CONNECTIVITY

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

First we show that if $p = c\frac{\log n}{n}$ with $c<1$ then $\mathcal{G}(n,p)$ 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. 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.

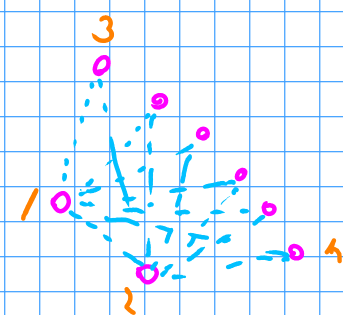
Clearly, $\mathbb{E}N = n(1-p)^{n-1} = n\left(1-\frac{c\ln n}{n}\right)^{n-1} \sim n^{1-c} \to \infty$

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

$$\mathbb{P}(N=0) \le \frac{\operatorname{var}(N)}{(\mathbb{E}N)^2} = \frac{\mathbb{E}(N^2)-(\mathbb{E}N)^2}{(\mathbb{E}N)^2} = \frac{\mathbb{E}(N^2)}{(\mathbb{E}N)^2} - 1$$

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

$$\begin{aligned}
\mathbb{E}(N^2) &= {\sum_{i,j}}' \mathbb{E}X_iX_j = \sum_{i=1}^{n} \mathbb{E}X_i + \sum_{i\neq j} \mathbb{E}(X_iX_j) \\
&= \mathbb{E}N + n(n-1)\mathbb{E}(X_1X_2) \\
&= n(1-p)^{n-1} + n(n-1)\cdot(1-p)^{2(n-2)+1} \\
&= (1+o(1))\left(n^{1-c} + n^2\cdot n^{-2c}\right) = (1+o(1))\, n^{2-2c}
\end{aligned}$$



as desired.

Suppose now $c > 1$. Then

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

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

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

union bound

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

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

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

$$\mathrm{I} = n \cdot \left(1 - \frac{c\log n}{n}\right)^{n-1} \le n\, e^{-c\frac{n-1}{n}\log n} \sim n^{1-c} \to 0$$

$$\mathrm{II} \le \sum_{k=2}^{\lfloor \frac{\varepsilon n}{2} \rfloor} \frac{n^k}{k!} (1-p)^{k \underbrace{n(1-\frac{\varepsilon}{2})}_{\ge n-k}}$$

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

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

$$\le e^{\alpha} - 1 - \frac{\alpha}{2} \approx \frac{\alpha^2}{2} \to 0$$

because $\alpha \le n\left(1 - \frac{c\ln n}{n}\right)^{n(1-\frac{\varepsilon}{2})} \le n e^{-c(1-\frac{\varepsilon}{2})\ln n}$

$= n^{1-c(1-\varepsilon/2)} \to 0$ by choice of ε.

$$\mathrm{III} = \sum_{k=\lceil \frac{n\varepsilon}{2}+1 \rceil}^{n} \binom{n}{k} (1-p)^{k(n-k)}$$

$$\le n \cdot 2^{n} e^{-pk(n-k)}$$

$$\le n 2^{n} e^{-p \frac{\varepsilon n}{2} \cdot \frac{n}{2}} = n 2^{n} e^{-\frac{c\varepsilon n \lg n}{4}} = n \cdot 2^{n} n^{-c\varepsilon n/4} \longrightarrow 0$$

$\boxed{\mathrm{III}}$

Remarks

- It follows from the proof above that for $c<1$, there are lots of isolated vertices: if $m_n = o(n^{1-c})$, then

$$G\left(n, \frac{c\log n}{n}\right) \text{ has } m \text{ isolated vertices w.h.p.}$$

- Erdős & Rényi showed that if

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

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

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

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

In particular, $\mathbb{P}(G(n,p) \text{ has no isolated vertices})$

$$\to e^{-\lambda} = e^{-e^{-c}}.$$

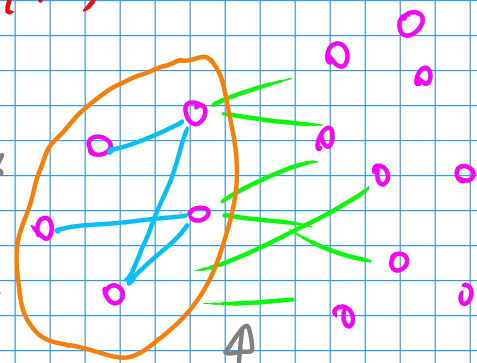
This is not a coincidence. One can prove that in the random graph process, the last edge added when the graph becomes connected connects an isolated vertex w.h.p. (Bollobás & Thomason '85).
In fact, at this moment $G(n,p)$ has a perfect matching (if n is even).

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 $2 \le k \le \frac{n}{2}$.

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

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



component of size k.

A spanning tree must exist

these $k(n-k)$ edges cannot be present.

$$\le \sqrt{2\pi} \sum_{k=2}^{n/2} \frac{n^k}{\left(\frac{k}{e}\right)^k} k^{k-3/2} p^{k-1} (1-p)^{k(n-k)}$$

comes from Stirling's formula

$$\le \sum_{k=2}^{n/2} \frac{1}{p} \left(nep(1-p)^{n-k}\right)^k = \sum_{k=2}^{\varepsilon n} + \sum_{k=\varepsilon n+1}^{n/2} = \mathrm{I} + \mathrm{II}$$

First term:

$$\mathrm{I} \le \sum_{k=2}^{n\varepsilon} \frac{n}{c \log n} \left(nec \frac{\log n}{n} \cdot e^{-\frac{c \log n}{n} \cdot n(1-\varepsilon)}\right)^k$$

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

$$= O\left(n^{1-2c(1-\varepsilon)} \log n\right) \longrightarrow 0 \text{ if } c > \tfrac{1}{2} \text{ and } \varepsilon \text{ is so small that } 1-2c(1-\varepsilon) < 0$$

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

$\left(\varepsilon < 1 - \frac{1}{2c}\right)$.

For the second term,

$$\underline{II} \le \sum_{k=\varepsilon n}^{n/2} \binom{n}{k} (kp)^{k-1} (1-p)^{k(n-k)}$$

$$\le n \cdot 2^n \left(\frac{n \cdot c \log n}{n} \right)^n \cdot (1-p)^{\varepsilon n \cdot \frac{n}{2}}$$

$$\le n \cdot 2^n (c \log n)^n e^{-c\varepsilon n \log n} \to 0. \quad \blacksquare$$

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!

$\sqrt{n}$

Remark Komlós & Szemerédi (1983) showed that if

$$p = \frac{\log n}{n} + \frac{\log \log n}{n} + \frac{c_n}{n},$$

then

$$\mathbb{P}(\mathcal{G}(n,p) \text{ has a Hamiltonian cycle}) \to \begin{cases} 1 & \text{if } c_n \to \infty \\ 0 & \text{if } c_n \to -\infty \\ e^{-e^{-c}} & \text{if } c_n \to c \end{cases}$$

Bollobás (1984) proved that in the random graph process that adds edges one by one, w.h.p. the graph becomes Hamiltonian exactly when the minimum degree becomes ≥ 2.

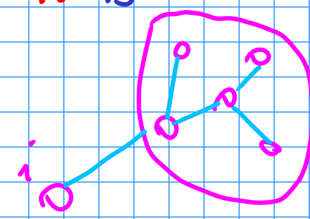
Cayley's formula

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 is 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,

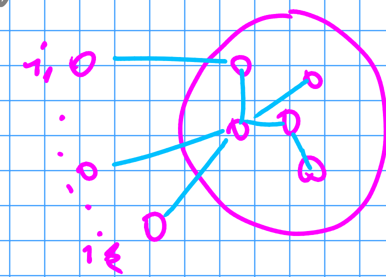


$$\mathbb{P}(i \text{ is a leaf}) = \mathbb{P}(A_i) = (n-1)\cdot T_{n-1} / T_n$$

# of vertices i can connect to

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

The same way,



$$\mathbb{P}(A_{i_1}, \ldots, A_{i_k}) = (n-k)^k \, T_{n-k} / T_n$$

By inclusion/exclusion,

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

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

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

because every tree has at least one leaf

We obtain the recursion

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

$$= \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} (n-k)^k (n-k)^{n-k-2} \leftarrow \text{by the induction hypothesis}$$

$$= \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} (n-k)^{n-2} = n^{n-2} \leftarrow \text{Exercise!}$$

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)