Frontiers of Network Science
Fall 2021

Class 17: Degree Correlations part II
(Chapter 7 in Textbook)

Boleslaw Szymanski

based on slides by Albert-László Barabási and Roberta Sinatra

www.BarabasiLab.com
(1) Difficult to extract information from a visual inspection of a matrix.

(2) Based on $e_{jk}$ and hence requires a large number of elements to inspect:

$$k_{\text{max}} \frac{(k_{\text{max}} - 1)}{2} - 1 - k_{\text{max}}$$

$\begin{pmatrix} \sum_{j,k} e_{jk} = 1 \\ \sum_{j=1,k_{\text{max}}} e_{jk} = q_k \end{pmatrix}$

Undirected network: $k_{\text{max}} \times k_{\text{max}}$ matrix

Nr. of independent elements

We need to find a way to reduce the information contained in $e_{jk}$

Average next neighbor degree

\( k_{\text{ann}}(k) \): average degree of the first neighbors of nodes with degree \( k \).

\[
k_{\text{ann}}(k) = \sum_{k'} k' P(k' | k) = \frac{\sum k' e_{kk'}}{\sum e_{kk'}}
\]

No degree correlations:

\[
k_{\text{ann}}(k) = \frac{\sum k' e_{kk'}}{\sum e_{kk'}} = \frac{\sum k' q_k q_{k'}}{q_k} = \sum_{k'} k' q_{k'} = \sum_{k'} k' p(k') = \left< k^2 \right> \left< k \right>
\]

If there are no degree correlations, \( k_{\text{ann}}(k) \) is independent of \( k \).

Network Science: Degree Correlations

$k_{anno}(k)$ FOR REAL NETWORKS

Astrophysics co-authorship network

\[ \langle k_{nn} \rangle \]

exponent: 0.37±0.11

Yeast PPI

\[ \langle k_{nn} \rangle \]

exponent: -0.27±0.03

Assortative

Disassortative
$k_{ann}(k)$: average degree of the first neighbors of nodes with degree $k$.

$$\sum_k k_{ann}(k) \cdot k Np_k = \sum_k k^2 \cdot Np_k$$

constraint:

$$\langle k_{ann}(k)k \rangle = \langle k^2 \rangle$$

$k_{ann}(k)$ is a $k$-dependent function, hence it has much fewer parameters,

and it is easier to interpret/read.

PEARSON CORRELATION

If there are degree correlations, $e_{jk}$ will differ from $q_j q_k$. The magnitude of the correlation is captured by $\langle jk \rangle - \langle j \rangle \langle k \rangle$ difference, which is:

$$\sum_{jk} e_{jk} - q_j q_k$$

$\langle jk \rangle - \langle j \rangle \langle k \rangle$ is expected to be:
- positive for assortative networks,
- zero for neutral networks,
- negative for disassortative networks

To compare different networks, we should normalize it with its maximum value; the maximum is reached for a perfectly assortative network, i.e. $e_{jk} = q_j q_k$

normalization: 

$$\sigma_r^2 = \max \sum_{jk} e_{jk} - q_j q_k = \sum_{jk} q_j q_k \delta_{jk} - q_j q_k$$

$$r = \frac{\sum_{jk} e_{jk} - q_j q_k}{\sigma_r^2}$$

$-1 \leq r \leq 1$

$r \leq 0$ disassortative

$r = 0$ neutral

$r \geq 0$ assortative

### Social networks are assortative

<table>
<thead>
<tr>
<th>Network</th>
<th>$n$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics coauthorship (a)</td>
<td>52909</td>
<td>0.363</td>
</tr>
<tr>
<td>Biology coauthorship (a)</td>
<td>1520251</td>
<td>0.127</td>
</tr>
<tr>
<td>Mathematics coauthorship (b)</td>
<td>253339</td>
<td>0.120</td>
</tr>
<tr>
<td>Film actor collaborations (c)</td>
<td>449913</td>
<td>0.208</td>
</tr>
<tr>
<td>Company directors (d)</td>
<td>7673</td>
<td>0.276</td>
</tr>
<tr>
<td>Internet (e)</td>
<td>10697</td>
<td>-0.189</td>
</tr>
<tr>
<td>World-Wide Web (f)</td>
<td>269504</td>
<td>-0.065</td>
</tr>
<tr>
<td>Protein interactions (g)</td>
<td>2115</td>
<td>-0.156</td>
</tr>
<tr>
<td>Neural network (h)</td>
<td>307</td>
<td>-0.163</td>
</tr>
<tr>
<td>Marine food web (i)</td>
<td>134</td>
<td>-0.247</td>
</tr>
<tr>
<td>Freshwater food web (j)</td>
<td>92</td>
<td>-0.276</td>
</tr>
<tr>
<td>Random graph (u)</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Callaway <em>et al.</em> (v)</td>
<td></td>
<td>$\delta/(1 + 2\delta)$</td>
</tr>
<tr>
<td>Barabási and Albert (w)</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

$r > 0$: assortative network:
- Hubs tend to connect to other hubs.

$r < 0$: disassortative network:
- Hubs tend to connect to small nodes.
RELATIONSHIP BETWEEN $r$ AND $k_{\text{annnd}}$

$$r = \frac{\sum_{kj} (e_{kj} - q_k q_j)}{\sigma_r^2} = \frac{\sum_k k q_k \sum_j j e_{kj} - \left( \sum_k k q_k \right)^2}{\sigma_r^2} = \frac{\sum_k k_{\text{annnd}}(k) q_k - \langle k^2 \rangle^2}{\sigma_r^2}$$

$$k_{\text{annnd}}(k) = \sum_{k'} k' P(k' \mid k) = \frac{\sum_k k' e_{kk'}}{\sum_k e_{kk'}} = \frac{\sum k' e_{kk'}}{q_k}$$

In general case we need to know $q_k$ and $k_{\text{annnd}}(k)$ to calculate $r$.

Assuming: $k_{\text{annnd}}(k) = a \cdot k + b$

Using the constraint for ANND:

$$\langle k^2 \rangle = \langle k_{\text{annnd}}(k) k \rangle = \sum_{k'} a \cdot k^2 p_k + b \cdot k p_k = a \langle k^2 \rangle + b \langle k \rangle$$

$$r = \frac{\sum_k (a \cdot k + b) q_k - \langle k^2 \rangle^2}{\sigma_r^2} = \sum_k \left( a \cdot k + \frac{(1-a) \langle k^2 \rangle}{\langle k \rangle} \right) p_k - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2}$$

$$b = \frac{(1-a) \langle k^2 \rangle}{\langle k \rangle}, \quad a = \frac{a \langle k^2 \rangle}{\langle k \rangle}$$
PROBLEM WITH THE PREVIOUS DEVIATION: $k_{\text{ann}}(k) \sim k^\beta$

Exponent: $0.37 \pm 0.11$

Exponent: $-0.27 \pm 0.03$

Astrophysics co-authorship network

Yeast PPI

Assortative

Disassortative
Assuming: \( k_{ann}(k) = a \cdot k^\beta \)

Using the constraint for ANND:

\[
\left\langle k^2 \right\rangle = \left\langle k_{ann}(k)k \right\rangle = \sum_k a \cdot k^{\beta+1} p_k = a \left\langle k^{\beta+1} \right\rangle \quad \Rightarrow \quad a = \frac{\left\langle k^2 \right\rangle}{\left\langle k^{\beta+1} \right\rangle}
\]

\[
r = \frac{\sum_k k \cdot a k^{\beta} \cdot q_k - \left\langle k^2 \right\rangle^2}{\left\langle k^2 \right\rangle} = \frac{\sum_k a \cdot k^{\beta+2} \cdot p_k - \left\langle k^2 \right\rangle^2}{\left\langle k^{\beta+1} \right\rangle} \quad \Rightarrow \quad \frac{\left\langle k^2 \right\rangle}{\left\langle k^{\beta+1} \right\rangle} - \frac{\left\langle k^2 \right\rangle}{\left\langle k \right\rangle} = \frac{\sigma_r^2}{\sigma_r^2}
\]

\[
= \frac{1}{\sigma_r^2} \left( \frac{\left\langle k^{\beta+2} \right\rangle}{\left\langle k^{\beta+1} \right\rangle} - \frac{\left\langle k^2 \right\rangle}{\left\langle k \right\rangle} \right)
\]

\[
\sigma_r^2 = \sum_{jk} jk (q_k \delta_{jk} - q_j q_k) = \frac{\left\langle k^3 \right\rangle}{\left\langle k \right\rangle} - \frac{\left\langle k^2 \right\rangle^2}{\left\langle k \right\rangle^2}
\]

\[
\beta < 0 \quad \Rightarrow \quad r < 0
\]

\[
\beta = 0 \quad \Rightarrow \quad r = 0
\]

\[
\beta > 0 \quad \Rightarrow \quad r > 0
\]
\[ \beta = 0 : \quad \frac{\langle k^{\beta+2} \rangle}{\langle k^{\beta+1} \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle} - \frac{\langle k^2 \rangle}{\langle k \rangle} = 0 \quad \Rightarrow \quad r = 0 \]

\[ \langle k^{\alpha+\beta} \rangle = \sum_{k_{\text{min}}}^{k_{\text{max}}} k^{\alpha+\beta} p_k = k_{\text{max}}^{\beta} \sum_{k_{\text{min}}}^{k_{\text{max}}} k^{\alpha} p_k = k_{\text{min}}^{\beta} \langle k^\alpha \rangle \]

\[ 0 > \beta > -1 : \quad \frac{\langle k^2 \rangle}{\langle k \rangle} > \left( \frac{k_{\text{min}}}{k_{\text{max}}} \right)^{\beta} \langle k^{\beta+2} \rangle \quad > \quad \frac{\langle k^{\beta+2} \rangle}{\langle k^{\beta+1} \rangle} \quad \Rightarrow \quad r < 0 \]

\[ +1 > \beta > 0 : \quad \frac{\langle k^{\beta+2} \rangle}{\langle k^{\beta+1} \rangle} > \left( \frac{k_{\text{min}}}{k_{\text{max}}} \right)^{\beta} \frac{\langle k^2 \rangle}{\langle k \rangle} > \frac{\langle k^2 \rangle}{\langle k \rangle} \quad \Rightarrow \quad r > 0 \]
DEGREE CORRELATION IN NETWORKS

\[ e_{jk} \]

\[ k_{\text{ann}}(k) \]

\[ r \]

\[ \frac{k_{\text{max}}(k_{\text{max}} - 1)}{2} - k_{\text{max}} - 1 \]
We have a desired $e_{jk}$ distribution, which also specifies $p_k$.

1. Generate a network with the desired degree distribution using the configuration model.
2. Choose two links at random from the network: $(v_1,w_1)$ and $(v_2,w_2)$.
3. Measure the degrees $j_1, k_1, j_2, k_2$ of nodes $v_1, w_1, v_2, w_2$. Replace the two selected links with two new ones $(v_1,v_2)$ and $(w_1,w_2)$ with probability

$$P = \begin{cases} 
\frac{e_{j_1j_2}e_{k_1k_2}}{e_{j_1k_1}e_{k_2j_2}} & \text{if } e_{j_1j_2}e_{k_1k_2} < e_{j_1k_1}e_{k_2j_2} \\
1 & \text{otherwise}
\end{cases}$$

1. Repeat from step 2.

The algorithm is ergodic and satisfies detailed balance, therefore in the long time limit it samples the desired network ensemble correctly.

2. Choose two edges random from the network: \((v_1, w_1)\) and \((v_2, w_2)\).
3. Measure the degrees \(j_1, k_1, j_2, k_2\) of vertices \(v_1, w_1, v_2, w_2\). Replace the two selected edges with two new ones \((v_1, v_2)\) and \((w_1, w_2)\) with probability

\[
P = \begin{cases} 
\frac{e_{j_1,j_2}e_{k_1,k_2}}{e_{j_1,k_1}e_{k_2,j_2}} & \text{if } e_{j_1,j_2}e_{k_1,k_2} < e_{j_1,k_1}e_{k_2,j_2} \\
1 & \text{otherwise}
\end{cases}
\]

Diagram:

1. Original network
2. Choose two edges \((v_1, w_1)\) and \((v_2, w_2)\)
3. Measure degrees \(j_1, k_1, j_2, k_2\)
4. Replace edges \((v_1, v_2)\) and \((w_1, w_2)\)
If we only specify $r$ we have great degree of freedom in choosing $e_{jk}$.

Possible choice for disassortative case:

$$e_{jk}^{(d)} = q_j x_k + x_j q_k - x_j x_k$$

Where $x_k$ is any normalized distribution.

This form satisfies the constraints on $e_{jk}$:

$$\sum_j e_{jk} = \sum_j q_k x_j + x_j q_k - x_j x_j = 1 + 1 - 1 = 1$$

$$\sum_j e_{jk} = \sum_j q_k x_j + x_j q_k - x_j x_j = q_k + x_k - x_k = q_k$$

The $r$ value can be easily calculated:

$$r_d = \frac{\sum_{jk} (q_k x_j + x_j q_k - x_j x_j - q_k q_j)}{\sigma_r^2} = \frac{2 \langle k \rangle_q \langle k \rangle_x - \langle k \rangle_x^2 - \langle k \rangle_q^2}{\sigma_r^2} = -\left(\frac{\langle k \rangle_x - \langle k \rangle_q}{\sigma_r}\right)^2$$

Assortative case:

$$e_{jk}^{(a)} = q_j q_k - e_{jk}^{(d)} \quad \rightarrow \quad r_a = -r_d$$

EXAMPLE: Erdős-Rényi

Network Science: Degree Correlations
EXAMPLE: Erdős-Rényi

Network Science: Degree Correlations
Structural cut-off

High assortativity $\rightarrow$ high number of links between the hubs.

If we allow only one link between two nodes, we can simply run out of hubs to connect to each other to satisfy the assortativity criteria.

Number of edges between the set of nodes with degree $k$ and degree $k'$:

$$E_{kk'} = e_{kk'} \langle k \rangle N$$

Maximum number of edges between the two groups:

$$m_{kk'} = \min \{ kN_k, k'N_{k'}, N_kN_{k'} \}$$

There cannot be more links between the two groups, than the overall number of edges joining the nodes with degree $k$.

If we only have simple edges, we cannot have more links between the two groups, than if we connect every node with degree $k$ to every node with degree $k'$ once.

This is true even if we allow multiple edges.

Structural cut-off

\[ E_{kk'} = e_{kk'} \langle k \rangle N \]

\[ m_{kk'} = \min \{ kN_k, k'N_{k'}, N_kN_{k'} \} \]

The ratio of \( E_{kk'} \) and \( m_{kk'} \) has to be \( \leq 1 \) in the physical region!

\[ r_{kk'} = \frac{E_{kk'}}{m_{kk'}} \leq 1 \]

\[ r_{k's_k} = 1 \] defines the structural cut-off

Uncorrelated networks: $m_{kk'} = \min\{kN_k, k'N_{k'}, N_k N_{k'}\}$

$\begin{align*}
m_{kk} &= k_s N_{k_s} = k_s N p_{k_s} \\
N_{k_s} &= k_s N p_{k_s}
\end{align*}$

$k_s(N)$ represents a structural cutoff:
One cannot have nodes with degree larger than $k_s(N)$.

→ If there are nodes with $k > k_s(N)$ we cannot find sufficient links between the highly connected nodes to maintain the neutral nature of the network.

Solution:
(a) Introduce a structural cutoff (i.e. do not allow nodes with $k > k_s(N)$)
(b) Let the network become more disassortative, having fewer links between hubs.
Example: Degree sequence introduces disassortativity

Scale-free network generated with the configuration model (N=300, L=450, $\gamma=2.2$).

**The measured $r=-0.19! \rightarrow$ Dissasortative!**

Red hub: 55 neighbors.
Blue hub: 46 neighbors.

Let's calculate the expectation number of links between red node ($k=55$) and blue node ($k=46$) for uncorrelated networks!

Here $N_{55}=N_{46}=1$, hence $m_{55,46}=1$ so $r_{55,46}=E_{55,46}$

$$E_{55,46} = \langle k \rangle N \cdot e_{55,46} = 900 \cdot \frac{55}{300} \cdot \frac{1}{3^2} \cdot \frac{46}{300} \cdot \frac{1}{3^2} \approx 2.8 > 1$$

In order for the network to be neutral, we need 2.8 links between these two hubs.
The largest nodes have $k_{nn} < < k_{nn}$. 

\[ 1 - CDF = P(k' > k) = 1 - \sum_{k'}^k p_{k'} \]
The effect is particularly clear for $N=10,000$:

The red curves are those of interest to us: one can see that a clear dissasortativity property is visible in this case.
Natural cutoffs in scale-free networks

All real networks are finite → let us explore its consequences.

→ We have an expected maximum degree, $K_{\text{max}}$

Estimating $K_{\text{max}}$

$$\int_{K_{\text{max}}}^{\infty} P(k)dk \approx \frac{1}{N}$$

Why: the probability to have a node larger than $K_{\text{max}}$ should not exceed the prob. to have one node, i.e. $1/N$ fraction of all nodes

$$\int_{K_{\text{max}}}^{\infty} P(k)dk = (\gamma - 1)K_{\text{min}}^{\gamma - 1} \int_{K_{\text{max}}}^{\infty} k^{-\gamma} dk = \frac{(\gamma - 1)}{(-\gamma + 1)}K_{\text{min}}^{\gamma - 1} \left[ k^{-\gamma + 1} \right]_{K_{\text{max}}}^{\infty} = \frac{K_{\text{min}}^{\gamma - 1}}{K_{\text{max}}^{\gamma - 1}} \approx \frac{1}{N}$$

Natural cutoff: $K_{\text{max}} = K_{\text{min}}N^{\gamma - 1}$
Structural cut-off for uncorrelated networks

Structural cutoff:  
\[ k_s(N) \sim \left( \langle k \rangle N \right)^{1/2} \]

Natural cut-off:  
\[ k_{\text{max}}(N) \sim N^{\gamma-1} \]

**\( \gamma = 3 \):** \( k_s(N) \) and \( k_{\text{max}}(N) \) scale the same way, i.e. \( \sim N^{1/2} \).

**\( \gamma < 3 \):**  
\( k_{\text{max}} > k_s \)

The size of the largest hub is above the structural cutoff, which means that it cannot have enough links to the other hubs to maintain its neutral status.  
\( \rightarrow \) **disassortative mixing**

\( \rightarrow \) a randomly wired network with \( \gamma < 3 \) will be  
(a) dissasortative  
(b) Or will have to have a cutoff at \( k_s(N) < k_{\text{max}}(N) \)
Example: introducing a structural cut-off

Scale-free network generated with the configuration model (N=300, L=450, γ=2.2) with structural cut-off $\sim N^{\frac{1}{2}}$.

$$r=0.005 \rightarrow \text{neutral}$$

Red hub: 12 neighbors.
Blue hubs: 11 neighbors.

Again we can calculate the expectation number of edges between the hubs.

$$E_{11,12} = \langle k \rangle N \cdot e_{11,12} = 900 \cdot \frac{12}{300} \cdot \frac{11}{300} = \frac{2}{3^2} \approx 0.3 < 1$$
The largest nodes have $k_{nn} \sim \langle k_{nn} \rangle$. 

\[ 1 - CDF = P(k' > k) = 1 - \sum_{k'}^k p_{k'} \]
The effect is particularly clear for $N=10,000$:

A clear case of neutral assortativity property is visible in this case thanks to imposing structural cut-off.
\[ r_{\alpha\beta} = \frac{\sum_{jk} (e_{jk}^{\alpha\beta} - q_j^{\alpha} q_k^{\beta})}{\sigma^\alpha \sigma^\beta} \]

\( \alpha, \beta: \{\text{in}, \text{out}\} \)

Pearson-correlation for directed networks

- **WWW**
- **Political Blogs**

$r_{\alpha\beta}$

- **in-in**
- **in-out**
- **out-in**
- **out-out**
P(k): not enough to characterize a network

Large degree nodes tend to connect to large degree nodes
Ex: social networks

Large degree nodes tend to connect to small degree nodes
Ex: technological networks
MULTIPOINT DEGREE CORRELATIONS

Measure of correlations:
$P(k',k'',...k^{(n)}|k)$: conditional probability that a node of degree $k$ is connected to nodes of degree $k'$, $k''$, ...

Simplest case:
$P(k'|k)$: conditional probability that a node of degree $k'$ is connected to a node of degree $k$
2-POINTS: CLUSTERING COEFFICIENT

- $P(k', k''|k)$: cumbersome, difficult to estimate from data

Do your friends know each other?

$C(i) = \frac{k(k - 1)}{2}$

$C = 0$

$C = 0.5$

$C = 1$
• Average clustering coefficient

\[ \bar{C} = \frac{1}{N} \sum_i C(i) \]

= average over nodes with very different characteristics
EMPIRICAL DATA FOR REAL NETWORKS

Pathlength

\[ l \approx N^{1/E} \]

Clustering

\[ C \sim \text{const} \]

Degree Distr.

\[ P(k) \sim k^{-\gamma} \]

**Regular network**

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

**Erdos-Renyi**

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

**Watts-Strogatz**

\[ C \sim \text{const} \]

**Barabasi-Albert**

\[ l \approx \frac{\ln N}{\ln \ln N} \]

\[ C \sim \frac{(\ln N)^2}{N} \]

\[ P(k) = \delta(k-k_d) \]

\[ P(k) = e^{-\langle k \rangle} \frac{<k^k>}{k!} \]

Exponential

\[ P(k) \sim k^{-\gamma} \]
CLUSTERING COEFFICIENT OF THE BA MODEL

Reminder: for a random graph we have:

\[ C_{rand} = \frac{\langle k \rangle}{N} \sim N^{-1} \]

The numerical results indicate a \textit{slightly} slower decay for BA network than for random networks.

But not slow \textit{enough}...

Konstantin Klemm, Victor M. Eguiluz,
Growing scale-free networks with small-world behavior,
MODULARITY IN THE METABOLISM

Clustering Coefficient:

\[ C(k) = \frac{\text{# links between } k \text{ neighbors}}{k(k-1)/2} \]

Metabolic network
(43 organisms)

Scale-free model
Existence of a high degree of local modularity in real networks, that is not captured by the current models.

C(N)— the average number of triangles around each node in a system of size N.

The fact that C(N) does not decrease means that the relative number of triangles around a node remains constant as the system size increases—in contrast with the ER and BA models, where the relative number of triangles around a node decreases. (here relative means relative to how many triangles we expected if all triangles that could be there would be there)

But C has some unexpected behavior, if we measure $C(k)$— the average clustering coefficient for nodes with degree $k$. 
CORRELATIONS: CLUSTER SPECTRUM

• Average clustering coefficient

\[ \bar{C} = \frac{1}{N} \sum_i C(i) \]

= average over nodes with very different characteristics

• Clustering spectrum:

\[ C(k) = \frac{1}{N_k} \sum_{i \in k_i - k} C(i) \]

putting together nodes which have the same degree

(link with hierarchical structures)
This is not true, however, for real networks. Let us look at some empirical data.
Hierarchical Networks

Society

Hollywood

Language

Human communication

The electronic skin

WWW
Eckmann & Moses, ‘02

Internet (AS)
Vazquez et al,’01
Cellular networks:

**GENOME**

- protein-gene interactions

**PROTEOME**

- protein-protein interactions

**METABOLISM**

- Bio-chemical reactions

---

**Citrate Cycle**

- 2-oxoglutarate
- succinate
- fumarate
- malate
- oxaloacetate
- L-glutamate
- aspartate

---

**Proteins**

- protein-gene interactions
- protein-protein interactions

**Bio-chemical reactions**

- Citrate Cycle
Protein-protein interaction

Regulatory networks
The metabolism forms a hierarchical network.

Geographically localized networks

- Internet (router)
- Power Grid
### SUMMARY OF EMPIRICAL RESULTS

#### $C(k) \sim k^{-\beta}$
- Internet (AS)
- WWW
- Metabolism
- Protein interaction network
- Regulatory network
- Language

#### $C(k)$ indep. of $k$
- Internet (router)
- Power grid

<table>
<thead>
<tr>
<th>Real systems</th>
<th>Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet (AS)</td>
<td>ER model</td>
</tr>
<tr>
<td>WWW</td>
<td>WS model</td>
</tr>
<tr>
<td>Metabolism</td>
<td>BA model</td>
</tr>
<tr>
<td>Protein interaction network</td>
<td></td>
</tr>
<tr>
<td>Regulatory network</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
</tr>
</tbody>
</table>

But there is a deeper issue as stake, that need to consider— that of modularity.
All models predict $C(k) \sim k^{-1}$

Is the exponent universal?

Or could we have for example: $C(k) \sim k^{-\beta}$
Randomly pick a $p$ fraction of the newly added nodes and connect each of them independently to the nodes belonging to the central module.
- Use preferential attachment to decide, to which central node the selected nodes link to.
- At the next level $p^2$ fraction will link back, then $p^3$, … $p^i$
1. Scale-free

\[ \gamma = 1 + \frac{\ln 5}{\ln 4} = 2.161 \]

2. Clustering coefficient independent of N

\[ C(N) = \text{const.} \]

3. Clustering spectrum

\[ C'(k) \sim k^{-1} \]

In real systems \( C(k) \) does not always decrease as a power law. What matters, however, that it decreases, i.e. it is not independent of \( k \).
Hierarchy is a new rather generic network property.

What does happen in real systems? Is a prediction that all systems with $\gamma<3$ should be automatically dissasortative, or have a cutoff – is this the case?
Let’s see: www, $\gamma=2.1$, no cutoff, dissasortative NICE
Actor network, no cutoff, but it is ASSORTATIVE (how is this possible?).
Internet: $\gamma=2.5$, disassortative, cutoff , NICE

Networks with $\gamma<3$ don’t have to be assortative:
Lets suppose we have a neutral network. High assortativity means a high degree nodes neighbors have high average degree. If we want to make it assortative we have to increase the degree of the neighbors of hubs. Even if the degree of the top neighbors cannot be increased because we used up all of the hubs, the low degree neighbors still can be replaced with higher ones, thus making the network assortative.
Anyway, the social networks checked (actor network, coauthorship network) have cut-offs according to Newman and Stanley.
http://samoa.santafe.edu/media/workingpapers/00-07-037.pdf
Static model used for examples

- Start with $N$ unconnected nodes.
- Assign a $w_i$ weight to each node $i$.
- Randomly select two nodes with probability proportional to $w_i$. Connect these nodes. Repeat $L$ times.

\[ w_i = \frac{1}{i^\alpha} \quad \rightarrow \quad p_k \sim k^{-1-1/\alpha} \]

Upper cut-off may be added by introducing $i_0$:  
\[ w_i = \frac{1}{(i+i_0)^\alpha} \]

For large $N$ this should be equivalent to the configuration model.
Static model used for examples

- Start with $N$ unconnected nodes.
- Assign a $w_i$ weight to each node $i$.
- Randomly select two nodes with probability proportional to $w_i$. Connect these nodes. Repeat $L$ times.

If $w_i = \frac{1}{i^\alpha} \rightarrow p_k \sim k^{-1-1/\alpha}$

Upper cut-off may be added by introducing $i_0$: $w_i = \frac{1}{(i + i_0)^\alpha}$

For large $N$ this should be equivalent to the configuration model.
A giant cluster exists if each node is connected to at least two other nodes.

The average degree of a node $i$ linked to the GC, must be 2, i.e.

$$< k_m | i \leftrightarrow j > = \sum_{k_m} k_m P(k_m | i \leftrightarrow j) = 2$$

$$P(k_m | i \leftrightarrow j) = \frac{P(k_m, i \leftrightarrow j)}{P(i \leftrightarrow j)} = \frac{P(i \leftrightarrow j \mid k_m) P(k_m)}{P(i \leftrightarrow j)}$$

Bayes’ theorem

$P(k_m | i \leftrightarrow j)$: joint probability that a node has degree $k_m$ and is connected to nodes $i$ and $j$.

For a randomly connected network (does NOT mean random network!) with $P(k)$:

$$P(i \leftrightarrow j) = \frac{2L}{N(N-1)} = \frac{< k >}{N-1}$$

$$P(i \leftrightarrow j | k_m) = \frac{k_m}{N-1}$$

$$\sum_{k_m} k_m P(k_m | i \leftrightarrow j) = \sum_{k_m} k_m \frac{P(i \leftrightarrow j | k_m) P(k_m)}{P(i \leftrightarrow j)} = \sum_{k_m} k_m \frac{k_m P(k_m)}{< k >} = \frac{\sum k_m^2 P(k_m)}{< k >}$$

$$\kappa \equiv \frac{< k^2 >}{< k >} = 2$$

$\kappa > 2$: a giant cluster exists

$\kappa < 2$: many disconnected clusters

Apply the Malloy-Reed Criteria to an Erdos-Renyi Network

Discrete Formulation
- binomial distribution -

\[ P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} \]

Continuum Formulation
- Poisson distribution -

\[ P(k) = e^{-<k>} \frac{<k>^k}{k!} \]

\[ <k> = (N-1)p \]
\[ <k^2> = p(1-p)(N-1) + p^2(N-1)^2 \]
\[ \sigma_k = (\sqrt{<k^2> - <k>^2})^{1/2} = [p(1-p)(N-1)]^{1/2} \]
A giant cluster exists if each node is connected to at least two other nodes.

\[ \kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = 2 \]

\( \kappa > 2: \) a giant cluster exists;

\( \kappa < 2: \) many disconnected clusters;

\[ \langle k \rangle = \langle k \rangle \]

\[ \langle k^2 \rangle = \langle k \rangle (1 + \langle k \rangle) \]

\[ \sigma_k = (\langle k^2 \rangle - \langle k \rangle^2)^{1/2} = \langle k \rangle^{1/2} \]