Properties common to many large-scale networks, independently of their origin and function:

1. The degree and betweenness distribution are decreasing functions, usually power-laws.
2. The distances scale logarithmically with the network size

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

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

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

As though all these networks were part of the same family/class.
Random networks

The average distance and clustering coefficient only depend on the number of nodes and edges in the network.

This suggests that general models based only on the number of nodes and edges in the network could be successful in describing the properties of an “expected” (characteristic) network.

Uniformly random network: distributes the edges uniformly among nodes.

Probabilistic interpretation:
There exists a set (ensemble) of networks with given number of nodes and edges. Select a random member of this set

What are the expected properties of this network? – studied by random graph theory.
Ex. 1
Start with 10 isolated nodes. For each pair of nodes, throw with a dice, and connect them if the number on the dice is 1. Describe the graph you obtained. How many edges are in the graph? Is it connected or not? What is the average degree and the degree distribution?

Ex. 2
Now connect node pairs if the number on the dice is 1 or 2. How is the graph different from the previous case?

Ex. 2
How many edges do you expect a graph with N nodes would have if each edge is selected with throwing with a dice?
Random graph theory


- fixed node number $N$
- connecting pairs of nodes with probability $p$

Expected number of edges: $E = p \frac{N(N-1)}{2}$

Random graph theory studies the expected properties of graphs with $N \rightarrow \infty$
The properties of random graphs depend on $p$

Properties studied:
- Is the graph connected?
- Does the graph contain a giant connected component?
- What is the diameter of the graph?
- Does the graph contain cliques (complete subgraphs)?

Probabilistic formulation: what is the probability that a graph with $N$ nodes and connection probability $p$ is connected?
Some of these properties appear suddenly, at a threshold $p_c(N)$

$$\lim_{N \to \infty} P_{N,p}(Q) = \begin{cases} 0 & \text{if } \frac{p(N)}{p_c(N)} \to 0 \\ 1 & \text{if } \frac{p(N)}{p_c(N)} \to \infty \end{cases}$$
Subgraphs of a random graph

Consider a subgraph with \( n \) nodes and \( e \) edges.

Expected number of of these subgraphs in a graph with \( N \) nodes and connection prob. \( p \)

\[
E(X) = C_N^n p^e \frac{n!}{a}
\]

Ways of selecting \( n \) nodes from \( N \)

Probability of having \( e \) edges

We can permute the n edges in any way we want...

but identical (isomorphic) graphs do not count

Isomorphic graphs: there exists a one-to-one mapping of the nodes in such a way that if (and only if) node \( i \) and \( j \) are connected in one then their images \( i' \) and \( j' \) are also connected
Special subgraphs

Consider a subgraph with $n$ nodes and $e$ edges.

Expected number of subgraphs with $n$ nodes and $e$ edges in a graph with $N$ nodes and connection prob. $p$

$$E(X) = C_N^n p^e \frac{n!}{a} \cong \frac{N^n p^e}{a}$$

If the connection probability is a function of the number of the nodes, we can find the condition of having a non-vanishing number of subgraphs.

$$\lim_{N \to \infty} p(N) N^{n/e} \neq 0$$

Ex. Find the condition of having a non-vanishing number of trees, cycles and completely connected subgraphs
Evolution of a random graph

Assume that the connection probability is a power-law of $N$, $p = cN^z$

Assume that $z$ increases from $-\infty$ to 0

Look for trees, cycles (circuits) and cliques in the graph.

Appearance thresholds:

$$z_c N^p = \infty$$

The graph contains cycles of any length if $z \geq -1$
Clusters in a random graph

• For $p < N^{-1}$ the graph contains only isolated trees.
• If $p = cN^{-1}$ with $c < 1$ the graph has isolated trees and cycles.
• At $p = cN^{-1}$ with $c = 1$ a giant cluster appears.
• The size of the giant cluster approaches $N$ rapidly as $c$ increases.

$$S = (f(1) - f(c))N$$

• The graph becomes connected if

$$\lim_{N \to \infty} \frac{p}{\ln N / N} = \infty$$
Node degrees in random graphs

- average degree: $\langle k \rangle = \frac{2E}{N} \approx pN$
- degree distribution:

$$P(k) \approx C_{N-1}^k p^k (1-p)^{N-1-k}$$

Most of the nodes have approximately the same degree. The probability of very highly connected nodes is exponentially small.
Distances in random graphs

Random graphs tend to have a tree-like topology with almost constant node degrees.

- nr. of first neighbors: $N_1 \approx \langle k \rangle$

- nr. of second neighbors: $N_2 \approx \langle k \rangle^2$

- estimate maximum distance:

\[
1 + \sum_{l=1}^{l_{\max}} \langle k \rangle^i = N \quad \Rightarrow \quad l_{\max} = \frac{\log N}{\log \langle k \rangle}
\]

This scaling was proven by Chung and Lu, Adv. Appl. Math 26, 257 (2001).
There is no local order in random graphs

Measure of local order: \[ C_i \equiv \frac{n_i}{k_i(k_i-1)/2} \]

Since edges are independent and have the same probability \( p \),

\[ n_i \approx p \frac{k_i(k_i-1)}{2} \quad \Rightarrow C \approx p = \frac{\langle k \rangle}{N} \]

The clustering coefficient of random graphs is small.
Are real networks like random graphs?

As quantitative data about real networks becomes available, we can compare their topology with that of random graphs.

Starting measures: $N$, $<k>$ for the real network.

Determine $l$, $C$ and $P(k)$ for a random graph with the same $N$ and $<k>$.

\[ l_{\text{rand}} \approx \frac{\log N}{\log \langle k \rangle} \quad C_{\text{rand}} = p = \frac{\langle k \rangle}{N} \]

\[ P_{\text{rand}}(k) \approx \binom{N}{N-k} p^k (1-p)^{N-1-k} \]

Measure $l$, $C$ and $P(k)$ for the real network. Compare.
Path length and order in real networks

\[ l_{\text{rand}} = \frac{\log N}{\log \langle k \rangle} \]

\[ C_{\text{rand}} = \frac{\langle k \rangle}{N} \]

Real networks have short distances like random graphs yet show signs of local order.
The degree distribution of the WWW is a power-law.

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


Power-law degree distributions were found in diverse networks

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

\[ P(k) \approx k^{-2.3} \quad (b) \]

The power-law degree distribution indicates a heterogeneous topology.

The average degree gives the characteristic scale (value) of the degree.

Large variability, the average degree not informative, no characteristic scale for the degree.

Scale-free
Idea: generate random graphs with a power-law degree distribution

Fixed $N$, $P(k) = Ak^{-\gamma}$, $k < k$

Network assembly - random edges, but enforcing the right $P(k)$

$$\sum_{k=1}^{K} P(k) = 1, \quad \Rightarrow \quad A = \frac{1}{\sum_{k=1}^{K} k^{-\gamma}}$$

$$< k > = \sum_{k=1}^{K} kP(k), \quad \Rightarrow \quad < k > = \frac{\sum_{k=1}^{K} k^{-\gamma+1}}{\sum_{k=1}^{K} k^{-\gamma}}$$

The number of edges increases as $\gamma$ decreases.
Constructing graphs with a given degree distribution

Configuration model:
• choose a degree sequence $N(k)=N P(k)$
• give the nodes $k$ “stubs” according to $N(k)$
• connect stubs randomly

M. E. J. Newman, S. H. Strogatz, and D. J. Watts,
Phys. Rev. E 64, 026118 (2001)

Ex. Construct a graph with 10 nodes and degree sequence $N(1)=4$, $N(2)=3$, $N(3)=2$, $N(4)=1$.
What is a necessary condition for the graph construction?
Theory of general random graphs

Looks at a characteristic member of the ensemble of graphs with given degree distribution.

Seeks the answers to the same questions as random graph theory
- is the graph connected?
- does the graph contain a giant component?
- what is the diameter of the graph?
- what is the clustering coefficient of the graph?

The theoretical concept needed for the analysis is the generating function

H. Wilf, Generatingfunctionology (1994)
Generating functions in a graph

Node degree generating function

\[ G_0(x) = \sum_{k=0}^{\infty} P(k)x^k, \quad |x| \leq 1 \]

Finding \( P(k) \) and degree moments from the generating function

\[ P(k) = \frac{1}{k!} \left. \frac{d^k G_0}{dx^k} \right|_{x=0} \quad \langle k^n \rangle = \sum_k k^n P(k) = \left[ \left( x \frac{d}{dx} \right)^n G_0(x) \right]_{x=1} \]

If a certain property is described by a gen. function, then its sum over \( m \) independent realizations is generated by the \( m \)th power of the gen. function

The generating function for the sum of the degrees of two nodes is \([G_0(x)]^2\)
Generating functions in a graph

Generating function for the degree of nodes at the end of a random edge

\[ G_1(x) = \frac{\sum_k kP(k)x^{k-1}}{\sum_k kP(k)} = \frac{G'_0(x)}{G'_0(1)} \]

discount the edge we arrived along

probability to find that node

normalization

Finding connected components in a random graph

Breadth-first-search algorithm:
• start from an arbitrary node,
• follow its (out) edges, record its first neighbors
• follow the (out) edges of first neighbors, record any new node found
• continue until no new nodes are found

The same step, “follow the edge” is repeated over and over.
Generating functions for connected components

\( P_s \) - distribution of component sizes, \( H_0(x) = \sum_{s=0}^{\infty} P_s x^s, \quad |x| \leq 1 \)

Generating function for the cluster we reach by following a randomly chosen edge

\[
\square = \bigcirc + \bigcirc + \bigcirc + \bigcirc + \ldots
\]

\( \sum kP(k)[H_1(x)]^{k-1} \)

\[
H_1(x) = \frac{k}{\sum kP(k)} = xG_1(H_1(x))
\]

Probability to find that node

Repeat the same step for each edge

Normalization

\[
H_0(x) = xG_0(H_1(x))
\]
Existence of a giant connected cluster

Average size of marked clusters

$$\langle s \rangle = H'_0(1) = 1 + \frac{G'_0(1)}{1 - G'_1(1)}$$

Diverges when

$$G'_1(1) \equiv \frac{\sum_k kP(k)}{\sum_k k(k-1)P(k)} = 1$$

Equivalent to

$$\frac{\langle k^2 \rangle}{\langle k \rangle} = 2$$

A giant connected component exists if the graph is sufficiently heterogeneous.
Theory of scale-free random graphs

The theory of scale-free random graphs is in many ways similar to that of Erdös-Rényi random graphs.

Graph properties depend on $\gamma$

- giant cluster: $\gamma \leq 3.47$
- connected: $\gamma \leq 2$


Average path length of scale-free random graphs

Network: \( N, \quad P(k) \approx k^{-\gamma} \quad \text{for} \quad k \leq \kappa \)

Prediction: \( l_{sf} = \frac{\ln N + B}{A} + 1 \quad A, B = f(\gamma, \kappa) \)


- qualitative agreement
- worse than a random graph
Clustering coefficient of scale-free random graphs

\[
C = \frac{\langle k \rangle}{N} \left( \frac{\langle k^2 \rangle - \langle k \rangle^2}{\langle k \rangle^2} \right)^2
\]

The second term depends on the variance of the degree distribution.

\[
P(k) \approx k^{-\gamma} \quad \quad \quad C \approx N^{-(3\gamma-7)/(\gamma-1)}
\]

For \( \gamma < 7/3 \) \( C \) increases with \( N \).

<table>
<thead>
<tr>
<th>Network</th>
<th>Size</th>
<th>$\langle k \rangle$</th>
<th>$\kappa$</th>
<th>$\gamma_{out}$</th>
<th>$\gamma_{in}$</th>
<th>$\ell_{real}$</th>
<th>$\ell_{rand}$</th>
<th>$\ell_{pow}$</th>
<th>Reference</th>
<th>Nr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>WWW</td>
<td>325 729</td>
<td>4.51</td>
<td>900</td>
<td>2.45</td>
<td>2.1</td>
<td>11.2</td>
<td>8.32</td>
<td>4.77</td>
<td>Albert, Jeong, and Barabási 1999</td>
<td>1</td>
</tr>
<tr>
<td>WWW</td>
<td>$4 \times 10^7$</td>
<td>7</td>
<td>2.38</td>
<td>2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kumar et al., 1999</td>
<td>2</td>
</tr>
<tr>
<td>WWW</td>
<td>$2 \times 10^8$</td>
<td>7.5</td>
<td>4000</td>
<td>2.72</td>
<td>2.1</td>
<td>16</td>
<td>8.85</td>
<td>7.61</td>
<td>Broder et al., 2000</td>
<td>3</td>
</tr>
<tr>
<td>WWW, site</td>
<td>260 000</td>
<td>1.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Huberman and Adamic, 2000</td>
<td>4</td>
</tr>
<tr>
<td>Internet, domain*</td>
<td>3015–4389</td>
<td>3.42–3.76</td>
<td>30–40</td>
<td>2.1–2.2</td>
<td>2.1–2.2</td>
<td>4</td>
<td>6.3</td>
<td>5.2</td>
<td>Faloutsos, 1999</td>
<td>5</td>
</tr>
<tr>
<td>Internet, router*</td>
<td>3888</td>
<td>2.57</td>
<td>30</td>
<td>2.48</td>
<td>2.48</td>
<td>12.15</td>
<td>8.75</td>
<td>7.67</td>
<td>Faloutsos, 1999</td>
<td>6</td>
</tr>
<tr>
<td>Internet, router*</td>
<td>150 000</td>
<td>2.66</td>
<td>60</td>
<td>2.4</td>
<td>2.4</td>
<td>11</td>
<td>12.8</td>
<td>7.47</td>
<td>Govindan, 2000</td>
<td>7</td>
</tr>
<tr>
<td>Movie actors*</td>
<td>212 250</td>
<td>28.78</td>
<td>900</td>
<td>2.3</td>
<td>2.3</td>
<td>4.54</td>
<td>3.65</td>
<td>4.01</td>
<td>Barabási and Albert, 1999</td>
<td>8</td>
</tr>
<tr>
<td>Co-authors, SPIRES*</td>
<td>56 627</td>
<td>173</td>
<td>1100</td>
<td>1.2</td>
<td>1.2</td>
<td>4</td>
<td>2.12</td>
<td>1.95</td>
<td>Newman, 2001b</td>
<td>9</td>
</tr>
<tr>
<td>Co-authors, neuro.*</td>
<td>209 293</td>
<td>11.54</td>
<td>400</td>
<td>2.1</td>
<td>2.1</td>
<td>6</td>
<td>5.01</td>
<td>3.86</td>
<td>Barabási et al., 2001</td>
<td>10</td>
</tr>
<tr>
<td>Co-authors, math.*</td>
<td>70 975</td>
<td>3.9</td>
<td>120</td>
<td>2.5</td>
<td>2.5</td>
<td>9.5</td>
<td>8.2</td>
<td>6.53</td>
<td>Barabási et al., 2001</td>
<td>11</td>
</tr>
<tr>
<td>Sexual contacts*</td>
<td>2810</td>
<td>7.4</td>
<td>110</td>
<td>2.2</td>
<td>2.2</td>
<td>3.2</td>
<td>3.32</td>
<td>2.89</td>
<td>Liljeros et al., 2001</td>
<td>12</td>
</tr>
<tr>
<td>Metabolic, <em>E. coli</em></td>
<td>778</td>
<td>7.4</td>
<td>110</td>
<td>2.2</td>
<td>2.2</td>
<td>3.2</td>
<td>3.32</td>
<td>2.89</td>
<td>Jeong et al., 2000</td>
<td>13</td>
</tr>
<tr>
<td>Protein, <em>S. cerevisiae</em></td>
<td>1870</td>
<td>2.39</td>
<td>24</td>
<td>2.4</td>
<td>2.4</td>
<td></td>
<td></td>
<td></td>
<td>Jeong, Mason, et al., 2001</td>
<td>14</td>
</tr>
<tr>
<td>Ythan estuary*</td>
<td>134</td>
<td>8.7</td>
<td>35</td>
<td>1.05</td>
<td>1.05</td>
<td>2.43</td>
<td>2.26</td>
<td>1.71</td>
<td>Montoya and Solé, 2000</td>
<td>14</td>
</tr>
<tr>
<td>Silwood Park*</td>
<td>154</td>
<td>4.75</td>
<td>27</td>
<td>1.13</td>
<td>1.13</td>
<td>3.4</td>
<td>3.23</td>
<td>2</td>
<td>Montoya and Solé, 2000</td>
<td>16</td>
</tr>
<tr>
<td>Citation</td>
<td>783 339</td>
<td>8.57</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Redner, 1998</td>
<td>17</td>
</tr>
<tr>
<td>Phone call</td>
<td>$53 \times 10^6$</td>
<td>3.16</td>
<td>2.1</td>
<td>2.1</td>
<td>2.1</td>
<td></td>
<td></td>
<td></td>
<td>Aiello et al., 2000</td>
<td>18</td>
</tr>
<tr>
<td>Words, co-occurrence*</td>
<td>460 902</td>
<td>70.13</td>
<td>2.7</td>
<td>2.7</td>
<td>2.7</td>
<td></td>
<td></td>
<td></td>
<td>Ferrer i Cancho and Solé, 2001</td>
<td>19</td>
</tr>
<tr>
<td>Words, synonyms*</td>
<td>22 311</td>
<td>13.48</td>
<td>2.8</td>
<td>2.8</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
<td>Yook et al., 2001b</td>
<td>20</td>
</tr>
</tbody>
</table>

**Expectations:**

$\langle k \rangle \geq 1$ giant connected component, $\langle k \rangle \geq \ln N$ connected

$\gamma \leq 3.47$ giant connected component, $\gamma \leq 2$ connected
Exponential random graphs

“Exponential” does not refer to the degree distribution but to the model construction!
This is a statistical method for generating a graph with $N$ nodes by specifying a distribution function over all graphs with $N$ nodes.

1. Select a set of informative network measures (e.g. number of edges, number of triangles, degree distribution)
2. Then select a network from the ensemble of all graphs using the probability

$$P(G) \sim \exp\left(-\sum \beta_i \varepsilon_i\right)$$

$\beta_i$ – parameters, $\varepsilon_i$ – network measures

3. Estimate the coefficients such that an observed (real) network corresponds to the most likely graph in that ensemble – maximum likelihood estimation

Markov graphs: edges that do not share a node are independent