1
Introduction

The aim of this first chapter is to motivate why stochastic processes, probability theory and graph theory are useful to solve problems in network science.

In any system or node in a network, there is always a non-zero probability of failure or of error penetration. A lot of problems in quantifying the failure rate, bit error rate or the computation of redundancy to recover from hazards are successfully treated by probability theory. Often we deal in communications with a large variety of signals, calls, source-destination pairs, messages, the number of customers per region, and so on. Often, precise information at any time is not available or, if it is available, deterministic studies or simulations are simply not feasible due to the large number of different parameters involved. For such problems, a stochastic approach is often a powerful vehicle, as has been demonstrated in the field of statistical physics or thermodynamics. Failure or attacks at the network level have reestablished the interest in network robustness analyses in relation to network security. In spite of the intuitively easy concept, a globally accepted definition as well as a framework to compute the robustness of a network is still lacking. Graph and probability theory are essential to address questions like: "Which are the vulnerable nodes?", "Is this network robust?", "Where do we need to add, remove or rewire links at minimum cost in order to maximize the network robustness?"

Perhaps the first impressing result of a stochastic approach was Boltzmann's and Maxwell's statistical theory. They studied the behavior of particles in an ideal gas and described how macroscopic quantities as pressure and temperature can be related to the microscopic motion of the huge amount of individual particles. Boltzmann also introduced the stochastic notion of the thermodynamic concept of entropy S,

$$S = k \log W$$

where W denotes the total number of ways in which the ensembles of particles can be distributed in thermal equilibrium and where $k = 1.380\ 65\ 10^{-23}$ J/K is a proportionality factor, afterwards attributed to Boltzmann as the Boltzmann constant. The pioneering work of these early physicists such as Boltzmann, Maxwell and others was the germ of a large number of breakthroughs in science. Shortly after

their introduction of a stochastic theory in classical physics, the theory of quantum mechanics (see e.g. Cohen-Tannoudji *et al.*, 1977) was established. This theory proposes that the elementary building blocks of nature, the atom and electrons, can only be described in a probabilistic sense. The conceptually difficult notion of a wave function whose squared modulus expresses the probability that a set of particles is in a certain state and the Heisenberg's uncertainty relation exclude in a dramatic way our deterministic, macroscopic view of nature at the fine atomic scale. The quantum computer – the successor of our current digital computer – is looming around the corner as small quantum computing devices are currently built and tested. The next step is quantum networking, which will require different design rules in networking and will create new challenges for network science.

At about the same time as the theory of quantum mechanics was being created, Erlang applied probability theory to the field of telecommunications. Erlang succeeded to determine the number of telephone input lines m of a switch in order to serve N_S customers with a certain probability p. Perhaps his most used formula is the Erlang B formula (14.18), derived in Section 14.2.2,

$$\Pr[N_S = m] = \frac{\frac{\rho^m}{m!}}{\sum_{j=0}^{m} \frac{\rho^j}{j!}}$$

where the load or traffic intensity ρ is the ratio of the arrival rate of calls to the telephone local exchange or switch over the processing rate of the switch per line. By equating the desired blocking probability $p = \Pr[N_S = m]$, say $p = 10^{-4}$, the number of input lines m can be computed for each load ρ. Shannon, another pioneer in the field of communications, explored the concept of entropy S. He introduced (see e.g. Cover and Thomas, 1991; Walrand, 1998) the notion of the Shannon capacity of a channel, the maximum rate at which bits can be transmitted with arbitrary small (but non-zero) probability of errors, and the concept of the entropy rate of a source, which is the minimum average number of bits per symbol required to encode the output of a source. Many others have extended his basic ideas and so it is fair to say that Shannon founded the field of information theory.

An important driver in telecommunication is the concept of quality of service (QoS). Customers can use the network to transmit different types of information, such as pictures, files, voice, etc., by requiring a specific level of service depending on the type of transmitted information. For example, a telephone conversation requires that the voice packets arrive at the receiver D ms later, while a file transfer is mostly not time critical but requires an extremely low information loss probability. The value of the mouth-to-ear delay D is clearly related to the perceived quality of the voice conversation. As long as $D < 150$ ms, the voice conversation has toll quality, which is, roughly speaking, the quality that we are used to in classical telephony. When D exceeds 150 ms, rapid degradation is experienced and when $D > 300$ ms, most of the test persons have great difficulty in understanding the conversation. However, perceived quality may change from person to person and

is difficult to determine, even for telephony. Therefore, QoS is both related to the nature of the information and to the individual's desire and perception. In the future, it is believed that customers may request a certain QoS for each type of information. Depending on the level of stringency, the network may either allow or refuse the customer. Since customers will also pay an amount related to this QoS stringency, the network function that determines to either accept or refuse a call for service will be of crucial interest to any network operator. Let us now state the connection admission control (CAC) problem for a voice conversation to illustrate the relation to stochastic analysis: "How many customers m are allowed in order to guarantee that the ensemble of all voice packets reaches the destination within D ms with probability p?" This problem is exceptionally difficult because it depends on the voice codecs used, the specifics of the network topology, the capacity of the individual network elements, the arrival process of calls from the customers, the duration of the conversation and other details. Therefore, we will simplify the question. Let us first assume that the delay is only caused by the waiting time of a voice packet in the queue of a router (or switch). As we will see in Chapter 13, this waiting time T of voice packets in a single queueing system depends on (a) the arrival process: the way voice packets arrive, and (b) the service process: how they are processed. Let us assume that the arrival process specified by the average arrival rate λ and the service process specified by the average service rate μ are known. Clearly, the arrival rate λ is connected to the number of customers m. A simplified statement of the CAC problem is, "What is the maximum λ allowed such that $\Pr[T > D] < \epsilon$?" In essence, the CAC problem consists in computing the tail probability of a quantity that depends on parameters of interest. We have elaborated on the CAC problem because it is a basic design problem that appears under several disguises. A related dimensioning problem is the determination of the buffer size in a router in order not to lose more than a certain number of packets with probability p, given the arrival and service process. The above-mentioned problem of Erlang is a third example. Another example treated in Chapter 19 is the server placement problem: "How many replicated servers m are needed to guarantee that any user can access the information within k hops with probability $\Pr[h_N(m) > k] \leq \epsilon$, where ϵ is certain level of stringency and $h_N(m)$ is the number of hops towards the most nearby of the m servers in a network with N routers.

Network science aims at understanding and at modeling complex networks such as the Internet, biological and brain networks, social networks and utility infrastructures for water, gas, electricity and transport (cars, trains, ships and airplanes). Since these networks consist of a huge number of nodes N and links L, classical and algebraic graph theory is often not suited to produce even approximate results. The beginning of probabilistic graph theory is commonly attributed to the appearance of papers by Erdős and Rényi in the late 1940s. They investigated a particularly simple growing model for a graph: start from N nodes and connect in each step an arbitrary random, not yet connected pair of nodes until all L links are used. After about $N/2$ steps, as shown in Section 16.9.1, they observed the

birth of a giant component that, in subsequent steps, swallows the smaller ones at a high rate. The link density $p = L/\binom{N}{2}$ plays a crucial role the Erdős-Rényi random graph. Around a critical value $p_c \sim \frac{\log N}{N}$ (see Sections 15.7.4 and 15.7.5), the probability of connectivity jumps sharply. This phenomenon is called a phase transition and often occurs in nature. In physics it is studied in, for example, percolation theory. The Internet is best regarded as a dynamic and growing network, whose graph is continuously changing. Yet, in order to deploy services over the Internet, an accurate graph model that captures the relevant structural properties is desirable. As shown in Part III, a probabilistic approach based on random graphs seems an efficient way to learn about the Internet's intriguing behavior. Although the Internet's topology is not a simple Erdős-Rényi random graph, results such as the hopcount of the shortest path and the size of a multicast tree deduced from the simple random graphs provide a first-order estimate for the Internet.

The popularity of the Internet gave birth to several developments in electronic banking, governing, publishing and other parts of society. On-line social networks, such as Twitter and Facebook, form open laboratories to study how humans use technology. Email forwarding and Twitter retweet times (see e.g. Doerr *et al.* (2013)) are not exponential, but rather lognormal, and thus refute a Markovian approach. Similarly as about 20 years earlier, Markovian-based design failed, when Internet traffic was shown to be "bursty" (long-range dependent, self-similar and even chaotic, non-Markovian (Veres and Boda, 2000)). As a consequence, new methods are needed to compute traffic or dimension servers and networks, based on non-Markovian human responses. Further interesting questions are: "How do we determine communities?", "Who are the key influential persons in a community?", "How does information spread, become popular and age?" or "What is the topological structure of communities and how do communities grow and change over time?". The social embedding of the Internet produces many more such inspiring questions that ask for methods discussed in this book.

We hope that this brief overview motivates sufficiently to surmount the mathematical barriers. Skill with probability theory is deemed necessary to understand complex phenomena in network science. Once mastered, the power and beauty of mathematics will be appreciated.

Part I

Probability theory

2

Random variables

This chapter reviews basic concepts from probability theory. A random variable (rv) is a variable that takes certain values by chance. Throughout this book, this imprecise and intuitive definition suffices. The precise definition involves axiomatic probability theory (Billingsley, 1995).

Here, a distinction between discrete and continuous random variables is made, although a unified approach including also mixed cases via the Stieltjes integral (Hardy *et al.*, 1999, pp. 152–157), $\int g(x)df(x)$, is possible. In general, the distribution $F_X(x) = \Pr[X \leq x]$ holds in both cases, and

$$\int g(x)dF_X(x) = \sum_k g(k)\Pr[X=k] \qquad \text{where } X \text{ is a discrete rv}$$
$$= \int g(x)\frac{dF_X(x)}{dx}dx \qquad \text{where } X \text{ is a continuous rv}$$

In most practical situations, the Stieltjes integral reduces to the Riemann integral, otherwise, Lesbesgue's theory of integration and measure theory (Royden, 1988) is required.

2.1 Probability theory and set theory

Pascal (1623–1662) is commonly regarded as one of the founders of probability theory. In his days, there was much interest in games of chance[1] and the likelihood of winning a game. In most of these games, there was a finite number n of possible outcomes and each of them was equally likely. The probability of the event A of interest was defined as

$$\Pr[A] = \frac{n_A}{n} \tag{2.1}$$

[1] "La règle des partis", a chapter in Pascal's mathematical work (Pascal, 1954), consists of a series of letters to Fermat that discuss the following problem (together with a more complex question that is essentially a variant of the probability of gambler's ruin treated in Section 11.2.2): Consider the game in which two dice are thrown n times. How many times n do we have to throw the two dice to throw double six with probability $p = \frac{1}{2}$?

where n_A is the number of favorable outcomes (samples points of A). If the number of outcomes of an experiment is not finite, this classical definition of probability no longer suffices. In order to establish a coherent and precise theory, probability theory employs concepts of group or set theory.

The set of all possible outcomes of an experiment is called the sample space Ω. A possible outcome of an experiment is called a sample point ω that is an element of the sample space Ω. An event A consists of a set of sample points. An event A is thus a subset of the sample space Ω. The *complement* A^c of an event A consists of all sample points of the sample space Ω that are not in (the set) A, thus $A^c = \Omega \backslash A$. Clearly, $(A^c)^c = A$ and the complement of the sample space is the empty set, $\Omega^c = \emptyset$ or, vice versa, $\emptyset^c = \Omega$. A family $\mathcal{F}$ of events is a set of events and thus a subset of the sample space Ω that possesses particular events as elements. More precisely, a family $\mathcal{F}$ of events satisfies the three conditions that define a σ-field[2]: (a) $\emptyset \in \mathcal{F}$; (b) if $A_1, A_2, \ldots \in \mathcal{F}$, then $\cup_{j=1}^{\infty} A_j \in \mathcal{F}$; and (c) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$. These conditions guarantee that $\mathcal{F}$ is closed under countable unions and intersections of events.

Events and the probability of these events are connected by a probability measure $\Pr[.]$ that assigns to each event of the family $\mathcal{F}$ of events of a sample space Ω a real number in the interval $[0, 1]$. As **Axiom 1**, we require that

$$\Pr[\Omega] = 1 \tag{2.2}$$

If $\Pr[A] = 0$, the occurrence of the event A is not possible, while $\Pr[A] = 1$ means that the event A is certain to occur. If $\Pr[A] = p$ with $0 < p < 1$, the event A occurs with probability p.

If the events A and B have no sample points in common, $A \cap B = \emptyset$, the events A and B are called *mutually exclusive events*. As an example, the event and its complement are mutually exclusive because $A \cap A^c = \emptyset$. **Axiom 2** of a probability measure is that, for mutually exclusive events A and B, it holds that $\Pr[A \cup B] = \Pr[A] + \Pr[B]$. The definition of a probability measure and the two axioms are sufficient to build a consistent framework on which probability theory is founded. Since $\Pr[\emptyset] = 0$ (which follows from Axiom 2 because $A \cap \emptyset = \emptyset$ and $A = A \cup \emptyset$), for mutually exclusive events A and B, it holds that $\Pr[A \cap B] = 0$.

As a classical example that explains the formal definitions, let us consider the experiment of throwing a fair die. The sample space consists of all possible outcomes: $\Omega = \{1, 2, 3, 4, 5, 6\}$. A particular outcome of the experiment, say $\omega = 3$, is

[2] A field $\mathcal{F}$ posseses the properties:

(i) $\emptyset \in \mathcal{F}$;
(ii) if $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$ and $A \cap B \in \mathcal{F}$;
(iii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

This definition is redundant. For, we have by (ii) and (iii) that $(A \cup B)^c \in \mathcal{F}$. Further, by De Morgan's law $(A \cup B)^c = A^c \cap B^c$, which can be deduced from Figure 2.1 and again by (iii), the argument shows that the reduced statement (ii), if $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$, is sufficient to also imply that $A \cap B \in \mathcal{F}$.

a sample point $\omega \in \Omega$. One may be interested in the event A where the outcome is even in which case $A = \{2, 4, 6\} \subset \Omega$ and $A^c = \{1, 3, 5\}$.

If A and B are events, the union of these events $A \cup B$ can be written using set theory as

$$A \cup B = (A \cap B) \cup (A^c \cap B) \cup (A \cap B^c)$$

because $A \cap B$, $A^c \cap B$ and $A \cap B^c$ are mutually exclusive events. The relation is immediately understood by drawing a Venn diagram as in Fig. 2.1. Taking the

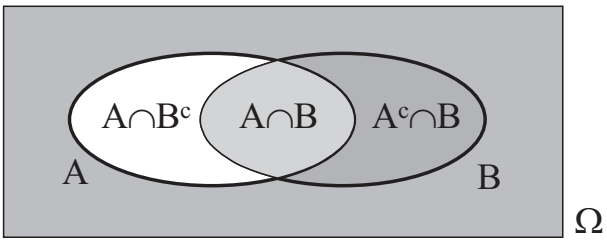

Fig. 2.1. A Venn diagram illustrating the union $A \cup B$.

probability measure of the union yields

$$\begin{aligned} \Pr[A \cup B] &= \Pr[(A \cap B) \cup (A^c \cap B) \cup (A \cap B^c)] \\ &= \Pr[A \cap B] + \Pr[A^c \cap B] + \Pr[A \cap B^c] \end{aligned} \tag{2.3}$$

where the last relation follows from Axiom 2. Figure 2.1 shows that $A = (A \cap B) \cup (A \cap B^c)$ and $B = (A \cap B) \cup (A^c \cap B)$. Since the events are mutually exclusive, Axiom 2 states that

$$\begin{aligned} \Pr[A] &= \Pr[A \cap B] + \Pr[A \cap B^c] \\ \Pr[B] &= \Pr[A \cap B] + \Pr[A^c \cap B] \end{aligned}$$

Substitution into (2.3) yields the important relation

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B] \tag{2.4}$$

Although derived for the measure $\Pr[.]$, relation (2.4) also holds for other measures, for example, the cardinality (the number of elements) of a set.

2.1.1 The inclusion-exclusion formula

A generalization of the relation (2.4) is the *inclusion-exclusion formula*,

$$\begin{aligned}\Pr\left[\cup_{k=1}^{n} A_k\right] = &\sum_{k_1=1}^{n} \Pr\left[A_{k_1}\right] - \sum_{k_1=1}^{n}\sum_{k_2=k_1+1}^{n} \Pr\left[A_{k_1}\cap A_{k_2}\right] \\ &+ \sum_{k_1=1}^{n}\sum_{k_2=k_1+1}^{n}\sum_{k_3=k_2+1}^{n} \Pr\left[A_{k_1}\cap A_{k_2}\cap A_{k_3}\right] \\ &+ \cdots + (-1)^{n-1}\sum_{k_1=1}^{n}\sum_{k_2=k_1+1}^{n}\cdots\sum_{k_n=k_{n-1}+1}^{n} \Pr\left[\cap_{j=1}^{n} A_{k_j}\right] \end{aligned} \tag{2.5}$$

The formula shows that the probability of the union consists of the sum of probabilities of the individual events (first term). Since sample points can belong to more than one event A_k, the first term possesses double countings. The second term removes all probabilities of sample points that belong to precisely two event sets. However, by doing so (draw a Venn diagram), we also subtract the probabilities of sample points that belong to three events sets more than needed. The third term adds these again, and so on. The inclusion-exclusion formula can be written more compactly as

$$\Pr\left[\cup_{k=1}^{n} A_k\right] = \sum_{j=1}^{n}(-1)^{j-1}\sum_{k_1=1}^{n}\sum_{k_2=k_1+1}^{n}\cdots\sum_{k_j=k_{j-1}+1}^{n} \Pr\left[\cap_{m=1}^{j} A_{k_m}\right] \tag{2.6}$$

or with

$$S_j = \sum_{1\le k_1<k_2<\cdots<k_j\le n} \Pr\left[\cap_{m=1}^{j} A_{k_m}\right]$$

as

$$\Pr\left[\cup_{k=1}^{n} A_k\right] = \sum_{j=1}^{n}(-1)^{j-1} S_j \tag{2.7}$$

Although impressive, the inclusion-exclusion formula is useful when dealing with dependent random variables because of its general nature. In particular, if

$$\Pr\left[\cap_{m=1}^{j} A_{k_m}\right] = a_j$$

and not a function of the specific indices k_m, the inclusion-exclusion formula (2.6) becomes more attractive,

$$\begin{aligned}\Pr\left[\cup_{k=1}^{n} A_k\right] &= \sum_{j=1}^{n}(-1)^{j-1} a_j \sum_{1\le k_1<k_2<\cdots<k_j\le n} 1 \\ &= \sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j} a_j\end{aligned}$$

An application of the latter formula to multicast can be found in Chapter 18 and many others are in Feller (1970, Chapter IV). Sometimes it is useful to reason with