707.000
Web Science and Web Technology
„Network Theory and Terminology“

Fundamentals!

Markus Strohmaier

Univ. Ass. / Assistant Professor
Knowledge Management Institute
Graz University of Technology, Austria

e-mail: markus.strohmaier@tugraz.at
web: http://www.kmi.tugraz.at/staff/markus

Network Theory and Terminology

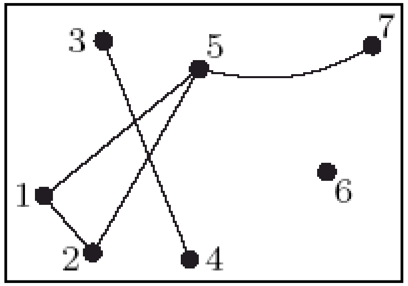
Terminology

http://www.cis.upenn.edu/~Emkearns/teaching/NetworkedLife/
[Diestel 2005]

Network

- A collection of individual or atomic entities
- Referred to as nodes or vertices (the “dots” or “points”)
- Collection of links or edges between vertices (the “lines”)
- Links can represent any pairwise relationship
- Links can be directed or undirected
- Network: entire collection of nodes and links
- For us, a network is an abstract object (list of pairs) and is separate from its visual layout
- that is, we will be interested in properties that are invariant
  - structural properties
  - statistical properties of families of networks

What different kinds of networks exist in the real world?



Social Networks

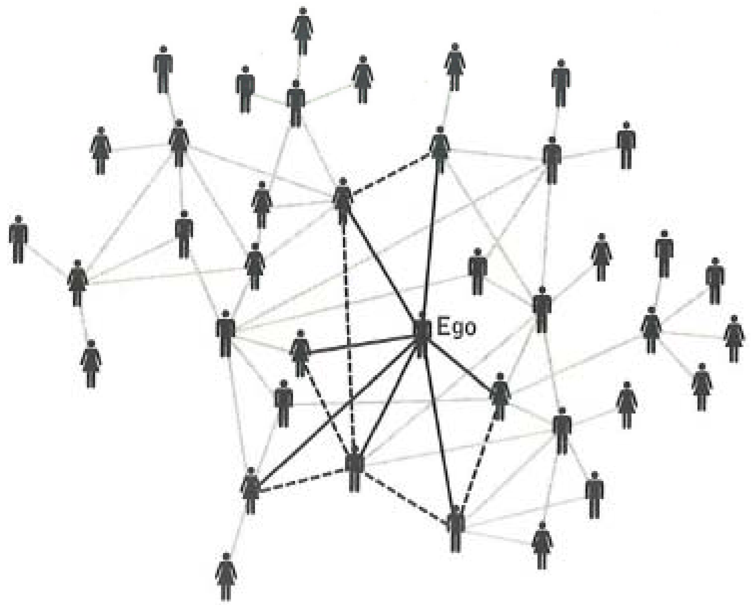
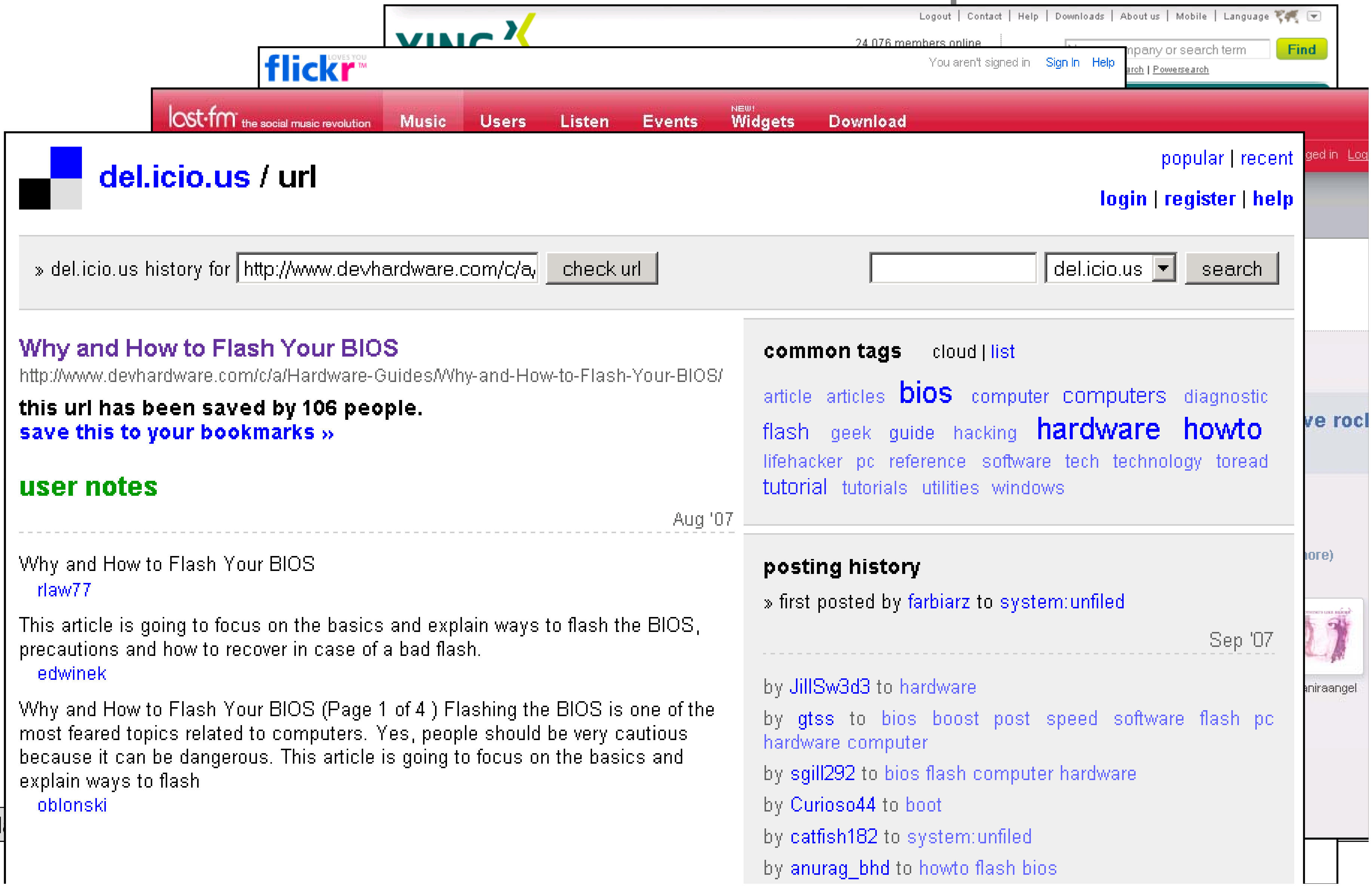


Figure 1.3. Real social networks exhibit clustering, the tendency of two individuals who share a mutual friend to be friends themselves. Here, Ego has six friends, each of whom is friends with at least one other.

Social Networks Examples



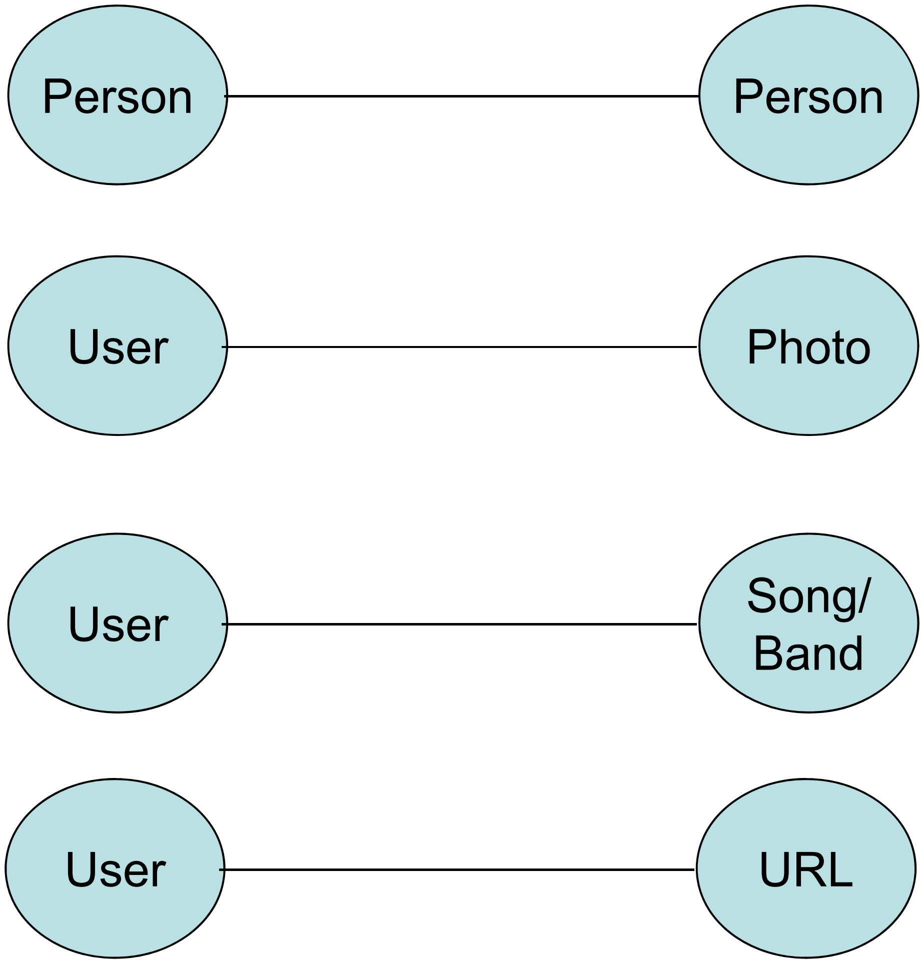
Social Networks Entities Simplified

Xing:

Flickr:

Last.fm:

Del.icio.us



Object-Centred Sociality [Knorr Cetina 1997]

- Suggests to extend the concept of sociality, which is primarily understood to exist between individuals, to objects
- Claims that in a knowledge society, object relations substitute for and become constitutive of social relations
- Promotes an „expanded conception of sociality" that includes (but is not limited to) material objects
- Objects of sociality are close to our interests
- From a more applied perspective, Zengestrom[1] argues that successful social software focuses on similiar objects of sociality (although the term is used slightly differently).
- These objects mediate social ties between people.

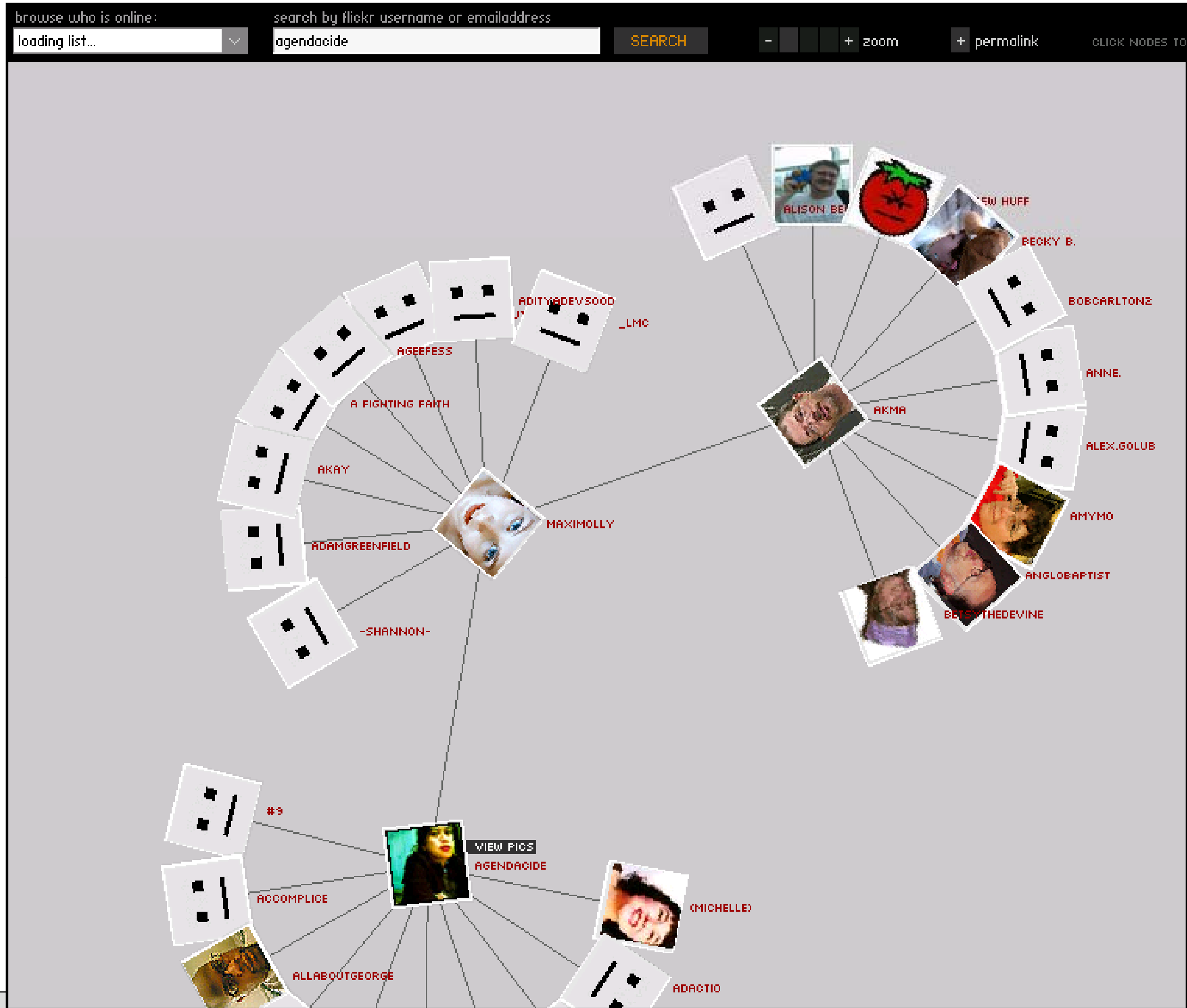
Can you name objects of sociality in existing social software?

Whats the object of sociality in, e.g. XING?

By altering the object of sociality, can you come up with new ideas for social software applications?

1 http://www.zengestrom.com/blog/2005/04/why_some_social.html

Flickr Graph



Network Examples [Newman 2003]

II Networks in the real world

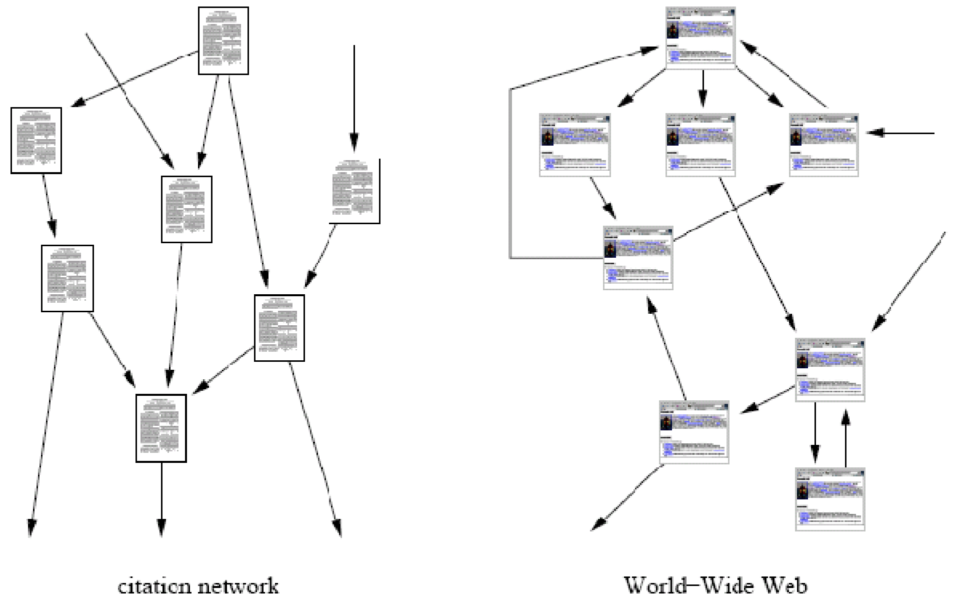


FIG. 4 The two best studied information networks. Left: the citation network of academic papers in which the vertices are papers and the directed edges are citations of one paper by another. Since papers can only cite those that came before them (lower down in the figure) the graph is acyclic—it has no closed loops. Right: the World Wide Web, a network of text pages accessible over the Internet, in which the vertices are pages and the directed edges are hyperlinks. There are no constraints on the Web that forbid cycles and hence it is in general cyclic.

Overview

Agenda

Technical preliminaries for your first course work:

- Network Preliminaries
  - One Mode and Two Mode Networks
  - Network Representation
  - Network Metrics
- Software Architecture Preliminaries
  - REST
  - JSON
- Release of Home Assignment 1.1

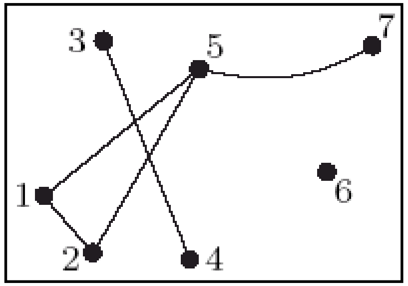
Terminology I

http://www.cis.upenn.edu/~Emkearns/teaching/NetworkedLife/
[Diestel 2005]

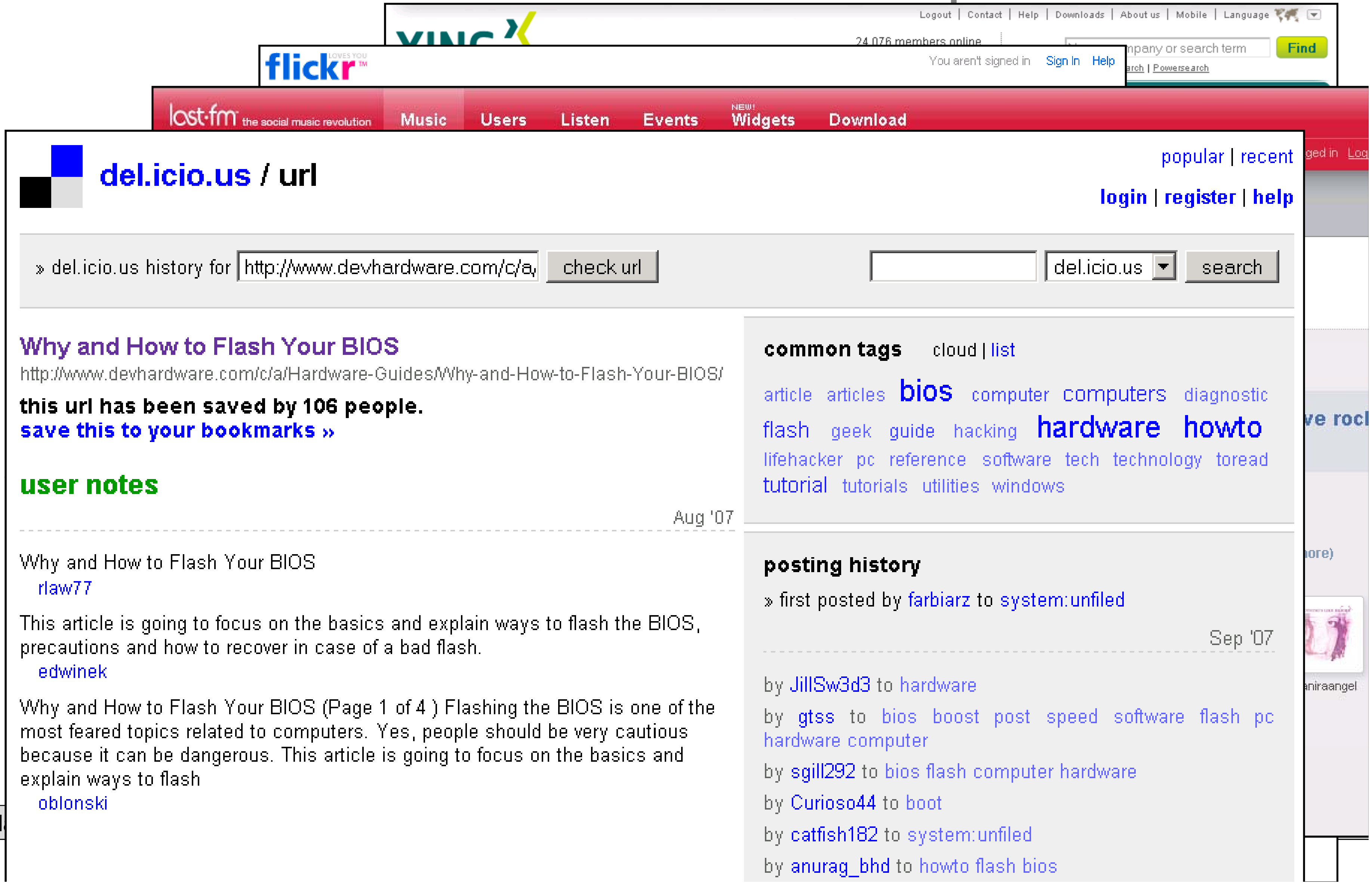
Network

- A collection of individual or atomic entities
- Referred to as nodes or vertices (the “dots” or “points”)
- Collection of links or edges between vertices (the “lines”)
- Links can represent any pairwise relationship
- Links can be directed or undirected
- Network: entire collection of nodes and links
- For us, a network is an abstract object (list of pairs) and is separate from its visual layout
- that is, we will be interested in properties that are invariant
  - structural properties
  - statistical properties of families of networks

What different kinds of networks exist in the real world?



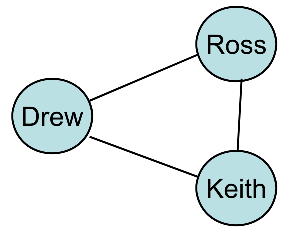
Social Networks Examples



One mode / two mode networks
(uni/bipartite graphs)

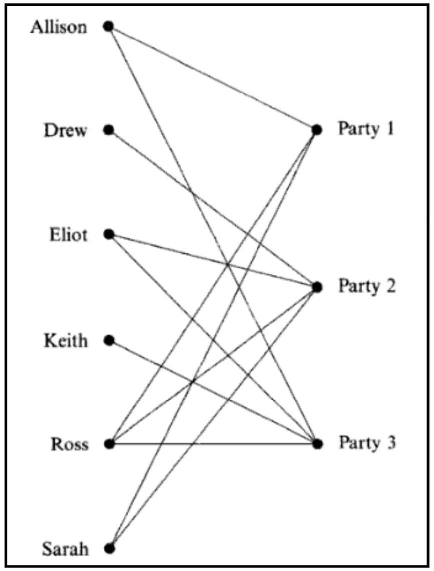
One mode network:

- A single type of nodes



Two mode network:

- Two types of nodes
- Edges are only possible between different types of nodes



How can we represent (social) networks?

We will discuss three basic forms:

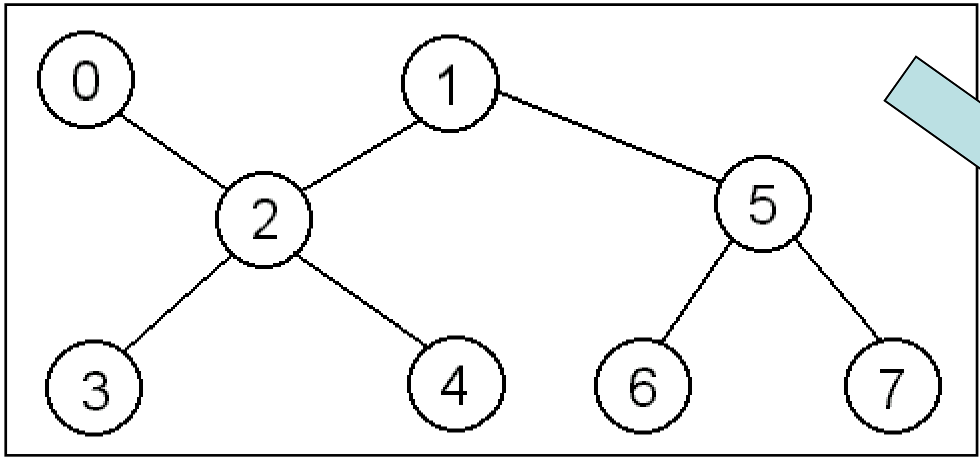
- Adjacency lists
- Adjacency matrices
- Incident matrices

Adjacency Matrix for one mode networks

- Complete description of a graph
- The matrix is symmetric for nondirectional graphs
- A row and a column for each node
- Of size g x g (g rows and g colums)

Adjacency matrices for One-Mode Networks

taken from http://courseweb.sp.cs.cmu.edu/~cs111/applications/ln/lecture18.html

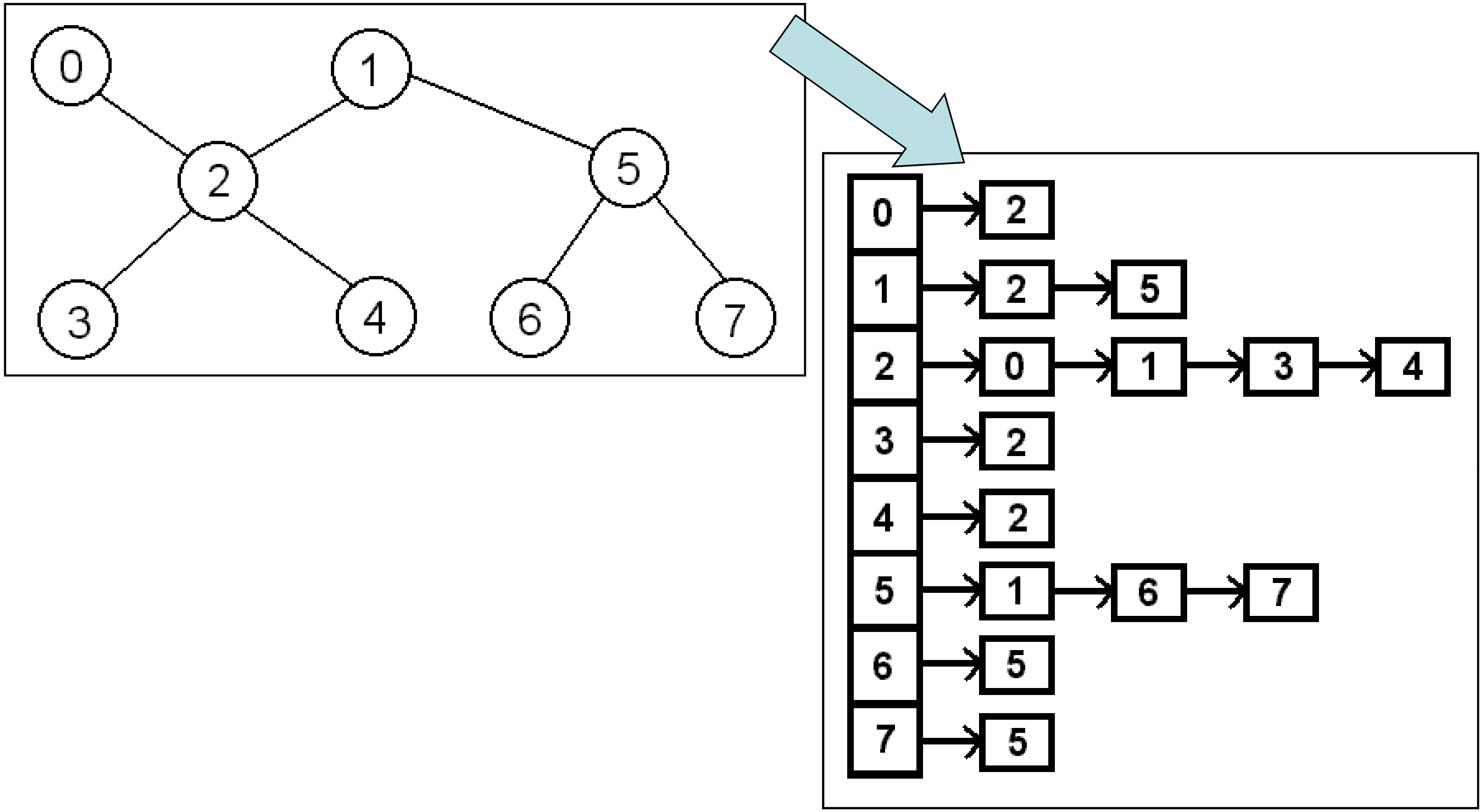


Adjacency matrix or *sociomatrix*

	0	1	2	3	4	5	6	7
0	0	-	0	-	-	-	-	-
1	-	0	0	-	-	0	-	-
2	0	0	0	0	0	-	-	-
3	-	-	0	0	-	-	-	-
4	-	-	0	-	0	-	-	-
5	-	0	-	-	-	0	0	0
6	-	-	-	-	-	0	0	-
7	-	-	-	-	-	0	-	0

Adjacency lists for One-Mode Networks

taken from http://courseweb.sp.cs.cmu.edu/~cs111/applications/ln/lecture18.html



Incidence Matrix for One-Mode Networks

- (Another) complete description of a graph
- Nodes indexing the rows, lines indexing the columns
- g nodes and L lines, the matrix I is of size g x L
- A „1“ indicates that a node n_i is incident with line l_j
- Each column has exactly two 1‘s in it

Table 4.3. *Example of an incidence matrix: “lives near” relation for six children*

[Dotted line]

I

	l_1	l_2	l_3	l_4	l_5	l_6
n_1	1	1	0	0	0	0
n_2	0	0	1	0	0	0
n_3	0	0	1	0	0	0
n_4	0	0	0	1	1	0
n_5	1	0	0	1	0	1
n_6	0	1	0	0	1	1

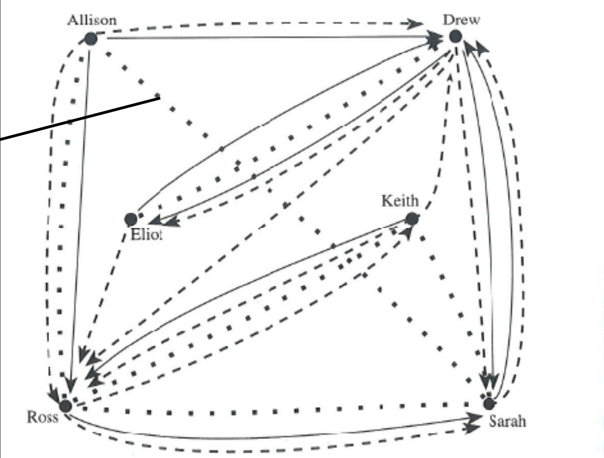


Fig. 3.2. The six actors and the three sets of directed lines — a multivariate directed graph

[Wasserman Faust 1994]

Adjacency lists vs. matrices

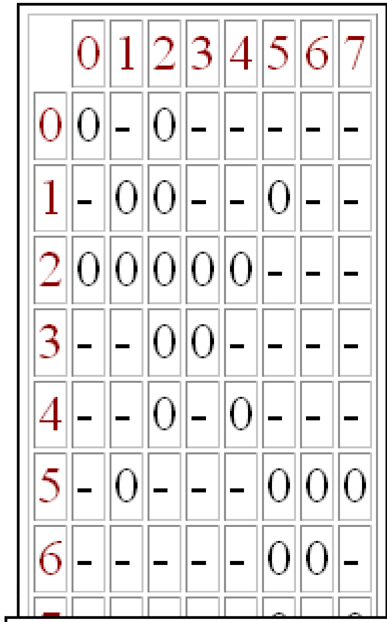
taken from http://courseweb.sp.cs.cmu.edu/~cs111/applications/ln/lecture18.html

Lists Vs. Matrices (I)

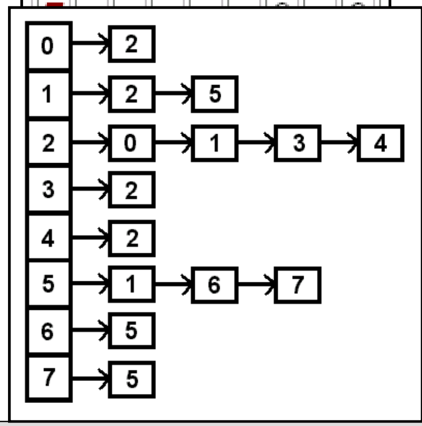
If the graph is sparse (there aren't many edges), then the **matrix will take up a lot of space** indication all of the pairs of vertices which don't have an edge between them, but the **adjacency list does not have that problem**, because it only keeps track of what edges are actually in the graph.

On the other hand, if there are a lot of edges in the graph, or if it is fully connected, then the list has a **lot of overhead** because of all of the references.

.



	0	1	2	3	4	5	6	7
0	0	-	0	-	-	-	-	-
1	-	0	0	-	-	0	-	-
2	0	0	0	0	0	-	-	-
3	-	-	0	0	-	-	-	-
4	-	-	0	-	0	-	-	-
5	-	0	-	-	-	0	0	0
6	-	-	-	-	-	0	0	-



Adjacency lists vs. matrices

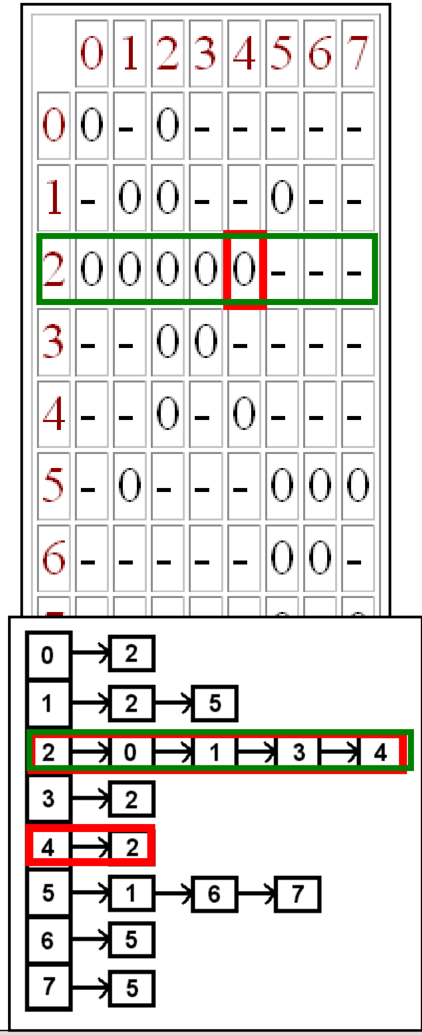
taken from http://courseweb.sp.cs.cmu.edu/~cs111/applications/ln/lecture18.html

Lists Vs. Matrices (II)

If we need to **look** specifically **at a given edge**, we can go right to that spot in the matrix, but in the list we might have to traverse a long linked list before we hit the end and find out that **it is not in the graph**.

If we need to **look at all of a vertex's neighbors**, if you use a matrix you will have to **scan through all of the vertices which aren't neighbors** as well, whereas in the list you **can just scan the linked-list** of neighbors.

.



Adjacency lists vs. matrices

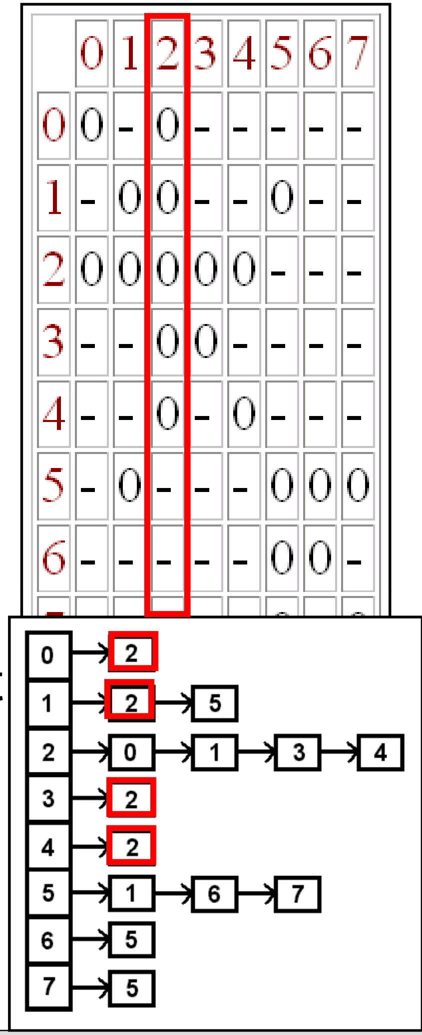
taken from http://courseweb.sp.cs.cmu.edu/~cs111/applications/ln/lecture18.html

Lists Vs. Matrices (III)

If, in a directed graph, we ask the question, "Which vertices have edges leading to vertex X?", the answer is **straight-forward to find in an adjacency matrix** - we just walk down column X and report all of the edges that are present. But, life isn't so easy with the adjacency list - we actually have to perform a brute-force search.

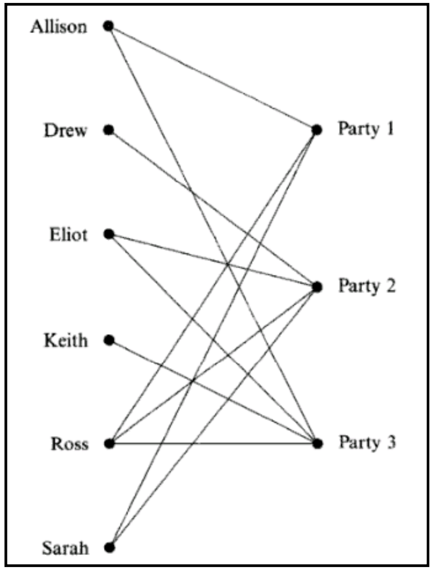
⇨ So which representation you use depends on what you are trying to represent and what you plan on doing with the graph

Illustration!



Adjacency matrices for for Two-Mode Networks

- Complete description of a graph
- A row and a column for each node
- Of size m x n (m rows and n colums)



	Allis on	Drew	Eliot	Keith	Ross	Sarah
Party 1	1	0	0	0	1	1
Party 2	0	1	1	0	1	1
Party 3	1	0	1	1	1	0

Network Metrics for One-Mode Networks

- If the distance between all pairs is finite, we say the **network is connected** (a single component); else it has multiple components
- **Degree of vertex v**: number of edges connected to v
- **Average degree of vertex v**: avg. number of edges connected to a vertex

Two Mode Networks - Rates of Participation

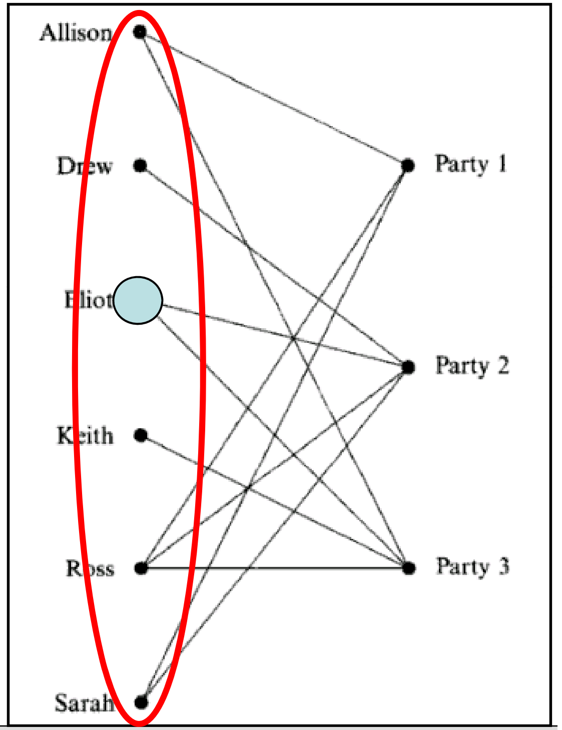
[Wasserman Faust 1994]

- The number of events with which each actor is affiliated.
- These quantities are either given by
  - the row totals of affiliation matrix A or
  - the entries on the main diagonal of the one-mode socio-matrice X^N
- Thus, the number of events with which actor i is affiliated is equal to the degree of the node representing the actor in the bipartite graph.
- Also interesting: **Average rate of participation**

Examples: What does the rate of participation relate to in the Netflix / Amazon bipartite graph of customer/movies or customer/products?

Example:

	Party 1	Party 2	Party 3
Allison	1	0	1
Drew	0	1	0
Eliot	0	1	1
Keith	0	0	1
Ross	1	1	1
Sarah	1	1	0



Two Mode Networks - Size of Events

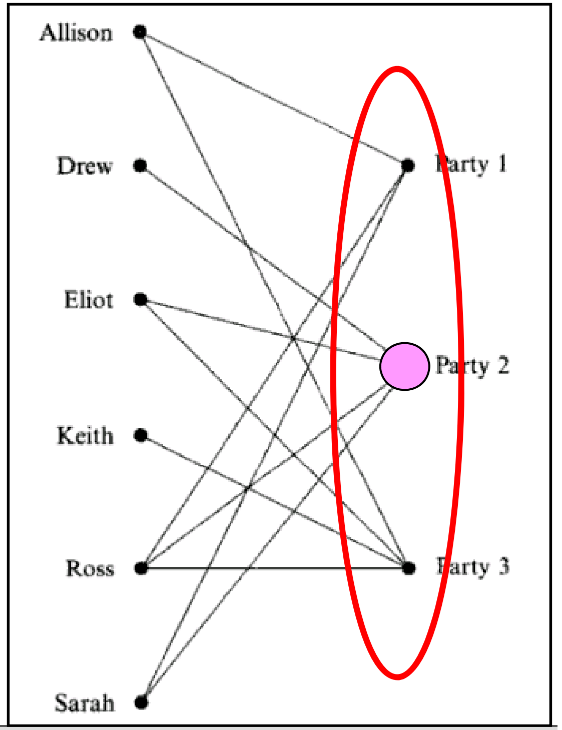
[Wasserman Faust 1994]

- The number of actors participating in each event.
- The size of each event is given by either
  - the column totals of the affiliation matrix A or
  - the entries on the main diagonal of the one-mode sociomatrix X^M.
- Thus, the size of each event is equal to the degree of the node representing the event in the bipartite graph.
- Also interesting: **Average size of events**
  - Sometimes useful to study average size of clubs or organizations
- Size of events might be constrained:
  - E.g. board of company directors are made up of a fixed number of people

Examples: What does the rate of participation relate to in the Netflix / Amazon bipartite graph of customer/movies or customer/products?

Example:

	Party 1	Party 2	Party 3
Allison	1	0	1
Drew	0	1	0
Eliot	0	1	1
Keith	0	0	1
Ross	1	1	1
Sarah	1	1	0



Terminology II

http://www.cis.upenn.edu/~Emkearns/teaching/NetworkedLife/

- Network size: total number of vertices (denoted N)
- Maximum number of edges (undirected): N(N-1)/2 ~ N^2/2
- Distance or geodesic path L between vertices u and v:
  - number of edges on the **shortest path** from u to v
  - can consider directed or undirected cases
  - infinite if there is no path from u to v
- Diameter of a network
  - worst-case diameter: largest distance between a pair
  - Diameter: longest shortest path between any two pairs
  - average-case diameter: average distance
- If the distance between all pairs is finite, we say the network is connected; else it has multiple components
- Degree of vertex v: number of edges connected to v
- Density: ratio of edges to vertices

Definitions

[Newman 2003]

Vertex (pl. vertices): The fundamental unit of a network, also called a site (physics), a node (computer science), or an actor (sociology).

Edge: The line connecting two vertices. Also called a bond (physics), a link (computer science), or a tie (sociology).

Directed/undirected: An edge is directed if it runs in only one direction (such as a one-way road between two points), and undirected if it runs in both directions. Directed edges, which are sometimes called *arcs*, can be thought of as sporting arrows indicating their orientation. A graph is directed if all of its edges are directed. An undirected graph can be represented by a directed one having two edges between each pair of connected vertices, one in each direction.

Degree: The number of edges connected to a vertex. Note that the degree is not necessarily equal to the number of vertices adjacent to a vertex, since there may be more than one edge between any two vertices. In a few recent articles, the degree is referred to as the "connectivity" of a vertex, but we avoid this usage because the word connectivity already has another meaning in graph theory. A directed graph has both an in-degree and an out-degree for each vertex, which are the numbers of in-coming and out-going edges respectively.

Component: The component to which a vertex belongs is that set of vertices that can be reached from it by paths running along edges of the graph. In a directed graph a vertex has both an in-component and an out-component, which are the sets of vertices from which the vertex can be reached and which can be reached from it.

Geodesic path: A geodesic path is the shortest path through the network from one vertex to another. Note that there may be and often is more than one geodesic path between two vertices.

Diameter: The diameter of a network is the length (in number of edges) of the longest geodesic path between any two vertices. A few authors have also used this term to mean the *average* geodesic distance in a graph, although strictly the two quantities are quite distinct.

Terminology III

http://www.infosci.cornell.edu/courses/info204/2007sp/

[Diestel 2005]

In undirected networks

- Paths
  - A sequence of nodes v_1, .., v_i, v_{i+1},…,v_k with the property that each consecutive pair v_i, v_{i+1} is joined by an edge in G
- Cycles (in undirected networks)
  - A path with $v_1 = v_k$ (Begin and end node are the same)
  - Cyclic vs. Acyclic (not containing any cycles: e.g. forests) networks

In directed networks

- Path or cycles must respect directionality of edges

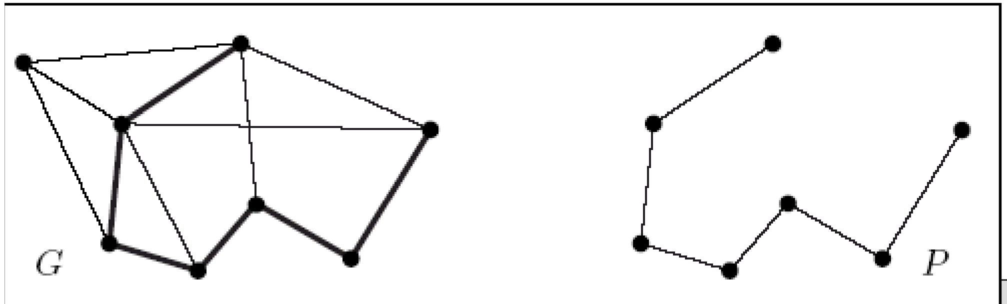


Fig. 1.3.1. A path $P = P^6$ in G

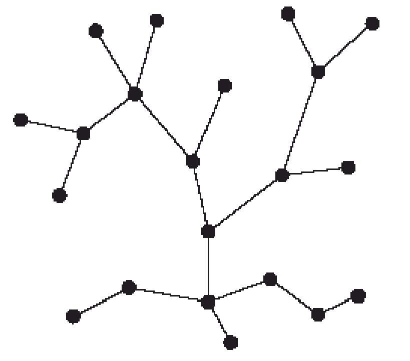


Fig. 1.5.1. A tree

Other types of networks

[Newman 2003]

Undirected, single edge and node type

Undirected, multiple edge and node types

Undirected, varying edge and node weights

Directed, each edge has a direction

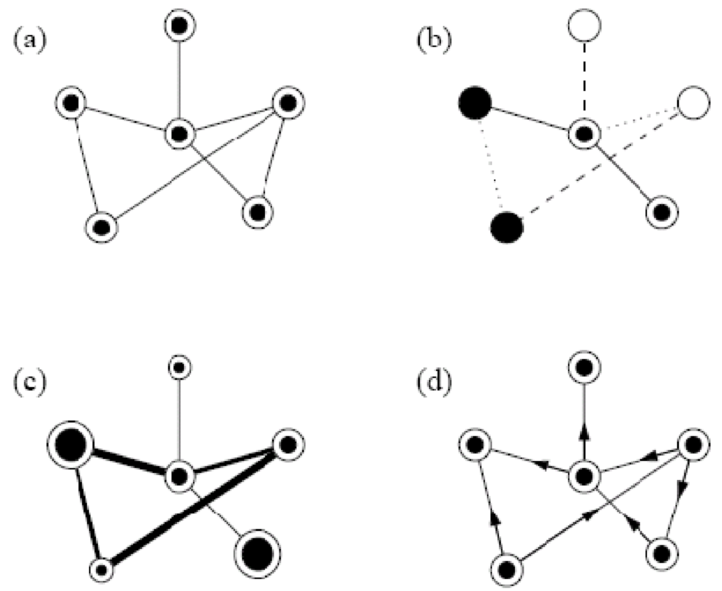


FIG. 3 Examples of various types of networks: (a) an undirected network with only a single type of vertex and a single type of edge; (b) a network with a number of discrete vertex and edge types; (c) a network with varying vertex and edge weights; (d) a directed network in which each edge has a direction.

Terminology IV

http://www.infosci.cornell.edu/courses/info204/2007sp/

- Average Pairwise Distance
  - The average distance between all pairs of nodes in a graph. If the graph is unconnected, the average distance between all pairs in the largest component.
- Connectivity
  - An undirected graph is connected if for every pair of nodes u and v, there is a path from u to v (there is not more than one component).
  - A directed graph is strongly connected if for every two nodes u and v, there is a path from u to v and a path from v to u
- Giant Component
  - A single connected component that accounts for a significant fraction of all nodes

Average degree k

http://www.infosci.cornell.edu/courses/info204/2007sp/

- Average degree k
  - Degree: The number of edges for which a node is an endpoint
  - In undirected graphs: number of edges
  - In directed graphs: k_{in} and k_{out}
  - Average degree: average of the degree of all nodes, a measure for the density of a graph

$$d(G) := \frac{1}{|V|} \sum_{v \in V} d(v)$$

Degree Distributions

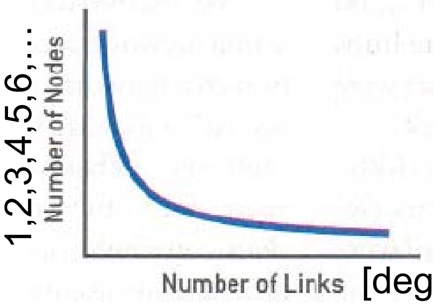
[Barabasi and Bonabeau 2003]

- Degree distribution p(k)
  - A plot showing the fraction of nodes in the graph of degree k, for each value of k

Related concepts

- Degree histogram
- Rank / frequency plot
- Cumulative Degree function (CDF)
- Pareto distribution

Example:



Degree Distributions Examples

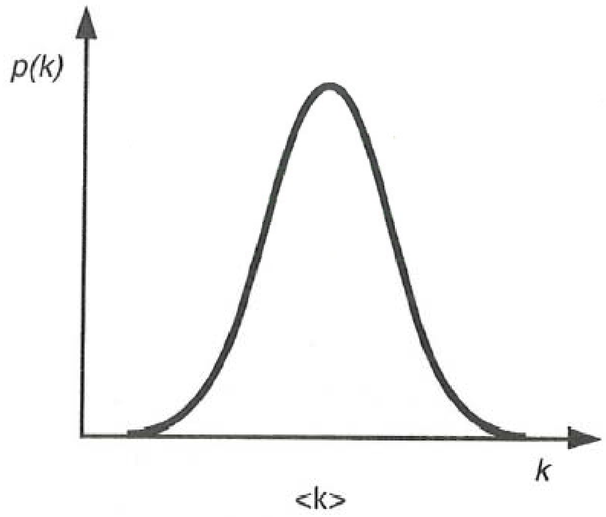


Figure 4.1. The normal distribution specifies the probability, *p(k)*, that a randomly selected node will have *k* neighbors. The average degree <k> lies at the peak of the distribution.

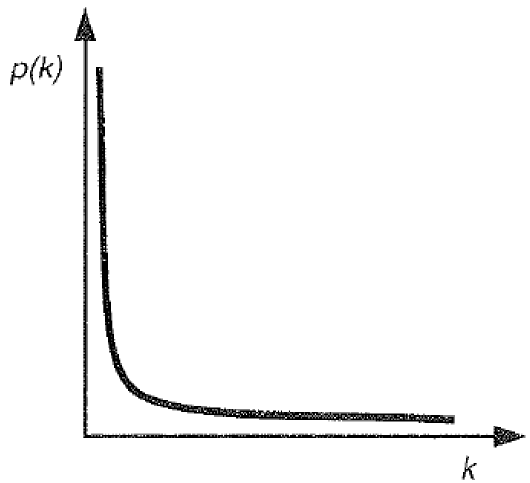
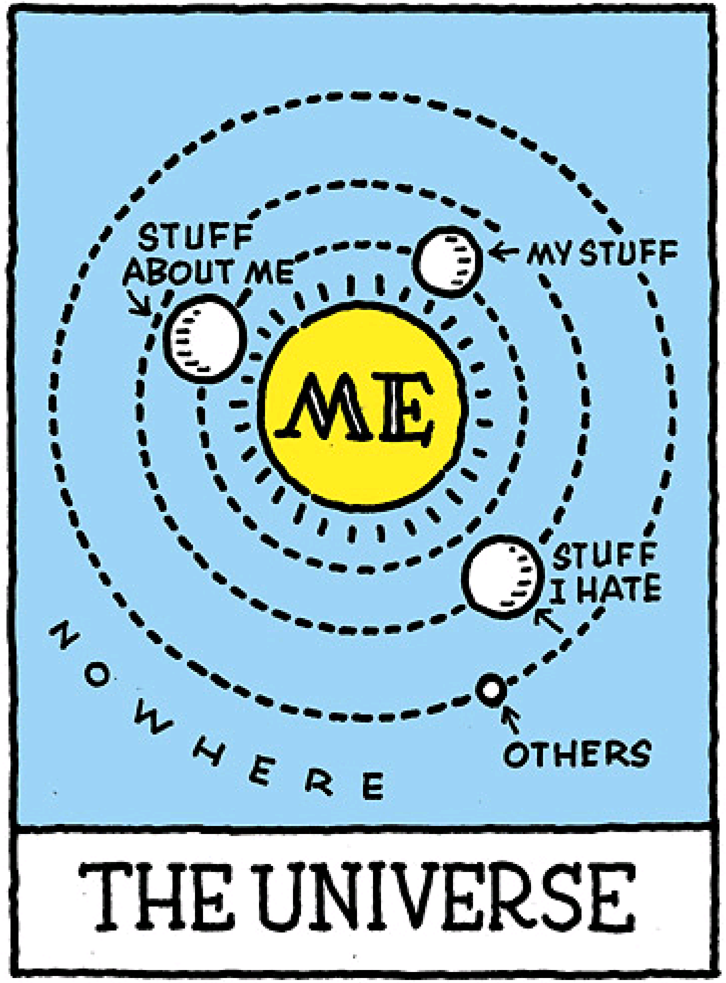


Figure 4.2. A power-law distribution. Although it decreases rapidly with *k*, it does so much slower than the normal distribution in figure 4.1, implying than large values of *k* are more likely.

Ego-Centred Networks

Definition:



http://www.nytimes.com/2008/07/20/fashion/20narcissist.html

Clustering Coefficient

http://www.infosci.cornell.edu/courses/info204/2007sp/

- Clustering Coefficient C
  - Triangles or closed triads: Three nodes with edges between all of them
  - over all sets of three nodes in the graph that form a connected set (i.e. one of the three nodes is connected to all the others), what fraction of these sets in fact form a triangle?
  - This fraction can range from 0 (when there are no triangles) to 1 (for example, in a graph where there is an edge between each pair of nodes — such a graph is called a clique, or a complete graph).
  - Or in other words: The clustering coefficient gives the fraction of pairs of neighbors of a vertex that are adjacent, averaged over all vertices of the graph. [p344, Brandes and Erlebach 2005]
  - Page 88, [Watts 2005]
  - Related: „Transitivity“

Clustering Coefficient

Images taken from http://en.wikipedia.org/w/index.php?title=Clustering_coefficient&oldid=152650779

- Number of edges between neighbours of a given node divided by the number of possible edges between neighbours
- Directed Graphs

$$C_i = \frac{|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{jk} \in E.$$

- Undirected Graphs

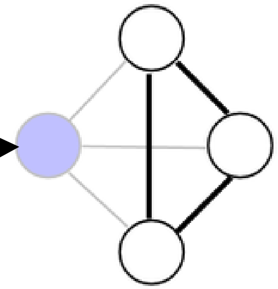
$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{ij} \in E.$$

Actual edges between neighbourhood nodes

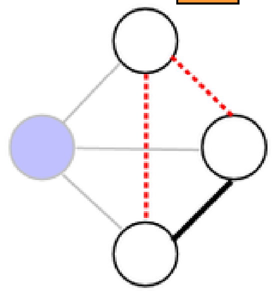
1/Number of potential edges between neighbours

Degree

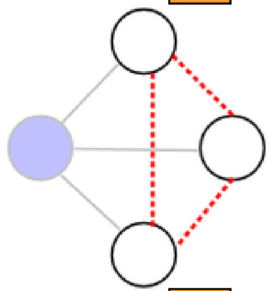
Neighbourhood nodes



c = ?



c = ?



c = ?

Graph Theory & Network Theory

- Graph Theory
  - Mathematics of graphs
  - Networks with pure structure with properties that are fixed over time
  - Focus on syntax rather than semantics
    - Nodes and edges do not have semantics
    - E.g. A node does not have a social identity
  - Concerned with characteristics of graphs
  - Proofs
  - Algorithms

Network Theory

- Relate to real-world phenomena
  - Social networks
  - Economic networks
  - Energy networks
- Networks are *doing something*
  - *Making new relations*
  - *Making money*
  - *Producing power*
- Are dynamic
  - Structure: Dynamics of the network
  - Agency: Dynamics in the network
- *Are active, which effects*
  - *Individual behavior*
  - *Behavior of the network as a whole*

Networks
[Watts 2003]

TABLE 3.2 STATISTICS OF SMALL WORLD NETWORKS

	L_{ACTUAL}	L_{RANDOM}	C_{ACTUAL}	C_{RANDOM}
MOVIE ACTORS	3.65	2.99	0.79	0.00027
POWER GRID	18.7	12.4	0.080	0.005
C. ELEGANS	2.65	2.25	0.28	0.05

L=Path Length; *C*=Clustering Coefficient.

C. elegans is a worm, one of the simplest organisms with a nervous system.

Compared to imaginery random networks

Network Theory

- Are there general statements we can make about *any* class of network?
- A Science of Networks

Random Networks

- Page 44/ff, Watts 2003, random graphs

Random graph: a network of nodes connected by links in a purely random fashion.

Analogy of Stuart Kaufmann: Throw a boxload of buttons onto the floor, then choose

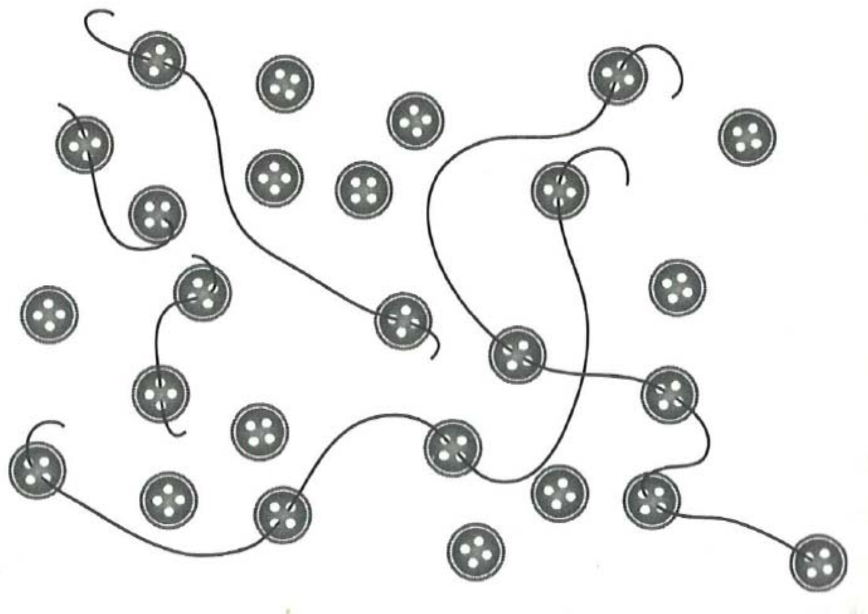


Figure 2.1. A random graph imagined as a collection of buttons tied by strings. Pairs of nodes (buttons) are connected at random by links or ties.

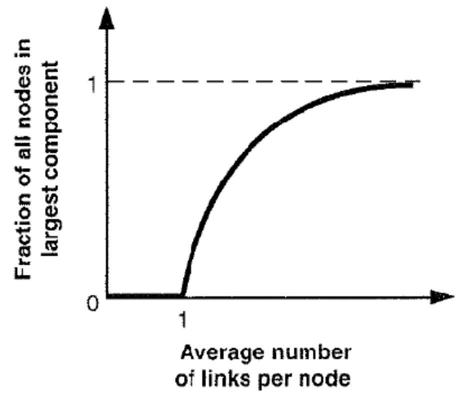
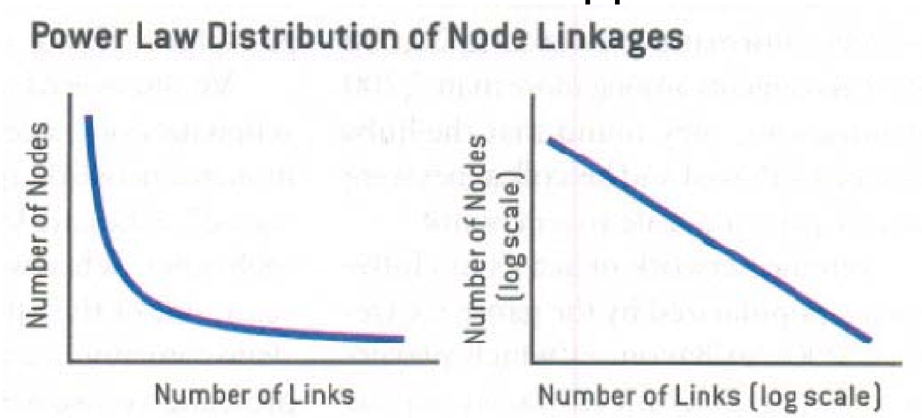


Figure 2.2. Connectivity of a random graph. The fraction of the nodes connected in a single component changes suddenly when the average number of links per node exceeds one.

Scale-Free Networks

[Barabasi and Bonabeau 2003]

- Some nodes have a tremendous number of connections to other nodes (hubs), whereas most nodes have just a handful
- Robust against accidental failures, but vulnerable to coordinated attacks (**DEMO**: http://projects.si.umich.edu/netlearn/GUESS/resiliencedegree.html)
- Popular nodes can have millions of links: The network appears to have no scale (no limit)
- Two prerequisites: [watts2003]
  - Growth
  - Preferential attachment



- Problem:
  - Scale-free networks are only ever truly scale-free when the network is infinitely large (whereas in practice, the are mostly not)
  - This introduces a cut off [page 111, watts 2003]

Scale-free Networks

[Watts 2003]

The alpha parameter

- ***y = C x***$^{-\alpha}$ (c, α being constants) or ***log(y) = log(C) - α log(x)***
- ***a power-law with exponent α is depicted as a straight line with slope -a on a log-log plot***

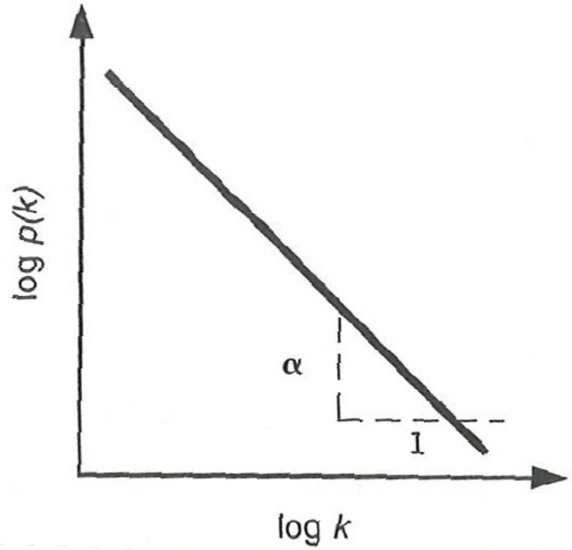


Figure 4.3. A power-law distribution on a log-log plot. The exponent alpha (α) is the slope of the line (the line drops by α for each unit on the horizontal axis).

Examples

•If the number of cities of a given size decreases in inverse proportion to the size, then we say the distribution has an exponent of [*one*/**two**]

That means, we are likely to see cities such as Graz (250.000) roughly [*ten*/**hundred**] times as frequently as cities like Vienna (including the Greater Vienna Area, *roughly* 10 times larger)

Networks [Newman 2003]

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	416, 421

TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n; total number of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out-degree exponents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r, Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

Scale-Free Networks

– cut off [page 111, watts 2003]

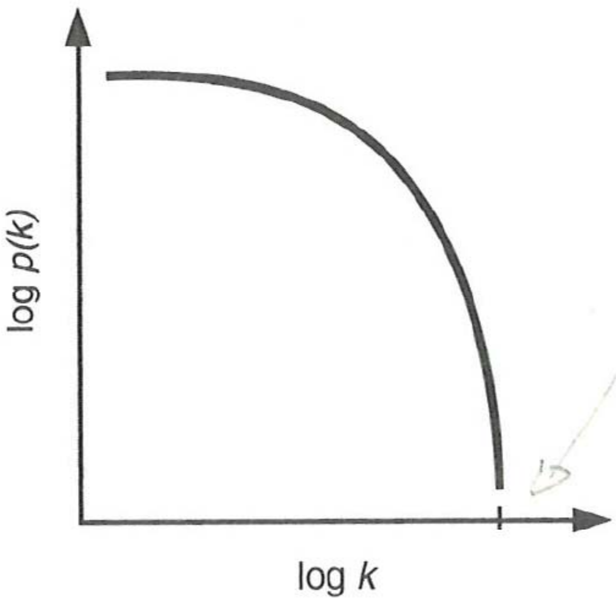


Figure 4.4. A normal-type distribution on a log-log plot. The *cutoff* occurs where the curve disappears into the horizontal axis.

Scale-Free Networks

Limited maximum degree because of e.g. the finite set of nodes in a network

– cut off [page 111, watts 2003]

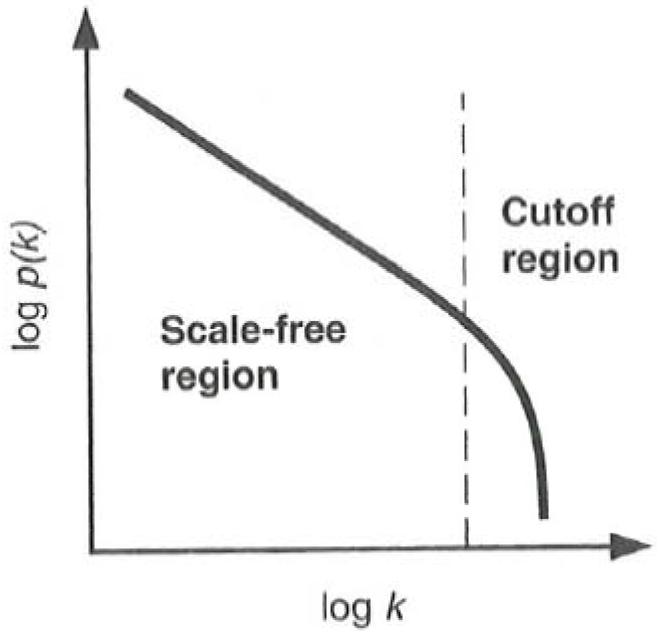


Figure 4.5. In practice, power-law distributions always display a characteristic cutoff because of the finite size of the system. The observed degree distribution, therefore, is only ever a straight line on a log-log plot, over some range.

Examples of Scale-Free Networks

[Newman 2003]

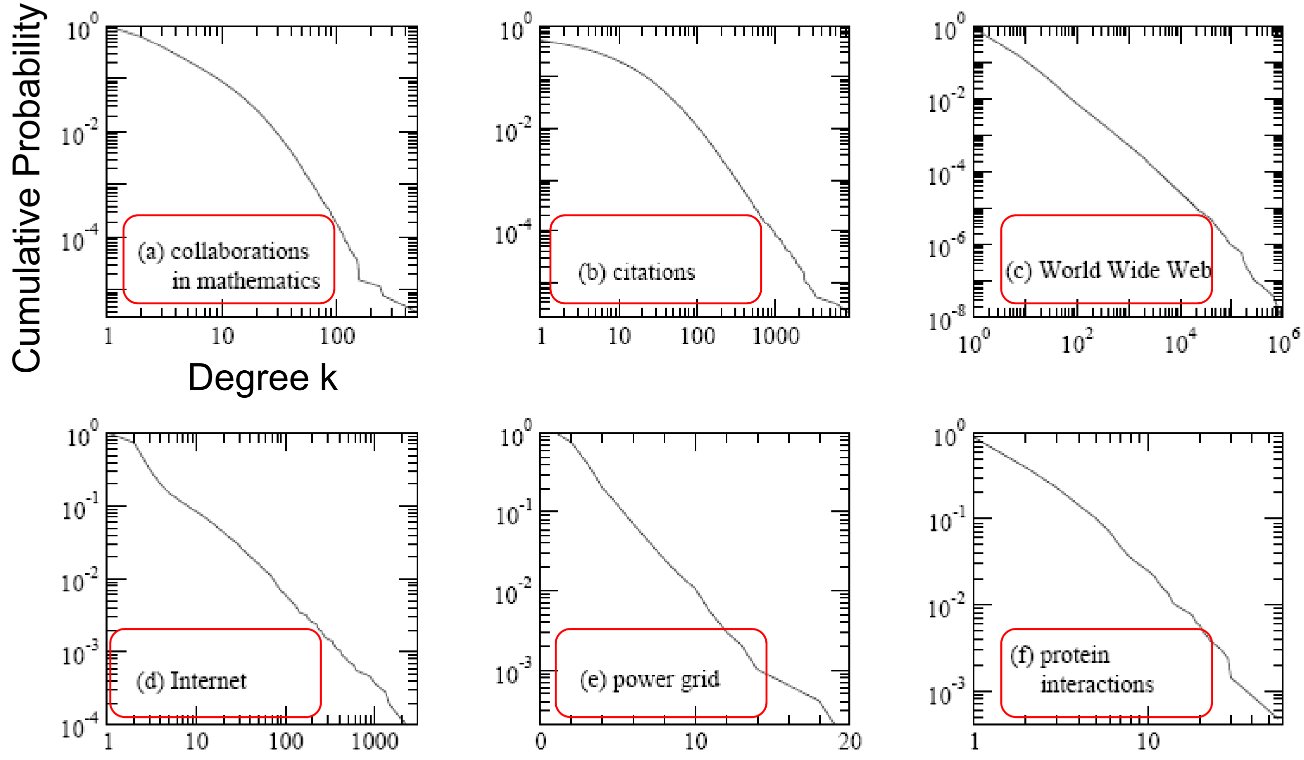


FIG. 6 Cumulative degree distributions for six different networks. The horizontal axis for each panel is vertex degree k (or in-degree for the citation and Web networks, which are directed) and the vertical axis is the cumulative probability distribution of degrees, i.e., the fraction of vertices that have degree greater than or equal to k. The networks shown are: (a) the collaboration network of mathematicians [182]; (b) citations between 1981 and 1997 to all papers cataloged by the Institute for Scientific Information [351]; (c) a 300 million vertex subset of the World Wide Web, *circa* 1999 [74]; (d) the Internet at the level of autonomous systems, April 1999 [86]; (e) the power grid of the western United States [416]; (f) the interaction network of proteins in the metabolism of the yeast *S. Cerevisiae* [212]. Of these networks, three of them, (c), (d) and (f), appear to have power-law degree distributions, as indicated by their approximately straight-line forms on the doubly logarithmic scales, and one (b) has a power-law tail but deviates markedly from power-law behavior for small degree. Network (e) has an exponential degree distribution (note the log-linear scales used in this panel) and network (a) appears to have a truncated power-law degree distribution of some type, or possibly two separate power-law regimes with different exponents.

Graph Structure in the Web

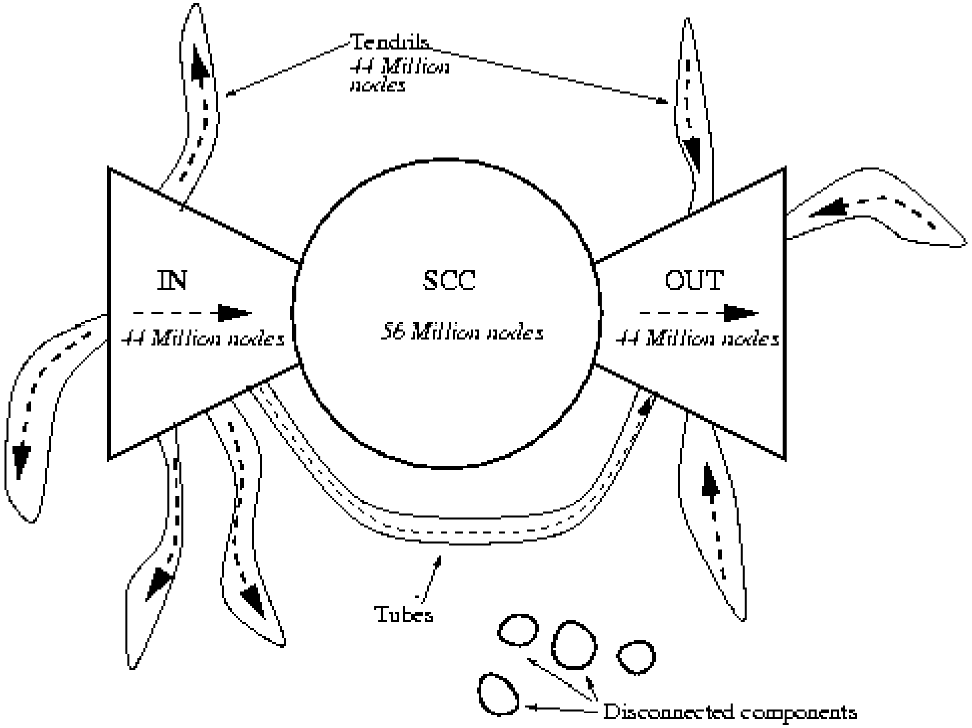
[Broder et al 2000]

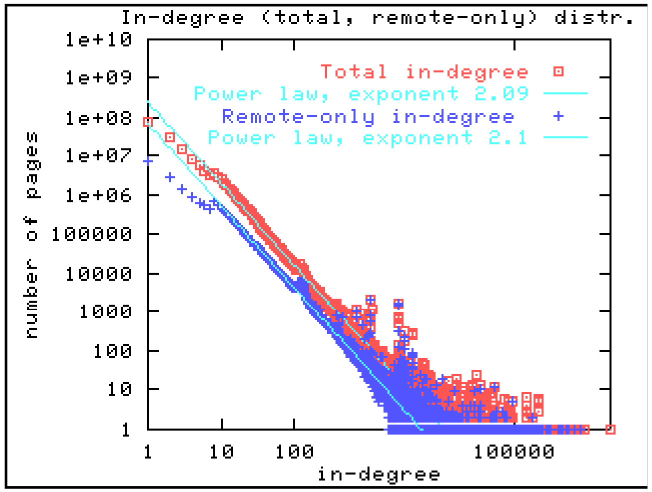
Most (over 90%) of the approximately 203 million nodes in a May 1999 crawl form a connected component if links are treated as *undirected* edges.

IN consists of pages that can reach the SCC, but cannot be reached from it

OUT consists of pages that are accessible from the SCC, but do not link back to it

TENDRILS contain pages that cannot reach the SCC, and cannot be reached from the SCC





Interesting Results

[Broder et al 2000]

• the diameter of the central core (*SCC*) is at least 28, and the diameter of the graph as a whole is over 500

• for randomly chosen source and destination pages, the probability that any path exists from the source to the destination is only 24%

• if a directed path exists, its average length will be about 16

•if an undirected path exists (i.e., links can be followed forwards or backwards), its average length will be about 6

Scale-Free vs. Random Networks

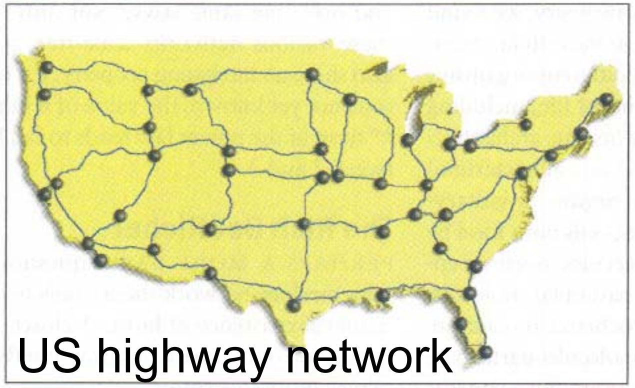
[Barabasi and Bonabeau 2003]

RANDOM VERSUS SCALE-FREE NETWORKS

RANDOM NETWORKS, which resemble the U.S. highway system (*simplified in left map*), consist of nodes with randomly placed connections. In such systems, a plot of the distribution of node linkages will follow a bell-shaped curve (*left graph*), with most nodes having approximately the same number of links.

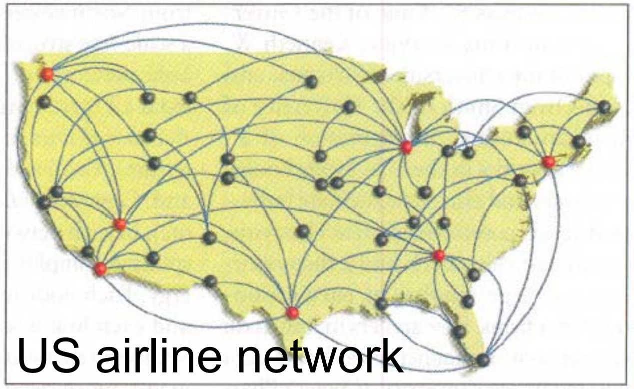
In contrast, scale-free networks, which resemble the U.S. airline system (*simplified in right map*), contain hubs (*red*)— nodes with a very high number of links. In such networks, the distribution of node linkages follows a power law (*center graph*) in that most nodes have just a few connections and some have a tremendous number of links. In that sense, the system has no "scale." The defining characteristic of such networks is that the distribution of links, if plotted on a double-logarithmic scale (*right graph*), results in a straight line.

Random Network



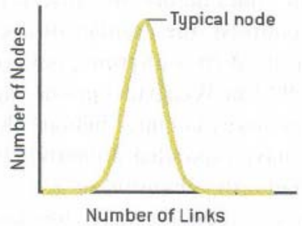
US highway network

Scale-Free Network

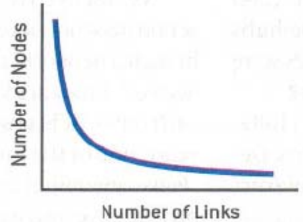


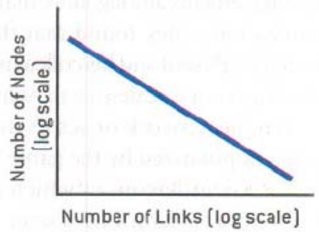
US airline network

Bell Curve Distribution of Node Linkages



Power Law Distribution of Node Linkages





A simple conceptual model for the Internet topology
[Tauro et al 2001]

Jellyfish Model:

- The Internet has a **core of nodes** that **form a clique** and this clique is located in the “middle” of the network.
- The **topological importance** of the nodes **decreases** as we move away from the center.
- The distribution of the **one-degree nodes** across the network **follows a power-law**.
- The Internet topology can be visualized as a **jellyfish**. The value of the model lies in its simplicity and its ability to represent graphically important topological properties.

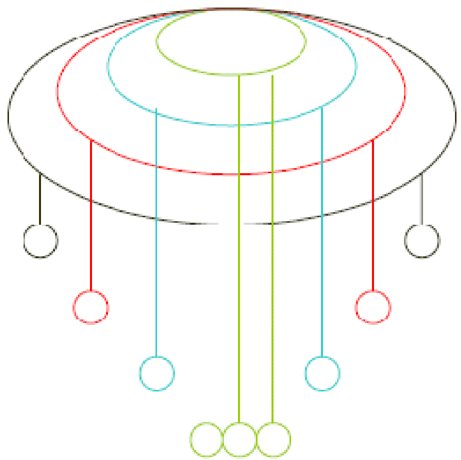


Figure 6: The jellyfish as a model for the AS topology.

[Based on inter-domain connectivity data]

A simple conceptual model for the Internet topology
[Tauro et al 2001]

- Core/Layer 0
  - The maximal clique that contains the highest-degree node
- Layer 1
  - All nodes that are neighbors of the core
- Layer 2
  - All nodes neighbouring layer 1 except for the core

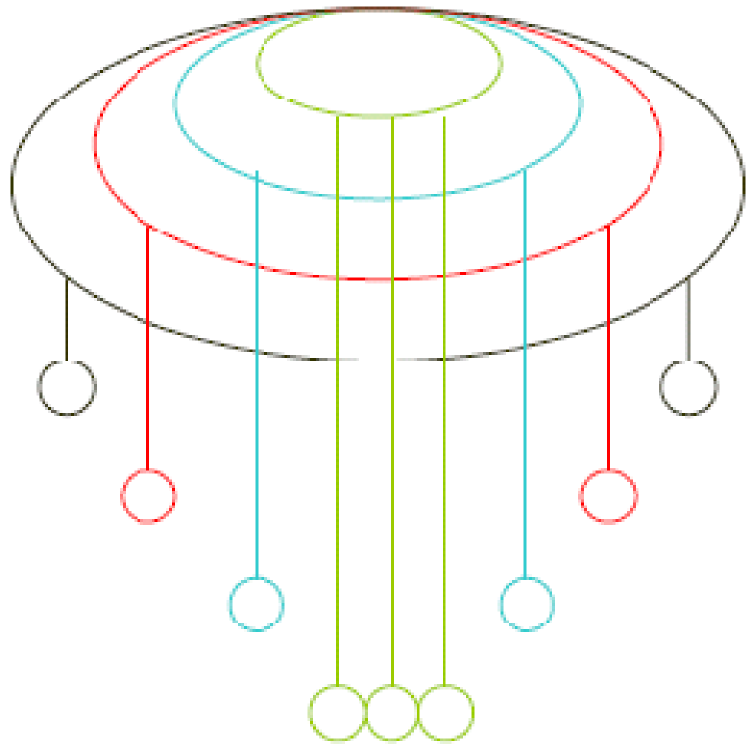


Figure 6: The jellyfish as a model for the AS Internet topology.

A simple conceptual model for the Internet topology
[Tauro et al 2001]

- Core is the center of the cap of the jellyfish
- Layers correspond to shells
- One-degree nodes connected to each shell is shown hanging forming the legs of the jellyfish (Hang-n)
  - Hang-1 has the one-degree nodes attached to Layer 1
- Length of legs represents the concentration of one-degree nodes per shell

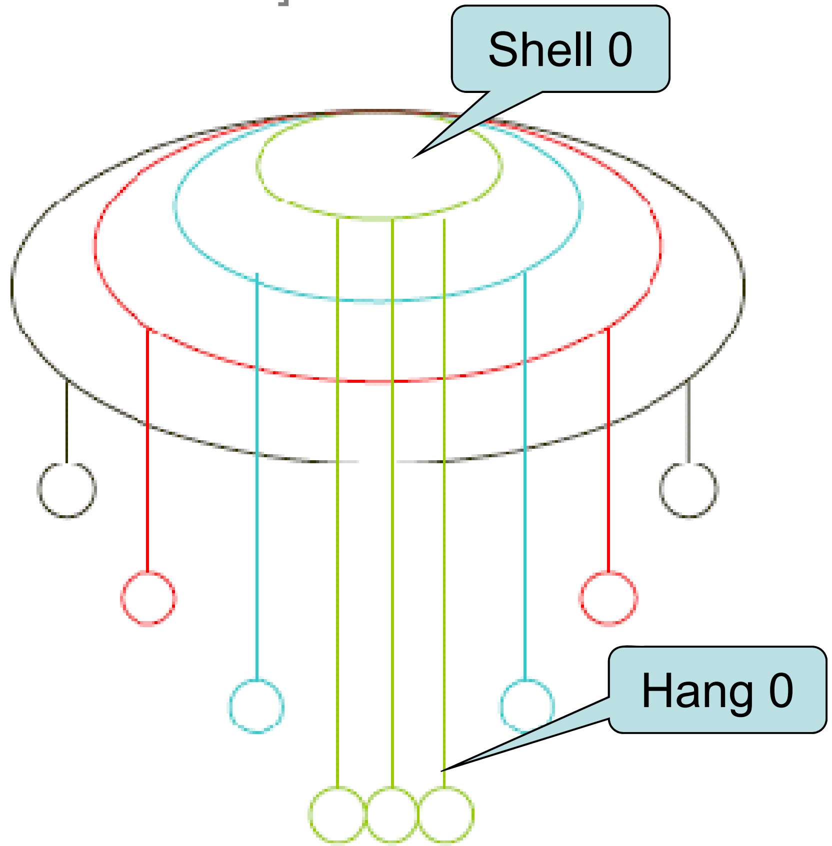


Figure 6: The jellyfish as a model for the AS Internet topology.

A simple conceptual model for the Internet topology
[Tauro et al 2001]

Real Graphs: We use three instances of the inter-domain Internet topology from the end of 1997 until the middle of 2000, which correspond to approximately three yearly intervals. The National Laboratory for Applied Network Research [9] provided the data.

1. Int-11-97: 3015 nodes and 5156 edges.
2. Int-10-98: 5896 nodes and 11424 edges.
3. Int-10-99: 7864 nodes and 15713 edges.

	Int-11-97	Int-10-98	Int-10-99
Core/Shell-0	8	9	13
Hang-0	465	514	808
Shell-1	889	1977	2820
Hang-1	623	1022	1243
Shell-2	579	1418	1812
Hang-2	299	526	683
Shell-3	97	317	394
Hang-3	41	95	67
Shell-4	2	13	14
Hang-4	12	5	10

Table 2: Distribution of nodes in Hanging layers and shells

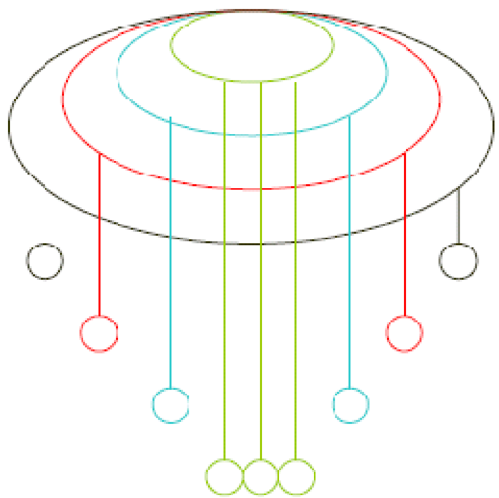


Figure 6: The jellyfish as a model for the AS Internet topology.

- Highest degree nodes exhibit the most relations to one-degree nodes
- Nodes at the layers need to go through the core for the majority of their shortest paths
- One-degree nodes useless in terms of connectivity
- The network is very sensitive to failures of the important nodes, while it is insensitive to random node failures

Bipartite Networks [Watts 2003, Page 120]

- Can always be represented as unipartite networks

Can you give examples for bipartite networks on the web?

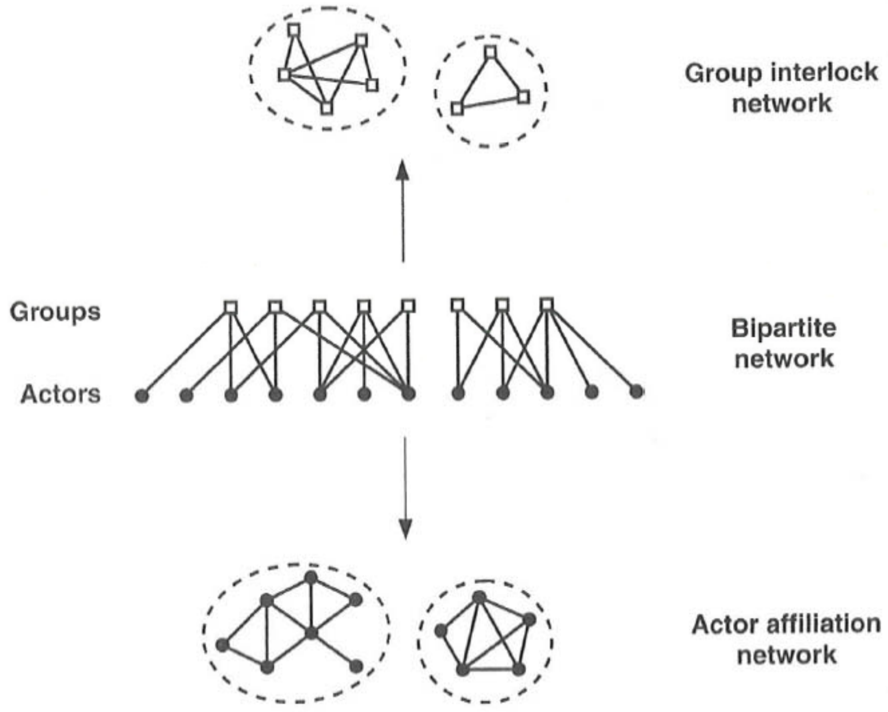


Figure 4.6. Affiliation networks are best represented as bipartite networks (center) in which actors and groups appear as distinct kinds of modes. Bipartite networks can always be projected onto one of two single-mode networks representing affiliations between the actors (bottom) or interlocks between groups (top).

Hierarchical Networks

- P39, [Watts2003]

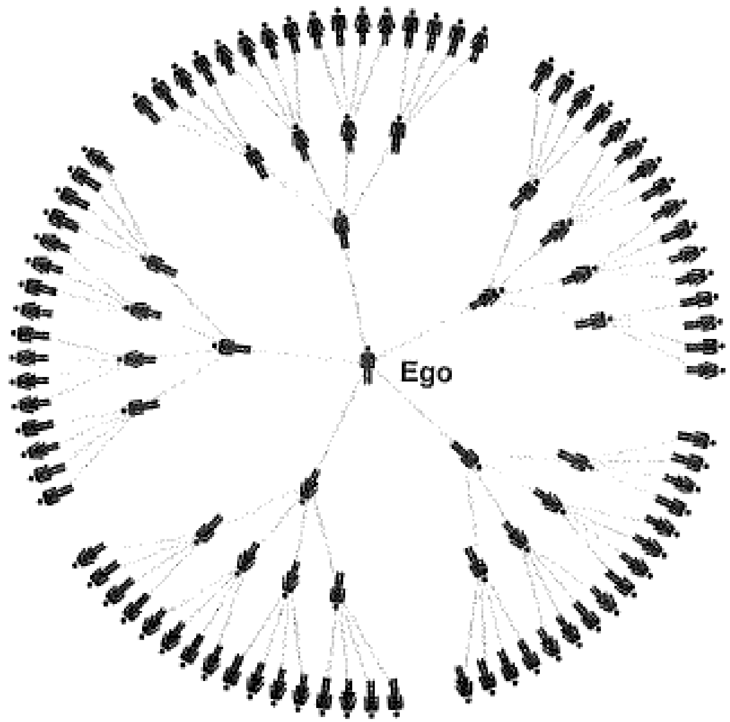


Figure 1.2. A pure branching network. Ego knows only 5 people, but within two degrees of separation, ego can reach 25; within three degrees, 105; and so on.

Home Assignment 1.2

- Released today
- http://www.kmi.tugraz.at/staff/markus/courses/SS2009/707.000_web-science/
- In case of any questions, do not hesitate to post to the newsgroup tu-graz.lv.web-science

Any questions?

See you next week!