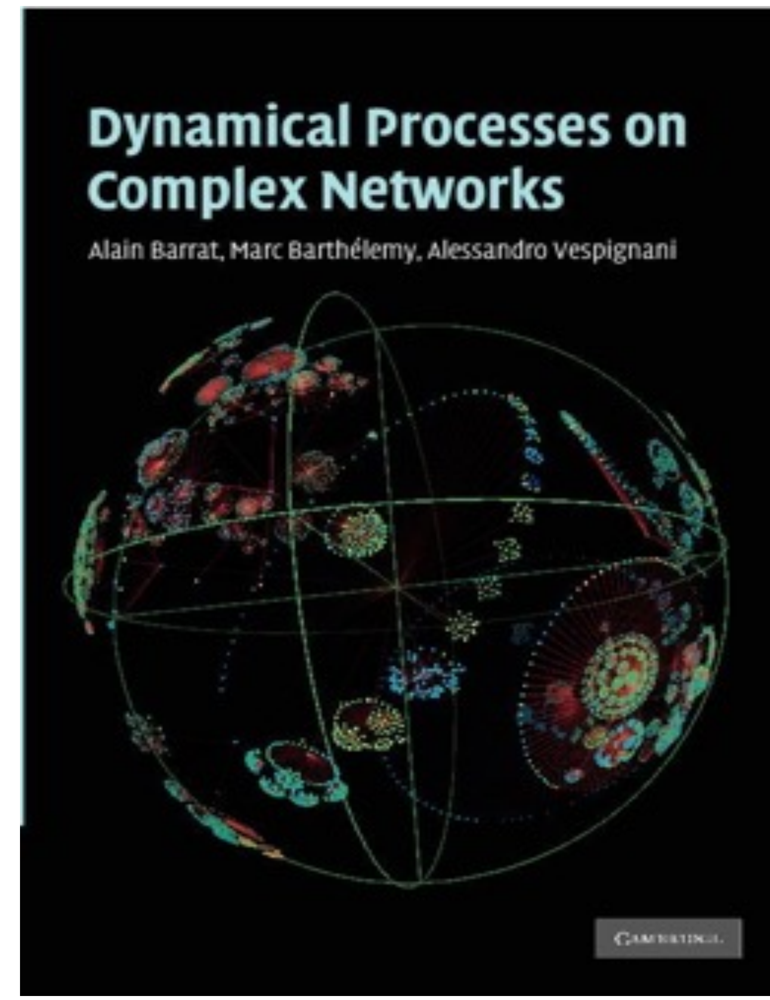
# An introduction to the physics of complex networks


**Alain Barrat**

CPT, Marseille, France

ISI, Turin, Italy


http://www.cpt.univ-mrs.fr/~barrat
http://www.cxnets.org
http://www.sociopatterns.org

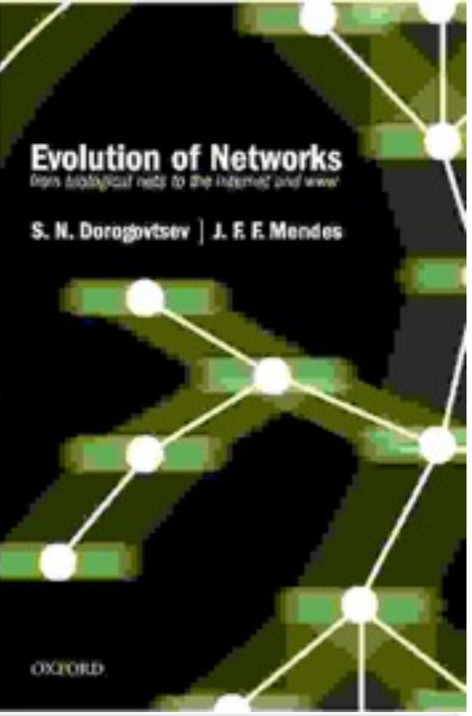
# REVIEWS:


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

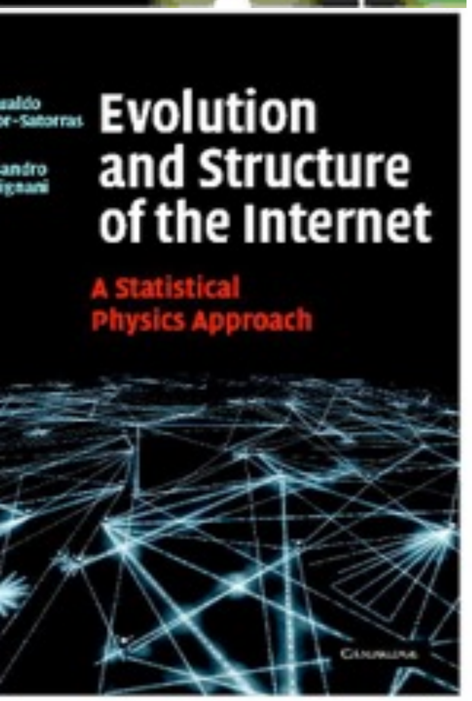
•**The structure and function of complex networks**
M. E. J. Newman, SIAM Review 45, 167-256 (2003), cond-mat/0303516

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

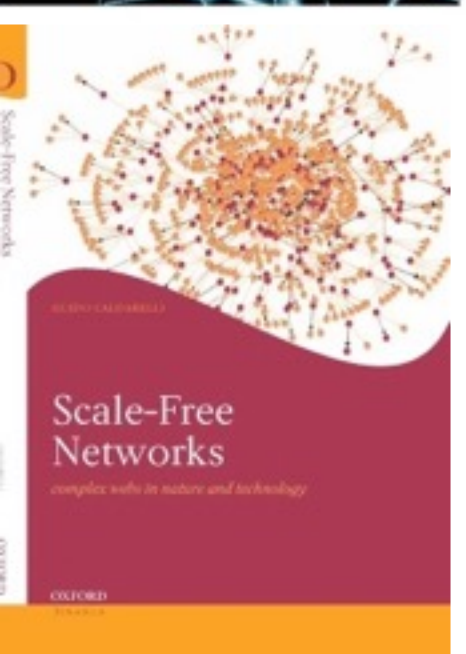
•**Complex Networks: Structure and Dynamics**
S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Physics Reports 424 (2006) 175



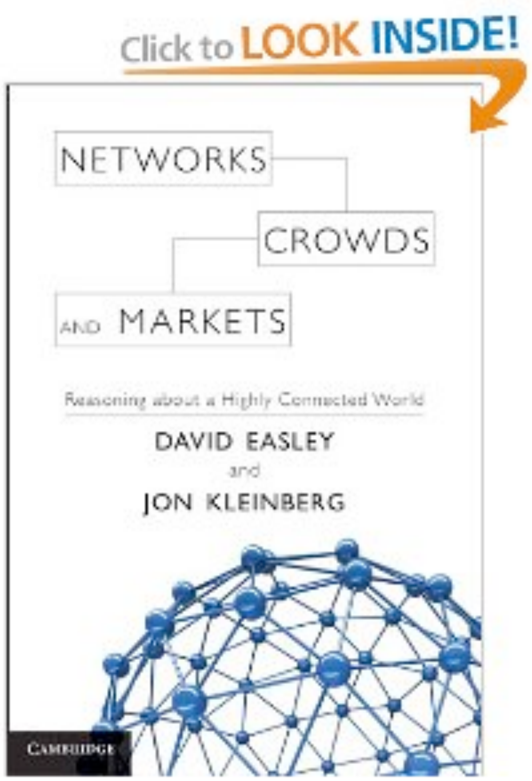

•***Evolution of Networks: From Biological Nets to the Internet and WWW,*** S.N. Dorogovtsev and J.F.F. Mendes. Oxford University Press, Oxford, 2003.




•***Evolution and Structure of the Internet: A Statistical Physics Approach***, R. Pastor-Satorras and A. Vespignani. Cambridge University Press, Cambridge, 2004.

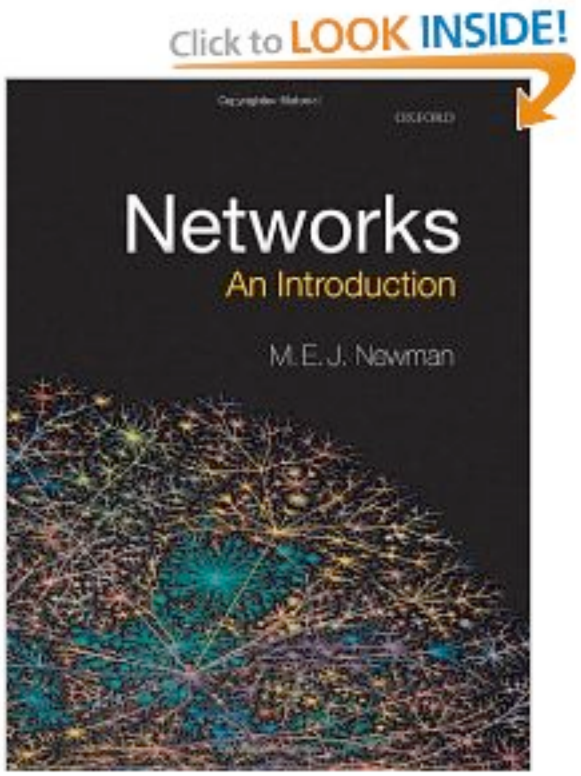



•***Scale-free networks: Complex Webs in Nature and Technology,*** G. Caldarelli. Oxford University Press, Oxford, 2007




**Networks, Crowds, and Markets: Reasoning About a Highly Connected World**

D. Easley, J. Kleinberg

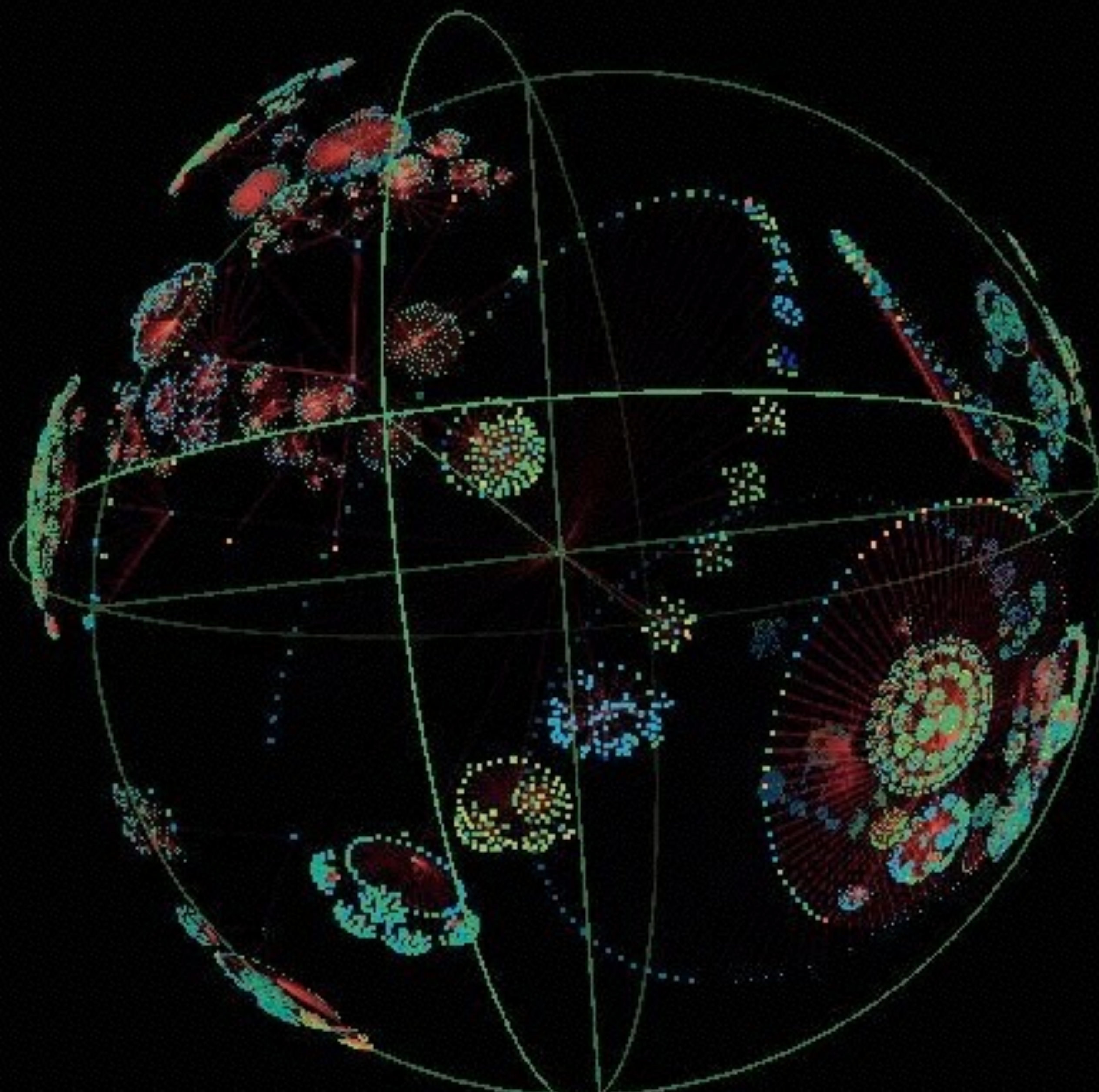



**Networks, An introduction**

M. Newman

# Dynamical Processes on Complex Networks

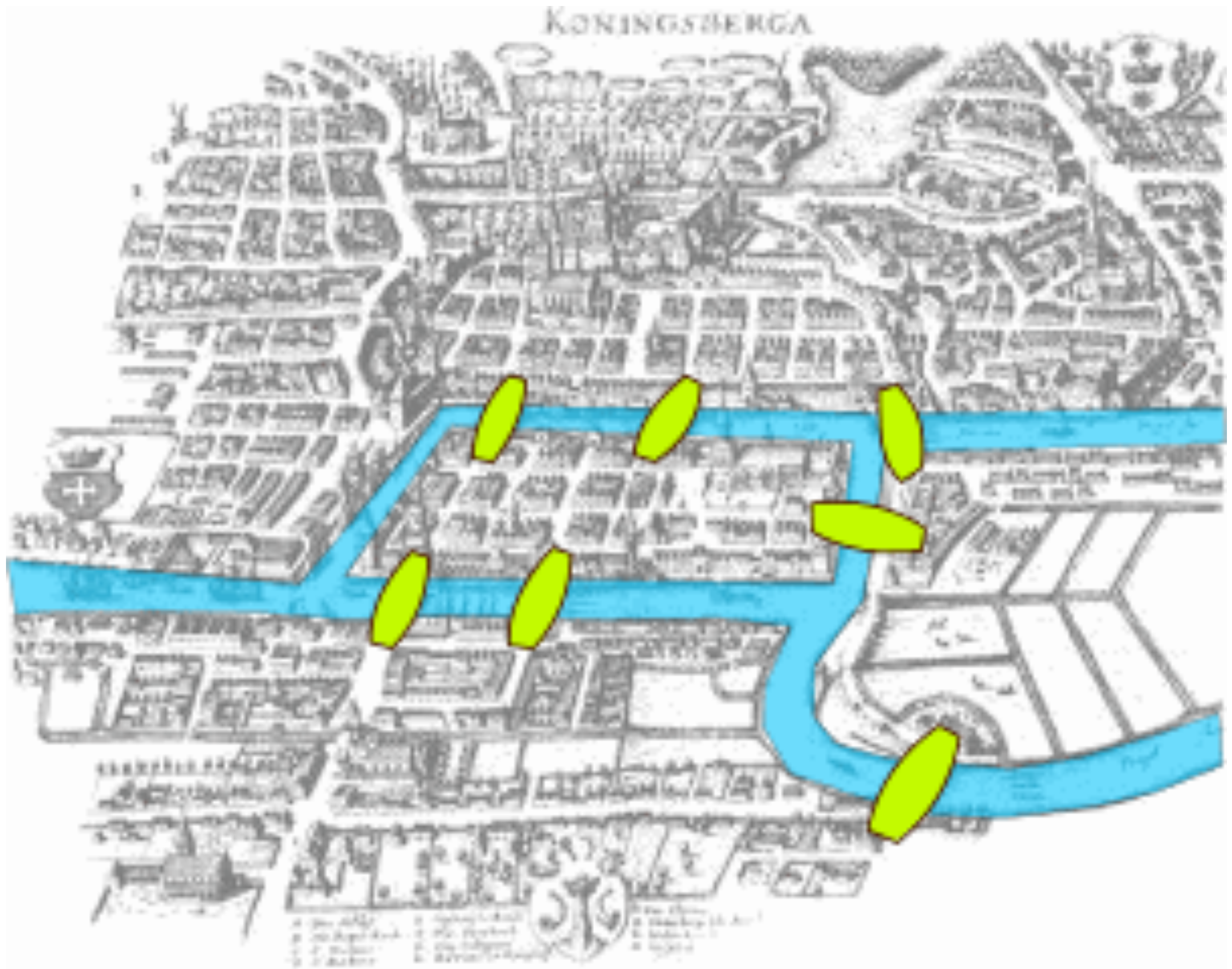
Alain Barrat, Marc Barthélemy, Alessandro Vespignani



CAMBRIDGE

- Introduction
  - Definitions
  - Network statistical characterisation
  - Empirics
- Models
- Processes on networks
  - Resilience
  - Epidemics
- Social Networks analysis

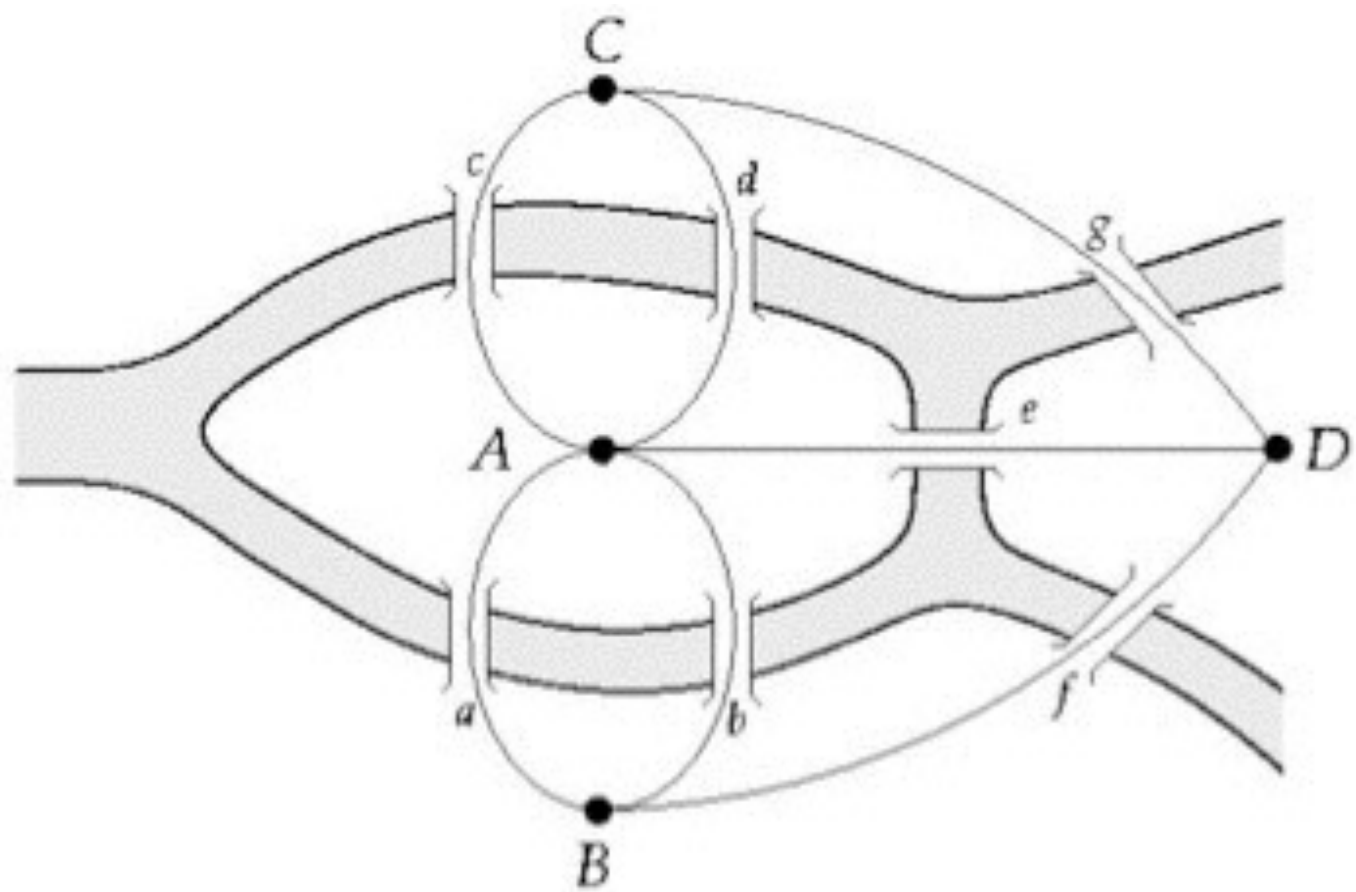
# The bridges of Koenigsberg




**L. Euler:**
**Can one walk once across each of the seven bridges, come back to the starting point and never cross the same bridge twice?**

# Representation of the question as a graph problem

areas = nodes
bridges = links



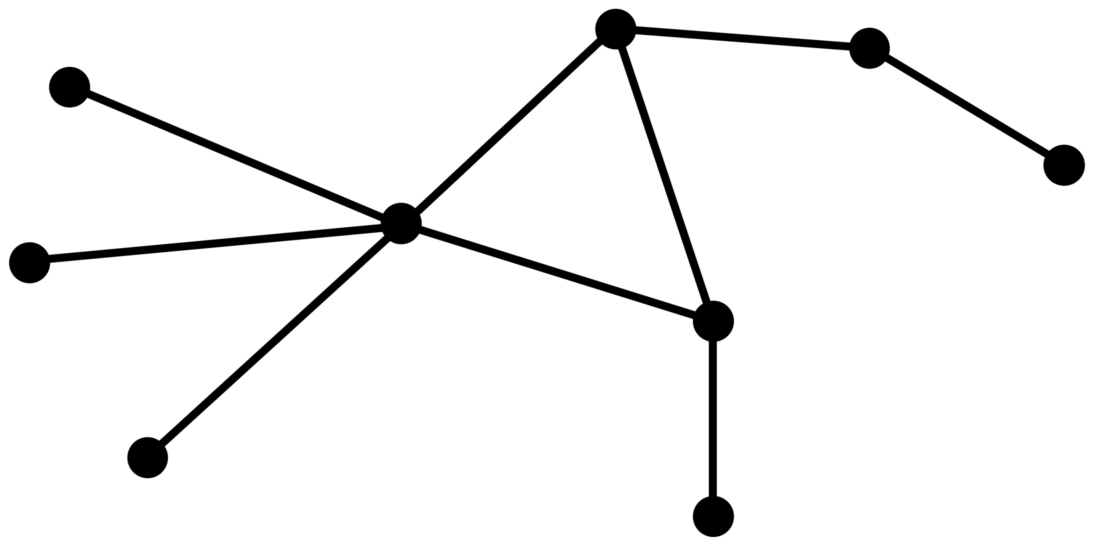

**1735**: **Leonhard Euler's theorem:**

(a) If a graph has nodes of odd degree, there is no path.

(b) If a graph is connected and has no odd degree nodes, it has at least one path.

# Graphs and networks

Graph=set V of nodes joined by links (set E)

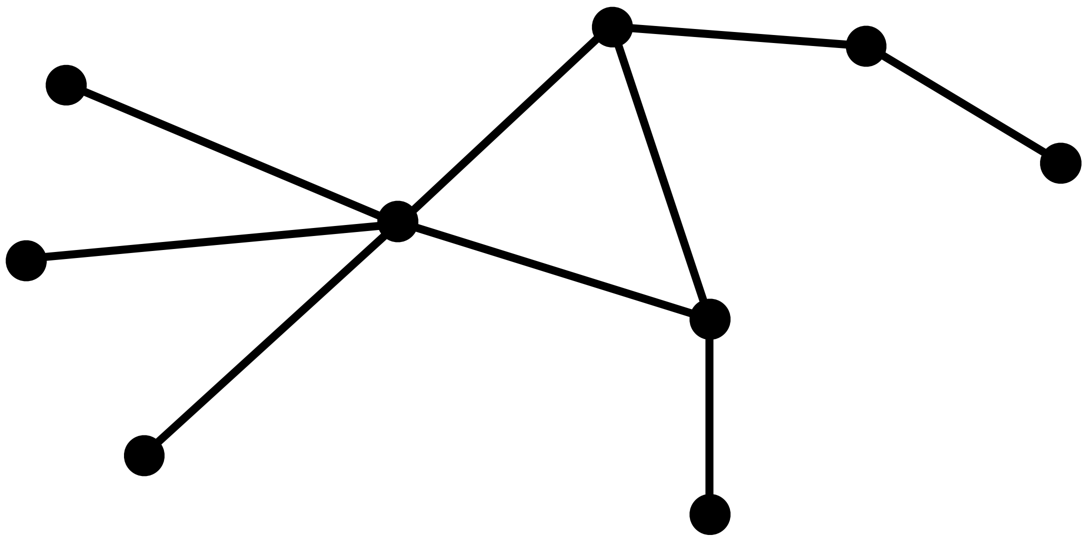


very abstract representation

very general

convenient to represent many different systems

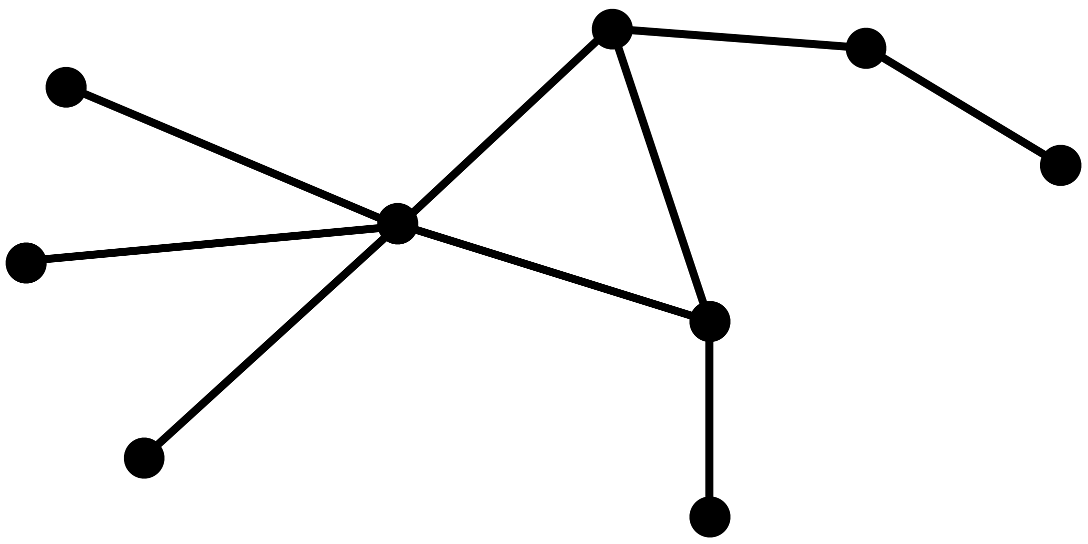
# Graphs



graph theory

abstract tools for the description of graphs (degrees, paths, distances, cliques, etc…)

# Networks

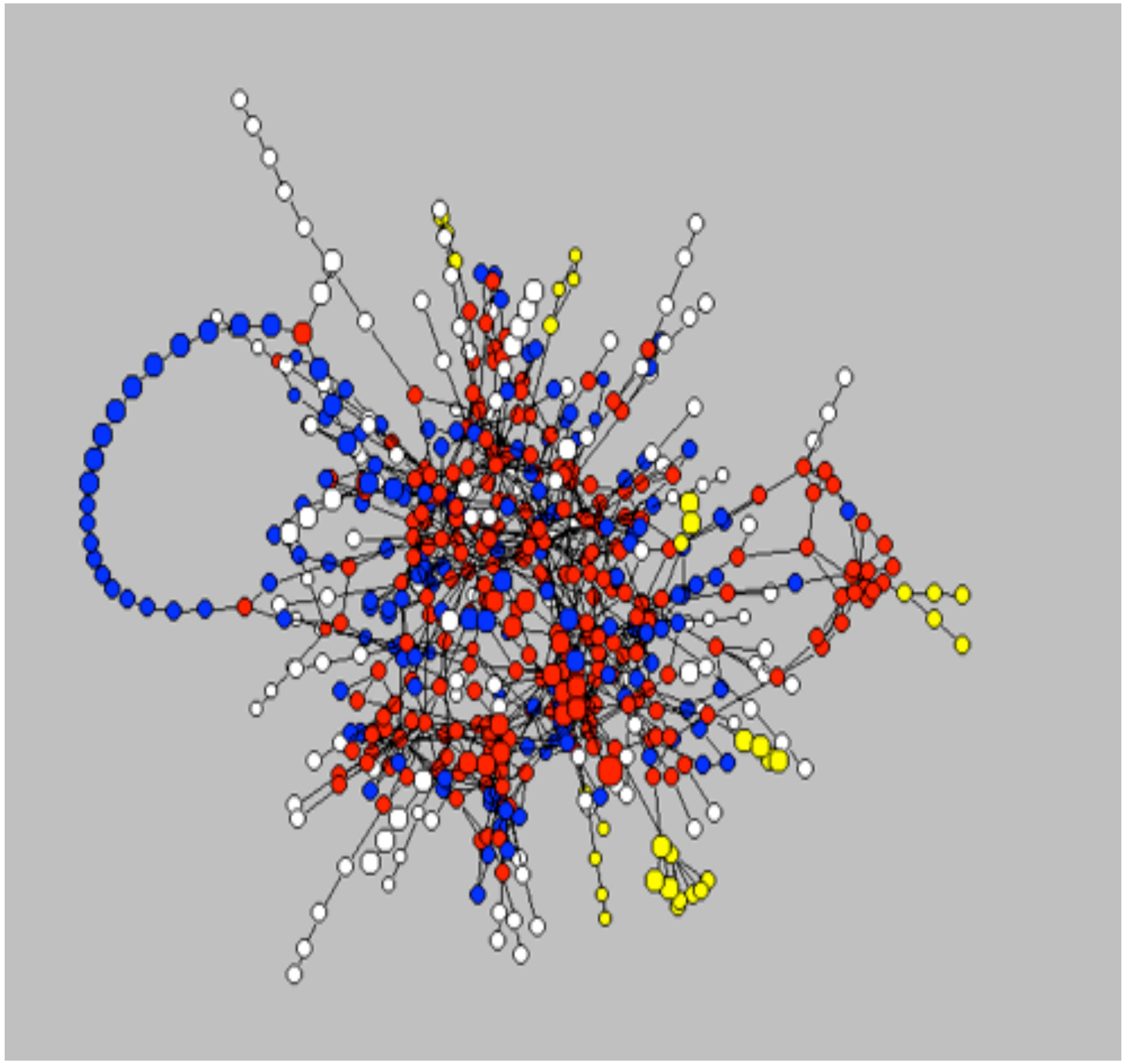


Nodes:
persons
computers
webpages
airports
molecules
....

Links:
social relationships
cables
hyperlinks
air-transportation
chemical reactions
....

## Metabolic Network

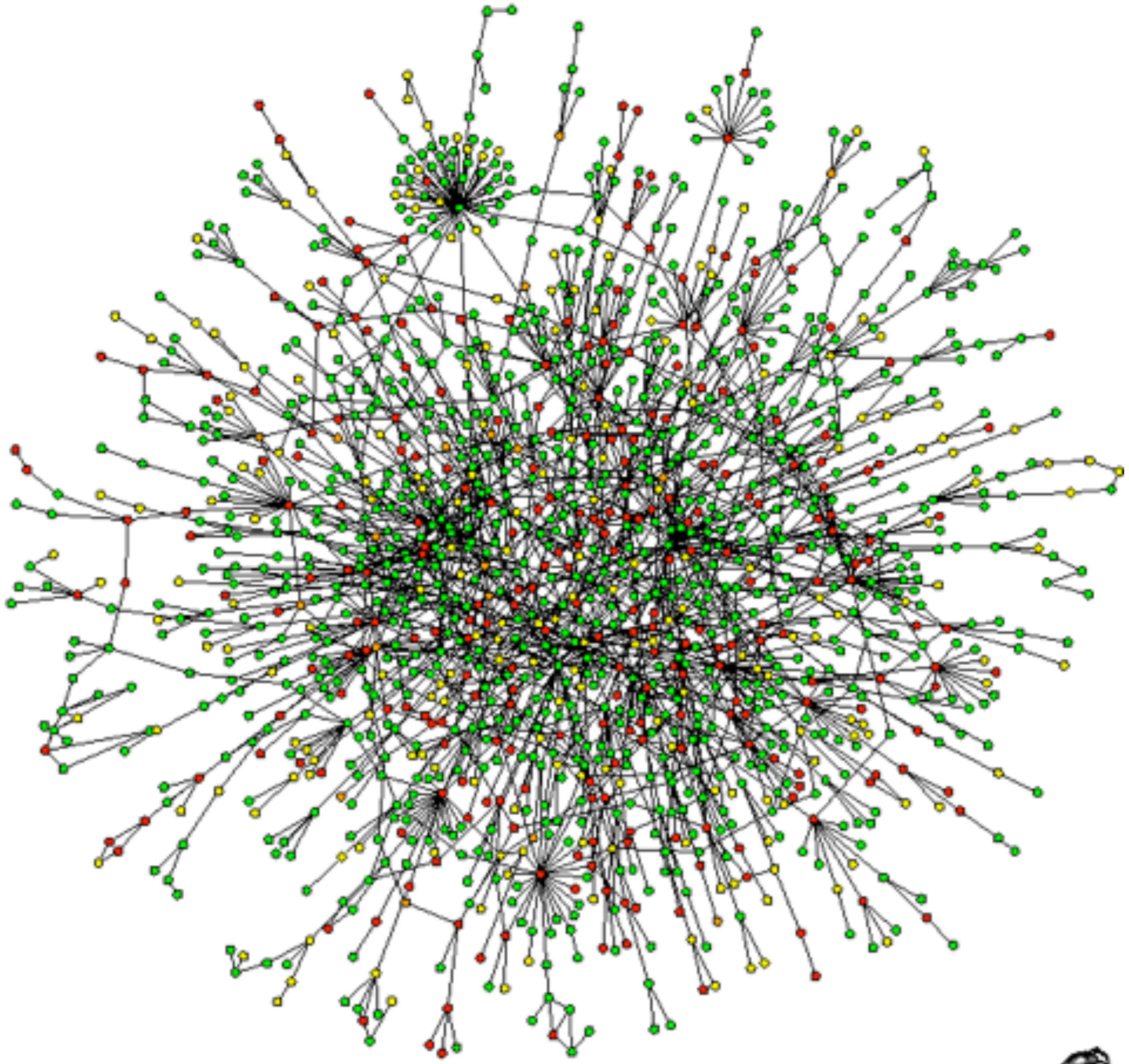
**Nodes**: metabolites
**Links**:chemical reactions



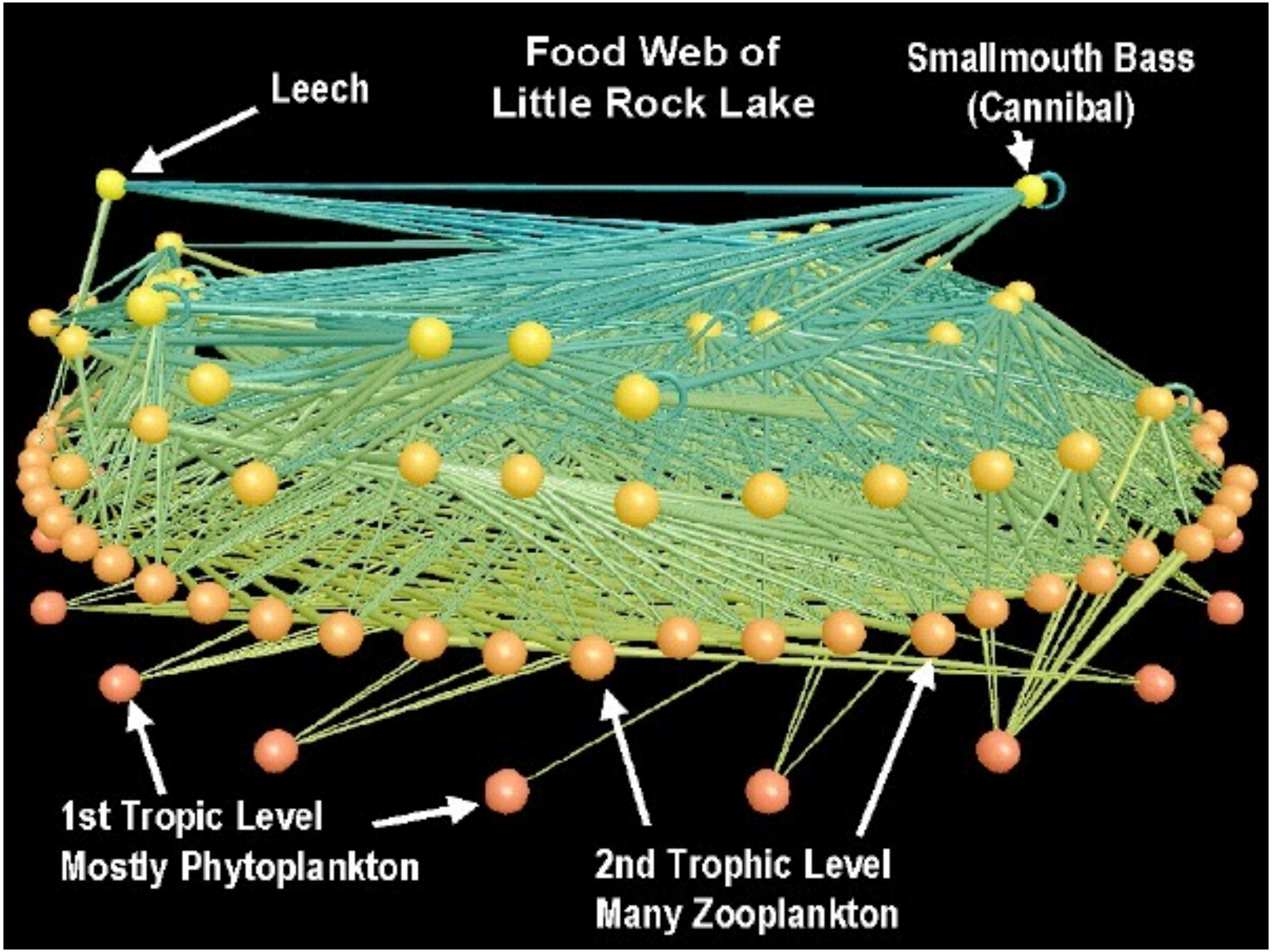
## Protein Interactions

**Nodes**: proteins

**Links**: interactions

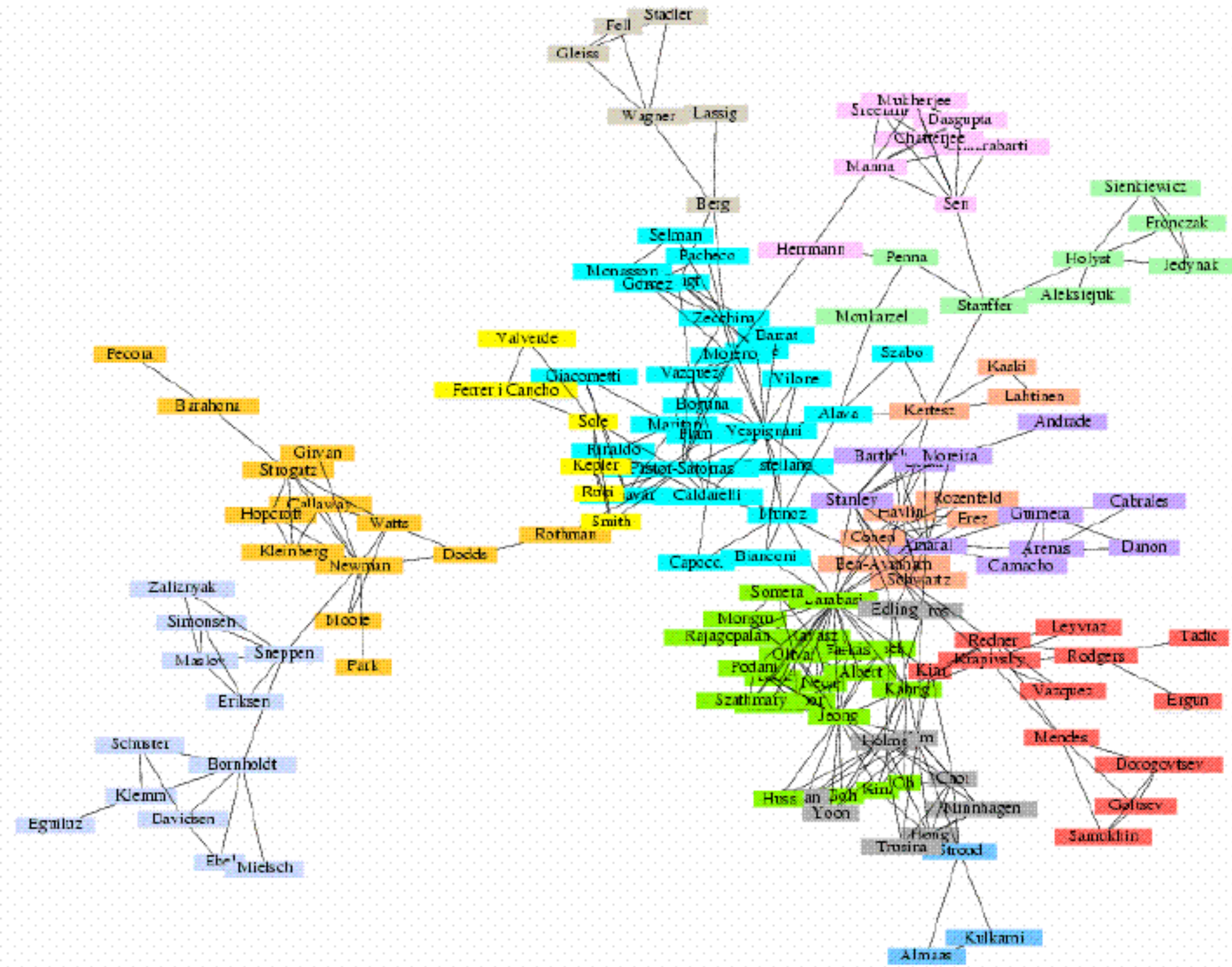


# Food-webs



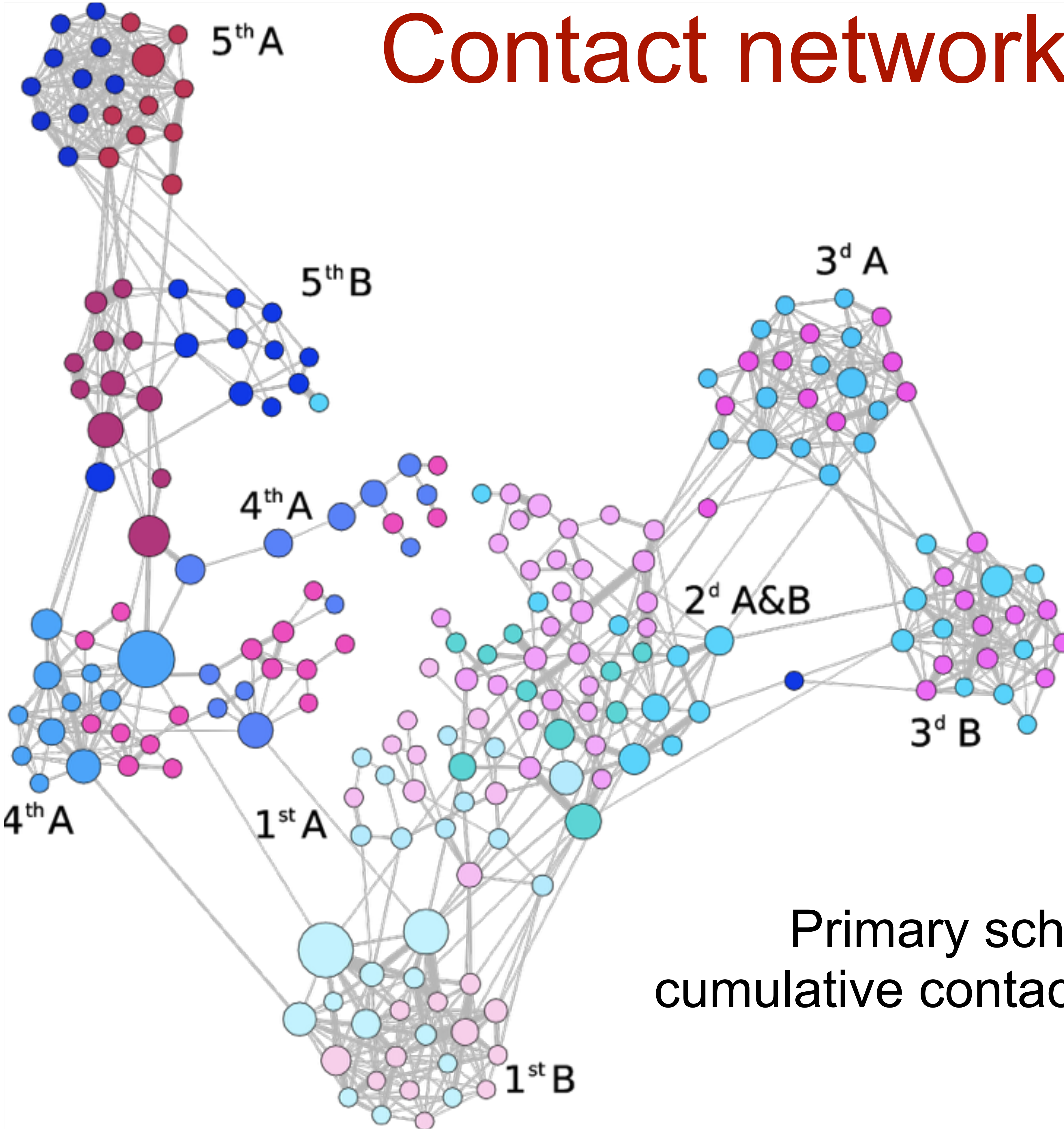

N. Martinez, R. Williams

# Scientific collaboration networks



M. E. J. Newman and M. Girvan, *Physical Review E* **69**, 026113 (2004).
Image: MEJ Newman, http://www-personal.umich.edu/~mejn/networks/

# Contact networks



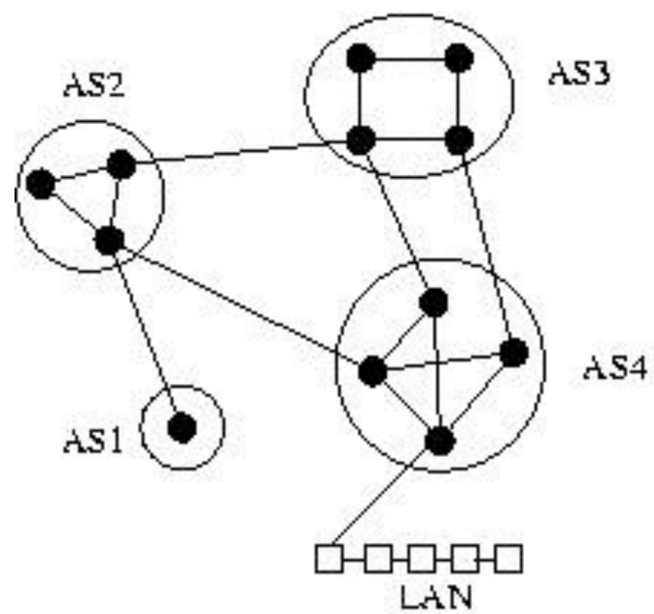

Primary school,
cumulative contact network

# World airport network

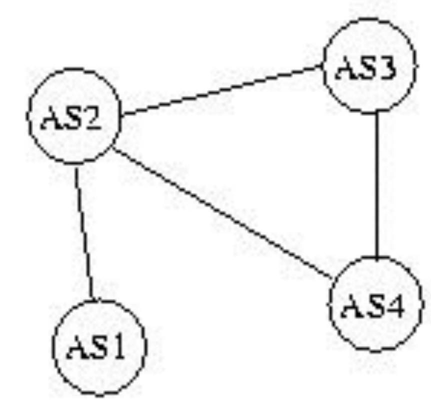
# Internet

## Graph representation

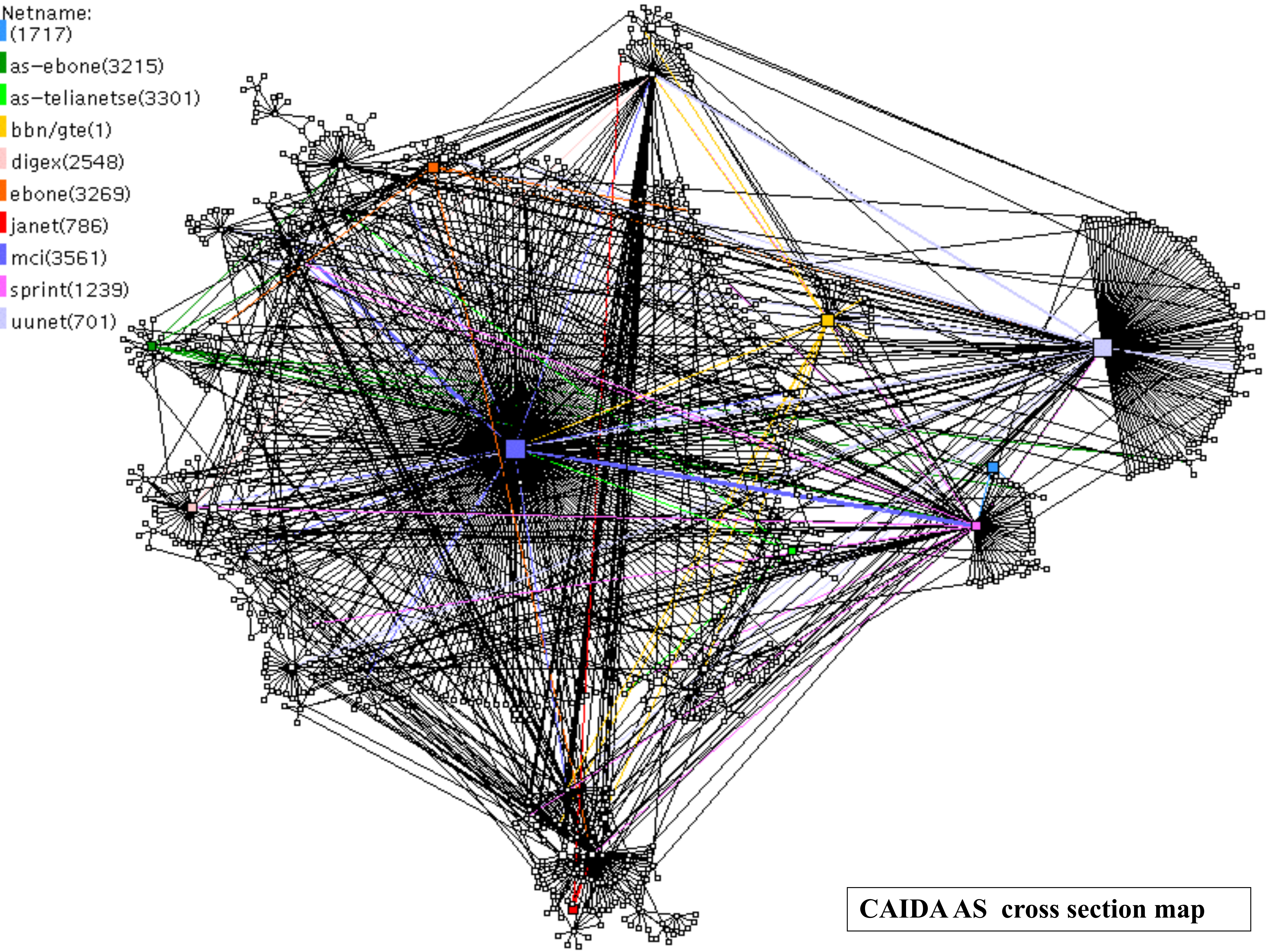
**different granularities**




Router Level




Autonomous System level




**CAIDA AS cross section map**

# Online (virtual) social networks

# Networks & Graphs

Networks: of very different origins

Do they have anything in common?
Possibility to find common properties?

the abstract character of the graph representation and graph theory allow to answer….

# Interdisciplinary science

Science of complex networks
("Network science")
-graph theory
-social sciences
-communication science
-biology
-physics
-computer science

Data-driven
Tools both from graph theory and outside graph theory

# Interdisciplinary science

Science of complex networks:

- Empirics
- Characterization
- Modeling
- Dynamical processes
- … and more…

Data-driven
Tools both from graph theory and outside graph theory

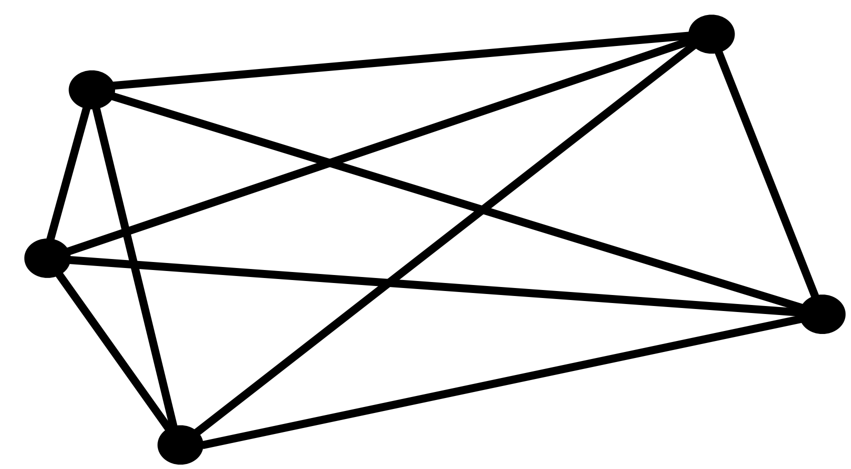
# Graph theory: basics

Graph: $G=(V,E)$ ; $|V|=N$

Maximum number of edges

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

Complete graph:



(all to all interaction/communication)

# How to represent a network

- List of nodes + list of edges

  i,j

- List of nodes + list of neighbors of each node (adjacency lists)

  1: 2,3,10,...

  2: 1,12,11

  3: 1,...

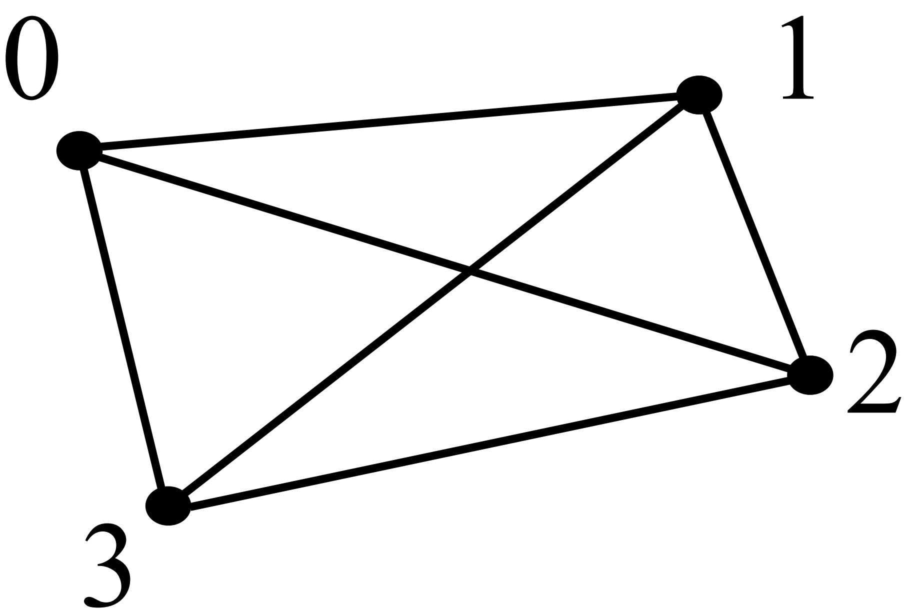
- Adjacency matrix

# Adjacency matrix

N nodes i=1,…,N

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

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



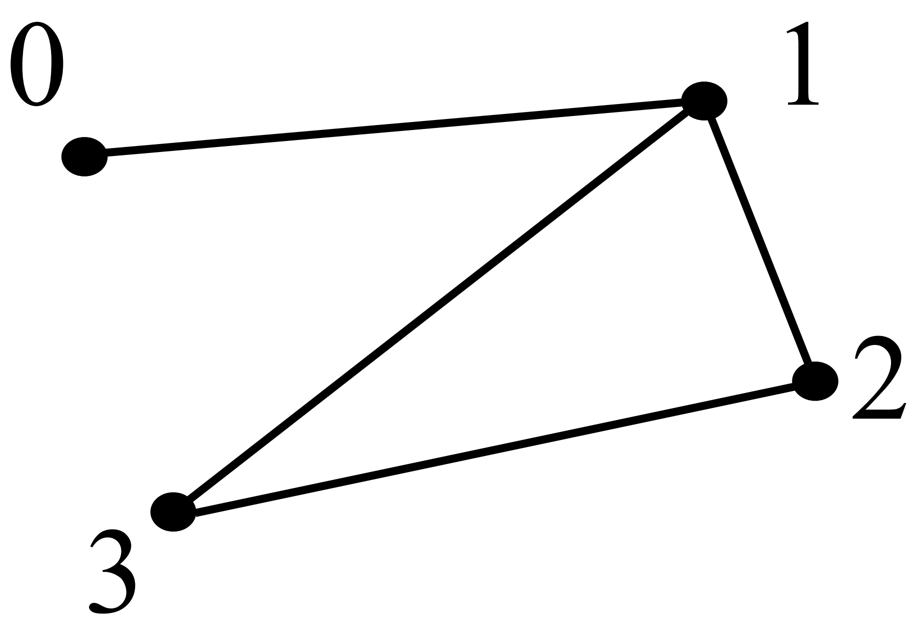
# Adjacency matrix

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 |



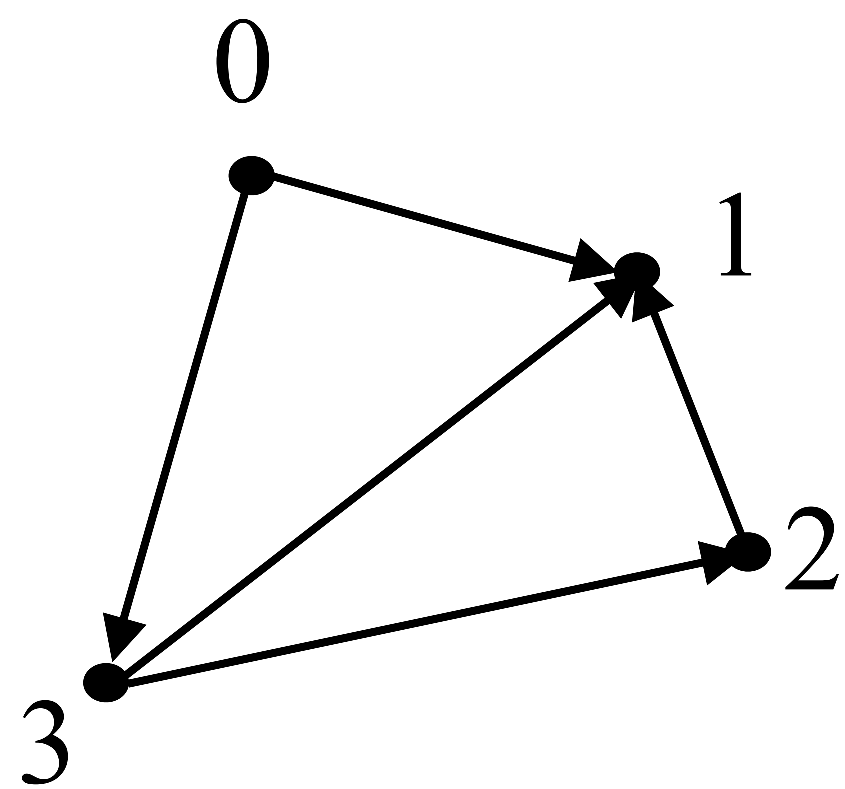
# Adjacency matrix

N nodes i=1,…,N

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

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

Non symmetric
for directed networks



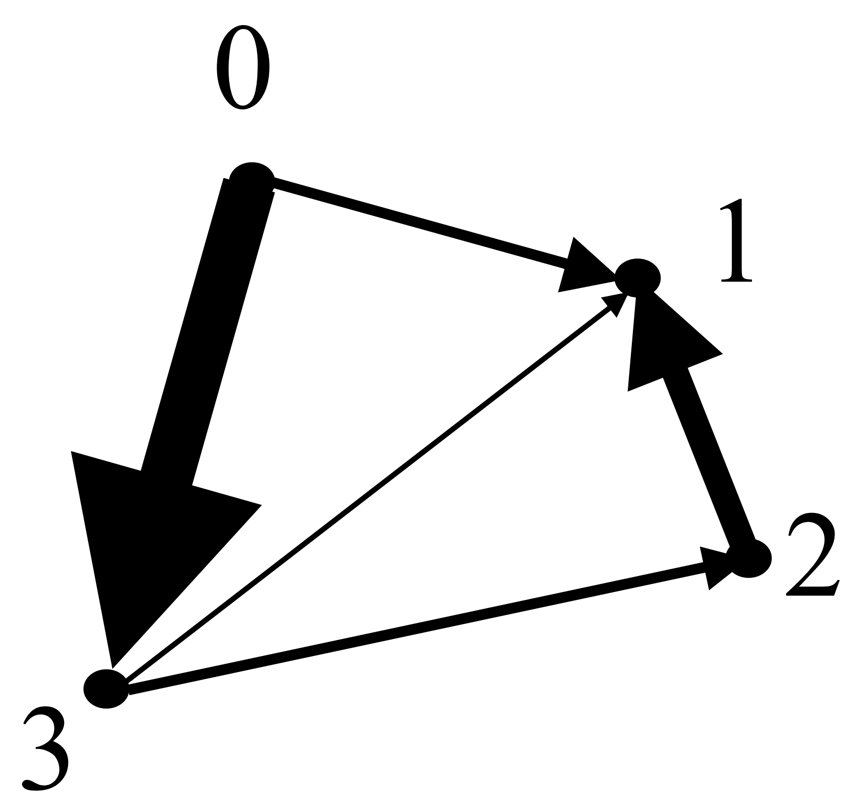
# Matrix of weights

N nodes i=1,…,N

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

(Non symmetric
for directed networks)

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 2 | 0 | 10 |
| 1 | 0 | 0 | 0 | 0 |
| 2 | 0 | 5 | 0 | 0 |
| 3 | 0 | 1 | 2 | 0 |



# Sparse graphs

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

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

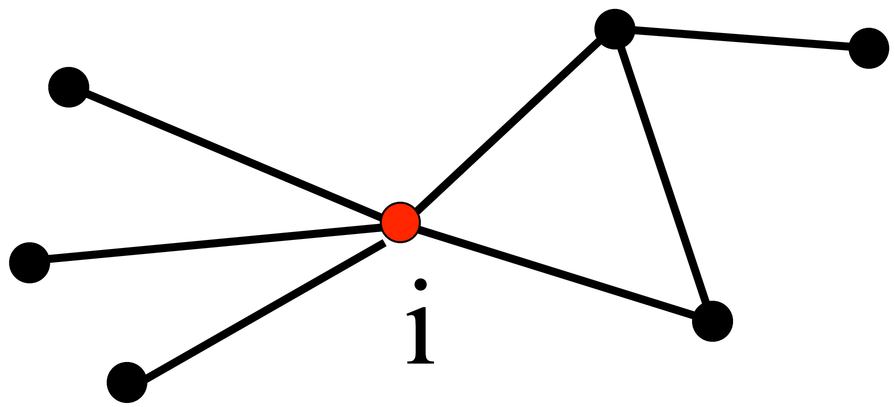
Sparse graph: D <<1 → Sparse adjacency matrix

→ Representation by lists of neighbours of each node (adjacency lists) better suited

# Node characteristics: Degrees and strengths

# Node characteristics

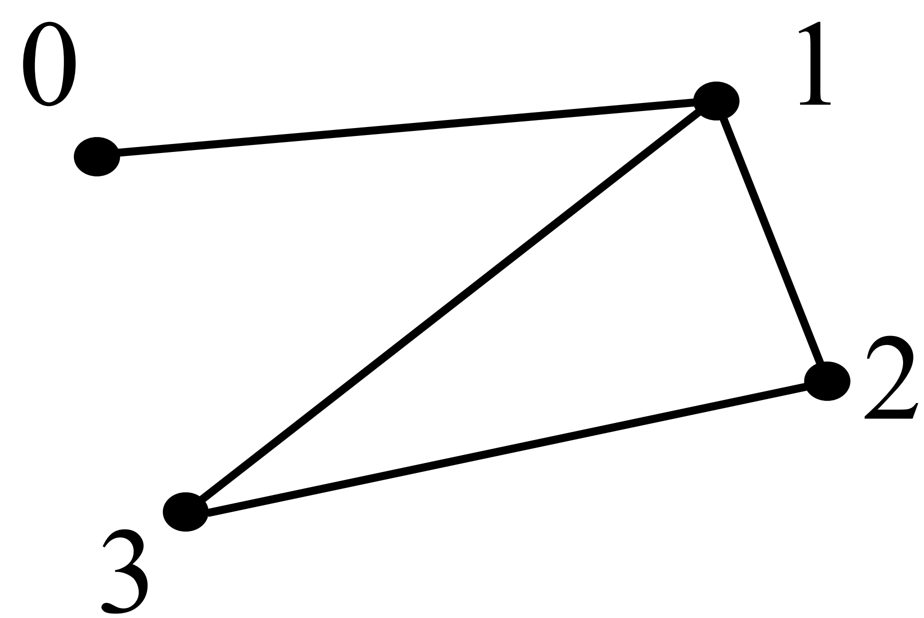
- Degree=number of neighbours=$\sum_j a_{ij}$




$k_i$=5

NB: in a sparse graph we expect $k_i << N$

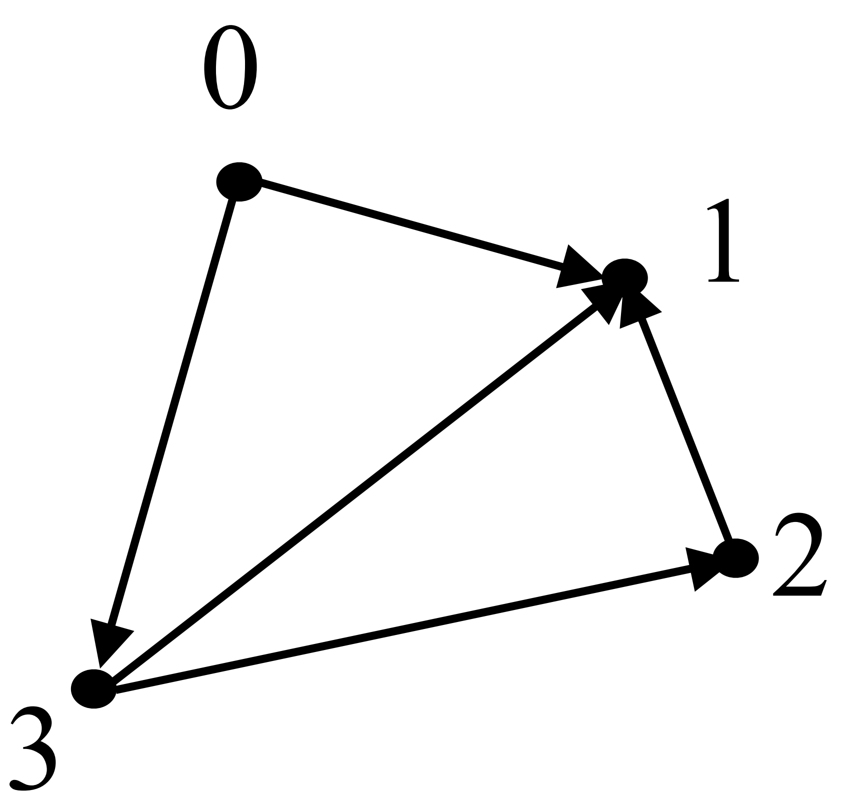
| | | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|
| | 0 | 0 | 1 | 0 | 0 |
| | 1 | 1 | 0 | 1 | 1 |
| i | 2 | 0 | 1 | 0 | 1 |
| | 3 | 0 | 1 | 1 | 0 |



# Node characteristics

- Degree in directed graphs:
  - in-degree= number of in-neighbours= $\sum_j a_{ji}$
  - out-degree= number of out-neighbours= $\sum_j a_{ij}$

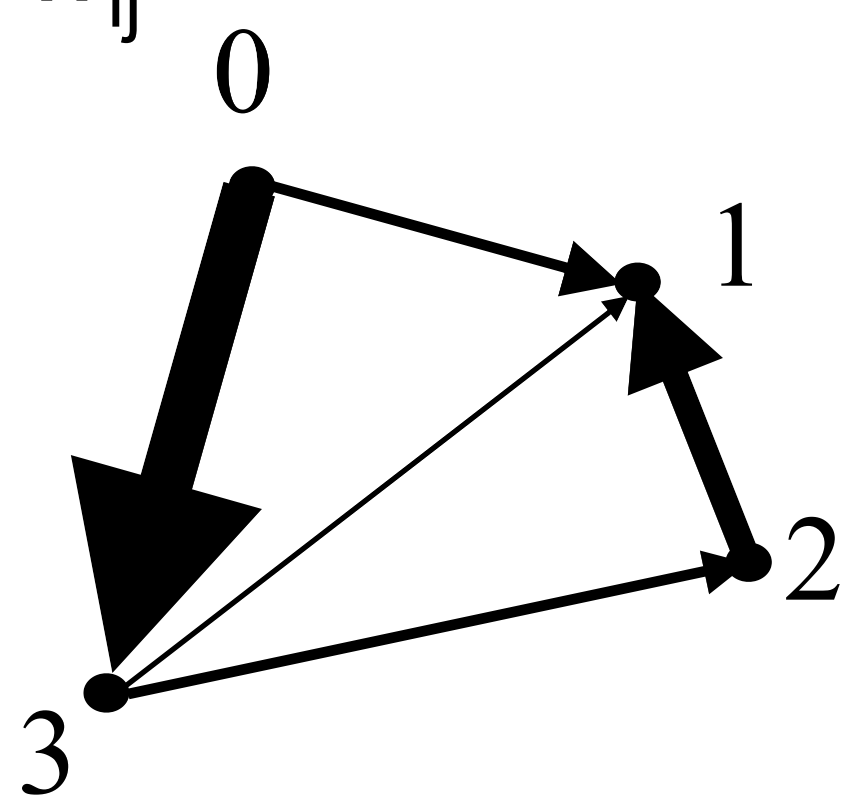
| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 1 |
| 1 | 0 | 0 | 0 | 0 |
| 2 | 0 | 1 | 0 | 0 |
| 3 | 0 | 1 | 1 | 0 |



# Node characteristics

- Weighted graphs: Strength $s_i = \sum_j w_{ij}$
- Directed Weighted graphs:
  - in-strength $s_i = \sum_j w_{ji}$
  - out-strength $s_i = \sum_j w_{ij}$

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0 | 2 | 0 | 10 |
| 1 | 0 | 0 | 0 | 0 |
| 2 | 0 | 5 | 0 | 0 |
| 3 | 0 | 1 | 2 | 0 |



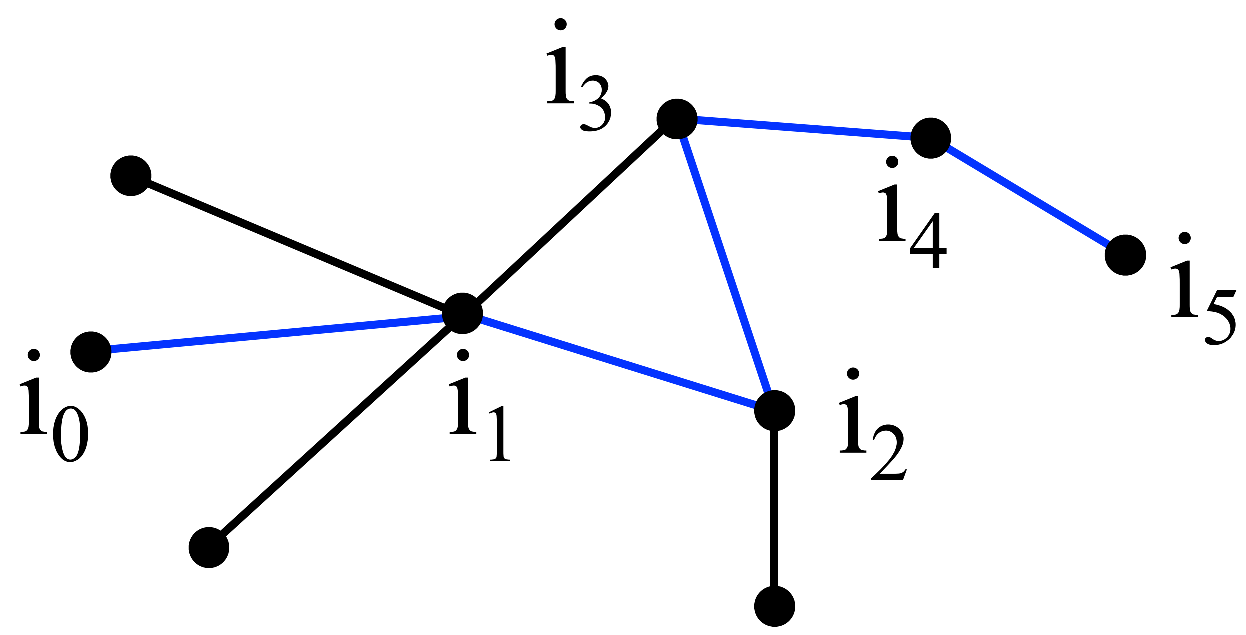
# Paths, connectedness, small-world effect

# Paths

G=(V,E)

Path of length n = ordered collection of

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

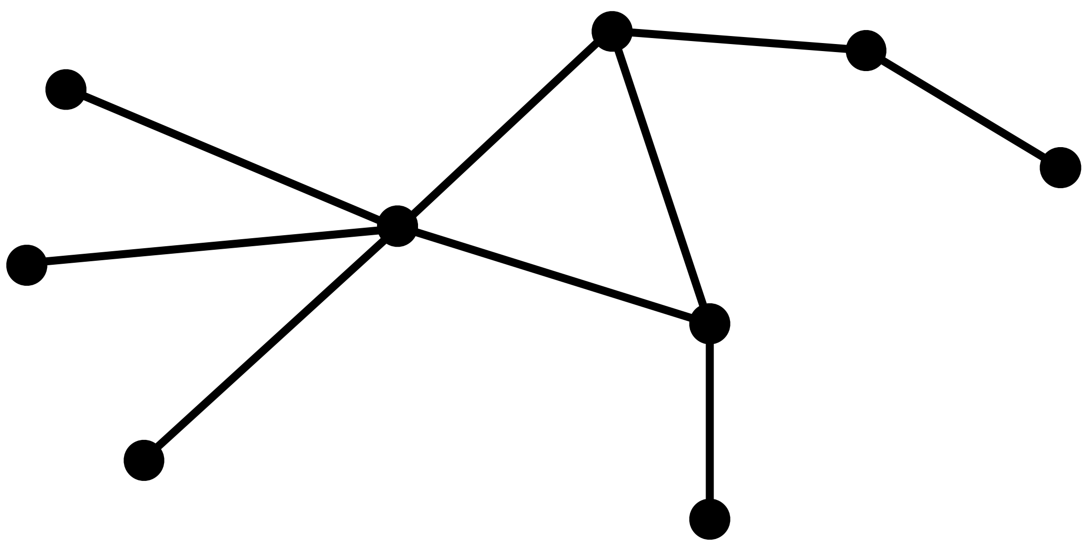



Cycle/loop = closed path ($i_0=i_n$)

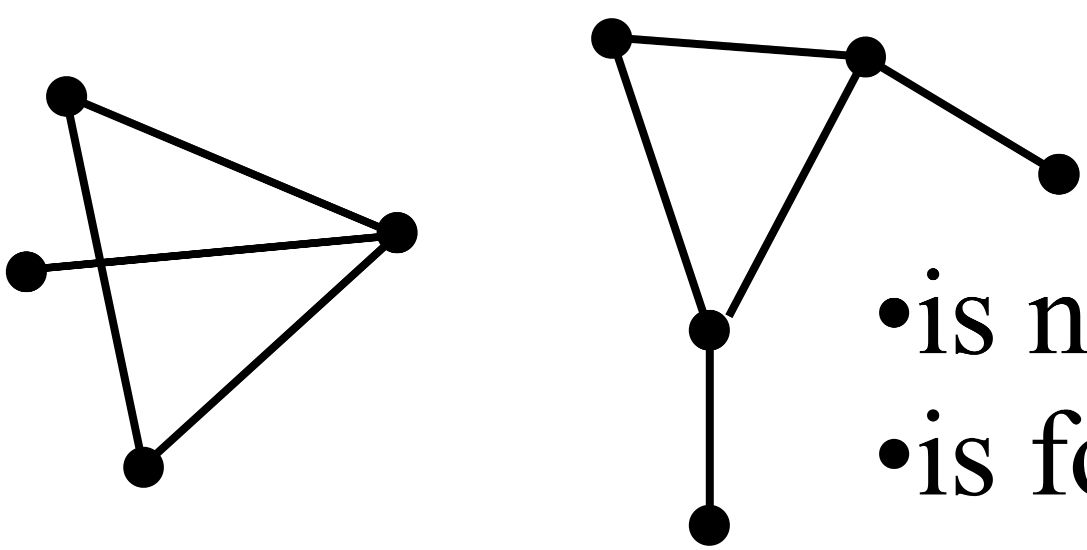
Tree=graph with no loops

# 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

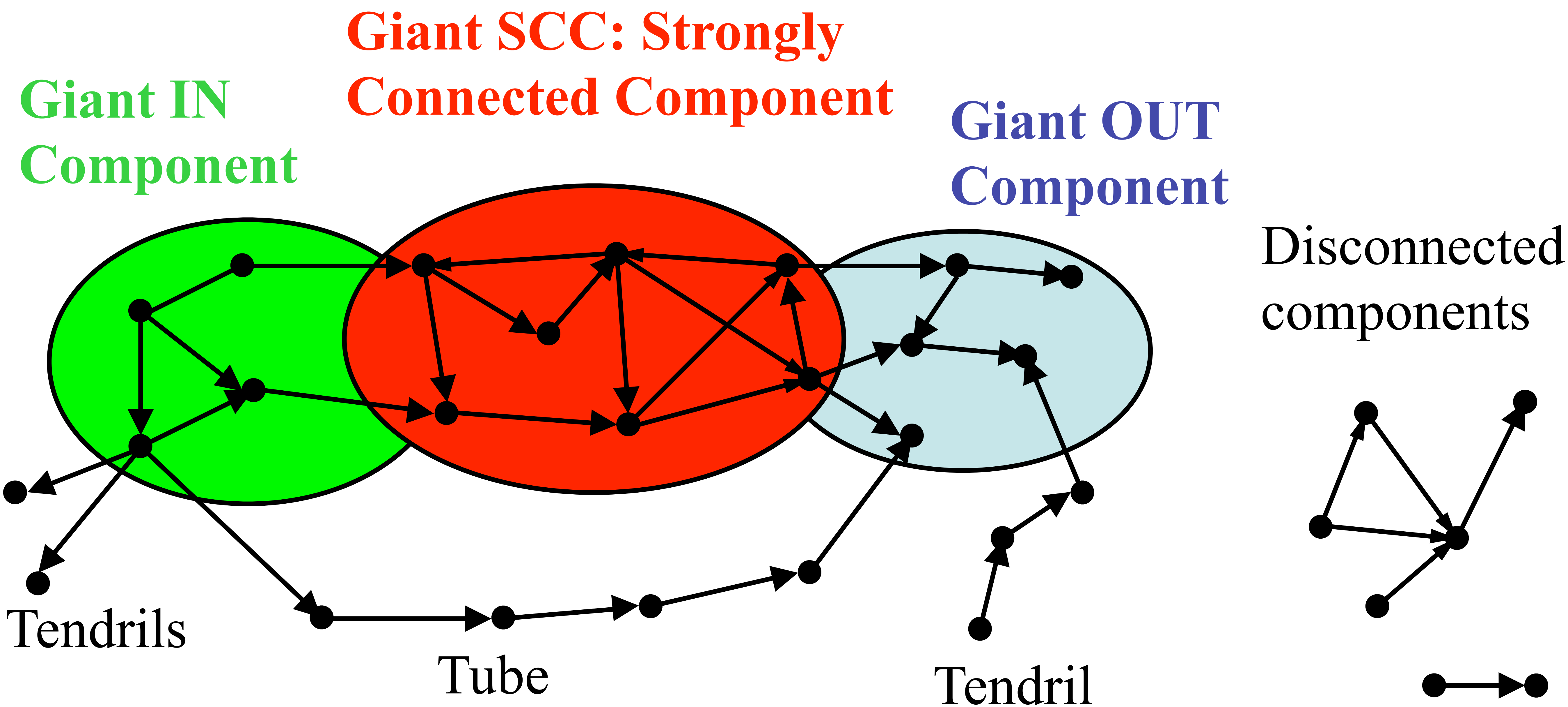
G=(V,E)=> distribution of components' sizes

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

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

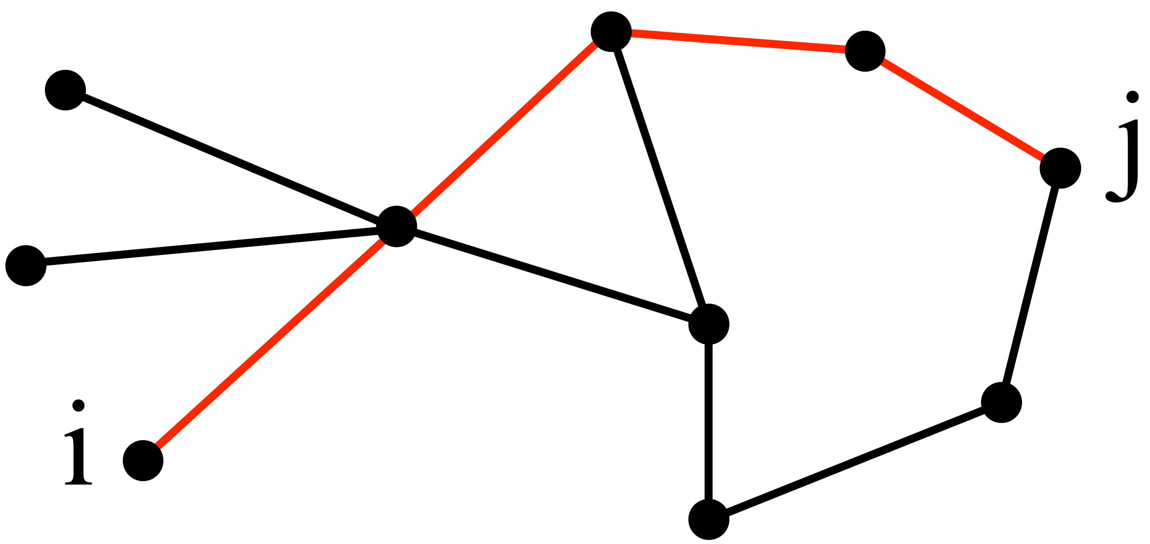
# Paths and connectedness: directed graphs

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

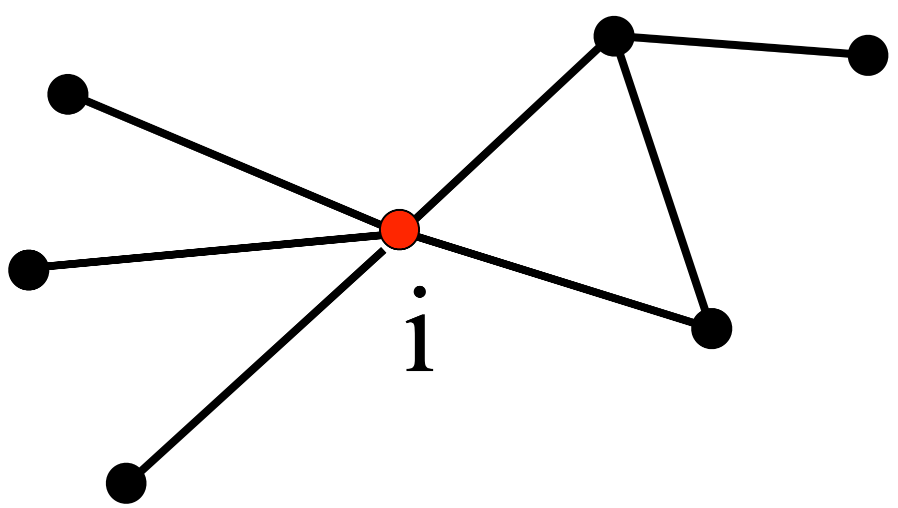
"Small-world": "small" diameter

# Ranking nodes

# Centrality measures

How to quantify the importance of a node?

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




$k_i=5$

- Large degree nodes=”hubs”
- Nodes with very large degree can be “peripheral”

# Path-based centrality measures

- Closeness centrality

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

Quantifies the reachability of other nodes from i

# Betweenness centrality

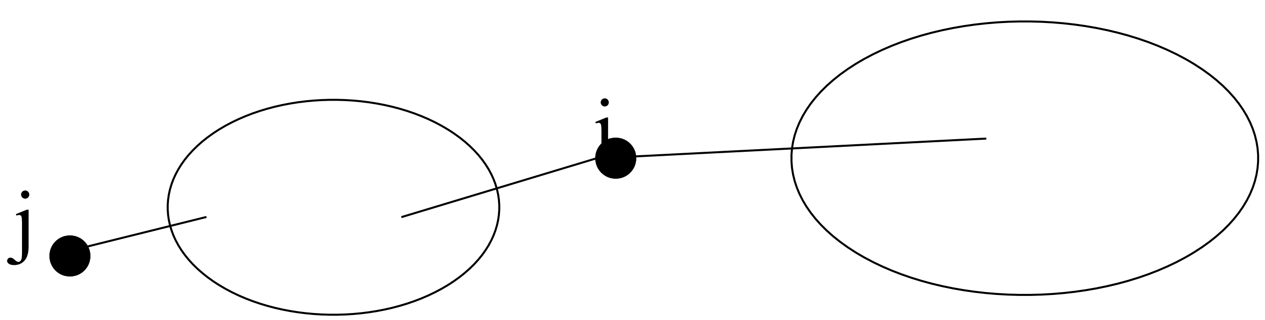
for each pair of nodes (l,m) in the graph, there are

$\sigma^{lm}$ shortest paths between l and m

$\sigma_i^{lm}$ 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*

# Betweenness centrality

path-based quantity => bc(i) depends on all the nodes that are connected to i by at least one path

non-local quantity

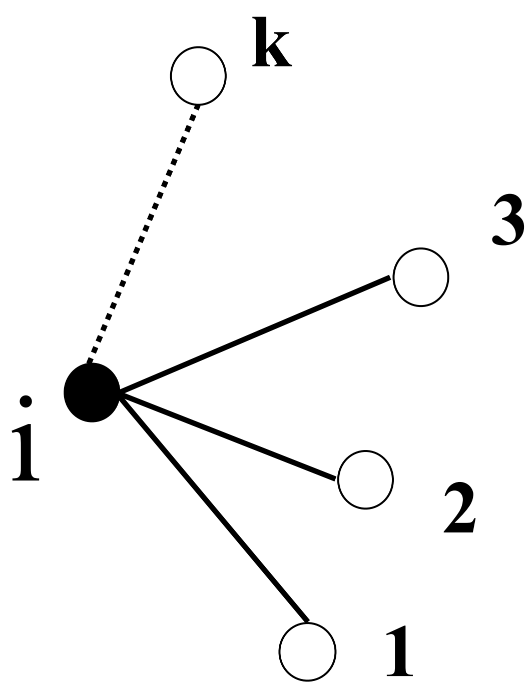
"hard" to compute

"naive" algorithm: $O(N^3)$

Brandes algorithm: $O(N*E)$

# Local structures; subgraphs; communities

# Structure of neighborhoods

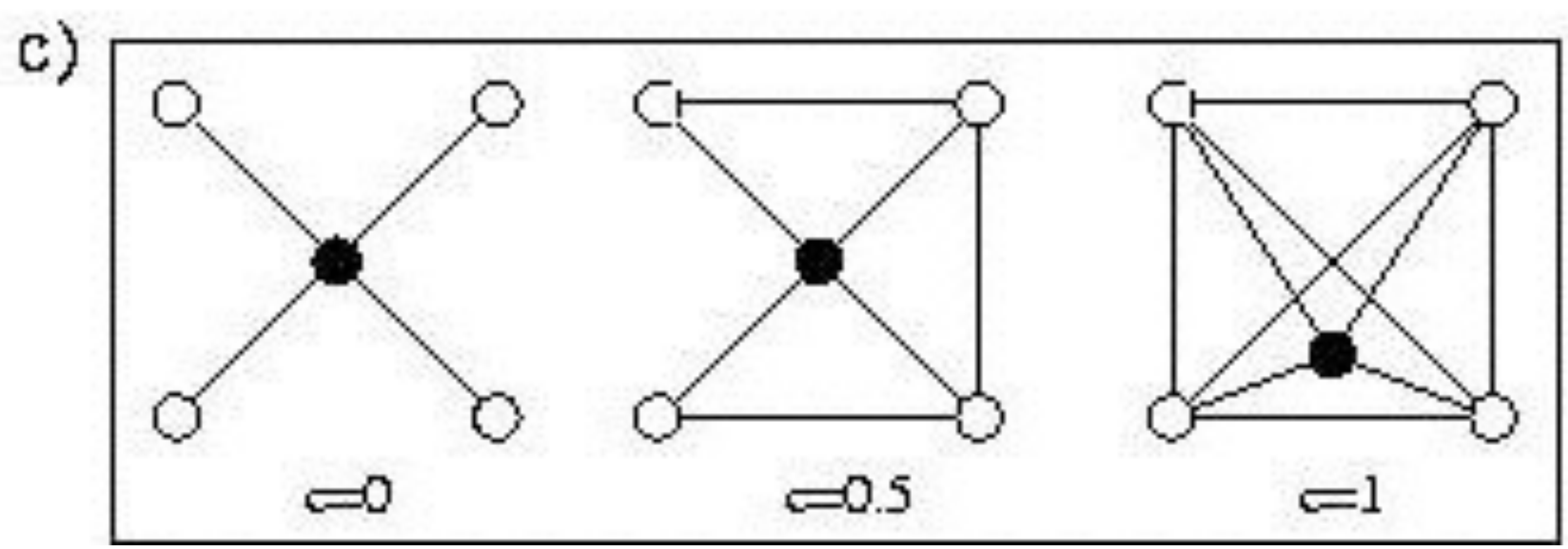



## Clustering coefficient of a node

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

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

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



# Subgraphs

A subgraph of G=(V,E) is a graph G'=(V',E')
such that
V' ⊆ V and E' ⊆ E
i.e., V' and E' are subsets of nodes and edges of G

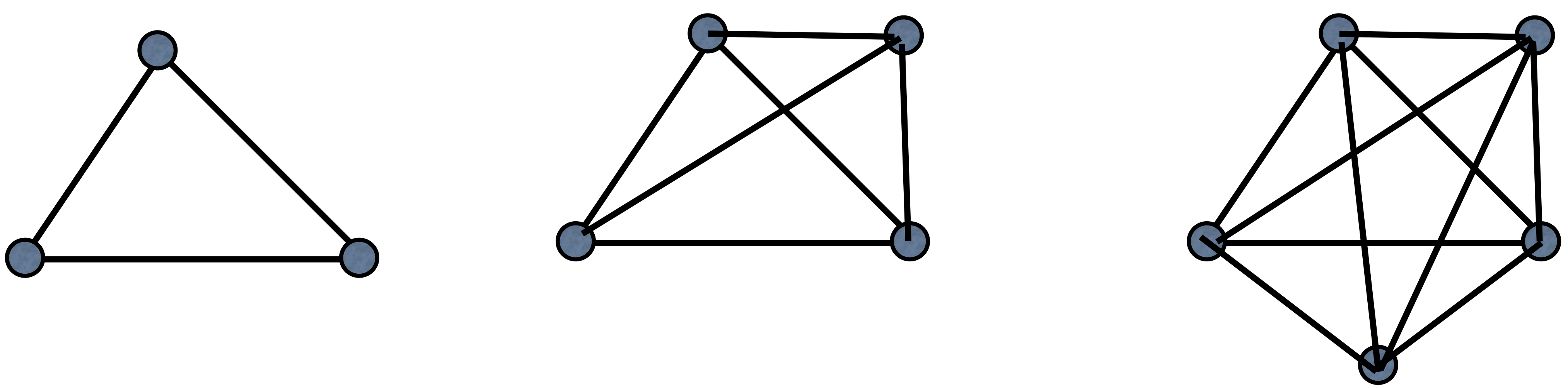
Special case: subgraph *induced* by a set of nodes=
-this set of nodes
-and all links of G between these nodes

Particular subgraphs=connected components

# Cliques

A clique is a set C of nodes of G=(V,E) such that for all i,j ∈ C, (i,j)∈ E

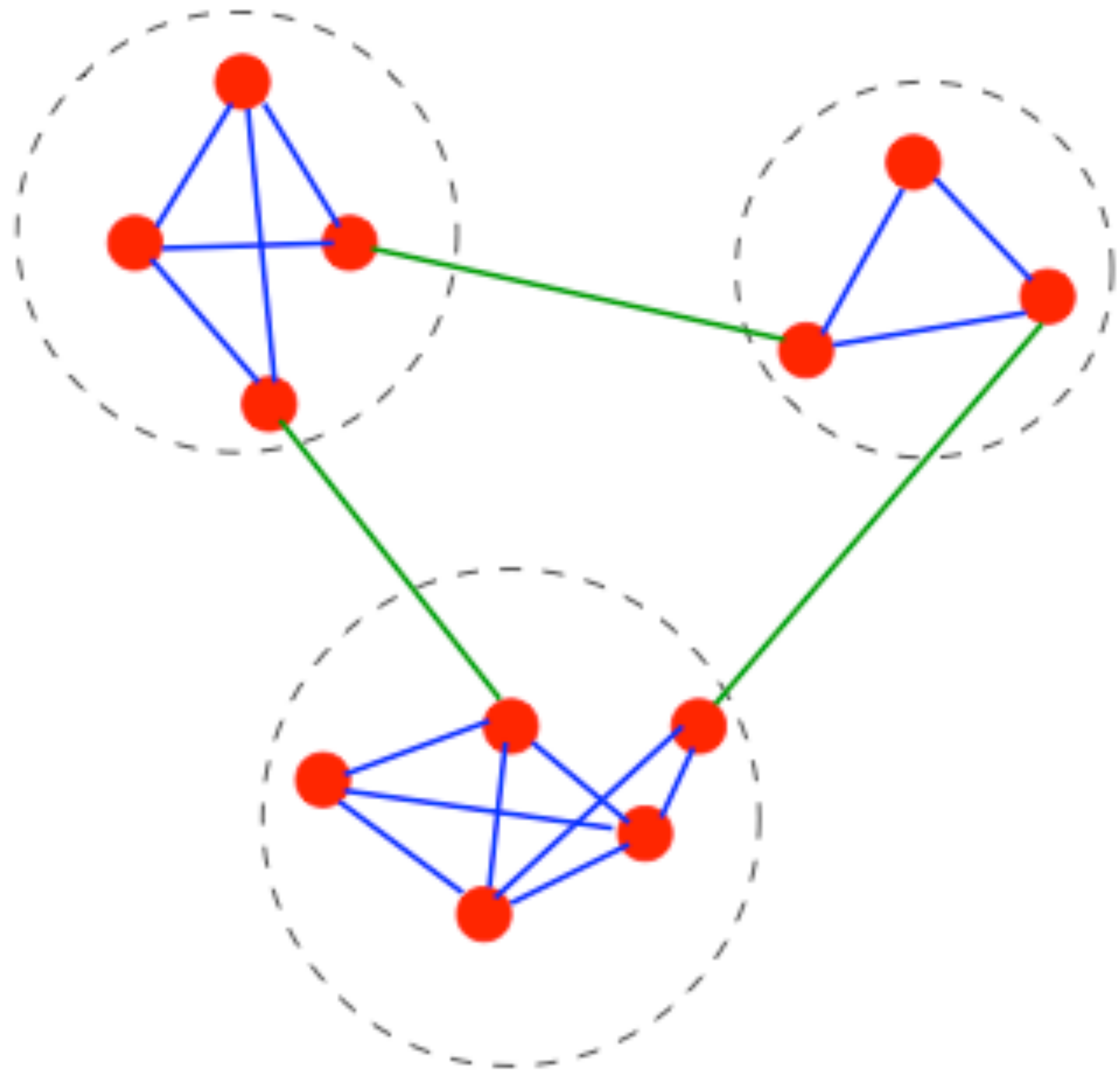
Examples:



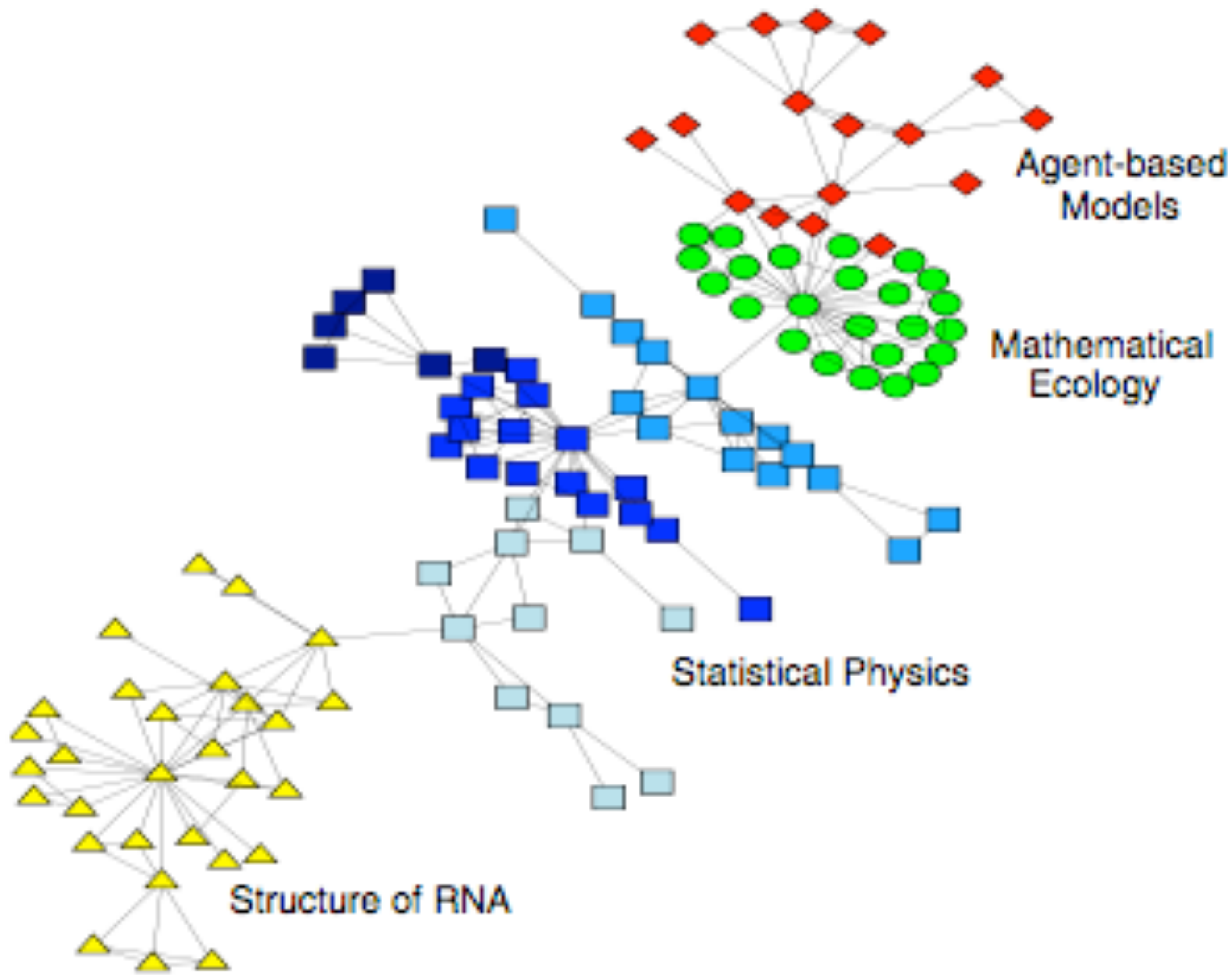
# Communities: (loose) definition

Group of nodes
that are more tightly linked together
than with the rest of the graph

# Communities: examples

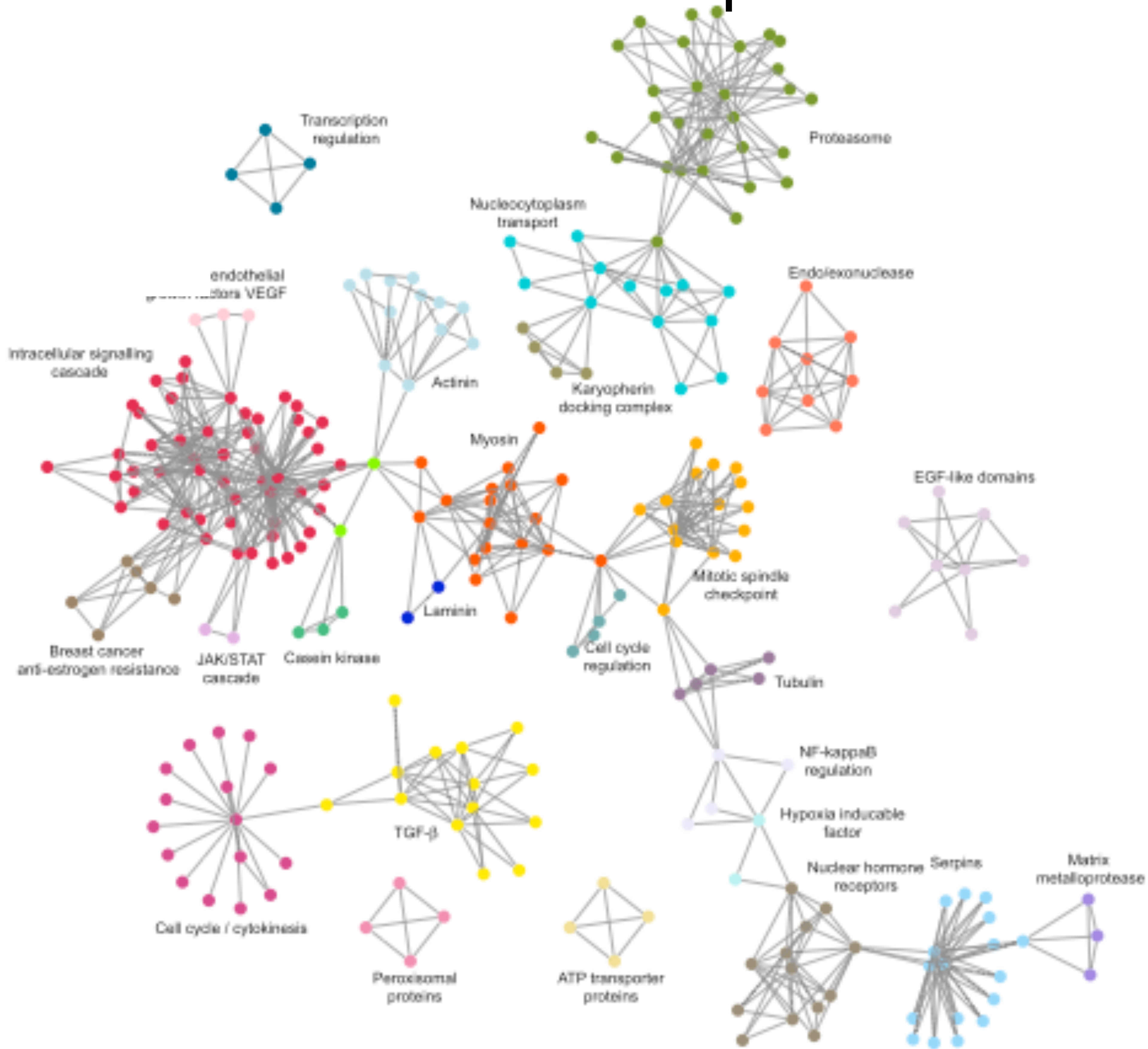


# Communities: examples




Scientist collaboration network
(Santa Fe Institute)

# Communities: examples




Protein-protein interaction network

# Why are communities interesting?

Node classification,
prediction of unknown characteristics/function

Discover groups in social networks,
bottom-up classification

Discover common interests
Recommendation systems

Understand role of communities in dynamical processes,
e.g. spreading or opinion formation mechanisms

# Community detection

Group of nodes
that are more tightly linked together
than with the rest of the graph

- How to (systematically) detect such groups?
- How to partition a graph into communities?
- How to check if it makes sense?

# Community detection

- Huge literature
- Tricky and much debated issue
- Many algorithms available, most often open source

http://www.cfinder.org/
http://www.oslom.org/
http://www.tp.umu.se/~rosvall/code.html

For a review
S. Fortunato, Phys. Rep. **486**, 75-174, 2010
(http://sites.google.com/site/santofortunato/)

# Hierarchies

# A way to measure hierarchies: K-core decomposition

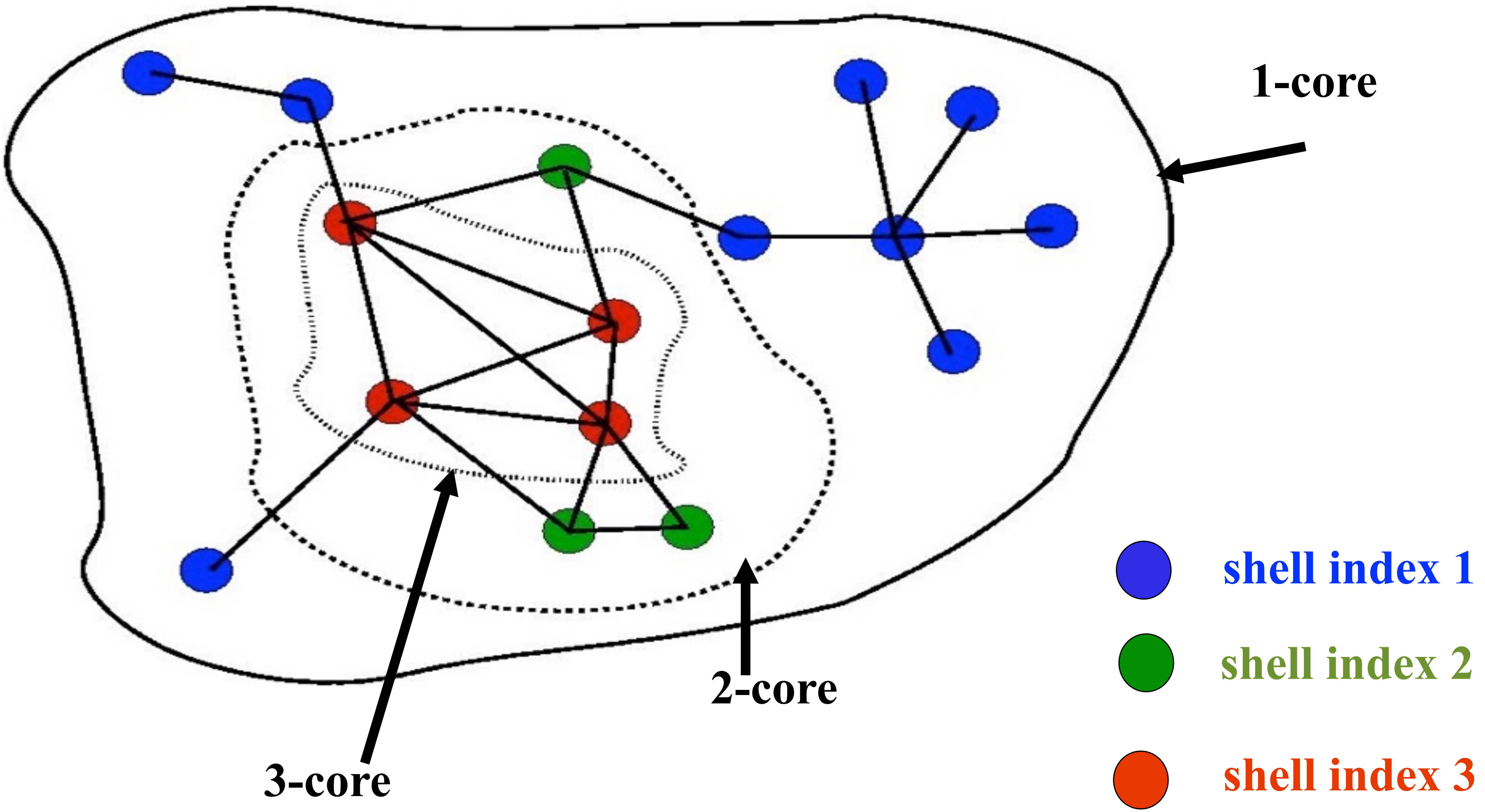
**graph G=(V,E)**

**–k-core of graph G: maximal subgraph such that for all vertices in this subgraph have degree at least k**

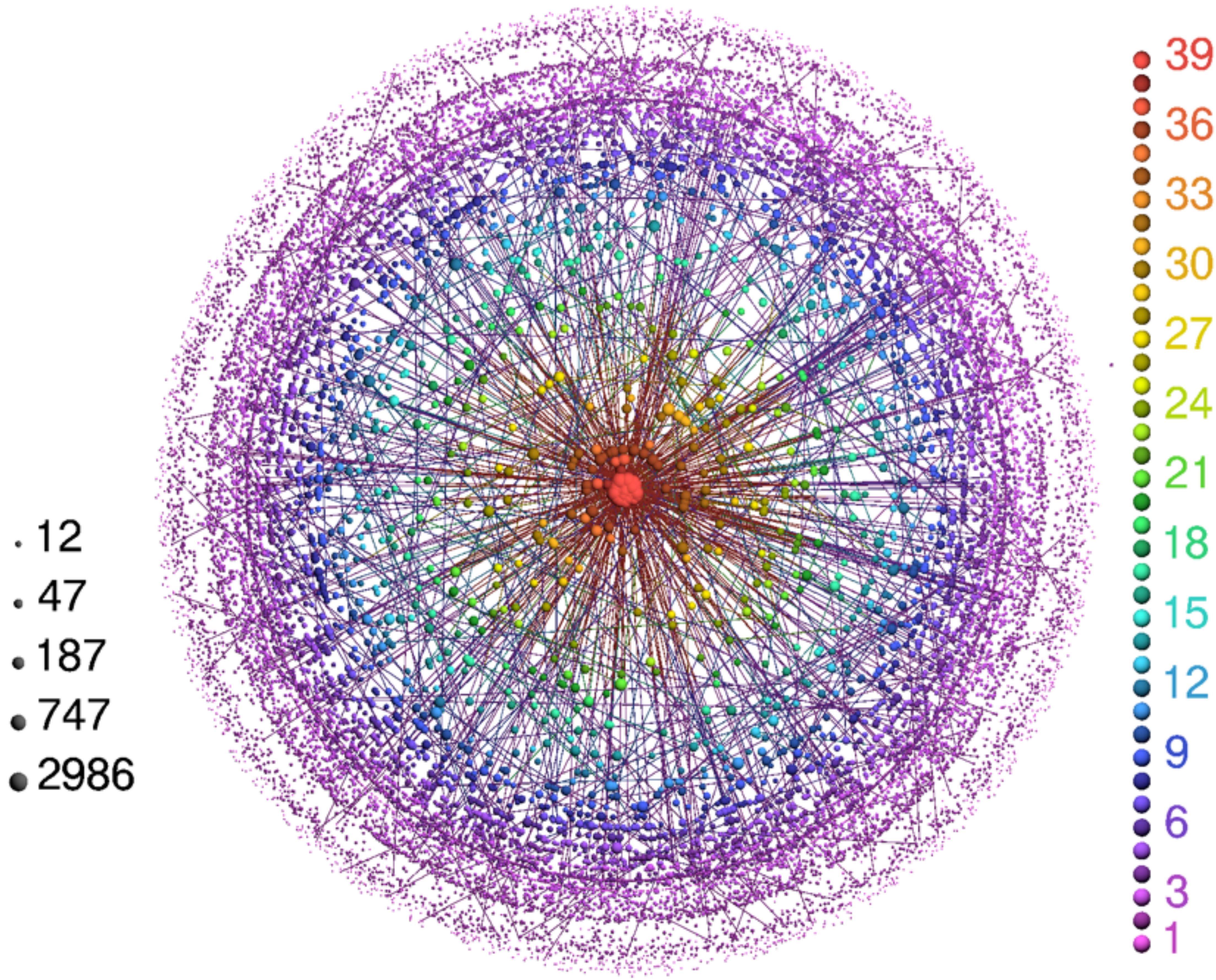
**–vertex i has shell index k iff it belongs to the k-core but not to the (k+1)-core**

**–k-shell: ensemble of all nodes of shell index k**

# Example



http://lanet-vi.fi.uba.ar/




NB: role in spreading processes

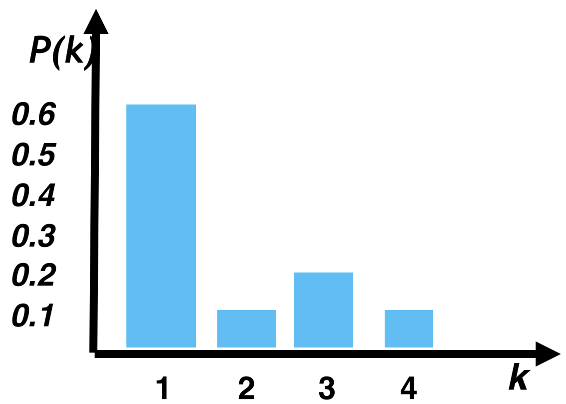
# Statistical characterization of networks

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

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

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

# Topological heterogeneity

Statistical analysis of centrality measures:

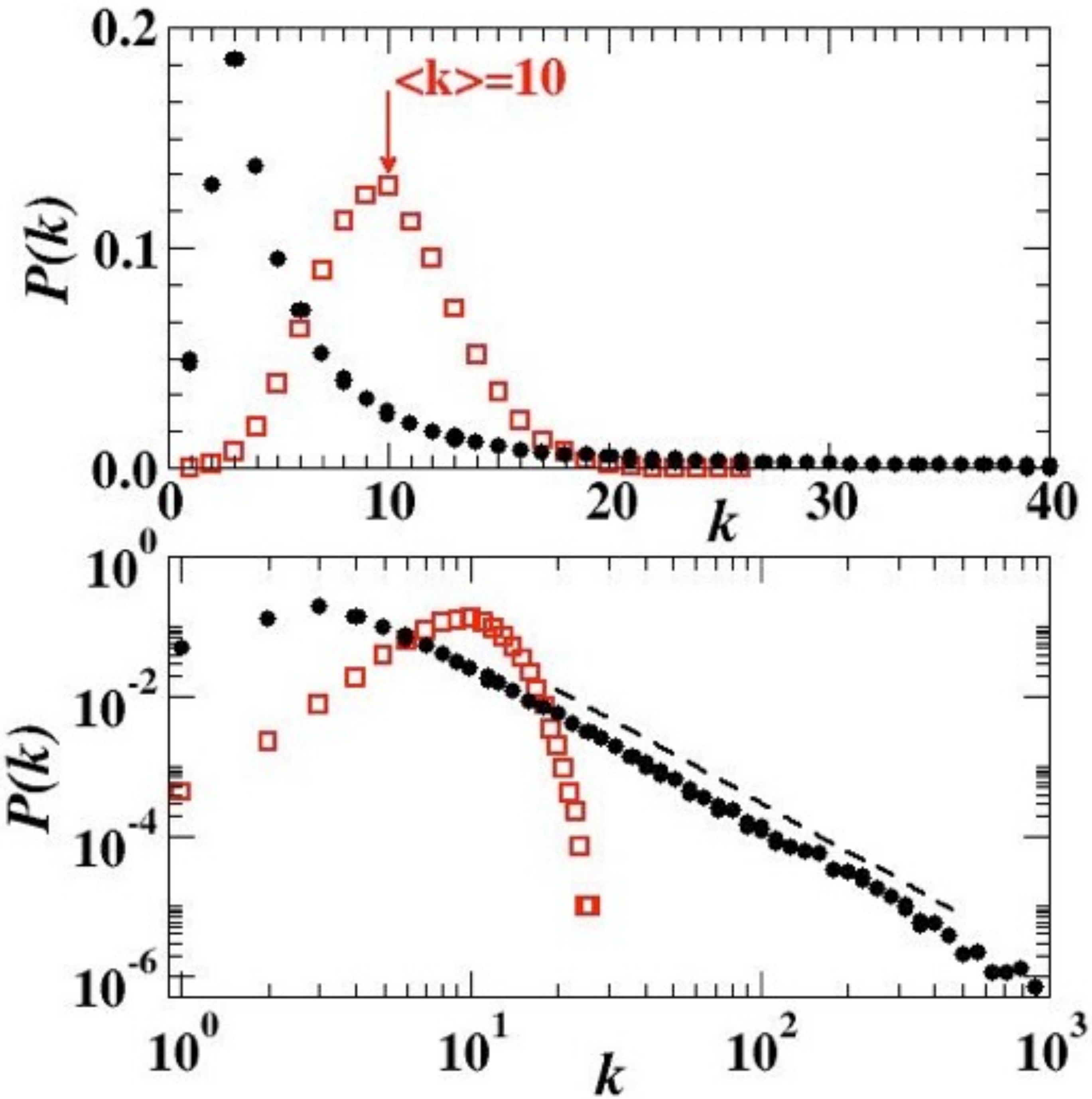
$P(k)=N_k/N=$probability that a randomly chosen node has degree k

Two broad classes

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

# Topological heterogeneity

Statistical analysis of centrality measures:




linear scale

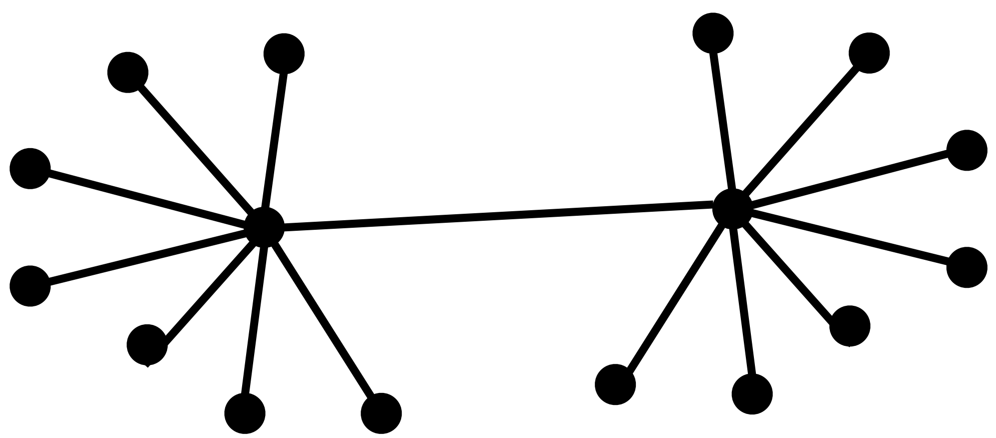
Poisson
vs.
Power-law

log-scale

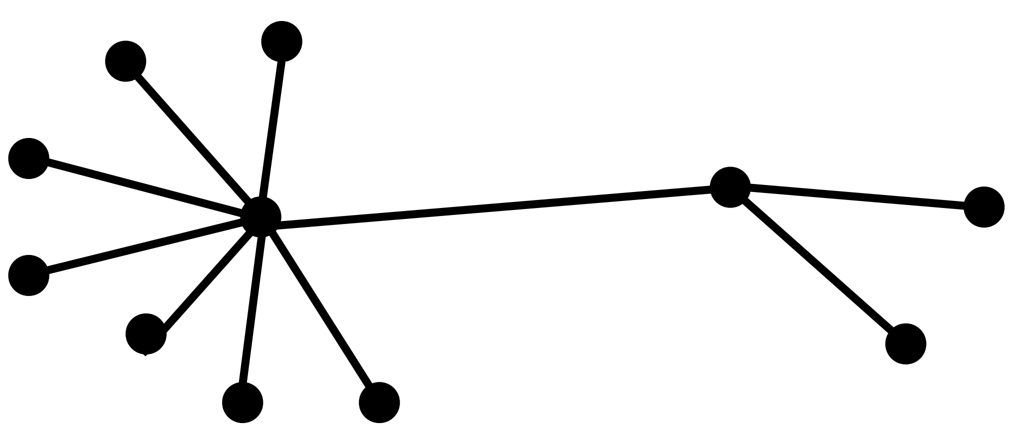
# Statistical characterization

## 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)

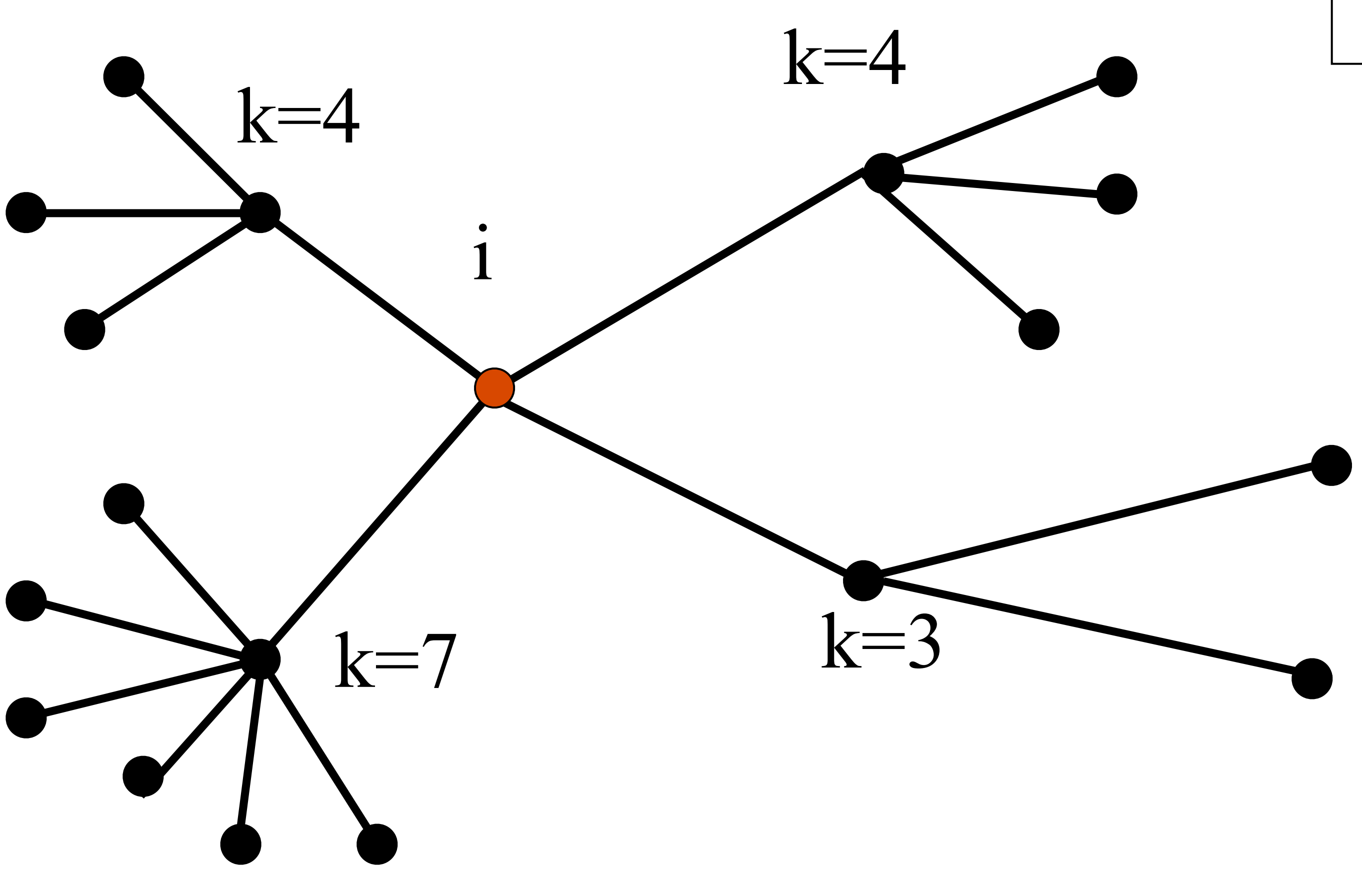
# Statistical characterization

## 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 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

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

# Statistical characterization

**case of *random uncorrelated* networks**

*P(k'|k)*

- independent of k
- proba that an edge points to a node 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

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

# Empirics

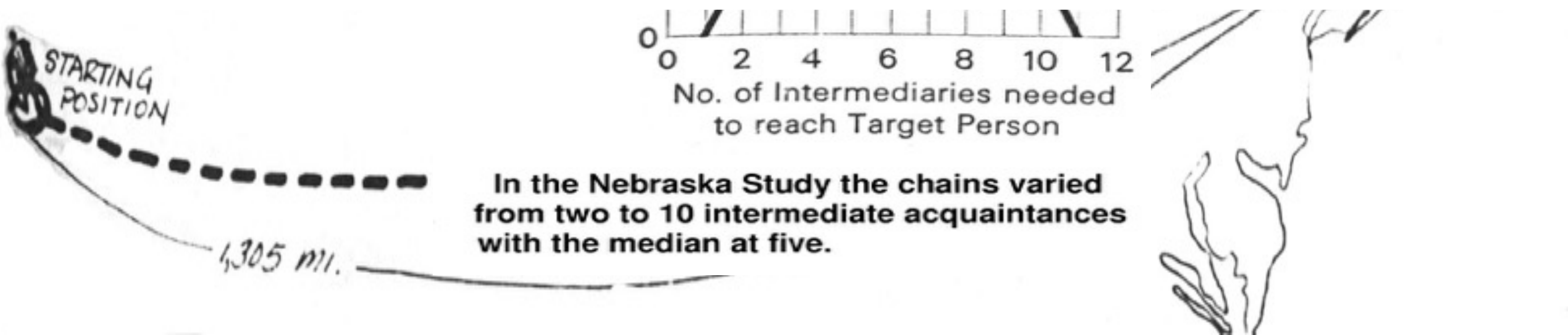
# Social networks: Milgram's experiment

Milgram, Psych Today **2**, 60 (1967)

Dodds et al., Science **301**, 827 (2003)

## "Six degrees of separation"

## SMALL-WORLD CHARACTER



# Social networks as small-worlds: Milgram's experiment, revisited

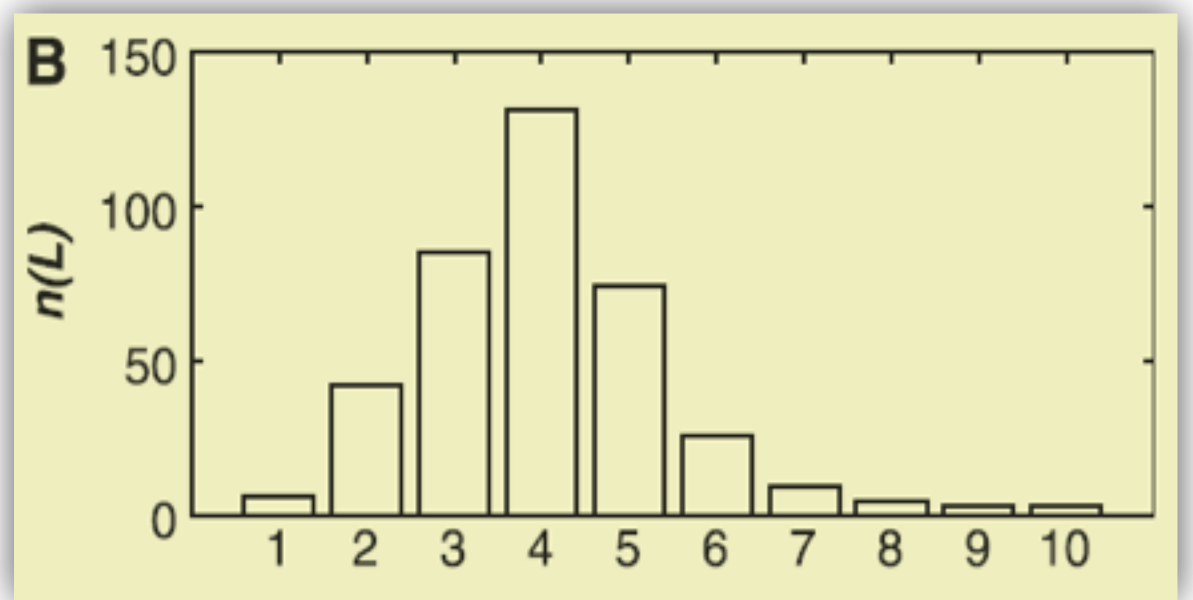
Dodds et al., Science **301**, 827 (2003)

email chains

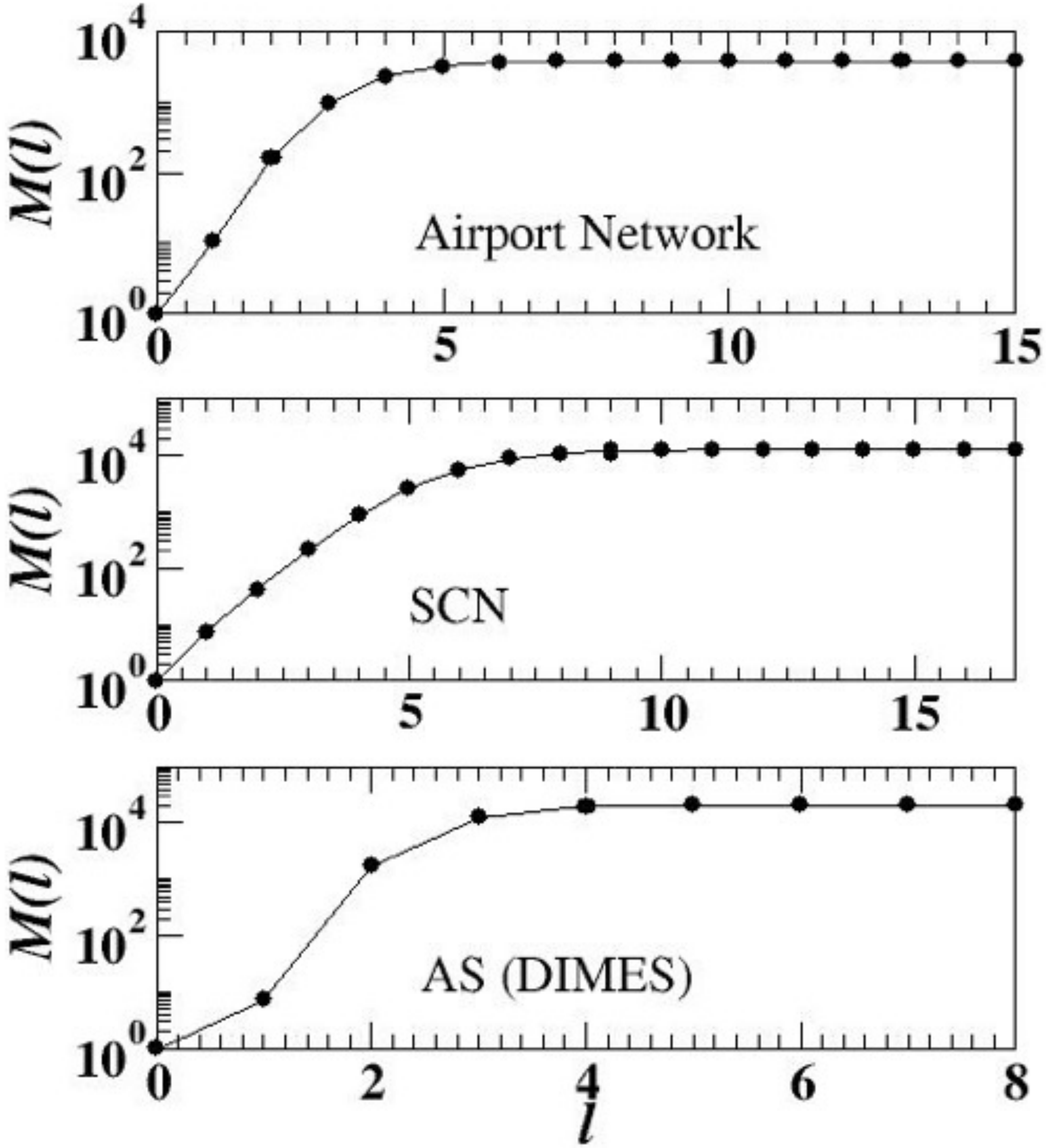
60000 start nodes

18 targets

384 completed chains



# Small-world properties

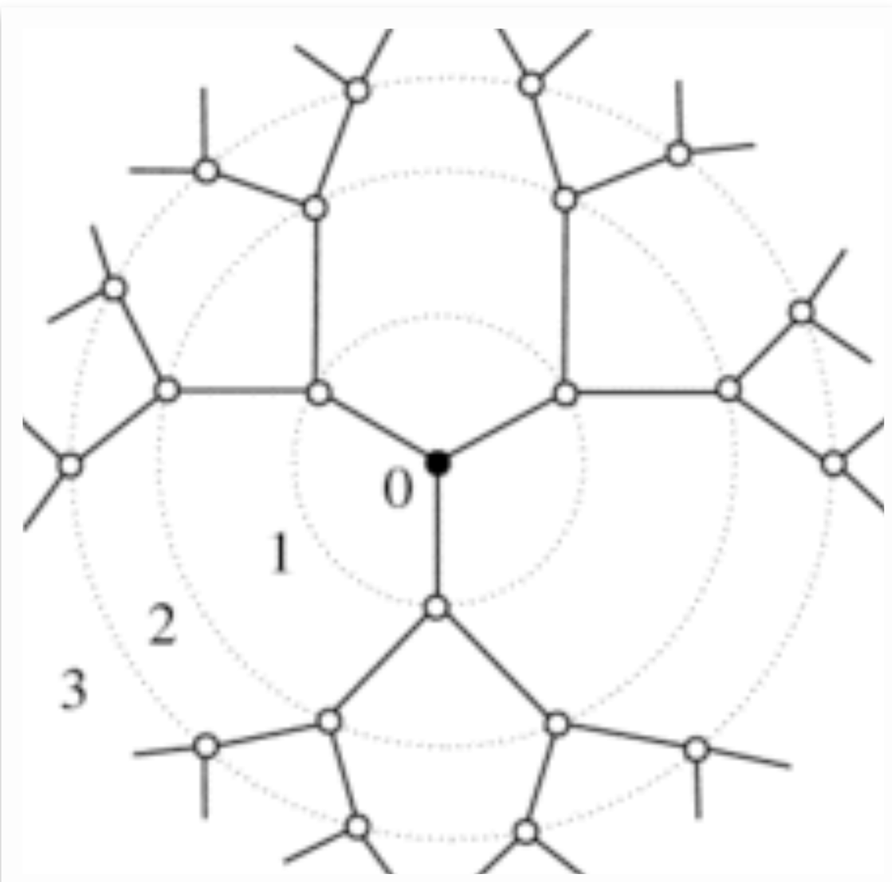



**Average number of nodes within a chemical distance l**

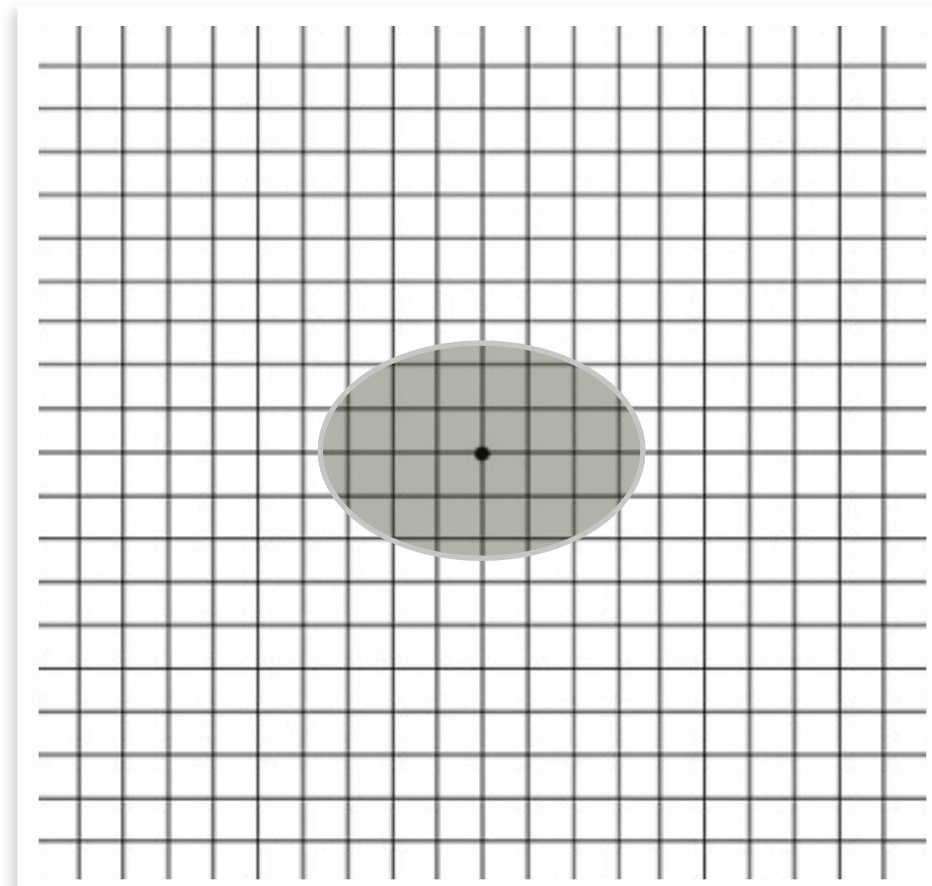
Scientific collaborations

Internet

# The intuition behind the small-world effect




versus



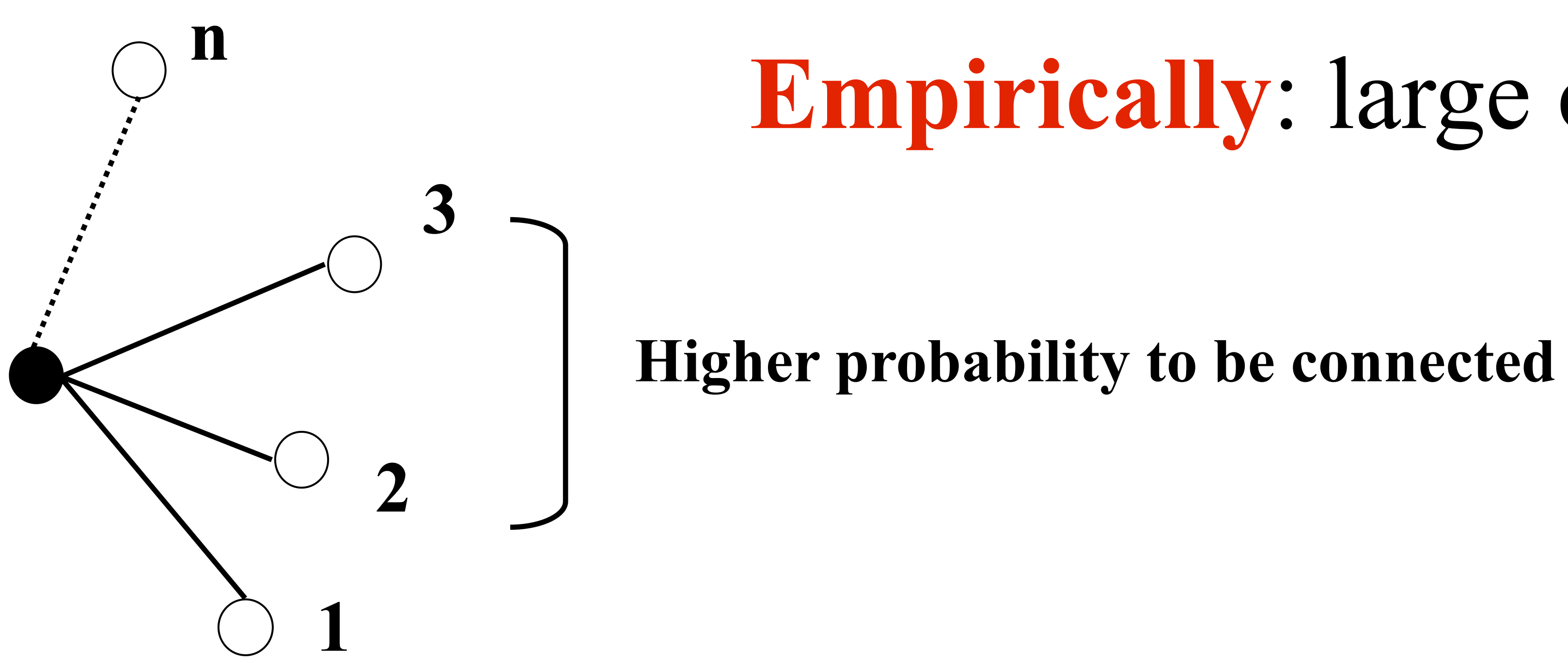
Tree:
number of reachable nodes grows very fast (exponentially) with the distance

(local) regular structure: slower growth of the number of reachable nodes (polynomial), because of path redundancy

Random networks: often locally tree-like

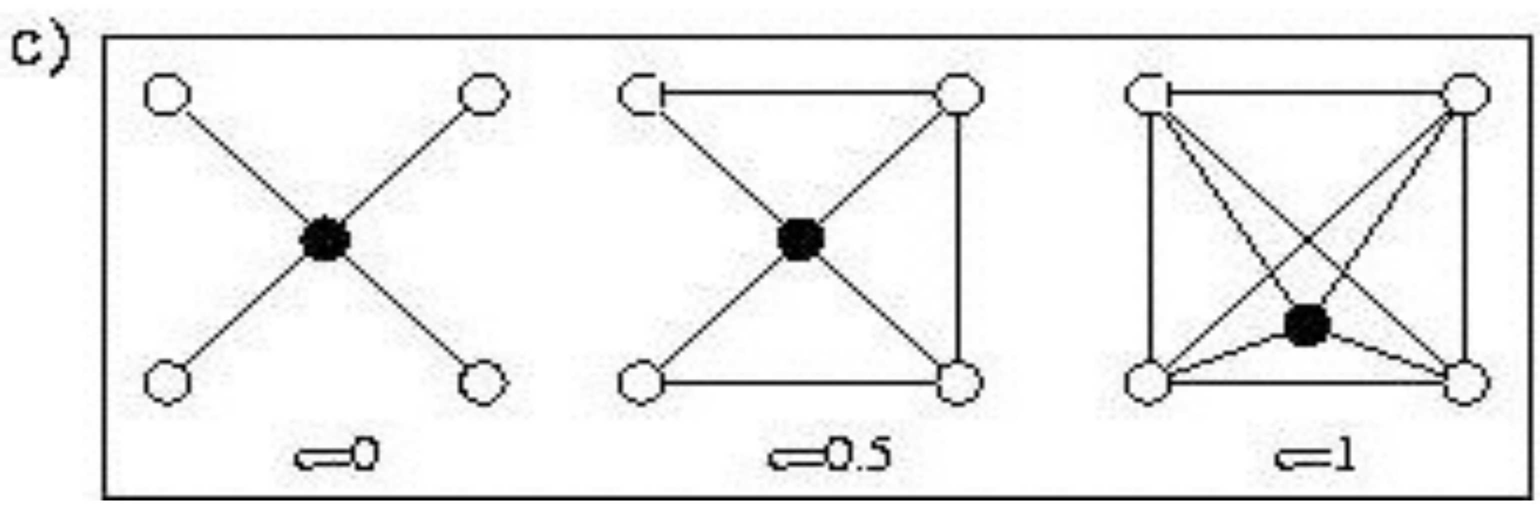
# Small-world yet clustered

# Clustering coefficient




**Empirically**: large clustering coefficients

**Clustering**: My friends will know each other with high probability
(typical example: social networks)
**Redundancy of paths**



# Topological heterogeneity

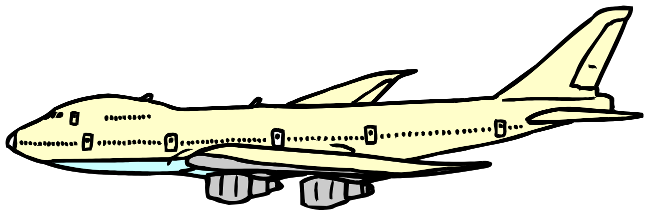
Statistical analysis of centrality measures:

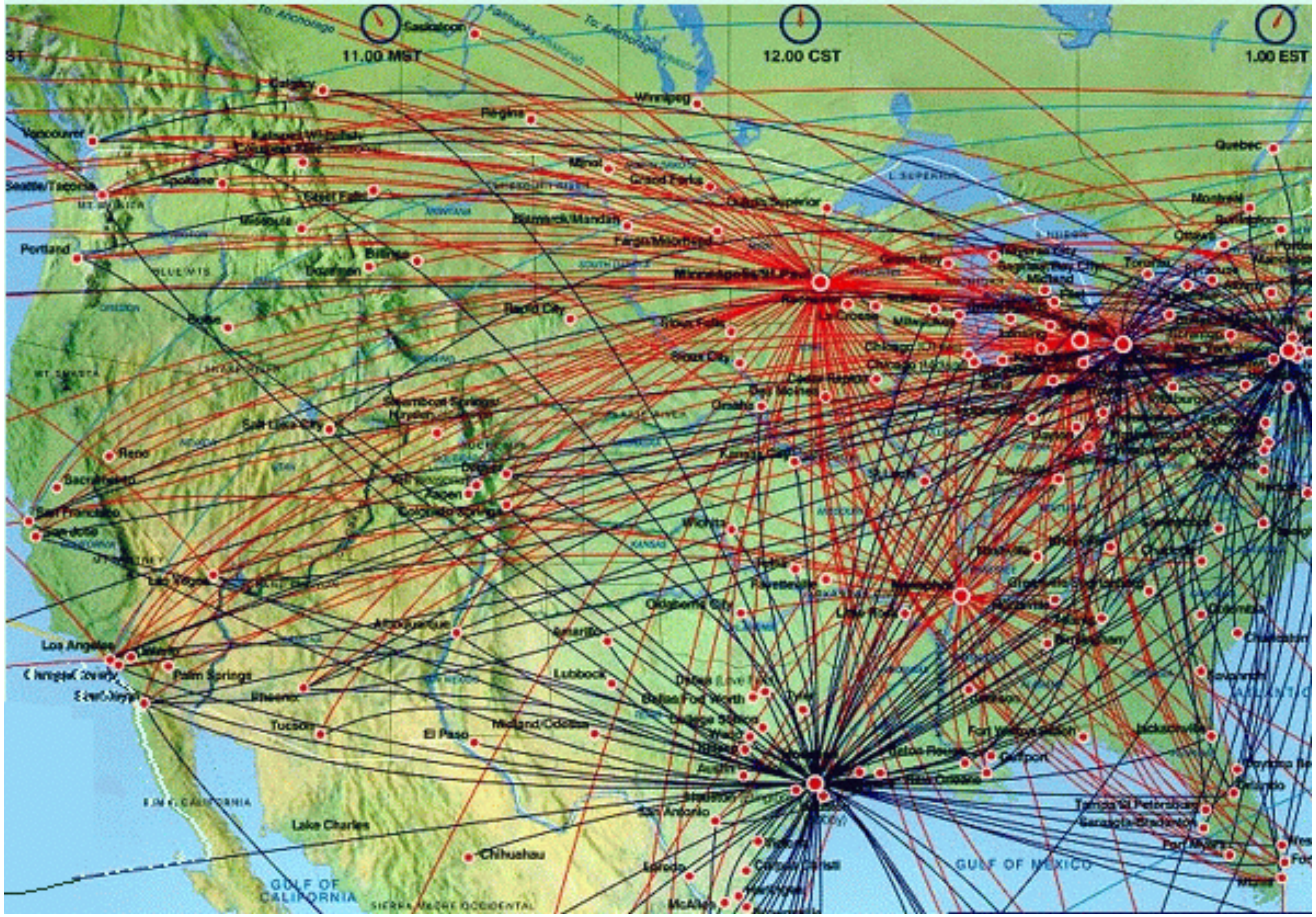
$P(k)=N_k/N=$probability that a randomly chosen node has degree k

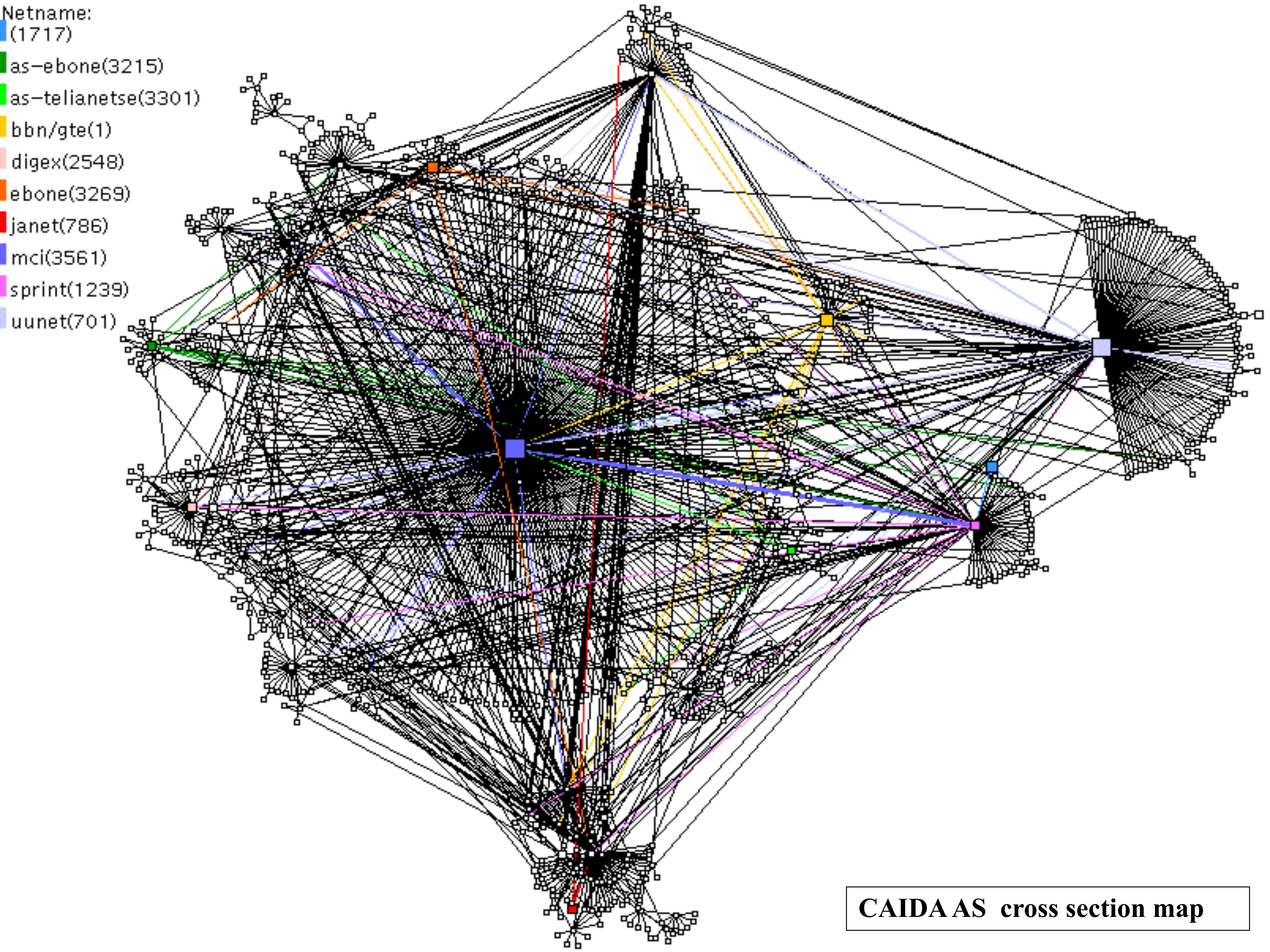
Two broad classes

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

# Airplane route network



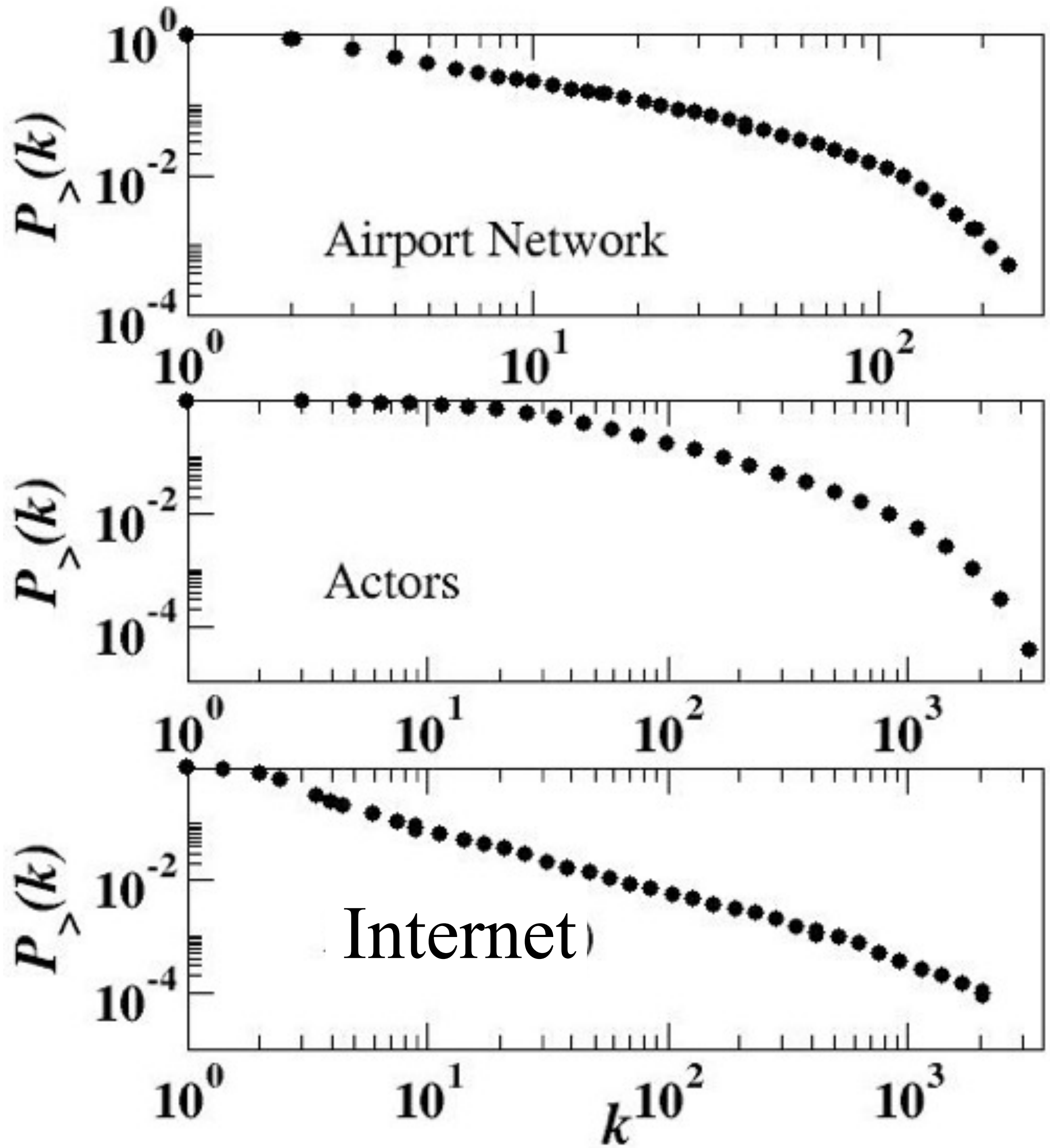





CAIDA AS cross section map

# Topological heterogeneity

## Statistical analysis of centrality measures




Broad degree distributions

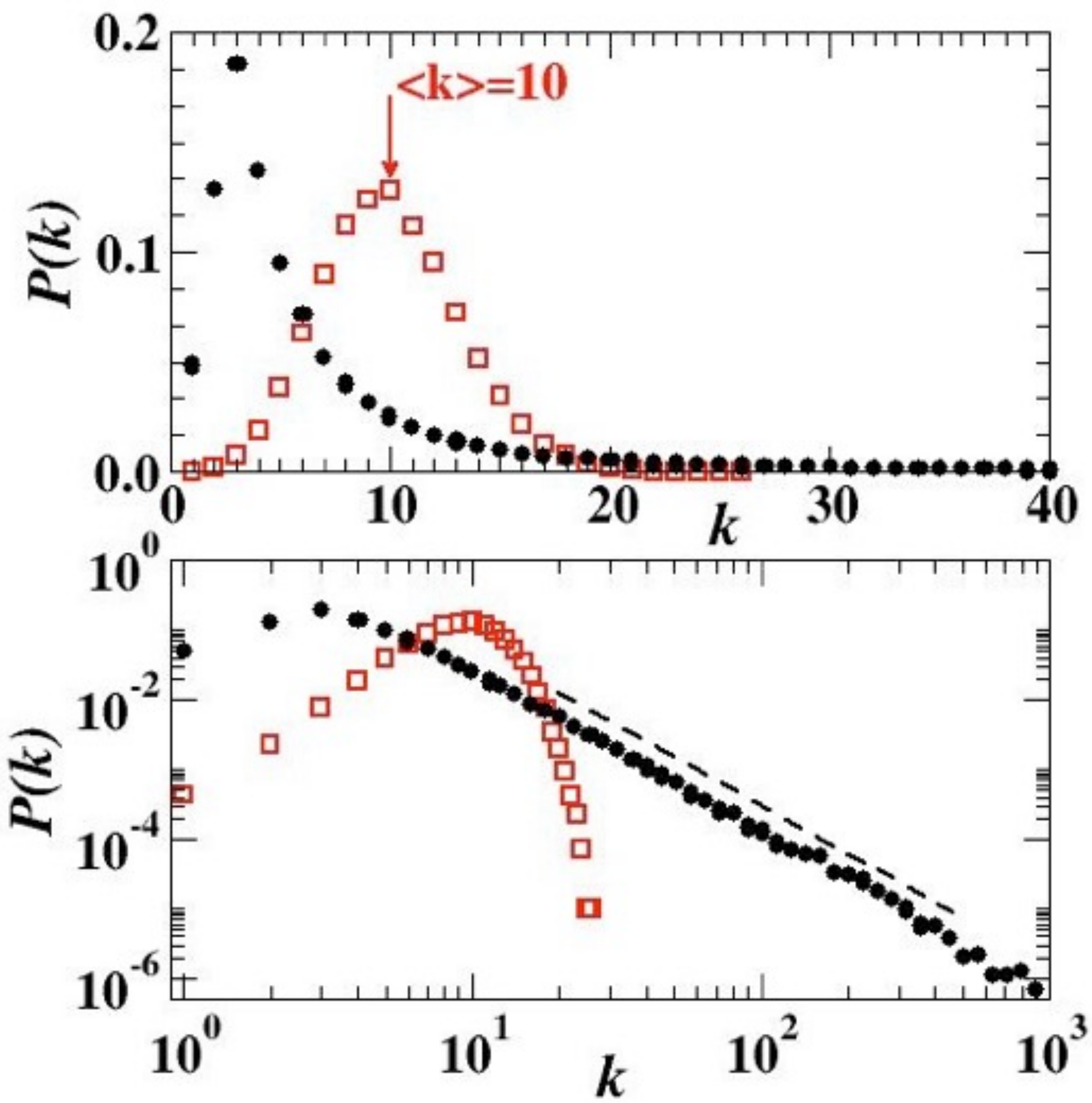
(often: power-law tails $P(k) \propto k^{-\gamma}$ , typically $2 < \gamma < 3$)

No particular *characteristic* scale

Unbounded fluctuations

# Topological heterogeneity

Statistical analysis of centrality measures:




linear scale

Poisson
vs.
Power-law

log-scale

# Consequences

Power-law tails

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

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

Fluctuations

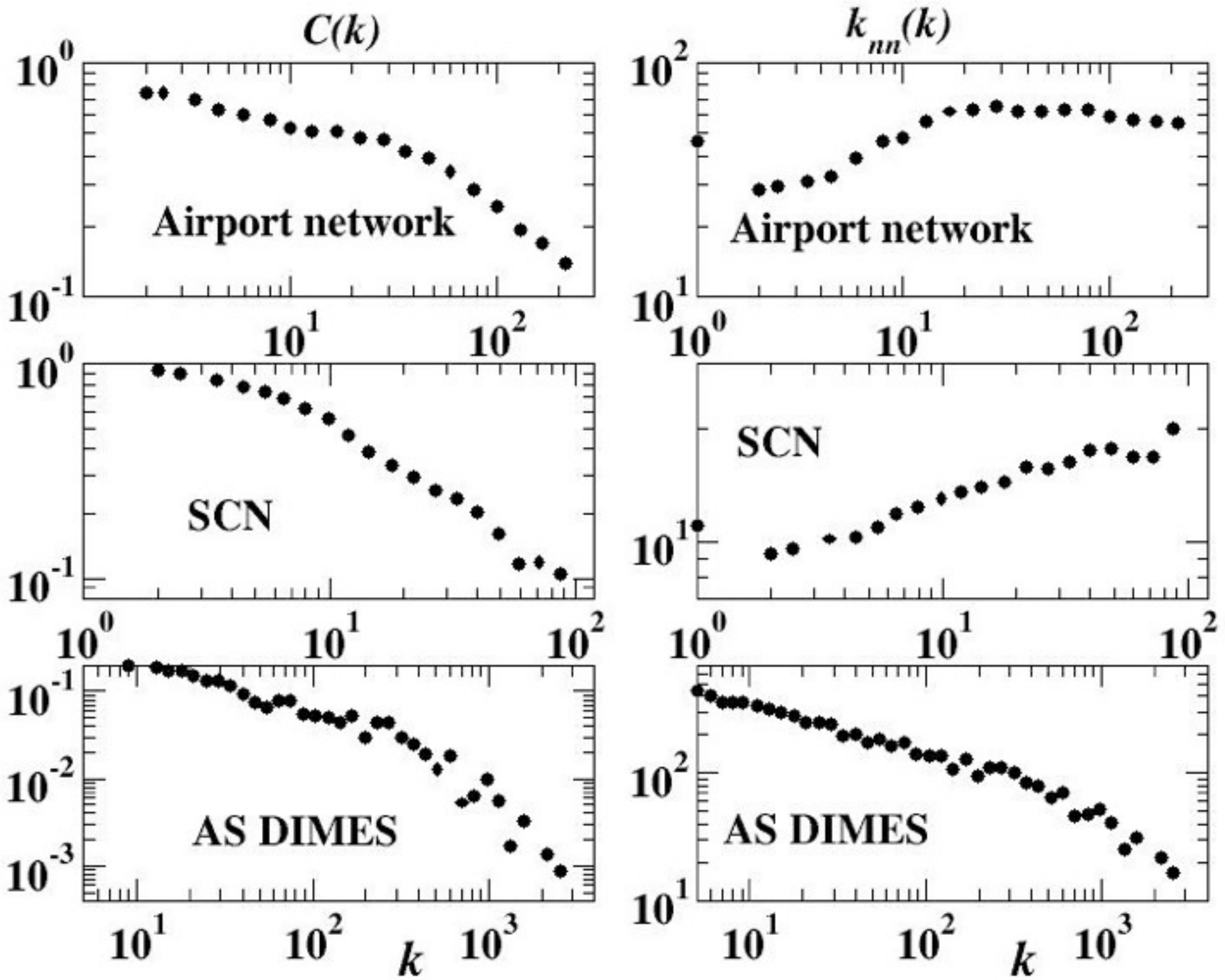
$\langle k^2 \rangle = \int k^2\, P(k)\, dk \propto k_c^{3-\gamma}$

$k_c$=cut-off due to finite-size

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

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

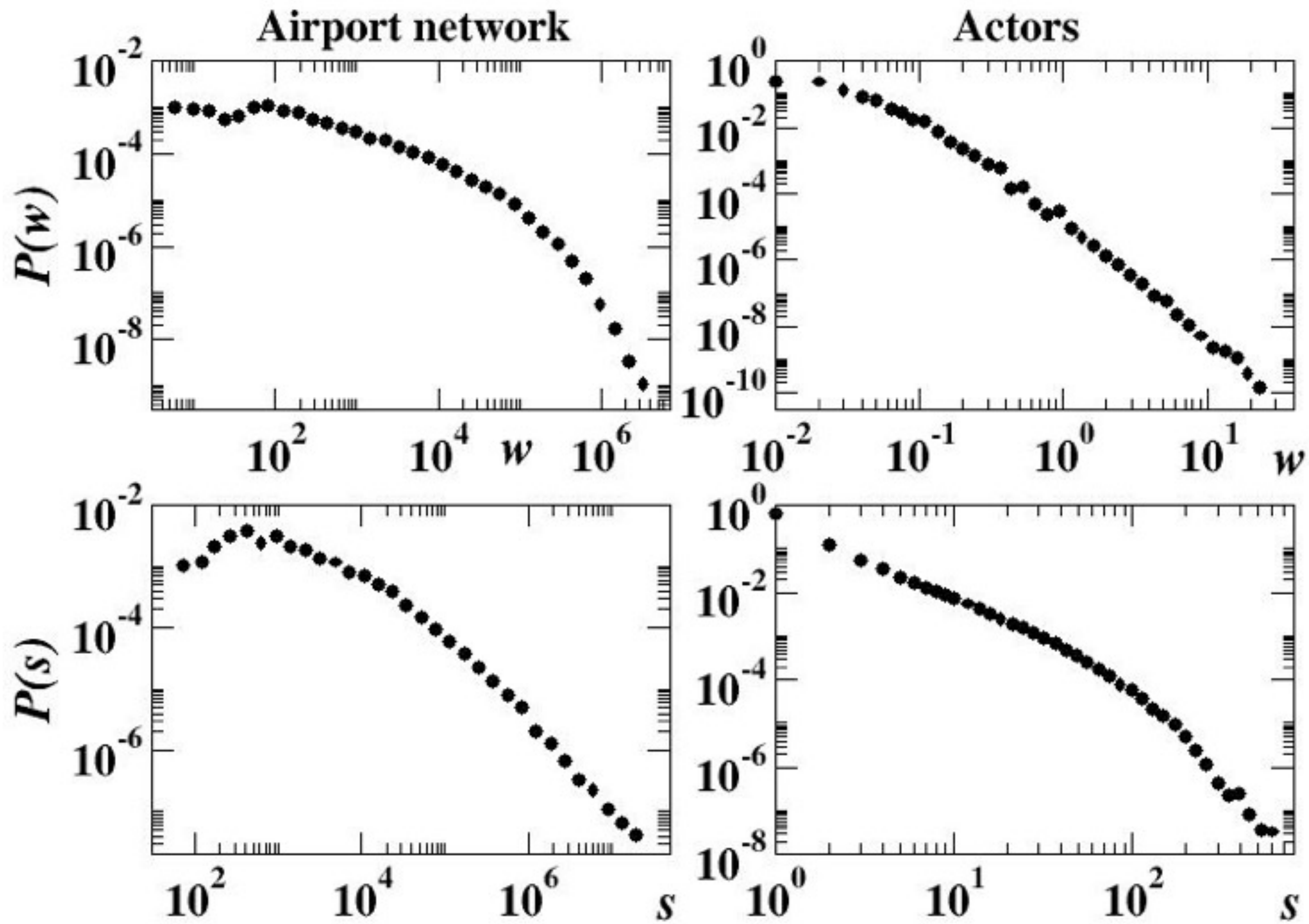
# Empirical clustering and correlations




non-trivial structures

No special scale

# Other heterogeneity levels




Weights

Strengths

# Main things to (immediately) measure in a network

- Degree distribution
- Distances, average shortest path, diameter
- Clustering coefficient
- (Weights/strengths distributions)

# Real-world networks characteristics

Most often:

- Small diameter
- Large local cohesiveness (clustering)
- Heterogeneities (broad degree distribution)
- Correlations
- Hierarchies
- Communities
- …

# Networks and complexity

# ***Complex*** networks

## Complex is not just "complicated"

Cars, airplanes…=> complicated, not complex

Complex (no unique definition):

- many interacting units
- no centralized authority, self-organized
- complicated at all scales
- evolving structures
- emerging properties (heavy-tails, hierarchies…)

Examples: Internet, WWW, Social nets, etc…

# Models

# The role of models

"All models are wrong, but some are useful"

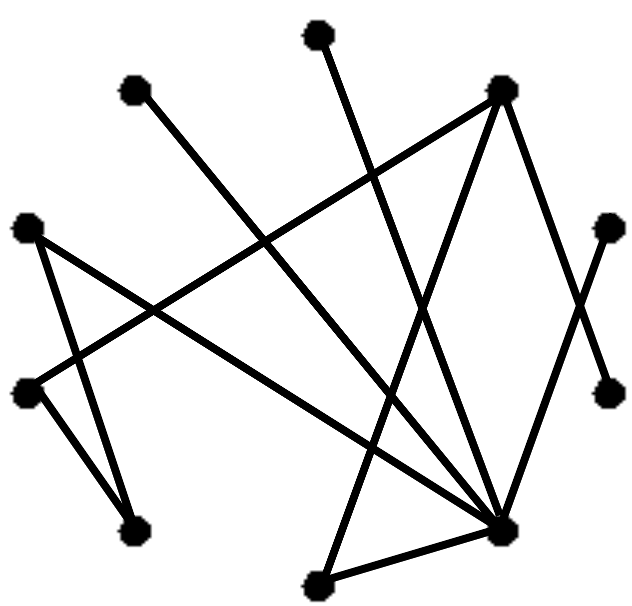
**(George E. P. Box)**

# The role of models

- Generative
- Explanatory
- Null models

# Erdös-Renyi random graph model (1960)

N points, links with proba p:
static random graphs



Average number of edges:
$\langle E \rangle = pN(N-1)/2$

Average degree:
$\langle k \rangle = p(N-1)$

$p = \langle k \rangle / N$ to have finite average degree as N grows

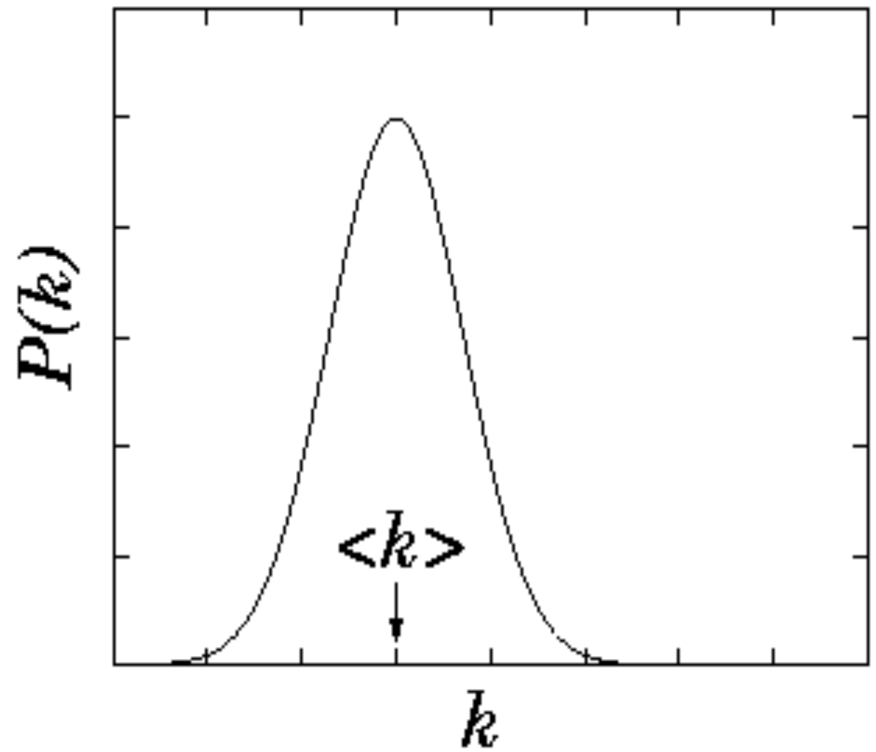
# Erdös-Renyi model (1960)

Proba to have a node of degree k=
•connected to k vertices,
•not connected to the other N-k-1

$$P(k) = C^k_{N-1}\, p^k\, (1-p)^{N-k-1}$$

Large N, fixed pN=< k > : Poisson distribution

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$



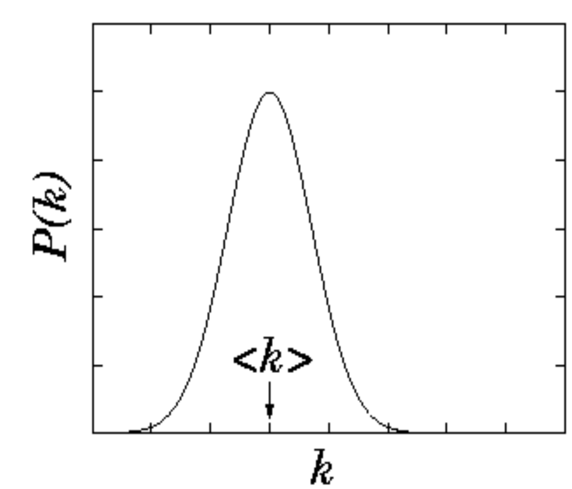

Exponential decay at large k

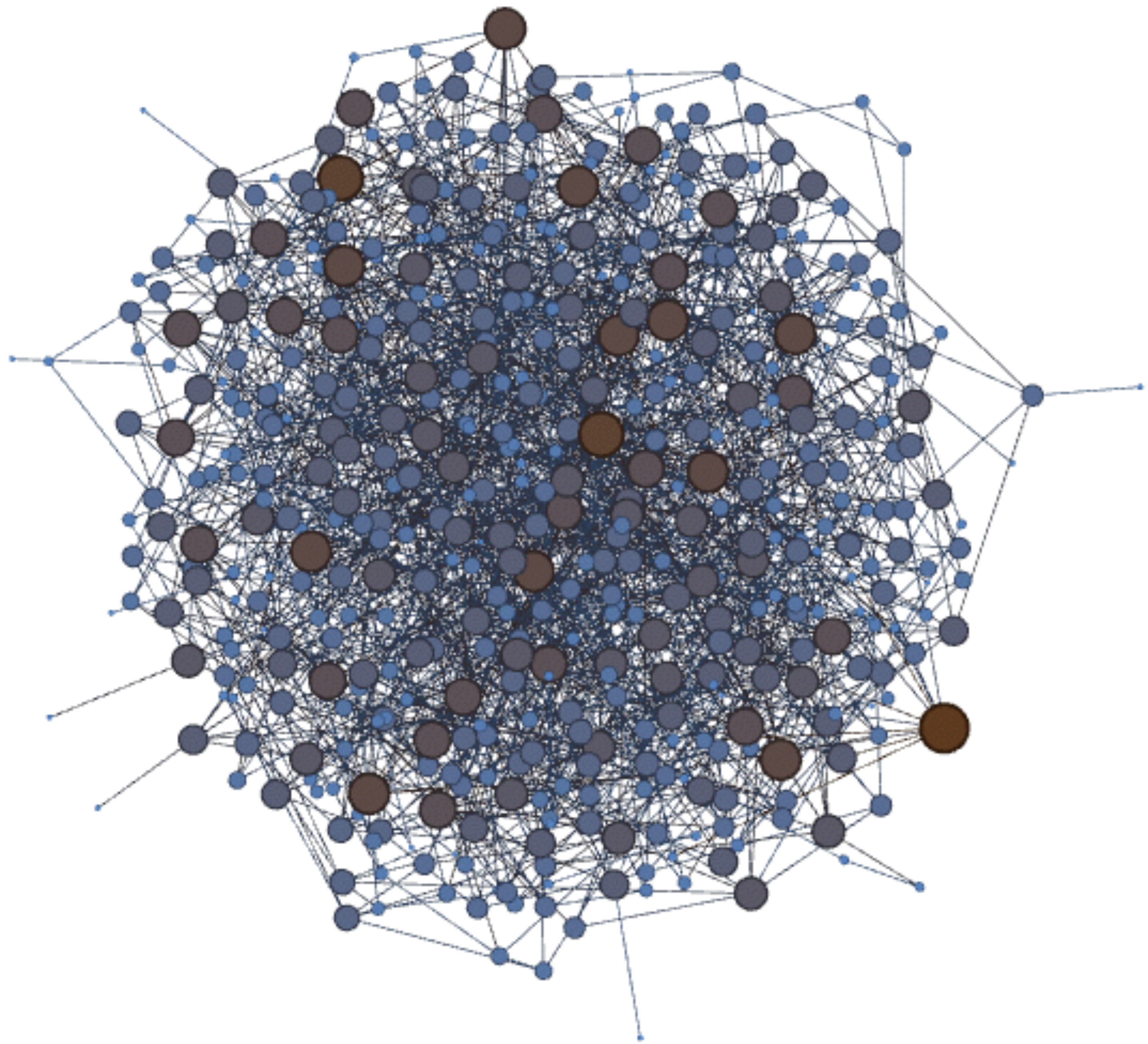
# Erdös-Renyi model (1960)

Short distances $l=\log(N)/\log(< k >)$
(number of neighbours at distance d: $< k >^d$ )

Small clustering: $< C > = p =< k > /N$

Poisson degree distribution



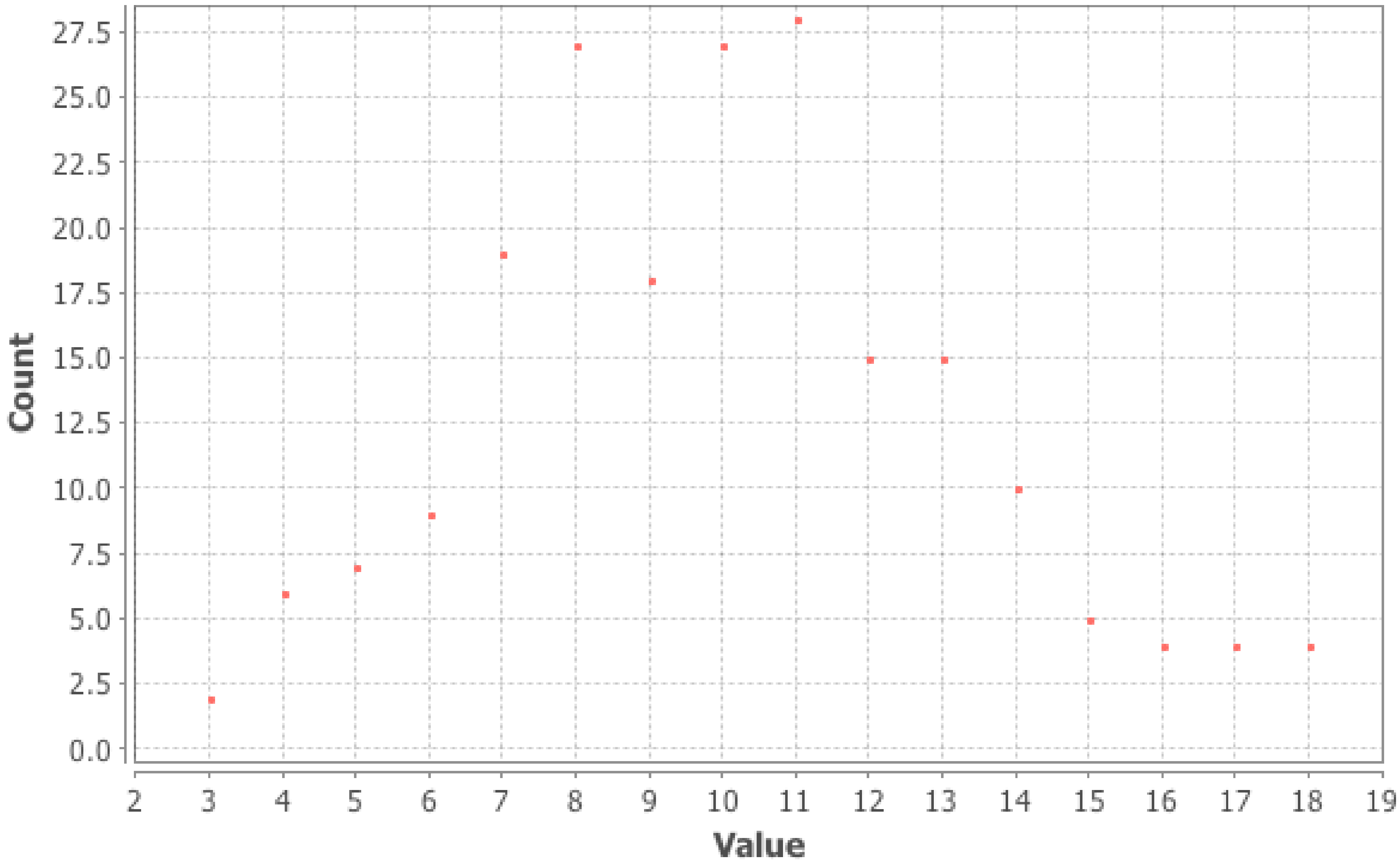


# Degree Report

**Results:**

Average Degree: 10.010

ER model,
N=200
p=0.05



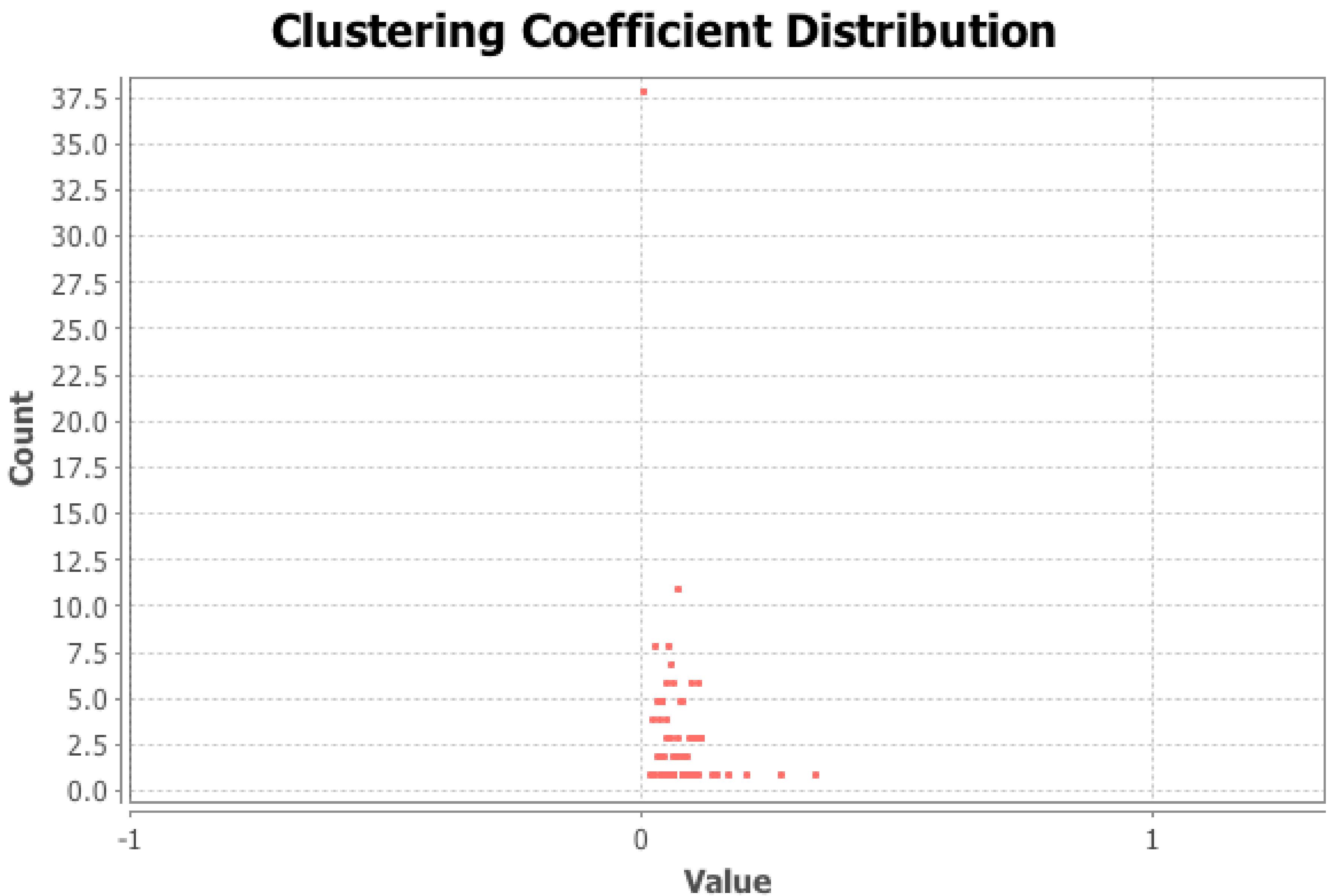
# Clustering Coefficient Metric Report

## Parameters:

Network Interpretation: undirected

## Results:

Average Clustering Coefficient: 0.052
Total triangles: 182
The Average Clustering Coefficient is the mean value of individual coefficients.




ER model,
N=200
p=0.05

**Airlines,**
N=235
<k>=11

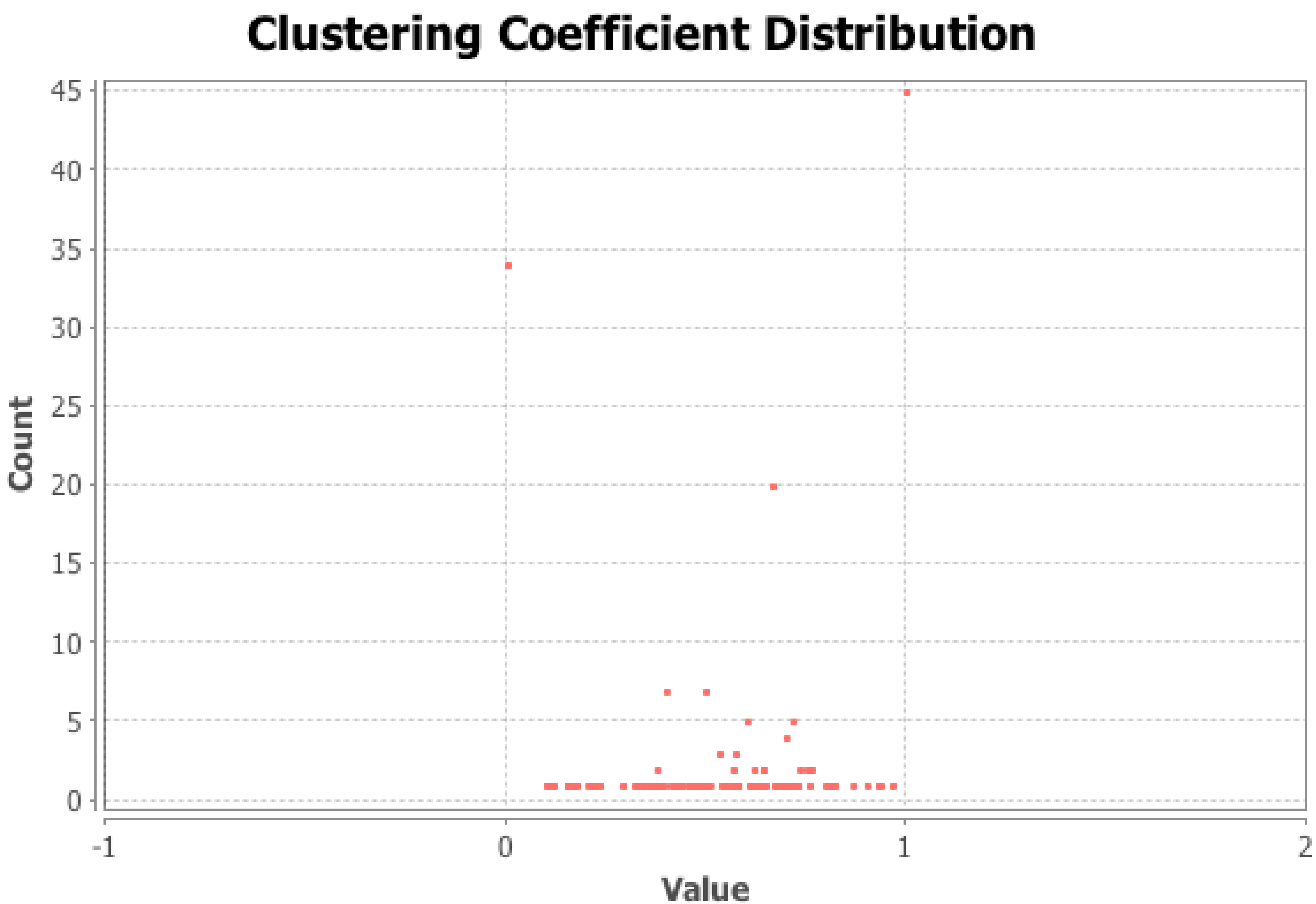
# Clustering Coefficient Metric Report

## Parameters:

Network Interpretation: undirected

## Results:

Average Clustering Coefficient: 0.652
Total triangles: 3688
The Average Clustering Coefficient is the mean value of individual coefficients.

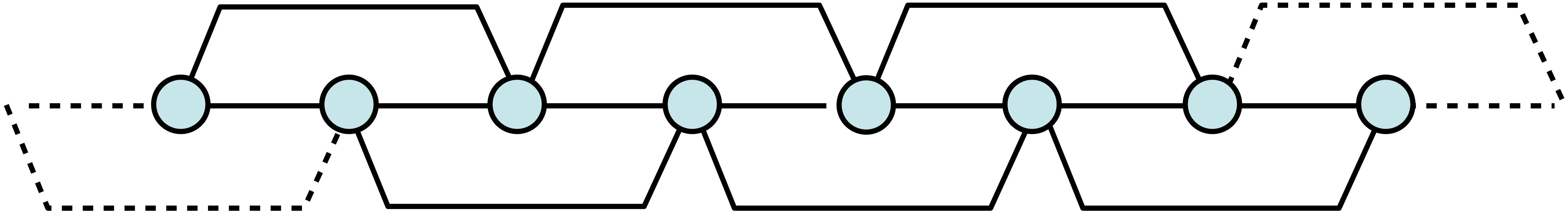


# Watts-Strogatz model

**Motivation:**

**-random graph: short distances but no clustering**

**-regular structure: large clustering but large distances**

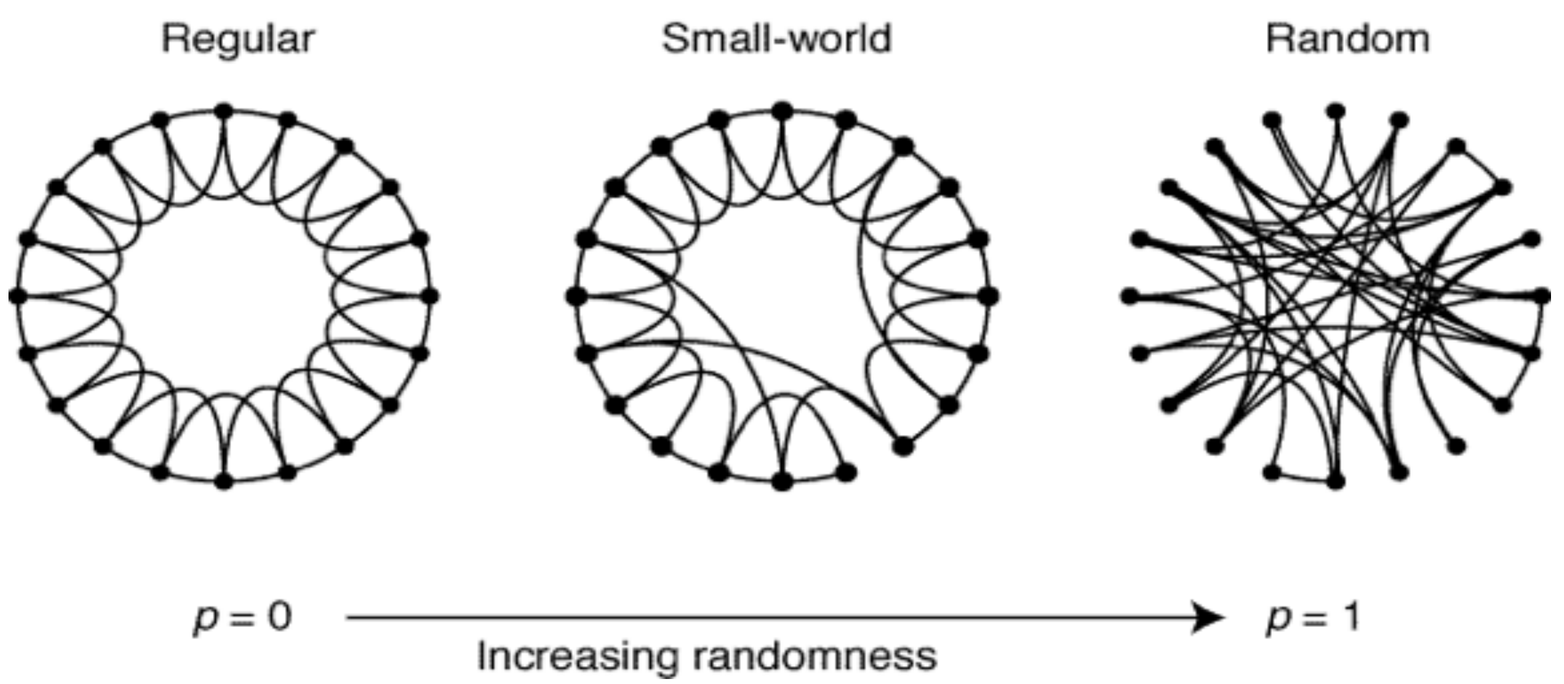


**=> how to have both small distances and large clustering?**

Watts & Strogatz,

Nature **393**, 440 (1998)

# Watts-Strogatz model




**1) N nodes arranged in a line/circle**

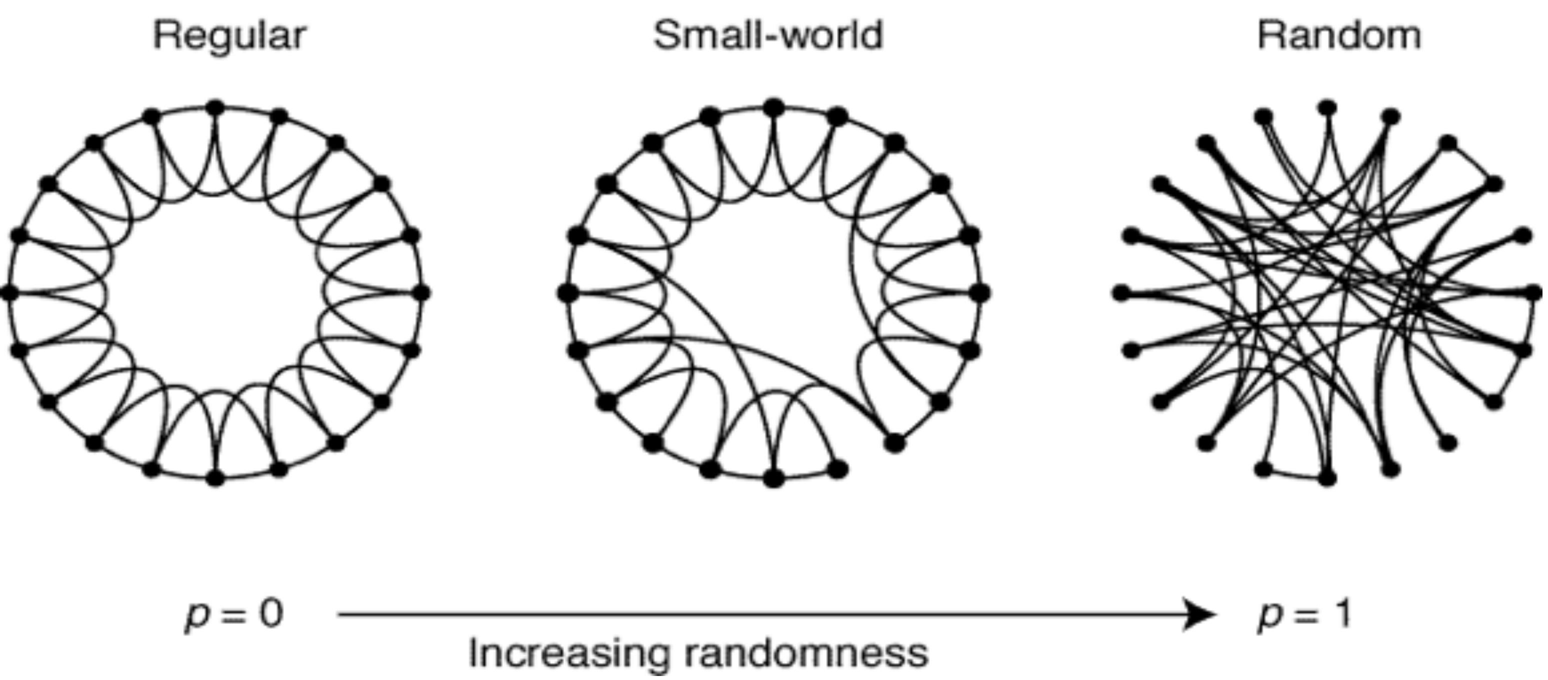
**2) Each node is linked to its 2k neighbors on the circle, k clockwise, k anticlockwise**

**2) Going through each node one after the other, each edge going clockwise is rewired towards a randomly chosen other node with probability p**

Watts & Strogatz,

Nature **393**, 440 (1998)

# Watts-Strogatz model

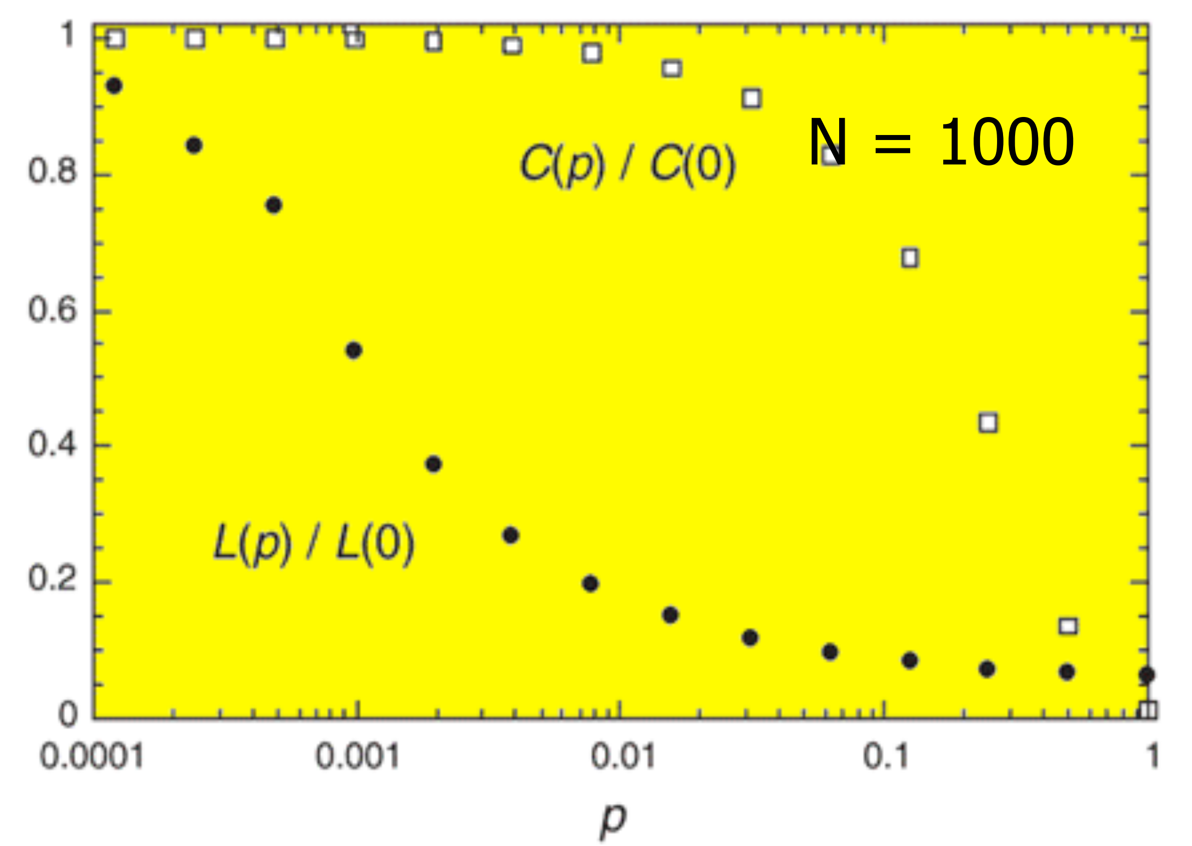



**N nodes forms a regular lattice.**
**With probability p,**
**each edge is rewired randomly**

**=>Shortcuts**

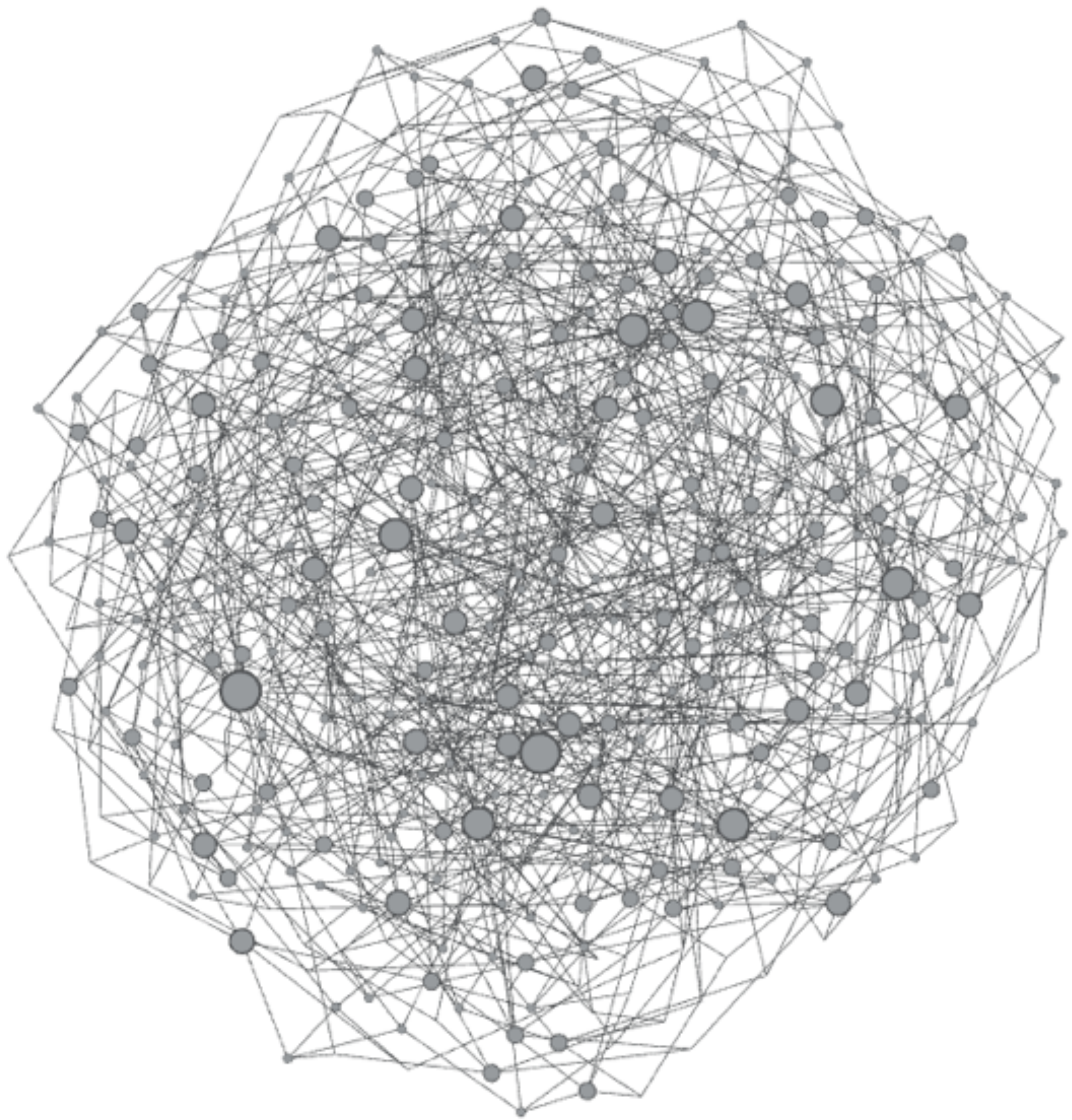
- Large clustering coeff.
- Short typical path

**It takes a lot of randomness to ruin the clustering, but a very small amount to overcome locality**




Watts & Strogatz, Nature **393**, 440 (1998)

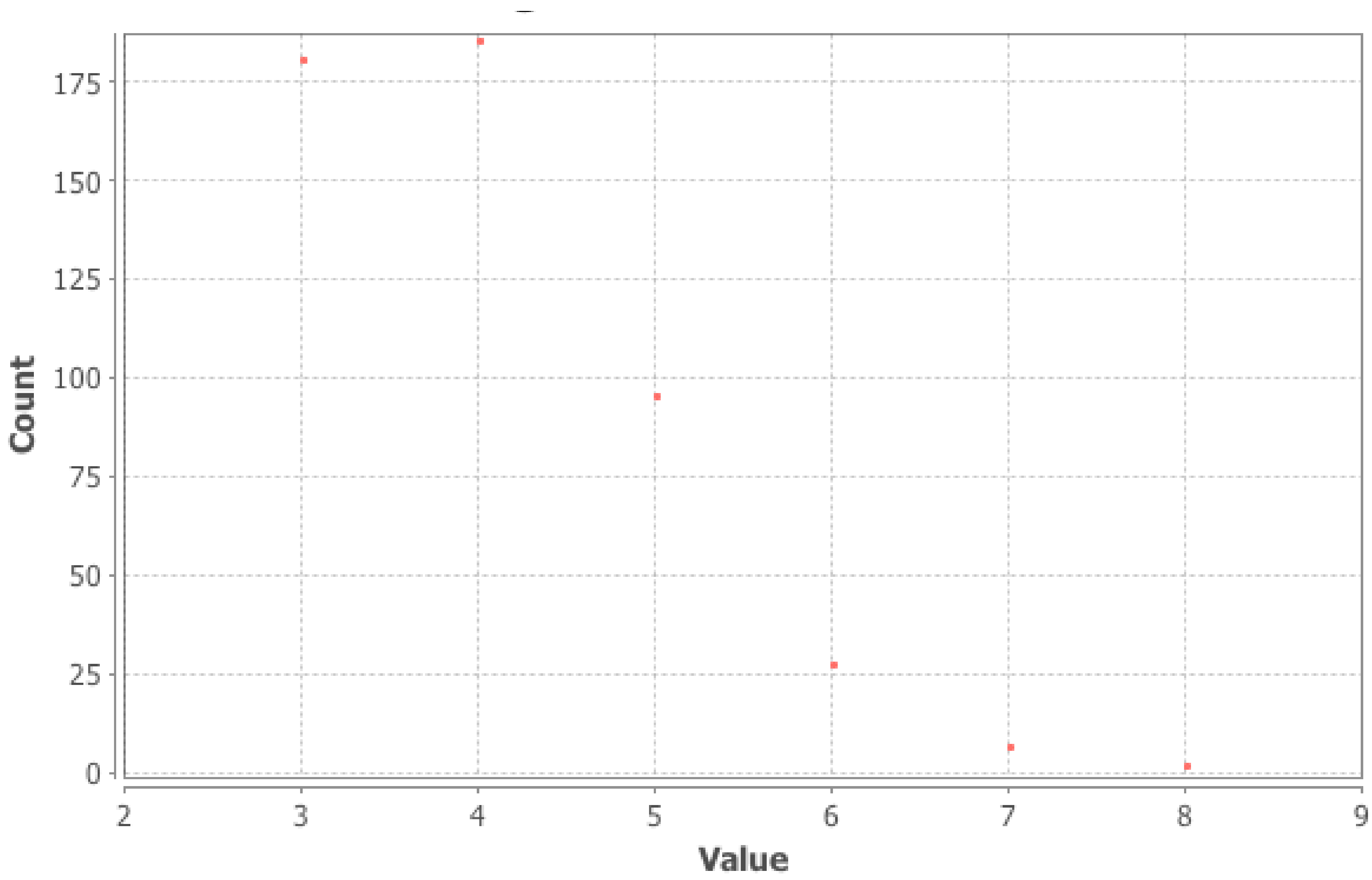
**BUT: still homogeneous degree distribution**



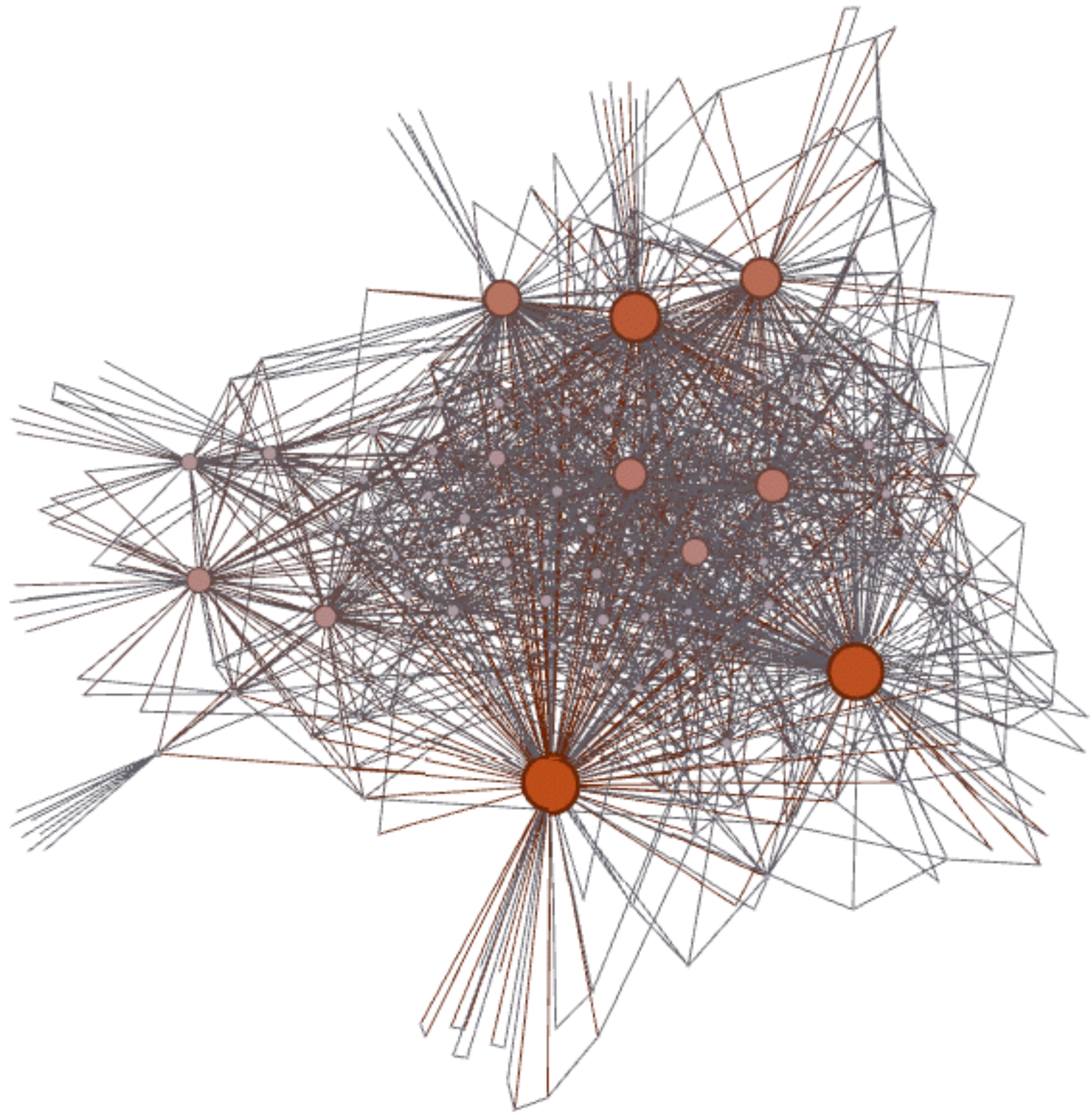
# Degree Report

## Results:

Average Degree: 4.000



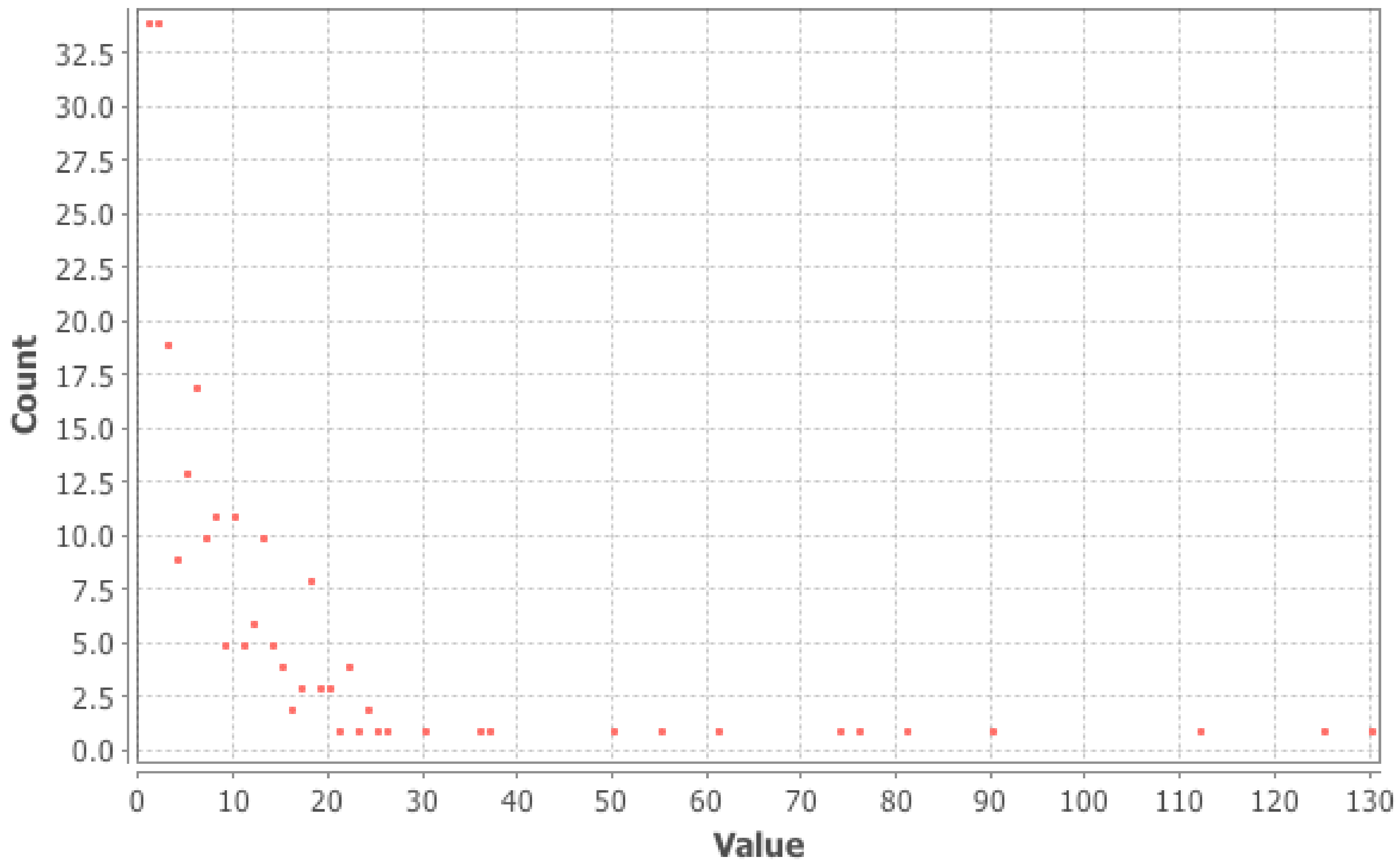
# Airlines



# Degree Report

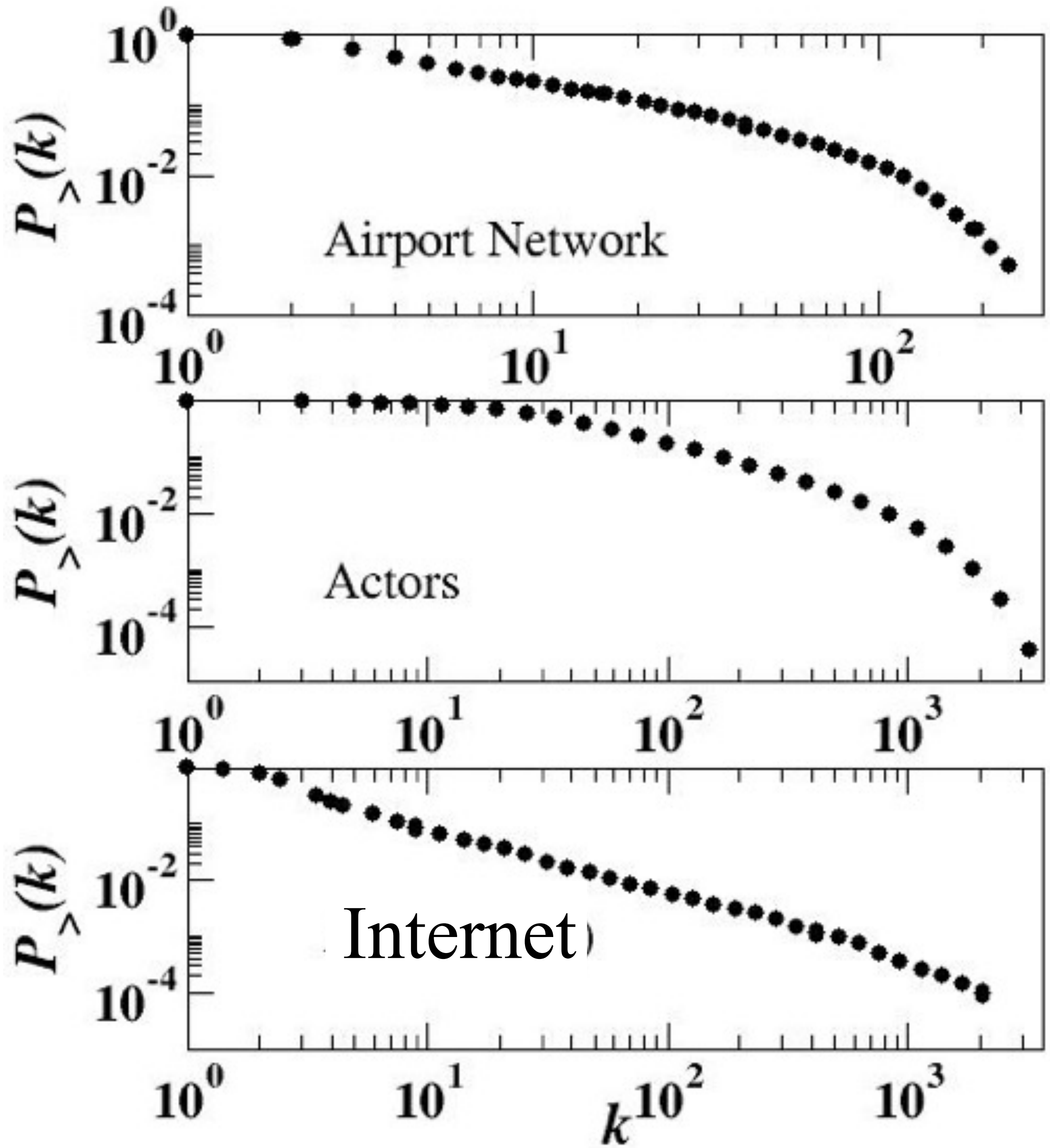
## Results:

Average Degree: 11.038



# Topological heterogeneity

## Statistical analysis of centrality measures




Broad degree distributions

(often: power-law tails $P(k) \propto k^{-\gamma}$ ,
typically $2 < \gamma < 3$)

No particular *characteristic* scale
Unbounded fluctuations

# Generalized random graphs

Desired degree distribution: P(k)

- Extract a sequence $k_i$ of degrees taken from P(k)
- Assign them to the nodes i=1,…,N
- Connect randomly the nodes together, according to their given degree

- =Configuration Model

# Statistical physics approach

**Microscopic processes of the many component units**

**Macroscopic statistical and dynamical properties of the system**

**Cooperative phenomena**
**Complex topology**

**Natural outcome of the dynamical evolution**

Find microscopic mechanisms

# Generative mechanisms

# Example of mechanism:
## preferential attachment

**(1) The number of nodes (N) is NOT fixed.**

Networks continuously expand by the addition of new nodes

Examples:
WWW: addition of new documents
Citation: publication of new papers

**(2) The attachment is NOT uniform.**

A node is linked with higher probability to a node that already has a large number of links.

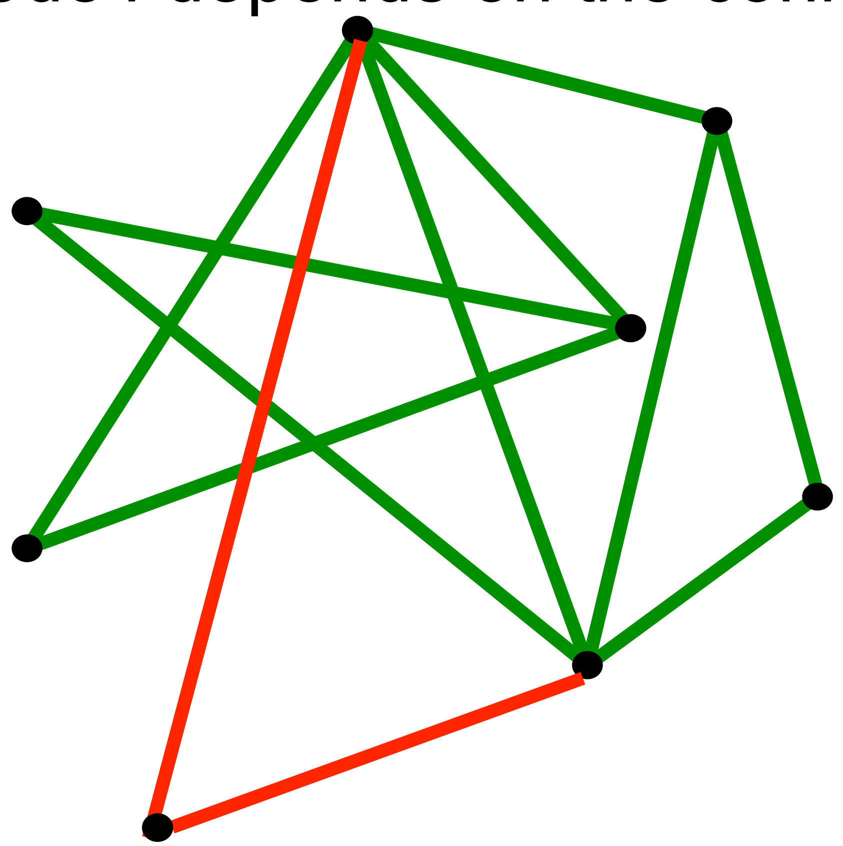
Examples :
WWW : new documents link to well known sites (CNN, YAHOO, NewYork Times, etc)
Citation : well cited papers are more likely to be cited again

# Example of mechanism:
## preferential attachment

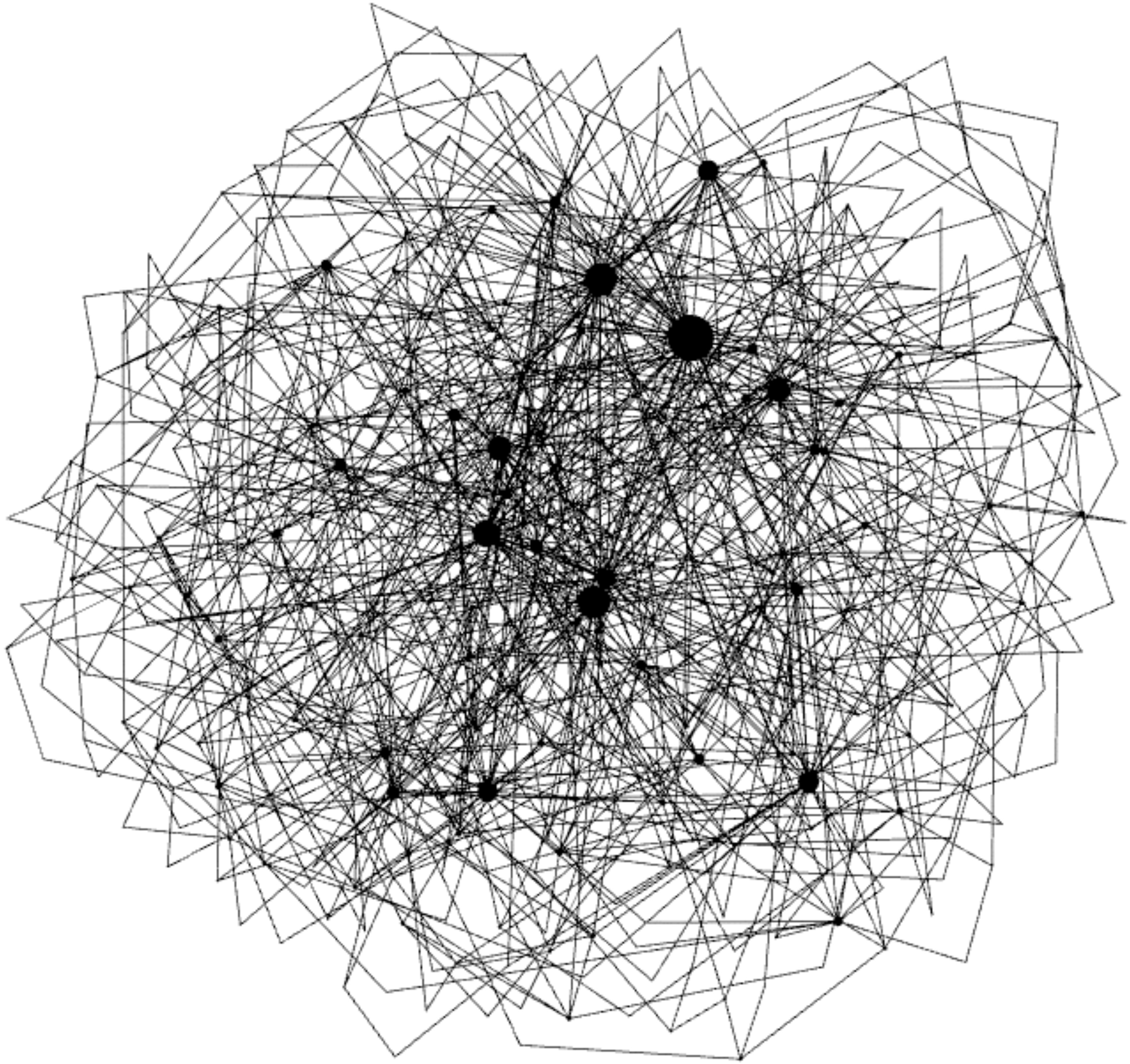
**(1) GROWTH** : At every timestep we add a new node with *m* edges (which have to connect to the nodes already present in the system).

**(2) PREFERENTIAL ATTACHMENT** :
The probability **Π** that a new node will be connected to node *i* depends on the connectivity $k_i$ of that node

$$\Pi(k_i) = \frac{k_i}{\Sigma_j k_j}$$



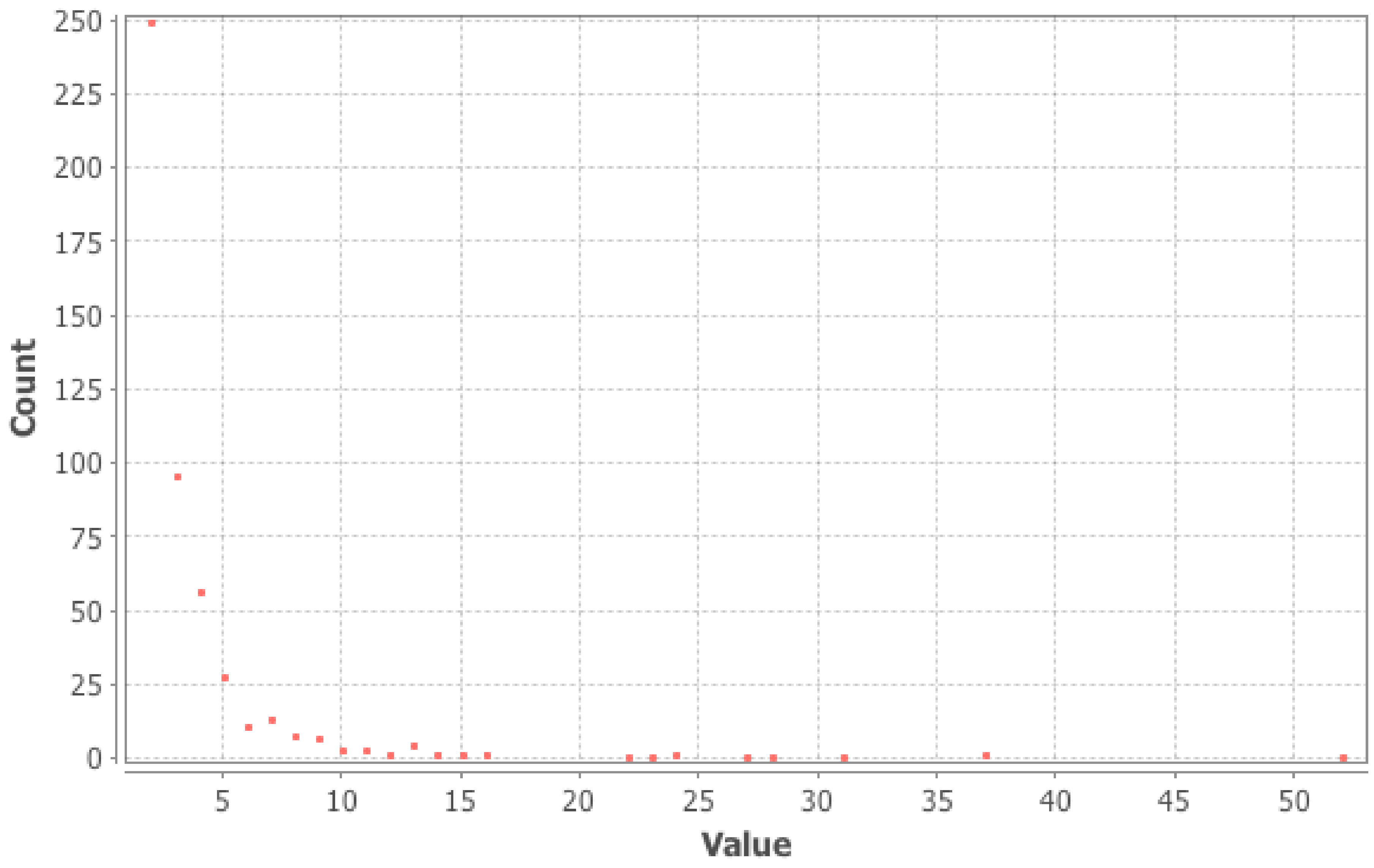
A.-L.Barabási, R. Albert, Science **286,** 509 (1999)



# Degree Report

## Results:

Average Degree: 3.988



# Example of mechanism:
## preferential attachment

**Result: scale-free degree distribution with exponent 3**

$$P(k,t) \sim \frac{2m^2}{k^3}$$

**ISSUES:**

- why linear?
- assumption: new node has full knowledge of nodes' degrees

- old nodes have larger degrees (=> fitness)
- trivial k-core decomposition (=> add other edge creation mechanisms)

# How to check if preferential attachment is empirically observed?

$T_k$=*a priori* probability for a new node to establish a link towards a node of degree k

P(k,t-1)=degree distribution of the N(t-1) nodes forming the network at time t-1

=> proba to observe the formation of a link to a node of degree k = $T_k$ *N(t-1)*P(k,t-1)

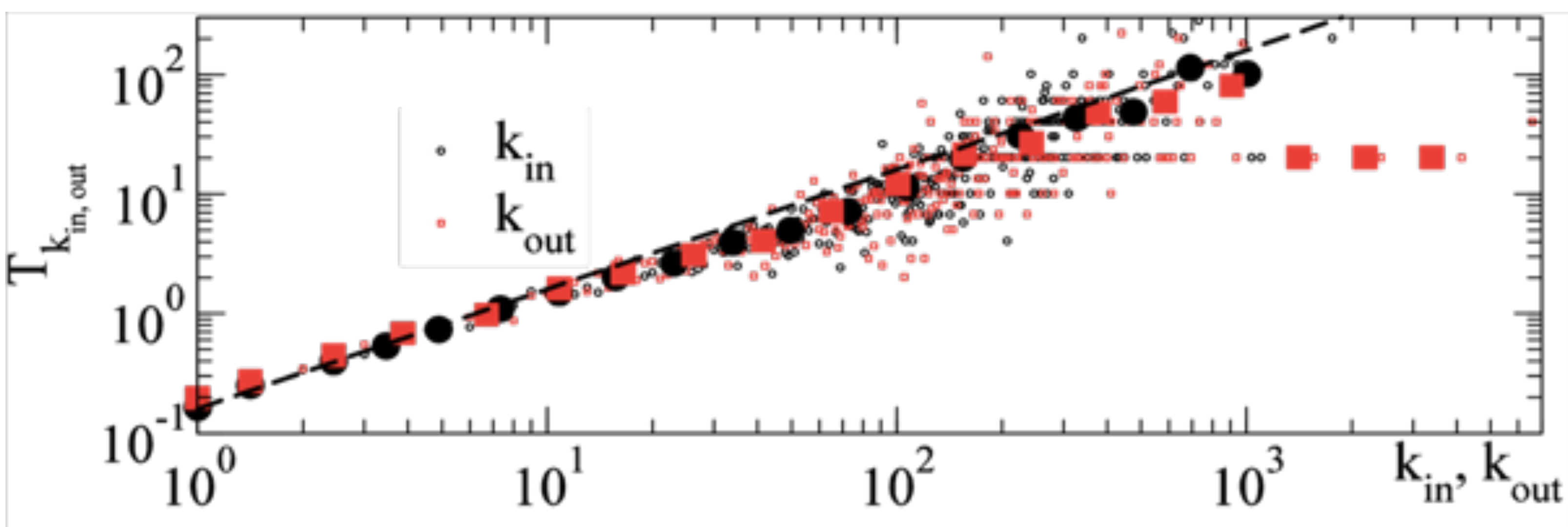
# How to measure the preferential attachment

Hence:

$T_k$= fraction of links created between t-1 and t that reach nodes of degree k, divided by N(t-1)P(k,t-1) (i.e., number of nodes of degree k at time t-1)

*Linear* $T_k$: sign of preferential attachment

Ex of an online social network:




*Where does it come from?*

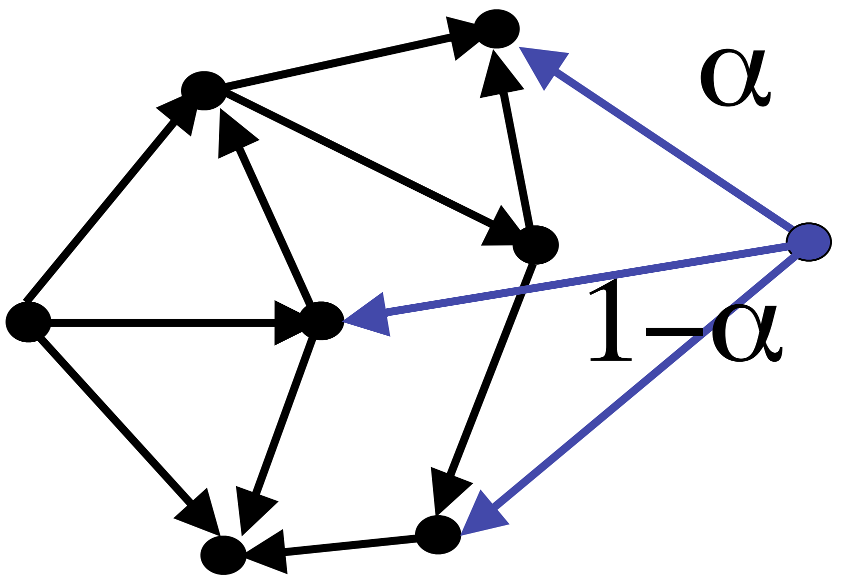
# Another mechanism:
## copying model

Growing network:

a. Introduction of a new vertex

b. Selection of a vertex

c. The new vertex copies m links of the selected one




d. Each new link is kept with proba α, rewired at random with proba 1-α

# Another mechanism:
## copying model

Probability for a vertex to receive a new link at time t:

- Due to random rewiring: $(1-\alpha)/t$

- Because it is neighbour of the selected vertex:

$k_{in}/(mt)$

effective preferential attachment, without a priori knowledge of degrees!

# Copying model

→ Power-law tail of degree distribution:

$$P(k,t) \sim (k+k_0)^{-1-\frac{1}{\alpha}}$$

(model for WWW and evolution of genetic networks)

- Many other proposed mechanisms in the literature,

  => modeling other attributes: weights, clustering, assortativity, spatial effects…

- Model validation:

  => comparison with (large scale) datasets:

  -degree distribution

  -degree correlations

  -clustering properties

  -k-core structure

  ...

# Model validation:

degree distribution, degree correlations, clustering properties, k-core structure, ...

Level of detail: depends on context/goal of study

- -find a very detailed model
- -find a model with qualitative similarities
- -show the plausibility of a formation mechanism
- -generate artificial data
- -study the influence of a particular ingredient
- -...

# Null models

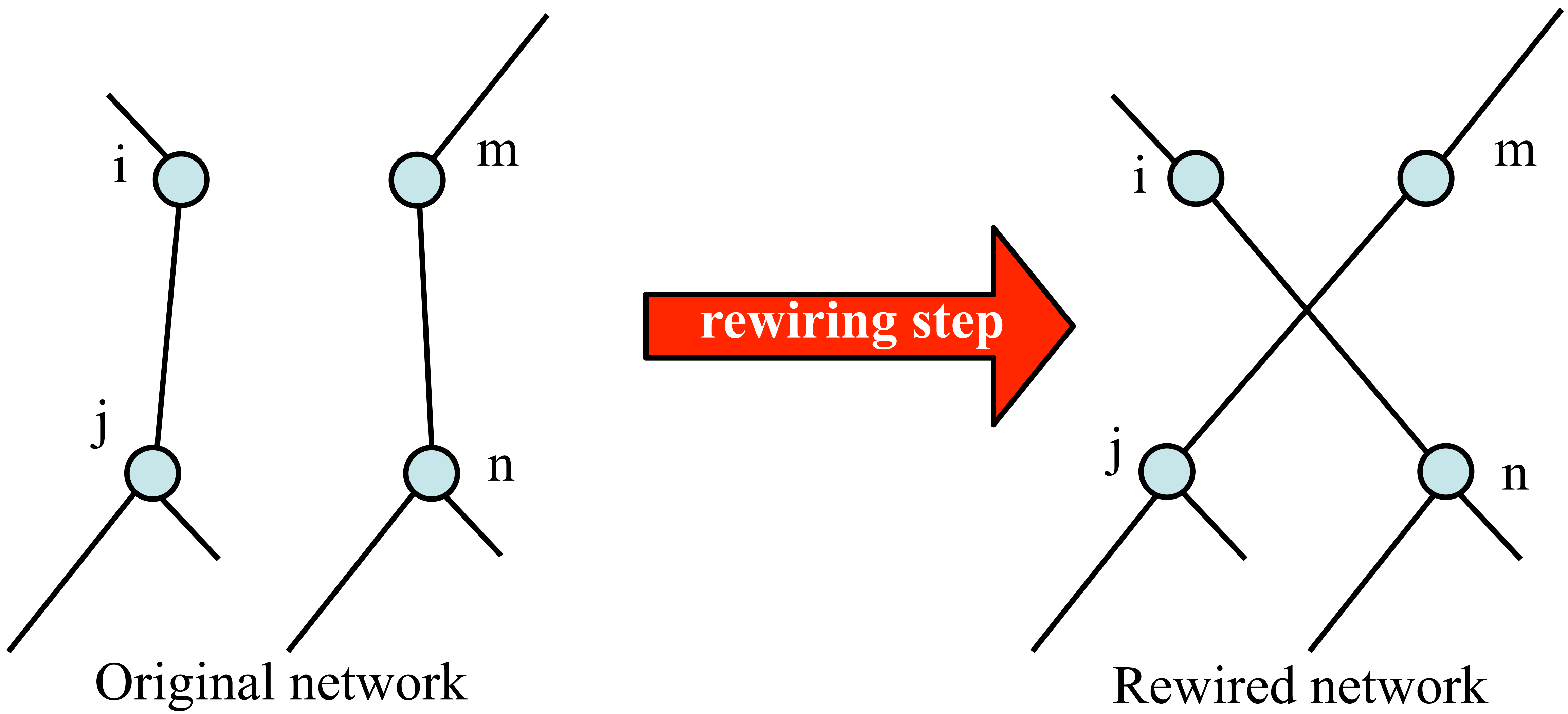
# What are null models?

- ensemble of instances of randomly built systems
- that preserve some properties of the studied systems

**Aim**:
• understand which properties of the studied system are simply random, and which ones denote an underlying mechanism or organizational principle
• compare measures with the known values of a random case

# Graph null models

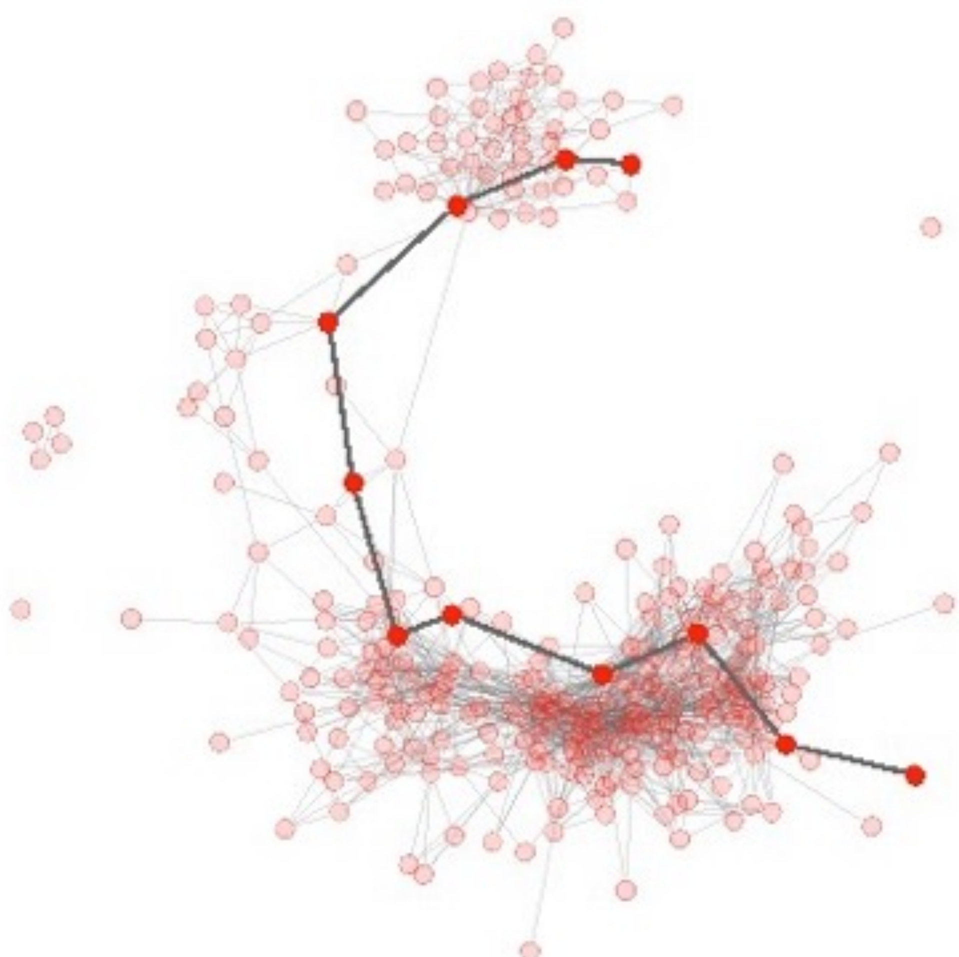
- Fixing size (N, E): random (Erdös-Renyi) graph
- Fixing degree sequence: **reshuffling/rewiring** methods




- preserves the degree of each node
- destroys topological correlations

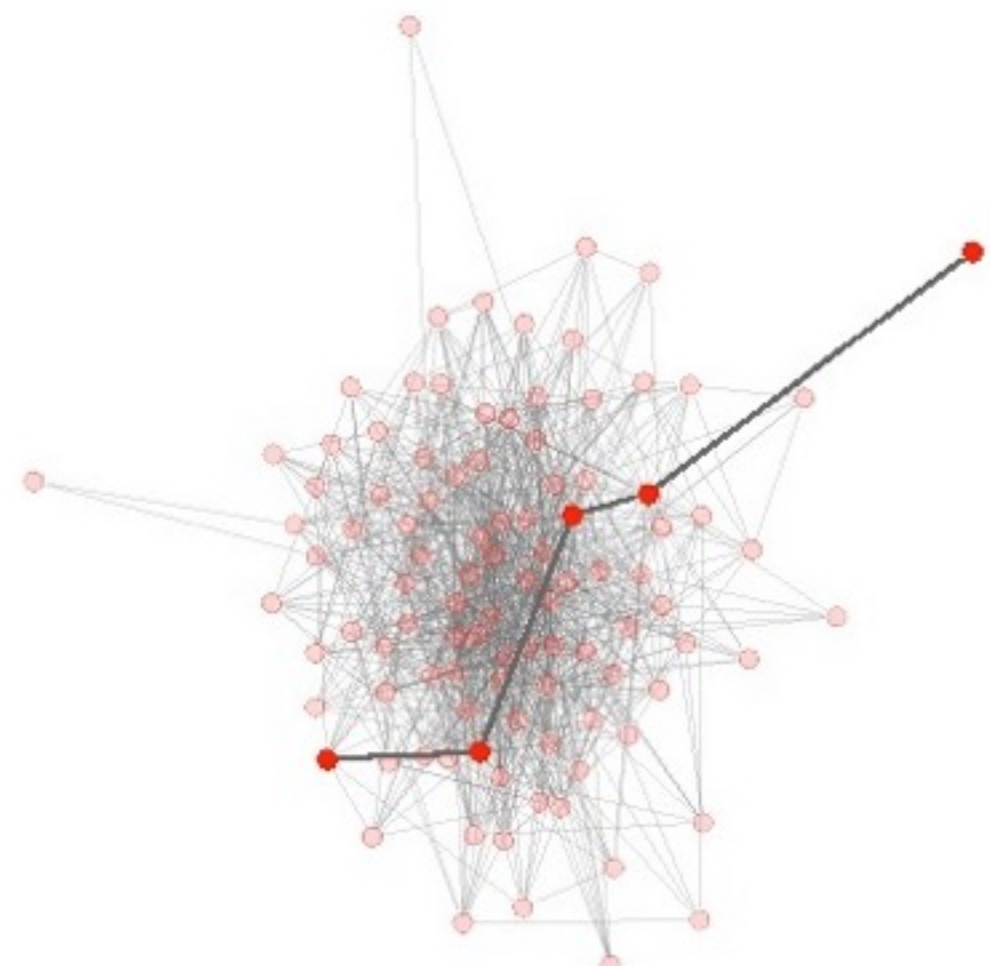
# An example:
# daily cumulated network of face-to-face interactions

Museum (SG)



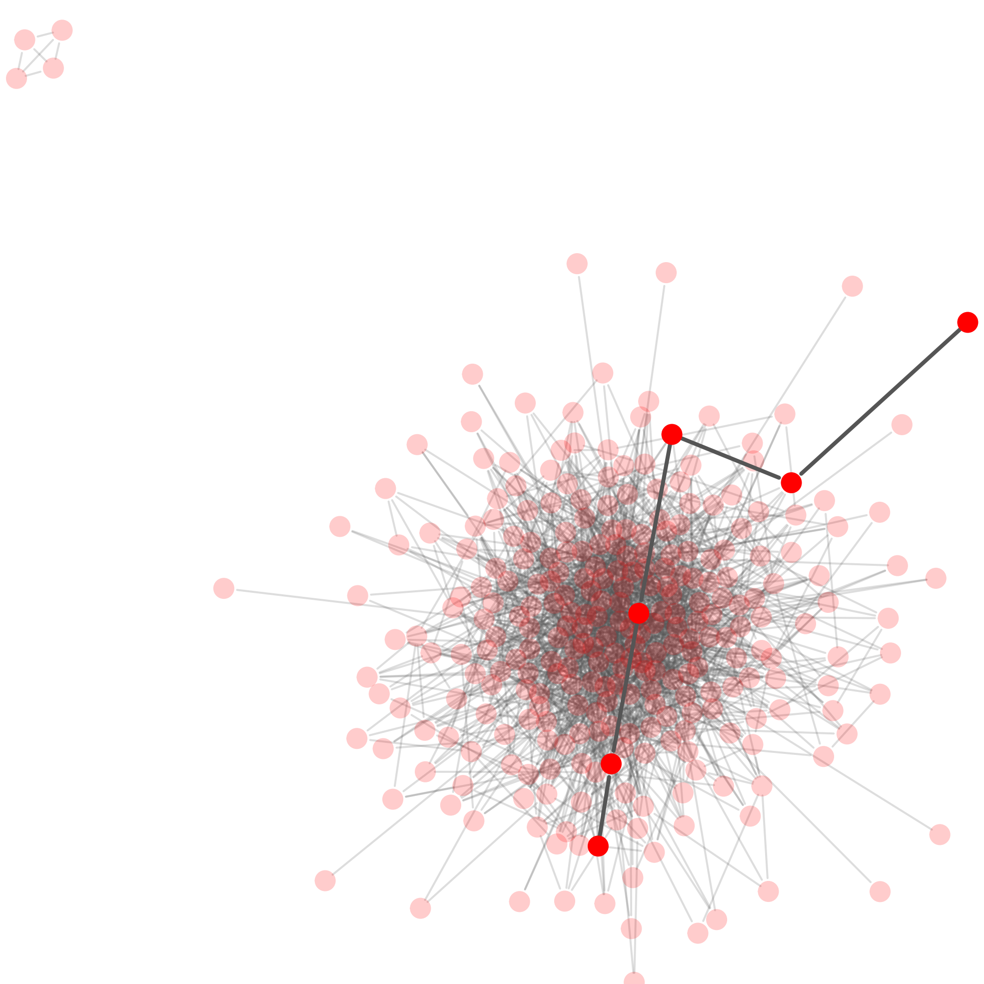
“seems” not to be
a small-world network

Conference (HT09)

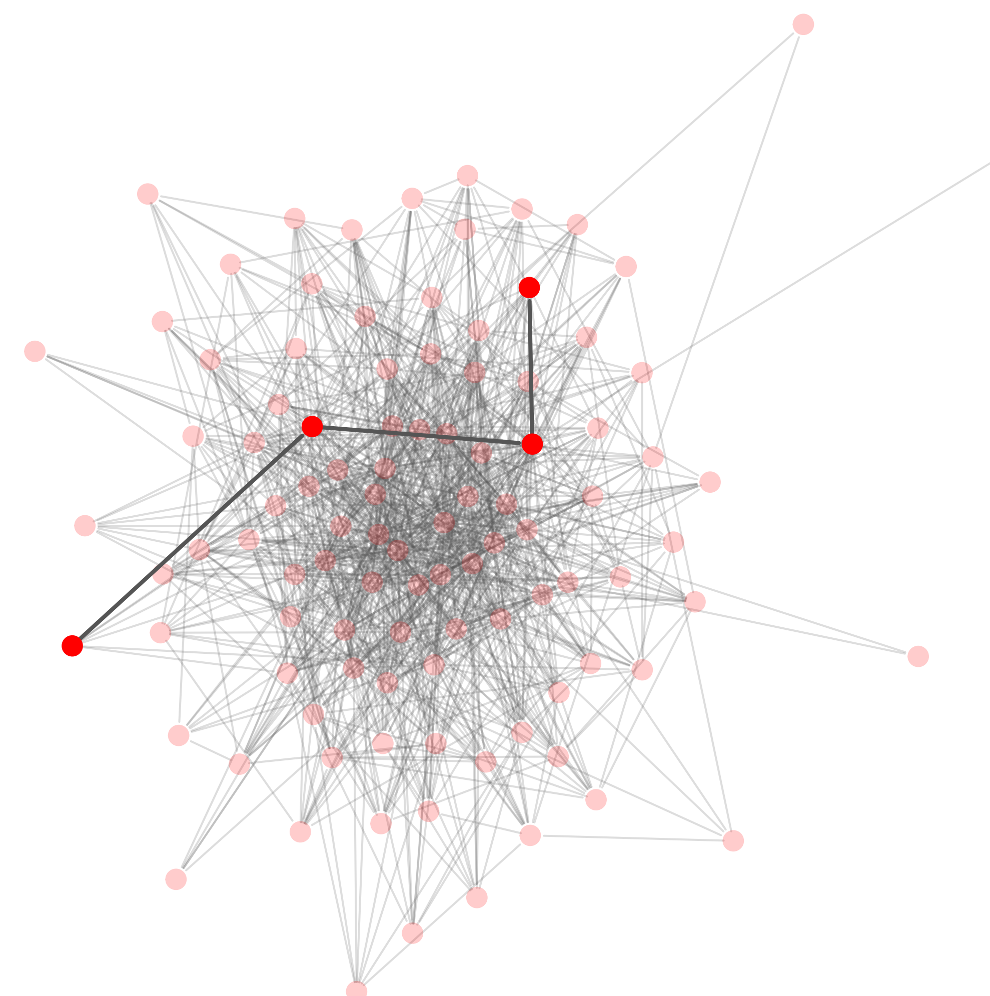


“seems” small-world

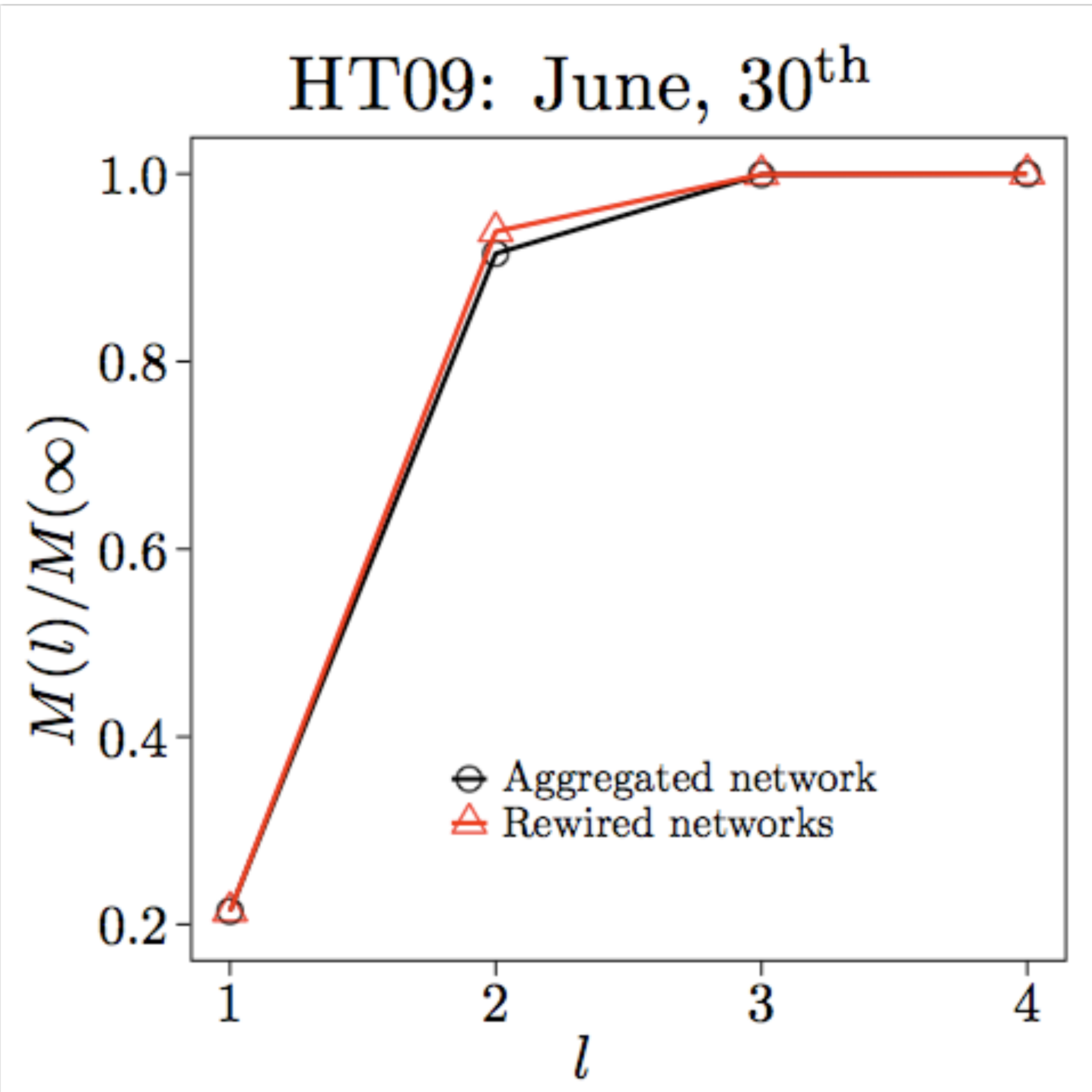
Museum (SG), rewired



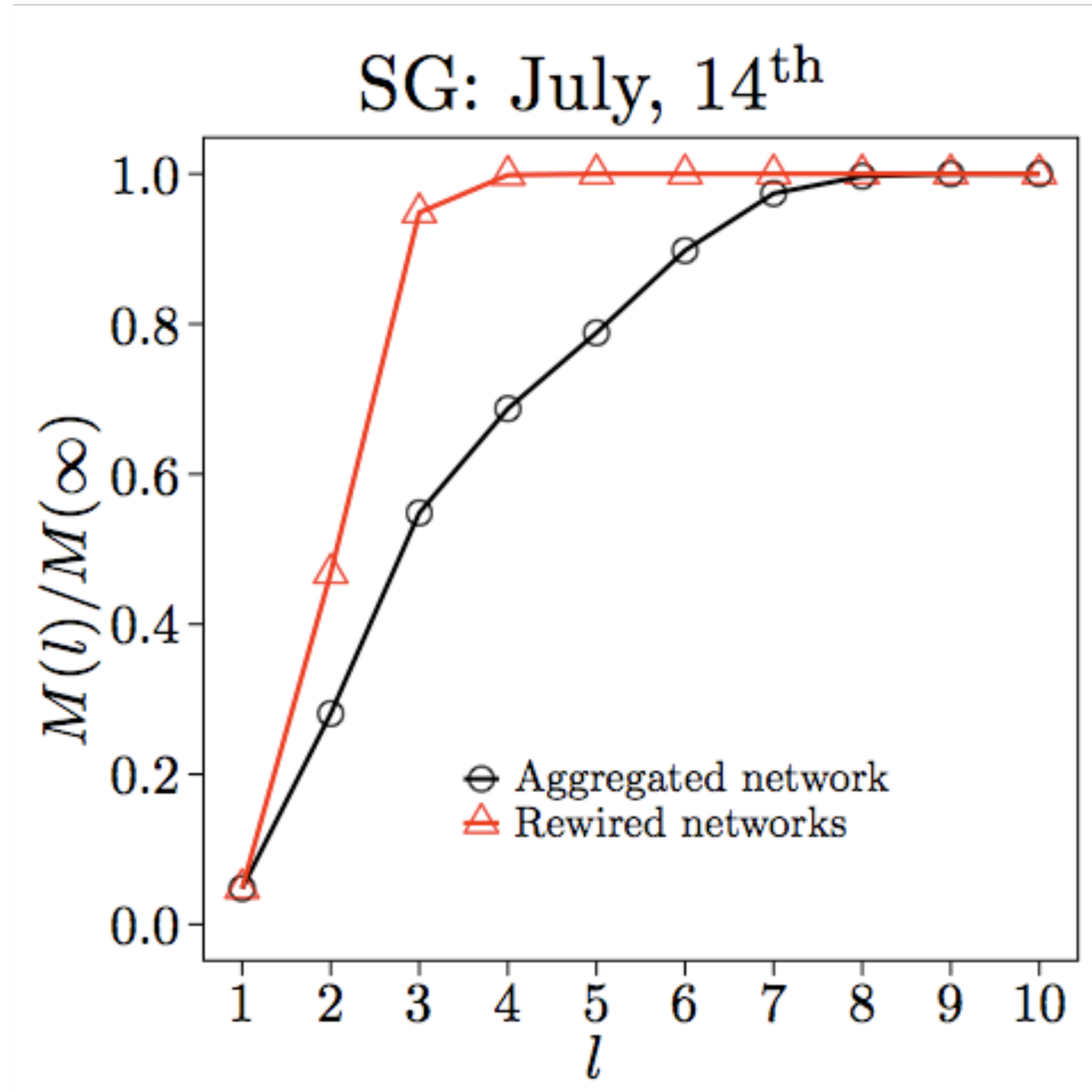
Conference (HT09), rewired



# (non) Small-worldness




Small-world




Non small-world