Frontiers of Network Science
Fall 2017

Class 15: Degree Correlations II
(Chapter 7 in Textbook)

Boleslaw Szymanski

based on slides by
Albert-László Barabási
and Roberta Sinatra

www.BarabasiLab.com
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:

$$m_{kk'} = \min \left\{ kN_k, k'N_{k'}, N_k N_{k'} \right\}$$

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!

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

\[ r_{k_s k_s} = 1 \] defines the structural cut-off

Uncorrelated networks:

- $m_{kk'} = \min\{kN_k, k'N_{k'}, N_k N_{k'}\}$
- $m_{k_s k_s} = k_s N_{k_s} = k_s N p_{k_s}$
- $m_{k_s k_s} = N_{k_s}^2 = N^2 p_{k_s}^2$

$e_{kk'} = q_k q_{k'} = \frac{kk' p_k p_{k'}}{\langle k \rangle^2}$

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

$r_{k_s k_s} = \frac{\langle k \rangle N \cdot k_s^2 \cdot p_{k_s}^2}{\langle k \rangle^2} = \frac{k_s p_{k_s}}{\langle k \rangle} = q_{k_s} < 1 \ \forall k_s$

$k_{s}(N) = \left(\langle k \rangle N\right)^{1/2}$

$k_{s}(N)$ represents a structural cutoff:

- one cannot have nodes with degree larger than $k_{s}(N)$,

$\Rightarrow$ 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 $r=-0.19! \rightarrow \textbf{Dissasortative!}$

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 $r_{55,46}=E_{55,46}$

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

In order for the network to be neutral, we need 2.8 links between these two hubs.
1 - CDF = P(k' > k) = 1 - \sum_{k'}^{k} p_{k'}.

The largest nodes have \( k_{nn} \ll \langle k_{nn} \rangle \).
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 \( \rightarrow \) let us explore its consequences.

\( \rightarrow \) 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 probability to have one node, i.e. \( 1/N \) fraction of all nodes

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

**Natural cutoff:**

\[
K_{\text{max}} = K_{\text{min}} N^{\gamma - 1}
\]
Structural cut-off for uncorrelated networks

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

\[ e_{kk'} = q_k q_{k'} = \frac{kk' p_k p_{k'}}{\langle k \rangle^2} \]

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.

\( \rightarrow \text{disassortative mixing} \)

\( \rightarrow \) 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.

$$E_{11,12} = \langle k \rangle N \cdot e_{11,12} = 900 \cdot \frac{12}{300} \cdot \frac{11}{300} \approx 0.3 < 1$$
\(1 - CDF = P(k' > k) = 1 - \sum_{k'} p_{k'}\)

The largest nodes have \(k_{nn} \sim \langle k_{nn} \rangle\)
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.
\[ r_{\alpha \beta} = \frac{\sum_{jk} (e_{jk}^{\alpha \beta} - q_j^{\alpha} q_k^{\beta})}{\sigma_{\alpha} \sigma_{\beta}} \]

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

DIRECTED NETWORKS

Pearson-correlation for directed networks

- www
- political blogs

Values of $r_{\alpha\beta}$ for different types of edges:
- 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**

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

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

**Erdos-Renyi**

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

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

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

**Watts-Strogatz**

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

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

**Barabasi-Albert**

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

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

\[ P(k) \sim k^{-\gamma} \]
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,
MODULARITY IN THE METABOLISM

Clustering Coefficient:

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

\[ C = 1 \quad \quad C = \frac{1}{2} \quad \quad C = 0 \]

Metabolic network (43 organisms)

Scale-free model
THE MEANING OF C(N)

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:

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

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

Human communication

Language

The electronic skin

WWW

Eckmann & Moses, ‘02

Internet (AS)

Vazquez et al, ‘01

Network Science: Degree Correlations
Cellular networks:

GENOME

protein-gene interactions

PROTEOME

protein-protein interactions

METABOLISM

Bio-chemical reactions
A GENE REGULATORY NETWORK

**INPUT**
- signal A
  - receptor proteins
  - cascade of interacting kinase proteins or other molecules
- inactive transcription factor A
- active transcription factor A
- DNA
  - cis-regulatory DNA sequence elements
  - RNA polymerase

**INPUT**
- signal B
  - receptor proteins
  - inactive transcription factor B
  - inhibitory factor
- active transcription factor B

**OUTPUT**
- mRNA
- target gene
- protein
  - cell functions

**OUTPUT**
- protein
  - feedbacks

GENOMES to LIFE

DEPARTMENT OF ENERGY
UNITED STATES OF AMERICA
Protein-protein interaction

Regulatory networks
The metabolism forms a hierarchical network.

ABSENCE OF HIERARCHY

Geographically localized networks

Network Science: Degree Correlations
### SUMMARY OF EMPIRICAL RESULTS

<table>
<thead>
<tr>
<th>$C(k) \sim k^{-\beta}$</th>
<th>$C(k)$ indep. of $k$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Real systems</strong></td>
<td></td>
</tr>
<tr>
<td>Internet (AS)</td>
<td>Internet (router)</td>
</tr>
<tr>
<td>WWW</td>
<td>Power grid</td>
</tr>
<tr>
<td>Metabolism</td>
<td></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>
<tr>
<td><strong>Models</strong></td>
<td></td>
</tr>
<tr>
<td>?</td>
<td>ER model</td>
</tr>
<tr>
<td></td>
<td>WS model</td>
</tr>
<tr>
<td></td>
<td>BA model</td>
</tr>
</tbody>
</table>

But there is a deeper issue as stake, that need to consider— that of modularity.
Real networks are fragmented into groups or modules


Traditional view of modularity:

MODULARITY VS. SCALE-FREE TOPOLOGY

(a) Scale-free

(b) Modular

Network Science: Degree Correlations
Clustering coefficient scales

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

\[ C(k) = \frac{\text{\# links between } k \text{ neighbors}}{k(k-1)/2} \]
WHAT DOES THE SCALING MEAN?

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

**Small \( k \) nodes:**
- high clustering coefficient;
- their neighbors tend to link to each other;
- in highly interlinked, compact communities.

**High \( k \) nodes (hubs):**
- small clustering coefficient;
- connect independent communities.
PROPERTIES OF THE MODEL

Degree distribution

\[ k_i(H_i) = \sum_{l=1}^{i} k^l = \frac{4}{3} \left( 4^i - 1 \right) \Rightarrow \ln k_i(H_i) \simeq i \cdot \ln 4 + \ln \frac{4}{3} \]

\[
\sum_{i=0}^{n} x' = \frac{x^{n+1} - 1}{x - 1} \quad \sum_{i=1}^{n} x' = \frac{x^{n+1} - 1}{x - 1} - 1 \quad \sum_{i=1}^{i} A' = \frac{4^{i+1} - 1}{4 - 1} - 1 = \frac{4^{i+1} - 4}{3}
\]

\[ N(H_i) = 4 \cdot 5^{n-i-1} \Rightarrow \ln N(H_i) = c_n - i \cdot \ln 5 \] (valid for i<n)

\[ \ln N(H_i) \simeq c'_n - \ln k_i \cdot \frac{\ln 5}{\ln 4} \Rightarrow N(H_i) \sim k_i^{-\frac{\ln 5}{\ln 4}} \]

\[ P(k_i) \sim \frac{N(H_i)}{(k_{i+1} - k_i)} \sim k_i^{-\gamma} \]

\[ k_{i+1} - k_i = \sum_{l=1}^{i+1} A' - \sum_{l=1}^{i} A' = 4^{i+1} = 3k_i + 4 \quad P(k_i) = \frac{k_i^{\ln 5}}{3k_i + 4} \sim k_i^{-\frac{\ln 5}{\ln 4}} \sim k_i^{-\gamma} \]

\[ \gamma = 1 + \frac{\ln 5}{\ln 4} \simeq 2.16 \]
PROPERTIES OF THE MODEL

Large average clustering

\[ C \approx 0.74 \]

Hierarchical clustering

\[ n_i \sim k_i \implies C(H_i) \sim \frac{2k_i}{k_i(k_i - 1)} \]

\[ C(k) \sim \frac{2}{k^*} \]
PROPERTIES OF HIERARCHICAL NETWORKS

1. Scale-free

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

2. Clustering coefficient independent of N

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

3. Scaling clustering coefficient (DGM)

\[ C(k) \sim k^{-1} \]
\[ \gamma = \frac{\ln 3}{\ln 2} \]

- Barabási, Ravasz, Vicsek, Physica A 2003

- Dorogovtsev, Goltsev, Mendes, 2001
All models predict $C(k) \sim k^{-1}$

Is the exponent universal?

Or could we have for example: $C(k) \sim k^{-\beta}$