Models

The role of models

“All models are wrong, but some are useful”

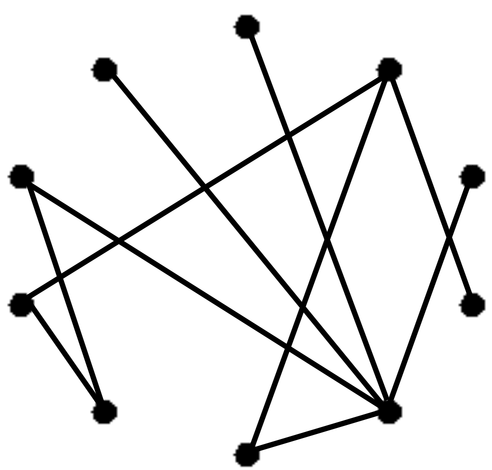
(George E. P. Box)

The role of models

- Generative
- Explanatory
- Null models

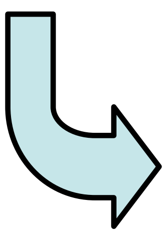
Erdös-Renyi random graph model (1960)

N points, links with proba p:
static random graphs



Average number of edges:
$< E > = pN(N-1)/2$

Average degree:
$< k > = p(N-1)$



$p = < k >/N$ to have finite average degree as N grows

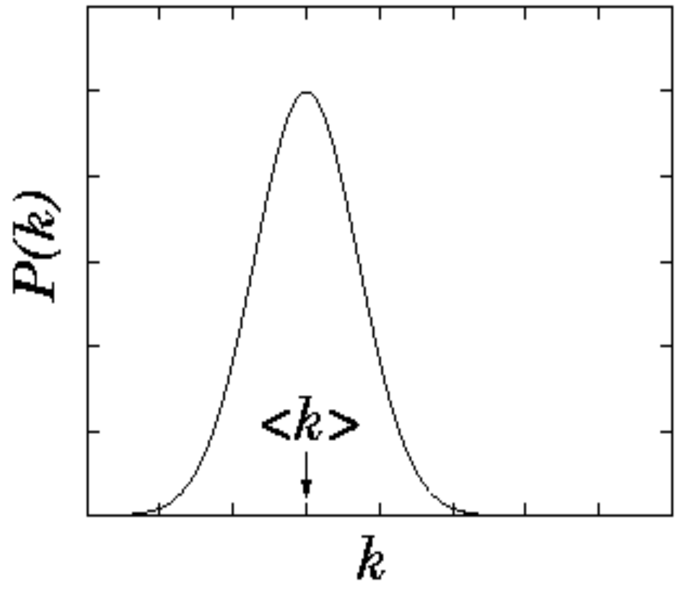
Erdös-Renyi model (1960)

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

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



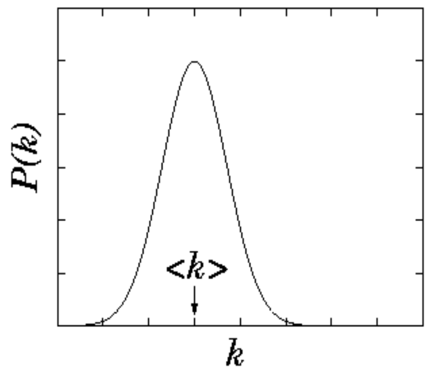
Exponential decay at large k

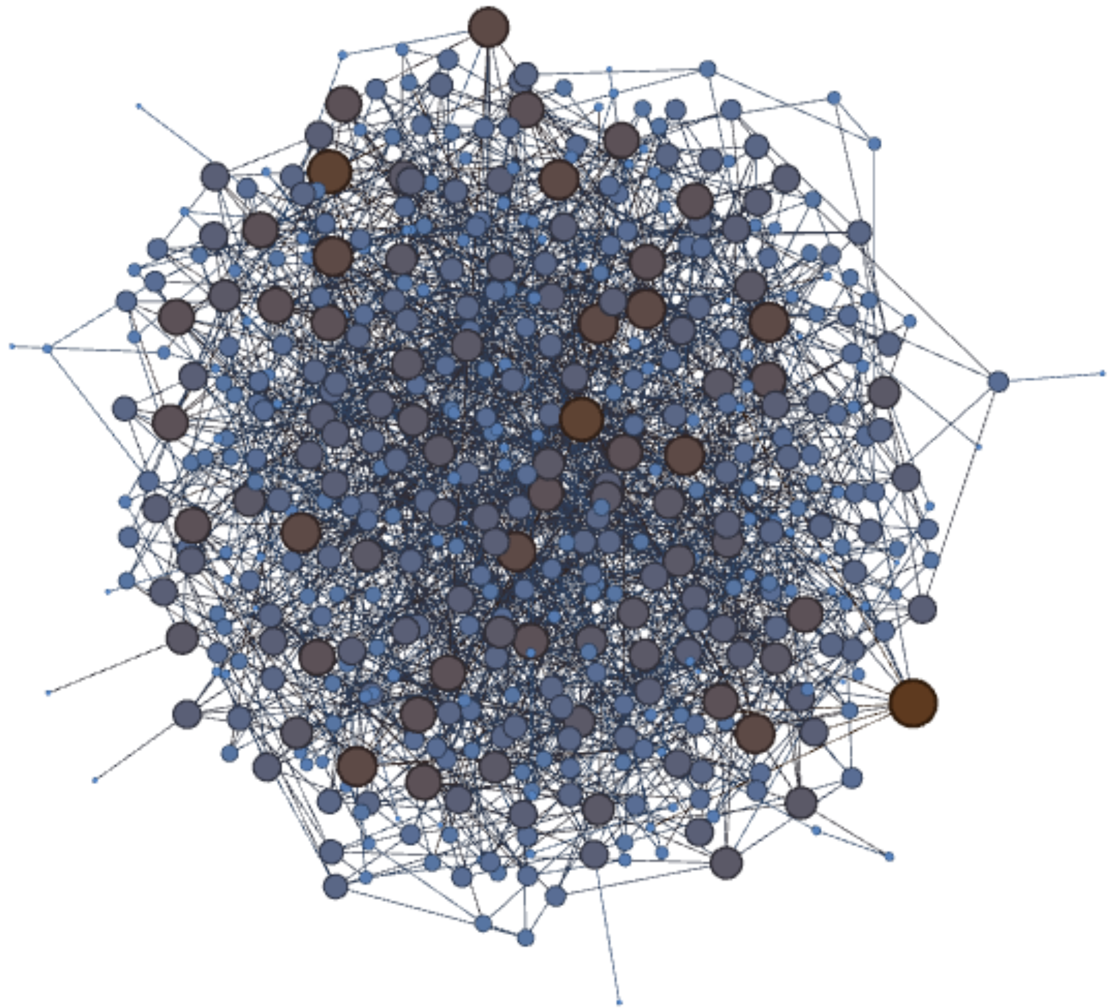
Erdös-Renyi model (1960)

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

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

Poisson degree distribution



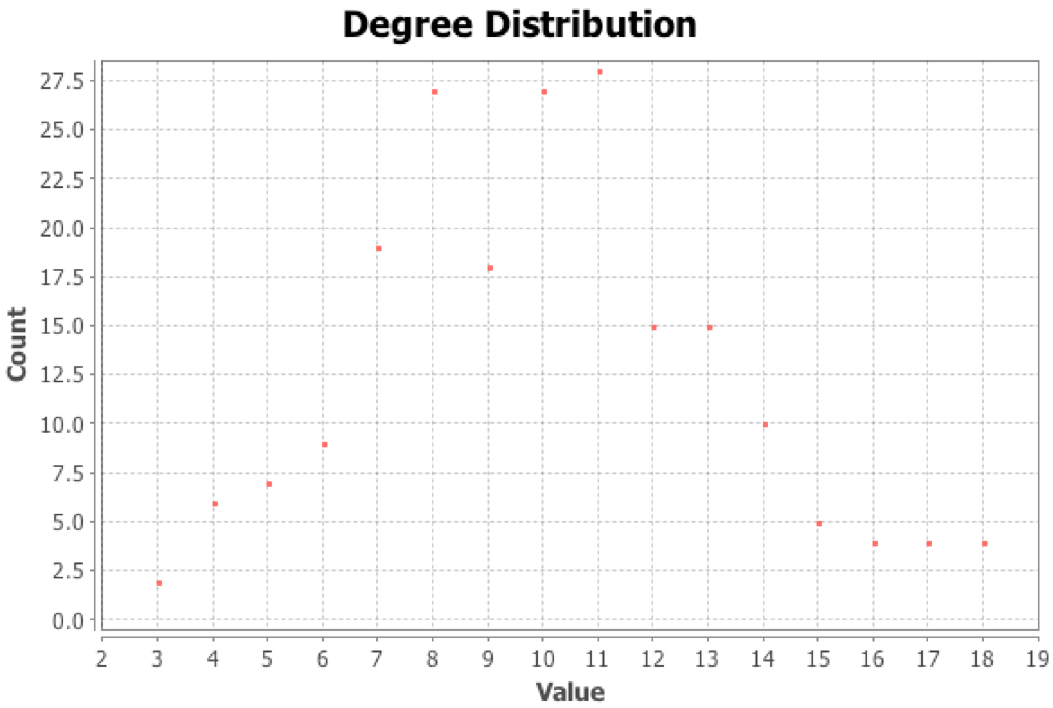


Gephi
makes graphs handy

Degree Report

Results:

Average Degree: 10.010



ER model,
N=200
p=0.05

Gephi
makes graphs handy

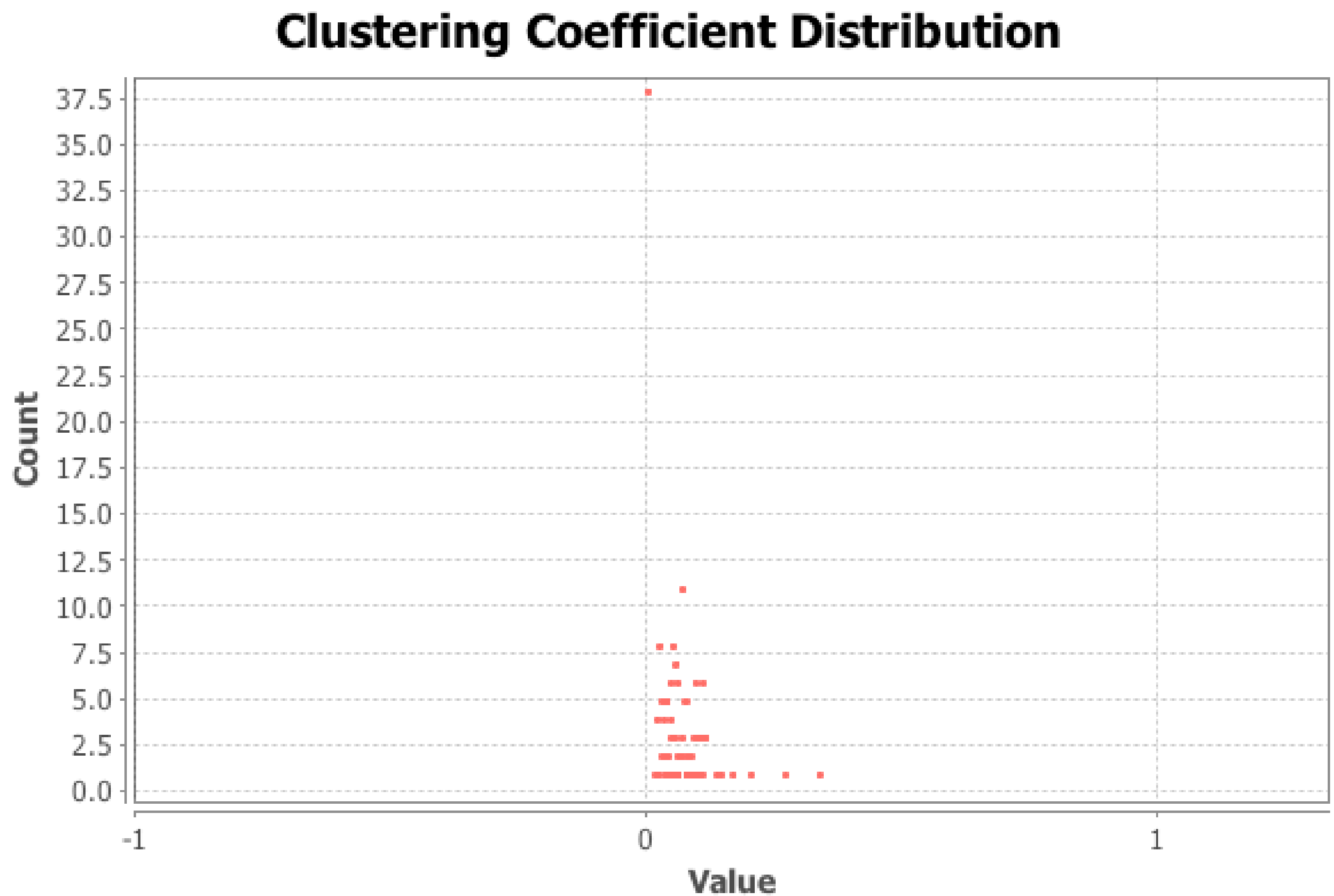
Clustering Coefficient Metric Report

Parameters:

Network Interpretation: undirected

Results:

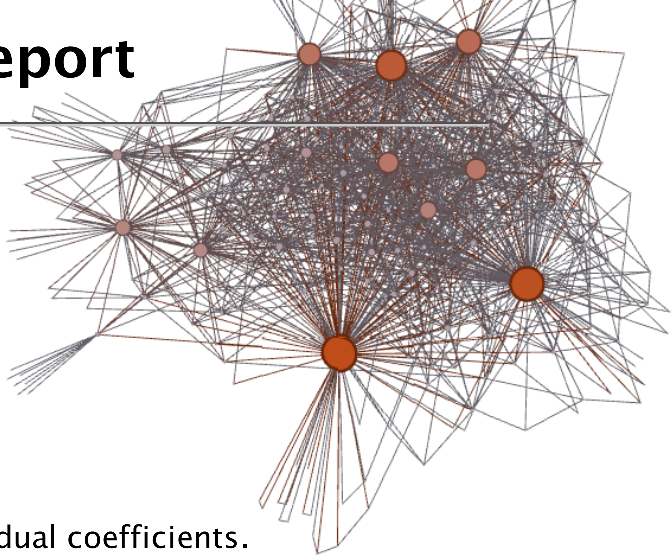
Average Clustering Coefficient: 0.052
Total triangles: 182
The Average Clustering Coefficient is the mean value of individual coefficients.



ER model,
N=200
p=0.05

Airlines,
N=235
<k>=11

Clustering Coefficient Metric Report

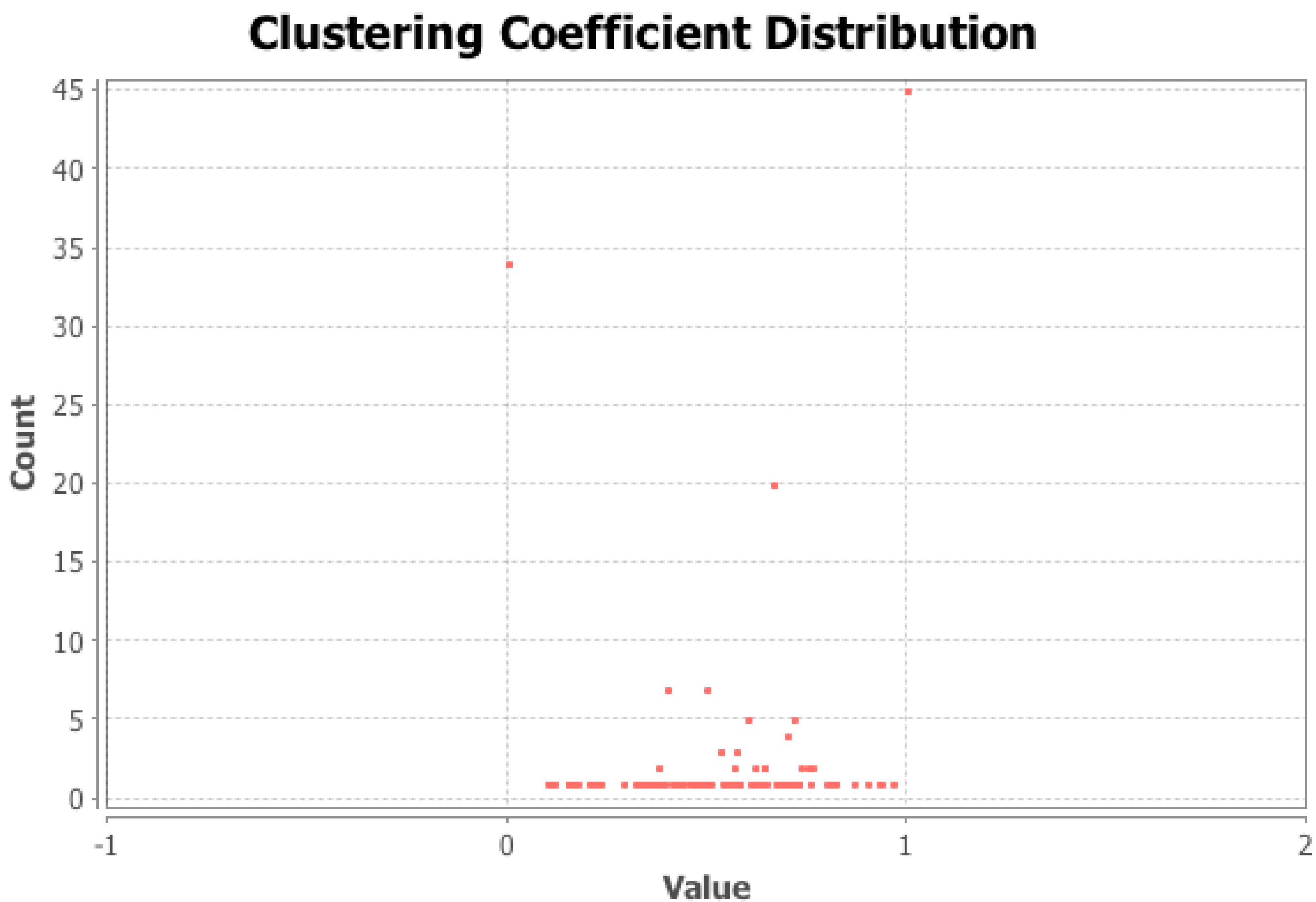


Parameters:

Network Interpretation: undirected

Results:

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



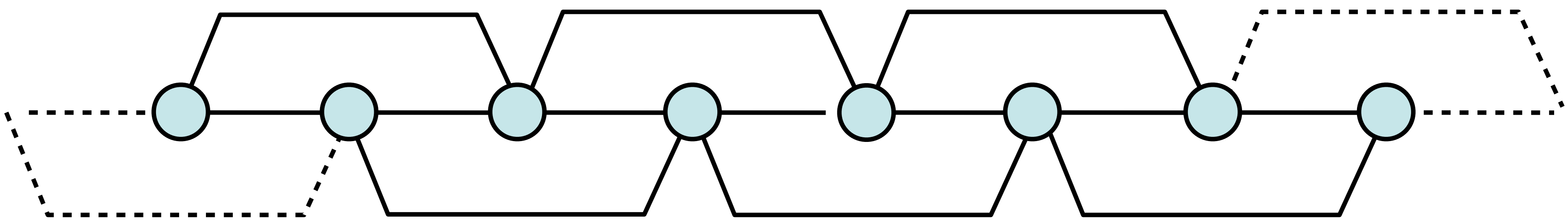
Clustering coefficient **much larger** than for an ER with same numbers of nodes and links

Watts-Strogatz model

Motivation:

-random graph: short distances but no clustering

-regular structure: large clustering but large distances

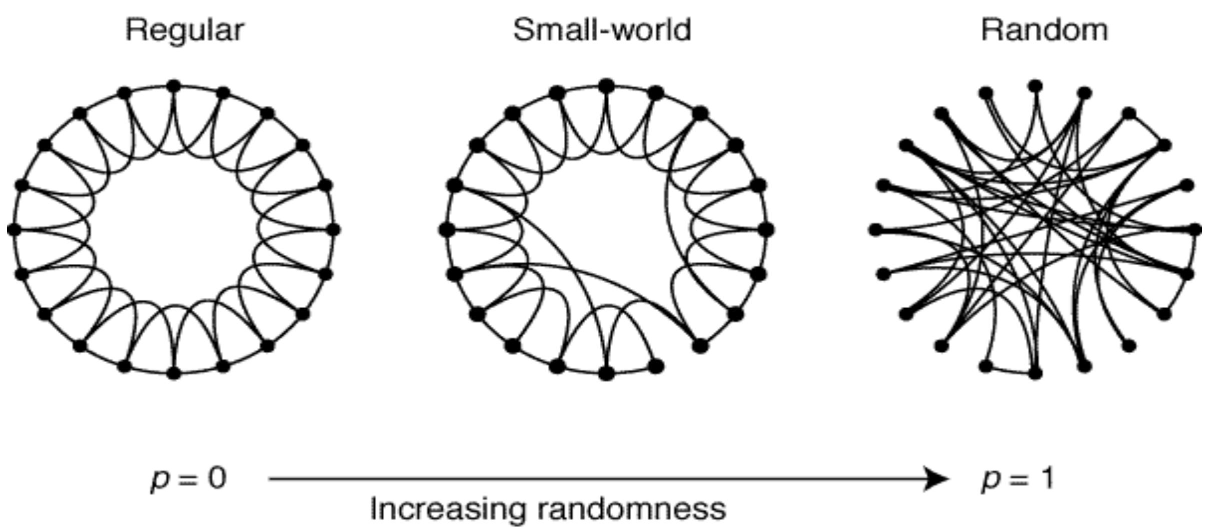


=> how to have both small distances and large clustering?

Watts & Strogatz,

Nature **393**, 440 (1998)

Watts-Strogatz model



1) N nodes arranged in a line/circle

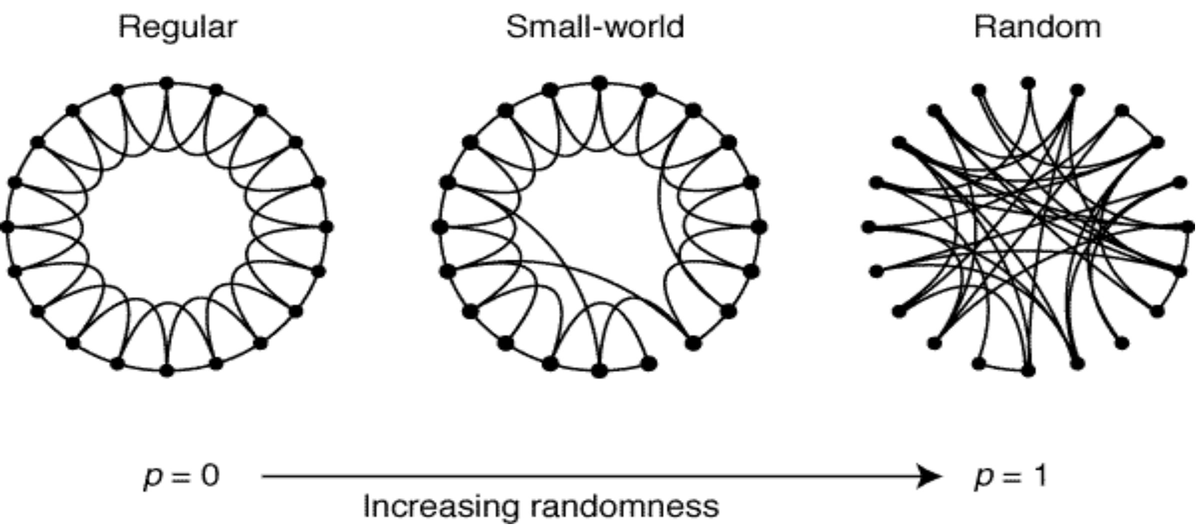
2) Each node is linked to its 2k neighbors on the circle, k clockwise, k anticlockwise

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

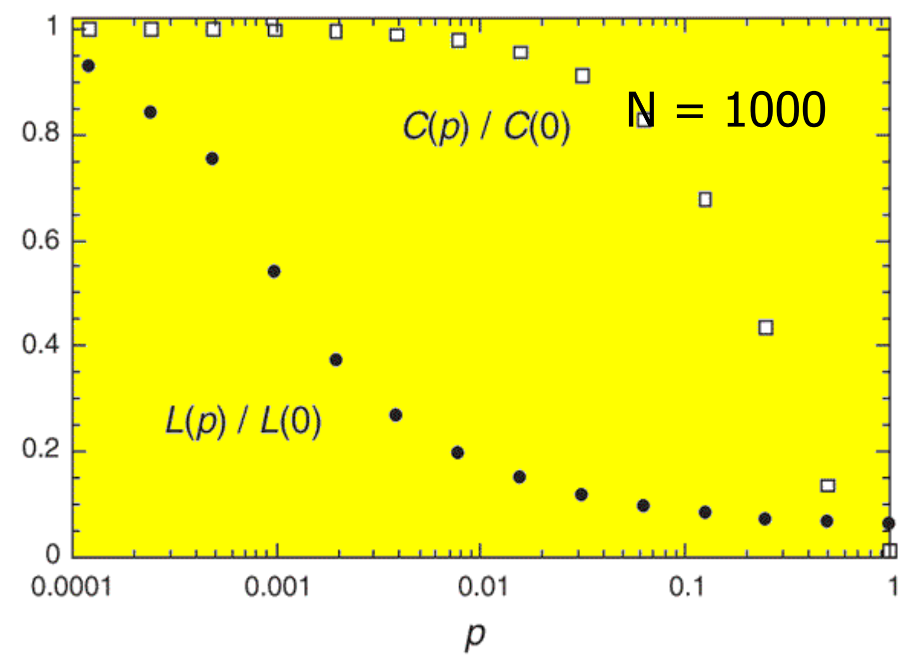


N nodes forms a regular lattice. With probability p, each edge is rewired randomly

=>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)

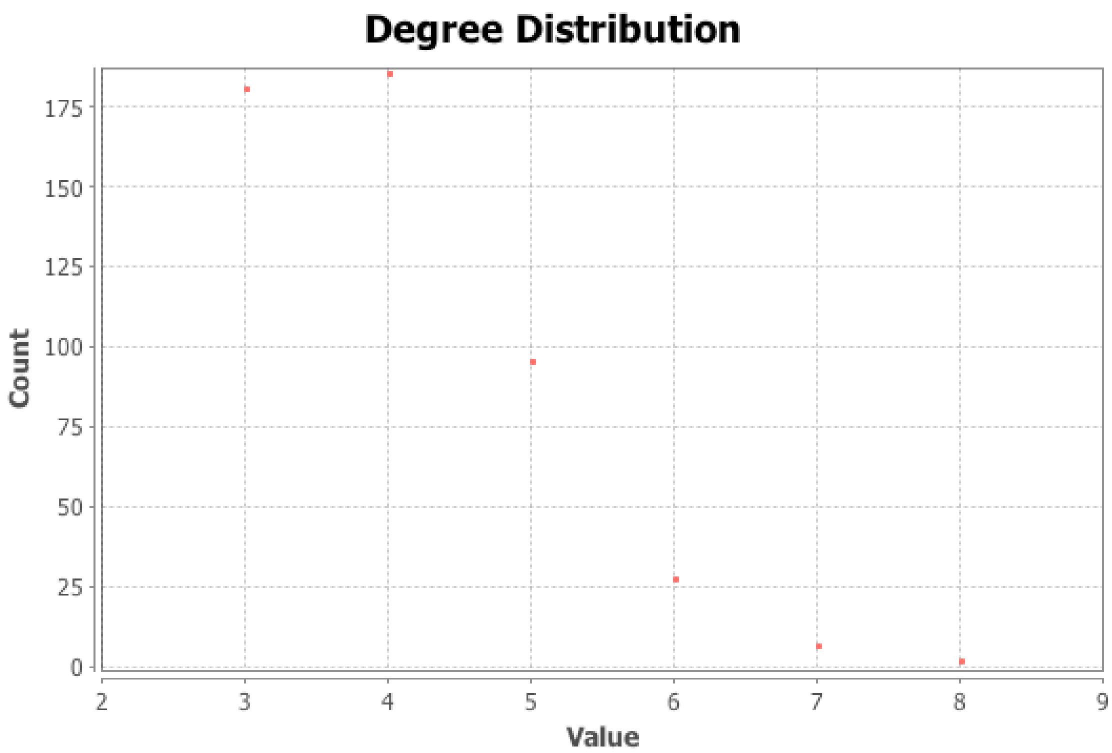
BUT: still homogeneous degree distribution

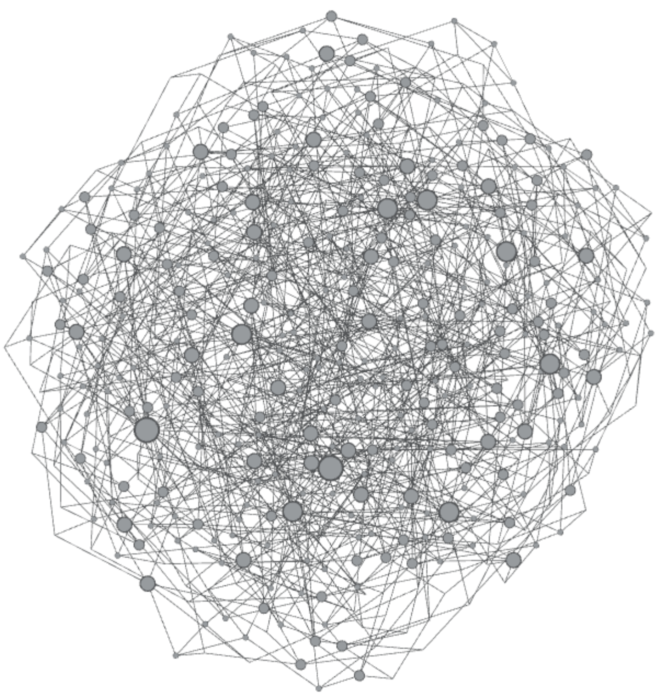
Degree Report

Watts-Strogatz network

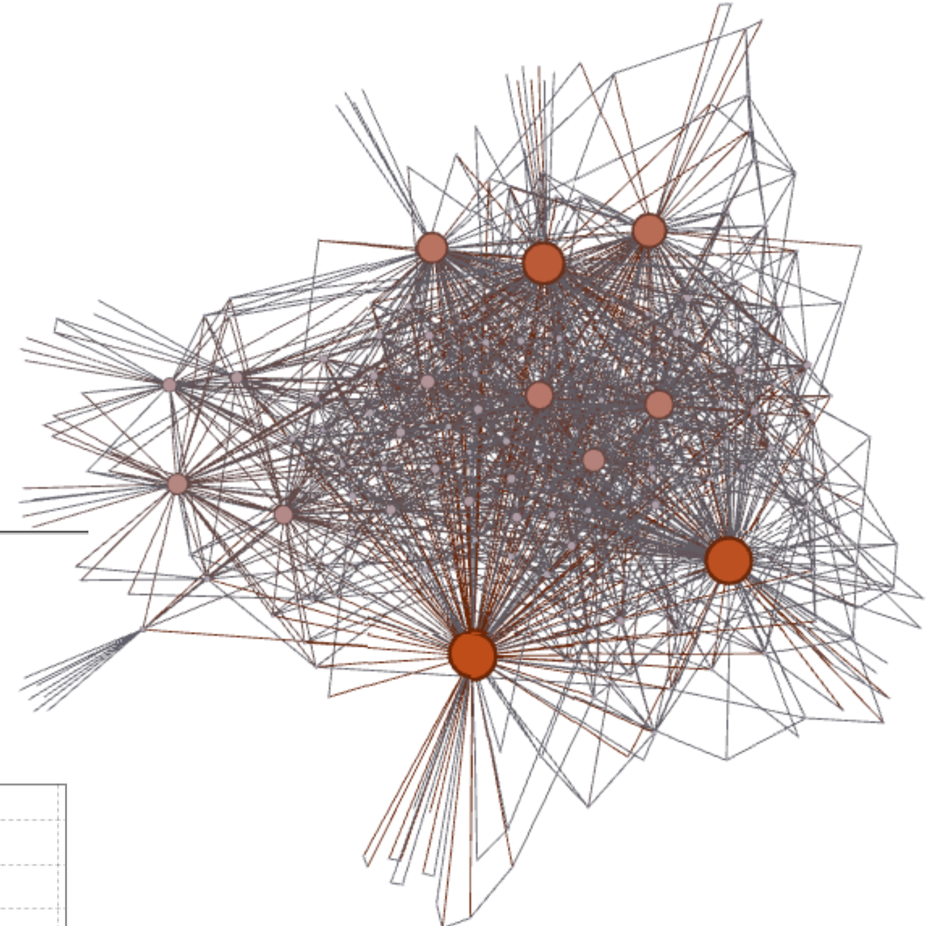
Results:

Average Degree: 4.000





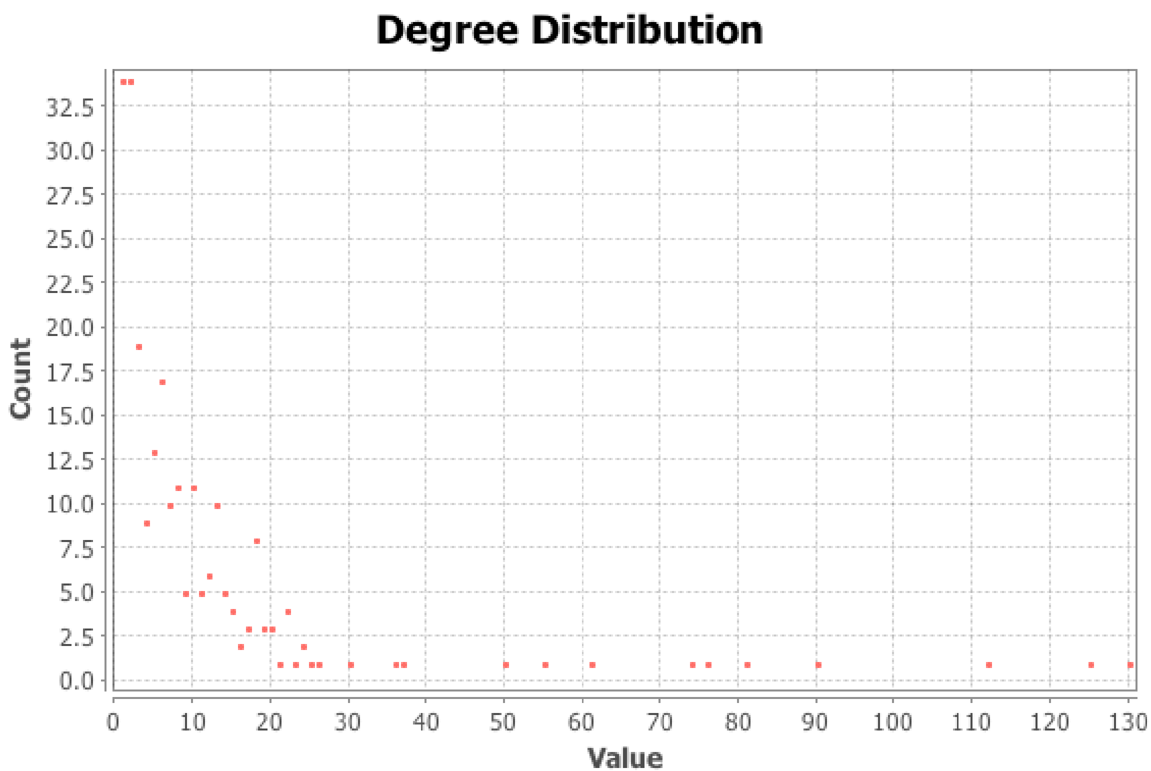
Airlines



Degree Report

Results:

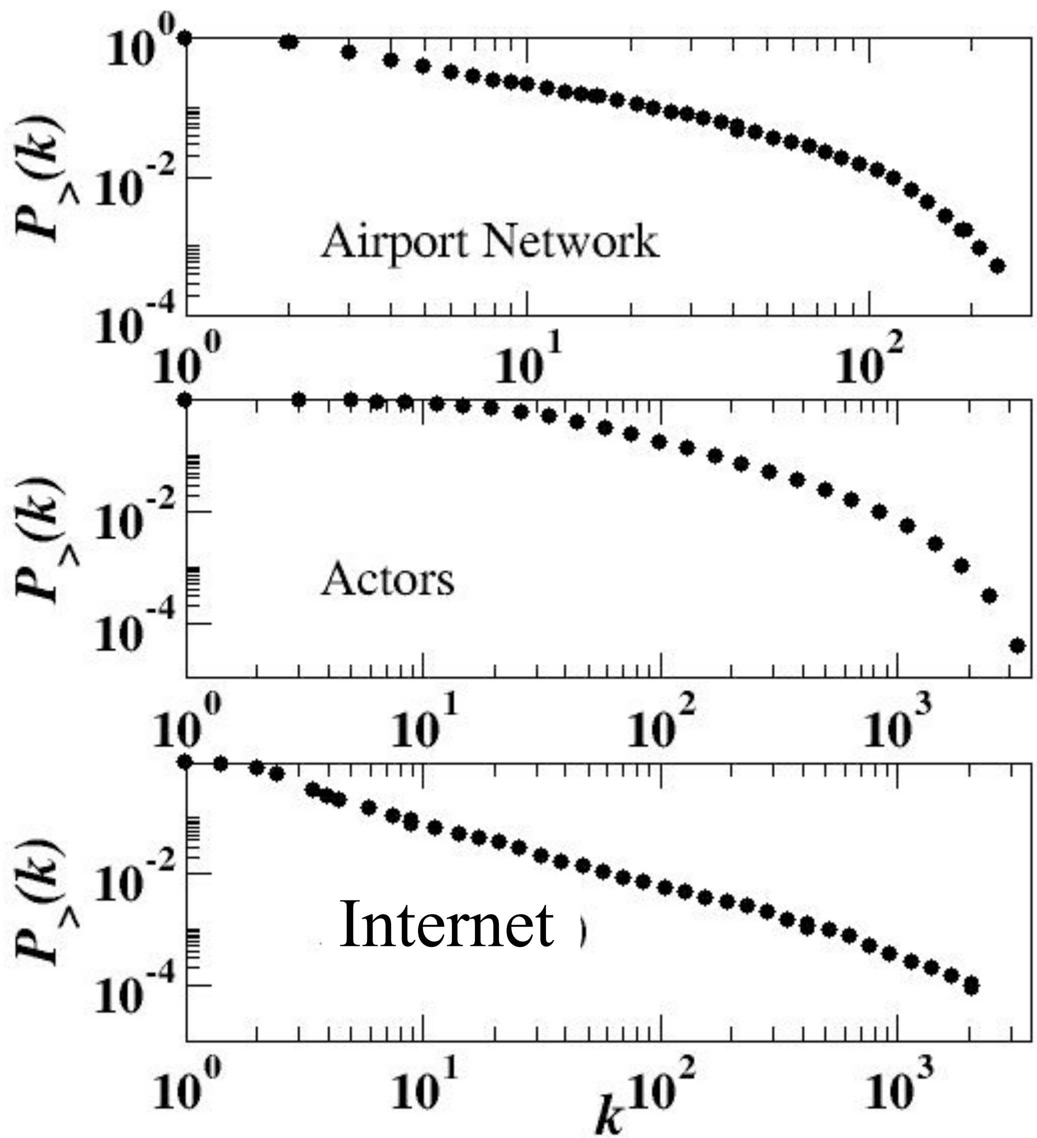
Average Degree: 11.038



Gephi makes graphs handy

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)

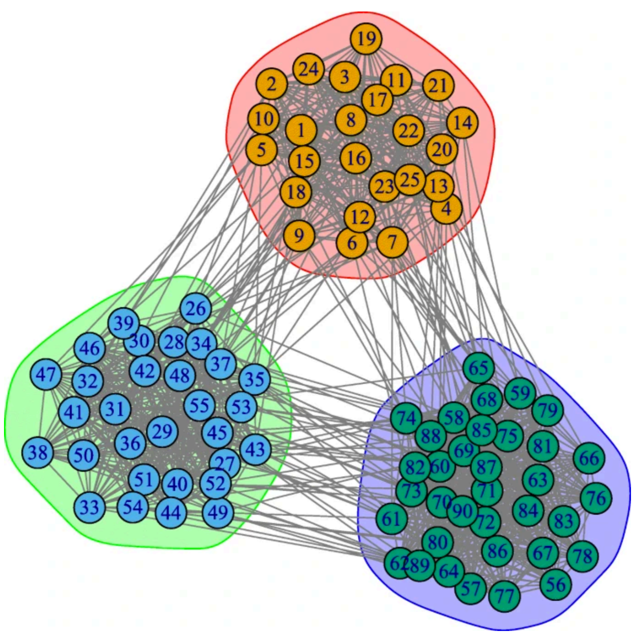
- Extract a sequence k_i of degrees taken from P(k)
- Assign them to the nodes i=1,…,N
- Connect randomly the nodes together, according to their given degree

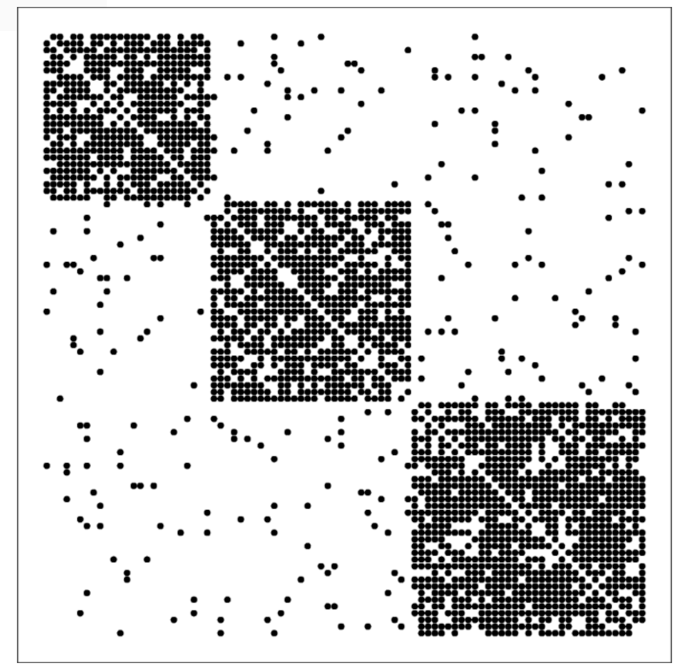
Review | Open Access | Published: 23 December 2019

A review of stochastic block models and extensions for graph clustering

Clement Lee & Darren J. Wilkinson

Applied Network Science **4**, Article number: 122 (2019) | Cite this article





$$i \in a, j \in b \Rightarrow \text{Proba(link i-j)} = p_{ab}$$

(also: degree-corrected block 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.

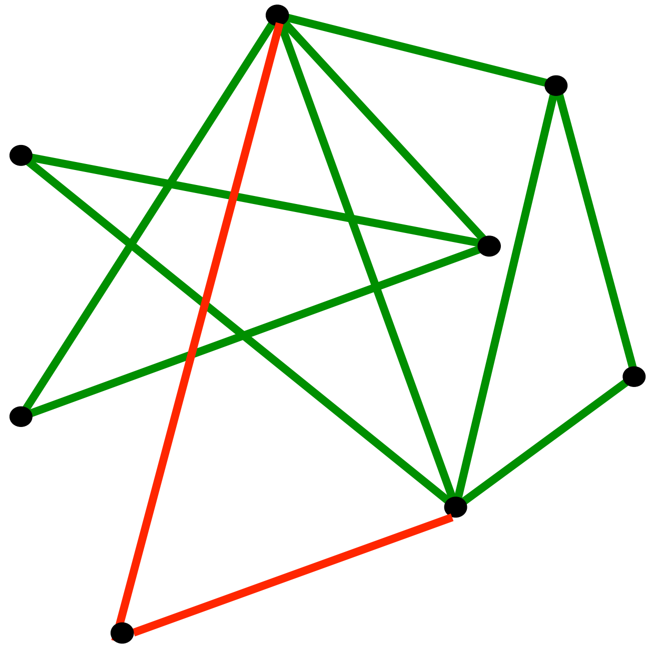
Examples :
WWW : new documents link to well known sites (CNN, YAHOO, NewYork Times, etc)
Citation : well cited papers are more likely to be cited again

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)

Microscopic mechanism:
An example

Continuous time and degree approximation

$$\Pi(k_i) = \frac{k_i}{\Sigma_j k_j} \Rightarrow \frac{dk_i(t)}{dt} = m\frac{k_i(t)}{\sum_{j=1}^{t} k_j(t)}$$

$$\sum_{j=1}^{t} k_j = 2mt \Rightarrow k_i(t) = m\sqrt{\frac{t}{i}}$$

Microscopic mechanism:
An example

$$P(k,t)dk = P(i,t)di = \frac{di}{t}$$

(proba to choose i at random among the t nodes)

$$k_i(t) = m\sqrt{\frac{t}{i}}$$

(nodes i=1,…,t)

$$i = m^2 t/k^2 \Rightarrow di = \frac{2m^2 t}{k^3} dk$$

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

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?
- unrealistic 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)

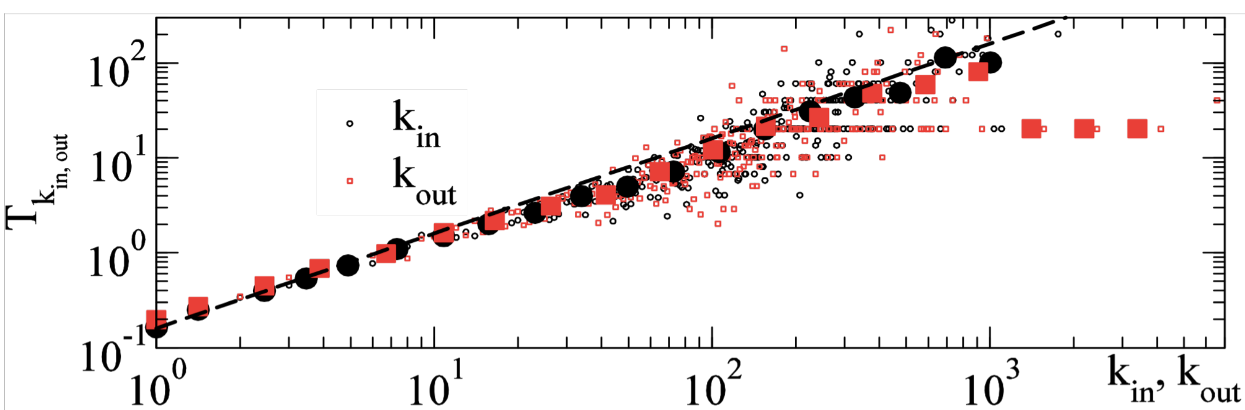
How to measure the preferential attachment

Hence:

T_k= number 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?

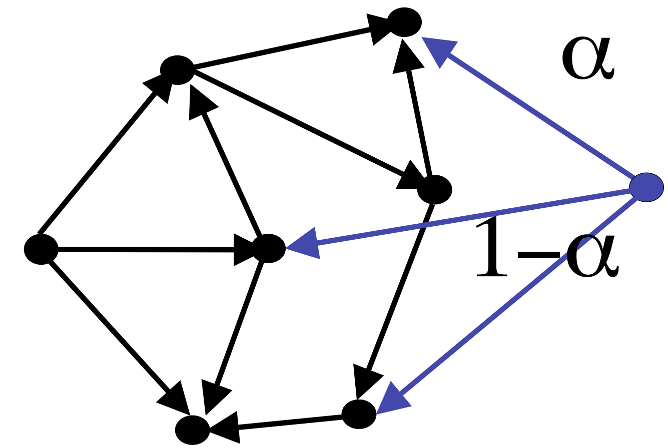
Another mechanism:
copying model

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

Continuous time and degree approximation

$$\frac{dk_i(t)}{dt} = m\frac{1-\alpha}{t} + m\alpha\frac{k_i(t)}{\sum_{j=1}^{t} k_j(t)}$$

(here we write k for k_{in})

$$\sum_{j=1}^{t} k_j = mt$$

$$k_i(t) = \frac{m}{\alpha}\left(\frac{t}{i}\right)^{\alpha} - k_0\,,\, k_0 = m(1-\alpha)/\alpha$$

Copying model

$$k_i(t) = \frac{m}{\alpha}\left(\frac{t}{i}\right)^{\alpha} - k_0$$

$$P(k,t)dk = P(i,t)di = \frac{di}{t}$$

$$di = \frac{t}{\alpha}\left(\frac{m}{\alpha}\right)^{1/\alpha}(k+k_0)^{-1-\frac{1}{\alpha}}dk$$

Power-law tail of degree distribution:

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

(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

  -hierarchical structures

  ...

- *Many other proposed mechanisms* in the literature => modeling other attributes: weights, clustering, assortativity, spatial effects…

- *Model validation*: degree distribution, degree correlations, clustering properties, hierarchies, …

- *Level of detail and type of model: 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/surrogate 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

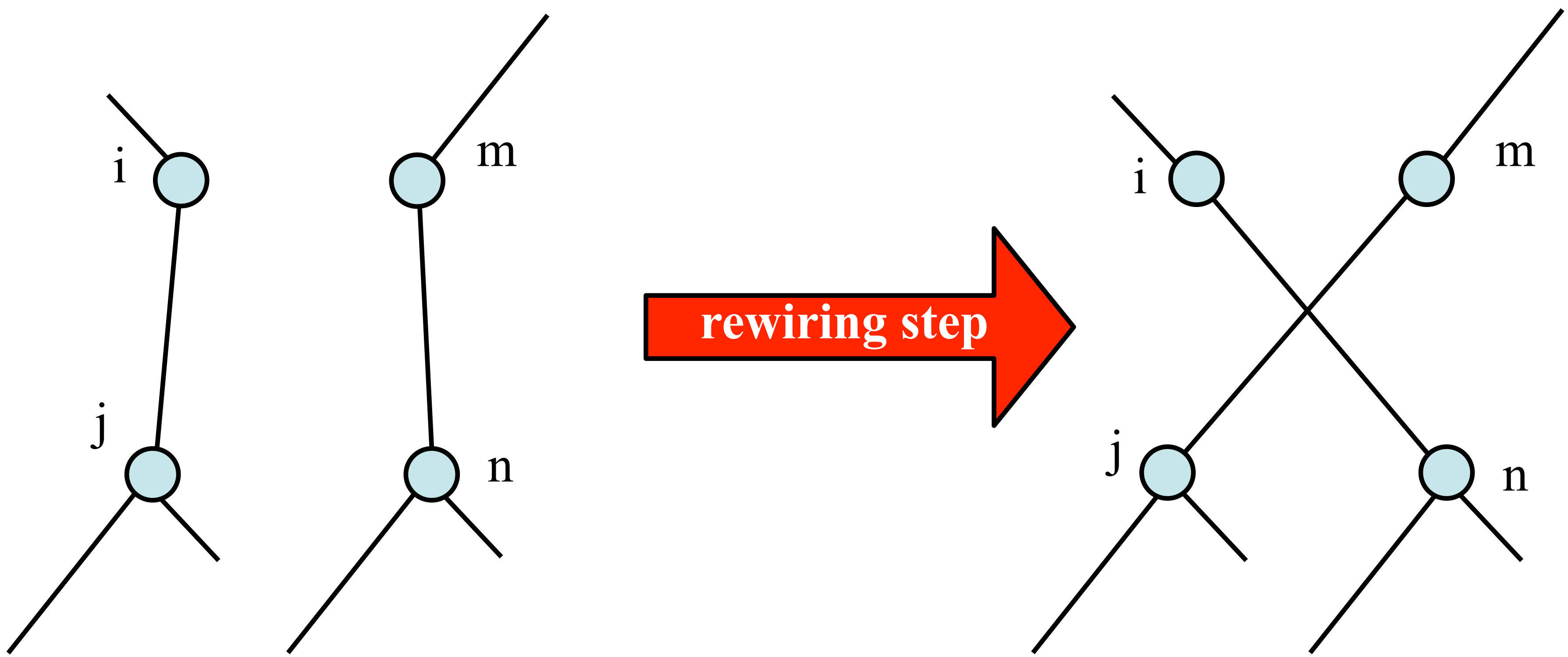
Aim:

• 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



Original network

Rewired network

Science

Current Issue First release papers Archi

HOME > SCIENCE > VOL. 296, NO. 5569 > SPECIFICITY AND STABILITY IN TOPOLOGY OF PROTEIN NETWORKS

REPORTS

Specificity and Stability in Topology of Protein Networks

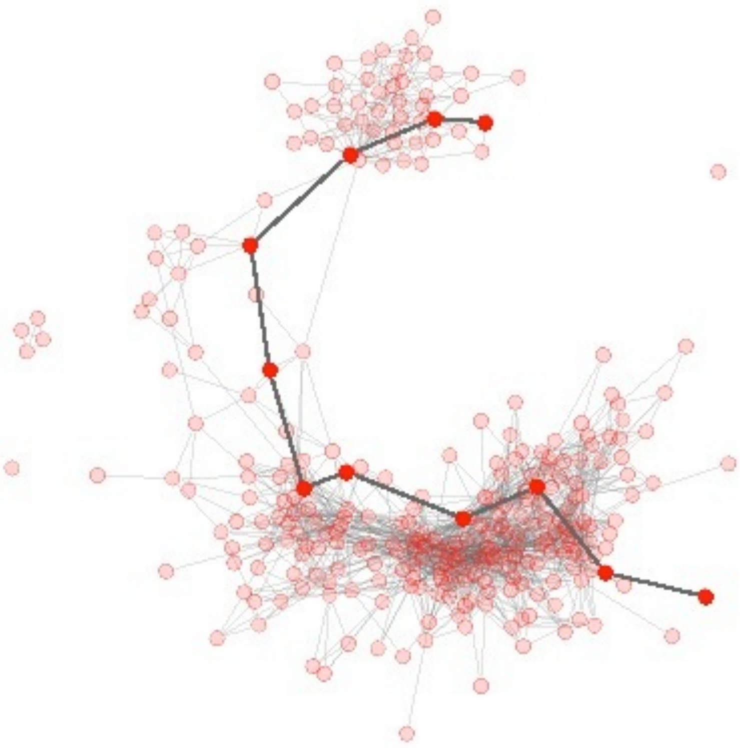
SERGEI MASLOV AND KIM SNEPPEN Authors Info & Affiliations

SCIENCE • 3 May 2002 • Vol 296, Issue 5569 • pp. 910-913 • DOI: 10.1126/science.1065103

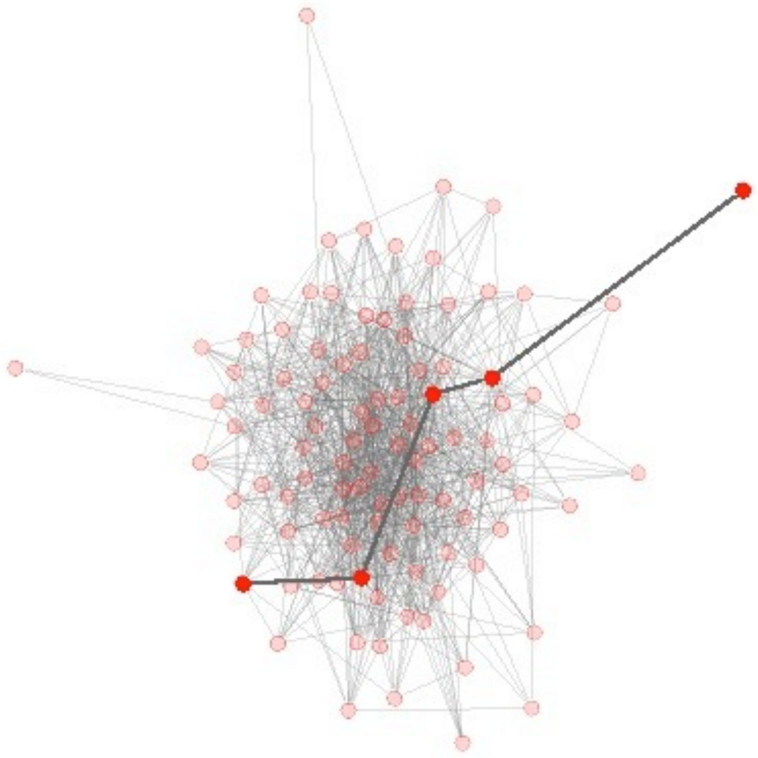
- preserves the degree of each node
- destroys topological correlations

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

Museum (SG)

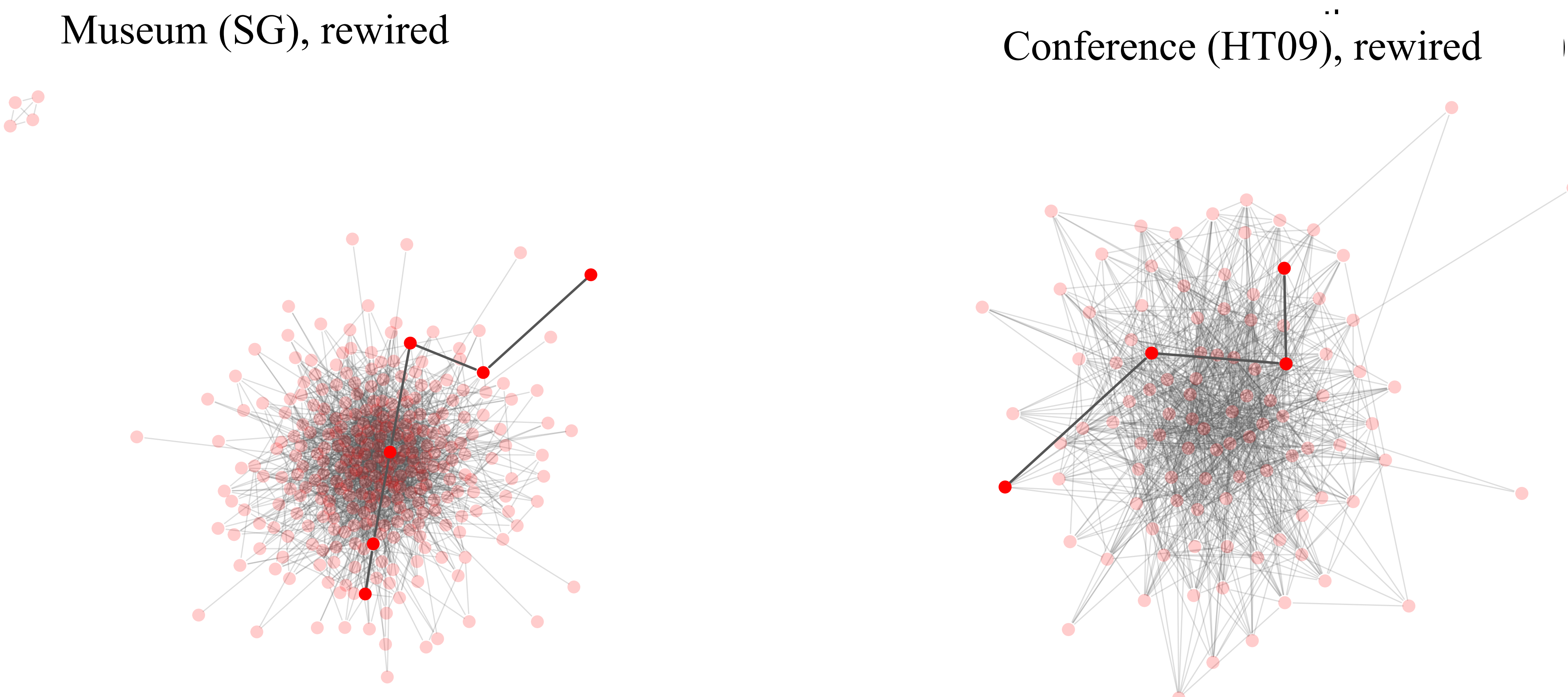


Conference (HT09)

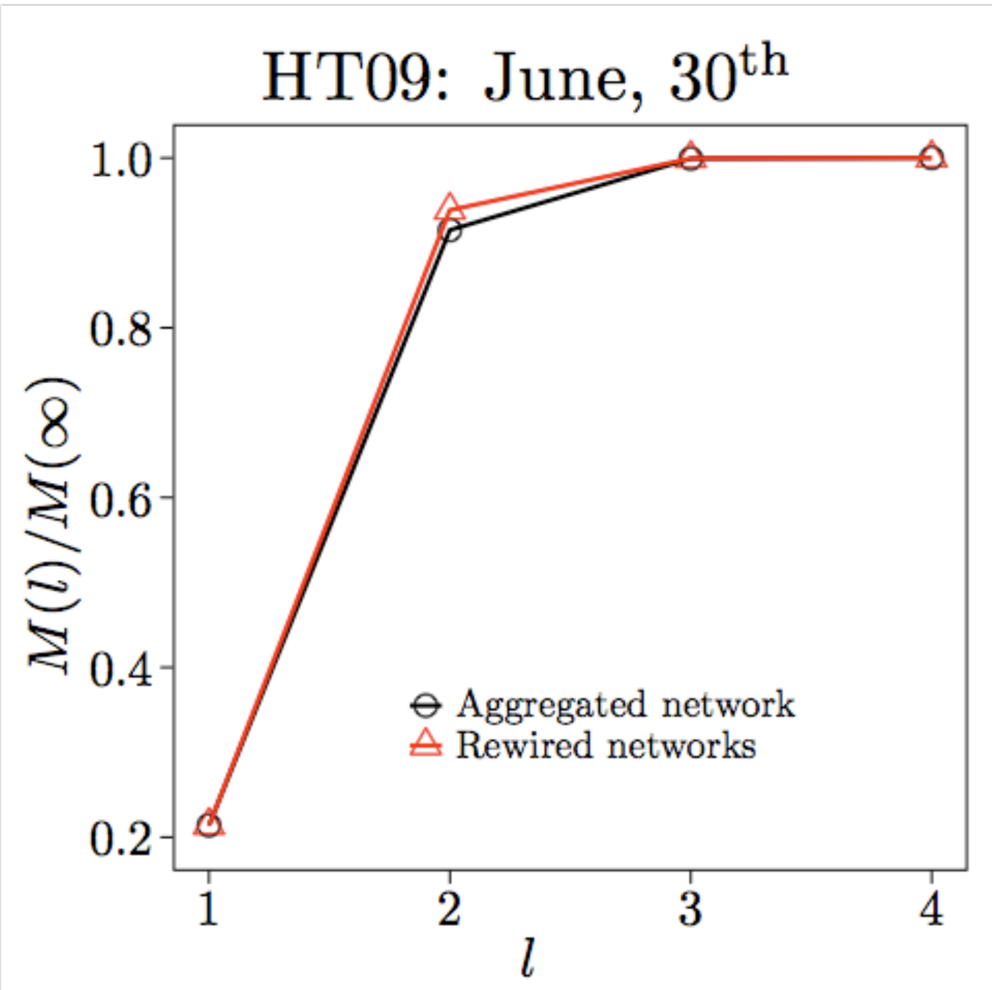


"seems" not to be
a small-world network

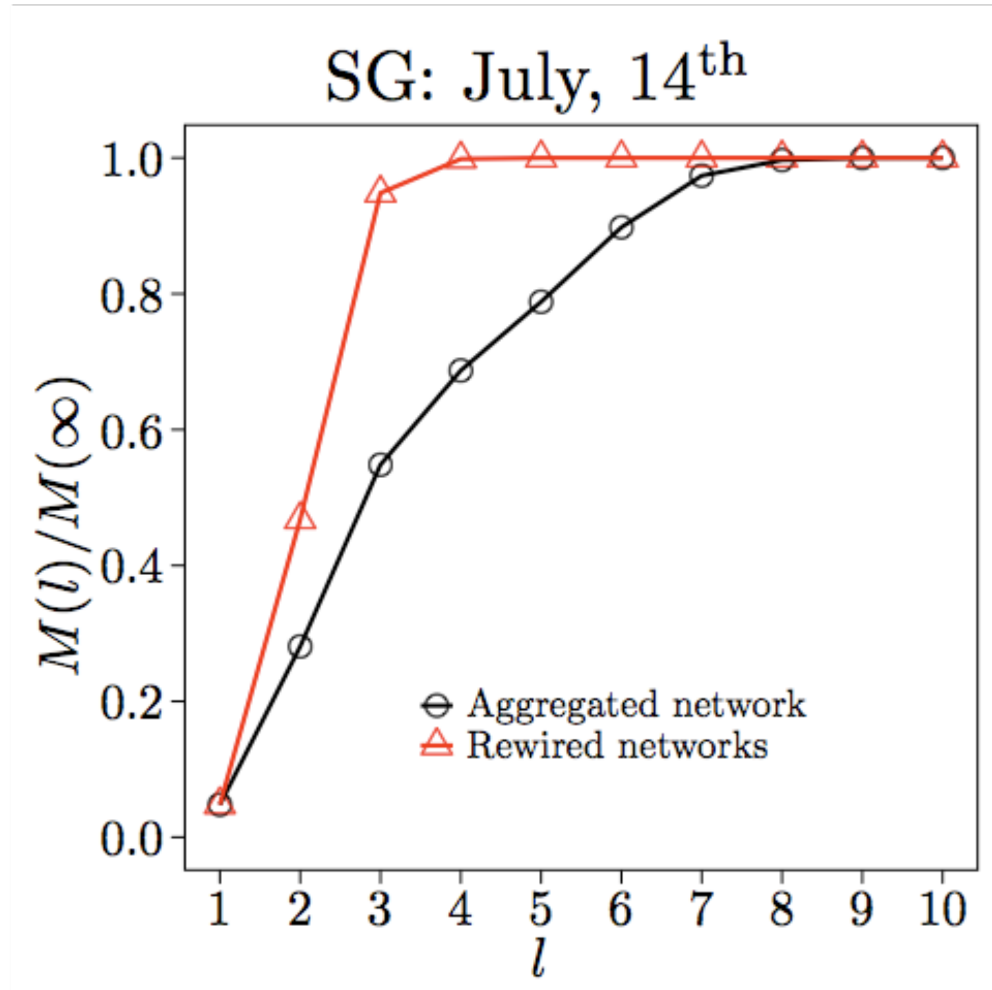
"seems" small-world


Museum (SG), rewired
Conference (HT09), rewired

(non) Small-worldness



Small-world



Non small-world

Outline of the lectures

I. Networks: definitions, statistical characterization, correlations, structures, hierarchies…

II. Modeling frameworks

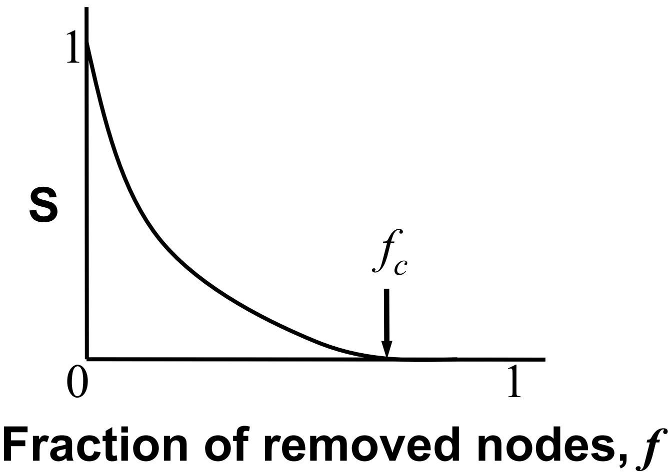
III. **Resilience, vulnerability**

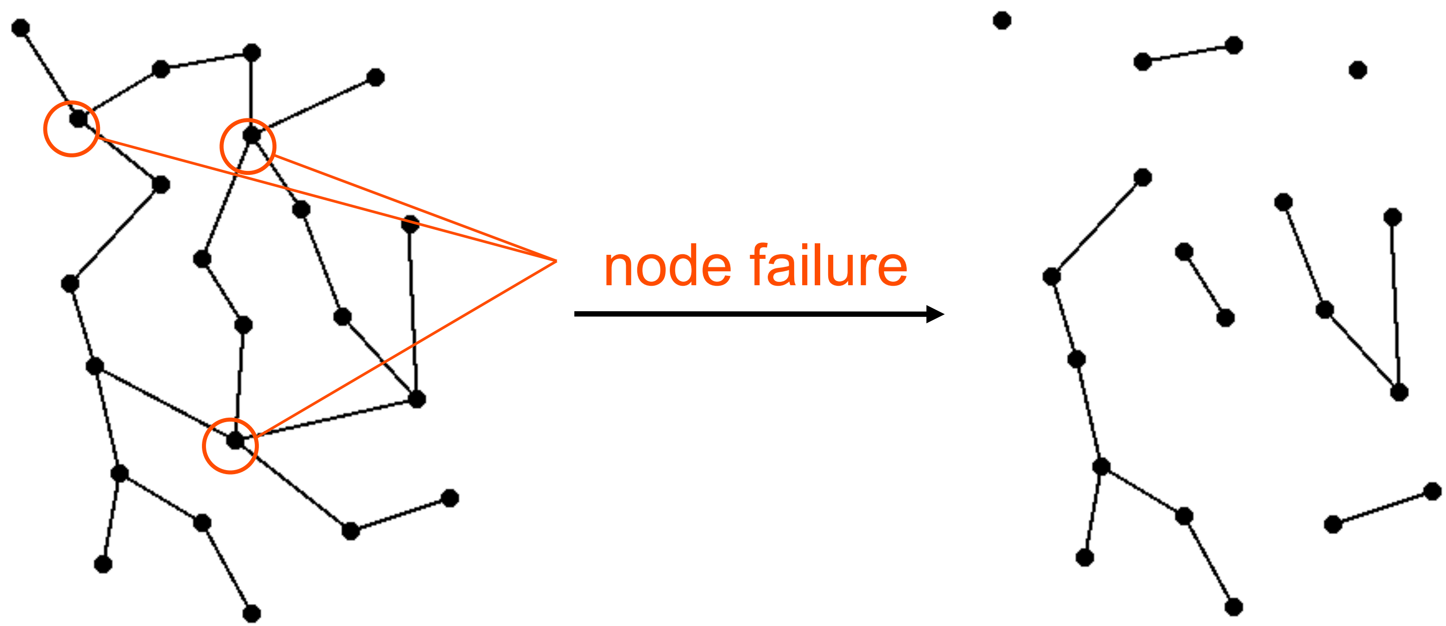
IV. Temporal networks

Robustness

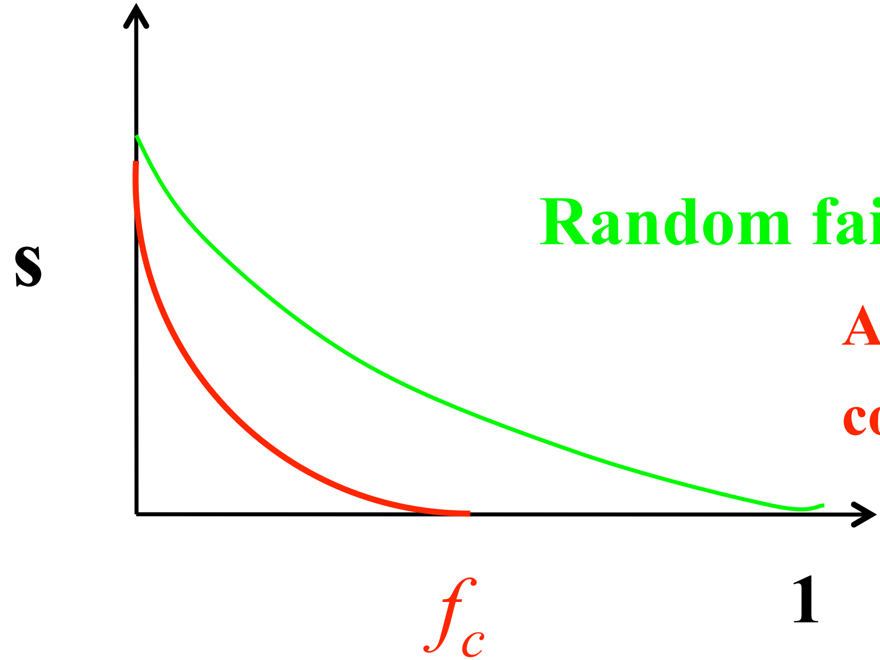
Complex systems maintain their basic functions
even under errors and failures
(cell → mutations; Internet → router breakdowns)

S: fraction of giant component





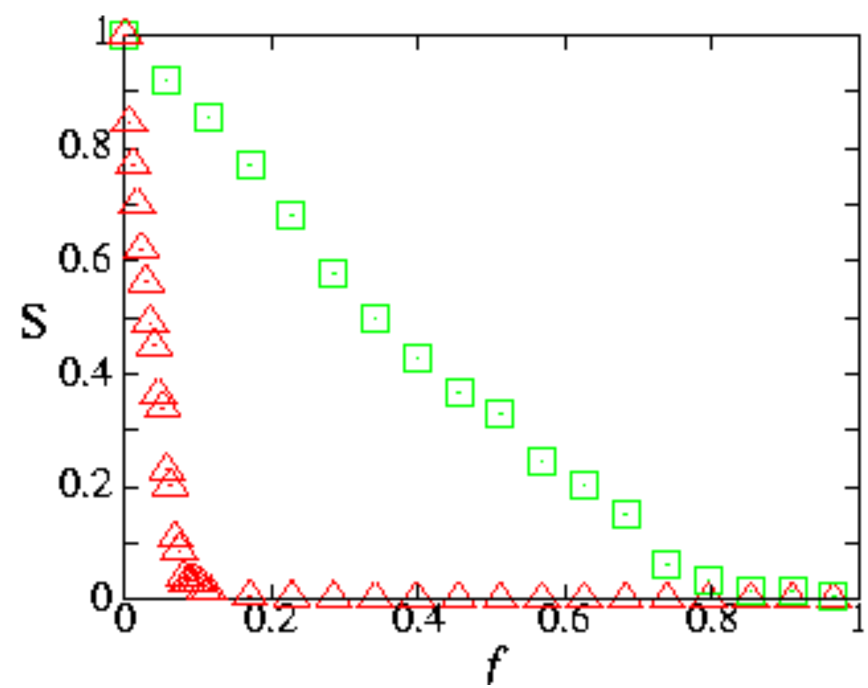
Case of Scale-free Networks



Random failure f_c =1 $(2 < \gamma \le 3)$

Attack =progressive failure of the most connected nodes f_c <1

Internet maps

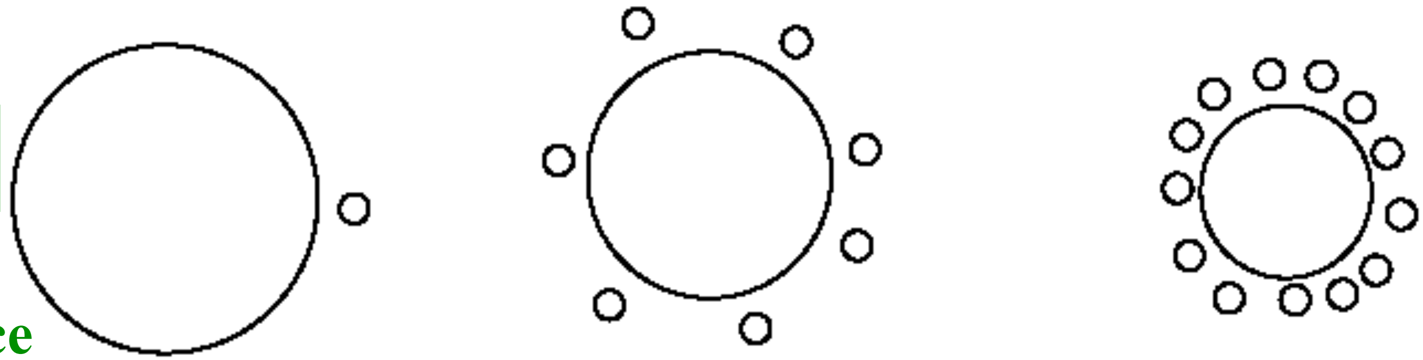


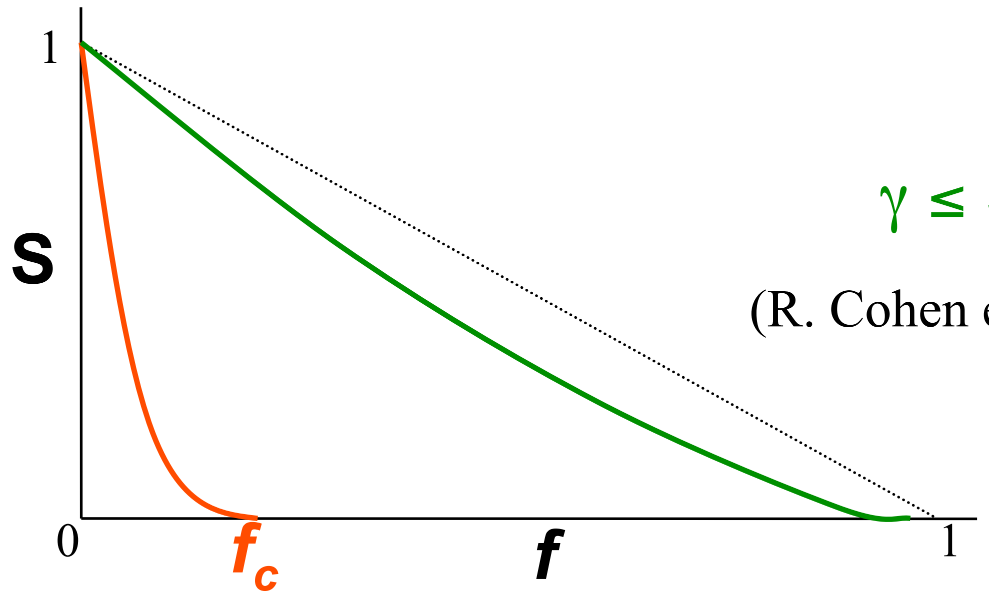
R. Albert, H. Jeong, A.L. Barabasi, Nature **406** 378 (2000)

Failures vs. attacks

Failures

Topological error tolerance

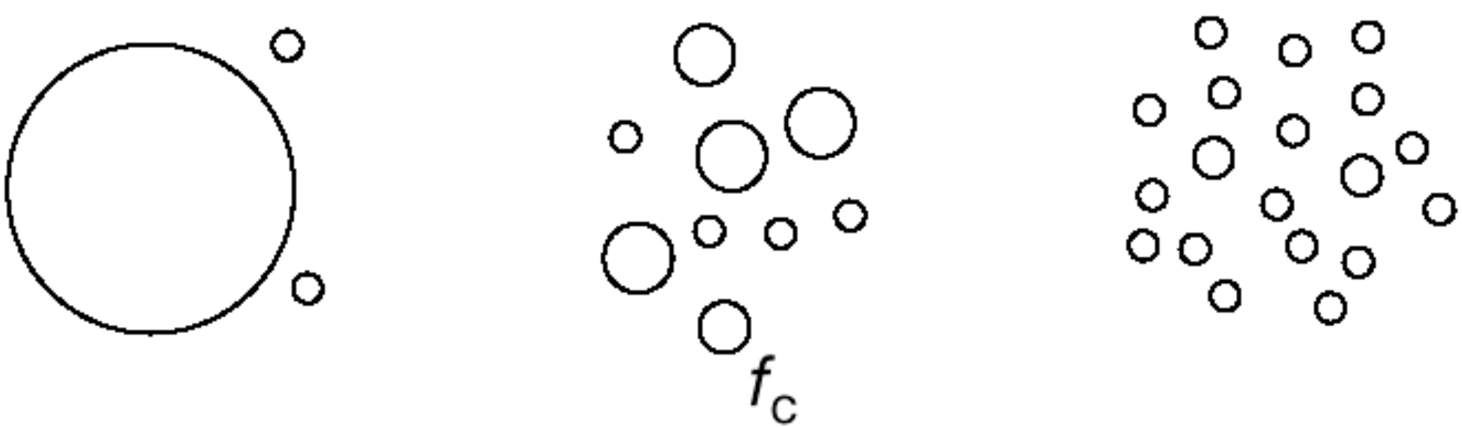




$\gamma \leq 3$: f_c=1

(R. Cohen et al PRL, 2000)

Attacks



Failures = percolation

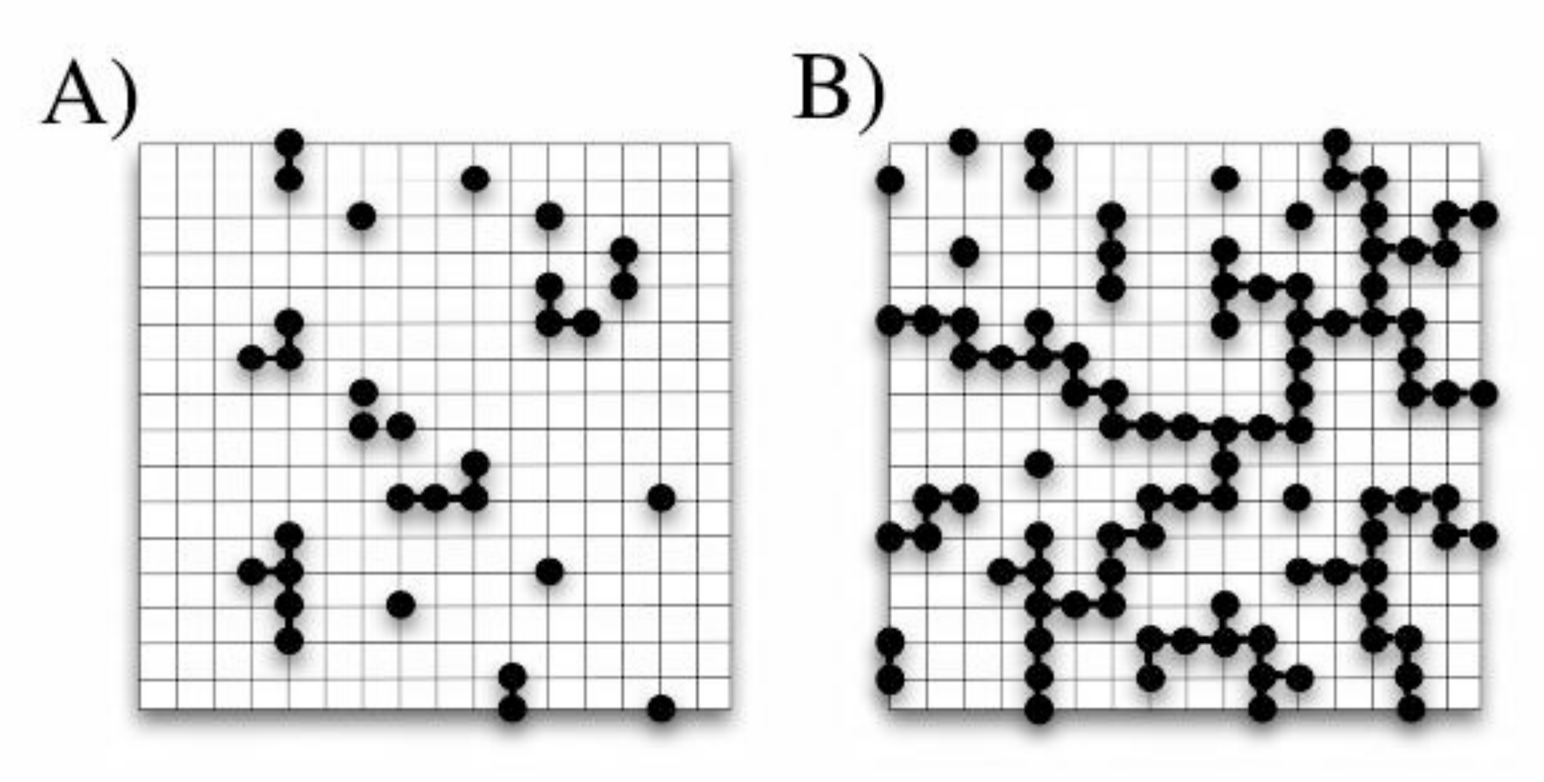
f=fraction of nodes removed because of failure

p=1-f

p=probability of a node to be present in a percolation problem

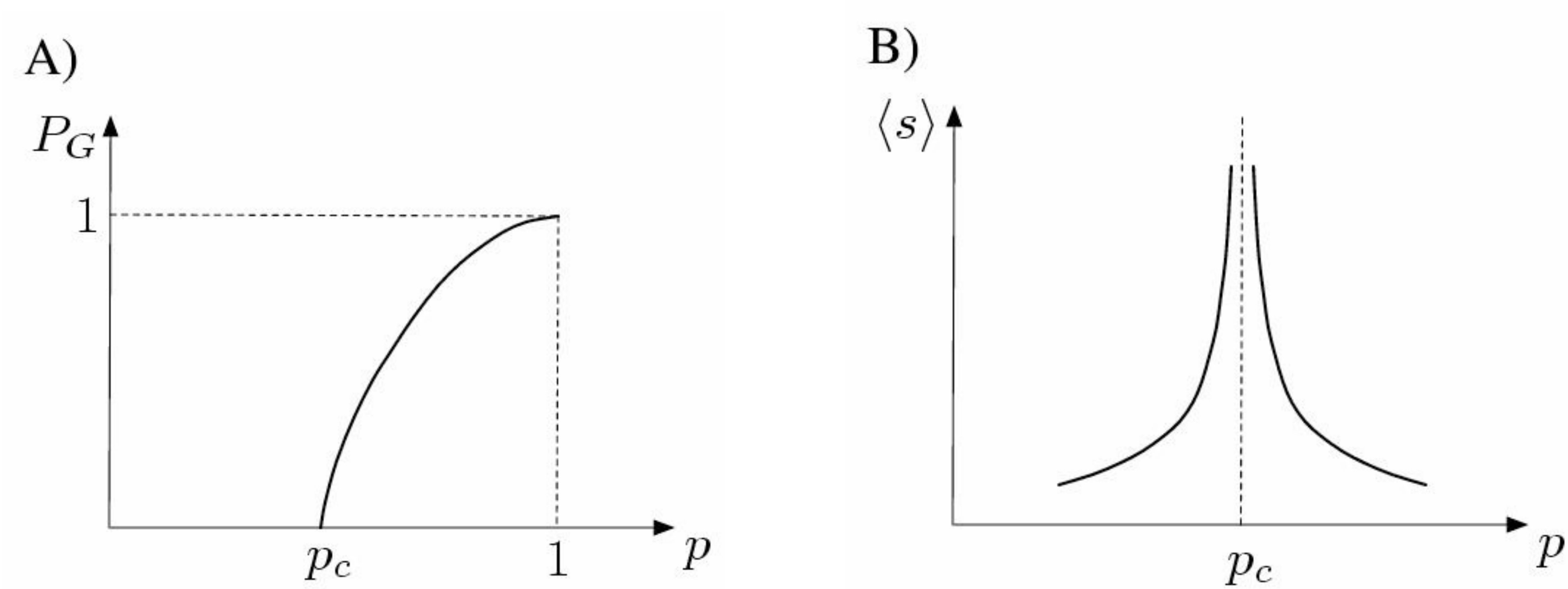
Question: existence or not of a giant/percolating cluster, i.e. of a connected cluster of nodes of size O(N)

Percolation



Question: existence or not of a giant/percolating cluster, i.e. of a connected cluster of nodes of size O(N)

Percolation



Question: existence or not of a giant/percolating cluster, i.e. of a connected cluster of nodes of size O(N)

Percolation in complex networks

q=probability that a *randomly chosen link* does **not** lead to a giant percolating cluster

$$q = \sum_k \frac{kP(k)}{\langle k \rangle} q^{k-1}$$

Sum over possible degrees

Proba that the link leads to a node of degree k

Proba that none of the outgoing k-1 links leads to a giant cluster

NB: uncorrelated random networks

Percolation in complex networks

q=probability that a randomly chosen link does **not** lead to a giant percolating cluster

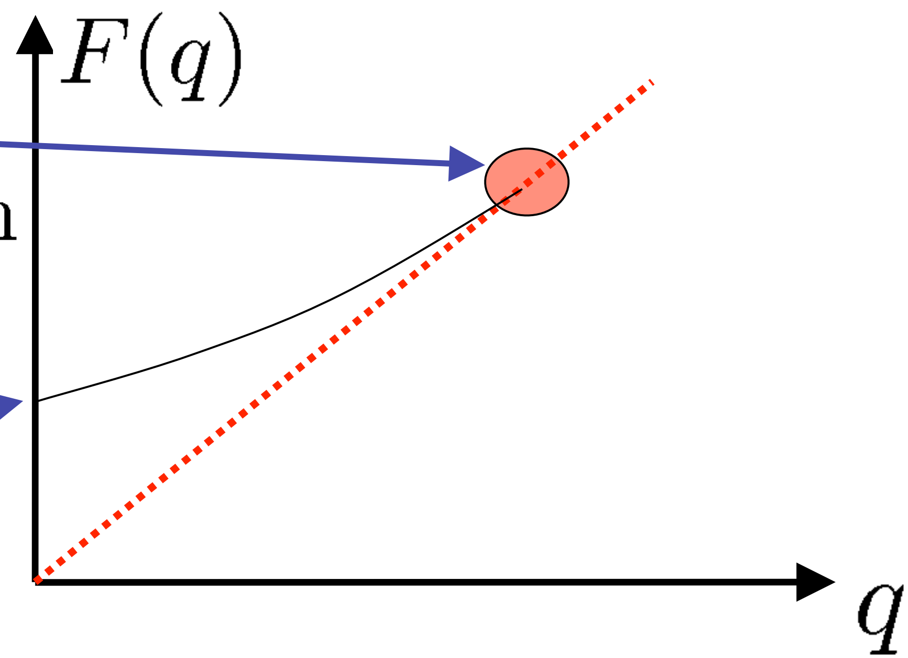
$$q = F(q), \text{with} \quad F(q) = \frac{1}{\langle k \rangle} \sum_k k P(k) q^{k-1}$$

$$F(0) > 0, F(1) = 1$$

$$F'(q), F''(q) > 0$$

$q = 1,$
always solution

No other solution if $F'(1) < 1$



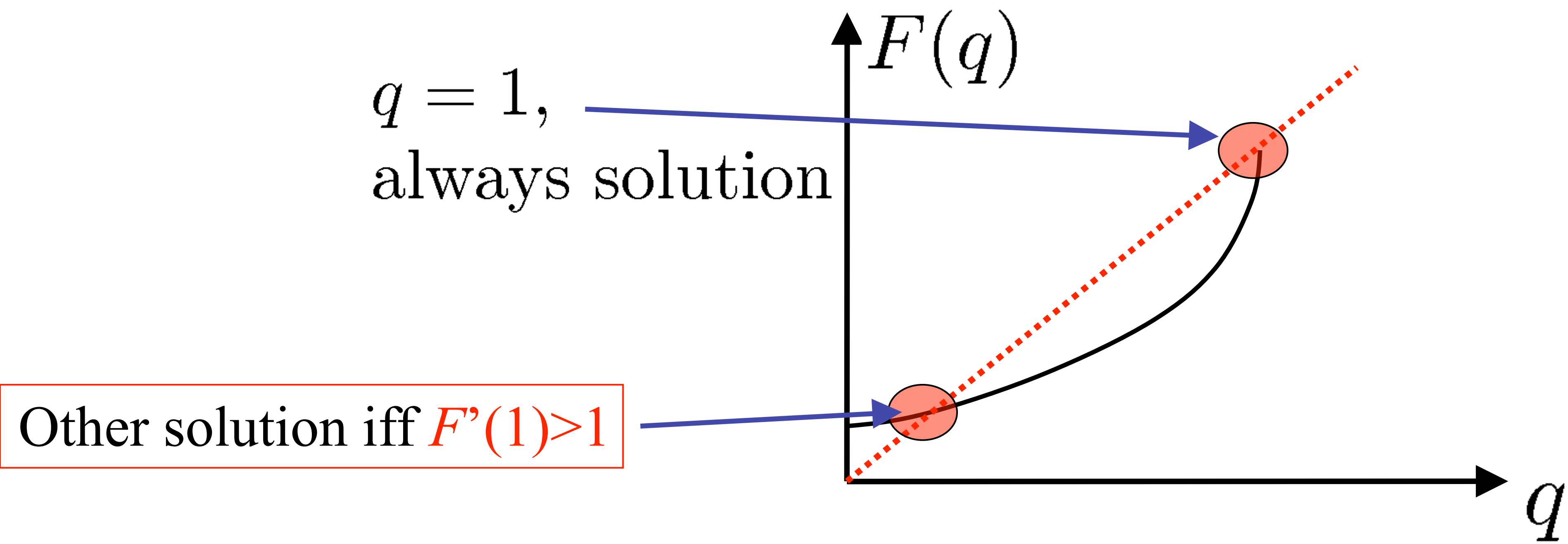
Percolation in complex networks

q=probability that a randomly chosen link does **not** lead to a giant percolating cluster

$$q = F(q), \text{with} \quad F(q) = \frac{1}{\langle k \rangle} \sum_k k P(k) q^{k-1}$$

$$F(0) > 0, F(1) = 1$$

$$F'(q), F''(q) > 0$$



Percolation in complex networks

q=probability that a randomly chosen link does **not** lead to a giant percolating cluster

$$q = F(q), \text{with} \qquad F(q) = \frac{1}{\langle k \rangle} \sum_k k P(k) q^{k-1}$$

$$F'(1) \geq 1 \quad \Longleftrightarrow \quad \langle k^2 \rangle \geq 2\langle k \rangle$$

“Molloy-Reed” criterion for the existence of a giant cluster in a random uncorrelated network

Back to random failures

Initial network: $P_0(k)$, $<k>_0$, $<k^2>_0$

After removal of fraction f of nodes: $P_f(k)$, $<k>_f$, $<k^2>_f$

Node of degree k_0 becomes of degree k with proba

$$C_{k_0}^{k}(1-f)^k f^{k_0-k}$$

$$P_f(k) = \sum_{k_0} P_0(k_0) C_{k_0}^{k}(1-f)^k f^{k_0-k}$$

Back to random failures

Initial network: $P_0(k)$, $<k>_0$, $<k^2>_0$

After removal of fraction f of nodes: $P_f(k)$, $< k>_f$, $<k^2>_f$

$$
\begin{cases}
\langle k \rangle_f = (1-f)\langle k \rangle_0 \\
\langle k^2 \rangle_f = (1-f)^2 \langle k^2 \rangle_0 + f(1-f)\langle k \rangle_0
\end{cases}
$$

Molloy-Reed criterion:
existence of a giant cluster iff $\langle k^2 \rangle_f \geq 2\langle k \rangle_f$

$$f \leq f_c, \text{with } f_c = 1 - \frac{\langle k \rangle_0}{\langle k^2 \rangle_0 - \langle k \rangle_0}$$

$\langle k^2 \rangle_0 \to \infty$ ⟹ $f_c \to 1$ ⟺ **Robustness!!!**

Finite-size effects

Finite number of nodes *N*

⇒*Finite cut-off for P(k)*

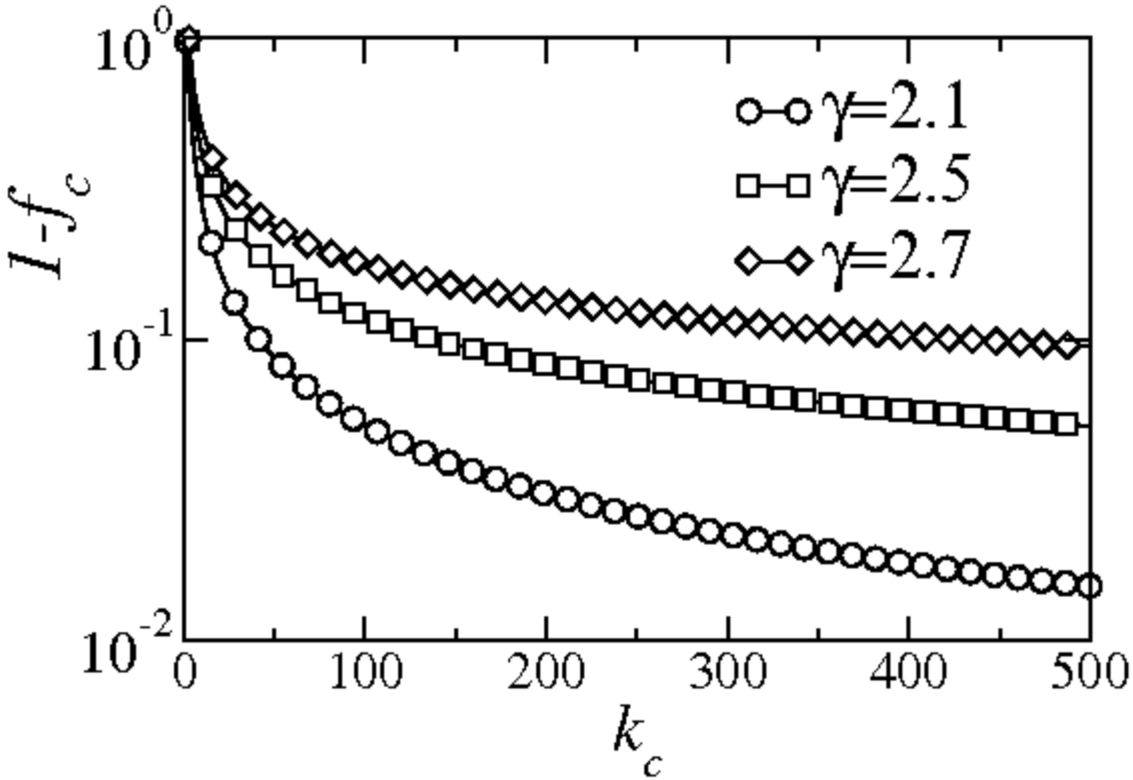
⇒*Finite* $\kappa = \langle k^2 \rangle / \langle k \rangle$

⇒f_c *strictly smaller than 1*

Ex: power-law

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

For $k < k_c$



Attacks: various strategies

- Most connected nodes
- Nodes with largest betweenness
- Removal of links linked to nodes with large k
- Removal of links with largest betweenness
- Cascades
- ...

Attacks: most connected nodes

Removal of a fraction f of nodes, such that these nodes are the most connected ones:

Implicit equation defining the largest degree after removal::

$$f = \sum_{k=k_c(f)+1}^{\infty} P(k)$$

=> Modification of the degree distribution of the remaining nodes

Attacks: most connected nodes

Removal of a fraction f of nodes, such that these nodes are the most connected ones

Modification of the degree distribution of the remaining nodes:

Probability that a neighbor of a given node has been removed=

probability that the neighbor has degree > $k_c(f)$ =

$$r(f) = \sum_{k=k_c(f)+1}^{\infty} \frac{kP(k)}{\langle k \rangle}$$

(in a random uncorrelated network)

Attacks: most connected nodes

Removal of a fraction f of nodes, such that these nodes are the most connected ones

Remaining network=

- ***Cut-off $k_c(f)$***
- ***Random removal with proba $r(f)$***

Molloy-Reed criterion => threshold f_c at which the giant component disappears

$$r(f_c) = 1 - \frac{1}{\kappa(f_c) - 1}$$

$$\kappa(f_c) = \frac{\sum_{k=1}^{k_c(f_c)} k^2 P(k)}{\sum_{k=1}^{k_c(f_c)} k P(k)}$$

Attacks: most connected nodes

Example: scale-free network, min. degree m $P(k) = ck^{-\gamma}$

$$f = \sum_{k=k_c(f)+1}^{\infty} P(k) \quad \Rightarrow \quad k_c(f) = m f^{1/(1-\gamma)}$$

$$r(f) = \sum_{k=k_c(f)+1}^{\infty} \frac{kP(k)}{\langle k \rangle} \approx f^{(2-\gamma)/(1-\gamma)} \qquad \kappa(f) = \frac{2-\gamma}{3-\gamma} \cdot \frac{k_c(f)^{3-\gamma} - m^{3-\gamma}}{k_c(f)^{2-\gamma} - m^{2-\gamma}}$$

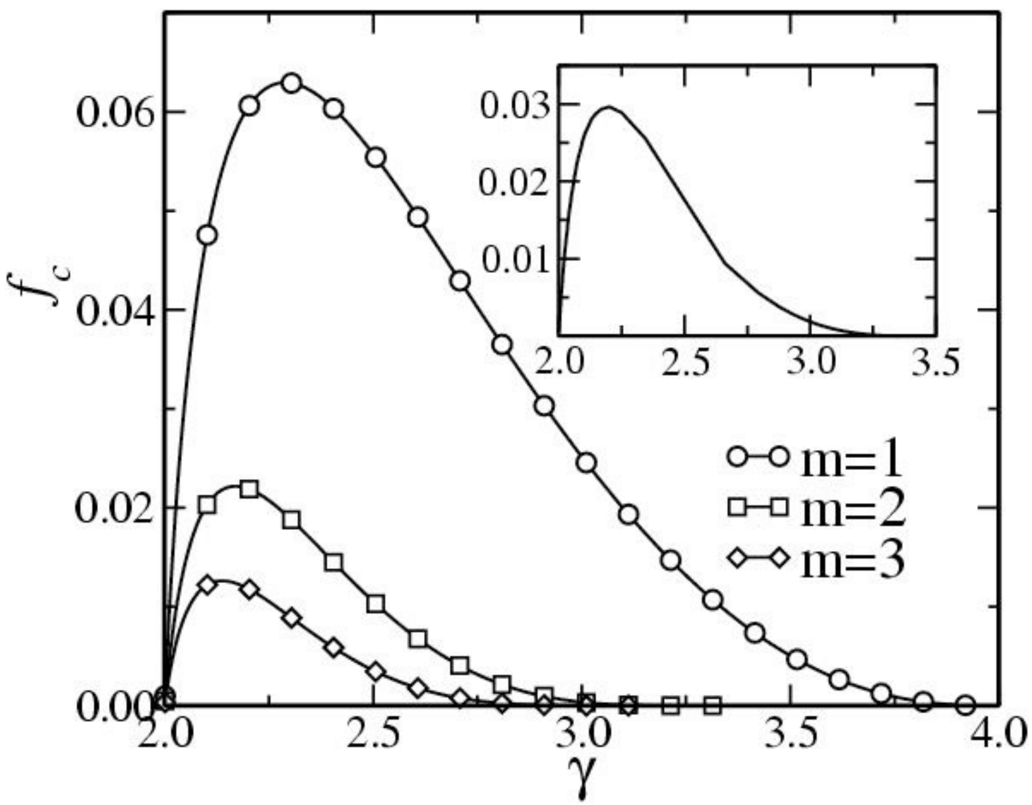
$$r(f_c) = 1 - \frac{1}{\kappa(f_c) - 1}$$

$$f_c^{(2-\gamma)/(1-\gamma)} \approx 2 + \frac{2-\gamma}{3-\gamma} m \left(f_c^{(3-\gamma)/(1-\gamma)} - 1\right)$$

Attacks: most connected nodes

Example: scale-free network, min. degree m $P(k) = ck^{-\gamma}$

$$f_c^{(2-\gamma)/(1-\gamma)} \approx 2 + \frac{2-\gamma}{3-\gamma} m \left(f_c^{(3-\gamma)/(1-\gamma)} - 1\right)$$



Attacks: other strategies

- Nodes with largest betweenness
- Removal of links linked to nodes with large k
- Removal of links with largest betweenness
- Cascades
- ...

Problem of reinforcement ?

P. Holme et al (2002); A. Motter et al. (2002);
D. Watts, PNAS (2002); Dall'Asta et al. (2006)…