The Abdus Salam International Centre for Theoretical Physics

United Nations Educational, Scientific and Cultural Organization

International Atomic Energy Agency

SMR.1656 - 3

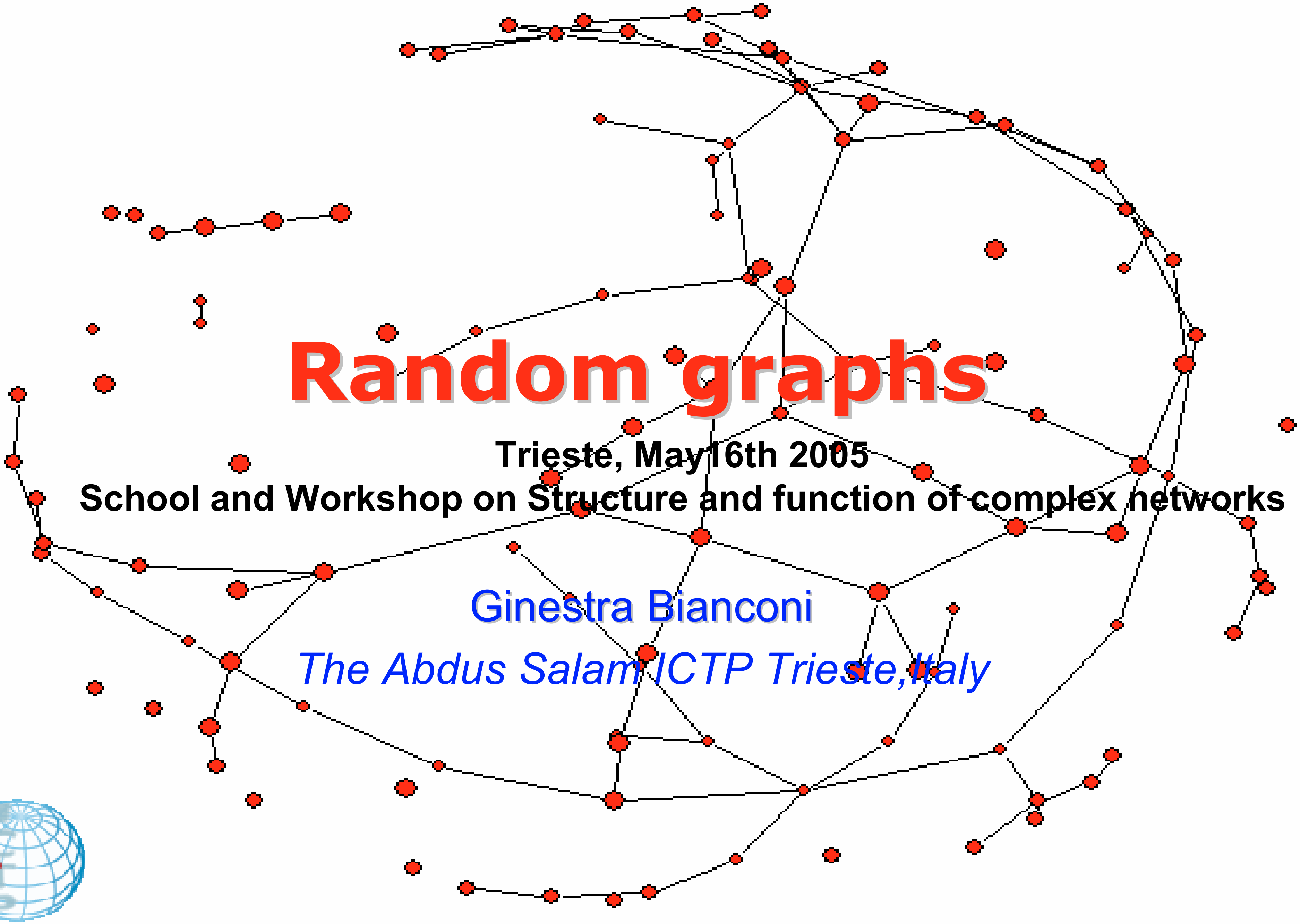
School and Workshop on Structure and Function of Complex Networks

16 - 28 May 2005

Random Graphs

Ginestra BIANCONI
the Abdus Salam ICTP
Strada Costiera 11
Trieste 34014
Italy

Strada Costiera 11, 34014 Trieste, Italy - Tel. +39 040 2240 111; Fax +39 040 224 163 - sci_info@ictp.it, www.ictp.it



Random graphs

Trieste, May 16th 2005

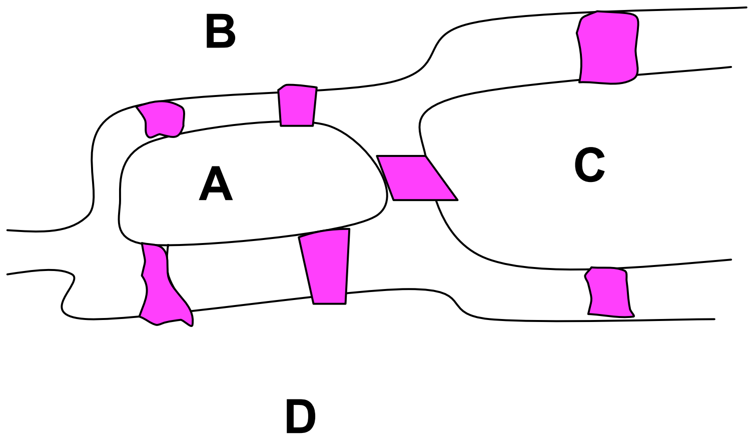
School and Workshop on Structure and function of complex networks

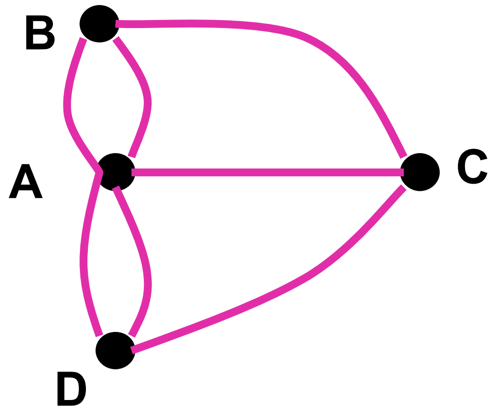
Ginestra Bianconi

The Abdus Salam ICTP Trieste, Italy

Graph theory

- 1736 Euler solved the problem of Konisberg bridges

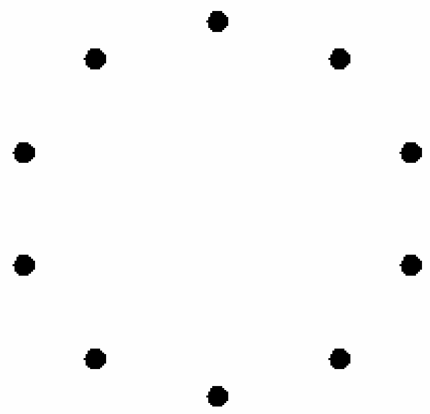




Does it exist a path that goes through each bridge only once and come back to the starting point ?

Random graphs

1947 Erdos paper introduce a **probability space** in graph theory.



- n nodes
- $n(n-1)/2$ total number of possible links
- $\mathbf{2}^{n(n-1)/2}$ total number of possible graphs

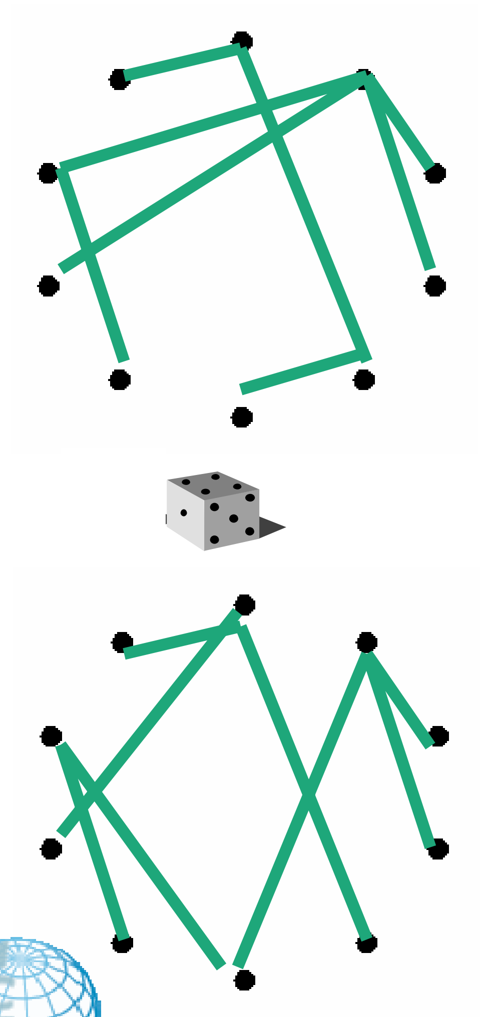
$(\Omega, \mathscr{F}, \mathbf{P})$ -probability space

Ω family of graphs with n nodes

$\mathscr{F}$ is the family of all subset of Ω

P is the probability measure of each graphs realization (G Ω) $\subset$

G(n,p) & G(n,M)



G(n,p) -Each couple of nodes are linked with probability p

n nodes
l links-random variable
<*l*>=*pn(n-1)/2*

$$P(G) = p^{l}(1-p)^{n(n-1)/2-l}$$

G(n,M)- Graphs with exactly M links

n nodes
M links

$$P(G) = \binom{n(n-1)/2}{M}^{-1}$$

Rule of thumb:
Asymptotic equivalence between G(n,p) and G(n,M) when *n* goes to infinity and *p(n)n(n-1)/2=M(n)*

In the following we will always refer to G(n,p) as Erdos and Renyi graphs

Small subgraphs appears abruptly when we increase p

$(p \approx n^{-z})$

or equivalently

the average number of links in the network

Small subgraphs in ER networks

The average number of subgraphs G of *v* nodes and *e* links is given by

$$E(X_G) = \binom{n}{v} \frac{v!}{aut(G)} p^e (1-p)^{v(v-1)/2-e} \approx n^v p^e \to \begin{cases} 0 & p << n^{-v/e} \\ \infty & p >> n^{-v/e} \end{cases}$$

Thus we have

$$P(X_G > 0) = o(1) \quad if \;\; p << n^{-v/e}$$

Probability of having a subgraph G

The probability that the number of subgraphs of type G is greater than zero satisfy a slightly more subtle condition

$$P(X_G > 0) \to \begin{cases} 0 & p << n^{-1/m(G)} \\ 1 & p >> n^{-1/m(G)} \end{cases}$$

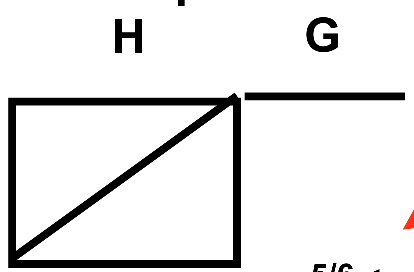
Abrupt change

Where *m(G)* is the density of the denser subset of G

$$m(G) = \max\left\{ \frac{e_H}{v_H} \quad for\, H \subset G \right\}$$

Inituition of the previous result

Suppose we want to look of subgraphs of type G : we know for sure that for $p \ll n^{-5/6}$ we don't have the network, what about $p \gg n^{-5/6}$?



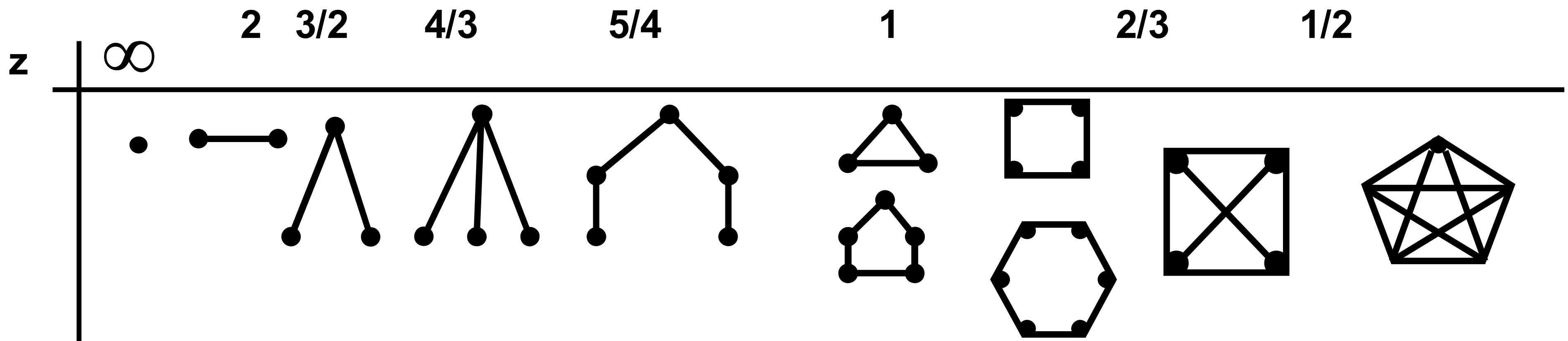
For these p value still this network G does not exist!

$n^{-5/6} < p < n^{-4/5}$

Let's consider the subgraph H we know for sure that this graph is not present in the network for $p \ll n^{-4/5}$ which is grater than $n^{-5/6}$!

Subgraph thresholds

$$p \approx n^{-z}$$



Subgraphs which appear at a given z value.

Finite connectivity

Finite connectivity random graphs
$p(n)=c/n$

Degree distribution of ER networks

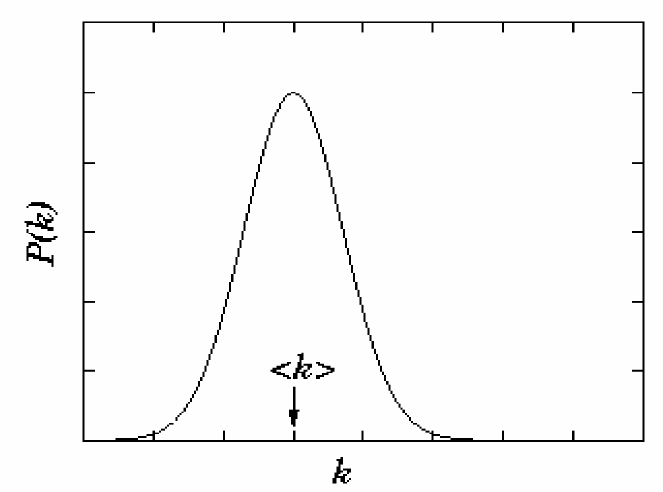
- For each node we extract *n-1* times a random number with probability *p*.
- The probability that the node has *k* links is then

$$P(k)=\binom{n-1}{k}p^{k}(1-p)^{n-1-k}\approx\frac{1}{k!}c^{k}e^{-c}$$

$$n\to\infty$$

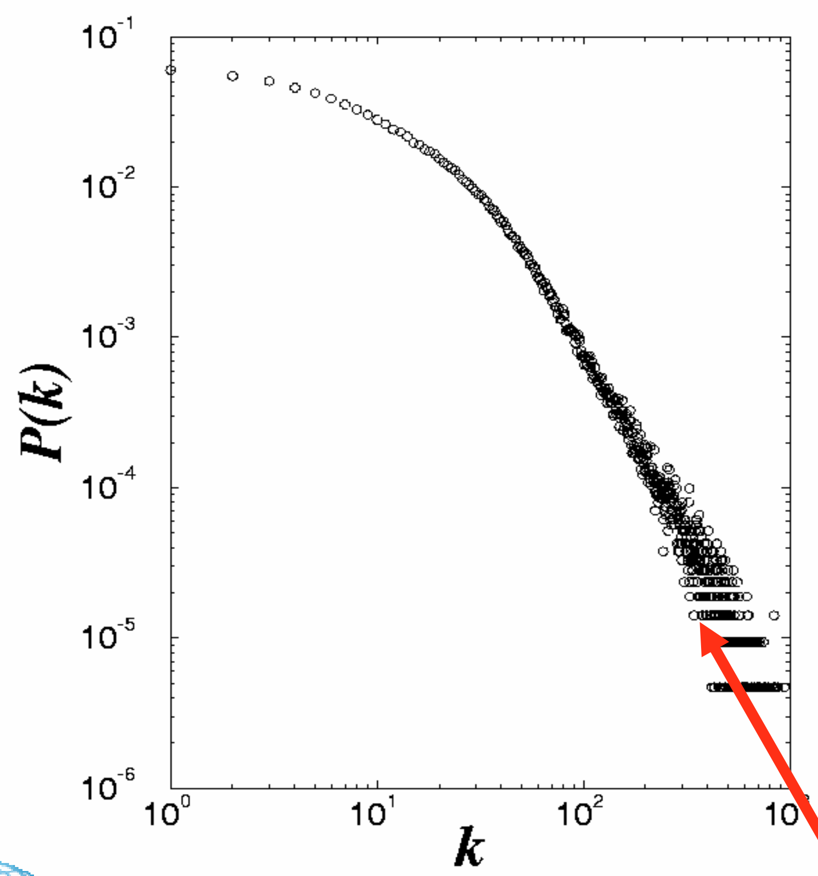
$$pn=c=\text{cost}$$

Poisson distribution



But the properties of real graph are different!

Scale-free degree distribution



Hubs

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

with $2 < \gamma < 3$

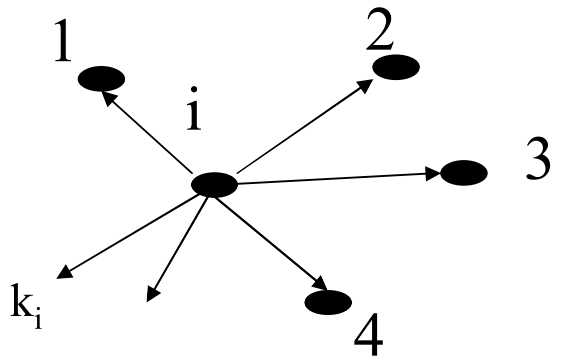
Well defined average connectivity $\langle k \rangle$
but diverging fluctuation around the mean $\langle k^2 \rangle$

with $1 < \gamma < 2$

Diverging average connectivity $\langle k \rangle$
and fluctuation around the mean $\langle k^2 \rangle$

Clustering Coefficient-Average distance

Clustering: My friends will likely know each other!



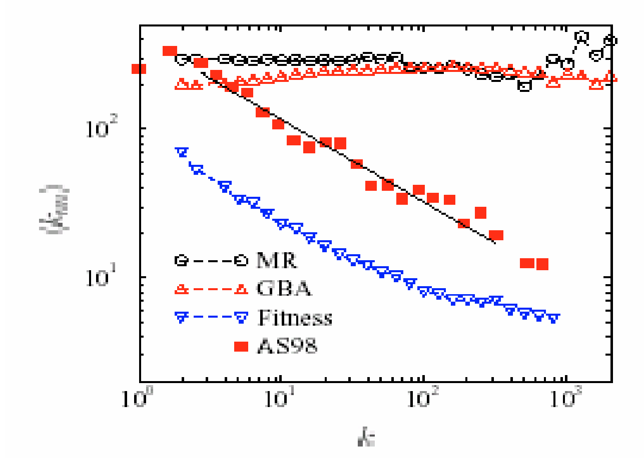
Probability to be connected C ≫ p

$$C = \frac{\text{\# of links between } 1,2,\ldots k_i \text{ neighbors}}{k_i(k_i-1)/2}$$

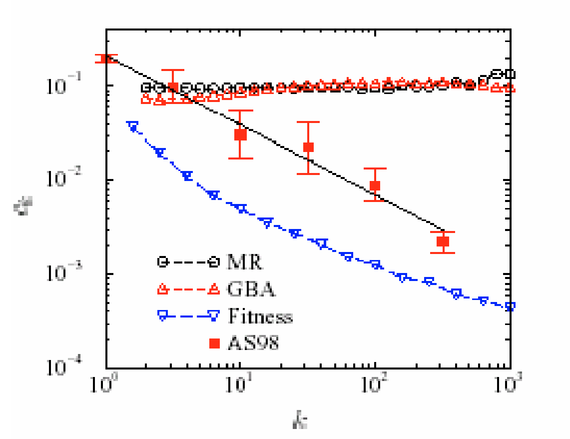
Network	C	C_{rand}	L	N
WWW	0.1078	0.00023	3.1	153127
Internet	0.18-0.3	0.001	3.7-3.76	3015-6209
Actor	0.79	0.00027	3.65	225226
Coauthorship	0.43	0.00018	5.9	52909
Metabolic	0.32	0.026	2.9	282
Foodweb	0.22	0.06	2.43	134
C. elegance	0.28	0.05	2.65	282

Degree-degree correlations for example in Internet

•$k_{nn}(k)$ **mean value of the connectivity of neighbors sites of a node with connectivity k**
If
and $\alpha<0$ the network is called disassortative while if $\alpha>0$ the network is assortative

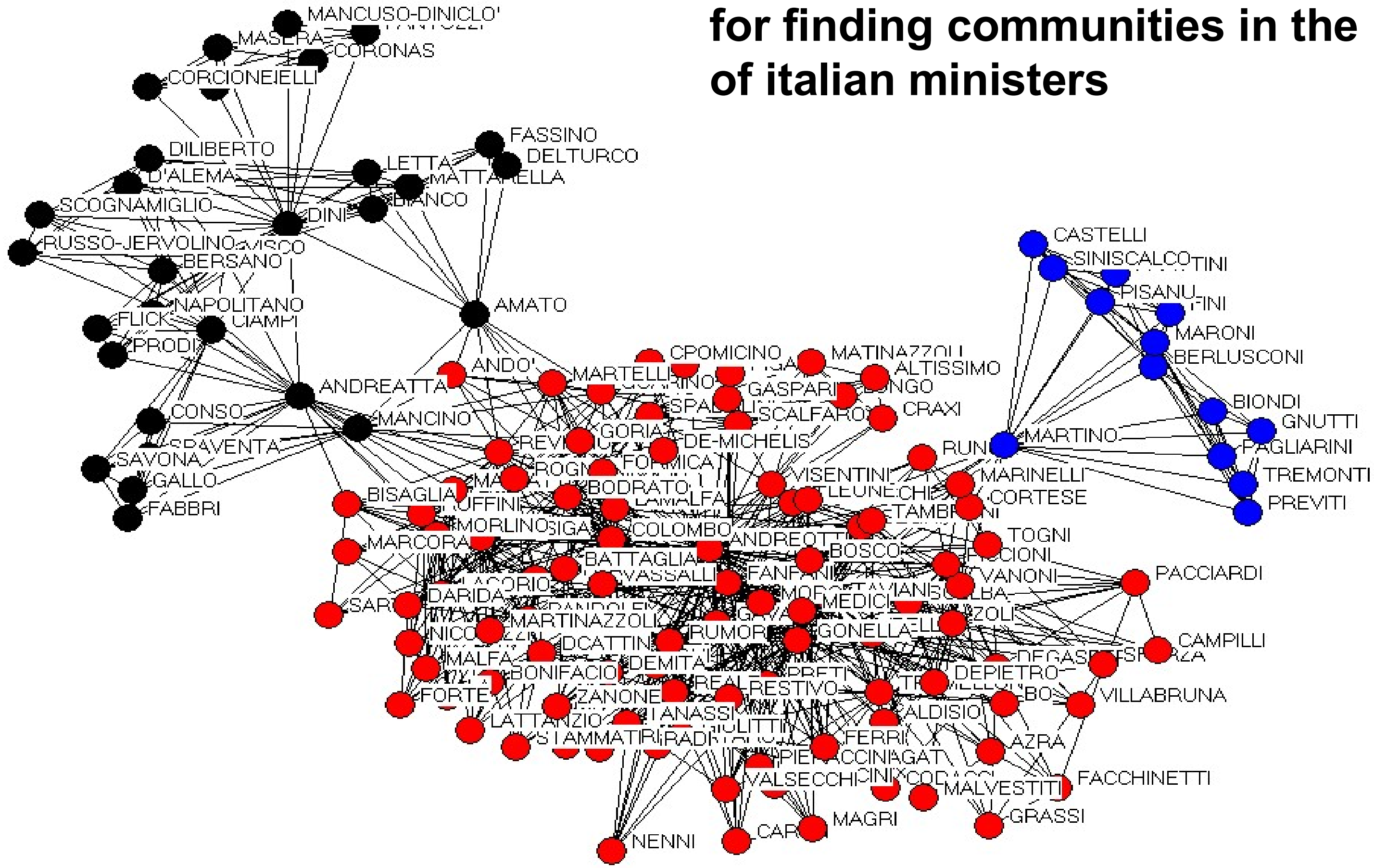


•*C(k)* average clustering coefficient of nodes with connectivity k.
If the network is called modular.



Community detection in the network of italian ministers

Givan, Newman algorithm for finding communities in the network of italian ministers



Lets building up a random graph which has at least one of all these properties: the degree distribution

Ensembles of random graphs with given degree distribution

- **Molloy-Reed ensemble**

To each node of the network *i* it is assigned a degree k_i from the desired degree distribution. Then edges are randomly matched.

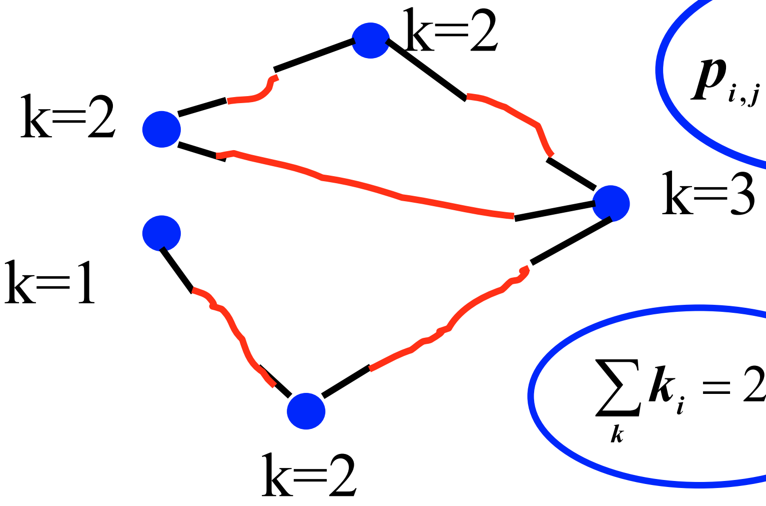
- **Hidden variable ensemble**

To each node it is assigned a random variable q_i from the desidered degree distribution. Each couple of nodes is linked with probability

$$p_{i,j} = \frac{q_i\, q_j}{\langle q \rangle n}$$

Ensembles of random graphs with given degree distribution

Molloy-Reed ensemble



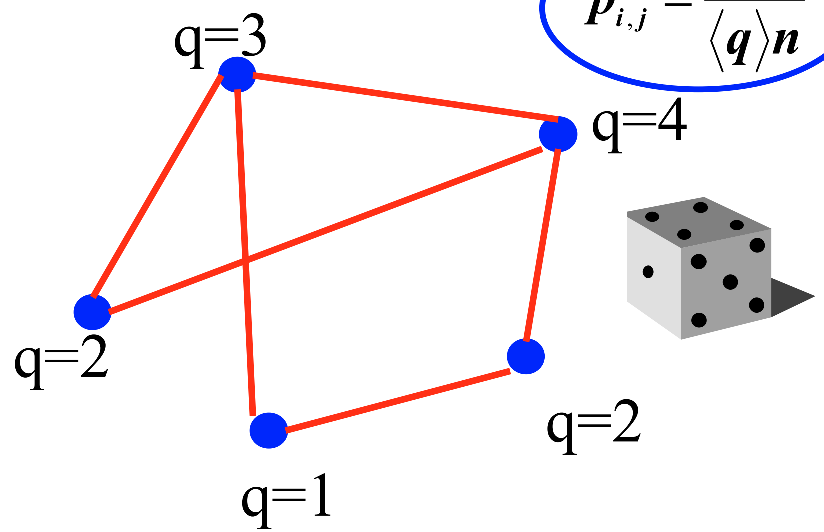
$$p_{i,j} = \frac{k_i\, k_j}{\langle k \rangle n}$$

$$\sum_k k_i = 2l$$

$$P(k) \quad k \in [m, K]$$

Hidden variable ensemble

$$p_{i,j} = \frac{q_i\, q_j}{\langle q \rangle n}$$



$$P(q) \quad q \in [m, K]$$

In these ensembles one gets the high clustering for free!

In fact the clustering coefficient of a node is given by the average number of triangles divided by the total number of possible triangles passing through that node.

$$C(k_i) = \frac{1}{k_i(k_i-1)} \left\langle \frac{(k_i-1)k'(k'-1)k''(k''-1)k_i}{\langle k \rangle^3} \right\rangle_{k',k''} = \frac{\langle k(k-1) \rangle^2}{\langle k \rangle^3}$$

But no correlations!!!

For $2<\gamma<3$ the numerator diverges and we have a clustering coefficient much higher than the one of random graphs

$$C(k) \approx \mathrm{cost} \approx N^{-(\gamma-1)/2} >> C_{ER} \approx N^{-1}$$

Giant component in random graphs with finite connectivity

Generating functions

$$P(k)$$

Probability distribution

$$F(z) = \sum_k P(k) z^k$$

Generating function

The derivatives of the generating function provides the all the moments of the distribution

$$F'(1) = \langle k \rangle$$

$$F''(1) = \langle k(k-1) \rangle$$

$$\vdots$$

Properties of generating functions

$\{f_s\}$ *F(z)*

$\{h_s\}$ *H(z)*

s=s1+s2

$$h_s = \sum_{s_1, s_2} f_{s_1} f_{s_2} \delta_{s_1+s_2, s}$$

$$H(z) = F(z)^2$$

Then it is clear that if

s=s1+s2+..sk ⇒ ***H(z)=F(z)^k***

Generating functions for the degree distribution

Generating function for the degree distribution

$$\{p_k\} \Rightarrow G_0(z) = \sum_k p_k z^k$$

Generating function for the degree of the nodes that we arrive at by following a link

$$\left\{p'_{k-1} = \frac{kp_k}{<k>}\right\} \Rightarrow G_1(z) = \frac{1}{<k>}\sum_k kp_k z^{k-1}$$

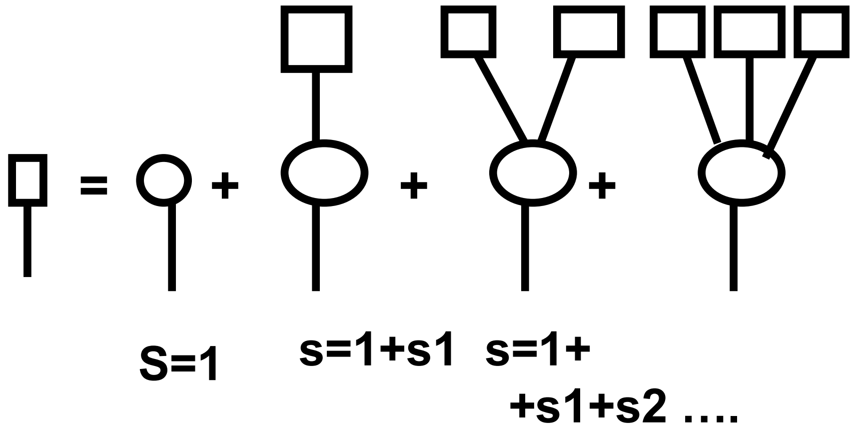
For a ER graph

$$G_0(z) = G_1(z) = e^{c(z-1)}$$

Generating functions for the cluster distribution

Generating function for the probability distribution that following a link we reach a cluster of finite size s

$$H_1(z) = zG_1(H_1(z))$$



Generating function for the probability distribution that choosing a node randomly we reach a cluster of finite size s

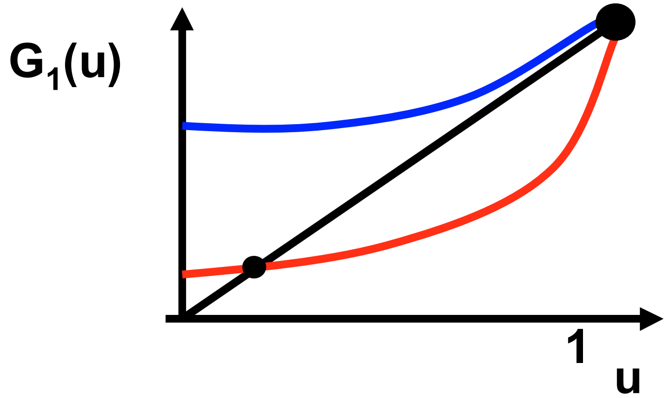
$$H_o(z) = zG_0(H_1(z))$$

Giant component

Probability that following a link we are not In the giant component *u* satisfy:

$$H_1(1) = u = G_1(u)$$

$$\begin{cases} G'(1) = \dfrac{<k(k-1)>}{<k>} > 1 \; u < 1 \quad \textit{there is a GC} \\ G'(1) = \dfrac{<k(k-1)>}{<k>} \leq 1 \; u = 1 \quad \textit{there is no GC} \end{cases}$$



- **In scale-free graphs with $\gamma<3$ there is always a GC**
- **In ER graphs there is a phase transition the birth of the GC at $c=<k>=1$**

Birth of the giant component in ER graph

As we change c in a random ER graph we find a phase transition at c=1:

one cluster of size of order *n* emerges.

This phase transition is exactly the same as percolation in infinite dimension and is naturally described by statistical mechanics methods.

Description of the phase transition

$$P \approx (c-1)^{\beta}$$

Size of the giant component

$$<s> \approx |c-1|^{-\gamma,\gamma'}$$

Average size of the finite components

$$n_s \approx s^{-\tau-1} e^{-s|c-1|^{\sigma}}$$

Distribution of the finite components

$$\beta = \gamma = \gamma' = 1$$

$$\tau = \frac{3}{2}$$

$$\sigma = 2$$

Size of the giant component in ER graph

- *S=1-u is* the probability that following a link we end in the giant component.
- *P* is the probability that choosing randomly a node it belongs to the giant component
- In a ER graph *S* satisfy

$$S = 1 - e^{-cS} \approx cS - \frac{1}{2}c^2 S^2 \quad \Leftarrow \quad \begin{array}{l} H_1(z) = zG_1(H_1(z)) \\ 1 - S = G_1(1 - S) \end{array}$$

$$\Downarrow$$

$$S \approx \frac{2(c-1)}{c^2}$$

$$\boxed{P \approx S \approx (c-1)^{\beta} \quad \beta = 1} \quad \Leftarrow \quad \begin{array}{l} H_o(z) = zG_0(H_1(z)) \\ 1 - P = G_0(1 - S) \end{array}$$

Average size of finite components

The average sizes of finite components are given by moment of $H'_0(1)$.

- Lets calculate how the average size of these components diverges with $(1-c)$ for $c<1$

$$< s >= 1 + \frac{G'_0(1)}{1 - G'_1(1)} = 1 + \frac{c}{1-c} \approx (1-c)^{-\gamma} \quad \gamma = 1$$

- For $c>1$ one has to be careful and calculate everything at $z=1$, and $H_1(1)=u$, one finds

$$< s > \approx \frac{1}{1 - G'_1(u)} = \frac{1}{1 - c(1-S)} \approx (c-1)^{-\gamma'} \quad \gamma' = 1$$

Distribution of the size of finite clusters

$$n_s \approx s^{-\tau-1}\exp(-s|c-1|^{\sigma})$$

Distribution of clusters

$$\Downarrow$$

$$P_0(s) \approx s^{-\tau}\exp(-s|c-1|^{\sigma})$$

Distribution of cluster of size s when choosing a random node

$$\Downarrow$$

$$P_1(s) \approx s^{-\tau}\exp(-s|c-1|^{\sigma})$$

Distribution of cluster of size s when choosing a random link

Distribution of clusters found following a link

If we suppose $$P_1(s) = s^{-\tau}\Phi(s|c-1|^{\sigma})$$

then $$H_1(e^{-\alpha}) = u - \alpha^{\tau-1}h(\alpha/|c-1|^{\sigma})$$

In fact we have

$$H_1(e^{-\alpha}) = u - \int ds\, s^{-\tau}\Phi(s\varepsilon^{\sigma})(1-e^{-\alpha s})$$

$$= u - \alpha^{\tau-1}h(\alpha/\varepsilon^{\sigma})$$

Where $$\varepsilon = |c-1|$$

τ and σ exponents

$$H_1(z) = zG_1(H_1(z))$$

$$1-\alpha^{\tau-1}h(x) = (1-\alpha)(1-c\alpha^{\tau-1}h(x)+\frac{1}{2}c^2\alpha^{2\tau-2}h^2(x)) + \mathrm{O}(\alpha^{3(\tau-1)})$$

$$(1-c)\alpha^{\tau-1}h(x)-\alpha+\frac{1}{2}c^2\alpha^{2\tau-2}h^2(x)+\ldots=0$$

$$x^{-\sigma}\alpha^{\tau-1+\sigma}h(x)-\alpha+\frac{1}{2}c^2\alpha^{2\tau-2}h^2(x)+\ldots=0$$

$$\begin{cases} \tau = \frac{3}{2} \\ \sigma = 2 \end{cases}$$

Average distance in ER graph

The network is locally a tree the number of nodes z_m at distance m is given by

$$z_m \approx c\left(\frac{z_2}{z_1}\right)^{m-1} \approx c^m$$

The average distance is given by *d* such that

$$\sum_{m=0,d} z_m \approx c^d = n$$

Thus *d* scales as the logarithm of the network size

$$d \approx \log(n)/\log(c)$$

Summary

- *Introduction to random graphs ensembles*
  - *ER graphs*
  - *Given degree distribution random graphs*
- *Abrupt appearance of small subgraphs in ER graphs*
- *Condition for existence of the giant component in any type of random graphs*
- *Description of the birth of the giant component in ER graphs*

Bibliography

- B. Bollobas *Random graphs*
- S. Janson,T. Luczac, A. Rucinski *Random graphs*
- R. Albert and A.L. Barabasi, *Statistical mechanics of complex networks Rev. Mod. Phys. (2002)*
- M.J.E. Newman, *The structure and function of complex networks* SIAM Rev. *(2003)*
- M. Molloy and B. Reed, *Rand. Struct. Algor*. 6,161 (1995).
- F. Chung and L. Lu, Applied Math. 26, 257 (2002).
- F. Chung and L. Lu, PNAS 99,15879 (2002).
- G. Caldarelli, A. Capocci, P. De Los Rios and M. A. Munoz, *PRL* 89 258702 (2002)