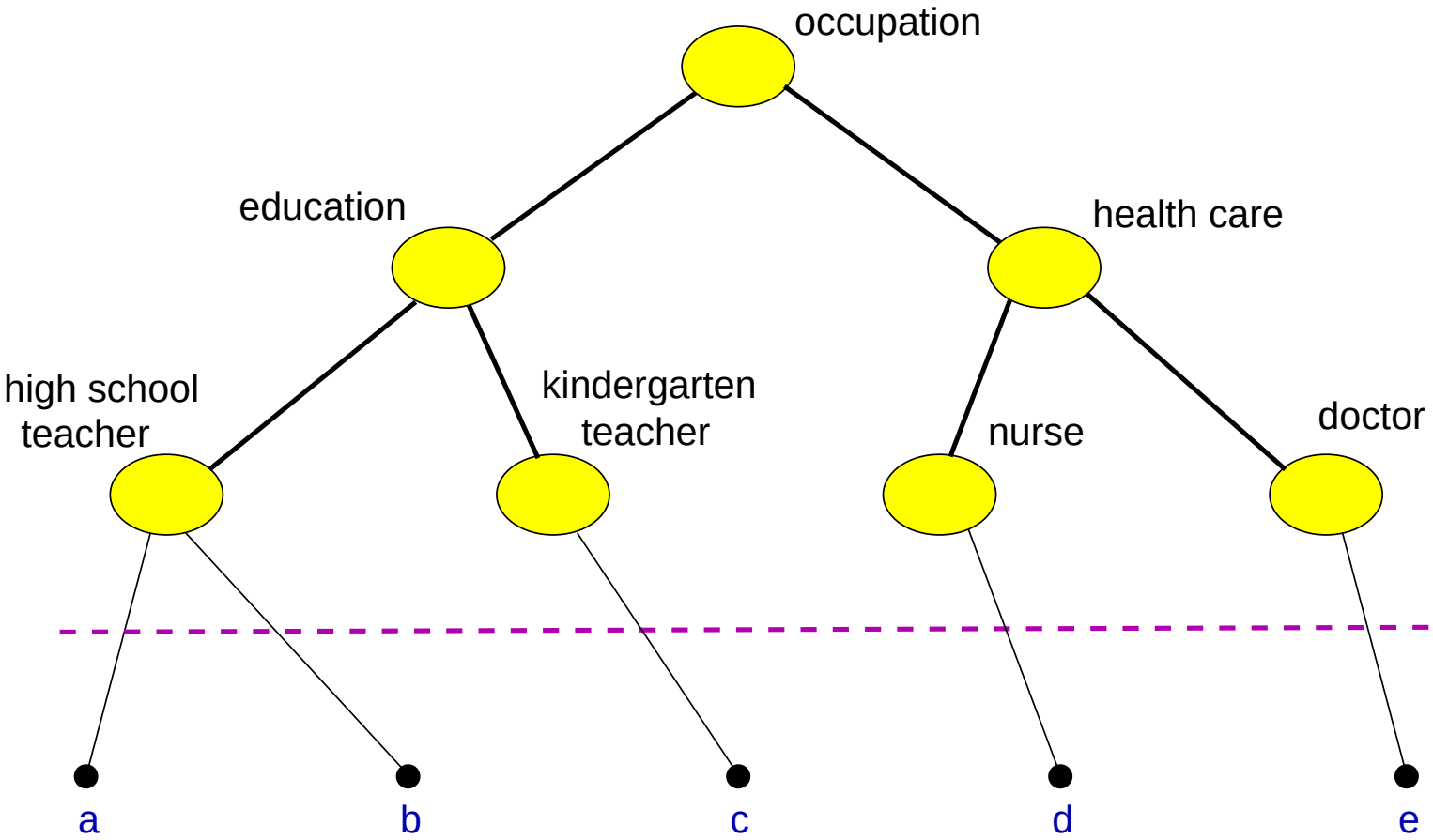
## 5. generalized affiliation networks




Bipartite affiliation networks: boards and directors, movies and actors.

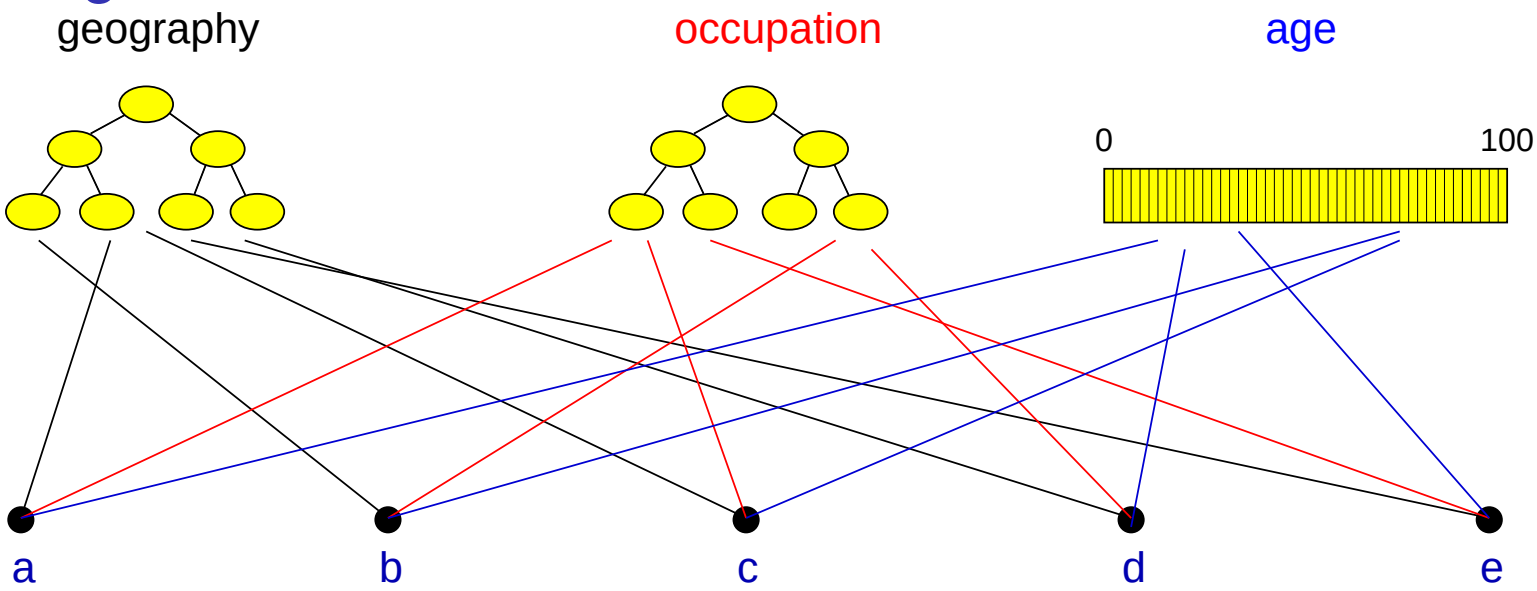
# Models

## 5. generalized affiliation networks



# Models

## 5. generalized affiliation networks




- Blau & Schwartz [2], Simmel [14], Breiger [4], Watts *et al.* [18]

# Popularity (according to ISI)

## "Collective dynamics of 'small-world' networks" [19]

- Watts and Strogatz
  Nature, 1998
- $\approx$ 3500 citations (as of Jan 13, 2009)
- 1100 citations in the last year

## "Emergence of scaling in random networks" [1]

- Barabási and Albert
  Science, 1999
- $\approx$ 3472 citations (as of Jan 13, 2009)
- 1172 citations in the last year

# Nutshell:

## Overview Key Points:

- The field of complex networks came into existence in the late 1990s.
- Explosion of papers and interest since 1998/99.
- Hardened up much thinking about complex systems.
- Specific focus on networks that are large-scale, sparse, natural or man-made, evolving and dynamic, and (crucially) measurable.
- Three main (blurred) categories:
  1. Physical (e.g., river networks),
  2. Interactional (e.g., social networks),
  3. Abstract (e.g., thesauri).

# Nutshell:

## Overview Key Points (cont.):

- Obvious connections with the vast extant field of graph theory.
- But focus on dynamics is more of a physics/stat-mech/comp-sci flavor.
- Two main areas of focus:
  1. Description: Characterizing very large networks
  2. Explanation: Micro story $\Rightarrow$ Macro features
- Some essential structural aspects are understood: degree distribution, clustering, assortativity, group structure, overall structure,...
- Still much work to be done, especially with respect to dynamics... exciting!

# References I


[1] A.-L. Barabási and R. Albert.
Emergence of scaling in random networks.
*Science*, 286:509–511, 1999. pdf (⊞)

[2] P. M. Blau and J. E. Schwartz.
*Crosscutting Social Circles*.
Academic Press, Orlando, FL, 1984.

[3] S. Bornholdt and H. G. Schuster, editors.
*Handbook of Graphs and Networks*.
Wiley-VCH, Berlin, 2003.

[4] R. L. Breiger.
The duality of persons and groups.
*Social Forces*, 53(2):181–190, 1974.

[5] A. Clauset, C. Moore, and M. E. J. Newman.
Structural inference of hierarchies in networks, 2006.
pdf (⊞)

# References II



[6] S. N. Dorogovtsev and J. F. F. Mendes.
*Evolution of Networks*.
Oxford University Press, Oxford, UK, 2003.


[7] M. Gladwell.
*The Tipping Point*.
Little, Brown and Company, New York, 2000.


[8] J. M. Kleinberg.
Authoritative sources in a hyperlinked environment.
*Proc. 9th ACM-SIAM Symposium on Discrete Algorithms*, 1998. pdf (⊞)


[9] M. Kretzschmar and M. Morris.
Measures of concurrency in networks and the spread of infectious disease.
*Math. Biosci.*, 133:165–95, 1996.

# References III



[10] M. Newman.
Assortative mixing in networks.
*Phys. Rev. Lett.*, 89:208701, 2002. pdf (⊞)


[11] M. E. J. Newman.
The structure and function of complex networks.
*SIAM Review*, 45(2):167–256, 2003. pdf (⊞)


[12] I. Rodríguez-Iturbe and A. Rinaldo.
*Fractal River Basins: Chance and Self-Organization*.
Cambridge University Press, Cambrigde, UK, 1997.


[13] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon.
Network motifs in the transcriptional regulation network of *Escherichia coli*.
*Nature Genetics*, pages 64–68, 2002. pdf (⊞)

# References IV



[14] G. Simmel.
The number of members as determining the sociological form of the group. I.
*American Journal of Sociology*, 8:1–46, 1902.


[15] F. Vega-Redondo.
*Complex Social Networks*.
Cambridge University Press, 2007.


[16] S. Wasserman and K. Faust.
*Social Network Analysis: Methods and Applications*.
Cambridge University Press, Cambridge, UK, 1994.


[17] D. J. Watts.
*Six Degrees*.
Norton, New York, 2003.

# References V


[18] D. J. Watts, P. S. Dodds, and M. E. J. Newman.
Identity and search in social networks.
*Science*, 296:1302–1305, 2002. pdf (⊞)

[19] D. J. Watts and S. J. Strogatz.
Collective dynamics of 'small-world' networks.
*Nature*, 393:440–442, 1998. pdf (⊞)