1

Introduction: Random Networks in Engineering

Scientists and engineers have to understand and reason about artifacts whose complexity and scale are often prohibitive. To make this task manageable, the reductionist approach calls for abstractions of reality that focus on the salient features of the problem. Such a *model* strips out all the details of the real artifact (an engineered system and its interactions with the environment) that are not crucial for understanding and reasoning. But beyond its use for solving problems, models impart on the scientist and engineer a "way of thinking", i.e., it shapes the way an engineer will approach a new problem, decompose it into manageable parts, and design a solution. In other words, the models taught to an engineering student constrains the design space that he will seek a solution in. The models are the engineer's toolbox.

The prevailing set of models in a field of engineering changes with technological progress, but is also influenced by academic tradition, intellectual elegance, etc. In the narrow disciplines of computer networking and telecommunications, the two main models and associated theories that students had to master are queueing theory and deterministic graph theory.

Queueing theory is the study of systems where clients try to obtain a service from a server. The key problem is that clients may have to await service if they request service when some other client is being served. In this case, a client is queued to await service. Queueing problems arise at various levels in networks. In packet-switched networks (such as the Internet), the arrival and departure of packets at intermediate nodes (routers) can be modeled as a queueing system to compute average packet delay and to dimension buffers. In the phone network, the calls established and terminated by users can be modeled as (another type of) queueing system to compute call block rates.

Graph theory is the study of structures consisting of nodes (vertices) that are connected by links (edges). A graph is a natural way to study global properties of a network. For example, delivering a packet or establishing a phone call involves finding an efficient path through the network to connect the originator to the destination. Graph theory provides elegant solutions to such shortest-path problems.

The goal of this class is to extend the toolbox of models of network engineers and researchers in the face of technological trends and new application scenarios that change the way we build, control, and use networks.

2

Tree percolation

2.1 Introduction

We will study different network models in this course.

A first class of models describes networks that are embedded in a geometrical space: nodes of these networks are directly connected by an edge if and only if they are close to each other in the geometric space. These models are studied by *percolation theory*, which is a discipline that emerged from studying disordered, random media. Take a large cubic block of such a material, and poor water at one of its faces. How deeply will the water seep in it? It turns out that there is a critical density of holes, below which the material will only get wet in surface, and above which the water will drain in the entire block, no matter how deep it is. This observation lead Broadbent and Hammersely to formulate it mathematically in 1957 [4], in a paper that gave birth in mathematics to the field of percolation theory. We will expose some of the main findings and techniques of percolation theory, mainly for the bond model, following closely the seminal book by Grimmett [7] on percolation theory.

We will consider different percolation models. The first one, and the easiest one, is the *bond model*, defined for the d-dimensional lattice $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$ or d-ary tree $\mathbb{T}$, where the set of edges $\mathbb{E}^d$ connects adjacent vertices. These edges represent the passages through which the water can flow. Each edge is randomly open (or maintained) with some probability p, and closed (or deleted) otherwise, independently of the others. The second one is the *site model*, defined for the d-dimensional lattice $\mathbb{L}^d$. Each vertex (or site) is now open with some probability p, and closed (or deleted) otherwise, independently of the others. The third model is the *Boolean model*, which is defined on the plane. Finally, we will move to a new model, closer to that used in wireless ad hoc/sensor networks. A common feature of percolation models is that the underlying graph is constrained by an underlying Euclidean embedding, or a lattice. We will use the term *random geometric graphs* (RGGs) for this class or random networks.

In contrast to percolation theory, which emerged from efforts to model physical phenomena such as the behavior of liquids in porous materials, *random graph theory* was devised originally as a mathematical tool in existence proofs of certain combinatorial objects. However, our goal is to study random graphs as models for networks, and this will govern our choice of results and insights we present here.

Both percolation theory and the theory of random graphs (RGs) have permeated and enriched many fields beyond the initial focus of the pioneers: mathematics, statistical physics, sociology, computer

science, biology. Another feature that random graph theory shares with percolation theory is that it remains a very active area of research, with many seemingly simple questions remaining open.

The key difference between percolation models and random graphs is that random graphs are not constrained by an underlying lattice or geometric space. Rather, every pair of vertices (nodes, sites) can potentially be connected by an edge (bond). As such, RGs are clearly not an adequate model for networks that "live" in a geometric space, such as ad hoc wireless networks. However, they may be a reasonable model for certain features of "logical" networks that live in high or infinite-dimensional spaces, such as peer-to-peer overlay networks or sociological networks. In addition, the results we derive and the methods we develop in this chapter will also serve as a basis to define and study small-world and scale-free networks.

Despite this key difference, many of the phenomena we have studied in percolation theory occur also in random graphs. There are sharp thresholds that separate regimes where various graph properties jump from being very unlikely to very likely. One of these thresholds concerns the emergence of a giant component, just as in percolation.

The simplest percolation model to study is a random tree. Contrary to square lattices and random geometric graphs (RGGs), the absence of circuit means that there is only at most one path connecting any pair of vertices and this makes the study of percolation much simpler. In addition, tree percolation is also an ingredient for the study of the giant component in a Random Graph (RG), hence it is useful to describe first the tree percolation process. We start first by listing some of terminology used in percolation theory, before computing these definitions for the tree model.

2.2 Definitions

Let $G(\mathbb{V}, \mathbb{E})$ be a arbitrary infinite connected graph, with edge set $\mathbb{E}$ and vertex set $\mathbb{V}$. Let $0 \leq p \leq 1$.

In *bond percolation*, we declare an edge of $\mathbb{E}$ to be *open* with probability p, and closed otherwise, independently of all other edges. This amounts to work on the probability space $(\Omega, \mathcal{F}, \mathbb{P}_p)$ with the sample space $\Omega = \prod_{e\in\mathbb{E}}\{0,1\}^e$ (its elements $\omega = (\omega(e) \mid e \in \mathbb{E})$ are called *configurations*, with $\omega(e) = 0$ if the edge e is closed and $\omega(e) = 1$ if the edge e is open); where $\mathcal{F}$ is the associated σ-field of subsets of Ω and where $\mathbb{P}_p$ is the product measure

$$\mathbb{P}_p = \prod_{e\in\mathbb{E}} \mu_e$$

where μ_e is a Bernoulli measure given by

$$\mu_e(\omega(e) = 0) = 1 - p, \qquad \mu_e(\omega(e) = 1) = p.$$

We denote by $\mathbb{E}_p$ the corresponding expectation operator.

There is a partial order on the set Ω of configurations, given by $\omega \leq \omega'$ if and only if $\omega(e) \leq \omega'(e)$ for all edges $e \in \mathbb{E}$.

Let us introduce some notations and terminology that will be used throughout the course. A *path* of G from vertex x_0 to vertex x_n is an alternating sequence $x_0, e_1, x_2, \ldots, e_{n-1}, x_n$ of distinct vertices x_i and edges $e_i = \langle x_{i-1}, x_i\rangle$. The length of this path is n. If all edges of the path are open, the path is an *open path*. Conversely, if all edges are closed, the path is *closed*. A circuit is a path whose first and last vertices are identical ($x_0 = x_n$).

We denote by $C(x)$ the part of G containing the set of vertices connected by open paths to vertex x and the open edges of $\mathbb{E}$ connecting such vertices. We denote by $|C(x)|$ the size (number of vertices) of $C(x)$.

We will often consider graphs that are d-dimensional lattices.

In Chapter 8, G will be a d-dimensional square lattice $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E})$, where the set of edges $\mathbb{E}$ connects sites $(x, y) = ((x_1, \ldots, x_d), (y_1, \ldots, y_d))$ located at the vertices of $\mathbb{Z}^d$ for which the Manhattan distance, defined by

$$\delta(x, y) = \sum_{i=1}^{d} |x_i - y_i|$$

is no more than one: $\delta(x, y) \leq 1$. The edges of $\mathbb{E}^d$ connect thus adjacent vertices of $\mathbb{Z}^d$. By translation invariance of the square lattice and of the probability measure $\mathbb{P}_p$, the distribution of $C(x)$ does not depend on the vertex x. We therefore take in general $x = 0$, and denote by C the open cluster at the origin: $C = C(0)$.

In this first chapter, G is a d-ary tree $\mathbb{T}$, whose root is the origin. The root is connected to d nodes, called its "children", each of its children is in turn connected to d new "grand-children", and so forth. All nodes are thus directly connected to their common ancestor and to their d direct children. Note that the origin is only connected to its d children, hence to make the tree regular, one would need to add a new child only for the origin, so that each node would then be connected to exactly $d+1$ other nodes. However this has no impact on the quantities that will be of interest to us, and therefore we do not need to care much about this detail. For the sake of simplification, we also consider only the binary tree ($d = 2$); there is no much difference with the general case. We therefore take in general, with no loss of generality, $x = 0$ in the definition of $C(x)$, and again denote by C the open cluster at the origin: $C = C(0)$.

2.3 Percolation probability

The main quantity of interest in percolation theory is the probability that the origin belongs to a cluster with an infinite number of vertices, which we denote by θ and call the *percolation probability*. With C denoting the cluster containing the origin, the percolation probability is thus defined as

$$\theta(p) = \mathbb{P}_p(|C| = \infty). \tag{2.1}$$

By space invariance, $\theta(p)$ is the probability that any node belongs to an infinite cluster.

Define the *critical* (or *percolation*) *threshold* as

$$p_c = \sup \{p \mid \theta(p) = 0\}. \tag{2.2}$$

In the one-dimensional lattices ($d = 1$), it is immediate to see that $p_c = 1$. Indeed, if $p_c < 1$, then walking along the lattice $\mathbb{L}^1 = \mathbb{T}^1$ in any direction, we will almost surely meet infinitely often an open edge, which yields that all clusters are almost surely finite. However, when $d \geq 2$, it is no longer the case. The main finding of percolation theory is that for a large class of graphs, including $\mathbb{L}^d$ or $\mathbb{T}^d$ with $d > 1$, we have that $0 < p_c < 1$, which implies that there are two phases: the *subcritical phase*, when $p < p_c$, where every vertex is almost surely in a finite open cluster, and the *supercritical phase*, when $p > p_c$, where each vertex has a non zero probability of belonging to an infinite cluster. Computing the exact value of p_c is a challenge, and still remains an open problem for most networks. In this course, we will compute it for 2-dim lattices and binary trees.

2.3.1 Percolation probability in the tree

Let us first compute $\theta(p) = \mathbb{P}_p(|C| = \infty)$. Observe that C is the family of all descendants of the root, according to a Galton-Watson branching process. Let $X(n)$ be the number of vertices belonging to C at the nth layer of the tree (with the origin being at layer 0), they form the nth generation of the

descendants of the root. Since $X(0) = 1$ (there is one node at the root of the tree), the probability that the branching process dies out for some finite n is

$$1 - \theta(p) = \mathbb{P}_p(X(n) = 0 \text{ for some } n \in \mathbb{N}^* \mid X(0) = 1).$$

Now, $X(n)$ is a homogeneous Markov chain, and we know from the theory of Markov chains that the above probability is the minimal solution of the set of equations, for all $i \in \mathbb{N}$,

$$\begin{array}{rcll} h_{i0} & = & 1 & \text{si} \quad i = 0 \\ h_{i0} & = & \sum_{j=0}^{\infty} p_{ij} h_{j0} & \text{if} \quad i \in \mathbb{N}^* \end{array} \tag{2.3}$$

where $h_{i0} = \mathbb{P}_p(X(n) = 0 \text{ for some } n \in \mathbb{N}^* \mid X(0) = i)$ and $p_{ij} = \mathbb{P}_p(X(n+1) = j \mid X(n) = i)$ are the transition probabilities of the chain. Here, $h_{i0} = h_{10}^i$ by independence between the i families born from the i ancestors, while for $j = 0, 1, 2$,

$$p_{1j} = \mathbb{P}_p(X(n+1) = j | X(n) = 1) = \left\{ \begin{array}{ll} (1-p)^2 & \text{if} \quad j = 0 \\ 2p(1-p) & \text{if} \quad j = 1 \\ p^2 & \text{if} \quad j = 2 \end{array} \right.$$

so that $1 - \theta(p) = h_{10}$ is the minimal solution of

$$h_{10} = \sum_{j=0}^{\infty} p_{1j} h_{10}^j.$$

Denoting by $G(z) = \sum_j z^j p_{1j} = (1 - p + pz)^2$ the probability generating function of the number of children of a given vertex, we see that $1 - \theta(p) = h_{10}$ is the minimal solution of

$$z = G(z) = (1 - p + pz)^2,$$

which is 1 if $p \leq 1/2$ and $((1-p)/p)^2$ if $p > 1/2$. This shows that the critical threshold for bond percolation on $\mathbb{T}$ is $p_c = 1/2$, and

$$\theta(p) = \left\{ \begin{array}{ll} 0 & \text{if} \quad p \leq 1/2 \\ 1 - \left(\frac{1-p}{p}\right)^2 & \text{if} \quad p > 1/2. \end{array} \right. \tag{2.4}$$

2.3.2 Mean cluster size

The next quantity of interest in percolation theory is the *mean size of an open cluster*, which is the expected number of vertices in the open cluster at the origin, and which we denote by

$$\chi(p) = \mathbb{E}_p[|C|]. \tag{2.5}$$

If $p > 1/2$, (2.4) yields that $|C| = \infty$ with a strictly positive probability, hence $\chi(p) = \infty$.

If $p < 1/2$, it is easy to compute the mean cluster size $\chi(p) = \mathbb{E}_p[|C|]$ for the binary tree. Indeed, any vertex v of $\mathbb{T}$ belongs to C if and only if every edge in the (unique) path connecting v to 0 is open. If the vertex v is at the nth layer of the tree, this path is open with probability p^n, and as there are 2^n vertices at the nth layer, we have that for $p < 1/2$

$$\chi(p) = \sum_{n=0}^{\infty} \sum_{v=1}^{2^n} \mathbb{E}_p \left[1_{\{0 \leftrightarrow v\}}\right] = \sum_{n=0}^{\infty} 2^n p^n = \frac{1}{2} \left(\frac{1}{2} - p\right)^{-1}. \tag{2.6}$$

We observe that as $p \uparrow 1/2$, $\chi(p) \to \infty$. At the critical threshold $p_c = 1/2$, the mean size of an open cluster is infinite although $\theta(1/2) = 0$: the next section will show that the open cluster size distribution is then heavy-tailed.

2.3.3 Cluster size distribution

The computation of $\mathbb{P}_p(|C| = n)$ is not difficult either, even though it may involve some tedious computations. Let $G_{|C|}(z) = \mathbb{E}[z^{|C|}]$ be the probability generating function of $|C|$. Since

$$\begin{aligned}
\mathbb{P}_p(|C| = 1) &= (1-p)^2 \\
\mathbb{P}_p(|C| = 2) &= 2p(1-p)^3 \\
\mathbb{P}_p(|C| = n) &= 2p(1-p)\mathbb{P}_p(|C| = n-1) + p^2 \sum_{k=1}^{n-2} \mathbb{P}_p(|C| = k)\mathbb{P}_p(|C| = n-k-1) \text{ for } n \geq 3,
\end{aligned}$$

we obtain after some computations, which we do not detail here, that

$$G_{|C|}(z) = \left(1 - 2p(1-p)z - \sqrt{1 - 4p(1-p)z}\right) / 2p^2 z$$

whose inverse z-transform yields that

$$\mathbb{P}_p(|C| = n) = \frac{1}{n} \begin{pmatrix} 2n \\ n-1 \end{pmatrix} p^{n-1}(1-p)^{n+1}. \tag{2.7}$$

Let us first use Stirling's formula $n! \approx n^n e^{-n}\sqrt{2\pi n}$ to replace

$$\begin{pmatrix} 2n \\ n-1 \end{pmatrix} = \frac{n}{n+1} \begin{pmatrix} 2n \\ n \end{pmatrix} \approx \frac{4^n}{\sqrt{\pi n}}.$$

Observe that if $p < 1/2$, then (2.7) yields that $\mathbb{P}_p(|C| = n)$ decreases exponentially fast with n, and is thus a light-tailed distribution. Conversely, if $p = 1/2$, (2.7) behaves like $\mathbb{P}_p(|C| = n) \approx 1/\sqrt{\pi n^3}$ for $n \to \infty$. Therefore, as $p \uparrow p_c$, the open cluster distribution becomes a heavy-tailed power law distribution (all its moments are infinite).