Random graph models

References The best reference for this chapter that fits very well into this course mindset is: Barabási et al. (2016). Note that an online and interactive version is available at: `http://networksciencebook.com/`. Other good references for this chapter are (by order of relevance): Hofstad (2016), Durrett (2007), and Chung and Lu (2006). Other traditional books (but heavier on the maths side) written by "big names" of the field are Janson et al. (2011) and Bollobás (1998).

1.1 Some reminders

1.1.1 Probability toolbox

Proposition 1.1.1.

- *For a, b constants, $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$*
- $\mathbb{E}(X_1 + \cdots + X_m) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_m)$*;*
- *Let X be a r.v., A an event and $1_A(X)$ the indicator that event is realized by X. Then:*

$$\mathbb{E}\big(1_A(X)\big) = \Pr(X \in A).$$

Definition 1.1.2. $\mathrm{Var}(X) = E\Big(\big(X - \mathbb{E}(X)\big)^2\Big)$.

Proposition 1.1.3.

- $\mathrm{Var}(X) = \mathbb{E}(X^2) - (E(X))^2$*;*
- *For a, b constants, $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$*
- *If $X_1, \ldots, X_m$ are mutually independent r.v., then $\mathrm{Var}(X_1 + \cdots + X_m) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_m)$.*
- *If we don't have this independence, then $\mathrm{Var}(X_1 + \cdots + X_m) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_m) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$, where $\mathrm{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)$.*

1.1.2 Basic probability laws

Definition 1.1.4. Let X be a random variable. We say X is generated around a Bernoulli law of parameter $p \in [0; 1]$, and denote $X \sim \mathrm{Ber}(p)$ if:

1. X takes values in $\{0; 1\}$ (almost surely);
2. $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p$.

Example 1.1.5. A $\mathrm{Ber}(p)$ model the result when we biased coin toss (p is the probability of winning the coin toss).

Proposition 1.1.6. *Let $X \sim \mathrm{Ber}(p)$. We have $\mathbb{E}X = p$ and $\operatorname{Var} X = p(1-p)$.*

Definition 1.1.7. The binomial distribution with parameters and p, denoted $\mathrm{Bin}(n,p)$, is the discrete probability distribution of the number of successes in a sequence of n independent Bernoulli trials of parameters p.

Proposition 1.1.8. *If $(X_i)_{i=1,\dots,n}$ is a sequence of n i.i.d. random variable distributed according to $\mathrm{Ber}(p)$, then $\sum_i X_i \sim \mathrm{Bin}(n,p)$.*

Corollary 1.1.9. *Let $X \sim \mathrm{Bin}(n,p)$. Then: $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. Moreover, $\mathbb{E}X = np$ and $\operatorname{Var} X = np(1-p)$.*

Definition 1.1.10. The geometric distribution of parameter p, denoted $\mathrm{Geo}(p)$ is the probability of the number of Bernoulli trials (of parameter p) needed to get one success. In particular, if $X \sim \mathrm{Geo}(p)$, then $\Pr(X = k) = (1-p)^{k-1}p$.

Exercise 1.1.11. Let $X \sim \mathrm{Geo}(p)$. Show that $\mathbb{E}X = \dfrac{1}{p}$ and $\operatorname{Var} p = \dfrac{1-p}{p^2}$.

1.1.3 Concentration of random variables

First moment inequalities

Proposition 1.1.12 (Markov's inequality)**.** *Let X be a random variable, with positive values, and $a \in \mathbb{R}$. We have:*

$$\Pr\big(X \geq a\big) \leq \frac{\mathbb{E}X}{a}.$$

Proof.

$$\mathbb{E}X \geq \mathbb{E}\Big(X 1_{X \geq a}\Big) \geq a\mathbb{E}\Big(1_{X \geq a}\Big) = a \Pr\Big(X \geq a\Big).$$

□

Remark 1.1.13. By letting $a = t\mathbb{E}X$, we obtain $\Pr(X \geq t\mathbb{E}X) \leq \dfrac{1}{t}$. The convergence speed in $1/t$ is rather slow, and depending on the situations might not be strong enough.

Corollary 1.1.14 (First moment method)**.** *Let X be a positive, integer valued random variable. We have:*

$$\Pr(X \neq 0) \leq \mathbb{E}(X).$$

The first moment is an upper bound on the probability that a integer random variable is not equal to zero.

Proof. Since X is integer valued, we have $\Pr(X \neq 0) = \Pr(X > 0) = \Pr(X \geq 1)$, and from there we can use Markov inequality. □

Application 1.1.15 (Union bound)**.** Let $A_1, \dots, A_m$ be a collection of events. Then,

$$\Pr(A_1 \cup \cdots \cup A_n) \leq \sum_{i=1}^{m} \Pr(A_i).$$

This can be shown by using the first moment method on $X = \sum_{i=1}^{m} 1_{A_i}$ and observe that $\{X > 0\} = A_1 \cup \cdots \cup A_m$.

Remark 1.1.16. The first moment method is generally used when we have a sequence of integer, positive r.v. X_n such that $\mathbb{E}X_n \to 0$. In that case, $X_n \to 0$ almost surely.

Remark 1.1.17. We could think that, if $\mathbb{E}X_n \to +\infty$, then $\Pr(X_n > 0) \to 1$, but this isn't true, as the next example shows.

Example 1.1.18. Let's take X_n such that $X_n = n^2$ with probability $1/n$ and $X_n = 0$ otherwise. Then $\mathbb{E}(X_n) = n \to +\infty$ but $X_n \to 0$. Loosely speaking, this can happen because the variance of X is very large.

Second moment inequalities

Proposition 1.1.19 (Chebyshev's inequality)**.** *Let X be a random variable, and $a > 0$. We have:*

$$\Pr\Big(\big|X - \mathbb{E}X\big| \geq a\Big) \leq \frac{\operatorname{Var} X}{a^2}.$$

Proof. Markov inequality to $Y = (X - \mathbb{E}X)^2$. □

Example 1.1.20. Let X be Gaussian $\mathcal{N}(0, \sigma^2)$. Then $\mathbb{E}|X| = \sigma\sqrt{\frac{2}{\pi}}$. Markov applied to $|X|$ gives us:

$$Pr(X \geq a) \leq \sqrt{\frac{2}{\pi}}\frac{\sigma}{b},$$

while Chebyshev gives:

$$\Pr(X \geq a) \leq \left(\frac{\sigma}{a}\right)^2.$$

If a is large, Chebyshev gives a stronger bound.

Application 1.1.21 (Weak law of Large Numbers)**.** $X_1, \dots X_n$ be independent r.v. with mean μ and variance $\sigma^2 < +\infty$. Then:

$$\Pr\left(\Big|\frac{X_1 + \cdots + X_n}{n} - \mu\Big| > \epsilon\right) \to 0.$$

Actually, with extra work, the condition $\sigma^2 < +\infty$ isn't needed. With lot of extra work, we can show that the convergence is almost surely (and not in probability, as we have here).

Proof. Chebychev at $U_n := \dfrac{X_1 + \cdots + X_n}{n}$, which has a mean μ and variance σ^2 gives:

$$\Pr\big(|U_n| \geq \epsilon\big) \leq \frac{\sigma^2}{n\epsilon^2} \to 0$$

□

Corollary 1.1.22 (Second moment method)**.** *Let X be a positive random variable. We have:*

$$\Pr\big(X=0\big) \leq \frac{\operatorname{Var} X}{\big(\mathbb{E}X\big)^2} = \frac{\mathbb{E}(X^2)}{(\mathbb{E}X^2)} - 1.$$

Proof. Chebychev with $a = \mathbb{E}X$.

$$\Pr\big(X=0\big) \leq \Pr\Big(\big|X-\mathbb{E}X\big| \geq \mathbb{E}X\Big) \leq \frac{\operatorname{Var} X}{\big(\mathbb{E}X\big)^2}.$$

The first inequality comes from $\big|X-\mathbb{E}X\big| \geq \mathbb{E}X \Rightarrow X \leq 0$ or $X \geq 2\mathbb{E}X$. □

Concentration sum of i.i.d. random variables

Proposition 1.1.23 (Hoeffding's inequality)**.** *Let X_i be some independent random values, such that $a_i \leq X_i \leq b_i$, and $S_n = \sum_{i=1}^n X_i$. For $t > 0$, we have:*

$$\Pr\Big(S_n \geq \mathbb{E}S_n + t\Big) \leq \exp\Big(-\frac{2t^2}{\sum_i (b_i-a_i)^2}\Big),$$
$$\Pr\Big(S_n \geq \mathbb{E}S_n - t\Big) \leq \exp\Big(-\frac{2t^2}{\sum_i (b_i-a_i)^2}\Big),$$
$$\Pr\Big(|S_n - \mathbb{E}S_n| \geq t\Big) \leq 2\exp\Big(-\frac{2t^2}{\sum_i (b_i-a_i)^2}\Big).$$

More details about concentration inequalities can be found in the chapter 2 of Vershynin (2018).

1.1.4 Exercises

Exercise 1.1.24. Let throw m balls into n bins. What can we say about the probability that one bin is empty:

1. If $m = (1+\epsilon)n\log n$;
2. If $m = (1-\epsilon)n\log n$.

Exercise 1.1.25.

1. What's the 'typical' position of a simple random walk after n steps?
2. What's the longest run of Heads in n flips of a fair coin?
3. What is the maximum of n independent standard normal RV's?

1.1.5 Graph Theory

Definition, vocabulary

Definition 1.1.26. A graph G is a pair (V, E), where V is a (finite) set, whose elements are called nodes (or vertices, or points) and E is a set of ordered node pairs called edges (or links, lines, bonds). Some vocabulary:

- If $(ij) \in E \iff (ji) \in E$, the graph is said undirected. (It means that if there is a link going from i to j, there exists the same link in opposite direction).

- If for all nodes i, $(ii) \notin E$, we say there is no self-loops.

Definition 1.1.27.
- We call path of G of length k a sequence $e_1, \ldots, e_k$ of edges $e_i = (v_{i-1}, v_i)$ where the v_i are distinct vertices.

- A k-cycle is a path of legnth k that starts and ends at the same vertex.

- Let suppose G is undirected. We say that two nodes u, v are connected if there exists a path going from u to v. We note $u \leftrightarrow v$.

Proposition 1.1.28. *$\leftrightarrow$ is an equivalence relationship. In particular, we can partition the nodes into equivalent classes, that we will call connected components.*

Proof. We have $u \leftrightarrow u$ (path of length 0), if $u \leftrightarrow v$ and $v \leftrightarrow z$, then $u \leftrightarrow z$ (by combining the two paths; transitivity). Finally, $u \leftrightarrow v$ implies $v \leftrightarrow u$ (the same path, on the opposite direction; symmetry). □

Definition 1.1.29. If G has only one equivalent class under the relation $\leftrightarrow$, we say G is connected. We say G is disconnected otherwise.

In particular, if a graph is connected, it means for every node i and j, there exists a path going from i to j.

Definition 1.1.30. Let i, j be two nodes. We call the distance between i and j, and denote $d(i, j)$ the length of the shortest path between i and j. If $i \not\leftrightarrow j$, then $d(i, j) := +\infty$.

Definition 1.1.31 (Diameter)**.** We call diameter of a graph the largest distance between any pair of connected vertices.

Graph Theory and Linear Algebra

Definition 1.1.32. We call adjacency matrix (denoted by A), the binary matrix such that $A_{ij} = 1$ iff $(ij) \in E$.

Proposition 1.1.33. *A is symmetric if and only if the graph is un-directed. Moreover, the diagonals elements of A are zeros iff the graph doesn't have self-loops.*

We can easily extend those definition to weighted graph (where the edges bear some weight).

Definition 1.1.34. We call degree matrix, denoted D, of a graph G the diagonal matrix whose element d_{ii} is the degree of node i.

Definition 1.1.35. We call Laplacian the matrix $L := D - A$.

1.1.6 Random Graph: vocabulary and notations

In the following notes, $G = (V, E)$ will denote a graph, where $V = \{1, \dots, n\}$ is the set of vertices (nodes) and E the set of edges.

When we say G is a random graph, we actually mean that G is a graph that was generated along some probability distribution. It is a slight abuse of notation, as once G is given, it is not random.

We will call $\mathbf{d} := (d_1, \dots, d_n)$ the degree sequence of nodes $1, \dots, n$.

Since G is random, the degree d_i of a node i is actually a random variable, which can be distributed along some probability distribution (binomial, etc.). When all the nodes degree are identically distributed (*i.e.,* $d_1, \dots, d_n$ are all distributed along the same probability distribution $\mathcal{D}$), we say that the degrees in the graph G are distributed along that degree distribution $\mathcal{D}$.

Most graph observed in nature exhibits a power law degree distribution Clauset et al. (2009).

Loosely speaking, the power law means that $P(d = k) := p_k \propto k^{-\tau}$. In particular, if we note $\Pr(X = k) = p_k$, we have $\log p_k = -\tau \log k + C$ for some constant C. In a log/log plot, we get a straight line of slope $-\tau$.

Definition 1.1.36 (Different power laws)**.** Let X be a random variable.

- X is said to be distributed according to a Zeta distribution of exponent τ if X takes integer values, and $\forall k \in \mathbb{N} : \Pr(X = k) = C_\tau k^{-\tau}$. Here C_τ is the normalization factor, equal to $\left(\sum_{k=1}^{\infty} k^{-\tau}\right)^{-1} = \dfrac{1}{\zeta(\tau)}$ (where ζ is the Riemann function).

- X is said to be distributed according to a Zipf law of parameter n and exponent τ if $\Pr(X = k) = C_{n,\tau} k^{-tau}$

- X is said to be distributed according to a (continuous) power law if X takes values in $[x_{min}; +\infty]$ where $x_{min} > 0$ and X has a density $f(x) = Cx^{-\tau}$

Exercise 1.1.37. Let X be a r.v. following a continuous power law distribution. Show that:

- X has a mean iff $\tau > 2$, and in that case $\mathbb{E}X = \dfrac{\tau - 1}{\tau - 2} x_{min}$;

- X has a second moment iff $\tau > 3$, and in that case $\mathbb{E}X^2 = \dfrac{\tau - 1}{\tau - 3} x_{min}^2$.

What would it be for a zeta distribution ?

Proposition 1.1.38. *Let $f(x) = ax^{-\tau}$ the density of a power law. Then f is scale invariant, that is to say $f(cx) \propto f(x)$ for any constant c.*

Because of this proposition, graphs whose degree distribution follow a power law are said to be *scale-free* (or *scale invariant*).

Small World property

Many real networks present the *small world property*, i.e., two nodes are not too far apart of each other (given the graph distance 1.1.30).

Definition 1.1.39. Let $(G_n)_{n\in\mathbb{N}}$ be a sequence of random graphs (G_n having n nodes), and let H_n be the distance among two connected nodes of G_n chosen uniformly at random.

Now what typical value should use to say that H_n is small or not?

Exercise 1.1.40.
- Consider the linear graph. What's the maximum distance between two nodes ? The typical distance H_n ?
- Consider the circular graph. Same questions.
- Same questions with the complete graph.
- The nearest-neighbors torus of width r and dimension d ?

Definition 1.1.41. We say that $(G_n)_{n\in\mathbb{N}}$ is a *small world* if there exists a constant $K < \infty$ such that:

$$\lim_{n\to+\infty} \Pr(H_n \leq K \log n) = 1$$

Moreover, we say that $(G_n)_{n\in\mathbb{N}}$ is *ultra small world* when, for every $\epsilon > 0$

$$\lim_{n\to+\infty} \Pr(H_n \leq \epsilon \log n) = 1$$

1.1.7 Miscellaneous

Lemma 1.1.42. $\left(1 - \frac{\lambda_n}{n}\right)^n \exp(\lambda_n) = \exp(o(\lambda_n))$.

1.2 Erdős-Rényi random graphs

1.2.1 Bernoulli random graphs

Definition 1.2.1. Let $P = (p_{ij})_{i,j\in\{1,\dots,n\}} \in [0;1]^{n\times n}$ be a symmetric matrix. A Bernoulli random graph G is a (undirected, unweighted) graph G where the edges (ij) are generated independently such that $\Pr\big((ij) \in E\big) = p_{ij}$. In that case, we note $G \sim G(n, (p_{ij}))$.

Remark 1.2.2. If $G \sim G(n, (p_{ij}))$ then the adjacency matrix of A is a symmetric random matrix, whose entries are independently distributed, and $A_{ij} \sim \text{Ber}(p_{ij})$.

Proposition 1.2.3. *Let $G \sim G(n, (p_{ij}))$. We have:*

$$\Pr(G) = \prod_{i<j} p_{ij}^{A_{ij}} (1 - p_{ij})^{1-A_{ij}}.$$

Example 1.2.4. Suppose that $\forall i, j : p_{ij} = p$. Then, $G(n, (p_{ij}))$ is called the Erdős-Rényi model[1], and traditionally denoted $G(n, p)$ or $G_{n,p}$.

Example 1.2.5. Assume the n nodes are separated into K distinct, non-overlapping communities, *i.e.*, there exists $\sigma : V \to \{1, \dots, K\}$, such that $\sigma(i)$ denotes the community of node i.

Assume that $p_{ij} = q_{\sigma_i, \sigma_j}$ (this means that the probability of observing an edge between i and j depends only on the community of node i and j).

Then in that case, $G(n, (p_{ij}))$ is called the Stochastic Block Model (SBM), and denoted $SBM(n, \sigma, Q)$. More precisely, $\Pr\Big((ij) \in E\Big) = q_{\sigma_i \sigma_j}$ ony depends on the community assignment of nodes i and j.

Moreover, $\Pr(G|\sigma) = \prod_{i<j} q_{\sigma_i \sigma_j}^{A_{ij}} (1 - q_{ij})^{1 - A_{ij}}$.

Remark 1.2.6. The adjacency matrix of SBM(n,Q,K) can be seen as a block matrix, whose blocks are ER graphs.

1.2.2 Degree distribution

Proposition 1.2.7. *Let $G \sim G(n, p)$, and let d_i be the degree of node i. Then, the sequence $(d_1, \dots, d_n)$ is i.i.d., and the d_i's are distributed according to* $\mathrm{Bin}(n, p)$.

Proof. Indeed, the degree of i, denoted d_i, is equal to $\sum_{j=1}^{n} A_{ij}$, where A_{ij} are i.i.d. Bernoulli random variable, of parameter p. Since the elements $(A_{ij})_{i<j}$ are independent, so are the d_i's. □

Proposition 1.2.8. *Let G be a homogeneous SBM graph, i.e., two communities of equal size $\frac{n}{2}$, with probabilities p_{in} and p_{out} of forming intra and inter-communities edge. Then, $d_i \sim \mathrm{Bin}\left(\frac{n}{2}, p_{in}\right) + \mathrm{Bin}\left(\frac{n}{2}, p_{out}\right)$ and:*

$$\mathbb{E}\, d_i = \frac{p_{in} + p_{out}}{2}.$$

Proof. Similar. □

Remark 1.2.9. It has been observed that many real graphs have a power law degree distribution, and not a binomial one. See for example Clauset et al. (2009). An intuitive explanation is that since binomial distributions are well concentrated, a Erdős-Rényi graph does not allows for many hubs (degree much higher than the average), which we tend to see in real networks (think of a social network).

1.2.3 Phase transition phenomena

Theorem 1.2.10 (Phase transition for giant component – constant degree regime)**.** *Let $G \sim G(n, p_n)$ be a Erdős-Rényi graph, with $d_n = np_n$ is a constant. We have:*

(a) If $d_n < 1$: a.s no connected component of size larger than $O(\log n)$;

(b) If $d_n = 1$: a.s there is one large component of size $O(n^{2/3})$;

(c) If $d_n > 1$ (d constant) : one giant component of size $O(n)$.

[1] But was introduced by Gilbert, in 1959.

Theorem 1.2.11 (Phase transition for connectivity – logarithmic degree regime)**.** *Let $G \sim G(n, p_n)$ be a Erdős-Rényi random graph, with $d_n = np_n$. We have:*

(a) If there exists $\omega_n \to +\infty$ such that $d_n < \log n - \omega_n$, then G is a.s. non connected (we in fact have a bit stronger: the graph contains a.s. isolated nodes);

(b) If there exists $\omega_n \to +\infty$ such that $d_n > \log n + \omega_n$, then G is a.s. connected.

We also say that the function $t(n) = \log n$ is a threshold function (of the degree) for the property "the graph is connected".

Example 1.2.12. Assume $d_n = \log n + \log\log n$, and let $G_n \sim G(n, p_n)$. Asymptotically, will G be connected ? Same question for $d_n = 1.00001 \log n$, $d_n = 0.9999 \log n$ and $d_n = \log n + 25$.

1.2.4 Beginning of a proof

Theorem 1.2.13 (Isolated nodes)**.** *The probability that a Erdős-Rényi graph G_n on node set $[n]$ with link probability p_n contains isolated nodes satisfies*

$$\Pr\big(\exists \text{ isolated node}\big) \to \begin{cases} 0 & \text{if} \quad p_n \geq \dfrac{\log n + \omega_n}{n} \text{ for some } \omega_n \to +\infty \\ 1 & \text{if} \quad p_n \leq \dfrac{\log n - \omega_n}{n} \text{ for some } \omega_n \to +\infty. \end{cases} \tag{1.2.1}$$

In particular, this result implies that if $p_n \leq \dfrac{\log n + \omega_n}{n}$, then the graph is a.s. not connected. This correspond to point (a) of Theorem 1.2.11.

Proof. To shorten notations, let $p_n^{\pm} := \dfrac{\log n \pm \omega_n}{n}$ (we deal with both case at once).

Let A_i be the event "node i is isolated", and let $I_n := \sum_{i=0}^{n} 1(A_i)$ be the number of isolated nodes. Recall we note $d_n = np_n$ the mean degree. We have:

$$\begin{aligned} \Pr(A_i) &= (1 - p_n)^n \\ &= \Big(1 - \frac{d_n}{n}\Big)^n \\ &\sim \exp\big(-d_n\big) \\ &\sim \frac{1}{n} \exp(\mp\omega_n) \end{aligned}$$

and thus

$$\begin{aligned} \mathbb{E}\big(I_n\big) &= \sum_{i=0}^{n} \Pr(A_i) \\ &= n \Pr(A_1) \\ &\sim \mathrm{e}^{\mp\omega_n}. \end{aligned}$$

So,

1. If $d_n = \log n + \omega_n$, we have $\mathbb{E}\big(I_n\big) \sim \mathrm{e}^{-\omega_n} \to 0$. Since the expected number of isolated nodes goes to 0, this imply (a) by the first moment method.

2. If $d_n = \log n - \omega_n$, we have $\mathbb{E}\big(I_n\big) \sim \mathrm{e}^{+\omega_n} \to +\infty$. the expected number of isolated nodes goes to infinity. Unfortunately, this isn't enough to imply (b), and we will need the second moment method.

Recall (a) is straightforward using Markov inequality. Assume $d_n = \log n - \omega_n$. Then:

$$\begin{aligned}\Pr\big(\exists \text{ isolated node}\big) &= \Pr(I_n \geq 1\big) \\ &\leq \frac{\mathbb{E}I_n}{1} \to 0.\end{aligned}$$

Now assume $d_n = \log n - \omega_n$. To get (b), we have to show that the random variable I_n is well-concentrated around its mean. Since its means diverges to infinity, the result will follow. For that, we will use Chebyshev's inequality (second moment method). We have:

$$\mathrm{Var}(I_n) = \mathbb{E}\big(I_n^2\big) - \big(\mathbb{E}I_n\big)^2.$$

Note that

$$\begin{aligned}\mathbb{E}\big(I_n^2\big) &= \mathbb{E}\Big(\sum_i \sum_j 1(A_i)1(A_j)\Big) \\ &= \sum_i \sum_j \Pr\big(A_i, A_j\big) \\ &= n \Pr\big(A_1\big) + n(n-1) \Pr\big(A_1, A_2\big).\end{aligned}$$

Here we need to be careful, since A_1 and A_2 are not independent. Indeed, knowing that node 1 is isolated means that there is no edge between 1 and 2, and thus increases a bit (weakly) the probability that 2 is isolated. We have:

$$\begin{aligned}\Pr\big(A_1, A_2\big) &= \Pr\big(A_1 \,|\, A_2\big) \Pr\big(A_2\big) \\ &= (1-p_n)^{n-1}(1-p_n)^n \\ &= (1-p_n)^{n-1} \Pr\big(A_1\big) \\ &= \frac{1}{1-p_n} \Pr\big(A_1\big),\end{aligned}$$

since $\Pr\big(A_1\big) = (1-p_n)^n$. Lastly,

$$\begin{aligned}\big(\mathbb{E}I_n\big)^2 &= \Big(\sum_i \Pr(A_i)\Big)^2 \\ &= \sum_i \sum_j \Pr(A_i) \Pr(A_j) \\ &= \sum_i \sum_j \Pr(A_1)^2 \\ &= n^2 \Pr(A_1)^2.\end{aligned}$$

By putting all pieces together, it leads to:

$$\begin{aligned}
\mathrm{Var}(I_n) &= n \Pr\left(A_1\right) + n(n-1) \Pr\left(A_1 \cup A_2\right) - n^2 \Pr(A_1)^2 \\
&= n \Pr\left(A_1\right) + \frac{n(n-1)}{1-p_n} \Pr\left(A_1\right)^2 - n^2 \Pr\left(A_1\right)^2 \\
&\leq n \Pr\left(A_1\right) + n^2 \Pr\left(A_1\right)^2 \frac{1}{1-p_n} - n^2 \Pr\left(A_1\right)^2 \\
&= n \Pr\left(A_1\right) + n^2 \Pr\left(A_1\right)^2 \Big(\frac{1}{1-p_n} - 1\Big) \\
&= \mathbb{E}(I_n) + \big(\mathbb{E}I_n\big)^2 \frac{p_n}{1-p_n}.
\end{aligned}$$

Thus:

$$\begin{aligned}
\Pr(I_n = 0) &\leq \frac{\mathrm{Var}(I_n)}{\Big(\mathbb{E}(I_n)\Big)^2} \\
&\leq \frac{1}{\mathbb{E}(I_n)} + \frac{p_n}{1-p_n}
\end{aligned}$$

and this last quantity goes to zero when n goes to infinity (when $p_n = (1-\epsilon)\log n$, we already saw that $\mathbb{E}I_n \to \infty$). □

1.3 Other models

1.3.1 Configuration model

Consider a sequence $d = (d_1, \ldots, d_n)$. We aim to construct a graph with n vertices, where node i has a degree d_i. Few remarks:

- We can suppose $d_i \geq 1$, as $d_i = 0$ means the node i is isolated and we can remove it.
- It is not obvious that there should exist a graph whose degree are given by a fixed sequence. In fact, such a graph does not necessarily exist. For example, if we assume the graph unweighted, then $\sum_{i=1} d_i$ should be even (since this sum corresponds to counting two times the edges).
- Even by adding the restraint $\sum_{i=1} d_i$, constructing such a graph isn't always possible.
- To avoid those issues, we will allow self loops and multi-edges.

Example 1.3.1. If $d_1 = \cdots = d_n = d$, then we obtain a random d-regular graph (*i.e.*, a graph for which all node have the same degree d).

Example 1.3.2. If the d_i follows a $\mathrm{Bin}(n, p)$, then when $n \to +\infty$, we recover a Erdős-Rényi graph.

Algorithm 1.3.3 (Configuration Model). *Let $\boldsymbol{d} := (d_1, \dots, d_n)$ a sequence such that $\ell_n := \sum_{i=1}^n d_i$ is even.*

At each node $i \in \{1, \dots, n\}$, we attach d_i half-edges. We thus have ℓ_n half-edges, that we can number in an arbitrary order. We pair the first half edge with another one, chosen uniformly at random with the $\ell_n - 1$: this pair gives us our first edge.

We then iterate the procedure, until all the half edges are connected.

The resulting graph is called the configuration model with degree sequence $\boldsymbol{d}$, abbreviated in $\mathrm{CM}_n(\boldsymbol{d})$.

Remark:

- As discussed ealier, this model allow multi-edges and self-loops.
- By convention, a self-loop counts for two in the degree of a node.
- The model does not depend on the numbering of the half-edge (since the procedure or pairing half edges is exchangeable).

Proposition 1.3.4 (The law of the Configuration Model). *Let $G = (x_{ij})_{i,j\in[n]}$ be a multigraph on the vertices $[n]$ such that:*

$$d_i = x_{ii} + \sum_{j=1}^{n} x_{ij}$$

Then,

$$\Pr\Big(\mathrm{CM}_n(\boldsymbol{d}) = G\Big) = \frac{1}{(\ell_n - 1)!!} \frac{\prod_{i=1}^n d_i!}{\prod_{i=1}^n 2^{x_{ii}} \prod_{1\le i\le j\le n} x_{ij}!}. \tag{1.3.1}$$

Proof. The total number of configurations is equal to $(\ell_n - 1)!!$. By construction of Algorithm 1.3.3, each configuration has equal probability, hence

$$Pr\Big(\mathrm{CM}_n(\mathbf{d}) = G\Big) = \frac{1}{(\ell_n - 1)!!} N(G) \tag{1.3.2}$$

where $N(G)$ is the number of configuration given rise to the same multigraph G, up to a permutation of the vertices labels.

Now, if we permutate the half-edges incident to a vertex, the resulting graph remains unchanged, but the configuration is ifferent. The factor $\prod_{i=1}^n d_i!$ accounts for this (it is the number of ways to permute the half-edges incident to all vertices).

Some of these permutations give the same configuration: this is taken into account by the factor $x_{ij}!$ (multiple edges between vertices i, j). The last factor $2^{x_{ii}}$ compensates the fact that pairing kl and lk gives the same overall configuration. □

Proposition 1.3.5. *Let $d_1, \dots, d_n$ be i.i.d., distributed according to some distribution d, and let $\gamma = \dfrac{\mathbb{E}(D(D-1))}{\mathbb{E}(D)}$. Then:*

- *the expected number of self loops is smaller than $\dfrac{\gamma}{2}$;*
- *the expected number of multi-edges is smaller than $\dfrac{\gamma}{4}$.*

Proof. We will admit it here. □

1.3.2 Preferential attachment model

Previous models are statics, in the sense that the number of nodes is fixed. Moreover, they don't really explain how interesting properties (degree distribution, etc.) can arise in the real graphs we observe. This section provides an example of random graphs where the nodes are added over time.

A first possibility is to imagine that G_n is an Erdős-Rényi graph $G(n,p)$, and a new node $n+1$ is added, and edges $(i, n+1)$ (for $i = 1, \dots, n$) are added independently with probability p. The new graph G_{n+1} is thus a $G(n+1,p)$. (Note that G_n is a subgraph of G_{n+1}). The problem is that the degree sequence is binomial, hence doesn't fit what we observe in most of real networks.

The *preferential attachment paradigm* offers an intuitive explanation behind the power law degree distribution that we seem to observe in reality. Indeed, a new node $n+1$ will be connected to the n previous nodes by some additional edges. Theses edges $(i, n+1)$ are drawn independently, but with a probability proportional to the degree of the vertex i at that time. Thus, the new node $n+1$ will be more likely to be linked to vertices with large degrees.

Definition 1.3.6 (Preferential attachment - Informal definition)**.** At time t, new node will be connected to an old node i with a probability proportional to the degree $d_i(t)$ of the old node (at time t).

With that definition, we can draw few remarks:

- The old nodes will tend to have higher degrees than the new ones;
- *The rich get richer.* Indeed, new nodes tend to be attached to high degree old nodes, thus we expect the formation of hubs.

The fact that the graph will have hubs tend to make us think that the degree distribution will not be binomial, but will exhibit a power law. We will show that in the next section, just after giving a proper definition of the model.

Remark 1.3.7. The term *preferential attachment* comes from Barabási and Albert (1999), who proposed a model (but not totally well defined). The model is actually close to earlier work of Yule (1925) For a more rigorous treatment, one can see Bollobás et al. (2001) (and of course Hofstad (2016)).

Model definition

Definition 1.3.8. A sequence of graph $\left\{G_t = (V_t, E_t), t \in \mathbb{N}\right\}$ is said to be drawn under the preferential attacheement model if:

- $|V_1| = 1$ and $|E_1| = 1$: at time $t = 1$, we have one node v_1 with a single self-loop;
- At timestep $t+1$, we add one node (let us call him v_{t+1}) to the graph. This node will be linked to ONE node. The probability that the new node is connected to node v_i is given by:

$$\Pr\left((v_{t+1}, v_i) \in E_{t+1} | G_t\right) = \begin{cases} \dfrac{1}{2t+1} & \text{if } v_i = v_{t+1} \\ \dfrac{D_i(t)}{2t+1} & \text{otherwise.} \end{cases} \tag{1.3.3}$$

Here $D_i(t)$ is the degree of node v_i at time t (recall that by convention, a self-loop increases the degree by 2).

Lemma 1.3.9. *After t timestep, the PA algorithm results in a network with $N = t$ nodes and t edges. In particular, the equation (1.3.3) defines a probability.*

Proof. Indeed, at each time step, we add one node, so $|V_t| = t$. Moreover, we add only one edge per time step, □

Remark 1.3.10. A more general version is given in Hofstad (2016). The one we gave here corresponds to the case $m = 1, \delta = 0$.

Study of the degree

Proposition 1.3.11. *When $t \to +\infty$, the Preferential Attachment model exhibits a power law degree distribution, of exponent 3.*

Proof. Let $p(k, s, t)$ be the probability that a vertex s has degree k at time t. The evolution of $p(k, s, t)$ is described by the *master equation*

$$p(k, s, t+1) = \frac{k-1}{2t+1} p(k-1, s, t) + \Big(1 - \frac{k}{2t+1}\Big) p(k, s, t) \tag{1.3.4}$$

with initial condition $p(k, s = 1, 1, t = 1) = \delta_{k,1}$ and boundary $p(k, t, t) = \delta_{k,1}$. Here the term $\dfrac{k-1}{2t+1}$ represents the probability that the new edge is linked to node s at time t (thus increasing the degree of s by 1), and $\big(1 - \dfrac{k}{2t+1}\big)$ the probability that the new edge is not linked to node s.

Let us sum over all nodes $s = 1, \dots, t$ in the networks at time t. By denoting $P(k, t)$ the total degree distribution of the entire network, *i.e.*:

$$P(k, t) = \frac{1}{t} \sum_{s=1}^{t} p(k, s, t), \tag{1.3.5}$$

we get:

$$(t+1)P(k, t+1) = \frac{k-1}{2t+1} tP(k-1, t) + \Big(1 - \frac{k}{2t+1}\Big) tP(k, t). \tag{1.3.6}$$

Thus the evolution of $P(k, t)$ can be written as:

$$(t+1)P(k, t+1) - tP(k, t) = \frac{t}{2t+1} \Big((k-1)P(k-1, t) - kP(k, t) \Big) + \delta_{k,1}. \tag{1.3.7}$$

When $t \to +\infty$, this equation for the stationary distribution reduces to

$$P(k) + \frac{1}{2} \Big(kP(k) - (k-1)P(k-1) \Big) = \delta_{k,1}, \tag{1.3.8}$$

where $P(k)$ would be $\lim\limits_{t \to +\infty} P(k, t)$. This last equation look like the differential equation

$$P(k) + \frac{1}{2} \frac{\mathrm{d}kP(k)}{\mathrm{d}k} = 0 \tag{1.3.9}$$

whose solution is

$$P(k) = Ck^{-3} \tag{1.3.10}$$

where C is a normalization factor, such that $\sum_k P(k) = 1$ ($C = \zeta(3)$ where ζ is the Riemann function). □

Remark 1.3.12. The previous proof is not totally rigorous, as it involved some approximations. One can refer to Hofstad (2016) for a more mathematically involved (but rigorous) proof, as well as deeper results on the Preferntial Attachment Model. Lastly, a nice note on the differential equation method can be found in Warnke (2019) (and references therein).

1.3.3 Random geometric graphs

Motivations : base station localisation in wireless and sensors networks ; features in a feature space.

Definition 1.3.13. Let $V = [n]$ be the set of vertices. Given a dimension d, for each vertex $u \in V$, we assign a vector $X_u \in \mathbb{R}^d$ chosen uniformly in $\mathcal{S}^{d-1}$ (the $d-1$ dimensional sphere of $\mathbb{R}^d$).

Then we assign an edge between two nodes i and j if an only if the distance $d(x_i, x_j)$ is less than some threshold $r > 0$.

Extension : Spatially Embedded Random Networks (SERN) - Barnett et al. (2007)

Definition 1.3.14 (SERN model)**.** A SERN ensemble of nodes is specified by the following:

1. A number n of nodes;
2. A metric space $(\mathcal{S}, d)$, where $\mathcal{S}$ is the space where the network resides;
3. A node distribution random variable X taking values in $\mathcal{S}$. X is to represent the location in $\mathcal{S}$ of a randomly (not necessarily uniformly) situated node. Note that the position of the nodes are not necessarily uniform;
4. A connectivity decay function $\gamma : \mathbb{R}^+ \to [0, 1]$. $\gamma(s)$ represents the probability of assigning an edge to a pair of nodes at a distance s apart.

Given a n independent realizations $(x_1, \dots, x_n)$ of the random variable X, we assign independently an edge between two nodes i and j with a probability $\gamma(d(x_i, x_j))$.

Example 1.3.15. If $\mathcal{S}$ is the sphere, X is uniformly distributed along $\mathcal{S}$ and $\gamma = 1_{d(s) \leq r}$ (for some $r \geq 0$), then we recover the Random Geometric Graphs model).

1.4 Exercises session

1. Propose a basic function (complexity in $O(n^2)$), who takes as parameter n and p, to generate the adjacency matrix of an Erdős-Rényi graph $G(n, p)$.
2. Now, propose an efficient algorithm (in complexity $O(|E|)$ where $|E|$ is the total number of edges) to generate large ER graph.
3. Find experimentally for an Erdős-Rényi graph the threshold of connectivity (for $n = 100$, $n = 1000$). You can use the *networkx* package to get basic function on graphs (such as *is_ connected*).

4. Write a function that generate a configuration model (given input parameters a degree sequence).
5. Simulate a graph generated according to Preferential Attachment Model, and find experimentally the exponent of the power law.
6. Simulate a Random Geometric Graph, find experimentally threshold of connectivity (consider the cases $d = 2$, $d = 3$).
7. Draw some Erdős-Rényi and Random Geometric Graphs using the package *networkx*.

Bibliography

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Barabási, A.-L. et al. (2016). *Network science*. Cambridge University Press.

Barnett, L., Di Paolo, E., and Bullock, S. (2007). Spatially embedded random networks. *Physical Review E*, 76(5):056115.

Bollobás, B. (1998). *Random Graphs*, pages 215–252. Springer New York, New York, NY.

Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290.

Chung, F. and Lu, L. (2006). *Complex Graphs and Networks (Cbms Regional Conference Series in Mathematics)*. American Mathematical Society, Boston, MA, USA.

Clauset, A., Shalizi, C., and Newman, M. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.

Durrett, R. (2007). Random graph dynamics. *Random Graph Dynamics*.

Hofstad, R. v. d. (2016). *Random Graphs and Complex Networks*, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.

Janson, S., Luczak, T., and Rucinski, A. (2011). *Random graphs*, volume 45. John Wiley & Sons.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press.

Warnke, L. (2019). On Wormald's differential equation method. *arXiv e-prints*, page arXiv:1905.08928.

Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of dr. j. c. willis, f. r. s. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 213(402-410):21–87.