Eberhard O. Voit

A First Course in
Systems Biology

Chapter 3
Static Network Models

Graph properties

Graph G

N: node

$e(N_i, N_j)$: edge connecting nodes N_i and N_j

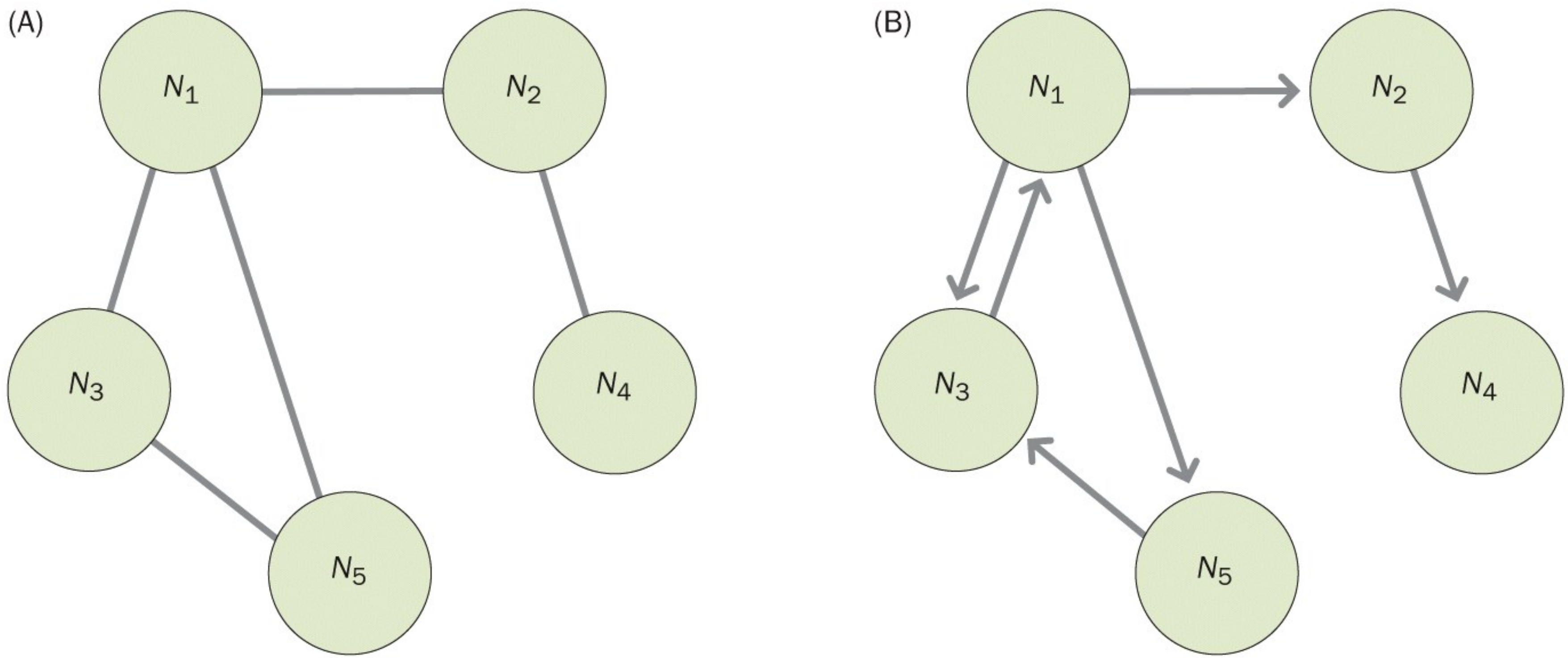


Figure 3.3 A First Course in Systems Biology (© Garland Science 2013)

Degree, $deg(N_i)$: # of linked nodes, or # of associated edges

Directed & undirected graphs; in- and out-degres

Degree of a node

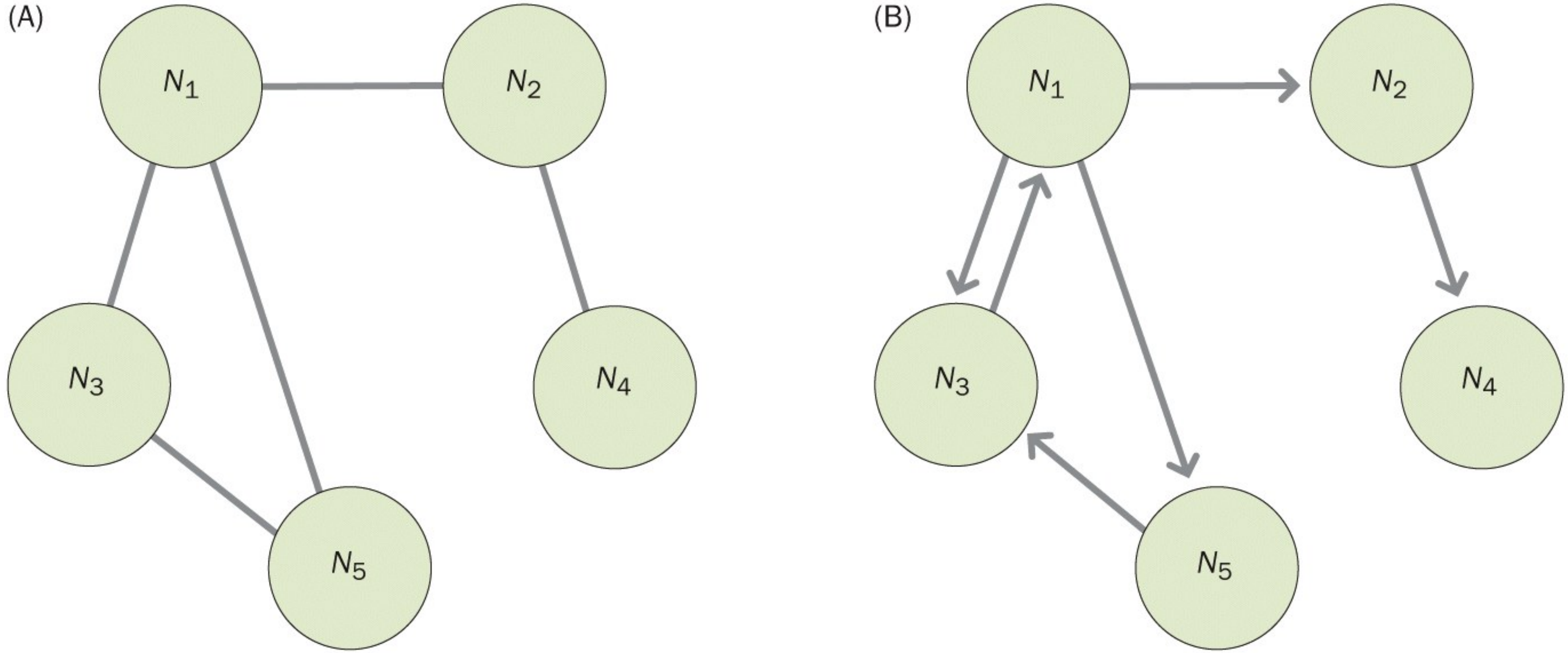


Figure 3.3 A First Course in Systems Biology (© Garland Science 2013)

<u>Undirected graph</u>

$\deg(N_1) = 3$

$\deg(N_2) = 2$

$\deg(N_3) = 2$

$\deg(N_4) = 1$

…

<u>Directed graph</u>

$\deg_{in}(N_1) = 1$

$\deg_{out}(N_1) = 3$

$\deg_{in}(N_2) = 1$

$\deg_{out}(N_2) = 1$

…

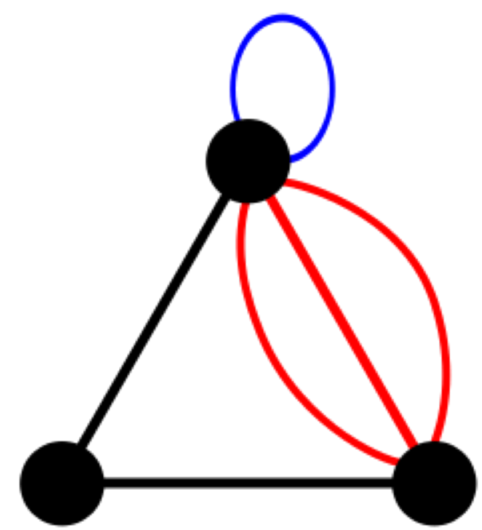
A weighted graph has edges with an additional property, a weight (can be integer or real value).

An edge colored graph has edges with an additional property, a color.

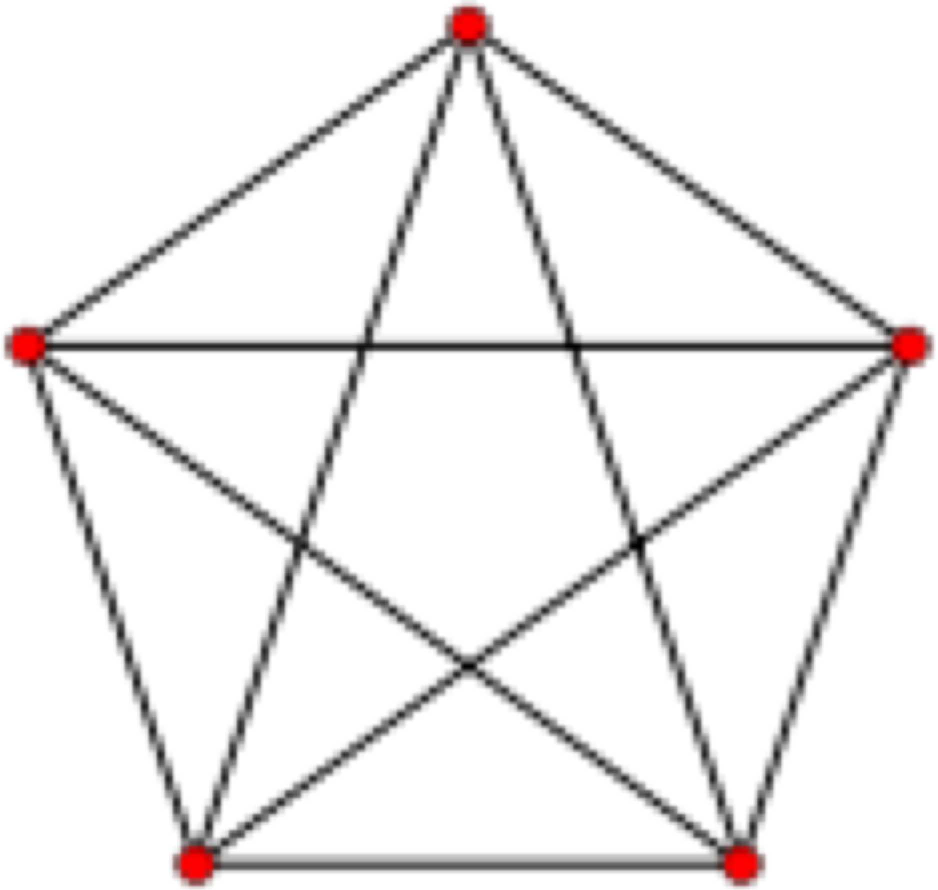
A subgraph of a graph G is a graph whose node & edge sets are subsets of those of G.

A general graph may have a self-loop which is an edge that goes from a node to itself, e.g. e(v3,v3).

A multi-graph may have more than one edge between the same pair of vertices. A directed graph is not a multi-graph, just two edges going different directions.



A graph is complete if there is an edge joining every pair of nodes



Clustering coefficient: a graph measure to determine whether or not a graph is a small-world network.

Small-world network: mathematical graph where most nodes can be reached by a small # of hops from all other nodes that are not neighbors.

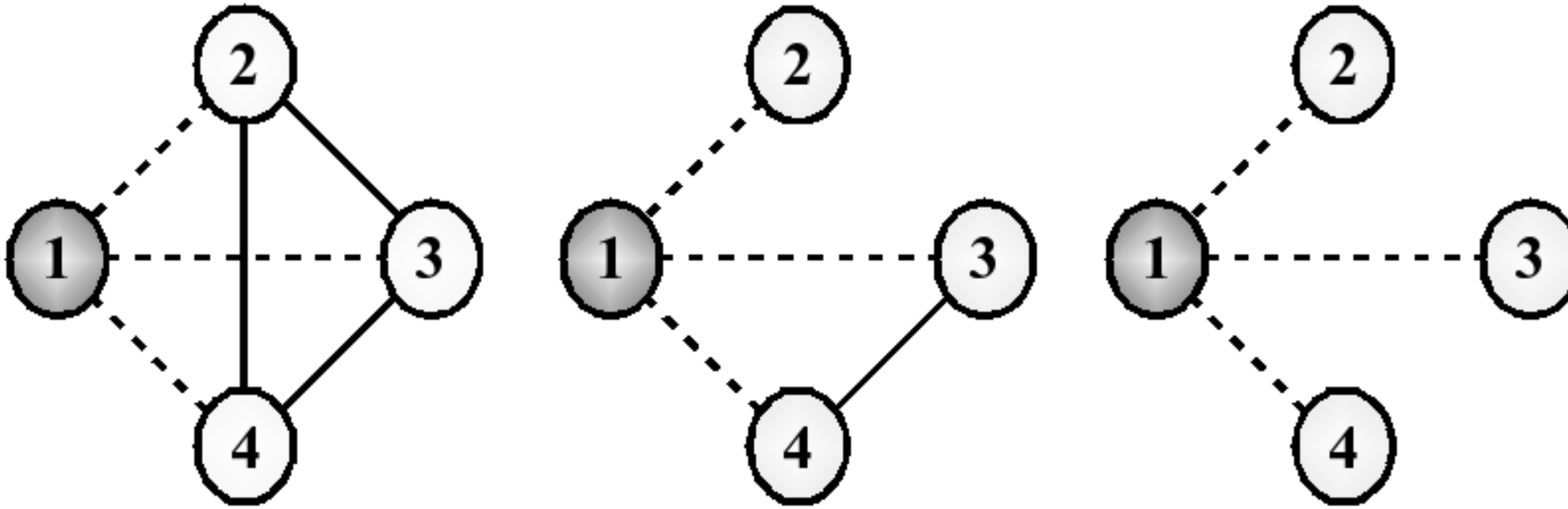
Example:

Consider the group formed by all your N friends.
The clustering coefficient describes whether your friends are also friends among themselves.

When your group of friends is a “clique”, all of its members will be friends with each other.

The clustering coefficient is defined by vertex i (you) & considers the connectivity of its neighbors (your friends) among each other.

Example illustrating the clustering coefficient for (shaded) vertex 1



c = 1 c = 1/3 c = 0

Fully-connected subgraph:
3 neighbor connections divided
by 3 connections of vertex 1
(or 3 possible triangles)

Dotted lines connect i to its neighbors; solid lines connect neighbors.

Clustering coefficient

C_N Describes how edges are statistically distributed within the graph G
or
characterizes the density of edges associated with a node N.

If deg(N) = k and if there are e edges among N's neighbors, then C_N is defined as

$$C_N = \frac{2e}{k(k-1)} \quad \text{(if G is undirected)}$$

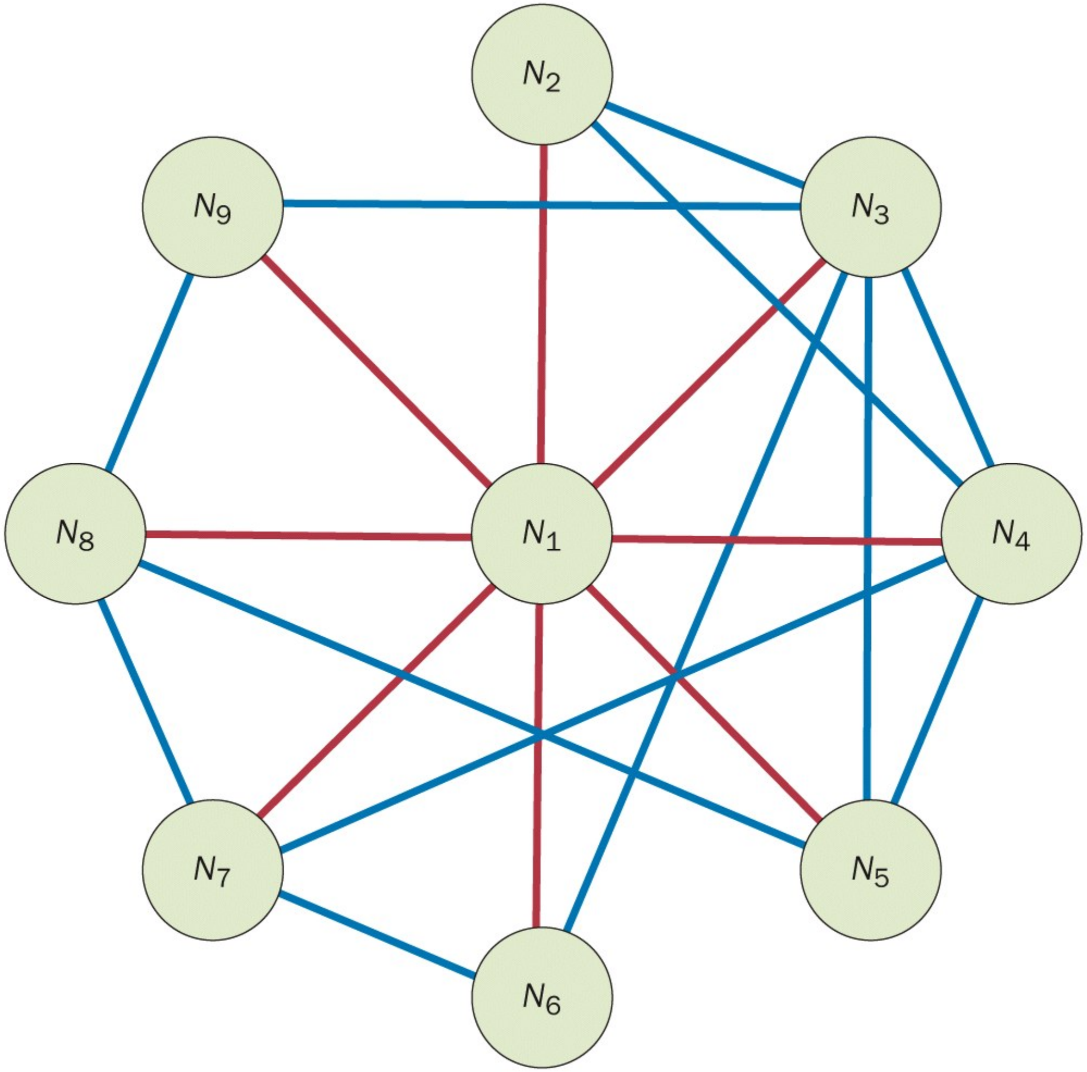
$$C_N = \frac{e}{k(k-1)} \quad \text{(if G is directed)}$$

Denominator: total possible number of edges among neighbors.

Example for calculation of clustering coefficient

Node N_1 has 8 neighbors (red arrows)

There are 12 connectivities among neighbors (blue arrows)



$$C_{N_1} = \frac{2e}{k(k-1)}$$

$$= \frac{2 \cdot 12}{8 \cdot 7} = 0.4286$$

Figure 3.5 A First Course in Systems Biology (© Garland Science 2013)

Average clustering coefficient of a graph

Overall measure of network's clustering: the arithmetic mean of the C_N for all nodes

$$C_G = \frac{1}{m}\sum_{N=1}^{m} C_N$$

Can be used as another measure of network topology

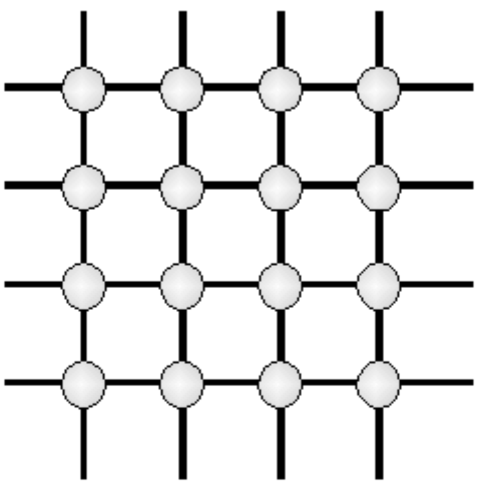
Degree distribution

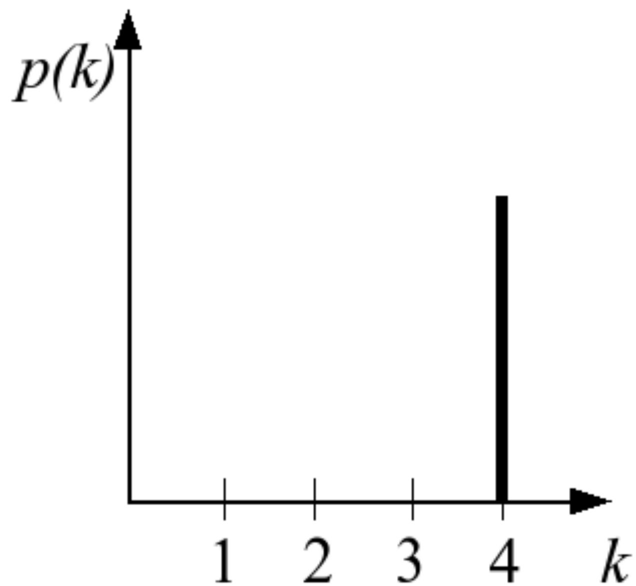
Degree of a vertex specifies the # of edges by which it is connected to other vertices

Degree distribution of a graph is a function measuring the total # of vertices in the graph with a given degree

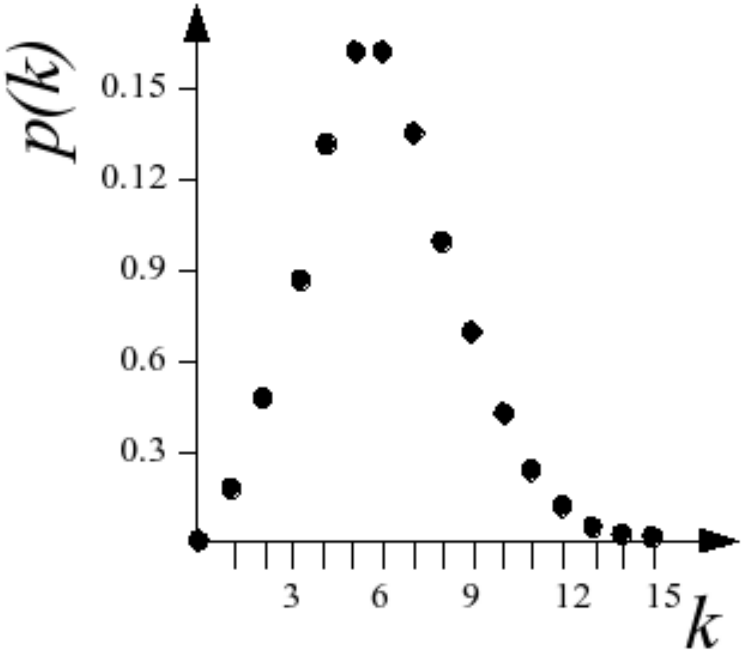
$$p(k) = \frac{1}{N} \sum_{v_i \in V | deg(v_i) = k} 1$$

Counts how many vertices have degree k

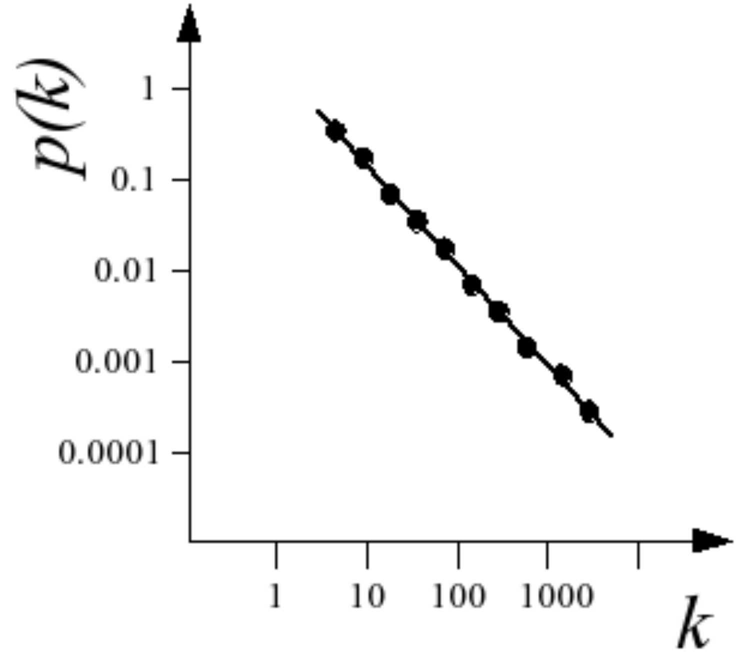




The degree distribution is a way to classify graphs into categories



Random network:
Poisson distribution



Scale-free network:
Power law → linear distribution log-log plot

Remember scale-free networks

Scale-free means that while the vast majority of vertices are weakly connected, there also exist some highly inter-connected super-vertices or hubs

The term scale-free expresses that the ratio of highly to weakly connected vertices remains the same irrespective of the total # of links in the network

Scale-free network is very simple, elegant & intuitive

To produce an artificial scale-free network possessing small-world properties, only 2 basic rules need to be followed:
Growth: the network is seeded with a small # of vertices
Preferential attachment: the more connectivities a vertex has, the more likely new vertices will be connected to it (equivalent to the rich become richer!)

Power law

A power law is any polynomial relationship that exhibits the property of **scale invariance**. The most common power laws relate two scalar quantities and have the form

$$f(x) = ax^k$$

where *a* and *k* are constants. *k* is typically called the *scaling exponent*, where the word "scaling" denotes the fact that a power-law function satisfies

$$f(cx) \propto f(x)$$

where *c* is a constant. Thus, a rescaling of the function's argument changes the constant of proportionality but preserves the shape of the function itself. Take the logarithm of both sides to clarify

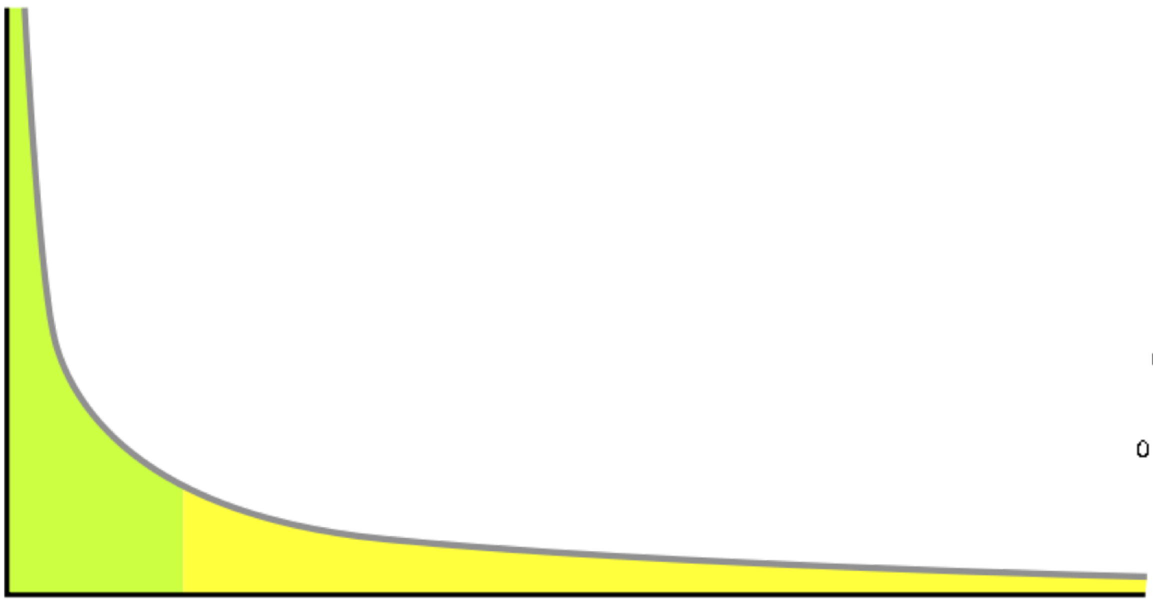
$$log[f(x)] = k\, log\, x + log\, a$$

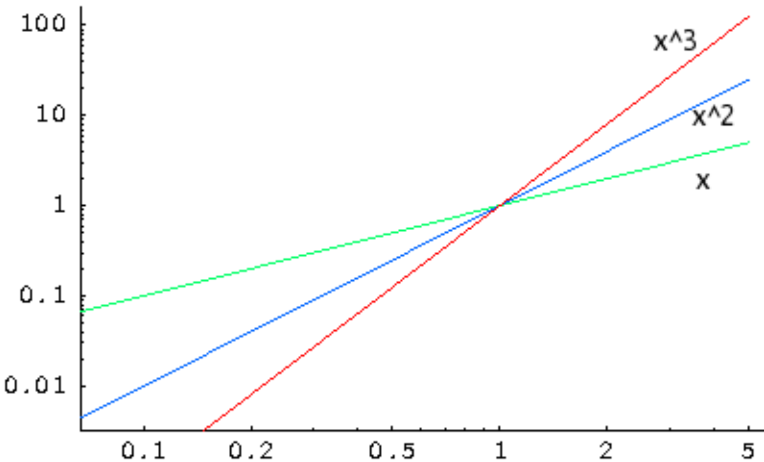
This expression has the form of a linear relationship with slope *k*. Rescaling the argument produces linear shift of the function up or down but leaves both the basic form and the slope *k* unchanged.

$$log[f(cx)] = k\, log\, x + k\, log\, c + log\, a$$

www.wikipedia.org

Power law





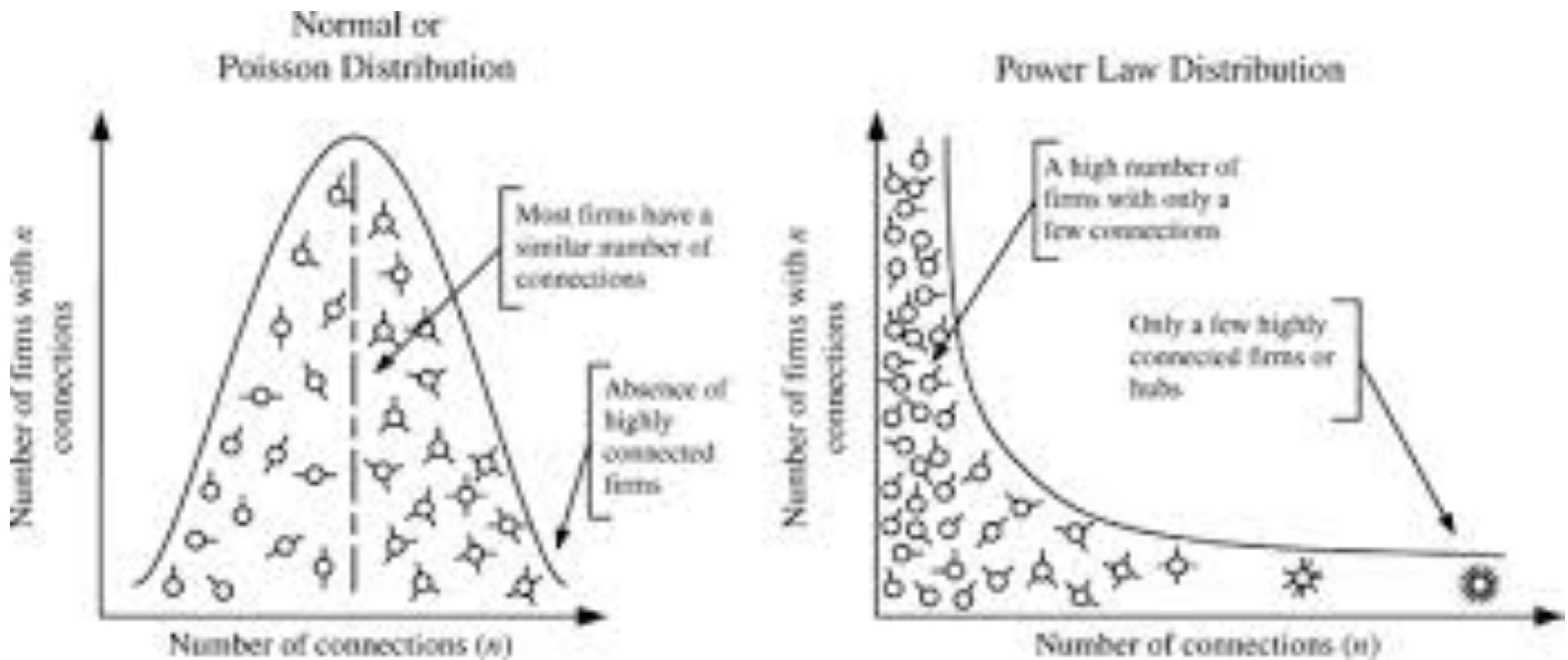
log-log plot

An example power law graph, being used to demonstrate ranking of popularity. To the right is the long tail, to the left are the few that dominate.

Also known as the 80-20 rule: for many events, roughly 80% of the effects come from 20% of the causes.

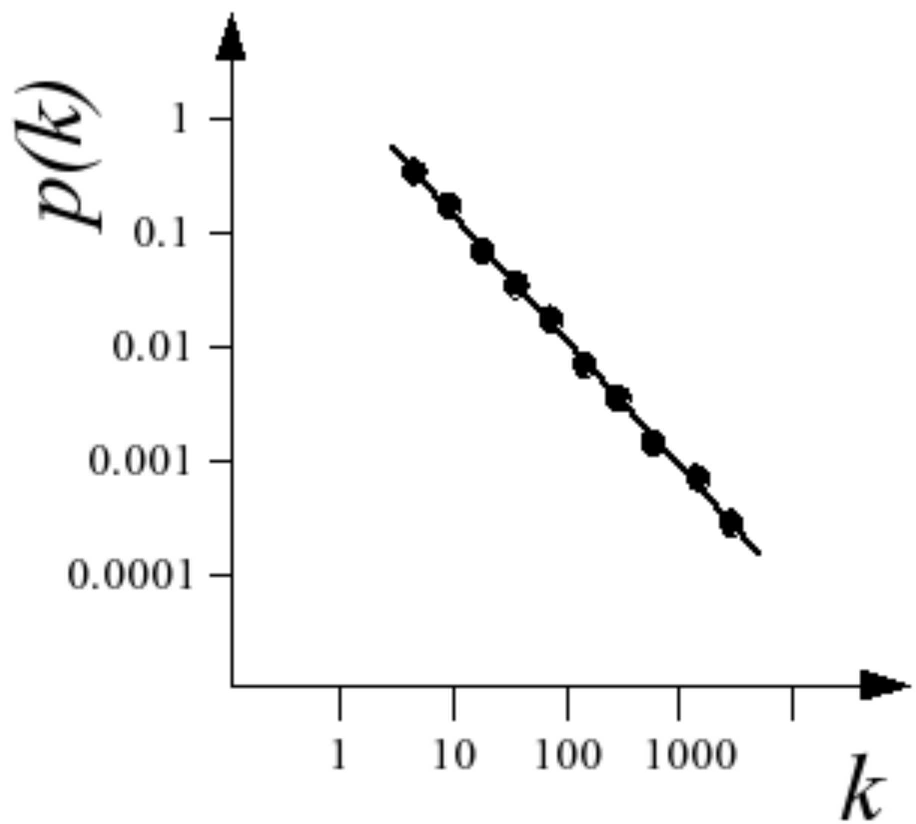
www.wikipedia.org

P(k) decays much slower for large k than the exponential decay of a Poisson distribution: highly connected hubs occur at much larger frequency than expected in random graphs.



In scale-free networks, the degree distribution follows a power law:

$$p(k) = k^{-\gamma}, \quad \gamma \in \Re^{+}$$



Degree distribution of a graph

Measures the proportion of nodes in G that have degree k.

$$P(k) = \frac{m_k}{m}$$

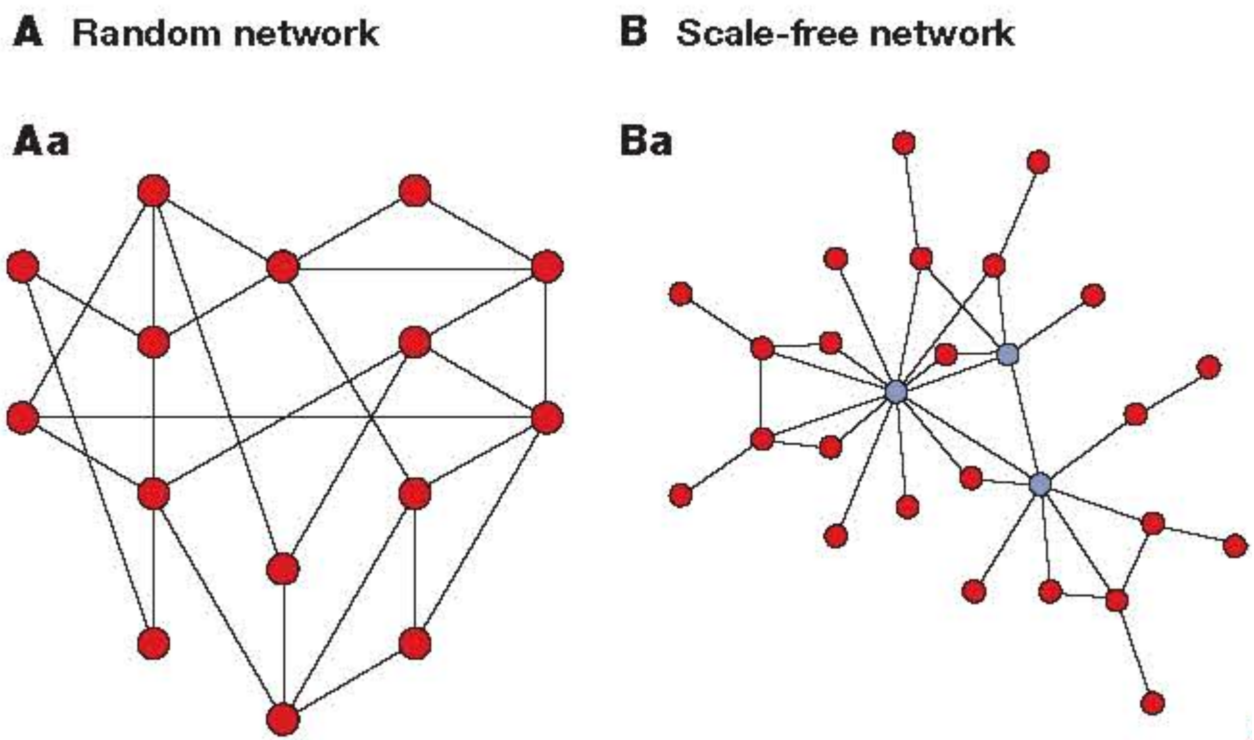
m_k is the number of nodes of degree k and m is the total number of all nodes

P(k) is not a single number, but a set of numbers, with one degree value for each k (k = 0, 1, 2, . . .)

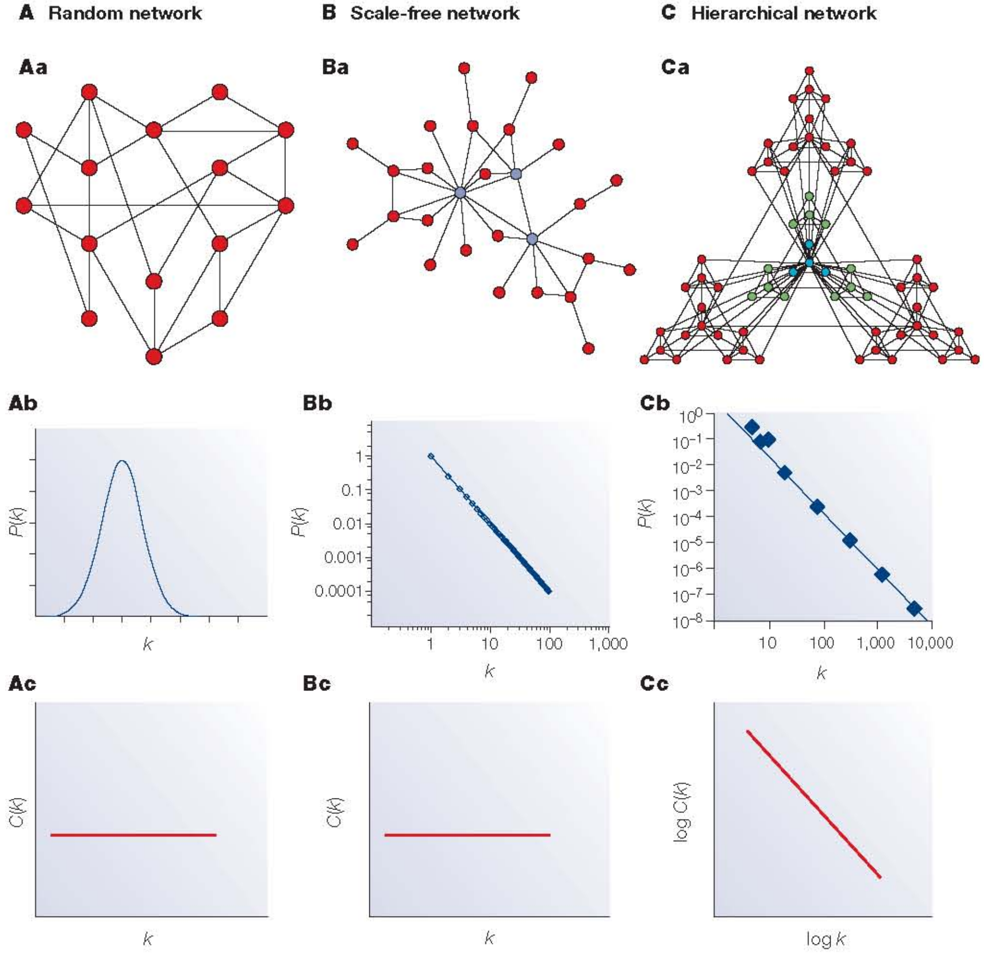
Division by m assures that the sum of P(k) over all values of k is equal to 1

Characterizations of graphs (degree distributions, clustering coefficients, etc) are performed using comparisons with random graphs.

Random graph: edges are randomly associated with nodes → degree distribution is a Poisson distribution (skinny bell curve with small variance), most nodes have a degree that is close to average.



Summary: network models



A hierarchical architecture implies that sparsely connected areas, with communication between the different highly clustered neighborhoods being maintained by a few nodes.