1

Introduction

In this chapter, we describe a few discrete probability models to which we will come back repeatedly throughout the book. While there exists a vast array of well-studied random combinatorial structures (permutations, partitions, urn models, Boolean functions, polytopes, etc.), our focus is primarily on a limited number of graph-based processes, namely, percolation, random graphs, the Ising model, and random walks on networks. We will not attempt to derive the theory of these models exhaustively here. Instead, we will employ them to illustrate some essential techniques from discrete probability. Note that the toolkit developed in this book is meant to apply to other probabilistic models of interest as well, and in fact many more will be encountered along the way. After a brief review of graph basics and Markov chains theory in Section 1.1, we formally introduce our main models in Section 1.2. We also formulate various questions about these models that will be answered (at least partially) later on. We assume that the reader is familiar with the measure-theoretic foundations of probability. A refresher of all required concepts and results is provided in Appendix B.

1.1 Background

We start with a brief review of graph terminology and standard countable-space Markov chains results.

1.1.1 Review of Graph Theory

Basic definitions An *undirected graph* (or *graph* for short) is a pair $G = (V, E)$, where V is the set of *vertices* (or *nodes* or *sites*) and GRAPH

$$E \subseteq \{\{u, v\} : u, v \in V\}$$

is the set of *edges* (or *bonds*). See Figure 1.1 for an example. We occasionally write $V(G)$ and $E(G)$ for the vertices and edges of the graph G. The set of vertices V is either finite or countably infinite. Edges of the form $\{u\}$ are called *self-loops*. In general, we do not allow E to be a multiset unless otherwise stated. But, when E is a multiset, G is called a *multigraph*. MULTIGRAPH

A vertex $v \in V$ is *incident* with an edge $e \in E$ (or vice versa) if $v \in e$. The incident vertices of an edge are called *endvertices*. Two vertices $u, v \in V$ are *adjacent* (or *neighbors*), denoted by $u \sim v$, if $\{u, v\} \in E$. The set of adjacent vertices of v, denoted by $N(v)$, is called the *neighborhood* of v and its size, that is, $\delta(v) := |N(v)|$, is the *degree* of v. A vertex v with $\delta(v) = 0$ is called *isolated*. A graph is called *d-regular* if all its degrees are d. A countable graph is *locally finite* if all its vertices have a finite degree.

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

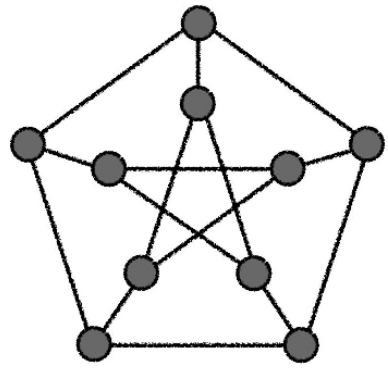

Figure 1.1 Petersen graph.

Example 1.1.1 (Petersen graph). All vertices in the Petersen graph in Figure 1.1 have degree 3, that is, it is a 3-regular graph. In particular, it has no isolated vertex. ◀

A convenient (and mathematically useful) way to specify a graph is the following matrix representation. Assume the graph $G = (V, E)$ has $n = |V|$ vertices. Assume that the vertices are numbered $1, \ldots, n$. The *adjacency matrix* A of G is the $n \times n$ symmetric matrix defined as

ADJACENCY MATRIX

$$A_{xy} = \begin{cases} 1 & \text{if } \{x, y\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Example 1.1.2 (Triangle). The adjacency matrix of a *triangle*, that is, a 3-vertex graph with all possible non-loop edges, is

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

◀

There exist other matrix representations. Here is one. Let $m = |E|$ and assume that the edges are labeled arbitrarily as $e_1, \ldots, e_m$. The *incidence matrix* of an undirected graph $G = (V, E)$ is the $n \times m$ matrix B such that $B_{ij} = 1$ if vertex i and edge e_j are incident and 0 otherwise.

INCIDENCE MATRIX

Subgraphs, paths, and cycles A *subgraph* of $G = (V, E)$ is a graph $G' = (V', E')$ with $V' \subseteq V$ and $E' \subseteq E$. Implied in this definition is the fact that the edges in E' are incident only to V'. The subgraph G' is said to be *induced* if

$$E' = \{\{x, y\} \colon x, y \in V',\ \{x, y\} \in E\},$$

that is, it contains exactly those edges of G that are between vertices in V'. In that case the notation $G' := G[V']$ is used. A subgraph is said to be *spanning* if $V' = V$. A subgraph containing all possible non-loop edges between its vertices is called a *clique* (or *complete subgraph*). A clique with k nodes is referred to as a k-clique.

CLIQUE

Example 1.1.3 (Petersen graph (continued)). The Petersen graph contains no triangle, that is, 3-clique, induced or not. ◀

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

A *walk* in G is a sequence of (not necessarily distinct) vertices $x_0 \sim x_1 \sim \cdots \sim x_k$. Note the requirement that consecutive vertices of a walk are adjacent. The number $k \geq 0$ is the *length* of the walk. If the *endvertices* x_0, x_k coincide, that is, $x_0 = x_k$, we refer to the walk as *closed*. If the vertices of a walk are all distinct, we call it a *path* (or *self-avoiding walk*). If the vertices of a closed walk are all distinct except for the endvertices and its length is at least 3, we call it a *cycle*. A path or cycle can be seen as a (not necessarily induced) subgraph of G. The length of the shortest path connecting two distinct vertices u, v is the *graph distance* between u and v, denoted by $d_G(u, v)$. It can be checked that the graph distance is a metric (and that, in particular, it satisfies the triangle inequality; see Exercise 1.6). The minimum length of a cycle in a graph is its *girth*.

GRAPH DISTANCE

We write $u \leftrightarrow v$ if there is a path between u and v. It can be checked that the binary relation $\leftrightarrow$ is an equivalence relation (i.e., it is reflexive, symmetric, and transitive; see Exercise 1.6). Its equivalence classes are called *connected components*. A graph is *connected* if any two vertices are linked by a path, that is, if $u \leftrightarrow v$ for all $u, v \in V$. Or put differently, if there is only one connected component.

Example 1.1.4 (Petersen graph (continued)). The Petersen graph is connected. ◀

Trees A *forest* is a graph with no cycle, or an *acyclic* graph. A *tree* is a connected forest. Vertices of degree 1 are called *leaves*. A spanning tree of G is a subgraph which is a tree and is also spanning. A tree is said to be *rooted* if it has a single distinguished vertex called the *root*.

TREE

Trees will play a key role and we collect several important facts about them (mostly without proofs). The following characterizations of trees will be useful. The proof is left as an exercise (see Exercise 1.8). We write $G + e$ (respectively $G - e$) to indicate the graph G with edge e added (respectively removed).

Theorem 1.1.5 (Trees: characterizations). *The following are equivalent.*

(i) The graph T is a tree.
(ii) For any two vertices in T, there is a unique path between them.
(iii) The graph T is connected, but $T - e$ is not for any edge e in T.
(iv) The graph T is acyclic, but $T + \{x, y\}$ is not for any pair of non-adjacent vertices x, y.

Here are two important implications.

Corollary 1.1.6 *If G is connected, then it has at least one spanning tree.*

Proof Indeed, from Theorem 1.1.5, a graph is a tree if and only if it is minimally connected, in the sense that removing any of its edges disconnects it. So a spanning tree can be obtained by removing edges of G that do not disconnect it until it is not possible anymore. ■

The following characterization is proved in Exercise 1.7.

Corollary 1.1.7 *A connected graph with n vertices is a tree if and only if it has $n - 1$ edges.*

And here is a related fact.

Corollary 1.1.8 *Let G be a graph with n vertices. If an acyclic subgraph H has n vertices and $n - 1$ edges, then it is a spanning tree of G.*

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

Proof If H is not connected, then it has at least two connected components. Each of them is acyclic and therefore a tree. By applying Corollary 1.1.7 to the connected components and summing up, we see that the total number of edges in H is $\leq n-2$, a contradiction. So H is connected and therefore a spanning tree. ■

Finally, a classical formula:

CAYLEY'S FORMULA

Theorem 1.1.9 (Cayley's formula). *There are k^{k-2} trees on a set of k labeled vertices.*

We give a proof of Cayley's formula based on branching processes in Exercise 6.19.

Some standard graphs Here are a few more examples of finite graphs.

COMPLETE GRAPH

- *Complete graph K_n*: This graph is made of n vertices with all possible non-loop edges.

CYCLE GRAPH

- *Cycle graph C_n (or n-cycle)*: The vertex set is $\{0, 1, \ldots, n-1\}$ and two vertices $i \neq j$ are adjacent if and only if $|i-j| = 1$ or $n-1$.

TORUS

- *Torus $\mathbb{L}_n^d$*: The vertex set is $\{0, 1, \ldots, n-1\}^d$ and two vertices $x \neq y$ are adjacent if and only if there is a coordinate i such that $|x_i - y_i| = 1$ or $n-1$ and all other coordinates $j \neq i$ satisfy $x_j = y_j$.

HYPERCUBE

- *Hypercube $\mathbb{Z}_2^n$ (or n-dimensional hypercube)*: The vertex set is $\{0,1\}^n$ and two vertices $x \neq y$ are adjacent if and only if $\|x-y\|_1 = 1$.
- *Rooted b-ary tree $\widehat{\mathbb{T}}_b^\ell$*: This graph is a tree with ℓ levels. The unique vertex on level 0 is called the root. For $j = 1, \ldots, \ell-1$, level j has b^j vertices, each of which has exactly one neighbor on level $j-1$ (its parent) and b neighbors on level $j+1$ (its children). The b^ℓ vertices on level ℓ are leaves.

INFINITE GRAPH

Here are a few examples of *infinite graphs*, that is, a graph with a countably infinite number of vertices and edges.

- *Infinite d-regular tree $\mathbb{T}_d$*: This is an infinite tree where each vertex has exactly d neighbors. The rooted version, that is, $\widehat{\mathbb{T}}_b^\ell$ with $\ell = +\infty$ levels, is denoted by $\widehat{\mathbb{T}}_b$.
- *Lattice $\mathbb{L}^d$*: The vertex set is $\mathbb{Z}^d$ and two vertices $\boldsymbol{x} \neq \boldsymbol{y}$ are adjacent if and only if $\|\boldsymbol{x}-\boldsymbol{y}\|_1 = 1$.

A *bipartite graph* $G = (L \cup R, E)$ is a graph whose vertex set is composed of the union of two disjoint sets L, R and whose edge set E is a subset of $\{\{\ell, r\} : \ell \in L,\ r \in R\}$. That is, there is no edge between vertices in L, and likewise for R.

Example 1.1.10 (Some bipartite graphs). The cycle graph C_{2n} is a bipartite graph. So is the *complete bipartite graph* $K_{n,m}$ with vertex set $\{\ell_1, \ldots, \ell_n\} \cup \{r_1, \ldots, r_m\}$ and edge set $\{\{\ell_i, r_j\} : i \in [n],\ j \in [m]\}$. ◄

In a bipartite graph $G = (L \cup R, E)$, a *perfect matching* is a collection of edges $M \subseteq E$ such that each vertex is incident to exactly one edge in M.

AUTOMORPHISM

An *automorphism* of a graph $G = (V, E)$ is a bijection ϕ of V to itself that preserves the edges, that is, such that $\{x, y\} \in E$ if and only if $\{\phi(x), \phi(y)\} \in E$. A graph $G = (V, E)$ is *vertex-transitive* if for any $u, v \in V$ there is an automorphism mapping u to v.

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

Example 1.1.11 (Petersen graph (continued)). For any $\ell \in \mathbb{Z}$, a $(2\pi\ell/5)$-rotation of the planar representation of the Petersen graph in Figure 1.1 corresponds to an automorphism. ◀

Example 1.1.12 (Trees). The graph $\mathbb{T}_d$ is vertex-transitive. The graph $\widehat{\mathbb{T}}_b^\ell$ on the other hand has many automorphisms, but is not vertex-transitive. ◀

Flows Let $G = (V, E)$ be a connected graph with two distinguished disjoint vertex sets, a *source-set* (or *source* for short) $A \subseteq V$ and a *sink-set* (or *sink* for short) Z. Let $\kappa : E \to \mathbb{R}_+$ be a *capacity* function.

Definition 1.1.13 (Flow). *A* flow *from source A to sink Z is a function $f : V \times V \to \mathbb{R}$ such that:* FLOW

F1 (Antisymmetry) $f(x,y) = -f(y,x), \forall x, y \in V$.
F2 (Capacity constraint) $|f(x,y)| \le \kappa(e), \forall e = \{x,y\} \in E$, *and* $f(x,y) = 0$ *otherwise.*
F3 (Flow-conservation constraint)

$$\sum_{y : y \sim x} f(x,y) = 0, \qquad \forall x \in V \setminus (A \cup Z).$$

For $U, W \subseteq V$, let $f(U,W) := \sum_{u \in U, w \in W} f(u,w)$. The strength *of f is $\|f\| := f(A, A^c)$.*

One useful consequence of antisymmetry is that, for any $U \subseteq V$, we have $f(U,U) = 0$ since each distinct pair $x \neq y \in U$ appears exactly twice in the sum, once in each ordering. Also if W_1 and W_2 are disjoint, then $f(U, W_1 \cup W_2) = f(U, W_1) + f(U, W_2)$. In particular, combining both observations, $f(U,W) = f(U, W \setminus U) = -f(W \setminus U, U)$.

For $F \subseteq E$, let $\kappa(F) := \sum_{e \in F} \kappa(e)$. We call F a *cutset separating* A and Z (or *cutset* for short) if all paths connecting A and Z include an edge in F. For such an F, let A_F be the set of vertices not separated from A by F, that is, vertices from which there is a path to A not crossing an edge in F. Clearly, $A \subseteq A_F$ but $A_F \cap Z = \emptyset$.

Lemma 1.1.14 (Max flow $\le$ min cut). *For any flow f and cutset F,*

$$\|f\| = f(A_F, A_F^c) \le \sum_{\{x,y\} \in F} |f(x,y)| \le \kappa(F). \tag{1.1.1}$$

Proof Since F is a cutset, $(A_F \setminus A) \cap (A \cup Z) = \emptyset$. So, by (F3),

$$\begin{aligned} f(A, A^c) &= f(A, A^c) + \sum_{u \in A_F \setminus A} f(u, V) \\ &= f(A, A_F \setminus A) + f(A, A_F^c) \\ &\quad + f(A_F \setminus A, A_F) + f(A_F \setminus A, A_F^c) \\ &= f(A, A_F \setminus A) + f(A, A_F^c) \\ &\quad + f(A_F \setminus A, A) + f(A_F \setminus A, A_F^c) \\ &= f(A_F, A_F^c) \\ &\le \sum_{\{x,y\} \in F} |f(x,y)|, \end{aligned}$$

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

where we used (F1) twice. The last line is justified by the fact that the edges between a vertex in A_F and a vertex in A_F^c have to be in F by definition of A_F. That proves the equality and the first inequality in the claim. Condition (F2) implies the second inequality. ∎

Remarkably, this bound is tight, in the following sense.

Theorem 1.1.15 (Max-flow min-cut theorem). *Let G be a finite connected graph with source A and sink Z, and let κ be a capacity function. Then the following holds*

$$\sup\{\|f\| : \textit{flow } f\} = \min\{\kappa(F) : \textit{cutset } F\}.$$

Proof Note that, by compactness, the supremum on the left-hand side is achieved. Let f be an optimal flow. The idea of the proof is to construct a "matching" cutset.

An *augmentable path* is a path $x_0 \sim \cdots \sim x_k$ with $x_0 \in A$, $x_i \notin A \cup Z$ for all $i \neq 0$ or k, and $f(x_{i-1}, x_i) < \kappa(\{x_{i-1}, x_i\})$ for all $i \neq 0$. By default, each vertex in A is an augmentable path. Moreover, by the optimality of f there cannot be an augmentable path with $x_k \in Z$. Indeed, otherwise, we could "push more flow through that path" and increase the strength of f – a contradiction.

Let $B \subseteq V$ be the set of all final vertices in some augmentable path and let F be the edge set between B and $B^c := V \setminus B$. Note that, again by contradiction, all vertices in B can be reached from A without crossing F and that $f(x, y) = \kappa(e)$ for all $e = \{x, y\} \in F$ with $x \in B$ and $y \in B^c$. Furthermore, F is a cutset separating A from Z: trivially $A \subseteq B$; $Z \subseteq B^c$ as argued above, and any path from A to Z must exit B and enter B^c through an edge in F. Thus, $A_F = B$ and we have equality in (1.1.1). That concludes the proof. ∎

COLORING

Colorings, independent sets, and matchings A *coloring* of a graph $G = (V, E)$ is an assignment of colors to each vertex in G. In a coloring, two vertices may share the same color. A coloring is *proper* if for every edge e in G the endvertices of e have distinct colors. The smallest number of colors in a proper coloring of a graph G is called the *chromatic number* $\chi(G)$ of G.

INDEPENDENT SET

MATCHING

An *independent vertex set* (or *independent set* for short) of $G = (V, E)$ is a subset of vertices $W \subseteq V$ such that all pairs of vertices in W are non-adjacent. Likewise, two edges are independent if they are not incident to the same vertex. A *matching* is a set of pairwise independent edges. A matching F is *perfect* if every vertex in G is incident to an edge of F.

Edge-weighted graphs We refer to an edge-weighted graph $G = (V, E, w)$ as a *network*. Here $w : E \to \mathbb{R}_+$ is a function that assigns positive real weights to the edges. Definitions can be generalized naturally. In particular, one defines the degree of a vertex i as

$$\delta(i) = \sum_{j : e = \{i,j\} \in E} w_e.$$

The adjacency matrix A of G is the $n \times n$ symmetric matrix defined as

$$A_{ij} = \begin{cases} w_e & \text{if } e = \{i,j\} \in E, \\ 0 & \text{otherwise,} \end{cases}$$

where we denote the vertices $\{1, \ldots, n\}$.

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

Directed graphs A *directed graph* (or *digraph* for short) is a pair $G = (V, E)$ where V is a set of *vertices* (or *nodes* or *sites*) and $E \subseteq V^2$ is a set of *directed edges* (or *arcs*). A directed edge from x to y is typically denoted by (x, y), or occasionally by $\langle x, y \rangle$. A *directed path* is a sequence of vertices $x_0, \ldots, x_k$, all distinct, with $(x_{i-1}, x_i) \in E$ for all $i = 1, \ldots, k$. We write $u \to v$ if there is such a directed path with $x_0 = u$ and $x_k = v$. We say that $u, v \in V$ *communicate*, denoted by $u \leftrightarrow v$, if $u \to v$ and $v \to u$. In particular, we always have $u \leftrightarrow u$ for every state u. The binary relation $\leftrightarrow$ relation is an equivalence relation (see Exercise 1.6). The equivalence classes of $\leftrightarrow$ are called the *strongly connected components* of G. DIGRAPH

The following definition will prove useful.

Definition 1.1.16 (Oriented incidence matrix). *Let $G = (V, E)$ be an undirected graph. Assume that the vertices of $G = (V, E)$ are numbered $1, \ldots, |V|$ and that the edges are labeled arbitrarily as $e_1, \ldots, e_{|E|}$. An* orientation *of G is the choice of a direction $\vec{e}_i$ for each edge e_i, turning it into a digraph $\vec{G}$. An* oriented incidence matrix *of G is the incidence matrix of an orientation, that is, the matrix B such that $B_{ij} = -1$ if edge $\vec{e}_j$ leaves vertex i, $B_{ij} = 1$ if edge $\vec{e}_j$ enters vertex i, and 0 otherwise.* ORIENTED INCIDENCE MATRIX

1.1.2 Review of Markov Chain Theory

Informally, a *Markov chain* (or *Markov process*) is a time-indexed stochastic process satisfying the property: conditioned on the present, the future is independent of the past. We restrict ourselves to the discrete-time, time-homogeneous, countable-space case, where such a process is characterized by its initial distribution and a transition matrix. MARKOV CHAIN

Construction of a Markov chain For our purposes, it will suffice to "define" a Markov chain through a particular construction. Let V be a finite or countable space. Recall that a *stochastic matrix* on V is a non-negative matrix $P = (P(i,j))_{i,j \in V}$ satisfying STOCHASTIC MATRIX

$$\sum_{j \in V} P(i,j) = 1, \qquad \forall i \in V.$$

We think of $P(i, \cdot)$ as a probability distribution on V. In particular, for a set of states $A \subseteq V$, we let

$$P(i, A) = \sum_{j \in A} P(i,j).$$

Let μ be a probability measure on V and let P be a stochastic matrix on V. One way to construct a Markov chain $(X_t)_{t \geq 0}$ on V with *transition matrix* P and *initial distribution* μ is the following:

- Pick $X_0 \sim \mu$ and let $(Y(i,n))_{i \in V, n \geq 1}$ be a mutually independent array of random variables with $Y(i,n) \sim P(i, \cdot)$.
- Set inductively $X_n := Y(X_{n-1}, n)$, $n \geq 1$.

So in particular:

$$\mathbb{P}[X_0 = x_0, \ldots, X_t = x_t] = \mu(x_0) P(x_0, x_1) \cdots P(x_{t-1}, x_t).$$

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

We use the notation $\mathbb{P}_x, \mathbb{E}_x$ for the probability distribution and expectation under the chain started at x. Similarly for $\mathbb{P}_\mu, \mathbb{E}_\mu$, where μ is a probability distribution.

SIMPLE RANDOM WALK ON A GRAPH

Example 1.1.17 (Simple random walk on a graph). Let $G = (V, E)$ be a finite or infinite, locally finite graph. *Simple random walk* on G is the Markov chain on V, started at an arbitrary vertex, which at each time picks a uniformly chosen neighbor of the current state. (Exercise 1.9 asks for the transition matrix.) ◀

Markov property Let $(X_t)_{t\geq 0}$ be a Markov chain (or *chain* for short) with transition matrix P and initial distribution μ. Define the filtration $(\mathcal{F}_t)_{t\geq 0}$ with $\mathcal{F}_t = \sigma(X_0, \ldots, X_t)$ (see Appendix B). As mentioned above, the defining property of Markov chains, known as the Markov property, is that given the present, the future is independent of the past. In its simplest form, that can be interpreted as $\mathbb{P}[X_{t+1} = y \mid \mathcal{F}_t] = \mathbb{P}_{X_t}[X_{t+1} = y] = P(X_t, y)$. More generally:

MARKOV PROPERTY

Theorem 1.1.18 (Markov property). *Let $f\colon V^\infty \to \mathbb{R}$ be bounded, measurable and let $F(x) := \mathbb{E}_x[f((X_t)_{t\geq 0})]$, then*

$$\mathbb{E}[f((X_{s+t})_{t\geq 0}) \mid \mathcal{F}_s] = F(X_s) \qquad a.s.$$

Remark 1.1.19 *We will come back to the "strong" Markov property in Chapter 3.*

We define $P^t(x, y) := \mathbb{P}_x[X_t = y]$. An important consequence of the Markov property (Theorem 1.1.18) is the following.

Theorem 1.1.20 (Chapman–Kolmogorov).

$$P^t(x, z) = \sum_{y\in V} P^s(x, y)P^{t-s}(y, z), \qquad s \in \{0, 1, \ldots, t\}.$$

Proof This follows from the Markov property. Indeed, note that $\mathbb{P}_x[X_t = z \mid \mathcal{F}_s] = F(X_s)$ with $F(y) := \mathbb{P}_y[X_{t-s} = z]$ and take $\mathbb{E}_x$ on each side. ■

Example 1.1.21 (Random walk on $\mathbb{Z}$). Let (X_t) be simple random walk on $\mathbb{Z}$ interpreted as a graph (i.e., $\mathbb{L}$) where $i \sim j$ if $|i-j| = 1$.[1] Then $P(0, x) = 1/2$ if $|x| = 1$. And $P^2(0, x) = 1/4$ if $|x| = 2$ and $P^2(0, 0) = 1/2$. ◀

If we write μ_s for the law of X_s as a row vector, then

$$\mu_s = \mu_0 P^s,$$

where P^s is the matrix product of P by itself s times. As is conventional in Markov chain theory, we think of probability distributions over the state space as row vectors. We will typically denote them by Greek letters (e.g., μ, π).

IRREDUCIBLE

Stationarity The *transition graph* of a chain is the directed graph on V whose edges are the transitions with *strictly positive probability*. A chain is *irreducible* if V is the unique (strongly) connected component of its transition graph, that is, if all pairs of states have a directed path between them in the transition graph.

[1] On $\mathbb{Z}$, simple random walk often refers to any nearest-neighbor random walk, whereas the example here is called simple symmetric random walk. We will not adopt this terminology here.

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

Example 1.1.22 (Simple random walk on a graph (continued)). Simple random walk on G is irreducible if and only if G is connected. ◀

A *stationary measure* π is a measure on V such that

$$\sum_{x \in V} \pi(x) P(x, y) = \pi(y), \qquad \forall y \in V,$$

or in matrix form $\pi = \pi P$. We say that π is a *stationary distribution* if in addition π is a probability measure. STATIONARY DISTRIBUTION

Example 1.1.23 (Random walk on $\mathbb{Z}^d$). The all-one measure $\pi \equiv 1$ is stationary for simple random walk on $\mathbb{L}^d$. ◀

Finite, irreducible chains always have a unique stationary distribution.

Theorem 1.1.24 (Existence and uniqueness: finite case). *If P is irreducible and has a finite state space, then:*

(i) (Existence) *it has a stationary distribution which, furthermore, is strictly positive;*
(ii) (Uniqueness) *the stationary distribution is unique.*

This result follows from Perron–Frobenius theory (a version of which is stated as Theorem 6.1.17). We give a self-contained proof.

Proof of Theorem 1.1.24 (i) We begin by proving existence. Denote by n the number of states. Because P is stochastic, we have by definition that $P\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the all-one vector. Put differently,

$$(P - I)\mathbf{1} = \mathbf{0}.$$

In particular, the columns of $P - I$ are linearly dependent, that is, the rank of $P - I$ is $< n$. That, in turn, implies that the rows of $P - I$ are linearly dependent since row rank and column rank are equal. Hence, there exists a non-zero row vector $\mathbf{z} \in \mathbb{R}^n$ such that $\mathbf{z}(P - I) = \mathbf{0}$, or after rearranging,

$$\mathbf{z}P = \mathbf{z}. \tag{1.1.2}$$

The rest of the proof is broken up into a series of lemmas. To take advantage of irreducibility, we first construct a positive stochastic matrix with $\mathbf{z}$ as a left eigenvector with eigenvalue 1. We then show that all entries of $\mathbf{z}$ have the same sign. Finally, we normalize $\mathbf{z}$.

Lemma 1.1.25 (Existence: Step 1). *There exists a non-negative integer h such that*

$$R = \frac{1}{h+1}[I + P + P^2 + \cdots + P^h]$$

is a stochastic matrix with strictly positive entries which satisfies

$$\mathbf{z}R = \mathbf{z}. \tag{1.1.3}$$

Lemma 1.1.26 (Existence: Step 2). *The entries of $\mathbf{z}$ are either all non-negative or all non-positive.*

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

Lemma 1.1.27 (Existence: Step 3). *Let*

$$\pi = \frac{\mathbf{z}}{\mathbf{z1}}.$$

Then π is a strictly positive stationary distribution.

We denote the entries of R and P^s by $R_{x,y}$ and $P^s_{x,y}$, $x, y = 1, \ldots, n$, respectively.

Proof of Lemma 1.1.25 By irreducibility (see Exercise 1.10), for any $x, y \in [n]$ there is h_{xy} such that $P^{h_{xy}}_{x,y} > 0$. Now define

$$h = \max_{x,y \in [n]} h_{xy}.$$

The matrix P^s, as a product of stochastic matrices, is a stochastic matrix for all s (see Exercise 1.11). In particular, it has non-negative entries. Hence, for each x, y,

$$\begin{aligned} R_{x,y} &= \frac{1}{h+1}[I_{x,y} + P_{x,y} + P^2_{x,y} + \cdots + P^h_{x,y}] \\ &\geq \frac{1}{h+1} P^{h_{x,y}}_{x,y} > 0. \end{aligned}$$

Moreover, the matrix R, as a convex combination of stochastic matrices, is a stochastic matrix (see Exercise 1.11).

Since $\mathbf{z}P = \mathbf{z}$, it follows by induction that $\mathbf{z}P^s = \mathbf{z}$ for all s. Therefore,

$$\begin{aligned} \mathbf{z}R &= \frac{1}{h+1}[\mathbf{z}I + \mathbf{z}P + \mathbf{z}P^2 + \cdots + \mathbf{z}P^h] \\ &= \frac{1}{h+1}[\mathbf{z} + \mathbf{z} + \mathbf{z} + \cdots + \mathbf{z}] \\ &= \mathbf{z}. \end{aligned}$$

That concludes the proof. ∎

Proof of Lemma 1.1.26 We argue by contradiction. Suppose that two entries of $\mathbf{z} = (z_1, \ldots, z_n)$ have different signs, say $z_i > 0$, while $z_j < 0$. By the previous lemma, $R_{x,y} > 0$ for all x, y. Therefore,

$$\begin{aligned} |z_y| &= \left| \sum_x z_x R_{x,y} \right| \\ &= \left| \sum_{x: z_x \geq 0} z_x R_{x,y} + \sum_{x: z_x < 0} z_x R_{x,y} \right|. \end{aligned}$$

The first term on the last line is strictly positive (since it is at least $z_i R_{i,y} > 0$), while the second term is strictly negative (since it is at most $z_j R_{j,y} < 0$). Hence, because of cancellations (see Exercise 1.13), the expression above is strictly smaller than the sum of the absolute values, that is,

$$|z_y| < \sum_x |z_x| R_{x,y}.$$

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

Since R is stochastic by the previous lemma, we deduce after summing over y that

$$\sum_y |z_y| < \sum_y \sum_x |z_x| R_{x,y} = \sum_x |z_x| \sum_y R_{x,y} = \sum_x |z_x|,$$

a contradiction, thereby proving the claim. ∎

Proof of Lemma 1.1.27 Now define π entrywise by

$$\pi_x = \frac{z_x}{\sum_i z_i} = \frac{|z_x|}{\sum_i |z_i|} \geq 0,$$

where the second equality comes from the previous lemma. We also used the fact that $\mathbf{z} \neq \mathbf{0}$.

For all y, by definition of $\mathbf{z}$,

$$\sum_x \pi_x P_{x,y} = \sum_x \frac{z_x}{\sum_i z_i} P_{x,y} = \frac{1}{\sum_i z_i} \sum_x z_x P_{x,y} = \frac{z_y}{\sum_i z_i} = \pi_y.$$

The same holds with $P_{x,y}$ replaced by $R_{x,y}$ from (1.1.3). Since $R_{x,y} > 0$ and $\mathbf{z} \neq \mathbf{0}$, it follows that $\pi_y > 0$ for all y. That proves the claim. ∎

That concludes the proof of the existence claim. ∎

It remains to prove uniqueness. See Exercise 1.14 for an alternative proof based on the maximum principle (to which we come back in Theorem 3.3.9 and Exercise 3.12).

Proof of Theorem 1.1.24 (ii) Suppose there are two distinct stationary distributions π_1 and π_2 (which must be strictly positive). Since they are distinct, they are not a multiple of each other and therefore are linearly independent. Apply the Gram–Schmidt procedure:

$$\mathbf{q}_1 = \frac{\pi_1}{\|\pi_1\|_2} \qquad \text{and} \qquad \mathbf{q}_2 = \frac{\pi_2 - \langle \pi_2, \mathbf{q}_1 \rangle \mathbf{q}_1}{\|\pi_2 - \langle \pi_2, \mathbf{q}_1 \rangle \mathbf{q}_1\|_2}.$$

Then

$$\mathbf{q}_1 P = \frac{\pi_1}{\|\pi_1\|_2} P = \frac{\pi_1 P}{\|\pi_1\|_2} = \frac{\pi_1}{\|\pi_1\|_2} = \mathbf{q}_1,$$

and all entries of $\mathbf{q}_1$ are strictly positive.

Similarly,

$$\begin{aligned} \mathbf{q}_2 P &= \frac{\pi_2 - \langle \pi_2, \mathbf{q}_1 \rangle \mathbf{q}_1}{\|\pi_2 - \langle \pi_2, \mathbf{q}_1 \rangle \mathbf{q}_1\|_2} P \\ &= \frac{\pi_2 P - \langle \pi_2, \mathbf{q}_1 \rangle \mathbf{q}_1 P}{\|\pi_2 - \langle \pi_2, \mathbf{q}_1 \rangle \mathbf{q}_1\|_2} \\ &= \frac{\pi_2 - \langle \pi_2, \mathbf{q}_1 \rangle \mathbf{q}_1}{\|\pi_2 - \langle \pi_2, \mathbf{q}_1 \rangle \mathbf{q}_1\|_2} \\ &= \mathbf{q}_2. \end{aligned}$$

Since $\mathbf{z} := \mathbf{q}_2$ satisfies (1.1.2), by Lemmas 1.1.25–1.1.27 there is a multiple of $\mathbf{q}_2$, say $\mathbf{q}_2' = \alpha \mathbf{q}_2$ with $\alpha \neq 0$, such that $\mathbf{q}_2' P = \mathbf{q}_2'$ and all entries of $\mathbf{q}_2'$ are strictly positive. By the Gram–Schmidt procedure,

$$\langle \mathbf{q}_1, \mathbf{q}_2' \rangle = \langle \mathbf{q}_1, \alpha \mathbf{q}_2 \rangle = \alpha \langle \mathbf{q}_1, \mathbf{q}_2 \rangle = 0.$$

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

But this is a contradiction since both vectors are strictly positive. That concludes the proof of the uniqueness claim. ■

REVERSIBLE **Reversibility** A transition matrix P is *reversible* with respect to (w.r.t.) a measure η if

$$\eta(x)P(x,y) = \eta(y)P(y,x)$$

for all $x, y \in V$. These equations are known as *detailed balance*. Here is the key observation: by summing over y and using the fact that P is stochastic, *such a measure is necessarily stationary*. (Exercise 1.12 explains the name.)

Example 1.1.28 (Random walk on $\mathbb{Z}^d$ (continued)). The measure $\eta \equiv 1$ is reversible for simple random walk on $\mathbb{L}^d$. ◀

Example 1.1.29 (Simple random walk on a graph (continued)). Let (X_t) be simple random walk on a connected graph $G = (V, E)$. Then (X_t) is reversible with respect to $\eta(v) := \delta(v)$, where recall that $\delta(v)$ is the degree of v. Indeed, for all $\{u, v\} \in E$,

$$\delta(u)P(u,v) = \delta(u)\frac{1}{\delta(u)} = 1 = \delta(v)\frac{1}{\delta(v)} = \delta(v)P(v,u).$$

(See Exercise 1.9 for the transition matrix of simple random walk on a graph.) ◀

METROPOLIS ALGORITHM **Example 1.1.30** (Metropolis chain). The Metropolis algorithm modifies an irreducible, symmetric (i.e., whose transition matrix is a symmetric matrix) chain Q to produce a new chain P with the same transition graph and a prescribed positive stationary distribution π. The idea is simple. For each pair $x \neq y$, either we multiply $Q(x,y)$ by $\pi(y)/\pi(x)$ and leave $Q(y,x)$ intact or vice versa. Detailed balance immediately follows. To ensure that the new transition matrix remains stochastic, for each pair we make the choice that lowers the transition probabilities; then we add the lost probability to the diagonal (i.e., to the probability of staying put).

Formally, the definition of the new chain is

$$P(x,y) := \begin{cases} Q(x,y)\left[\frac{\pi(y)}{\pi(x)} \wedge 1\right] & \text{if } x \neq y, \\ 1 - \sum_{z \neq x} Q(x,z)\left[\frac{\pi(z)}{\pi(x)} \wedge 1\right] & \text{otherwise.} \end{cases}$$

Note that, by definition of P and the fact that Q is stochastic, we have $P(x,y) \leq Q(x,y)$ for all $x \neq y$, so

$$\sum_{y \neq x} P(x,y) \leq 1,$$

and hence P is well defined as a transition matrix. We claim further that P is reversible with respect to π. Suppose $x \neq y$, and assume without loss of generality that $\pi(x) \geq \pi(y)$. Then, by definition of P, we have

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

$$\begin{aligned}\pi(x)P(x,y) &= \pi(x)Q(x,y)\frac{\pi(y)}{\pi(x)}\\ &= Q(x,y)\pi(y)\\ &= Q(y,x)\pi(y)\\ &= P(y,x)\pi(y),\end{aligned}$$

where we used the symmetry of Q. ◀

Convergence and mixing time A key property of Markov chains is that, under suitable assumptions, they converge to a stationary regime. We need one more definition before stating the theorem. A chain is said to be *aperiodic* if, for all $x \in V$, the greatest common divisor of $\{t : P^t(x,x) > 0\}$ is 1. APERIODIC

Example 1.1.31 (Lazy random walk on a graph). The *lazy simple random walk* on G is the Markov chain such that, at each time, it stays put with probability $1/2$ or chooses a uniformly random neighbor of the current state otherwise. Such a chain is aperiodic. ◀ LAZY

Lemma 1.1.32 (Consequence of aperiodicity). *If P is aperiodic, irreducible, and has a finite state space, then there is a positive integer t_0 such that for all $t \geq t_0$ the matrix P^t has strictly positive entries.*

We can now state the convergence theorem. For probability measures μ, ν on V, their *total variation distance* is TOTAL VARIATION DISTANCE

$$\|\mu - \nu\|_{\mathrm{TV}} := \sup_{A \subseteq V} |\mu(A) - \nu(A)|. \tag{1.1.4}$$

Theorem 1.1.33 (Convergence theorem). *Suppose P is irreducible, aperiodic, and has stationary distribution π. Then, for all x,*

$$\|P^t(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \to 0,$$

as $t \to +\infty$.

We give a proof in the finite case in Example 4.3.3. In particular, the convergence theorem implies that for all x, y,

$$P^t(x,y) \to \pi(y).$$

Without aperiodicity, we have the weaker claim

$$\frac{1}{t}\sum_{s=1}^{t} P^s(x,y) \to \pi(y), \tag{1.1.5}$$

as $t \to +\infty$.

We will be interested in quantifying the speed of convergence in Theorem 1.1.33. For this purpose, we define

$$d(t) := \sup_{x \in V} \|P^t(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}}. \tag{1.1.6}$$

Lemma 1.1.34 (Monotonicity of $d(t)$). *The function $d(t)$ is non-increasing in t.*

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

Proof Note that, by definition of P^{t+1},

$$\begin{aligned} d(t+1) &= \sup_{x\in V}\sup_{A\subseteq V}|P^{t+1}(x,A)-\pi(A)| \\ &= \sup_{x\in V}\sup_{A\subseteq V}\left|\sum_z P(x,z)(P^t(z,A)-\pi(A))\right| \\ &\le \sup_{x\in V}\sum_z P(x,z)\sup_{A\subseteq V}|P^t(z,A)-\pi(A)| \\ &\le \sup_{z\in V}\sup_{A\subseteq V}|P^t(z,A)-\pi(A)| \\ &= d(t), \end{aligned}$$

where on the second and fourth line we used that P is a stochastic matrix. ∎

The following concept will play a key role.

MIXING TIME **Definition 1.1.35** (Mixing time). *For a fixed $\varepsilon > 0$, the* mixing time *is defined as*

$$t_{\text{mix}}(\varepsilon) := \inf\{t \ge 0 : d(t) \le \varepsilon\}.$$

1.2 Some Discrete Probability Models

With the necessary background covered, we are now in a position to define formally a few important discrete probability models that will be ubiquitous in this book. These are all graph-based processes. Many more interesting random discrete structures and other related probabilistic models will be encountered throughout (and defined where needed).

Percolation Percolation processes are meant to model the movement of a fluid through a porous medium. There are several types of percolation models. We focus here on bond percolation. In words, edges of a graph are "open" at random, indicating that fluid is passing through. We are interested in the "open clusters," that is, the regions reached by the fluid.

BOND PERCOLATION **Definition 1.2.1** (Bond percolation). *Let $G = (V, E)$ be a finite or infinite graph. The* bond percolation *process on G with density $p \in [0,1]$, whose measure is denoted by $\mathbb{P}_p$, is defined as follows: each edge of G is independently set to* open *with probability p, otherwise it is set to* closed*. Write $x \Leftrightarrow y$ if $x, y \in V$ are connected by a path all of whose edges are open. The* open cluster *of x is*

$$\mathcal{C}_x := \{y \in V : x \Leftrightarrow y\}.$$

We will mostly consider bond percolation on the infinite graphs $\mathbb{L}^d$ or $\mathbb{T}_d$. The main question we will ask is: *For which values of p is there an infinite open cluster?*

Random graphs Random graphs provide a natural framework to study complex networks. Different behaviors are observed depending on the modeling choices made. Perhaps the simplest and most studied is the Erdős–Rényi random graph model. We consider the version due to Gilbert. Here the edges are present independently with a fixed probability. Despite its simplicity, this model exhibits a rich set of phenomena that make it a prime example for the use of a variety of probabilistic techniques.

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

Definition 1.2.2 (Erdős–Rényi graph model). *Let $n \in \mathbb{N}$ and $p \in [0,1]$. Set $V := [n]$. Under the* Erdős–Rényi graph model on n vertices with density p, *a random graph $G = (V, E)$ is generated as follows: for each pair $x \neq y$ in V, the edge $\{x, y\}$ is in E with probability p independently of all other edges. We write $G \sim \mathbb{G}_{n,p}$ and we denote the corresponding measure by $\mathbb{P}_{n,p}$.*

ERDŐS–RÉNYI GRAPH MODEL

Typical questions regarding the Erdős–Rényi graph model (and random graphs more generally) include: *How are degrees distributed? Is G connected? What is the (asymptotic) probability of observing a particular subgraph, for example, a triangle? What is the typical chromatic number?*

As one alternative to the Erdős–Rényi model, we will also encounter preferential attachment graphs. These are meant to model the growth of a network where new edges are more likely to be incident with vertices of high degree, a reasonable assumption in some applied settings. Such a process produces graphs with properties that differ from those of the Erdős–Rényi model; in particular, they tend to have a "fatter" degree distribution tail. In the definition of preferential attachment graphs, we restrict ourselves to the tree case (see Exercise 1.15).

Definition 1.2.3 (Preferential attachment graph). *The* preferential attachment graph process *produces a sequence of graphs $(G_t)_{t\geq 1}$ as follows: we start at time 1 with two vertices, denoted v_0 and v_1, connected by an edge. At time t, we add vertex v_t with a single edge connecting it to an old vertex, which is picked proportionally to its degree. We write $(G_t)_{t\geq 1} \sim \mathrm{PA}_1$.*

PREFERENTIAL ATTACHMENT GRAPHS

Markov random fields Another common class of graph-based processes involves the assignment of random "states" to the vertices of a *fixed graph*. The state distribution is typically specified through "interactions between neighboring vertices." Such models are widely studied in statistical physics and also have important applications in statistics. We focus on models with a Markovian (i.e., conditional independence) structure that makes them particularly amenable to rigorous analysis and computational methods. We start with Gibbs random fields, a broad class of such models.

Definition 1.2.4 (Gibbs random field). *Let S be a finite set and let $G = (V, E)$ be a finite graph. Denote by $\mathcal{K}$ the set of all cliques of G. A positive probability measure μ on $\mathcal{X} := S^V$ is called a* Gibbs random field *if there exist* clique potentials $\phi_K \colon S^K \to \mathbb{R}$, $K \in \mathcal{K}$, *such that*

GIBBS RANDOM FIELD

$$\mu(x) = \frac{1}{\mathcal{Z}} \exp\left(\sum_{K\in\mathcal{K}} \phi_K(x_K)\right),$$

where x_K is x restricted to the vertices of K and $\mathcal{Z}$ is a normalizing constant.

The following example introduces the primary Gibbs random field we will encounter.

Example 1.2.5 (Ising model). For $\beta > 0$, the *(ferromagnetic) Ising model* with inverse temperature β is the Gibbs random field with $S := \{-1, +1\}$, $\phi_{\{i,j\}}(\sigma_{\{i,j\}}) = \beta\sigma_i\sigma_j$ and $\phi_K \equiv 0$ if $|K| \neq 2$. The function $\mathcal{H}(\sigma) := -\sum_{\{i,j\}\in E} \sigma_i\sigma_j$ is known as the *Hamiltonian*. The normalizing constant $\mathcal{Z} := \mathcal{Z}(\beta)$ is called the *partition function*. The states $(\sigma_i)_{i\in V}$ are referred to as *spins*. ◀

ISING MODEL

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

Typical questions regarding Ising models include: *How fast is correlation decaying down the graph? How well can one guess the state at an unobserved vertex?* We will also consider certain Markov chains related to Ising models (see Definition 1.2.8).

Random walks on graphs and reversible Markov chains The last class of processes we focus on are random walks on graphs and their generalizations. Recall the following definition.

SIMPLE RANDOM WALK ON A GRAPH

Definition 1.2.6 (Simple random walk on a graph). *Let $G = (V, E)$ be a finite or countable, locally finite graph.* Simple random walk *on G is the Markov chain on V, started at an arbitrary vertex, which at each time picks a uniformly chosen neighbor of the current state.*

We generalize the definition by adding weights to the edges. In this context, we denote edge weights by $c(e)$ for "conductance" (see Section 3.3).

RANDOM WALK ON A NETWORK

Definition 1.2.7 (Random walk on a network). *Let $G = (V, E)$ be a finite or countably infinite graph. Let $c\colon E \to \mathbb{R}_+$ be a positive edge weight function on G. Recall that we call $\mathcal{N} = (G, c)$ a network. We assume that for all $u \in V$,*

$$c(u) := \sum_{e=\{u,v\}\in E} c(e) < +\infty.$$

Random walk on network $\mathcal{N}$ *is the Markov chain on V, started at an arbitrary vertex, which at each time picks a neighbor of the current state proportionally to the weight of the corresponding edge. That is, the transition matrix is given by*

$$P(u,v) = \begin{cases} \frac{c(\{u,v\})}{c(u)} & \text{if } \{u,v\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

By definition of P, it is immediate that this Markov chain is reversible with respect to the measure $\eta(u) := c(u)$. In fact, conversely, any countable reversible Markov chain can be seen as a random walk on a network by setting $c(e) := \pi(x)P(x,y) = \pi(y)P(y,x)$ for all x, y such that $P(x,y) > 0$.

Typical questions include: *How long does it take to visit all vertices at least once or a particular subset of vertices for the first time? How fast does the walk approach stationarity? How often does the walk return to its starting point?*

We will also encounter a particular class of Markov chains related to Ising models, the Glauber dynamics.

GLAUBER DYNAMICS

Definition 1.2.8 (Glauber dynamics of the Ising model). *Let μ_β be the Ising model with inverse temperature $\beta > 0$ on a graph $G = (V, E)$. The* (single-site) Glauber dynamics *is the Markov chain on $\mathcal{X} := \{-1, +1\}^V$, which at each time*

- *selects a site $i \in V$ uniformly at random, and*
- *updates the spin at i according to μ_β conditioned on agreeing with the current state at all sites in $V \setminus \{i\}$.*

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

Specifically, for $\gamma \in \{-1,+1\}$, $i \in V$, and $\sigma \in \mathcal{X}$, let $\sigma^{i,\gamma}$ be the configuration σ with the spin at i being set to γ. Let $n = |V|$ and $S_i(\sigma) := \sum_{j\sim i} \sigma_j$. The non-zero entries of the transition matrix are

$$Q_\beta(\sigma, \sigma^{i,\gamma}) := \frac{1}{n} \cdot \frac{e^{\gamma\beta S_i(\sigma)}}{e^{-\beta S_i(\sigma)} + e^{\beta S_i(\sigma)}}.$$

This chain is irreducible since we can flip each site one by one to go from any state to any other state. It is straightforward to check that $Q_\beta(\sigma, \sigma^{i,\gamma})$ is a stochastic matrix. The next theorem shows that μ_β is its stationary distribution.

Theorem 1.2.9 *The Glauber dynamics is reversible with respect to μ_β.*

Proof For all $\sigma \in \mathcal{X}$ and $i \in V$, let

$$S_{\neq i}(\sigma) := \mathcal{H}(\sigma^{i,+}) + S_i(\sigma) = \mathcal{H}(\sigma^{i,-}) - S_i(\sigma).$$

We have

$$\begin{aligned}
\mu_\beta(\sigma^{i,-})\, Q_\beta(\sigma^{i,-}, \sigma^{i,+}) &= \frac{e^{-\beta S_{\neq i}(\sigma)} e^{-\beta S_i(\sigma)}}{\mathcal{Z}(\beta)} \cdot \frac{e^{\beta S_i(\sigma)}}{n[e^{-\beta S_i(\sigma)} + e^{\beta S_i(\sigma)}]} \\
&= \frac{e^{-\beta S_{\neq i}(\sigma)}}{n\mathcal{Z}(\beta)[e^{-\beta S_i(\sigma)} + e^{\beta S_i(\sigma)}]} \\
&= \frac{e^{-\beta S_{\neq i}(\sigma)} e^{\beta S_i(\sigma)}}{\mathcal{Z}(\beta)} \cdot \frac{e^{-\beta S_i(\sigma)}}{n[e^{-\beta S_i(\sigma)} + e^{\beta S_i(\sigma)}]} \\
&= \mu_\beta(\sigma^{i,+})\, Q_\beta(\sigma^{i,+}, \sigma^{i,-}).
\end{aligned}$$

That concludes the proof. ∎

Exercises

Exercise 1.1 (0-norm). Show that $\|\mathbf{u}\|_0$ does not define a norm.

Exercise 1.2 (A factorial bound: one way). Let ℓ be a positive integer.

(i) Use the bound $1 + x \le e^x$ to show that

$$\frac{k+1}{k} \le e^{1/k}$$

and

$$\frac{k}{k+1} \le e^{1/(k+1)}$$

for all positive integers k.

(ii) Use part (i) and the quantity

$$\prod_{k=1}^{\ell-1} \frac{(k+1)^k}{k^k}$$

to show that

$$\ell! \ge \frac{\ell^\ell}{e^{\ell-1}}.$$

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

(iii) Use part (i) and the quantity

$$\prod_{k=1}^{\ell-1} \frac{k^{k+1}}{(k+1)^{k+1}}$$

to show that

$$\ell! \leq \frac{\ell^{\ell+1}}{e^{\ell-1}}.$$

Exercise 1.3 (A factorial bound: another way). Let ℓ be a positive integer. Show that

$$\frac{\ell^{\ell}}{e^{\ell-1}} \leq \ell! \leq \frac{\ell^{\ell+1}}{e^{\ell-1}}$$

by considering the logarithm of $\ell!$, interpreting the resulting quantity as a Riemann sum, and bounding above and below by an integral.

Exercise 1.4 (A binomial bound). Show that for integers $0 < d \leq n$,

$$\sum_{k=0}^{d} \binom{n}{k} \leq \left(\frac{en}{d}\right)^{d}.$$

(Hint: Multiply the left-hand side of the inequality by $(d/n)^d \leq (d/n)^k$ and use the binomial theorem.)

Exercise 1.5 (Powers of the adjacency matrix). Let A^n be the nth matrix power of the adjacency matrix A of a graph $G = (V, E)$. Prove that the (i,j)th entry a_{ij}^n is the number of walks of length exactly n between vertices i and j in G. (Hint: Use induction on n.)

Exercise 1.6 (Paths). Let u, v be vertices of a graph $G = (V, E)$.

(i) Show that the graph distance $d_G(u, v)$ is a metric.
(ii) Show that the binary relation $u \leftrightarrow v$ is an equivalence relation.
(iii) Prove (ii) in the directed case.

Exercise 1.7 (Trees: number of edges). Prove that a connected graph with n vertices is a tree if and only if it has $n - 1$ edges. (Hint: Proceed by induction. Then use Corollary 1.1.6.)

Exercise 1.8 (Trees: characterizations). Prove Theorem 1.1.5.

Exercise 1.9 (Simple random walk on a graph). Let $(X_t)_{t\geq 0}$ be simple random walk on a finite graph $G = (V, E)$. Suppose the vertex set is $V = [n]$. Write down an expression for the transition matrix of (X_t).

Exercise 1.10 (Communication lemma). Let (X_t) be a finite Markov chain. Show that if $x \to y$, then there is an integer $r \geq 1$ such that

$$\mathbb{P}[X_r = y \,|\, X_0 = x] = (P^r)_{x,y} > 0.$$

Exercise 1.11 (Stochastic matrices from stochastic matrices). Let

$$P^{(1)}, P^{(2)}, \ldots, P^{(r)} \in \mathbb{R}^{n\times n},$$

be stochastic matrices.

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

(i) Show that $P^{(1)}P^{(2)}$ is a stochastic matrix. That is, a product of stochastic matrices is a stochastic matrix.
(ii) Show that for any $\alpha_1, \ldots, \alpha_r \in [0,1]$ with $\sum_{i=1}^r \alpha_i = 1$,

$$\sum_{i=1}^{r} \alpha_i P^{(i)}$$

is stochastic. That is, a convex combination of stochastic matrices is a stochastic matrix.

Exercise 1.12 (Reversing time). Let (X_t) be a finite Markov chain with transition matrix P. Assume P is reversible with respect to a probability distribution π. Assume that the initial distribution is π. Show that for any sequence of states $z_0, \ldots, z_s$, the reversed sequence has the same probability, that is,

$$\mathbb{P}[X_s = z_0, \ldots, X_0 = z_s] = \mathbb{P}[X_s = z_s, \ldots, X_0 = z_0].$$

Exercise 1.13 (A strict inequality). Let $a, b \in \mathbb{R}$ with $a < 0$ and $b > 0$. Show that

$$|a + b| < |a| + |b|.$$

(Hint: Consider the cases $a + b \geq 0$ and $a + b < 0$ separately.)

Exercise 1.14 (Uniqueness: maximum principle). Let $P = (P_{i,j})_{i,j=1}^n \in \mathbb{R}^n$ be a transition matrix.

(i) Let $\alpha_1, \ldots, \alpha_m > 0$ such that $\sum_{i=1}^m \alpha_i = 1$. Let $\mathbf{x} = (x_1, \ldots, x_m) \in \mathbb{R}^n$. Show that

$$\sum_{i=1}^{m} \alpha_i x_i \leq \max_i x_i,$$

and that equality holds if and only if $x_1 = x_2 = \cdots = x_m$.
(ii) Let $\mathbf{0} \neq \mathbf{y} \in \mathbb{R}^n$ be a right eigenvector of P with eigenvalue 1, that is, $P\mathbf{y} = \mathbf{y}$. Assume that $\mathbf{y}$ is not a constant vector, that is, there is $i \neq j$ such that $y_i \neq y_j$. Let k be such that $y_k = \max_{i \in [n]} y_i$. Show that for any ℓ such that $P_{k,\ell} > 0$, we necessarily have $y_\ell = y_k$. (Hint: Use that $y_k = \sum_{i=1}^n P_{k,i} y_i$ and apply (i).)
(iii) Assume that P is irreducible. Let $\mathbf{0} \neq \mathbf{y} \in \mathbb{R}^n$ again be a right eigenvector of P with eigenvalue 1. Use (ii) to show that $\mathbf{y}$ is necessarily a constant vector.
(iv) Use (iii) to conclude that the dimension of the null space of $P^T - I$ is exactly 1. (Hint: Use the Rank–Nullity Theorem.)

Exercise 1.15 (Preferential attachment trees). Let $(G_t)_{t \geq 1} \sim \mathrm{PA}_1$ as in Definition 1.2.3. Show that G_t is a tree with $t + 1$ vertices for all $t \geq 1$.

Exercise 1.16 (Warm-up: a little calculus). Prove the following inequalities which we will encounter throughout. (Hint: Basic calculus should do.)

(i) Show that $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$.
(ii) Show that $1 - x \geq e^{-x-x^2}$ for $x \in [0, 1/2]$.
(iii) Show that $\log(1 + x) \leq x - x^2/4$ for $x \in [0, 1]$.
(iv) Show that $\log x \leq x - 1$ for all $x \in \mathbb{R}_+$.

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press

(v) Show that $\cos x \leq e^{-x^2/2}$ for $x \in [0, \pi/2)$. (Hint: Consider the function $h(x) = \log(e^{x^2/2} \cos x)$ and recall that the derivative of $\tan x$ is $1 + \tan^2 x$.)

Bibliographic Remarks

Section 1.1 For an introduction to graphs, see, for example, [Die10] or [Bol98]. Four different proofs of Cayley's formula are detailed in the delightful [AZ18]. Markov chain theory is covered in detail in [Dur10, chapter 6]. For a more gentle introduction, see, for example, [Dur12, chapter 1], [Nor98, chapter 1], or [Res92, chapter 2].

Section 1.2 For book-length treatments of percolation theory, see [BR06a, Gri10b]. The version of the Erdős–Rényi random graph model we consider here is due to Gilbert [Gil59]. For much deeper accounts of the theory of random graphs and related processes, see, for example, [Bol01, Dur06, JLR11, vdH17, FK16]. Two standard references on finite Markov chains and mixing times are [AF, LPW06].

https://doi.org/10.1017/9781009305129.002 Published online by Cambridge University Press