Clustering Coefficient:

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

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

= average over nodes with very different characteristics

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

• Clustering spectrum:

putting together nodes which have the same degree

(putting together nodes which have the same degree)

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

class of degree k
C(k) for the ER and BA models

This is not true, however, for real networks. Let us look at some empirical data.
Cellular networks:

**GENOME**
- protein-gene interactions

**PROTEOME**
- protein-protein interactions

**METABOLISM**
- Bio-chemical reactions

Citrate Cycle
A GENE REGULATORY NETWORK

INPUT
signal A
receptor proteins
cascade of interacting kinase proteins or other molecules
active transcription factor A
inhibitory factor
OUTPUT
mRNA
OUTPUT
protein
cell functions
DNA
cis-regulatory DNA sequence elements
RNA polymerase
target gene

INPUT
signal B
receptor proteins
inactive transcription factor B
inactive transcription factor A
feedbacks

YGG 01-0083
Protein-protein interaction

Regulatory networks
The metabolism forms a hierarchical network.

ABSENCE OF HIERARCHY

Geographically localized networks

[Graph: Internet (router) and Power Grid, showing 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>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>Models</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.
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 \( p^3 \), \( p^4 \), \ldots, \( p^i \)

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

\[
C(k) \sim k^{-\beta}
\]
In real systems $C(k)$ does not always decrease as a power law. What matters, however, is that it decreases, i.e., it is not independent of $k$. 

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

$$C'(k) \sim k^{-1}$$
Hierarchy is a new rather generic network property.

FINAL REMARKS: EFFECT OF ASSORTATIVE MIXING: PERCOLATION

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
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.

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

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 \( P(k) \):

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

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

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

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

\( \kappa > 2 \): 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-

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

\[ <k> = (N-1)p \]

\[ <k^2> = p(1-p)(N-1) + p^2(N-1)^2 \]

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

\[ <k^2> = <k>(1+<k>) \]

\[ \sigma_k = (<k^2> - <k>^2)^{1/2} = <k>^{1/2} \]
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;

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

Random network: Damage is modeled as an inverse percolation process.

Component structure

Graph

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

\( f_c : \langle k \rangle = 1 \)

\( f = \text{fraction of removed nodes} \)

(Inverse percolation phase transition)