Complex Networks II
Basic Definitions

Plan of the lectures

I. Introduction

II. Networks: basic definitions

III. Models

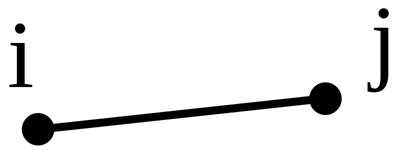
IV. Community Detection

Graph theory: basics

Graph G=(V,E)

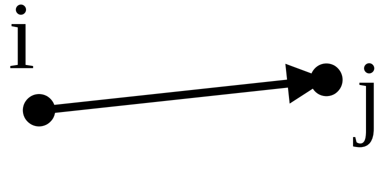
- V=set of nodes/vertices i=1,…,N
- E=set of links/edges (i,j)

Undirected edge:



Bidirectional communication/ interaction

Directed edge:

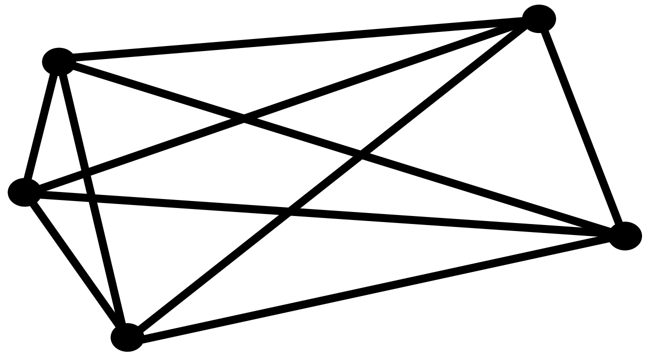


Graph theory: basics

Maximum number of edges

- Undirected: N(N-1)/2
- Directed: N(N-1)

Complete graph:



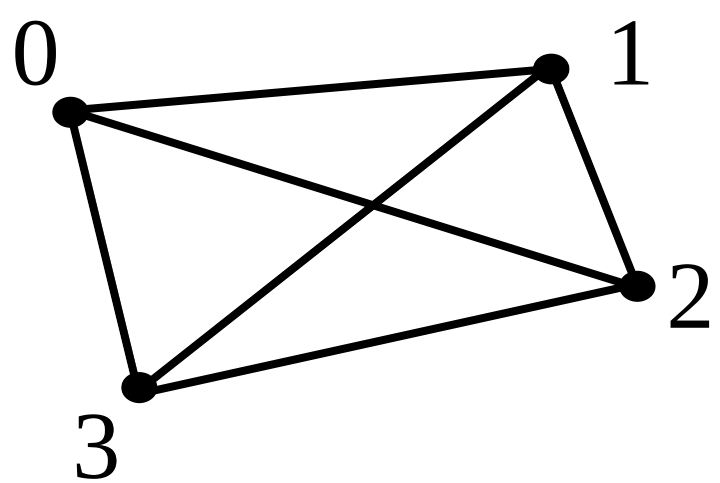
(all to all interaction/communication)

Adjacency matrix

N nodes i=1,…,N

$$a_{ij} = \begin{cases} 1 \text{ if } (i,j) \ \hat{I} \ E \\ 0 \text{ if } (i,j) \notin E \end{cases}$$

	0	1	2	3
0	0	1	1	1
1	1	0	1	1
2	1	1	0	1
3	1	1	1	0



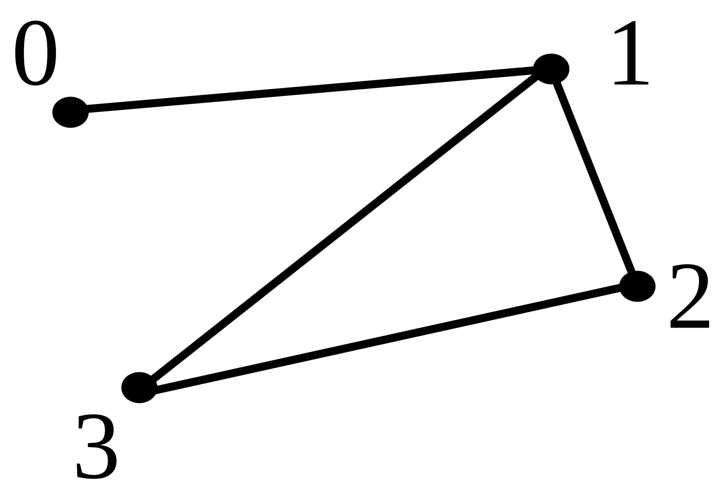
Adjacency matrix

N nodes i=1,....,N

$$a_{ij} = \begin{cases} 1 \text{ if } (i,j) \in E \\ 0 \text{ if } (i,j) \notin E \end{cases}$$

Symmetric
for undirected networks

	0	1	2	3
0	0	1	0	0
1	1	0	1	1
2	0	1	0	1
3	0	1	1	0



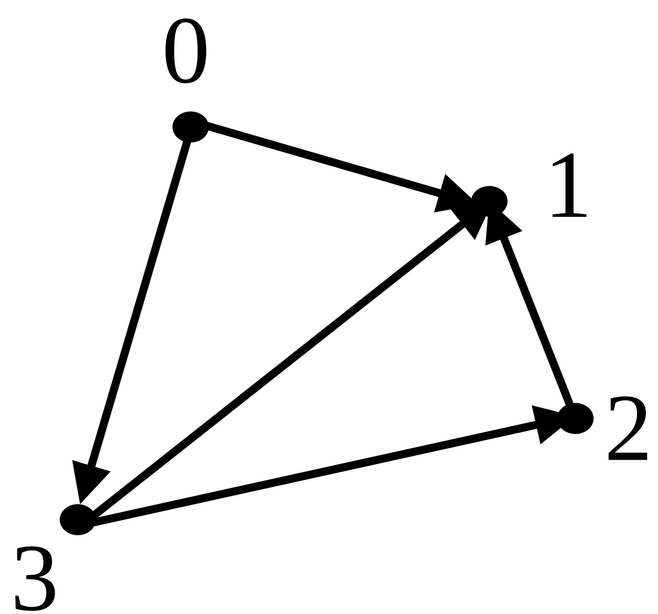
Adjacency matrix

N nodes i=1,…,N

$$a_{ij} = \begin{cases} 1 \text{ if } (i,j) \in E \\ 0 \text{ if } (i,j) \notin E \end{cases}$$

Non symmetric
for directed networks

	0	1	2	3
0	0	1	0	1
1	0	0	0	0
2	0	1	0	0
3	0	1	1	0



Sparse graphs

Density of a graph $D=|E|/(N(N-1)/2)$

$$D = \frac{\text{Number of edges}}{\text{Maximal number of edges}}$$

Sparse graph: $D \ll 1$ → Sparse adjacency matrix

→ Representation: lists of neighbours of each node

$$l(i, V(i))$$

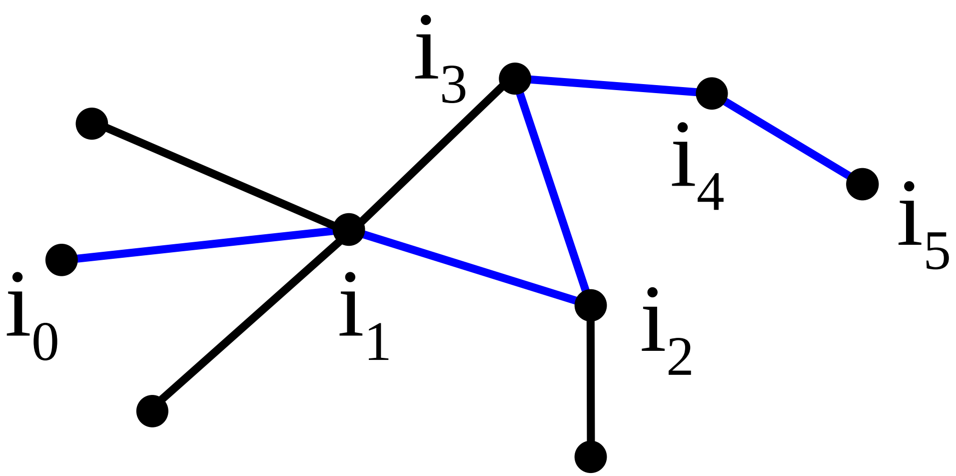
$V(i)$=neighbourhood of i

Paths

G=(V,E)

Path of length n = ordered collection of

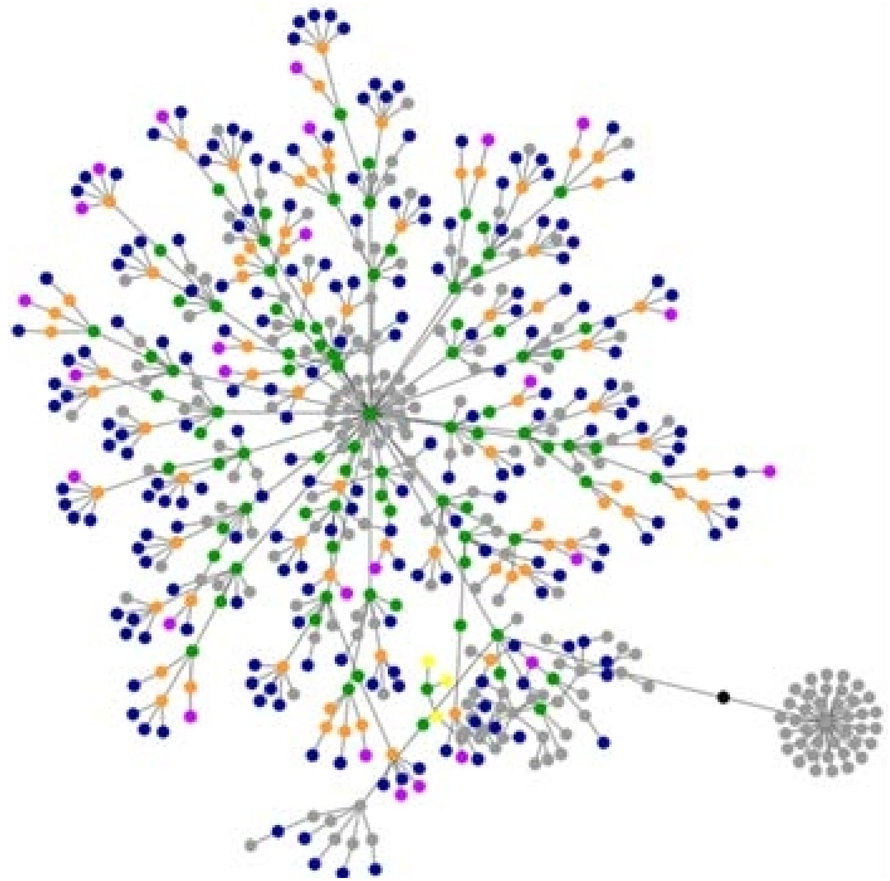
- n+1 vertices $i_0, i_1, \ldots, i_n$ Î V
- n edges $(i_0,i_1), (i_1,i_2)\ldots,(i_{n-1},i_n)$ Î E



Cycle/loop = closed path ($i_0=i_n$)

Trees

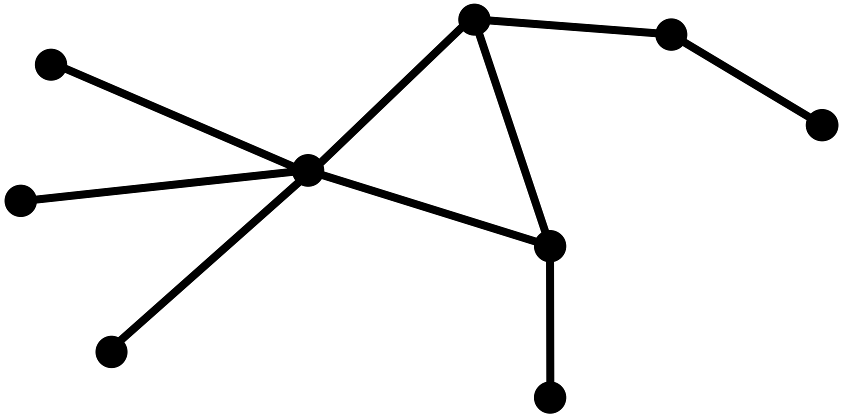
A tree is a graph without loops/cycles



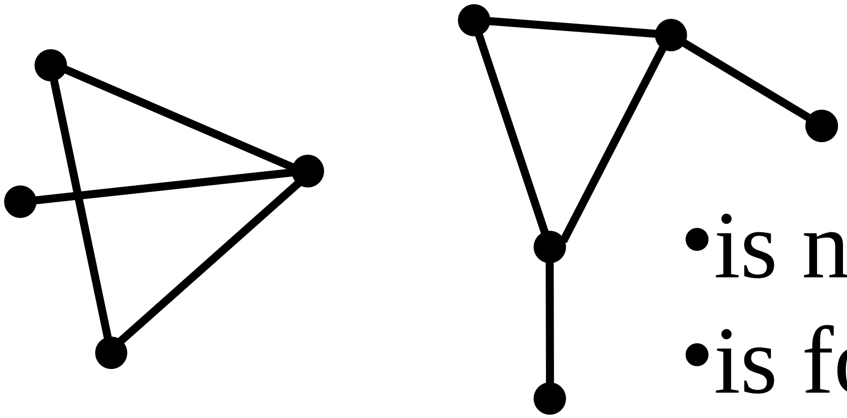
- N nodes, N-1 links
- Maximal loopless graph
- Minimal connected graph

Paths and connectedness

G=(V,E) is connected if and only if there exists a path connecting any two nodes in G



is connected



- is not connected
- is formed by two components

Paths and connectedness

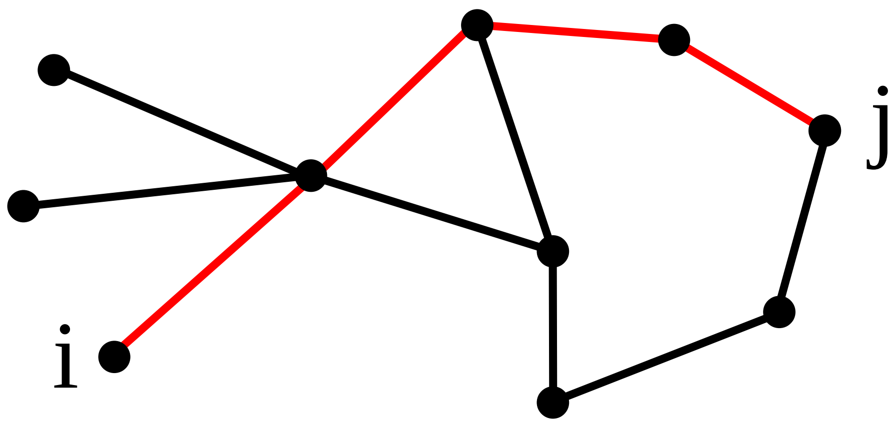
G=(V,E)=> distribution of components' sizes

Giant component= component whose size scales with the number of vertices N

Existence of a giant component $\Longleftrightarrow$ Macroscopic fraction of the graph is connected

Shortest paths

Shortest path between i and j: minimum number of traversed edges



distance $l(i,j)$=minimum number of edges traversed on a path between i and j

Diameter of the graph= $\max(l(i,j))$

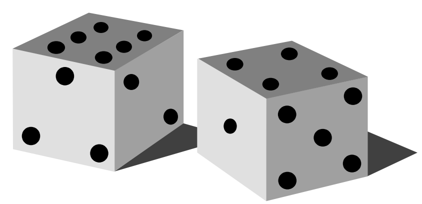
Average shortest path= $\sum_{ij} l(i,j)/(N(N-1)/2)$

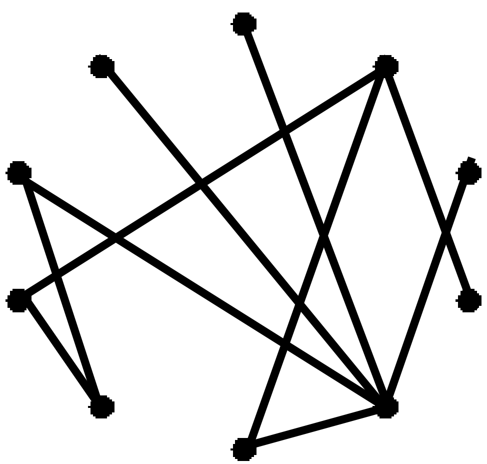
Complete graph: $l(i,j)=1$ for all i,j

“Small-world”: “small” diameter

Small-world

N points, links with probability p:
static random graphs

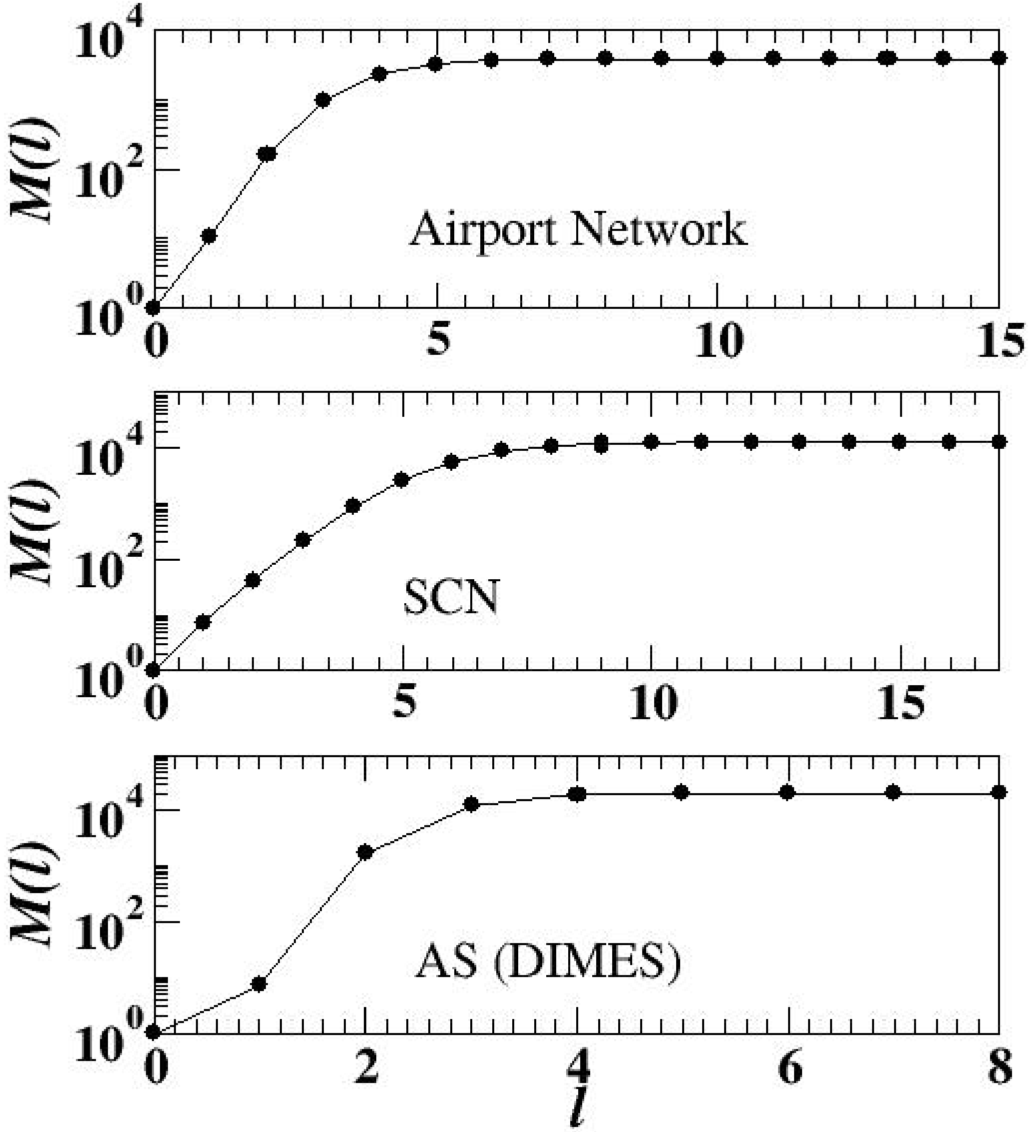




short distances

(log N)

Small-world



Average number of nodes within a distance l

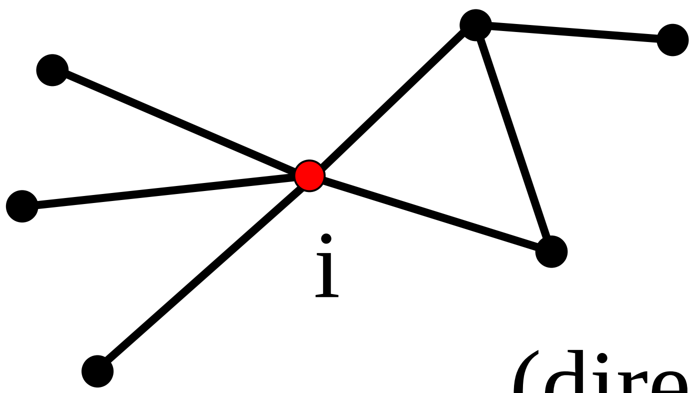
Scientific collaborations

Internet

Centrality measures

How to quantify the importance of a node?

- Degree=number of neighbours=$\Sigma_j\ a_{ij}$



$k_i=5$

(directed graphs: k_{in}, k_{out})

Betweenness centrality

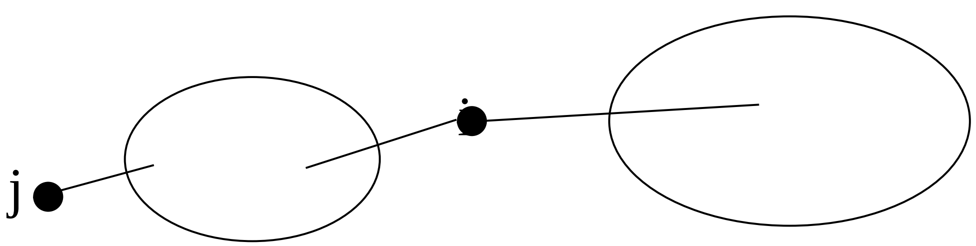
for each pair of nodes (l,m) in the graph, there are

σ^{lm} shortest paths between l and m

σ_i^{lm} shortest paths going through i

b_i is the sum of $\sigma_i^{lm} / \sigma^{lm}$ over all pairs (l,m)

path-based quantity



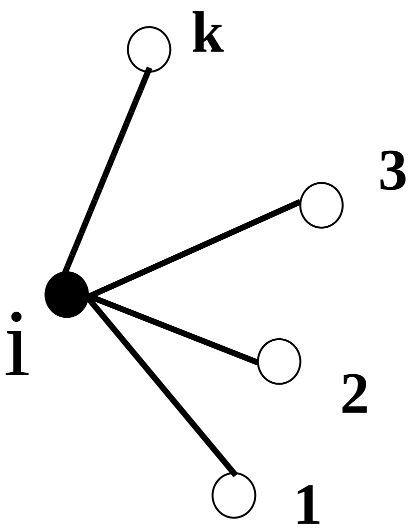
b_i is large

b_j is small

NB: similar quantity= **load $l_i = \sum \sigma_i^{lm}$**

NB: generalization to *edge betweenness centrality*

Structure of neighborhoods



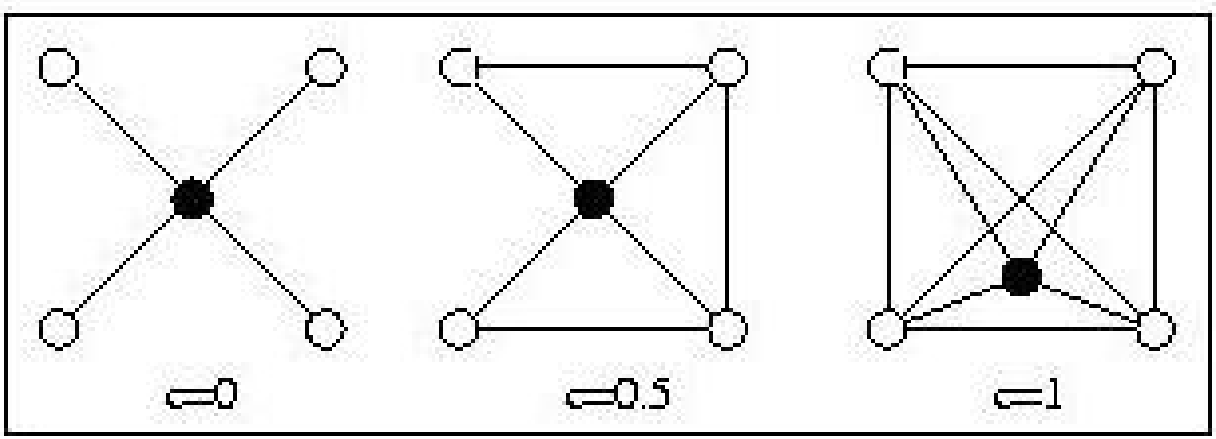
Clustering coefficient of a node

$$C(i) = \frac{\text{\# of links between 1,2,…n neighbors}}{k(k-1)/2}$$

$$C(i) = \frac{1}{k_i(k_i-1)} \sum_{j \neq k} a_{ij} a_{jk} a_{ik}$$

Clustering: My friends will know each other with high probability!
(typical example: social networks)

c)



Statistical characterization

Degree distribution

- List of degrees $k_1, k_2, \ldots, k_N$ ← Not very useful!
- Histogram:

  N_k = number of nodes with degree k
- Distribution:

  $P(k)=N_k/N$=probability that a randomly chosen node has degree k
- Cumulative distribution:

  $P^{>}(k)$=probability that a randomly chosen node has degree at least k

Statistical characterization

Degree distribution

$P(k)=N_k/N=$probability that a randomly chosen node has degree k

Average$= \langle k \rangle = \sum_i k_i/N = \sum_k k\, P(k)=2|E|/N$

Sparse graphs: $\langle k \rangle \ll N$

Fluctuations: $\langle k^2 \rangle - \langle k \rangle^2$

$\langle k^2 \rangle = \sum_i k_i^2/N = \sum_k k^2\, P(k)$

$\langle k^n \rangle = \sum_k k^n\, P(k)$

Topological heterogeneity

Statistical analysis of centrality measures:

$P(k)=N_k/N=$probability that a randomly chosen node has degree k

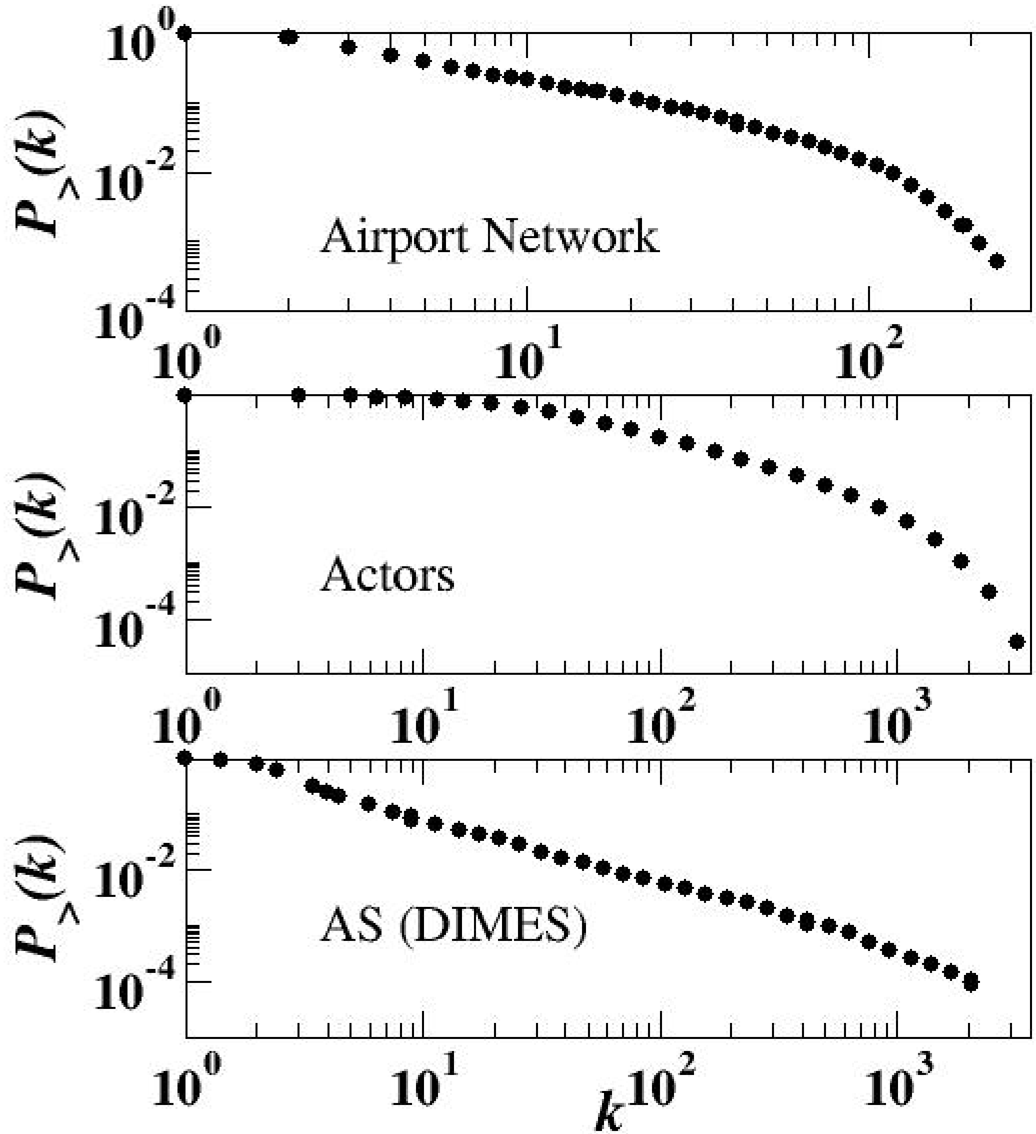
also: P(b), P(c)….

Two broad classes

- homogeneous networks: light tails
- heterogeneous networks: skewed, heavy tails

Topological heterogeneity

Statistical analysis of centrality measures

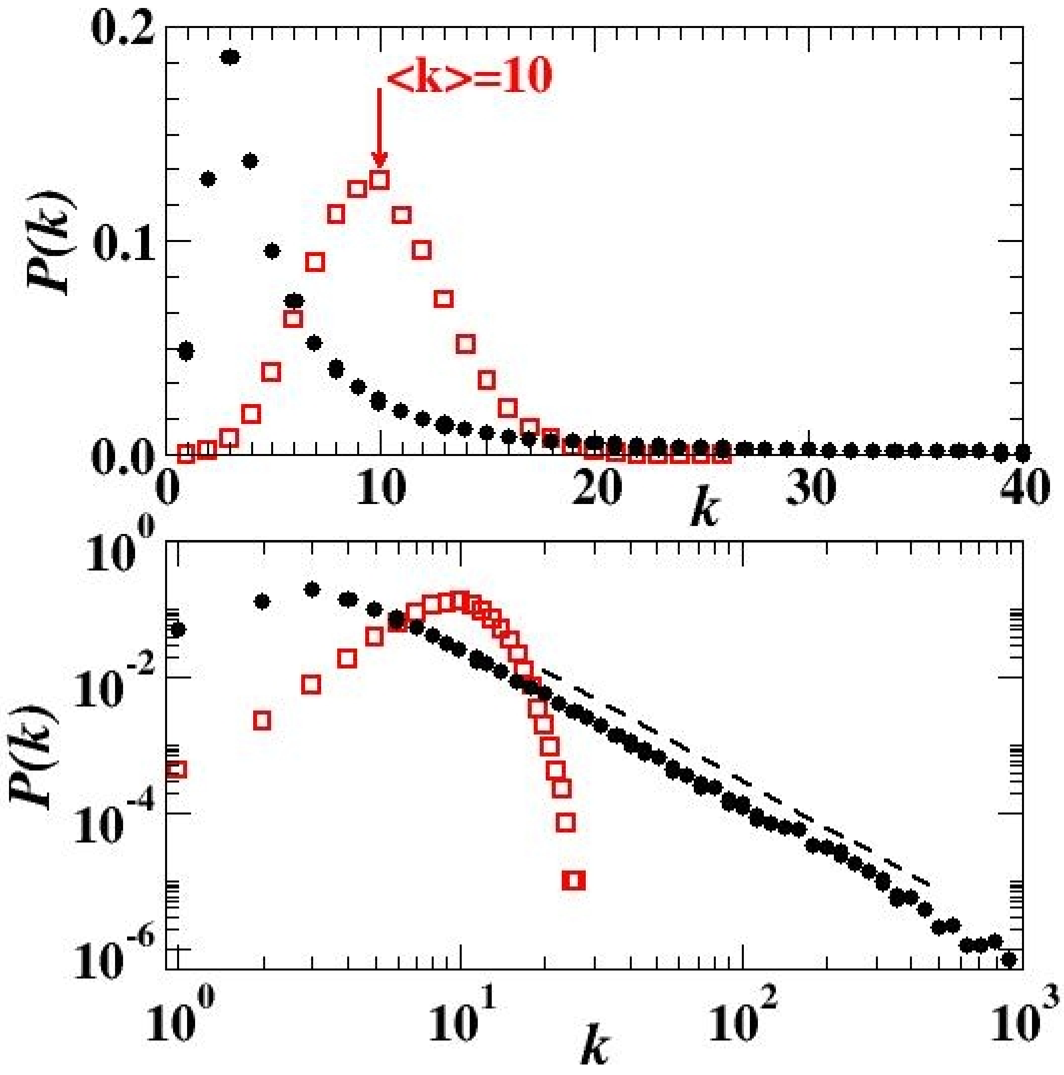


Broad degree distributions

Power-law tails $P(k) \sim k^{-\gamma}$, typically $2 < \gamma < 3$

Topological heterogeneity

Statistical analysis of centrality measures:



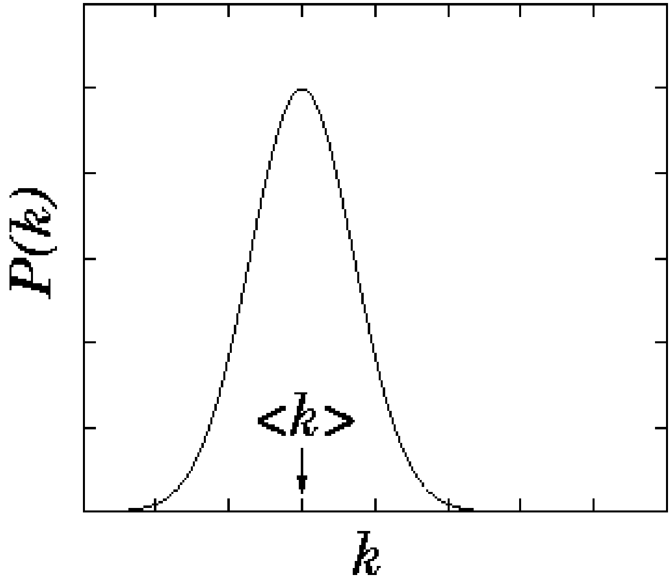
linear scale

Poisson
vs.
Power-law

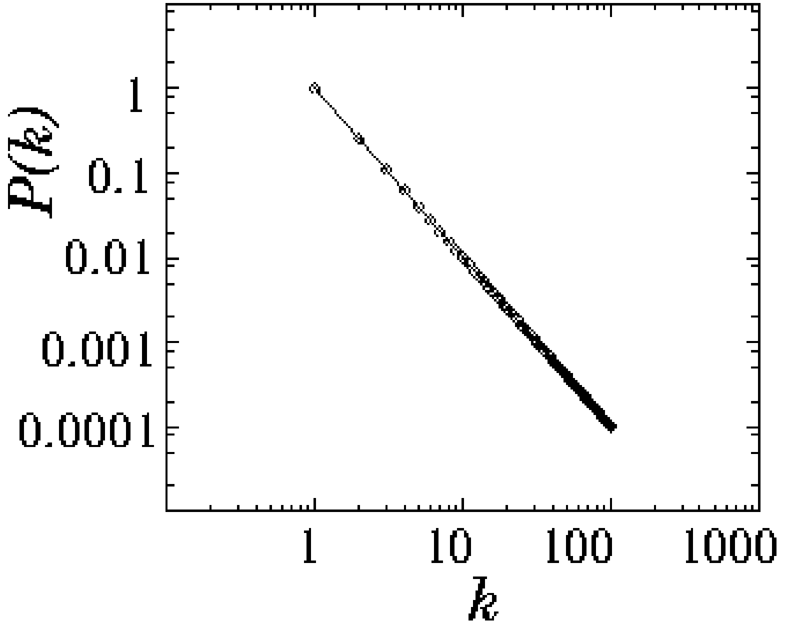
log-scale

Exp. vs. Scale-Free

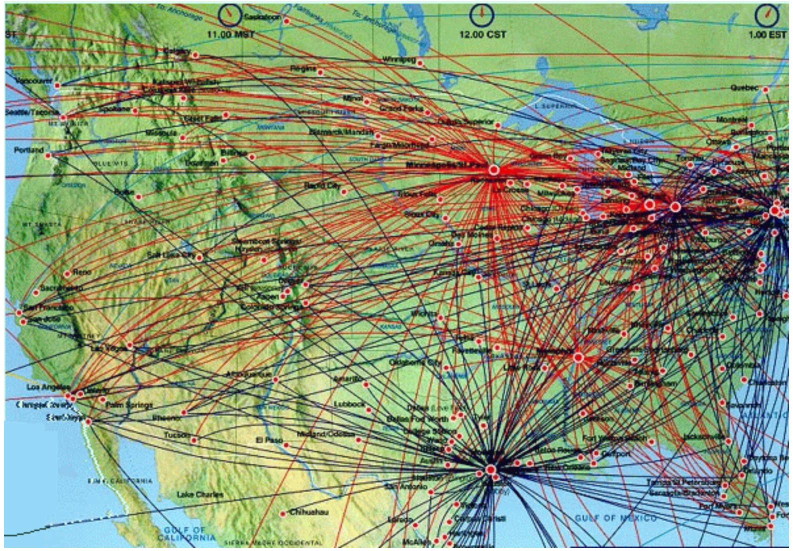
Poisson distribution



Power-law distribution



Exponential



Scale-free

Consequences

Power-law tails

$P(k) \sim k^{-\gamma}$

Average $= \langle k \rangle = \int k\, P(k) dk$

Fluctuations

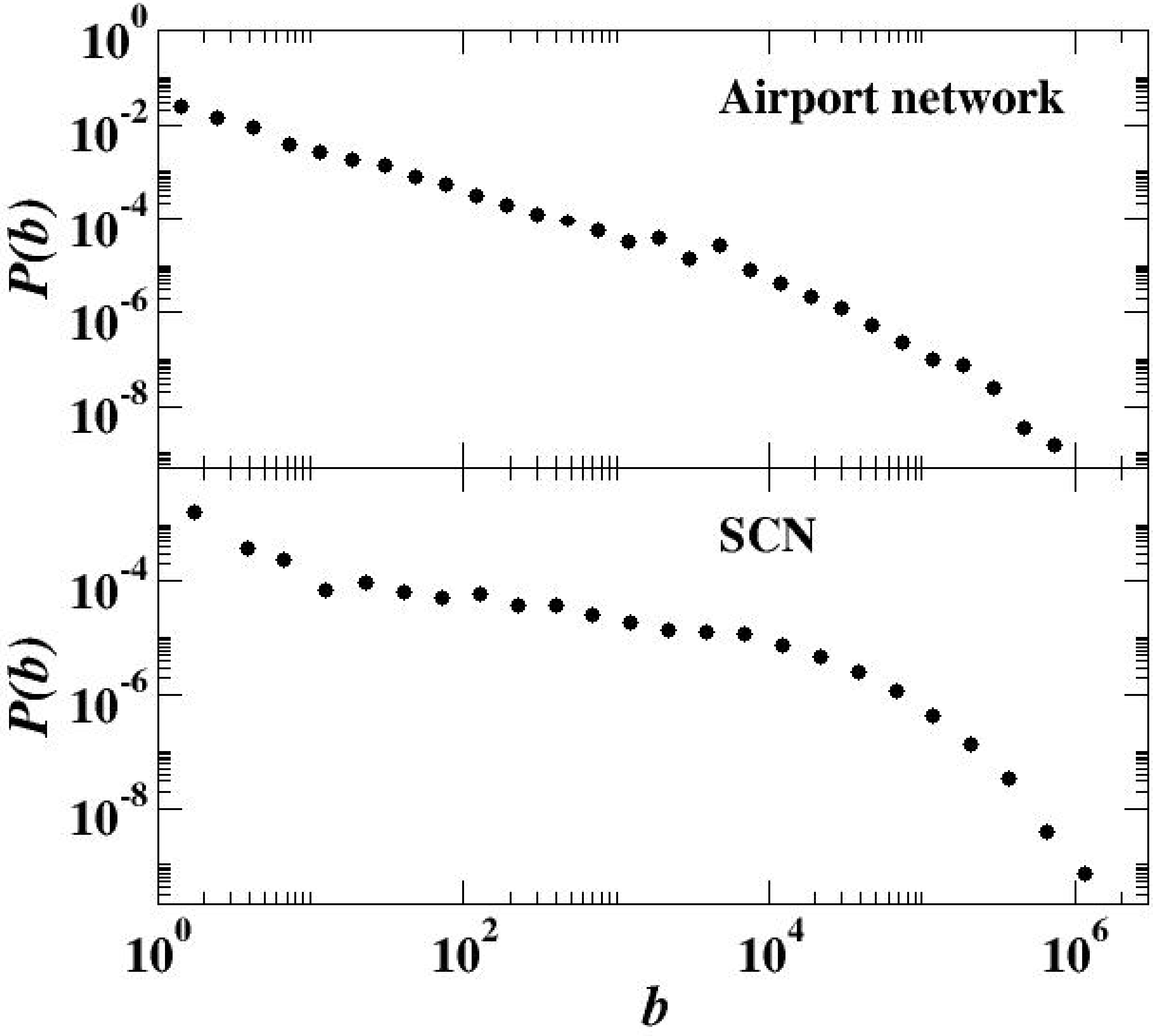
$\langle k^2 \rangle = \int k^2\, P(k)\, dk \sim k_c^{3-\gamma}$

k_c=cut-off due to finite-size

$N \rightarrow$ infinity => diverging degree fluctuations for $\gamma < 3$

Level of heterogeneity: $\boxed{\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle}}$

Other heterogeneity levels

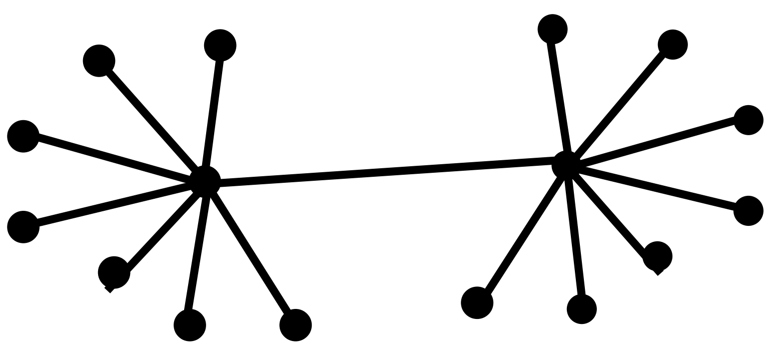


Betweenness centrality

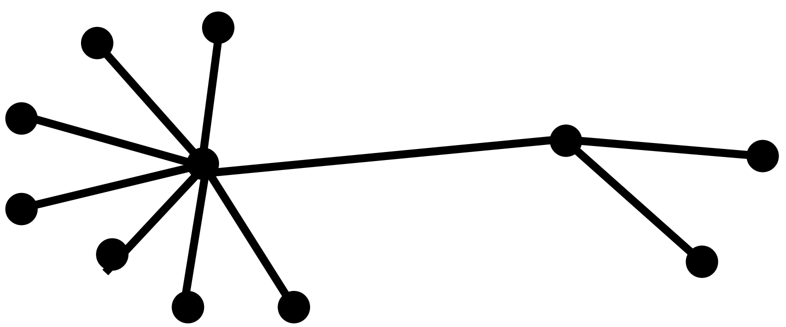
Degree Assortativity

Multipoint degree correlations

P(k): not enough to characterize a network



Large degree nodes tend to
connect to large degree nodes
Ex: social networks



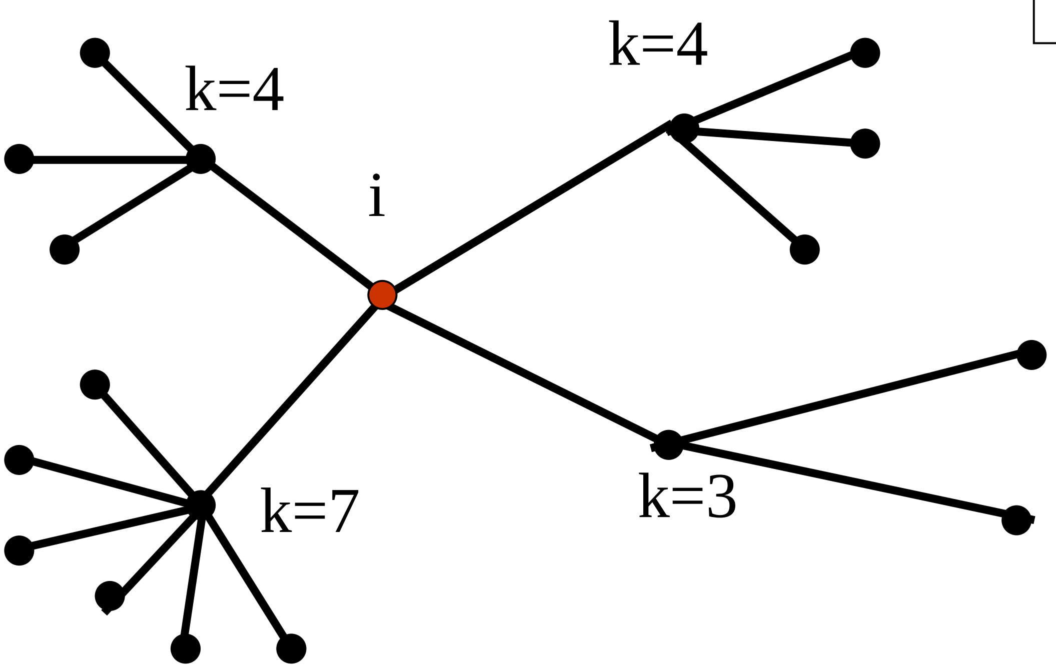
Large degree nodes tend to
connect to small degree nodes
Ex: technological networks

Multipoint degree correlations

Practical measure of correlations:

average degree of nearest neighbors

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} k_j$$



k_i=4

$k_{nn,i}$=(3+4+4+7)/4=4.5

Statistical characterization

average degree of nearest neighbors

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} k_j$$

Correlation spectrum:

putting together nodes which have the same degree

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i/k_i = k} k_{nn,i}$$

class of degree k

$$\boxed{k_{nn}(k) = \sum_{k'} k' P(k'|k)}$$

Typical correlations

- Assortative behaviour: growing $k_{nn}(k)$

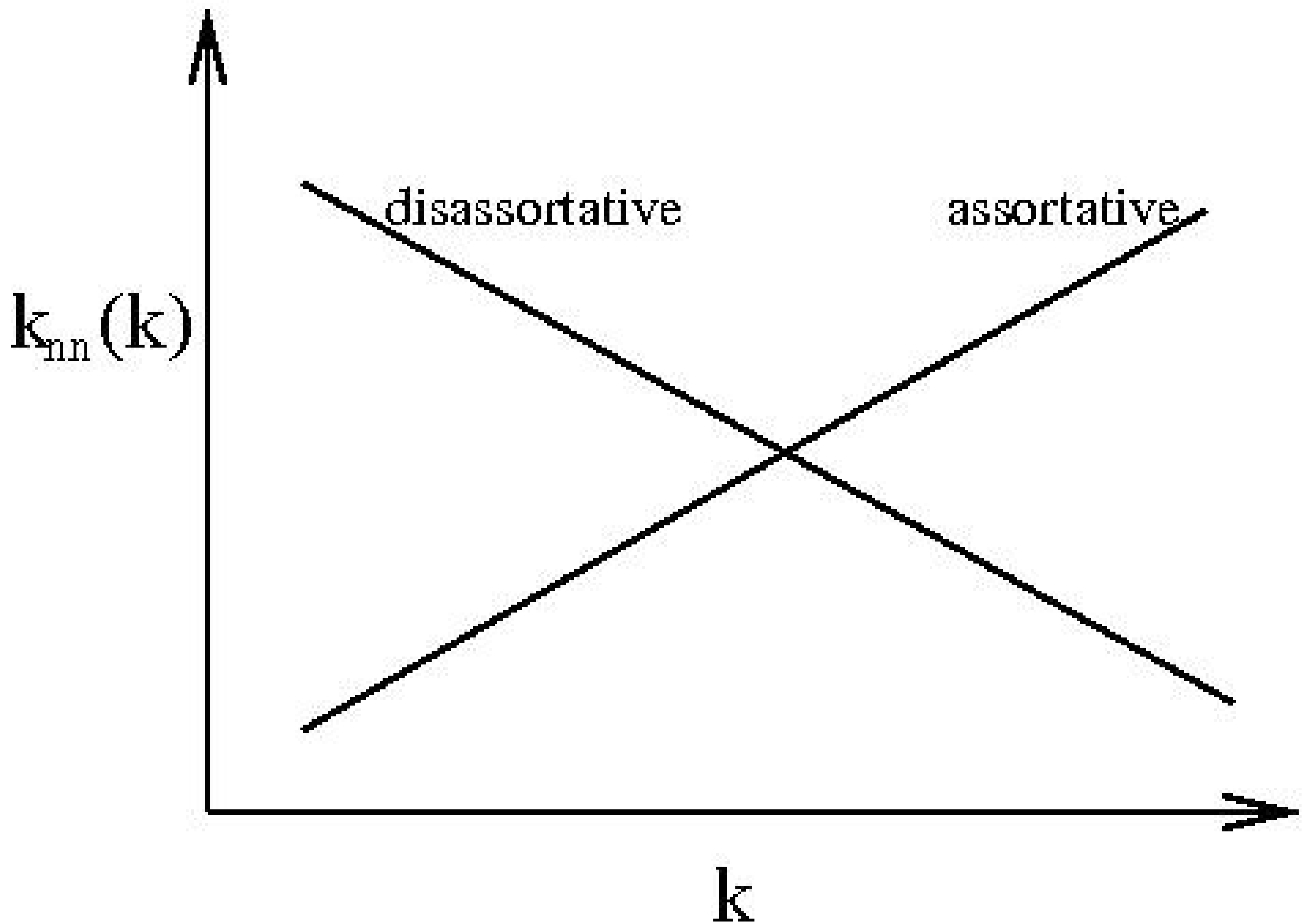
Example: social networks

Large sites are connected with large sites

- Disassortative behaviour: decreasing $k_{nn}(k)$

Example: internet

Large sites connected with small sites, hierarchical structure



$k_{nn}(k)$
disassortative
assortative
k

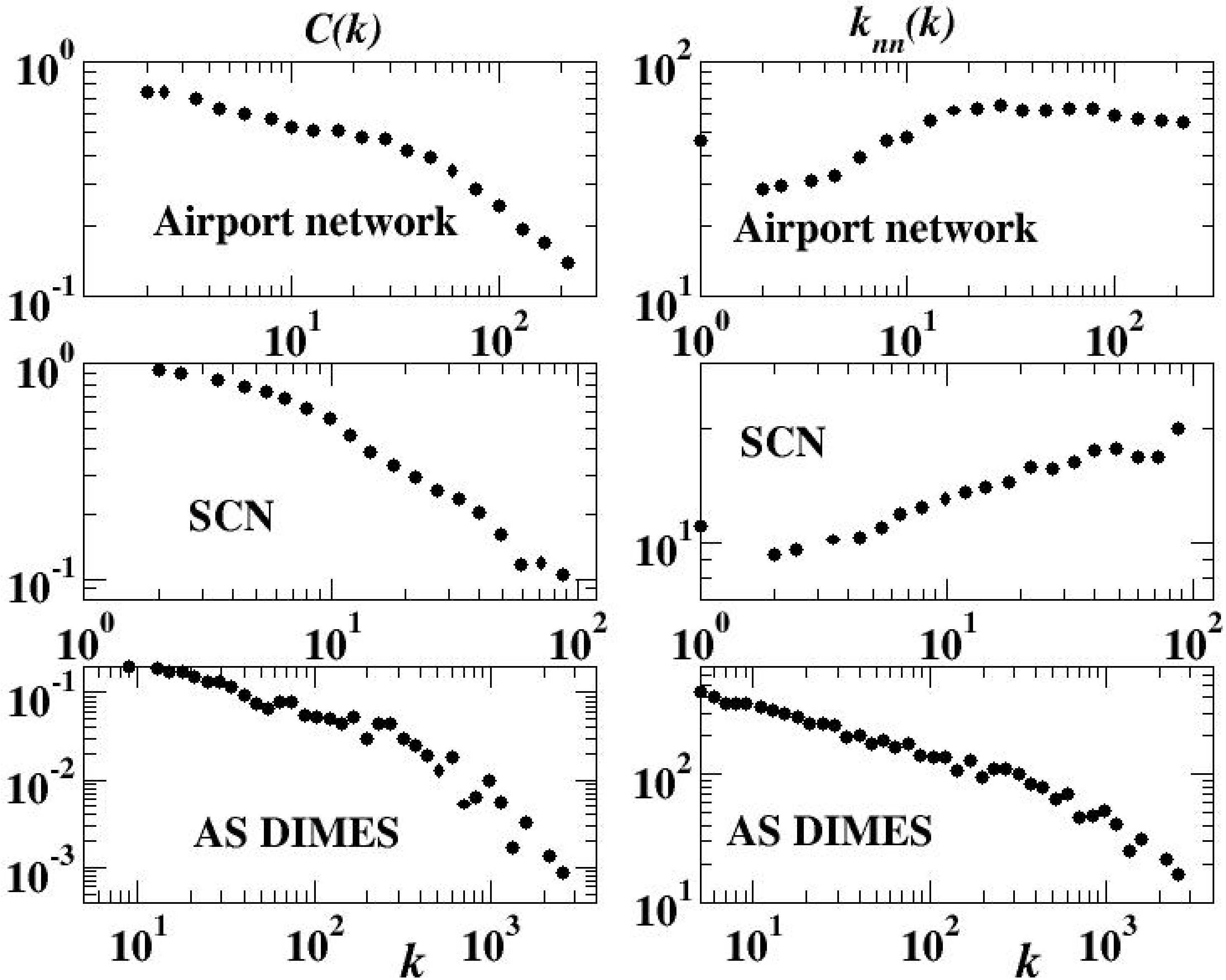
Clustering Spectrum

Average clustering coefficient

$C=\sum_i C(i)/N$

$$\boxed{C(k) = \frac{1}{N_k} \sum_{i/k_i=k} C(i)}$$

Clustering and correlations



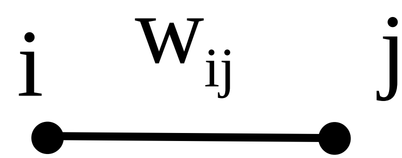
non-trivial structures

Weighted networks

Real world networks: links

- carry traffic (transport networks, Internet…)
- have different intensities (social networks…)

General description: weights



a_{ij}: 0 or 1

w_{ij}: continuous variable

Weighted networks

Weights: on the links

Strength of a node:

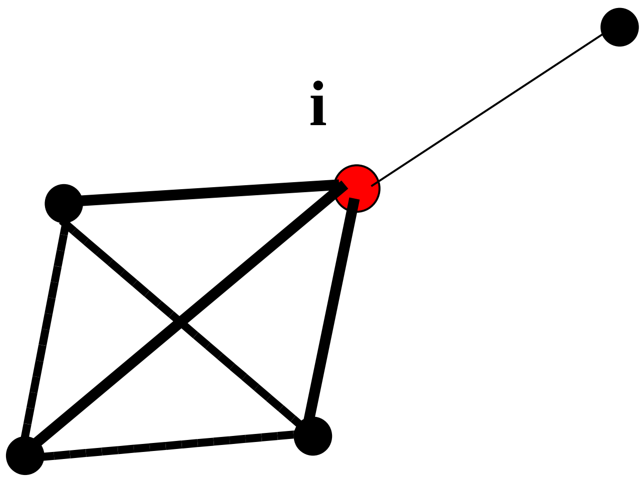
$$s_i = \sum_{j \in V(i)} w_{ij}$$

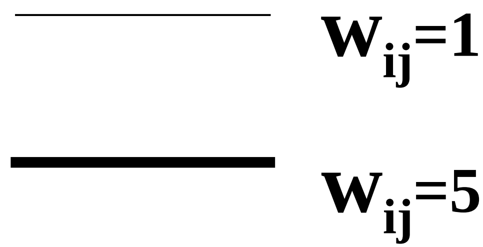
=>Naturally generalizes the degree to weighted networks

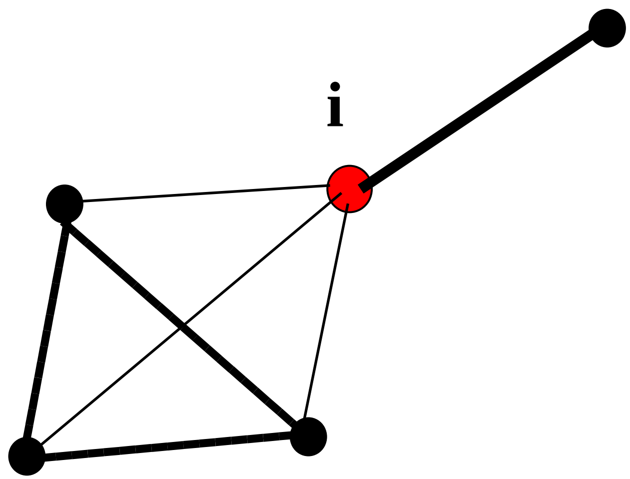
=>Quantifies for example the total traffic at a node

Weighted clustering coefficient

$$C(i) = \frac{1}{k_i(k_i-1)} \sum_{j \neq h} a_{ij} a_{jh} a_{ih}$$







$$C^w(i) = \frac{1}{s_i(k_i-1)} \sum_{j \neq h} a_{ij} a_{jh} a_{ih} \frac{w_{ij}+w_{ih}}{2}$$

s_i=16

c_i^w=0.625 > c_i

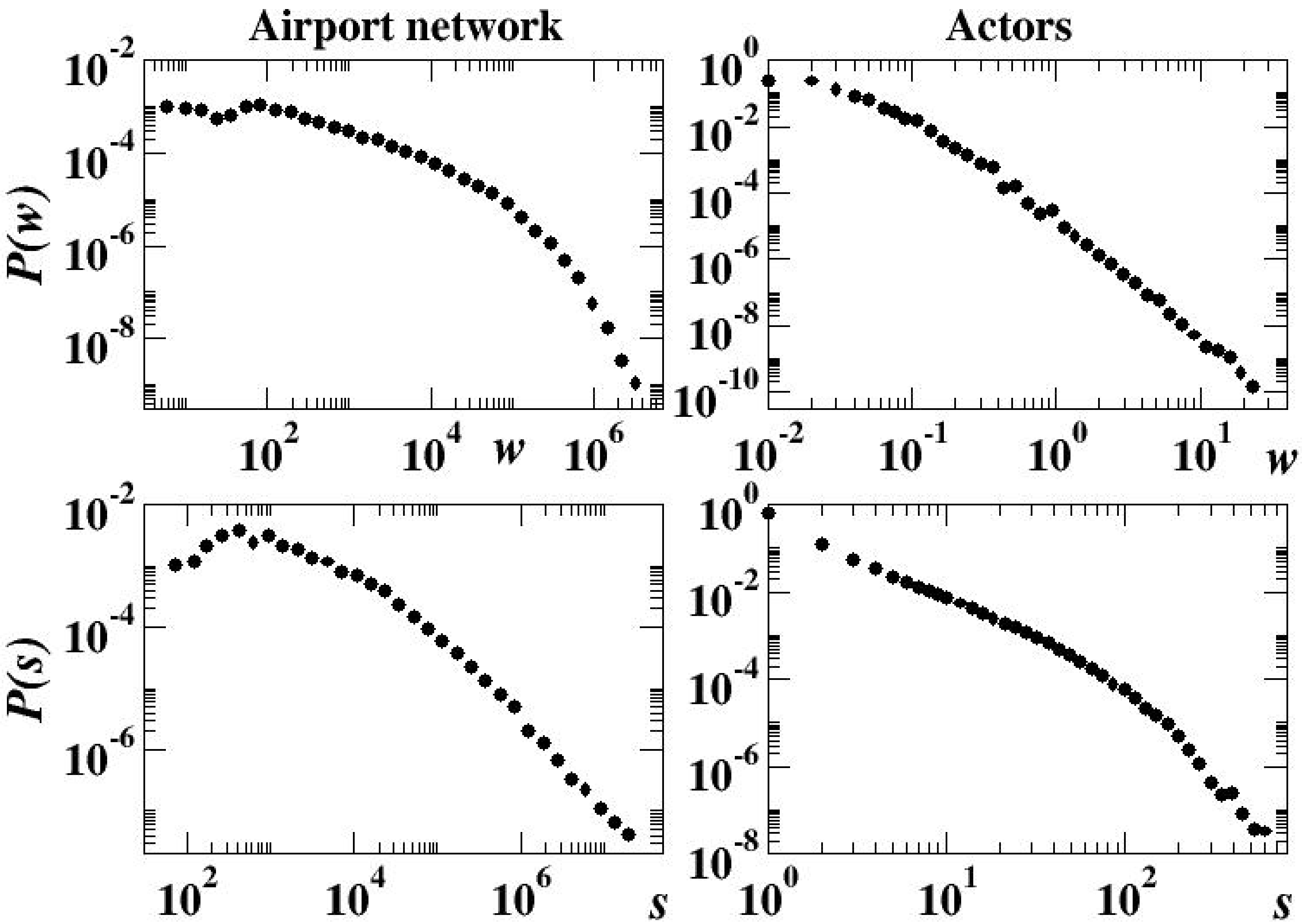
k_i=4

c_i=0.5

s_i=8

c_i^w=0.25 < c_i

Other heterogeneity levels



Weights

Strengths