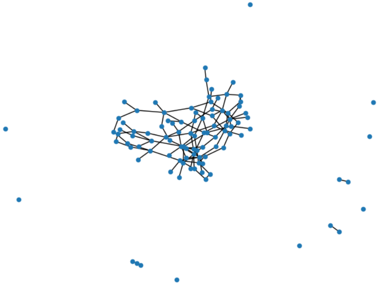
CS4423: Networks

Week 9, Part 1: Properties of the ER models

Dr Niall Madden

School of Maths, University of Galway

12+13 March 2025)



Homework Assignment 2

Homework Assignment 2 has started

- **Part 1**: A written (i.e., Python-free) assignment. You can find the details at `https://www.niallmadden.ie/2425-CS4423/`. Specifically, the questions are at `https://www.niallmadden.ie/2425-CS4423/CS4423-HW2-1.pdf`. To help you work on that, I've also prepared a "*tutorial sheet*" for Questions 5-9, which you can work on in classes this week. See `https://www.niallmadden.ie/2425-CS4423/CS4423-HW2-1-tutorial.pdf`
- **Part 2**: A programming/`networkx`-based assignment, which will be posted Thursday morning, and which are can work on next week.
- **Deadline:** 5pm. Friday, 28 March.

Questions?

Outline

This weeks notes are split between PDF slides, and a Jupyter Notebook.

Slides are at:

https://www.niallmadden.ie/2425-CS4423

Recall: the Erdös-Rényi $G_{ER}(n, m)$ model

Last week we met:

ER Model $G_{ER}(n, m)$: Uniform Random Graphs

Let $n \geq 1$, let $N = \binom{n}{2}$ and let $0 \leq m \leq N$.
The model $G_{ER}(n, m)$ consists of the ensemble of graphs G on the n nodes $X = \{0, 1, \ldots, n-1\}$, and m randomly selected edges, chosen uniformly from the $N = \binom{n}{2}$ possible edges.

Equivalently, one can choose uniformly at random one network in the **set** $\mathcal{G}(n, m)$ of *all* networks on a given set of n nodes with *exactly* m edges.

Recall: the Erdös-Rényi $G_{ER}(n, m)$ model

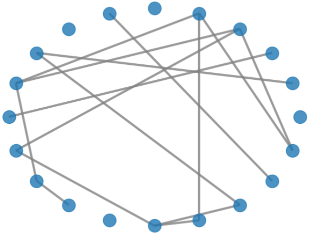
One could think of $G(n, m)$ as a probability distribution $P\colon G(n, m) \to \mathbb{R}$, that assigns to each network $G \in G(n, m)$ the same probability

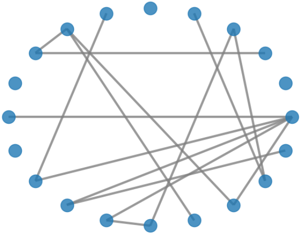
$$P(G) = \binom{N}{m}^{-1},$$

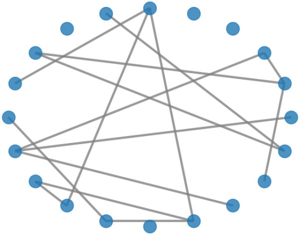
where $N = \binom{n}{2}$.

Recall: the Erdös-Rényi $G_{ER}(n, m)$ model

Some networks drawn from $G_{ER}(20, 15)$.







Model B: $G_{ER}(n, p)$

Erdös-Rényi: Randomly selected edges

ER Model $G_{ER}(n, p)$: Random Edges

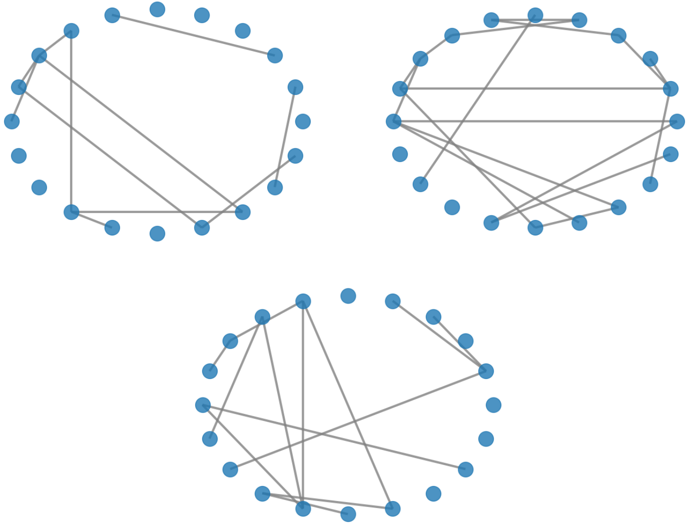
Let $n \geq 1$, let $N = \binom{n}{2}$ and let $0 \leq p \leq 1$.
The model $G_{ER}(n, p)$ consists of the ensemble of graphs G on the n nodes $X = \{0, 1, \ldots, n-1\}$, with each of the possible $N = \binom{n}{2}$ edges chosen with probability p.

The probability $P(G)$ of a particular graph $G = (X, E)$ with $X = \{0, 1, \ldots, n-1\}$ and $m = |E|$ edges in the $G_{ER}(n, p)$ model is

$$P(G) = p^m (1 - p)^{N-m}.$$

Model B: $G_{ER}(n, p)$

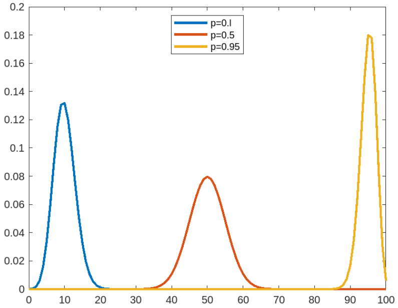
Some networks drawn from $G_{ER}(20, 0.5)$.



Model B: $G_{ER}(n, p)$

Of the two models, $G_{ER}(n, p)$ is the more studied. They are many similarities, but they do differ. For example:

1. $G_{ER}(n, m)$ will have m edges with probability 1.
2. A graph in $G_{ER}(n, p)$ with have m edges with probability $\binom{N}{m} p^m (p-1)^{N-m}$.



Properties

We'd like to investigate (theoretically and computationally) the properties of such graphs. For example

- When might it be a tree?
- Does it contain a tree, or other cycles? If so, how many?
- When does it contain a small complete graph?
- When does it contain a **large component**, larger than all other components?
- When does the network form a single **connected component**?
- How do these properties depend on n and m (or p)?

Denote by $\mathcal{G}_n$ the set of *all* graphs on the n points $X = \{0, \ldots, n-1\}$.

Set $N = \binom{n}{2}$, the maximal number of edges of a graph $G \in \mathcal{G}_n$.

Regard the ER models A and B as **probability distributions** $P \colon \mathcal{G}_n \to \mathbb{R}$.

Notation: m Denote $m(G)$: the number of edges of a graph G.

As we have seen, the probability of a specific graph G to be sampled from the model $G(n, m)$ is:

$$P(G) = \begin{cases} \binom{N}{m}^{-1}, & \text{if } m(G) = m, \\ 0, & \text{else.} \end{cases}$$

And the probability of a specific graph G to be sampled from the model $G(n, p)$ is:

$$P(G) = p^m(1-p)^{N-m}$$

Expected size and average degree

Let's use the following notation:

- $\bar{a}$ is the expected value of property a (that is, as the graphs vary across the ensemble produced by the model).
- $\langle a \rangle$ is the average of property a over all the nodes of a graph.

In $G(n, m)$, the expected **size** is

$$\bar{m} = m,$$

as every graph G in $G(n, m)$ has exactly m edges. The expected **average degree** is

$$\langle k \rangle = \frac{2m}{n},$$

as every graph has average degree $2m/n$.

Other properties of $G(n, m)$ are less straightforward, and it is easier to work with the $G(n, p)$.

In $G(n,p)$, with $N = \binom{n}{2}$,

- the **expected size** is
$$\bar{m} = pN$$
(Also: variance is $\sigma_m^2 = Np(1-p)$).
- the expected **average degree** is (we'll see why soon):
$$\langle k \rangle = p(n-1).$$
with standard deviation $\sigma_k = \sqrt{p(1-p)(n-1)}$.
- In particular, the *relative standard deviation* of the size of a random model B graph is
$$\frac{\sigma_m}{\bar{m}} = \sqrt{\frac{1-p}{pN}} = \sqrt{\frac{2(1-p)}{pn(n-1)}} = \sqrt{\frac{2}{n\langle k \rangle} - \frac{2}{n(n-1)}},$$
a quantity that converges to 0 as $n \to \infty$ if $p(n-1) = \langle k \rangle$, the average node degree, is kept constant.

Degree Distribution

Definition (Degree distribution)

The **degree distribution** $p\colon \mathbb{N}_0 \to \mathbb{R}$, $k \mapsto p_k$ of a graph G is defined as

$$p_k = \frac{n_k}{n},$$

where, for $k \geq 0$, n_k is the number of nodes of degree k in G.

This definition can be extended to ensembles of graphs with n nodes (like the random graphs $G(n, m)$ and $G(n, p)$), by setting

$$p_k = \bar{n}_k / n,$$

where $\bar{n}_k$ denotes the expected value of the random variable n_k over the ensemble of graphs.

Degree Distribution

The degree distribution in a random graph $G(n, p)$ is a **binomial distribution**

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k} = \mathrm{Bin}(n-1, p, k)$$

That is, in the $G(n, p)$ model, the *probability that a node has degree* k *is* p_k.

Also, the *average degree* of a randomly chosen node is

$$\langle k \rangle = \sum_{k=0}^{n-1} k p_k = p(n-1)$$

(with standard deviation $\sigma_k = \sqrt{p(1-p)(n-1)}$).

Example (Q3(c) from 2023/24 exam)

Suppose one constructed a graph G on 120 nodes by tossing a (fair, 6-sided) die once for each possible edge, adding the edge only if the die shows 3 or 6. Then pick a node at random in this graph. What is the probability that this node has degree 50? (You do not need to return a numerical value. It is enough to give an explicit formula in terms of the given data.)

In general, it is not so easy to compute

$$\binom{n-1}{k} p^k (1-p)^{n-1-k}$$

However, in the limit $n \to \infty$, with $\langle k \rangle = p(n-1)$ kept constant, the binomial distribution $\mathrm{Bin}(n-1, p, k)$ is well approximated by the **Poisson distribution**

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!} = \mathrm{Pois}(\lambda, k),$$

where $\lambda = p(n-1)$.