

## CONNECTIVITY

(4)

Another natural question is: for what values of  $p$  does  $G(n, p)$  become connected w.h.p.?

First we show that if  $p = c \log n$  with  $c < 1$  then  $G(n, p)$  contains an isolated point w.h.p. and therefore the graph cannot possibly be connected.

Proof Once again, we resort to the second moment method.

$$\text{Let } N = \sum_{i=1}^n \mathbb{1}_{\{\text{vertex } i \text{ is isolated}\}} = \sum_{i=1}^n X_i$$

be the number of isolated vertices.

$$\text{Clearly, } \mathbb{E}N = n(1-p)^{n-1} = n\left(1 - \frac{c \log n}{n}\right)^{n-1} \sim n^{1-c} \rightarrow \infty$$

This doesn't say anything yet. However, by the second moment method,

$$P(N=0) \leq \frac{\text{Var}(N)}{(\mathbb{E}N)^2} = \frac{\mathbb{E}(N^2) - (\mathbb{E}N)^2}{(\mathbb{E}N)^2} = \frac{\mathbb{E}(N^2)}{(\mathbb{E}N)^2} - 1$$

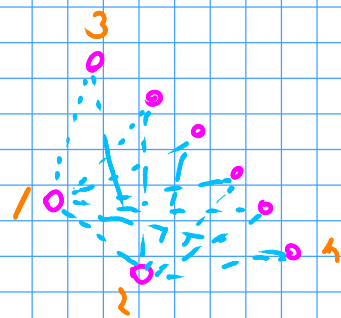
This goes to 0 if  $\frac{\mathbb{E}(N^2)}{(\mathbb{E}N)^2} \rightarrow 1$ . But

$$\mathbb{E}(N^2) = \sum_{i,j} \mathbb{E}X_i X_j = \sum_{i=1}^n \mathbb{E}X_i + \sum_{i \neq j} \mathbb{E}(X_i X_j)$$

$$= \mathbb{E}N + n(n-1)\mathbb{E}(X_i X_j)$$

$$= n(1-p)^{n-1} + n(n-1) \cdot (1-p)^{2(n-1)+1}$$

$$= (1+o(1))\left(n^{1-c} + n^2 \cdot n^{-2c}\right) = (1+o(1))n^{2-2c} \text{ as desired.}$$



Suppose now  $c > 1$ . Then

$$\mathbb{P}(G(n, \frac{c \log n}{n}) \text{ is not connected})$$

$$= \mathbb{P}(\exists K \in \{1, \dots, \lfloor \frac{n}{2} \rfloor\} \text{ and a vertex set } V \text{ with } |V|=K \text{ such that there is no edge between } V \text{ and } V^c)$$

$$\leq \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{k} (1-p)^{k(n-k)} = \sum_{k=1}^1 + \sum_{k=2}^{\lfloor \frac{\varepsilon n}{2} \rfloor} + \sum_{k=\lfloor \frac{\varepsilon n}{2} \rfloor + 1}^{\lfloor \frac{n}{2} \rfloor}$$

↑  
union bound

where we choose  $\varepsilon < \frac{2(c-1)}{c}$

$$= \text{I} + \text{II} + \text{III}$$

We estimate the three terms on the right-hand side:

$$\text{I} = n \cdot (1 - \frac{c \log n}{n})^{n-1} \leq n e^{-c \frac{\log n}{n} n} \sim n^{1-c} \rightarrow 0$$

$$\text{II} \leq \sum_{k=2}^{\lfloor \frac{\varepsilon n}{2} \rfloor} \frac{n^k}{k!} (1-p)^{k(n-k)} \leq \sum_{k=2}^{\lfloor \frac{\varepsilon n}{2} \rfloor} \frac{(n(1-p))^{n(1-\frac{\varepsilon}{2})}}{k!} = \sum_{k=2}^{\lfloor \frac{\varepsilon n}{2} \rfloor} \frac{\alpha^k}{k!}$$

we define  $\alpha = n(1-p)^{n(1-\frac{\varepsilon}{2})}$

$$\leq e^\alpha - 1 - \frac{\alpha}{2} \approx \frac{\alpha^2}{2} \rightarrow 0$$

because

$$\alpha \leq n \left(1 - \frac{c \log n}{n}\right)^{n(1-\frac{\varepsilon}{2})} \leq n e^{-c(1-\frac{\varepsilon}{2}) \log n} = n^{1-c(1-\frac{\varepsilon}{2})} \rightarrow 0 \text{ by choice of } \varepsilon$$

$$\begin{aligned}
 \text{III} &= \sum_{k=\lfloor \frac{n\varepsilon}{2} + 1 \rfloor}^n \binom{n}{k} (1-p)^{k(n-k)} && (16) \\
 &\leq n \cdot 2^n e^{-pk(n-k)} \\
 &\leq n 2^n e^{-p \frac{\varepsilon n}{2} \cdot \frac{n}{2}} = n 2^n e^{-\frac{c\varepsilon n \lg n}{4}} = n 2^n n^{-c\varepsilon n/4} \rightarrow 0
 \end{aligned}$$

□

## Remarks

(17)

- It follows from the proof above that for  $c < 1$ , there are lots of isolated vertices: if  $m_n = o(n^{1-c})$ , then  $G(n, \frac{c \log n}{n})$  has  $m$  isolated vertices w.h.p.
- Erdős & Rényi showed that if

$$p = \frac{\log n}{n} + \frac{c}{n} \text{ then}$$

$$P(G(n,p) \text{ is connected}) \rightarrow e^{-e^{-c}}$$

$$\text{also, } P(G(n,p) \text{ has } k \text{ isolated vertices}) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}$$

where  $\lambda = e^{-c}$

$$\text{in particular, } P(G(n,p) \text{ has no isolated vertices}) \rightarrow e^{-\lambda} = e^{-e^{-c}}$$

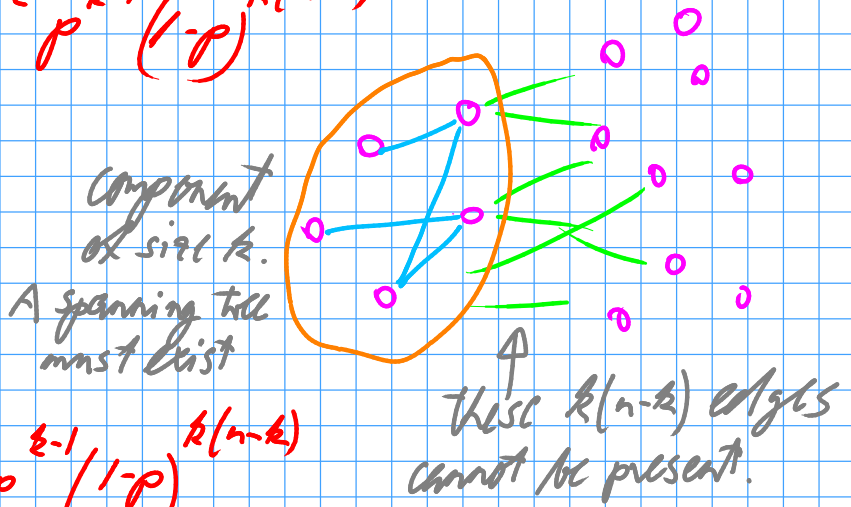
This is not a coincidence. One can prove that in the random graph process, the last edge added when the graph becomes connected connects an isolated vertex w.h.p. (Bollobás & Thomason '85). In fact, at this moment  $G(n,p)$  has a perfect matching.

In fact, apart from isolated vertices,  $G(n,p)$  becomes connected for values of  $p$  much smaller than  $\frac{\log n}{n}$ .

Indeed, let  $p = \frac{c \log n}{n}$  and let's examine the probability that there is an isolated component of size  $k$  with  $2 \leq k \leq \frac{n}{2}$ .

$$P(\exists \text{ isolated component of size } k \in [2, \frac{n}{2}])$$

$$\leq \sum_{k=2}^{n/2} \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)}$$



$$\leq \sqrt{n} \sum_{k=2}^{n/2} \frac{n^k}{k!} k^{k-3} p^{k-1} (1-p)^{k(n-k)}$$

comes from Stirling's formula

$$\leq \sum_{k=2}^{n/2} \frac{1}{p} \left( n e p (1-p)^{n-k} \right)^k = \sum_{k=2}^{\epsilon n} + \sum_{k=\epsilon n+1}^{n/2} = I + II$$

First term:

$$I \leq \sum_{k=2}^{\epsilon n} \frac{n}{c \log n} \left( n e c \frac{\log n}{n} \cdot e^{-\frac{c \log n}{n} \cdot n(1-\epsilon)} \right)^k$$

$$= \frac{n}{c \log n} \sum_{k=2}^{\epsilon n} \left( e c \log n \cdot n^{-c(1-\epsilon)} \right)^k$$

$$= O\left( n^{1-2c(1-\epsilon)} \log n \right) \rightarrow 0 \text{ if } c > \frac{1}{2} \text{ and } \epsilon \text{ is so small that } 1-2c(1-\epsilon) < 0 \left( \epsilon < 1 - \frac{1}{2c} \right).$$

(the sum is of the order of the first term)

For the second term,

(13)

$$\begin{aligned} II &\leq \sum_{k=\varepsilon n}^{n/2} \binom{n}{k} (kp)^{k-1} (1-p)^{k(n-k)} \\ &\leq n \cdot 2^n \left( \frac{n \cdot c \log n}{n} \right)^n \cdot (1-p)^{\varepsilon n \cdot \frac{n}{2}} \\ &\leq n \cdot 2^n (c \log n)^n e^{-c \varepsilon n \log n} \rightarrow 0. \quad \square \end{aligned}$$

Thus, if  $p = \frac{c \log n}{n}$  with  $c = \frac{1}{2} + \varepsilon$ , the only remaining disconnected components are isolated vertices! They won't get connected to the main component until we get to  $c=1$ , that is, we need the same number of edges to connect the last few remaining isolated vertices as to connect all other components!

### Cayley's formula

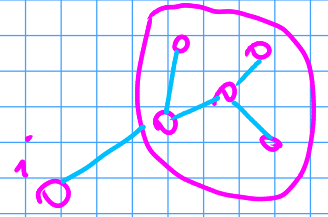
In the proof above we used the fact that the number of trees on  $n$  (labeled) vertices equals  $T_n = n^{n-2}$ . This is Cayley's formula.

Proof The proof goes by induction. For  $n=2,3$  the formula is easy to check.

Now choose a tree uniformly at random among all trees on  $n$  vertices.

What's the probability that vertex  $i$  is a leaf?  $i$  is a leaf if and only if by removing it and the corresponding edge, the remainder is a tree on  $n-1$  vertices. Thus,

(20)



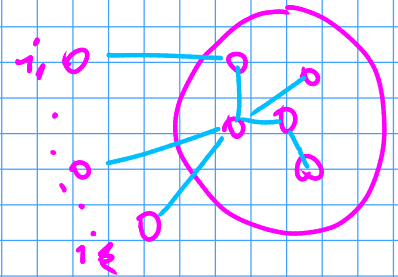
$$P(i \text{ is a leaf}) = P(A_i) = \frac{\binom{n-1}{\uparrow} \cdot T_{n-1}}{T_n}$$

$\uparrow$   
# of vertices  $i$   
can connect to

Now take  $k$  vertices  $i_1, \dots, i_k$ .

The same way,

$$P(A_{i_1}, \dots, A_{i_k}) = \frac{\binom{n-k}{\uparrow} T_{n-k}}{T_n}$$



By inclusion/exclusion,

$$P(\exists \text{ vertex } i \text{ that is a leaf})$$

$$= \sum_{k=1}^n (-1)^{k-1} \sum_{i_1 < i_2 < \dots < i_k} P(A_{i_1}, \dots, A_{i_k})$$

$$= \frac{1}{T_n} \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} (n-k)^k T_{n-k} = 1$$

$\uparrow$   
because every tree has  
at least one leaf

We obtain the recursion

$$T_n = \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} (n-k)^k T_{n-k}$$

$$= \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} (n-k)^k (n-k)^{n-k-2}$$

$$= \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} (n-k)^{n-2} = n^{n-2}$$

$\leftarrow$  by the induction hypothesis

$\leftarrow$  Exercise!

Exercise Let  $T_{k,n}$  denote the number of forests of  $k$  components such that component  $i$  contains vertex  $i$  for  $i=1, \dots, k$ . Prove that

$$T_{k,n} = k n^{n-k-1}$$

(Take  $k=1$  to recover Cayley's formula.)

(11)