- ER, Small World, Core Periphery, Preferential Attachment...
They are all random graphs, we need to extract random numbers to generate them sooner or later.

Today we talk about random graphs in a general way, suppose we want to generate a random graph with a certain degree distribution.
Is there any way to generate a random graph with exactly that degree distribution?
YES, and the method is quite simple.

We can have two networks with exactly the same degree distribution, so "the degree is preserved globally", but certain nodes can have different degree, so "the degree is not preserved locally" at the level of each node.

We can generate the canonical degree sequence:
We know the degree distribution p_k, this is normalized, so to know how many nodes have k degree: $N_k = N \cdot p_k$

When we know this, we can put them in lists based on degree:

$\{$ a list long N_1 $\}$, $\{$ a list long N_2 $\}$, ...

here p_k is the normalized counting of how many nodes have k degree.

How many networks can we generate with the same N and the same degree distribution?

$$\boxed{\frac{N!}{\prod_i N_i!}}$$

$N_i \Rightarrow$ number of lists of the canonical degree sequence

Molloy and Reed criterion for the existence of a giant component:

$$\boxed{\langle k^2 \rangle - 2\langle k \rangle > 0}$$

if this is respected, I will have almost surely a single connected component

second moment of the distribution

average degree of the distribution

- Dim:

In average, the number of nodes that can be reached in one step from a node, is

$$z_1 = \langle k \rangle$$

the number of nodes that can be reached in two steps...

$$z_2 = \rho \langle k \rangle \qquad \rho = k_{max} - 1$$

the same of assortativity, it's the k nearest neighbours

because we need to remove each node we get to in one step

$$\boxed{z_n = \rho^{n-1} \cdot z_1 = \rho^{n-1} \cdot \langle k \rangle}$$

this is how many nodes we can reach in n steps from the starting point.

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

$$\rho = \frac{\langle k^2 \rangle}{\langle k \rangle} - 1$$

can be QED

see from the prof. slides

If we have a giant component, we want to reach $z_n = N$, doesn't matter if we have to reach a number of steps n very high.

But, according to our formula, for very high n...

$$z_n \xrightarrow[\rho < 1,\ n \to \infty]{} 0$$

And that makes no sense!

So, if we want a giant component...

$$\boxed{\rho > 1} \implies \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 > 1$$

$$\implies \boxed{\langle k^2 \rangle - 2\langle k \rangle > 0} \qquad \square \quad \text{Q.E.D.}$$

With very few informations (just the degree distribution), we can get a condition to have a giant component

Starting from this result, we can get equations for the average path lenght and the clustering coefficient:

$$L = 1 + \frac{\ln\left(\frac{N}{\langle k \rangle}\right)}{\ln\left(\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}\right)} \qquad C = \frac{\langle k \rangle}{N}\left(\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}\right)^2$$

$$\langle k^2 \rangle = \frac{1}{N} \sum_i^N (k_i^2)$$

it's easy to calculate second and first momentum of a distribution if we know the degree of each node

$$\langle k \rangle = \frac{1}{N} \sum_i^N k_i$$

If we have a continous distribution:

$$\langle k^2 \rangle = \int_{-\infty}^{\infty} dx \; k^2 \; p(x)$$

$$\langle k \rangle = \int_{-\infty}^{\infty} dx \; k \; p(x)$$

- If we have a Scale free random graph:

$$p_k \propto \frac{1}{k^\gamma}$$

Specifically, $p_k = \frac{1}{k^\gamma} \cdot \frac{1}{\zeta(k)}$ — normalization is the riemann ZETA function

$$\Rightarrow \langle k \rangle \propto \int dx \, \frac{1}{k^\gamma} \cdot k < \infty \quad \text{if} \quad \gamma - 1 > 1 \Rightarrow \gamma > 2$$

$$\Rightarrow \langle k^2 \rangle \propto \int dx \, \frac{1}{k^\gamma} \cdot k^2 < \infty \quad \text{if} \quad \gamma - 2 > 1 \Rightarrow \gamma > 3$$

Between $2 < \gamma < 3$ $\quad \langle k^2 \rangle \to \infty \quad \langle k \rangle < \infty$

$\Rightarrow$ we have for sure a giant component, the molley-reed is respected

Specifically, we see that the riemann zeta has a 0 in 3.47875...
We have for sure a giant component for:

$$2 < \gamma < 3.47875$$

- In the case of a binomial distribution

$$\langle k^2 \rangle = p^2N^2 + pN \qquad \langle k \rangle = pN$$

$$\Rightarrow \quad p^2N^2 + pN - 2pN > 0$$

$$p^2N^2 - pN > 0$$

$$\underbrace{pN}_{\text{for sure greater than 0}}\underbrace{(pN-1)}_{} > 0 \Rightarrow pN - 1 > 0 \Rightarrow pN > 1$$

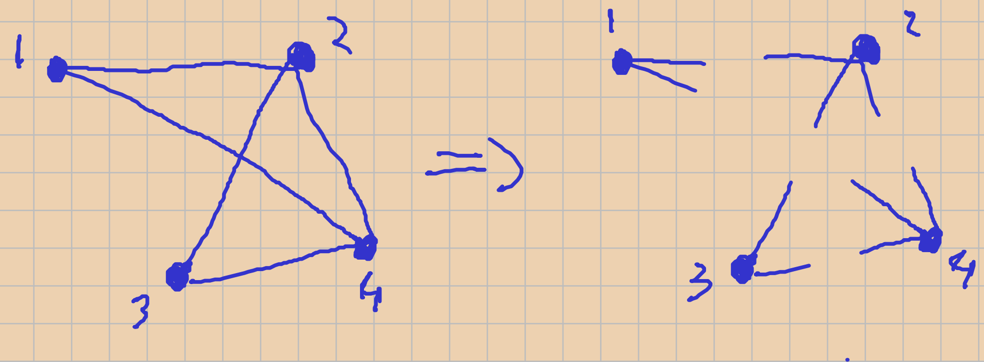
$$\Rightarrow \boxed{\langle k \rangle > 1}$$

this is something we already defined for ER without justifying it!

condition to have a giant component!

Going back to the beginning of the lecture, what if we want to preserve k LOCALLY?
Ho many networks with the same N and same degree distribution AND also every node mantains the same specifical degree?

we use the configuration model



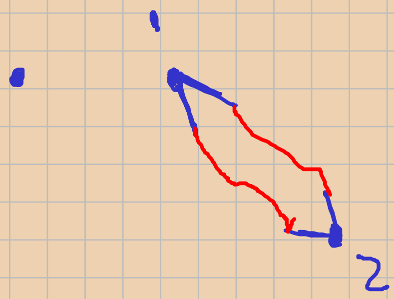
we connect back those STUBS in a random way, this way the k is locally preserved.

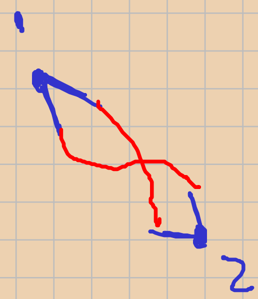
We write a list where each node appears a number of times equal to its degree

$$\{1, 1, 2, 2, 2, 3, 3, 4, 4, 4\}$$

the len of this list is the number of stubs

to generate a new graph, we extract 2 elements randomly on this list.
We can also have self loops, if the len of this list is odd we must introduce self loops





are those 2 the same?
No if we label both the nodes and the links!

The number of stubs we have in a graph with N nodes and K links is:

$$2K$$

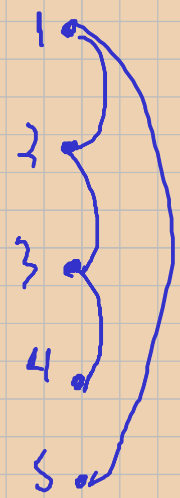
The number of totale graphs we can generate is:

$$\underbrace{(2K-1)!!}_{\text{double factorial}}$$

if we take into account the labels of both nodes and link!

We can groups the possible graphs based on topology (so the general shapes), unlabeled graphs without self-loops and multilinks are equiprobable! Otherwise no

The number of ways to create a graph like :



$\prod_i k_i = 2\cdot 2\cdot 2\cdot 1\cdot 1 = 8$ we have 8 different ways to generate it

It's possible to prove that for topologies with self-loops and multilinks the different ways to generate them is always smaller than the ones without (8 in the precedent example)

Important: It's possible to prove that the number of ways to generate networks with multilinks and self-loops is negligeable compared to the ones without, when the number of nodes N gets larger and larger