# Lecture I
# Introduction to complex networks

Santo Fortunato

INSTITUTE
FOR SCIENTIFIC INTERCHANGE
FOUNDATION

# References


- **Evolution of networks**
  S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. 51, 1079 (2002),
  cond-mat/0106144

- **Statistical mechanics of complex networks**
  R. Albert, A-L Barabasi
  Reviews of Modern Physics 74, 47 (2002),
  cond-mat/0106096

- **The structure and function of complex networks**
  M. E. J. Newman, SIAM Review 45, 167-256 (2003),
  cond-mat/0303516

- **Complex networks: structure and dynamics**
  S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang
  Physics Reports 424, 175-308 (2006)

- **Community detection in graphs**
  S. Fortunato
  arXiv: 0906.0612

# Plan of the course

I. **Networks: definitions, characteristics, basic concepts in graph theory**

II. Real world networks: basic properties. Models I

III. Models II

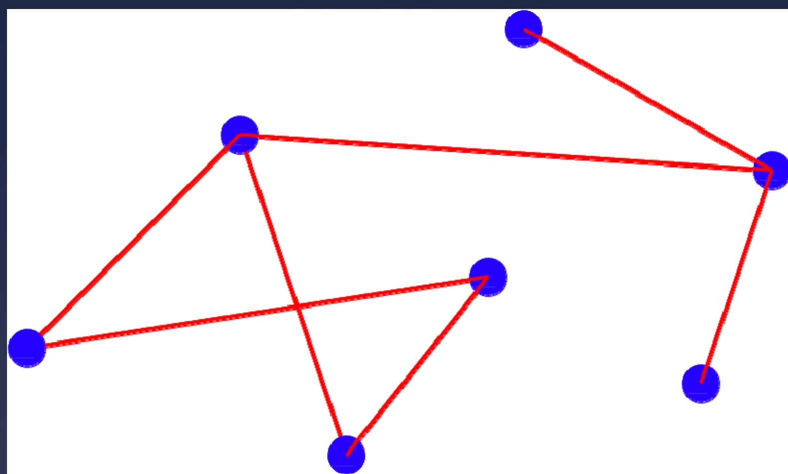
IV. Community structure I

V. Community structure II

VI. Dynamic processes in networks

# What is a network?

Network or graph=set of vertices joined by edges



very abstract representation

very general

convenient to describe many different systems

# Some examples

| | Nodes | Links |
|---|---|---|
| Social networks | Individuals | Social relations |
| Internet | Routers<br>AS | Cables<br>Commercial agreements |
| WWW | Webpages | Hyperlinks |
| Protein interaction networks | Proteins | Chemical reactions |

and many more (email, P2P, foodwebs, transport....)

# Interdisciplinary science

Science of complex networks:

-graph theory

-sociology

-communication science

-biology

-physics

-computer science

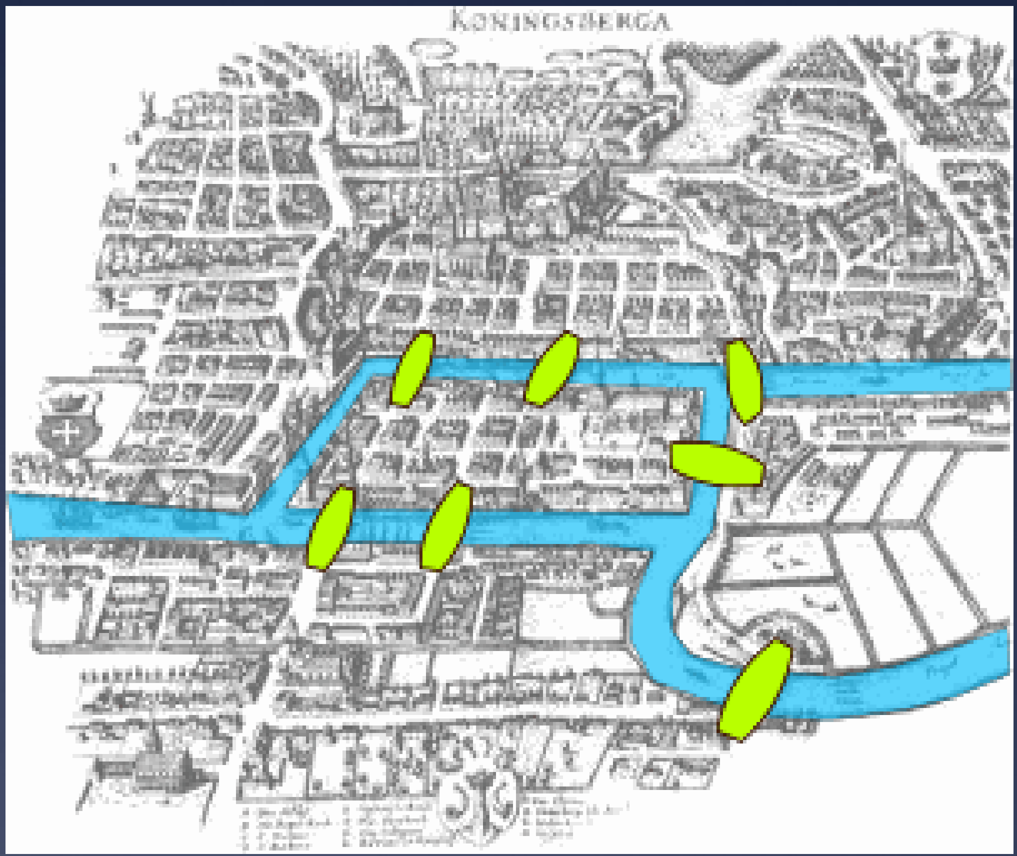
# Interdisciplinary science

Science of complex networks:

* Empirics
* Characterization
* Modeling
* Dynamical processes

# Graph Theory

Origin: Leonhard Euler (1736)



# Graph theory: basics

Graph G=(V,E)

* V=set of nodes/vertices i=1,…,n
* E=set of links/edges (i,j), m

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:

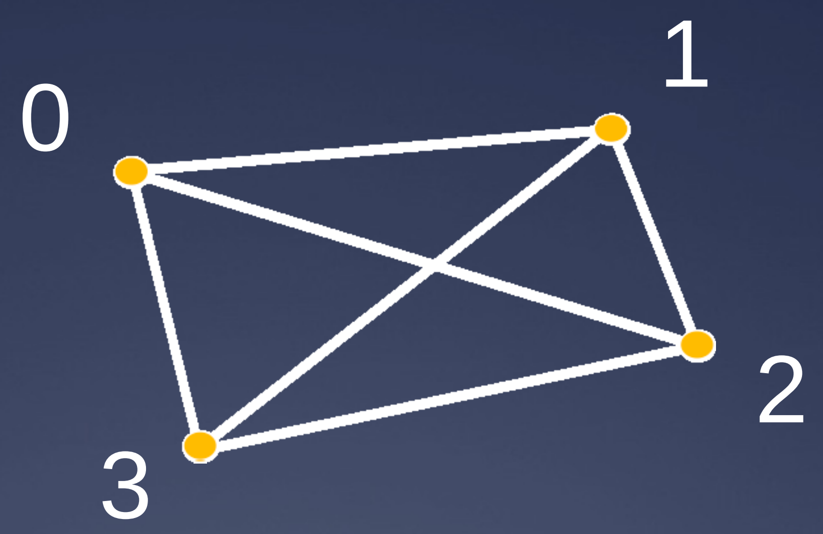
(all to all interaction/communication)

# Adjacency matrix

n vertices i=1,...,n

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

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 1 | 1 | 1 |
| 1 | 1 | 0 | 1 | 1 |
| 2 | 1 | 1 | 0 | 1 |
| 3 | 1 | 1 | 1 | 0 |



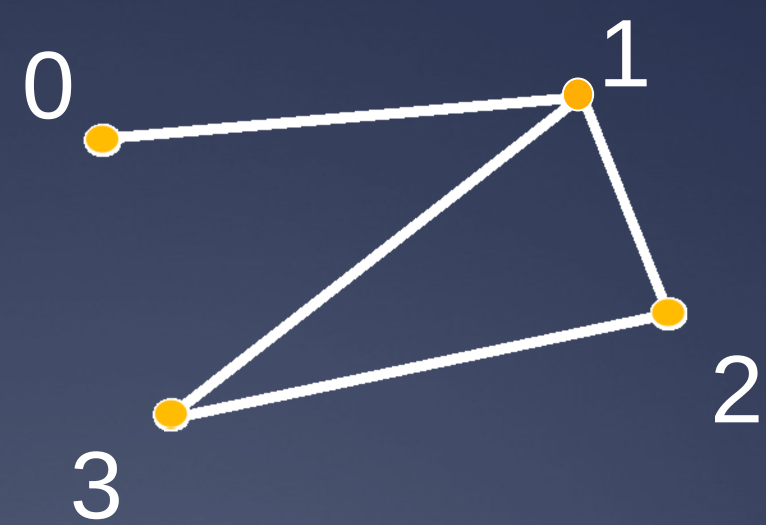
# Adjacency matrix

n vertices 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 |



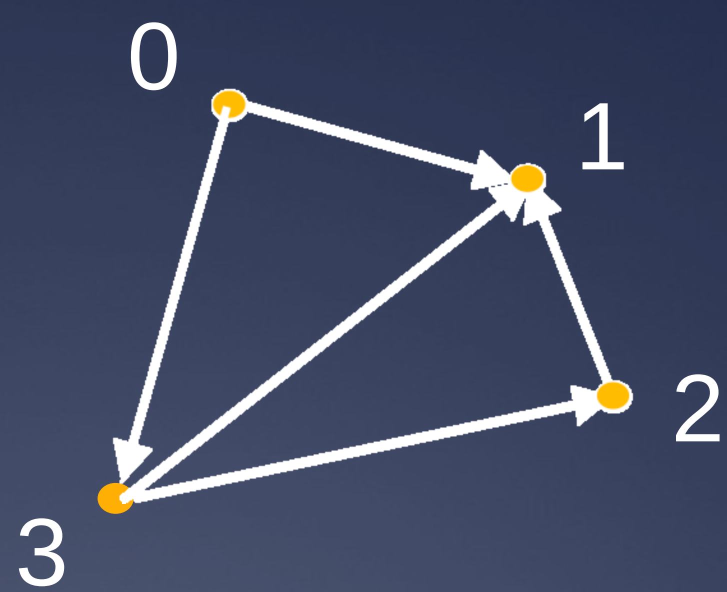
# Adjacency matrix

n vertices 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 <<1 → Sparse adjacency matrix

→ Representation: lists of neighbours of each node

$l(i, V(i))$

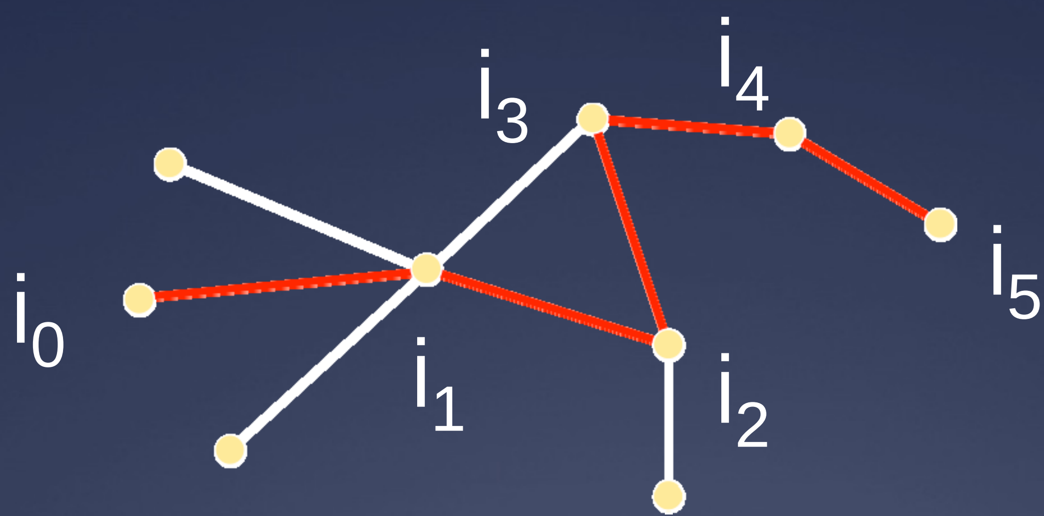
V(i)=neighbourhood of i

# Paths

G=(V,E)

Path of length l = ordered collection of

- l+1 vertices $i_0, i_1, \ldots, i_l \; \varepsilon \; V$
- l edges $(i_0,i_1), (i_1,i_2) \ldots, (i_{l-1},i_l) \; \varepsilon \; E$

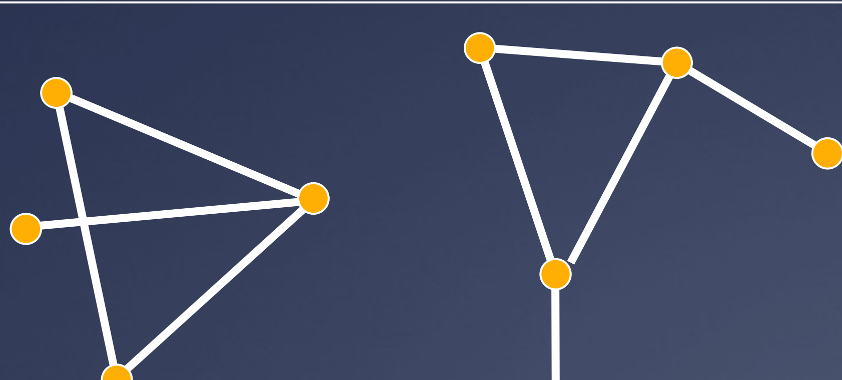



Cycle/loop = closed path ($i_0 = i_l$) with all other vertices and edges distinct

# Paths and connectedness

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

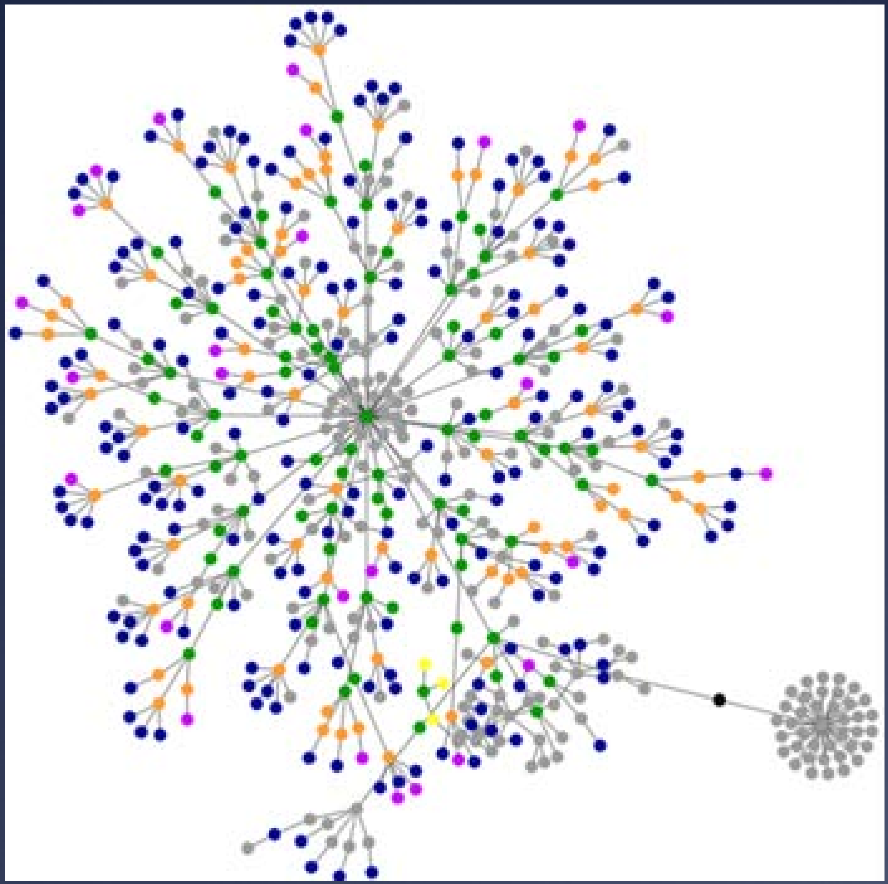
is connected



- is not connected
- is formed by two components

# Trees

A tree is a connected graph without loops/cycles



* n nodes, n-1 links
* Maximal loopless graph
* Minimal connected graph

# Paths and connectedness

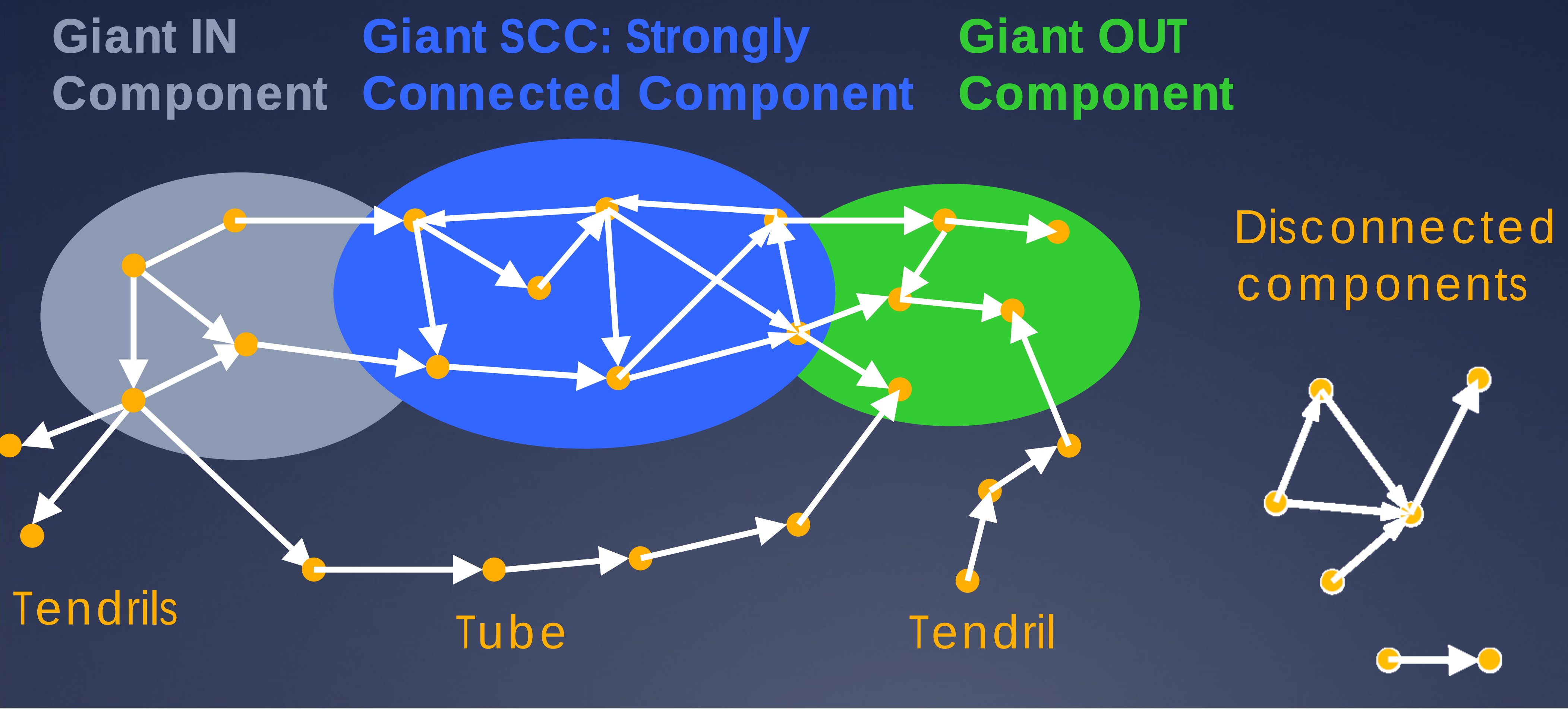
G=(V,E)=> distribution of components' sizes

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

Existence of a giant component ⟺ Macroscopic fraction of the graph is connected

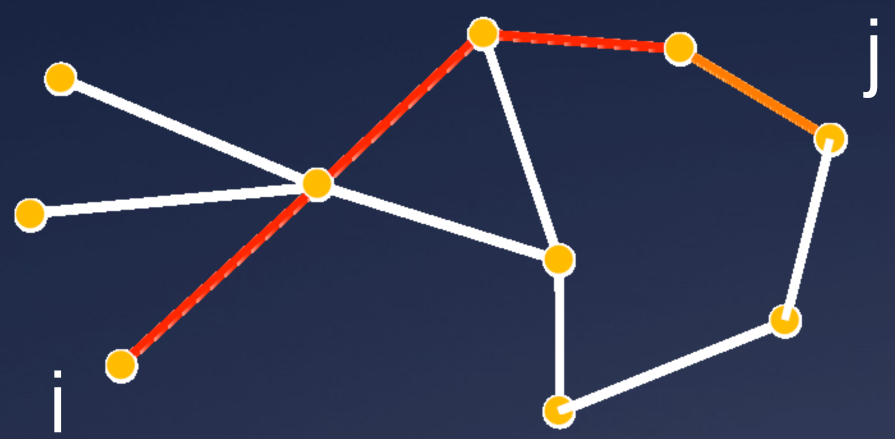
# Paths and connectedness: directed graphs

Paths are ***directed***



# 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

# Graph spectra

Spectrum of a graph: set of eigenvalues of adjacency matrix A

If A is symmetric (undirected graph), n real eigenvalues with real orthogonal eigenvectors

If A is asymmetric, some eigenvalues may be complex

Perron-Frobenius theorem: any graph has (at least) one real eigenvalue $\mu_n$ with one non-negative eigenvector, such that $|\mu| \leq \mu_n$ for any eigenvalue $\mu$. If the graph is connected, the multiplicity of $\mu_n$ is one.

Consequence: on an undirected graph there is only one eigenvector with positive components, the others have mixed-signed components

# Graph spectra

Spectral density

$$\rho(\mu) = \frac{1}{n}\sum_{i=1}^{n}\delta(\mu - \mu_i)$$

Continuous function in the limit $n \longrightarrow \infty$

k-th moment of spectral density

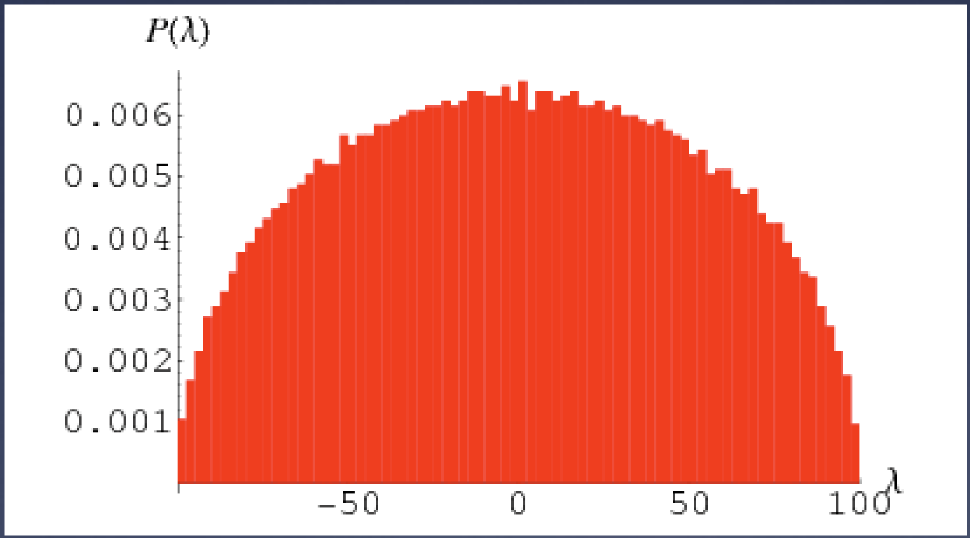
$$M_k = \frac{1}{n}\sum_{j=1}^{n}(\lambda_j)^k = \frac{1}{n}Tr(A^k) = \frac{1}{n}\sum_{i_1,i_2,\ldots,i_k} A_{i_1,i_2}A_{i_2,i_3}\cdots A_{i_k,i_1}$$

# Wigner's semicircle law

For real symmetric uncorrelated random matrices whose elements have finite moments in the limit $n \longrightarrow \infty$

$$\langle A_{ij} \rangle \equiv 0 \text{ and } \langle A_{ij}^2 \rangle = \sigma^2$$

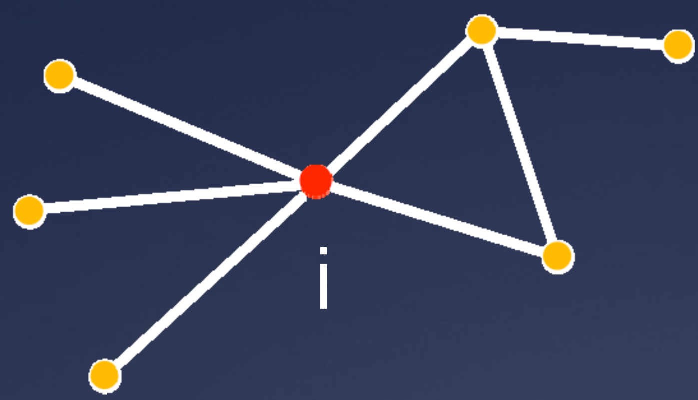
$$\rho(\lambda) = \begin{cases} (2\pi\sigma^2)^{-1}\sqrt{4\sigma^2 - \lambda^2} & \text{if } |\lambda| < 2\sigma \\ 0 & \text{otherwise.} \end{cases}$$



# Centrality measures

How to quantify the importance of a node?

* Degree=number of neighbours=$\sum_j a_{ij}$




$k_i=5$

For directed graphs: $k_{in}$, $k_{out}$

- Closeness centrality

$$g_i = 1 / \sum_j l(i,j)$$

# Betweenness centrality

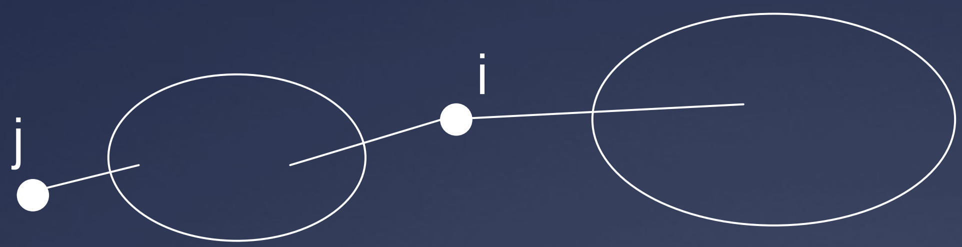
for each pair of nodes (l,m) in the graph, there are

$s^{lm}$ shortest paths between l and m

$s_i^{lm}$ shortest paths going through i

$b_i$ is the sum of $s_i^{lm} / s^{lm}$ over all pairs (l,m)

## Path-based quantity

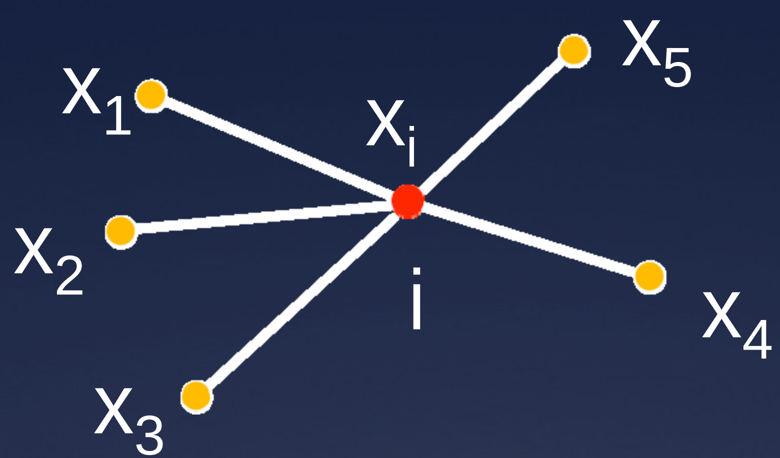



$b_i$ is large
$b_j$ is small

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

NB: generalization to *edge betweenness centrality*

# Eigenvector centrality




Basic principle = the importance of a vertex is proportional to the sum of the importances of its neighbors

$$\lambda x_i = \sum_j A_{ij} x_j \quad \rightarrow \quad \mathbf{Ax} = \lambda \mathbf{x}$$

Solution: eigenvectors of adjacency matrix!

# Eigenvector centrality

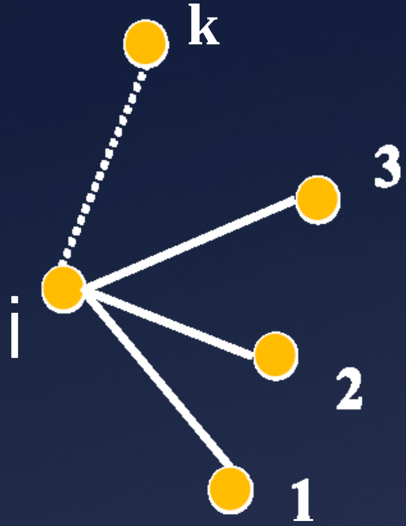
Not all eigenvectors are good solutions!

Requirement: the values of the centrality measure have to be positive

Because of Perron-Frobenius theorem only the eigenvector with largest eigenvalue (principal eigenvector) is a good solution!

The principal eigenvector can be quickly computed with the power method!

# Structure of neighborhoods

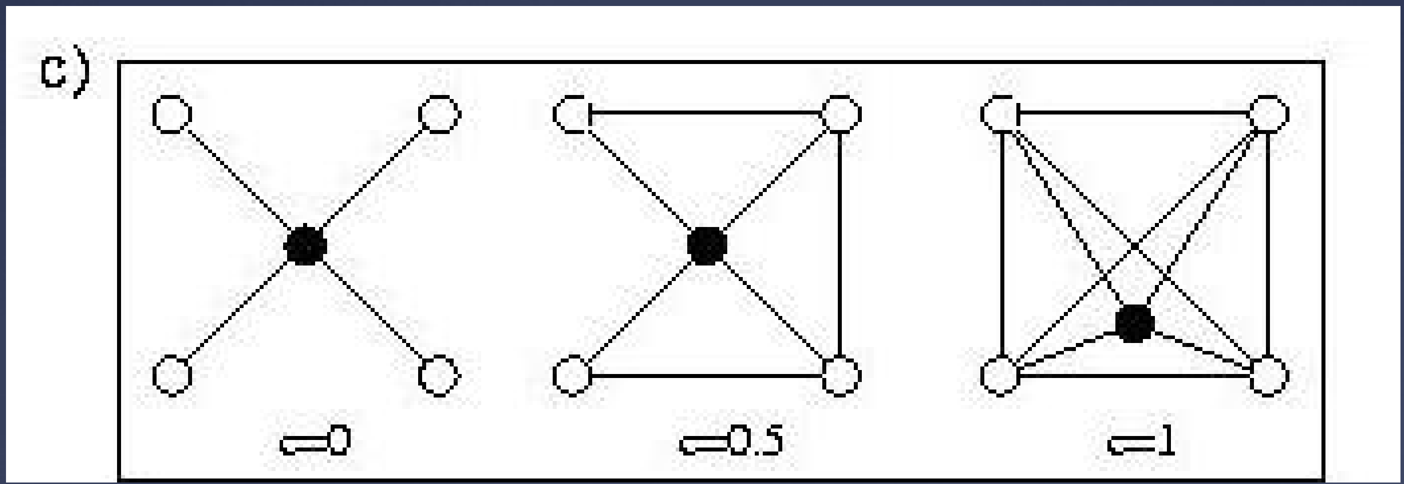



## Clustering coefficient of a node

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

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

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



# Structure of neighborhoods

Average clustering coefficient of a graph

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

# 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

## Cumulative degree distribution

$$P(k) \sim k^{-\gamma} \rightarrow P^{>}(k) \sim \sum_{k'=k}^{\infty} k'^{-\gamma} \sim k^{-(\gamma-1)}$$

$$P(k) \sim e^{-k/\kappa} \rightarrow P^{>}(k) \sim \sum_{k'=k}^{\infty} e^{-k'/\kappa} \sim e^{-k/\kappa}$$

Conclusion: power laws and exponentials can be easily recognized

# 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**: $\langle k^2 \rangle - \langle k \rangle^2$

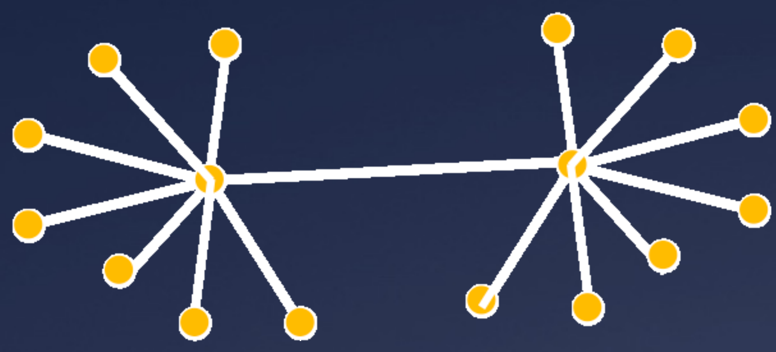
$\langle k^2 \rangle = \sum_i k^2_i/n = \sum_k k^2\,P(k)$

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

# Statistical characterization

## 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

# Statistical characterization

## Multipoint degree correlations

Measure of correlations:
$P(k',k'',\ldots k^{(n)} | k)$: conditional probability that a node of degree k is connected to nodes of degree k', k'',...

Simplest case:
$P(k' | k)$: conditional probability that a node of degree k is connected to a node of degree k'

often inconvenient (statistical fluctuations)

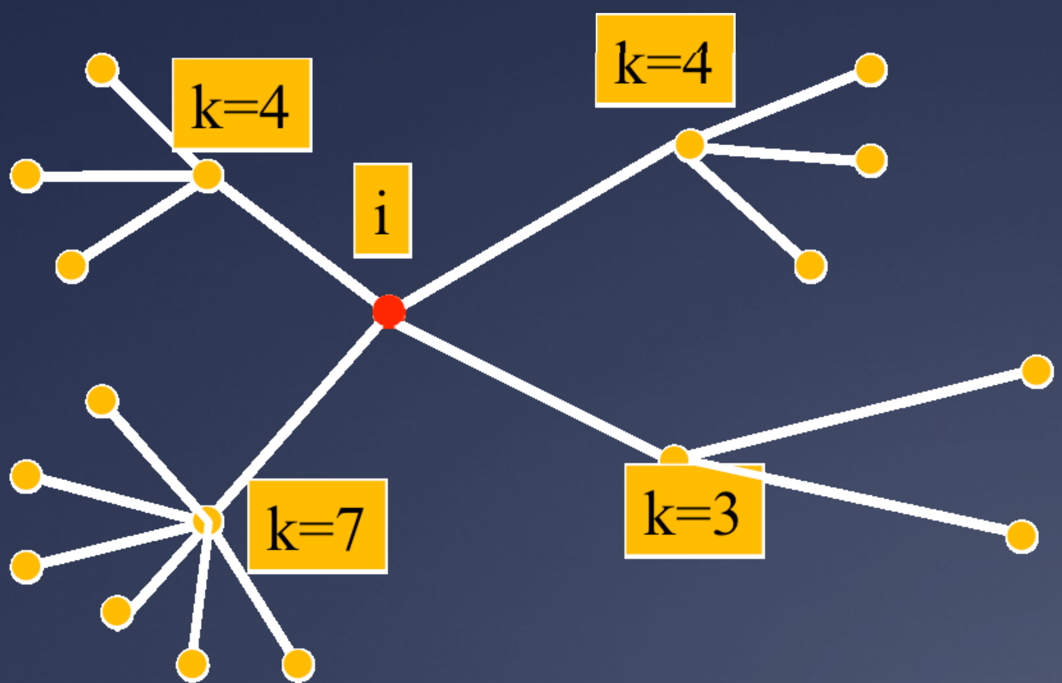
# Statistical characterization

## 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 vertices having the same degree

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

class of degree k

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

# Statistical characterization

## Case of random uncorrelated networks

*P(k' | k)*

- independent of k
- prob. that an edge points to a vertex of degree k'

$$\frac{\text{number of edges from nodes of degree } k'}{\text{number of edges from nodes of any degree}} = \frac{k' n_{k'}}{\sum_{k''} k'' n_{k''}}$$

$$P^{unc}(k'|k) = k' P(k') / \langle k \rangle$$

proportional to k' itself

$$k_{nn}^{unc}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

# Typical correlations

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

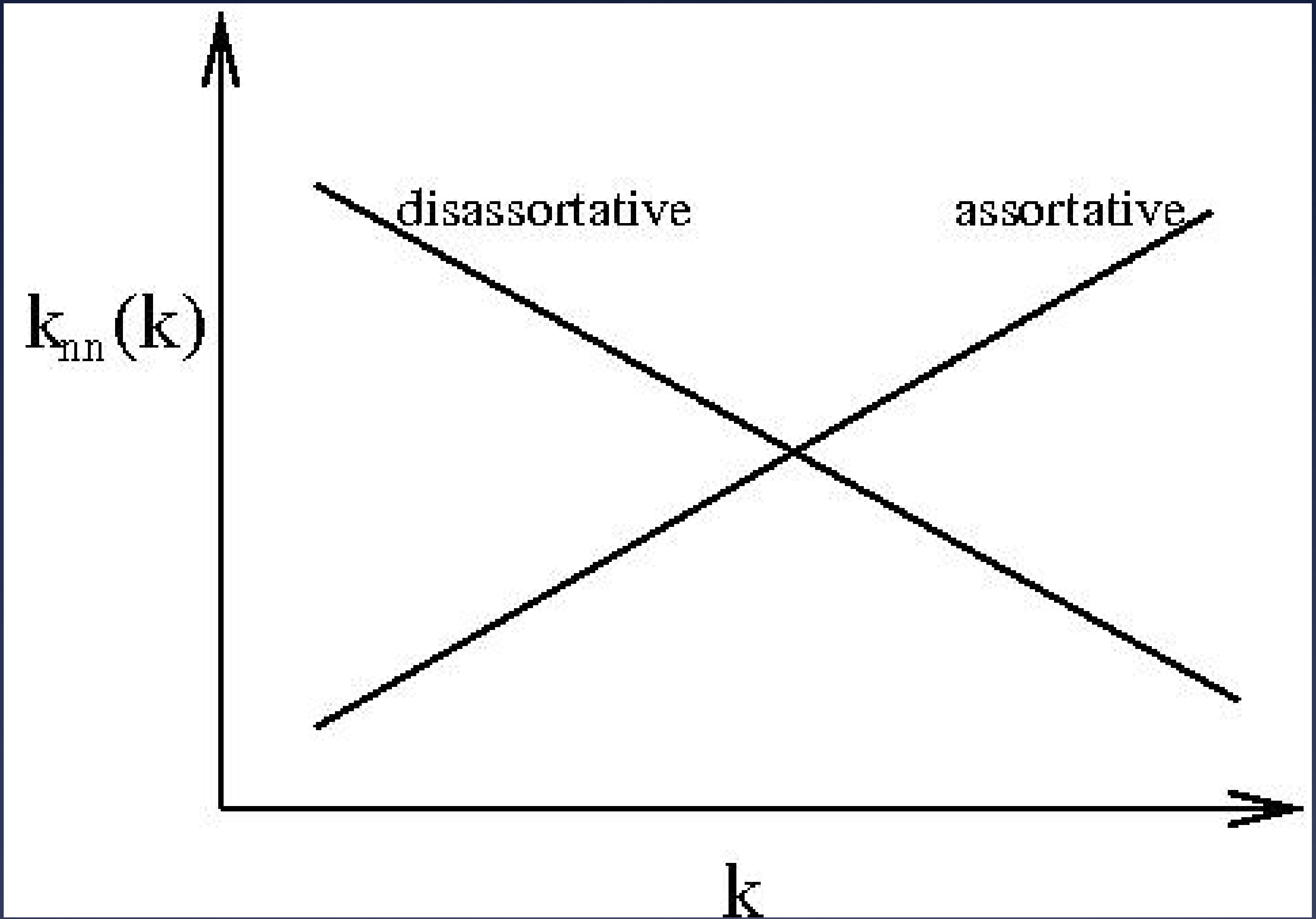
Example: social networks

Large sites are connected with large sites

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

Example: internet

Large sites connected with small sites, hierarchical structure


disassortative
assortative
k_nn(k)
k

# Correlations: Clustering spectrum

- P(k',k''|k): cumbersome, difficult to estimate from data
- Average clustering coefficient C=average over nodes with very different characteristics

Clustering spectrum:

putting together nodes which have the same degree

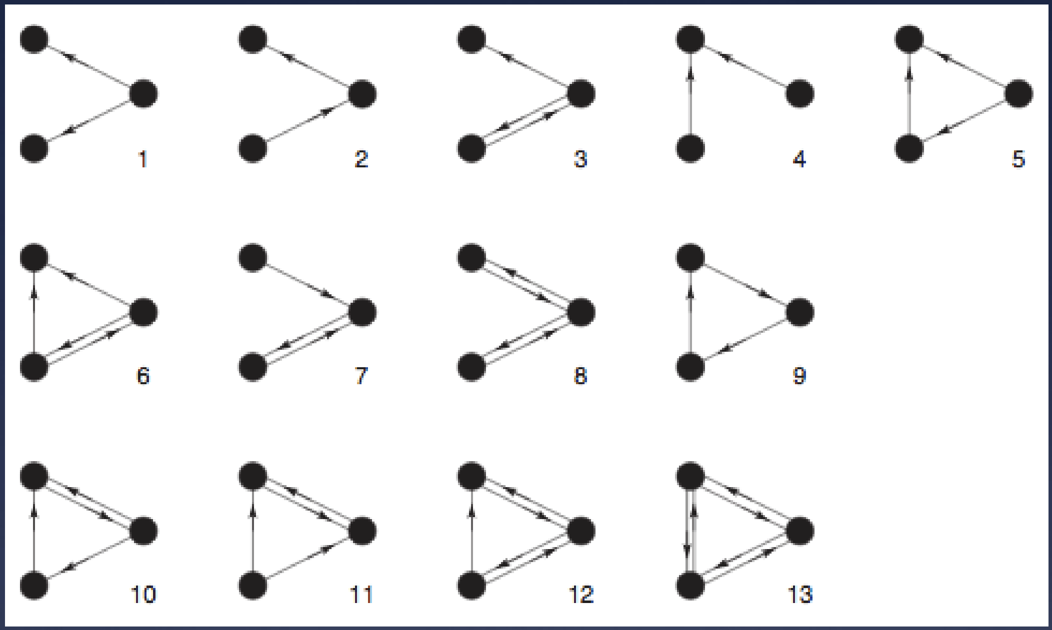
$$C(k) = \frac{1}{n_k} \sum_{i/k_i=k} C(i)$$

class of degree k

(link with hierarchical structures)

# Motifs

Motifs: subgraphs occurring more often than on random versions of the graph




Significance of motifs: Z-score!

$$Z_M = \frac{n_M - \langle n_M^{\mathrm{rand}} \rangle}{\sigma_{n_M}^{\mathrm{rand}}}$$

# 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

# Weights: examples

- Scientific collaborations: number of common papers
- Internet, emails: traffic, number of exchanged emails
- Airports: number of passengers
- Metabolic networks: fluxes
- Financial networks: shares
- ...

usually $w_{ii}=0$

symmetric: $w_{ij}=w_{ji}$

# Weighted networks

Weights: on the links

Strength of a vertex:

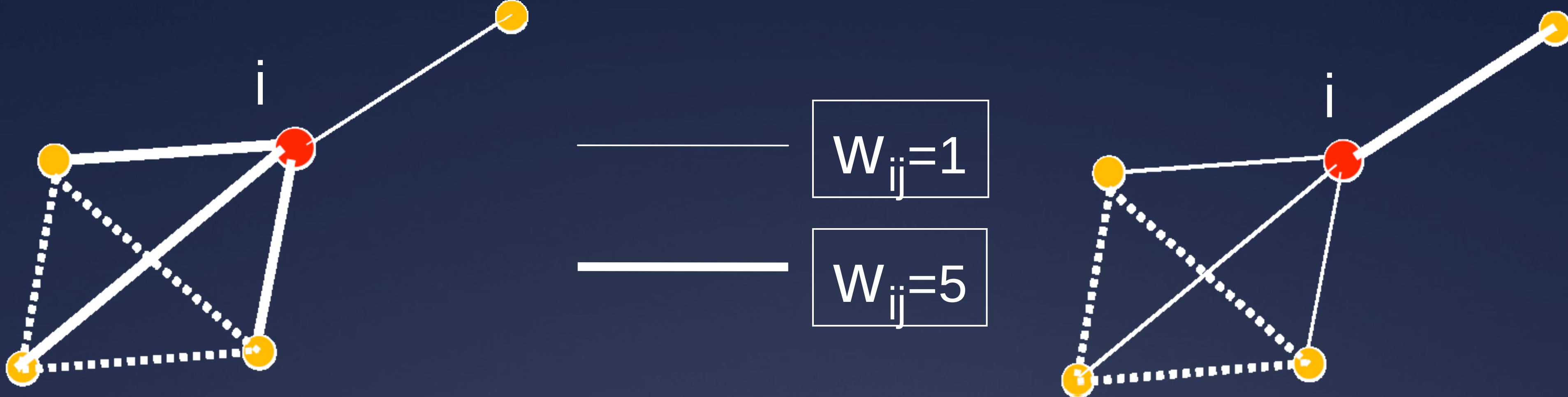
$$s_i = \sum_{j \in V(i)} w_{ij}$$

=>Naturally generalizes the degree to weighted networks

=>Quantifies for example the total traffic at a node

# Weighted clustering coefficient I

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




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

A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, PNAS 101, 3747 (2004)

$s_i=16$

$c_i^w=0.625 > c_i$

$k_i=4$

$c_i=0.5$

$s_i=8$

$c_i^w=0.25 < c_i$

# Weighted clustering coefficient II

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

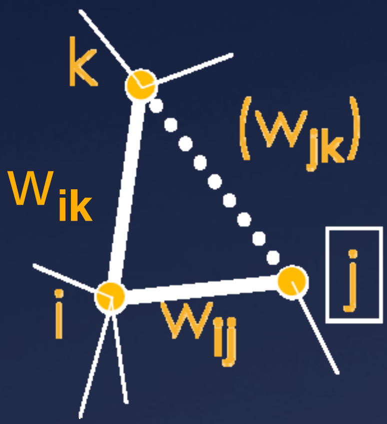
Definition based on subgraph intensity

$$\tilde{C}_{i,O} = \frac{1}{k_i\left(k_i-1\right)} \sum_{j,k} \left(\hat{w}_{ij} \hat{w}_{ik} \hat{w}_{jk}\right)^{1/3}$$

$$\hat{w}_{ij} = w_{ij} / max(w)$$

J. Saramäki, M. Kivela, J.-P. Onnela, K. Kaski, J. Kertész, Phys. Rev. E 75, 027105 (2007)

# Weighted clustering coefficient




Average clustering coefficient

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

$C^w = \sum_i C^w(i)/n$

Random(ized) weights: $C = C_w$

$C < C_w$ : more weights on cliques

$C > C_w$ : less weights on cliques

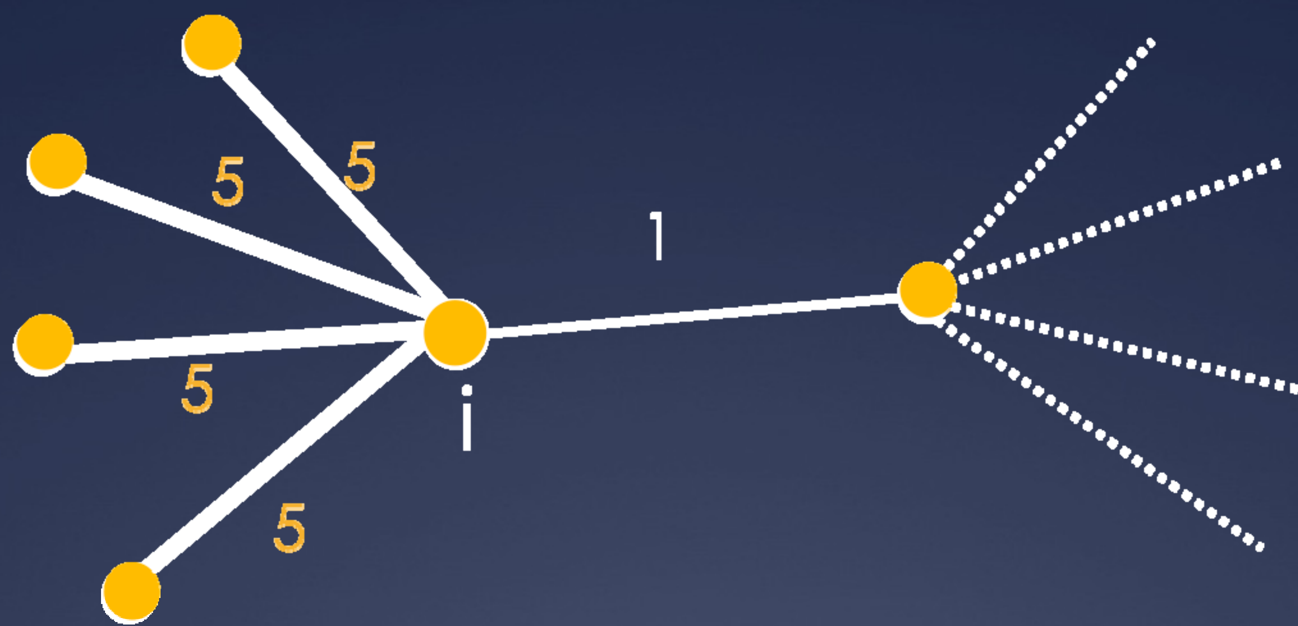
Clustering spectra

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

$$C^w(k) = \frac{1}{n_k} \sum_{i/k_i=k} C^w(i)$$

# Weighted assortativity

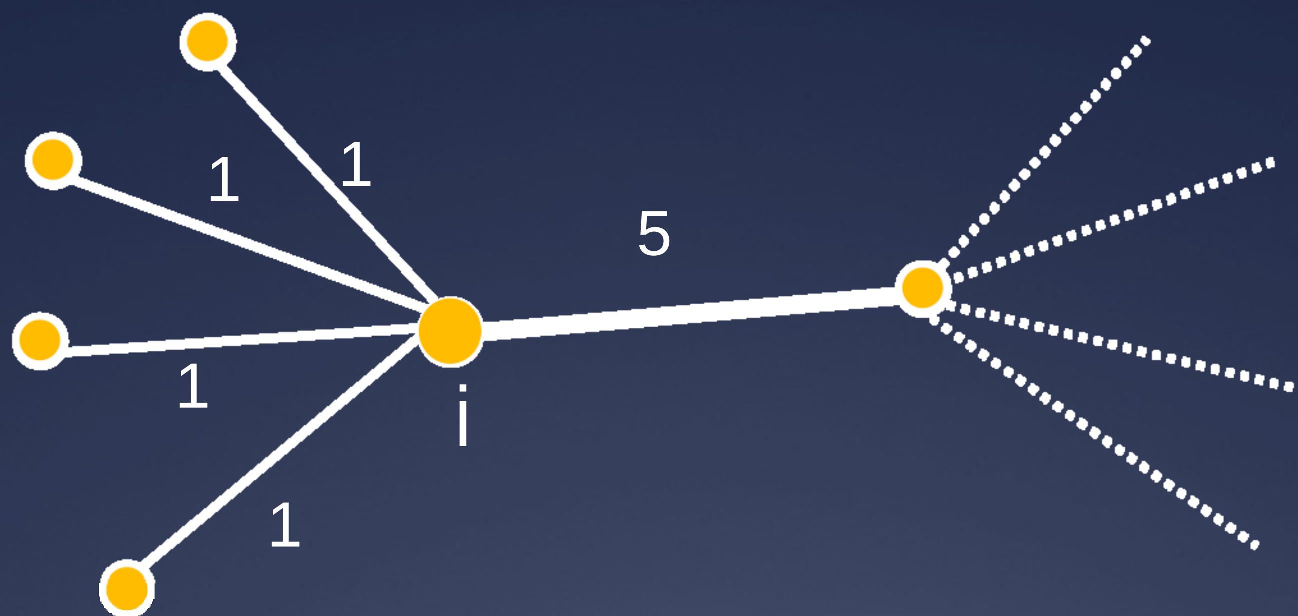
$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} k_j = \frac{1}{k_i} \sum_j a_{ij} k_j$$




$k_i$=5; $k_{nn,i}$=1.8

# Weighted assortativity

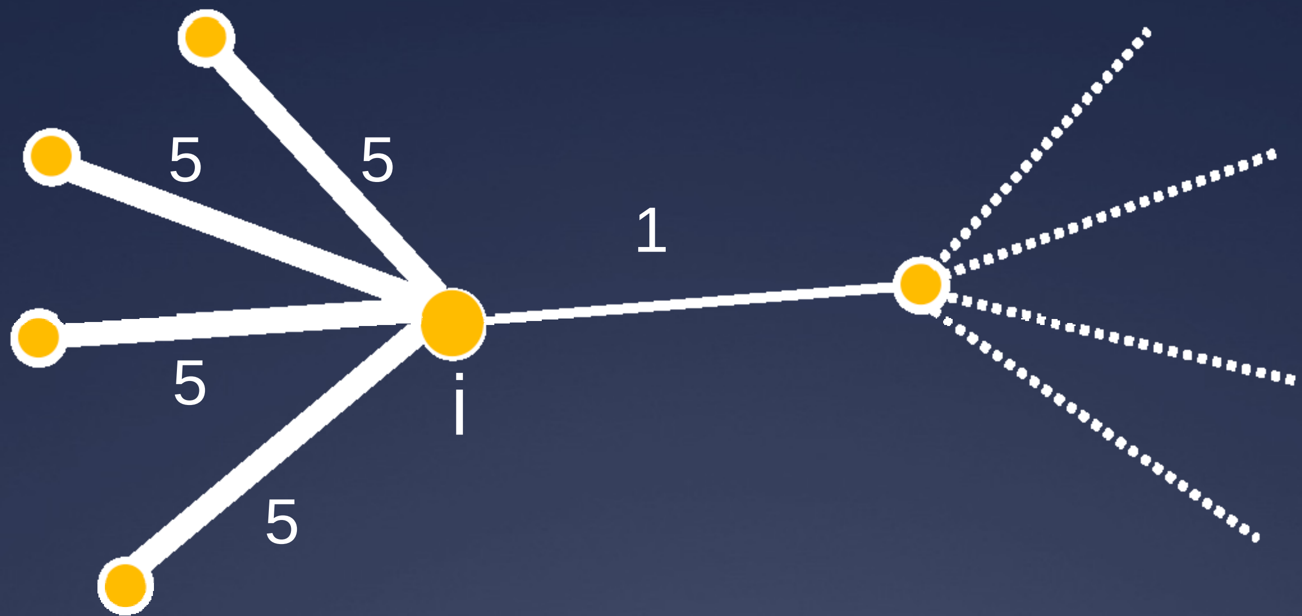
$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} k_j = \frac{1}{k_i} \sum_j a_{ij} k_j$$




$k_i=5$; $k_{nn,i}=1.8$

# Weighted assortativity

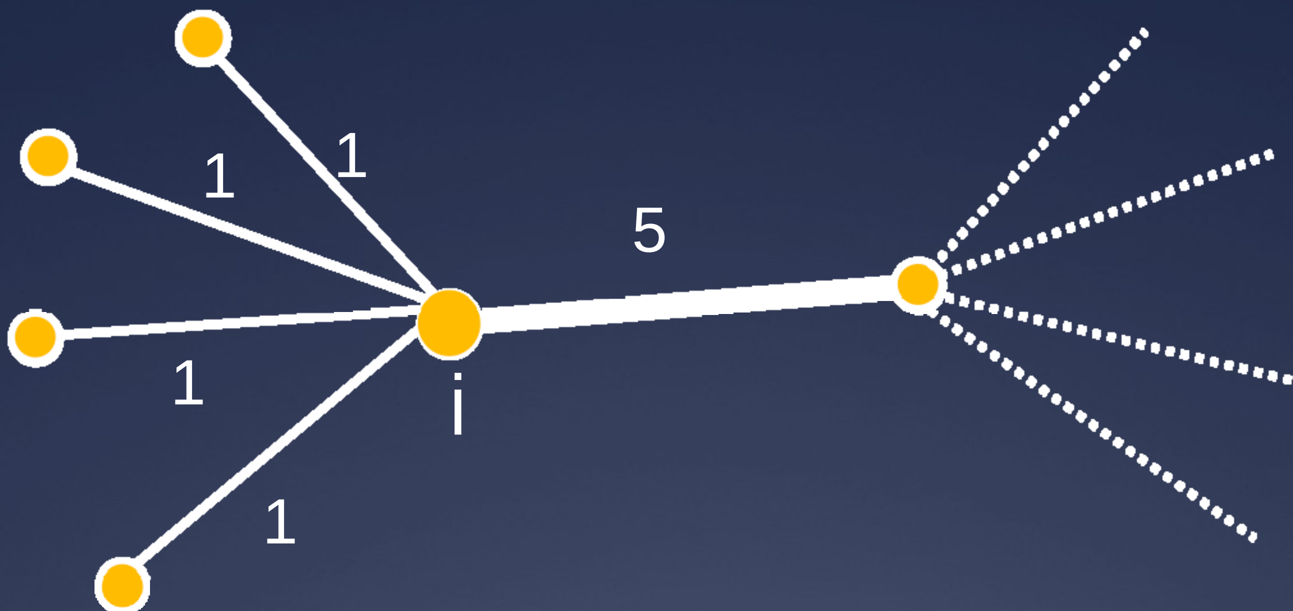
$$k_{nn,i}^{w} = \frac{1}{s_i} \sum_{j \in V(i)} w_{ij} k_j$$




$k_i=5$; $s_i=21$; $k_{nn,i}=1.8$ ; $k_{nn,i}^{w}=1.2$: $k_{nn,i} > k_{nn,i}^{w}$

# Weighted assortativity

$$k^{w}_{nn,i} = \frac{1}{s_i} \sum_{j \in V(i)} w_{ij} k_j$$

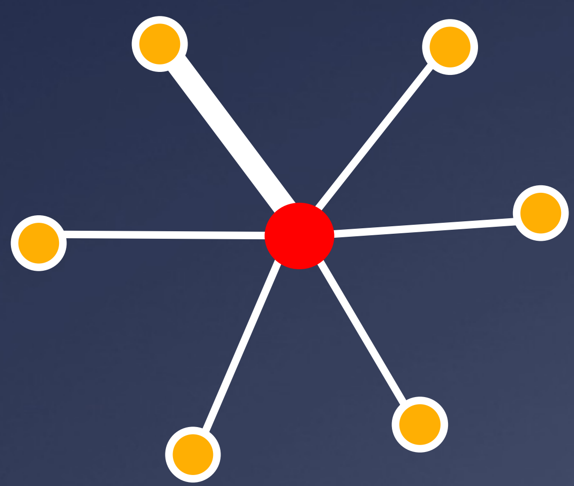


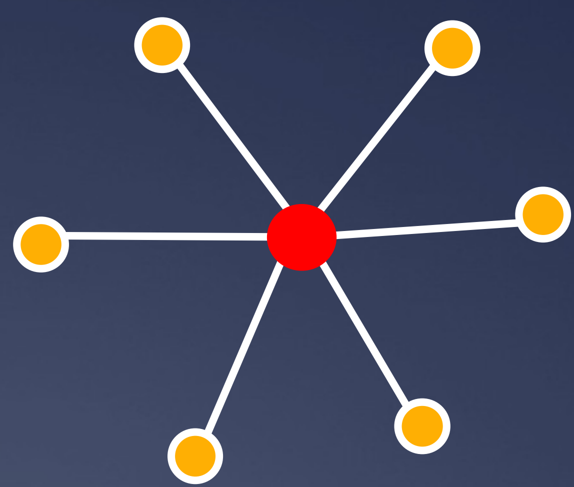

$k_i$=5; $s_i$=9; $k_{nn,i}$=1.8 ; $k_{nn,i}{}^{w}$=3.2: $k_{nn,i} < k_{nn,i}{}^{w}$

# Participation ratio

$$Y_2(i) = \sum_{j \in V(i)} \left[ \frac{w_{ij}}{s_i} \right]^2$$

- $1/k_i$ if all weights equal
- close to 1 if few weights dominate





# Plan of the course

I. Networks: definitions, characteristics, basic concepts in graph theory

**II. Real world networks: basic properties. Models I**

III. Models II

IV. Community structure I

V. Community structure II

VI. Dynamic processes in networks