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

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

Why and How to Flash Your BIOS
this url has been saved by 106 people. 
save this to your bookmarks »

Why and How to Flash Your BIOS
Aug '07
This article is going to focus on the basics and explain ways to flash the BIOS, precautions and how to recover in case of a bad flash.

posting history
» first posted by farbizar to system:unfiled
by JillSw3d3 to hardware
by gts to bios boost post speed software flash pc hardware computer
by agil292 to bios flash computer hardware
by Curioso44 to boot
by catfish182 to system:unfiled
by anurag_bhd to howto flash bios
Social Networks Entities
Simplified

Xing:
- Person

Flickr:
- User
  - Photo

Last.fm:
- User
  - Song/Band

Del.icio.us:
- User
  - URL
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, Zengestrom¹ argues that successful social software focuses on similar 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? What’s 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
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

• Release of Home Assignment 1.1
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
Social Networks Examples

Why and How to Flash Your BIOS
this url has been saved by 106 people.
save this to your bookmarks »

user notes

Why and How to Flash Your BIOS

Aug '07

This article is going to focus on the basics and explain ways to flash the BIOS, precautions and how to recover in case of a bad flash.

edwinek

Why and How to Flash Your BIOS (Page 1 of 4) Flashing the BIOS is one of the most feared topics related to computers. Yes, people should be very cautious because it can be dangerous. This article is going to focus on the basics and explain ways to flash

oblonski

common tags
cloud list article articles bios computer computers diagnostic flash geek guide hacking hardware howto lifehacker pc reference software tech technology toread tutorial tutorials utilities windows

posting history

» first posted by farbians to system:unfiled

Sep '07

by JillSw3c3 to hardware
by gtss to bios boost post speed software flash pc
hardware computer
by agil292 to bios flash computer hardware
by Curioso44 to boot
by catfish182 to system:unfiled
by anurag_bhd to howto flash bios
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 $g \times g$ ($g$ rows and $g$ columns)
Adjacency matrices for One-Mode Networks

taken from http://courseweb.sp.cs.cmu.edu/~cs111/applications/Ln/lecture18.html
Adjacency lists for One-Mode Networks

taken from http://courseweb.sp.cs.cmu.edu/~cs111/applications/ln/lecture18.html
Incidience 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 \( g \times L \)
• 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

<table>
<thead>
<tr>
<th></th>
<th>( l_1 )</th>
<th>( l_2 )</th>
<th>( l_3 )</th>
<th>( l_4 )</th>
<th>( l_5 )</th>
<th>( l_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 )</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( n_2 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( n_3 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( n_4 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( n_5 )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( n_6 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

[Wasserman Faust 1994]

Fig. 3.2. The six actors and the three sets of directed lines — a multivariate directed graph
Adjacency lists vs. matrices
taken from http://courseweb.sp.cs.cmu.edu/~cs111/applications/ln/lecture18.html

Lists Vs. Matrices (l)

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.
Lists Vs. Matrices (II)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

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.
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!

```
<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
```

```
0 → 2 → 5
1 → 0 → 3 → 4
2 → 0
3 → 2
4 → 2
5 → 1 → 6 → 7
6 → 5
7 → 5
```
Adjacency matrices 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 columns)

<table>
<thead>
<tr>
<th></th>
<th>Allison</th>
<th>Drew</th>
<th>Eliot</th>
<th>Keith</th>
<th>Ross</th>
<th>Sarah</th>
</tr>
</thead>
<tbody>
<tr>
<td>Party 1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Party 2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Party 3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
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?
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?
• Network size: total number of vertices (denoted N)
• Maximum number of edges (undirected): \( N(N-1)/2 \sim N^2/2 \)
• 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.
In undirected networks

• Paths
  – A sequence of nodes $v_1, \ldots, v_i, v_{i+1}, \ldots, v_k$ 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 $G$
Other types of networks
[Newman 2003]

Undirected, single edge and node type

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 $k$

- **Degree**: The number of edges for which a node is an endpoint
- **In undirected graphs**: number of edges
- **In directed graphs**: $k_{\text{in}}$ and $k_{\text{out}}$
- **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:

```
1, 2, 3, 4, 5, 6, ...
or: 6, 5, 4, 3, 2, 1
```
Degree Distributions Examples

Figure 4.1. The normal distribution specifies the probability, $\rho(k)$, 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 that large values of $k$ are more likely.
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


- **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**

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

- Degree
- Neighbourhood nodes
- Actual edges between neighbourhood nodes

1/Number of potential edges between neighbours

Markus Strohmaier 2013
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>
<thead>
<tr>
<th></th>
<th>$L_{\text{ACTUAL}}$</th>
<th>$L_{\text{RANDOM}}$</th>
<th>$C_{\text{ACTUAL}}$</th>
<th>$C_{\text{RANDOM}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOVIE ACTORS</td>
<td>3.65</td>
<td>2.99</td>
<td>0.79</td>
<td>0.00027</td>
</tr>
<tr>
<td>POWER GRID</td>
<td>18.7</td>
<td>12.4</td>
<td>0.080</td>
<td>0.005</td>
</tr>
<tr>
<td>C. ELEGANS</td>
<td>2.65</td>
<td>2.25</td>
<td>0.28</td>
<td>0.05</td>
</tr>
</tbody>
</table>

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

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

Compared to imaginary 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

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.
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.
- 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, they are mostly not).
  - This introduces a cut off [page 111, watts 2003]
The alpha parameter

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

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 \(\text{one/two}\)

That means, we are likely to see cities such as Graz (250,000) roughly \(\text{ten/hundred}\) times as frequently as cities like Vienna (including the Greater Vienna Area, \(\text{roughly} 10 \text{ times larger}\)
# Networks [Newman 2003]

<table>
<thead>
<tr>
<th>network</th>
<th>type</th>
<th>$n$</th>
<th>$m$</th>
<th>$z$</th>
<th>$\ell$</th>
<th>$\alpha$</th>
<th>$C^{(1)}$</th>
<th>$C^{(2)}$</th>
<th>$r$</th>
<th>Ref(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>film actors</td>
<td>undirected</td>
<td>449913</td>
<td>25516482</td>
<td>113.43</td>
<td>3.48</td>
<td>2.3</td>
<td>0.20</td>
<td>0.78</td>
<td>0.208</td>
<td>20, 416</td>
</tr>
<tr>
<td>company directors</td>
<td>undirected</td>
<td>7673</td>
<td>55392</td>
<td>14.44</td>
<td>4.60</td>
<td>-</td>
<td>0.59</td>
<td>0.88</td>
<td>0.276</td>
<td>105, 323</td>
</tr>
<tr>
<td>math coauthorship</td>
<td>undirected</td>
<td>253330</td>
<td>406489</td>
<td>3.92</td>
<td>7.57</td>
<td>-</td>
<td>0.15</td>
<td>0.34</td>
<td>0.120</td>
<td>107, 182</td>
</tr>
<tr>
<td>physics coauthorship</td>
<td>undirected</td>
<td>52900</td>
<td>245300</td>
<td>9.27</td>
<td>6.19</td>
<td>-</td>
<td>0.45</td>
<td>0.56</td>
<td>0.363</td>
<td>311, 313</td>
</tr>
<tr>
<td>biology coauthorship</td>
<td>undirected</td>
<td>1520251</td>
<td>11803064</td>
<td>15.53</td>
<td>4.92</td>
<td>-</td>
<td>0.088</td>
<td>0.60</td>
<td>0.127</td>
<td>311, 313</td>
</tr>
<tr>
<td>telephone call graph</td>
<td>undirected</td>
<td>47000000</td>
<td>80000000</td>
<td>3.16</td>
<td>2.1</td>
<td>-</td>
<td>0.005</td>
<td>0.001</td>
<td>-0.029</td>
<td>45</td>
</tr>
<tr>
<td>email messages</td>
<td>directed</td>
<td>59912</td>
<td>86300</td>
<td>1.44</td>
<td>4.95</td>
<td>1.5/2.0</td>
<td>0.16</td>
<td>-</td>
<td>0.106</td>
<td>136</td>
</tr>
<tr>
<td>email address books</td>
<td>directed</td>
<td>16881</td>
<td>57029</td>
<td>3.38</td>
<td>5.22</td>
<td>-</td>
<td>0.17</td>
<td>0.13</td>
<td>0.092</td>
<td>321</td>
</tr>
<tr>
<td>student relationships</td>
<td>undirected</td>
<td>573</td>
<td>477</td>
<td>1.66</td>
<td>16.01</td>
<td>-</td>
<td>0.005</td>
<td>0.001</td>
<td>-0.029</td>
<td>45</td>
</tr>
<tr>
<td>sexual contacts</td>
<td>undirected</td>
<td>2840</td>
<td>32</td>
<td>3.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WWW ns.edu</td>
<td>directed</td>
<td>260504</td>
<td>1497135</td>
<td>5.55</td>
<td>11.27</td>
<td>2.1/2.4</td>
<td>0.11</td>
<td>0.29</td>
<td>-0.067</td>
<td>14, 34</td>
</tr>
<tr>
<td>WWW Altavista</td>
<td>directed</td>
<td>28354904</td>
<td>213000000</td>
<td>10.46</td>
<td>16.18</td>
<td>2.1/2.7</td>
<td></td>
<td></td>
<td></td>
<td>74</td>
</tr>
<tr>
<td>citation network</td>
<td>directed</td>
<td>782339</td>
<td>6716198</td>
<td>8.57</td>
<td></td>
<td>3.0/-</td>
<td></td>
<td></td>
<td></td>
<td>351</td>
</tr>
<tr>
<td>Roget’s Thesaurus</td>
<td>directed</td>
<td>1022</td>
<td>5103</td>
<td>4.99</td>
<td>4.87</td>
<td>-</td>
<td>0.13</td>
<td>0.15</td>
<td>0.157</td>
<td>244</td>
</tr>
<tr>
<td>word co-occurrence</td>
<td>undirected</td>
<td>400902</td>
<td>17000000</td>
<td>70.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>119, 157</td>
</tr>
<tr>
<td>Internet</td>
<td>undirected</td>
<td>10697</td>
<td>31992</td>
<td>5.98</td>
<td>3.31</td>
<td>2.5</td>
<td>0.035</td>
<td>0.39</td>
<td>-0.189</td>
<td>86, 148</td>
</tr>
<tr>
<td>power grid</td>
<td>undirected</td>
<td>4941</td>
<td>6594</td>
<td>2.67</td>
<td>18.99</td>
<td>-</td>
<td>0.10</td>
<td>0.080</td>
<td>-0.003</td>
<td>416</td>
</tr>
<tr>
<td>train routes</td>
<td>undirected</td>
<td>587</td>
<td>19603</td>
<td>66.79</td>
<td>2.16</td>
<td>-</td>
<td>0.09</td>
<td>0.69</td>
<td>-0.033</td>
<td>306</td>
</tr>
<tr>
<td>software packages</td>
<td>directed</td>
<td>1439</td>
<td>1723</td>
<td>1.20</td>
<td>2.42</td>
<td>1.6/1.4</td>
<td>0.070</td>
<td>0.082</td>
<td>-0.016</td>
<td>318</td>
</tr>
<tr>
<td>software classes</td>
<td>directed</td>
<td>1377</td>
<td>2213</td>
<td>1.61</td>
<td>1.51</td>
<td>-</td>
<td>0.033</td>
<td>0.012</td>
<td>-0.119</td>
<td>395</td>
</tr>
<tr>
<td>electronic circuits</td>
<td>undirected</td>
<td>24097</td>
<td>53248</td>
<td>4.34</td>
<td>11.05</td>
<td>3.0</td>
<td>0.010</td>
<td>0.030</td>
<td>-0.154</td>
<td>155</td>
</tr>
<tr>
<td>peer-to-peer network</td>
<td>undirected</td>
<td>880</td>
<td>1296</td>
<td>1.47</td>
<td>4.28</td>
<td>2.1</td>
<td>0.012</td>
<td>0.011</td>
<td>-0.366</td>
<td>6, 354</td>
</tr>
<tr>
<td>metabolic network</td>
<td>undirected</td>
<td>765</td>
<td>3686</td>
<td>9.64</td>
<td>2.56</td>
<td>2.2</td>
<td>0.090</td>
<td>0.67</td>
<td>-0.240</td>
<td>214</td>
</tr>
<tr>
<td>protein interactions</td>
<td>undirected</td>
<td>2115</td>
<td>2240</td>
<td>2.12</td>
<td>6.80</td>
<td>2.4</td>
<td>0.072</td>
<td>0.071</td>
<td>-0.156</td>
<td>212</td>
</tr>
<tr>
<td>marine food web</td>
<td>directed</td>
<td>135</td>
<td>508</td>
<td>4.43</td>
<td>2.05</td>
<td>-</td>
<td>0.16</td>
<td>0.23</td>
<td>-0.263</td>
<td>204</td>
</tr>
<tr>
<td>freshwater food web</td>
<td>directed</td>
<td>92</td>
<td>997</td>
<td>10.84</td>
<td>1.90</td>
<td>-</td>
<td>0.20</td>
<td>0.087</td>
<td>-0.326</td>
<td>272</td>
</tr>
<tr>
<td>neural network</td>
<td>directed</td>
<td>307</td>
<td>2359</td>
<td>7.68</td>
<td>3.97</td>
<td>-</td>
<td>0.18</td>
<td>0.28</td>
<td>-0.226</td>
<td>416, 421</td>
</tr>
</tbody>
</table>

**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 $\ell$; exponent $\alpha$ of degree distribution if the distribution follows a power law (or $\ell^{-\alpha}$ 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 4.5. 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.

Limited maximum degree because of e.g. the finite set of nodes in a network.
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 [381]; (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.
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 1.1

- Released today
- [http://www.kmi.tugraz.at/staff/markus/courses/SS2013/707.000_web-science/](http://www.kmi.tugraz.at/staff/markus/courses/SS2013/707.000_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!