Chaos, Complexity, and Inference (36-462)

Lecture 20: Network Basics

Cosma Shalizi

26 March 2009

Networks: The Basics

Basics and Examples

Some Examples

Bipartite Networks

Network Properties

Further reading: Newman (2003) (assigned); Watts (2004) (assigned); Scott (2000) (old-school social network theory); Wasserman and Faust (1994) (the Bible, but, like the Bible, can be very detailed and very dull. . .)

Basic Defintions

Network/graph consists of **nodes** and **edges**

Nodes/vertices things of some sort; say n of them

Edges/links/ties binary relationship between nodes; **directed** or **undirected**

In-degree/Out-degree number of links to/from a node

Adjacency matrix $n \times n$ binary matrix, $A_{ij} = 1$ if there is an edge from i to j, $= 0$ otherwise

Sub-graph subset of nodes, plus all the edges between them

Path contiguous series of edges (respecting direction)

The adjacency matrix A says which nodes are directly linked

The powers of A are linked by paths: $A^k_{ij} = 0$ iff there is no path of length k from i to j; otherwise A^k_{ij} counts the number of paths

Nodes are **connected** when there's a path linking them

Networks break up into **connected components** (possibly just one), which are sub-graphs

(**geodesic**) **Distance** between nodes = number of edges in shortest path; ∞ if no such path

Betweenness of a node/edge: how many shortest paths between pairs of (other) nodes go through this?

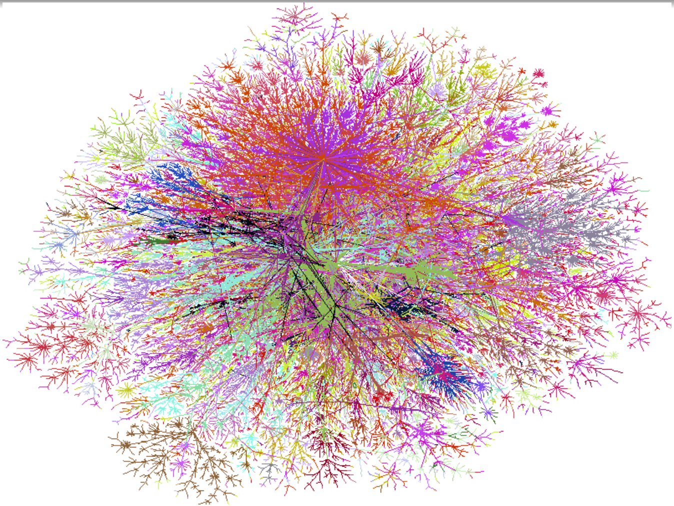
Discounted for number of shortest paths between a given pair; formula is messy

Some Examples

Unless otherwise noted, pictures snarfed from `http://www-personal.umich.edu/~mejn/networks/`, see there for full credits

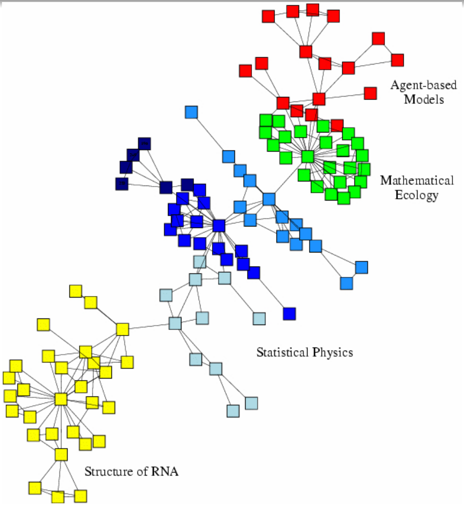
Use the GraphViz programs to draw your own `graphviz.org`

Several `R` packages for networks, mostly called "social networks"; `igraph` and `statnet` (on CRAN) may be best



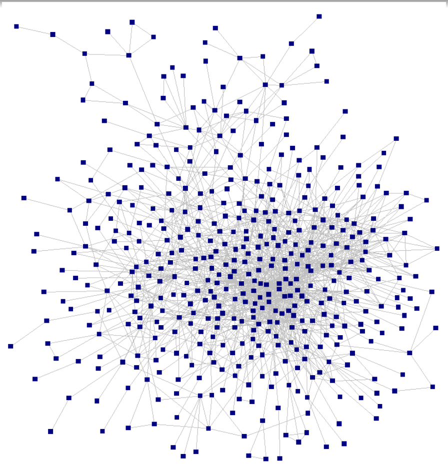
Nodes: autonomous systems on the Internet

Edge relationship: "passes packets to"



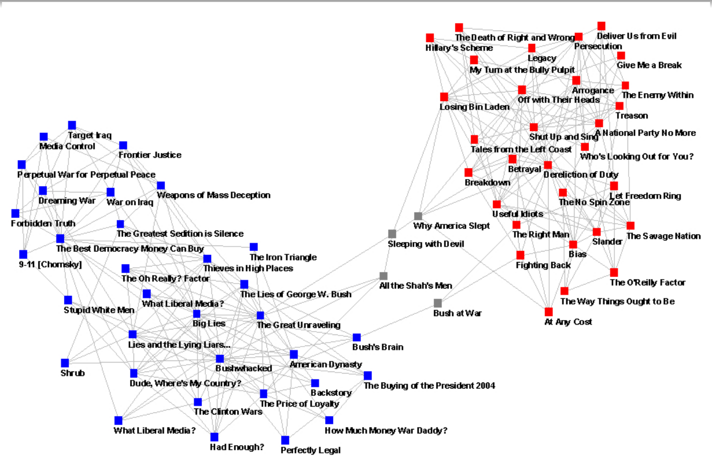
Nodes: scientists at Santa Fe Institute, late 1990s-early 2000s

Edge relationship: “wrote paper with”



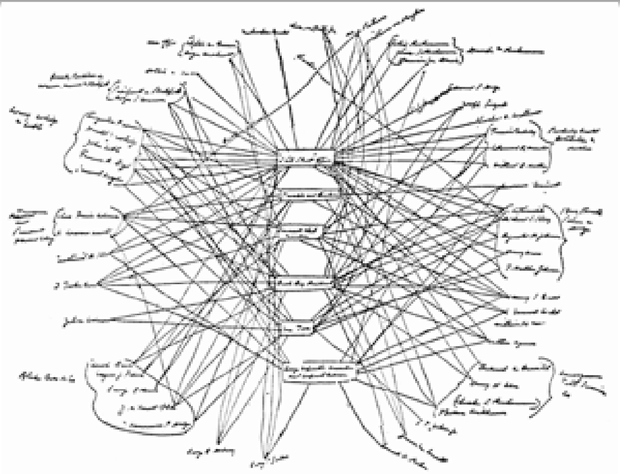
Nodes: Paul Erdős and co-authors
Edge relationship: “wrote paper with”

From Vladis Krebs, `http://www.orgnet.com/Erdos.html`



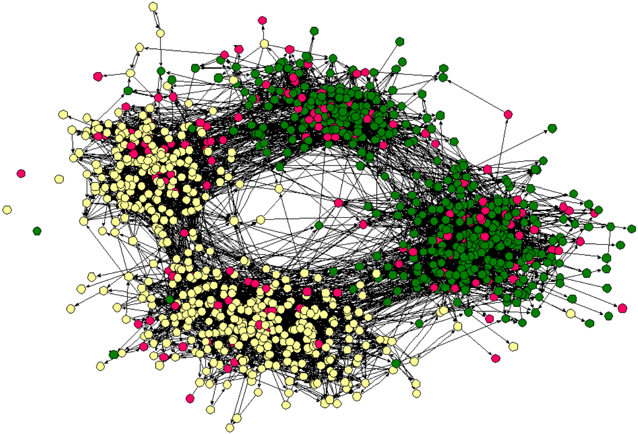
Nodes: Top-selling political books on Amazon, 2004
Edge relationship: "customers also bought ..."

Also by Krebs



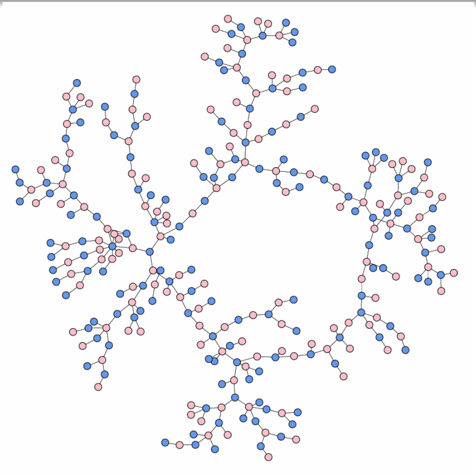
Opponents of the nomination of Louis Brandeis to the Supreme Court, 1916; diagram by James Butler Studley; via Eric Rauchway's blog

Apparently oldest known social network diagram



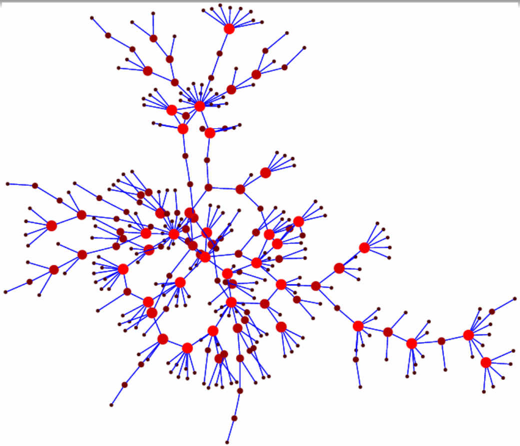
Nodes: high school students (colored by race)

Edge relationship: “claims to be friends with”



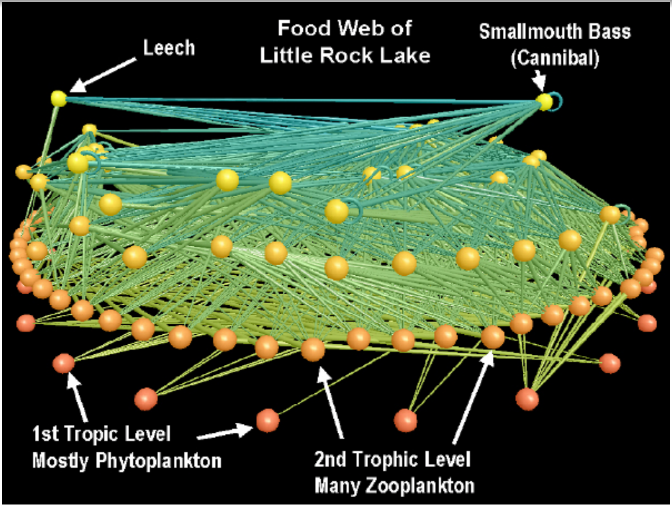
Nodes: high school students
Edge relationship: “dates”

Limited to largest connected component



Nodes: people in Colorado Springs, early 1980s (color = HIV status)
Edge relationship: “bonks and/or shares needle”
Limited to largest connected component

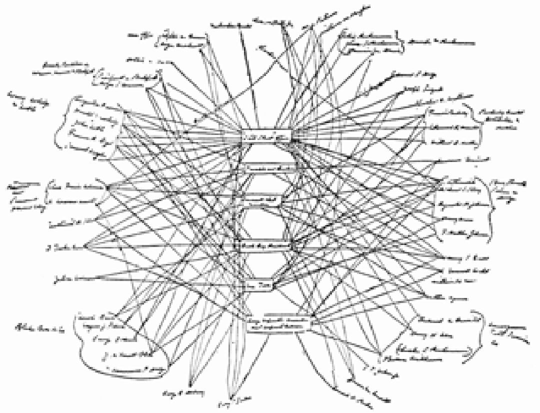
Re-drawn by Newman from Potterat *et al.* (2002)



Nodes: plant and animal species in lake

Edge relationship: “eats”

Back to the anti-Brandeis network



Two kinds of nodes: people and institutions
Multi-component network: here two components, so also called **bipartite**

Bipartite Graphs: Collaboration networks

- women in Natchez, MS. in 1930s/social events ("Southern Women" data, Davis *et al.* (1941) as cited by Freeman (2003))
- actors/movies ("Kevin Bacon game")
- scientists/papers (many papers by Newman et al.)
- musicians/albums (several papers on jazz)
- superheroes/comic books (Alberich *et al.*, 2002)
- company directors/corporate boards (a.k.a. "the power elite")
- campaign donors/politicians
- words/documents

Analyzing Bipartite Networks

1. "project down" to one component, nodes linked if they have a common partner in other component

as in SFI and Erdős collaboration graphs

2. special techniques for bipartite networks, based on **Galois lattices**:

- smaller and smaller groups of people who have more and more in common
- smaller and smaller sets of projects common to more and more people
- hierarchies coincide

Good at describing community structure, may revisit in later lecture

Freeman and White (1993); White and Duquenne (1996); Roth and Bourgine (2003, 2005)

Small World Property

Diameter: maximum distance between two nodes
Six degrees of separation: The diameter of the social network is no more than 6.

What exactly would that mean?

Small world property: diameter is $O(\log n)$, $n =$ number of nodes

Made famous by Milgram, apparently on rather dubious evidence (Kleinfeld, 2002)

The small world property is *mathematically* easy:

- Assume each node has about k neighbors
- Assume those neighbors have few neighbors in common (≈ 1)
- Pick an arbitrary node; how many nodes can be reached in t steps?
- Clearly $\approx (k-1)^t$
- To find diameter set $n \approx (k-1)^d$
- $d \approx \log n / \log k - 1$

Argument runs in to trouble when paths from the starting node begin to cross each other
We'll revisit this later when talking about contagion

Random Walks and Centrality

Random walk on a network:

1. Start at an arbitrary node
2. Pick a neighbor, uniformly at random, and go there
3. Go to step 2

This is a Markov chain…

EXERCISE: Explain how to get its transition matrix from the adjacency matrix

…on a finite, connected state space…

at least on each connected component of the graph …so it goes to a unique invariant distribution (ergodic theorem)

What is this invariant distribution like?

$$p_i = \sum_{j:A_{ji}=1} p_j \frac{1}{\sum_{k=1}^{n} A_{jk}}$$

$p_i \uparrow$ in-degree of i (many places to reach it)
$\Pr(j \rightarrow i) \downarrow$ out-degree of j (many places it could go)
$\Pr(j \rightarrow i) \uparrow$ probability of j

Centrality

Important nodes are ones which are major neighbors of other important nodes

Sounds like: “Celebrities are people who are famous for being well-known”

but not *viciously* circular

This probability is (**Bonacich**) **centrality** (Scott, 2000, pp. 87–88, 97–99)

There are other centrality measures, see Scott

In essence, this is page-rank

See also: `eigenfactor.org` for ranking scientific journals

Simple versus Complex Networks

Rough notion of “complex”: many *strongly interdependent* parts

Networks clearly have many parts...

Simple networks by way of contrast to complex ones

1. Completely regular, deterministic lattices (grids, etc.)
2. Completely random graphs (Erdős-Rényi model)

Erdős-Rényi Model

Erdős: "A mathematician is a machine for turning amphetamines into proofs" often bowlderized into "coffee"

Actually also done by Solomonoff/Rapoport, possibly others. . .

Not realistic but (1) cute math and (2) gives a kind of baseline

Model specification:

- n nodes (fixed)
- Each possible edge exists with probability p, independent of all other edges

Degree of node i = K_i

$$K_i \sim \mathrm{Binom}(n-1, p)$$

Why $n-1$?

Take limit $N \to \infty$, $p \to 0$, $np = \lambda = \mathit{constant}$

$$K_i \rightsquigarrow \mathrm{Pois}(\lambda)$$

If $\lambda > \lambda_c$, one connected component has size $\propto n$ (“giant component”), small world property in giant component

THOUGHT EXERCISE: Try to guess λ_c

Limitations

Degree distribution: Rarely binomial/Poisson; often highly skewed; sometimes, arguably, power-law tailed

Reciprocity: In directed networks, $A_{ij} = A_{ji}$ more often than you'd expect from chance

Transitivity: If $A_{ij} = 1$ and $A_{jk} = 1$, higher odds that $A_{ik} = 1$
clustering coefficients measure this transitivity (counting triangles)

Homophily/Assortativeness: $A_{ij} = 1$ is more likely if i and j are similar — or, in some networks, dis-similar
Social networks tend to be assortative by degree, technological networks tend to be *dis*-assortative (Newman and Park, 2003)

Can make some of these limitations go away in **inhomogeneous** Erdős-Rényi models, with different p between different *types* of nodes (Clauset *et al.*, 2007)

Will see other models of networks, with more complexity, next time

Alberich, R., J. Miro-Julia and F. Rossello (2002). "Marvel Universe looks almost like a real social network." E-print, arxiv.org, cond-mat/0202174. URL http://arxiv.org/abs/cond-mat/0202174.

Clauset, Aaron, Cristopher Moore and Mark E. J. Newman (2007). "Structural Inference of Hierarchies in Networks." In *Statistical Network Analysis: Models, Issues, and New Directions* (Edo Airoldi and David M. Blei and Stephen E. Fienberg and Anna Goldenberg and Eric P. Xing and Alice X. Zheng, eds.), vol. 4503 of *Lecture Notes in Computer Science*, pp. 1–13. New York: Springer-Verlag. URL http://arxiv.org/abs/physics/0610051.

Davis, Allison, Burleigh B. Gardner and Mary R. Gardner (1941). *Deep South*. Chicago: University of Chicago Press.

Freeman, Linton C. (2003). "Finding Social Groups: A Meta-Analysis of the Southern Women Data." In *Dynamic*

Social Network Modeling and Analysis (Ronald Breiger and Kathleen Carley and Philippa Pattison, eds.), pp. 39–77. Washington, D.C.: National Academies Press. URL http://www.nap.edu/openbook.php?record_id=10735&page=39.

Freeman, Linton C. and Douglas R. White (1993). "Using Galois Lattices to Represent Network Data." *Sociological Methodology*, **23**: 127–146. URL http://eclectic.ss.uci.edu/~drwhite/pw/Galois.pdf.

Kleinfeld, Judith (2002). "Could It Be a Big World After All? What the Milgram Papers in the Yale Archive Reveal About the Original Small World Study." *Society*, **39**: 61–66. URL http://www.uaf.edu/northern/big_world.html.

Newman, M. E. J. and Juyong Park (2003). "Why social networks are different from other types of networks." *Physical*

Review E, **68**. URL
http://arxiv.org/abs/cond-mat/0305612/.

Newman, Mark E. J. (2003). "The Structure and Function of Complex Networks." *SIAM Review*, **45**: 167–256. URL
http://arxiv.org/abs/cond-mat/0303516.

Potterat, J. J., L. Phillips-Plummer, S. Q. Muth, R. B. Rothenberg, D. E. Woodhouse, T. S. Maldonado-Long, H. P. Zimmerman and J. B. Muth (2002). "Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs." *Sexually Transmitted Infections*, **78**: 159–163. URL
http://sti.bmj.com/cgi/content/abstract/78/suppl_1/i159.

Roth, Camille and Paul Bourgine (2003). "Binding Social and Cultural Networks: A Modelization." Electronic pre-print. URL
http://arxiv.org/abs/nlin.AO/0309035.

— (2005). "Epistemic communities: description and hierarchic categorization." *Mathematical Population Studies*, **12**: 107–130. URL http://arxiv.org/abs/nlin.AO/0409013.

Scott, John (2000). *Social Network Analysis: A Handbook*. Thousand Oaks, California: Sage Publications, 2nd edn.

Wasserman, Stanley and Katherine Faust (1994). *Social Network Analysis: Methods and Applications*. Cambridge, England: Cambridge University Press.

Watts, Duncan J. (2004). "The "New" Science of Networks." *Annual Review of Sociology*, **30**: 243–270. doi:10.1146/annurev.soc.30.020404.104342.

White, Douglas R. and Vincent Duquenne (1996). "Special Issue on "Social Network and Discrete Structure Analysis"." *Social Networks*, **18**: 169–318.