Networks and Epidemics

Moez Draief and Laurent Massoulié

May 15, 2006

Inaugurated by D. Bernoulli (1700-1782), the analysis of epidemics and their dissemination have been studied by various mathematicians. The benefits from epidemic modelling are three-fold: understanding mechanisms of spread of epidemics, predicting their future course and developing strategies to control them.

Recently, epidemic algorithms (also known as gossip algorithms) have been proposed as means to disseminate information in large scale settings, such as the Internet, or "Peer-to-Peer" networks. Such algorithms operate by letting desired information spread in a distributed system as an epidemic would spread throughout a group of susceptible individuals. Their study has provided a renewed impetus in the study of epidemics.

This course will give an introduction to deterministic and stochastic models of epidemics; with a special focus on coupling methods, Poisson approximation and branching processes. This will be illustrated by examples of human diseases, spread of rumours and dissemination of information.

Contents

Chapter 1

Differential equation approximation

1.1 Introduction

Differential equations and Markov procecsses are the basic models of dynamical systems in a deterministic and, respectively, probabilistic context. Since the analysis, both mathematical and computational, of differential equations is often more feasible and efficient, it is of interest to understand in some generality when the sample paths of a Markov chain can be guaranteed to lie, with high probability (whp), close to the solution of a differential equation.

We are interested in certain families of jump processes depending on a parameter n usually interpreted as the total population size. In what follows we approximate certain jump Markov processes as the parameter n becomes large. It is worth mentioning that the techniques presented here can be applied to a wide range of problems such as epidemic models, models for chemical reactions and population genetics, as well as other processes. More precisely the states of these systems can be normalised and interpreted as measuring population densities.

1.2 Kurtz's theorem

We begin with the definition of a density dependent family of Markov processes. Consider $(e_l)_{l=1,\dots,I}$ such that $e_l \in \mathbb{Z}^d$ $I \in \mathbb{N}$, and let $\beta_l : \mathbb{R}^d \to \mathbb{R}_+$, for $l \in \{1, \dots, I\}$. We define the Markov jump process X on $\mathbb{Z}^d$ with jump direction $(e_l)_{l=1,\dots,I}$ and transition rates $\lambda_{x,y} = \beta_{(e_l)}(x)$, $x \in \mathbb{Z}^d$, and $y = x + e_l$, i.e.

$$\begin{aligned} \mathbf{P}(X(t+h) = x + e_l \mid X(t) = x) &= h\beta_l(x/n) + o(h) \\ \mathbf{P}(X(t+h) = x \mid X(t) = x) &= 1 - h\sum_{l \in I} \beta_l(x/n) + o(h)\,. \end{aligned}$$

Let $(\mathcal{N}_l)_{l \in \mathbb{Z}^d}$ be independent standard Poisson processes. Then we can construct the

process $(X(t))_{t\in\mathbb{R}}$ as follows

$$X(t) = X(0) + \sum_{l=1}^{I} e_l \mathcal{N}_l\Big(\int_0^t \beta_l\big(X(s)/n\big)ds\Big)\,. \tag{1.1}$$

Indeed, if $X(t) = x$, then there will be a jump in $\mathcal{N}_l\Big(\int_0^t \beta_l\big(X(s)/n\big)ds\Big)$ during $(t, t+h)$ with probability $h\beta_l(x/n) + o(h)$.

Now let $F(x) = \sum_{l=1}^{I} e_l\beta_l(x)$ and $X_n(t) = X(nt)/n$, we have

$$X_n(t) = X_n(0) + \sum_{l=1}^{I} \frac{1}{n} e_l\, \overline{\mathcal{N}}_l\Big(n\int_0^t \beta_l\big(X_n(s)\big)ds\Big) + \int_0^t F(X_n(s))ds\,, \tag{1.2}$$

where $\overline{\mathcal{N}}_l(u) = \mathcal{N}_l(u) - u$ is a centered Poisson process.

Theorem 1.2.1 (Kurtz's theorem). *Suppose that* $\max_{l=1,\dots,I}|e_l| = \overline{e} \in (0,\infty)$ *and*

$$\max_{l=1,\dots,I} \sup_{x\in\mathbb{R}^d} \beta_l(x) = \overline{\beta}\,,$$

is finite and that there exists a non-negative constant M *such that* F *is* M*-lipschitz, i.e.*

$$|F(x) - F(y)| \le M|x-y|\,, \quad \forall x,\, y.$$

Assume that $\lim_{n\to\infty} X_n(0) = x(0)$, *a.s. and let* $x : \mathbb{R} \to \mathbb{R}^d$ *be the solution to the integral equation*

$$x(t) = x(0) + \int_0^t F\big(x(s)\big)ds\,. \tag{1.3}$$

Then, almost surely,

$$\mathbf{P}\big(\sup_{0\le s\le t}|X_n(s) - x(s)| \ge \epsilon\big) \le 2e^{-nT\overline{\beta}h\left(\frac{\epsilon e^{-MT}}{2IT\overline{\beta}\overline{e}}\right)}\,,$$

where $h(t) = (1+t)\log(1+t) - t$. *Moreover,*

$$\lim_{n\to\infty} \sup_{0\le s\le t} |X_n(s) - x(s)| = 0, \quad a.s. \tag{1.4}$$

Kurtz's work provides a law of large numbers for density dependent families. Indeed imagine starting both the deterministic and stochastic system from the same point for a small period of time. Since the jump rates given by F are initially the same, they will have nearly the same behaviour. Now suppose that if two points are close in the stochastic system then their transition rates are also close, due to the Lipschitz condition. Therefore even after the two processes separate, if they remain close, they will still have nearly the same behaviour. We now give a proof for Kurtz's theorem.

LEMMA 1.2.1: Let $\mathcal{N}$ be a standard Poisson process. For $\epsilon > 0$ and $T > 0$

$$\mathbf{P}\big(\sup_{0\le t\le T} |\,\mathcal{N}(t) - t| \ge \epsilon\big) \le 2e^{-Th(\epsilon/T)}\,,$$

where $h(t) = (1+t)\log(1+t) - t$.

Proof. Let $\theta > 0$, we have

$$\begin{aligned}\mathbf{P}(\sup_{0\le t\le T} |\mathcal{N}(t)-t| \ge \epsilon) &\le \mathbf{P}(\sup_{0\le t\le T} (\mathcal{N}(t)-t) \ge \epsilon) + \mathbf{P}(\sup_{0\le t\le T} (t-\mathcal{N}(t)) \ge \epsilon) \\ &= \mathbf{P}(\sup_{0\le t\le T} e^{\theta(\mathcal{N}(t)-t)} \ge e^{\theta\epsilon}) + \mathbf{P}(\sup_{0\le t\le T} e^{\theta(t-\mathcal{N}(t))} \ge e^{\theta\epsilon}) .\end{aligned}$$

Noting that both $(\mathcal{N}(t)-t)_{t\ge 0}$ and $(t-\mathcal{N}(t))_{t\ge 0}$ are martingales, we have by Jensen's inequality that the processes $(e^{\theta(\mathcal{N}(t)-t)})_{t\ge 0}$ and $(e^{\theta(t-\mathcal{N}(t))})_{t\ge 0}$ are non-negative submartingales. Therefore, by applying Doob's inequality, we obtain that

$$\begin{aligned}\mathbf{P}(e^{\theta(\mathcal{N}(t)-t)} \ge e^{\theta\epsilon}) &\le e^{-\theta\epsilon}\mathbf{E}(e^{\theta(\mathcal{N}(T)-T)}) \\ \mathbf{P}(e^{-\theta(\mathcal{N}(t)-t)} \ge e^{\theta\epsilon}) &\le e^{-\theta\epsilon}\mathbf{E}(e^{-\theta(\mathcal{N}(T)-T)}) . \qquad (1.5)\end{aligned}$$

We are going to focus our attention on the submartingale $(e^{\theta(\mathcal{N}(t)-t)})_{t\ge 0}$ for the bound in (1.5) can be derived following the same lines,

$$\begin{aligned}\mathbf{P}(e^{\theta(\mathcal{N}(t)-t)} \ge e^{\theta\epsilon}) &\le e^{-\theta(\epsilon+T)}\mathbf{E}(e^{\theta\mathcal{N}(T)}) \\ &\le e^{-\theta(\epsilon+T)}e^{T(e^\theta-1)} \\ &= \exp\Big(-\theta(\epsilon+T)+T(e^\theta-1)\Big) .\end{aligned}$$

By optimising $-\theta(\epsilon+T)+T(e^\theta-1)$ over the positive parameter θ, we easily see that $\theta = \log(1+\epsilon/T)$ realises the minimum. Hence,

$$\mathbf{P}(\sup_{0\le t\le T} (\mathcal{N}(t)-t) \ge \epsilon) \le e^{-Th(\epsilon/T)} ,$$

where $h(t) = (1+t)\log(1+t) - t$. Similarly, one can derive that

$$\mathbf{P}(\sup_{0\le t\le T} (t-\mathcal{N}(t)) \ge \epsilon) \le e^{-Th(-\epsilon/T)} ,$$

and we conclude combining these bounds and the fact that $h(-x) \ge h(x)$, for $x \ge 0$. □

Proof. First, note that

$$|X_n(t)-x(t)| \le |X_n(0)-x(0)| + \int_0^t |F\big(X_n(s)\big)-F\big(x(s)\big)|ds + \sum_{l=1}^{I} \frac{|e_l|}{n}|\overline{\mathcal{N}}\Big(n\int_0^t \beta_l\big(X_n(s)\big)ds\Big).$$

Using the fact that $\lim_{n\to\infty} X_n(0) = x(0)$, a.s. and that F is M-Lipschitz, we have, for large n

$$|X_n(t)-x(t)| \le \epsilon + M\int_0^t |X_n(s)-x(s)| + \sum_{l=1}^{I}\frac{1}{n}\epsilon_{n,l}(t), \text{ a.s} ,$$

where $\epsilon_{n,l}(t) = |e_l| \,.\Big|\overline{\mathcal{N}}\Big(n\int_0^t \beta_l\big(X_n(s)\big)ds\Big)\Big|$.

Thence,

$$|X_n(t)) - x(t)| - M\int_0^t |X_n(s)) - x(s)|ds \leq \epsilon + \sum_{l=1}^{I} \frac{1}{n}\epsilon_{n,l}(t)\,. \tag{1.6}$$

Applying Lemma 1.2.1, we can derive the following bound,

$$\begin{aligned} \mathbf{P}\Big(\sup_{0\leq t\leq T}\sum_{l=1}^{I}\frac{1}{n}\epsilon_{n,l}(t) \geq \epsilon\Big) &\leq \mathbf{P}(\sup_{0\leq t\leq nT\overline{\beta}} I\overline{e}|\overline{\mathcal{N}}(t)| \geq n\epsilon) \\ &\leq 2e^{-nT\overline{\beta}h\left(\frac{\epsilon}{IT\overline{\beta}\overline{e}}\right)} \qquad (1.7) \\ & \qquad (1.8) \end{aligned}$$

Combining Lemma 1.2.1 and (1.7) yields for large n

$$\mathbf{P}\Big(\sup_{0\leq t\leq T}|X_n(t)) - x(t)| - M\int_0^t |X_n(s)) - x(s)|ds \geq 2\epsilon\Big) \leq 2e^{-nT\overline{\beta}h\left(\frac{\epsilon}{IT\overline{\beta}\overline{e}}\right)}\,.$$

By Gronwall's Lemma applied to the function $|X_n(t)) - x(t)|$ we conclude that

$$\mathbf{P}\Big(\sup_{0\leq t\leq T}|X_n(t) - x(t)| \geq 2\epsilon e^{MT}\Big) \leq 2e^{-nT\overline{\beta}h\left(\frac{\epsilon}{IT\overline{\beta}\overline{e}}\right)}\,,$$

or alternatively,

$$\mathbf{P}\Big(\sup_{0\leq t\leq T}|X_n(t)) - x(t)| \geq \epsilon\Big) \leq 2e^{-nT\overline{\beta}h\left(\frac{\epsilon e^{-MT}}{2IT\overline{\beta}\overline{e}}\right)}\,.$$

Hence $\sum_{n\in\mathbb{N}} \mathbf{P}\big(\sup_{0\leq t\leq T}|X_n(t)) - x(t)| \geq \epsilon\big) < \infty$ and the almost sure convergence in (1.4) follows immediately from (1.3) and Borel-Cantelli lemma.

□

1.3 Applications

1.3.1 Logistic model

The populations consists of only susceptible and infectives. A susceptible that is infected will stay infectious indefinitely. This model is usual referred to as the logistic model of population growth.

In this context n can be seen as the area of a region occupied by a certain population. If the population size is k then the population density is k/n. We assume that birth and

deaths occur singly with intensities proportional to the population size. More precisely the intensities are given by

$$\begin{aligned}\mathbf{P}\big(X_n(t+h) = k+1 \mid X_n(t) = k\big) &= k\lambda\big(\frac{k}{n}\big)\\ \mathbf{P}\big(X_n(t+h) = k-1 \mid X_n(t) = k\big) &= k\mu\big(\frac{k}{n}\big)\end{aligned}$$

If we let $\lambda(x) = a$ and $\mu(x) = b + cx$, we recover the deterministic logistic model given by

$$\frac{d}{dt}x = (a-b)x - cx^2\,.$$

Alternatively, one can suppose that we have a closed population of size n. Let $x(t)$ and $y(t)$ denote the number of susceptibles and infectives at time t. Hence $x(t) + y(t) = n$ and we assume that $y(0) \geq 1$. Then assuming that the population is homogeneously mixing, i.e. a given individual has contact with any other individual in the population, we have

$$\frac{dy}{dt} = \beta xy = \beta y(n-y)\,, \tag{1.9}$$

where β is the infection rate. This differential equation known as the logistic growth equation can be transformed into

$$\frac{dy}{y(n-y)} = \Big(\frac{1}{y} + \frac{1}{n-y}\Big)\frac{dy}{n} = \beta dt\,,$$

integrating,

$$\log\frac{y(t)}{n-y(t)} - \log\frac{y(0)}{n-y(0)} = \beta nt\,,$$

or alternatively

$$y(t) = \frac{y(0)n}{y(0) + (n-y(0))e^{-\beta nt}}\,.$$

First notice that as expected when t goes to ∞, $y(t) \to n$, meaning that all individual become infected, causing the end of the epidemic.

Moreover we can derive the actual time T_1 when all the individuals are infected given by

$$T_1 = \inf\{t \,:\, y(t) > n-1\}\,,$$

which solves the following equation, since y is increasing,

$$\frac{y(0)n}{y(0) + (n-y(0))e^{-\beta nT_1}} = n-1\,,$$

and

$$T_1 = \frac{1}{\beta n}\log\Big(\frac{(n-1)(n-y(0)}{y(0)}\Big)\,.$$

1.3.2 Contact model

Denote by $Y_n(t)$ the number of infectives at time t. Then Y_n is a simple continuous time birth and death process having the following transitions

$$\begin{aligned}\mathbf{P}\big(Y_n(t+h) = k+1 \mid Y_n(t) = k\big) &= \lambda k \frac{n-k}{n})\\ \mathbf{P}\big(Y_n(t+h) = k-1 \mid Y_n(t) = k\big) &= \gamma k\end{aligned}$$

Applying the above result, we have

$$\frac{dy}{dt} = \lambda(1-y)y - \gamma\,.$$

or

$$\Big(\frac{1}{y} + \frac{\lambda}{(\lambda-\gamma)-\lambda y}\Big)dy = (\lambda-\gamma)dt\,,$$

integrating

$$\log\frac{y(t)}{(\lambda-\gamma)-\lambda y(t)} - \log\frac{y(0)}{(\lambda-\gamma)-\lambda y(0)} = (\lambda-\gamma)t\,,$$

alternatively

$$y(t) = \frac{(\lambda-\gamma)e^{(\lambda-\gamma)t}y(0)}{(\lambda-\gamma)-\lambda y(0)(1-e^{(\lambda-\gamma)t})}\,.$$

Hence

- if $\lambda > \gamma$, then $y(t) \to (1-\frac{\gamma}{\lambda})$ as $t \to \infty$,
- while if $\lambda < \gamma$, then $y(t) \to 0$ as $t \to \infty$.

1.3.3 General epidemic

We assume that initially there are m infectives ad n scuceptibles. The infection period of different infectives are i.i.d.During its infectious period an infective makes contact with a given individual at the time points of a time homogeneous Poisson process with intensity λ/n [1]. Once its infectious period terminated an individual is removed and plays no further role in the epidemic spread.

We are now going to derive a triangular system of equations for $P^n = (P_1^n, \ldots, P_n^n)$, where P_k^n is the probability that k of the initial susceptibles are ultimately infected.

Theorem 1.3.1. *Consider the SIR epidemic $E_{n,m}(\lambda, I)$. Denote by P_k^n the probability that the final size of the epidemic is equal to k. Then*

$$\sum_{k=0}^{l}\binom{n-k}{l-k}\frac{P_k^n}{\Big[\phi\Big(\frac{\lambda(n-l)}{n}\Big)\Big]^{k+m}} = \binom{n}{l}\,,\quad 0 \le l \le n\,.$$

[1] in order to keep the rate at which a given infective makes contact with other (initially susceptible) individuals constant equals λ, independently of the size of the population.

Proof. Recall that Z be the final size of the epidemic, an define A as the total pressure of the epidemic, i.e.

$$A = \frac{\lambda}{n}\int_0^\infty Y(u)du = \frac{\lambda}{n}\sum_{j=-(m-1)}^{Z} I_j\,.$$

Define $\phi(\theta) = E(e^{-\theta I})$ the Laplace transform of I. Then

$$\begin{aligned}(\phi(\lambda\theta/n))^{n+m} &= E\Big[\exp\Big(-\frac{\lambda\theta}{n}\sum_{j=-(m-1)}^{n} I_j\Big)\Big]\\ &= E\Big[\exp\Big(-\theta\Big(A+\frac{\lambda}{n}\sum_{j=Z+1}^{n} I_j\Big)\Big)\Big]\\ &= E\Big[e^{-\theta A}(\phi(\lambda\theta/n))^{n-Z}\Big]\,,\end{aligned}$$

where the last identity follows since $\{I_j,\ j \geq Z+1\}$ is independent of Z and A. This yields the following Wald's identity

$$E\Big[\frac{e^{-\theta A}}{(\phi(\lambda\theta/n))^{Z+m}}\Big] = 1\,, \quad \theta \geq 0\,. \tag{1.10}$$

We are now in position to derive the system of equations for P^n. To this end, we need to introduce some more notations. First let $[k] = \{1,\ldots,k\}$ and note that if we denote by $P^n_{[k]}$, then by symmetry $P^n_k = \binom{n}{k}P^n_{[k]}$. Now fix $0 \leq k \leq l \leq n$. the event that an epidemic within $[n]$ infects precisely the set $[k]$ is the same as the event that a sub-epidemic within $[l]$ infects precisely $[k]$ and that these k new infectives together with the initial m infectives fail to infect the set $[n]$ $[l]$. It follows that

$$P^n_{[k]} = P^l_{[k]}E[e^{-A^l(n-l)}|Z^l = k]\,,$$

where Z^l and A^l are respectively the final size of the sub-epidemic and the total epidemic pressure within $[l]$. This in turn is equivalent to

$$\frac{\binom{l}{k}P^n_k}{\binom{n}{k}} = P^l_k E[e^{-A^l(n-l)}|Z^l = k]\,. \tag{1.11}$$

Using Wald's identity (1.10) with $\theta = n-l$ and conditioning on the final size Z^l, we get

$$\sum_{k=0}^{l}\frac{P^l_k E[\exp\left(-A^l(n-l)\right)|Z^l = k]}{[\phi(\lambda(n-l)/n)]^{k+m}} = 1\,. \tag{1.12}$$

Combining (1.11) and (1.12) gives us

$$\sum_{k=0}^{l}\frac{\binom{l}{k}P^n_k}{\binom{n}{k}[\phi(\lambda(n-l)/n)]^{k+m}} = 1\,.$$

We conclude by noticing that $\binom{l}{k}/\binom{n}{k} = \binom{n-k}{l-k}/\binom{n}{l}$. □

If the infectious period has the lack-of-memory property (the infectious period $I$is exponential with intensity γ for example) then the process (X_n, Y_n), where $X_n(t)$ and $Y_n(t)$ is the number of susceptibles and the number of infectives, respectively, at time t, is a Markov process. Then the process (X, Y) is governed by the following transition intensities

$$\begin{aligned}\mathbf{P}\big((X_n(t+h), Y_n(t+h)) = (i-1, j+1) \mid (X_n(t), Y_n(t)) = (i,j)\big) &= h\lambda\frac{ij}{n} + o(h)\\ \mathbf{P}\big(X_n(t+h), Y_n(t+h)) = (i, j-1) \mid (X_n(t), Y_n(t)) = (i,j)\big) &= h\gamma j + o(h)\,.\end{aligned}$$

Let $X_n(0) = n$ and $Y_n(0) = \mu n$Thence the deterministic solution that approximates the trajectory of (X_n, Y_n) is governed by the the following pair of differential equations

$$\begin{aligned}\frac{dx}{dt} &= -\lambda xy\,, \quad x(0) = 1\\ \frac{dy}{dt} &= \lambda xy - \gamma y\,, \quad y(0) = \mu\,.\end{aligned}$$

This differentiaConsider a general equation was the first deterministic epidemic model presented by Kermack and McKendrick in 1927.

Theorem 1.3.2 (Kermack and McKendrick (1927))**.** *Consider the general epidemic model evolving according to*

$$\begin{aligned}\frac{dx}{dt} &= -\lambda xy\,, \quad x(0) = 1\\ \frac{dy}{dt} &= \lambda xy - \gamma y\,, \quad y(0) = \mu\\ \frac{dz}{dt} &= \gamma y\,, \quad z(0) = 0\end{aligned}$$

- *When the infection ultimately ceases spreading, a positive number x_∞ of susceptibles remain uninfected, and the total number z_∞ od individuals ultimately infected and removed equals $n - x_\infty$ and it is the unique root of the equation*

$$n - z_\infty = x(0)e^{-z_\infty/rho}\,, \tag{1.13}$$

  where $\rho = \gamma/\lambda$ is the relative removal rate, and $y(0) < z_\infty < n$.

- *A major outbreak occurs if and only if $\frac{dy}{dt}(0) > 0$; this happens only if the initial number of susceptibles $x(0) > \rho$.*

- *It $x(0) = \rho + \nu$, and if the initial number of infectives $y(0)$ is small relative to ν, then the final number of susceptibles left in the population is approximately $\rho - \nu$ and $\frac{d}{dt}x$ is roughly 2ν*

1.3.4 Rumour models

Chapter 2

Galton-Watson branching processes

2.1 Introduction

Model introduced by Sir Francis Galton in 1873. Characterised by offspring distribution on $\mathbb{N}$, $\{p_k\}_{k\in\mathbb{N}}$. Starting from one individual at generation 0, note Z_n: number of individuals at generation n. Then:

$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_{n,i},$$

where $\xi_{n,i}$: number of offspring from i-th individual of $n-1$-th generation. The $\{\xi_{i,n}\}_{i,n\in\mathbb{N}}$ are i.i.d., distributed according to $\{p_k\}$.

2.2 Depth-first view

Draw the tree, as in Figure 2.1, where there is one edge from each individual to each of its offspring. Call $\mathcal{T}$ the corresponding tree, rooted at the ancestor. Conditionally on Z_1 (the number of daughters of the ancestor), $\mathcal{T}$ is obtained by connecting the root to the individual roots of Z_1 rooted trees, $\mathcal{T}_1, \ldots, \mathcal{T}_{Z_1}$, that are mutually independent, and distributed as $\mathcal{T}$. Note p_{ext} the extinction probability, i.e. the probability that after some finite n, $Z_n = 0$. Note $|\mathcal{T}|$ the number of nodes of a tree $\mathcal{T}$. We thus have

$$\begin{aligned} p_{ext} &= \mathbf{P}(|\mathcal{T}| < \infty) \\ &= \textstyle\sum_{k=0}^{\infty} p_k \mathbf{P}(|\mathcal{T}_1| < \infty, \ldots, |\mathcal{T}_k| < \infty) \\ &= \textstyle\sum_{k=0}^{\infty} p_k p_{ext}^k. \end{aligned}$$

Thus, defining $\phi(s) := \sum_{k\in\mathbb{N}} p_k s^k$, the characteristic function of the offspring distribution, we have

$$p_{ext} = \phi(p_{ext}). \tag{2.1}$$

Figure 2.1: Galton-Watson tree

2.3 Breadth-first view

i.e., look at process generation after generation. Define

$$\phi_n(s) := \mathbf{E}[s^{Z_n}].$$

Then:

$$\phi_n(s) = \mathbf{E}\left[s^{\sum_{i=1}^{Z_{n-1}} \xi_{i,n}}\right] = \phi_{n-1}\left(\phi(s)\right). \tag{2.2}$$

This will be used to show the following:

Theorem 2.3.1. *If $p_0 = 0$, then $p_{ext} = 0$. Assume thus $p_0 > 0$.*
If $\mathbf{E}[\xi] \leq 1$, then $p_{ext} = 1$. If $\mathbf{E}[\xi] > 1$, p_{ext} is the unique solution in $]0,1[$ of Equation (2.1).

Proof. Assume $p_0 > 0$. ϕ is convex, and increases from p_0 to 1 as s goes from 0 to 1. The slope of $s \to \phi(s)$ at $s = 1$ is $\mathbf{E}[\xi]$. Thus, if $\mathbf{E}[\xi] < 1$, the graph of ϕ is strictly above the line $y = x$ on $[0,1[$. The unique root to (2.1) in $[0,1]$ is thus $p_{ext} = 1$.

If $\mathbf{E}[\xi] = 1$, then ϕ is strictly convex; it is strictly above its tangent at $s = 1$ on $[0,1[$ and again the unique root of (2.1) on $[0,1]$ is $p_{ext} = 1$.

If $\mathbf{E}[\xi] > 1$, necessarily the graph of ϕ crosses the line $y = x$ in $]0,1[$, at one and only one point; denote it p^*. Let $p_{ext,n} = \mathbf{P}(Z_n = 0)$. This is the probability that extinction has occured in less than n generations. By monotone convergence, $p_{ext,n} \to p_{ext}$ as $n \to \infty$. Note that $p_{ext,n} = \phi_n(0)$, and thus by (2.2), $p_{ext,n} = \phi(p_{ext,n-1})$. Using $p_{ext,0} = 0$, it follows by induction that $p_{ext,n} \leq p^*$. Thus necessarily,

$$\lim_{n\to\infty} p_{ext,n} = p_{ext} \leq p^*.$$

Thus p_{ext} must coincide with p^*. □

Define $\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n)$. Then (exercise):

$$\mathbf{E}[Z_n | \mathcal{F}_{n-1}] = (\mathbf{E}[\xi])\, Z_{n-1},\ n > 0.$$

Equivalently,

$$M_n := (\mathbf{E}[\xi])^{-n}\, Z_n \tag{2.3}$$

is a martingale. Recall:

Definition 2.3.1. *A family $\{Z_s\}_{s\in S}$ of random variables, indexed by some arbitrary set S, is equi-integrable if:*

$$\lim_{a\to+\infty} \sup_{s\in S} \mathbf{E}\left(|Z_s| 1_{|Z_s|>a}\right) = 0.$$

LEMMA 2.3.1: A sufficient condition for equi-integrability of $\{Z_s\}_{s\in S}$ is that, for some $\epsilon > 0$,

$$\sup_{s\in S} \mathbf{E}\left(|Z_s|^{1+\epsilon}\right) < +\infty.$$

Proof. Note that:

$$\mathbf{E}\left(|Z_s| 1_{|Z_s|>a}\right) \leq \mathbf{E}\left(|Z_s|^{1+\epsilon}\right) a^{-\epsilon}.$$

□

Theorem 2.3.2. *(martingale convergence theorem) A discrete time martingale $\{M_n\}_{n\geq 0}$ is such that M_n converges almost surely, as $n \to \infty$, to a limiting random variable M_∞. Furthermore, if the family of random variables $\{M_n\}_{n>0}$ is equi-integrable, one has the identity*

$$M_n = \mathbf{E}\left(M_\infty | \mathcal{F}_n\right).$$

(for a proof see XXX). Application to "branching martingale" (2.3): for subcritical branching (i.e. $\mathbf{E}[\xi] < 1$) we know that $p_{ext} = 1$. Hence necessarily, $M_\infty = 0$ a.s.; martingale can therefore not be equi-integrable.

Exercise 2.3.1. *1) Show that*

$$Var(Z_n) = \begin{cases} n var(\xi) & \text{if } \mathbf{E}[\xi] = 1, \\ var(\xi)\left(\mathbf{E}(\xi)\right)^{n-1} \frac{(\mathbf{E}\xi)^n - 1}{\mathbf{E}\xi - 1} & \text{if } \mathbf{E}(\xi) \neq 1. \end{cases}$$

2) Show that the branching martingale $\{M_n\}_{n\geq 0}$ as defined in (2.3) is equi-integrable when $\mathbf{E}(\xi) > 1$.

2.4 The one-by-one, or random walk view

Starting with the ancestor node at step 1, at each step: pick a node that has been unveiled in preceding steps, but whose offspring has not yet been unveiled; then unveil its offspring nodes. Let a node be active at step n if it has already been unveiled, but not its offspring. Denoting by A_n the number of active nodes at step n, one has the recursion:

$$A_n = A_{n-1} - 1 + \xi_n,$$

initialised with $A_0 = 1$, and where the ξ_n are iid, distributed according to the offspring distribution.

Let

$$T = \inf\{n > 0 : A_n = 0\}.$$

Proposition 2.4.1. *The total population size of the branching process equals T.*

Proof. For all $t \leq T$, it holds that

$$A_t = A_0 - t + \sum_{i=1}^{t} \xi_i.$$

At T, for the first time we exhaust active nodes; the total population size must then be the number of offspring distributions seen so far, that is $\sum_{i=1}^{T} \xi_i$, plus $A_0 = 1$ to account for the initial ancestor. However, since $A_T = 0$, from the previous equation we have that:

$$T = A_0 + \sum_{i=1}^{T} \xi_i.$$

□

The first application of the random walk approach is to characterise the behaviour of super-critical (i.e. with $\mathbf{E}\xi > 1$) branching processes, conditionally on extinction taking place.

Definition 2.4.1. *A history of the branching process is a sequence* $\{A_0, \ldots, A_T\}$ *such that* $A_0 = 1$*, for all* $t = 1, \ldots, T$*,*

$$A_t = A_{t-1} - 1 + \xi_t$$

for some $\xi_t \in \mathbb{N}$*,* $A_t > 0$ *for* $t < T$*, and* $A_T = 0$*.*

Let a reference offspring distribution $\{q_k\}_{k\geq 0}$ be given, such that $q_0 > 0$, and the corresponding mean offspring size $\mathbf{E}_q(\xi)$ equals 1. Define for positive parameter $\lambda > 0$ the tilted offspring distribution

$$p_k(\lambda) := q_k \frac{\lambda^k}{\phi_1(\lambda)},\ k \geq 0,$$

where

$$\phi_1(\lambda) = \sum_{k\geq 0} q_k \lambda^k.$$

Note that the $\{p_k(\lambda)\}_{k\geq 0}$ define a subcritical, or super-critical branching process according to whether $\lambda > 1$ or $\lambda < 1$ respectively.

Remark 2.4.1. *If instead a super-critical distribution* $\{p_k\}_{k\geq 0}$ *is given initially, provided* $p_0 > 0$*, one can always determine* $\{q_k\}_{k\geq 0}$ *satisfying the above assumptions and such that* $\{p_k\}_{k\geq 0}$ *is an exponential tilting of* $\{q_k\}_{k\geq 0}$ *with some tilting parameter* $\lambda > 1$*.*

EXAMPLE 2.4.1: With $q_k = e^{-1}/k!$ the unit mean Poisson distribution, one has: $\phi_1(\lambda) = e^{-1+\lambda}$, and thus the tilted distribution $p_k(\lambda)$ is the Poisson distribution with parameter λ.

Definition 2.4.2. *Given the reference distribution $\{q_k\}_{k\geq 0}$, and $\lambda > 1$, denote by $p_{ext}(\lambda)$ the extinction probability associated with offspring distribution $\{p_k(\lambda)\}_{k\geq 0}$. One says that parameter μ is conjugate to λ if*

$$\frac{\phi_1(\lambda)}{\lambda} = \frac{\phi_1(\mu)}{\mu}.$$

LEMMA 2.4.1: Given $\lambda > 1$, there is a unique conjugate parameter $\mu \neq \lambda$. It satisfies: $\mu < 1$, and is given by

$$\mu = \lambda p_{ext}(\lambda). \tag{2.4}$$

Proof. Define

$$f(x) := \phi_1(x)/x = \sum_{k\geq 0} q_k x^{k-1}.$$

This is a strictly convex function, and such that $f'(1) = \phi_1'(1) - \phi_1(1) = 0$. Also, $f(0^+) = +\infty$ since $q_0 > 0$. This implies existence and uniqueness of a conjugate parameter $\mu \neq \lambda$, and the fact that $\mu < 1$. Moreover, the extinction probability $p_{ext}(\lambda)$ satisfies, in view of (2.1):

$$p_{ext}(\lambda) = \sum_{k\geq 0} q_k \frac{[\lambda p_{ext}(\lambda)]^k}{\phi_1(\lambda)}.$$

Equivalently, $\phi_1(\lambda p_{ext}(\lambda)) = \phi_1(\lambda) p_{ext}(\lambda)$. This implies that $\mu := \lambda p_{ext}(\lambda)$ is conjugate to λ. □

We can now characterise the distribution of the history of a super-critical branching process conditionally on its extinction:

Theorem 2.4.1. *The distribution of the history $\{A_1, \ldots, A_T\}$ under offspring distribution $\{p_k(\lambda)\}_{k\geq 0}$, conditioned on extinction, coincides with the distribution of the history under offspring distribution $\{p_k(\mu)\}_{k\geq 0}$, where μ is the conjugate parameter of λ.*

Proof. Given a finite history $H = \{A_1, \ldots, A_T\}$, let $\{\xi_1, \ldots, \xi_T\}$ be the offspring numbers associated to it via

$$A_t = A_{t-1} - 1 + \xi_t.$$

According to the definition of a history, one has $\sum_{i=1}^T \xi_i = T - 1$. Thus the probability of this history under offspring distribution $\{p_k(\lambda)\}_{k\geq 0}$, conditioned on extinction, is:

$$\begin{aligned}
\mathbf{P}_\lambda(H|\text{Extinction}) &= \tfrac{1}{p_{ext}(\lambda)} \textstyle\prod_{i=1}^T p_{\xi_i}(\lambda) \\
&= \tfrac{1}{p_{ext}(\lambda)} \textstyle\prod_{i=1}^T q_{\xi_i} \tfrac{\lambda^{\xi_i}}{\phi_1(\lambda)} \\
&= \tfrac{1}{p_{ext}(\lambda)} \tfrac{\lambda^{T-1}}{\phi_1(\lambda)^T} \textstyle\prod_{i=1}^T q_{\xi_i} \\
&= \tfrac{\mu^{T-1}}{\phi_1(\mu)^T} \textstyle\prod_{i=1}^T q_{\xi_i} \\
&= \textstyle\prod_{i=1}^T p_{\xi_i}(\mu),
\end{aligned}$$

where we used the conjugacy relation between λ and μ. This last expression is precisely the probability of history H under offspring distribution $\{p_k(\mu)\}_{k\geq 0}$. □

The next application of the random walk representation of branching processes is to obtain bounds on the total population in the sub-critical case $\mathbf{E}(\xi) < 1$. The main tool is the Chernoff bound:

LEMMA 2.4.2: (Chernoff bound) Given i.i.d. random variables $X_1, \ldots, X_n$, for any $a \in \mathbb{R}$, one has:

$$\mathbf{P}\left(\sum_{i=1}^{n} X_i \geq na\right) \leq e^{-nh(a)},$$

where

$$h(a) = \sup_{\theta \geq 0} \left[\theta a - \log \mathbf{E}\left(e^{\theta X_1}\right)\right].$$

Proof. For all $\theta \geq 0$, by Tchebitchev's inequality, one has:

$$\begin{array}{rcl} \mathbf{P}(\sum_{i=1}^{n} X_i \geq na) & = & \mathbf{P}\left(\exp(\theta \sum_{i=1}^{n} X_i) \geq \exp(n\theta a)\right) \\ & \leq & \mathbf{E}[\exp(\theta \sum_{i=1}^{n} X_i)] e^{-n\theta a} \\ & \leq & \exp\left(-n\left[\theta a - \log \mathbf{E}(e^{\theta X_1})\right]\right). \end{array}$$

The result follows by optimising this bound over $\theta \geq 0$. □

This bound is useful when there exists some $b > 0$ such that $\mathbf{E}(\exp(bX_1)) < +\infty$, and $a > \mathbf{E}(X_1)$. Under the first condition, the function $\theta \to \theta a - \log \mathbf{E}[\exp(\theta X_1)]$ is finite and differentiable in the interval $[0, b[$. Its derivative at $\theta = 0^+$ equals $a - \mathbf{E}(X_1)$, which is positive under the second assumption. Thus under the two conditions, the exponent $h(a)$ is strictly positive.

A direct application of Chernoff's bound yields:

LEMMA 2.4.3: The total population T of the branching process verifies

$$\mathbf{P}(T \geq t) \leq \exp(-th(1)),$$

where $h(x) = \sup_{\theta \geq 0} [\theta x - \log \mathbf{E}(\exp(\theta \xi))]$.

Proof. One has:

$$\begin{array}{rcl} \mathbf{P}(T \geq t) & = & \mathbf{P}(A_1 > 0, \ldots, A_t > 0) \\ & \leq & \mathbf{P}(A_t > 0) \\ & = & \mathbf{P}(A_0 + \sum_{i=1}^{t} \xi_i > t) \\ & = & \mathbf{P}(\sum_{i=1}^{t} \xi_i \geq t). \end{array}$$

The claimed bound then follows directly from Chernoff's bound. □

EXAMPLE 2.4.2: If the offspring distribution is Poisson(λ), one has:

$$\begin{aligned} h(x) &= \sup_{\theta \geq 0} \left[\theta a - \log \left(\sum_{k \geq 0} e^{-\lambda} \frac{(\lambda e^{\theta})^k}{k!} \right) \right] \\ &= \sup_{\theta \geq 0} \left[\theta a - \lambda (e^{\theta} - 1) \right]. \end{aligned}$$

Upon differentiating, one finds that the optimal value is at $\theta = \log(a/\lambda)$, giving:

$$h(a) = \lambda h_1(\lambda/a),$$

where $h_1(u) = u \log(u) - u + 1$.

2.5 Yule process

Chapter 3

Erdős-Rényi graphs and Reed-Frost epidemics

3.1 Introduction

The Reed-Frost model is a particular example of an SIR (susceptible-infective-removed) epidemic process. Its basic version is as follows. A set of N individuals is given, indexed by $i \in \{1, \ldots, N\}$. At step 0, a single individual is infected (say). Once infected, an individual is infectious during the subsequent time slot, after which it is removed (either dead, or immunised). While infectious, it will succeed in infecting a target node with probability p, and this independently for all target nodes and infectious nodes. Thus the model's parameters are the number of nodes, N, and the infection probability, p.

A more formal description is as follows. Let $Z_u(t) \in \{S, I, R\}$ denote the state of node u during step t. Then the process $Z(t) = \{Z_u(t)\}_{u \in \{1,\ldots,N\}}$ is a homogeneous discrete time Markov process. Given two states z, z' in $\{S, I, R\}^N$, a transition from z to z' can take place only if for all $i = 1, \ldots, N$, $z_i \in \{I, R\} \Rightarrow z'_i = R$: all nodes are removed after being infected, and remain removed afterwards; and $z_i = S \Rightarrow z'_i \in \{S, I\}$. Noting $I(z)$ (respectively, $S(z)$, $R(z)$) the number of infectious (respectively, susceptible, removed) nodes in state z, provided the pair of states z, z' satisfies the above constraints, the transition probability is given by:

$$\mathbf{P}(Z(t+1) = z' | Z(t) = z) = (1-p)^{I(z)S(z')}[1 - (1-p)^{I(z)}]^{I(z')}.$$

A related model is the Erdős-Rényi (ER) random graph model. The $\mathcal{G}(N, p)$ random graph model is a graph on N nodes $\{1, \ldots, N\}$, in which for each pair (u, v) of nodes, $u < v$, the edge (u, v) is present with probability p, independently of the presence of other edges. Let $\xi_{uv} = 1$ if edge (u, v) is present and $\xi_{uv} = 0$ otherwise.

The Reed-Frost epidemics can then be constructed from the $\mathcal{G}(N, p)$ model, in the following manner. Assume node u is infected during step t. Then for any other node v, u will successfully infect node v during step $t + 1$ if $\xi_{uv} = 1$ (in case $u < v$; if $v < u$, use variable ξ_{vu} instead). Note that if the target node v is already infected or removed,

it will already have "used" the random variable ξ_{uv}, but this is of no consequence since then node u can't infect v anymore.

The corresponding state of the Reed-Frost epidemics at time t, if initiated from a single infectious node $u \in \{1, \ldots, N\}$, can be described purely in graph-theoretic properties of $\mathcal{G}(N,p)$.

Let $d_G(u,v)$ denote the shortest number of hops in a path connecting u to v in graph G. Define the i-neighborhood of node u as

$$\Gamma_i(u) := \{v : d_{\mathcal{G}(N,p)}(u,v) = i\}.$$

Then the state of the Reed-Frost epidemics started at u, at time t is given by: $Z_v(t) = R$ if $d_{\mathcal{G}(N,p)}(u,v) < t$, $Z_v(t) = I$ if $d_{\mathcal{G}(N,p)}(u,v) = t$, that is if $v \in \Gamma_t(u)$, and $Z_v(t) = S$ otherwise.

3.2 Emergence of the giant component

We shall study the size of the largest component of the graph $\mathcal{G}(N,p)$ as $N \to \infty$, assuming the particular scaling that p also depends on N, in such a way that $Np \equiv \lambda > 0$ as $N \to \infty$. We denote by C_1 the largest connected component of $\mathcal{G}(N,p)$ (the size being measured in number of constituent vertices), by C_2 the second largest component, etc. We then have:

Theorem 3.2.1. *i) Consider the subcritical regime, that is when $\lambda < 1$. Then for some constant A depending on λ, we have the following:*

$$\lim_{N\to\infty} \mathbf{P}\left(|C_1| \geq A\log(N)\right) = 0. \tag{3.1}$$

ii)Consider the supercritical regime, that is when $\lambda > 1$. Denote by $p_{ext}(\lambda)$ the extinction probability of a Galton-Watson branching process with Poisson(λ) offspring distribution, that is the unique root in $]0,1[$ of the equation $x = \exp(-\lambda(1-x))$. Then for some constant $A' > 0$ depending on λ, and all $\delta > 0$, one has the following:

$$\lim_{N\to\infty} \mathbf{P}\left(\left|\frac{|C_1|}{N} - (1 - p_{ext}(\lambda))\right| \leq \delta \text{ and } |C_2| \leq A'\log(N)\right) = 1. \tag{3.2}$$

Thus, in the subcritical regime ($\lambda < 1$) all connected components are of logarithmic size; in the supercritical regime ($\lambda > 1$) a "giant component" has appeared, whose size rescaled by N converges in probability as $N \to \infty$ to $1 - p_{ext}(\lambda)$, while other components remain of logarithmic size.

The next lemma will be used in the proof of the theorem:

LEMMA 3.2.1: Let $L(1)$ denote the size of the connected component to which node 1

belongs. One then has:

(i) If $\lambda < 1, \exists \beta > 0$ such that $\mathbf{P}(L(1) \geq t) \leq e^{-\beta t},\ t \in \mathbb{N}$;
(ii) $\mathbf{P}(L(1) = t) \leq \mathbf{P}(\mathrm{Bin}(N-1, 1-(1-p)^t) = t-1),\ t \in \mathbb{N}$.

Proof. We construct the size of the connected component $C(1)$ in the following manner, reminiscent of the random walk approach to the exploration of Galton-Watson branching processes.

Denote by A_t the set of active nodes after the exploration of t nodes, starting with $A_0 = \{1\}$. Denote by B_t the set of nodes that have been expored after t steps, starting with $B_0 = \emptyset$.

To go from step $t-1$ to step t we pick an arbitrary node (say the one with smallest label), u_t, in A_{t-1}, and determine the set of its neighbours in $\{1, \ldots, N\} \setminus \{A_{t-1} \cup B_{t-1}\}$, that we denote by D_t. We then update the sets as follows:

$$A_t = A_{t-1} \cup D_t \setminus \{u_t\}, \quad B_t = B_{t-1} \cup \{u_t\}.$$

Let Y_t be the size of A_t, and ξ_t the size of D_t. We thus have:

$$Y_t = Y_{t-1} - 1 + \xi_t.$$

The size L of the connected component is determined by:

$$L = \inf\{t > 0 : Y_t = 0\}.$$

Indeed, this exploration process terminates when for the first time t, there is no active node to consider, that is when for the first time, $Y_t = 0$. The size of the component is then given by 1 plus the sum of the numbers of neighbours discovered at each step, that is:

$$L = 1 + \sum_{i=1}^{t} \xi_i.$$

Since $A_t = 0 = 1 + \sum_{i=1}^{t} \xi_i - t$, the announced characterisation of L follows.

The distribution of the number of newly discovered neighbours, ξ_t, conditionally on the outcome of the previous explorations, $D_1, \ldots, D_{t-1}$, has a binomial distribution with parameters $(p, N - Y_{t-1} - t + 1)$: indeed, there are Y_{t-1} active nodes yet to explore, and $|B_{t-1}| = t-1$ nodes already explored within the cluster.

The first half of the lemma follows by an application of the Chernoff bounding technique. For fixed $\theta > 0$, one has:

$$\begin{aligned} \mathbf{P}(L > t) \leq \mathbf{P}(Y_t > 0) = \mathbf{P}(\xi_1 + \ldots + \xi_t \geq t) \quad &\leq \mathbf{E}\left(e^{\theta \sum_1^t \xi_i - \theta t}\right) \\ &\leq \left(1 - p + pe^{\theta}\right)^{tN} e^{-\theta t} \\ &\leq \exp\left(-Npt\left(1 - e^{\theta}\right)\right) e^{-\theta t} \\ &= \exp\left(-t\left(\theta - \lambda\left(1 - e^{\theta}\right)\right)\right). \end{aligned}$$

In the second step we increased the parameter of the binomial variables ξ_i, and in the third step we used the inequality $1 + x \leq \exp(x)$, valid for all $x \in \mathbb{R}$. For $\lambda < 1$ and $\theta > 0$ small enough, the exponent $\theta - \lambda(1 - \exp(\theta))$ is positive; it provides the announced coefficient $\beta > 0$ in the first part of the lemma.

To establish the second half, consider the following specific construction of the consecutive sets A_i, B_i. Let i.i.d. Bernoulli random variables $\{\xi_{i,v}\}_{i>0,v=1,\ldots,N}$ be given, all with parameter p. A specific node v will be incorporated in the set D_i if and only if $\xi_{i,v}$ equals 1, and it has not been incorporated in any of $D_1, l \ldots, D_{i-1}$.

Let X_t denote the number of nodes not incorporated in the connected component $C(1)$ by step t. It then follows that X_t is Binomial, with parameters $(N-1, (1-p)^t)$. Indeed, for each node $v = 2, \ldots, N$, it fails to be incorporated by t if and only if $\xi_{1,v} = \ldots = \xi_{v,t} = 0$, which happens with probability $(1-p)^t$, and this independently for all such v.

Then, using the identity $N = X_t + Y_t + t$, we finally have that $Y_t + t - 1$ is Binomial with parameters $(N - 1, 1 - (1-p)^t)$.

It follows that:

$$\begin{aligned} \mathbf{P}(L = t) &= \mathbf{P}(A_1, \ldots, A_{t-1} > 0, A_t = 0) \\ &\leq \mathbf{P}(A_t = 0) \\ &= \mathbf{P}\left(\mathrm{Bin}(N-1, 1-(1-p)^t) = t-1\right), \end{aligned}$$

as announced. □

3.2.1 Subcritical phase ($\lambda < 1$)

The first half of Theorem 3.2.1 is established as follows. The number of connected components of size at least k is not larger than

$$\sum_{i=1}^{N} \mathbf{1}_{|C(i)| \geq k}.$$

The expectation of the latter sum is $N\mathbf{P}(L \geq k)$. For $k \geq \beta^{-1}(1+\delta)\log(N)$, where β is obtained from the Lemma, this expectation tends to 0, hence the probability that there is a connected component of size larger than $\beta^{-1}(1+\delta)\log(N)$ tends to 0.

3.2.2 Supercritical phase ($\lambda > 1$)

We shall use the following version of Chernoff's bound:

LEMMA 3.2.2: Le X be a sum of independent, not necessarily identically distributed, $\{0,1\}$-valued random variables. Let $\bar{X} = \mathbf{E}(X)$. Then for all $\epsilon > 0$, it holds that

$$\mathbf{P}(X - \bar{X} \geq \epsilon\bar{X}) \leq e^{-\bar{X}h(\epsilon)}, \quad \mathbf{P}(X - \bar{X} \leq \epsilon\bar{X}) \leq e^{-\bar{X}h(-\epsilon)}, \tag{3.3}$$

where $h(x) := (1+x)\log(1+x) - x$.

Proof is a combination of the Chernoff technique, and the upper bound $1 + x \leq e^x$.

The main step in the analysis of the supercritical case is the following:

LEMMA 3.2.3: Let $\lambda > 1$ be given. Then for all $\epsilon > 0$, $\delta > 0$, there exists some $\delta' > 0$ and some $t_0 > 0$ such that, for N large enough, and all p such that $|Np - \lambda| \leq \delta'$:

$$p_{ext}(\lambda) - \epsilon \leq \mathbf{P}(C(1) < t_0) \leq p_{ext}(\lambda) + \epsilon,$$
$$1 - p_{ext}(\lambda) - \epsilon \leq \mathbf{P}(|C(1) - N(1 - p_{ext}(\lambda))| \leq \delta N) \leq 1 - p_{ext}(\lambda) + \epsilon.$$

Proof. Remark that in the above statement we allow for the possibility that Np is actually distinct from λ. Let ϵ, $\delta > 0$ be fixed.

Let $\delta' > 0$ be chosen so that $|p_{ext}(\lambda \pm \delta') - p_{ext}(\lambda)| \leq \epsilon/2$. Recall the previous construction of a connected component $C(1)$, and the fact that the conditional distribution of ξ_t given $\xi_1, \ldots, \xi_{t-1}$ is Binomial with parameters $(N - 1 - \xi_1 - \ldots - \xi_{t-1}, p)$. The probability that the corresponding sequence $Y_1, \ldots, Y_t$ first hits zero before some t_0 is monotonic decreasing in p, for given N. Denote by $p^{\pm} = \lambda^{\pm}/N$ where $\lambda^{\pm} = \lambda \pm \delta'$ are the two end-points of the range in which Np will be allowed to vary.

For any fixed t, the vector $(\xi_1, \ldots, \xi_t)$ under parameters $N, p^{\pm}$ converges in distribution to a vector of independent Poisson random variables with common parameter $\lambda^{\pm}$. Indeed, this follows from the evaluation:

$$\begin{aligned} \mathbf{P}_{N,p^{\pm}}(\xi_1 = a_1, \ldots, \xi_t = a_t) &= \prod_{i=1}^{t} \binom{N-1-\sum_{j=1}^{i-1} a_j}{a_i} (p^{\pm})^{a_i} (1-p^{\pm})^{N-1-\sum_{j=1}^{i-1} a_j} \\ &\to \prod_{i=1}^{t} e^{-\lambda^{\pm}} \frac{(\lambda^{\pm})^{a_i}}{a_i!} \text{ as } N \to \infty, \end{aligned}$$

valid for all $(a_1, \ldots, a_t) \in \mathbb{N}^t$. This weak convergence ensures the following, for all fixed $t_0 > 0$:

$$\begin{aligned} \mathbf{P}_{N,p^{\pm}}(|C(1)| \leq t_0) &= \mathbf{P}_{N,p^{\pm}}(\text{for some } t \leq t_0, \xi_1 + \ldots + \xi_t \leq t - 1). \\ &\to \mathbf{P}(\text{population size of G.W. process with Poisson}(\lambda^{\pm}) \text{ offspring} \leq t_0) \end{aligned}$$

as $N \to \infty$. Thus, as this limiting value converges to $p_{ext}(\lambda^{\pm})$ when $t_0 \to \infty$, there exists t_0 such that, for sufficiently large N,

$$\mathbf{P}_{N,p^{\pm}}(|C(1)| \leq t_0) \in [p_{ext}(\lambda^{\pm}) - \epsilon/2, p_{ext}(\lambda^{\pm}) + \epsilon/2] \subset [p_{ext}(\lambda) - \epsilon, p_{ext}(\lambda) + \epsilon],$$

by our choice of $\lambda^{\pm}$. Finally, for any $p \in [p^-, p^+]$, one has:

$$\mathbf{P}_{N,p}(|C(1)| \leq t_0) \in \left[\mathbf{P}_{N,p^+}(|C(1)| \leq t_0), \mathbf{P}_{N,p^-}(|C(1)| \leq t_0)\right] \subset [p_{ext}(\lambda) - \epsilon, p_{ext}(\lambda) + \epsilon].$$

Lemma 3.2.1 (ii) says that

$$\mathbf{P}(|C(1)| = t) \leq \mathbf{P}(\text{Bin}(N - 1, 1 - (1 - p)^t) = t - 1).$$

The ratio of the mean $\bar{X}_t := (N-1)[1-(1-p)^t]$ of this variable to the value $t-1$ reads

$$(1 + o(1))\phi(t/N),$$

where $\phi(x) := x^{-1}[1 - e^{-(Np)x}]$. By differentiating it is easily seen that the function ϕ is decreasing from $Np > 1$ to $1 - e^{-Np}$ as x increases from 0 to 1. This readily implies the existence of a positive constant $\kappa > 0$ such that, for all $p \in [p^-, p^+]$, and all $t \leq [1 - p_{ext}(\lambda)) - \delta]N$, $(t-1)/\bar{X}_t < 1 - \kappa$, and for all $t \geq [1 - p_{ext}(\lambda)) + \delta]N$, $(t-1)/\bar{X}_t > 1 + \kappa$. Thus Lemma 3.2.2 yields:

$$\mathbf{P}\left(|C(1)| > t_0, ||C(1)| - (1 - p_{ext}(\lambda))N| > \delta N\right) \leq \sum_{t \geq t_0} 2e^{-\kappa' t},$$

for some other positive constant κ'. For sufficiently large t_0, this quantity can be made arbitrarily small; the second result of the lemma follows. □

All the ingredients are there to conclude the proof of Theorem 3.2.1. Pick an arbitrary node (say node 1) of the original graph, and extract its connected component, $C(1)$. If $|C(1)| \leq t_0$ proceed to next step; if $||C(1)| - (1 - p_{ext}(\lambda))N| \leq \delta N$, stop the procedure, claiming a successful identification of the giant component. If none of the two conditions prevails, stop the procedure, admitting failure of the identification of the giant component.

Repeatedly extract such connected components, until either success (i.e. component of size L: $|L - (1 - p_{ext}(\lambda))N| \leq \delta N$) or failure (not success, and a component of size $> t_0$) occurs, up to k times, for some k to be determined. After i extractions, which have been neither failures nor successes, we are left with an E-R graph with parameters N' and p, where

$$N' \in [N - it_0, N - i],$$

since we have removed at most t_0 nodes in each extraction. Thus at each step, the previous lemma applies to the remaining graph, since for bounded k, the product $N'p$ remains in the range $[\lambda, \lambda^+]$. Note that the probability of a success is at each step at least $1 - p_{ext}(\lambda) - \epsilon$, and the probability of failure at each step is at most 2ϵ. Thus the probability of success in at most k steps is at least

$$\begin{aligned} \mathbf{P}(\text{success}) &\geq \textstyle\sum_{i=1}^{k} (1 - p_{ext}(\lambda) - \epsilon)(p_{ext}(\lambda) - \epsilon)^{i-1} \\ &= (1 - p_{ext}(\lambda) - \epsilon)\tfrac{1-(p_{ext}(\lambda)-\epsilon)^k}{1-p_{ext}(\lambda)+\epsilon}. \end{aligned}$$

By choosing k such that $(p_{ext}(\lambda) - \epsilon)^k \leq \epsilon$, we ensure that this probability of success is at least $1 - O(\epsilon)$.

To conclude, conditionally on success, we have one giant component of the right size, up to k components of size at most t_0, and the remaining graph is an E-R graph with N' nodes, where

$$N' \in [N(p_{ext}(\lambda) - 2\delta), N(p_{ext}(\lambda) + \delta)].$$

Note now that the corresponding product $N'p$ is close to $p_{ext}(\lambda)\lambda$, that is the conjugate parameter of λ, from the discussion of conjugate branching processes. Therefore, this remaining E-R graph is subcritical, and by the first half of Theorem 3.2.1 its largest component is of logarithmic size.

3.2.3 The critical case ($\lambda = 1$)

Theorem 3.2.2. *Let $(C_j)_{j\geq 1}$ denote the size of the connected components of $\mathcal{G}(N, 1/N)$ ordered in decreasing order. For any N large enough, we have*

$$\mathbf{E}\sum_{j\geq 1} C_j^2 \leq 3N^{4/3}\,.$$

In particular, for $A > 0$

$$\mathbf{P}(L_1 \geq AN^{2/3}) \leq \frac{3}{A^2}\,.$$

Proof. From the exploration process

$$Y_{i+1} = Y_i - 1 + \xi_{i+1}\,. \tag{3.4}$$

Let $\tau = \min\{i \geq 1 : Y_i = 0\}$ which gives the size of the giant component. To prove the theorem, we will couple $(Y_i)_i$ to a random walk with shifted binomial increments. More precisely, We consider a sequence $(X_i)_{i\geq 1}$ of i.i.d. $Bin(N, 1/N)$ random variables, such that $\xi_i \geq Z_i$ for all i.

Let $(S_i)_{i\geq 1}$ the random walks defined by $S_0 = 1$ and for $i \geq 1$

$$S_i = S_{i-1} + \xi_i - 1 = 1 + \sum_{j=1}^{i}(X_j - 1)\,. \tag{3.5}$$

Fix $H > 0$ and define γ by

$$\gamma = \min\{i \geq 1\,,\ S_i \geq H \text{ or } S_i = 0\}\,.$$

By induction, we check that $S_i \geq Y_i$ for all $i \leq \gamma$.

Since $(S_i)_i$ is a martingale, the optional stopping theorem gives

$$\mathbf{E}(S_0) = \mathrm{E}(S_\gamma) \geq H\mathbf{P}(S_\gamma \geq H)\,.$$

Consequently,

$$\mathbf{P}(S_\gamma \geq H) \leq \frac{1}{H}\,. \tag{3.6}$$

We will now need the following result.

LEMMA 3.2.4: Let X be distributed according to $Bin(N, 1/N)$ and let f be an increasing function. Then

$$\mathbf{E}[f(S_\gamma - H) \mid S_\gamma \geq H] \leq \mathbf{E}f(X)\,. \tag{3.7}$$

Proof. Let X be a random variable with $Bin(N, 1/N)$ distribution. We can write X as a sum of n Bernoulli independent random variables $(I_j)_{j=1\ldots N}$ with parameter $1/N$. Let J be the minimal index such that $\sum_{k=1}^{J} I_k = r+1$, i.e.

$$J = \min\{j \geq 1, \sum_{k=1}^{j} I_k = r+1\}\,.$$

Given J the conditional distribution of $X - r + 1$ is $Bin(N - J, 1/N)$ which is obviously dominated by $Bin(N, 1/N)$. Let f be an increasing real function, using the fact that $\{X \geq r+1\} = \{J \leq N\}$

$$\begin{aligned}
\mathbf{E}[f(X-H) \mid X \geq r+1] &= \frac{1}{\mathbf{P}(X \geq r+1)}\mathbf{E}(f(\sum_{j=J+1}^{N} I_j)1_{\{X \geq r+1\}}) \\
&= \frac{1}{\mathbf{P}(J \leq N)}\mathbf{E}(\sum_{i=1}^{N} f(\sum_{j=i+1}^{N} I_j)1_{\{J=i\}}) \\
&\leq \frac{1}{\mathbf{P}(J \leq N)}\sum_{i=1}^{N}\mathbf{E}(f(\sum_{j=1}^{i} I'_j + \sum_{j=i+1}^{N} I_j)1_{\{J=i\}})\,,
\end{aligned}$$

where $(I'_j)_{j\geq 1}$ are i.i.d. bernoulli r.v. with parameter $1/N$ which are independent of $(I_j)_{j\geq 1}$. Therefore, by the independence property,

$$\mathbf{E}[f(X-(r+1)) \mid X \geq r+1] \leq \frac{1}{\mathbf{P}(J \leq N)}\mathbf{E}(f(\sum_{j=1}^{N} I_j)\sum_{i=1}^{N}\mathbf{E}(1_{\{J=i\}}) = \mathbf{E}f(X)\,.$$

Now let $l \in \mathbb{Z}_+$ and, $r \in \{1, \ldots, H-1\}$. Given $\{\gamma = l\} \cap \{S_\gamma \geq H\} \cap \{S_{\gamma-1} = H-r\}$, we have

$$\begin{aligned}
\mathbf{E}[f(S_\gamma - H) \mid S_\gamma \geq H] &= \mathbf{E}[f(X-(r+1)) \mid X \geq r+1] \\
&\leq \mathbf{E}[f(X)]\,.
\end{aligned}$$

□

Apply (3.7) with $f(x) = 2Hx + x^2$ to obtain

$$\begin{aligned}
\mathbf{E}[2H(S_\gamma - H) + (S_\gamma - H)^2 \mid S_\gamma \geq H] &\leq \mathbf{E}[\xi^2 + 2H\xi] \\
&= 1 + \frac{n-1}{n} + 2H \\
&\leq 2H + 2\,.
\end{aligned}$$

Write $S_\gamma^2 = H^2 + 2H(S_\gamma - H) + (S_\gamma - H)^2$ and

$$\mathbf{E}(S_\gamma^2 \mid S_\gamma \geq H) \leq H^2 + 2H + 2 \leq H^2 + 3H\,. \tag{3.8}$$

Noting that $S_i^2 - (1-\frac{1}{n})i$ is a martingale and applying the optional stopping theorem we have

$$\mathbf{E}[S_\gamma^2 - (1-\frac{1}{n})\gamma] = 1\,.$$

Combing this with (3.6) and (3.8) we obtain

$$(1-\frac{1}{n})\mathbf{E}\gamma = \mathbf{E}(S_\gamma^2) = \mathbf{P}(S_\gamma \geq H)\mathbf{E}[S_\gamma^2 \mid S_\gamma \geq H] \leq H+3\,.$$

Hence for $H < n-3$ we have

$$\mathbf{E}\gamma \leq \frac{n-1}{n}(H+3) \leq H+3\,. \tag{3.9}$$

Let $\tau_0 = \min\{i \geq 0\,,\ Y_{\gamma+i} = 0\}$. First note that if $S_\gamma = 0$ then $\tau \leq \gamma$. Therefore $\tau \leq \gamma + \tau_0 1_{\{S_\gamma \geq H\}}$, so

$$\mathbf{E}\tau \leq \mathbf{E}\gamma + \mathbf{E}(\tau_0 \mid S_\gamma \geq H)\mathbf{P}(S_\gamma \geq H)\,. \tag{3.10}$$

Since ξ_i is distributed according to $Bin(n - Y_i - i - 1, 1/n)$ which is stochastically dominated by $Bin(n-i, 1/n)$ we have

$$\mathbf{E}(\xi_i - 1) \leq \frac{n-i}{n} - 1 = -\frac{i}{n}\,.$$

Thus the process $\eta_i = Y_{\gamma+i} + \sum_{j=1}^{i} \frac{i}{n}$ is a supermatingale. By the optional stopping theorem

$$\begin{aligned}\mathbf{E}(\eta_{\tau_0} \mid S_\gamma \geq H) \leq \mathbf{E}(\eta_0 \mid S_\gamma \geq H) &= \mathbf{E}(Y_\gamma \mid S_\gamma \geq H) \\ &\leq \mathbf{E}(S_\gamma \mid S_\gamma \geq H) \leq H+1\,.\end{aligned}$$

By the obvious inequality

$$\eta_i \geq \sum_{j=1}^{i} \frac{j}{n} = \frac{i(i+1)}{2n} \geq \frac{i^2}{2n}\,,$$

we have

$$\frac{\mathbf{E}(\tau_0^2 \mid S_\gamma \geq H)}{2n} \leq H+1\,.$$

By Jensen's inequality $(\mathbf{E}[\tau 1_{\{S_\gamma \geq H\}}])^2 \leq \mathbf{E}[\tau^2 1_{\{S_\gamma \geq H\}}]$. Hence,

$$\mathbf{E}(\tau_0 \mid S_\gamma \geq H) \leq \sqrt{2n(H+1)}\,. \tag{3.11}$$

By (3.6), (3.9), (3.10) and (3.11)

$$\mathbf{E}\tau \leq H+3+\frac{1}{H}\sqrt{2n(H+1)} \leq H+2\sqrt{n/H}-1\,, \tag{3.12}$$

where the second inequality holds if n/H is large. To minimise the right-hand side, take $H\lceil n^{1/3}\rceil$ so that (3.11) is valid for large n ($n > 1000$ suffices). This yields

$$\mathbf{E}[|C(v)|] = \mathrm{E}\tau \leq 3n^{1/3}\,. \tag{3.13}$$

Finally, let $\{v_1, \ldots, v_n\}$ be the set of all nodes. By symmetry,

$$\mathbf{E}[|C(v)|] = \frac{1}{n}\mathbf{E}[\sum_{i=1}^{n} |C(v_i)|] = \frac{1}{n}\mathbf{E}[\sum_j C_j^2]\,,$$

where the second inequality follows from the fact that in the middle term we C_j exactly C_j times, for every $j \geq 1$. Thus, by (3.13),

$$\mathbf{E}[C_1^2] \leq \mathbf{E}[\sum_j C_j^2] \leq 3n^{4/3}\,.$$

By the Markov's inequality

$$\mathbf{P}(C_1 \geq An^{2/3}) \leq \frac{\mathrm{E}[C_1^2]}{A^2 n^{4/3}} \leq \frac{3}{A^2}\,.$$

□

Chapter 4

Connectivity and Poisson approximation

To analyse the emergence of connectivity in th E-R graph, we first introduce the Stein-Chen method for approximation of sums of $\{0,1\}$-valued random variables by Poisson distributions.

4.1 The Stein-Chen method

First recall the following

Definition 4.1.1. *The variation distance between two probability measures μ_1, μ_2 on the same measurable space $(\Omega, \mathcal{F})$ is defined as*

$$d_{var}(\mu_1, \mu_2) = 2 \sup_{A \in \mathcal{F}} |\mu_1(A) - \mu_2(A)|.$$

This admits the alternative characterisation:

Proposition 4.1.1. *Let a measure μ be such that both μ_1 and μ_2 are absolutely continuous with respect to μ (by definition this means that for all $A \in \mathcal{F}$, $\mu_i(A) > 0 \Rightarrow \mu(A) > 0$; e.g. the measure $\mu = \mu_1 + \mu_2$ satisfies this). Then the variation distance verifies*

$$d_{var}(\mu_1, \mu_2) = \int_\Omega \left| \frac{d\mu_1}{d\mu}(\omega) - \frac{d\mu_2}{d\mu}(\omega) \right| \mu(d\omega),$$

where $d\mu_i/d\mu$ is the Radon-Nykodim derivative of μ_i with respect to μ.

In particular, for probability measures on $\mathbb{N}$, taking for μ the counting measure $\sum_{n \geq 0} \delta_n$, it holds that

$$d_{var}(\mu_1, \mu_2) = \sum_{n \in \mathbb{N}} |\mu_1(n) - \mu_2(n)| .$$

Proof. For any $A \in \mathcal{F}$, it holds that

$$\begin{aligned} 2\,|\mu_1(A) - \mu_2(A)| &= |\mu_1(A) - \mu_2(A)| + |\mu_2(\bar{A}) - \mu_1(\bar{A})| \\ &= \left|\int_\Omega \mathbf{1}_A(\omega)\left[\tfrac{d\mu_1}{d\mu} - \tfrac{d\mu_2}{d\mu}\right] d\mu\right| \\ &\quad + \left|\int_\Omega \mathbf{1}_{\bar{A}}(\omega)\left[\tfrac{d\mu_1}{d\mu} - \tfrac{d\mu_2}{d\mu}\right] d\mu\right| \\ &\leq \int_\Omega \left|\tfrac{d\mu_1}{d\mu} - \tfrac{d\mu_2}{d\mu}\right| d\mu. \end{aligned}$$

Conversely, defining

$$A = \left\{\omega : \frac{d\mu_1}{d\mu}(\omega) > \frac{d\mu_2}{d\mu}(\omega)\right\},$$

one has

$$\begin{aligned} \int_\Omega \left|\tfrac{d\mu_1}{d\mu} - \tfrac{d\mu_2}{d\mu}\right| d\mu &= \mu_1(A) - \mu_2(A) + \mu_2(\bar{A}) - \mu_1(\bar{A}) \\ &= 2|\mu_1(A) - \mu_2(A)|. \end{aligned}$$

□

Another useful property of variation distance is the following:

Proposition 4.1.2. *Assume that for a sequence of probability measures $\{\mu_n\}_{n>0}$, there exists a probability measure μ_∞ such that*

$$\lim_{n\to\infty} d_{var}(\mu_n, \mu_\infty) = 0.$$

Then μ_n converges weakly to μ_∞ as $n \to \infty$.

Proof. Indeed, the announced weak convergence holds by definition if, for all bounded and continuous function f, it holds that:

$$\lim_{n\to\infty} \int_\Omega f(\omega)\mu_n(d\omega) = \int_\Omega f(\omega)\mu_\infty(d\omega).$$

However, for any two measures ν_1, ν_2, and any bounded measurable function g,

$$\left|\int_\Omega g(\omega)\mu(d\omega) - \int_\Omega g(\omega)\nu(d\omega)\right| \leq \sup_{\omega\in\Omega} |g(\omega)| d_{var}(\mu, \nu).$$

Indeed, this follows by writing, for some measure μ such that both ν_1 and ν_2 are absolutely continuous with respect to it:

$$\begin{aligned} \left|\int_\Omega g(\omega)\mu(d\omega) - \int_\Omega g(\omega)\nu(d\omega)\right| &\leq \int_\Omega |g(\omega)| \left|\tfrac{d\nu_1}{d\mu}(\omega) - \tfrac{d\nu_2}{d\mu}(\omega)\right| \mu(d\omega) \\ &\leq \sup_{\omega\in\Omega} |g(\omega)| d_{var}(\nu_1, \nu_2), \end{aligned}$$

where we used the previous characterisation of variation distance. Thus convergence of variation implies convergence in distribution. □

Let $X = \sum_{v \in V} I_v$, where V is a countable set, and each I_v is $\{0,1\}$-valued. We then have

Theorem 4.1.1. *Let*

$$\pi_v := \mathbf{E}(I_v), \quad \lambda = \mathbf{E}(X) = \sum_v \pi_v. \tag{4.1}$$

Assume that there exist random variables $J_{v,w}$ defined on the same probability space as the random variables I_v and such that for all $v \in V$ we have the equality of distributions:

$$\mathcal{L}\left(\{J_{vw}\}_{w \neq v}\right) = \mathcal{L}\left(\{I_w\}_{w \neq v} | \, I_v = 1\right). \tag{4.2}$$

Then the following bound on the variation distance between (the distribution of) X and the Poisson distribution with parameter λ, that is P_λ, holds:

$$d_{var}\left(X, P_\lambda\right) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{v \in V} \pi_v \left\{ \pi_v + \sum_{w \neq v} \mathbf{E}\left|J_{vw} - I_w\right| \right\}. \tag{4.3}$$

Remark 4.1.1. *Note that the term $(1 - e^{-\lambda})/\lambda$ in the right-hand side of (4.3) is upper-bounded by $\min(1, \lambda^{-1})$. The result is often stated with the latter term instead of the former.*

Proof. Define for all $A \subset \mathbb{N}$ the function f_A by $f_A(0) = 0$ and

$$f_A(i+1) = \lambda^{-1} \frac{P_\lambda(A \cap C_i) - P_\lambda(A) P_\lambda(C_i)}{P_\lambda(\{i\})}, \quad i \geq 0, \tag{4.4}$$

where $C_i = \{0, \ldots, i\}$. We then have the following lemma, whose proof will be given after the end of the current proof:

LEMMA 4.1.1: For all $A \subset \mathbb{N}$, the function $f = f_A$ verifies the following properties:

$$\begin{array}{ll} (i) & \sup_{i \geq 0} |f(i)| < \infty, \\ (ii) & \Delta f := \sup_{i \geq 0} |f(i+1) - f(i)| \leq \frac{1 - e^{-\lambda}}{\lambda}, \\ (iii) & \lambda f(i+1) - i f(i) = \mathbf{1}_A(i) - P_\lambda(A), \; i \in \mathbb{N}. \end{array}$$

In view of (iii), for all $A \subset \mathbb{N}$, one has

$$\begin{array}{rcl} \mathbf{P}(X \in A) - P_\lambda(A) & = & \mathbf{E}\left[\lambda f(X+1) - X f(X)\right] \\ & = & \sum_{v \in V} \pi_v \mathbf{E}\left[f(X+1) - f(\sum_{w \neq v} J_{vw} + 1) \right]. \end{array}$$

The Lipschitz property (ii) of f then implies:

$$\begin{array}{rcl} \left|\mathbf{P}(X \in A) - P_\lambda(A)\right| & \leq & \frac{1 - e^{-\lambda}}{\lambda} \sum_{v \in V} \pi_v \mathbf{E}\left| \sum_{w \in V} I_w - \sum_{w \neq v} J_{vw} \right| \\ & \leq & \frac{1 - e^{-\lambda}}{\lambda} \sum_{v \in V} \pi_v \left\{ \pi_v + \sum_{w \neq v} \mathbf{E}\left|I_w - J_{vw}\right| \right\}. \end{array}$$

The upper bound does not depend on A, and is thus an upper bound on the variation distance $d_{var}(X, P_\lambda)$. □

Proof of Lemma 4.1.1 (i) Rearrangin terms in definition (4.4) de f yields:

$$f(i+1) = \frac{1}{\lambda P_\lambda(i)} \left\{ P_\lambda(A \cap C_i) P_\lambda(\overline{C_i}) - P_\lambda(A \cap \overline{C_i}) P_\lambda(C_i) \right\},$$

from which expression it easily follows that

$$|f(i+1)| \leq \frac{2\mathrm{P}_\lambda(\overline{C_i})}{\lambda P_\lambda(i)},$$

hence

$$|f(i+1)| \leq 2 \sum_{j>i} \lambda^{j-i-1} \frac{i!}{j!} \leq 2e^\lambda.$$

(ii) Consider first the case $i = 0$, for which:

$$f(1) - f(0) = f(1) = \frac{1}{\lambda} \left(\mathbf{1}_A(0) - P_\lambda(A) \right),$$

hence

$$|f(1) - f(0)| \leq \frac{1 - e^{-\lambda}}{\lambda}.$$

Next, for $i > 0$, write:

$$\begin{array}{rcl} f(i+1) - f(i) & = & \frac{P_\lambda(A \cap C_i) - P_\lambda(A) P_\lambda(C_i) - \frac{\lambda}{i} [P_\lambda(A \cap C_{i-1}) - P_\lambda(A) P_\lambda(C_{i-1})]}{\lambda P_\lambda(i)} \\ & = & \frac{P_\lambda(A \cap C_{i-1})}{\lambda P_\lambda(i)} \left\{ P_\lambda(\overline{C_i}) - \frac{\lambda}{i} P_\lambda(\overline{C_{i-1}}) \right\} \\ & & + \frac{P_\lambda(A \cap \{i\})}{\lambda P_\lambda(i)} \left\{ P_\lambda(\overline{C_i}) + \frac{\lambda}{i} P_\lambda(C_{i-1}) \right\} \\ & & + \frac{P_\lambda(A \cap \overline{C_i})}{\lambda P_\lambda(i)} \left\{ -P_\lambda(C_i) + \frac{\lambda}{i} P_\lambda(C_{i-1}) \right\}. \end{array}$$

Remark that the terms in curly brackets have the following signs:

$$\begin{array}{l} P_\lambda(\overline{C_i}) - \frac{\lambda}{i} P_\lambda(\overline{C_{i-1}}) = \sum_{j>i} e^{-\lambda} \lambda^j \left[\frac{1}{j!} - \frac{1}{(j-1)!} \frac{1}{i} \right] < 0, \\ P_\lambda(\overline{C_i}) + \frac{\lambda}{i} P_\lambda(C_{i-1}) > 0, \\ -P_\lambda(C_i) + \frac{\lambda}{i} P_\lambda(C_{i-1}) = e^{-\lambda} \left\{ -\sum_{j=0}^{i} \frac{\lambda^j}{j!} + \sum_{j=0}^{i-1} \frac{\lambda^{j+1}}{j!i} \right\} \leq e^{-\lambda} \sum_{j=1}^{i} \frac{\lambda^j}{(j-1)!} \left[\frac{1}{i} - \frac{1}{j} \right] < 0. \end{array}$$

This implies that the diffrence $f(i+1) - f(i)$ is maximal for $A = \{i\}$, and minimal for $A = \mathbb{N} \setminus \{i\}$. Thus:

$$f(i+1) - f(i) \leq \frac{1}{\lambda} \left\{ 1 - e^{-\lambda} + \sum_{j=1}^{i} e^{-\lambda} \frac{\lambda^j}{(j-1)!} \left[\frac{1}{j} - \frac{1}{i} \right] \right\} \leq \frac{1 - e^{-\lambda}}{\lambda},$$

and:

$$\begin{aligned} f(i+1) - f(i) &\geq -\frac{P_\lambda(\overline{C_i}) + \frac{\lambda}{i} P_\lambda(C_{i-1})}{\lambda} \\ &\geq -\frac{1-e^{-\lambda}}{\lambda}, \end{aligned}$$

which together imply (ii).

(iii) For $i = 0$, the relation holds in view of

$$\lambda f(1) = P_\lambda(A \cap \{0\})/P_\lambda(0) - P_\lambda(A).$$

For $i > 0$, one has:

$$\lambda f(i+1) - i f(i) = \frac{1}{P_\lambda(i)} \left\{ P_\lambda(A \cap C_i) - P_\lambda(A) P_\lambda(C_i) - P_\lambda(A \cap C_{i-1}) + P_\lambda(A) P_\lambda(C_{i-1}) \right\},$$

hence the result.

4.2 Emergence of connectivity in E-R graphs

4.2.1 The number of isolated nodes

Recall that the *degree* of a node is the number of its graph neighbours, and that a node is *isolated* if it has zero degree, i.e. no neighbours. Clearly, when the graph contains an isolated node, the graph cannot be connected; the study of the presence or not of isolated nodes will thus give us upper bounds on the probability that the graph is connected.

Let I_v equal 1 if node v is isolated, and 0 otherwise. Recalling that $\xi_{uv} \in \{0, 1\}$ indicates the presence of edge (u, v), it thus holds that

$$I_v = \prod_{w \neq v} \mathbf{1}_{\xi_{vw}=0},$$

The number of isolated nodes is thus:

$$X = \sum_v I_v.$$

We then have the following

Theorem 4.2.1. *Assume that for some fixed $c \in \mathbb{R}$, $Np = \log(N) + c + o(1)$. Then the distribution of X converges in variation, as $N \to \infty$, to the Poisson distribution with parameter e^{-c}.*

Proof. We use Stein-Chen's method. Set

$$J_{vw} = \prod_{u \neq v, w} \mathbf{1}_{\xi_{uw}=0},\ v \in \{1, \ldots, N\},\ w \neq v.$$

Note that conditioning on $I_v = 1$ is equivalent to conditioning on $\xi_{vw} = 0$ for all $w \neq v$. Conditionally on this event, all other edge indicator variables are left unaffected, by independence. We therefore have the equality of distributions

$$\mathcal{L}(\{J_{vw}\}_{w \neq v}) = \mathcal{L}(\{I_w\}_{w \neq v} | I_v = 1).$$

Let $\pi = \mathbf{E}(I_v) = (1-p)^{N-1}$, and $\lambda = \mathbf{E}(X) = N\pi$. Equation (4.3) gives us

$$d_{var}(X, P_\lambda) \leq \min(1, \lambda^{-1})\, N\pi \left(\pi + \sum_{w \neq v} \mathbf{E}|I_w - J_{vw}| \right).$$

Note that

$$\mathbf{E}|I_w - J_{vw}| = \mathbf{E}\left(\xi_{vw} \prod_{u \neq v,w} (1 - \xi_{uw}) \right) = p(1-p)^{N-2}.$$

This then yields

$$d_{var}(X, P_\lambda) \leq \min(1, \lambda^{-1})\lambda \left(\pi + p(N-1)\pi/(1-p)\right) \leq \pi + \lambda \frac{p}{1-p}.$$

As $N \to \infty$, $N\pi \to e^{-c}$, and thus this upper bound is of order $e^{-c}\log(N)/N$ as $N \to \infty$, hence the convergence

$$\lim_{N \to \infty} d_{var}(X, P_\lambda) = 0.$$

The triangle inequality further yields

$$d_{var}(X, P_{e^{-c}}) \leq d_{var}(P_{e^{-c}}, P_\lambda) + d_{var}(X, P_\lambda).$$

It remains to show that $d_{var}(P_{e^{-c}}, P_\lambda) \to 0$ as $N \to \infty$. This will follow from the Lemma below. □

LEMMA 4.2.1: For $\lambda, \lambda' \geq 0$, it holds that:

$$d_{var}(P_\lambda, P_{\lambda'}) \leq 2|\lambda - \lambda'|.$$

Proof. Assume without loss of generality that $\lambda \geq \lambda'$. Let X, Y be two independent Poisson random variables with respective parameters λ', $\lambda - \lambda'$. Thus $X' := X + Y$ is Poisson with parameter λ. For any $A \subset \mathbb{N}$, write

$$\begin{aligned} 2|P_\lambda(A) - P_{\lambda'}(A)| &= 2|\mathbf{P}(X + Y \in A) - \mathbf{P}(X \in A)| \\ &= 2|\mathbf{P}(Y = 0)\mathbf{P}(X \in A) - \mathbf{P}(X \in A) + \mathbf{P}(Y > 0, X + Y \in A)| \\ &\leq 2\mathbf{P}(Y > 0). \end{aligned}$$

The latter upper bound is independent of A, and equal to $2(1 - e^{-|\lambda - \lambda'|})$. The result follows. □

4.2.2 Connectivity

Theorem 4.2.2. *Let $c \in \mathbb{R}$ be given, and assume that p is such that $Np = \log(N) + c + o(1)$. One then has the limit*

$$\lim_{N\to\infty} \mathbf{P}(\mathcal{G}(N,p) \text{ connected}) = e^{-e^{-c}}. \tag{4.5}$$

Proof. We first show that the probability that the graph contains connected components of sizes between 2 and $N/2$ goes to zero under the theorem's assumptions. This will entail that the probability that the graph is connected is asymptotically equivalent to the probability that it has no isolated edges. By the previous theorem 4.2.1, it holds that

$$\mathbf{P}(\mathcal{G}(N,p) \text{ has no isolated nodes}) = \mathbf{P}(X=0) \sim e^{-e^{-c}},$$

hence the announced result.

Let us first check that with high probability, the graph has no connected components of size 2:

$$\begin{aligned}\mathbf{P}(\exists\text{connected component of size 2}) &\leq \tfrac{N(N-1)}{2}\mathbf{P}((u,v)\text{connected component of } \mathcal{G}(N,p)) \\ &= \tfrac{N(N-1)}{2}p(1-p)^{2(N-2)}.\end{aligned}$$

By the inequality $1-x \leq e^{-x}$, this expression is not larger than

$$N^2 \frac{p}{(1-p)^4} e^{-2Np}.$$

This last term is of order $p \sim \log(N)/N$, and thus asymptotically negligible.

We now show that the probability of having a connected component of size between 3 and $N/2$ vanishes asymptotically. To this end, we shall rely on Cayley's theorem (see below) according to which the number of trees on a set of r labelled nodes is exactly r^{r-2}. For any $r \in \{3,\dots,N/2\}$, and an arbitrary set C of r nodes, the probability that this set is a connected component of $\mathcal{G}(N,p)$ is then not larger than

$$\sum_{\text{trees } \mathcal{T} \text{ on } C} \mathbf{P}(\text{edges in } \mathcal{T} \text{ present and no edge between } C \text{ and } \bar{C}) \leq r^{r-2}p^{r-1}(1-p)^{r(N-r)}.$$

Summing over all $r \in \{3,\dots,N/2\}$ and all size r components C, one obtains the upper bounded on the probability π that there is a connected component of such a size in $\mathcal{G}(N,p)$:

$$\pi \leq \sum_{r=3}^{\lceil N/2\rceil} \binom{N}{r} r^{r-2}p^{r-1}(1-p)^{r(N-r)}.$$

Using $\binom{N}{r} \leq N^r/r!$ and Stirling's formula, one obtains:

$$\pi \leq \sum_{r=3}^{\lceil N/2\rceil} N^r \frac{1}{\sqrt{r}} \left(\frac{e}{r}\right)^r r^{r-2}p^{r-1}e^{-pr(N-r)}.$$

Upon simplification, using the fact that $N - r \geq N/2$ for r in the summation range, this yields:

$$\pi \leq \frac{1}{p} \sum_{r=3}^{\lceil N/2 \rceil} r^{-5/2} e^{r\{1+\log(Np)-(1/2)Np\}}.$$

The exponent in curly brackets is equivalent to $(1/2)Np$; also, $1/p \leq N$. Thus, for all $\epsilon > 0$, one has:

$$\begin{aligned} \pi &\leq N \sum_{r=3}^{\lceil N/2 \rceil} r^{-5/2} e^{-r(1/2-\epsilon)Np} \\ &\leq N \sum_{r \geq 3} e^{-r(1/2-\epsilon)Np} \\ &= N \frac{e^{-3(1/2-\epsilon)Np}}{1-e^{-(1/2-\epsilon)Np}} \\ &= O\left(N^{-1/2+3\epsilon}\right), \end{aligned}$$

which concludes the proof of the theorem. □

4.2.3 Cayley's theorem

We now show that the number of trees on a set $\{0, \ldots, r-1\}$ is r^{r-2}, that is the statement of Cayley's theorem. To this end we construct a bijection between sequences of $r-2$ integers in $\{0, \ldots, r-1\}$ and the set of trees on this set.

Given a tree $\mathcal{T}$, one determines $\phi(\mathcal{T}) = \{w_1, \ldots, w_{r-2}\}$ as follows. By convention, node 0 is the root of the tree. The degree 1 nodes distinct from the root are called the "leaves" of the tree. One starts by picking the leaf with the smallest label, say v_1, and let w_1 be the label of the node v_1 attaches to. One then examines the leaves of the reduced tree $\mathcal{T} \setminus \{v_1\}$, picks its smallest leaf, say v_2, and then w_2 is the label of the neighbour of v_2 in $\mathcal{T} \setminus \{v_1\}$. One repeats this procedure until only one leaf remains, that is after having extracted $r-2$ labels.

To show that function ψ is bijective, let us see how the tree $\mathcal{T}$ is determined by $\psi(\mathcal{T})$. The first leaf v_1 that has been removed must be the smallest integer in $\{1, \ldots, r-1\}$ which does not appear in the sequence $\psi(\mathcal{T}) = \{w_1, \ldots, w_{r-2}\}$. One can thus reconstruct the edge (v_1, w_1). Similarly, the node removed next, v_2, is the smallest integer in $\{1, \ldots, r-1\} \setminus \{v_1\}$ which does not appear in $\{w_2, \ldots, w_{r-2}\}$, and we can thus reconstruct the edge (v_2, w_2). Iterating, we next reconstruct $(v_3, w_3), \ldots, (v_{r-2}, w_{r-2})$. There only remains one element in $\{1, \ldots, r-1\} \setminus \{v_1, \ldots, v_{r-2}\}$, say v_{r-1}. Necessarily, this last element is connected to the root 0, hence the edge $(v_{r-1}, 0)$ is also in tree $\mathcal{T}$. We have thus reconstructed the $r-1$ edges of the tree $\mathcal{T}$.

Chapter 5

Diameter of Erdős-Rényi graphs

5.1 Introduction

Given a graph G, and two vertices u, v, the graph distance $d_G(u,v)$ is by definition the minimal length (in number of hops) of a path connecting u to v. The diameter of the graph G is then defined as the supremum over pairs of nodes u, v of the distance $d_G(u,v)$. It is denoted $D(G)$:

$$D(G) = sup_{\text{vertices } u,v} d_G(u,v).$$

In the case of the E-R graph, recall that it captures the dynamics of the Reed-Frost epidemics started at a source node u, with $\Gamma_t(u) := \{v : d_G(u,v) = t\}$ representing the set of infectious nodes after t time steps. The epidemics will infect all nodes if and only if the graph is connected, which is equivalent to the diameter being finite. In that case, the diameter provides an upper bound on the time it takes for the Reed-Frost epidemics to reach all nodes.

The diameter of a graph is also interesting when the graph represents a network over which goods (information e.g.) need to be transported: it then gives an upper bound on the time for goods to travel from any location u to any other location v, provided shortest paths between locations are used.

The following result illustrates an interesting relation between the diameter, the number of nodes and the maximal node degree of a graph:

LEMMA 5.1.1: Given a graph G on N nodes, such that the maximal node degree is at most Δ, its diameter D verifies

$$N \leq 1 + \Delta \frac{(\Delta - 1)^D - 1}{\Delta - 2}.$$

Equivalently, one has

$$D \geq \frac{\log\left(N\left[1 - \frac{2}{\Delta}\right] + \frac{2}{\Delta}\right)}{\log(\Delta - 1)}.$$

Proof. Recall that $d_i(u) = |\Gamma_i(u)|$. It is easily seen that, when the maximal node degree is at most Δ, for any u, $d_1(u) \leq \Delta$, and for all $i \geq 2$, $d_i(u) \leq \Delta(\Delta-1)^{i-1}$. If the diameter is not larger than D, then necessarily,

$$N = 1 + d_1(u) + \ldots + d_D(u) \leq 1 + \Delta\left\{1 + (\Delta-1) + \ldots + (\Delta-1)^{D-1}\right\}.$$

The result follows. □

An important class of graphs that approaches this bound is that of de Bruijn graphs. For two integers k, ℓ, the de Bruijn graph $B(k,\ell)$ has k^ℓ nodes, that are identified with the ℓ-letter words on the alphabet $\{1,\ldots,k\}$. A node $x_1 \ldots x_\ell$ is connected to nodes $yx_1 \ldots x_{\ell-1}$ and $x_2 \ldots x_\ell y$ for all $y \in \{1,\ldots,k\}$. Thus its degree is at most $\Delta = 2k$. Finally, between any two nodes $x_1 \ldots x_\ell$, $y_1 \ldots y_\ell$ there is an ℓ-hop path, going through $x_2 \ldots x_\ell y_1$, $x_3 \ldots x_\ell y_1 y_2$, …, $x\ell y_1 \ldots y_{\ell-1}$.

For the de Bruijn graph $B(k,\ell)$ we thus have

$$D \leq \ell = \frac{\log(N)}{\log(k)} = \frac{\log(N)}{\log(\Delta) - \log(2)}.$$

As we shall see now, E-R graphs also achieve a diameter close to optimal, given their maximal degree and number of nodes.

5.2 Diameter of E-R graphs

Theorem 5.2.1. *Let $\delta = (N-1)p$ denote the average node degree of $\mathcal{G}(N,p)$. Assume that*

$$\log(N) << \delta << \sqrt{N}. \tag{5.1}$$

Noting

$$D' = \left\lceil \frac{\log(N)}{2\log(\delta)} \right\rceil, \tag{5.2}$$

it holds that:

$$\lim_{N\to\infty} \mathbf{P}\left(D(\mathcal{G}(N,p)) \in \{2D'-3, 2D'-2, \ldots, 2D'+1\}\right) = 1.$$

Thus, the diameter of $\mathcal{G}(N,p)$ takes with high probability at most 5 distinct values in the assumed parameter range.

Remark 5.2.1. *The best possible result, given in Bollobás [4], establishes that in fact D can take at most two values, and identifies the probability of each value occuring.*

5.3 Key lemma: control of neighborhood growth

Given some $\epsilon > 0$, define the quantities

$$d_j^{\pm} = \begin{cases} (1 \pm \epsilon)^j \delta^j & \text{if } j = 1, 2, \\ (1 \pm \epsilon)^2 \left(1 \pm \frac{\epsilon}{\delta}\right)^{j-2} \delta^j & \text{if } j = 3, \ldots, D'. \end{cases} \tag{5.3}$$

The key ingredient in the proof of the theorem is the following:

LEMMA 5.3.1: Let $\epsilon > 0$ be fixed. Define for all $u \in \{1, \ldots, N\}$ and all $i = 1, \ldots, D'$, the event $\mathcal{E}_i(u)$ by

$$\mathcal{E}_i(u) = \{d_i^- \leq d_i(u) \leq d_i^+\}.$$

Assume that condition (5.1) holds. Then for any fixed $K > 0$, for large enough N, it holds that:

$$\mathbf{P}(\mathcal{E}_i(u)) \geq 1 - N^{-K}, \quad u \in \{1, \ldots, N\},\ i = 1, \ldots, D'. \tag{5.4}$$

Proof. Let $u \in \{1, \ldots, N\}$ and $i \in \{1, \ldots, D'\}$ be fixed. Note that, conditionally on $d_1(u), \ldots, d_{i-1}(u)$, $d_i(u)$ admits a binomial distribution with parameters:

$$\mathcal{L}(d_i(u) | d_1(u), \ldots, d_{i-1}(u)) = \text{Bin}\left(N - 1 - d_1(u) - \ldots - d_{i-1}(u), 1 - (1-p)^{d_{i-1}(u)}\right).$$

Denote by $\bar{\mathcal{E}}_i(u)$ the complementary event of $\mathcal{E}_i(u)$. From the above it readily follows that:

$$\begin{aligned} \mathbf{P}(\bar{\mathcal{E}}_i(u) | \mathcal{E}_1(u), \ldots, \mathcal{E}_{i-1}(u)) \leq \quad & \mathbf{P}\left(\text{Bin}(N, 1 - (1-p)^{d_{i-1}^+}) \geq d_i^+\right) \\ & + \mathbf{P}\left(\text{Bin}(N - 1 - d_1^+ - \ldots - d_{i-1}^+, 1 - (1-p)^{d_{i-1}^-}) \leq d_i^-\right). \end{aligned} \tag{5.5}$$

Note that, for all $j < D'$, one has:

$$\begin{aligned} d_j^- \leq d_j^+ \leq d_{D'-1}^+ \quad & \leq \left(\tfrac{\delta(1+\epsilon)}{\delta+\epsilon}\right)^2 (\delta + \epsilon)^{\log(N)/(2\log(\delta))} \\ & \leq 4\sqrt{N} e^{\log(N)/(\delta \log(\delta))} \\ & = O(\sqrt{N}), \end{aligned} \tag{5.6}$$

in view of the assumption that $\delta >> \log(N)$. This can be used to establish the following estimates as $N \to \infty$:

$$\begin{aligned} N\left(1 - (1-p)^{d_{i-1}^+}\right) & = (1 + o(1))\delta d_{i-1}^+, \\ \left(N - 1 - d_1^+ - \ldots - d_{i-1}^+\right)\left(1 - (1-p)^{d_{i-1}^-}\right) & = (1 + o(1))\delta d_{i-1}^-, \end{aligned}$$

where the $o(1)$ term can be chosen independently of $i \leq D'$. Indeed, the term $1-(1-p)^{d_{i-1}^{\pm}}$ also reads

$$\begin{aligned} 1 - \exp\left(-p d_{i-1}^{\pm} + O(p^2 d_{i-1}^{\pm})\right) & = p d_{i-1}^{\pm} + O\left((p d_{i-1}^{\pm})^2\right) \\ & = p d_{i-1}^{\pm} + O(p^2 N) \\ & = p d_{i-1}^{\pm} + O(\delta^2 / N), \end{aligned}$$

and the latter term is $o(1)$ as we assumed $\delta = o(\sqrt{N})$. Furthermore, the term $N - 1 - d_1^+ - \ldots d_{i-1}^+$ is larger than $N - D'O(\sqrt{N}) = N - O(\sqrt{N}\log(N))$, and thus equals $N(1 + o(1))$.

The Chernoff bound lemma 3.2.2 for sums of independent $\{0,1\}$-valued random variables, applied to the right-hand side of (5.5), thus yields

$$\mathbf{P}(\bar{\mathcal{E}}_i(u)|\mathcal{E}_1(u),\ldots,\mathcal{E}_{i-1}(u)) \leq e^{-(1+o(1))\delta d_{i-1}^- h(-\epsilon_i)} + e^{-(1+o(1))\delta d_{i-1}^+ h(\epsilon_i)}, \tag{5.7}$$

where $h(x) = (1+x)\log(1+x) - x$, and

$$\epsilon_i = \frac{d_i^+}{\delta d_{i-1}^+} - 1 = 1 - \frac{d_i^-}{\delta d_{i-1}^-},$$

and is given by ϵ for $i = 1, 2$, and by ϵ/δ for $i > 2$.

For $i = 1$ or 2, the exponents in the right-hand side of (5.7) are larger than $c\delta$ for some constant $c > 0$. Since we have assumed that $\delta >> \log(N)$, for any fixed $K > 0$, the right-hand side of (5.7) is, in this case, less than N^{-K} for N large enough.

For $i > 2$, using the fact that $h(x) = (1 + 0(x))x^2/2$, the exponents in the right-hand side of (5.7) are in this case larger than $c\delta\delta_{i-1}^{\pm}\delta^{-2}$ for some constant $c > 0$, and thus again of order at least δ. Thus for $i > 2$ as well, for any fixed $K > 0$, the right-hand side of (5.7) is less than N^{-K} for N large enough.

The claim of the lemma is finally established by writing:

$$\begin{array}{ll} \mathbf{P}(\mathcal{E}_1(u),\ldots,\mathcal{E}_i(u)) & \geq \mathbf{P}(\mathcal{E}_1(u),\ldots,\mathcal{E}_{i-1}(u)) - \mathbf{P}(\bar{\mathcal{E}}_i(u)|\mathcal{E}_1(u),\ldots,\mathcal{E}_{i-1}(u)) \\ & \geq 1 - \sum_{j=1}^{i} \mathbf{P}(\bar{\mathcal{E}}_j(u)|\mathcal{E}_1(u),\ldots,\mathcal{E}_{j-1}(u)) \\ & \geq 1 - D'N^{-K} \end{array}$$

for N large enough. The result follows. □

5.4 Proof of Theorem 5.2.1

5.4.1 The upper bound

Let us first establish that, with high probability, $D(\mathcal{G}(N,p)) \leq 2D'+1$. For two arbitrary nodes u, v, note that:

$$\mathbf{P}(d_G(u,v) > 2D'+1|\Gamma_1(u),\ldots,\Gamma_{D'}(u),\Gamma_1(v),\ldots,\Gamma_{D'}(v)) \leq (1-p)^{d_{D'}(u)d_{D'}(v)}.$$

Indeed, given the neighborhoods $\Gamma_i(u), \Gamma_i(v)$ for $i = 1,\ldots,v$, either they have non empty intersection, in which case $d_G(u,v) \leq 2D'$, or they do not intersect, in which case all of the $d_{D'}(u)d_{D'}(v)$ edges must be absent for $d_G(u,v) > 2D'+1$ to hold. We thus obtain:

$$\mathbf{P}(d_G(u,v) > 2D'+1) \leq \mathbf{P}(\bar{\mathcal{E}}_{D'}(u)) + \mathbf{P}(\bar{\mathcal{E}}_{D'}(v)) + (1-p)^{2(d_{D'}^-)^2}.$$

The last term in the right-hand side is evaluated as follows:

$$\begin{aligned}(1-p)^{2(d^-_{D'})^2} &\leq \exp\left(-2p\left[\tfrac{1-\epsilon}{1-\epsilon/\delta}\right]^4(\delta-\epsilon)^{\log(N)/\log(\delta)}\right)\\ &\leq \exp\left(-\Omega(1)pN^{1-\epsilon/(\delta\log(\delta))(1+o(1))}\right)\\ &\leq \exp\left(-\delta\Omega(1)\right).\end{aligned}$$

Since $\delta >> \log(N)$, combined with the Lemma's result, we get that for all $K > 0$, and sufficiently large N,

$$\mathbf{P}(d_G(u,v) > 2D' + 1) \leq N^{-K}.$$

Finally,

$$\mathbf{P}(D(\mathcal{G}(N,p)) > 2D' + 1) \leq \sum_{u\neq v}\mathbf{P}(d_G(u,v) > 2D' + 1) \leq N^2 \times N^{-K}.$$

By choosing $K > 2$, the desired upper bound follows.

5.4.2 The lower bound

Its proof relies on the following lemma:

LEMMA 5.4.1: Given a set of N items, and two subsets C_1, C_2 both of size $r << N$, selected independently, and uniformly at random from sets of that size, it holds that

$$\mathbf{P}(C_1 \cap C_2 = \emptyset) = (1+o(1))\exp\left(-\frac{r^2}{N} + O(r^3/(N-2r)^2))\right).$$

Proof. The probability that the intersection is empty equals

$$\frac{\binom{N-r}{r}}{\binom{N}{r}} = \frac{(N-r)!(N-r)!}{N!(N-2r)!}.$$

Stirling's formula yields the following equivalent:

$$\mathbf{P}(C_1 \cap C_2 = \emptyset) = (1+o(1))\exp\left((N-2r)\log(1+r/(N-2r)) + N\log(1-r/N)\right),$$

hence the result. □

Let $C = D'-2$. Conditioning on the neighborhood sizes $d_1(u),\ldots,d_C(u), d_1(v),\ldots,d_C(v)$, we have that:

$$\mathbf{P}(d_G(u,v) \leq 2C | d_1(u),\ldots,d_C(u),d_1(v),\ldots,d_C(v)) \leq \mathbf{P}(C_1 \cap C_2 \neq \emptyset),$$

where C_1 and C_2 are sampled as in the previous lemma from an N-items set, and have sizes

$$|C_1| = 1 + d_1(u) + \ldots + d_C(u),\ |C_2| = 1 + d_1(v) + \ldots + d_C(v).$$

Indeed, we may construct the successive neighborhoods for the two nodes u, v as follows. First determine $\Gamma_1(u), \ldots, \Gamma_C(u)$. Let $\mathcal{N}'$ be the set of nodes not in any of these neighborhoods. Conditionally on these neighborhoods, there are no edges between $\Gamma_i(u)$ and $\mathcal{N}'$ for all $i = 1, \ldots, C-1$; the distribution of the edges internal to $\mathcal{N}'$, and between $\mathcal{N}'$ and $\Gamma_C(u)$ is unaffected by the conditioning. For convenience, we also introduce the notation $\mathcal{N}'' := \mathcal{N}' \cup \Gamma_C(u)$.

Introduce the notation:

$$\mathcal{B}_k(v) = \cup_{0 \leq i \leq k} \Gamma_i(v).$$

Pick now v at random from the total node set. If it falls in $\mathcal{N}'$, construct its first neighborhood $\Gamma_1(v)$ by picking at random a set of size $d_1(v)$ from $\mathcal{N}'' \setminus \{v\}$. If this does not intersect $\Gamma_C(v)$, choose $\Gamma_2(v)$ as a random set of size $d_2(v)$ taken from $\mathcal{N}'' \setminus \mathcal{B}_1(v)$. Proceed until either all C neighborhoods are constructed, or one does intersect $\Gamma_C(v)$.

The probability that this occurs is easily seen to be as claimed the probability of intersection of two independently, randomly selected sets of sizes $|\mathcal{B}_C(u)|$ and $|\mathcal{B}_C(v)|$ respectively.

To conclude, by Lemmas 5.4.1 and 5.3.1, it holds that:

$$\begin{aligned}
\mathbf{P}(d_G(u,v) \leq 2C) &\leq \textstyle\sum_{i=1}^{C} \mathbf{P}(\bar{\mathcal{E}}_i(u)) + \mathbf{P}(\bar{\mathcal{E}}_i(v)) + O\left(\frac{(1+d_1^+ + \ldots + d_C^+)^2}{N}\right) \\
&\leq 2\log(N) N^{-K} + O\left(\frac{(d_{D'-2}^+)^2}{N}\right) \\
&\leq \log(N) N^{-K} + O((\delta+\epsilon)^{-2}),
\end{aligned}$$

where we have used the fact (5.6) that $d_{D'-1}^+ = O(\sqrt{N})$. Thus the right-hand side in the last display goes to zero as $N \to \infty$.

Chapter 6

Small worlds

6.1 Background

In 1967, the sociologist Stanley Milgram published [11] results of a letter-relaying experiment of his design. The now famous experiment required a source individual to forward a letter to a destination individual, about whom were disclosed informations such as address, name and profession. However, each source individual was forbidden to post the letter directly to the target person. Instead she was required to forward the letter to someone known on a first-name basis, who in turn was allowed to forward it only to such familiar contacts.

The outcome was that a significant fraction of letters reached their destination. Moreover, they did so in at most six hops, justifying the term "six degrees of separation". This fact is also often referred to as the "small world phenomenon".

Viewing the social world as a graph, with edges between acquainted persons, certainly if any individual could relay information to any other in a small number of hops as in Milgram's experiment, the corresponding graph must have a small diameter. As we just saw, the E-R graph does have a small diameter (logarithmic in the number of nodes). However this is not a realistic model of social graphs, as it does not possess any structure, whereas social graphs are certainly affected by geographic location of individuals and area of professional activity among many factors. The model we discuss in this section is a variant of some of the models studied by physicists Strogatz and Watts (see e.g. [15]) to illustrate how graphs with spatial structure can also exhibit small diameters, hence providing more realistic models of social graphs.

6.2 Small world according to Strogatz and Watts

Given some integer $m > 0$, the node set of the graph is the set of points $(i, j) \in \{1, \ldots, m\} \times \{1, \ldots, m\}$. There are thus $N = m^2$ nodes. These nodes are connected via two types of edges: local edges to grid neighbours (see Figure 6.1), and thus between 2 and 4 such edges per node (accounting for boundary effects), and "shortcut" edges. The latter are generated as follows. The model features a second parameter, $p \in]0, 1[$.

Figure 6.1: Strogatz-Watts graph, and induced graph G'

Each node u, with probability p, creates a shortcut edge, whose other end-point is chosen uniformly at random from the node set. We denote by $SW(N,p)$ the corresponding random graph.

Note that without the addition of shortcuts, the diameter of the graph would be $2m = 2\sqrt{N}$. We now show that for any fixed $p > 0$, the presence of shortcuts will radically reduce the graph's diameter:

Theorem 6.2.1. *Let $p > 0$ be fixed. Then for some constant A depending on p, the diameter $D(SW)$ of the graph SW, verifies*

$$\lim_{N\to\infty} \mathbf{P}(D(SW(N,p)) \leq A\log(N)) = 1.$$

Proof. For some integer k such that $kp > 1$, cut the grid into squares of k nodes (we assume here that $\sqrt{k}$ is an integer which divides m; this restriction looses no generality; exercise...). We shall work on the graph G' whose nodes correspond to these squares; neighboring squares are connected by local edges; each square generates a random number of shortcut edges, whose distribution is now Binomial with parameters (k,p). The other end-points of these shortcuts are uniformly drawn from the node set of G'. We shall denote by $\mathcal{N}'$ the vertex set of G', and let $N' = N/k$ be the number of vertices.

Note that the diameters of SW and G' satisfy:

$$D(SW) \leq 2\sqrt{k}(2 + D(G')).$$

Indeed, given two nodes u', v' of SW, let u, v denote the corresponding nodes in G'. Given a path in G' of length not larger than $D(G')$, one can construct from it a path with not more than $2\sqrt{k}(D(G')+2)$ hops, hence the result. From now on we aim to bound the diameter of G'.

Let $\Gamma_1(u)$ denote a group of nodes containing u, of size $C\log(N)$ for some suitable constant C, and such that the nodes of $\Gamma_1(u)$ are all connected via grid edges. We shall denote by $\Gamma_2(u)$ the nodes reached from nodes in $\Gamma_1(u)$ via shortcuts generated from $\Gamma_1(u)$, and similarly we define $\Gamma_i(u)$ for $i > 2$ as the sets of new nodes reached by shortcuts created within $\Gamma_{i-1}(u)$. As in the study of the E-R graph, for vertices $u \in \mathcal{N}'$ of G', and $i \geq 0$, we note

$$d_i(u) = |\Gamma_i(u)|.$$

The following result is the counterpart of Lemma 5.3.1 in the present context:

LEMMA 6.2.1: Let $\epsilon > 0$ be fixed, such that

$$kp(1-\epsilon) > 1, \tag{6.1}$$

and

$$\frac{\log(kp(1+\epsilon))}{\log(kp(1-\epsilon))} < 2. \tag{6.2}$$

The constant $C > 0$ can be chosen so that, for all $u \in \mathcal{N}'$, with probability $1 - o(N^{-2})$, the following inequalities hold:

$$kp(1-\epsilon) \leq \frac{d_i(u)}{d_{i-1}(u)} \leq kp(1+\epsilon), \quad i = 2, \ldots, D, \tag{6.3}$$

where $D = \lceil \log(N)/2\log(kp(1-\epsilon)) \rceil + 1$.

Before giving the Lemma's proof, we state the so-called Azuma-Hoeffding inequality, which will relay the Chernoff bound 3.2.2, since we need to deal with correlated indicator variables rather than independent ones.

Theorem 6.2.2. *(Azuma-Hoeffding inequality) Let $\{M_t\}_{t=0,\ldots,T}$ be a martingale such that for all $t = 1, \ldots, T$,*

$$|M_t - M_{t-1}| \leq c_T \text{ almost surely} \tag{6.4}$$

for positive constants $c_1, \ldots, c_T$. Then for all $x > 0$, one has

$$\mathbf{P}(M_T - M_0 \geq x) \leq \exp\left(-\frac{x^2}{2\sum_{t=1}^{T} c_t^2}\right). \tag{6.5}$$

Proof. Fix $\theta > 0$. Tchebitchev's inequality yields:

$$\mathbf{P}(M_T - M_0 \geq x) \leq \mathbf{E}\left[\exp(\theta(M_{T-1} - M_0))\mathbf{E}\left(\exp(\theta(M_T - M_{T-1}))|\mathcal{F}_{T-1}\right)\right] e^{-\theta x}, \tag{6.6}$$

where $\mathcal{F}_t = \sigma(M_0, \ldots, M_t\}$. Note that, in view of (6.4), for some $Z \in [0,1]$, one has:

$\left(N' - 1 - d_1(u) - \ldots - d_{i-1}(u)\right)(1 - (1 - 1/N')^T).M_T - M_{T-1} = Zc_T + (1-Z)(-c_T)$.

Thus by convexity of the function $y \to e^{\theta y}$, one has

$$\exp(\theta(M_T - M_{T-1})) \leq Ze^{\theta c_T} + (1-Z)e^{-\theta c_T}.$$

Furthermore, by the martingale property $\mathbf{E}(M_T - M_{T-1}|\mathcal{F}_{T-1}) = 0$. Equivalently, $\mathbf{E}(Z|\mathcal{F}_{T-1}) = 1/2$, so that

$$\mathbf{E}\left(\exp(\theta(M_T - M_{T-1}))|\mathcal{F}_{T-1}\right) \leq \frac{e^{\theta c_T} + e^{-\theta c_T}}{2}.$$

Expanding the right-hand side in power series of θc_T, it can be seen that this is less than or equal to $\exp((\theta c_T)^2/2)$. Repeating the argument, we obtain from (6.6):

$$\mathbf{P}(M_T - M_0 \geq x) \leq \exp\left(\frac{\theta^2 \sum_{t=1}^{T} c_t^2}{2}\right)\exp(-\theta x).$$

Optimising over $\theta > 0$ yields the result (6.5) □

A useful corollary is the following:

Corollary 6.2.1. *Let $f : \Omega_1 \times \cdots \times \Omega_T \to \mathbb{R}$ be a measurable function such that, for all $x_1, \ldots, x_T \in \Omega_1 \times \cdots \times \Omega_T$, all $t \in \{1, \ldots, T\}$, and all $y_t \in \Omega_T$, one has:*

$$|f(x) - f(x_1, \ldots, x_{t-1}, y_t, x_{t+1}, \ldots, x_T)| \leq c_t. \tag{6.7}$$

Then, given independent random variables $X_1, \ldots, X_T$ taking their values in $\Omega_1, \ldots, \Omega_T$ respectively, the random variable $Y := f(X_1, \ldots, X_T)$ verifies for all $x > 0$:

$$\mathbf{P}(Y - \mathbf{E}(Y) \geq x) \leq \exp\left(-\frac{x^2}{2\sum_{t=1}^T c_t^2}\right).$$

Proof. Let for all $t = 0, \ldots T$: $M_t = \mathbf{E}(Y|X_1, \ldots, X_T)$. In particular, $M_0 = \mathbf{E}(Y)$ and $M_T = Y$. Let us verify that the martingale $\{M_t\}_{0 \leq t \leq T}$ satisfies assumption (6.4). Let $p_t(\cdot)$ denote the probability distribution of X_t. For all $t = 1, \ldots, T$, one has

$$\begin{array}{rl} |M_t - M_{t-1}| & = \left| \int_{\Omega_t \times \cdots \times \Omega_T} p_t(dy_t) \times \cdots \times p_T(dy_T) \left[f(X_1^t y_{t+1}^T) - f(X_1^{t-1} y_t^T) \right] \right| \\ & \leq c_t \end{array}$$

in view of assumption (6.7), where we used the notation

$$(x_1^t y_{t+1}^T) = (x_1, \ldots, x_t, y_{t+1}, \ldots, y_T).$$

The result follows. □

We now return to the proof of Lemma 6.2.1. Note that, conditionally on $d_1(u), \ldots, d_{i-1}(u)$, the number of shortcuts generated from $\Gamma_{i-1}(u)$ follows a binomial distribution with parameters $(kd_i(u), p)$. Let T denote this number of shortcuts.

Condition on this number T, and on the sets $\Gamma_1(u), \ldots, \Gamma_{i-1}(u\ )$, and let $A_t, \ldots, A_T$ denote the locations to which these shortcuts connect. Writing $d_i(u) = f(A_1, \ldots, A_T)$, it is readily seen that function f satisfies condition 6.7 of Corollary 6.2.1 with $c_t = 1$. Indeed, $d_i(u)$ counts the number of occupied bins, among $N' - 1 - d_1(u) - \ldots - d_{i-1}(u)$ available, after throwing T balls at random. Changing the location of one ball can change the number of occupied bins by at most 1. We thus have:

$$\mathbf{P}(d_i(u) - \bar{d}_i(u) \geq x | T, \Gamma_1(u), \ldots, \Gamma_{i-1}(u)) \leq \exp\left(-\frac{x^2}{2T}\right),$$

where $\bar{d}_i(u) = \mathbf{E}(d_i(u)|T, \Gamma_1(u), \ldots, \Gamma_{i-1}(u))$. The same upper bound holds for deviations of $d_i(u) - \bar{d}_i(u)$ below $-x$, by symmetry. Note that

$$\bar{d}_i(u) = \big(N' - 1 - d_1(u) - \ldots - d_{i-1}(u)\big)\,(1 - (1 - 1/N')^T).$$

Denote by $\mathcal{E}_i(u)$ the event

$$\mathcal{E}_i(u) = \{(1 - \epsilon)kp \leq d_i(u)/d_{i-1}(u) \leq (1 + \epsilon)kp\}.$$

Let $\mathcal{I} := [kpd_{i-1}(u)(1-\epsilon/2), kpd_{i-1}(u)(1+\epsilon/2)]$. We then have:

$$\begin{aligned}\mathbf{P}\left(\bar{\mathcal{E}}_i|\mathcal{E}_2(u),\ldots,\mathcal{E}_{i-1}(u)\right) \leq\ & \mathbf{P}(T\notin\mathcal{I}|\mathcal{E}_2(u),\ldots,\mathcal{E}_{i-1}(u)) \\ & +\mathbf{P}\left(\tfrac{d_i(u)}{d_{i-1}(u)}\geq kp(1+\epsilon)|T\in\mathcal{I},\mathcal{E}_2(u),\ldots,\mathcal{E}_{i-1}(u)\right) \\ & +\mathbf{P}\left(\tfrac{d_i(u)}{d_{i-1}(u)}\leq kp(1-\epsilon)|T\in\mathcal{I},\mathcal{E}_2(u),\ldots,\mathcal{E}_{i-1}(u)\right).\end{aligned} \tag{6.8}$$

On the event $\cap_{j=2}^{i-1}\mathcal{E}_j(u)$, since $kp(1-\epsilon)>1$ by assumption (6.1), it holds that $d_{i-1}(u)\geq C\log(N)$. Thus, by Lemma 3.2.2, the first term in the right-hand side of the above expression is bounded by $\exp(-C\log(N)h(\epsilon/2))$. For suitably large C, this is less than N^{-K} for any desired $K>0$.

Azuma-Hoeffding inequality yields the following bound on the second term in the right-hand side:

$$\sup_{T\in\mathcal{I}}\exp(-x^2/(2T)),$$

where

$$x = d_{i-1}(u)kp(1+\epsilon) - \left(N' - 1 - d_1(u) - \ldots - d_{i-1}(u)\right)\left(1-(1-1/N')^T\right).$$

Using Condition (6.2), it can be shown that for all $i\leq D$, on the event $\cap_{j=2}^{i-1}\mathcal{E}_j(u)$, one has $d_{i-1}(u)=o(N)$. It follows that x, in above expression, is lower-bounded by $d_{i-1}(u)kp(1+\epsilon)-(1+o(1))T$, and eventually the Azuma-Hoeffding bound is in turn upper-bounded by $\exp -\epsilon' C\log(N)$ for some suitable $\epsilon'>0$. The third term in the right-hand side of (6.8) is dealt with in a similar manner. Eventually, we obtain that for any $K>0$, and for a suitable choice of C, for all u and all $i=1,\ldots,D$,

$$\mathbf{P}(\bar{\mathcal{E}}_i|\mathcal{E}_2(u),\ldots,\mathcal{E}_{i-1}(u))\leq N^{-K}.$$

The result of Lemma 6.2.1 readily follows.

The proof of Theorem 6.2.1 is now concluded as follows. Given any two nodes u,v, we have:

$$\mathbf{P}(d_{G'}(u,v)>2D+2C\log(N))\leq\mathbf{P}\left(\overline{\cap_{i=2}^{D}\mathcal{E}_i(u)}\right)+\mathbf{P}\left(\overline{\cap_{i=2}^{D}\mathcal{E}_i(v)}\right)+\pi,$$

where π is the probability that two sets of sizes $C\log(N)kp(1-\epsilon)^D$, picked uniformly at random from a set of N nodes, have an empty intersection. Note that

$$C\log(N)kp(1-\epsilon)^D\geq r = C\log(N)\sqrt{N},$$

by our choice of D. In view of Lemma 5.4.1, we thus have that

$$\pi\leq(1+o(1))\exp\left(-C^2\log(N)^2(1+o(1))\right).$$

Thus certainly, $\mathbf{P}(d_{G'}(u,v)>2D+2C\log(N))=o(N^{-2})$. Summing this evaluation over all pairs (u,v) ensures that with high probability, the diameter of G' is $O(\log(N))$. □

6.3 Small world according to Kleinberg

In a recent article [7], Kleinberg revisited the small world phenomenon. His key observation is that while models such as that of Strogatz and Watts can explain the presence of short paths between nodes, they do not explain how individuals managed to efficiently determine such short paths in Milgram's experiment.

He proposes a model related to that of Strogatz and Watts, but where the distribution of the destination of shortcuts is not necessarily uniform. A critical parameter, α, characterizes this distribution of shortcuts. Kleinberg showed that, for a critical value α^* of α, individuals can determine paths of length at most a power of the logarithm of the number of nodes. He also established that, for $\alpha < \alpha^*$, short paths exist, but individuals cannot determine them in a decentralised manner. When $\alpha > \alpha^*$, short paths no longer exist. As a result, unless $\alpha = \alpha^*$, in this model it takes a number of steps that is of the order of a power of the number of nodes, instead of its logarithm, for an individual to determine another one.

6.3.1 The model

Nodes are again identified with the points of the grid $\{1, \ldots, m\} \times \{1, \ldots, m\}$. Nodes u and v are grid neighbours if their L^1 distance $|u - v| = |u_1 - v_1| + |u_2 - v_2|$ equals 1. An additional parameter q determines the number of shortcuts generated by each individual. Finally, a node u chooses another node v as the destination of a shortcut with probability $|u-v|^{-\alpha} / \sum_{w \neq u} |u-w|^{-\alpha}$, for some parameter $\alpha \geq 0$. Here, $|u-v|$ refers to the minimal number of hops from u to v using grid edges only, or equivalently to L^1 distance.

Note that when $\alpha = 0$, shortcuts are again selected uniformly at random. Thus, by the analysis of the model of Strogatz and Watts, we know that, in this case, the diameter is logarithmic in the number of nodes $N = m^2$.

6.3.2 Efficient routing for critical α

We now show the following

Theorem 6.3.1. *Assume $\alpha = 2$. Consider the following greedy routing scheme. A node u, trying to reach a node v, forwards the request to the node w from its grid and shortcut neighbours which is closest (according to L^1 distance) to target v. Then, for all nodes u, v, the number of steps $T_{greedy}(u, v)$ used by this scheme to reach v from u verifies*

$$\mathbf{E}(T_{greedy}(u, v)) \leq A \log(N)^2 \tag{6.9}$$

for some constant A.

Proof. Let the source and destination nodes u, v be fixed. Let $u(t)$ be the node at which the greedy algorithm is after t steps. We shall say that the algorithm is in phase j at time t if

$$2^j < |u(t) - v| \leq 2^{j+1}.$$

At each step spent in phase j, at least one new shortcut is discovered (recall that there $q \geq 1$ shortcuts created from each node; also, the presence of grid edges guarantees that the greedy algorithm moves closer to destination at each step). Given that $u(t)$ belongs to phase j, the probability that a shortcut towards a node w leads to a phase $k < j$ admits the lower bound

$$\min_{u(t):2^j<|u(t)-v|\leq 2^{j+1}} \frac{\sum_{w:|v-w|\leq 2^j} |u(t)-w|^{-2}}{\sum_{u'\neq u(t)} |u(t)-u'|^{-2}}.$$

The upper term is in turn lower-bounded by

$$(2^{j+1}+2^j)^{-2} \sum_{i=1}^{j-1} \geq 1/36,$$

while the lower summation is upper-bounded by

$$\sum_{i=1}^{2m} (4i) i^{-2} \leq 4\left(1+\int_1^{2m} \frac{1}{x} dx\right) \leq 4(1+\log(2m)).$$

Thus the probability of moving to a better phase is at least

$$\frac{1}{144\,(1+\log(2m))}.$$

This entails that the number of steps spent in a given phase j is dominated by a geometric random variable with parameter $1/[144(1+\log(2m))]$. Since there are at most $\log_2(2m)$ phases, the average number of steps to reach the destination verifies

$$\mathbf{E}(T(u,v)] \leq 144\,(1+\log(2m))\,\frac{\log(2m)}{\log(2)} = O((\log(N))^2),$$

as announced. □

6.3.3 Impossibility of efficient routing, $\alpha < 2$

In the present context, we say that a routing algorithm is decentralised if the routing decision made at step t depends only on knowledge of the nodes $u(0), \ldots, u(t)$ visited so far, and of the knowledge of the coordinates of shortcuts generated at these nodes. Clearly, the greedy algorithm used in the previous theorem is decentralised in this sense.

We now show that for $\alpha < 2$, no decentralised algorithm can perform efficiently:

Theorem 6.3.2. *Assume $\alpha \in [0,2)$. Then for any decentralised algorithm alg, the average number of steps $E[T_{alg}(u,v)]$ it takes to reach a destination v from a source u verifies, for "most" pairs (u,v),*

$$\mathbf{E}[T_{alg}(u,v)] = \Omega\left(m^{\frac{2-\alpha}{3}}\right). \tag{6.10}$$

Proof. Consider the neighborhood $\mathcal{V} = \{w : |v-w| \leq C\}$, for some C to be specified, and let $t = \epsilon C$, for some fixed $\epsilon \in (0,1)$. Assume that $|u-v| > C$. Then, if the algorithm is to reach v from u in t steps, necessarily the last shortcut used by the algorithm must end in neighborhood $\mathcal{V}$. Indeed, otherwise after the last shortcut is taken, the current node is outside $\mathcal{V}$; however there remain at most $t < C$ steps to take, and using only grid edges this will not reach v from outside $\mathcal{V}$, by definition of $\mathcal{V}$.

Thus the algorithm will fail to route to destination in t steps if it does not discover a shortcut leading into $\mathcal{V}$ in the first t visited locations. From any node w, the probability that a shortcut generated at that node reaches $\mathcal{V}$ reads

$$\frac{\sum_{v':|v'-v|\leq C} |v'-w|^{-\alpha}}{\sum_{v' \neq w} |v'-w|^{-\alpha}}.$$

The upper term is bounded from above, uniformly in w, by $|\mathcal{V}|$, which is no larger than $1+4C(C+1)/2$ itself not larger than $3C^2$ for $C \geq 1$. The denominator is lower-bounded by:

$$\sum_{i=1}^{m/2} i \times i^{-\alpha} \geq \int_1^{m/2} x^{1-\alpha} dx = \frac{m^{2-\alpha}-1}{2-\alpha} \geq \frac{1}{2} m^{2-\alpha}$$

for m large enough. Thus the probability of a shortcut reaching $\mathcal{V}$ is at most $6C^2 m^{\alpha-2}$. Finally, since at each step at most q new shortcuts are discovered, by the above argument the probability of failing to route to destination in t steps is at least:

$$1 - qt \sup_w \mathbf{P}(\text{shortcut generated from } w \text{ fails to reach } \mathcal{V}) \geq 1 - q\epsilon(6C^3 m^{\alpha-2}).$$

Setting $C = m^{\frac{\alpha-2}{3}}$, and $\epsilon = 1/(12q)$, the right-hand side simplifies to $1/2$. Finally, we obtain that, for u, v such that $|u-v| > C = m^{\frac{2-\alpha}{3}}$,

$$\mathbf{E}[T_{alg}(u,v)] \geq \frac{1}{2} \epsilon m^{\frac{2-\alpha}{3}} = \frac{1}{24q} m^{\frac{2-\alpha}{3}}.$$

Since the fraction of pairs of nodes (u,v) such that $|u-v| > m^{\frac{2-\alpha}{3}}$ approaches 1 as $m \to \infty$, the claim follows. □

6.3.4 Impossibility of efficient routing, $\alpha > 2$

Theorem 6.3.3. *Assume that $\alpha > 2$. There is an increasing function $f : \mathbb{R}^+ \to \mathbb{R}^+$ such that, for any distributed algorithm alg, and any two source and destination nodes u, v, the corresponding expected number of steps verifies*

$$\mathbf{E}[T_{alg}(u,v)] \geq f\left(|u-v|/m\right) m^{\gamma}, \tag{6.11}$$

where $\gamma := (\alpha-2)/(\alpha-1)$. Thus, for more than one half (say) of pairs of nodes u, v, decentralised routing takes of the order of m^{γ} steps on average.

Proof. For some d to be specified, note that the probability that a shortcut generated at a node w reaches a target w' such $|w - w'|$ exceeds d is at most

$$\frac{\sum_{i=d+1}^{\infty} 4i \times i^{-\alpha}}{\sum_{w' \neq w} |w - w'|^{-\alpha}} \leq 4 \int_d^{\infty} x^{1-\alpha} dx = \frac{4}{2-\alpha} d^{\alpha-2}.$$

Given a target number of steps t, and two source nodes u, v, such that $td < |u - v|$, routing will fail to reach destination v in t steps if all shortcuts found in these t steps all have length not larger than d. This entails that:

$$\mathbf{E}[T_{alg}(u, v)] \geq t \left[1 - qt \frac{4}{\alpha - 2} d^{\alpha-2}\right]. \tag{6.12}$$

Now choose t, d to ensure:

$$td = \frac{|u - v|}{2},$$

and

$$qt \frac{4}{\alpha - 2} d^{2-\alpha} = 1/2,$$

the latter ensuring that the right-hand side in (6.12) equals 1/2. These two equations solve to give

$$d = |u - v|^{1/(\alpha-1)} \left[\frac{4q}{\alpha - 2}\right]^{1/(\alpha-1)}, \; t = |u - v|^{(\alpha-2)/(\alpha-1)} \frac{1}{2} \left[\frac{4q}{\alpha - 2}\right]^{-1/(\alpha-1)}.$$

The evaluation (6.11) now follows from (6.12) by setting

$$f(x) = x^{\gamma} \frac{1}{4} \left[\frac{4q}{\alpha - 2}\right]^{-1/(\alpha-1)}.$$

□

Chapter 7

Power laws via preferential attachment

So far we have considered E-R random graphs, de Bruijn graphs, and small world graphs à la Srogatz and Watts or à la Kleinberg. In all these examples, the degree distribution of nodes is sharply concentrated around its mean. For E-R graphs, using the Chernoff bound 3.2.2, denoting by δ the average degree $\delta = (N-1)p$, it holds that:

$$\mathbf{P}(\exists i \in \{1, \ldots, N\} : |d_i - \delta| \geq \epsilon\delta) \leq N \left(\exp(-\delta h(\epsilon)) + \exp(-\delta h(-\epsilon))\right).$$

Thus, assuming $\delta >> \log(N)$, if we take $\epsilon = 2\sqrt{\log(N)/\delta}$, since $h(x) = x^2/2(1 + o(x))$, we find that the exponents in the right-hand side above are equivalent to $2\log(N)$, hence the right-hand side is of order $1/N$: with high probability, no node degree deviates from its mean δ by a factor larger than $2\sqrt{\log(N)/\delta} << 1$.

The degrees in all the other graphs we have studied so far are, with high probability, bounded by a constant multiple of the logarithm $\log(N)$ of the number of nodes N. In contrast, many examples of graphs display very different degree distributions. In particular, it is common to have graphs where the number of nodes of degree larger than i is roughly proportional to $i^{-\beta}$, for some exponent β, and over a significant range of values i. We do not try to provide a rigorous definition, and to make precise what is meant by "roughly proportional", and "significant range of values". Examples where this behaviour appears are: the graph of the Internet topology, when viewed at the router level and also at the autonomous system level; the web graph, in which nodes are web pages and links are hyperlinks; the Hollywood graph where nodes are actors and links indicate that two actors played in the same movie; etc... For a survey with many more examples, see Newman [14].

This feature of graphs has implications on the behaviour of processes such as epidemics that may evolve on them. This chapter is concerned with generative models that produce graphs with such properties. One mechanism that has been proposed for explaining the presence of power laws is that of preferential attachment. We shall first illustrate this on a simple graph formation process, which has been popularised by Barabási and Albert [3]. We will then present a precursor of this model, namely the model of Yule [16]

for explaining power law distributions for the number of species within genera of plants (among other broad families of living things).

7.1 Barabási-Albert random graphs

The model we consider is as follows. A graph is grown over time, starting with an initial graph $G(0)$, and adding one vertex $u(t)$ at each time step $t = 1, 2, \ldots$. The resulting graph by the end of step t is denoted $G(t)$, has

$$N(t) = N(0) + t$$

vertices. The new vertex $u(t)$ is attached to the previous graph $G(t-1)$ by a single edge. Thus the total number of edges in $G(t)$, that we denote $E(t)$, verifies

$$E(t) = E(0) + t.$$

The node to which $u(t+1)$ attaches in $G(t)$ is chosen as follows. For some parameter $\alpha \in (0,1)$, the anchor node is chosen uniformly at random from the $N(t)$ nodes of $G(t)$. With probability $1-\alpha$, a node v is selected with probability $d_t(v)/2E(t)$, where $d_t(v)$ denotes the degree of node u in graph $G(t)$. Note that this defines a proper probability distribution since the sum of node degrees in $G(t)$ is indeed twice the number of edges $E(t)$.

A descriptor of the graph $G(t)$ suitable for the identification of power laws is the following. Denote by $X_i(t)$ the number of nodes with degree i in $G(t)$. Denote by $\mathcal{F}_t$ the sigma-field containing all the information about the graphs $G(0), \ldots, G(t)$. The above preferential attachment rule entails the following properties for the vector $X(t) = \{X_i(t)\}_{i \geq 1}$:

$$\begin{aligned} &\mathbf{P}(X_1(t+1) = X_1(t) | \mathcal{F}_t) = \alpha \tfrac{X_1(t)}{N(t)} + (1-\alpha)\tfrac{X_1(t)}{2E(t)}, \\ &\mathbf{P}(X_1(t+1) = X_1(t) + 1 | \mathcal{F}_t) = 1 - \alpha \tfrac{X_1(t)}{N(t)} - (1-\alpha)\tfrac{X_1(t)}{2E(t)}. \end{aligned} \tag{7.1}$$

Similarly, for all $i > 1$, one has:

$$\begin{aligned} &\mathbf{P}(X_i(t+1) = X_i(t) + 1 | \mathcal{F}_t) = \alpha \tfrac{X_{i-1}(t)}{N(t)} + (1-\alpha)\tfrac{(i-1)X_{i-1}(t)}{2E(t)}, \\ &\mathbf{P}(X_i(t+1) = X_i(t) - 1 | \mathcal{F}_t) = \alpha \tfrac{X_i(t)}{N(t)} + (1-\alpha)\tfrac{(i\,X_i(t)}{2E(t)}, \\ &\mathbf{P}(X_i(t+1) = X_i(t) | \mathcal{F}_t) = 1 - \alpha \tfrac{X_i(t)+X_{i-1}(t)}{N(t)} - (1-\alpha)\tfrac{i\,X_i(t)+(i-1)X_{i-1}(t))}{2E(t)}. \end{aligned} \tag{7.2}$$

The main result of this section is the following

Theorem 7.1.1. *Let*

$$c_1 = \frac{2}{3+\alpha}, \quad \frac{c_i}{c_{i-1}} = \frac{\alpha + \frac{1-\alpha}{2}(i-1)}{1+\alpha+\frac{1-\alpha}{2}i}, \; i > 1. \tag{7.3}$$

Then for all $i \geq 1$, it holds that

$$\frac{X_i(t)}{t} \to c_i \text{ almost surely as } t \to \infty. \tag{7.4}$$

A consequence of the theorem is that the graph G_t is approximately a power-law random graph. Indeed, for large $i > 1$, one has:

$$\frac{c_i}{c_{i-1}} = 1 - \frac{3-\alpha}{2+2\alpha+(1-\alpha)i} = 1 - \frac{1}{i}\frac{3-\alpha}{1-\alpha} + O(i^{-2}).$$

Consequently,

$$c_i = c_1 \prod_{j=2}^{i} \left(1 - \frac{1}{i}\frac{3-\alpha}{1-\alpha} + O(i^{-2})\right) \sim A i^{-\frac{3-\alpha}{1-\alpha}} \text{ as } i \to \infty,$$

for some constant $A > 0$.

[Here give more details about graph being power-law]

The proof proceeds by first controlling the average values $\bar{X}_i(t) := \mathbf{E}(X_i(t))$, which is done in the following:

Theorem 7.1.2. *For all $\epsilon > 0$, and all $i \geq 1$, it holds that*

$$\bar{X}_i(t) = c_i t + o(t^\epsilon). \tag{7.5}$$

Proof. Let $\epsilon > 0$ be fixed. Introduce the notation

$$\Delta_i(t) = \bar{X}_i(t) - c_i(t), \quad i \geq 1,\ t \geq 1.$$

By Equation (7.1), it holds that:

$$\begin{aligned}\Delta_1(t+1) &= \Delta_1(t) - c_1 + 1 - \alpha\tfrac{\bar{X}_1}{N(t)} - (1-\alpha)\tfrac{\bar{X}_1(t)}{2E(t)} \\ &= \Delta_1(t)\left[1 - \tfrac{\alpha}{N(t)} - \tfrac{1-\alpha}{2E(t)}\right] - c_1 + 1 - c_1 t\left(\tfrac{\alpha}{N(t)} + \tfrac{1-\alpha}{2E(t)}\right).\end{aligned}$$

Note that the term $[\alpha/N(t) + (1-\alpha)/2E(t)]$ is in the interval $[0,1]$, and furthermore it equals $t^{-1}[\alpha + (1-\alpha)/2] + O(t^{-2})$. This yields

$$\begin{aligned}\Delta_1(t+1) &= \Delta_1(t)\left[1 - \tfrac{\alpha}{N(t)} - \tfrac{1-\alpha}{2E(t)}\right] - c_1 + 1 - c_1(\alpha + (1-\alpha)/2) + O(t^{-1}) \\ &= \Delta_1(t)\left[1 - \tfrac{\alpha}{N(t)} - \tfrac{1-\alpha}{2E(t)}\right] + O(t^{-1}),\end{aligned}$$

by our choice of c_1. It thus follows that

$$|\Delta_1(t+1))| \leq |\Delta_1(t)| + O(t^{-1}) \leq O(\log(t)),$$

the latter evaluation being obtained by induction on t. Thus certainly, $\Delta_1(t) = \bar{X}_1(t) - c_1 t = o(t^\epsilon)$.

Let us now consider $i > 1$, and assume that a similar condition (7.5) holds for all $j < i$. Using (7.2), write

$$\begin{aligned}\Delta_i(t+1) &= \Delta_i(t) - c_i + \bar{X}_{i-1}(t)\left[\tfrac{\alpha}{N(t)} + \tfrac{(1-\alpha)(i-1)}{2E(t)}\right] - \bar{X}_i(t)\left[\tfrac{\alpha}{N(t)} + \tfrac{(1-\alpha)i}{2E(t)}\right] \\ &= -c_i + \Delta_i(t)\left[1 - \tfrac{\alpha}{N(t)} + \tfrac{(1-\alpha)i}{2E(t)}\right] + \Delta_{i-1}(t)\left[\tfrac{\alpha}{N(t)} + \tfrac{(1-\alpha)(i-1)}{2E(t)}\right] \\ &\quad + c_{i-1} t\left[1 - \tfrac{\alpha}{N(t)} + \tfrac{(1-\alpha)i}{2E(t)}\right] - c_i t\left[\tfrac{\alpha}{N(t)} + \tfrac{(1-\alpha)i}{2E(t)}\right] \\ &= \Delta_i(t)\left[1 - \tfrac{\alpha}{N(t)} + \tfrac{(1-\alpha)i}{2E(t)}\right] + O(\Delta_{i-1}(t)/t) + O(t^{-1}).\end{aligned}$$

By the induction hypothesis, $\Delta_{i-1}(t)/t = o(t^{\epsilon-1})$. We thus arrive at

$$|\Delta_i(t)| = o\left(\sum_{s=1}^{t} s^{\epsilon-1}\right) = o(t^{\epsilon}),$$

which is the claimed property. □

We next need the following lemma:

LEMMA 7.1.1: For all $i, t \geq 1$, and all $M > 0$, it holds that

$$\mathbf{P}\left(|X_i(t) - \bar{X}_i(t)| \geq M\right) \leq 2\exp\left(-\frac{M^2}{8t}\right). \tag{7.6}$$

Before giving the proof of this Lemma, we show how, together with Theorem 7.1.2, it implies the result of Theorem 7.1.1. Take $M = 4\sqrt{t\log(t)}$ in (7.6). This yields

$$\mathbf{P}\left(|X_i(t) - \bar{X}_i(t)| \geq 4\sqrt{t\log(t)}\right) \leq 2t^{-2}.$$

The sum over $t \geq 0$ of the right-hand side $2t^{-2}$ is finite. Thus, by Borel-Cantelli's lemma, the event

$$|X_i(t) - \bar{X}_i(t)| \geq 4\sqrt{t\log(t)}$$

occurs for only finitely many t's. Thus combined with Theorem 7.1.2, this entails that for all $\epsilon > 0$, and all large enough t, it holds that

$$|X_i(t) - c_i t| \leq t^{\epsilon} + 4\sqrt{t\log(t)},$$

and the result of Theorem 7.1.1 follows. Let us now give the proof of Lemma 7.1.1.

Proof. Let $t, i \geq 1$ be fixed. Denote by $v(s)$ the node in $G(s-1)$ to which the node $u(s)$ attaches to. We make the dependency of $X_i(t)$ on the consecutive choices $v(1), \ldots, v(t)$ explicit by writing

$$X_i(t) = f(v(1), \ldots, v(t)).$$

We further define the martingale $\{M(s)\}_{0 \leq s \leq t}$ by letting

$$M(s) = \mathbf{E}[X_i(t)|v(1), \ldots, v(s)], \quad s = 0, \ldots, t.$$

Let us show that this martingale satisfies the following property:

$$|M(s) - M(s-1)| \leq 2 \text{ almost surely}, \quad s = 1, \ldots, t. \tag{7.7}$$

To this end, let $s \in \{1, \ldots, t\}$ be fixed. Let the sequence $v(1), \ldots, v(s)$ be given, and let another random node $V'(s)$ of $G(s-1)$ be given, that is distributed as the anchor node in $G(s)$, given that the previous anchor nodes are $v(1), \ldots, v(s-1)$. We now construct jointly random node sequences $V(s+1), \ldots, V(t)$, and $V'(s+1), \ldots, V'(t)$ with the following two properties:

- The distribution of $V(s+1),\dots,V(t)$ (respectively, $V'(s+1),\dots,V'(t)$) is that of the $s+1$-th to t-th anchor nodes in the graph growth model under consideration, conditionally on the first s anchor nodes being $v(1),\dots,v(s)$ (respectively, $v(1),\dots,v(s-1),V'(s)$);
- Denoting by $G(s),\dots G(t)$ (respectively, $G'(s),\dots,G'(t)$) the corresponding sequence of growing graphs, for all $u=s,\dots,t$, and any node w in the node set of $G(u)$ and $G'(u)$, then the degree $d_u(w)$ of w within graph $G(u)$ coincides with the degree $d'_u(w)$ within graph $G'(u)$, unless w equals either $v(s)$ or $V'(s)$.

Let us prove by induction on $u=s,\dots,t$ that this can indeed be achieved. Note first that the graphs $G(s)$ and $G'(s)$ only differ by the attaching point $v(s)$ or $v'(s)$ of newly added node $u(s)$. Thus the degrees of all other nodes agree. Assuming that this property holds for $G(u)$ and $G'(u)$, sample the new anchor nodes $V(u+1)$ and $V'(u+1)$ as follows.

With probability α, pick the same anchor node $V(u+1)=V'(u+1)$, uniformly at random from the node set of $G(u)$ and $G'(u)$.

With probability $1-\alpha$, sample the anchor nodes $V(u+1)$, $V'(u+1)$ as follows. Let $V(u+1)=V'(u+1)=w$ with probability $d_u(w)/2E(u)$ if $w\neq v(s),V'(s)$. With probability $(d_u(v(s))+d_u(V'(s))/2E(u) = (d'_u(v(s))+d'_u(V'(s)))/2E(s)$, take $V(s+1),V'(s+1)\in\{v(s),v'(s)\}$, ensuring that they take the proper values $v(s)$ or $V'(s)$ with the relative probabilities consistent with the preferential attachment model.

Note that this joint, or coupled construction, is feasible since at each step both graphs $G(u)$ and $G'(u)$ have the same number of edges, given that the degrees of all nodes but $v(s)$, $v'(s)$ agree in both graphs, then the sums of degrees $d_u(v(s))+d_u(V'(s))$ and $d'_u(v(s))+d'_u(V'(s))$ coincide. Now, equation (7.7) follows by writing

$$\begin{aligned}|M(s)-M(s-1)| &\leq \left|\sum_{v_{s+1}^t,v_s'^t}\mathbf{P}(V_{s+1}^t=v_{s+1}^t,V_s'^t=v_s'^t)\left[f(v_1^t)-f(v_1^{s-1}v_s'^t)\right]\right| \\ &\leq \sum_{v_{s+1}^t,v_s'^t}\mathbf{P}(V_{s+1}^t=v_{s+1}^t,V_s'^t=v_s'^t)\left|f(v_1^t)-f(v_1^{s-1}v_s'^t)\right|.\end{aligned}$$

However, the coupling construction ensures that in the graphs $G(t)$ and $G'(t)$, the degrees of all but at most two nodes disagree. Recall that f counts the number of degree i nodes. Then clearly, the absolute value in the right-hand side of the above does not exceed 2. Since the sum of probabilities equals 1, Equation (7.7) follows. □

7.2 Yule process

A precursor of the previous model is the so-called Yule process, introduced in [16] as a plausible model of the evolution of the number of species within *genera*, that are families of species. The model is as follows. Species are organised in families. Each species gives birth, at some fixed rate, to a new species, thanks to mutations. Species don't go extinct in the basic version of the model. The new species is, with probability α, so different from any other species that it creates a new family of species of its own. With the complementary probability $(1-\alpha)$, it is a member of the same family as the species from which it originates.

Let us denote by $X_i(t)$ the number of families comprising exactly i species, after the t-th new species has appeared. The initial condition is specified by the vector $\{X_i(0)\}_{i\geq 1}$. Denoting by $\mathcal{F}_t$ the sigma-field $\sigma(X(0),\ldots,X(t)$, the dynamics are then such that:

$$\begin{array}{lcl} \mathbf{P}(X_1(t+1)=X_1(t)+1|\mathcal{F}_t) & = & \alpha, \\ \mathbf{P}(X_1(t+1)=X_1(t)-1|\mathcal{F}_t) & = & (1-\alpha)\frac{X_1(t)}{N(t)}, \\ \mathbf{P}(X_1(t+1)=X_1(t)|\mathcal{F}_t) & = & 1-\alpha-(1-\alpha)\frac{X_1(t)}{N(t)}, \end{array} \tag{7.8}$$

where $N(t)$ denote the total number of species. Note that

$$N(t) = N(0) + t.$$

Similarly, we have for all $i > 1$ and $t \geq 0$:

$$\begin{array}{lcl} \mathbf{P}(X_i(t+1)=X_i(t)+1|\mathcal{F}_t) & = & (1-\alpha)\frac{(i-1)X_{i-1}(t)}{N(t)}, \\ \mathbf{P}(X_i(t+1)=X_i(t)-1|\mathcal{F}_t) & = & (1-\alpha)\frac{iX_i(t)}{N(t)}, \\ \mathbf{P}(X_i(t+1)=X_i(t)|\mathcal{F}_t) & = & \alpha+(1-\alpha)\left[1-\frac{iX_i(t)+(i-1)X_{i-1}(t)}{N(t)}\right]. \end{array} \tag{7.9}$$

We now establish the following

Theorem 7.2.1. *Let*

$$c_1 = \frac{\alpha}{2-\alpha}, \quad \frac{c_i}{c_{i-1}} = 1 - \frac{2-\alpha}{1+i(1-\alpha)},\ i > 1. \tag{7.10}$$

Then for all $i \geq 1$, it holds that

$$\frac{X_i(t)}{t} \to c_i \text{ almost surely as } t \to \infty. \tag{7.11}$$

The proof proceeds along the same lines as that of Theorem 7.1.1. We first analyse the asymptotic behaviour of the expected values $\bar{X}_i(t) := \mathbf{E}(X_i(t))$:

Theorem 7.2.2. *For all $\epsilon > 0$, and all $i \geq 1$, it holds that*

$$\bar{X}_i(t) = c_i t + o(t^\epsilon). \tag{7.12}$$

Proof. Define as previously $\Delta_i(t) = \bar{X}_i(t) - c_i t$, for all $i, t \geq 1$. For $i = 1$, we have that

$$\begin{array}{lcl} \Delta_1(t+1) & = & \Delta_1(t) - c_1 + \alpha - (1-\alpha)\frac{\bar{X}_1(t)}{N(t)} \\ & = & \Delta_1(t)\left(1-\frac{1-\alpha}{N(t)}\right) - c_1 + \alpha - (1-\alpha)\frac{c_1 t}{N(t)} \\ & = & \Delta_1(t)\left(1-\frac{1-\alpha}{N(t)}\right) + O(t^{-1}), \end{array}$$

which again ensures that $\Delta_1(t) = O(\log(t))$.

Let us assume that for all $j < i$, and all $\epsilon > 0$, $\Delta_j(t) = o(t^\epsilon)$. Write then

$$\begin{aligned}\Delta_i(t+1) &= \Delta_i(t) - c_i + \tfrac{1-\alpha}{N(t)}\left[(i-1)\bar{X}_{i-1}(t) - i\bar{X}_i(t)\right] \\ &= \Delta_i(t)\left[1 - \tfrac{(1-\alpha)i}{N(t)}\right] + \tfrac{(1-\alpha)(i-1)}{N(t)}\Delta_{i-1}(t) \\ &\quad -c_i\left(1 + \tfrac{(1-\alpha)it}{N(t)}\right) + c_{i-1}\tfrac{(1-\alpha)(i-1)t}{N(t)} \\ &= \Delta_i(t)\left[1 - \tfrac{(1-\alpha)i}{N(t)}\right] + O(t^{\epsilon-1}),\end{aligned}$$

where we have used the induction assumption $\Delta_{i-1}(t) = o(t^\epsilon)$, and cancellation of terms based on the expression of c_i/c_{i-1} in (7.10). The result (7.12) easily follows. □

In order to conclude the proof of Theorem 7.2.1, let us establish that the statement of Lemma 7.1.1 is true with the state variables $X_i(t)$ for the Yule process rather than for the preferential attachment graph. Let then i, t be fixed. Denote by $v(s)$ the label of the species that gave birth to a new species at the s-th such birth, and let $\xi_s \in \{0, 1\}$ equal 1 if this new species starts a new family, and 0 otherwise. Denote by $x(s)$ the pair $(v(s), \xi_s)$. Note that the variables $\{x(s)\}_{s=1,\ldots,t}$ are independent, $v(s)$ being uniform on the label set of species at time $s-1$ (for definiteness, take it to be $\{1, \ldots, N(s-1)\}$), and independent of ξ_s, which is a Bernoulli random variable, equal to 1 with probability α. Writing

$$X_i(t) = f(x(1), \ldots, x(t))$$

we now show that we can use the second form of Azuma-Hoeffding's inequality, Corollary 6.2.1, to control the deviations of this random variable from its mean. Indeed, consider a sequence $x_1^t = (x_1, \ldots, x_t)$, and another sequence $x_1^{s-1} y_s x_{s+1}^t$, differing from the first only in its s-th coordinate. Taking the graphical representation, in which a new species is identified to a graph node, and is connected to its generating species by an edge if it belongs to the same family, and to no other family otherwise (as in Figure 7.1), by changing the coordinate x_s to y_s in the sequence x_1^t we simply modify the end point of the edge connecting the s-th node (we may either remove this edge, or create it if it was absent). Now, the number of families consisting of i species is exactly the set of connected components with i nodes inthe graphical representation.

By removing one edge of a graph, we may split a connected component into two. This can remove at most one connected component of size i, and at most create two connected components of size i. Thus addition/removal of one edge modifies the number of size i components by at most 2.

Similarly, by changing the end-point of one edge, we can remove at most two size i components. By symmetry, we can add at most two such components. This shows that, for the function f above, it holds that:

$$|f(x_1^t) - f(x_1^{s-1} y_s x_{s+1}^t)| \leq 2,$$

for all possible choices of x_1^t, $s = 1, \ldots, t$, and y_s. Thus Corollary 6.2.1 applies, and the proof of Theorem 7.2.1 then parallels that of Theorem 7.1.1.

Figure 7.1: Graphical representation of Yule process. Squares are species starting new family; circles are species connected to its originating species, hence part of the same family.

7.3 Notes

For further reading: Mitzenmacher [13] discusses other models for generating power laws, as well as lognormal distributions. He also reviews a justification completely different from the preferential attachment model, due to Mandelbrot [12] for the fact that the distribution of word occurences in a text follows a power law. According to Mandelbrot, such power laws yield the maximal information rate per length of text, counted in letters rather than words.

A richer model than that of Barabási-Albert, featuring multiple edge additions per node arrival, and oriented edges, is presented in [5], together with an analysis of resulting power laws for both in-degrees and out-degrees.

Chapter 8

Epidemics on general graphs

8.1 Model and motivation

In the present chapter we investigate the behaviour of Susceptible-Infective-Susceptible epidemics on general finite graphs. The model we consider is also known as the contact process. It is described as follows. A graph G is given, with a finite node set $\{1, \ldots, N\}$. Variable $X_i(t)$ tracks the health status of node i: it is infected if $X_i(t) = 1$, and healthy if $X_i(t) = 0$. Infected nodes return to susceptible state at unit rate, while susceptible nodes become infected at a rate that is the product of the base infection rate, $\beta > 0$, and the number of graph neighbours that are infected.

Such dynamics could be plausible models of the following situations:

- Epidemics of mutating viruses, where a new mutant can re-infect an individual previously infected by another version of the virus (think of influenza);
- A crude information storage system. Here, nodes correspond to storage locations; they remove stored information at unit rate, while nodes holding some information replicate it at neighbour nodes at some rate β.

In these two scenarios, a quantity of interest is the time to recovery from the epidemics (in the second situation, this would correspond to information loss by the system). Note that, when the graph is finite, extinction is bound to happen. This is in contrast with the case of infinite graphs, where infection can survive forever. For a survey of results for the contact process on infinite graphs, such as regular grids and trees, see [8].

A more formal description of the contact process is as follows. It is a Markov jump process on $\{0,1\}^N$, with non-zero transition rates $q(x, y)$ between states $x, y \in \{0,1\}^N$ given by

$$\begin{array}{ll} q(x, x + e_i) & = \beta(1 - x_i) \sum_{j \sim i} x_j,\ x \in \{0,1\}^N, i \in \{1, \ldots, N\}, \\ q(x, x - e_i) & = x_i,\ x \in \{0,1\}^N, i \in \{1, \ldots, N\}. \end{array} \tag{8.1}$$

In the above, $i \sim j$ refers to i and j being graph neighbours, and e_i denotes the vector with its i-th coordinate equal to 1, and all other coordinates equal to 0.

The chapter is organised as follows. We first give a sufficient condition for fast extinction (absorption at 0) of the process. We next give a sufficient condition for long survival of the process. We then apply these two results to several graph models of interest.

8.2 Fast extinction and spectral radius for the SIS epidemics

Recall that the adjacency matrix A of a graph G is determined by $A_{ij} = 1$ if $i \sim j$, and 0 otherwise. Also recall that the spectral radius of a matrix is the maximum of the absolute value of its eigenvalues. We shall establish the following

Theorem 8.2.1. *Let A denote the adjacency matrix of graph G, and ρ denote the spectral of this matrix. Then for any initial condition $X(0) = \{X_i(0)\}_{1=1,\dots,N}$, and all $t \geq 0$, one has the following:*

$$\mathbf{P}(X(t) \neq 0) \leq \sqrt{N \sum_{i=1}^{N} X_i(0)} \exp\left((\beta\rho - 1)t\right), \tag{8.2}$$

where $X(t) := \{X_i(t)\}_{1=1,\dots,N}$ denotes the state of the contact process with parameter β, on graph G, at time t.

In order to establish this result, we shall rely on a general coupling technique, which allows to relate the trajectories of different Markov processes. This will be phrased in the context of skip-free Markov jump processes, which we now define.

Definition 8.2.1. *Let $K > 0$ be some fixed integer. A skip-free Markov jump process on the state $\mathbb{N}^K$, is by definition a Markov jump process on this state space, whose transition rates $q(x, y)$, for $x \neq y \in \mathbb{N}^K$, are all zero except when $y = x + e_i$ or $y = x - e_i$ for some $i \in \{1, \dots, K\}$. The transition rate $q(x, x + e_i)$ is also referred to as the birth rate at site i when in state x. Similarly, the transition rate $q(x, x - e_i)$ is the death rate at site i when in state x.*

The basic coupling result we shall use is the following

Theorem 8.2.2. *Consider two skip-free Markov jump processes X, X' defined on the state space $\mathbb{N}^K$, with respective birth rates $\beta_i(x)$, $\beta'_i(x)$, and death rates $\delta_i(x)$, $\delta'_i(x)$, for $x \in \mathbb{N}^K$, and $i \in \{1, \dots, K\}$.*
Assume that for all $x, y \in \mathbb{N}^K$ such that $x \leq y$ (i.e., $x_i \leq y_i$ for all $i = \{1, \dots, K\}$), the following holds:

$$x_i = y_i \Rightarrow \beta_i(x) \leq \beta'_i(y) \text{ and } \delta_i(x) \geq \delta'_i(y). \tag{8.3}$$

Then, for initial conditions $X(0)$ and $X'(0)$ verifying $X(0) \leq X'(0)$, one can construct the two processes X, X' jointly so that for all $t \geq 0$, the ordering is preserved, that is $X(t) \leq X'(t)$.

Proof. Consider the Markov process on the state space $\{(x, x') \in \mathbb{N}^K \times \mathbb{N}^K : x \leq x'\}$, with only non-zero transition rates given as follows. For any state (x, x'), and $i \in \{1, \ldots, K\}$, if $x_i < x'_i$ the non-zero transition rates are

$$\begin{array}{ll} q((x,x'),(x+e_i,x')) & = \beta_i(x), \\ q((x,x'),(x,x'+e_i)) & = \beta'_i(x'), \\ q((x,x'),(x-e_i,x')) & = \delta_i(x), \\ q((x,x'),(x,x'-e_i)) & = \delta'_i(x'). \end{array} \tag{8.4}$$

When $x_i = x'_i$, the non-zero transition rates are given by

$$\begin{array}{ll} q((x,x'),(x+e_i,x'+e_i)) & = \beta_i(x), \\ q((x,x'),(x,x'+e_i)) & = \beta'_i(x') - \beta_i(x), \\ q((x,x'),(x-e_i,x'-e_i)) & = \delta'_i(x'), \\ q((x,x'),(x-e_i,x')) & = \delta_i(x) - \delta'_i(x'). \end{array} \tag{8.5}$$

Note that these terms are non-negative when Condition (8.3) holds. The proof of Theorem 8.2.2 will be concluded by establishing that the Markov process $(X(t), X'(t))_{t>0}$ started from initial condition $(X(0), X'(0))$ and whose dynamics are specified by these transition rates is such that the component processes $(X(t))_{t>0}$ and $(X'(t))_{t>0}$ are skip-free Markov jump processes with the desired birth an death rates given by the functions (β, δ) and (β', δ') respectively. Since by construction, $X(t) \leq X'(t)$ for all $t > 0$, the result will follow.

To establish that the component processes indeed have the desired dynamics, we use the following result.

LEMMA 8.2.1: Let a Markov jump process $\{Y(t)\}_{t\geq 0}$ on a countable state space E, with transition rates $q(x, y)$, $x, y \in E$, and a function $f : E \to F$ be given. Assume that there exists a function $\tilde{q}(u, v)$ defined on $F \times F$ such that, for all $i \in E$ and all $v \in F$, one has

$$\sum_{j:f(j)=v} q(i, j) = \tilde{q}(f(i), v). \tag{8.6}$$

Then the image process $Z(t) := f(Y(t))$ is a Markov jump process on F, with transition rates $\tilde{q}(u, v)$.

Proof will be included in the appendix. Let us now use this result to show that the componeThnt processes of the above-defined coupled process have the desired dynamics. Let $f : \mathbb{N}^K \times \mathbb{N}^K \to \mathbb{N}^K$ be defined by $f(x, x') = x$.

For any x, and any $x' \geq x$, and any $y \in \mathbb{N}^K$, we need to check the following identity:

$$\sum_{z\in\mathbb{N}^K} q((x,x'),(y,z)) = \begin{cases} \beta_i(x) & \text{if } y = x + e_i, \\ \delta_i(x) & \text{if } y = x - e_i, \\ 0 & \text{otherwise.} \end{cases}$$

This is easily verified from the rate specifications (8.4) and (8.5). The same needs to be done for the second component; again this is straightforward. □

We now return to the proof of Theorem 8.2.1. Define the so-called *branching random walk* process on $\mathbb{N}^N$ as the skip-free Markov jump process with birth and death rates:

$$\beta_i^{brw}(x) = \beta \sum_{j \sim i} x_j, \quad \delta_i^{brw}(x) = x_i, \ i \in \{1, \ldots, N\}.$$

Also, view the contact process as a skip-free Markov jump process on $\mathbb{N}^N$, by extending the definition of its birth and death rates β^c, δ^c to $\mathbb{N}^N$ as follows:

$$\beta_i^c(x) = \mathbf{1}_{x_i=0} \beta \sum_{j \sim i} x_j, \quad \delta_i^c(x) = x_i, \ i \in \{1, \ldots, N\}.$$

Next, we verify that the branching random walk and the contact process thus defined verify the assumptions of Theorem 8.2.2. To this end, let $x \leq x'$, and let i be such that $x_i = x'_i$. For all such parameter choices, we need to verify:

$$\delta_i^c(x) \geq \delta_i^{brw}(x').$$

This trivially holds, as the two terms equal x_i. It only remains to check that:

$$\beta_i^c(x) \leq \beta_i^{brw}(x').$$

The left-hand side is less than $\beta \sum_{j \sim i} x_j$, which is clearly less than $\beta \sum_{j \sim i} x'_j = \beta_i^{brw}(x')$ when $x \leq x'$. Theorem 8.2.2 thus applies.

Using the coupled construction of the two processes $(X^c(t), X^{brw}(t))_{t \geq 0}$, started from the same initial condition $X(0) \in \{0,1\}^N$, provided by this theorem, write:

$$\begin{aligned} \mathbf{P}(X^c(t) \neq 0) &\leq \mathbf{P}(X^{brw}(t) \neq 0) \\ &\leq \mathbf{E}(X^{brw}(t)). \end{aligned}$$

The linear structure of the transition rates of the branching random walk entail that:

$$\frac{d}{dt} \mathbf{E}(X^{brw}(t)) = \beta A \mathbf{E}(X^{brw}(t)) - \mathbf{E}(X^{brw}(t)),$$

which solves to give

$$\mathbf{E}(X^{brw}(t)) = \exp\left(t(\beta A - I)\right) X(0),$$

where I denotes the identity matrix, A is the adjacency matrix of G, and $\exp(t(\beta A - I))$ is the matrix exponential of the matrix $t(\beta A - I)$. We thus have:

$$\mathbf{P}(X^c(t) \neq 0) \leq e' \exp\left(t(\beta A - I)\right) X(0),$$

where $e = (1, \ldots, 1)^T$. By Cauchy-Schwarz inequality, the right-hand side of the above is not larger than

$$||e|| \times ||\exp\left(t(\beta A - I)\right) X(0)||.$$

However, since the matrix involved in the second term is symmetric, this term is not larger than $||X(0)||$ times the spectral radius of this matrix. The latter equals $\exp((\beta\rho - 1)t)$, which thus yields

$$\begin{aligned}\mathbf{P}(X^c(t) \neq 0) &\leq ||e|| \exp((\beta\rho - 1)t)||X(0)|| \\ &= \sqrt{N \sum_{i=1}^N X_i^2(0)} \exp((\beta\rho - 1)t) \\ &= \sqrt{N \sum_{i=1}^N X_i(0)} \exp((\beta\rho - 1)t),\end{aligned}$$

that is the announced result. The last equality holds because $X_i^2(0) = X_i(0)$, since $X_i(0) \in \{0, 1\}$. Th The main application of Theorem 8.2.1 is the following

Corollary 8.2.1. *Consider the branching process on a finite graph G on N nodes, with base infection rate β and arbitrary initial condition $X(0) \in \{0, 1\}^N$. Let τ denote the time to absorption at 0 by the process. Then, under the condition*

$$\beta\rho < 1, \tag{8.7}$$

where ρ is the spectral radius of the adjacency matrix of G, it holds that:

$$\mathbf{E}(\tau) \leq \frac{\log(N) + 1}{1 - \beta\rho}. \tag{8.8}$$

Proof. Write

$$\begin{aligned}\mathbf{E}(\tau) &= \int_0^\infty \mathbf{P}(\tau > 0)dt \\ &= \int_0^\infty \mathbf{P}(X(t) \neq 0)dt \\ &\leq \int_0^\infty \min\left(1, N \exp(-(1 - \beta\rho)t)\right) dt \\ &= t^* + \int_{t^*}^\infty N \exp(-(1 - \beta\rho)t)dt,\end{aligned}$$

where $t^* = \log(N)/(1 - \beta\rho)$. We thus obtain:

$$\mathbf{E}(\tau) \leq t^* + \frac{N}{1 - \beta\rho} \exp(-(1 - \beta\rho)t^*) = \frac{\log(N) + 1}{1 - \beta\rho}.$$

□

8.3 Small outbreaks and spectral radius for the SIR epidemics

We consider a closed population of n individuals, connected by a neighbourhood structure which is represented by an undirected, labelled graph $\mathcal{G} = (V, E)$ with node set V and edge set E. Each node can be in one of three possibly states, susceptible (S), infective (I) or removed (R). The initial set of infectives at time 0 is assumed to be non-empty, and all other nodes are assumed to be susceptible at time 0. The evolution of the epidemic is described by the following discrete-time model. Let $X_v(t)$ denote the indicator that node v is infected at the beginning of time slot t and $Y_v(t)$ the indicator that it is removed.

Each node that is infected at the beginnning of a time slot attempts to infect each of its neighbours; each infection attempt is successful with probability β independent of other infection attempts. Each infected node is removed at the end of the time slot. Thus, the probability that a susceptible node u becomes infected at the end of time slot t is given by $1 - \prod_{v \sim u}(1 - \beta X_v(t))$, where we write $v \sim u$ to mean that $(u, v) \in E$. Note that the evolution stops when there are no more infectives in the population. At this time, we want to know how many nodes are removed.

The above model is known as the Reed-Frost model. It corresponds to a deterministic infectious period which is the same at every node. It is one of the earliest stochastic SIR models to be studied in depth, because of its analytical tractability. Note that the evolution can be described by a Markov chain in this case. Another commonly used model assumes that infectious periods are iid and exponentially distributed, so that the system evolves as a continuous time Markov process. General infectious periods give rise to non-Markovian systems. These are outside the scope of this work.

The object of interest is the number of nodes that eventually become infected (and removed) compared to the number initially infected. As noted earlier, in mean field models of SIR epidemics, the number of nodes removed exhibits a sharp threshold; as β is increased, it suddenly jumps from a constant (which doesn't depend on n) to a non-zero fraction of n, the number of nodes in the system. We wish to ask if a similar threshold is exhibited on general graphs and, if so, how the critical value of β is related to properties of the graph.

We now state general conditions for the number of nodes removed to be small. Let A denote the adjacency matrix of the undirected graph G, i.e., $a_{ij} = 1$ if $(i, j) \in E$ and $a_{ij} = 0$ otherwise. Since A is a symmetric, non-negative matrix, all its eigenvalues are real, the eigenvalue with the largest absolute value is positive and its associated eigenvector has non-negative entries (by the Perron-Frobenius theorem). If the graph is connected, as we shall assume, then this eigenvalue has multiplicity one, and the corresponding eigenvector is the only one with all entries being non-negative.

Theorem 8.3.1. *Suppose $\beta\lambda_1 < 1$. Then, the total number of nodes removed, $|Y(\infty)|$, satisfies*

$$\mathrm{E}[|Y(\infty)|] \leq \frac{1}{1 - \beta\lambda_1}\sqrt{n|X(0)|},$$

where $|X(0)|$ is the number of initial infectives. Morevoer, if the graph G is regular (i.e., each node has the same number of neighbours), then

$$\mathrm{E}[|Y(\infty)|] \leq \frac{1}{1 - \beta\lambda_1}|X(0)|.$$

Proof. In order for an arbitrary node v to be infected at the start of time slot t, there must be a chain of distinct nodes $u_0, u_1, \ldots, u_t = v$ along which the infection passes from some initial infective u_0 to v. Thus, by the union bound,

$$\mathrm{P}(X_v(t) = 1) \leq \sum_{u_0, \ldots, u_{t-1}} \beta^t X_{u_0}(0),$$

where the sum is taken over nodes $u_0, \ldots, u_{t-1}$ such that $(u_{i-1}, u_i) \in E$ for all $i = 1, \ldots, t$, where we take $u_t = v$. Note that we have not imposed the requirement that the u_i be distinct as we are only seeking an upper bound. Consequently, the probability that node v ever gets infected (and hence that $Y_v(\infty) = 1$) is bounded above by

$$\mathrm{P}(Y_v(\infty) = 1) \leq \sum_{t=0}^{\infty} \sum_{u \in V} (\beta A)^t_{uv} X_u(0),$$

since the uv^{th} entry of the matrix A^t is simply the number of paths of length t between nodes u and v. It is immediate from the above that

$$\mathrm{E}[|Y(\infty)|] = \sum_{v \in V} \mathrm{P}(Y_v(\infty) = 1) \leq \sum_{t=0}^{\infty} \mathbf{1}^T (\beta A)^t X(0),$$

where $\mathbf{1}$ denotes the vector of ones. Now, if $\beta\lambda_1 < 1$, then we can rewrite the above as

$$\begin{aligned} \mathrm{E}[|Y(\infty)|] &\leq \mathbf{1}^T (I - \beta A)^{-1} X(0) \\ &\leq \|\mathbf{1}\| \, \|(I - \beta A)^{-1}\| \, \|X(0)\|, \end{aligned} \tag{8.9}$$

where $\|\cdot\|$ denotes the Euclidean norm in the case of a vector, and the matrix or operator norm in the case of a matrix. Now the operator norm of a symmetric matrix is its spectral radius, the largest of its eigenvalues in absolute value. Hence $\|(I-\beta A)^{-1}\| = (1-\beta\lambda_1)^{-1}$. Moreover, $\|X(0)\| = \sqrt{\sum_{v \in V} X_v^2(0)} = \sqrt{|X(0)|}$. Likewise, $\|\mathbf{1}\| = \sqrt{n}$. Substituting these in (8.9) yields

$$\mathrm{E}[|Y(\infty)|] \leq \frac{1}{1 - \beta\lambda_1} \sqrt{n|X(0)|},$$

which is the first claim of the theorem.

Next, note that by using the spectral decomposition

$$(I - \beta A)^{-1} = \sum_{i=1}^{n} \frac{1}{1 - \beta\lambda_i} x_i x_i^T,$$

where x_i denotes the eigenvector corresponding to the eigenvalue λ_i of A, and x_i^T its transpose, we can rewrite (8.9) as

$$\mathrm{E}[|Y(\infty)|] \leq \sum_{i=1}^{n} \frac{1}{1 - \beta\lambda_i} \mathbf{1}^T x_i x_i^T X(0). \tag{8.10}$$

Now, if G is a regular graph and each node has degree d (i.e., has exactly d neighbours), then each row sum of its adjacency matrix A is equal to d. Hence, it is clear that the positive vector $\frac{1}{\sqrt{n}}\mathbf{1}$ is an eigenvector of A corresponding to the eigenvalue d. By the Perron-Frobenius theorem, this is therefore the largest eigenvalue. Hence, $\lambda_1 = d$, $x_1 = \frac{1}{\sqrt{n}}\mathbf{1}$, and all other eigenvectors $x_2, \ldots, x_n$ are orthogonal to $\mathbf{1}$. Hence, by (8.10),

$$\begin{aligned} \mathrm{E}[|Y(\infty)|] &\leq \frac{1}{1 - \beta\lambda_1} \mathbf{1}^T x_1 x_1^T X(0) \\ &= \frac{1}{n(1 - \beta\lambda_1)} \mathbf{1}^T \mathbf{1}\mathbf{1}^T X(0) = \frac{1}{1 - \beta\lambda_1} |X(0)|. \end{aligned}$$

This is the second claim of the theorem. □

Actually, there is an easier proof. Let $\nu(i) = \mathrm{P}(\text{node } i \text{ is ever infected})$. Then, $\nu(i) = 1$ if $i \in I$, where I denotes the set of initial infectives, and otherwise $\nu(i) \leq \sum_{j\sim i} \beta\nu(j)$, where we write $j \sim i$ to mean that (i, j) is an edge. Thus,

$$(I - \beta A)\nu \leq \mathbf{1}_I, \tag{8.11}$$

where $\mathbf{1}_I$ denotes the vector with components 1 for $i \in I$ and 0 for $i \notin I$, and the inequality holds in the usual partial order, namely componentwise. Now, if $\beta\lambda_1(A) < 1$, then we have the power series expansion

$$(I - \beta A)^{-1} = \sum_{k=0}^{\infty} \beta^k A^k,$$

from which it is immediate that $(I - \beta A)^{-1}$ is a non-negative matrix. Therefore, we can multiply both sides of the inequality in (8.11) by $(I - \beta A)^{-1}$ to obtain

$$\nu = \mathrm{E}[Y(\infty)] \leq (I - \beta A)^{-1} X(0),$$

and so

$$\mathrm{E}[|Y(\infty)|] \leq \mathbf{1}(I - \beta A)^{-1} X(0).$$

This is the same as (8.9), and the proof carries on the same way from there.

The theorem says that, if $\beta\lambda_1 < 1$, then starting from a 'small' population of initial infectives, the final size of the epidemic is small. For example, if $|X(0)| = 1$, then the final size of the epidemic is bounded by a constant in the case of regular graphs, and by a multiple of $\sqrt{n}$ in general. Thus, the fraction of nodes infected goes to zero as n tends to infinity.

Note that the proof of the theorem above doesn't require us to assume that the epidemic be of Reed-Frost type. It works for general infectious periods J since we are only using expectations throughout, which don't require independence assumptions. Therefore, following the steps of the above proof and replacing β by the probability that a node gets infected bu an infected neighbour.

In turn, if node u is infected, it will infect j, if they are connected and if the time it takes to contact this node given by an exponential random variable with parameter λ is less than J.

Theorem 8.3.2. *Suppose that J is such that $\mathrm{E}[e^{-\lambda J}] < \infty$ and let*

$$p_J = 1 - \mathrm{E}[e^{-\lambda J}]\,.$$

If $p_J\lambda_1 < 1$ then the total number of nodes removed, $|Y(\infty)|$, satisfies

$$\mathrm{E}[|Y(\infty)|] \leq \frac{1}{1 - p_J\lambda_1}\sqrt{n|X(0)|},$$

where $|X(0)|$ is the number of initial infectives. Morevover, if the graph G is regular (i.e., each node has the same number of neighbours), then

$$\mathrm{E}[|Y(\infty)|] \leq \frac{1}{1 - p_J\lambda_1}|X(0)|.$$

8.4 Long survival and isoperimetric constants

We now provide a sufficient condition for long survival of the epidemics. This is phrased in terms of isoperimetric constants of the supporting graph, which we now define:

Definition 8.4.1. *For a graph G on the node set $\{1,\dots,N\}$, and any integer $m < N$, the isoperimetric constant $\eta(m)$ of graph G is defined by*

$$\eta(m) = \min_{S\subset\{1,\dots,N\},|S|\le m} \frac{E(S,\bar{S})}{|S|}, \tag{8.12}$$

where $\bar{S}$ denote the complementary set $\{1,\dots,N\}\setminus S$, and $E(S,T)$ denotes the number of edges with one end point in set S and the other in set T.

The main result of this section is the following

Theorem 8.4.1. *Let a finite graph G on N nodes be given, and assume that for some $m < N$, and some $r \in (0,1)$, it holds that*

$$\beta\eta(m) \ge \frac{1}{r}, \tag{8.13}$$

where $\eta(m)$ denote the isoperimetric constant of G. Then, denoting by τ the time to absorption of the contact process on G, for any initial condition $X(0) \neq 0$, it holds that:

$$\mathbf{P}(\tau \ge \frac{s}{2m}) \ge \frac{1-r}{1-r^m}\left(\frac{1-r^{m-1}}{1-r^m}\right)^s \left(1-o(s^{-1})\right), \quad s\in\mathbb{N}, \tag{8.14}$$

where the term $o(s^{-1})$ is independent of the model parameters.

Proof. Consider the Markov jump process $\{Z(t)\}_{t\ge0}$ defined on the state space $\{0,\dots,m\}$, with non-zero transition rates

$$\begin{aligned} q(z,z+1) &= \tfrac{z}{r}\mathbf{1}_{z<m},\ z\in\{0,\dots,m\},\\ q(z,z-1) &= z,\ z\in\{0,\dots,m\}. \end{aligned}$$

We now show that for any initial condition $X(0)\neq 0$, the contact process on G can be coupled with the process $\{Z(t)\}_{t\ge0}$ with initial condition $Z(0)=1$, in such a way that $\sum_{i=1}^N X_i(t) \ge Z(t)$ for all $t\ge0$. To this end, we define the joint process (X,Z) on the state space $\{(x,z)\in\{0,1\}^N\times\{0,\dots,m\}, z\le\sum_{i=1}^N x_i\}$ as follows. For any state (x,z), any $i\in\{1,\dots,N\}$, if $\sum_{i=1}^N x_i > z$, we have the non-zero transition rates

$$\begin{aligned} q((x,z),(x+e_i,z)) &= \beta(1-x_i)\textstyle\sum_{j\sim i}x_j,\\ q((x,z),(x-e_i,z)) &= x_i,\\ q((x,z),(x,z+1)) &= r^{-1}z\mathbf{1}_{z<m},\\ q((x,z),(x,z-1)) &= z. \end{aligned}$$

If $\sum_{i=1}^N x_i = z$, the non-zero transition rates are:

$$\begin{array}{ll} q((x,z),(x+e_i,z+1) & = c_i(x), \\ q((x,z),(x+e_i,z) & = \beta(1-x_i)\sum_{j\sim i} x_j - c_i(x), \\ q((x,z),(x-e_i,z-1)) & = x_i, \end{array}$$

where the rates $c_i(x)$ are chosen to satisfy the following conditions:

$$0 \leq c_i(x) \leq \beta(1-x_i)\sum_{j\sim i} x_j,\ i \in \{1,\ldots,N\},$$

which ensures that the transition rates are non-negative, and

$$\sum_{i=1}^N c_i(x) = r^{-1} z \mathbf{1}_{z<m}.$$

Let us show that such rates $c_i(x)$ exist. This will be the case if we have

$$\sum_{i=1}^N \beta(1-x_i)\sum_{j\sim i} x_j \geq r^{-1} z \mathbf{1}_{z<m}.$$

Note now that the left-hand side of this equation also reads $\beta E(S,\bar{S})$, where S denotes the set of sites $j \in \{1,\ldots,N\}$ such that $x_j = 1$. Note also that $|S| = \sum_j x_j = z \leq m$; hence, by the definition of the isoperimetric constant $\eta(m)$, the left-hand side is larger than or equal to $\beta\eta z$. In view of Condition (8.13), this is indeed larger than z/r.

One can then easily verify that the component processes have the desired dynamics by checking that Lemma 8.2.1 applies. This coupling entails that

$$\mathbf{P}(\tau > s) \geq \mathbf{P}(Z(s) = 0).$$

To evaluate the right-hand side of the above, consider the discrete time embedded Markov chain $\{Y(n)\}_{n\geq 0}$ keeping track of the states visited by process $\{Z_t\}_{t\geq 0}$. Its non-zero transition probabilities are given by:

$$\begin{array}{ll} \mathbf{P}(Y(n+1) = y+1 | Y(n) = y) & = \frac{y/r}{y/r+y} = \frac{1}{1+r}, \quad y \in \{1,\ldots,m-1\}, \\ \mathbf{P}(Y(n+1) = y-1 | Y(n) = y) & = \frac{y}{y/r+y} = \frac{r}{r+1}, \quad y \in \{1,\ldots,m-1\}, \\ \mathbf{P}(Y(n+1) = m-1 | Y(n) = m) & = 1, \\ \mathbf{P}(Y(n+1) = 0 | Y(n) = 0) & = 1. \end{array}$$

The probability π_k that, starting from state $k \in \{0,\ldots,m\}$, the chain $\{Y(n)\}_{n\geq 0}$ hits m before it is absorbed at 0 is given by:

$$\pi_k = \frac{1-r^k}{1-r^m}.$$

This is a classical result, which is the solution of the so-called gambler's ruin problem. To establish this, note that necessarily,

$$\pi_0 = 0,\ \pi_m = 1,\ (1+r)\pi_k = r\pi_{k+1} + \pi_{k-1},\ k \in \{1, \ldots, m-1\},$$

and verify that the only solution to these relations is the one given above.

Thus, the probability that process $\{Z_t\}_{t\geq 0}$ pays at least s visits to state m before being absorbed at 0 is the probability that the chain $\{Y(n)\}_{n\geq 0}$ pays at least s visits to state m. By the above formula, this reads

$$\mathbf{P}(\{Y(n)\}_{n\geq 0} \text{ visits state } m \text{ at least } s \text{ times}) = \frac{1-r}{1-r^m}\left(\frac{1-r^{m-1}}{1-r^m}\right)^s.$$

After each entrance into state m, process $\{Z_t\}_{t\geq 0}$ remains there for an exponentially distributed sojourn time, with mean $1/m$. Thus, the probability that process $\{Z_t\}_{t\geq 0}$ is not absorbed by time $s/2m$ verifies

$$\mathbf{P}(Z_{s/(2m)} > 0) \geq \mathbf{P}(\sum_{i=1}^{s} E_i \geq s/2)\frac{1-r}{1-r^m}\left(\frac{1-r^{m-1}}{1-r^m}\right)^s,$$

where the random variables E_i are i.i.d., exponentially distributed with mean 1. Chernoff's Lemma 2.4.2 entails that the first term in the right-hand side verifies

$$\mathbf{P}(\sum_{i=1}^{s} E_i \geq s/2) \geq 1 - \exp(-sh_{exp}(1/2)),$$

where

$$\begin{aligned} h_{exp}(x) &= \sup_{\theta\in\mathbb{R}}(\theta x - \log \mathbf{E}(\exp(\theta E_1))) \\ &= \sup_{\theta\in\mathbb{R}}(\theta x - \log(1/(1-\theta))) \\ &= x - 1 - \log(x). \end{aligned}$$

The term $\exp(-sh_{exp}(1/2))$ is clearly $o(s^{-1})$, and the result (8.14) follows. □

The main application of this result is the following

Corollary 8.4.1. *Assume that for an infinite sequence of integers N, one is given a finite graph G_N on N nodes, a base infection rate β_N, and an integer $m_N \geq N^a$, where a is a fixed positive constant, such that*

$$\beta_N \eta(m_N, G_N) \geq \frac{1}{r},$$

where $r \in (0,1)$ is fixed. Then, denoting by τ_N the time to extinction of the contact process on G_N, with parameter β_N, it holds that

$$\mathbf{E}(\tau_N) \geq \exp(bN^a), \tag{8.15}$$

for some positive constant $b > 0$.

Proof. Let $N > 0$ be fixed. By (8.14), it holds that for all $s \in \mathbb{N}$,

$$\mathbf{E}(\tau_N) \geq \frac{s}{2m} \frac{1-r}{1-r^m} \left(\frac{1-r^{m-1}}{1-r^m} \right)^s \left(1 - o(s^{-1})\right).$$

where $m = m_N$. Take now $s = \lfloor r^{-m+1} \rfloor$ in the above expression to obtain

$$\begin{aligned} \mathbf{E}(\tau_N) &\geq \tfrac{\lfloor r^{-m+1} \rfloor}{2m} \tfrac{1-r}{1-r^m} \left(\tfrac{1-r^{m-1}}{1-r^m} \right)^s \left(1 - o(s^{-1})\right) \\ &\geq (1-r)(1-O(r^m)) \tfrac{\lfloor r^{-m+1} \rfloor}{2m} \exp(-(1-r)/(1-r^m)) \\ &\geq \tfrac{1-r}{2e}(1-O(r^m)) \exp(\log(1/r)(m-1) - \log(2m)). \end{aligned}$$

For $m \geq N^a$, the exponent $\log(1/r)(m-1) - \log(2m)$ is clearly larger than bN^a for some suitable constant $b > 0$ (taking e.g. $b = \log(1/r)/2$), and the result follows. □

8.5 Epidemics on specific graphs

8.5.1 Application to hypercubes

8.5.2 Application to E-R graphs

8.5.3 Application to star networks

Bibliography

[1] N. Alon, J. Spencer, *The probabilistic method*, 2nd ed., Wiley, 2000.

[2] Athreya and Ney, *Branching processes*, Dover, 2004.

[3] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks", Science 286, 509-512, 1999.

[4] B. Bollobás, *Random Graphs*, Cambridge University Press, 2nd Edition, 2001.

[5] B. Bollobás, C. Borgs, J. T. Chayes and O. Riordan, "Directed Scale-Free Graphs". Proceedings of the 14th ACM-SIAM Symposium on Discrete Algorithms (SODA), 132-139, 2003.

[6] S. Janson, T. Luczak, A. Ruciński, *Random Graphs*, Wiley Interscience, 2000.

[7] J. Kleinberg, "The Small-World Phenomenon: An Algorithmic Perspective", Proc. 32nd ACM Symposium on Theory of Computing, 2000.

[8] T.M. Liggett, *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*. Springer, 1999.

[9] T. Lindvall, *Lectures on the Coupling Method*, Wiley & Sons, New York 1992.

[10] L. Lovász, K. Vesztergombi, *Discrete Mathematics*, available at http://research.microsoft.com/users/lovasz/dmbook.ps

[11] S. Milgram, "The small world problem", Psychology Today 1(61), 1967.

[12] B.Mandelbrot, "An information theory of the structure of languages". In Communication Theory, ed. W. Jackson, 486-502. Woburn, MA: Butterworth, 1953.

[13] M. Mitzenmacher, "A Brief History of Generative Models for Power Law and Lognormal Distributions". Internet Mathematics 1(2), 226-251, 2004.

[14] M.E.J. Newman, "Power laws, Pareto distributions and Zipf's law", 2004, available at http://arxiv.org/abs/cond-mat/0412004

[15] D. Watts, *Small worlds*, Wiley, 1999.

[16] G.U. Yule, "A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis". Philos. Trans. R. Soc. London B 213, 21-87, 1925.