QUANTIFYING NETWORK PROPERTIES

Carlo PICCARDI

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

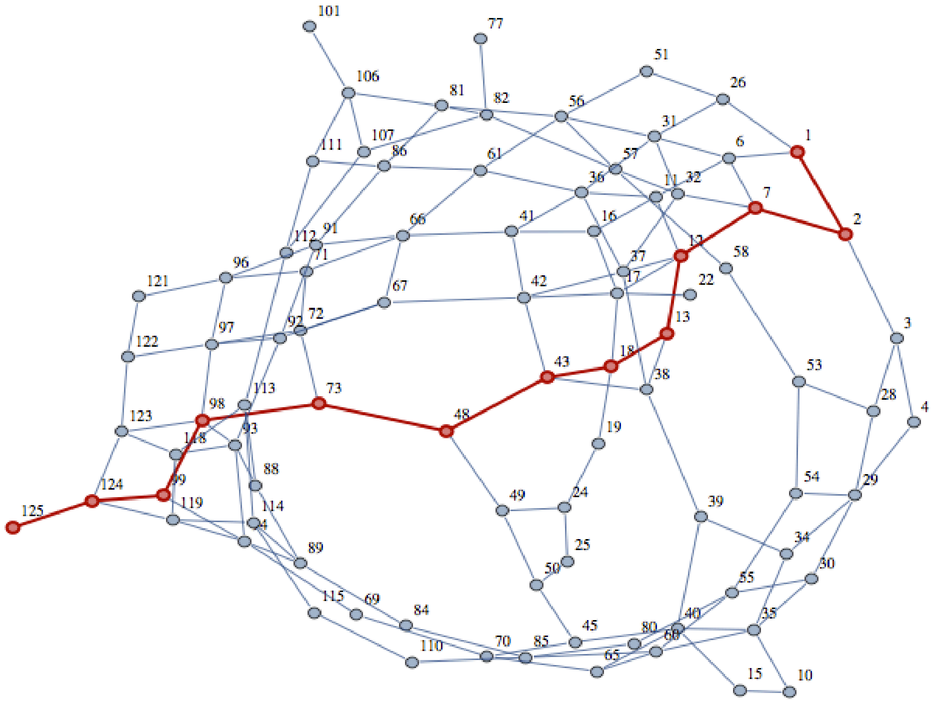
email carlo.piccardi@polimi.it
https://piccardi.faculty.polimi.it/

DISTANCE AND DIAMETER

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

In a directed network $d_{ij} \neq d_{ji}$, in general.

In the network is connected (=all paths $i \to j$ exist) we can define the diameter D and the average distance d :



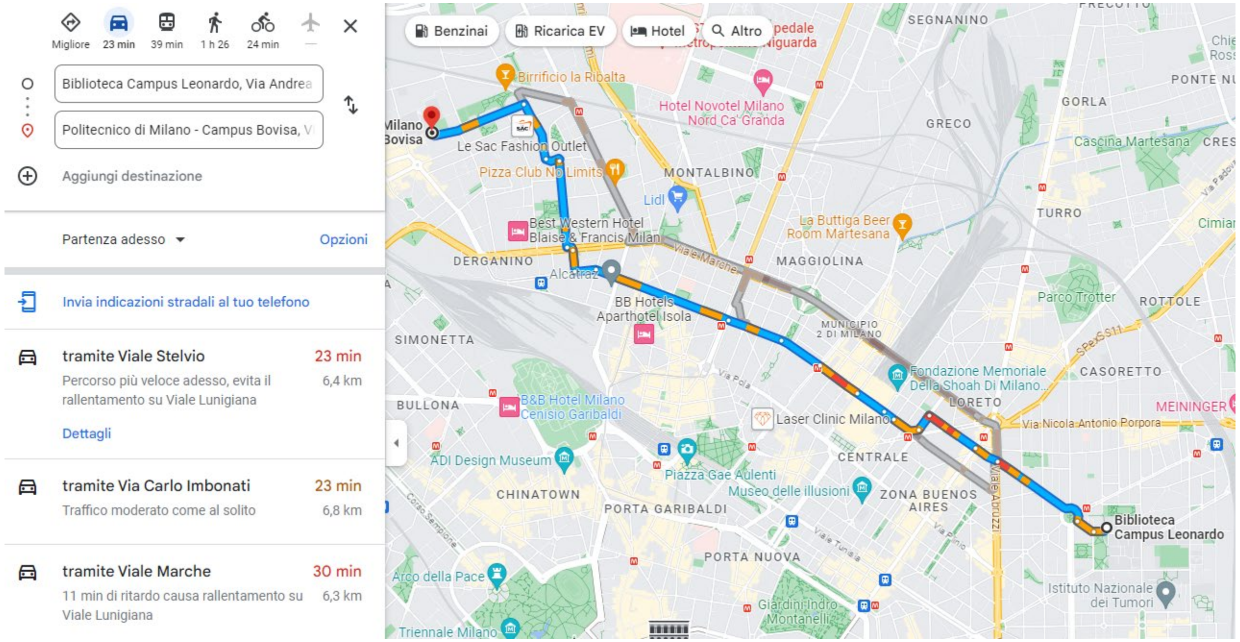
from mathematica.stackexchange.com

$$D = \max_{i,j} d_{ij} \qquad\qquad d = \langle d_{ij} \rangle = \frac{1}{N(N-1)} \sum_{i,j\,(i\neq j)} d_{ij}$$

The efficiency E can be used even if the network is not connected:

$$E = \langle \frac{1}{d_{ij}} \rangle = \frac{1}{N(N-1)} \sum_{i,j(i\neq j)} \frac{1}{d_{ij}}$$

by letting $1/d_{ij} = 0$ when the path $i \to j$ does not exist ($d_{ij} = \infty$).

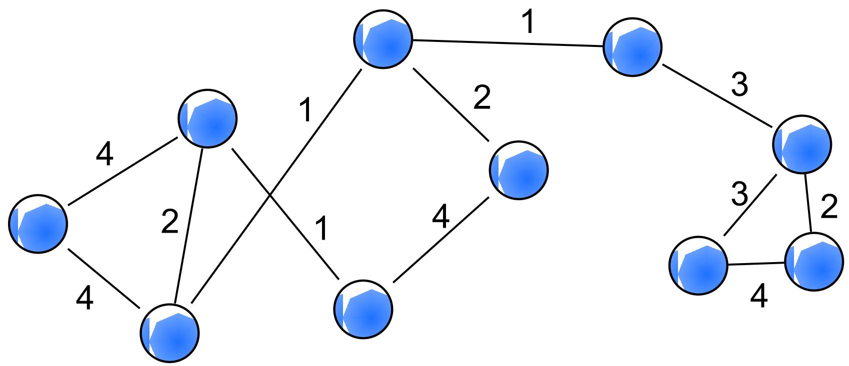


If the network is weighted, the shortest path connecting $i \rightarrow j$ is the path with the smallest total weight $d_{ij} = w_{ih_1} + w_{h_1 h_2} + \cdots + w_{h_n j}$.

Remark: contrary to the above, often a large weight models a strong relationship, affinity, proximity, etc. (messages through a social network, trade relationships, etc.).

d_{ij} should be the largest total weight connecting $i \rightarrow j$.

To use shortest path algorithms, replace $w_{ij} \rightarrow 1/w_{ij}$.



https://commons.wikimedia.org/w/index.php?curid=31358616

CLUSTERING (or TRANSITIVITY) COEFFICIENT

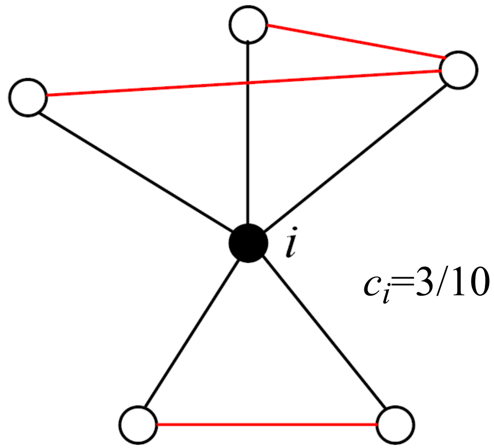
It quantifies the "local link density" by counting the triangles in the network.

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 each other?)*

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 $k_i > 1$ is the degree of i, and e_i the number of links directly connecting neighbours of i (at most $k_i(k_i - 1)/2$) [conventionally, $c_i = 0$ if $k_i \le 1$].

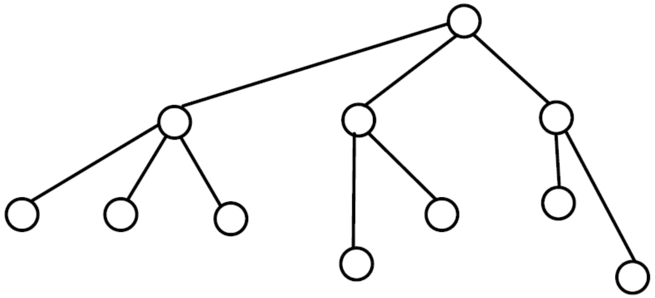


"ego-network" of node i

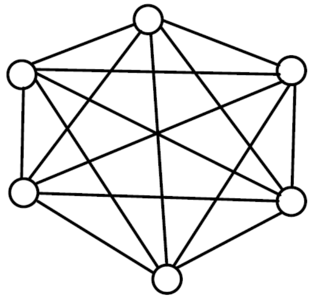
Large c_i, large (local) redundancy

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: $C = 0$



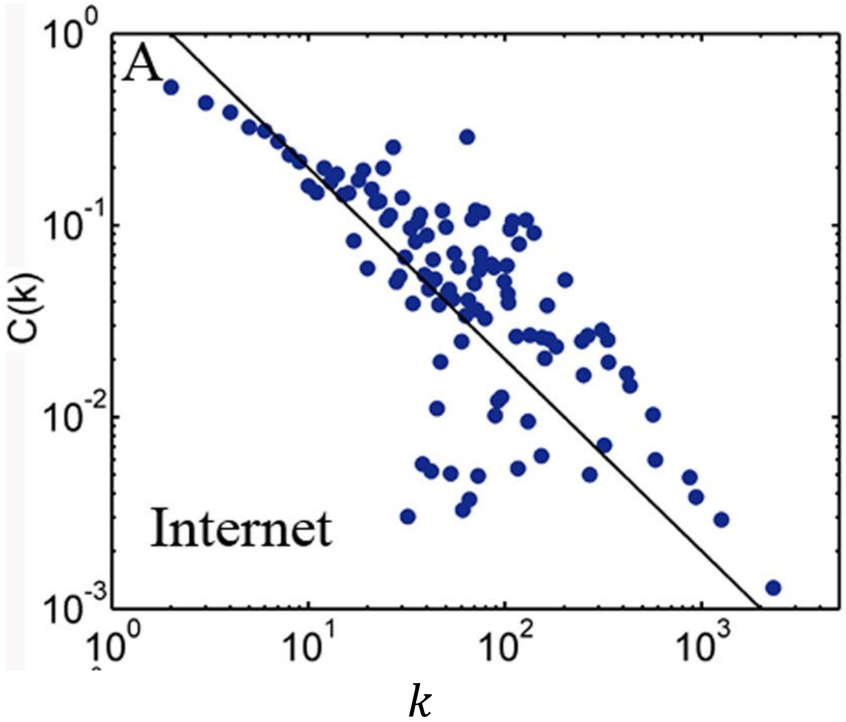
complete network: $C = 1$

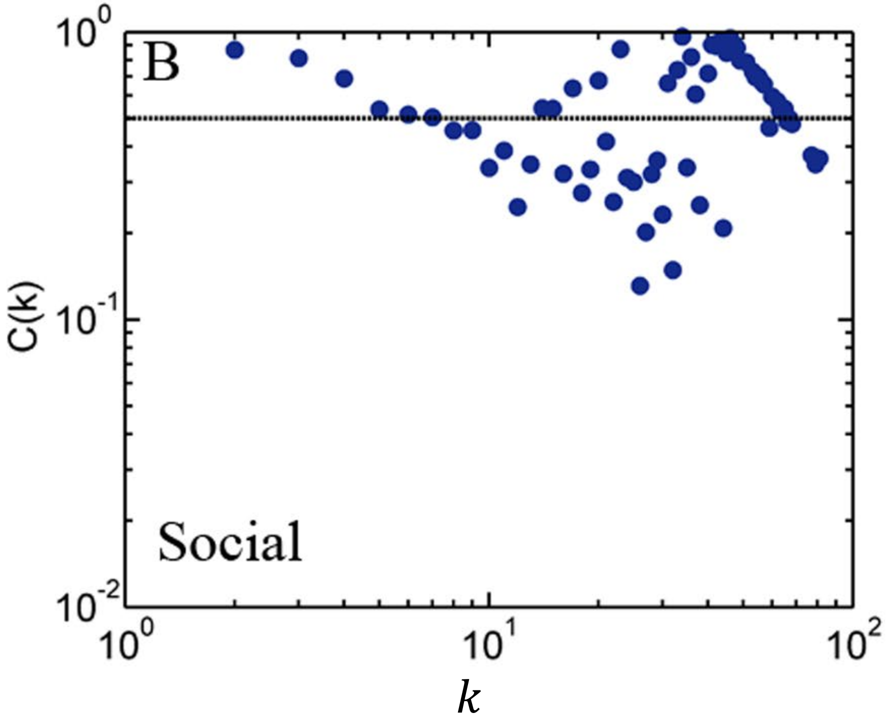
Network	Size	Clustering coefficient	Average path length
Internet, domain level [13]	32711	0.24	3.56
Internet, router level [13]	228298	0.03	9.51
WWW [14]	153127	0.11	3.1
E-mail [15]	56969	0.03	4.95
Software [16]	1376	0.06	6.39
Electronic circuits [17]	329	0.34	3.17
Language [18]	460902	0.437	2.67
Movie 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 $k_i = k$:

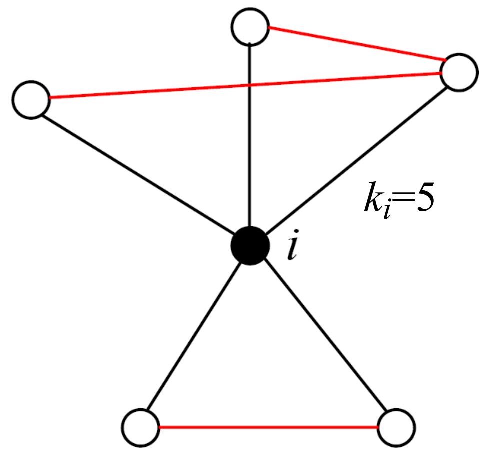
$$C(k) = \frac{1}{N_k} \sum_{i|k_i=k} c_i$$



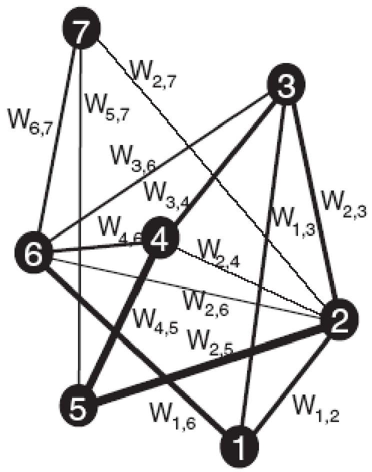


DEGREE AND STRENGTH OF A NODE

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:

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

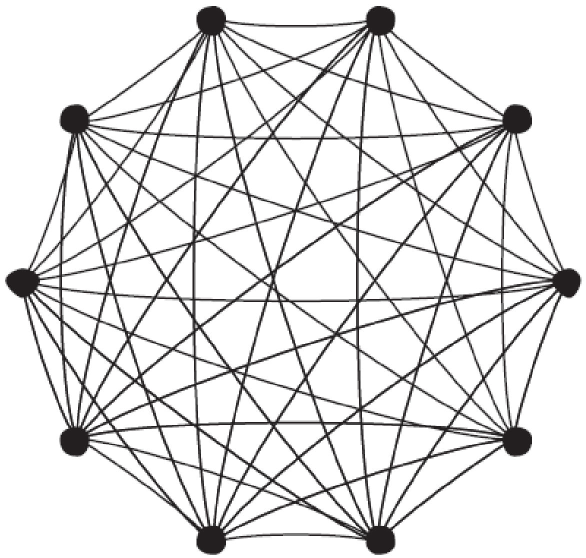
The r-moments of the degree distribution $P(k)$ are:

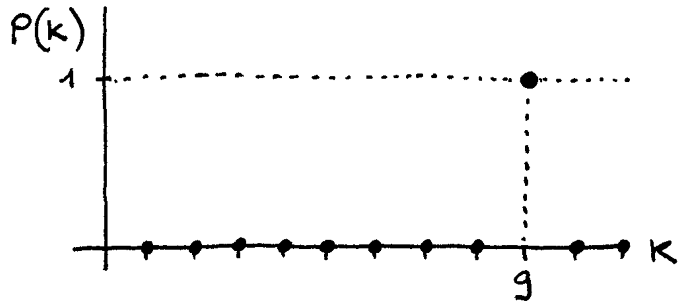
$$< k^r >= \sum_k k^r P(k) \quad , \quad r = 1,2,\ldots$$

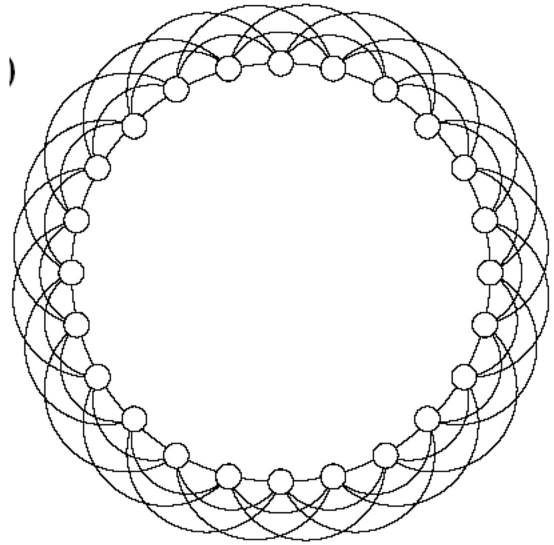
The first moment ($r = 1$) is the average degree $< k >= \sum_k kP(k) = \frac{1}{N}\sum_i k_i = \frac{2L}{N}$.

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

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

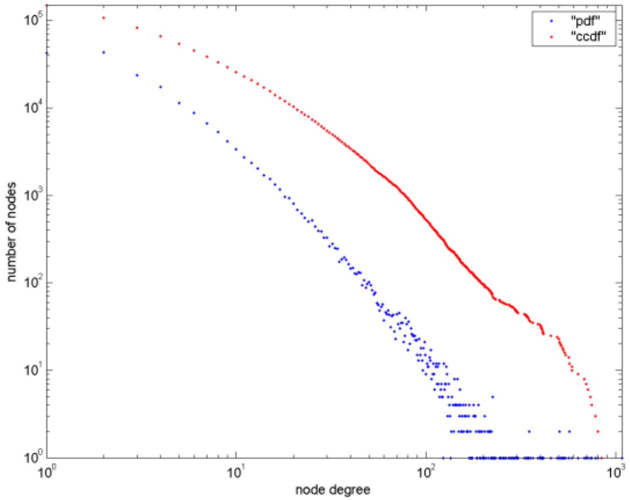






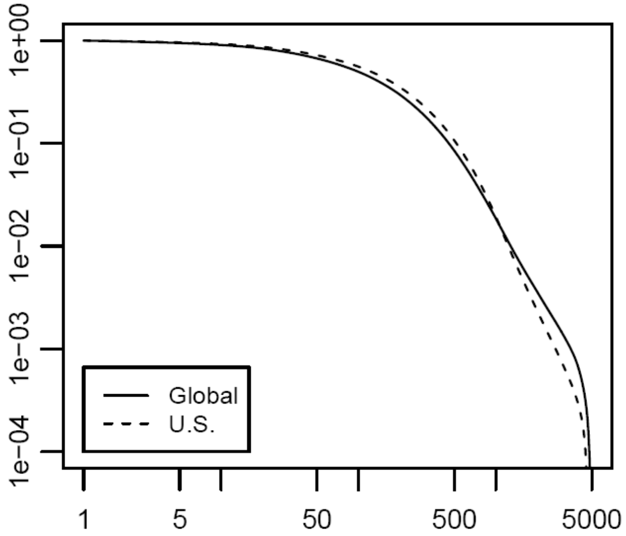
Example: a homogeneous network with $P(6) = 1$.

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

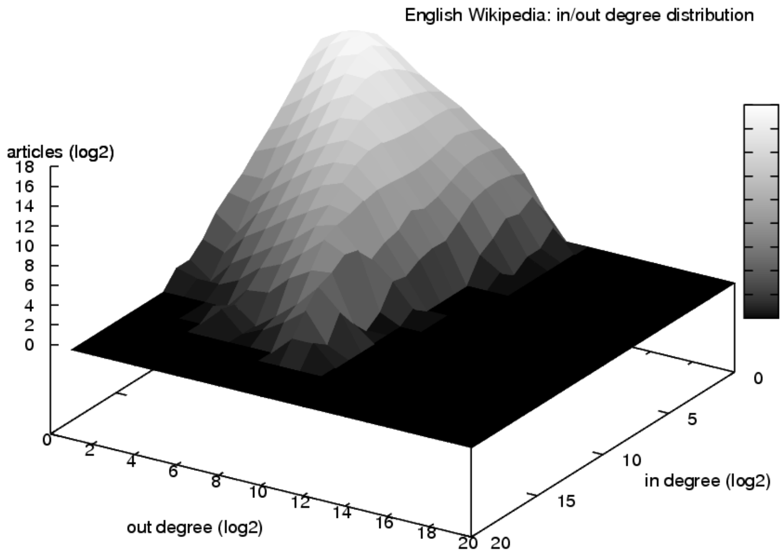


Internet

Facebook



Wikipedia (directed)



The degree distribution of nearest neighbours $Q(h)$ specifies the fraction of nodes' neighbours having exactly degree h (=the probability that a randomly selected link ends at a node of degree h):

It is not $P(k)$ but it is biased towards highest degrees:

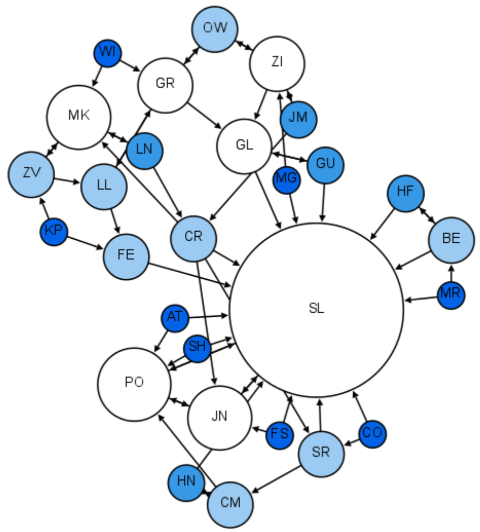
$$Q(h) = \frac{\text{n. of links from nodes of degree } h}{\text{n. of links from nodes of any degree}} = \frac{h(P(h)N)}{\sum_k k(P(k)N)} = \frac{hP(h)}{< k >}$$

Thus the average degree of nearest neighbours k_{nn} is:

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

which is larger than $< k >$ provided $\sigma^2 \neq 0$
(non strictly homogeneous network).

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



https://commons.wikimedia.org/wiki/File:Moreno_Sociogram_2nd_Grade.png

CORRELATED NETWORKS

A network is (degree-) 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:

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

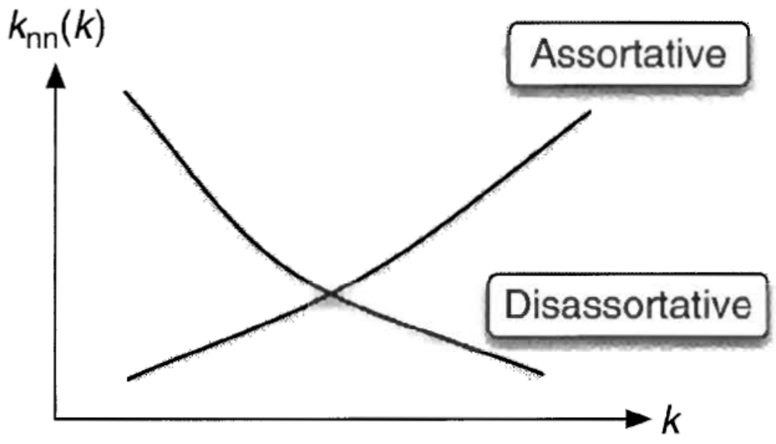
which in practice is computed by

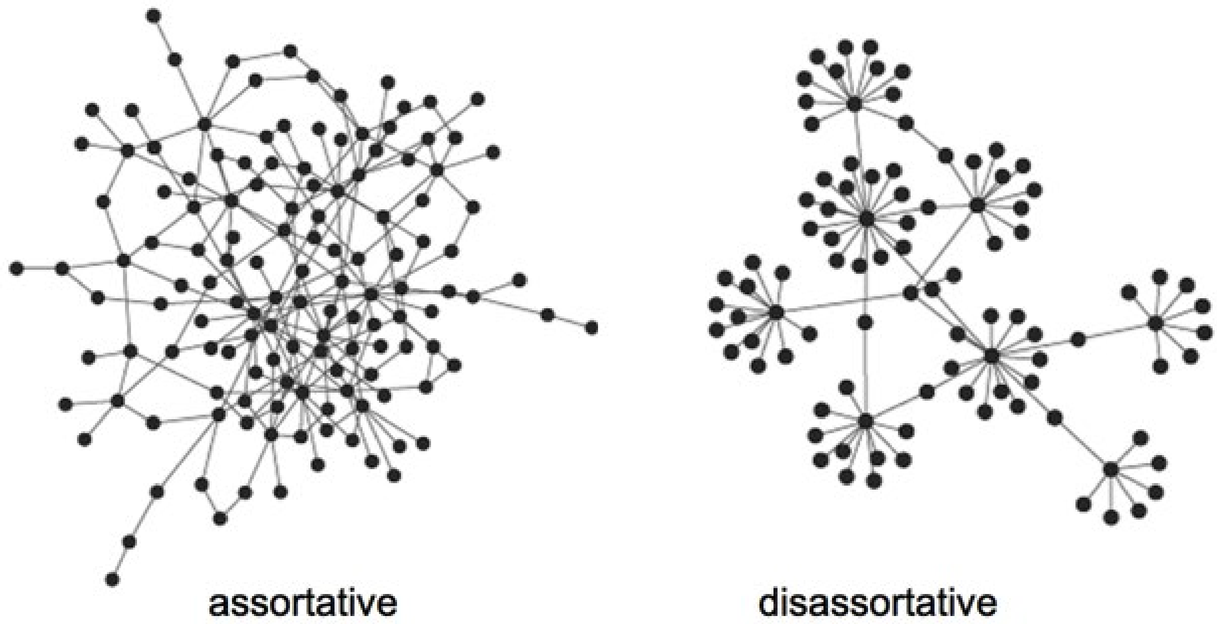
$$k_{nn}(k) = \frac{1}{N_k} \sum_{i|k_i=k} \frac{1}{k} \sum_j a_{ij} k_j$$

where N_k is the number of nodes with degree k ($\sum_k N_k = N$).

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].

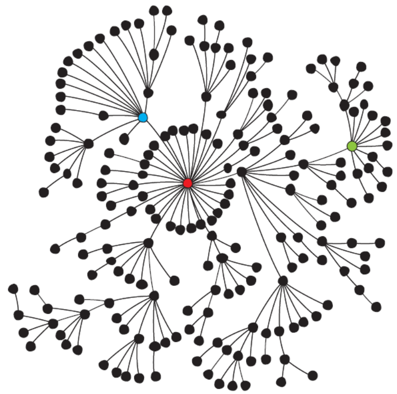




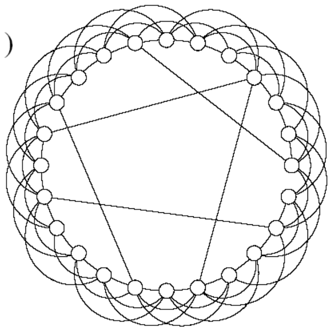
Hao, Li, Plos One, 2011

Two artificial networks with resp. assortative and disassortative patterns.

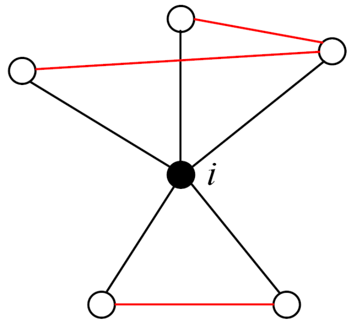
REMARK: The term "complex network" is mostly used to define graphs with non-trivial features, including:



- *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.