QUANTIFYING NETWORK PROPERTIES

Carlo PICCARDI

DEIB - Department of Electronics, Information and Bioengineering
Politecnico di Milano, Italy

email carlo.piccardi@polimi.it
http://home.deib.polimi.it/piccardi

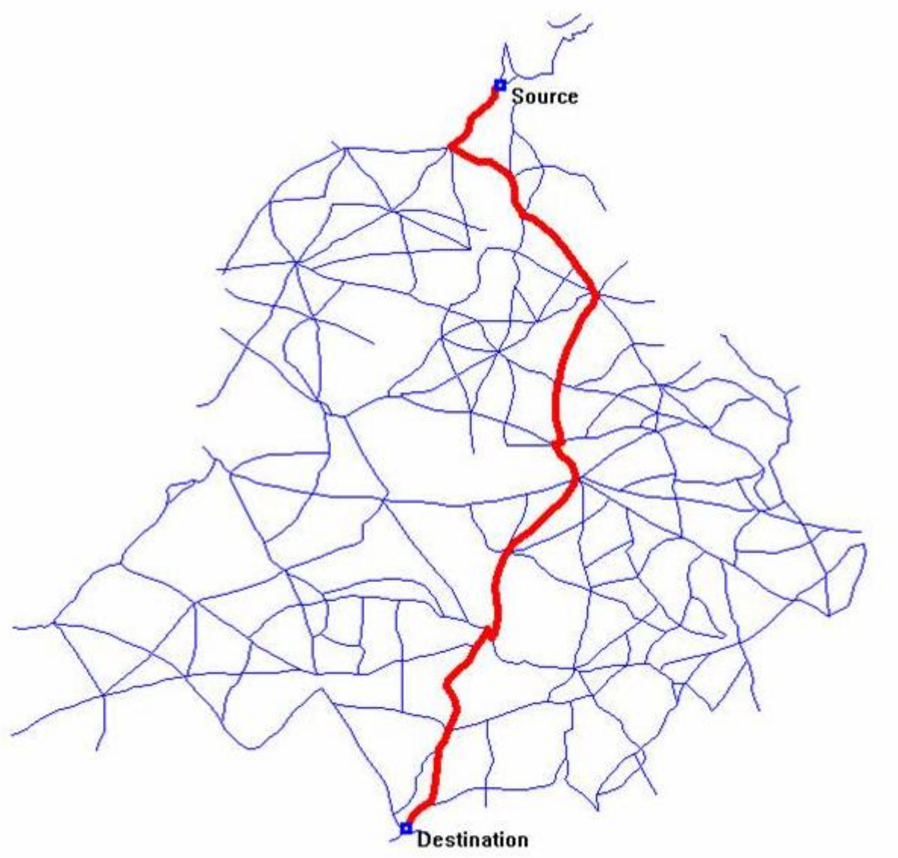
DISTANCE AND DIAMETER

The distance d_{ij} is the length (measured in number of links) of the shortest path connecting $i \rightarrow j$.

For a connected network, the diameter D and the average distance d are:

$$D = \max_{i,j} d_{ij}$$

$$d = < d_{ij} > = \frac{1}{N(N-1)} \sum_{i,j\,(i \neq j)} d_{ij}$$



If the network is weighted, several (non trivial) generalized definitions are available.

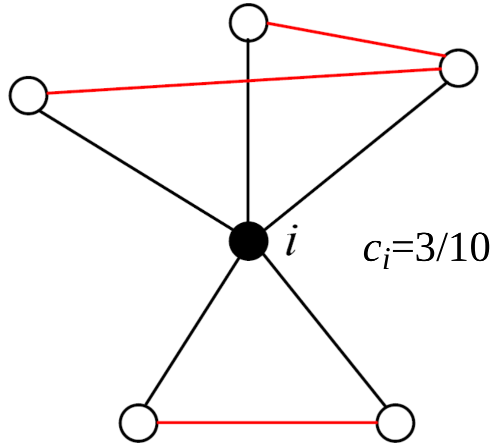
CLUSTERING (or TRANSITIVITY) COEFFICIENT

It quantifies the "local link density" by counting the triangles in the network.

How frequently, if we have the links $j \leftrightarrow i$ and $i \leftrightarrow l$, then we also have $j \leftrightarrow l$ (thus the triangle j,i,l) ? *(Or: how frequently two friends of mine are also friends?)*

The (local) clustering coefficient $0 \le c_i \le 1$ of node i is:

$$c_i = \frac{\#\,\text{triangles connected to } i}{\#\,\text{triples } j,i,l \text{ centered on } i} = \frac{e_i}{k_i(k_i - 1)/2}$$

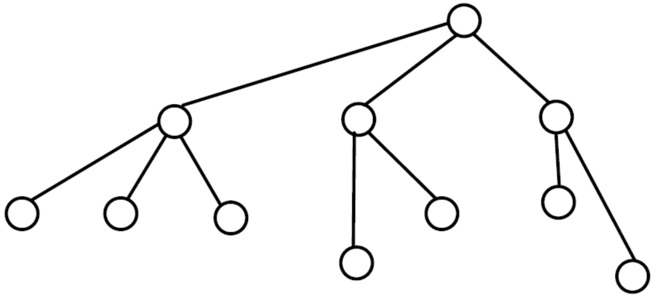


where k_i is the degree of i, and e_i the number of links directly connecting neighbors of i (at most $k_i(k_i - 1)/2$).

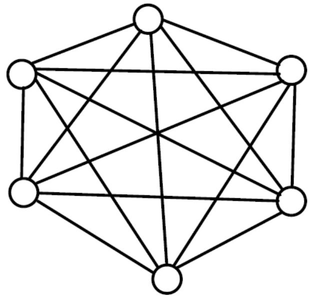
Large c_i, large (local) redundancy

Small c_i, large (local) influence of i

The (global) clustering coefficient C is the average c_i : $C = \langle c_i \rangle = \frac{1}{N}\sum_i c_i$



tree network: $C = 0$



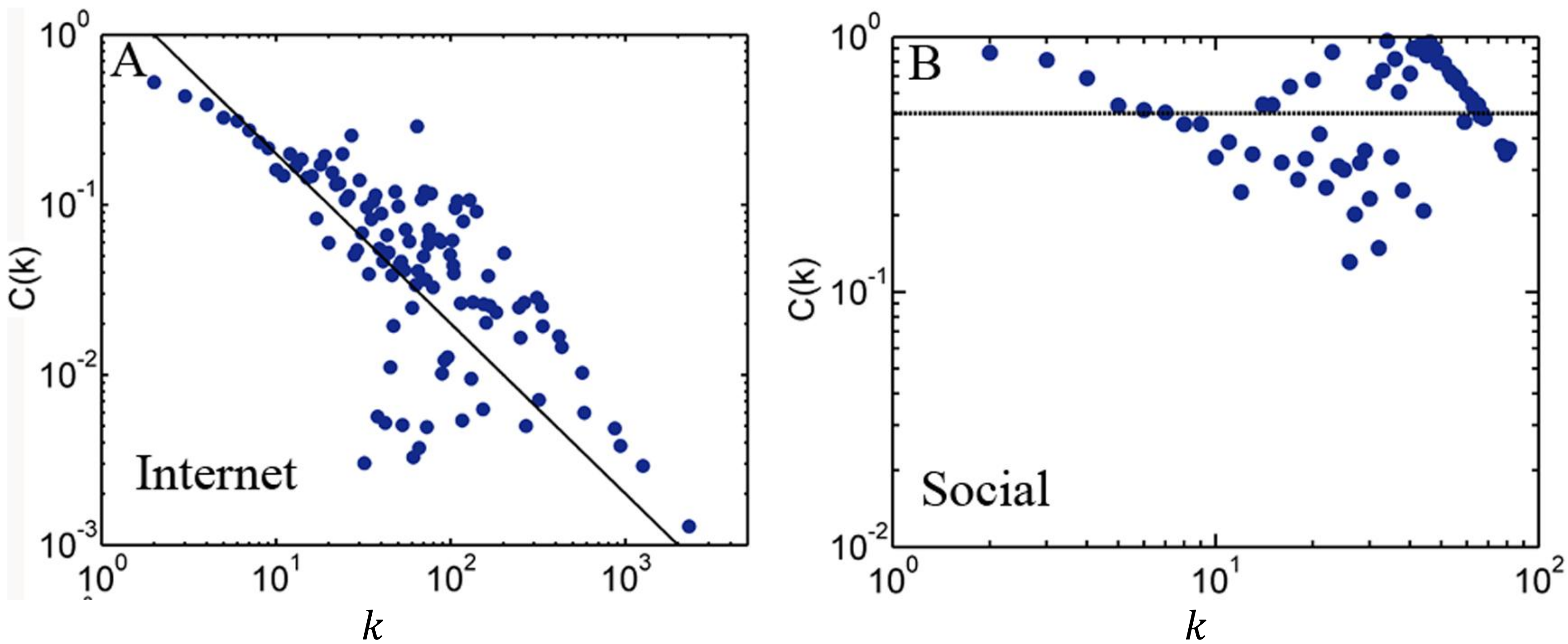
complete network: $C = 1$

Network	Size	Clustering coefficient	Average path length
Internet, domain level [13]	32711	0.24	3.56
Internet, router level [13]	228298	0.03	9.51
WWW [14]	153127	0.11	3.1
E-mail [15]	56969	0.03	4.95
Software [16]	1376	0.06	6.39
Electronic circuits [17]	329	0.34	3.17
Language [18]	460902	0.437	2.67
Movie actors [5, 7]	225226	0.79	3.65
Math. co-authorship [19]	70975	0.59	9.50
Food web [20, 21]	154	0.15	3.40
Metabolic system [22]	778	–	3.2

A more refined analysis: clustering coefficient c_i vs. degree k_i.

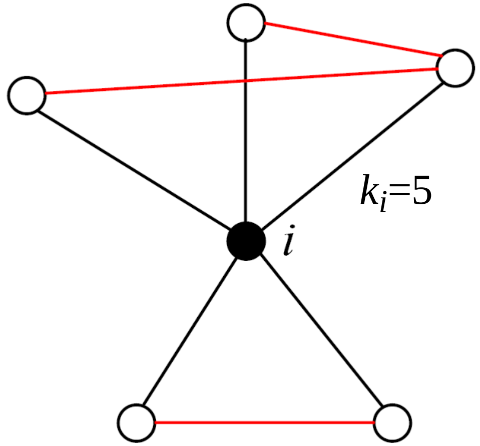
Grouping together the N_k nodes with the same degree $k_i = k$:

$$C(k) = \frac{1}{N_k} \sum_{i|k_i=k} c_i$$

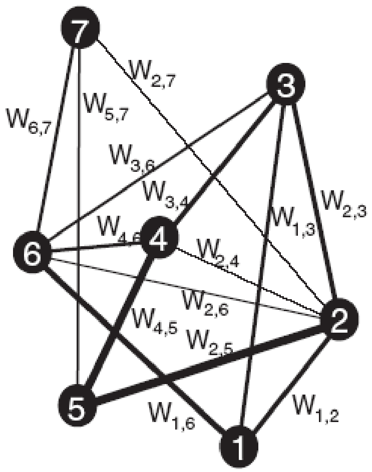


DEGREE AND STRENGTH OF A NODE

In an undirected network, the degree k_i of node i is the number of links connected to i (=the number of neighbors of i):



$$k_i = \sum_j a_{ij}$$



In a (undirected) weighted network, the strength s_i of node i is the total weight of the links connected to i:

$$s_i = \sum_j w_{ij}$$

If the network is directed, we must distinguish among in-, out-, and total degree, and in-, out-, and total strength of node i.

The degree distribution $P(k)$ of a network specifies the fraction of nodes having exactly degree k (=the probability that a randomly selected node has degree k):

$$P(k) = \frac{\#\,\text{nodes with degree } k}{N} \quad , \qquad \sum_k P(k) = 1$$

It is often more practical to consider the cumulative degree distribution:

$$\overline{P}(k) = \frac{\#\,\text{nodes with degree} \geq k}{N} = \sum_{h=k}^{k_{\max}} P(h) \quad , \qquad \overline{P}(k_{\min}) = 1$$

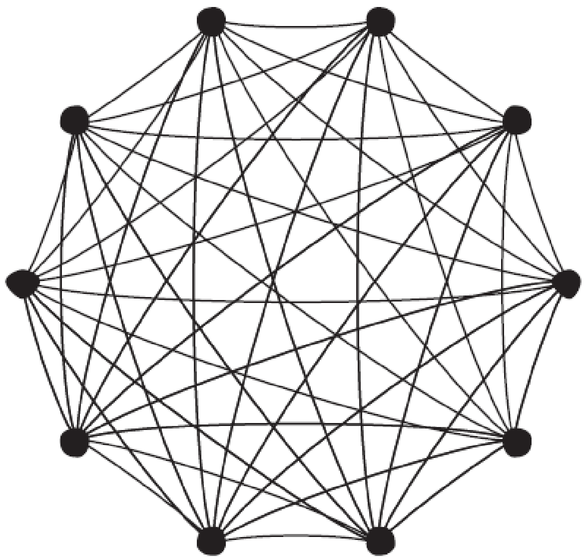
The r-moments of the degree distribution $P(k)$ are:

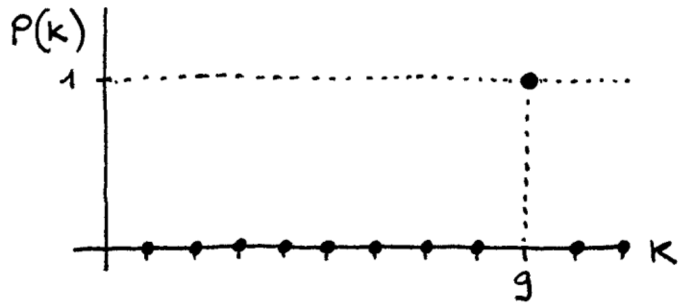
$$< k^r >= \sum_k k^r P(k) \quad , \quad r = 1,2,\ldots$$

The first moment ($r = 1$) is the average degree $< k >= \sum_k kP(k) = \frac{1}{N} \sum_i k_i = \frac{2L}{N}$.

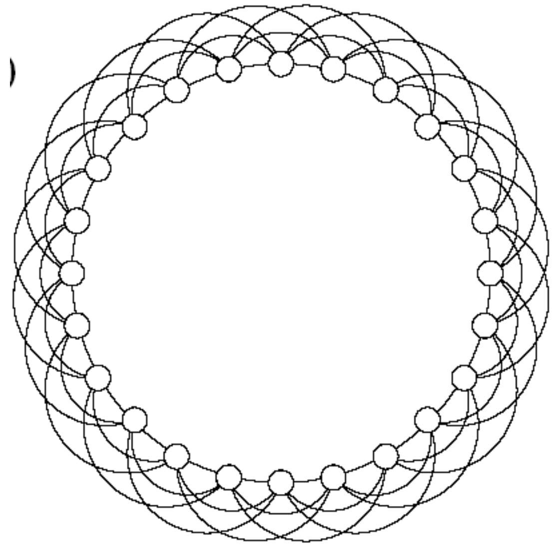
In a (strictly) homogeneous network all nodes have the same degree...

Example: a complete (=all-to-all) network with $N = 10$ and $k_i = \langle k \rangle = 9 \ \forall i$.

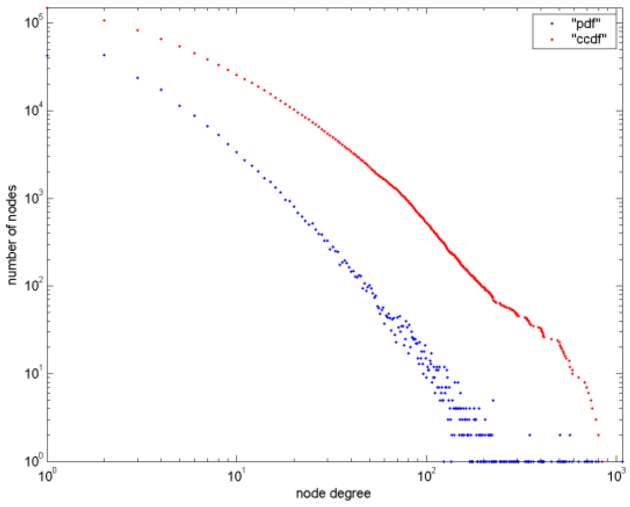




Example: a homogeneous network with $P(6) = 1$.

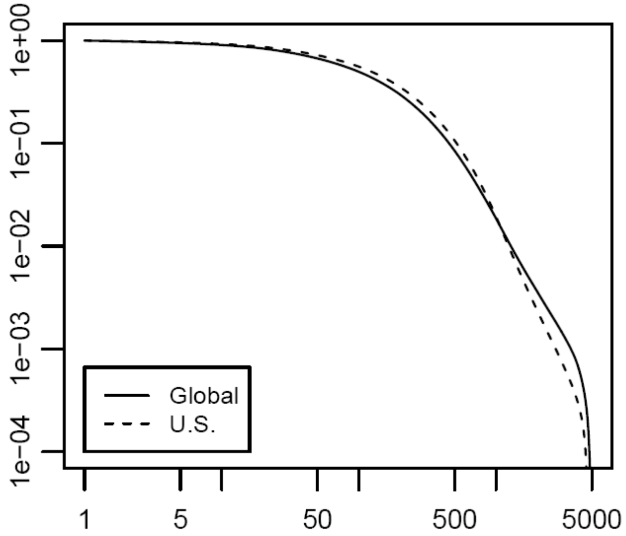


...but the degree distribution of real-world networks is typically very different.

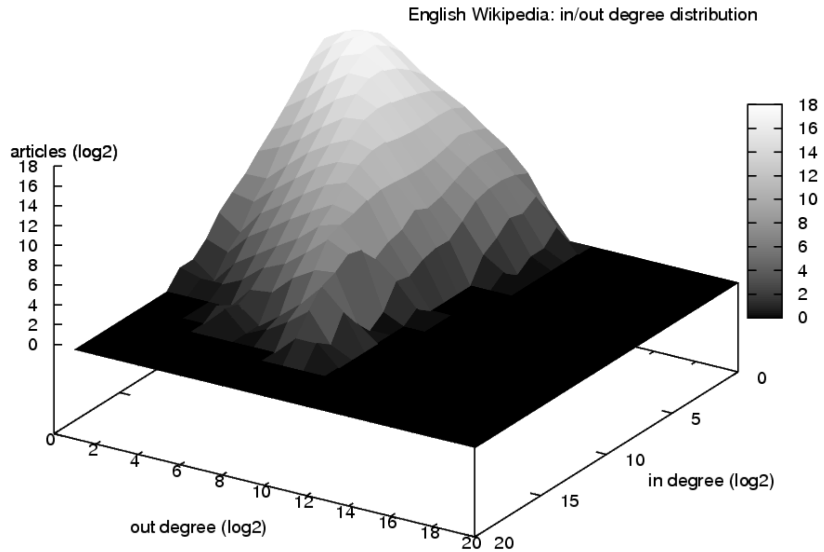


Internet

Facebook



Wikipedia (directed)



CORRELATED NETWORKS

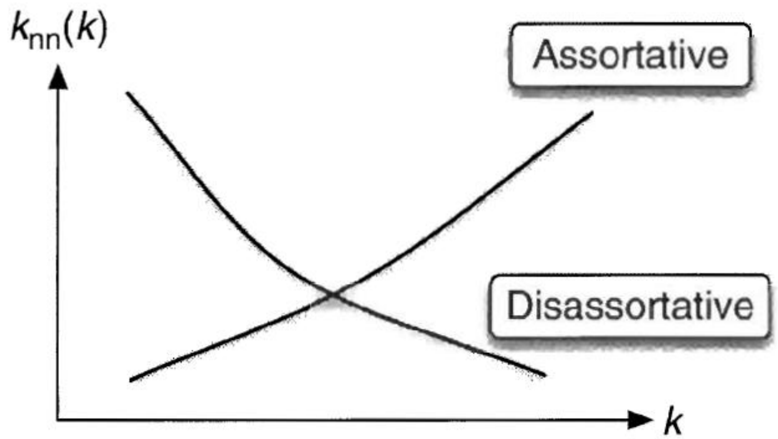
A network is correlated if the probability $Q(h|k)$ that the neighbour of a degree-k node has degree-h does depend on k.

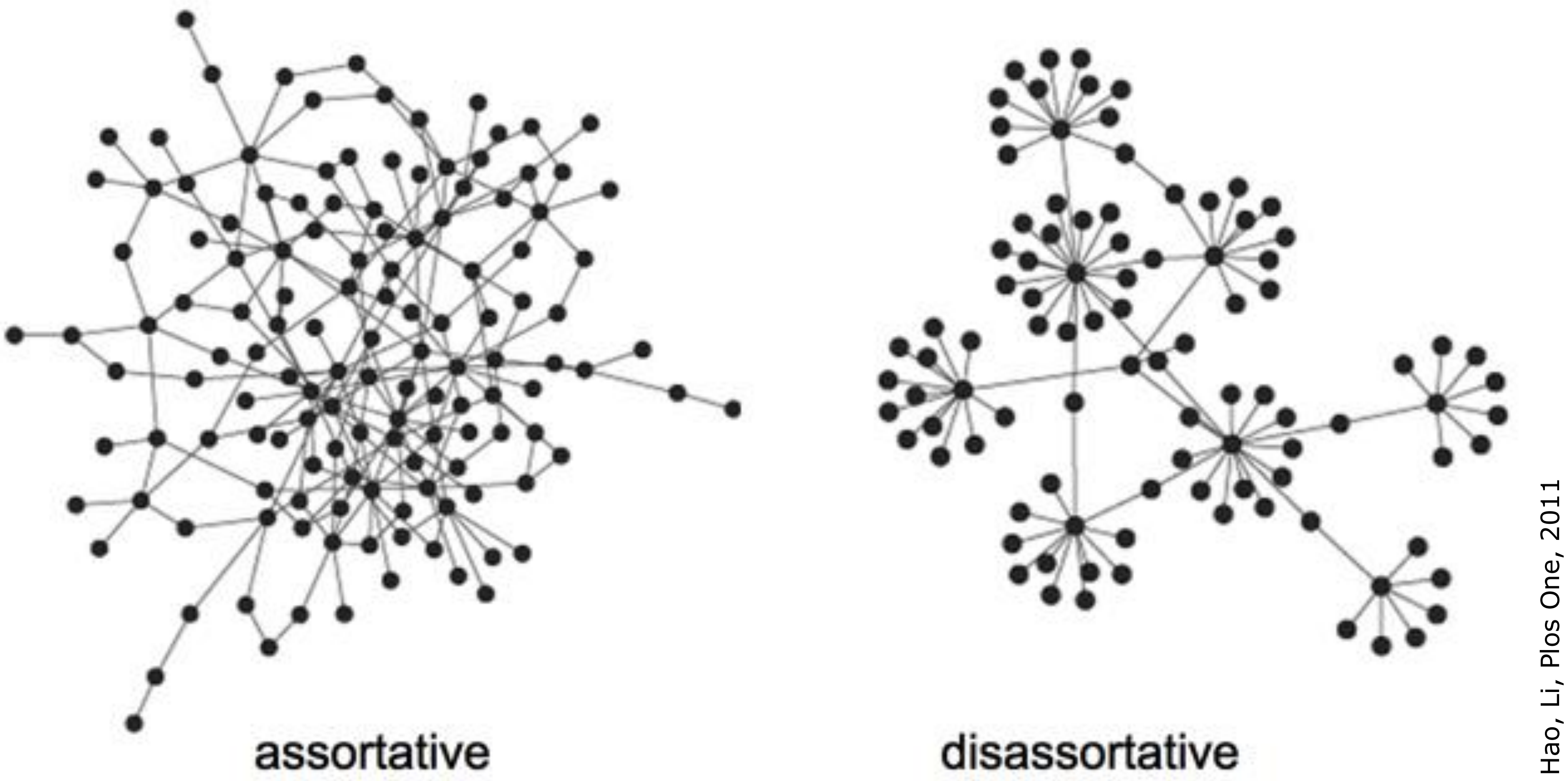
Correlations can be captured by the average nearest neighbours degree function:

$$k_{nn}(k) = \sum_h hQ(h|k)$$

In an assortative [disassortative] network, high-degree nodes tend to connect to high-degree [low-degree] nodes.

Examples: social [technological] networks are typically assortative [disassortative].





Two artificial networks with resp. assortative and disassortative patterns.

If the network is uncorrelated (or we assume it is) then $Q(h|k) = Q(h)$, the degree distribution of neighbours does not depend on k.

- The degree distribution of neighbours is not $P(k)$ but it is biased towards highest degrees:

$$Q(h) = \frac{\text{n. of links from nodes of degree } h}{\text{n. of links from nodes of any degree}} = \frac{hP(h)}{\sum_k kP(k)} = \frac{hP(h)}{< k >}$$

- The average nearest neighbours degree $k_{nn}(k)$ is constant...

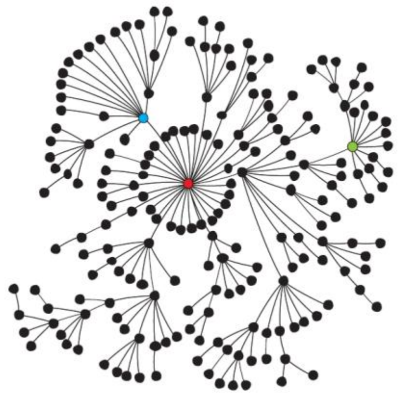
$$k_{nn}(k) = \sum_h hQ(h|k) = \sum_h \frac{h^2 P(h)}{< k >} = \frac{< k^2 >}{< k >}$$

- ... and larger than $< k >$:

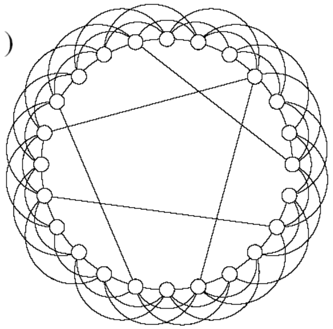
$$k_{nn} = \frac{< k^2 >}{< k >} = \frac{< k >^2 + \sigma^2}{< k >} = < k > + \frac{\sigma^2}{< k >}$$

The "friendship paradox" (*my friends have more friends than I have*): applications in finding hub nodes.

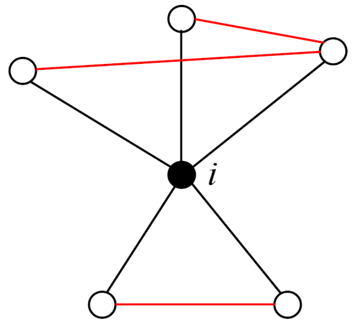
REMARK: The term "complex network" is mostly used to define graphs with non-trivial features, including:



- *The degree distribution is broad: the network has no characteristic scale (scale-free networks).*



- *The average shortest path increases slowly (=logarithmically) with the number of nodes (small-world property).*



- *The clustering coefficient is much larger than in randomly generated networks.*

The above features are found in most real-world networks, and have important effects on a number of (static and dynamic) network properties.