Complex Networks

Principles of Complex Systems
Course CSYS/MATH 300, Fall, 2009

Prof. Peter Dodds

Dept. of Mathematics & Statistics
Center for Complex Systems :: Vermont Advanced Computing Center
University of Vermont

Outline

net•work |ˈnetˌwərk|

noun

1 an arrangement of intersecting horizontal and vertical lines.

- a complex system of roads, railroads, or other transportation routes : *a network of railroads.*

2 a group or system of interconnected people or things : *a trade network.*

- a group of people who exchange information, contacts, and experience for professional or social purposes : *a support network.*
- a group of broadcasting stations that connect for the simultaneous broadcast of a program : *the introduction of a second TV network* | [as adj.] *network television.*
- a number of interconnected computers, machines, or operations : *specialized computers that manage multiple outside connections to a network* | *a local cellular phone network.*
- a system of connected electrical conductors.

verb [trans.]

connect as or operate with a network : *the stock exchanges have proven to be resourceful in networking these deals.*

- link (machines, esp. computers) to operate interactively : [as adj.] (**networked**) *networked workstations.*
- [intrans.] [often as n.] (**networking**) interact with other people to exchange information and develop contacts, esp. to further one's career : *the skills of networking, bargaining, and negotiation.*

Thesaurus deliciousness:

network

noun

1 *a network of arteries* WEB, lattice, net, matrix, mesh, crisscross, grid, reticulum, reticulation; Anatomy plexus.

2 *a network of lanes* MAZE, labyrinth, warren, tangle.

3 *a network of friends* SYSTEM, complex, nexus, web, webwork.

Ancestry:

From Keith Briggs's excellent etymological investigation: (⊞)

- Opus reticulatum:
- A Latin origin?

[http://serialconsign.com/2007/11/we-put-net-network]

Ancestry:

First known use: Geneva Bible, 1560

'And thou shalt make unto it a grate like networke of brass (Exodus xxvii 4).'

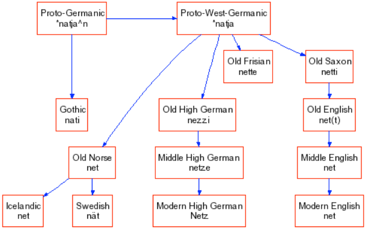
From the OED via Briggs:

- 1658–: reticulate structures in animals
- 1839–: rivers and canals
- 1869–: railways
- 1883–: distribution network of electrical cables
- 1914–: wireless broadcasting networks

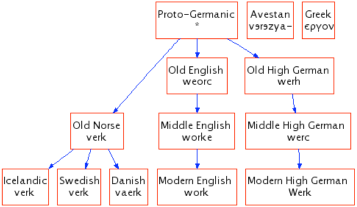
Ancestry:

Net and Work are venerable old words:

- ‘Net’ first used to mean spider web (King Ælfréd, 888).
- ‘Work’ appear to have long meant purposeful action.



The network of Germanic ‘net’ words

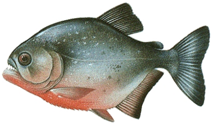


The network of ‘work’ words

- ‘Network’ = something built based on the idea of natural, flexible lattice or web.
- c.f., ironwork, stonework, fretwork.

Key Observation:

- Many complex systems
  can be viewed 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.
- Mindboggling amount of work published on complex networks since 1998...
- ... largely due to your typical theoretical physicist:



- *Piranha physicus*
- Hunt in packs.
- Feast on new and interesting ideas (see chaos, cellular automata, ...)

Popularity (according to ISI)

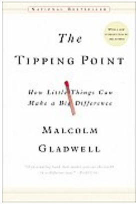
"Collective dynamics of 'small-world' networks" [28]

- Watts and Strogatz
  Nature, 1998
- ≈ 3752 citations (as of June 5, 2009)
- Over 1100 citations in 2008 alone.

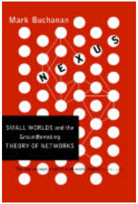
"Emergence of scaling in random networks" [3]

- Barabási and Albert
  Science, 1999
- ≈ 3860 citations (as of June 5, 2009)
- Over 1100 citations in 2008 alone.

Popularity according to books:

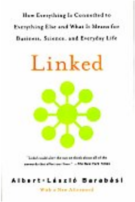


The Tipping Point: How Little Things can make a Big Difference—Malcolm Gladwell [12]

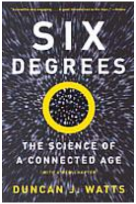


Nexus: Small Worlds and the Groundbreaking Science of Networks—Mark Buchanan

Popularity according to 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 [26]

Numerous others:

- Complex Social Networks—F. Vega-Redondo [25]
- Fractal River Basins: Chance and Self-Organization—I. Rodríguez-Iturbe and A. Rinaldo [20]
- 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
- Social Network Analysis—Stanley Wasserman and Kathleen Faust
- Handbook of Graphs and Networks—Eds: Stefan Bornholdt and H. G. Schuster [6]
- Evolution of Networks—S. N. Dorogovtsev and J. F. F. Mendes [11]

More observations

- But surely networks aren't new...
- Graph theory is well established...
- Study of social networks started in the 1930's...
- So why all this 'new' research on networks?
- Answer: Oodles of Easily Accessible Data.
- We can now inform (alas) our theories with a much more measurable reality.*
- A worthy goal: establish mechanistic explanations.

* *If this is upsetting, maybe string theory is for you...*

More observations

- Web-scale data sets can be overly exciting.

Witness:

- The End of Theory: The Data Deluge Makes the Scientific Theory Obsolete (Anderson, Wired) (⊞)
- "The Unreasonable Effectiveness of Data," Halevy et al. [13].

But:

- For scientists, description is only part of the battle.
- We still need to understand.

Super 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,...

Links = Connections between nodes

- Links may be directed or undirected.
- Links may be binary or weighted.

Other spiffing words: vertices and edges.

Super 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 z)
- Connection between number of edges m and average degree:

$$\langle k \rangle = \frac{2m}{N}.$$

- Defn: $\mathcal{N}_i$ = the set of i's k_i neighbors

Super Basic definitions

Adjacency matrix:

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

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

- (n.b., for numerical work, we always use sparse matrices.)

Examples

So what passes for a complex network?

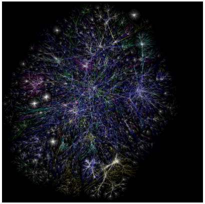
- Complex networks are large (in node number)
- Complex networks are sparse (low edge to node ratio)
- Complex networks are usually dynamic and evolving
- Complex networks can be social, economic, natural, informational, abstract, ...

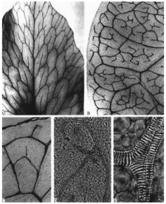
Examples

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

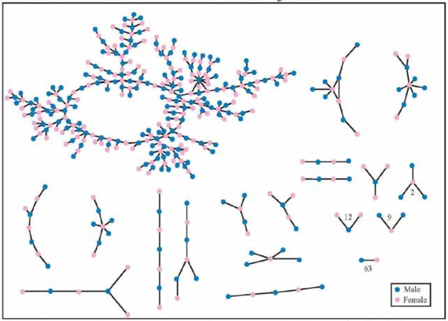
datamining.typepad.com (⊞)

Examples

Interaction networks: social networks

- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- facebook (⊞) twitter (⊞),

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



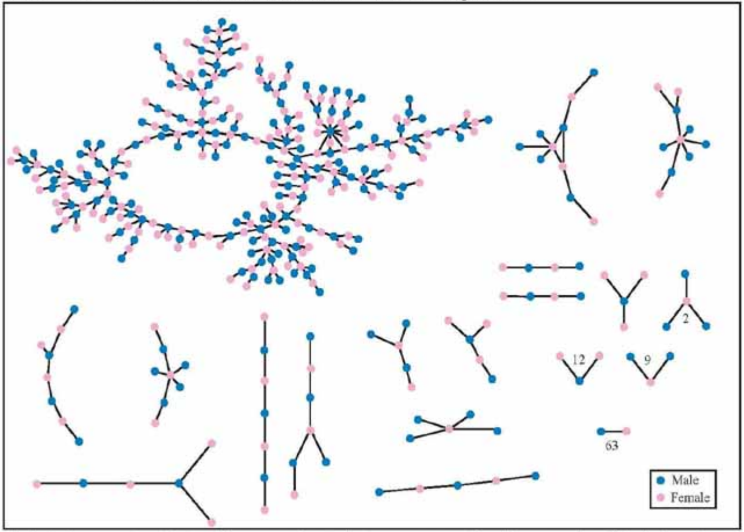
Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

(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

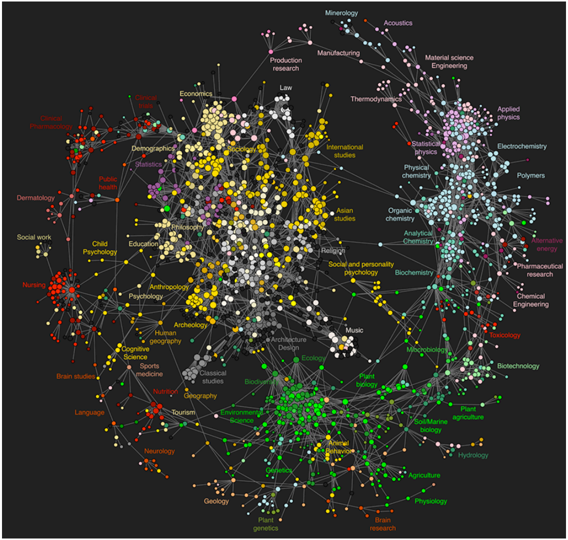
Relational networks

- 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

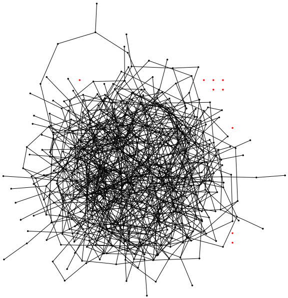
Clickworthy Science:



Bollen et al. [5]

A notable feature of large-scale networks:

- Graphical renderings are often just a big mess.



$\Leftarrow$ Typical hairball

- number of nodes $N = 500$
- number of edges $m = 1000$
- average degree $\langle k \rangle = 4$

- And even when renderings somehow look good:
  "That is a very graphic analogy which aids understanding wonderfully while being, strictly speaking, wrong in every possible way"
  said Ponder [Stibbons] —*Making Money*, T. Pratchett.
- We need to extract digestible, meaningful aspects.

Properties

Some key features 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
- Big deal: Form of P_k key to network's behavior
- ex 1: Erdös-Rényi random networks have a Poisson distribution:

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

- ex 2: "Scale-free" networks: $P_k \propto k^{-\gamma} \Rightarrow$ 'hubs'
- We'll come back to this business soon...

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: [18] 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, protein interactions, neural networks, food webs.*

Properties

4. Clustering:

- Your friends tend to know each other.
- Two measures:

$$C_1 = \left\langle \frac{\sum_{j_1 j_2 \in \mathcal{N}_i} a_{j_1 j_2}}{k_i(k_i - 1)/2} \right\rangle_i \text{ due to Watts \& Strogatz [28]}$$

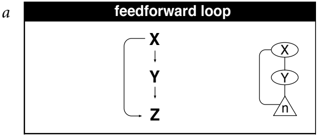
$$C_2 = \frac{3 \times \#\text{triangles}}{\#\text{triples}} \text{ due to Newman [19]}$$

- C_1 is the average fraction of pairs of neighbors who are connected.
- Interpret C_2 as probability two of a node's friends know each other.

Properties

5. Motifs:

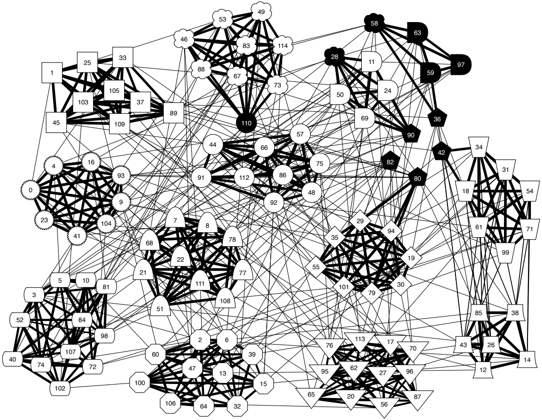
- Small, recurring functional subnetworks
- e.g., Feed Forward Loop:



Shen-Orr, Uri Alon, *et al.* [21]

Properties

6. modularity:



Clauset *et al.*, 2006 [8]: NCAA football

Overview of Complex Networks

Basic definitions

Examples of Complex Networks

Properties of Complex Networks

Nutshell

Basic models of complex networks

Generalized random networks

Scale-free networks

Small-world networks

Generalized affiliation networks

References

Frame 31/122

Properties

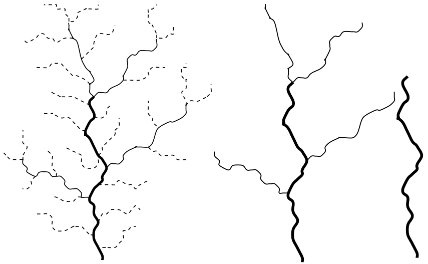
7. Concurrency:

- Transmission of a contagious element only occurs during contact [16]
- 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!

Properties

8. Horton-Strahler stream ordering:

- Metrics for branching networks:
  - Method for ordering streams hierarchically
  - Reveals fractal nature of natural branching networks
  - Hierarchy is not pure but mixed (Tokunaga). [23, 10]
  - Major examples: rivers and blood networks.



(a) (b) (c)

- Beautifully described but poorly explained.

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:

(c) Network diameter $d_{\max}$:

- Maximum shortest path length in network.

(d) 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.

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: Edge ℓ's betweenness
  = fraction of shortest paths that travel along ℓ.
- ex 4: Recursive centrality: Hubs and Authorities (Jon Kleinberg [15])

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...

Models

Some important models:

1. generalized random networks
2. scale-free networks
3. small-world networks
4. statistical generative models (p^*)
5. generalized affiliation networks

Models

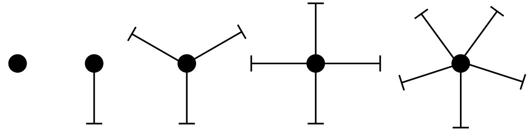
Generalized random networks:

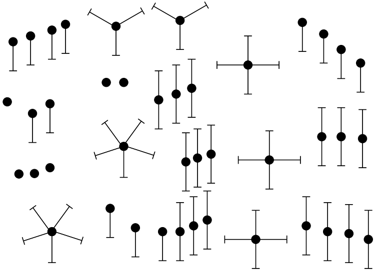
- Arbitrary degree distribution P_k.
- Create (unconnected) nodes with degrees sampled from P_k.
- Wire nodes together randomly.
- Create ensemble to test deviations from randomness.

Building random networks: Stubs

Phase 1:

- Idea: start with a soup of unconnected nodes with stubs (half-edges):





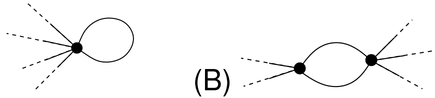
- Randomly select stubs (not nodes!) and connect them.
- Must have an even number of stubs.
- Initially allow self- and repeat connections.

Building random networks: First rewiring

Phase 2:

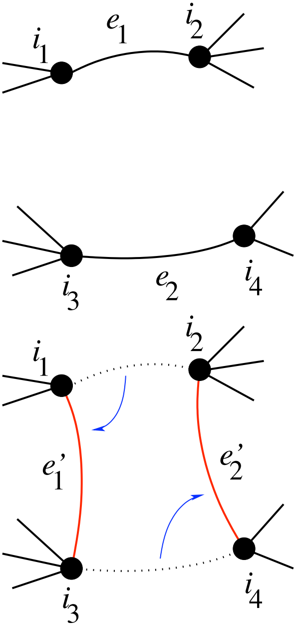
- Now find any (A) self-loops and (B) repeat edges and randomly rewire them.

(A) (B)



- Being careful: we can't change the degree of any node, so we can't simply move links around.
- Simplest solution: randomly rewire two edges at a time.

General random rewiring algorithm



- Randomly choose two edges. (Or choose problem edge and a random edge)
- Check to make sure edges are disjoint.

- Rewire one end of each edge.
- Node degrees do not change.
- Works if e_1 is a self-loop or repeated edge.
- Same as finding on/off/on/off 4-cycles. and rotating them.

Sampling random networks

Phase 2:

- Use rewiring algorithm to remove all self and repeat loops.

Phase 3:

- Randomize network wiring by applying rewiring algorithm liberally.
- Rule of thumb: # Rewirings $\simeq 10 \times$ # edges [17].

Scale-free networks

- Networks with power-law degree distributions have become known as scale-free networks.
- Scale-free refers specifically to the degree distribution having a power-law decay in its tail:

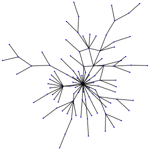
$$P_k \sim k^{-\gamma} \text{ for `large' } k$$

- One of the seminal works in complex networks: Laszlo Barabási and Reka Albert, Science, 1999: "Emergence of scaling in random networks" [3]
- Somewhat misleading nomenclature...

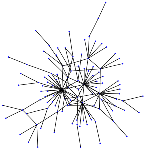
Scale-free networks

- Scale-free networks are not fractal in any sense.
- Usually talking about networks whose links are abstract, relational, informational, . . . (non-physical)
- Primary example: hyperlink network of the Web
- Much arguing about whether or networks are 'scale-free' or not. . .

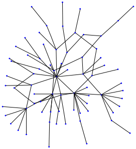
Random networks: largest components



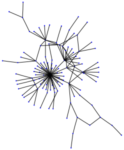
$\gamma = 2.5$
$\langle k \rangle = 1.8$



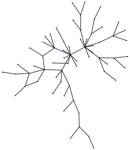
$\gamma = 2.5$
$\langle k \rangle = 2.05333$



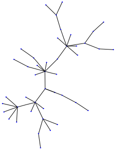
$\gamma = 2.5$
$\langle k \rangle = 1.66667$



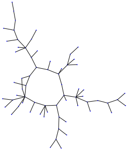
$\gamma = 2.5$
$\langle k \rangle = 1.92$



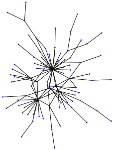
$\gamma = 2.5$
$\langle k \rangle = 1.6$



$\gamma = 2.5$
$\langle k \rangle = 1.50667$



$\gamma = 2.5$
$\langle k \rangle = 1.62667$



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

Scale-free networks

The big deal:

- We move beyond describing networks to finding mechanisms for why certain networks are the way they are.

A big deal for scale-free networks:

- How does the exponent γ depend on the mechanism?
- Do the mechanism details matter?

BA model

- Barabási-Albert model = BA model.
- Key ingredients:
  Growth and Preferential Attachment (PA).
- Step 1: start with m_0 disconnected nodes.
- Step 2:
  1. Growth—a new node appears at each time step $t = 0, 1, 2, \ldots$.
  2. Each new node makes m links to nodes already present.
  3. Preferential attachment—Probability of connecting to ith node is $\propto k_i$.
- In essence, we have a rich-gets-richer scheme.

BA model

- Definition: A_k is the attachment kernel for a node with degree k.
- For the original model:

$$A_k = k$$

- Definition: $P_{\text{attach}}(k, t)$ is the attachment probability.
- For the original model:

$$P_{\text{attach}}(\text{node } i, t) = \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = \frac{k_i(t)}{\sum_{k=0}^{k_{\max}(t)} k N_k(t)}$$

where $N(t) = m_0 + t$ is # nodes at time t
and $N_k(t)$ is # degree k nodes at time t.

Approximate analysis

- When $(N+1)$th node is added, the expected increase in the degree of node i is

$$E(k_{i,N+1} - k_{i,N}) \simeq m \frac{k_{i,N}}{\sum_{j=1}^{N(t)} k_j(t)}.$$

- Assumes probability of being connected to is small.
- Dispense with Expectation by assuming (hoping) that over longer time frames, degree growth will be smooth and stable.
- Approximate $k_{i,N+1} - k_{i,N}$ with $\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t}$:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)}$$

where $t = N(t) - m_0$.

Approximate analysis

- Deal with denominator: each added node brings m new edges.

$$\therefore \sum_{j=1}^{N(t)} k_j(t) = 2tm$$

- The node degree equation now simplifies:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = m \frac{k_i(t)}{2mt} = \frac{1}{2t} k_i(t)$$

- Rearrange and solve:

$$\frac{\mathrm{d}k_i(t)}{k_i(t)} = \frac{\mathrm{d}t}{2t} \Rightarrow \boxed{k_i(t) = c_i\, t^{1/2}.}$$

- Next find c_i ...

Approximate analysis

- Know ith node appears at time

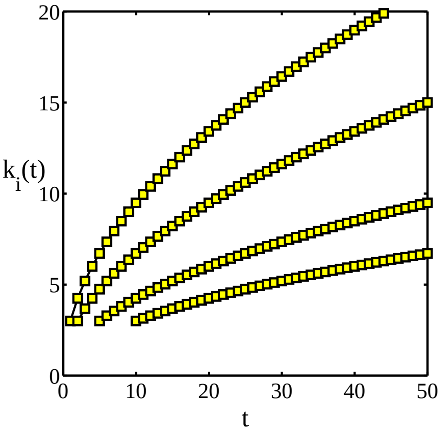
$$t_{i,\text{start}} = \begin{cases} i - m_0 & \text{for } i > m_0 \\ 0 & \text{for } i \le m_0 \end{cases}$$

- So for $i > m_0$ (exclude initial nodes), we must have

$$k_i(t) = m \left(\frac{t}{t_{i,\text{start}}} \right)^{1/2} \text{ for } t \ge t_{i,\text{start}}.$$

- All node degrees grow as $t^{1/2}$ but later nodes have larger $t_{i,\text{start}}$ which flattens out growth curve.
- Early nodes do best (First-mover advantage).

Approximate analysis



- $m = 3$
- $t_{i,\text{start}} = 1, 2, 5,$ and 10.

Degree distribution

- So what's the degree distribution at time t?
- Use fact that birth time for added nodes is distributed uniformly:

$$\mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}t_{i,\mathrm{start}} \simeq \frac{\mathrm{d}t_{i,\mathrm{start}}}{t}$$

- Also use

$$k_i(t) = m\left(\frac{t}{t_{i,\mathrm{start}}}\right)^{1/2} \Rightarrow t_{i,\mathrm{start}} = \frac{m^2 t}{k_i(t)^2}.$$

Transform variables—Jacobian:

$$\frac{\mathrm{d}t_{i,\mathrm{start}}}{\mathrm{d}k_i} = -2\frac{m^2 t}{k_i(t)^3}.$$

Degree distribution

- $$\mathbf{Pr}(k_i)\mathrm{d}k_i = \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}t_{i,\mathrm{start}}$$
- $$= \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}k_i \left|\frac{\mathrm{d}t_{i,\mathrm{start}}}{\mathrm{d}k_i}\right|$$
- $$= \frac{1}{t}\mathrm{d}k_i\, 2\frac{m^2 t}{k_i(t)^3}$$
- $$= 2\frac{m^2}{k_i(t)^3}\mathrm{d}k_i$$
- $$\propto k_i^{-3}\mathrm{d}k_i\,.$$

Degree distribution

- We thus have a very specific prediction of $\mathbf{Pr}(k) \sim k^{-\gamma}$ with $\gamma = 3$.
- Typical for real networks: $2 < \gamma < 3$.
- Range true more generally for events with size distributions that have power-law tails.
- $2 < \gamma < 3$: finite mean and 'infinite' variance (wild)
- In practice, $\gamma < 3$ means variance is governed by upper cutoff.
- $\gamma > 3$: finite mean and variance (mild)

Examples

WWW	$\gamma \simeq 2.1$ for in-degree
WWW	$\gamma \simeq 2.45$ for out-degree
Movie actors	$\gamma \simeq 2.3$
Words (synonyms)	$\gamma \simeq 2.8$

The Internets is a different business...

Real data

From Barabási and Albert's original paper [3]:

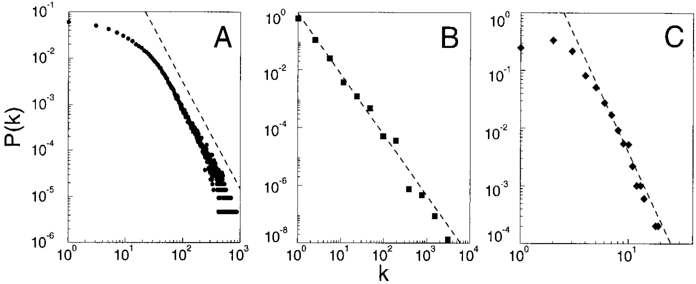


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

Things to do and questions

- Vary attachment kernel.
- Vary mechanisms:
  1. Add edge deletion
  2. Add node deletion
  3. Add edge rewiring
- Deal with directed versus undirected networks.
- Important Q.: Are there distinct universality classes for these networks?
- Q.: How does changing the model affect γ?
- Q.: Do we need preferential attachment and growth?
- Q.: Do model details matter?
- The answer is (surprisingly) yes. More later re Zipf.

Preferential attachment

- Let's look at preferential attachment (PA) a little more closely.
- PA implies arriving nodes have complete knowledge of the existing network's degree distribution.
- For example: If $P_{\mathrm{attach}}(k) \propto k$, we need to determine the constant of proportionality.
- We need to know what everyone's degree is...
- PA is $\therefore$ an outrageous assumption of node capability.
- But a very simple mechanism saves the day...

Preferential attachment through randomness

- Instead of attaching preferentially, allow new nodes to attach randomly.
- Now add an extra step: new nodes then connect to some of their friends' friends.
- Can also do this at random.
- Assuming the existing network is random, we know probability of a random friend having degree k is

$$Q_k \propto kP_k$$

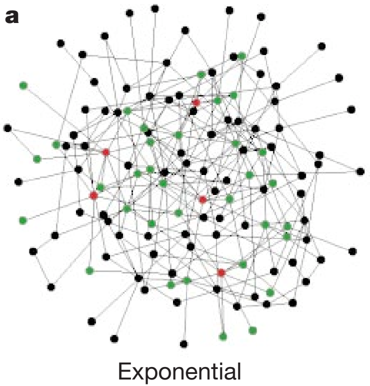
- So rich-gets-richer scheme can now be seen to work in a natural way.

Robustness

- System robustness and system robustness.
- Albert et al., Nature, 2000:
  "Error and attack tolerance of complex networks" [2]

Robustness

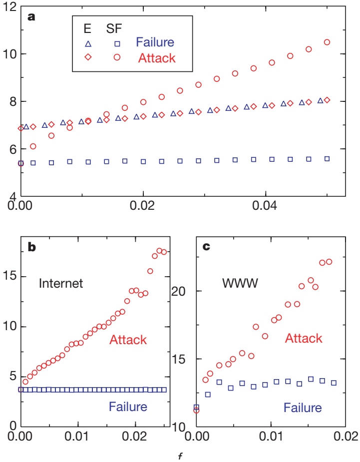
- Standard random networks (Erdös-Rényi)
  versus
  Scale-free networks



from

Albert et al., 2000

Robustness



from

Albert et al., 2000

- Plots of network diameter as a function of fraction of nodes removed
- Erdös-Rényi versus scale-free networks
- blue symbols = random removal
- red symbols = targeted removal (most connected first)

Robustness

- Scale-free networks are thus robust to random failures yet fragile to targeted ones.
- All very reasonable: Hubs are a big deal.
- But: next issue is whether hubs are vulnerable or not.
- Representing all webpages as the same size node is obviously a stretch (e.g., google vs. a random person's webpage)
- Most connected nodes are either:
  1. Physically larger nodes that may be harder to 'target'
  2. or subnetworks of smaller, normal-sized nodes.
- Need to explore cost of various targeting schemes.

Some problems for people thinking about people?:

How are social networks structured?

- How do we define connections?
- How do we measure connections?
- (remote sensing, self-reporting)

What about the dynamics of social networks?

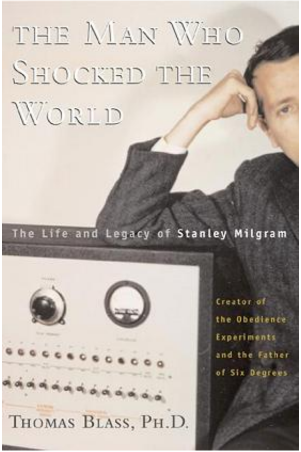
- How do social networks evolve?
- How do social movements begin?
- How does collective problem solving work?
- How is information transmitted through social networks?

Social Search

A small slice of the pie:

- Q. Can people pass messages between distant individuals using only their existing social connections?
- A. Apparently yes...

Milgram's social search experiment (1960s)



http://www.stanleymilgram.com

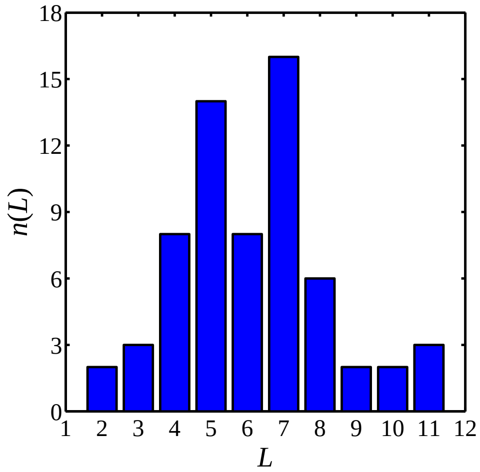
- Target person = Boston stockbroker.
- 296 senders from Boston and Omaha.
- 20% of senders reached target.
- chain length $\simeq$ 6.5.

Popular terms:

- The Small World Phenomenon;
- "Six Degrees of Separation."

The problem

Lengths of successful chains:



From Travers and Milgram (1969) in Sociometry: [24] "An Experimental Study of the Small World Problem."

The problem

Two features characterize a social 'Small World':

1. Short paths exist

   and

2. People are good at finding them.

Social Search

Milgram's small world experiment with e-mail [9]

Social search—the Columbia experiment

- 60,000+ participants in 166 countries
- 18 targets in 13 countries including
  - a professor at an Ivy League university,
  - an archival inspector in Estonia,
  - a technology consultant in India,
  - a policeman in Australia,
    and
  - a veterinarian in the Norwegian army.
- 24,000+ chains

Social search—the Columbia experiment

- Milgram's participation rate was roughly 75%
- Email version: Approximately 37% participation rate.
- Probability of a chain of length 10 getting through:

$$.37^{10} \simeq 5 \times 10^{-5}$$

- $\Rightarrow$ 384 completed chains (1.6% of all chains).

Social search—the Columbia experiment

- Motivation/Incentives/Perception matter.
- If target *seems* reachable
  $\Rightarrow$ participation more likely.
- Small changes in attrition rates
  $\Rightarrow$ large changes in completion rates
- e.g., $\searrow$ 15% in attrition rate
  $\Rightarrow$ $\nearrow$ 800% in completion rate

Social search—the Columbia experiment

Successful chains disproportionately used

- weak ties (Granovetter)
- professional ties (34% vs. 13%)
- ties originating at work/college
- target's work (65% vs. 40%)

. . . and disproportionately avoided

- hubs (8% vs. 1%) (+ no evidence of funnels)
- family/friendship ties (60% vs. 83%)

Geography → Work

Social search—the Columbia experiment

Senders of successful messages showed little absolute dependency on

- age, gender
- country of residence
- income
- religion
- relationship to recipient

Range of completion rates for subpopulations:

30% to 40%

Social search—the Columbia experiment

Nevertheless, some weak discrepencies do exist...

An above average connector:

Norwegian, secular male, aged 30-39, earning over $100K, with graduate level education working in mass media or science, who uses relatively weak ties to people they met in college or at work.

A below average connector:

Italian, Islamic or Christian female earning less than $2K, with elementary school education and retired, who uses strong ties to family members.

Social search—the Columbia experiment

Mildly bad for continuing chain:

choosing recipients because “they have lots of friends” or because they will “likely continue the chain.”

Why:

- Specificity important
- Successful links used relevant information. (e.g. connecting to someone who shares same profession as target.)

Social search—the Columbia experiment

Basic results:

- $\langle L \rangle = 4.05$ for all completed chains
- L_* = Estimated 'true' median chain length (zero attrition)
- Intra-country chains: $L_* = 5$
- Inter-country chains: $L_* = 7$
- All chains: $L_* = 7$
- Milgram: $L_* \simeq 9$

The social world appears to be small...

- Connected random networks have short average path lengths:

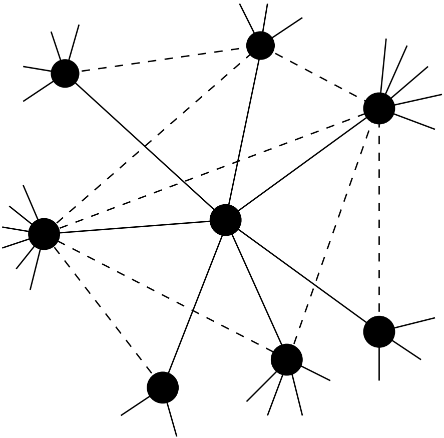
$$\langle d_{AB} \rangle \sim \log(N)$$

N = population size,

d_{AB} = distance between nodes A and B.

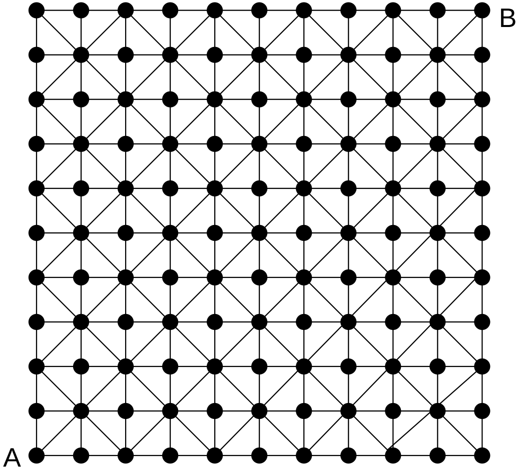
- But: social networks aren't random...

Simple socialness in a network:



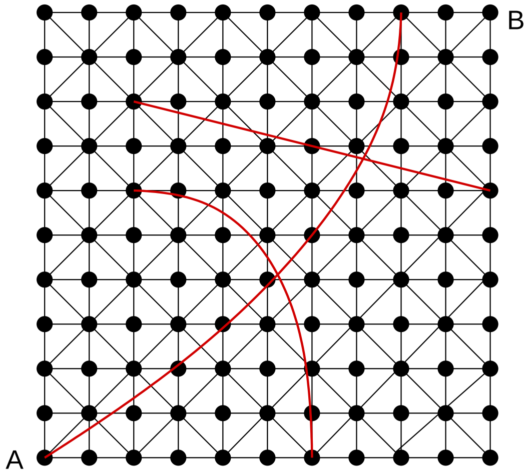
Need “clustering” (your friends are likely to know each other):

Non-randomness gives clustering:



$d_{AB} = 10 \rightarrow$ too many long paths.

Randomness + regularity



Now have $d_{AB} = 3$

$\langle d \rangle$ decreases overall

Small-world networks

Introduced by Watts and Strogatz (Nature, 1998) [28]
"Collective dynamics of 'small-world' networks."

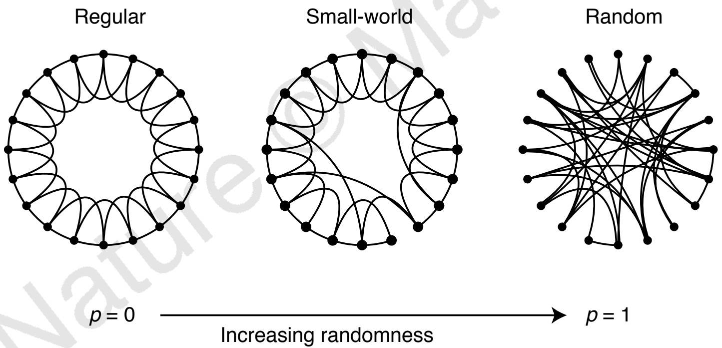
Small-world networks were found everywhere:

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph,
- food webs,
- social networks of comic book characters,...

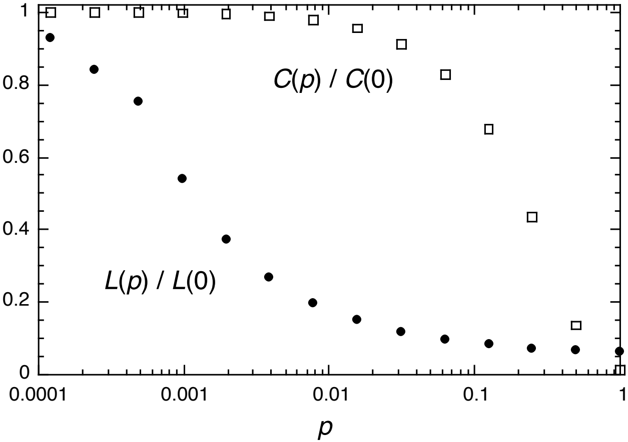
Very weak requirements:

- local regularity + random short cuts

Toy model:



The structural small-world property:



- $L(p)$ = average shortest path length as a function of p
- $C(p)$ = average clustring as a function of p

Previous work—finding short paths

But are these short cuts findable?

Nope.

Nodes cannot find each other quickly with any local search method.

Need a more sophisticated model...

Previous work—finding short paths

- What can a local search method reasonably use?
- How to find things without a map?
- Need some measure of distance between friends and the target.

Some possible knowledge:

- Target's identity
- Friends' popularity
- Friends' identities
- Where message has been

Previous work—finding short paths

Jon Kleinberg (Nature, 2000) [14]
"Navigation in a small world."

Allowed to vary:

1. local search algorithm

   and

2. network structure.

Previous work—finding short paths

Kleinberg's Network:

1. Start with regular d-dimensional cubic lattice.
2. Add local links so nodes know all nodes within a distance q.
3. Add m short cuts per node.
4. Connect i to j with probability

$$p_{ij} \propto {x_{ij}}^{-\alpha}.$$

- $\alpha = 0$: random connections.
- α large: reinforce local connections.
- $\alpha = d$: same number of connections at all scales.

Previous work—finding short paths

Theoretical optimal search:

- "Greedy" algorithm.
- Same number of connections at all scales: $\alpha = d$.

Search time grows slowly with system size (like $\log^2 N$).

But: social networks aren't lattices plus links.

Previous work—finding short paths

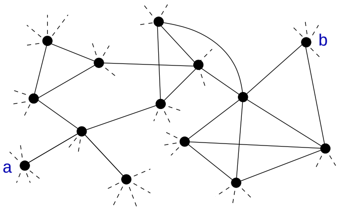
- If networks have hubs can also search well: Adamic et al. (2001) [1]

$$P(k_i) \propto k_i^{-\gamma}$$

  where k = degree of node i (number of friends).
- Basic idea: get to hubs first (airline networks).
- But: hubs in social networks are limited.

The problem

If there are no hubs and no underlying lattice, how can search be efficient?



Which friend of a is closest to the target b?

What does ‘closest’ mean?

What is ‘social distance’?

Models

One approach: incorporate identity.

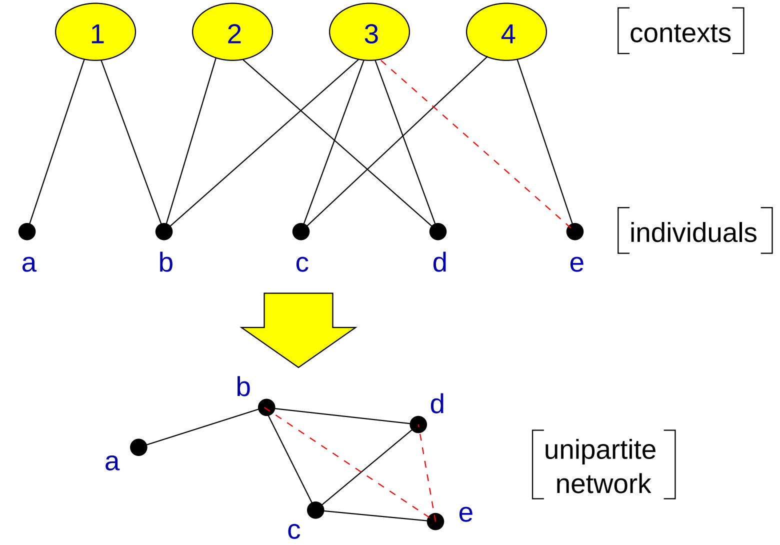
Identity is formed from attributes such as:

- Geographic location
- Type of employment
- Religious beliefs
- Recreational activities.

Groups are formed by people with at least one similar attribute.

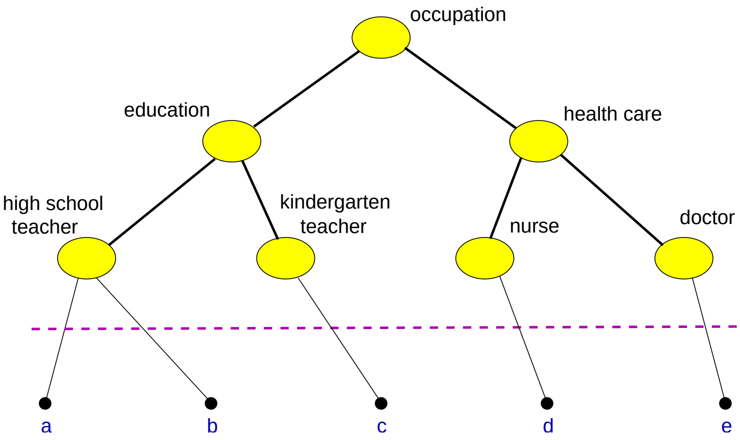
Attributes $\Leftrightarrow$ Contexts $\Leftrightarrow$ Interactions $\Leftrightarrow$ Networks.

Social distance—Bipartite affiliation networks



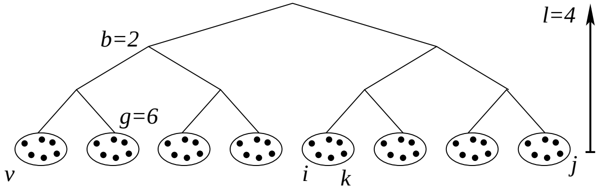
Bipartite affiliation networks: boards and directors, movies and actors.

Social distance—Context distance



Models

Distance between two individuals x_{ij} is the height of lowest common ancestor.



$x_{ij} = 3$, $x_{ik} = 1$, $x_{iv} = 4$.

Models

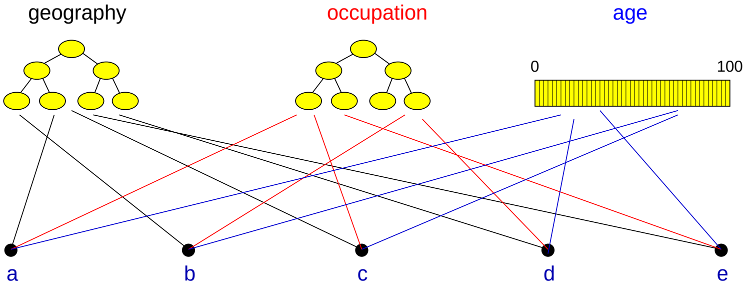
- Individuals are more likely to know each other the closer they are within a hierarchy.
- Construct z connections for each node using

$$p_{ij} = c\,\exp\{-\alpha x_{ij}\}.$$

- $\alpha = 0$: random connections.
- α large: local connections.

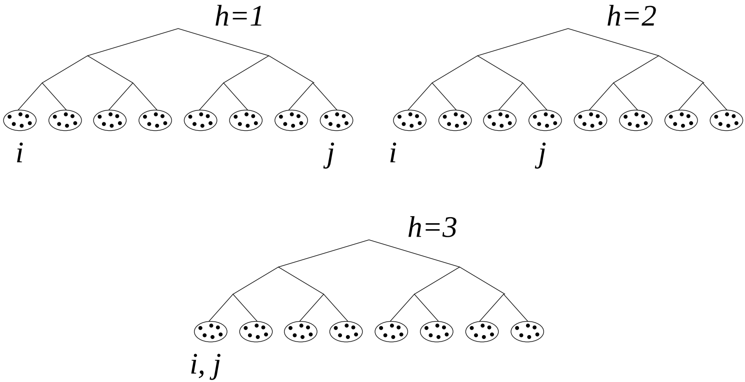
Models

Generalized affiliation networks



- Blau & Schwartz [4], Simmel [22], Breiger [7], Watts *et al.* [27]

The model



$\vec{v}_i = [1\ 1\ 1]^T$, $\vec{v}_j = [8\ 4\ 1]^T$

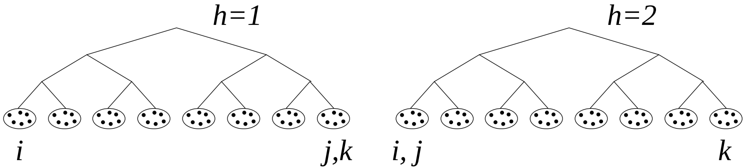
$x_{ij}^1 = 4,\ \ x_{ij}^2 = 3,\ \ x_{ij}^3 = 1.$

Social distance:

$$y_{ij} = \min_h x_{ij}^h .$$

The model

Triangle inequality doesn't hold:



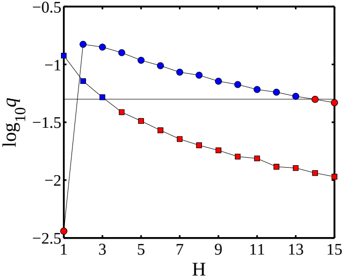
$$y_{ik} = 4 > y_{ij} + y_{jk} = 1 + 1 = 2.$$

The model

- Individuals know the identity vectors of
  1. themselves,
  2. their friends, and
  3. the target.
- Individuals can estimate the social distance between their friends and the target.
- Use a greedy algorithm + allow searches to fail randomly.

The model-results—searchable networks

$\alpha = 0$ versus $\alpha = 2$ for $N \simeq 10^5$:



$q \geq r$

$q < r$

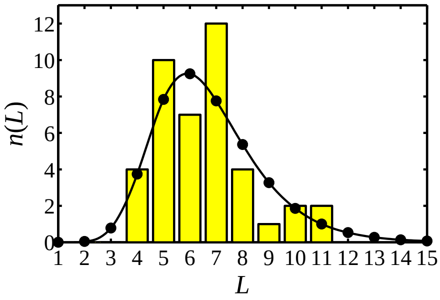
$r = 0.05$

q = probability an arbitrary message chain reaches a target.

- A few dimensions help.
- Searchability decreases as population increases.
- Precise form of hierarchy largely doesn't matter.

The model-results

Milgram's Nebraska-Boston data:



Model parameters:

- $N = 10^8$,
- $z = 300$, $g = 100$,
- $b = 10$,
- $\alpha = 1$, $H = 2$;

- $\langle L_{\mathrm{model}} \rangle \simeq 6.7$
- $L_{\mathrm{data}} \simeq 6.5$

Social search—Data

Adamic and Adar (2003)

- For HP Labs, found probability of connection as function of organization distance well fit by exponential distribution.
- Probability of connection as function of real distance $\propto 1/r$.

Social Search—Real world uses

- Tags create identities for objects
- Website tagging: `http://www.del.icio.us`
- (e.g., Wikipedia)
- Photo tagging: `http://www.flickr.com`
- Dynamic creation of metadata plus links between information objects.
- Folksonomy: collaborative creation of metadata

Social Search—Real world uses

Recommender systems:

- Amazon uses people's actions to build effective connections between books.
- Conflict between 'expert judgments' and tagging of the hoi polloi.

Nutshell

- Bare networks are typically unsearchable.
- Paths are findable if nodes understand how network is formed.
- Importance of identity (interaction contexts).
- Improved social network models.
- Construction of peer-to-peer networks.
- Construction of searchable information databases.

References I

L. Adamic, R. Lukose, A. Puniyani, and B. Huberman.
Search in power-law networks.
Phys. Rev. E, 64:046135, 2001. pdf (⊞)

R. Albert, H. Jeong, and A.-L. Barabási.
Error and attack tolerance of complex networks.
Nature, 406:378–382, July 2000. pdf (⊞)

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

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

References II

J. Bollen, H. Van de Sompel, A. Hagberg, L. Bettencourt, R. Chute, M. A. Rodriguez, and B. Lyudmila.
Clickstream data yields high-resolution maps of science.
PLoS ONE, 4:e4803, 2009. pdf (⊞)

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

R. L. Breiger.
The duality of persons and groups.
Social Forces, 53(2):181–190, 1974. pdf (⊞)

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

References III

P. S. Dodds, R. Muhamad, and D. J. Watts.
An experimental study of search in global social networks.
Science, 301:827–829, 2003. pdf (⊞)

P. S. Dodds and D. H. Rothman.
Unified view of scaling laws for river networks.
Physical Review E, 59(5):4865–4877, 1999. pdf (⊞)

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

M. Gladwell.
The Tipping Point.
Little, Brown and Company, New York, 2000.

References IV

A. Halevy, P. Norvig, and F. Pereira.
The unreasonable effectiveness of data.
IEEE Intelligent Systems, 24:8–12, 2009. pdf (⊞)

J. Kleinberg.
Navigation in a small world.
Nature, 406:845, 2000. pdf (⊞)

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

M. Kretzschmar and M. Morris.
Measures of concurrency in networks and the spread of infectious disease.
Math. Biosci., 133:165–95, 1996. pdf (⊞)

References V

R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon.
On the uniform generation of random graphs with prescribed degree sequences, 2003. pdf (⊞)

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

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

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

References VI

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 (⊞)

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

E. Tokunaga.
The composition of drainage network in Toyohira River Basin and the valuation of Horton's first law.
Geophysical Bulletin of Hokkaido University, 15:1–19, 1966.

References VII

J. Travers and S. Milgram.
An experimental study of the small world problem.
Sociometry, 32:425–443, 1969. pdf (⊞)

F. Vega-Redondo.
Complex Social Networks.
Cambridge University Press, 2007.

D. J. Watts.
Six Degrees.
Norton, New York, 2003.

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

References VIII

D. J. Watts and S. J. Strogatz.
Collective dynamics of ‘small-world’ networks.
Nature, 393:440–442, 1998. pdf (⊞)