Properties of real networks: degree distribution

Nodes with small degrees are most frequent. The fraction of highly connected nodes decreases, but is not zero. Look closer: use a logarithmic plot.
power law $P(k) \sim k^{-\gamma}$

exponential $P(k) \sim \exp(-k/\kappa)$
Logarithmic binning

Finding the exponent of a power law may be difficult

Logarithmic binning: average $P(k)$ values within bins of exponentially increasing size

$$\Delta \kappa^i = \kappa^i - \kappa^{i-1}$$

OR: plot the cumulative degree distribution

$$P(k > \kappa) = 1 - \sum_{k=0}^{\kappa} P(k)$$
\[ \log(P(X > x)) \]

Probability that node has degree greater than \( x \).

\[ \log(P(x)) \]

Probability that a node has degree \( x \).

\[ \approx C x^{-\alpha + 1} \]

\[ \approx C x^{-\alpha} \]

Probability that a node has a degree bigger than \( x \).
The in- and out-degree distribution of the WWW are power-laws

\[ P_{\text{out}}(k) \approx k^{-2.45} \]
\[ P_{\text{in}}(k) \approx k^{-2.1} \]

nodes: webpages
edges: hyperlinks

Power-law degree distributions were found in diverse networks

\[ P(k) \approx k^{-2.4} \]

\[ P(k) \approx (31 + k)^{-3} \]

Networks of science collaborations also have power-law degree distributions

\[ P(k) \approx k^{-1.2} \]

\[ P(k) \approx k^{-2.1} \]


A.-L. Barabási et al., cond-mat/0104162 (2001)
Metabolic networks have a power-law degree distribution

$P_{in}(k) \approx k^{-2.2}$

$P_{out}(k) \approx k^{-2.2}$

C. elegans

Archaeoglobus f.

E. coli

bipartite

nodes: metabolites, reactions

directed edges,

out: reactant (substrate)
in: product of reaction

Broad degree distributions in semantic networks


Power grid has exponential degree distribution

$P(k > K) \propto \exp(0.5K)$

nodes: generators, power stations
edges: power lines

Path length and order in real networks

\[ l \approx \frac{\log N}{\log \langle k \rangle} \]

\[ C \propto \langle k \rangle \]

Apparent scaling with the network size and average degree - as though these different networks were members of the same family.
Distribution of betweenness centrality

$P_B(g) \approx g^{-2.2}$

Coauthorship
Protein interaction
Metabolic netw.

World-wide Web
Internet (AS level)

$P_B(g) \approx g^{-2}$

Betweenness centrality (load) distribution of the power grid

\[ P(l > L) \approx (2500 + L)^{-0.7} \]

<table>
<thead>
<tr>
<th>Network</th>
<th>Nodes</th>
<th>Edges</th>
<th>$N_{real}$</th>
<th>$N_{rand} \pm SD$</th>
<th>$Z$ score</th>
<th>$N_{real}$</th>
<th>$N_{rand} \pm SD$</th>
<th>$Z$ score</th>
<th>$N_{real}$</th>
<th>$N_{rand} \pm SD$</th>
<th>$Z$ score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene regulation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(transcription)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli</td>
<td>424</td>
<td>519</td>
<td>40</td>
<td>7 ± 3</td>
<td>10</td>
<td>203</td>
<td>47 ± 12</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. cerevisiae*</td>
<td>685</td>
<td>1,052</td>
<td>70</td>
<td>11 ± 4</td>
<td>14</td>
<td>18,12</td>
<td>300 ± 40</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Neurons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli</td>
<td>252</td>
<td>509</td>
<td>125</td>
<td>90 ± 10</td>
<td>3.7</td>
<td>127</td>
<td>55 ± 13</td>
<td>5.3</td>
<td>227</td>
<td>35 ± 10</td>
<td>20</td>
</tr>
<tr>
<td><strong>Food webs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. cerevisiae*</td>
<td>92</td>
<td>984</td>
<td>3219</td>
<td>3120 ± 50</td>
<td>2.1</td>
<td>7295</td>
<td>2220 ± 210</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Electronic circuits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(forward logic chips)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s15850</td>
<td>10,383</td>
<td>14,240</td>
<td>424</td>
<td>2 ± 2</td>
<td>285</td>
<td>1040</td>
<td>1 ± 1</td>
<td>1200</td>
<td>480</td>
<td>2 ± 1</td>
<td>335</td>
</tr>
<tr>
<td><strong>Electronic circuits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(digital fractional multipliers)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s208</td>
<td>122</td>
<td>189</td>
<td>10</td>
<td>1 ± 1</td>
<td>9</td>
<td>4</td>
<td>1 ± 1</td>
<td>3.8</td>
<td>5</td>
<td>1 ± 1</td>
<td>5</td>
</tr>
<tr>
<td>s420</td>
<td>252</td>
<td>399</td>
<td>20</td>
<td>1 ± 1</td>
<td>18</td>
<td>10</td>
<td>1 ± 1</td>
<td>10</td>
<td>11</td>
<td>1 ± 1</td>
<td>11</td>
</tr>
<tr>
<td>s838‡</td>
<td>512</td>
<td>819</td>
<td>40</td>
<td>1 ± 1</td>
<td>63</td>
<td>42</td>
<td>1 ± 1</td>
<td>20</td>
<td>23</td>
<td>1 ± 1</td>
<td>25</td>
</tr>
<tr>
<td><strong>World Wide Web</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nd.edu§</td>
<td>325,729</td>
<td>1.46e6</td>
<td>1.1e5</td>
<td>8e3 ± 1e2</td>
<td>800</td>
<td>6.8e6</td>
<td>5e4 ± 4e2</td>
<td>15,000</td>
<td>1.2e6</td>
<td>1e4 ± 2e2</td>
<td>5000</td>
</tr>
</tbody>
</table>
Mixing patterns in networks

Mixing in social networks
- assortative: people prefer to associate with others who are like them
- disassortative: people prefer to associate with others who are different

In general mixing is defined in terms of node characteristics (age, race) or classification.
Mixing with respect of node degree:
- assortative: high degree nodes tend to be connected to high degree nodes
- disassortative: high degree nodes tend to be connected to low degree nodes

Focus on edge $i$, denote the excess in-degree of its starting point with $j_i$ and the excess out-degree of its endpoint with $k_i$
Mixing is quantified by the correlation between $j_i$ and $k_i$ over all $i$

Positive correlation - assortative,
Negative correlation - disassortative
Assortativity coefficient

Focus on edge $i$, denote the excess in-degree of its starting point with $j_i$ and the excess out-degree of its endpoint with $k_i$. Mixing is quantified by the assortativity coefficient characterizing the correlation between $j_i$ and $k_i$ over all $i$.

Similar to a Pearson correlation

$$ r = \frac{\sum_i j_i k_i - \sum_i j_i \sum_i k_i}{\sqrt{\left(\sum_i j_i^2 - \left(\sum_i j_i\right)^2/N\right)^{0.5} \left(\sum_i k_i^2 - \left(\sum_i k_i\right)^2/N\right)^{0.5}}/N} $$

$-1 \leq r \leq 1$

Positive $r$ means assortativity, $r=0$ means neutral, negative $r$ means disassortativity.
Mixing in real networks

<table>
<thead>
<tr>
<th>Network</th>
<th>Type</th>
<th>Size $n$</th>
<th>Assortativity $r$</th>
<th>Error $\sigma_r$</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics coauthorship</td>
<td>undirected</td>
<td>52909</td>
<td>0.363</td>
<td>0.002</td>
<td>a</td>
</tr>
<tr>
<td>Biology coauthorship</td>
<td>undirected</td>
<td>1520251</td>
<td>0.127</td>
<td>0.0004</td>
<td>a</td>
</tr>
<tr>
<td>Mathematics coauthorship</td>
<td>undirected</td>
<td>253339</td>
<td>0.120</td>
<td>0.002</td>
<td>b</td>
</tr>
<tr>
<td>Film actor collaborations</td>
<td>undirected</td>
<td>449913</td>
<td>0.208</td>
<td>0.0002</td>
<td>c</td>
</tr>
<tr>
<td>Company directors</td>
<td>undirected</td>
<td>7673</td>
<td>0.276</td>
<td>0.004</td>
<td>d</td>
</tr>
<tr>
<td>Student relationships</td>
<td>undirected</td>
<td>573</td>
<td>-0.029</td>
<td>0.037</td>
<td>e</td>
</tr>
<tr>
<td>Email address books</td>
<td>directed</td>
<td>16881</td>
<td>0.092</td>
<td>0.004</td>
<td>f</td>
</tr>
<tr>
<td>Power grid</td>
<td>undirected</td>
<td>4941</td>
<td>-0.003</td>
<td>0.013</td>
<td>g</td>
</tr>
<tr>
<td>Internet</td>
<td>undirected</td>
<td>10697</td>
<td>-0.189</td>
<td>0.002</td>
<td>h</td>
</tr>
<tr>
<td>World-Wide Web</td>
<td>directed</td>
<td>269504</td>
<td>-0.067</td>
<td>0.0002</td>
<td>i</td>
</tr>
<tr>
<td>Software dependencies</td>
<td>directed</td>
<td>3162</td>
<td>-0.016</td>
<td>0.020</td>
<td>j</td>
</tr>
<tr>
<td>Protein interactions</td>
<td>undirected</td>
<td>2115</td>
<td>-0.156</td>
<td>0.010</td>
<td>k</td>
</tr>
<tr>
<td>Metabolic network</td>
<td>undirected</td>
<td>765</td>
<td>-0.240</td>
<td>0.007</td>
<td>l</td>
</tr>
<tr>
<td>Neural network</td>
<td>directed</td>
<td>307</td>
<td>-0.226</td>
<td>0.016</td>
<td>m</td>
</tr>
<tr>
<td>Marine food web</td>
<td>directed</td>
<td>134</td>
<td>-0.263</td>
<td>0.037</td>
<td>n</td>
</tr>
<tr>
<td>Freshwater food web</td>
<td>directed</td>
<td>92</td>
<td>-0.326</td>
<td>0.031</td>
<td>o</td>
</tr>
</tbody>
</table>

Social networks tend to be assortative, technological and biological networks tend to be disassortative.
Mechanisms that influence mixing properties

- In social relationships there is a documented attraction between individuals of similar temperament
- Group affiliations can create assortativity
- On the WWW or Internet service relationships (directories, connectivity providers) can create disassortativity
- Constraints on the network assembly process (e.g. no multiple edges among pairs of nodes) can cause disassortativity

Universality in large-scale networks

The degree distribution is a decreasing function, usually a power-law. The betweenness centrality distribution is a power law as well. Both indicate heterogeneity and the existence of hubs.

The distances scale logarithmically with the network size

\[ l \approx \frac{\log N}{\log \langle k \rangle} \]

The clustering coefficient does not seem to depend on the network size

\[ C \propto \langle k \rangle \]

Frequent subgraphs – not universal but common to several networks.