Frontiers of Network Science
Fall 2020

Class 18: Degree Correlations II
(Chapter 7 in Textbook)

Boleslaw Szymanski

based on slides by Albert-László Barabási and Roberta Sinatra

www.BarabasiLab.com
\[ r_{\alpha\beta} = \frac{\sum_{jk} (e_{jk}^\alpha - q_j^\alpha q_k^\beta)}{\sigma^\alpha \sigma^\beta} \]

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

**J. G. Foster, D. V. Foster, P. Grassberger, M. Paczuski, PNAS 107, 10815 (2010)**
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

Do your friends know each other?

P(k’,k’’|k): cumbersome, difficult to estimate from data

# of links between neighbors

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

\[ C = 0 \]
\[ C = 0.5 \]
\[ C = 1 \]
• Average clustering coefficient

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

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

Pathlength

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

Clustering

\[ C \sim const \]

Degree Distr.

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

- Regular network
- Erdos-Renyi
- Watts-Strogatz
- Barabasi-Albert

- \[ l_{rand} \approx \frac{\log N}{\log \langle k \rangle} \]
- \[ C_{rand} = p = \frac{\langle k \rangle}{N} \]
- \[ C \sim \frac{(\ln N)^2}{N} \]
- \[ P(k) = e^{-\kappa} \frac{\kappa^k}{k!} \]
- \[ P(k) = \delta(k-k_d) \]
Reminder: for a random graph we have:

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

The numerical results indicate a slightly slower decay for BA network than for random networks.

But not slow enough...

Konstantin Klemm, Victor M. Eguiluz,
Growing scale-free networks with small-world behavior,
Clustering Coefficient:

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

Hollywood, Society

The electronic skin

WWW, Internet (AS)

Language

Eckmann & Moses, ‘02

Vazquez et al, ‘01
Cellular networks:

PROTEOME

protein-protein interactions

METABOLISM

Bio-chemical reactions

GENOME

protein-gene interactions
Protein-protein interaction

Regulatory networks
The metabolism forms a hierarchical network.

ABSENCE OF HIERARCHY

Geographically localized networks

[Graphs showing the distribution of nodes and degree distribution for Internet (router) and Power Grid.]
But there is a deeper issue as stake, that need to consider— that of modularity.

<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>

$C(k) \sim k^{-\beta}$  

$C(k)$ indep. of $k$

Internet (router)  
Power grid
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:
Let’s 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
http://viseu.chem-eng.northwestern.edu/site_media/publication_pdf/00-07-037.pdf
http://viseu.chem-eng.northwestern.edu/site_media/publication_pdf/00-07-038.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.

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

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.

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

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.

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

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

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 $P(k)$:

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

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

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

$\kappa > 2$: a giant cluster exists

$\kappa < 2$: many disconnected clusters

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

Discrete Formulation
-binomial distribution-

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

\[ \langle k \rangle = (N - 1) p \]

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

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

Continuum Formulation
-Poisson distribution-

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

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

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

\[ \sigma_k = (\langle k^2 \rangle - \langle k \rangle^2)^{1/2} = \langle k \rangle^{1/2} \]
A giant cluster exists if each node is connected to at least two other nodes.

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

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

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

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

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

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

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

\[ \kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\langle k \rangle (1 + \langle k \rangle)}{\langle k \rangle} = 1 + \langle k \rangle = 2 \]

\[ \langle k \rangle \geq 1 \]