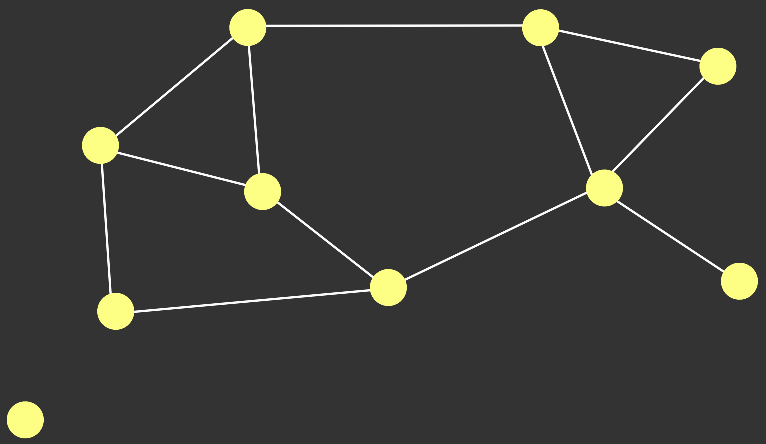
Random graphs



Erdős - Rényi random graph model

Two variants:

1. In the $\mathcal{G}(n, N)$ model a graph is chosen uniformly at random from the collection of all graphs which have n labeled nodes and N edges.

2. In $\mathcal{G}(n, p)$ model, a graph is constructed by connecting labeled nodes randomly. Each edge is included in the graph with probability p, independently from every other edge.

Comparison between two models.

Consider the second model $\mathcal{G}(n, p)$.

What is the probability P_N that a random graph has exactly N edges?

Comparison between two models.

Consider the second model $\mathcal{G}(n, p)$.

What is the probability P_N that a random graph has exactly N edges?

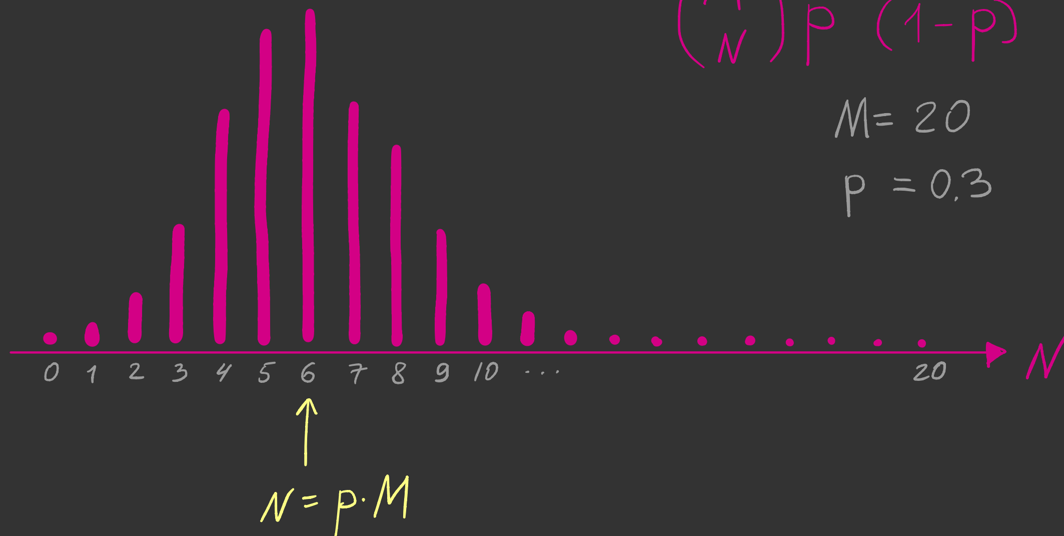
$$\binom{\frac{n(n-1)}{2}}{N} p^N (1-p)^{\frac{n(n-1)}{2} - N}$$

Bell curve again

$$\binom{M}{N} p^N (1-p)^{M-N}$$

$M = 20$

$p = 0.3$



Heuristically, if $pn^2 \to \infty$, then $\mathcal{G}(n,p)$ behaves similarly to $\mathcal{G}(n,N)$ with $N = \binom{n}{2} \cdot p$.

Question:

What is the probability that a random graph is connected?

The answer depends on parameters n, N or n, p.

Answer (Erdös - Renyi 1959)

As n goes to infinity

$$N = \frac{1}{2} n \cdot \log(n) \qquad p = \frac{\log(n)}{n}$$

is a sharp threshold for the connectedness of

$$\mathcal{G}(n, N) \qquad \mathcal{G}(n, p)$$

Let c be an arbitrary fixed real number.

We set

$$N_c = \left[\frac{1}{2} n \log(n) + c \cdot n \right].$$

Theorem (Erdös - Rényi)

Let $P(n, N_c)$ denote the probability of $\mathcal{G}_{n,N_c}$ being connected.

Then we have

$$\lim_{n \to \infty} P(n, N_c) = e^{-e^{-2c}}.$$

Other questions:

1. What is the probability that $\mathcal{G}(n, N)$ has exactly k connected components

2. What is a typical size of the biggest connected component

...

Definition: A graph $G=(V,E)$ is of **type A** if it consists of k isolated points and a connected graph having $n-k$ vertices.

Any graph which is not of type A is called to be of **type $\bar{A}$**.

Lemma: Let $P(\bar{A}, n, N_c)$ denote the probability of $\mathcal{G}_{n,N_c}$ being of type $\bar{A}$. Then we have

$$\lim_{n\to\infty} P(\bar{A}, n, N_c) = 0.$$

Thus for a large n almost all graphs $\mathcal{G}_{n,N_c}$ are of type $\bar{A}$.

Proof of the lemma:

Let M be a large positive number, which will be chosen later.

E_M - the set of graphs in $\mathcal{G}(n, N_c)$ such that the greatest connected component consists of not less then $n-M$ points.

$\overline{E}_M$ - complement of E_M.

Let $\mathcal{N}(\overline{E}_M, n, N_c)$ denote the size of $\overline{E}_M$.

If the graph consists of r connected components having ℓ_i points $(i=1,2,\dots,r)$ then

$$\sum_{i=1}^{r} \ell_i = n \quad \text{and} \quad \sum_{i=1}^{r} \binom{\ell_i}{2} \geqslant N_c.$$

Therefore if $L := \max_i \ell_i$ we have

$$\frac{L-1}{2} \geqslant \frac{N_c}{n} \quad \text{and thus} \quad L > \frac{2N_c}{n}.$$

This implies $M \leq n - \frac{2N_c}{n}$.

Therefore we have

$$\eta(\bar{E}_M, n, N_c) \leq \sum_{M < s < n - \frac{2N_c}{n}} \binom{n}{s} \binom{\binom{n}{2} - s(n-s)}{N_c}.$$

Probability of Γ_{n,N_c} belonging to the class $\bar{E}_M$

$$P(\bar{E}_M, n, N_c) = \frac{\eta(\bar{E}_M, n, N_c)}{\binom{\binom{n}{2}}{N_c}} \leq$$

$$\leq \sum_{M < s < n - \frac{2N_c}{n}} \binom{n}{s} \frac{\binom{\binom{n}{2} - s(n-s)}{N_c}}{\binom{\binom{n}{2}}{N_c}} .$$

We will use the following estimates for $n \geq n_0$

$$\binom{n}{s}\frac{\binom{\binom{n}{2}-s(n-s)}{N_c}}{\binom{\binom{n}{2}}{N_c}} \leq \frac{e^{(3-2c)s}}{s!} \quad \text{for} \quad s \leq \frac{n}{2}$$

and

$$\binom{n}{s}\frac{\binom{\binom{n}{2}-s(n-s)}{N_c}}{\binom{\binom{n}{2}}{N_c}} \leq \frac{e^{(3-2c)(n-s)}}{(n-s)!} \quad \text{for} \quad s \geq \frac{n}{2}.$$

$$P(\bar{E}_M, n, N_c) \leq \sum_{s > \frac{2N_c}{n}} \frac{e^{(3-2c)s}}{s!} + \sum_{s > M} \frac{e^{(3-2c)s}}{s!} \quad \text{for } n \geq n_0.$$

Thus

$$\lim_{n \to \infty} P(\bar{E}_{\log\log n}, n, N_c) = 0.$$

It suffices to show that

$$\lim_{n \to \infty} P(\bar{A} \cap E_{\log\log n}, n, N_c) = 0.$$

We have

$$P(\bar{A}\, E_{\log\log n}, n, N_c) \leq \frac{1}{\binom{\binom{n}{2}}{N_c}} \sum_{s=2}^{\log\log n} \binom{n}{s} \left(\sum_{r=1}^{\binom{s}{2}} \binom{\binom{s}{2}}{r} \binom{\binom{n-s}{2}}{N_c - r} \right)$$

because

if the $n-s$ points forming the greatest connected component of a graph belonging to the set $\bar{A} \cap E_{\log\log n}$ are fixed, then if r is the number of edges connecting some of s points outside this connected component we must have $r \geq 1$, and these r edges can be chosen in $\binom{\binom{s}{2}}{r}$ ways, the remaining $N_c - r$ edges can be chosen in less than $\binom{\binom{n-s}{2}}{N_c - r}$ ways.

Thus we obtain

$$P(\bar{A}\, E_{\log\log n,\, n,\, N_c}) \leq \frac{\log n}{n} \sum_{s=2}^{\log\log n} \frac{2^{\binom{s}{2}} e^{-2sc}}{s!} \leq$$

$$\leq \frac{e^{e^{-2c}} \cdot e^{\frac{1}{2}(\log\log n)^2}}{n}.$$

Therefore $\lim\limits_{n\to\infty} P(\bar{A}\, E_{\log\log n},\, n,\, N_c) = 0.$

This finishes the proof of the lemma $\blacksquare$

Proof of Theorem:

$\eta_0(n, N_c)$ – number of connected graphs in $\mathcal{G}(n, N_c)$.

$\eta_0'(n, N_c)$ – number of graphs in $\mathcal{G}(n, N_c)$ having no isolated points.

$$\eta_0'(n, N_c) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \binom{\binom{n-k}{2}}{N_c}$$

Exercise: Prove this statement

Moreover

$$\sum_{k=0}^{2s+1} (-1)^k \binom{n}{k} \binom{\binom{n-k}{2}}{N_c} \leq \eta_0' \leq \sum_{k=0}^{2s} (-1)^k \binom{n}{k} \binom{\binom{n-k}{2}}{N_c}.$$

Exercise: Use Bonferroni theorem to prove this statement

For any fixed value of k we have

$$\lim_{n\to\infty} \frac{\binom{n}{k}\binom{\binom{n-k}{2}}{N_c}}{\binom{\binom{n}{2}}{N_c}} = \frac{e^{-2kc}}{k!}.$$

Exercise: Complete this computation.

Thus we obtain

$$\lim_{n\to\infty} \frac{\mathcal{N}_0'(n,N_c)}{|\mathcal{G}(n,N_c)|} = \sum_{k=0}^{\infty} \frac{(-1)^k e^{-2kc}}{k!} = e^{-e^{-2c}}.$$

Clearly

number of graphs having no isolated points

number of connected graphs

graphs ~~not~~ of the form $\{\text{isolated pts}\} \amalg \{\text{one connected component}\}$

$$0 \leq \frac{\eta_0'(n, N_c) - \eta_0(n, N_c)}{|\mathcal{G}(n, N_c)|} \leq \frac{\eta(\bar{A}, n, N_c)}{|\mathcal{G}(n, N_c)|}.$$

By the lemma $\dfrac{\eta(\bar{A}, n, N_c)}{|\mathcal{G}(n, N_c)|} \to 0$ as $n \to \infty$.

This finishes the proof of the theorem. ∎