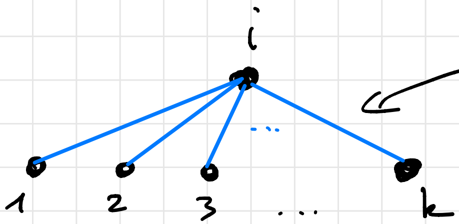
TD 4 : Erdös - Rényi random graphs

1) Local properties of E.R. random graphs.

1.1.)



exists w. proba p

$$q_k = \underbrace{\binom{N-1}{k}}_{\substack{\text{nber of ways to chose } k \text{ vertices} \\ \text{among } (N-1)}} \underbrace{p^k}_{k \text{ vertices are connected to `}i\text{'}} \underbrace{(1-p)^{N-1-k}}_{N-1-k \text{ vertices are not connected to `}i\text{'}}$$

Large N behavior: $\binom{N-1}{k} = \frac{1}{k!}(N-1)(N-2)\dots(N-k)$

$$\sim \frac{N^k}{k!}$$

Setting $p = \frac{c}{N}$ one gets for $N \gg 1$:

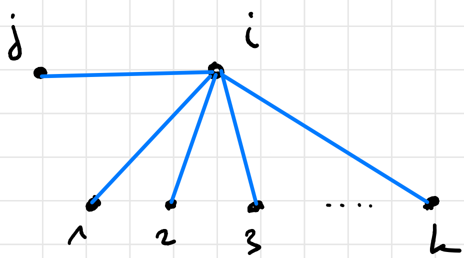
$$q_k \simeq \frac{N^k}{k!}\left(\frac{c}{N}\right)^k\left(1-\frac{c}{N}\right)^N$$

hence

$$q_k \xrightarrow[N\to\infty]{} \frac{1}{k!} c^k e^{-c}, \quad k \text{ fixed}$$

which is a Poisson distribution w. parameter c.

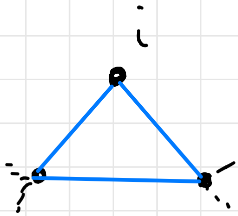
1.2) We compute the proba. of the event:



$$\mathbb{P}\left(i \text{ has } (k+1) \text{ neighbours} \;\middle|\; j \text{—} i \text{ is present}\right)$$

$$= \binom{N-2}{k} p^k (1-p)^{N-2-k} \xrightarrow[N\to\infty]{} e^{-c}\frac{c^k}{k!}$$

1.3) We now consider the proba. of the event:

$$\mathbb{P}\left(\right.$$  $$\left.\right) = \binom{N-1}{2} p^3 (1-p)^{N-3}$$

$$\underset{N \to \infty}{\simeq} \frac{N^2}{2!} \left(\frac{c}{N}\right)^3 e^{-c} = O\left(\frac{1}{N}\right)$$

More generally, the proba. that the neighbourhood of site 'i' is a given connected graph H is

$$= O\left(N^{(\# \text{vertices of } H - 1)} \left(\frac{1}{N}\right)^{\# \text{edges of } H} \right)$$

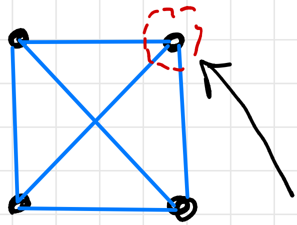
$$= O\left(N^{\# \text{vertices of } H - \# \text{edges of } H - 1} \right)$$

Let us now show that for a connected graph H:

① $\# \text{edges} \geq \# \text{vertices} - 1$

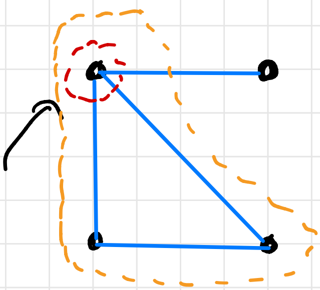
Let us start with the complete graph with # vertices $= n$, there are

edges $= \frac{n(n-1)}{2}$.



connected to $(n-1)$ vertices

To keep it connected the maximum number of edges that I can remove from that site is $(n-2)$. We are left with:

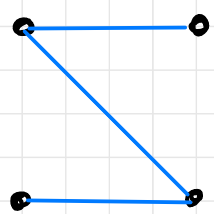


connected to $(n-2)$ edges of this complete sub-graph

We iterate the procedure on this complete sub-graph.

The max. nber of edges that one can remove is $n-3$.

One can iterate the procedure until the remaining graph is :



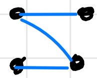
The max nber of edges that can be removed is then:

$$(n-2)+(n-3)+\dots+1=\frac{(n-2)(n-1)}{2}$$

And the minimum number of remaining edge is : $\frac{n(n-1)}{2}-\frac{(n-2)(n-1)}{2}=(n-1)$.

$\Rightarrow$ # edges $\geqslant n-1=$ # vertices -1, which proves the inequality 1 .

Note that the equality

$\#\text{ edges} = \#\text{ Vertices} - 1$ is satisfied only for  which is the tree

Therefore:

one gets that the proba. that the neighbourhood of site 'i' is a given connected graph H

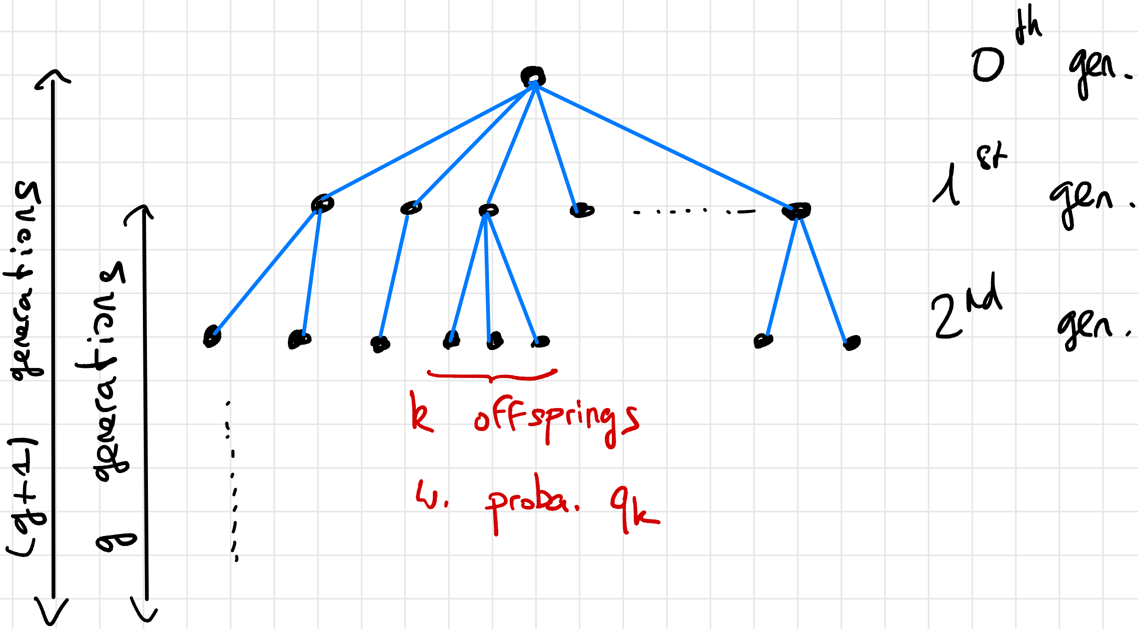
is $= O\left(N^{(\#\text{ vertices of } H - 1)} \left(\frac{1}{N}\right)^{\#\text{ edges of } H}\right)$

$$= O\left(N^{\#\text{ vertices of } H - \#\text{ edges of } H - 1}\right)$$

$\xrightarrow[N\to\infty]{}$ constant > 0 iff H is a tree.

2) Branching processes

Galton-Watson tree / process

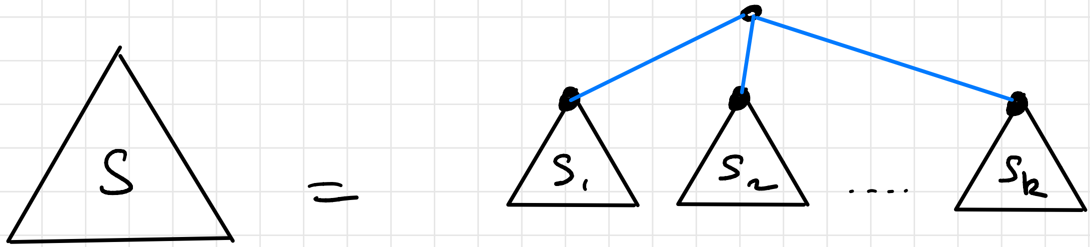


Note that $(g+1)^{th}$ generation of the full tree is the g^{th} generation of the subtree.

S = total nber of individuals

$\mathcal{N}^{(g)}$ = nber of individuals in the g^{th} gen.

2.1) Graphically one has the relation:



$$\Rightarrow \quad S = 1 + S_1 + S_2 + \dots + S_k$$

where k is a random number

and $\quad S_i \overset{d}{=} S \quad$ (in distribution)

$\Rightarrow$ Similarly one has:

$$N^{(g+1)} = N_1^{(g)} + N_2^{(g)} + \dots + N_k^{(g)}$$

2.2) Here we have: $\mathbb{E}\left(N^{(g+1)}\right) = \underbrace{\mathbb{E}(k)}_{= c}\, \mathbb{E}\left(N^{(g)}\right)$

$$\Rightarrow \mathbb{E}\left(N^{(g)}\right) = c^{g}$$

$\Rightarrow$ Transition at $c = 1$.

Indeed:

* For $c<1$: $\mathbb{E}(\mathcal{N}^{(g)}) \xrightarrow[g\to+\infty]{} 0$

and since $\mathbb{P}(\mathcal{N}^{(g)} > 0) \leq \mathbb{E}(\mathcal{N}^{(g)})$

(Markov inequality)

with high proba. $\mathcal{N}^{(g)} = 0$ for $g \to \infty$.

* For $c>1$: $\mathbb{E}(\mathcal{N}^{(g)})$ grows exp. with g and with finite probability the process persists as $g \to \infty$.

Interpretation in terms of E.R. graphs:

* $c<1$: connected components are small trees

* $c>1$: this description can not be exact up to $g \to \infty$ when one has "discovered" $O(N)$ sites: loops appear $\Rightarrow$ appearance of a giant component.

2.3) Consider the case $c<1$ (subcritical)

a) $\mathbb{E}(S) = 1 + \underbrace{\mathbb{E}(k)}_{c}\,\mathbb{E}(S)$

$$\Rightarrow \boxed{\mathbb{E}(S) = \frac{1}{1-c}}$$

An alternative derivation is:

$$\mathbb{E}(S) = \sum_{g=0}^{\infty} \mathbb{E}\left(\partial^{(g)}\right) = \sum_{g=0}^{\infty} c^{g} = \frac{1}{1-c}$$

for $c<1$.

b) Critical exponent γ_{ER}:

$$\mathbb{E}(S) \underset{c \to 1}{\sim} \frac{1}{(1-S)^{\gamma_{ER}}} \qquad \gamma_{ER} = 1.$$

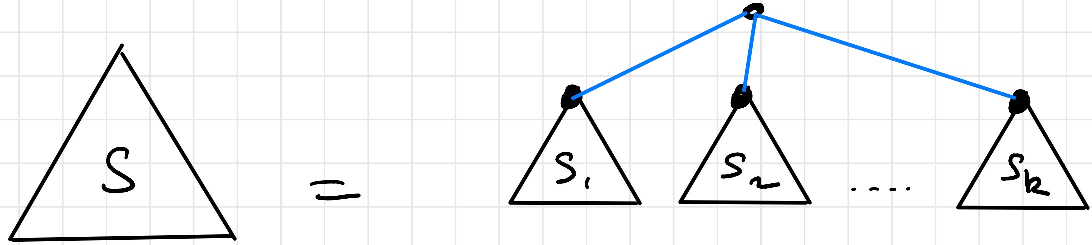
2.4) We now consider the case $c > 1$ and study $f(c) = \mathbb{P}(`S = \infty')$

a) It is convenient to study instead

$$\mathbb{P}(`S < \infty') = 1 - \mathbb{P}(`S = \infty')$$

$$= \sum_{k=0}^{\infty} p_k \, \mathbb{P}(S_1 < \infty)\, \mathbb{P}(S_2 < \infty) \dots \mathbb{P}(S_k < \infty)$$

having in mind this picture:



Hence one has:

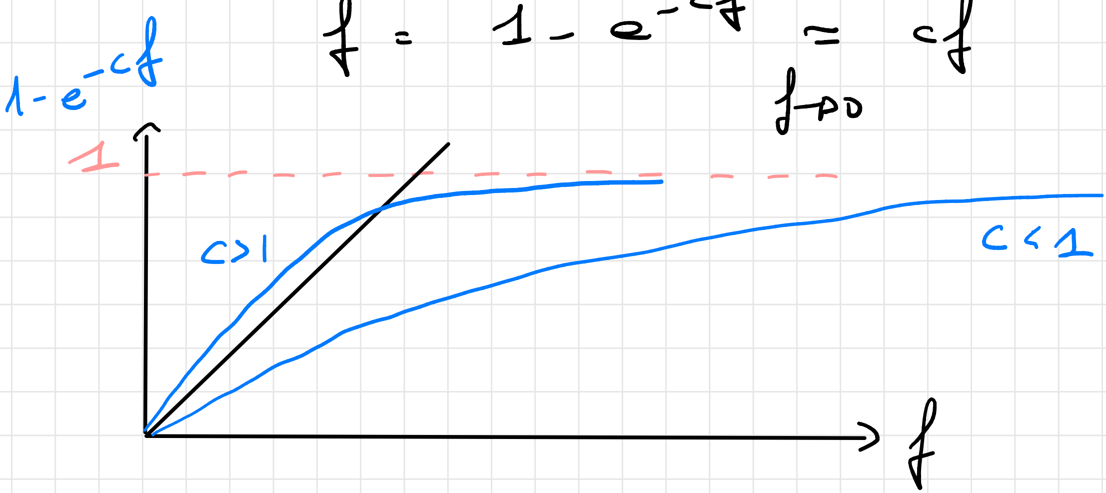
$$1 - f(c) = \sum_{k=0}^{\infty} e^{-c} \frac{c^k}{k!} \left(1 - f(c)\right)^k$$

$$1 - f(c) = e^{-c} e^{c(1 - f(c))}$$

$$\Rightarrow \qquad 1 - f(c) = e^{-c f(c)}$$

b) Hence $f \equiv f(c)$ is solution of

$$f = 1 - e^{-cf} \underset{f \to 0}{\simeq} cf$$



* For $c < 1$: the only solution is $f = 0$

* For $c > 1$: there exists a nontrivial solution $f \neq 0$.

c) For $c = 1 + \epsilon$, $\epsilon \ll 1$

one expects that the nontrivial solution for f is also small.

Expanding: $f = 1 - e^{-cf}$

with $c = 1 + \epsilon$ and small f one finds:

$$\boxed{f = 2\epsilon + O(\epsilon^2)}$$

d) Interpretation in terms of E.R.:

one sees the tree as an exploration of the E.R. graph from a given node:

* $\mathbb{P}(S < \infty)$ is the proba. that it belongs to a finite tree (not the giant component).
* $\mathbb{P}(S = \infty)$ is the proba. that it belongs to the giant component (and this is precisely the proba. that a site selected at random belongs to the giant component).

2.5) We now consider the critical case $c = 1$.

a) We introduce the generating function

$$g(x) = \sum_{s=1}^{\infty} \mathbb{P}(S = s)\, x^s = \mathbb{E}(x^S)$$

Using again $s = 1 + s_1 + \dots + s_k$

one finds:

$$\mathbb{E}(x^s) = \sum_{k=0}^{\infty} \mathbb{E}\left(x^{1+s_1+\dots+s_k}\right) p_k$$

$$= x \sum_{k=0}^{\infty} [g(x)]^k \frac{c^k}{k!} e^{-c}$$

$$\Rightarrow \quad g(x) = x\, e^{-c + c\,g(x)}$$

$$\boxed{g(x) = x\, e^{-c(1-g(x))}}$$

Remark: (i) $g(x=1) = \mathbb{P}(s < \infty) = 1 - f(c)$

$= 1$ for $c = 1$.

(ii) $g'(x) = \mathbb{E}(s\, x^{s-1})$

$$\Rightarrow \quad g'(1) = \mathbb{E}(s)$$

For $c=1$, $g(x) = x e^{g(x)-1}$

$$g(1) = e^{g(1)-1}$$

$$\Rightarrow \boxed{g(1) = 1}$$

And $g'(x) = e^{g(x)-1} + x g'(x) e^{g(x)-1}$

$\Rightarrow g'(1) = 1 + g'(1)$

which indicates that $g'(x) \xrightarrow[x \to 1^-]{} \infty$,

hence $E(S) = +\infty$ for $c=1$.

ICFP M2 - Statistical physics 2 – TD n° 4
Erdös-Rényi random graphs – Solution to the last questions

Grégory Schehr, Francesco Zamponi

2 Branching processes

5. Study of the critical case $c = 1$.

(a) The equation satisfied by the generating function of S, $g(x) = \mathbb{E}(x^S)$ is

$$g(x) = x e^{-c(1-g(x))} \,, \tag{1}$$

together with (for $c = 1$) $g(1) = 1$ and $\lim_{x \underset{<}{\to} 1} g'(x) = +\infty$.

(b) We set $x = 1 - \epsilon$ and write $g(x = 1 - \epsilon) = 1 - a\epsilon^\mu + o(\epsilon^\mu)$ as $\epsilon \to 0$. By inserting this ansatz in (1) with $c = 1$ one obtains

$$1 - a\epsilon^\mu + o(\epsilon) = (1 - \epsilon)\left(1 - a\epsilon^\mu + \frac{1}{2}a^2\epsilon^{2\mu} + o(\epsilon^{2\mu})\right) \,. \tag{2}$$

At lowest order in ϵ, one can check that the only consistent solution to (2) is

$$0 = \frac{a^2}{2}\epsilon^{2\mu} - \epsilon \Longrightarrow \begin{cases} \mu = \frac{1}{2} \,, \\ a = \sqrt{2} \,. \end{cases} \tag{3}$$

(c) One thus obtained $g(x) = 1 - \sqrt{2(1-x)}$, as $x \to 1$, which implies

$$g'(x) = \frac{1}{\sqrt{2}}(1-x)^{-1/2} + o((1-x)^{-1/2}) \,. \tag{4}$$

Recalling the definition of $g(x)$, Eq. (4) reads

$$g'(x) = \sum_{s=1}^{\infty} s\,P(s)\,x^{s-1} = \frac{1}{\sqrt{2}}(1-x)^{-1/2} + o((1-x)^{-1/2}) \,, \tag{5}$$

where $P(s) = \mathbb{P}(S = s)$. The fact that $g'(x)$ diverges as $x \to 1$ indicates that $s\,P(s)$ decays slower that $1/s$ for large s, such that the series $\sum_{s\geq 1} s\,P(s)$ is divergent. We thus assume that $P(s) \sim A/s^{\tau-1}$ for $s \to \infty$ with some amplitude A and exponent τ to be determined from (5). For this purpose, it is convenient to set $x = e^{-p}$, such that $x \to 1$ corresponds to $p \to 0$. In terms of p, the above relation (5) reads, for small p

$$e^p \sum_{s=1}^{\infty} s\,\mathbb{P}(S = s)e^{-sp} = \frac{1}{\sqrt{2}}p^{-1/2} + o(p^{-1/2}) \,. \tag{6}$$

In the small p limit, one can replace the discrete sum over s by a (Riemann) integral $e^p \sum_{s=1}^{\infty} s\,\mathbb{P}(S = s)e^{-sp} \simeq \int_1^\infty s\,\mathbb{P}(S = s)e^{-sp}ds$. By performing the change of variable $u = ps$, and substituting $P(s = u/p) \sim A(u/p)^{1-\tau}$ in the limit $p \to 0$, Eq. (7) finally leads to

$$A\,\Gamma(3-\tau)p^{\tau-3} = \frac{1}{\sqrt{2}}p^{-1/2} \,,\ p \to 0 \,, \tag{7}$$

where we have used that $\int_0^\infty u^{2-\tau} e^{-u}\, du = \Gamma(3-\tau)$. Hence Eq. (7) gives

$$\tau = \frac{5}{2} \,, \qquad A = \frac{1}{\sqrt{2\pi}} \,. \tag{8}$$

(d) The probability $\hat{P}(s)$ is the average fraction of components of an Erdös-Rényi random graph that contains exactly s vertices. On the other hand $P(s)$, that we just computed, is the average fraction of sites that belongs to a component of size s, hence one has $P(s) \propto s\hat{P}(s)$, i.e. (since it is normalised)

$$P(s) = \frac{s\hat{P}(s)}{\sum_{s'=1}^{\infty} s'\hat{P}(s')} \,. \tag{9}$$

If needed, it might be useful to convince yourself of this relation (9) on a simple example (for instance the case $N = 7$ with one component of size $S = 3$ and two of size $S = 2$). Hence, from (9), one obtains $\hat{P}(s) \propto s^{-5/2}$ for large s.

(e) To estimate the scaling (with $N \gg 1$) of the size $\mathcal{S}_N$ of the largest component at the critical point $c = 1$, let us assume that the sizes of the different connected components are i.i.d. variables, their common distribution being $\hat{P}(s)$ – this is an approximation since the sizes of the different connected components are actually correlated. In addition, we use the fact that there are $\mathcal{O}(N)$ connected components. Hence, from the results of the first lecture on extreme value statistics, since $\hat{P}(s)$ has an algebraic tail with exponent $\tau = 5/2$, one finds that $\mathcal{S}_N \sim N^{2/3}$, which coincides with the exact result.