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
Overview

Agenda

• Overview of the Google Online Marketing Challenge
• A *selection* of relevant concepts from Network Theory
The Google Marketing Challenge
Objective

Develop an effective online marketing campaign with Google AdWords for a business of your choice
Basic Data

Groups of 3-4 students

200 US-Dollar available for an AdWords Campaign
  – This campaign runs for 3 consecutive weeks

Two documents required
  – Pre-Campaign Document (4 pages)
  – Post-Campaign Document (8 pages)
What do you get?

Introduction to the AdWords System
Insight into online marketing
If your team performs best:
  – Local winners get a trip to the local Google office
  – Global winners get a trip to the Google HQs in Mountain View, CA
Which business should you choose?

Small to medium business
The selected business must agree to a provided letter
Should take active interest in the campaign
Web Site is required!
AdWords was not used for the past 6 months
What kind of businesses should be avoided?

Web Hosting
Web Design Agencies
Insurance Companies
Mortgage Agencies
Debt Consolidation Companies
Multi-level Marketers
Distributors
Affiliate Companies
# Timeline

<table>
<thead>
<tr>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Campaign Document</td>
<td>Until May 1\textsuperscript{st} 2009</td>
</tr>
<tr>
<td>Campaign</td>
<td>Starting on May 1\textsuperscript{st} for 3 consecutive weeks</td>
</tr>
<tr>
<td>Post-Campaign Document</td>
<td>After campaign, before June 12 2009</td>
</tr>
</tbody>
</table>
Want to participate?

Nominate team captain for your team

Team captain sends an email to
christian.koerner@tugraz.at

- Subject: [GMC] – “Your_team_name”
- Includes all names and matriculation numbers of the team members
- DEADLINE for registration: March 31st 2009 - 23:59
  - Everyone who is not registered for GMC until this date participates in the “KMI Yahoo Boss Search Challenge”
Questions?
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 »

user notes

Why and How to Flash Your BIOS
rlaw77
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.
edwined

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

posting history
» first posted by farbiarz to system:unfiled

by JillISw3d3 to hardware
by gtss to bios boost post speed software flash pc
by sgill292 to bios flash computer hardware
by Curioso44 to boot
by catfish162 to system:unfiled
by anurag_bhd to howto flash bios
Social Networks Entities Simplified

Xing:
- Person
- 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, Zengestrom\(^1\) 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.

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

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?
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.
Terminology II

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

- Network size: total number of vertices (denoted N)
- Maximum number of edges (undirected): \( N(N-1)/2 \sim N^2/2 \)
- 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.
Terminology III

http://www.infosci.cornell.edu/courses/info204/2007sp/  
[Diestel 2005]

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

\begin{figure}[h]
  \centering
  \includegraphics[width=\textwidth]{figure.png}
  \caption{A path $P = P^6$ in $G$}
  \end{figure}
Other types of networks
[Newman 2003]

Undirected, single edge and node type

Undirected, varying edge and node weights

Directed, each edge has a direction

Undirected, multiple edge and node types

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: $k_{\text{in}}$ 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, $p(k)$, 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 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

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

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, \textit{roughly} 10 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>449,913</td>
<td>25,516,482</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>7,673</td>
<td>53,302</td>
<td>14.44</td>
<td>4.60</td>
<td>–</td>
<td>0.50</td>
<td>0.88</td>
<td>0.276</td>
<td>105, 323</td>
</tr>
<tr>
<td>math coauthorship</td>
<td>undirected</td>
<td>253,339</td>
<td>496,489</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>52,000</td>
<td>245,300</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>1,520,251</td>
<td>11,803,064</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>47,000,000</td>
<td>80,000,000</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>59,912</td>
<td>86,300</td>
<td>1.44</td>
<td>4.95</td>
<td>1.5/2.0</td>
<td>0.16</td>
<td>0.13</td>
<td>0.092</td>
<td>321</td>
</tr>
<tr>
<td>email address books</td>
<td>directed</td>
<td>16,881</td>
<td>57,029</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>2,810</td>
<td></td>
<td>3.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>265, 266</td>
</tr>
<tr>
<td>WWW nd.edu</td>
<td>directed</td>
<td>269,504</td>
<td>1,497,135</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>203,549,046</td>
<td>2,130,000,000</td>
<td>10.46</td>
<td>16.18</td>
<td>2.1/2.7</td>
<td>0.36</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>citation network</td>
<td>directed</td>
<td>783,339</td>
<td>6,716,188</td>
<td>8.57</td>
<td>3.0/-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>351</td>
</tr>
<tr>
<td>Roget’s Thesaurus</td>
<td>directed</td>
<td>1,022</td>
<td>5,103</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>460,902</td>
<td>17,000,000</td>
<td>70.13</td>
<td>2.7</td>
<td></td>
<td>0.44</td>
<td></td>
<td></td>
<td>119, 157</td>
</tr>
<tr>
<td>Internet</td>
<td>undirected</td>
<td>10,697</td>
<td>31,992</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>4,944</td>
<td>6,594</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>19,603</td>
<td>66.79</td>
<td>2.16</td>
<td>–</td>
<td>0.69</td>
<td>-0.033</td>
<td>366</td>
<td></td>
</tr>
<tr>
<td>software packages</td>
<td>directed</td>
<td>1,439</td>
<td>1,723</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>1,377</td>
<td>2,213</td>
<td>1.61</td>
<td>1.54</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>24,997</td>
<td>53,248</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>1,206</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>3,866</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>2,115</td>
<td>2,240</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>598</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>2,359</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 “-” 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
[Norman 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 *Saccharomyces* [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.

US highway network
US airline network

Bell Curve Distribution of Node Linkages
Power Law Distribution of Node Linkages
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 nodes. 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]
Home Assignment 1.2

• Released today

• In case of any questions, do not hesitate to post to the newsgroup tu-graz.lv.web-science
Any questions?

See you next week!