Random Graphs (Ch. 2 Kaminski, Pralat, Théberge)

Why study random graphs?

(i) interesting mathematical objects that have many applications in other parts of mathematics.

(ii) they can be used to model and understand real-world networks.
- the definition of the random graph sometimes explains why real-world networks behave the way they do.
  e.g. Preferential attachment ("rich get richer") helps explain the degree distribution observed in social networks.

(iii) they can be used to create synthetic networks for testing purposes. This is easier and cheaper than measuring (or modifying) real-world networks. You can also adjust their parameters to test ~~hypo~~ scenarios. (Example: What happens when the number of users in my 10,000-member social network grows to 10,000,000?)

(iv) Benchmarking & hypothesis testing: By comparing a real network to a random graph, we can learn things about the real network. E.g. A subgraph that is more dense than expected might correspond to a community.

<u>Question</u>: What is a "random" graph?

"Random graph" really means a probability space $(\Omega, \Pr)$ where Ω is a set of graphs.

<u>Erdős-Renyi</u>: $G_{n,\frac{1}{2}}$ uniform over all undirected graphs with vertex set $V(G) = \{1, \dots, n\}$.

$|\Omega| = 2^{\binom{n}{2}}$ since, for each $\{i,j\} \subseteq \binom{[n]}{2}$ there are two possibilities:

ij is ~~an~~ edge of the graph; or

ij is not an edge of the graph.

$\Pr(\omega) = \frac{1}{|\Omega|} = \frac{1}{2^{\binom{n}{2}}}$ for each $\omega \in \Omega$ [uniform distribution].

The properties of $G_{n,\frac{1}{2}}$ have been studied extensively for (at least) 70 years. It has been the source of many counterexamples to conjectures in graph theory.

For any $\{i,j\} \in \binom{[n]}{2}$, $\Pr(ij \in E(G_{n,\frac{1}{2}})) = \frac{1}{2}$, since there are $2^{\binom{n}{2}-1}$ graphs with $V(G) = [n]$ that contain the edge ij (and $2^{\binom{n}{2}-1}$ graphs that don't.).

$$\text{So } \Pr(ij \in E(G_{n,\frac{1}{2}})) = \frac{2^{\binom{n}{2}-1}}{|\Omega|} = \frac{2^{\binom{n}{2}-1}}{2^{\binom{n}{2}}} = \frac{1}{2}.$$

• For each $\{i,j\} \in \binom{[n]}{2}$ define $X_{ij} = \begin{cases} 1 & \text{if } ij \in E(G_{n,\frac{1}{2}}) \\ 0 & \text{otherwise.} \end{cases}$

$$\text{Then } \mathbb{E}(E(G_{n,\frac{1}{2}})) = \mathbb{E}\Big(\sum_{\{i,j\} \in \binom{n}{2}} X_{ij}\Big) = \sum_{\{i,j\} \in \binom{[n]}{2}} \mathbb{E}(X_{ij}) = \sum_{\{i,j\} \in \binom{[n]}{2}} \Pr(X_{ij} = 1) = \binom{n}{2}/2$$

Alternative Definition of $G_{n,\frac{1}{2}}$:

- For each $\{ij\} \subseteq \binom{[n]}{2}$ toss a coin
  - heads: include the edge ij
  - tails: don't include the edge ij

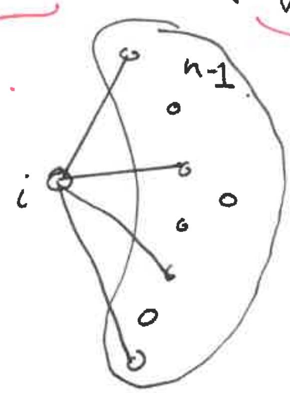
For any G with $V(G) = [n]$, $\Pr(G_{n,\frac{1}{2}} = G) = \left(\frac{1}{2}\right)^{\binom{n}{2}} = \frac{1}{2^{\binom{n}{2}}} = \frac{1}{|\Omega|}$.

$G_{n,\frac{1}{2}}$ is not a good model for most real-world networks.

First problem: It has too many edges!

$$\mathbb{E}(|E(G_{n,\frac{1}{2}})|) = \underbrace{\frac{1}{2}\cdot\binom{n}{2}}_{\text{edge density.}}$$

$$\underbrace{\mathbb{E}\left(\text{avg-degree}(G_{n,\frac{1}{2}})\right)}_{\text{average degree.}} = \mathbb{E}\left(\frac{2|E(G_{n,\frac{1}{2}})|}{n}\right) = 2\cdot\left(\frac{n(n-1)}{2}\right)\cdot\frac{1}{n} = \frac{n-1}{2}.$$



Facebook has $\gtrsim 3$ billion active users.

No Facebook users have 1.5 billion friends.

$G_{n,p}$:

$V(G_{n,p}) = [n]$.

For each $\{i,j\} \in \binom{[n]}{2}$, $\Pr(ij \in E(G_{n,p})) = p$.

~~Let~~ For each $\{i,j\} \in \binom{[n]}{2}$, let A_{ij} denote the event $ij \in E(G_{n,p})$. The events $\{A_{ij} : \{i,j\} \in \binom{[n]}{2}\}$ are mutually independent.

For each potential edge $\{i,j\} \in \binom{[n]}{2}$ we toss a biased coin to determine if $ij \in E(G_{n,p})$. All coin tosses are mutually independent.

Now: $\mathbb{E}(|E(G_{n,p})|) = \mathbb{E}\left(\sum_{\{i,j\} \in \binom{[n]}{2}} X_{ij}\right) = p \cdot \binom{n}{2}$

$$\mathbb{E}(\text{avg-degree}(G_{n,p})) = p(n-1).$$

- By using small values of p, we can control the average degree.
- [In 2026] the average number of friends of a Facebook user is ≈ 338, so we can set $\hat{p} = \frac{338}{n}$, to get

$$\mathbb{E}(\text{avg-degree}(G_{n,\hat{p}})) = 338 - \frac{338}{n} \approx 338.$$

$\frac{338}{3\text{billion}} \approx 0$

- $G_{n,p}$ is sometimes called a binomial random graph, because

  $|E(G_{n,p})|$ is a binomial$\left(p, \binom{n}{2}\right)$ random variable and, for any $i \in [n]$ $\deg_{G_{n,p}}(i)$ is a binomial$(p, n-1)$ random variable.

- Similar Model:

  $G_{n,m}$ is a graph with $V(G_{n,m}) = [n]$ chosen uniformly at random from all graphs with $V(G) = [n]$ that have exactly m edges.

  Very similar to $G_{n,p}$ when $p = \frac{m}{n^2}$

Digression: Efficient Algorithm for Generating $G_{n,m}$ (when $m \le \frac{1}{2} \cdot \binom{n}{2}$).

```
edges = ∅
While ~~edge~~ size(edges) < m :
    choose random {i,j} ∈ ([n] choose 2)
    if ij ∉ edges :
        edges ← edges ∪ {ij}
```

true with probability $1 - \frac{|edges|}{\binom{n}{2}} > 1 - \frac{m}{\binom{n}{2}} \ge 1 - \frac{1}{2} = \frac{1}{2}$.

Generates $G_{n,m}$ in $O(m)$ iterations (in expectation).

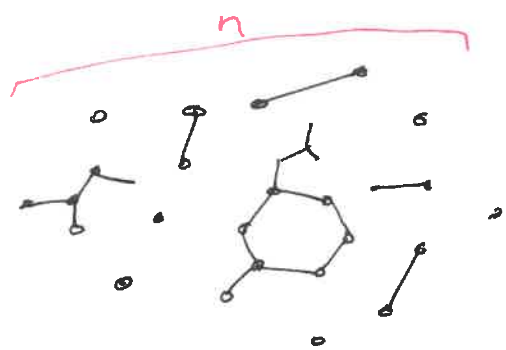
To generate $G_{n,p}$, ~~choose~~ generate a binomial$\left(p, \binom{n}{2}\right)$ random variable $\tilde{m}$ and then use the preceding algorithm to generate $G_{n,\tilde{m}}$.

Evolution of Random Graphs.

Let $e_1, e_2, \dots, e_{\binom{n}{2}}$ be a random permutation of $\binom{[n]}{2}$.

Then $G_{n,m} \simeq G$ with $V(G_m) := [n]$ and $E(G_m) := \{e_1, e_2, \dots, e_m\}$.

So $G_0, G_1, \dots, G_{\binom{n}{2}}$ is a sequence of ~~denser and~~ increasingly dense graphs.

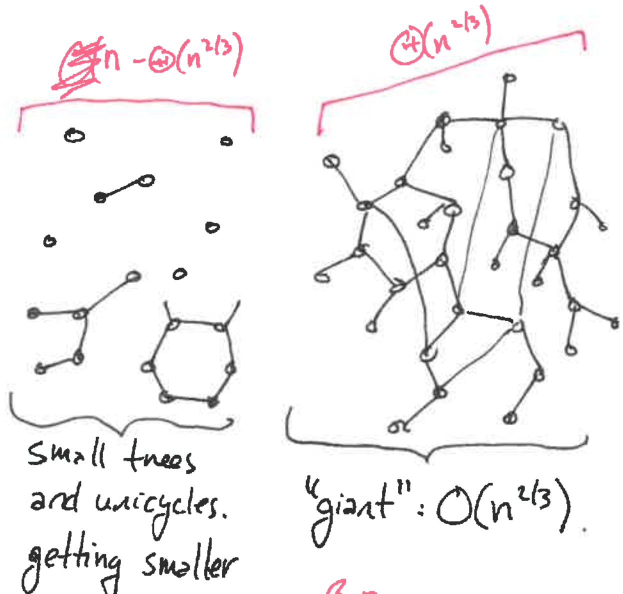


1: Subcritical Phase: $\frac{m}{n} < 1-\varepsilon$

$G_{n,m}$ consits of small trees and unicycles [of size at most $O(\log n)$].

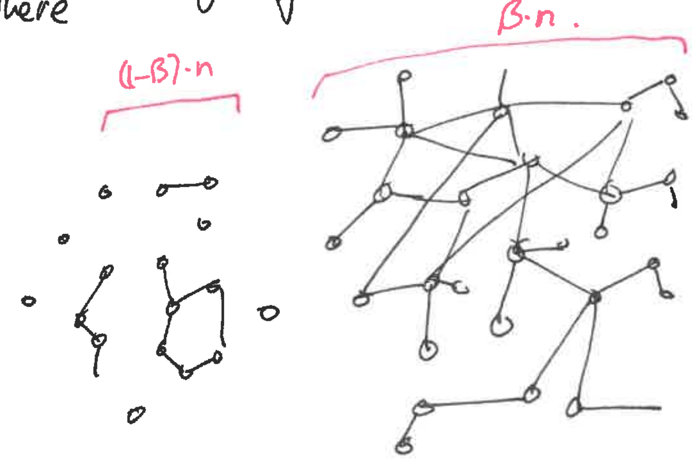
2: Critical Phase $\frac{m}{n} \simeq 1$

$G_{n,m}$ has one component of size $\Theta(n^{2/3})$ and all other componnts are small trees and unicycles.



3: (Very) Supercritical phase: $\frac{m}{n} \geq 1+\varepsilon$

Giant component has size $\approx \beta \cdot n$ where β is solution to $\beta + e^{-\beta \cdot \frac{m}{n}} = 1$



Connectivity of $G_{n,m}$. (or $G_{n,\frac{m}{n^2}}$)

Note that, for each $i \in [n]$,

$$\Pr\left(\deg_{G_{n,p}}(i) = 0\right) = (1-p)^{n-1} \approx \left(\frac{1}{e}\right)^{pn}$$

$$\mathbb{E}(\#\text{degree-zero vertices}) \approx \binom{n}{2} \cdot \left(\frac{1}{e}\right)^{pn} \approx \frac{n^2}{2} \cdot \left(\frac{1}{e}\right)^{pn} = \cancel{\frac{1}{2}\left(\frac{1}{e}\right)} \frac{1}{2} \cdot e^{2\ln n - pn}$$

Connectivity of $G_{n,p}$

Even when $p = \frac{m}{n^2} \gtrapprox \frac{(1+\varepsilon)}{n}$ we still get isolation vertices.

$$\Pr\left(\deg_{G_{n,p}}(i) = 0\right) = (1-p)^{n-1} \approx e^{-pn}.$$

[For small p, $1-p \approx e^{-p}$].

$$\mathbb{E}\begin{pmatrix}\#\text{degree-zero}\\ \text{vertices}\end{pmatrix} \approx n \cdot e^{-pn} = e^{\ln n - p \cdot n}$$

For $p = \frac{\ln n + c}{n}$,

$$\mathbb{E}\begin{pmatrix}\#\text{degree-zero}\\ \text{vertices}\end{pmatrix} = e^{\ln n - (\ln n + c)} = e^{-c}$$

So $\mathbb{E}(\#\text{degree zero}) < 1$ a soon as $pn \geq \ln n + c$.

It turns out that this is exactly when G becomes connected

For $p = \ln + c$

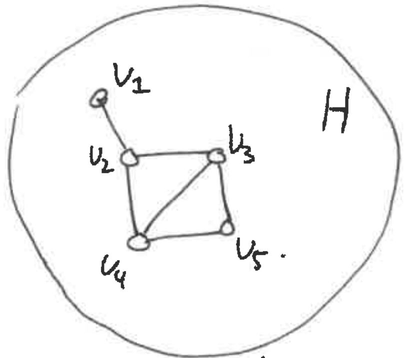
$$\Pr\left(G_{n,p} \text{ is connected}\right) \approx \begin{cases} 0 & \text{if } c \to -\infty \\ e^{-e^{-c}} & \text{if } c \in \mathbb{R} \\ 1 & \text{if } c \to +\infty \end{cases}.$$

In particular, if $c = (1+\varepsilon)\ln n$, then

$$\lim_{n\to\infty} \Pr\left(G_{n,p} \text{ is connected}\right) = 1.$$

Strange Behaviour of $G_{n,p}$.

Consider a small graph like:

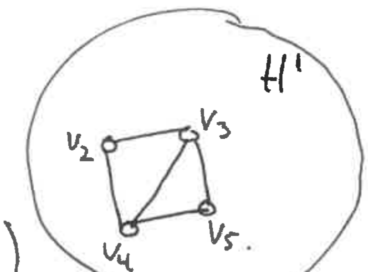


Fix 5 vertices. $v_1 \dots v_5$.

$$q := \Pr(\{v_1 \dots v_5\} \text{ form } \cancel{\text{this subgraph}}\ H.) = p^6 \cdot (1-p)^{\binom{5}{2}-6} \approx p^6 \text{ (for small } p).$$

$$\mathbb{E}\left(\begin{matrix}\text{\#occurrences}\\ \text{of } H \text{ in } G_{n,p}\end{matrix}\right) = \binom{n}{5} \cdot 5! \cdot q = \Theta(n^5 \cdot p^6).$$

Now consider this graph:



$$\mathbb{E}\left(\begin{matrix}\text{\#occurrences}\\ \text{of } H'\end{matrix}\right) = \binom{n}{4} 4!\, p^5 (1-p)^{\binom{4}{2}-5} = \Theta(n^4 p^5).$$

Look what happens when $p = \frac{1}{n^{9/11}} = n^{-9/11}$

$$\mathbb{E}\left(\begin{matrix}\text{\#occurences}\\ \text{of } H\end{matrix}\right) = \Theta(n^5 p^6) = \Theta(n^{5-\frac{54}{11}}) = \Theta(n^{\frac{55}{11}-\frac{54}{11}}) = \Theta(n^{\frac{1}{11}}) \xrightarrow[n]{} \infty$$

$$\mathbb{E}\left(\begin{matrix}\text{\#occurences}\\ \text{of } H'\end{matrix}\right) = \Theta(n^4 p^5) = \Theta(n^{4-\frac{45}{11}}) = \Theta(n^{\frac{44}{11}-\frac{45}{11}}) = \Theta(n^{-\frac{1}{11}}) \xrightarrow[n]{} 0$$

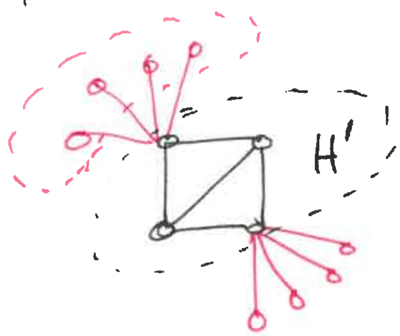
But H' is a subgraph of H, so we can't see H if we don't see H'. (!)

$\Pr(H' \text{ appears}) \to 0$, so $\Pr(H \text{ appears}) \to 0$, but $\mathbb{E}\left(\begin{matrix}\text{\#appearances}\\ \text{of } H\end{matrix}\right) \to \infty$ (!!)

W

Why does this happen?

Because H' is unlikely to appear, but when it does many copies ($\Theta(n^{2/11})$) of H appear that all share the same copy of H'!



When we study a graph like H, we need to consider its ~~most dense~~ highest average degree subgraph (H' in this case)

Here, $\text{avg-degree}(H) = \frac{6}{5} < \frac{5}{4} = \text{avg-degree}(H')$.

If H has a subgraph of ~~density~~ average-degree d and $p \cdot n^{1/d} \xrightarrow[n]{} 0$ then $\Pr(G_{n,p} \text{ contains } H) \to 0$.

Other Problems with $G_{n,p}$:

Local clustering coefficient: Edges of $G_{n,p}$ are chosen independently, so

For any $i \in [n]$, $\mathbb{E}\left(\frac{|E(G_{n,p}[N_{G_{n,p}}(i)])|}{\binom{|N_{G_{n,p}}(i)|}{2}}\right) = p = \text{edge-density}(G_{n,p}))$

(with $\deg_{G_{n,p}}(i) \geq 2$)

We saw that for social networks $c(v)$ is typically bigger than $\frac{m_G}{\binom{n}{2}} = \text{edge-density}(G)$.

Degree Distribution: The degree of each vertex in $G_{n,p}$ is a binomial(p,n) random variable. We are very unlikely to see nodes of degree much greater than $(1+\varepsilon)np$. In Facebook, ~~we would~~ (where $\frac{m}{n} \approx 336$, so $p \approx \frac{336}{n}$) we would not see anyone with ≥ 5000 friends.