10 Epidemic Processes on Random Graphs

10.1 Introduction

How far and how fast does information propagate on a network, running a local distributed algorithm?

Suppose that one node (the origin, without loss of generality) of a graph $G(V, E)$, with $|V| = n$, the *source* of information, wants to broadcast a message to all the $n-1$ other nodes in V. The information can be obtained by a node i only from its direct neighbors, without any form of coordination other than the possible synchronization of the time clocks if the algorithms run under a discrete-time model. There are different popular models for information dissemination, among which epidemic processes play a central role. We will consider is a class of model known, as *SI(R) epidemics model*, where each node relays the message (or virus) at most one time with some fixed probability z to all its neighbors, and either remains infected (SI model) or recovers and is removed from the epidemics (SIR).

Roughly speaking, one can distinguish the three main canonical models, in which nodes involved in the epidemic process can be *susceptible* (S) of being contaminated by a virus but are still healthy, or can be *infected* (I) with the virus, or can have *recovered* (R) from the infection:

- **SI** (Susceptible-Infected): this is the simplest model, where all nodes are in state S, except for the source of the virus, which can transmit the virus to all its susceptible neighbors who then become infected, and, in turn, contaminate their neighbors. Transitions from S to I are irreversible. The question is then to know the size of the component of infected nodes; and in particular if it will be proportional to the total number of vertices n.
- **SIR** (Susceptible-Infected-Recovered): this is an extension of the SI model, where infected nodes (I) remain so during only some amount of time (during which the can contaminate their susceptible neighbors), but after which they recover (R) and are removed from the epidemics, and therefore can no longer be susceptible to the virus. Transitions from S to I and finally to R are irreversible, and the process becomes a "wave" that propagates in the graph.
- **SIS** (Susceptible-Infected-Susceptible): here, nodes switch from a susceptible state to an infected state, and vice-versa. The behavior of interest here is the steady state. Such processes are also close to the so-called contact model in Physics.

We will focus here on the SI/SIR model, where all nodes are initially susceptible (except the source, which is already infected). We first briefly describe the most frequent formulation of the SI and SIR model in a population model, where every individual can be in contact any other individual. In terms other words, the contact network is a full, complete network (with an edge between every pair of nodes). In this case, the model is to set of coupled nonlinear ordinary differential equations. When the network is random tree (or lattice), the SI/SIR epidemics boils down to a percolation problem, see Chapters 2 and 8. We will consider contact network that are fixed degree network, under the pairing/configuration model in Chapter 4. The problem is to find out whether there is a giant component in the network of infected nodes, which is a problem which we have already addressed in a number of scenarios, depending on the node degree distribution.

10.2 Epidemics on Complete Contact Networks

10.2.1 SI Epidemics

Let $N_S(t)$ (respectively, $N_I(t)$) denote the number of individuals in the susceptible (resp., infected) state at time $t \in \mathbb{R}^+$, which we model by a continuous-time Markov chain. Clearly, in the SI model $N_S(t) + N_I(t) = n$.

Suppose that people meet each other uniformly at random, at an average rate of β contacts per time unit. Since the virus can only be transmitted from an infected individual to a susceptible one, the probability that a susceptible individual at time t gets infected by time $t + \Delta t$ is proportional to the proportion of infected individuals in the population at time t, which is $N_I(t)/n$. Therefore the probability that any of the $N_S(t)$ susceptible individuals becomes infected Δt time units later is taken as

$$\mathbb{P}(N_I(t + \Delta t) = i + 1 \mid N_I(t) = i, N_S(t) = s) = \beta \frac{i}{n} s\Delta t + o(\Delta t),$$

where $o(\Delta t)$ is a function such that $o(\Delta t)/\Delta t \to 0$ when $\Delta t \to 0$. Since Δt is very small, one can assume that for all $j > i + 1$

$$\mathbb{P}(N_I(t + \Delta t) = j \mid N_I(t) = i, N_S(t) = s) = o(\Delta t),$$

whereas for $j < i - 1$,

$$\mathbb{P}(N_I(t + \Delta t) = j \mid N_I(t) = i, N_S(t) = s) = 0.$$

Therefore

$$\mathbb{P}(N_I(t + \Delta t) = i \mid N_I(t) = i, N_S(t) = s) = 1 - \beta \frac{i}{n} s\Delta t + o(\Delta t).$$

Using these transition probabilities and writing $p_i(t) = \mathbb{P}(N_I(t) = i)$, we get

$$p_i(t + \Delta t) = \beta \, \frac{(i-1)(n-i+1)}{n} \Delta t \; p_{i-1}(t) + \left(1 - \beta \frac{i(n-i)}{n} \Delta t\right) p_i(t) + o(\Delta t),$$

whence

$$\frac{p_i(t + \Delta t) - p_i(t)}{\Delta t} = \beta \frac{(i-1)(n-i+1)}{n} p_{i-1}(t) - \beta \frac{i(n-i)}{n} p_i(t) + \frac{o(\Delta t)}{\Delta t}.$$

Letting $\Delta t \to 0$, one gets the Kolmogorov equation

$$\frac{dp_i}{dt}(t) = \beta \, \frac{(i-1)(n-i+1)}{n} p_{i-1}(t) - \beta \, \frac{i(n-i)}{n} p_i(t) \tag{10.1}$$

We solve this equation using the probability generating function

$$G_I(z; t) = \mathbb{E}\left[z^{N_I(t)}\right] = \sum_{i=0}^{n} z^i p_i(t). \tag{10.2}$$

We multiply both sides of (10.1) by z^i and add all the equations for $1 \leq i \leq n$, to find, after a few manipulations,

$$\frac{dG_I}{dt}(z;t) = \frac{\beta\,(n-1)}{n} z(z-1)\frac{dG_I}{dz}(z;t) - \frac{\beta}{n} z^2(z-1)\frac{d^2G_I}{dz^2}(z;t) \tag{10.3}$$

where we used

$$\begin{aligned} \frac{dG_I}{dz}(z;t) &= \sum_{i=0}^{n} i z^{i-1} p_i(t) \\ \frac{d^2G_I}{dz^2}(z;t) &= \sum_{i=0}^{n} i(i-1) z^{i-1} p_i(t). \end{aligned}$$

In $z = 1$, these quantities become

$$\begin{aligned} \frac{dG_I}{dz}(1;t) &= \mathbb{E}[N_I(t)] \\ \frac{d^2G_I}{dz^2}(1;t) &= \mathbb{E}[N_I^2(t)] - \mathbb{E}[N_I(t)]. \end{aligned}$$

Therefore, taking the derivative of (10.3) with respect to z, and setting $z = 1$ we find

$$\frac{d\mathbb{E}[N_I(t)]}{dt}(t) = \beta\mathbb{E}[N_I(t)] - \frac{\beta}{n}\,\mathbb{E}[N_I^2(t)]. \tag{10.4}$$

We note that this differential equation depends on the first and on the second moments of $N_I(t)$. A mean-field *approximation* amounts to replace $\mathbb{E}[N_I^2(t)]$ by $\mathbb{E}^2[N_I(t)]$, which allows to replace (10.4) by a simpler deterministic differential equation. let $I(t) = \mathbb{E}[N_I(t)]/n$ (respectively, $S(t) = \mathbb{E}[N_S(t)]/n$) be the average proportion of individuals in the infected (resp., susceptible) state. Then (10.4) becomes

$$\frac{dI}{dt}(t) = \beta I(t)(1 - I(t)), \tag{10.5}$$

which is known as the *logistic equation*, and whose solution is

$$I(t) = \frac{I_0 \exp(\beta t)}{1 - I_0 + I_0 \exp(\beta t)},$$

where $I_0 = I(0)$ is the initial proportion of infected individuals at time 0. The resulting logistic growth curve $I(t)$ is plotted as a function of t in Figure 10.1 (left).

Note that we can recast (10.4) to show that $I(t)$ and $S(t) = 1 - I(t)$ evolve as

$$\frac{dS}{dt}(t) = -\beta S(t) I(t) \tag{10.6}$$

$$\frac{dI}{dt}(t) = \beta S(t) I(t). \tag{10.7}$$

10.2.2 SIR Epidemics

The SIR model requires an additional parameter, γ, which is the rate at which an infected individual recovers from the virus (more precisely, an infected individual remains in this state during an exponentially distributed time, with mean $1/\gamma$. The SIR model extends (10.6) and (10.7) to

$$\frac{dS}{dt}(t) = -\beta S(t) I(t) \tag{10.8}$$

$$\frac{dI}{dt}(t) = \beta S(t) I(t) - \gamma I(t) \tag{10.9}$$

$$\frac{dR}{dt}(t) = \gamma I(t), \tag{10.10}$$

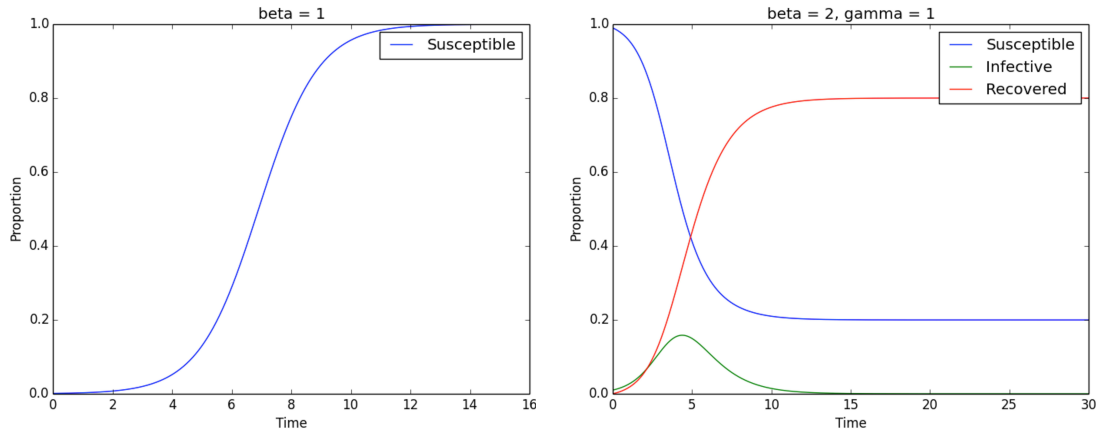

Figure 10.1: Time evolution of $I(t)$ in the SI model, with $\beta = 2$ and a complete contact graph (left). Time evolution of $S(t)$, $I(t)$ and $R(t)$ in the SIR model (right) with complete contact graph, with $\beta = 2$ and $\gamma = 1$.

where $R(t) = 1 - S(t) - I(t)$ is the fraction of recovered individuals. Eliminating I between (10.8) and (10.10), we get

$$\frac{1}{S}\frac{dS}{dt} = -\frac{\beta}{\gamma}\frac{dR}{dt},$$

and we integrate both sides with respect to t to get $S(t) = S_0 \exp(-\beta R(t)/\gamma)$, where $S_0 = S(0)$ is the initial proportion of susceptible individuals, and assuming $R(0) = 0$. Combining this relation with $I(t) = 1 - R(t) - S(t)$, we eventually replace (10.10) by

$$\frac{dR}{dt}(t) = \gamma\left(1 - R(t) - S_0 e^{-\beta R(t)/\gamma}\right).$$

We cannot solve analytically this o.d.e., but only numerically. The three curves $S(t)$, $I(t)$ and $R(t)$ are plotted as a function of t on Figure 10.1 (right), and their shape depend on the ratio $\mathcal{R}_0 = \beta/\gamma$, which is known in epidemiology as the basic *reproduction number*.

10.3 Epidemics on Fixed Degree Contact Networks

We now move to epidemics on random graphs. We assume that the infection time is constant and without loss of generality take it equal to one unit of time. We assume that the infection time is constant and without loss of generality take it equal to one unit of time. Random infection times can be handled as well [7, 19]. Each infected node has a probability ρ of contaminating any of its direct neighbors during one time unit, independently for each of them.

Let us first consider a random graph with a fixed degree distribution $G(n, \lambda)$ (See Section 4.7) where the empirical degree distribution $\lambda = \{\lambda_i\}$ is given a priori, and which generalizes the pairing (or configuration) model studied for random regular graphs $G(n, r)$. The epidemic process can be viewed as the iterative construction of a set of saturated and active nodes, a construction that we used before to prove the existence of giant components in the pairing model.

Starting from an initially infected node of $G(n, \lambda)$, the probability that j of its neighbors get infected is thus

$$p'_j = \sum_{i=j}^{\infty} \lambda_i \binom{i}{j} \rho^j (1-\rho)^{i-j},$$

and the average number of infected first neighbors is

$$\mathbb{E}[N'_{\text{infected}}] = \sum_{i=1}^{\infty} \lambda_i i\rho = \overline{\lambda}\rho,$$

where $\overline{\lambda}$ is the average node degree.

Let us now consider a second nearest neighbor of the source. Remember that in the configuration (or pairing) model, the probability of hitting a vertex of degree i is not proportional to λ_i, but to $i\lambda_i$. More precisely, the probability that a vertex has $(i+1)$ neighbors (the set of its i children and 1 parent, in the construction of the configuration model) is, for $i \geq 0$,

$$\nu_i = \frac{(i+1)\lambda_{i+1}}{\sum_{k=0}^{\infty}(k+1)\lambda_{k+1}} = \frac{(i+1)\lambda_{i+1}}{\overline{\lambda}}, \tag{10.11}$$

where the index i appears on the left because one edge was used to connect the parent to the vertex. As a result, the probability that j of its neighbors get infected is thus

$$p''_j = \sum_{i=j}^{\infty} \nu_i \left(\begin{array}{c} i \\ j \end{array} \right) \rho^j (1-\rho)^{i-j},$$

and the average number of infected descendants of a second generation (as well as any subsequent generation) is

$$\mathbb{E}[N''_{\text{infected}}] = \sum_{i=1}^{\infty} \nu_i i\rho = \overline{\nu}\rho,$$

where $\overline{\nu} = \sum_{i=1}^{\infty} i\nu_i$. Using results from branching processes (as we can do for tree percolation, but with some additional conditions to make the result rigorous), we find that the condition for the virus to continue spreading is that $\mathbb{E}[N''_{\text{infected}}] > 1$, and therefore that $\sum_{i=1}^{\infty} \nu_i i\rho > 1$, which can be recast as

$$\frac{\sum_{i=1}^{\infty} i(i-1)\lambda_i \rho}{\sum_{i=1}^{\infty} i\lambda_i} > 1, \tag{10.12}$$

using (10.11).

For a random regular graph $G(n,r)$, the degree distribution boils down to a constant, i.e.

$$\lambda_i = \left\{ \begin{array}{lll} 1 & \text{if} & i = r \\ 0 & \text{if} & i \neq r \end{array} \right.$$

and (10.12) becomes

$$\rho > \frac{1}{r-1},$$

which coincides with the percolation threshold of the $(r-1)$-ary tree $\mathbb{T}^{r-1}$: this can be excepted, since the epidemics initially propagates along a tree in the graph.

10.3.1 Scale-free networks

In this section, we consider networks whose node degrees follow a power law:

$$\lambda_i = Ci^{-\gamma},$$

for some $2 < \gamma \leq 3$, which are typical values found in empirical scale-free networks (we remember from Chapter 6 that we need $\gamma > 2$ for having a finite variance of the node degree, and that the BA model works with $\gamma = 3$).

For all $2 < \gamma \leq 3$ and all $\rho > 0$, $\sum_{i=1}^{\infty} i(i-1)\lambda_i \rho = \sum_{i=1}^{\infty} (i-1) i^{1-\gamma} \rho$ diverges and $\sum_{i=1}^{\infty} i\lambda_i < \infty$. Therefore any epidemics will propagate to a giant component, for any value $\rho > 0$, which has prompted a lot of attention, since it means that for any virus would spread on a scale free network even when the infection probability is arbitrarily small. This should be balanced however with the fact that the size of the giant component is very small. Bollobas and Riordan [4, 18] found that the size of the giant component in the BA model scales as $(\xi(\rho) + o(1))n$ w.h.p., with $\xi(\rho) = \exp(-C/\rho)$ for some constant C. In other words, if ρ is small, there is a giant component whose size is proportional to n, but with a very small proportionality constant [7].

We can also consider that $f = 1 - \rho$ is the fraction of deleted edges at random in the network: the analysis above indicates some form of robustness of scale-free networks to random failures. In contrast, it is also usually claimed that scale-free networks are vulnerable to intentional attacks, and one way to model this is by removing vertices with degree than some i_0, where i_0 is chosen so that a given fraction $1 - \rho$ of vertices is removed [7].