Complex networks: an introduction

Alain Barrat

CPT, Marseille, France

ISI, Turin, Italy

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

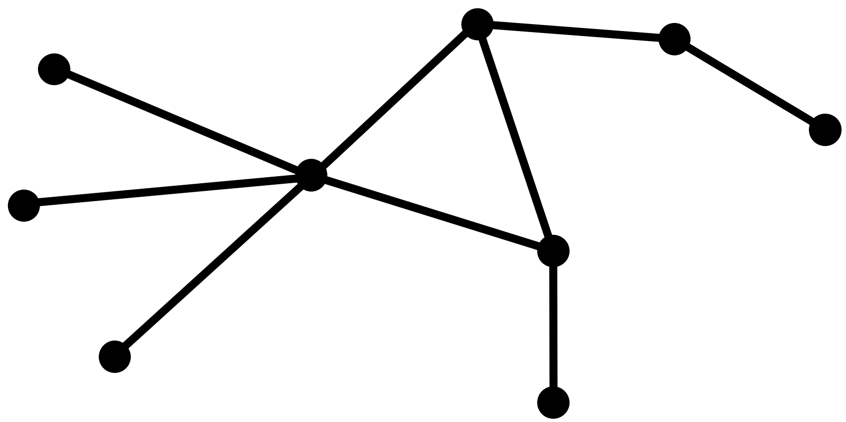
http://cxnets.googlepages.com

Plan of the lecture

I. **INTRODUCTION**
   I. **Networks: definitions, statistical characterization**
   II. Real world networks
II. DYNAMICAL PROCESSES
   I. Resilience, vulnerability
   II. Random walks
   III. Epidemic processes
   IV. (Social phenomena)
   V. Some perspectives

What is a network

Network=set of nodes joined by links



very abstract representation

→ very general

→ convenient to describe many different systems

Some examples

	Nodes	Links
Social networks	Individuals	Social relations
Internet	Routers AS	Cables 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:

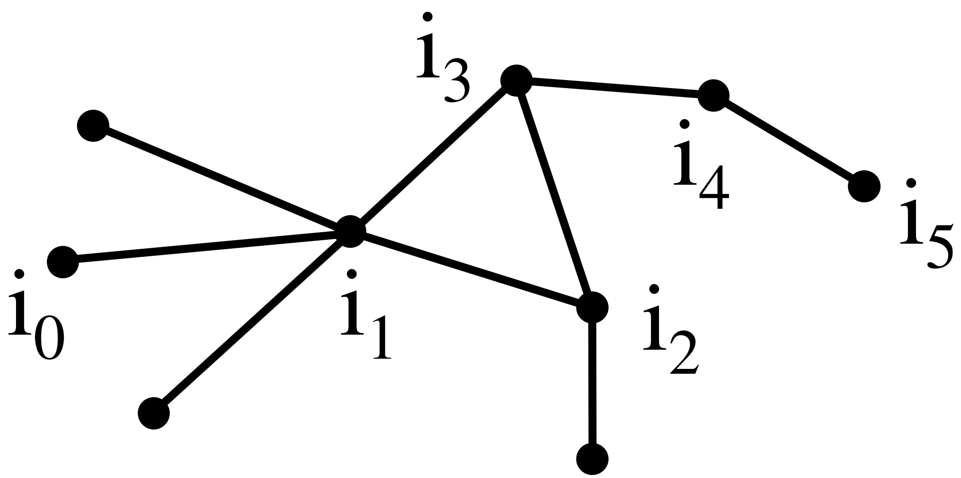
- Empirics
- Characterization
- Modeling
- Dynamical processes

Paths

G=(V,E)

Path of length n = ordered collection of

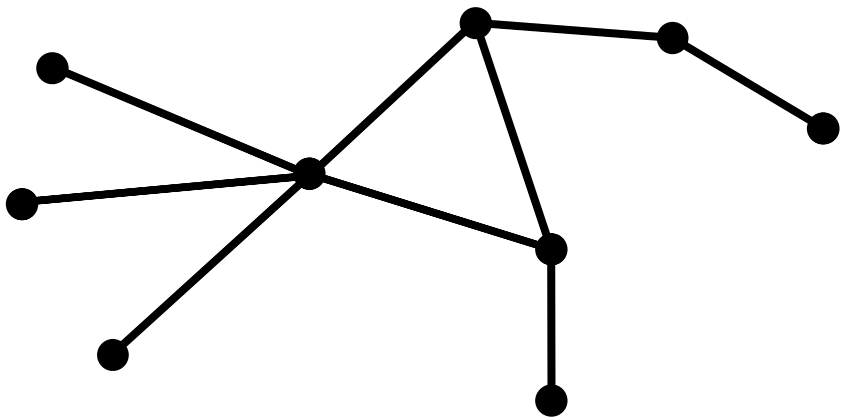
- n+1 vertices $i_0, i_1, \ldots, i_n \in V$
- n edges $(i_0,i_1), (i_1,i_2) \ldots, (i_{n-1},i_n) \in E$



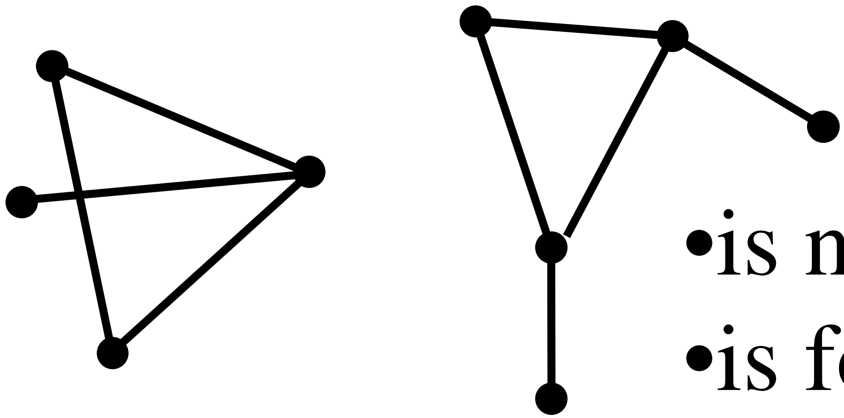
Cycle/loop = closed path (i_0=i_n)

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

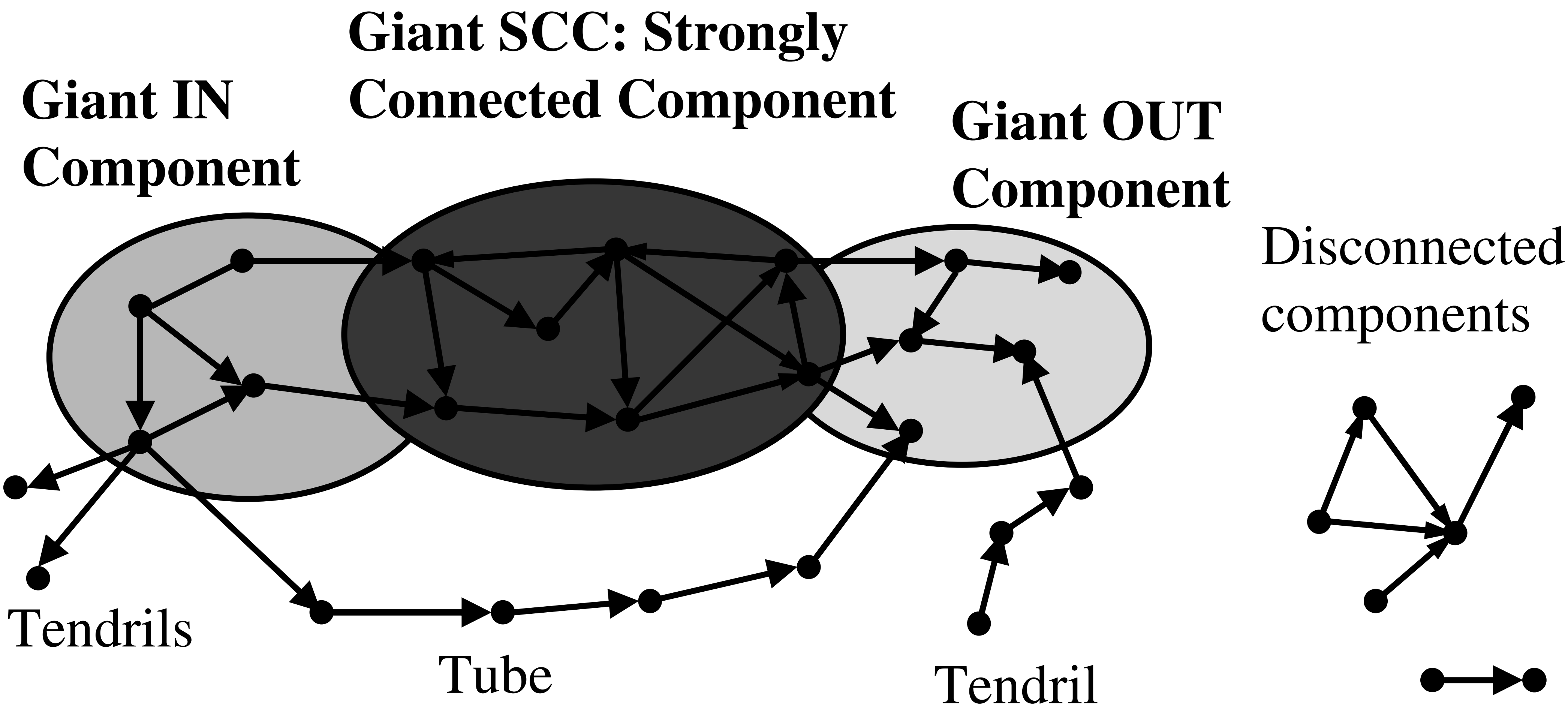
G=(V,E)=> distribution of components' sizes

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

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

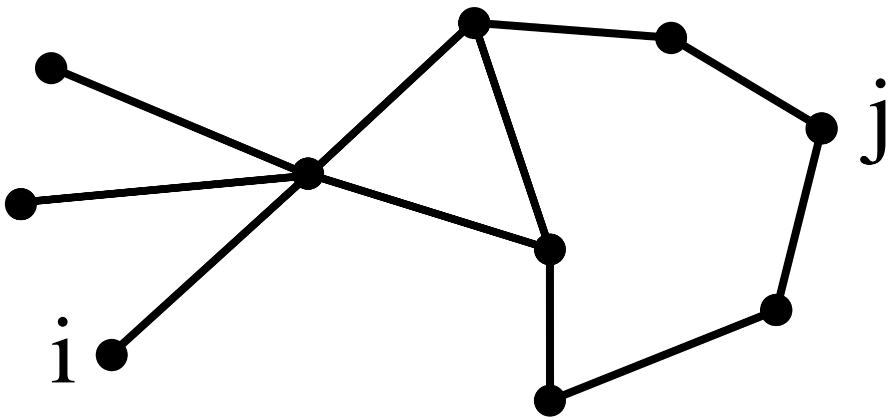
Paths and connectedness: directed graphs

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)

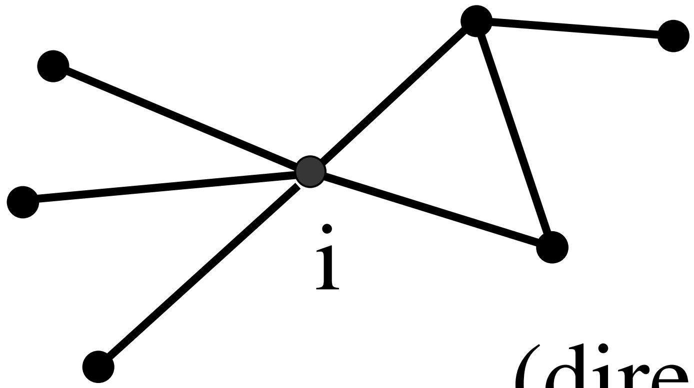
Complete graph: l(i,j)=1 for all i,j

“Small-world”: “small” diameter

Centrality measures

How to quantify the importance of a node?

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



k_i=5

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

- Closeness centrality

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

Betweenness centrality

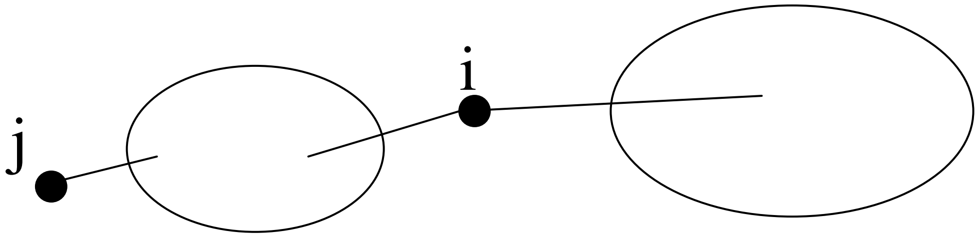
for each pair of nodes (l,m) in the graph, there are

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



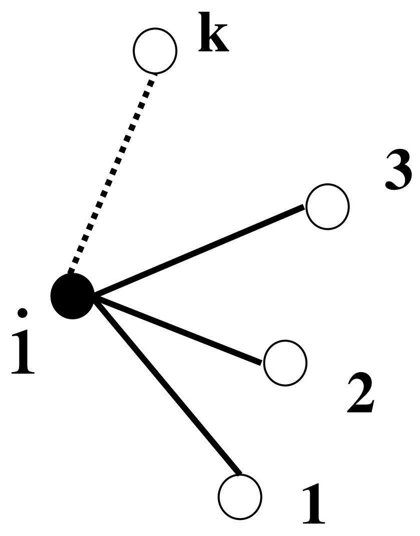
b_i is large

b_j is small

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

NB: generalization to *edge betweenness centrality*

Structure of neighborhoods

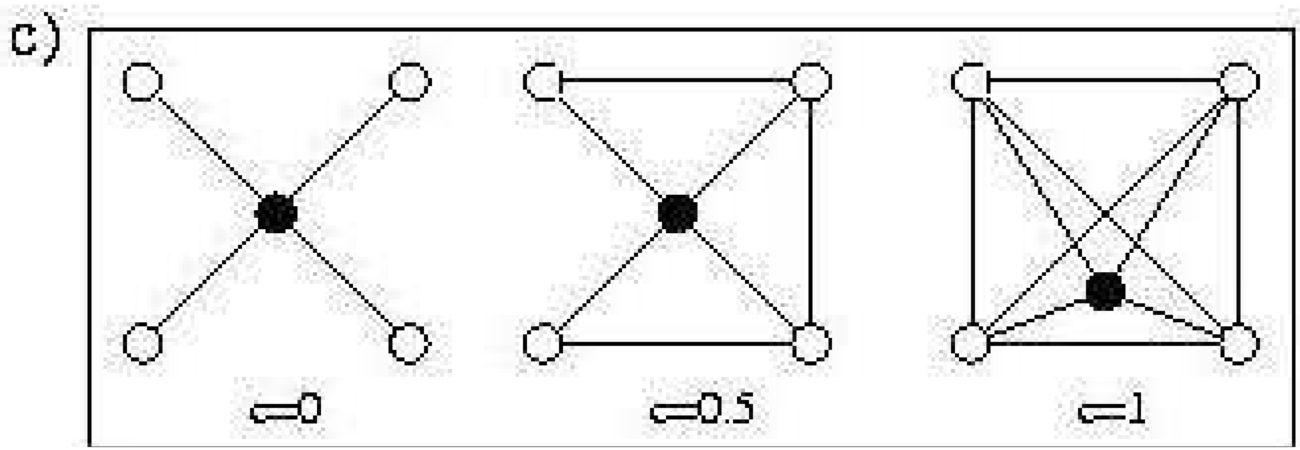


Clustering coefficient of a node

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

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

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



Structure of neighborhoods

Average clustering coefficient of a graph

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

NB: slightly different definition from the fraction of transitive triples:

$$C' = \frac{3 \text{ x number of fully connected triples}}{\text{number of triples}}$$

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

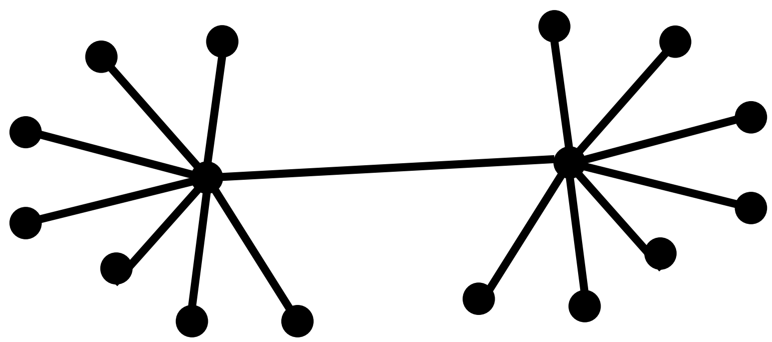
$\langle k^2 \rangle = \sum_i k^2_i/N = \sum_k k^2\, P(k)$

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

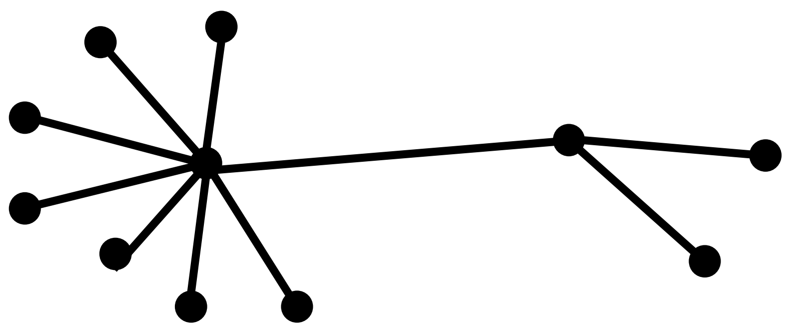
Statistical characterization

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)

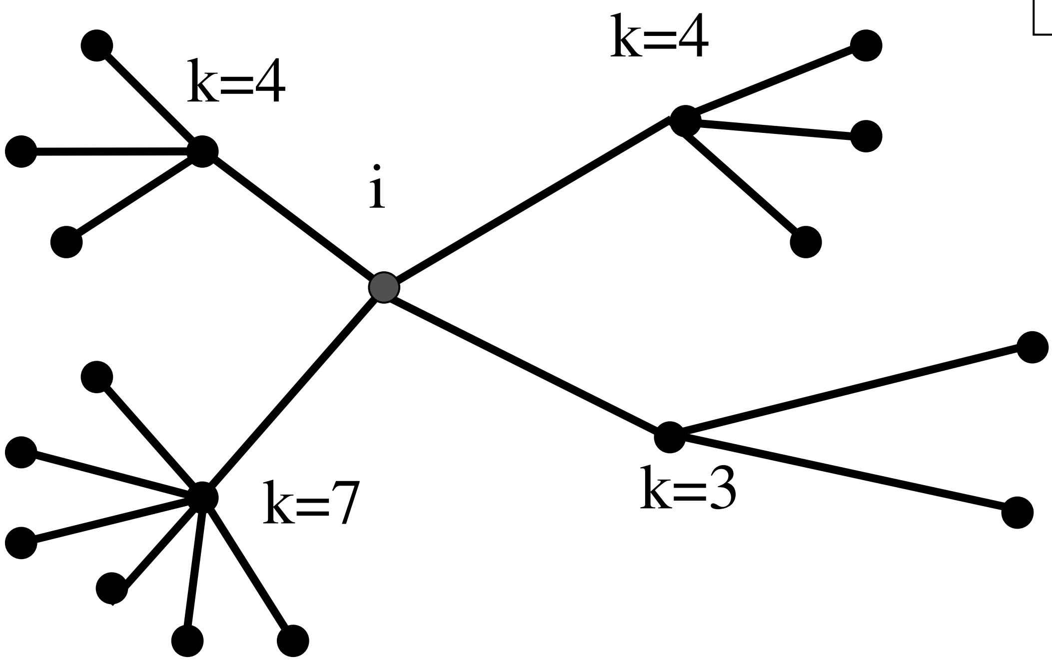
Statistical characterization

Multipoint degree correlations

Practical measure of correlations:

average degree of nearest neighbors

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



$\mathbf{k_i=4}$

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

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

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

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

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

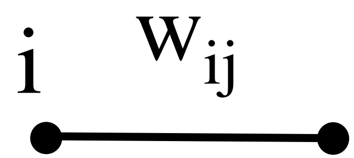
(link with hierarchical structures)

Weighted networks

Real world networks: links

- carry trafic (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 papaers
- Internet, emails: traffic, number of exchanged emails
- Airports: number of passengers
- Metabolic networks: fluxes
- Financial networks: shares
- …

usually $w_{ii}=0$
symetric: $w_{ij}=w_{ji}$

Weighted networks

Weights: on the links

Strength of a node:

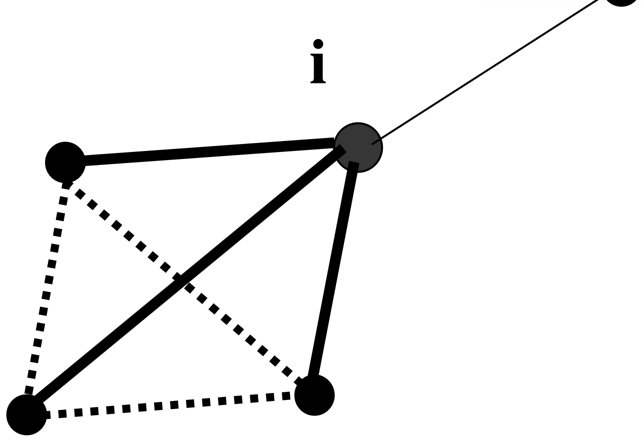
$$s_i = \sum_{j \in V(i)} w_{ij}$$

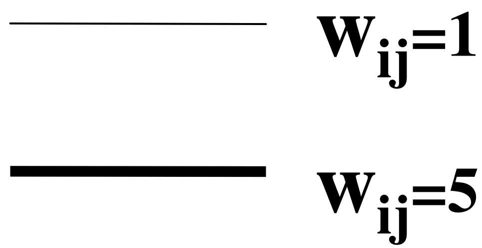
=>Naturally generalizes the degree to weighted networks

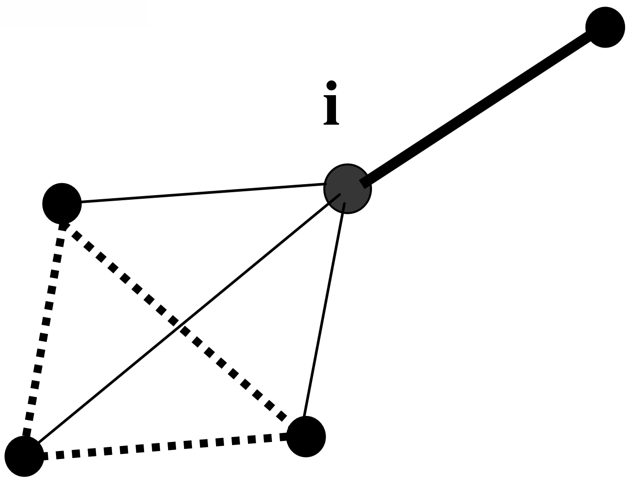
=>Quantifies for example the total trafic at a node

Weighted clustering coefficient

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







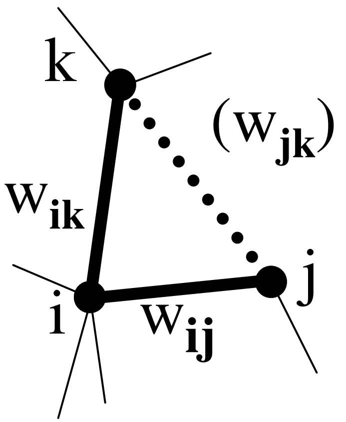
$$C^w(i) = \frac{1}{s_i(k_i-1)} \sum_{j \neq h} a_{ij} a_{jh} a_{ih} \frac{w_{ij} + w_{ih}}{2}$$

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

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

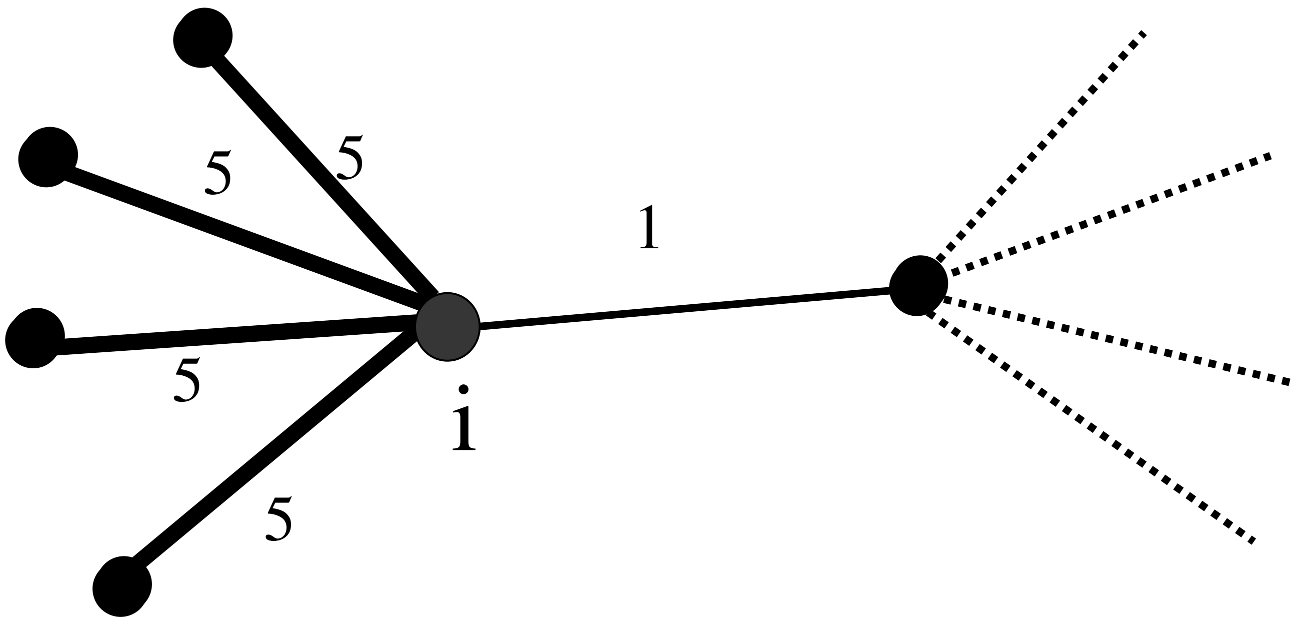
$C > C_w$: less weights on cliques

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

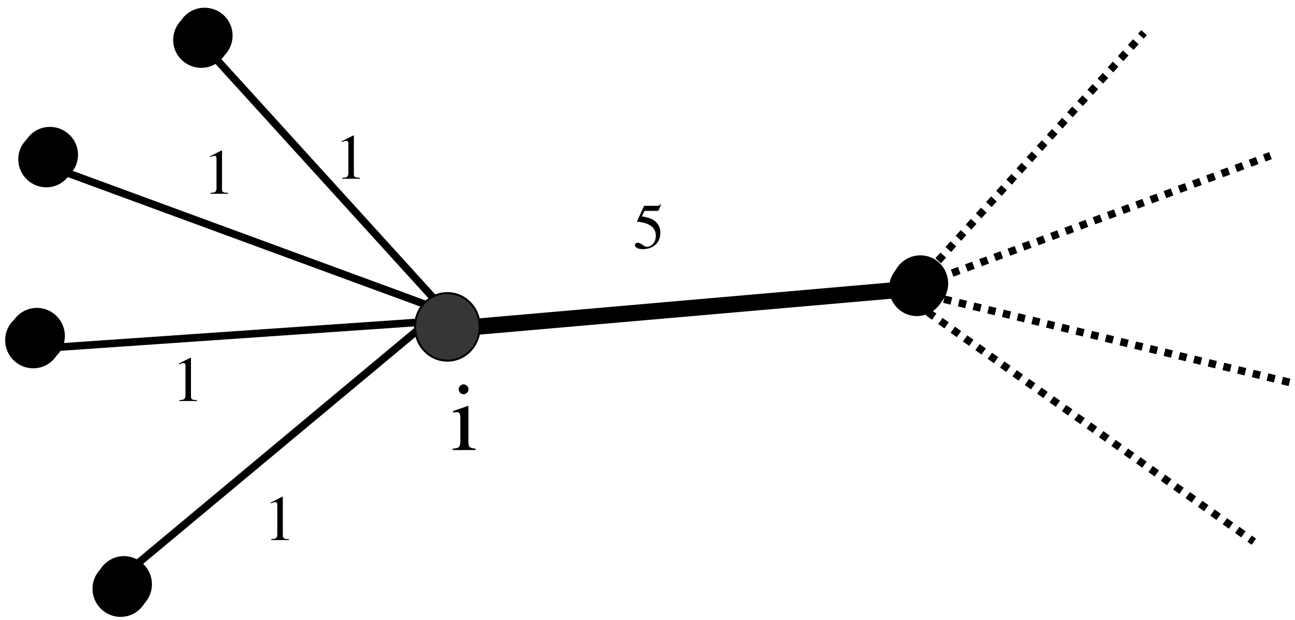
$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} k_j = \frac{1}{k_i} \sum_{j \in V(i)} a_{ij} k_j$$



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

Weighted assortativity

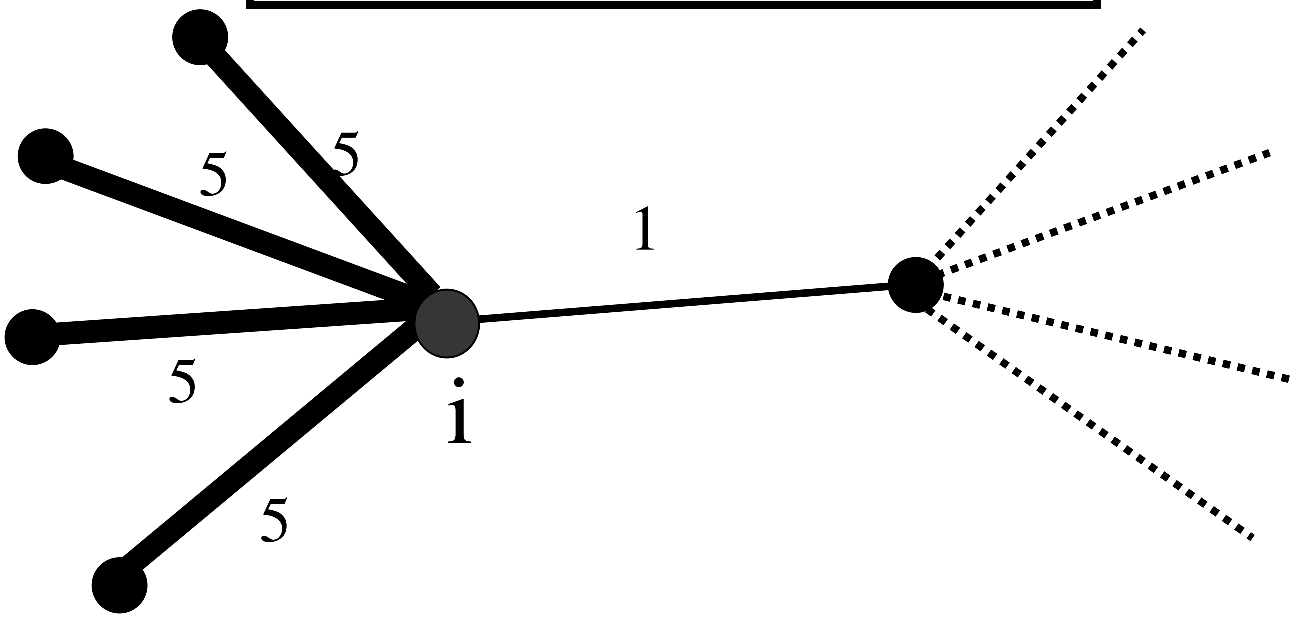
$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} k_j = \frac{1}{k_i} \sum_{j \in V(i)} a_{ij} k_j$$



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

Weighted assortativity

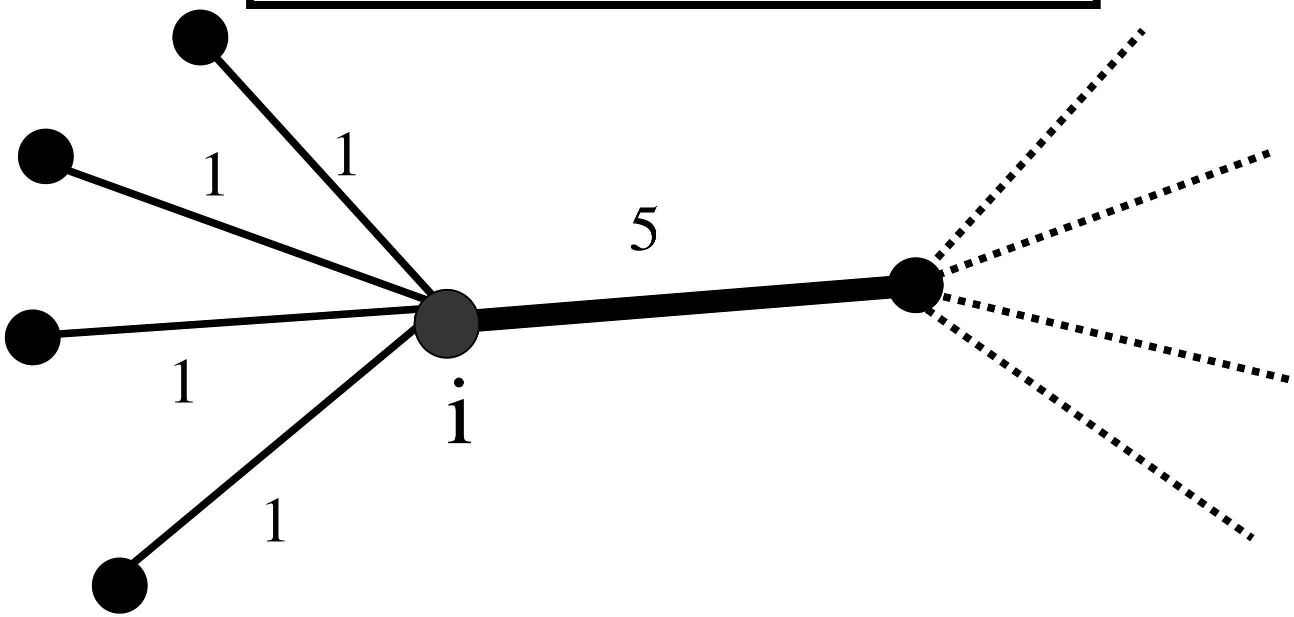
$$k_{nn,i}^{w} = \frac{1}{s_i} \sum_{j \in V(i)} w_{ij} k_j$$



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

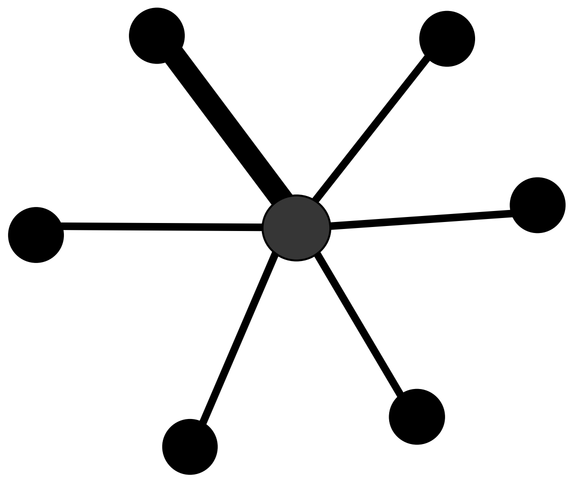
$$k_{nn,i}^{w} = \frac{1}{s_i} \sum_{j \in V(i)} w_{ij} k_j$$

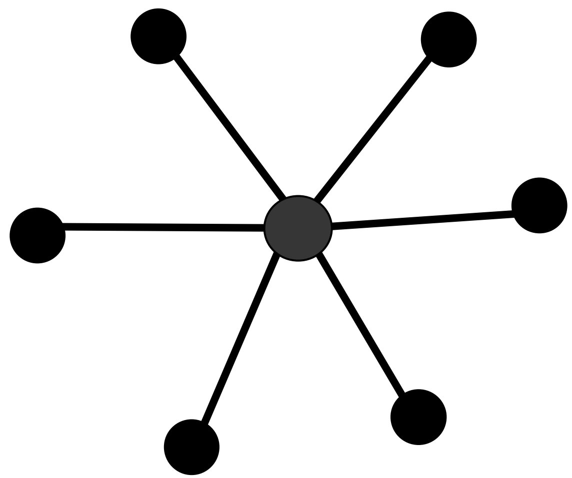


k_i=5; s_i=9; $k_{nn,i}$=1.8 ; $k_{nn,i}^{w}$=3.2: $k_{nn,i} < k_{nn,i}^{w}$

Participation ratio

$$Y_2(i) = \sum_{j \in V(i)} \left[\frac{w_{ij}}{s_i}\right]^2 \begin{cases} 1/k_i \text{ if all weights equal} \\ \text{close to 1 if few weights dominate} \end{cases}$$





Plan of the lecture

I. **INTRODUCTION**
   I. **Networks: definitions, statistical characterization**
   II. **Real world networks**

II. DYNAMICAL PROCESSES
   I. Resilience, vulnerability
   II. Random walks
   III. Epidemic processes
   IV. (Social phenomena)
   V. Some perspectives

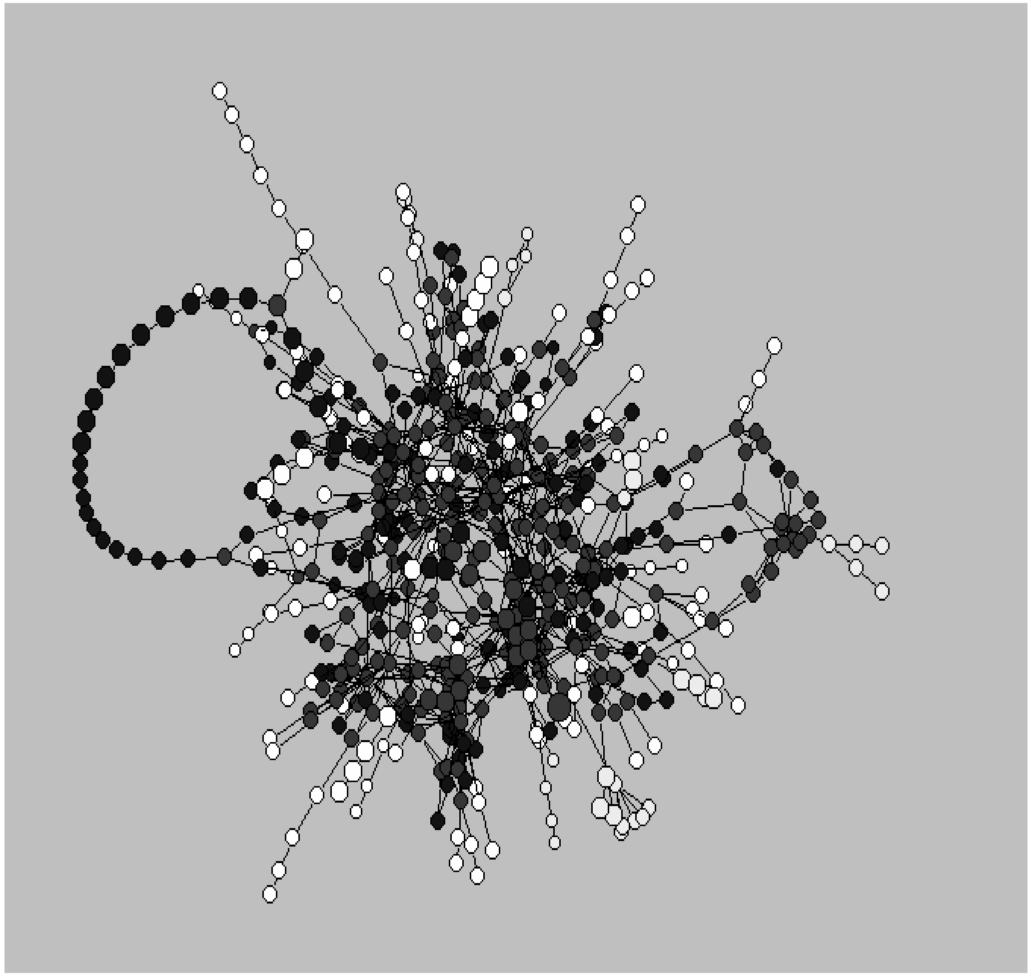
Two main classes

Natural systems:
Biological networks: genes, proteins…
Foodwebs
Social networks

Infrastructure networks:
Virtual: web, email, P2P
Physical: Internet, power grids, transport…

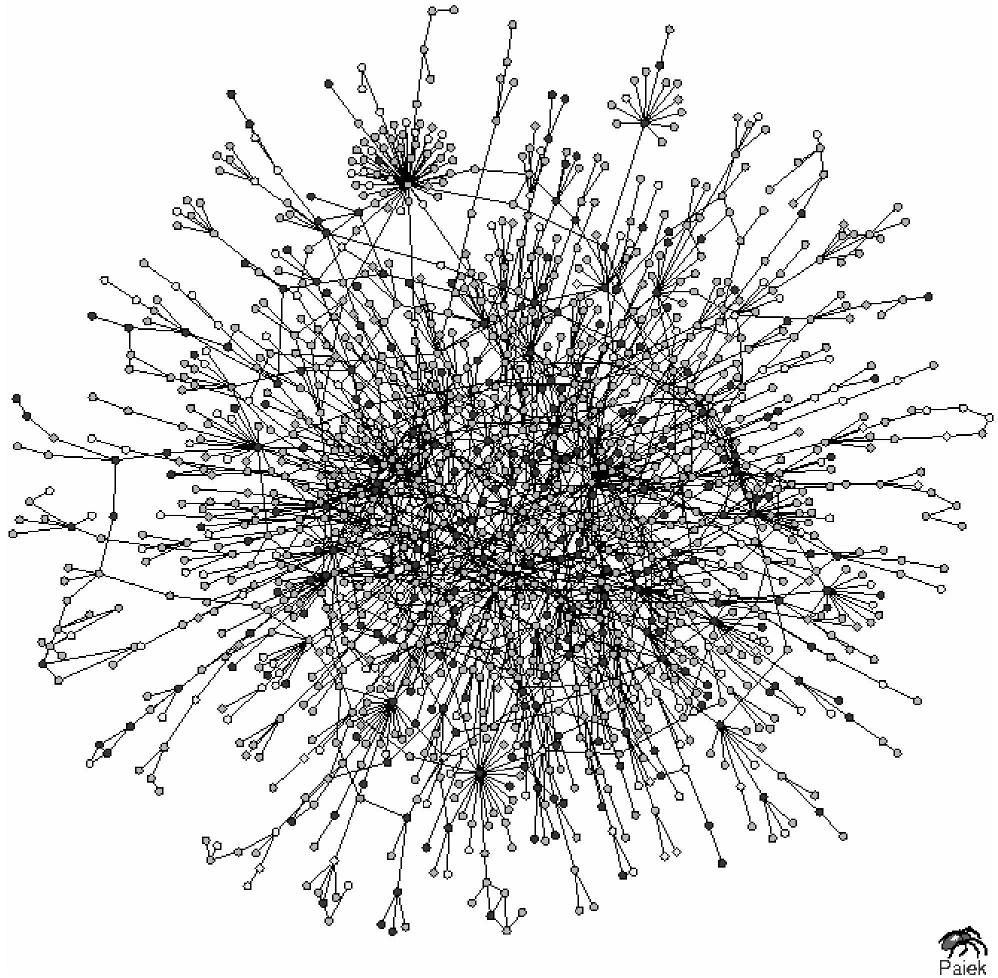
Metabolic Network

Nodes: metabolites
Links:chemical reactions



Protein Interactions

Nodes: proteins
Links: interactions



Scientific collaboration network

Nodes: scientists

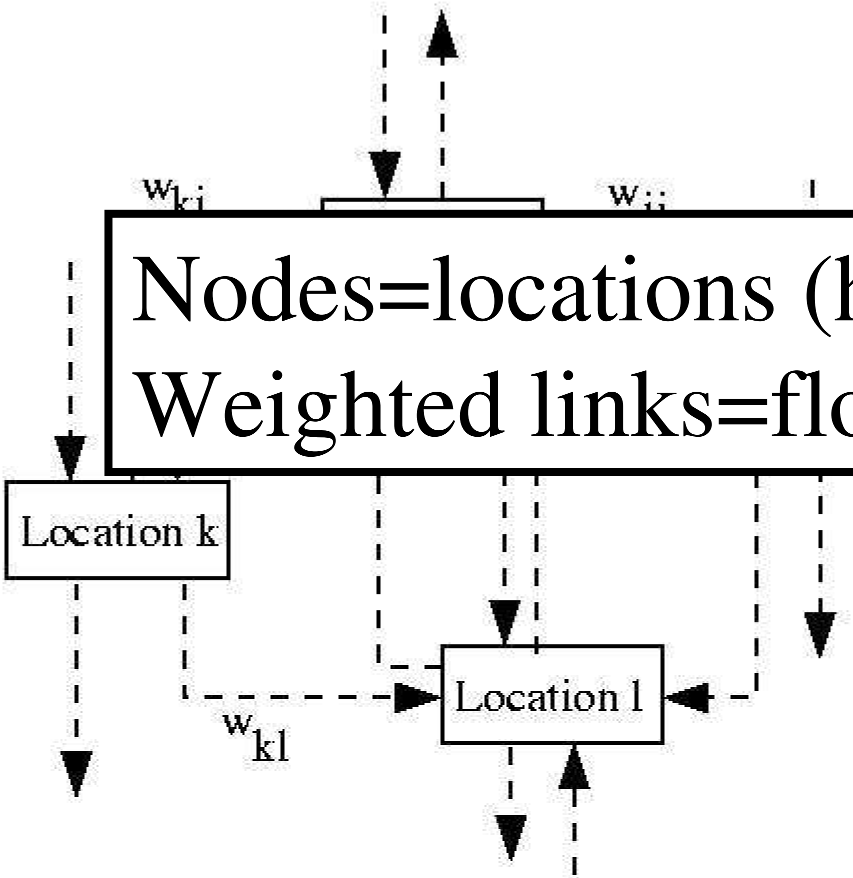
Links: co-authored papers

Weights: depending on

- number of co-authored papers
- number of authors of each paper
- number of citations…

Transportation network: Urban level

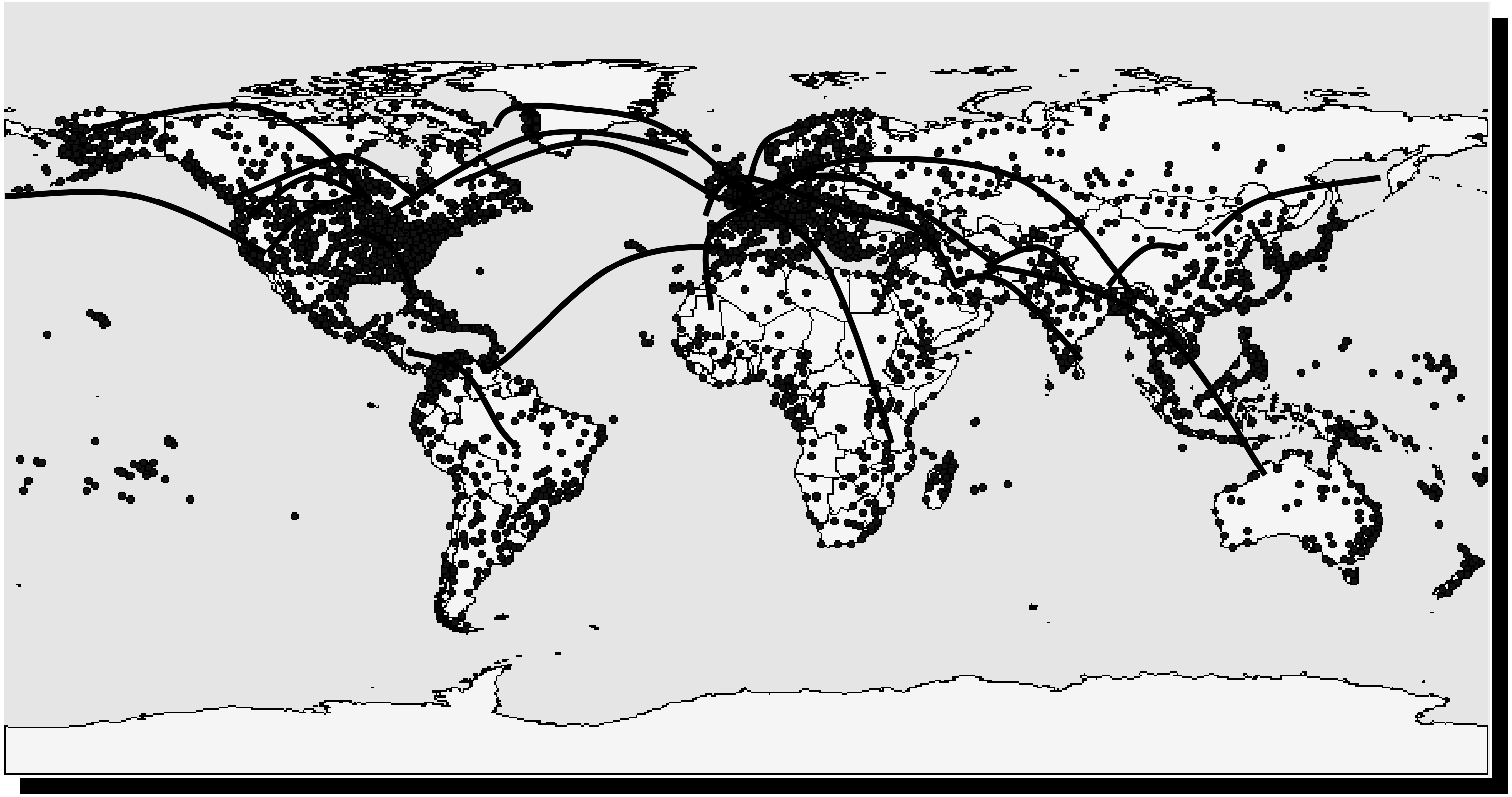
TRANSIMS project



Person ID	Location ID	Location type	Arrival time	Departure time
116	4356	Home	00:00	07:00
116	21343	Social	19:30	21:00
116	4356	Home	21:00	07:00
324	12679	Home	00:00	07:00
324	431	School	08:00	14:00
324	12679	Home	14:30	19:00

Nodes=locations (homes, shops, offices…)
Weighted links=flow of individuals

World airport network



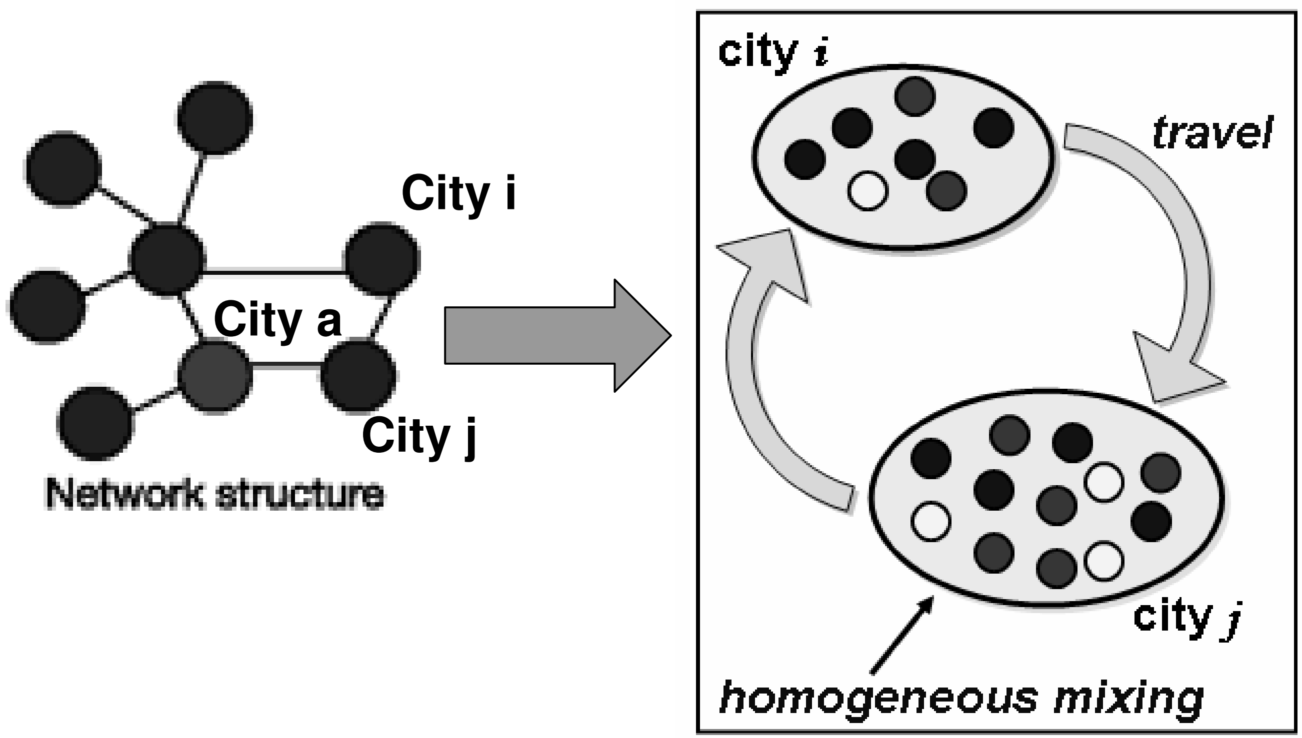
complete IATA database

- ***V* = *3100* airports**
- ***E* = *17182* weighted edges**
- **w_{ij} #seats / (time scale)**

> 99% of total traffic

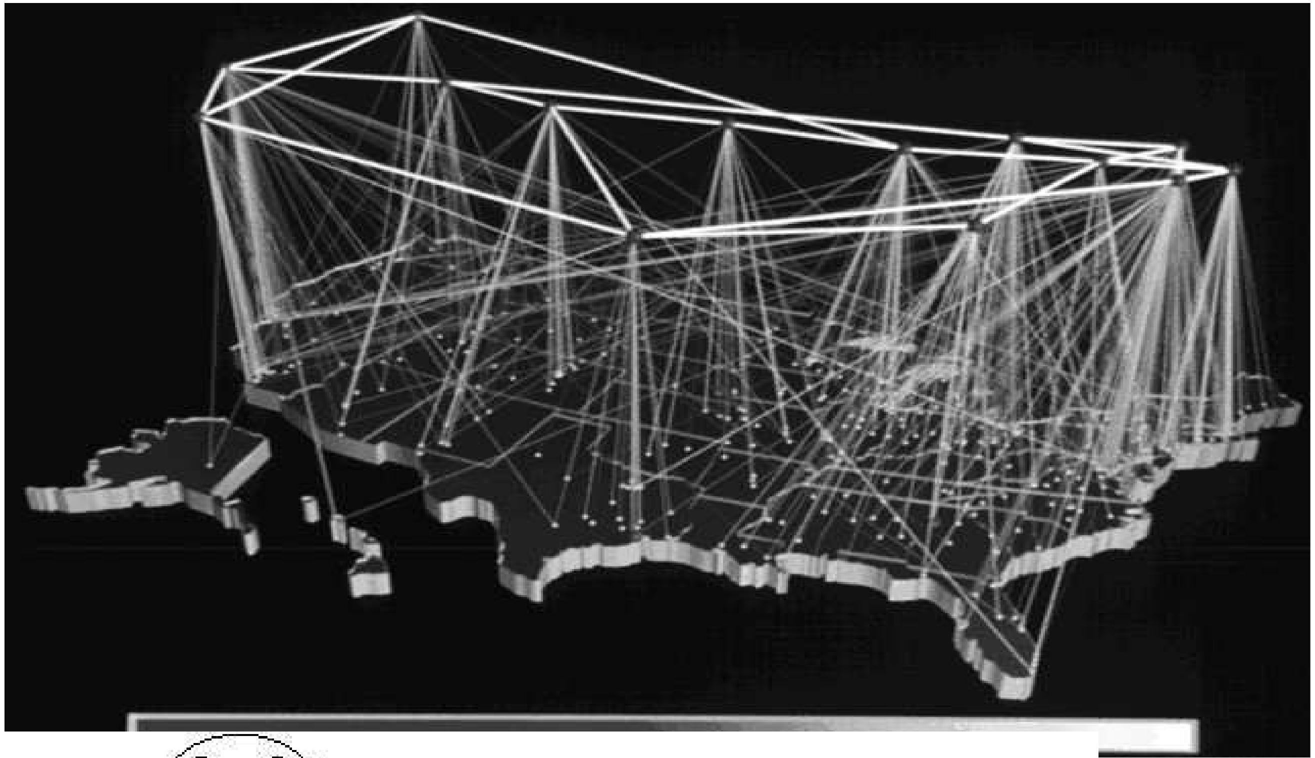
Meta-population networks

Each node: internal structure
Links: transport/traffic

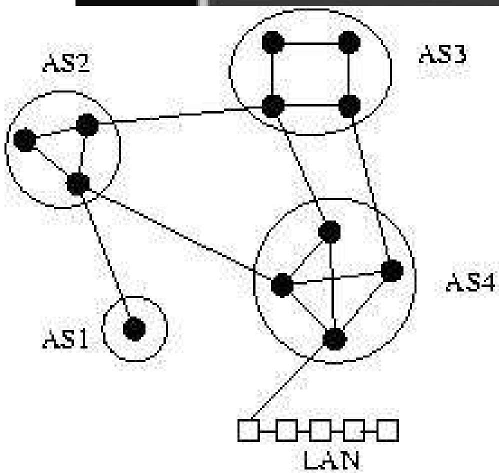


Internet

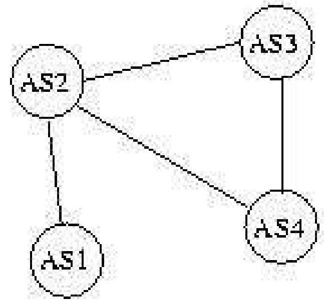
- **Computers (routers)**
- **Satellites**
- **Modems**
- **Phone cables**
- **Optic fibers**
- **EM waves**



different granularities



Router Level



Autonomous System level

Internet mapping

- continuously evolving and growing
- intrinsic heterogeneity
- self-organizing

→ Largely unknown topology/properties

Mapping projects:

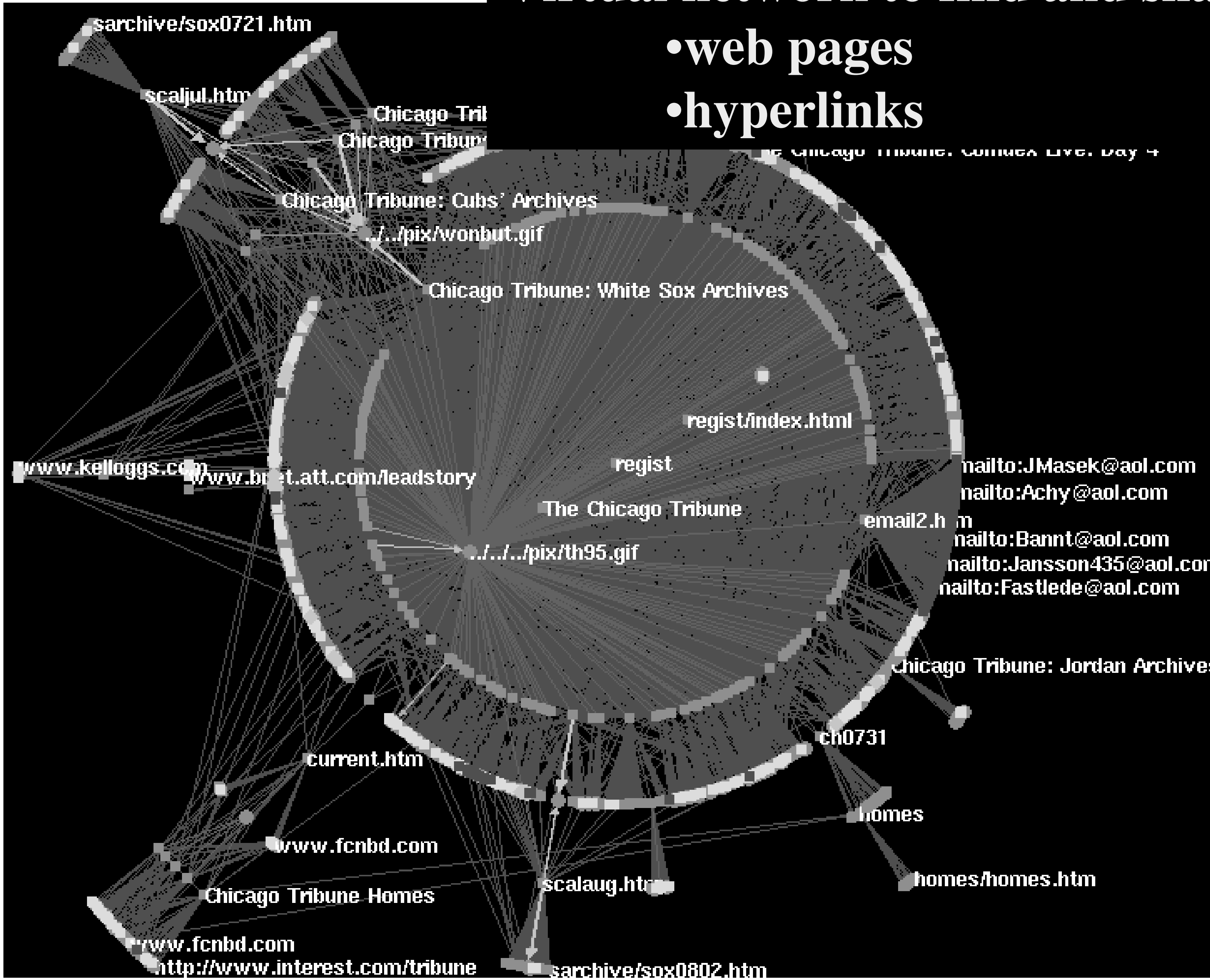
- **Multi-probe reconstruction (router-level): traceroute**
- **Use of BGP tables for the Autonomous System level (domains)**

•CAIDA, NLANR, RIPE, IPM, PingER, DIMES → **Topology and performance measurements**

The World-Wide-Web

Virtual network to find and share informations

- **web pages**
- **hyperlinks**



CRAWLS

Sampling issues

- social networks: various samplings/networks
- transportation network: reliable data
- biological networks: incomplete samplings
- Internet: various (incomplete) mapping processes
- WWW: regular crawls
- ...

possibility of introducing biases in the measured network characteristics

Networks characteristics

Networks: of very different origins

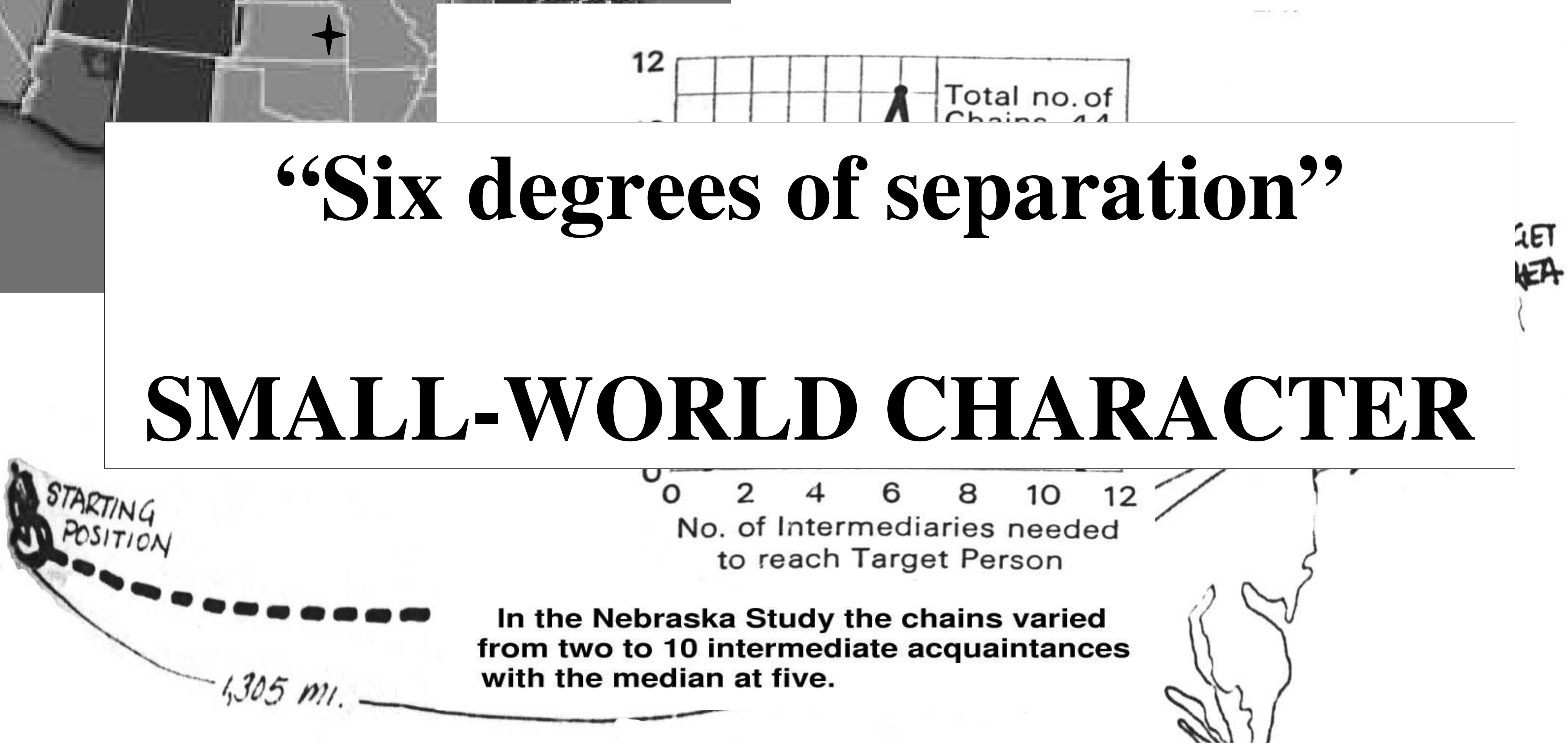
Do they have anything in common?
Possibility to find common properties?

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

Social networks: Milgram's experiment

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

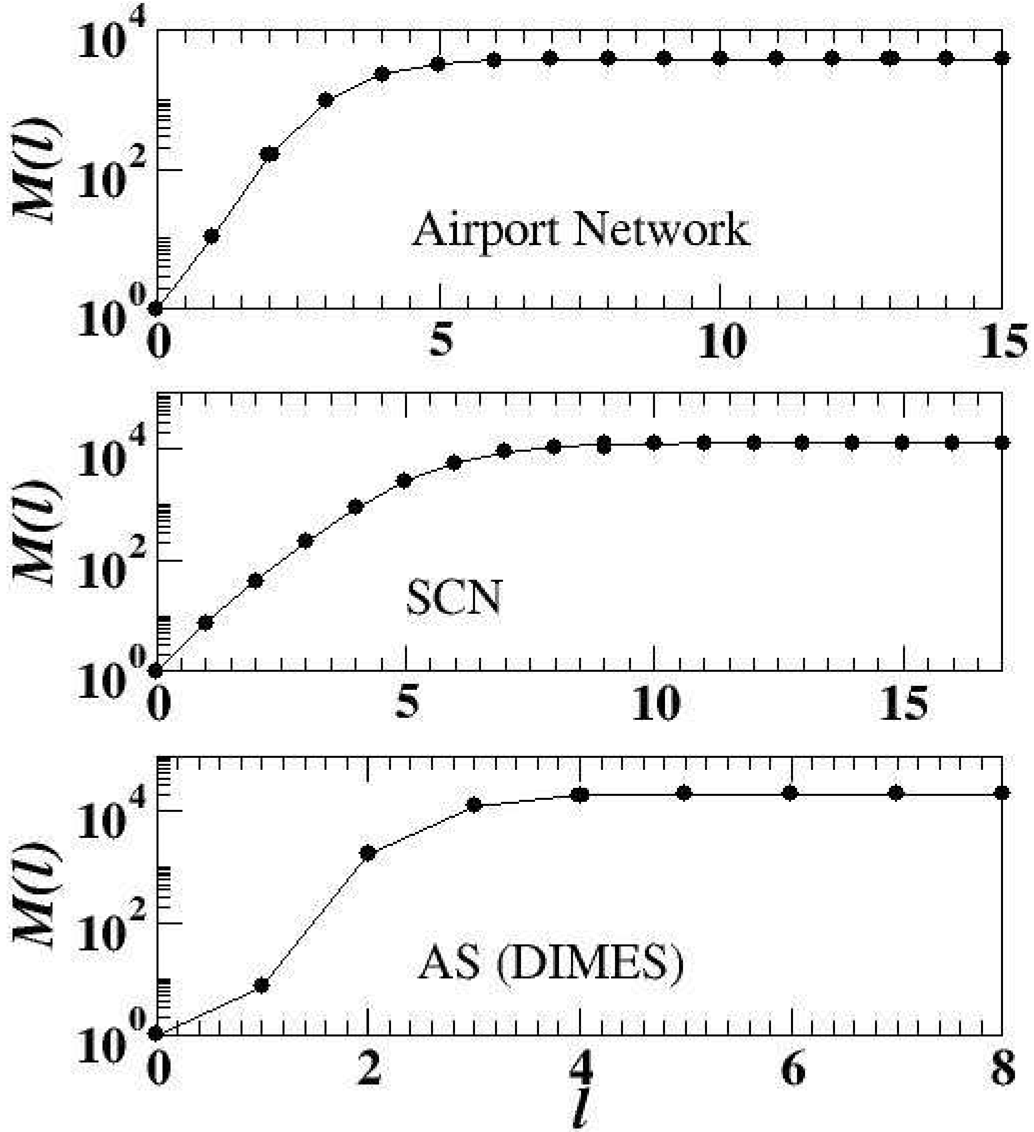
Dodds et al., *Science* 301, 827 (2003)



"Six degrees of separation"

SMALL-WORLD CHARACTER

Small-world properties



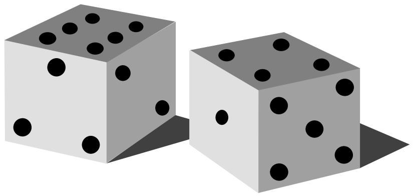
Average number of nodes within a chemical distance l

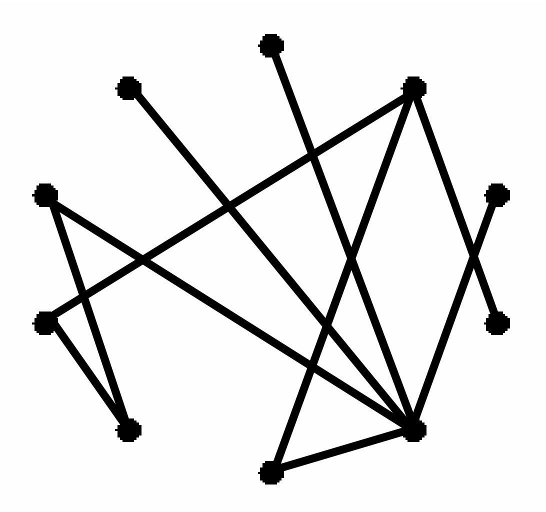
Scientific collaborations

Internet

Small-world properties

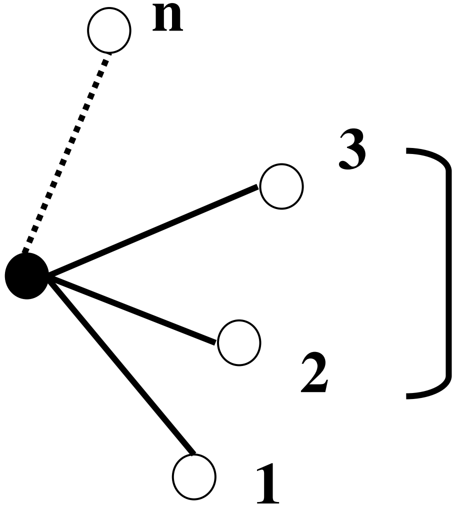
N points, links with proba p:
static random graphs





short distances
(log N)

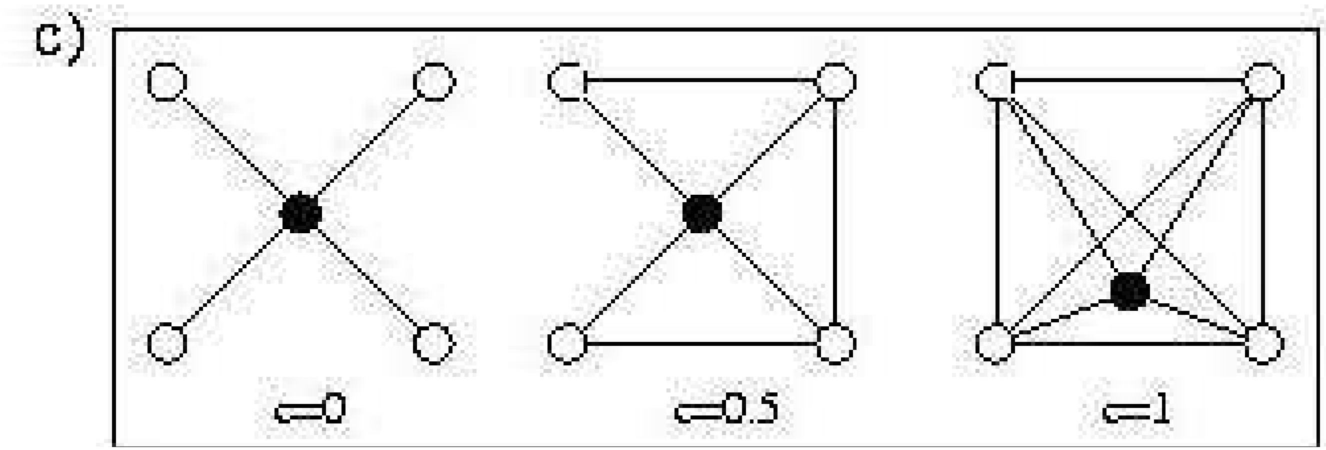
Clustering coefficient



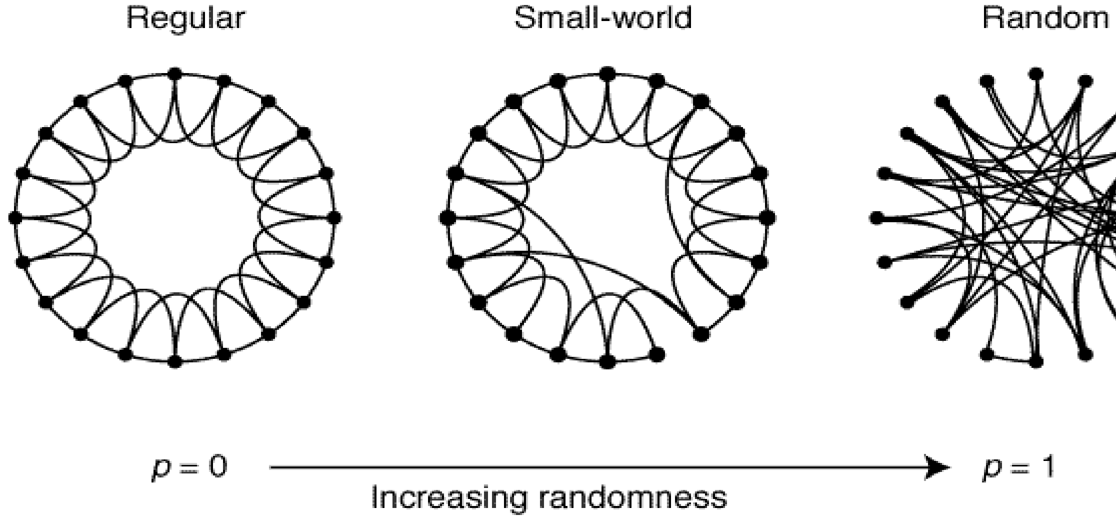
Empirically: large clustering coefficients

Higher probability to be connected

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

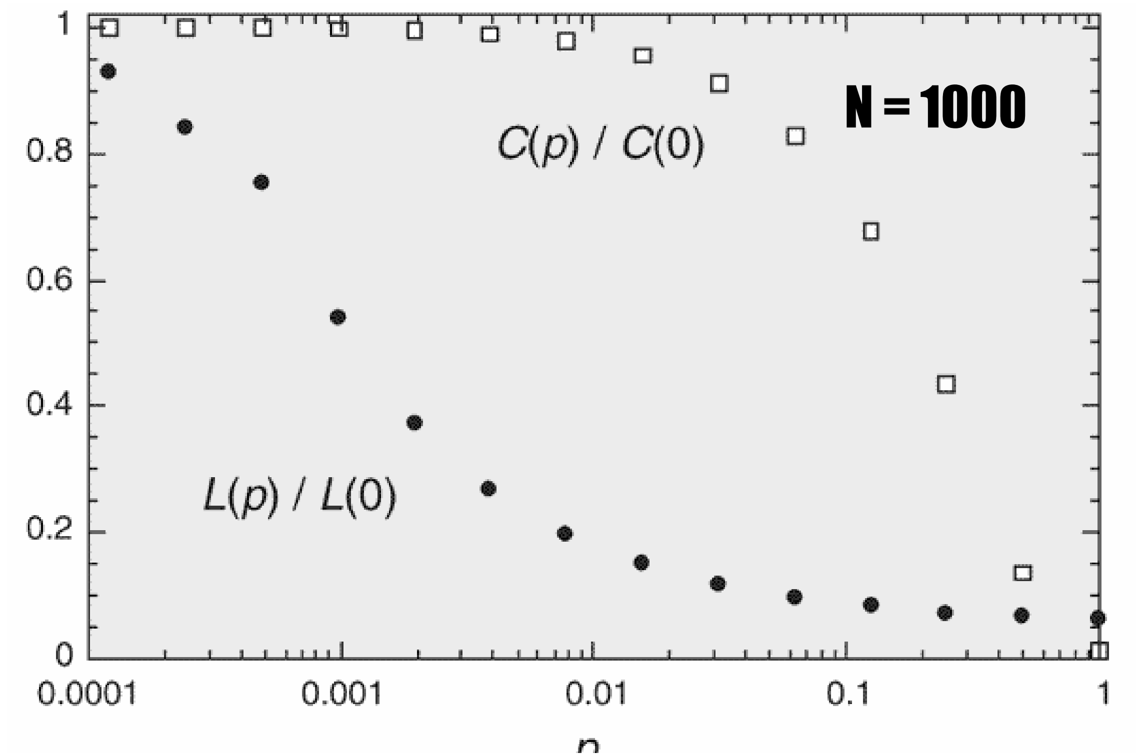


Small-world networks



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

=>Shortcuts



- **Large clustering coeff.**
- **Short typical path**

Watts & Strogatz,

***Nature* 393, 440 (1998)**

Topological heterogeneity

Statistical analysis of centrality measures:

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

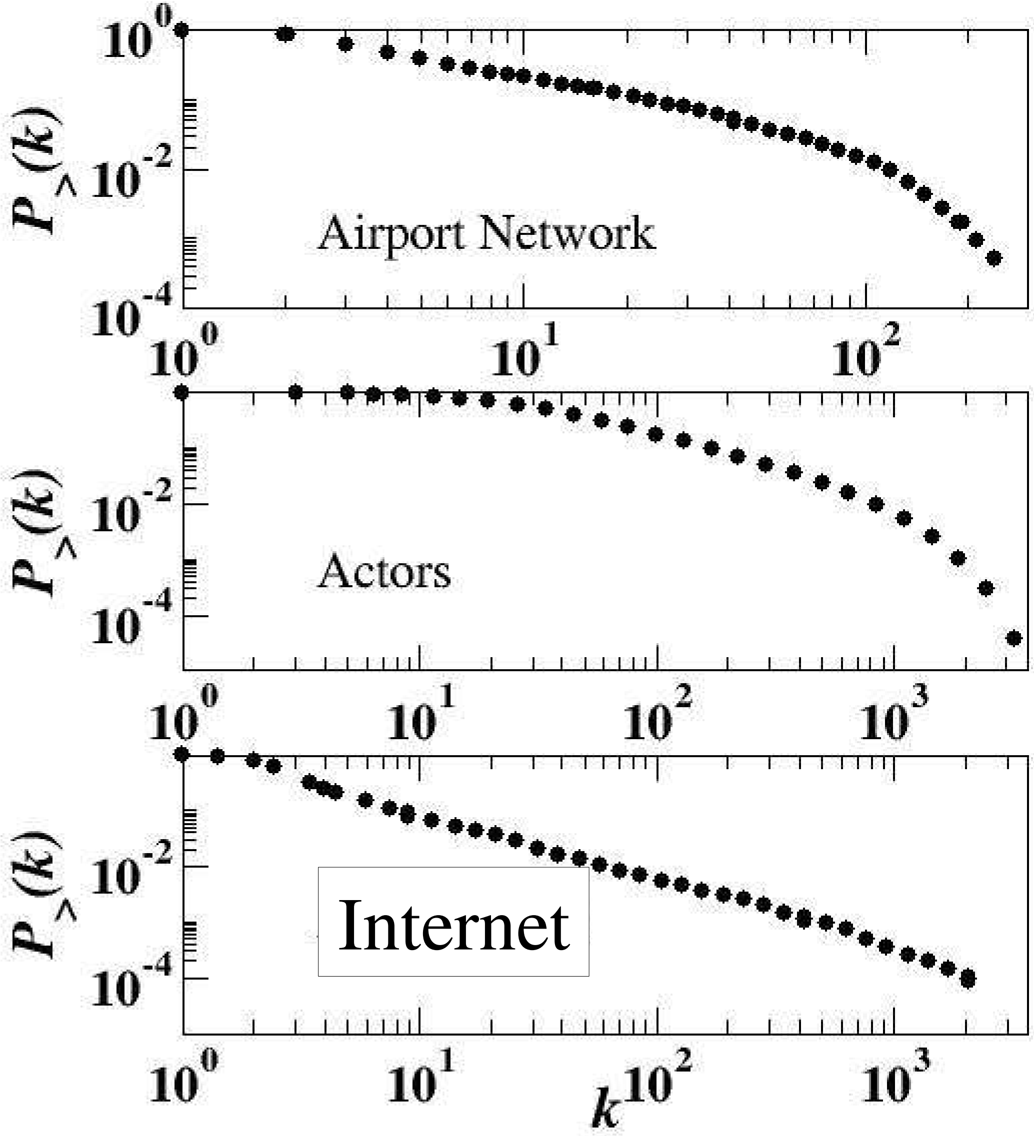
also: P(b), P(w)….

Two broad classes

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

Topological heterogeneity

Statistical analysis of centrality measures



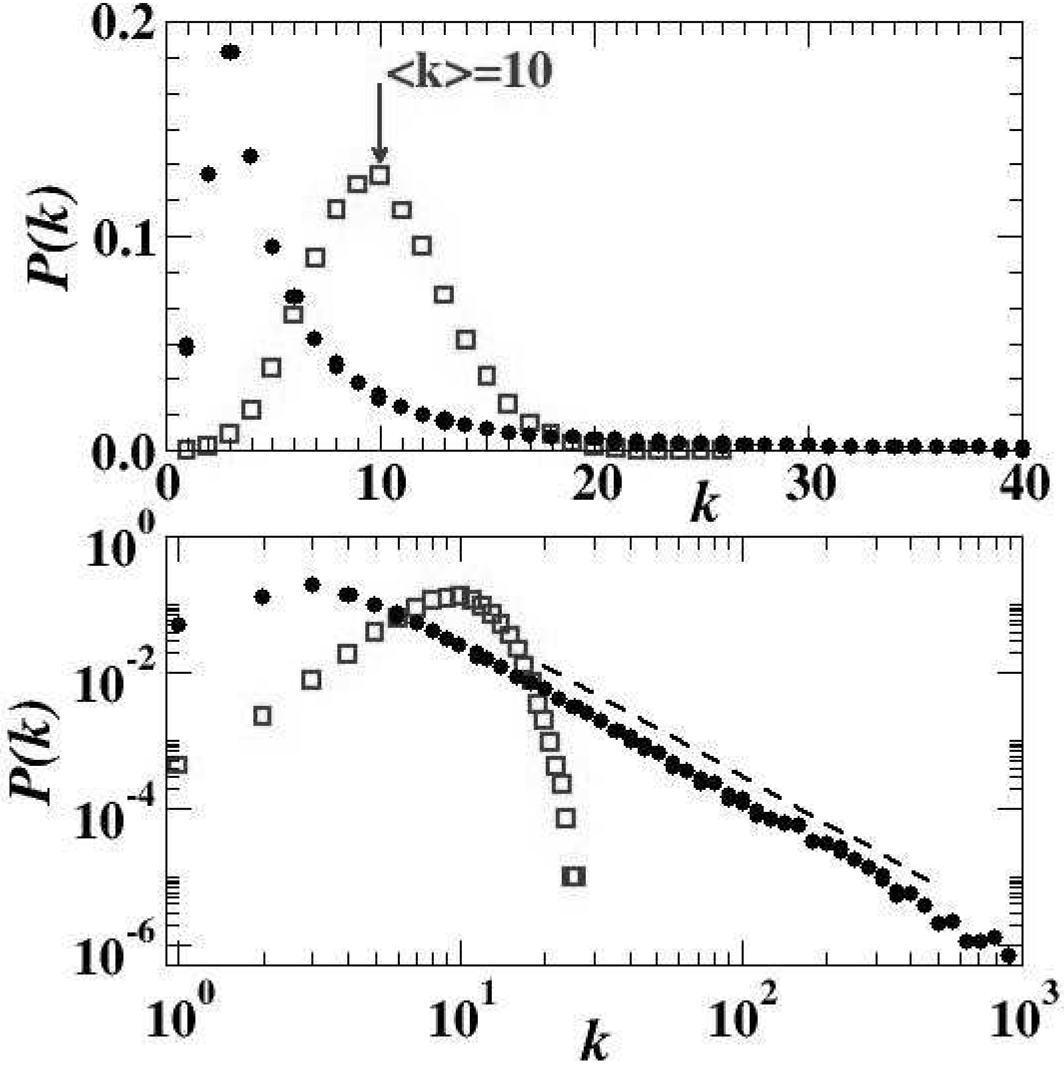
Broad degree distributions

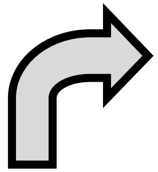
(often: power-law tails $P(k) \sim k^{-\gamma}$, typically $2 < \gamma < 3$)

No particular *characteristic* scale

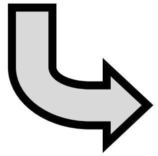
Topological heterogeneity

Statistical analysis of centrality measures:



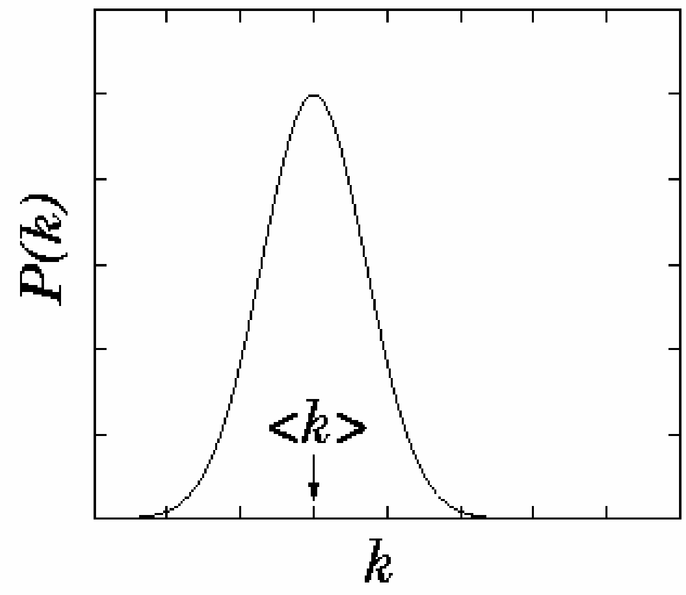


Poisson
vs.
Power-law

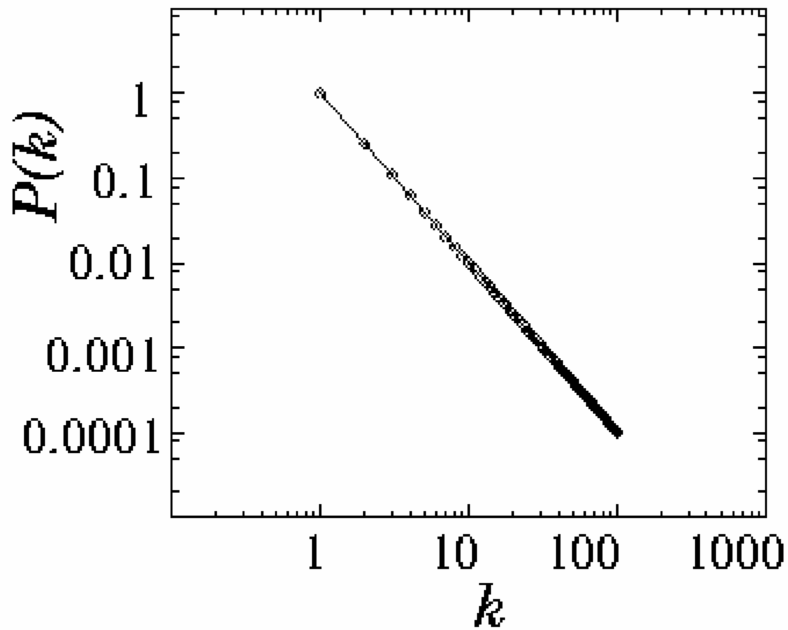


Exp. vs. Scale-Free

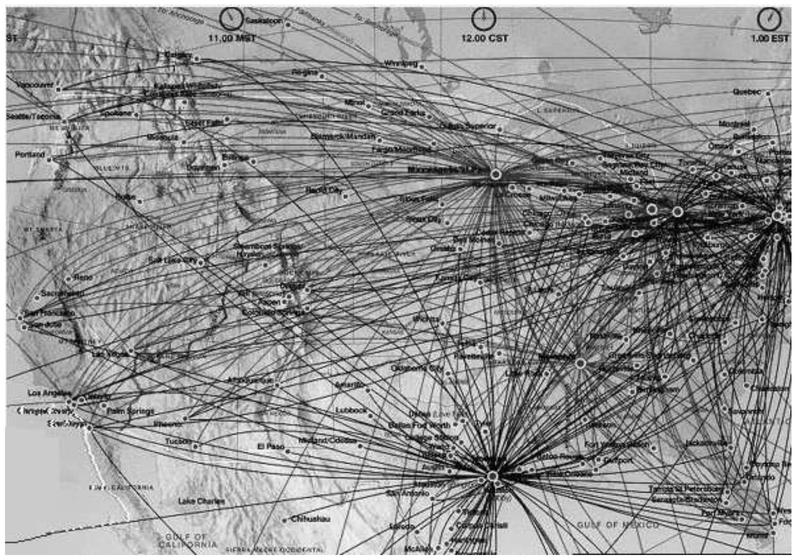
Poisson distribution



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

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

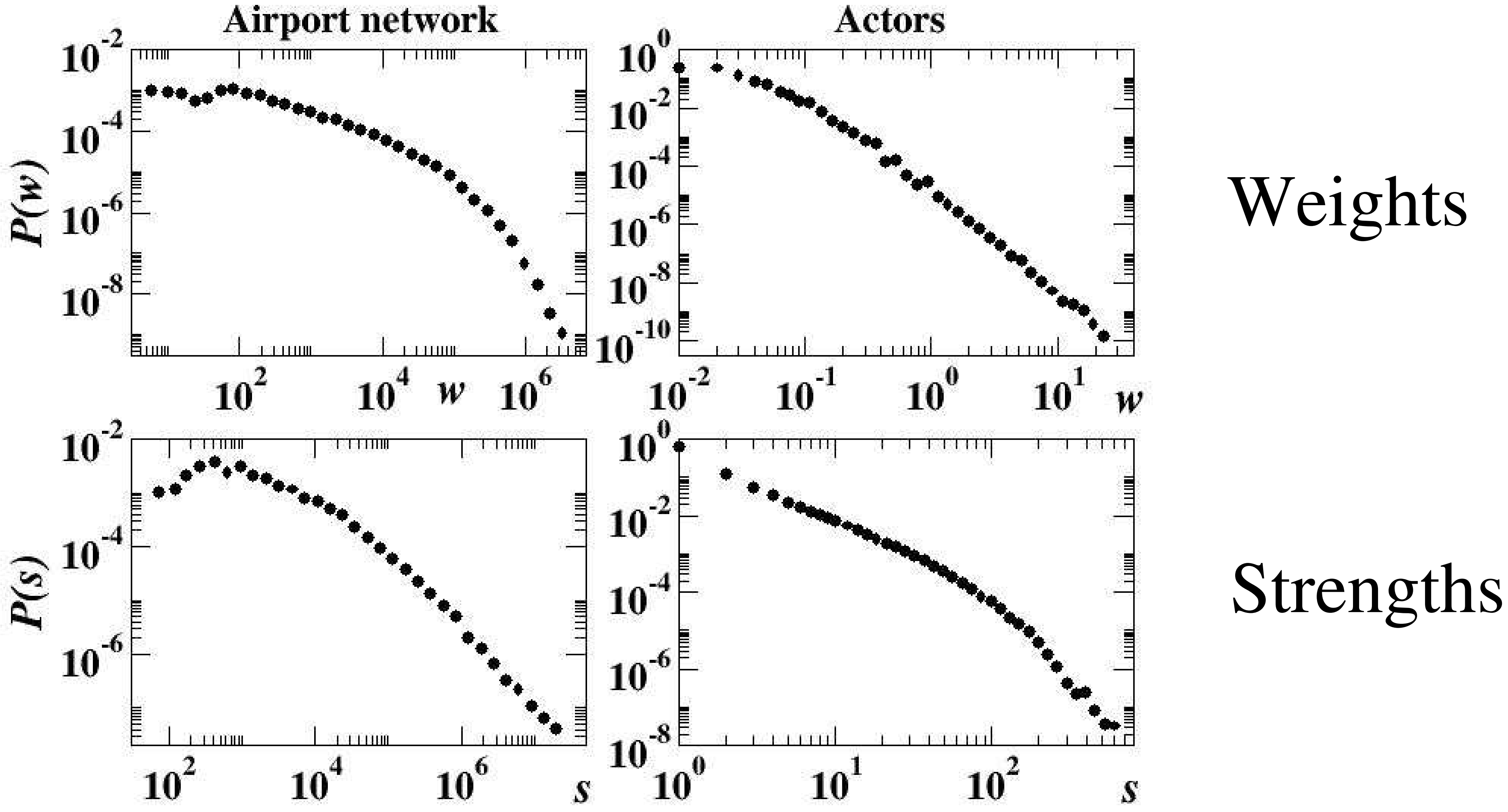
k_c=cut-off due to finite-size

$N \to \infty$ => diverging degree fluctuations

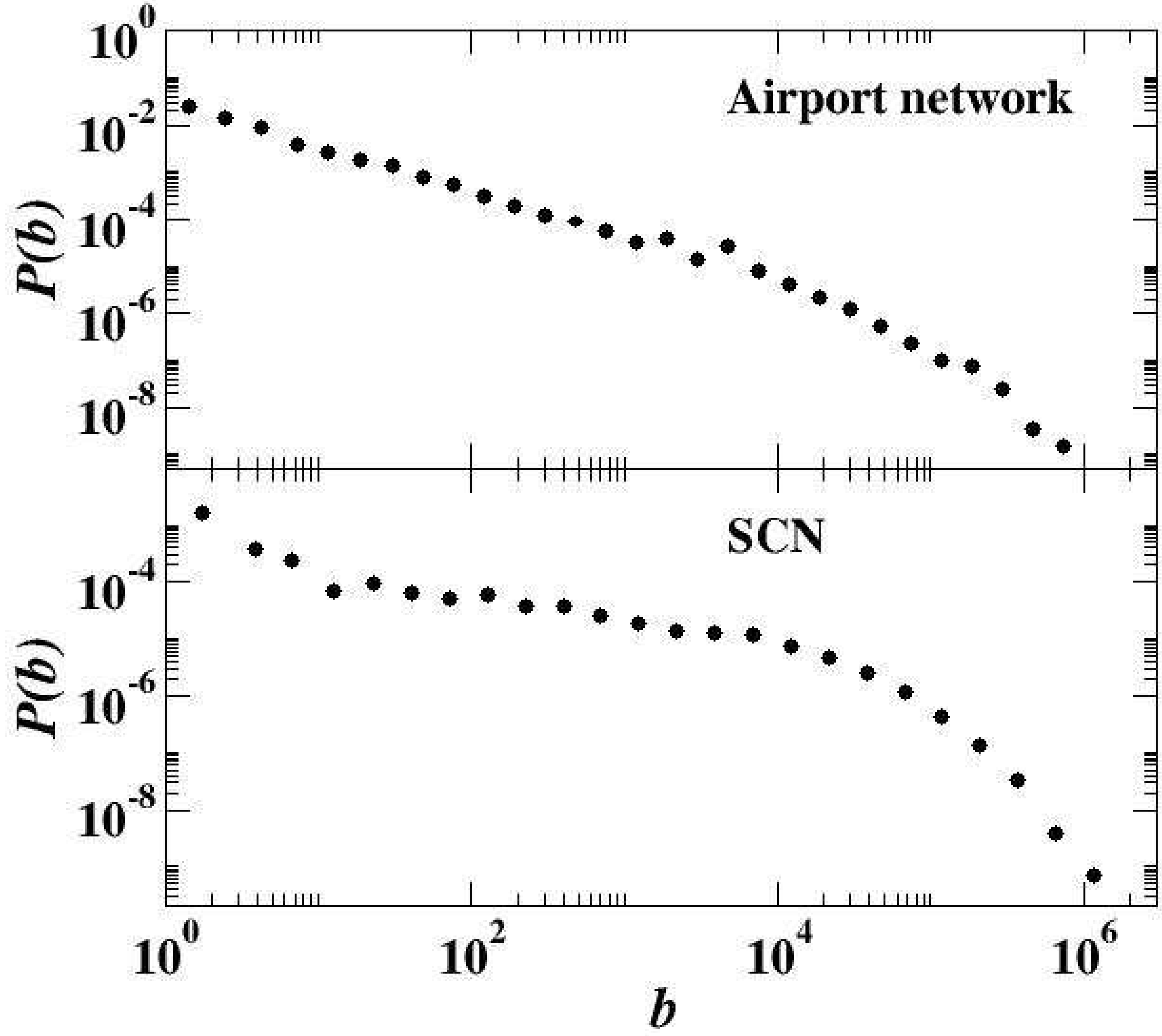
for $\gamma < 3$

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

Other heterogeneity levels

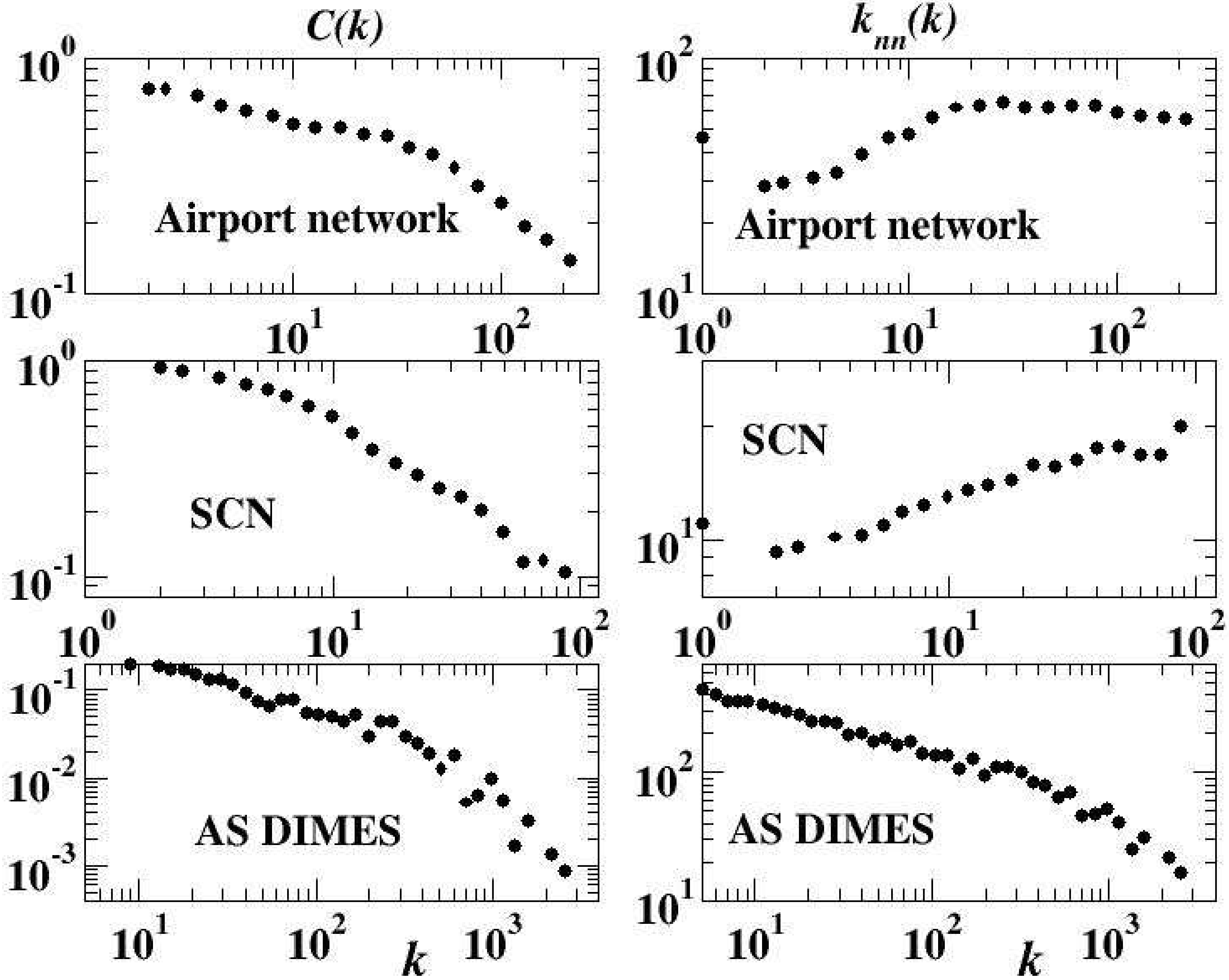


Other heterogeneity levels



Betweenness centrality

Clustering and correlations



non-trivial structures

Complex networks

Complex is not just “complicated”

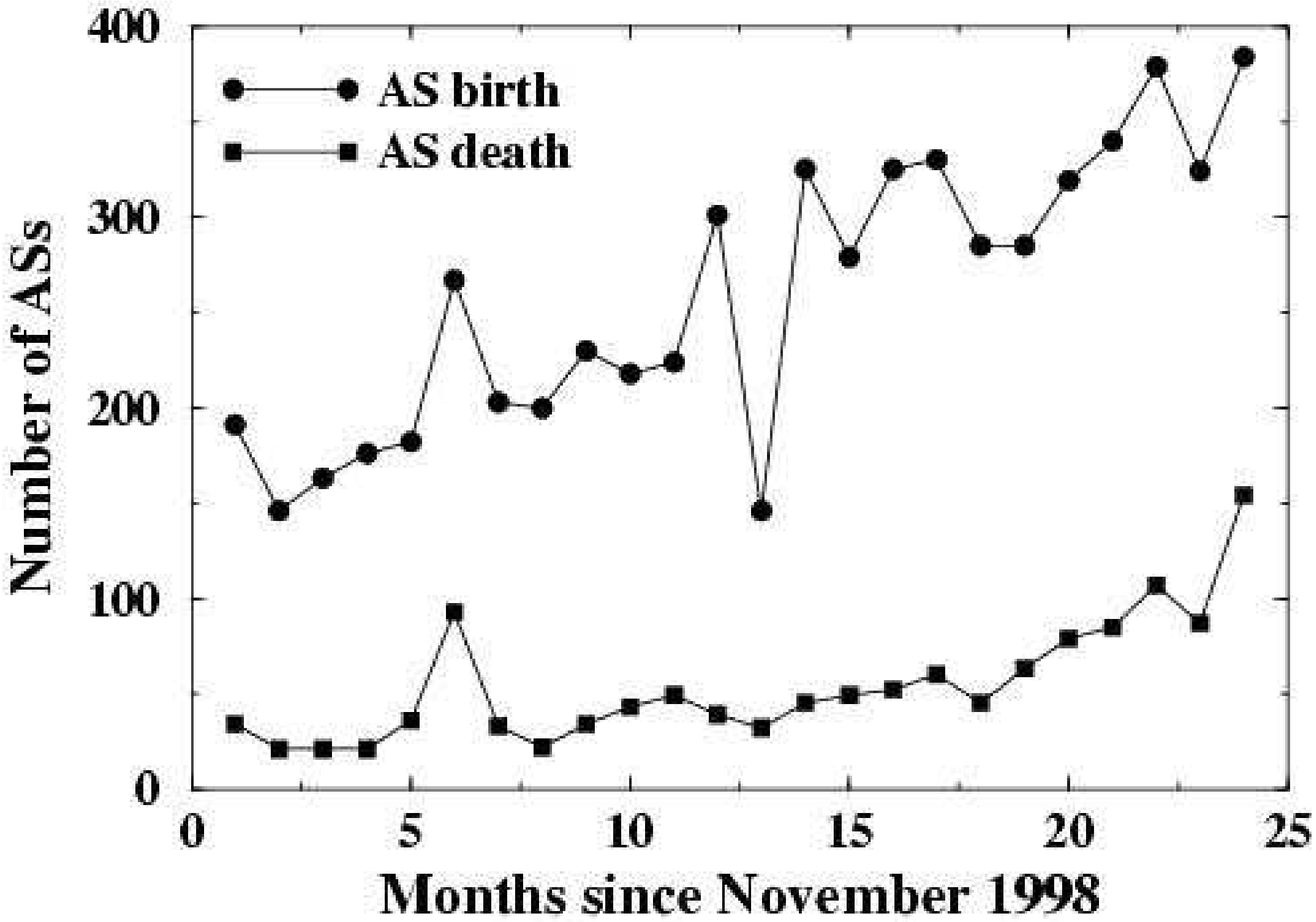
Cars, airplanes…=> complicated, not complex

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…

Example: Internet growth



Main features of complex networks

- **Many interacting units**
- **Self-organization**
- **Small-world**
- **Scale-free heterogeneity**
- **Dynamical evolution**

Standard graph theory

Random graphs

- **Static**
- **Ad-hoc topology**

Example: Internet topology generators
Modeling of the Internet structure with ad-hoc algorithms tailored on the properties we consider more relevant

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**

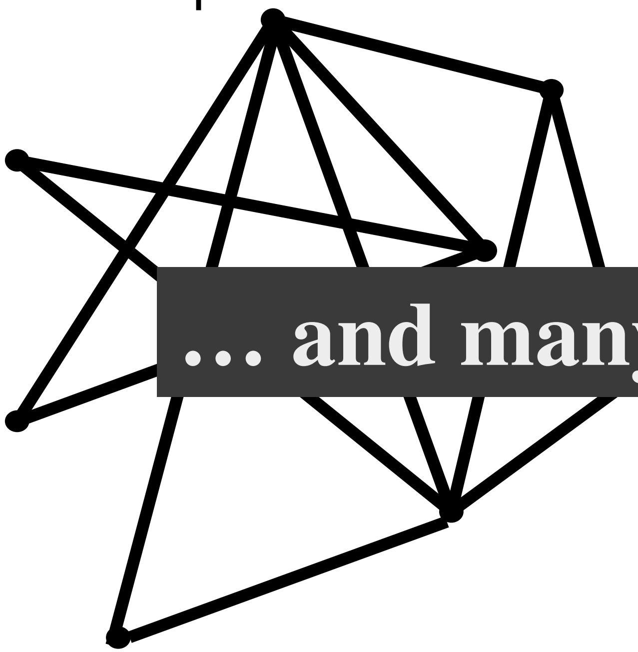
Development of new modeling frameworks

New modeling frameworks
Example: preferential attachment

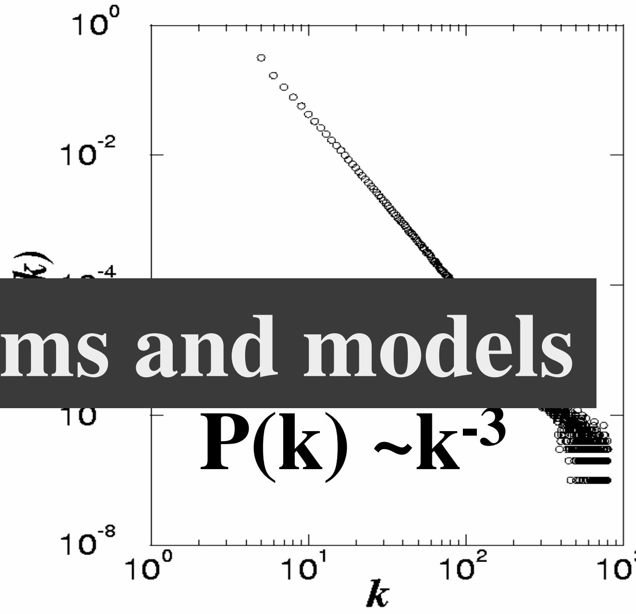
(1) GROWTH : At every timestep we add a new node with *m* edges (connected to the nodes already present in the system).

(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

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



… and many other mechanisms and models



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