SC9 Probability on Graphs and Lattices

Christina Goldschmidt and Joost Jorritsma

Michaelmas Term 2025
(Version of October 13, 2025)

SC9 Probability on Graphs and Lattices

Recommended Prerequisites

Discrete and continuous time Markov processes on countable state space, as covered for example in A8 Probability and SB3.1 Applied Probability.

Aims and Objectives

The aim is to introduce fundamental probabilistic and combinatorial tools, as well as key models, in the theory of discrete disordered systems. We will examine the large-scale behaviour of systems containing many interacting components, subject to some random noise. Models of this type have a wealth of applications in statistical physics, biology and beyond, and we will see several key examples in the course. Many of the tools we will discuss are also of independent theoretical interest, and have far reaching applications. In the final part of the course we will study how long it takes for a random system to get close to equilibrium (its mixing time). This concept is also important in many statistical applications, such as studying the run time of MCMC methods.

Synopsis

- Uniform spanning trees, loop-erased random walks, Wilson's algorithm, the Aldous-Broder algorithm.
- Percolation, phase transitions in $\mathbb{Z}^d$, specific tools in $\mathbb{Z}^2$.
- Ising model, random-cluster model and other models from statistical mechanics (e.g. Potts model, hard-core model).
- Glauber dynamics, mixing times, couplings.

Reading

G. Grimmett, *Probability on Graphs: Random Processes on Graphs and Lattices*, Cambridge University Press, 2010; 2017 (2nd edition).

B. Bollobás, O. Riordan, *Percolation*, Cambridge University Press, 2006.

H. Duminil-Copin, *Introduction to Bernoulli percolation*, notes 2017. Available at `https://www.ihes.fr/~duminil/publi/2017percolation.pdf`.

H. Duminil-Copin, *Lectures on the Ising and Potts models on the hypercubic lattice*, 2017. Available at `https://arxiv.org/abs/1707.00520`.

S. Friedli, Y. Velenik, *Statistical Mechanics of Lattice Systems*, Cambridge University Press, 2018. Available at `https://www.unige.ch/math/folks/velenik/smbook/`.

D. A. Levin, Y. Peres, E. L. Wilmer, *Markov Chains and Mixing Times*, American Mathematical Society, 2009.

T. Liggett, *Continuous time Markov Processes: an introduction*, American Mathematical Society, 2010.

R. Lyons, Y. Peres, *Probability on trees and networks*, Cambridge University Press, 2016.

Introduction

The purpose of this course is to introduce you to a number of topics from modern probability which have something of a discrete flavour. They all involve random structures or random processes built from (or on top of) a fixed graph G, which may be finite or infinite. Many of the models we will study have their origins in physics, for example in the study of magnetism. However, because of their intrinsic mathematical interest, they have become some of the key touchstones of modern probability theory. Moreover, they are sufficiently general as to have applications in, for example, biology, the study of networks, and beyond. It turns out that some fascinating and highly non-trivial phenomena may occur even in relatively simple settings. The percolation model we will study in Chapter 2, for example, simply consists of keeping each edge of some underlying graph with probability p and removing it otherwise, independently for different edges. We often take the base graph to be the lattice $\mathbb{Z}^d$ for $d \geq 2$ and, in that context (as well as many others), the model undergoes a *phase transition*: there exists a critical probability p_c such that if $p < p_c$ then with probability 1 there is no infinite connected component, whereas if $p > p_c$ there is an infinite connected component, again with probability 1. More physically relevant models involve notions of the *energy* of a configuration, and correlations between what happens in different parts of the graph, but similar phenomena, including phase transitions, arise there too.

The random objects that we will study often take their values in spaces of *graphs*, and so methods for describing and "getting our hands on" their distributions are a key issue. (We no longer have recourse to anything as simple as a cumulative distribution function!) This issue turns out to be intimately related to the question of how to *generate* (or, indeed, *simulate*) our random objects.

Some other unifying questions are as follows:

- What can we learn about a graph by studying random structures defined on that graph?
- How effectively can we study random structures on infinite graphs via large finite approximations? Does the randomness 'feel the effect of the boundary'?
- When behaviours on different parts of the graph are not independent, are they positively or negatively correlated? What mathematical tools can be deployed in each case?
- How does the "shape" of the graph determine the speed of random processes on the graph?

Along the way, we will encounter concepts and techniques which enable you to develop your "probabilistic toolkit" for use more generally.

We very much hope you enjoy the course!

Acknowledgements

Previous versions of this course have been given by Paul Chleboun, Dominic Yeo, Alessandra Caraceni, James Martin, and Brett Kolesnik. We are very grateful to them for their notes, on which these are based. Any errors are our responsibility. Please send comments or corrections to `christina.goldschmidt@stats.ox.ac.uk` for the first half of the course and to `joost.jorritsma@stats.ox.ac.uk` for the second half of the course.

Some definitions and notation

We will assume very little graph theory beyond the following definitions.

Definition 0.1. A *graph* G consists of a set of vertices $V = V(G)$, and a set of edges $E = E(G) \subseteq \binom{V(G)}{2}$, the set of pairs of distinct vertices.

(In particular, in this course, graphs will be *simple*, so include no *self-loops* nor *multiple edges*.)

- A *rooted graph* (G, ρ) is a graph G, together with a choice of *root* $\rho \in V(G)$.
- Graph H is a *subgraph* of graph G when $V(H) \subseteq V(G)$, and $E(H) \subseteq E(G)$.
- Graph H is an *induced subgraph* of graph G when $V(H) \subseteq V(G)$, and $E(H)$ consists of precisely those edges in $E(G)$ which are incident to two vertices in $V(H)$.
- Graph G is *connected* if any pair of distinct vertices $x, y \in V(G)$ may be joined by a path made up of edges from $E(G)$. Otherwise, G has at least two *connected components*.
- The degree $\deg_G(v)$ of a vertex $v \in V(G)$ is the number of edges incident to v. A graph G is *locally finite* if all the degrees are finite.
- For vertices $v, w \in V(G)$, we will write $v \sim w$ when v and w are neighbours, that is, when $\{v, w\} \in E$.

We observe that $\sum_{v \in V(G)} \deg_G(v) = 2|E(G)|$, since each edge has two incident vertices.

Example. *Useful graphs which will appear frequently, especially as subgraphs or induced subgraphs, include:*

- *The complete graph K_n on n vertices, where every edge is present. Also called a* clique.
- *The empty graph E_n on n vertices, with no edges. (This is the* complement *of K_n.)*
- *The path P_n on n vertices (or of length $n-1$).*
- *The cycle C_n on n vertices.*

*We will often work on the d-*dimensional integer lattice*, that is, the graph $\mathbb{Z}^d$ with nearest-neighbour edges $\big\{\{x, y\} : ||x - y||_1 = 1\big\}$. We will make use of large finite approximations:*

- *The d-dimensional* box $[0, n]^d \cap \mathbb{Z}^d$ *or* $[-n, n]^d \cap \mathbb{Z}^d$*;*
- *The d-dimensional* torus *consisting of the points in $[0, n]^d \cap \mathbb{Z}^d$ but where vertices $\mathbf{x} = (x_1, \ldots, x_d)$ and $\mathbf{x}' = (x'_1, \ldots, x'_d)$ are identified when there exists $1 \leq j \leq d$ such that $x_j = 0$ and $x'_j = n$ and $x_i = x'_i$ for all $i \neq j$. Informally, we glue together "opposite edges" of the box.*

Definition 0.2. A graph T is a *tree* if it is connected, and contains no cycles. A *forest* is a graph each of whose connected components is a tree. For a connected graph G, a *spanning tree* of G is a subgraph T on vertex set $V(T) = V(G)$, which is a tree. (Note that if G is not connected then it has no spanning trees, but it does have a *spanning forest*.)

We observe the standard fact that a tree T with n vertices has exactly $n - 1$ edges.

(There are many standard texts on graph theory; for example, *Modern Graph Theory* by Bollobás is available online via the Bodleian Library.)

Chapter 1

Uniform spanning trees

1.1 Background: spanning trees and random walks

Definition 1.1. Let $\mathcal{T}$ be the set of all spanning trees of a finite connected graph G. Then, we call T a *uniform spanning tree* (UST) of G if it is distributed according to the uniform measure on $\mathcal{T}$. That is, for each $t \in \mathcal{T}$, we have $\mathbb{P}(T = t) = \frac{1}{|\mathcal{T}|}$.

Perhaps the simplest case is when $G = K_n$.

Theorem 1.2 (Cayley's formula)**.** *There are n^{n-2} labelled trees with vertex set $[n] := \{1, 2, \ldots, n\}$. In other words, there are n^{n-2} spanning trees of the complete graph K_n.*

You can see a simulation of the UST of K_n for a large n in Figure 1.1.

Except for very unusual (in particular, very *sparse*) graphs, the number of spanning trees will be large, and in almost all cases, it will be infeasible to compute the number precisely. In light of this, a key question is how can we generate/simulate a UST on a large graph. Moreover, how can we do so as efficiently as possible? The question of whether we can generate/simulate a random structure is often intimately related to the question of whether we can analyse it or understand its properties. We will also be interested in how can one extend this framework to infinite graphs, where $|\mathcal{T}|$ will normally be infinite, and so picking an element uniformly at random no longer makes sense.

An important role in answering these questions turns out to be played by random walks. We recall some definitions from Part A Probability and Part B Applied Probability.

Definition 1.3. The simple random walk (SRW) $(X_n)_{n\geq 0}$ on a locally finite graph G is the Markov chain with state space $V(G)$ and transition probabilities $p_{vw} = \frac{1}{\deg_G(v)}$ when $\{v, w\} \in E(G)$, and $p_{vw} = 0$ otherwise. In words, the chain moves to its next state by choosing uniformly from among the neighbours of its current state, independently of its history.

Let τ_v be the *hitting time* of a vertex $v \in V(G)$,

$$\tau_v := \inf\{n \geq 0 : X_n = v\}$$

and τ_v^+ be the *first return time* to v,

$$\tau_v^+ = \inf\{n \geq 1 : X_n = v\}.$$

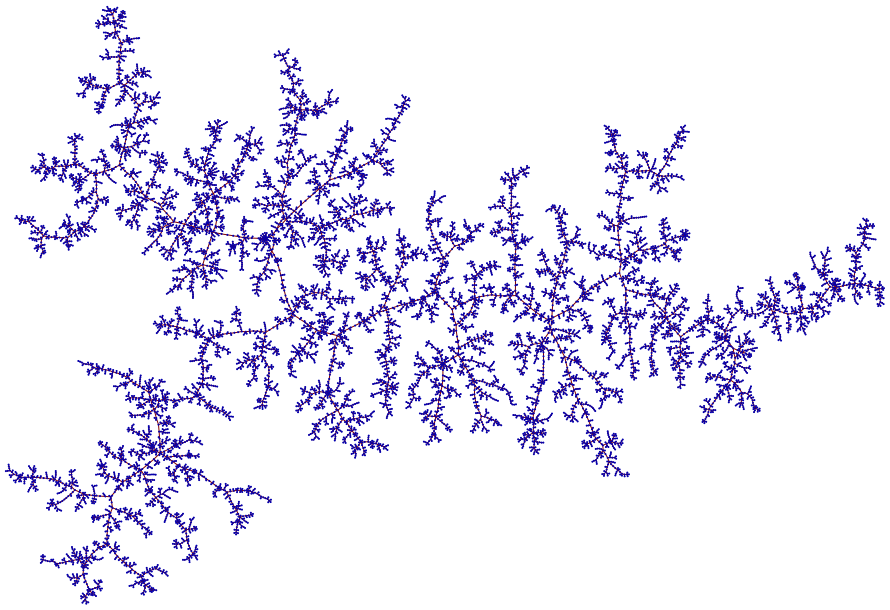


Figure 1.1: A UST of the complete graph on 27179 vertices. Picture by Igor Kortchemski.

Note that τ_v and τ_v^+ are stopping times. The *cover time* t_{cov} is the time (if any) at which the walk has first visited *every* vertex:

$$t_{\text{cov}} := \sup_{v \in V(G)} \tau_v = \inf \Big\{ n \geq 0 \,:\, \{X_0, X_1, \ldots, X_n\} = V(G) \Big\}.$$

Recall that a graph G is *recurrent* if the simple random walk started from any vertex $v \in V(G)$ visits v infinitely many times, almost surely. Equivalently, the RW returns to v almost surely. On the other hand, an infinite graph G is *transient* if simple random walk started from v returns to v with probability strictly less than 1, and so visits v finitely many times, almost surely.

When G is connected and recurrent (including when G is finite), all the hitting times τ_v are finite almost surely. When G is infinite, the cover time is infinite almost surely, although it remains valid when G is recurrent to say 'the random walk visits every site almost surely'. Note that the distributions of the hitting times and the cover time will, in general, depend on the initial state X_0 of the walk, which may be random or deterministic.

Recall that a Markov chain is *reversible* if, in equilibrium, the chain run forwards in time and the chain run backwards in time are distributionally indistinguishable (which is equivalent to the *detailed balance equations* having a solution).

Proposition 1.4. *The SRW on a finite connected graph G is* reversible *and has stationary distribution satisfying $\pi_v \propto \deg_G(v)$ for $v \in V(G)$.*

Proof. Since G is finite and connected the SRW is irreducible and positive recurrent. The non-trivial detailed balance equations are

$$\pi_v \frac{1}{\deg_G(v)} = \pi_w \frac{1}{\deg_G(w)}$$

for $v \sim w$. So any measure proportional to vertex-degree provides a solution and is also stationary. It

follows that the unique stationary distribution is

$$\pi_v = \frac{\deg_G(v)}{2|E(G)|}, \quad v \in V(G),$$

and that the chain is reversible in equilibrium. □

If $(X_n)_{n\geq 0}$ is a reversible Markov chain, it is sometimes useful to have a stationary version where time is indexed by all of $\mathbb{Z}$ (we call this an *eternal stationary version* of the Markov chain). We can build such a process $(X_n)_{n\in\mathbb{Z}}$ by taking copies $(X'_n)_{n\geq 0}$ and $(X''_n)_{n\geq 0}$ of the Markov chain which are conditionally independent given their starting points, and with $X'_0 = X''_0$ distributed according to the stationary distribution. Then set

$$X_n = \begin{cases} X'_n & \text{if } n \geq 0 \\ X''_{-n} & \text{if } n < 0. \end{cases}$$

Note that the time 0 does not actually play any special role here: X_n has the invariant distribution for all $n \in \mathbb{Z}$, so we could have done the same construction from any other fixed time.

1.2 Generating uniform spanning trees

We will describe two algorithms for generating a UST on a (for now, finite) graph G.

- The *Aldous–Broder algorithm* (1989), based on a random walk on G.
- *Wilson's algorithm* (1996), based on a series of *loop-erased random walks* on G.

1.2.1 The Aldous–Broder algorithm

Our first algorithm, which was introduced roughly simultaneously by David Aldous and Andrei Broder, builds a UST from a SRW on G.

Definition 1.5 (Aldous–Broder algorithm)**.** For a connected, recurrent graph G, run a SRW on G from any initial state $X_0 \in V(G)$. Let T be the subgraph of G with $V(T) = V(G)$ and, for each vertex apart from X_0, including the edge along we first reach that vertex, i.e.

$$E(T) := \Big\{ \{X_{\tau_v - 1}, X_{\tau_v}\} \,:\, v \in V(G)\backslash\{X_0\} \Big\}. \tag{1.1}$$

Theorem 1.6 (Aldous (1990), Broder (1989))**.** *For G finite, the random subgraph T generated by the Aldous–Broder algorithm is a UST on G.*

Proof. (non-examinable) Since the graph is finite and connected, the cover time is finite with probability 1. It is an exercise on Problem Sheet 1 to show that T is indeed a spanning tree of G. It remains to prove that it is uniform.

Now let $\mathcal{R}$ be the set of rooted spanning trees of G. Consider an eternal stationary version of the SRW on G, $(X_n)_{n\in\mathbb{Z}}$, and for $k \in \mathbb{Z}$, let (T_k, X_k) be the rooted tree obtained by running the Aldous–Broder algorithm using the random walk from time k onwards. We claim that $(T_n, X_n)_{n\in\mathbb{Z}}$ is an irreducible stationary Markov chain with (finite) state-space $\mathcal{R}$. To see that the evolution is Markovian, note first that T_k is a measurable functional of $X_k, X_{k+1}, X_{k+2}, \ldots$. Moreover, starting at a later time-point can only mean that first hitting times occur *later*. So, given (T_k, X_k), we have that (T_{k+1}, X_{k+1}) is conditionally independent of $(T_{k-1}, X_{k-1}), (T_{k-2}, X_{k-2}), \ldots$.

To see that the chain is irreducible, first fix any pair of rooted trees (t, ρ) and (t', ρ'). Now observe that there exists a possible trajectory $S(t, \rho)$ of the SRW from ρ to a vertex v of t such that, along this trajectory, the Aldous–Broder algorithm produces the rooted tree (t, ρ). Similarly, along some possible trajectory $S(t', \rho')$ of the SRW starting from ρ', the Aldous–Broder algorithm produces the rooted tree (t', ρ'). Since G is connected, there exists a path $P(v, \rho')$ from v to ρ'. Now if the SRW follows the concatenation of the three trajectories $S(t, \rho)$, $P(v, \rho')$ and $S(t', \rho')$, the tree-chain will transition from (t, ρ) to (t', ρ'). Since the concatenated trajectory occurs with strictly positive probability, the chain is irreducible. Finally, if the SRW is stationary, the rooted tree must also be.

We now investigate the stationary distribution for (T, X) on $\mathcal{R}$. It turns out to be more straightforward to do this by considering the *time-reversal* of the chain $(T_n, X_n)_{n\in\mathbb{Z}}$, which is necessarily a Markov chain with the same stationary distribution. For $(t, \rho), (t', \rho') \in \mathcal{R}$, let

$$q((t, \rho), (t', \rho')) = \mathbb{P}\left(T_{-1} = t', X_{-1} = \rho' | T_0 = t, X_0 = \rho\right),$$

the transition probabilities for the time-reversal. The advantage of this backward-in-time perspective is that we learn nothing more from the rooted tree (T_0, X_0) about the value of X_{-1} than we do from knowing X_0 alone. In particular, for a fixed (t, ρ), given that $T_0 = t, X_0 = \rho$, there are $\deg_G(\rho)$ possible different values for (T_{-1}, X_{-1}), which are all equally likely i.e. $q((t, \rho), (t', \rho')) = 1/\deg_G(\rho)$ for $\deg_G(\rho)$ values of (t', ρ') and $q((t, \rho), (t', \rho')) = 0$ for the rest.

For fixed (t', ρ'), on the other hand, there are $\deg_G(\rho')$ possible rooted trees (t, ρ) from which we could have made a (reversed-time) transition with positive probability (which is then necessarily equal to $1/\deg_G(\rho)$ by the previous observation). It follows that for $(t', \rho') \in \mathcal{R}$,

$$\sum_{(t,\rho)\in\mathcal{R}} q((t, \rho), (t', \rho'))\deg_G(\rho) = \deg_G(\rho'),$$

and so the unique stationary distribution $(\eta_{(t,\rho)} : (t, \rho) \in \mathcal{R})$ must be such that

$$\eta_{(t,\rho)} \propto \deg_G(\rho).$$

It follows that $\mathbb{P}\left(T_0 = t | X_0 = \rho\right)$ is independent of the choice of t and so the *unrooted* tree T_0 is uniform on the set of all unrooted spanning trees of G. (Indeed, the tree is *independent* of the root X_0.) □

1.2.2 Wilson's algorithm

In the Aldous–Broder algorithm, we could not directly use the path taken by the random walk, because this would have contained cycles. The next algorithm is based instead on the *loop-erased random walk* on G, which is constructed by removing the cycles from the trajectory of a random walk in the order in which they appear. As we shall discuss later, each algorithm has its advantages and its disadvantages.

Definition 1.7. Let $\gamma = (x_0, x_1, \ldots, x_n) \subseteq V(G)$ be a finite, possibly self-intersecting path in a graph G. For every $v \in \{x_0, x_1, \ldots, x_n\}$, define $\ell_v := \max\{0 \le m \le n : x_m = v\}$, the index of the *last* visit to v.

Now, set $v_0 = x_0$, and then inductively define $v_{i+1} = x_{\ell_{v_i}+1}$ for $i \ge 1$, terminating at v_k for which $v_k = x_n$. In other words, v_{i+1} is the next vertex visited after visiting v_i for the last time. The path $(v_0, v_1, \ldots, v_k)$ does not self-intersect, and is called the *loop-erasure* of γ.

Note that this definition applies equally to an *infinite* path $(x_0, x_1, x_2, \ldots)$, provided no vertex $v \in V(G)$ occurs infinitely many times in the path.

Definition 1.8. Let G be a graph. The loop-erasure of a finite-length path of a simple random walk on G, or an infinite path if G is transient, is called *loop-erased random walk* (LERW).

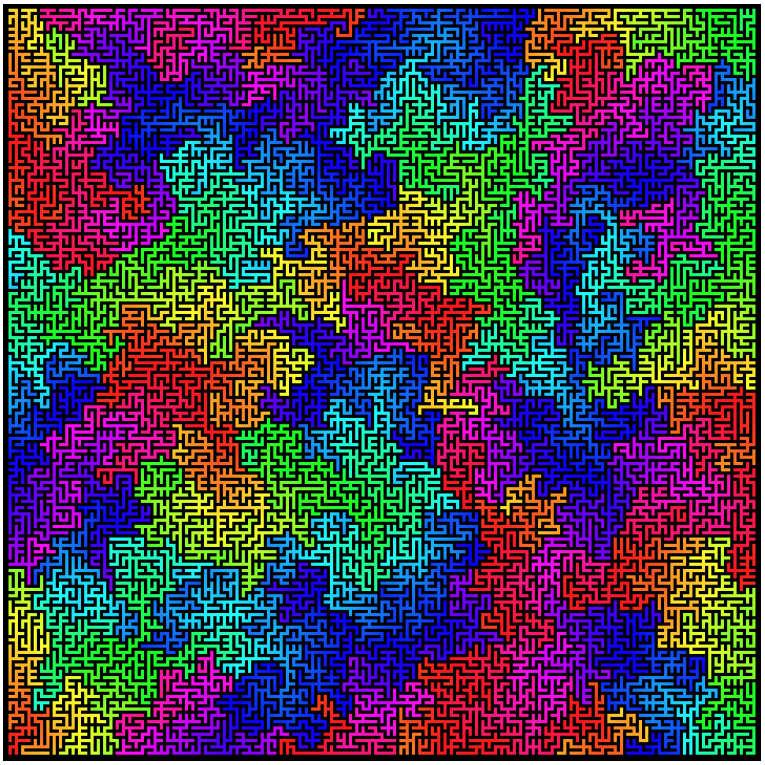


Figure 1.2: A UST on a box in $\mathbb{Z}^2$, simulated using Wilson's algorithm (JavaScript implementation at `https://bl.ocks.org/mbostock/c03ee31334ee89abad83`). The colours represent the ℓ^1 distance from the starting vertices of the branches.

Remark. *A word of caution. It can be tempting to think of LERW as a* process*, like the random walk itself. This is normally not appropriate. For example, as a random process with the same time-indexing as the underlying random walk, LERW does not live in $V(G)$, but instead in the space of paths on G.*

Definition 1.9 (Wilson's algorithm)**.** Given a finite connected graph G with $|V(G)| = n$, let $(v_0, v_1, \dots, v_{n-1})$ be an enumeration of the vertices. We define a sequence of random trees $T_0 \subset T_1 \subset T_2 \subset \cdots$ as follows:

- Let T_0 be the tree with a single vertex $\{v_0\}$.
- When $i \geq 0$, and the vertex set of T_i is a strict subset of $V(G)$, independently of what has happened before, define T_{i+1} as follows:
  - Let j be the smallest label in $\{1, 2, \dots, n-1\}$ such that $v_j \notin T_i$.
  - Let τ_{T_i} be the hitting time of T_i for a random walk $(X_0, X_1, \dots)$ on G with initial vertex $X_0 = v_j$, and let $\gamma = (X_0, \dots, X_{\tau_{T_i}})$.
  - Define the edge set of T_{i+1} to be the union of the edge set of T_i and the loop-erasure of γ. Note that T_{i+1} is also a tree.
- When T_i spans $V(G)$, declare $T = T_i$.

For a simulation of Wilson's algorithm on a box in $\mathbb{Z}^2$, see `https://bl.ocks.org/mbostock/11357811`, and for the end result see Figure 1.2.

Theorem 1.10 (Wilson (1996))**.** *Given a finite connected graph G, the random spanning tree T given by Wilson's algorithm is a UST.*

We start by introducing a "naive" algorithm as a building block along the way to a proof of Wilson's algorithm. Given a finite connected graph G with n vertices, and an identified vertex $v_0 \in V(G)$, by a *spanning tree directed towards* v_0 we mean a collection of edges which form a spanning tree, where each edge is given the (unique) orientation towards v_0 in this tree.

Now, for every vertex v_i in $V(G)\backslash\{v_0\}$, select an edge $\overrightarrow{v_i w_i}$ directed away from v_i.

Claim: such a collection of directed edges is either a spanning tree directed towards v_0, or includes a directed cycle.

The proof of this claim is an exercise on Problem Sheet 1. It motivates the following use of rejection sampling.

Naive algorithm: For each vertex $v_i \in V(G)\backslash\{v_0\}$ independently, select an edge $\overrightarrow{v_i w_i}$ uniformly from those incident to v_i.

- If the resulting collection of directed edges is a spanning tree directed towards v_0, stop, and return the undirected version of this spanning tree.
- If the resulting collection of directed edges includes a directed cycle, start again.

Each time we do the sampling step, a particular spanning tree t occurs with probability

$$\prod_{v_i \in V(G)\backslash\{v_0\}} \frac{1}{\deg_G(v_i)}.$$

Since this is equal for all spanning trees t, the naive algorithm generates a UST on G.

Wilson's algorithm may be viewed as adapting the naive algorithm above so that resampling the directed edges happens only at those vertices which are part of a problematic directed cycle.

Proof of Theorem 1.10. We begin by defining a larger probability space, rich enough to support several realisations of LERW.

We are given a vertex $v_0 \in V(G)$. For every $v_i \in V(G)\backslash\{v_0\}$, let $(w_i^{(1)}, w_i^{(2)}, \ldots)$ be any sequence of choices from among the neighbours of v_i in G. We construct a walk $(X_0, X_1, \ldots)$ on G, *absorbed at* v_0, as a function of this information, as follows.

Suppose that we are given $X_0, X_1, \ldots, X_k$ with $X_k = v$, and that $\#\{0 \le i \le k : X_i = v\} = \ell$ (i.e. time k is the time of the ℓth visit to v). If $v = v_0$, simply set $X_{k+1} = v_0$. Otherwise, if $v = v_i \neq v_0$, set $X_{k+1} = w_i^{(\ell)}$. In other words, at the ℓth visit to $v_i \neq v_0$, the random walk moves along the directed edge $\overrightarrow{v_i w_i^{(\ell)}}$ to vertex $w_i^{(\ell)}$.

Similarly, one may construct a walk on G absorbed at any collection of vertices, and also its loop-erasure, so long as the path is finite.

We think of the collection $(w_i^{(\ell)})$ as a *stack of instructions* left at vertex v_i, with $w_i^{(1)}$ initially at the top of the stack. In the following construction, we will iteratively discard the top element of some stacks, so the top-most (or *visible*) instruction on some stacks changes from $w_i^{(\ell)}$ to $w_i^{(\ell+1)}$.

The visible instructions always define a set of directed edges, just as in the claim and naive algorithm above. When this includes at least one directed cycle, we may try to remove one of these directed cycles, by discarding the visible instruction from every stack corresponding to a vertex in the cycle. The collection of visible instructions thus changes, and maybe further directed cycles may be removed in subsequent steps.

This process, which is referred to as *cycle popping*, may be iterated until the visible instructions induce no directed cycle. Let N be the number of cycles 'popped' in the procedure. Either $N = \infty$, or $N < \infty$, and in the latter case (by the earlier claim) what remains is a spanning tree t directed towards v_0.

Obviously, we have to choose some order in which to pop the cycles. It is not clear *a priori* whether t and N will necessarily be independent of this choice of ordering. Fortunately, this turns out to be true.

Lemma 1.11 (Cycle-popping lemma)**.** *For any choice of* $(w_i^{(k)}, i = 1, \dots, n-1, k \geq 1)$*, for any order of cycle-popping,* N *and* t *are the same. That is:*

- *either* $N = \infty$ *for all orderings of cycle-popping;*
- *or* $N < \infty$ *is the same for all orderings of cycle-popping, and all such orderings give the same directed spanning tree* t*.*

Proof. Note that this is a *deterministic* result, not a probabilistic one. The proof is left as an exercise here, or alternatively see Section 2.2 of Grimmett's *Probability on Graphs*. □

Now, we will consider the situation where, independently for each $v_i \neq v_0$, $(w_i^{(1)}, w_i^{(2)}, \dots)$ is a sequence of i.i.d. uniform picks from among the neighbours of v_i. Note then, that we can use this (random) information to perform Wilson's algorithm (where for LERW, cycle-popping happens in 'chronological' order), and finite termination of Wilson's algorithm at a tree t exactly corresponds to cycle-popping with $N < \infty$ and t directed towards $\{v_0\}$.

It is clear by construction (of Wilson's algorithm) that in this setting, N is almost surely finite. So it remains to show that t is a UST on G. Each cycle which is popped consists of a collection of directed edges of the form $\overrightarrow{v_i w_i^{(k)}}$, where k indexes its position in the stack. Let $\mathcal{C}_1, \dots, \mathcal{C}_M$ be some collection of these indexed directed cycles, which can be popped in this order (which is certainly independent of whatever is left underneath). Note that, for example, if $w_1^{(1)} = w_1^{(2)} = v_2$ and $w_2^{(1)} = w_2^{(3)} = v_1$, we view the 'cycles' $\{v_1 w_1^{(1)}, v_2 w_2^{(1)}\}$ and $\{v_1 w_1^{(2)}, v_2 w_2^{(3)}\}$ as distinct.

Then the probability that the random stacks of instructions $(w_i^{(k)})$ give these indexed, directed cycles, and leave the spanning tree t directed towards $\{v_0\}$ as the visible instructions underneath the popped cycles, is precisely

$$\prod_{m=1}^{M} \prod_{v_i w_i^{(k)} \in \mathcal{C}_m} \frac{1}{\deg_G(v_i)} \times \prod_{v_i \in V(G) \setminus \{v_0\}} \frac{1}{\deg_G(v_i)}, \tag{1.2}$$

where the second term arises for exactly the same reasons as in the naive algorithm discussed earlier.

Note that (1.2) factorises as a product of a term depending on $(\mathcal{C}_1, \dots, \mathcal{C}_M)$ and a term depending on t (in fact, it's a constant). So, in fact, we have the stronger result that *conditional* on any $(\mathcal{C}_1, \dots, \mathcal{C}_M)$, t has the uniform distribution on spanning trees. The required statement follows immediately by the law of total probability, where we partition according to the sequence of cycles. □

The proof shows that we may choose which vertex to start the 'next LERW' from in Wilson's algorithm with considerable flexibility.

Remark. *The time taken to run the Aldous–Broder algorithm is, of course, proportional to the cover time of the graph. Let us write* $\mathbb{P}_v$ *for the probability measure governing the SRW started from a fixed vertex* v*, and* $\mathbb{E}_v$ *for the corresponding expectation. On Problem Sheet 1, we show that* $\max_{v \in V(G)} \mathbb{E}_v[t_{cov}]$ *is bounded above by* n^3 *for* any *graph, although in many cases it is much smaller. It turns out that the expected time taken to run Wilson's algorithm, on the other hand, is bounded above by a constant times* $\sum_{v,w \in V(G)} \pi_v \pi_w \mathbb{E}_v[\tau_w]$*, which is always smaller than* $\max_{v \in V(G)} \mathbb{E}_v[t_{cov}]$*. So Wilson's algorithm is more efficient, but at the cost of being somewhat more complicated to implement.*

The construction used in this proof of the validity of Wilson's algorithm is an example of the *coupling method*. This is a crucial and recurring tool in discrete probability theory, where one compares two or more random structures by constructing them jointly on the same probability space.

Formally, given random variables X_1 and X_2 defined on probability spaces $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$, respectively, a *coupling* of X_1, X_2 is a pair of random variables (Y_1, Y_2) defined on some new probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for which $Y_1 \stackrel{d}{=} X_1$ and $Y_2 \stackrel{d}{=} X_2$.

You have probably seen coupling proofs in earlier probability courses (although they may have been non-examinable!). For example, the standard proof of convergence to equilibrium for an irreducible aperiodic positive-recurrent discrete-time Markov chain makes use of a coupling between a copy of the Markov chain which starts from a particular vertex and another which starts in equilibrium.

We will see a number of examples of couplings throughout the course.

1.3 Negative association

Since a spanning tree of a graph G must include a fixed number of edges, it is intuitively reasonable to imagine that the events that the edges e and e' are present in a UST T will be negatively correlated. Similarly, if we have fewer edges from which to assemble a spanning tree, then any given edge e is more likely to appear in a UST.

We make these statements precise in the following result.

Proposition 1.12. *Given a finite, connected graph G, let T be a UST on G.*

(a) For any pair e, e' of distinct edges in $E(G)$,

$$\mathbb{P}\left(e \in E(T) \,\middle|\, e' \in E(T)\right) \leq \mathbb{P}\left(e \in E(T)\right). \tag{1.3}$$

More generally, if A and B are disjoint, acyclic collections of edges in $E(G)$, then

$$\mathbb{P}\left(B \subset E(T) \,\middle|\, A \subset E(T)\right) \leq \mathbb{P}\left(B \subset E(T)\right). \tag{1.4}$$

(b) Let H be a connected subgraph of G, and let T_H be a UST on H. Then, for any collection A of edges in $E(H)$,

$$\mathbb{P}\left(A \subset E(T_H)\right) \geq \mathbb{P}\left(A \subset E(T)\right). \tag{1.5}$$

(Note that if A contains a cycle, then both probabilities in (1.5) are zero.)

Sketch proof. We shall give a sketch proof of (a) in the case where e and e' share no vertices. Let us write $e = \{u, v\}$ and consider the Aldous–Broder algorithm started from the vertex u. Recall that τ_v is the first hitting time of v and that τ_u^+ is the first return time to u. The edge e is included in $E(T)$ if and only if e is the first entry edge into v. This either occurs immediately, or requires the SRW to return to u before hitting v, so that by the Strong Markov property we get

$$\mathbb{P}\left(e \in E(T)\right) = \frac{1}{\deg_G(u)} + \mathbb{P}_u\left(\tau_u^+ < \tau_v\right)\mathbb{P}\left(e \in E(T)\right).$$

On rearranging (note that $\mathbb{P}_u(\tau_u^+ = \tau_v) = 0$) we obtain that

$$\mathbb{P}\left(e \in E(T)\right) = \frac{1}{\deg_G(u)\mathbb{P}_u(\tau_v < \tau_u^+)}.$$

Consider now the graph G/e' obtained by removing e' from the graph and identifying its two end-points into a single vertex. (This is the *contraction of G along e'* and, in general, may be a multigraph. There

is, however, no difficulty in extending everything we have done so far to that setting.) Now observe that spanning trees of G which contain the edge e' are in bijection with spanning trees of G/e'. Hence, if we write $\tilde{T}$ for a UST of G/e',

$$\mathbb{P}\left(e \in E(T) | e' \in E(T)\right) = \mathbb{P}\left(e \in E(\tilde{T})\right).$$

Since we have assumed that e and e' do not share any vertices, $\deg_{G/e'}(u) = \deg_G(u)$ and so what we need to show is that

$$\mathbb{P}_u^G(\tau_v < \tau_u^+) \leq \mathbb{P}_u^{G/e'}(\tau_v < \tau_u^+)$$

i.e. that contracting along e' can only *increase* the probability that the random walk started from u hits v before returning to u. (The superscripts on the probability measure indicate the graph under consideration.) This can be shown directly via somewhat involved Markov chain methods (see Section 1.4.3 of *Random walks and electric networks* by Doyle and Snell; we will investigate some special cases in Problem Sheet 1). However, it is most elegantly proved using a deep connection between reversible Markov chains and electrical networks on graphs. Briefly, we can interpret the quantity $\frac{1}{\deg_G(u)\mathbb{P}_u(\tau_v < \tau_u^+)}$ as the "effective resistance" between u and v if we treat the edges of the graph as unit resistors and connect a 1-volt battery between u and v. The theory of electrical networks originated with Kirchhoff in the 1840s, and provides a very powerful method for proving results about reversible Markov chains. See Doyle and Snell (pdf file linked from the course Canvas page) for a very accessible account. See Grimmett's *Probability on Graphs* for a briefer account of electrical networks, and a proof of this Proposition. □

Remark 1.13. *Negative dependence results are often very hard, and several intuitively 'obvious' results are in fact famous open problems. The theory of positive dependence, on the other hand, is much better understood, and will be introduced properly in the next chapter.*

Proposition 1.12 has some striking consequences, as we will see in the next section.

1.4 Infinite graphs

When G has a countably infinite vertex set, it is not clear what the analogue of the UST will be. While the definition of *spanning tree* extends immediately, one cannot make a uniform choice directly from the set of spanning trees since this set will be infinite, except in special cases. We could consider using the Aldous–Broder algorithm directly on G. However, this will only generate a *spanning* tree with probability one if G is recurrent. Alternatively, we could consider using Wilson's algorithm on a transient graph G, if we allow the possibility to run the LERW paths for infinitely many steps (if they don't hit the currently-established tree within finitely many steps). However, the resulting subgraph will not necessarily be connected.

Definition 1.14. Let G be an infinite connected graph. An *exhaustion* of G is a sequence (G_n) of finite connected graphs which is induced by an increasing sequence $V_1 \subset V_2 \subset \cdots \subset \bigcup_{n\geq 1} V_n = V(G)$ of subsets of the vertices.

Example 1.15.
- *The d-dimensional boxes $[-n, n]^d \cap \mathbb{Z}^d$ for $n \geq 1$ form an exhaustion of $\mathbb{Z}^d$.*
- *Consider the infinite rooted binary tree, in which the root has degree 2 and every other vertex 3. The depth-n rooted binary trees for $n \geq 1$ form an exhaustion of the infinite tree.*

Motivation: Given an infinite graph G, and an exhaustion (G_n), let T_n be the UST on G_n. By (1.5), for any *finite* collection of edges $A \subset E(G)$, we have

$$\mathbb{P}\left(A \subset E(T_n)\right) \geq \mathbb{P}\left(A \subset E(T_{n+1})\right), \tag{1.6}$$

whenever n is large enough that $A \subset E(G_n)$. We might therefore conjecture the existence of an infinite random 'tree' T on G for which

$$\mathbb{P}\left(A \subset E(T)\right) = \lim_{n\to\infty} \mathbb{P}\left(A \subset E(T_n)\right),$$

since this limit exists by monotonicity.

Before we can make this precise, we need to introduce some formalism, and some notions from the theory of weak convergence. (The standard reference, which you might like to refer to if you want to know more, is Billingsley's *Convergence of Probability Measures*.)

Let us suppose that G is a graph with a countably infinite edge-set $E(G) = \{e_1, e_2, \ldots\}$, where the labelling is arbitrary. We will work in the space $\Omega_G := \{0,1\}^{E(G)}$, whose elements can be thought of as possible edge-sets of subgraphs of G: 1 corresponds to the presence of the edge, and 0 to its absence. As is common, we view the elements of Ω_G as functions $\omega : E(G) \to \{0,1\}$. So for $\omega \in \Omega_G$, we will write $\omega(e) = 1$ to mean the edge $e \in E(G)$ is present in ω and $\omega(e) = 0$ otherwise.

A topological aside. *The space Ω_G may be viewed as a countable product of copies of the discrete topological space $\{0,1\}$, and so is compact by Tychonoff's theorem. It is metrisable by*

$$d(\omega, \omega') = \sum_{i \geq 1} 2^{-i} |\omega(e_i) - \omega'(e_i)|.$$

A subset $\mathcal{C}$ of Ω_G is called a *cylinder event* or *cylinder set* if for some finite subset $F \subset E(G)$ and some $\omega' \in \{0,1\}^F$, we have

$$\mathcal{C} = \Big\{ \omega \in \Omega_G \,:\, \omega|_F = \omega' \Big\}. \tag{1.7}$$

Note that this may be written equivalently as

$$\Big\{ \omega \in \Omega_G \,:\, \omega(e) = 1,\, e \in A;\, \omega(e) = 0,\, e \in B \Big\},$$

for some finite disjoint subsets $A, B \subset E(G)$. In words, a cylinder event says that a particular finite collection of edges is present, and a distinct particular finite collection of edges is not present.

Cylinder events are both open and closed in the product topology. The *product σ-algebra* $\mathcal{F}_G$ is defined to be the σ-algebra generated by these cylinder events. The *Borel σ-algebra* $\mathcal{B}_G$ is defined to be the minimal σ-algebra containing all the open sets. Because of the particular nature of the open sets here, we in fact have $\mathcal{F}_G = \mathcal{B}_G$.

An *increasing* cylinder event has the form

$$\mathcal{C}_A := \Big\{ \omega \in \Omega_G \,:\, \omega(e) = 1,\, e \in A \Big\}, \tag{1.8}$$

for some finite set $A \subset E(G)$. In words, an increasing cylinder event says that a particular finite collection of edges is present. It is called increasing because if $\omega \leq \omega'$ in Ω_G, and $\omega \in \mathcal{C}_A$, then $\omega' \in \mathcal{C}_A$ i.e. opening additional edges cannot cause $\mathcal{C}_A$ to fail.

Definition 1.16. Let $(\mu_n)_{n \geq 1}$ and μ be probability measures on $(\Omega_G, \mathcal{B}_G)$. We say that μ_n *converges weakly* to μ, written $\mu_n \Rightarrow \mu$, if

$$\mu(\mathcal{C}) = \lim_{n \to \infty} \mu_n(\mathcal{C}), \quad \text{for all cylinder events } \mathcal{C}. \tag{1.9}$$

Furthermore, if the limits in (1.9) all exist, then a limit probability measure μ on $(\Omega_G, \mathcal{B}_G)$ exists.

Remark. *Weak convergence is usually defined in terms of the convergence of the expectations of bounded continuous functionals. The* Portmanteau theorem *(see, for example, Theorem 2.1 of Billingsley's* Convergence of Probability Measures*) gives various equivalent definitions. In our setting, the cylinder events (which generate the Borel σ-algebra) are both open and closed, and we do not need to have separate versions of* (1.9) *for open and closed events in the generating class, as we would normally in the Portmanteau theorem.*

Next, we observe that it suffices to restrict our attention to increasing cylinder events.

Proposition 1.17. *Let (μ_n) be a sequence of probability measures on $(\Omega_G, \mathcal{B}_G)$ such that $\lim_{n\to\infty} \mu_n(\mathcal{C})$ exists for all* increasing *cylinder events $\mathcal{C}$. Then $\lim_{n\to\infty} \mu_n(\mathcal{C})$ exists for* all *cylinder events $\mathcal{C}$.*

Proof. The result follows from the principle of inclusion-exclusion and the observation that any cylinder set can be expressed, via unions and set differences, in terms only of increasing cylinder sets. This is perhaps best illustrated with an example. Let a, b, c, d, f be distinct edges in $E(G)$, and let

$$\mathcal{C} = \Big\{\omega \in \Omega_G \,:\, \omega(a) = \omega(b) = \omega(c) = 1,\, \omega(d) = \omega(f) = 0\Big\}.$$

Then, for any measure μ on Ω_G, we have

$$\mu(\mathcal{C}) = \mu\left(\mathcal{C}_{\{a,b,c\}}\right) - \mu\left(\mathcal{C}_{\{a,b,c,d\}}\right) - \mu\left(\mathcal{C}_{\{a,b,c,f\}}\right) + \mu\left(\mathcal{C}_{\{a,b,c,d,f\}}\right).$$

The case for a general cylinder set $\mathcal{C}$ is similar, and implies the required result. □

1.5 The free and wired uniform spanning forests

1.5.1 The free uniform spanning forest as a weak limit

For now, fix G, a countably infinite, locally-finite connected graph and an exhaustion (G_n) of G by finite graphs, for which T_n is the UST on G_n. We view T_n as a random element of Ω_G, and let μ_{T_n} be its distribution, so that $\mathbb{P}(T_n = \omega) = \mu_{T_n}(\omega)$ for $\omega \in \Omega_G$ (and note that this can only be non-zero for ω such that $\omega|_{E(G)\setminus E(G_n)} = 0$). The negative association result (1.6) asserts that $\mu_{T_{n+1}}(\mathcal{C}_A) \le \mu_{T_n}(\mathcal{C}_A)$, where $\mathcal{C}_A$ is the increasing cylinder set defined by some finite subset $A \subset E(G)$. Therefore $\lim_{n\to\infty} \mu_{T_n}(\mathcal{C}_A)$ exists for all such increasing cylinder sets and so, by Proposition 1.17, there is a probability measure μ^F on Ω_G such that $\mu_{T_n} \Rightarrow \mu^F$.

Proposition 1.18. *The measure μ^F does not depend on the choice of exhaustion (G_n) of G.*

Proof. Consider two sequences $V_1 \subset V_2 \subset \cdots \subset \bigcup_{n\ge1} V_n = V(G)$ and $V_1' \subset V_2' \subset \cdots \subset \bigcup_{n\ge1} V_n' = V(G)$ of subsets of the vertices, inducing exhaustions $(\bar{G}_n)$ and (G_n') of G.

Note that vertex set $V_n \cap V_n'$ does not necessarily induce a *connected* graph on G. We may assume WLOG (by shifting all the indices) that $V_1 \cap V_1'$ is non-empty, so choose some $v_1 \in V_1 \cap V_1'$. Then, define $\bar{G}_n$ to be the connected component of the induced graph on $V_n \cap V_n'$ which contains v_1, with $\bar{V}_n$ the corresponding vertex set. Then $(\bar{G}_n)$ is also an exhaustion of G. However, each $\bar{G}_n$ is certainly a subgraph of G_n, and so, taking $\bar{T}_n$ to be a UST on $\bar{G}_n$, we have, again by (1.5) that $\mu_{\bar{T}_n}(\mathcal{C}_A) \ge \mu_{T_n}(\mathcal{C}_A)$ whenever n is large enough that $A \subset E(\bar{G}_n)$. Similarly, writing T_n' for the UST of G_n', we have $\mu_{\bar{T}_n}(\mathcal{C}_A) \ge \mu_{T_n'}(\mathcal{C}_A)$ whenever n is large enough that $A \subset E(\bar{G}_n)$

Now, since $\bigcup_{n\ge1} \bar{V}_n = V(G)$, there exists an increasing sequence (α_n) of integers such that $V_n \cup V_n' \subset \bar{V}_{\alpha_n}$ for all n. For such a sequence, G_n and G_n' are both subgraphs of $\bar{G}_{\alpha_n}$ and so again we may apply (1.5) to conclude that $\mu_{T_n}(\mathcal{C}_A) \ge \mu_{\bar{T}_{\alpha_n}}(\mathcal{C}_A)$ and $\mu_{T_n'}(\mathcal{C}_A) \ge \mu_{\bar{T}_{\alpha_n}}(\mathcal{C}_A)$, whenever n is large enough that $A \subset E(G_n) \cap E(G_n')$. Combining these inequalities, we have

$$\mu_{\bar{T}_{\alpha_n}}(\mathcal{C}_A) \le \mu_{T_n}(\mathcal{C}_A) \le \mu_{\bar{T}_n}(\mathcal{C}_A)$$

and, similarly,

$$\mu_{\bar{T}_{\alpha_n}}(\mathcal{C}_A) \le \mu_{T_n'}(\mathcal{C}_A) \le \mu_{\bar{T}_n}(\mathcal{C}_A).$$

Since $\lim_{n\to\infty} \mu_{\bar{T}_n}(\mathcal{C}_A)$ exists and is equal to $\lim_{n\to\infty} \mu_{\bar{T}_{\alpha_n}}(\mathcal{C}_A)$, we obtain

$$\lim_{n\to\infty} \mu_{T_n}(\mathcal{C}_A) = \lim_{n\to\infty} \mu_{T_n'}(\mathcal{C}_A),$$

by sandwiching. So the measure μ^F on Ω_G, which is determined by its values on increasing cylinder sets $\mathcal{C}_A$, is the same for both exhaustions (G_n) and (G_n'), as required. □

We call a Ω_G-valued random variable with distribution μ^F the *free uniform spanning forest* (FUSF) on G. We have changed terminology from tree to forest because, as we will prove below, the property of having no cycles passes to the limit, but the property of being connected may not.

1.5.2 Properties of the free uniform spanning forest

Proposition 1.19. *Almost surely,*

(i) the FUSF contains no cycles;

(ii) the FUSF has no finite components.

Proof. Let (G_n) be some exhaustion of G, with associated USTs (T_n).

(i) Let A be the edge set of a (finite) cycle in G. Since T_n is a tree, $\mu_{T_n}(\mathcal{C}_A) = 0$, and so $\mu^F(\mathcal{C}_A) = 0$ also. The result follows by countable additivity.

(ii) See Problem Sheet 1. □

Definition 1.20. A *graph automorphism* of G is a bijection $\varphi : V(G) \to V(G)$ such that $\{\varphi(v), \varphi(w)\} \in E(G)$ iff $\{v, w\} \in E(G)$. So (abusing notation) we may extend $\varphi : E(G) \to E(G)$, which then induces a map $\varphi : \Omega_G \to \Omega_G$ as $e \in \varphi(\omega)$ iff $\varphi^{-1}(e) \in \omega$.

Example 1.21.

- *Translations and reflections in $\mathbb{Z}^d$ are graph automorphisms.*
- *The infinite rooted binary tree has many automorphisms, by exchanging the two subtrees below a given vertex. However, all automorphisms must preserve the root since this has degree 2, while all other vertices have degree 3.*

Definition 1.22. A measure μ on Ω_G is *invariant* under a graph automorphism φ if $\mu(\mathcal{C}_A) = \mu(\mathcal{C}_{\varphi(A)})$ for every finite $A \subset E(G)$.

Proposition 1.23. *The distribution of the FUSF is invariant under any graph automorphism φ of G.*

Proof. We may use Proposition 1.18. Given a sequence $V_1 \subset V_2 \subset \cdots$ of vertex sets inducing an exhaustion (G_n) of G, the sequence $\varphi(V_1) \subset \varphi(V_2) \subset \cdots$ also induces an exhaustion. Note that φ induces a bijection between the set of spanning trees of G_n and the set of spanning trees of $\varphi(G_n)$. To avoid confusion, we refer to the USTs on G_n and $\varphi(G_n)$ as T_{G_n} and $T_{\varphi(G_n)}$, respectively. For any finite subset $A \subset E(G)$, we have $\mu_{T_{\varphi(G_n)}}(\mathcal{C}_{\varphi(A)}) = \mu_{T_{G_n}}(\mathcal{C}_A)$. Hence,

$$\mu^F\left(\mathcal{C}_{\varphi(A)}\right) = \lim_{n\to\infty} \mu_{T_{\varphi(G_n)}}\left(\mathcal{C}_{\varphi(A)}\right) = \lim_{n\to\infty} \mu_{T_{G_n}}\left(\mathcal{C}_A\right) = \mu^F(\mathcal{C}_A),$$

as required. □

1.5.3 The wired uniform spanning forest

Given an (infinite) graph G and an exhaustion (G_n), define the wired subgraph G_n^W with *wired boundary* from G_n as follows. Let $\partial G_n \subset V_n$ be the set of *boundary* vertices of V_n, namely those vertices of V_n which are connected by an edge (of G) to $V(G)\backslash V_n$. We construct G_n^W from G_n by identifying all vertices of $V(G)\backslash V_n$ to a single vertex w_n (and removing any self-loops this creates). Note that if $v \in \partial G_n$ has k edges to $V(G)\backslash V_n$ then there are k edges from v to w_n in G_n^W.

Let T_n^W be a UST of G_n^W, and let $\mu_{T_n}^W$ be the distribution of T_n^W, *viewed as a random element of* Ω_G (by ignoring the edges between ∂G_n and w_n).

Proposition 1.24. *With the conditions above, and $A \subset E(G)$ finite, we have*

$$\mu^W_{T_n}(\mathcal{C}_A) \leq \mu^W_{T_{n+1}}(\mathcal{C}_A). \tag{1.10}$$

(Note that the inequality is reversed compared with the free case.) We omit the proof, but some special cases are discussed on Problem Sheet 1.

Proposition 1.25. *The limit $\lim_{n\to\infty} \mu^W_{T_n}(\mathcal{C}_A)$ exists on all increasing cylinder events $\mathcal{C}_A$, and furthermore*

$$\lim_{n\to\infty} \mu^W_{T_n}(\mathcal{C}_A) \leq \mu^F(\mathcal{C}_A). \tag{1.11}$$

Proof. The sequence $\mu^W_{T_n}(\mathcal{C}_A)$ is *increasing*, by (1.10). It is bounded above by 1, and so possesses a limit. Since in general G_n is a subgraph of G^W_n, we also have that $\mu^W_{T_n}(\mathcal{C}_A) \leq \mu_{T_n}(\mathcal{C}_A)$ by (1.5). Hence, $\lim_{n\to\infty} \mu^W_{T_n}(\mathcal{C}_A) \leq \lim_{n\to\infty} \mu_{T_n}(\mathcal{C}_A) = \mu^F(\mathcal{C}_A)$, and the result follows. □

As before, it follows that $\lim_{n\to\infty} \mu^W_{T_n}(\mathcal{C})$ exists for all cylinder sets $\mathcal{C}$, and so there is a measure μ^W on Ω_G such that $\mu^W_{T_n} \Rightarrow \mu^W$. One may also prove a version of Proposition 1.18, stating that μ^W does not depend on the choice of exhaustion (G_n) of G.

An Ω_G-valued random variable with distribution μ^W is called the *wired uniform spanning forest* (WUSF) on G. For similar reasons to the FUSF, the WUSF almost surely has no cycles, nor finite components, and is invariant under graph automorphisms of G.

1.5.4 Comparing the FUSF and the WUSF

A few natural questions:

- For which graphs G do the FUSF and WUSF have the same law?
- In Proposition 1.19, we showed that the FUSF has no cycles almost surely, and thus is a forest. But for what G is the FUSF connected (with some probability) and thus actually a spanning tree?
- If we extend the Aldous–Broder algorithm or Wilson's algorithm to infinite graphs G, how does this relate to the FUSF and WUSF?

We will give at least partial answers to all these questions.

1.5.5 Recurrent graphs

When G is infinite and recurrent, both Wilson's algorithm and the Aldous–Broder algorithm are well-defined with hardly any adjustments required.

- The Aldous–Broder algorithm on recurrent G: we now need to run the random walk for infinitely many steps (since the cover time is not finite). As before, let the edge set of T be

$$E(T) := \Big\{ \{X_{\tau_v - 1}, X_{\tau_v}\} \,:\, v \in V(G)\backslash\{X_0\} \Big\}.$$

- Wilson's algorithm on recurrent G: as before, construct a sequence of trees $T_0 \subset T_1 \subset \ldots$, where T_{i+1} is obtained from T_i by adding the edges of the loop-erasure of random walk started from a new vertex and stopped at τ_{T_i}. This process will not terminate when G is infinite and recurrent, so declare $T := \bigcup_{i\geq 1} T_i$.

Proposition 1.26. *Let G be a* recurrent*, infinite, connected graph. Then the FUSF and WUSF have the same law, and each is connected almost surely. Furthermore, they are generated by Wilson's algorithm and by the Aldous–Broder algorithm on G.*

In this case, we may refer to a *uniform spanning tree* or *UST* of G. A key example is the two-dimensional lattice $\mathbb{Z}^2$. The UST on $\mathbb{Z}^2$, in particular, is a much-studied object, and some of its properties are investigated on Problem Sheet 1.

Proof. Because the underlying random walk is recurrent, each loop-erased random walk hits the tree that has already been built in a finite number of steps with probability 1. So Wilson's algorithm for G generates a spanning tree T (in particular, T is connected) with probability 1. We will show that we can view Wilson's algorithm on G as a 'limit' of Wilson's algorithm on G_n or G_n^W, in a sense we shall make precise via a coupling argument.

Let (G_n) be an exhaustion of G, where G_n has vertex set V_n. Recall that ∂G_n denotes the boundary vertices in V_n. Let $A \subset E(G)$ be finite, with $\mathcal{C}_A$ the corresponding increasing cylinder event. Let $\mathcal{A} \subset V(G)$ be the finite set of vertices incident to A and, for convenience, assume that $\mathcal{A}$ forms an initial segment of the ordered list $(v_0, v_1, v_2, \ldots)$ of vertices which is used to drive Wilson's algorithm. We will assume that n is large enough that $\mathcal{A} \subset V_n$.

Now, run Wilson's algorithm on G. Note that we need to start at most $|\mathcal{A}| - 1$ new LERWs in order to determine T restricted to $\mathcal{A}$: we require exactly $|\mathcal{A}| - 1$ LERWs if we need to start a new one from each of $v_1, v_2, \ldots, v_{|\mathcal{A}|-1}$. Note also that we can run Wilson's algorithm on G_n using the *same* LERWs, and the partial subtrees generated will be the same *until* $\tau_{\partial G_n}$, the first time at which one of the SRWs (generating the LERWs) hits ∂G_n.

Suppose we now generate independent SRWs on G starting from each of $v_1, v_2, \ldots, v_{|\mathcal{A}|-1}$, each stopped when it hits v_0. Certainly this contains all of the random walk steps that we need to run Wilson's algorithm, since v_0 lies in every partial subtree. (In general this will be much more information than we need.) Suppose now that none of these random walks hits ∂G_n. Then certainly we will have that the restrictions of T and T_n to $\mathcal{A}$ will be the same. Thinking of the complementary events, it follows that if the restrictions of T and T_n to $\mathcal{A}$ are different, we must have that one of the random walks hits ∂G_n before v_0. Recall that the symmetric difference $B \triangle C$ of two events is the event that *exactly* one happens. We can therefore bound as follows:

$$\begin{aligned}
|\mathbb{P}(A \subset E(T)) - \mathbb{P}(A \subset E(T_n))| &\leq \mathbb{P}(\{A \subset E(T)\} \triangle \{A \subset E(T_n)\}) \\
&\leq \mathbb{P}(T \text{ restricted to } \mathcal{A} \text{ is not built before } \tau_{\partial G_n}) \\
&\leq \mathbb{P}\left(\bigcup_{i=1}^{|\mathcal{A}|-1} \{\text{SRW from } v_i \text{ hits } \partial G_n \text{ before } v_0\} \right) \\
&\leq \sum_{i=1}^{|\mathcal{A}|-1} \mathbb{P}(\text{SRW from } v_i \text{ hits } \partial G_n \text{ before } v_0),
\end{aligned}$$

by the union bound. Each of the probabilities in this sum vanishes when $n \to \infty$, since G is recurrent, and so

$$\lim_{n \to \infty} |\mathbb{P}(A \subset E(T)) - \mathbb{P}(A \subset E(T_n))| = 0. \tag{1.12}$$

Crucially, exactly the same argument holds for the wired graphs G_n^W, and so we find

$$\mu^F(\mathcal{C}_A) = \mu^W(\mathcal{C}_A) = \mathbb{P}(A \subset E(T)).$$

Since cylinder events characterise a probability measure on Ω_G, we have shown that FUSF and WUSF have the same law and are generated by Wilson's algorithm, hence are almost surely connected.

It is an exercise on Problem Sheet 1 to check that the Aldous–Broder algorithm can be handled by a similar coupling. □

1.5.6 Transient graphs

Example 1.27. *The rooted infinite binary tree $T_\infty^{(2)}$ is a transient graph. It is exhausted by the finite-depth rooted binary trees $(T_n^{(2)})$.*

- *The only uniform spanning tree of $T_n^{(2)}$ is $T_n^{(2)}$ itself, so the FUSF of $T_\infty^{(2)}$ is just $T_\infty^{(2)}$.*
- *On the other hand, the WUSF has infinitely many tree components almost surely. (This is an exercise on Problem Sheet 1.)*

So WUSF $\neq$ FUSF in this case.

Definition 1.28. We extend Wilson's algorithm to *transient* graphs G with vertices enumerated as $(v_0, v_1, v_2, \ldots)$.

- Let T_0 be the loop-erasure of a random walk path started from v_0. (Since G is transient, we may loop-erase the infinite SRW path almost surely, since it visits each vertex only finitely many times.)
- When $i \geq 0$, and the vertex set of T_i is a strict subset of $V(G)$, define T_{i+1} as follows (independently from what has happened before):
  - Let j be the smallest label in $\{1, 2, \ldots\}$ such that $v_j \notin T_i$.
  - Let τ_{T_i} be the hitting time of T_i for a random walk $(X_0, X_1, \ldots)$ on G with initial vertex $X_0 = v_j$. Note that since G is transient, τ_{T_i} may be infinite with positive probability. When $\tau_{T_i} < \infty$, we let $\gamma = (X_0, \ldots, X_{\tau_{T_i}})$, and when $\tau_{T_i} = \infty$, we let $\gamma = (X_0, X_1, \ldots)$.
  - Define the edge set of T_{i+1} to be the union of the edge set of T_i and the loop-erasure of γ.
- Let T be the union $\bigcup_{i \geq 0} T_i$.

This algorithm is sometimes referred to as *Wilson's algorithm rooted at infinity.*

Proposition 1.29. *On a transient graph G, Wilson's algorithm rooted at infinity generates the WUSF on G.*

Proof. The proof is non-examinable, and we give only a sketch. We exploit a similar coupling argument to the proof of Proposition 1.26. The rough idea of the proof is that when n is large enough, a finite collection of LERWs either (pairwise) meet inside G_n, or never meet at all. So we can couple Wilson's algorithm on G rooted at infinity with Wilson's algorithm on G_n^W started at the wiring vertex w_n. □

Remark. *Proposition 1.29 gives a sufficient condition for connectivity of the WUSF on a transient graph G. In particular, if a SRW started from any $v \neq v_0$ almost surely intersects an independent infinite LERW started from v_0 then the WUSF is almost surely connected.*

A theorem due to Benjamini, Lyons, Peres, Schramm (2001) shows that, in fact, whether two independent SRWs intersect almost surely precisely determines whether the WUSF is connected almost surely. This remains a very active research topic.

Chapter 2

Percolation

Much of the language and setting of our second topic come from the physical sciences: percolation originated as a model for the flow of a fluid through a porous medium. The fluid *percolates* if it is able to pass completely through the medium.

In contrast to the previous section, here we study the behaviour on infinite graphs directly, rather than as a limit of finite models. In the percolation models we study, each edge is present independently, and when there is dependence between events, it is generally *positive* dependence, rather than the negative dependence we saw in the context of uniform spanning trees. These mathematically-tractable properties will allow us to address a broader range of more advanced questions, especially about the existence and structure of infinite components.

2.1 Background: definitions and phase transition

We will generally work with an infinite, locally finite connected graph G (recall that locally finite means that all vertex degrees are finite). By far the most important example for this course will be the *d-dimensional integer lattice* $G = \mathbb{Z}^d$, for $d \geq 2$.

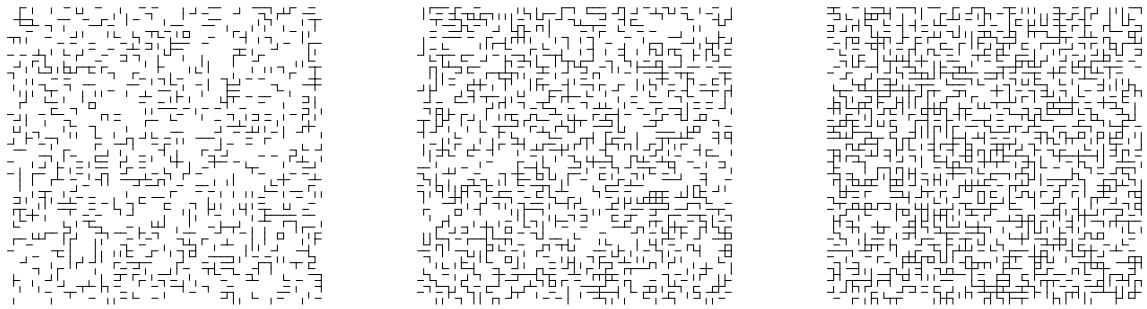
As in Chapter 1, we will work with $\Omega_G := \{0,1\}^{E(G)}$, the set of all possible collections of edges. For historical reasons stemming from the model's origins and applications, we use slightly different terminology to describe $\omega \in \Omega_G$:

- if $\omega(e) = 1$, we say edge e is *open* (i.e. present);
- if $\omega(e) = 0$, we say edge e is *closed* (i.e. absent).

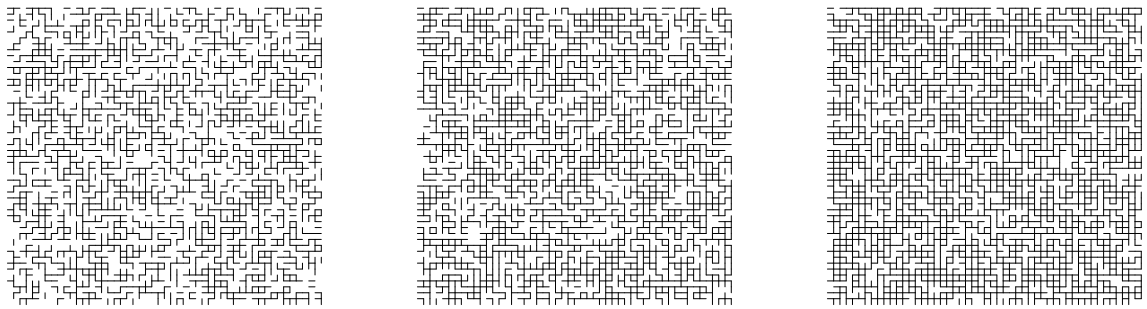
Percolation concerns the structure of the graph consisting of the open edges (see Figure 2.1). We denote by $\mathbb{P}_p$ the product (probability) measure on Ω_G with probability p. That is, every edge $e \in E(G)$ is *open* independently with probability p and closed otherwise. There are variant models in which, for example, the *vertices* are open or closed, rather than edges (sometimes called *bonds*); the version we are studying is known as *bond percolation*.

We will aim to clarify and answer some of the following questions:

- Does the structure of the collection of open edges look very different for different values of $p \in [0,1]$?
- Under what circumstances are there infinite components, and how many can there be?



$p = 0.2$ $p = 0.3$ $p = 0.4$



$p = 0.5$ $p = 0.6$ $p = 0.7$

Figure 2.1: Bond percolation on $\mathbb{Z}^2$ at different values of p.

- How much dependence is there between the components containing vertices which are far apart?
- When they exist, what do the large (or infinite) components look like?

Definition 2.1. With respect to an element $\omega \in \Omega_G$, for $x, y \in V(G)$, we write $x \leftrightarrow y$ if there is a path of open edges joining x and y. For every $x \in V(G)$, the *cluster* of x is defined to be

$$\mathcal{C}(x) = \{y \in V(G) : x \leftrightarrow y\}.$$

(Note that this is just the component containing x in subgraph consisting of precisely the open edges.)

When $|\mathcal{C}(x)| = \infty$, we will sometimes write $x \leftrightarrow \infty$ as an abbreviation. This event will sometimes be referred to as '*percolation*' or that '*x percolates*'.

Definition 2.2. A graph G is *vertex-transitive* if $\forall x, y \in V(G)$, there exists an automorphism φ of G such that $\varphi(x) = y$.

Informally, a graph is *vertex-transitive* if "it looks the same from all vertices".

Example 2.3. *The lattice $\mathbb{Z}^d$ is vertex-transitive, for example by considering a translation. Note that in fact there are four automorphisms of $\mathbb{Z}^2$ mapping any* edge *e to any other edge e', and further symmetries in higher dimensions.*

The infinite rooted binary tree $\mathbb{T}_\infty^{(2)}$ is not *vertex-transitive since the degree of the root is different to the degree of all other vertices.*

In particular, when G is vertex-transitive, we have $|\mathcal{C}(x)| \stackrel{d}{=} |\mathcal{C}(y)|$ under $\mathbb{P}_p$ for all $x, y \in V(G)$ (and, indeed, more strongly, $\mathcal{C}(x) \stackrel{d}{=} \mathcal{C}(y)$ in any appropriate space of locally finite graphs considered up to isomorphism).

ASSUMPTION: Unless specified, we will now assume $G = \mathbb{Z}^d$.

We will focus on the study of $\mathcal{C}(0)$.

Definition 2.4. The *percolation probability* is

$$\theta(p) := \mathbb{P}_p(|\mathcal{C}(0)| = \infty) = \mathbb{P}_p(0 \leftrightarrow \infty).$$

It is clear that $\theta(0) = 0$ and $\theta(1) = 1$, but the behaviour of θ on $(0, 1)$ is not obvious. Intuitively, it is reasonable to assume that when p gets larger, the graph will include more edges in some sense, and so $\theta(p)$ will be larger. We make this precise in the following lemma.

Lemma 2.5. *The percolation probability $\theta(p)$ is weakly increasing as a function of p.*

Proof. We use a coupling of $(\mathbb{P}_p : p \in [0, 1])$ on a single probability space with probability measure $\mathbb{P}$ as follows. Define random variables $(U(e), e \in E(G))$ on this probability space which are independent and identically distributed with $U[0, 1]$ distribution. For every $p \in [0, 1]$, define a random element ω_p of Ω_G by

$$\omega_p(e) = \mathbf{1}_{\{U(e) \leq p\}}.$$

So $\mathbb{P}(\omega_p(e) = 1) = p$, independently over $e \in E(G)$. By construction, whenever $p \leq q$, we have $\omega_p(e) \leq \omega_q(e)$ for every $e \in E(G)$, $\mathbb{P}$-a.s. (Note, this is really just the infinite product over $E(G)$ of a coupling of Bernoulli(p) and Bernoulli(q) random variables.)

Note that

$$\left\{0 \overset{\omega_p}{\leftrightarrow} \infty\right\} \subset \left\{0 \overset{\omega_q}{\leftrightarrow} \infty\right\},$$

since adding more open edges can only increase the size of the cluster of 0. Hence,

$$\theta(p) = \mathbb{P}_p(0 \leftrightarrow \infty) = \mathbb{P}\left(0 \overset{\omega_p}{\leftrightarrow} \infty\right) \leq \mathbb{P}\left(0 \overset{\omega_q}{\leftrightarrow} \infty\right) = \mathbb{P}_q(0 \leftrightarrow \infty) = \theta(q). \qquad \square$$

In view of Lemma 2.5, there must at some point be a transition from $\theta(p) = 0$ to $\theta(p) > 0$.

Definition 2.6. The *critical probability* p_c (or sometimes $p_c(d)$ when we wish to specify that we work on $\mathbb{Z}^d$) is

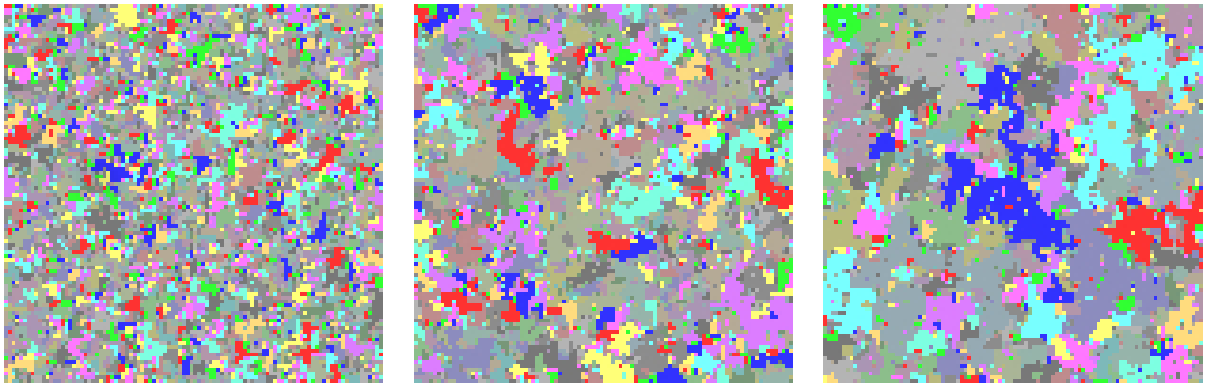
$$p_c := \sup\{p \in [0, 1] : \theta(p) = 0\}. \tag{2.1}$$

It is an exercise on Problem Sheet 2 to show that $p_c(1) = 1$. The following proposition, which we will prove in the next section, shows that for $d \geq 2$, on the other hand, $p_c(d)$ lies in the interior of the interval $[0, 1]$.

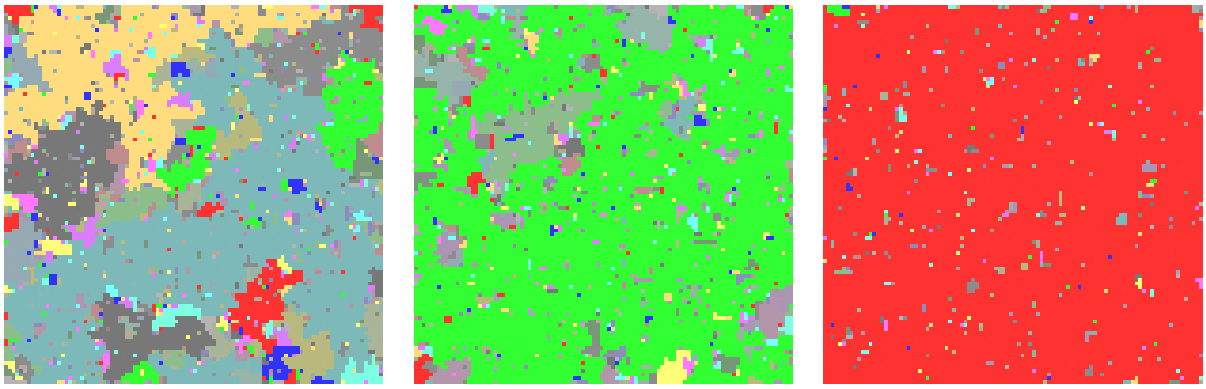
Proposition 2.7. *For each $d \geq 2$, there exist $p, p' \in (0, 1)$ with $p < p'$ such that $\theta(p) = 0$ and $\theta(p') > 0$. It follows that $p_c(d) \in (0, 1)$.*

It is possible to show that $\theta(\cdot)$ is continuous away from p_c and is *right-continuous* at p_c, that is $\theta(p_c) = \lim_{p \downarrow p_c} \theta(p)$. It is, in fact, conjectured that for $\mathbb{Z}^d$ with $d \geq 2$, one should have *continuity of the phase transition*, that is, $\theta(p_c(d)) = 0$.

As we will see in this course, in two dimensions, we have a duality relation leading to $p_c(2) = \frac{1}{2}$ and $\theta(\frac{1}{2}) = 0$. This is illustrated in Figure 2.2. The so-called *Lace expansion* has been used to prove $\theta(p_c(d)) = 0$ for all 'large' $d \geq 11$. (The most recent improvement of this result occurred as recently as 2017!) Continuity of the percolation phase transition remains a **major open problem** for $d = 3, 4, \ldots, 10$, and is an active research area.



$p = 0.3$ $p = 0.4$ $p = 0.46$



$p = 0.5$ $p = 0.52$ $p = 0.6$

Figure 2.2: The percolation clusters in $\mathbb{Z}^2$ for different values of p.

2.2 Proofs: self-avoiding walks and a Peierls argument

In the proof of Proposition 2.7, we will make essential use of the planar nature of $\mathbb{Z}^2$. Let G be any planar graph (that is, a graph which may be embedded in the plane without any of its edges crossing). We imagine drawing it in the plane, so that every edge is incident to one or two *faces*. The *dual* of G is obtained by putting a vertex in every face of G (including the infinite face if there is one) and then putting an edge between two dual vertices whenever the corresponding faces share an edge in the primal graph. An important example is the dual graph of $\mathbb{Z}^2$, which we can think of as having vertex set $(\mathbb{Z} + \frac{1}{2})^2$, and nearest-neighbour edges. (This is clearly isomorphic to $\mathbb{Z}^2$.) For the purposes of this course, we refer to this new graph as $\mathbb{Z}^2_{\text{dual}}$.

Proof of Proposition 2.7. First, we prove that there exists $p \in (0,1)$ such that $\theta(p) = 0$. The proof proceeds by showing that when p is small enough, the probability of the origin being part of a long path decays exponentially in the length of the path.

A *self-avoiding walk* (SAW) on a graph is a vertex-disjoint nearest-neighbour path. Let σ_n be the number of SAWs on $\mathbb{Z}^d$ with length n, starting from the origin. Since there are exactly $2d$ choices for the 'first step', and then, iteratively, at most $2d-1$ choices for each subsequent step, we have $\sigma_n \leq 2d(2d-1)^{n-1}$. (This is a very crude bound, but it will suffice for our purposes. Finding exact asymptotics for SAWs of length n is a very hard problem!)

Now, let N_n be the number of SAWs of length n, starting from the origin, all of whose edges are open. Since there are exactly n edges in such a path, and each edge is open independently, we have $\mathbb{E}_p[N_n] = p^n \sigma_n$. Certainly, if $0 \leftrightarrow \infty$, then one of these SAWs of length n is open, so

$$\theta(p) = \mathbb{P}_p(0 \leftrightarrow \infty) \leq \mathbb{P}_p(N_n \geq 1) \leq \mathbb{E}_p[N_n] = p^n \sigma_n, \tag{2.2}$$

using Markov's inequality. If $p < \frac{1}{2d-1}$, then $p^n \sigma_n \to 0$ as $n \to \infty$. Since (2.2) holds for all $n \geq 1$, we may take the limit in n to get that $\theta(p) = 0$ for $p < \frac{1}{2d-1}$. Note that this also shows that $p_c \geq \frac{1}{2d-1}$.

Now, we must prove that there exists $p \in (0,1)$ such that $\theta(p) > 0$. We will use a so-called *Peierls argument*. (Later in the course, we will study the Ising model which was the original setting for such an argument, and we'll see another one there.)

We begin with the observation that $\mathbb{Z}^d$ is a subgraph of $\mathbb{Z}^{d+1}$. (Formally, this is up to isomorphism i.e. properly we should identify $\mathbb{Z}^d$ with $\mathbb{Z}^d \times \{0\} \subset \mathbb{Z}^{d+1}$.) In particular,

$$\left\{ 0 \overset{\mathbb{Z}^d}{\leftrightarrow} \infty \right\} \subset \left\{ 0 \overset{\mathbb{Z}^{d+1}}{\leftrightarrow} \infty \right\},$$

and so $\theta(p)$ is increasing *as a function of d*. Consequently, $p_c(d)$ is decreasing as a function of d.

In particular, it suffices to prove the existence of a p such that $\theta(p) > 0$ in the case $d = 2$. The idea of the rest of the proof is that for $|\mathcal{C}(0)| < \infty$ to hold, there must be a collection of boundary edges which are all *closed*. By controlling the number of such collections, we can control the probability that one such collection is entirely closed.

Now consider the dual graph $\mathbb{Z}^2_{\text{dual}}$ of $\mathbb{Z}^2$. For any edge e of $\mathbb{Z}^2$, let e_d be the dual edge which crosses it, and declare that

$$\text{dual edge } e_d \text{ open} \quad \Longleftrightarrow \quad \text{primal edge } e \text{ closed}. \tag{2.3}$$

(Warning: some books take the opposite convention!) So in our setting, the product measure $\mathbb{P}_p$ on the primal lattice $\mathbb{Z}^2$ induces the product measure with probability $1-p$ on the dual lattice $\mathbb{Z}^2_{\text{dual}}$. From (2.3) we have the relation

$$|\mathcal{C}(0)| < \infty \text{ in } \mathbb{Z}^2 \quad \Longleftrightarrow \quad \exists \text{ cycle of open edges in } \mathbb{Z}^2_{\text{dual}} \text{ around } 0. \tag{2.4}$$

We will compute the number of such dual cycles, and then control the probability of the event described by (2.4).

Let m_n be the number of cycles in $\mathbb{Z}^2_{\text{dual}}$ around 0 with length n. Note that a dual cycle of length n must cross the positive y-axis (for the first time) at one of the points $(0, \frac{1}{2}), (0, \frac{3}{2}), \ldots, (0, \frac{n-3}{2})$. There are then at most 4^{n-1} choices for the remaining $n-1$ edges of the cycle (viewed as a path in $\mathbb{Z}^2_{\text{dual}}$). Let M_n be the number of these cycles which have all their edges open, and note that $\mathbb{E}_p[M_n] = m_n(1-p)^n$. We obtain

$$\begin{aligned}
\mathbb{P}_p\left(|\mathcal{C}(0)| < \infty\right) &= \mathbb{P}_p\left(\exists \text{ open cycle in } \mathbb{Z}^2_{\text{dual}} \text{ around } 0\right) \\
&= \sum_{n \geq 4} \mathbb{P}_p\left(M_n \geq 1\right) \\
&\leq \sum_{n \geq 4} \mathbb{E}_p\left[M_n\right] \\
&\leq \sum_{n \geq 4} (1-p)^n \, m_n \\
&\leq \sum_{n \geq 4} (1-p)^n \cdot \tfrac{n}{2} \cdot 4^{n-1}.
\end{aligned}$$

This sum converges for $1 - p < \frac{1}{4}$. So as $p \uparrow 1$, the sum tends to 0. In particular, we have

$$\theta(p) = 1 - \mathbb{P}_p\left(|\mathcal{C}(0)| < \infty\right) \to 1, \quad p \uparrow 1.$$

So there exists $p \in (0,1)$ for which $\theta(p) > 0$, which completes the proof. □

Remark. *The reduction to $d = 2$ was crucial for this argument, which could not have been formulated in higher dimensions.*

Proposition 2.7 concerns the probability of existence of an infinite cluster *containing the origin*. Although the events $\{x \leftrightarrow \infty\}$ and $\{y \leftrightarrow \infty\}$ are not independent, they should be 'almost independent' when x, y are far apart. So when $\theta(p) > 0$, and each such event has positive probability, our intuition might suggest that there should be an infinite cluster somewhere almost surely. The following proposition asserts that this intuition is valid.

Let V and E denote the vertex and edge set of $\mathbb{Z}^d$ respectively.

Proposition 2.8. *(a) If $\theta(p) = 0$, then $\mathbb{P}_p\left(\exists\, x \in V \,:\, x \leftrightarrow \infty\right) = 0$.*

(b) If $\theta(p) > 0$, then $\mathbb{P}_p\left(\exists\, x \in V \,:\, x \leftrightarrow \infty\right) = 1$.

(This dichotomy is phrased in terms of $\theta(p)$ rather than $p < p_c$ since, as we observed at the end of the previous section, it is not always known whether $\theta(p_c) = 0$.)

In the course of the proof we will need to make use of the following standard theorem (see, for example, Part B Probability, Measure and Martingales).

Theorem 2.9 (Kolmogorov's 0-1 Law)**.** *Let $X_1, X_2, \ldots$ be independent random variables, generating a σ-algebra $\mathcal{F}$. We say that $A \in \mathcal{F}$ is a* tail event *if A is independent of $(X_1, \ldots, X_n)$ for any finite n. If A is a tail event then $\mathbb{P}(A) \in \{0, 1\}$.*

Proof of Proposition 2.8. (a) Since $\mathbb{Z}^d$ is vertex-transitive, we have $\mathbb{P}_p(x \leftrightarrow \infty) = \mathbb{P}_p(0 \leftrightarrow \infty) = 0$ for all $x \in V$. But then

$$\mathbb{P}_p\left(\exists\, x \in V \,:\, x \leftrightarrow \infty\right) = \mathbb{P}_p\left(\bigcup_{x \in V} \{x \leftrightarrow \infty\}\right) \leq \sum_{x \in V} \mathbb{P}_p\left(x \leftrightarrow \infty\right) = 0,$$

since the vertex-set is countable.

(b) Let $e_1, e_2, \ldots$ be some enumeration of the (countable) set E, and note that under $\mathbb{P}_p$, the elements of the sequence $\omega(e_1), \omega(e_2), \ldots$ are i.i.d. random variables.

Note that removing finitely many edges from an infinite cluster leaves finitely many clusters, at least one of which is still infinite. In particular, this means that the event $\{\exists$ infinite cluster$\}$ is independent of $(\omega(e_1), \ldots, \omega(e_n))$ for any finite n. By Kolmogorov's 0-1 Law, we have

$$\mathbb{P}_p\left(\exists\, x \in V \,:\, x \leftrightarrow \infty\right) \in \{0, 1\}.$$

But

$$\mathbb{P}_p\left(\exists\, x \in V \,:\, x \leftrightarrow \infty\right) \geq \mathbb{P}_p\left(0 \leftrightarrow \infty\right) = \theta(p) > 0,$$

and so $\mathbb{P}_p\left(\exists\, x \in V \,:\, x \leftrightarrow \infty\right) = 1$ as required. □

2.3 Positive association and increasing events

As we have seen, in general, collections of events in Ω_G will not be independent. However, in many circumstances it will useful to exploit *positive dependence* of various events of interest. In this section we introduce the two main examples of this phenomenon, which will form a useful toolkit for more advanced analysis of percolation models.

2.3.1 The Harris inequality

Observe that Ω_G is not a totally ordered set, but it has a natural partial order: we have $\omega \leq \omega'$ if $\omega(e) \leq \omega'(e)$ for all $e \in E(G)$. We begin by recalling the definition of increasing events and functions in this setting.

Definition 2.10. An event $A \subset \Omega_G$ is *increasing* if whenever $\omega \leq \omega'$ and $\omega \in A$, then $\omega' \in A$ (i.e. the event continues to hold is more edges are added). A function $f : \Omega_G \to \mathbb{R}$ is *increasing* if whenever $\omega \leq \omega'$, we have $f(\omega) \leq f(\omega')$. In a slight abuse of notation, we will write $\mathbb{E}_p[f]$ for $\mathbb{E}_p[f(\omega)]$ where ω has law $\mathbb{P}_p$.

Example 2.11. *$\{\omega$ is connected$\}$ and $\{0 \leftrightarrow \infty\}$ are increasing events.*

Proposition 2.12 (Harris inequality)**.** *(a) Let $A, B \subset \Omega_G$ be increasing events. Then*

$$\mathbb{P}_p\left(A \cap B\right) \geq \mathbb{P}_p(A)\mathbb{P}_p(B). \tag{2.5}$$

(b) More generally, let f and g be bounded increasing functions. Then

$$\mathbb{E}_p\left[fg\right] \geq \mathbb{E}_p\left[f\right]\mathbb{E}_p\left[g\right]. \tag{2.6}$$

The Harris inequality is a special case of the *FKG inequality* (due to Fortuin, Kasteleyn and Ginibre) when we have a product measure, and we will often simply refer to it as "FKG". The general case will be introduced in the next chapter.

(The proof makes use of a small amount of martingale theory, which some of you will have encountered in Part B Probability, Measure and Martingales. A quick primer, containing all you need to know for these purposes, may be found in the appendix to these notes.)

Proof. Note that (a) follows immediately from (b) by taking $f = \mathbf{1}_A$ and $g = \mathbf{1}_B$. We will prove (b) in two steps. Throughout, we will use an arbitrary enumeration $e_1, e_2, \ldots$ of the edges of G.

Step I: First consider the special case where f and g depend on only on the states $\{\omega(e_1), \ldots, \omega(e_n)\}$ of a finite subset $\{e_1, \ldots, e_n\}$ of the edges. The proof is by induction.

When $n = 1$, we have that f and g are functions only of $\omega(e_1)$ i.e. there exist increasing functions $f_1, g_1 : \{0, 1\} \to \mathbb{R}$ such that

$$f(\omega) = f_1(\omega(e_1)), \quad g(\omega) = g_1(\omega(e_1)).$$

The fact that both f and g are increasing implies that

$$[f_1(a) - f_1(b)]\,[g_1(a) - g_1(b)] \geq 0, \quad \forall a, b \in \{0, 1\}.$$

Taking a weighted sum over all pairs of outcomes (a, b) we obtain

$$0 \leq \sum_{a,b \in \{0,1\}} \mathbb{P}_p(\omega(e_1) = a)\mathbb{P}_p(\omega(e_1) = b)\,[f_1(a) - f_1(b)]\,[g_1(a) - g_1(b)] = 2\mathbb{E}_p\,[fg] - 2\mathbb{E}_p\,[f]\,\mathbb{E}_p\,[g]\,,$$

which concludes the case $n = 1$. (The $n = 1$ case is sometimes known as Chebyshev's other inequality.)

Suppose now the result holds for $n < k$ and that f and g are increasing functions of the states $\omega(e_1), \ldots, \omega(e_k)$ i.e. there exist increasing functions $f_k, g_k : \{0, 1\}^k \to \mathbb{R}$ such that

$$f(\omega) = f_k(\omega(e_1), \ldots, \omega(e_k)), \quad g(\omega) = g_k(\omega(e_1), \ldots, \omega(e_k)).$$

Using the tower law,

$$\mathbb{E}_p\,[fg] = \mathbb{E}_{\omega(e_1),\ldots,\omega(e_{k-1})}\Big[\,\mathbb{E}_{\omega(e_k)}\,[f_k(\omega(e_1), \ldots, \omega(e_k))g_k(\omega(e_1), \ldots, \omega(e_k))]\,\Big].$$

For fixed $x_1, \ldots, x_{k-1}$, we have that $f_k(x_1, \ldots, x_{k-1}, x)$ and $g_k(x_1, \ldots, x_{k-1}, x)$ are increasing functions of x. So if we just take the expectation over the random variable $\omega(e_k)$, applying the $n = 1$ case we get

$$\mathbb{E}_{\omega(e_k)}\,[f_k(\omega(e_1), \ldots, \omega(e_k))g_k(\omega(e_1), \ldots, \omega(e_k))] \geq \mathbb{E}_{\omega(e_k)}\,[f_k(\omega(e_1), \ldots, \omega(e_k))]\,\mathbb{E}_{\omega(e_k)}\,[g_k(\omega(e_1), \ldots, \omega(e_k))]$$

Now taking the expectation over $\omega(e_1), \ldots, \omega(e_{k-1})$, we may apply the statement for $n = k-1$ to obtain

$$\begin{aligned}
&\mathbb{E}_{\omega(e_1),\ldots,\omega(e_{k-1})}\Big[\mathbb{E}_{\omega(e_k)}\,[f_k(\omega(e_1), \ldots, \omega(e_k))]\,\mathbb{E}_{\omega(e_k)}\,[g_k(\omega(e_1), \ldots, \omega(e_k))]\Big] \\
&\quad \geq \mathbb{E}_{\omega(e_1),\ldots,\omega(e_{k-1})}\Big[\mathbb{E}_{\omega(e_k)}\,[f_k(\omega(e_1), \ldots, \omega(e_k))]\Big]\,\mathbb{E}_{\omega(e_1),\ldots,\omega(e_{k-1})}\Big[\mathbb{E}_{\omega(e_k)}\,[g_k(\omega(e_1), \ldots, \omega(e_k))]\Big] \\
&\quad = \mathbb{E}_p[f]\mathbb{E}_p[g],
\end{aligned}$$

by the tower law again.

Hence, by induction on n, we obtain the claimed result in the setting where f, g depend on a finite subset of the edges.

Step II: We now assume that f, g are general bounded increasing functions $\Omega_G \to \mathbb{R}$. We introduce the natural filtration $(\mathcal{F}_n)_{n\geq 1}$ as $\mathcal{F}_n := \sigma(\omega(e_1), \ldots, \omega(e_n))$ and set

$$F_n := \mathbb{E}_p\left[f(\omega)\,\middle|\,\mathcal{F}_n\right], \quad G_n := \mathbb{E}_p\left[g(\omega)\,\middle|\,\mathcal{F}_n\right].$$

Observe that

$$\mathbb{E}_p\left[F_{n+1}\,\middle|\,\mathcal{F}_n\right] = \mathbb{E}_p\left[\,\mathbb{E}_p\left[f(\omega)\,\middle|\,\mathcal{F}_{n+1}\right]\,\middle|\,\mathcal{F}_n\right] = \mathbb{E}_p\left[f(\omega)\,\middle|\,\mathcal{F}_n\right] = F_n,$$

and so $(F_n)_{n\geq 1}$ and $(G_n)_{n\geq 1}$ are $(\mathcal{F}_n)_{n\geq 0}$-martingales. Since they are also bounded, we may apply the L^1 martingale convergence theorem, which asserts that $F_n \to \mathbb{E}_p\,[f(\omega)|\mathcal{F}_\infty] = f(\omega)$ and, similarly,

$G_n \to g(\omega)$ as $n \to \infty$, $\mathbb{P}_p$-a.s. Using the triangle inequality and boundedness, we have that for some constant $C > 0$,

$$\begin{aligned}\mathbb{E}_p\left[|F_nG_n - f(\omega)g(\omega)|\right] &\leq \mathbb{E}_p\left[|F_nG_n - f(\omega)G_n|\right] + \mathbb{E}_p\left[|f(\omega)G_n - f(\omega)g(\omega)|\right] \\ &\leq C\left(\mathbb{E}_p\left[|F_n - f(\omega)|\right] + \mathbb{E}_p\left[|G_n - g(\omega)|\right]\right) \to 0.\end{aligned}$$

So we have $\mathbb{E}_p\left[f(\omega)g(\omega)\right] = \lim_{n\to\infty}\mathbb{E}_p\left[F_nG_n\right]$. Now F_n and G_n depend only on $(\omega(e_1), \ldots, \omega(e_n))$ and so by the result of *Step I*, we have

$$\mathbb{E}_p\left[F_nG_n\right] \geq \mathbb{E}_p\left[F_n\right]\mathbb{E}_p\left[G_n\right],$$

and so taking limits we obtain

$$\mathbb{E}_p\left[f(\omega)g(\omega)\right] \geq \mathbb{E}_p\left[f(\omega)\right]\mathbb{E}_p\left[g(\omega)\right]. \qquad \square$$

Corollary 2.13. *When A is an increasing event, f an increasing function, then $\mathbb{E}_p\left[f\middle|A\right] \geq \mathbb{E}_p[f]$.*

Let us immediately give a percolation consequence of the Harris inequality. Consider a general (not necessarily vertex-transitive) connected infinite graph G. The percolation probability

$$\theta^x(p) := \mathbb{P}_p(x \leftrightarrow \infty), \quad x \in V(G),$$

will generally depend on the reference vertex x. The critical probability,

$$p_c^x := \sup\{p \in [0,1] \,:\, \theta^x(p) = 0\},$$

however, does not depend on the reference vertex, as we show in the following proposition.

Proposition 2.14. *$p_c^x = p_c^y$ for all $x, y \in V(G)$.*

Proof. For any $x, y \in V(G)$, we have

$$\theta^x(p) \geq \mathbb{P}_p\Big(\{x \leftrightarrow y\} \cap \{y \leftrightarrow \infty\}\Big) \geq \mathbb{P}_p(x \leftrightarrow y)\theta^y(p),$$

by the Harris inequality, since the events $\{x \leftrightarrow y\}$ and $\{y \leftrightarrow \infty\}$ are both increasing. In any case, we have $\theta^x(p) > 0 \iff \theta^y(p) > 0$, and so $p_c^x = p_c^y$ for all $x, y \in V(G)$. $\square$

2.3.2 The BK inequality

In this subsection, we give a partial converse to the Harris inequality.

Intuitively, the probability that there exist (edge-)disjoint open paths between u and v, and between w and x is less than the probability that such paths exist independently: once one path is declared, there are fewer options for how to construct the other path.

Definition 2.15. Given an event $A \subset \Omega_G$ and an element $\omega \in A$, say $I(\omega) \subset E(G)$ is a *witness set* of A for ω if

$$\forall\, \omega' \in \Omega_G \text{ such that } \omega(e) = \omega'(e)\, \forall e \in I(\omega), \text{ we have } \omega' \in A.$$

In words, "the restriction $\omega|_{I(\omega)}$ provides sufficient information to conclude that $\omega \in A$".

Example 2.16. *For the event $A = \{x \leftrightarrow y\}$, and $\omega \in A$, then the edges of* any *open path from x to y in ω constitute a witness set of A for ω.*

Definition 2.17. Given events $A, B \subset \Omega_G$, we define the *disjoint occurrence* of A and B as:

$$A \circ B := \Big\{\omega \in \Omega_G \,:\, \omega \in A \cap B, \exists \text{ witness sets } I(\omega) \text{ of } A \text{ and } J(\omega) \text{ of } B \text{ s.t. } I(\omega) \cap J(\omega) = \varnothing\Big\}. \quad (2.7)$$

The key example is $\{u \leftrightarrow v\} \circ \{w \leftrightarrow x\}$, the event described in words in the introduction to this section. The following inequality, named after van den Berg and Kesten, confirms our intuition for the probability of two events occurring disjointly, under certain conditions. We will see in the next section how this is used in practice.

Proposition 2.18 (BK Inequality)**.** *Let $A, B \subset \Omega_G$ be increasing events, which each depend on only finitely many edges. Then*

$$\mathbb{P}_p(A \circ B) \leq \mathbb{P}_p(A)\mathbb{P}_p(B). \tag{2.8}$$

The proof is omitted, but the interested reader may find it in Grimmett's *Probability on Graphs*.

Note. *The condition that A and B are increasing turns out to be unnecessary. The corresponding stronger result is known as* Reimer's inequality*. Its proof is considerably more involved, and the extra strength has not, in practice, been used much in the literature. The condition that A, B depend on only finitely many edges can sometimes, with sufficient care, be bypassed.*

2.4 The subcritical regime $p < p_c$: exponential tails for cluster sizes

Throughout this section, we will study $\mathbb{Z}^d$ for $d \geq 2$. In Proposition 2.7, we showed that there is a non-trivial critical probability $p_c(d) \in (0, 1)$. In this section we will study $p \in (0, p_c)$, and show that in this regime the probability that a cluster is large decays exponentially.

We will consider

$$\Lambda(n) := [-n, n]^d \cap \mathbb{Z}^d, \quad \partial\Lambda(n) = \Lambda(n) \backslash \Lambda(n-1) = \left\{x \in \mathbb{Z}^d \,:\, ||x||_\infty = n\right\}, \tag{2.9}$$

the d-dimensional box, and its boundary.

Theorem 2.19. *For all $p \in (0, p_c)$, there exists a constant $\gamma = \gamma(p) > 0$ such that*

$$\mathbb{P}_p(0 \leftrightarrow \partial\Lambda(n)) \leq e^{-\gamma n}. \tag{2.10}$$

This theorem was proved by Aizenmann and Barsky in 1987 and, roughly simultaneously, by Menshikov. Both proofs are long and involved, and introduce techniques which are valuable more generally. A considerably shorter proof was provided by Duminil-Copin and Tassion in 2016, using influence theory. Interested readers are encouraged to consult the paper *A new proof of the sharpness of the phase transition for Bernoulli percolation on $\mathbb{Z}^d$*, which is only 8 pages long, and is available at `https://www.e-periodica.ch/cntmng?pid=ens-001%3A2016%3A62%3A%3A357` .

Note. *Since $\mathbb{P}_p(0 \leftrightarrow \partial\Lambda(n)) \geq p^n$, there is a lower bound for (2.10) of the same form.*

Corollary 2.20. *For all $p \in (0, p_c)$, we have $\mathbb{E}_p[|\mathcal{C}(0)|] < \infty$.*

Proof. We have $|\partial\Lambda(n)| \sim C_d n^{d-1}$, for some constant C_d. So, given (2.10), we get

$$\mathbb{E}_p[|\mathcal{C}(0)|] \leq \sum_{n \geq 0} C_d n^{d-1} e^{-\gamma n} < \infty. \qquad \square$$

We will not give the full proof of Theorem 2.19, but rather give a proof *assuming* that we already know that $\mathbb{E}_p[|\mathcal{C}(0)|] < \infty$.

Conditional proof of Theorem 2.19. Denote by $R_n := |C(0) \cap \partial\Lambda(n)|$ the number of vertices on the boundary of the n-box which are connected to the origin by open paths. So $\sum_{n\geq 0} R_n = |\mathcal{C}(0)|$. To demonstrate exponential decay, we might try to relate $\partial\Lambda(m)$ and $\partial\Lambda(m+k)$ in a multiplicative way, for example by

$$\mathbb{P}_p\left(0 \leftrightarrow \partial\Lambda(m+k)\right) \leq \sum_{x\in\partial\Lambda(m)} \mathbb{P}_p\left(0 \leftrightarrow x \text{ and } x \leftrightarrow \partial\Lambda(m+k)\right). \tag{2.11}$$

However, applying FKG to the RHS of (2.11) gives an inequality in the wrong direction. On the other hand, if there is a path from 0 to $\partial\Lambda(m+k)$, there must be *disjoint* paths from 0 to some $x \in \partial\Lambda(m)$, and from x to $\partial\Lambda(m+k)$, so we may instead use the BK inequality.

$$\begin{aligned}
\mathbb{P}_p\left(0 \leftrightarrow \partial\Lambda(m+k)\right) &\leq \sum_{x\in\partial\Lambda(m)} \mathbb{P}_p\big(\{0 \leftrightarrow x\} \circ \{x \leftrightarrow \partial\Lambda(m+k)\}\big) \\
&\leq \sum_{x\in\partial\Lambda(m)} \mathbb{P}_p(0 \leftrightarrow x)\mathbb{P}_p\left(x \leftrightarrow \partial\Lambda(m+k)\right) \\
&\leq \sum_{x\in\partial\Lambda(m)} \mathbb{P}_p(0 \leftrightarrow x)\mathbb{P}_p(0 \leftrightarrow \partial\Lambda(k)) \\
&= \mathbb{P}_p(0 \leftrightarrow \partial\Lambda(k))\ \mathbb{E}_p\left[R_m\right].
\end{aligned} \tag{2.12}$$

Now, since $\sum_{m\geq 0} \mathbb{E}_p\left[R_m\right] = \mathbb{E}_p\left[\mathcal{C}(0)\right] < \infty$, there must exist some m such that $\mathbb{E}_p\left[R_m\right] < 1$. Fix that m. We may apply (2.12) for each $0 \leq k \leq m-1$, and iterating we find that for any $r \geq 1$,

$$\mathbb{P}_p\Big(0 \leftrightarrow \partial\Lambda(mr+k)\Big) \leq \mathbb{P}_p\big(0 \leftrightarrow \partial\Lambda(k)\big)\Big(\mathbb{E}_p\left[R_m\right]\Big)^r.$$

Then we may find a constant $\gamma > 0$ for the exponential decay appropriately to handle all of these cases simultaneously, and obtain (2.10). □

2.5 The supercritical regime $p > p_c$: infinite cluster(s)

We now turn our attention to $p > p_c$, for which $\theta(p) > 0$. We have already seen that there is an infinite cluster with probability 1 in this regime. But how many infinite clusters are there?

(Note that the results will apply equally well at p_c if $\theta(p_c) > 0$, though this is conjectured *not* to be true for all $\mathbb{Z}^d$, $d \geq 2$, as discussed earlier.)

Recall from Section 1.4 that $\mathcal{B}_G$ denotes the Borel σ-algebra on Ω_G. The following lemma is a consequence of the fact that the cylinder events (which depend on only finitely many edges) generate $\mathcal{B}_G$.

Lemma 2.21. *Let $A \in \mathcal{B}_G$ be an event. Then for any $\epsilon > 0$, there exists an event $B \in \mathcal{B}_G$ that depends on only finitely many edges*[1] *and such that the symmetric difference of A and B satisfies $\mathbb{P}_p(A \triangle B) < \epsilon$.*

Example 2.22. *We have $\{0 \leftrightarrow \infty\} = \cap_{n\geq 1}\{0 \leftrightarrow \partial\Lambda(n)\}$ and so $\{0 \leftrightarrow \infty\}$ can be approximated by $\{0 \leftrightarrow \partial\Lambda(n)\}$ for large n.*

Lemma 2.23. *Let A be an event which is invariant under automorphisms of $\mathbb{Z}^d$. Then $\mathbb{P}_p(A) = 0$ or 1.*

Proof. The idea of the proof is to exploit the fact that A is invariant under all automorphisms, while an approximating B is strongly affected by some automorphisms, e.g. a large translation.

Fix $\epsilon > 0$, and an approximating event $B \in \mathcal{B}_G$ depending only on finitely many edges and such that $\mathbb{P}_p(A \triangle B) < \epsilon$, as guaranteed by Lemma 2.21. There exists an automorphism $\varphi : G \to G$ such that B

[1] Formally, one could say: $\exists$ finite $E_B \subset E$ such that for any $\omega \in B$, given $\omega' \in \Omega_G$ satisfying $\omega'(e) = \omega(e)$, $\forall e \in E_B$, we also have $\omega' \in B$.

and $\varphi(B)$ depend on *disjoint* sets of edges. So, using this property, and the automorphism invariance of the product measure $\mathbb{P}_p$, we have

$$\mathbb{P}_p\left(B \cap \varphi(B)\right) = \mathbb{P}_p(B)\mathbb{P}_p(\varphi(B)) = \mathbb{P}_p(B)^2.$$

When we attempt to do the same thing to A, it will be useful to observe that

$$\Big(A \cap \varphi(A)\Big) \triangle \Big(B \cap \varphi(B)\Big) \subset \Big(A \triangle B\Big) \cup \Big(\varphi(A) \triangle \varphi(B)\Big),$$

and the two symmetric differences on the RHS are equally probable under $\mathbb{P}_p$. By the automorphism-invariance of the event A, we have that $A = A \cap \varphi(A)$. So then

$$\mathbb{P}_p(A) = \mathbb{P}_p(A \cap \varphi(A)) \leq \mathbb{P}_p\left(B \cap \varphi(B)\right) + 2\epsilon = \mathbb{P}_p(B)^2 + 2\epsilon \leq \Big(\mathbb{P}_p(A) + \epsilon\Big)^2 + 2\epsilon \leq \mathbb{P}_p(A)^2 + 4\epsilon + \epsilon^2.$$

Now, letting $\epsilon \downarrow 0$, we obtain $\mathbb{P}_p(A) \leq \mathbb{P}_p(A)^2$, and so $\mathbb{P}_p(A) = 0$ or 1, as required. □

Note that Lemma 2.23 provides an alternative argument for Proposition 2.8, that there is an infinite cluster almost surely. It is, similarly, a useful tool to study N_∞, the number of infinite clusters, since the event $\{N_\infty = k\}$ is invariant under automorphisms of $\mathbb{Z}^d$. Indeed, we have an immediate corollary.

Corollary 2.24. *For all $p \in [0,1]$, there exists $k = k(p) \in \{0, 1, 2, \ldots\} \cup \{\infty\}$ such that $\mathbb{P}_p(N_\infty = k) = 1$.*

The following proposition narrows down the set of possible values for $k(p)$ and, for example, eliminates the (intuituively, rather implausible!) possibility that there are almost surely exactly *eleven* infinite clusters.

Proposition 2.25. *For all $p \in [0,1]$, $k(p)$ is equal to 0 or 1 or ∞.*

Proof. We have already seen in Proposition 2.8 that when $\theta(p) = 0$, we have $k(p) = 0$. From now on, we assume $\theta(p) > 0$, which is certainly the case for $p > p_c$, and assume for a contradiction that $k = k(p) \in \{2, 3, \ldots\}$, which implies that $\mathbb{P}_p(N_\infty = 0 \text{ or } 1) = 0$.

We will study the infinite clusters via their interactions with the boundary of a large box. Note that if $\mathbb{P}_p(N_\infty = k) = 1$, then

$$\mathbb{P}_p\Big(\{N_\infty = k\} \cap \{\text{each infinite cluster meets } \partial\Lambda(n)\}\Big) \to 1 \quad \text{as } n \to \infty.$$

In particular, there exists some n such that this probability is strictly positive, which will be sufficient for our purposes.

Note that

$$\mathbb{P}_p\Big(\{N_\infty = k\} \cap \{\text{each infinite cluster meets } \partial\Lambda(n)\}\Big) \leq \mathbb{P}_p\Big(\text{each infinite cluster meets } \partial\Lambda(n)\Big),$$

and the latter event depends only on edges *outside* $\Lambda(n)$. Indeed, we have the following decomposition of events:

$$\Big\{N_\infty = 0 \text{ or } 1\Big\} \supset \Big\{\text{all edges inside } \Lambda(n) \text{ are open}\Big\} \cap \Big\{\text{each infinite cluster meets } \partial\Lambda(n)\Big\}, \tag{2.13}$$

since if every edge inside $\Lambda(n)$ is open, then any infinite clusters are in fact the same, as they are joined by a path within $\Lambda(n)$. In any case, the two events on the RHS of (2.13) depend on, respectively, the edges inside $\Lambda(n)$, and the edges outside $\Lambda(n)$, and are thus independent. As a result

$$\begin{aligned}\mathbb{P}_p\big(N_\infty = 0 \text{ or } 1\big) &\geq \mathbb{P}_p\big(\text{all edges inside } \Lambda(n) \text{ open}\big)\mathbb{P}_p\big(\text{each infinite cluster meets } \partial\Lambda(n)\big) \\ &= p^{|E(\Lambda(n))|}\,\mathbb{P}_p\big(\text{each infinite cluster meets } \partial\Lambda(n)\big) > 0,\end{aligned}$$

by our choice of n, which is a contradiction. □

We omit the proof of the following stronger result, which states that, in fact, the infinite cluster is almost surely unique when it exists. The interested reader will find it in Section 5.3 of Grimmett's *Probability on Graphs*.

Proposition 2.26. *For all $p \in [0,1]$, $k(p)$ is equal to 0 or 1. In particular, for $p > p_c$, there is a unique infinite cluster.*

2.6 The critical probability in two dimensions

As we have discussed, it is in general hard to study the critical probability $p_c(d)$ concretely for general $d \geq 3$. However, in the case $d = 2$, there is the opportunity to use the powerful planar duality framework introduced earlier. In fact, the critical probability is $p_c = 1/2$ and for this choice of probability, the percolation measure on $\mathbb{Z}^2$ and the measure induced on its dual $\mathbb{Z}^2_{\text{dual}}$ are equivalent, a fact which will be used in both sections of the proof.

Theorem 2.27 (Kesten (1980))**.** *For $\mathbb{Z}^2$, the critical probability $p_c = \frac{1}{2}$ and, furthermore, $\theta(\frac{1}{2}) = 0$.*

In particular, the phase transition for $d = 2$ is *continuous*.

Proof. The proof has two parts.

Part I: $p_c \leq 1/2$.

We will show that $p = 1/2$ is *not* subcritical, by demonstrating that the long-range path probabilities do not decay as fast as specified by Theorem 2.19.

Consider the rectangle $J_n := ([0,n] \cap \mathbb{Z}) \times ([0,n-1] \cap \mathbb{Z})$, with L_n and R_n denoting the vertices of the left and right boundaries. We will study the event $H_n := \{L_n \overset{J_n}{\leftrightarrow} R_n\}$ that there exists an open horizontal crossing within the rectangle. It is convenient to introduce also the rectangle $J'_n := (\frac{1}{2}, -\frac{1}{2}) + ([0,n-1] \cap \mathbb{Z}) \times ([0,n] \cap \mathbb{Z})$ in the dual graph $\mathbb{Z}^2_{\text{dual}}$ that meshes exactly with J_n. Denoting by T_n and B_n the top and bottom boundaries of J'_n, we will study, simultaneously, the event $V_n := \{T_n \overset{J'_n}{\leftrightarrow} B_n\}$ that there exists an open vertical crossing of J'_n in the dual graph.

Recall that a dual edge e_d is open precisely when the corresponding primal edge e (that is, the unique primal edge which crosses e_d) is closed. It follows that there cannot exist both a left-to-right open crossing in J_n *and* a top-to-bottom open crossing in J'_n. Moreover, there must exist one or the other of these. To see this, suppose there is no left-to-right crossing. Then consider the set of vertices connected L_n by paths of open edges. This set of vertices cannot contain any element of R_n. Now trace its right-hand boundary: this gives a collection of open dual edges which must contain an open path from the top to the bottom of J'_n. Hence,

$$H_n = V_n^c$$

and so $\mathbb{P}_{1/2}(H_n) + \mathbb{P}_{1/2}(V_n) = 1$.

However, under the product measure $\mathbb{P}_{1/2}$, these two events must have the same probability, since the dual edges are also open independently with probability $1/2$. Hence $\mathbb{P}_{1/2}(H_n) = 1/2$.

Now note that

$$\begin{aligned}\mathbb{P}_{1/2}(H_n) = \mathbb{P}_{1/2}\left(\bigcup_{x \in L_n} \{x \overset{J_n}{\leftrightarrow} R_n\}\right) \leq \sum_{x \in L_n} \mathbb{P}_{1/2}\left(x \overset{J_n}{\leftrightarrow} R_n\right) &\leq \sum_{x \in L_n} \mathbb{P}_{1/2}\big(0 \leftrightarrow \partial\Lambda(n)\big) \\ &= n\mathbb{P}_{1/2}\big(0 \leftrightarrow \partial\Lambda(n)\big),\end{aligned}$$

from which we find

$$\mathbb{P}_{1/2}\big(0 \leftrightarrow \partial\Lambda(n)\big) \geq \frac{1}{2n}.$$

In particular, we have shown that $\mathbb{P}_{1/2}\big(0 \leftrightarrow \partial\Lambda(n)\big)$ does *not* decay exponentially, and so with reference to Theorem 2.19, we know that $p = 1/2$ is *not* subcritical. In other words, $p_c \leq 1/2$.

Part II: To complete the argument that $p_c \geq 1/2$, it suffices to show that $\theta(\frac{1}{2}) = 0$.

Suppose, for a contradiction, that $\theta(\frac{1}{2}) > 0$, and so $\mathbb{P}_{1/2}\,(\exists \text{ infinite cluster}) = 1$. Then there is also almost surely an infinite open dual cluster.

As before, we will study interactions with a large box, this time $\Lambda(n)$. Note that if there exists an infinite cluster somewhere almost surely then

$$\lim_{n\to\infty} \mathbb{P}_{1/2}\Big(\partial\Lambda(n) \leftrightarrow \infty\Big) = 1.$$

Let us now denote by L_n, R_n, T_n, B_n the left, right, top, and bottom boundaries of the box $\Lambda(n)$, respectively. We have

$$\{\partial\Lambda(n) \leftrightarrow \infty\} = \{L_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\} \cup \{R_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\} \cup \{T_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\} \cup \{B_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\}.$$

By Question 3 on Problem Sheet 2, if $A_1, \ldots, A_k$ are increasing events then

$$\max_{1\leq i\leq k} \mathbb{P}_p(A_i) \geq 1 - \big(1 - \mathbb{P}_p(A_1 \cup \ldots \cup A_k)\big)^{1/k}.$$

The events $\{L_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\}$, $\{R_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\}$, $\{T_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\}$ and $\{B_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\}$ are all increasing and have equal probability under $\mathbb{P}_{1/2}$. So

$$\begin{aligned}
&\mathbb{P}_{1/2}\left(L_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\right) \\
&\geq 1 - \left(1 - \mathbb{P}_{1/2}\left(\{L_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\} \cup \{R_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\} \cup \{T_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\} \cup \{B_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\}\right)\right)^{1/4} \\
&= 1 - \Big(1 - \mathbb{P}_{1/2}\big(\partial\Lambda(n) \leftrightarrow \infty\big)\Big)^{1/4}.
\end{aligned}$$

So we must have

$$\lim_{n\to\infty} \mathbb{P}_{1/2}\left(L_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\right) = 1,$$

and similarly for the other three events. We can apply a similar argument in the *dual* box $\Lambda'(n) := (\frac{1}{2}, \frac{1}{2}) + [-n, n-1]^2 \cap \mathbb{Z}^2$ which sits just inside $\Lambda(n)$. Write L'_n, R'_n, T'_n, B'_n for the boundaries of $\Lambda'(n)$, so that

$$\lim_{n\to\infty} \mathbb{P}_{1/2}\left(T'_n \overset{\Lambda'(n)^c}{\underset{\text{dual}}{\leftrightarrow}} \infty\right) = 1.$$

Consider two primal infinite paths and two dual infinite paths simultaneously. Then we have

$$\lim_{n\to\infty} \mathbb{P}_{1/2}\left(\{L_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\} \cap \{R_n \overset{\Lambda(n)^c}{\leftrightarrow} \infty\} \cap \{T'_n \overset{\Lambda'(n)^c}{\underset{\text{dual}}{\leftrightarrow}} \infty\} \cap \{B'_n \overset{\Lambda'(n)^c}{\underset{\text{dual}}{\leftrightarrow}} \infty\}\right) = 1, \tag{2.14}$$

since each of the events in the intersection has probability tending to 1. Now note that none of these infinite paths can cross *outside* $\Lambda(n)$, since open primal and open dual paths never cross. Let us refer to these infinite paths as the "left infinite cluster" and so on. Suppose that there is a path of open edges joining the left infinite cluster to the right infinite cluster. Then the top and bottom dual infinite clusters must not be connected. Similarly, if there is a path of open dual edges joining the top and bottom dual infinite clusters then the left and right infinite clusters must be disconnected from one another in the primal. So, on the event studied in (2.14), we must have either at least two distinct infinite primal clusters in $\mathbb{Z}^2$, or at least two distinct infinite dual clusters in $\mathbb{Z}^2_{\text{dual}}$. Hence,

$$\mathbb{P}_{1/2}\left(\{\geq 2 \text{ infinite clusters in } \mathbb{Z}^2\} \cup \{\geq 2 \text{ infinite dual clusters in } \mathbb{Z}^2_{\text{dual}}\}\right) = 1.$$

But this is a contradiction, since the probabilities of both events in the union are zero by Proposition 2.26. □

Chapter 3

The Ising model

In this section, we will discuss one of the most important and well-studied models in statistical mechanics, called the *Ising model*. This model was invented by Wilhelm Lenz in 1920. However, it is named after his student Ernst Ising, who was the first to publish results on this model in 1925, following his 1924 thesis, titled "Contribution to the Theory of Ferromagnetism."

The Ising model describes the physical setting of ferromagnetic material immersed in a magnetic field.

We note that ferromagnetism is a common phenomenon encountered frequently in everyday life. For example, a fridge magnet is an example of a "permanent magnet" obtained by exposing a ferromagnetic material (e.g., iron) to a strong magnetic field in such a way that it remains magnetized after it is removed from the field. It will thereafter be attracted to other ferromagnetic material (e.g., the steel in a fridge door).

Mathematically, a material of interest is modeled as a set of particles arranged in a graph (often a Euclidean lattice $\mathbb{Z}^d$ or some subset of it). Each particle is given one of two possible spins, *up* or *down*. Two parameters govern the behaviour of the system, the *inverse temperature* $1/\beta$ and the strength h of the *external magnetic field*. The effects of and interplay between these parameters is as follows: As the inverse temperature is increased (that is, the temperature is decreased), spins of neighbouring particles tend more to align with each other. On the other hand, as the strength of the external magnetic field is increased, spins tend more to align with the field.

In this course, we will focus on an intriguing physical phenomenon called *spontaneous magnetisation,* observed at low temperatures when the magnetic field is zero (absent). More specifically, in spite of the natural symmetry of up/down spins, it is possible under certain conditions to observe a preference of the system for either up or down spins, even in the absence of an external magnetic field. Indeed, suppose we expose a ferromagnetic material to a positive external magnetic field, and then decrease the intensity of this field until it reaches zero. Provided the temperature is lower than a certain threshold (which depends on the material) called the *Curie temperature,* the system will continue to exhibit a strictly positive *average magnetisation* (corresponding to a majority of up spins) even after the material is removed from the field. On the other hand, at higher temperatures, the magnetisation vanishes entirely when the field is removed (called *paramagnetic* behaviour).

3.1 The Ising model Hamiltonian

Let $G = (V, E)$ be a finite graph. Let $\beta \in \mathbb{R}_{\geq 0}$ and $h \in \mathbb{R}$. We will think of β as the *inverse temperature* and h as the *external magnetic field*.

A *configuration* ω of the Ising model on G is an element of the set $\Omega_G = \{\pm 1\}^V$. This represents a so-called *microstate* of the system, in which vertices $i \in V$ represent particles, and edges $(i,j) \in E$ represent bonds between neighbouring particles. Moreover, each vertex $i \in V$ is assigned a "spin" $\omega_i \in \{0,1\}$, that is, either a $+1$ (*up/positive* spin) or -1 (*down/negative* spin).

The *energy* associated with a given configuration ω is quantified using the so-called Ising *Hamiltonian*

$$\mathcal{H}_{G;\beta,h}(\omega) = -\beta \sum_{(i,j)\in E} \sigma_i(\omega)\sigma_j(\omega) - h\sum_{i\in V} \sigma_i(\omega),$$

where $\sigma_i(\omega) = \omega_i$ is the ith projection map. Note that we use the Greek letter σ here, as $\sigma_i(\omega)$ gives us the spin at location i in the configuration ω.

We can then define a *Gibbs measure* on the set of all possible configurations, by setting

$$\mu_{G;\beta,h}(\omega) = \frac{e^{-\mathcal{H}_{G;\beta,h}(\omega)}}{Z_{G;\beta,h}} = \frac{1}{Z_{G;\beta,h}} \exp\left[\beta \sum_{(i,j)\in E} \sigma_i(\omega)\sigma_j(\omega) + h\sum_{i\in V} \sigma_i(\omega)\right]$$

where

$$Z_{G;\beta,h} = \sum_{\omega\in\Omega_G} e^{-\mathcal{H}_{G;\beta,h}(\omega)}$$

is the normalising constant, called the *partition function* of the model. As we shall soon discover, the normalising constant $Z_{G;\beta,h}$, in spite of its apparently minor technical role, has deep physical significance. It will play a crucial role in describing the macroscopic features of the model.

In practice, one does not generally know the specific microstate of the system (that is, the spins of all particles) but rather only its macrostate, that is, a probability distribution on the set of microstates (that is, configurations) which accounts for random microscopic fluctuations. In this context, the measure $\mu_{G;\beta,h}$ is the macrostate of the model at *inverse temperature* β and with *external magnetic field* h.

Note that, under the measure $\mu_{G;\beta,h}$, microstates ω are more likely the more often that spins of neighbouring particles are aligned with each other and the external magnetic field. The first effect is amplified at high values of β (that is, low temperatures) and the second effect is proportional to the value of $|h|$.

We will often want to measure the value of some observable, that is, find the expected value $\mathbb{E}_{G;\beta,h}(f)$ of some real random variable f on the probability space $(\Omega_G, \mathcal{P}(\Omega_G), \mu_{G;\beta,h})$. However, in this section of the lecture notes (and in accordance with much of the statistical physics literature), we will use the "angle brackets/physics" notation for expected value,

$$\langle f\rangle_{G;\beta,h} = \mathbb{E}_{G;\beta,h}(f).$$

That being said, you are welcome to use either notation in your solutions to class and exam problems, etc., as you prefer.

For example, we can define the *total magnetisation* of a configuration ω as

$$\sum_{i\in V} \sigma_i(\omega)$$

and the *average magnetisation* as

$$m_G(\omega) = \frac{1}{|V|}\sum_{i\in V} \sigma_i(\omega).$$

The average magnetisation of the system (that is, the expected value of the function $m_G : \Omega_G \to \mathbb{R}$) is

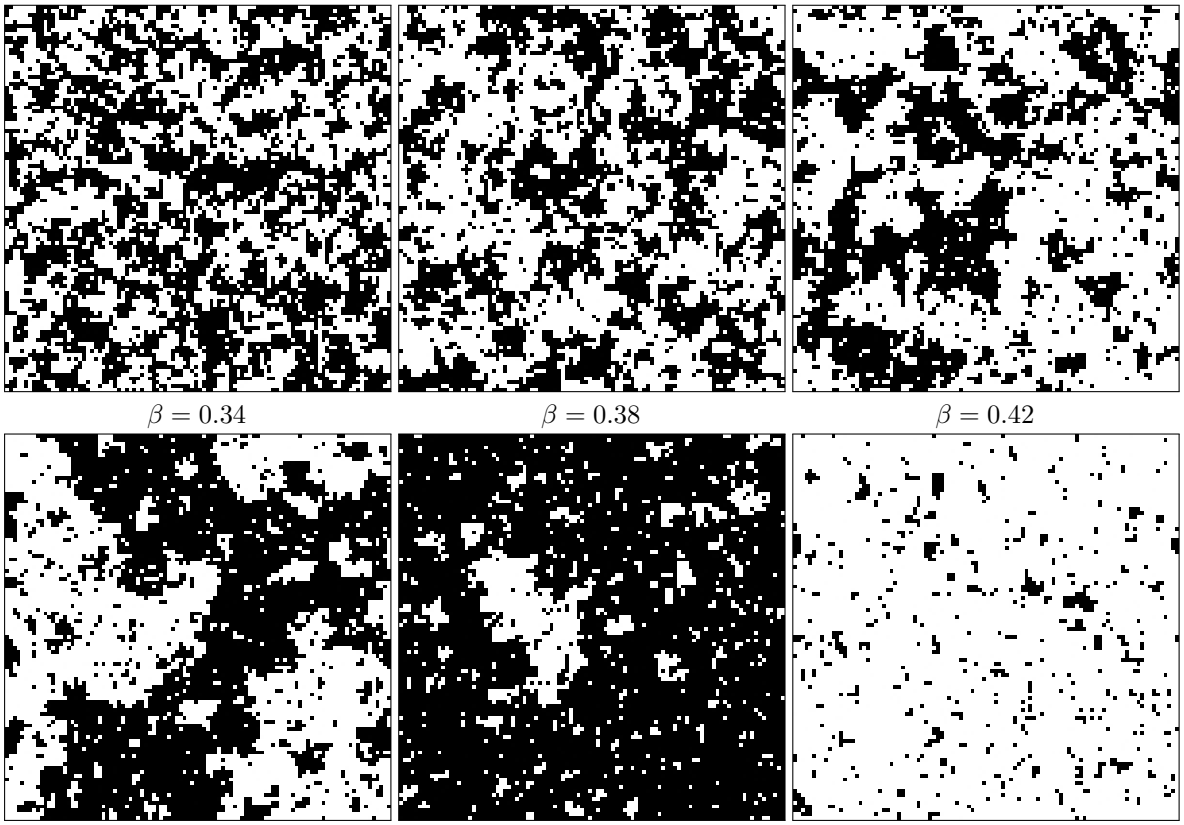


Figure 3.1: Simulations of the Ising model on a 100×100 box, at various values of the inverse temperature β. Black cells represent $+$ spins and white cells represent $-$ spins. The "critical point" for the Ising model on $\mathbb{Z}^2$ (the inverse Curie temperature) is $\beta_c \approx 0.4407$.

then the quantity

$$\begin{aligned}
\langle m_G \rangle_{G;\beta,h} &= \sum_{\omega \in \Omega_G} m_G(\omega) \mu_{G;\beta,h}(\omega) \\
&= \frac{1}{|V| Z_{G;\beta,h}} \sum_{\omega \in \Omega_G} e^{-\mathcal{H}_{G;\beta,h}(\omega)} \sum_{i \in V} \sigma_i(\omega) \\
&= \frac{1}{|V| Z_{G;\beta,h}} \sum_{\omega \in \Omega_G} \exp\left[\beta \sum_{(i,j) \in E} \sigma_i(\omega) \sigma_j(\omega) + h \sum_{i \in V} \sigma_i(\omega) \right] \sum_{i \in V} \sigma_i(\omega). \qquad (3.1)
\end{aligned}$$

3.2 Average magnetisation, pressure and phase transitions

As discussed above, part of our motivation for defining the Ising model was to capture the phenomenon of spontaneous magnetisation. We are therefore led to wonder whether the theoretical setup outlined in the previous section enables us to observe, for some values of the inverse temperature β, some peculiar behaviour of the average magnetisation $\langle m_G \rangle_{G;\beta,h}$ when h is at or near 0. More specifically, with reference to our previous discussions of the phenomenon, we expect to see something like that in Figure 3.2. That is, spontaneous magnetisation corresponds to $\langle m_G \rangle_{G;\beta,h}$ being discontinuous at $h = 0$.

As mentioned above, the partition function is useful in expressing macroscopic observables of the model. Let us now describe how this works in the case of the average magnetisation.

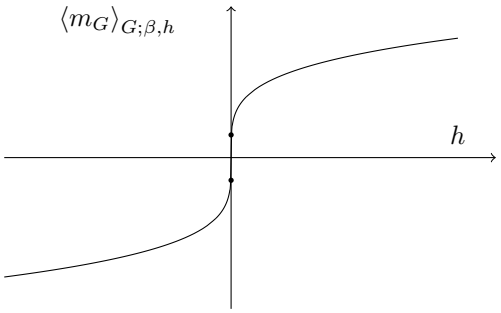


Figure 3.2: If the temperature $1/\beta$ (that is, the inverse of the inverse temperature β) of the system is lower than the Curie temperature then (1) as $h \downarrow 0$, the average magnetisation tends to a positive value, and (2) as $h \uparrow 0$, it tends to a strictly negative value.

We define the *pressure* of the model as the quantity

$$\Psi_{G;\beta,h} = \frac{1}{|V|} \log Z_{G;\beta,h}.$$

Observe (exercise) that by differentiating the pressure with respect to h, we obtain the average magnetisation (3.1), that is,

$$\frac{\partial}{\partial h} \Psi_{G;\beta,h} = \langle m_G \rangle_{G;\beta,h}$$

In fact, all moments of the random variable m_G can be expressed in terms of partial derivatives of the pressure. In this way, we see that fine features of the distribution of the magnetisation can be recovered using the partition function of the model.

However, observe that, the expression (3.1) is a ratio of finite sums of functions of the form Ae^{hB}. Therefore, within our current framework, the magnetisation is *not* discontinuous in h. Indeed, the pressure itself is a real analytic function in h (all of its partial derivatives in h are continuous).

The reason for this failure to capture spontaneous magnetisation is that the system is finite (that is, $|V| < \infty$). However, as the number of particles increases, the graph of the magnetisation will increasingly *look* like it has a discontinuity. Therefore, in order to mathematically observe spontaneous magnetisation, we will need to take a limit, as the number of particles $|V| \to \infty$.

To this end, let $\Lambda(n) \subset \mathbb{Z}^d$ be the subgraph of the d-dimensional Euclidean lattice induced by the "box" $\{-n, \ldots, n\}^d$. The *thermodynamic limit pressure* is defined as the limiting pressure

$$\Psi(\beta, h) = \lim_{n \to \infty} \Psi_{\Lambda(n);\beta,h}.$$

It can be shown that $\Psi(\beta, h)$ is a real convex function in β, h on all of its domain $\mathbb{R}_{\geq 0} \times \mathbb{R}$.

Motivated by the above discussion, we make the following definition.

Definition 3.1. The Ising model on $\mathbb{Z}^d$ has a first-order phase transition at (β, h) if $\Psi(\beta, h)$ is not differentiable at (β, h).

3.3 Configuration boundary conditions

As a first step towards obtaining infinite-volume Gibbs measures, by taking limits of finite-volume measures $\mu_{G;\beta,h}$, we define slightly more general models that allow for boundary conditions. In order to simplify notation during our limiting procedure, we shall redefine finite-volume Gibbs measures as probability measures on the space $(\{\pm 1\}^{\mathbb{Z}^d}, \mathcal{B})$, where $\mathcal{B}$ is the σ-algebra of Borel sets for the product topology,

that are concentrated on configurations of the form $\Omega_G \times \prod_{i \in \mathbb{Z}^d \setminus V} \{\eta_i\}$ for some finite subgraph $G = (V, E)$ of $\mathbb{Z}^d$ and a given configuration $\eta \in \{\pm 1\}^{\mathbb{Z}^d}$, the latter of which is used to specify the boundary conditions in what follows.

More specifically, for such a G and η, let

$$\Omega_G^\eta = \{\omega \in \{\pm 1\}^{\mathbb{Z}^d} \mid \omega_i = \eta_i \ \forall i \in \mathbb{Z}^d \setminus V\}$$

denote the set of configurations which agree with η outside of V.

Likewise, we define the Ising Hamiltonian on G with boundary conditions η as

$$\mathcal{H}^\eta_{G;\beta,h}(\omega) = -\beta \sum_{(i,j) \in E \cup \partial G} \sigma_i(\omega)\sigma_j(\omega) - h \sum_{i \in V} \sigma_i(\omega),$$

where ∂G is the set of "boundary edges" in $\mathbb{Z}^d$ with exactly one endpoint in V. Note that the only difference between $\mathcal{H}^\eta_{G;\beta,h}$ and $\mathcal{H}_{G;\beta,h}$ is a summand

$$-\beta \sum_{(i,j) \in \partial G} \sigma_i(\omega)\sigma_j(\omega),$$

related to interactions between particles in V and their neighbors outside of V. The spins of particles outside of V are considered to be "frozen" (that is, non-random), as prescribed η, and as such do not appear in the above expression $\mathcal{H}^\eta_{G;\beta,h}$.

The finite-volume Gibbs measure $\mu^\eta_{G;\beta,h}$ is the probability measure on $(\{\pm 1\}^{\mathbb{Z}^d}, \mathcal{B})$ given by

$$\mu^\eta_{G;\beta,h}(A) = \sum_{\omega \in A \cap \Omega^\eta_G} \mu^\eta_{G;\beta,h}(\omega), \quad A \in \mathcal{B}.$$

Here, similarly as before,

$$\mu^\eta_{G;\beta,h}(\omega) = \frac{e^{-\mathcal{H}^\eta_{G;\beta,h}(\omega)}}{Z^\eta_{G;\beta,h}}, \quad \omega \in \Omega^\eta_G,$$

where $Z^\eta_{G;\beta,h}$ is the partition function (that is, the appropriate normalising factor).

Given a function $f : \{\pm 1\}^{\mathbb{Z}^d} \to \mathbb{R}$ (that is measurable with respect to $\mathcal{B}$) we write $\langle f \rangle^\eta_{G;\beta,h}$ for its expected value under the measure $\mu^\eta_{G;\beta,h}$.

Of particular importance are the extreme cases where η is either $\{+1\}^{\mathbb{Z}^d}$ or $\{-1\}^{\mathbb{Z}^d}$. In these cases, we will simply write $\mu^+_{G;\beta,h}$ and $\mu^-_{G;\beta,h}$ to denote the finite-volume Gibbs measures on G with "all positive" and "all negative" boundary conditions.

Note that, at this stage, the σ-algebra $\mathcal{B}$ is not so relevant. However, our construction of the free and wired uniform spanning forests in Chapter 1 should make the reason for our selection of the appropriate measurable space clear. Namely, we plan to investigate limits of the form

$$\mu^{\eta_n}_{G_n;\beta,h} \Rightarrow \mu,$$

where $(G_n)_{n \geq 1}$ is an exhaustion of $\mathbb{Z}^d$, $(\eta_n)_{n \geq 1}$ is a sequence of configurations in $\{\pm 1\}^{\mathbb{Z}^d}$, and the convergence $\Rightarrow$ is weak convergence of measures on the measurable topological space $(\{\pm 1\}^{\mathbb{Z}^d}, \mathcal{B})$. We will discuss this in more detail in the next section.

3.4 Infinite-volume Gibbs measures

In this section, we establish the existence of two (potentially equal) infinite-volume Gibbs measures $\mu^\pm_{\beta,h}$ on $\mathbb{Z}^d$ of fundamental importance. These measures are obtained as weak limits of finite-volume Gibbs

measures with all positive/negative boundary conditions.

Given some sequence $(\eta_n)_{n\geq 1}$ of configurations in $\{\pm 1\}^{\mathbb{Z}^d}$ and an exhaustion $(G_n)_{n\geq 1}$ of $\mathbb{Z}^d$, it is natural to ask whether the sequence of probability measures $(\mu^{\eta_n}_{G_n;\beta,h})_{n\geq 1}$ converges weakly to some limiting measure on the space $(\{\pm 1\}^{\mathbb{Z}^d}, \mathcal{B})$. The appropriate setting has essentially already been put in place in Section 1.4, the only difference being that the measurable space there was $(\{0,1\}^{E(G)}, \mathcal{B}_G)$ for some infinite graph G.

In this instance, we can again define

- a *partial order* $\leq$ on $\{\pm 1\}^{\mathbb{Z}^d}$, where $\omega \leq \omega'$ if $\omega_i \leq \omega'_i$ for all sites i in $\mathbb{Z}^d$;
- *increasing cylinder events* sets of the form

$$\mathcal{C}_A = \{\omega \in \{\pm 1\}^{\mathbb{Z}^d} \mid \omega_i = +1 \ \forall i \in A\},$$

  where A is a finite subset of $\mathbb{Z}^d$;
- *increasing functions* $f : \{\pm 1\}^{\mathbb{Z}^d} \to \mathbb{R}$ such that $f(\omega) \leq f(\omega')$, for all $\omega \leq \omega'$.

As in Section 1.4, in studying weak convergence, it suffices to consider cylinder events.

Proposition 3.2. *Suppose that, for all finite subsets A of $\mathbb{Z}^d$, the limit*

$$\mu(\mathcal{C}_A) = \lim_{n\to\infty} \mu^{\eta_n}_{G_n;\beta,h}(\mathcal{C}_A)$$

exists. Then the mapping $\mathcal{C}_A \mapsto \mu(\mathcal{C}_A)$ can be uniquely extended to a probability measure μ on $(\{\pm 1\}^{\mathbb{Z}^d}, \mathcal{B})$. Moreover, the sequence $(\mu^{\eta_n}_{G_n;\beta,h})_{n\geq 1}$ converges weakly to μ, that is, $\mu^{\eta_n}_{G_n;\beta,h} \Rightarrow \mu$.

Definition 3.3. Let $\mathcal{G}(\beta, h)$ denote the set *infinite-volume Gibbs measures* on $\mathbb{Z}^d$, that is, the set of all probability measures on $(\{\pm 1\}^{\mathbb{Z}^d}, \mathcal{B})$ which are weak limits of sequences of the form $(\mu^{\eta_n}_{G_n;\beta,h})_{n\geq 1}$.

In this section, our objective is to construct a measure in $\mathcal{G}(\beta, h)$. In the next section, we will discuss how uniqueness of the infinite-volume Gibbs measure relates to the existence of a phase transition. More specifically, we turn now aim to construct the measures $\mu^{\pm}_{\beta,h} \in \mathcal{G}(\beta, h)$, which arise as limits when η is either $\{+1\}^{\mathbb{Z}^d}$ or $\{-1\}^{\mathbb{Z}^d}$. Our construction makes use of the following two fundamental properties of finite-volume Gibbs measures (which, as it turns out, also hold for infinite-volume Gibbs measures).

The first result we require is the FKG inequality, which has already been discussed in Section 2.3.1 above, in the context of percolation.

Theorem 3.4 (FKG inequality)**.** *Let G be a finite subgraph of $\mathbb{Z}^d$. Then, for all increasing functions $f, g : \{\pm 1\}^{\mathbb{Z}^d} \to \mathbb{R}$, we have that*

$$\langle fg\rangle^{\eta}_{G;\beta,h} \geq \langle f\rangle^{\eta}_{G;\beta,h}\langle g\rangle^{\eta}_{G;\beta,h}.$$

We omit the proof of the FKG inequality, but the interested reader can find it in many places. For instance, see Section 4 in Grimmett's text *Probability on Graphs: Random Processes on Graphs and Lattices*. Therein, it is shown that the FKG inequality holds for any positive probability measure μ on $\Omega = \{0,1\}^S$, for a finite set S, provided that the so-called *FKG lattice condition* holds:

$$\mu(\omega \vee \omega')\mu(\omega \wedge \omega') \geq \mu(\omega)\mu(\omega'), \quad \omega, \omega' \in \Omega.$$

Here the configurations on the left hand side are defined coordinate-wise by $(\omega \vee \omega)_i = \max\{\omega_i, \omega'_i\}$ and $(\omega \wedge \omega)_i = \min\{\omega_i, \omega'_i\}$. (In the present case, we are considering a measure $\mu^{\eta}_{G;\beta,h}$ on $\{\pm 1\}^V$, but we can

encode each $\omega_i = 2\hat{\omega}_i - 1$ using some $\hat{\omega}_i \in \{0, 1\}$.) Intuitively, $\mu^{\eta}_{G;\beta,h}$ satisfies the FKG lattice condition since spins will be more aligned in the configurations $\omega \vee \omega'$ and $\omega \wedge \omega'$ than in ω and ω'.

The second result is straightforward (exercise) to show, using the explicit form of the Hamiltonian $\mathcal{H}^{\eta}_{G;\beta,h}(\omega)$.

Theorem 3.5 (Spatial Markov property)**.** *For finite subgraphs $G' \subseteq G$ of $\mathbb{Z}^d$, we have that*

$$\mu^{\eta}_{G;\beta,h}(\cdot \mid \sigma_i = \eta_i \, \forall i \in V(G) \setminus V(G')) = \mu^{\eta}_{G';\beta,h}(\cdot).$$

Using these results, we will prove the following.

Proposition 3.6. *Let $(G_n)_{n\geq 1}$ be an exhaustion of $\mathbb{Z}^d$. The sequence of probability measures $(\mu^{+}_{G_n;\beta,h})_{n\geq 1}$ converges weakly to a probability measure $\mu^{+}_{\beta,h}$ on $(\{\pm 1\}^{\mathbb{Z}^d}, \mathcal{B})$. Moreover, the measure $\mu^{+}_{\beta,h}$ does not depend on the exhaustion $(G_n)_{n\geq 1}$ and is translation invariant.*

Proof. By Proposition 3.2, in order to show convergence, we need to show the existence of

$$\lim_{n\to\infty} \mu^{+}_{G_n;\beta,h}(\mathcal{C}_A)$$

for all finite subsets A of $\mathbb{Z}^d$.

To this end, fix such an A. Let $V_n = V(G_n)$. Since (G_n) is an exhaustion, for some N_A we have that $A \subseteq V_n$ for $n \geq N_A$. Let $n \geq N_A$.

Note that

$$\mu^{+}_{G_n;\beta,h}(\mathcal{C}_A) = \mu^{+}_{G_{n+1};\beta,h}(\mathcal{C}_A \mid \sigma_i = +1 \, \forall i \in V_{n+1} \setminus V_n)$$

by Theorem 3.5. Next, note that

$$\mu^{+}_{G_{n+1};\beta,h}(\mathcal{C}_A \mid \sigma_i = +1 \, \forall i \in V_{n+1} \setminus V_n) = \frac{\langle 1_{\mathcal{C}_A} 1_{\mathcal{C}_{V_{n+1}\setminus V_n}} \rangle^{+}_{G_{n+1};\beta,h}}{\langle 1_{\mathcal{C}_{V_{n+1}\setminus V_n}} \rangle^{+}_{G_{n+1};\beta,h}}.$$

Then, applying Theorem 3.4 to the increasing functions $1_{\mathcal{C}_A}$ and $1_{\mathcal{C}_{V_{n+1}\setminus V_n}}$, we find that

$$\begin{aligned}
\frac{\langle 1_{\mathcal{C}_A} 1_{\mathcal{C}_{V_{n+1}\setminus V_n}} \rangle^{+}_{G_{n+1};\beta,h}}{\langle 1_{\mathcal{C}_{V_{n+1}\setminus V_n}} \rangle^{+}_{G_{n+1};\beta,h}} &\geq \frac{\langle 1_{\mathcal{C}_A} \rangle^{+}_{G_{n+1};\beta,h} \langle 1_{\mathcal{C}_{V_{n+1}\setminus V_n}} \rangle^{+}_{G_{n+1};\beta,h}}{\langle 1_{\mathcal{C}_{V_{n+1}\setminus V_n}} \rangle^{+}_{G_{n+1};\beta,h}} \\
&= \langle 1_{\mathcal{C}_A} \rangle^{+}_{G_{n+1};\beta,h} \\
&= \mu^{+}_{G_{n+1};\beta,h}(\mathcal{C}_A).
\end{aligned}$$

Altogether, we conclude that, for all $n \geq N_A$,

$$\mu^{+}_{G_{n+1};\beta,h}(\mathcal{C}_A) \leq \mu^{+}_{G_n;\beta,h}(\mathcal{C}_A).$$

Therefore, by monotonicity, we obtain a limit

$$\mu^{+}_{G_n;\beta,h}(\mathcal{C}_{\mathcal{A}}) \downarrow \mu^{+}_{\beta,h}(\mathcal{C}_A),$$

as required.

Finally, arguing as in the proof of Proposition 1.18, it can be shown that this limit does not depend on the exhaustion. Likewise, the fact that it is translation invariant can be proven in the same vein as Proposition 1.23. □

By symmetry (replacing all spins ± 1 by ∓ 1), we obtain the analogous result for negative boundary conditions.

Corollary 3.7. *Let $(G_n)_{n\geq 1}$ be an exhaustion of $\mathbb{Z}^d$. The sequence of probability measures $(\mu^{-}_{G_n;\beta,h})_{n\geq 1}$ converges weakly to a probability measure $\mu^{-}_{\beta,h}$ on $(\{\pm 1\}^{\mathbb{Z}^d}, \mathcal{B})$. Moreover, the measure $\mu^{-}_{\beta,h}$ does not depend on the exhaustion $(G_n)_{n\geq 1}$ and is translation invariant.*

3.5 Uniqueness of the infinite-volume Gibbs measure and phase transitions

In the previous section, we obtained the infinite-volume Gibbs measures $\mu^{\pm}_{\beta,h}$. In this section, we show that inequality $\mu^{+}_{\beta,h} \neq \mu^{-}_{\beta,h}$ of these measures corresponds to a phase transition for the Ising model at the point (β, h).

First, we note that the measures $\mu^{+}_{\beta,h}$ and $\mu^{-}_{\beta,h}$ are extreme points of the set of all infinite-volume Gibbs measures $\mathcal{G}(\beta, h)$ on $\mathbb{Z}^d$.

Lemma 3.8. *Let $\mu \in \mathcal{G}(\beta, h)$ be an infinite-volume Gibbs measure on $\mathbb{Z}^d$ and let $f : \{\pm 1\}^{\mathbb{Z}^d} \to \mathbb{R}$ be a local increasing function (i.e., an increasing function that depends on only finitely many spins). Then we have that*

$$\langle f \rangle^{-}_{\beta,h} \leq \langle f \rangle_{\mu} \leq \langle f \rangle^{+}_{\beta,h},$$

where $\langle f \rangle_{\mu}$ is the expected value of f under μ.

Proof. This lemma follows immediately from the finite-volume version which appears on the third problem sheet. □

As a consequence, we have

Proposition 3.9. *The following are equivalent:*

1. *There is a unique infinite-volume Gibbs measure, $|\mathcal{G}(\beta, h)| = 1$*
2. *$\mu^{+}_{\beta,h} = \mu^{-}_{\beta,h}$*
3. *$\langle \sigma_0 \rangle^{+}_{\beta,h} = \langle \sigma_0 \rangle^{-}_{\beta,h}$, where σ_0 is the spin at the origin.*

Provided that $h = 0$, statements (1-3) are equivalent with $\mu^{+}_{\beta,h}(\sigma_0 = +1) = \mu^{+}_{\beta,h}(\sigma_0 = -1) = 1/2$.

The equivalence of (1) and (2) are intuitively clear by Lemma 3.8, and (3) is equivalent by the translation invariance of $\mu^{\pm}_{\beta,h}$. Item (3) and (4) are equivalent by the symmetry of the model.

Proof. Clearly (1) ⇒ (2) ⇒ (3). Also, the fact that (2) ⇒ (1) follows immediately by Lemma 3.8. Indeed, given a finite subset A of $\mathbb{Z}^d$ and a measure $\mu \in \mathcal{G}(\beta, h)$, we have

$$\langle 1_{\mathcal{C}_A} \rangle^{-}_{\beta,h} \leq \langle 1_{\mathcal{C}_A} \rangle_{\mu} \leq \langle 1_{\mathcal{C}_A} \rangle^{+}_{\beta,h}.$$

Hence

$$\mu^{-}_{\beta,h}(\mathcal{C}_A) \leq \mu(\mathcal{C}_A) \leq \mu^{+}_{\beta,h}(\mathcal{C}_A).$$

Therefore, if (2) holds, holds then by Proposition 3.2 we find that (1) holds.

We next show that (3) ⇒ (2). Fix a finite subset A of $\mathbb{Z}^d$ and consider the local increasing function

$$f_A = \sum_{i \in A} 1_{\mathcal{C}_{\{i\}}} - 1_{\mathcal{C}_A}$$

on $\{\pm 1\}^{\mathbb{Z}^d}$. Applying Lemma 3.8, we have

$$\langle f_A \rangle^{-}_{\beta,h} \leq \langle f_A \rangle^{+}_{\beta,h},$$

and so, by linearity,

$$\sum_{i\in A}\Big[\langle 1_{\mathcal{C}_{\{i\}}}\rangle^-_{\beta,h} - \langle 1_{\mathcal{C}_{\{i\}}}\rangle^+_{\beta,h}\Big] \leq \langle 1_{\mathcal{C}_A}\rangle^-_{\beta,h} - \langle 1_{\mathcal{C}_A}\rangle^+_{\beta,h} = \mu^-_{\beta,h}(\mathcal{C}_A) - \mu^+_{\beta,h}(\mathcal{C}_A).$$

Next, observe that

$$1_{\mathcal{C}_{\{i\}}} = \frac{1}{2}(1+\sigma_i).$$

As such,

$$\sum_{i\in A}\Big[\langle 1_{\mathcal{C}_{\{i\}}}\rangle^-_{\beta,h} - \langle 1_{\mathcal{C}_{\{i\}}}\rangle^+_{\beta,h}\Big] = \frac{1}{2}\sum_{i\in A}\Big[\langle \sigma_i\rangle^-_{\beta,h} - \langle \sigma_i\rangle^+_{\beta,h}\Big].$$

Then, assuming (3), it follows by translation invariance that

$$\frac{1}{2}\sum_{i\in A}\Big[\langle \sigma_i\rangle^-_{\beta,h} - \langle \sigma_i\rangle^+_{\beta,h}\Big] = \frac{1}{2}\sum_{i\in A}\Big[\langle \sigma_0\rangle^-_{\beta,h} - \langle \sigma_0\rangle^+_{\beta,h}\Big] = 0.$$

It then follows that

$$\mu^+_{\beta,h}(\mathcal{C}_A) \leq \mu^-_{\beta,h}(\mathcal{C}_A).$$

But since, by Lemma 3.8, the reverse inequality also holds, we conclude that

$$\mu^+_{\beta,h}(\mathcal{C}_A) = \mu^-_{\beta,h}(\mathcal{C}_A),$$

and so, by Proposition 3.2, that (2) holds.

To finish the proof, we show that the last statement is implied by (3) when $h = 0$ (and the argument can be reversed to show the other direction). Writing out the expectations $\langle\sigma_0\rangle^+_{\beta,h}$, and $\langle\sigma_0\rangle^-_{\beta,h}$, (3) is equivalent to

$$\mu^+(\sigma_0 = +1) - \mu^+(\sigma_0 = -1) = \mu^-(\sigma_0 = +1) - \mu^-(\sigma_0 = -1).$$

By symmetry of the model $\mu^-(\sigma_0 = +1) = \mu^+(\sigma_0 = -1)$ and $\mu^-(\sigma_0 = -1) = \mu^+(\sigma_0 = +1)$. After rearranging, the above is equivalent to

$$\mu^+(\sigma_0 = +1) = \mu^+(\sigma_0 = -1).$$

As, $\mu^+(\sigma_0 = +1) + \mu^+(\sigma_0 = -1) = \mu^+(\Omega) = 1$, this finishes the proof. □

Proposition 3.9 draws a connection between the uniqueness of the infinite-volume Gibbs measure and the equality of the two quantities $\langle\sigma_0\rangle^+_{\beta,h}$ and $\langle\sigma_0\rangle^-_{\beta,h}$, that is, the expected values of the spin at the origin under the extremal measures $\mu^+_{\beta,h}$ and $\mu^-_{\beta,h}$. These quantities are are naturally related to the average magnetisation $\langle m_G\rangle_{G;\beta,h}$, see (3.1) above. Indeed, in the third problem sheet we will show that the quantities $\langle\sigma_0\rangle^\pm_{\beta,h}$ are equal to the thermodynamic limit of the average magnetisation with $\pm$ boundary conditions, that is,

$$\langle\sigma_0\rangle^\pm_{\beta,h} = \lim_{n\to\infty}\frac{1}{|\Lambda(n)|}\Big\langle\sum_{i\in\Lambda(n)}\sigma_i\Big\rangle^\pm_{\Lambda(n),\beta,h}.$$

However, one can say even more.

Theorem 3.10. *The expected values $\langle\sigma_0\rangle^\pm_{\beta,h}$ of the spin at the origin are equal to the one-sided partial derivative with respect to h of the pressure. That is,*

$$\langle\sigma_0\rangle^\pm_{\beta,h} = \frac{\partial\Psi(\beta,h)}{\partial h^\pm}.$$

Therefore, by Proposition 3.9, the Ising model on $\mathbb{Z}^d$ has a first-order phase transition at (β, h) if and only if there are at least two distinct infinite-volume Gibbs measures on $\mathbb{Z}^d$.

In closing, we note that, from a more probabilistic rather than physical point of view, it is quite natural to take non-uniqueness of the infinite-volume Gibbs measure as the *definition* of a phase transition.

3.6 The phase diagram of the Ising model

Finally, in this section, we show that the Ising model on $\mathbb{Z}^d$ displays a phase transition.

Ising, in his thesis, derived an explicit expression in the one-dimensional case for the thermodynamic limit pressure $\Psi(\beta, h)$. As it turns out, when $d = 1$ it is real analytic, excluding the presence of a phase transition. The third problem sheet contains a guided exercise exemplifying the use of the transfer matrix method in order to obtain this result. In fact, Ising conjectured the absence of a phase transition also when $d > 1$, however, we shall see that this is *not* the case.

In summary, when $d \geq 2$, the situation is as follows.

1. The limit pressure $\Psi(\beta, h)$ is analytic for all $\beta \geq 0$ and $h \neq 0$. Hence, if $h \neq 0$, there is no phase transition at (β, h).
2. When $h = 0$, there is a *critical inverse temperature* β_c such that for $\beta \leq \beta_c$ we have uniqueness and for $\beta > \beta_c$ we have non-uniqueness of the infinite-volume Gibbs measure. In other words, we have a first-order phase transition at $(\beta, 0)$ for all $\beta > \beta_c$, and thus our model captures the phenomenon of spontaneous magnetisation, as desired.

We note that $1/\beta_c$ is the so-called *Curie temperature* discussed above.

We will not show (1), as it is best suited for (complex) analytic methods that are outside the scope of this course. However, note that this is in accordance with our expectations with regards to the physical phenomenon of spontaneous magnetisation described in Section 3.2.

We turn our attention to (2).

When $d = 2$, it is still possible to "completely solve" the Ising model by computing the thermodynamic pressure. This was done by Lars Onsager in 1944, a celebrated achievement in the field.

We will not discuss Onsager's solution in this course, but rather present the first proof, given by Rudolf Peierls in 1936, of the fact that there *is* a phase transition for the Ising model in dimension $d = 2$. Specifically, we will show that $\mu^+_{\beta,0}(\sigma_0 = -1) < 1/2$ for large enough β. Though certainly superseded by Onsager's solution, the Peierls' argument is comparatively elementary, and has the added advantage of being generalizable to higher dimensions.

Theorem 3.11. *There exists a $\beta_c \geq 0$ such that the Ising model on $\mathbb{Z}^2$ has a first-order phase transition at $(\beta, 0)$ for all $\beta > \beta_c$. That is, $\langle\sigma_0\rangle^+_{\beta,0} \neq \langle\sigma_0\rangle^-_{\beta,0}$ for all $\beta > \beta_c$.*

Proof. Consider the Ising model on the 2-dimensional box $\Lambda(n) \subset \mathbb{Z}^2$ with the all positive boundary condition. Let ω be a configuration $\Omega^+_{\Lambda(n)}$. Consider the subgraph $B(\omega)$ of the dual lattice, whose edge set consists of all edges in the dual lattice which correspond to edges in $E(\Lambda(n)) \cup \partial\Lambda(n)$ which join sites with opposing spins. Intuitively, each edge in $B(\omega)$ "separates" two vertices with opposing signs, and so the set of edges in $B(\omega)$ together partition the spins into regions with the same sign. Indeed, note that $B(\omega)$ is an *even* subgraph of the dual lattice, that is, each of its vertices have even degree 0, 2 or 4. Therefore $B(\omega)$ can be canonically decomposed as a union of disjoint cycles $\gamma_1, \ldots, \gamma_k$ (by resolving each crossing in a canonical way) as in Figure 3.3. As such, we claim that

$$\mu^+_{\Lambda(n);\beta,0}(\omega) \propto \exp\left[-2\beta \sum_{i=1}^{k} |\gamma_i|\right],$$

where $|\gamma_i|$ is the number of edges in the ith cycle. Indeed, to see this, let $E_n = E(\Lambda(n)) \cup \partial\Lambda(n)$, and

then simply observe that (noting $h = 0$)

$$\begin{aligned}
-\frac{1}{\beta}\mathcal{H}^+_{\Lambda(n);\beta,0}(\omega) &= |\{(i,j) \in E_n \mid \omega_i = \omega_j\}| - |\{(i,j) \in E_n \mid \omega_i \neq \omega_j\}| \\
&= |E_n| - 2|\{(i,j) \in E_n \mid \omega_i \neq \omega_j\}| \\
&= |E_n| - 2|E(B(\omega))| \\
&= |E_n| - 2\sum_{i=1}^{k} |\gamma_i|.
\end{aligned}$$

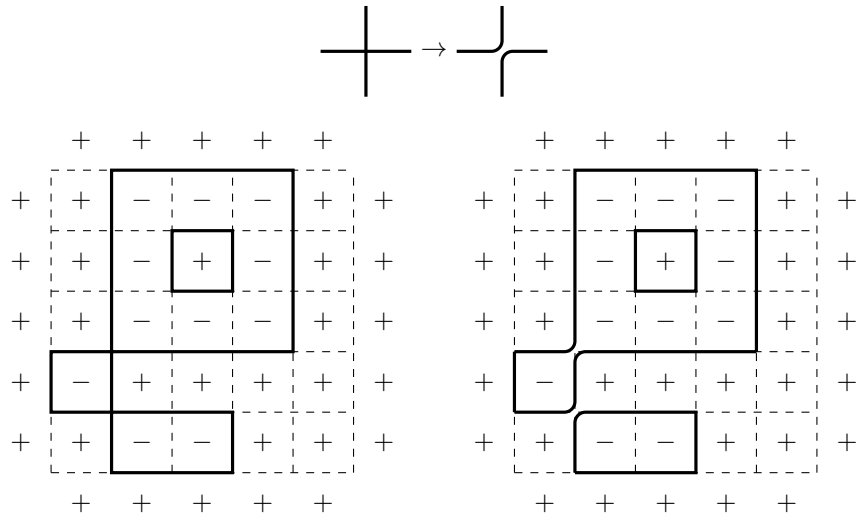


Figure 3.3: Resolving "crossings" in order to express $B(\omega)$ as a union of cycles.

Next, consider the probability $\mu^+_{\Lambda(n);\beta,0}(\sigma_0 = -1)$. If $\sigma_0(\omega) = -1$ then, since $\omega \in \Omega^+_{\Lambda(n)}$, one of the cycles of $B(\omega)$ contains the origin. Let Γ_n be the set of all cycles in the dual lattice of $\Lambda(n)$ that contain the origin. For each such $\gamma \in \Gamma_n$, let

$$E_\gamma = \{\omega \in \Omega^+_{\Lambda(n)} \mid \gamma \subseteq B(\omega)\}.$$

Taking a union bound,

$$\mu^+_{\Lambda(n);\beta,0}(\sigma_0 = -1) \leq \sum_{\gamma \in \Gamma_n} \mu^+_{\Lambda(n);\beta,0}(E_\gamma).$$

To finish the proof, we bound $\mu^+_{\Lambda(n);\beta,0}(E_\gamma)$ as follows. Consider the "switch" map $s_\gamma : E_\gamma \to \Omega^+_{\Lambda(n)}$ that changes the spin of every particle *inside* the cycle γ. Note that s_γ is a bijection between E_γ and its image $s_\gamma(E_\gamma)$ in $\Omega^+_{\Lambda(n)}$. Moreover, observe that

$$B(s_\gamma(\omega)) = B(\omega) \setminus \gamma.$$

Therefore

$$\mu^+_{\Lambda(n);\beta,0}(\omega) = e^{-2\beta|\gamma|}\mu^+_{\Lambda(n);\beta,0}(s_\gamma(\omega)).$$

Consequently,

$$\begin{aligned}\mu^+_{\Lambda(n);\beta,0}(\sigma_0=-1) &\le \sum_{\gamma\in\Gamma_n}\sum_{\omega\in E_\gamma}\mu^+_{\Lambda(n);\beta,0}(\omega)\\ &= \sum_{\gamma\in\Gamma_n} e^{-2\beta|\gamma|}\sum_{\omega\in E_\gamma}\mu^+_{\Lambda(n);\beta,0}(s_\gamma(\omega))\\ &= \sum_{\gamma\in\Gamma_n} e^{-2\beta|\gamma|}\mu^+_{\Lambda(n);\beta,0}(s_\gamma(E_\gamma))\\ &\le \sum_{\gamma\in\Gamma_n} e^{-2\beta|\gamma|}.\end{aligned}$$

Finally, note that we can bound the number of $\gamma\in\Gamma_n$ of length $|\gamma|=\ell$ by $(\ell/2)3^{\ell-1}$, using the same argument as in the proof of Proposition 2.7 in (**??**). Therefore

$$\mu^+_{\beta,0}(\sigma_0=-1)=\lim_{n\to\infty}\mu^+_{\Lambda(n);\beta,0}(\sigma_0=-1)\le\sum_{\ell=1}^{\infty}(\ell/6)(3e^{-2\beta})^\ell=\frac{1}{6}\frac{3e^{-2\beta}}{(1-3e^{-2\beta})^2}<\frac{1}{2}$$

for all

$$\beta>\log\sqrt{\frac{18}{7-\sqrt{13}}}\approx 0.8314.$$

We conclude that $\mu^+_{\beta,0}(\sigma_0=-1)<1/2$ for all sufficiently large β, which is equivalent to $\langle\sigma_0\rangle^+_{\beta,0}\neq\langle\sigma_0\rangle^-_{\beta,0}$ by Proposition 3.9. Hence, by monotonicity, there is some critical $\beta_c\ge 0$ at which point $\langle\sigma_0\rangle^+_{\beta,0}\neq\langle\sigma_0\rangle^-_{\beta,0}$ for all $\beta>\beta_c$. □

Note that this argument gives an upper bound $\beta_c\le 0.8341$, however, it is known that the inverse Curie temperature in dimension $d=2$ is $\beta_c\approx 0.4407$, see Figure 3.1 above.

3.7 Related models: Potts and random cluster

The *Potts model* (introduced by Renfrey Potts in 1951) is a generalisation of the Ising model where, instead of having two possible spins, particles may be assigned any one of q distinct "colours" in the set $[q]=\{1,2,\dots,q\}$. Given a finite graph $G=(V,E)$, the Potts model on G with parameter q is the probability measure on $[q]^V$ given by

$$\pi_{G;q,\beta}(\omega)=\frac{1}{Z^{\text{Potts}}_{G;q,\beta}}\exp\left[\beta\sum_{(i,j)\in E}1_{\omega_i=\omega_j}\right],$$

where $Z^{\text{Potts}}_{G;q,\beta}$ is the partition function of the model. Note that, when $q=2$, the Potts model directly corresponds to the Ising model with $h=0$ (and a certain inverse temperature that one can compute as a function of the Potts parameter β).

The next generalization that we will discuss has connections with several of the topics of in this course so far: the Ising and Potts models, percolation, and even uniform spanning trees. Once again, let $G=(V,E)$ be a finite graph. The *random cluster model,* sometimes called the *RC model* or the *FK representation/percolation model* after Kees Fortuin and Piet Kasteleyn, is a probability measure on $\{0,1\}^E$ defined as follows. For $\omega\in\{0,1\}^E$, let $k(\omega)$ be the number of connected components in the

configuration ω. Then, with parameters $q \in \mathbb{R}^+$ and $p \in (0,1)$, the RC probability measure is given by

$$\begin{aligned}\phi_{p,q}(\omega) &= \frac{1}{Z^{\mathrm{RC}}_{G;p,q}} q^{k(\omega)} \prod_{e \in E} p^{\omega(e)}(1-p)^{1-\omega(e)} \\ &= \frac{1}{Z^{\mathrm{RC}}_{G;p,q}} p^{n_{\mathrm{open}}(\omega)}(1-p)^{n_{\mathrm{closed}}(\omega)} q^{k(\omega)},\end{aligned}$$

where

$$\begin{aligned}n_{\mathrm{open}}(\omega) &= |\{e \in E \mid \omega(e) = 1\}| \\ n_{\mathrm{closed}}(\omega) &= |\{e \in E \mid \omega(e) = 0\}|\end{aligned}$$

are the numbers of open and closed edges in ω, and $Z^{\mathrm{RC}}_{G;p,q}$ is the appropriate normalising constant.

Note that $\phi_{p,1}$ is the probability measure corresponding to bond percolation on G. On the other hand, for $q > 1$, the RC measure favours configurations with a larger number of connection components. For $q < 1$, configurations with fewer connected components receive more weight. One can also obtain the uniform spanning tree (UST) model as a limiting case of the RC model. More specifically, as discussed in Problem Sheet 3, the UST probability measure on $\{0,1\}^E$ is the limit of the measures $\phi_{p,q}$, as $p, q \to 0$ and $q/p \to 0$.

Finally, let us discuss a coupling between the Potts model (including the Ising case $q = 2$) and the random cluster model on a finite graph $G = (V, E)$. We can generate a configuration in $[q]^V$ according to the following two-step process:

1. First, generate an edge configuration $\omega \in \{0,1\}^E$ according to the RC measure $\phi_{p,q}$, with $p = 1 - e^{-\beta}$.
2. Then, given the edge configuration ω, assign to each component C of ω a colour chosen independently and uniformly at random from the set $[q]$, This induces a configuration $\sigma \in [q]^V$.

Theorem 3.12. *The configuration σ is distributed according to the Potts measure $\pi_{G;q,\beta}$.*

Proof. Fix a configuration $\sigma \in [q]^V$. We show that the probability of obtaining σ via the two-step procedure above is equal to $\pi_{G;q,\beta}(\sigma)$ in the following way. First, we partition the edge set

$$E = E^{=}(\sigma) \cup E^{\neq}(\sigma),$$

where, for each $e = (i,j) \in E$, we put $e \in E^{=}(\sigma)$ if $\sigma_i = \sigma_j$ and $e \in E^{\neq}(\sigma)$ otherwise. We say that a configuration ω is *compatible* with σ, and write $\omega \sim \sigma$, if every edge $e \in E^{\neq}(\sigma)$ is closed (that is, $\omega(e) = 0$) in ω. Note that edges $e \in E^{=}(\sigma)$ can be open or closed, with no effect upon whether $\omega \sim \sigma$. The reason for these definitions is that, if ω is generated as in step (1) above, then the probability of obtaining σ from ω in step (2) is simply $1/q^{k(\omega)}$ if $\omega \sim \sigma$, and 0 otherwise. Therefore, the probability with the two-step process produces σ is

$$\begin{aligned}\sum_{\omega:\omega\sim\sigma} \phi_{p,q}(\omega)\frac{1}{q^{k(\omega)}} &\propto \sum_{\omega:\omega\sim\sigma} (1-p)^{n_{\mathrm{closed}}(\omega)} p^{n_{\mathrm{open}}(\omega)} q^{k(\omega)} \frac{1}{q^{k(\omega)}} \\ &= \sum_{\omega:\omega\sim\sigma} (1-p)^{n_{\mathrm{closed}}(\omega)} p^{n_{\mathrm{open}}(\omega)} \\ &= (1-p)^{|E^{\neq}(\sigma)|} \\ &\propto (1-p)^{-|E^{=}(\sigma)|} \\ &\propto \exp\left[-\log(1-p) \sum_{(i,j)\in E} 1_{\sigma_i=\sigma_j}\right] \\ &\propto \pi_{G;q,\beta}(\sigma),\end{aligned}$$

noting that $\beta = -\log(1-p)$ since $p = 1 - e^{-\beta}$, as required. □

Just as we did for spanning trees, we can consider a *wired* version of the RC model. For a finite region $G \subset \mathbb{Z}^d$, consider edge configurations in $\omega \in \{0,1\}^{E \cup \partial G}$. For such an ω, let $k^{\text{wired}}(\omega)$ be the number of connected components in ω which do not intersect $\mathbb{Z}^d \setminus V$ (that is, which do not contain a boundary edge $e \in \partial G$). We then define the wired RC model on G as

$$\phi^{\text{wired}}_{p,q}(\omega) = \frac{1}{Z^{\text{wired}}_{G;p,q}} q^{k^{\text{wired}}(\omega)} \prod_{e \in E(G) \cup \partial G} p^{\omega(e)}(1-p)^{1-\omega(e)}.$$

Just as we used the free RC model above to obtain free boundary Potts configurations, we can used this wired RC model to obtain Ising configurations from the extremal Ising measures $\mu^{\pm}_{G;\beta,0}$ (and their Potts model equivalents). For step (2) above in this case, we fix some particular boundary value (either $+$ or $-$ in the Ising model, or more generally some colour $b \in \{1,\dots,q\}$ in the Potts model) and if any component of ω includes a boundary edge then all its sites receive the boundary value in the configuration σ.

Results like Theorem 3.12 make it clear that connectivity properties in the RC model are related to spin correlations in the Ising/Potts model. See Problem Sheet 3 for further details.

Chapter 4

Mixing times

Let $(X_t)_{t\geq 0}$ be an irreducible and aperiodic discrete-time Markov chain on a finite state space S. Recall that (X_t) has a unique stationary distribution π, which is its limiting/equillibrium distribution. From any initial state (or starting distribution, if the initial state itself is random) the distribution of X_t converges to π as $t \to \infty$. But how quickly does that convergence occur? In other words, how long do we need to wait until the chain is (approximately) in equilibrium? The theory of *mixing times* provides a mathematical framework in which to ask such questions precisely, and develops techniques to answer them.

We will write P_x^t for the probability measure on S describing the distribution of the state of the chain at time t, when started in state x. That is,

$$P_x^t(y) = \mathbb{P}(X_t = y \mid X_0 = x), \quad x, y \in S.$$

Recall that we are ultimately interested in determining how close the distribution of a Markov chain is to its equillibrium π by some time t. The following gives us a way of measuring the distance between two measures.

Definition 4.1. Let μ, ν be two measures on a set S. The *total variation distance* between μ and ν is

$$|\mu - \nu|_{\mathrm{TV}} = \max_{A \subseteq S} |\mu(A) - \nu(A)|.$$

The total variation distance is a metric on the set of all probability measures on S (exercise).

We use this metric to measure closeness to stationarity by letting

$$d(t) = \max_{x \in S} |P_x^t - \pi|_{\mathrm{TV}}.$$

This gives a uniform upper bound, over all possible starting positions x. Note that

$$\lim_{t \to \infty} d(t) = 0.$$

Definition 4.2. Let $\epsilon > 0$. The *ϵ-mixing time*

$$t_{\mathrm{mix}}(\epsilon) = \min\{t \geq 0 \mid d(t) \leq \epsilon\}$$

is the time by which, regardless of its starting position, the law of the position X_t of the Markov chain at time t is at most distance ϵ from its stationary/equilibrium distribution π.

The dependence on ϵ is usually rather straightforward, and so we often put $t_{\mathrm{mix}} = t_{\mathrm{mix}}(1/4)$, fixing some ϵ arbitrarily as $1/4$.

The value $t_{\text{mix}}(\epsilon)$ for specific Markov chains of fixed size is often of interest. For example, we might be interested in this quantity in the natural context of shuffling a standard deck of 52 playing cards. However, our focus in this course is on mixing time asymptotics for sequences of Markov chains $(X_t^n)_{t\in\mathbb{N}}$ on state spaces S_n, where $|S_n| \to \infty$ and $n \to \infty$. In this context, we study the behaviour of the sequence t_{mix}^n (for some fixed $\epsilon > 0$) as $n \to \infty$. For example, one might wonder how the mixing time of a random walk on a cycle graph C_n of length n depends on n. Moreover, we will mainly concentrate on proving *upper* bounds for the mixing time.

Of particular interest for us are Markov chains on product spaces S_n, such as $\{0,1\}^{E_n}$ or $[q]^{V_n}$, for some sequence of graphs $G_n = (V_n, E_n)$, whose states can be interpreted as configurations in a (for example) percolation, random cluster, or Ising/Potts model on G_n.

Note that often (as in the examples $\{0,1\}^{E_n}$ and $[q]^{V_n}$ above) the cardinality of the state space S_n grows exponentially. However, when the mixing times t_{mix}^n for such a family of Markov chains grow polynomially, the family of chains is called *fast mixing*. Such Markov chains can be very useful, for example, in efficiently generating random objects according to some target distribution.

4.1 Coupling and mixing times

Recall that a *coupling* of a pair of probability measures (μ, ν) is a pair of random variables (X, Y) defined on the same probability space in such a way that the marginal distribution of X is μ and the marginal distribution of Y is ν.

Lemma 4.3 (Coupling Lemma)**.** *Let μ, ν be probability measures on a finite set S. Then*

$$|\mu - \nu|_{\text{TV}} = \inf\{\mathbb{P}(X \neq Y) \mid (X, Y) \text{ is a coupling of } \mu, \nu\}.$$

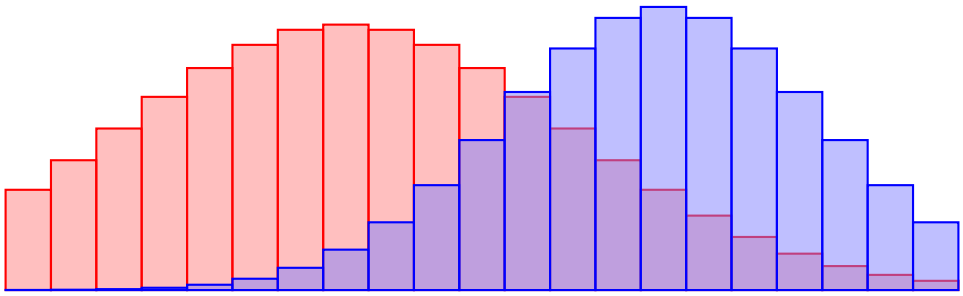


Figure 4.1: The two graphs (red and blue) represent the measures μ and ν as functions from S to $[0, 1]$. The area of each graph is 1, and one can show (do you see why?) that the area of the region in which the graphs overlap is exactly $1 - |\mu - \nu|_{\text{TV}}$. Suppose we wish to choose uniform points in the two regions in such a way that if one of the two falls within the intersection, then the two points coincide. Can we do that? What does that say about a coupling of (μ, ν)?

Proof. We start by showing that the right hand side is an upper bound for the total variation distance. Let (X, Y) be a coupling of (μ, ν). Let $A \subseteq S$. Note that

$$\begin{aligned}
|\mu(A) - \nu(A)| &= |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \\
&= |\mathbb{P}(X \in A, Y \neq X) - \mathbb{P}(Y \in A, Y \neq X)| \\
&\leq \max\{\mathbb{P}(X \in A, Y \neq X), \mathbb{P}(Y \in A, Y \neq X)\} \\
&\leq \mathbb{P}(Y \neq X).
\end{aligned}$$

Therefore, recalling that

$$|\mu - \nu|_{\mathrm{TV}} = \max_{A \subseteq S} |\mu(A) - \nu(A)|,$$

it follows that, for any such coupling (X, Y),

$$|\mu - \nu|_{\mathrm{TV}} \leq \mathbb{P}(Y \neq X),$$

yielding the desired upper bound.

To complete the proof, one can find (exercise) a specific *optimal coupling* (X^*, Y^*) for which

$$\mathbb{P}(X^* \neq Y^*) = |\mu - \nu|_{\mathrm{TV}}. \qquad \square$$

Next, let us note the following useful fact.

Lemma 4.4. *For all t, we have that*

$$d(t) \leq \rho(t) \leq 2d(t),$$

where

$$\rho(t) = \max_{x,y \in S} |P_x^t - P_y^t|_{\mathrm{TV}}.$$

Proof. The second inequality follows by the triangle inequality,

$$\begin{aligned}
\rho(t) &= \max_{x,y \in S} |P_x^t - P_y^t|_{\mathrm{TV}} \\
&\leq \max_{x,y \in S} (|P_x^t - \pi|_{\mathrm{TV}} + |P_y^t - \pi|_{\mathrm{TV}}) \\
&= \max_{x \in S} |P_x^t - \pi|_{\mathrm{TV}} + \max_{y \in S} |P_y^t - \pi|_{\mathrm{TV}} \\
&= 2d(t).
\end{aligned}$$

As for the other inequality, first note that, since π is stationary,

$$\pi(x) = \sum_{y \in S} \pi(y) P_y^t(x),$$

for any $x \in S$. Therefore, in particular,

$$\pi(A) = \sum_{y \in S} \pi(y) P_y^t(A).$$

Hence

$$\begin{aligned}
|P_x^t - \pi|_{\mathrm{TV}} &= \max_{A \subseteq S} |P_x^t(A) - \pi(A)| \\
&= \max_{A \subseteq S} \left| P_x^t(A) - \sum_{y \in S} \pi(y) P_y^t(A) \right| \\
&= \max_{A \subseteq S} \left| \sum_{y \in S} \pi(y) [P_x^t(A) - P_y^t(A)] \right| \\
&\leq \max_{A \subseteq S} \max_{y \in S} |P_x^t(A) - P_y^t(A)| \\
&= \max_{y \in S} |P_x^t - P_y^t|_{\mathrm{TV}}.
\end{aligned}$$

Finally, by maximizing over $x \in S$, we obtain $d(t) \leq \rho(t)$. $\square$

Using the coupling lemma, one can show (exercise) that ρ is submultiplicative, that is,

$$\rho(t+s) \leq \rho(t)\rho(s).$$

Given this fact, let us discuss how $t_{\text{mix}}(\epsilon)$ depends on ϵ. By Lemma 4.4 and the submultiplicativity of ρ, we have that

$$d(ct) \leq \rho(ct) \leq \rho(t)^c \leq (2d(t))^c,$$

for all integers $c \geq 1$. From now on, we put

$$t_{\text{mix}} = t_{\text{mix}}(1/4).$$

Then, by the above inequality and the definition of t_{mix}, it follows that

$$d(ct_{\text{mix}}) \leq [2d(t_{\text{mix}})]^c \leq 2^{-c}.$$

Hence,

$$d\left(-\frac{\log(\epsilon)}{\log(2)}t_{\text{mix}}\right) \leq \epsilon,$$

and therefore, for all $\epsilon \in (0, 1/4)$, we have

$$t_{\text{mix}} \leq t_{\text{mix}}(\epsilon) \leq -\frac{\log(\epsilon)}{\log(2)}t_{\text{mix}}.$$

For this reason, we often estimate t_{mix}, rather than derive explicit estimates for $t_{\text{mix}}(\epsilon)$ involving ϵ. Indeed, for any given ϵ, $t_{\text{mix}}(\epsilon)$ will differ from t_{mix} by at most a constant factor (depending on ϵ).

In the next two sections, we will study two examples of families of Markov chains. Using Lemmas 4.4 and 4.3, we will estimate their mixing times and show that they are "fast mixing."

4.2 The lazy random walk on the hypercube

Recall that the *hypercube* on $\{0,1\}^n$ is the graph where $x, y \in S$ are joined by an edge if and only if $x_i = y_i$ for all except exactly one "bit" (that is, coordinate) $i \in [n]$. In this section, we consider *lazy random walk* on the hypercube. A Markov chain (X_t) is called "lazy" if at any given time t there a chance that the chain will remain where it is for another time step, $X_{t+1} = X_t$. Note that all lazy Markov chains are aperiodic.

Consider the Markov chain $(X_t)_{t\geq 1}$ on state space $S = \{0,1\}^n$, constructed as follows. Given $X_t = x$,

1. choose a coordinate $i \in \{1, \ldots, n\}$ uniformly at random and independently of X_t, and then
2. conditionally on i, set

$$X_{t+1} = \begin{cases} x & \text{with probability } 1/2 \\ x^{(i)} & \text{with probability } 1/2, \end{cases}$$

where $x^{(i)}$ is is obtained from x by "flipping" its ith bit, that is,

$$(x^{(i)})_j = \begin{cases} x_j & j \neq i \\ 1 - x_i & j = i. \end{cases}$$

Hence the transition probabilities for the Markov chain are

$$P(x,y) = \begin{cases} 1/(2n) & y = x^{(i)} \\ 1/2 & y = x \\ 0 & \text{otherwise.} \end{cases}$$

This Markov chain is clearly aperiodic and irreducible. Moreover, since $P(x,y) = P(y,x)$, it follows that its unique stationary distribution π is the uniform measure on S.

Our goal in this section is to estimate the mixing time t_{mix}, via the "coupling method."

We proceed as follows. Fix two starting points $x, y \in S$. We define couplings (X_t, Y_t) of (P_x^t, P_y^t) by building a Markov chain $(X_t, Y_t)_{t\geq 0}$ such that $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are distributed as the original Markov chain described above, started at $X_0 = x$ and at $Y_0 = y$. We achieve this using a common source of randomness: let $(Z_t)_{t\geq 0}$ be an IID sequence of uniform random variables on $[n]$, and let $(B_t)_{t\geq 0}$ be independent IID sequence of Bernoulli$(1/2)$ random variables. We set $(X_0, Y_0) = (x, y)$. For subsequent times, given (X_t, Y_t), we define (X_{t+1}, Y_{t+1}) by changing the (Z_t)th coordinates of X_t and Y_t to B_t. Note that, by construction, $X_t \sim P_x^t$ and $Y_t \sim P_y^t$ for all $t \geq 0$. Indeed, note that the transition probabilities of (X_t) satisfy

$$\mathbb{P}(X_{t+1} = s^{(i)} \mid X_t = s) = \mathbb{P}(Z_t = i, B_t = 1 - s_i) = 1/(2n)$$

and

$$\mathbb{P}(X_{t+1} = s \mid X_t = s) = \mathbb{P}(B_t = s_{Z_t}) = 1/2,$$

and similarly for (Y_t).

Next, observe that once a coordinate is selected by the process (Z_t), the the corresponding coordinates of X_t and Y_t will agree for all times thereafter. In other words, by time $t \geq 1$, we have that $(X_t)_i = (Y_t)_i$ for all coordinates $i \in \{Z_{t'} : t' < t\}$. As such, there is evidently a connection with the famous *coupon collector* problem (see Problem Sheet 1). More specifically, note that

$$\mathbb{P}(X_t \neq Y_t) \leq \mathbb{P}(\{Z_0, \ldots, Z_{t-1}\} \neq \{1, \ldots, n\}) = \mathbb{P}(T > t),$$

where

$$T = \max_i \min\{t : Z_t = i\}$$

is the "coupon collector time" by which all n coordinates have been selected at least once. As is well-known, it takes on average about (up to constants) $n \log n$ trials to collect n coupons. More specifically, it can be shown that

$$\mathbb{P}(T > t(c,n)) \leq e^{-c},$$

where $t(c,n) = \lceil n \log n + cn \rceil$. Note that this upper bound does not depend on x, y. Therefore, applying Lemmas 4.4 and 4.3, we find that

$$\begin{aligned} d(t(c,n)) &\leq \rho(t(c,n)) \\ &= \max_{x,y\in S} |P_x^{t(c,n)} - P_y^{t(c,n)}|_{\text{TV}} \\ &\leq \max_{x,y\in S} \mathbb{P}(X_{t(c,n)} \neq Y_{t(c,n)}) \\ &\leq \mathbb{P}(T > t(c,n)) \\ &\leq e^{-c}. \end{aligned}$$

Taking $c = -\log(\epsilon)$ yields

$$t_{\text{mix}}(\epsilon) \leq t(-\log(\epsilon), n) = \lceil n \log n - n \log(\epsilon) \rceil \leq O(n \log n).$$

In closing, let us briefly discuss how a better estimate

$$t_{\text{mix}} \sim \frac{1}{2} n \log n.$$

can be obtained. Note that, by the symmetry of the hypercube, the distance $|P_t^x - \pi|_{\text{TV}}$ does not depend on the starting position x. Hence, without loss of generality, suppose that $x = (0, \ldots, 0) \in \mathbb{Z}^n$. Then in bounding $|P_t^x - \pi|_{\text{TV}}$ it is enough to simply consider the number $\sum_i (X_t)_i$ of coordinates of X_t which take the value 1. Note that the limiting distribution of $\sum_i (X_t)_i$ is Binomial$(n, 1/2)$, whose fluctuations

are of order $O(\sqrt{n})$ by the Central Limit Theorem. Therefore, in order to be at TV distance at most ϵ from stationarity it is not necessary to "refresh" all coordinates (that is, all collect coupons) but rather only $n - o(\sqrt{n})$ of them. At this point, any lack of randomness due to the fact that we have yet to visit all coordinates will be overwhelmed by the "natural" randomness/fluctuations of the equilibrium distribution itself.

4.3 Proper colourings

Consider a graph $G = (V, E)$ with $|V| = n$ vertices. A *proper q-colouring* of G is an assignment $\eta \in [q]^V$ of one of q colours to each vertex in V such that neighbours are assigned different colours. That is, $\eta(u) \neq \eta(v)$ for all $(u, v) \in E$.

Let S be the set of all such colourings of G. Determining whether $S \neq \emptyset$ is a nontrivial problem in general. Indeed, deciding if there exist a proper q-colouring is an NP problem for graphs G with maximal degree $\Delta > q$. However, if $q > \Delta$ then it is clear that proper q-colourings exist. For instance, G can be properly q-coloured using a simple greedy algorithm, which colours each vertex v in turn, while avoiding any colours that have already be used in colouring any neighbours of v (of which there are at most $\Delta < q$).

Suppose that we have managed (e.g., by using a greedy algorithm) to produce one fixed q-colouring $x \in S$. Consider the Markov chain $(\eta_t)_{t \geq 0}$ with $\eta_0 = x$ on S with steps defined as follows:

1. Given η_t, select a uniformly random vertex $v_t \in V$ and a uniformly random colour $c_t \in [q]$.
2. Obtain η'_{t+1} from η_t by recolouring vertex v_t using colour c_t.
3. If η'_{t+1} is a proper colouring, then put $\eta_{t+1} = \eta'_{t+1}$, and otherwise put $\eta_{t+1} = \eta_t$.

We note that the Markov chain (η_t) is *not* irreducible in general (exercise). However, it can be shown (exercise) that if $q \geq \Delta + 2$ then it is irreducible, and we hereafter make this assumption. Since $P(\eta, \eta) > 0$, the chain is aperiodic. Finally, to see that equilibrium distribution is uniform, note that

$$P(\eta, \eta^{v,c}) = \frac{1}{qn} = P(\eta^{v,c}, \eta)$$

for every possible transition from some η to $\eta^{v,c}$, in which the colour of v changed to c. As such, we have that, as $t \in \infty$,

$$P_x^t(y) \to \frac{1}{|S|}$$

for any $x, y \in S$.

This provides us with a way of approximately sampling a uniformly random proper q-colouring of G (when $q \geq \Delta + 2$). Namely, if we run the chain for a long time, then its state will be approximately uniform on S. In order to quantify how long we will need to wait, we estimate the mixing time t_{mix} of the chain. The question of whether t_{mix} is polynomial in n when $q \geq \Delta + 2$ is an open problem. We will instead show the following.

Theorem 4.5. *Suppose that $q > 4\Delta$. Then $t_{mix} \leq O(n \log n)$.*

Proof. We will use the same "coupling method" that we used in our analysis of the hypercube in the previous section. Fix $x, y \in S$. Let $(v_t)_{t \geq 0}$ be an IID sequence of uniform random variables on V, and let $(c_t)_{t \geq 0}$ be an independent IID sequence of uniform random variables on $[q]$. Let $x, y \in S$. Consider the Markov chain $(X_t, Y_t)_{t \geq 0}$ started from (x, y) that transitions as follows. Given (X_t, Y_t), we attempt to recolour vertex v_t in the colourings X_t and Y_t using colour c_t. Call these colourings X'_{t+1} and Y'_{t+1}.

We put $X_{t+1} = X'_{t+1}$ if X'_{t+1} is a proper colouring, and put $X_{t+1} = X_t$ otherwise, and similarly for Y_{t+1}. Clearly, $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are both distributed as the original Markov chain described above.

The analysis of this coupling is more involved that it was in the case of lazy random walk on the hypercube. Recall that in that previous example, the analysis boiled down to the classical coupon collector problem. However, our current situation is not that simple. Indeed, once a vertex is "refreshed" (that is, recoloured) it is *not* necessarily the case that these vertices in the two colourings will thereafter be equal, since for example it is possible that the recolouring of v_t will succeed (that is, produce a proper colouring) in X_t but not in Y_t. As a result, we will need to estimate the probability $\mathbb{P}(X_t \neq Y_t)$ differently in our current example. To this end, consider the number

$$\delta_t = |\{v \in V \mid X_t(v) \neq Y_t(v)\}|$$

of vertices of different colour in the two colourings X_t and Y_t. Notice that, given (X_t, Y_t) (of which δ_t is a function), there are only three possibilities. Namely, either:

1. $\delta_{t+1} = \delta_t - 1$ (*good move*). If $X_t(v_t) \neq Y_t(v_t)$ and recolouring v_t using c_t is legitimate in both cases then the total number of differently coloured vertices will decrease by 1 at time $t+1$. Note that there are at least $(q - 2\Delta)\delta_t$ possibilites (v, c) for (v_t, c_t) that would result in a good move.
2. $\delta_{t+1} = \delta_t + 1$ (*bad move*). If $X_t(v_t) = Y_t(v_t)$ and some vertex in the neighbourhood of v_t is coloured c_t in exactly one of the colourings X_t or Y_t, then the total number of differently coloured vertices will increase by 1 at time $t+1$. Note that are at most $2\Delta\delta_t$ possibilites (v, c) for (v_t, c_t) that would result in a bad move.
3. $\delta_{t+1} = \delta_t$ (*neutral move*).

Hence

$$\mathbb{E}\left(\delta_{t+1} - \delta_t \mid (X_t, Y_t)\right) \leq \frac{2\Delta\delta_t}{qn} - \frac{(q-2\Delta)\delta_t}{qn} = -\frac{q-4\Delta}{qn}\delta_t,$$

and so

$$\mathbb{E}\left(\delta_{t+1} \mid (X_t, Y_t)\right) \leq (1-\alpha)\delta_t,$$

where (recalling that $q > 4\Delta$)

$$\alpha = \frac{q-4\Delta}{qn} > 0.$$

By the Law of Total Expectation,

$$\mathbb{E}(\delta_{t+1}) \leq (1-\alpha)\mathbb{E}(\delta_t).$$

Therefore, regardless of the starting point $(X_0, Y_0) = (x, y)$,

$$\mathbb{E}(\delta_t) \leq (1-\alpha)^t \delta_0 \leq e^{-\alpha t} n.$$

To conclude, we apply Lemmas 4.4 and 4.3 and Markov's Inequality, to find that

$$\begin{aligned} d(t) &\leq \max_{x,y} \mathbb{P}(X_t \neq Y_t) \\ &\leq \max_{x,y} \mathbb{P}(\delta_t \geq 1) \\ &\leq \max_{x,y} \mathbb{E}(\delta_t) \\ &\leq e^{-\alpha t} n. \end{aligned}$$

Therefore

$$t_{\text{mix}}(\epsilon) \leq \left\lceil \frac{\log n - \log(\epsilon)}{\alpha} \right\rceil.$$

In particular, $t_{\text{mix}} = O(n \log n)$, as claimed. □

Our assumption above that $q > 4\Delta$ is quite strong. Indeed, it is know that $t_{\text{mix}} = O(n \log n)$ when $q > 2\Delta$, however, the original proof of this (see Jerrum 1995) involves a complicated coupling. On the problem sheet, we will see how to prove this result using the *path coupling* technique developed in the next section.

4.4 Path coupling

The previous two examples of the "coupling method" have demonstrated the following consequence Lemmas of 4.4 and 4.3.

Corollary 4.6. *Suppose that for some Markov chain (satisfying the assumptions and notation stated at the beginning of Section 4) we have*

$$d(t) \leq \max_{x,y \in S} \mathbb{P}(X_t^{(x)}, Y_t^{(y)}),$$

for some coupling $(X_t^{(x)}, Y_t^{(y)})_{t \geq 0}$ *of* $(P_x^t, P_y^t)_{t \geq 0}$*, such that* $(X_0^{(x)}, Y_0^{(y)}) = (x, y)$ *and*

$$\mathbb{P}(X_t^{(x)} \neq Y_t^{(y)}) \leq e^{-f(t)},$$

for some increasing function f*. Then*

$$t_{mix}(\epsilon) \leq \lceil f^{-1}(-\log(\epsilon)) \rceil.$$

For instance, in the previous example, we had

$$f(t) = \alpha t - \log n, \qquad f^{-1}(s) = \frac{\log n + s}{\alpha}.$$

However, in our two examples, we have only used one part of Lemma 4.3. More specifically, we have not taken advantage of the fact (which was left as an exercise in the proof of Lemma 4.3 above) that there exists an *optimal coupling* that attains the TV distance. Let us highlight this fact once again below.

Lemma 4.7. *There is an optimal coupling* (X^*, Y^*) *of* (μ, ν) *such that* $|\mu - \nu|_{\text{TV}} = \mathbb{P}(X^* \neq Y^*)$.

As a result, there *exists* an optimal coupling $(X_t^{(x),*}, Y_t^{(y),*})$ of $(P_x^t,\, P_x^y)$ such that

$$d(t) \leq \rho(t) = \max_{x,y \in S} \mathbb{P}(X_t^{(x),*} \neq Y_t^{(y),*}).$$

Therefore, since $d(t) \geq \rho(t)/2$ by Lemma 4.4, we achieve optimal bounds for the mixing time by working with the optimal coupling.

But, of course, there is a "catch". Namely, it can be difficult to *find* the optimal coupling. One reason for the challenge is that $(X_t^{(x),*}, Y_t^{(y),*})$ is not necessarily Markovian. Indeed, observe that in our two examples, the couplings (X_t, Y_t) constructed were Markov chains on S^2. Most natural couplings will have this property, since we want the marginals to be Markovian (that is, both distributed as the original Markov chain). However, there is no reason, a priori, that the *optimal* coupling should be Markovian, and indeed, it is not always the case (see Kumar and Ramesh 1999).

The following theorem (see Bubley and Dyer 1999) summarises the "path coupling method," which sharpens the regular "coupling method."

Theorem 4.8 (Path Coupling Theorem)**.** *Consider an irreducible and aperiodic Markov chain on a finite state space S. Let (S, E_S) be a connected graph on S, with graph distance d_S. Suppose that for some $\alpha > 0$ and all edges $(x, y) \in E_S$ we have a coupling (X, Y) of (P_x^1, P_y^1) for which*

$$\mathbb{E}(d_S(X, Y)) \leq e^{-\alpha} d_S(x, y) = e^{-\alpha}. \tag{4.1}$$

Then the mixing time of the chain satisfies

$$t_{mix}(\epsilon) \leq \left\lceil \frac{1}{\alpha} \log\Big(\frac{\mathrm{diam}(S)}{\varepsilon}\Big) \right\rceil,$$

where

$$\mathrm{diam}(S) = \max_{x,y \in S} d_S(x, y).$$

We will give a sketch of the theorem. A precise proof and more general setting of the statement can be found in Chapter 14 the book by Levin, Peres, and Wilmer, on which the following is based. Let us start with a more general version of the Coupling Lemma.

Lemma 4.9. *Let μ and ν be distributions on a finite state space S, and let $d_S(x, y)$ be the graph distance on (S, E_S). Define*

$$\delta_K(\mu, \nu) := \inf\Big\{\mathbb{E}\big[d_S(X, Y)\big] : (X, Y) \text{ is a coupling of } (\mu, \nu)\Big\}. \tag{4.2}$$

There exists an optimal coupling (X^*, Y^*) such that $\delta_K(\mu, \nu) = \mathbb{E}[d_S(X^*, Y^*)] \geq d_{\mathrm{TV}}(\mu, \nu)$. In particular, when $d_S(x, y) = \mathbb{1}_{\{x \neq y\}}$ (corresponding to (S, E_S) being the complete graph), then $\delta_K(\mu, \nu) = d_{\mathrm{TV}}(\mu, \nu)$). Moreover, $\delta_K(\cdot, \cdot)$ in (4.2) defines a metric on the space of probability measures on S.

A proof of the lemma can be found in Levin, Peres, and Wilmer, and we only give a few comments here. The result when $d_S(x, y) = \mathbb{1}_{\{x \neq y\}}$ is yet another phrasing of Coupling Lemma. We note that the lower bound of d_{TV} follows immediately from the the lower bound $d_S(x, y) \geq \mathbb{1}_{\{x \neq y\}}$ for the graph distance and the Coupling Lemma.

Thus, instead of aiming to bound the total variation distance between P_x^t and P_y^t from above when proving Theorem 4.8, we may instead bound $\delta_K(P_x^t, P_y^t)$. The assumed structure of an underlying graph on S will make it easier to bound $\delta_K(P_x^t, P_y^t)$ from above. While doing so, we heavily rely on the fact that δ_K is a metric and thus satisfies the triangle inequality (we do not construct a coupling of (P_x^t, P_y^t), but only prove its existence).

We state a last lemma before giving the proof sketch of Theorem 4.8.

Lemma 4.10. *Consider the setting of Theorem 4.8. For any two probability measures μ and ν on S,*

$$\delta_K(\mu P, \nu P) \leq e^{-\alpha} \delta_K(\mu, \nu).$$

We comment on the proof after the sketch of Theorem 4.8 as it relies on similar reasoning.

Proof sketch of Theorem 4.8. The main idea is that it is enough to show the existence of couplings (X, Y) satisfying (4.1) for (P_x^1, P_y^1) with $(x, y) \in E_S$ in order to obtain the existence of couplings satisfying the inequality in (4.1) for *all* $x, y \in S$.

We first apply Lemmas 4.4 and 4.9, and then apply Lemma 4.10 $(t-1)$ times, with $\mu = P_x^s$ and $\nu = P_y^s$

for $s \in \{t-1, t-2, ..., 1\}$. This yields,

$$\begin{aligned} d(t) &\leq \max_{x,y \in S} |P_x^t - P_y^t|_{\mathrm{TV}} \\ &\leq \max_{x,y \in S} \delta_K(P_x^t, P_y^t) \\ &= \max_{x,y} \delta_K(P_x^{t-1}P, P_y^{t-1}P) \\ &\leq \max_{x,y} \delta_K(P_x^{t-1}, P_y^{t-1})e^{-\alpha} \\ &\vdots \\ &\leq \max_{x,y} \delta_K(P_x^1, P_y^1)e^{-\alpha(t-1)}. \end{aligned} \tag{4.3}$$

The above step illustrates a contraction between P_x^t and P_y^t over time. In the next step, we will show that there exists a coupling (X, Y) of (P_x^1, P_y^1) such that $\mathbb{E}_{x,y}[d_S(X, Y)] \leq d_S(x, y)e^{-\alpha}$, which applies for x and y at arbitrary distance from each other in (S, E_S). We will show the existence of this coupling (without explicitly constructing it!). Let $x, y \in S$ be given and consider a shortest path $x = z_0, z_1, \ldots, z_\ell = y$ in the graph (S, E_S) of length $\ell = d_S(x, y)$. Since δ_K is a metric by Lemma 4.9, it follows by the triangle inequality that

$$\delta_K(P_x^1, P_y^1) \leq \sum_{i=0}^{\ell-1} \delta_K(P_{z_i}^1, P_{z_{i+1}}^1). \tag{4.4}$$

We assumed that the contraction property 4.1 holds for all neighbouring (z_i, z_{i+1}), and therefore we can apply it to all the terms in the sum above. Combining this with (4.3) we obtain

$$d(t) \leq e^{-\alpha(t-1)} \max_{x,y} \sum_{i=0}^{\ell-1} e^{-\alpha} = e^{-\alpha t} \max_{x,y} d_S(x, y) = \mathrm{diam}(S)e^{-\alpha t}.$$

For a bound on $t_{\mathrm{mix}}(\varepsilon)$, we aim to find the smallest t such that the right-hand side is at most ε. Solving this bound yields that

$$t_{\mathrm{mix}}(\varepsilon) \leq \left\lceil \frac{1}{\alpha} \log\Big(\frac{\mathrm{diam}(S)}{\varepsilon}\Big) \right\rceil. \qquad \square$$

Let us comment briefly on the proof of Lemma 4.10 (see for instance Levin, Peres, and Wilmer): one first proves (4.4) using the triangle inequality. Then using the existence of optimal couplings in Lemma 4.9 of (μ, ν) and (P_x^1, P_y^1), one can use these couplings to construct a coupling (Z_μ, Z_ν) of $(\mu P, \nu P)$ for which $\mathbb{E}[d_S(Z_\mu, Z_\nu)] \leq \delta_K(\mu, \nu)e^{-\alpha}$, so that the optimal coupling (Z_μ^*, Z_ν^*) (and thus $\delta_K(\mu P, \nu P)$ must satisfy $\delta_K(\mu P, \nu P) = \mathbb{E}[d_S(Z_\mu^*, Z_\nu^*)] \leq \delta_K(\mu, \nu)e^{-\alpha}$).

Theorem 4.8 can be used to improve on our previous Theorem 4.5. Roughly speaking, using the method of path coupling, we can consider a generic pair (x, y) of proper q-colourings that differ only on a single vertex, and try to optimise a coupling of (P_x^1, P_y^1) to have as few "bad moves" as possible. On the problem sheet, you are asked to come up with such a path coupling to obtain the following.

Theorem 4.11. *Suppose that $q > 2\Delta$. Then the proper colouring Markov chain from Section 4.3 satisifes $t_{mix} \leq O(n \log n)$.*

This theorem was first proved by Jerrum 1995 without relying on path coupling, which was introduced a couple of years later.

The path coupling approach works naturally in cases where the state space $S = \Lambda^X$, for some finite sets Λ and X. Taking E_S to be the set of all (x, y) such that, for some $j \in X$, we have that $x(j) \neq y(j)$ and $x(i) = y(i)$ for all $i \neq j$ corresponds to the so-called *Hamming distance* d_S on S. Then if (as is often the case) a single step of the chain we are considering consists of updating the value of a single coordinate

of the current configuration, then pairs (x, y) of states that have nonzero transition probability in the chain are adjacent in the graph (S, E_S). We will see an example of such a situation in the next section.

However, the utility of the path coupling method is much more far-reaching. Indeed, part of the strength of this method comes from the fact that the edges in E_S need not have anything to do with the relation of adjacency given by the Markov chain. There is a wide range of distances d_S induced by connected graphs on S, and one can generalise Theorem 4.8 accordingly. For example, a natural choice in some contexts is the distance induced by a (weighted) graph whose edges have lengths at least 1.

4.5 Glauber dynamics on the Ising model

We conclude these notes with an application of the path coupling method in the context of statistical mechanics. Specifically, we will study the so-called *Glauber dynamics* on the Ising model, in which the spins in the Ising model are resampled according to a certain Markov chain on the set of all possible configurations of the model.

Consider the finite-volume Gibbs measure $\mu = \mu_{G;\beta,0}$ for the Ising model on $\Omega = \Omega_G$ from Chapter 3, where $G = (V, E)$ is a finite subgraph of $\mathbb{Z}^d$. For convenience, we relabel $V = [n]$. For simplicity, we are considering zero external field $h = 0$ and no boundary conditions η (but both can be incorporated into the approach below, if desired). We are interested in sampling a random configuration from Ω distributed according to μ. Note that, given that $|\Omega| = 2^n$, any sampling scheme that involves some sort of straightforward enumeration of the set of possible configurations becomes impractical, for even rather small graphs.

Our approach is to design a Markov chain on Ω with equilibrium μ. Given this, we can then approximately sample from μ by running the chain for a suitably large (depending on t_{mix}) amount of time. The chain is known as *Glauber dynamics,* and is a simple example of a Markov chain Monte Carlo (MCMC) algorithm. Given that the chain is in a state x at time t, we obtain its state at time $t + 1$ by selecting a uniformly random vertex i of G and resetting the spin x_i to ± 1 with probability $\mu(\sigma_i = \pm 1 \mid \sigma_j = x_j \ \forall j \neq i)$. In other words, we "resample" the spin at position i according the conditional probability with which it would have a given spin, given the rest of the spin configuration at positions $j \neq i$. Therefore, the transition probabilities for this chain are

$$P(x, y) = \begin{cases} n^{-1}\mu(\sigma_i = +1 \mid \sigma_j = x_j \ \forall j \neq i) & \text{if } y_i = +1 \text{ and } y_j = x_j \ \forall j \neq i \\ n^{-1}\mu(\sigma_i = -1 \mid \sigma_j = x_j \ \forall j \neq i) & \text{if } y_i = -1 \text{ and } y_j = x_j \ \forall j \neq i \\ 0 & \text{otherwise.} \end{cases}$$

Note that, by the spatial Markov property (Theorem 3.5),

$$\mu(\sigma_i = \pm 1 \mid \sigma_j = x_j \ \forall j \neq i) = \mu^x_{\{i\};\beta,0}(\sigma_i = \pm 1) = \frac{1}{Z^x_{\{i\};\beta,0}} \exp\left(\pm\beta \sum_{j \sim i} x_j\right).$$

However, $\Omega^x_{\{i\};\beta,0}$ contains only has two configurations. Namely, the ones in which $\sigma_i = \pm 1$. Hence,

$$Z^x_{\{i\};\beta,0} = \exp\left(\beta \sum_{j \sim i} x_j\right) + \exp\left(-\beta \sum_{j \sim i} x_j\right).$$

Therefore,

$$\mu(\sigma_i = \pm 1 \mid \sigma_j = x_j \ \forall j \neq i) = \frac{e^{\pm\beta C(x,i)}}{e^{\beta C(x,i)} + e^{-\beta C(x,i)}},$$

where

$$C(x, i) = \sum_{j \sim i} x_j.$$

Let us note that $|C(x,i)| \le 2d$, since each vertex in $G \subset \mathbb{Z}^d$ has at most $2d$ neighbours.

Clearly, the Glauber dynamics chain is irreducible and aperiodic. Moreover, the Gibbs measure μ is its stationary distribution, as can be checked using the detailed balance equations. For this reason, such a chain is sometimes called a *Gibbs sampler*. We conclude this course with a proof that the Glauber dynamics are fast-mixing at high enough temperatures.

Theorem 4.12. *There is a $\beta' > 0$, depending only on d, such that for all $\beta \in (0, \beta']$ the mixing time of the Glauber dynamics satisfies $t_{mix} = O(n \log n)$.*

Proof. The proof is by the path coupling method Theorem 4.8, with the Hamming distance

$$d(x,y) = |\{i \in V : x_i \neq y_i\}|$$

on Ω. For a configuration $z \in \Omega$ and vertex $i \in V$, let $z^{i,\sigma}$ denote the configuration obtained by replacing the spin at location i in z with σ. Thus, for $x, y \in \{0,1\}^n$ at distance 1, there exists an index $i \in [n]$ such that $y = x^{i,-x_i}$. Pick such x and y, and let (X,Y) denote a pair of random variables after running one time step of the Markov chains starting from x and y under a coupling that we construct next, i.e., (X,Y) is a coupling of (P^1_x, P^1_y). First, let $J = \mathrm{Unif}\{1,...,n\}$ denote the index of the spin that we resample. We distinguish three cases for J:

1. If $J = i$, we resample the spins at the only vertex in which x and y disagree. Since the transition probabilities depend only on the spins of neighbours of i (and not on the current state of i), we can couple the Markov chains such that the spin of i agrees in X and Y.
2. If $J \neq i$ and J is not a neighbour of i, then the spins of the neighbours of vertex J agree in x and y, so again we can couple the Markov chains such that the spins at index J agree in X and Y.
3. The only remaining option is when $J \sim i$. Let $U \sim \mathrm{Unif}[0,1]$, and put

$$X = \begin{cases} x^{J,+1} & \text{if } U \le \frac{e^{\beta C(x,J)}}{e^{\beta C(x,J)} + e^{-\beta C(x,J)}} =: q_{x,J}, \\ x^{J,-1} & \text{otherwise,} \end{cases} \qquad Y = \begin{cases} y^{J,+1} & \text{if } U \le \frac{e^{\beta C(y,J)}}{e^{\beta C(y,J)} + e^{-\beta C(y,J)}} =: q_{y,J}, \\ y^{J,-1} & \text{otherwise.} \end{cases}$$

Note that this indeed constructs a coupling (X,Y). Let us now analyse $d(X,Y)$. In case 1 (when $J = i$, which happens with probability $1/n$), the distance drops to 0. In case 2 (there are $n - \deg(i) - 1$ vertices that are not equal to i and not a neighbour of i), the only spin-disagreement between X and Y is at i. In the last case, one extra disagreement between X and Y is created if and only if $U \in [\min(q_{x,J}, q_{y,J}), \max(q_{x,J}, q_{y,J})]$. Combining these observations, we obtain

$$\begin{aligned} \mathbb{E}[d(X,Y)] &= \frac{1}{n} \cdot 0 + \frac{n - \deg(i) - 1}{n} \cdot 1 + \frac{1}{n} \sum_{j \sim i} \big((1 - |q_{x,j} - q_{y,j}|) \cdot 1 + |q_{x,j} - q_{y,j}| \cdot 2 \big) \\ &= 1 - \frac{1}{n} + \frac{1}{n} \sum_{j \sim i} |q_{x,j} - q_{y,j}| = 1 - \frac{1}{n} + \frac{1}{n} \sum_{j \sim i} \left| \frac{e^{\beta C(x,j)}}{e^{\beta C(x,j)} + e^{-\beta C(x,j)}} - \frac{e^{\beta C(y,j)}}{e^{\beta C(y,j)} + e^{-\beta C(y,j)}} \right|. \end{aligned}$$

Note that $|C(x,j) - C(y,j)| = 2$ for $j \sim i$. Moreover, $|C(\cdot,\cdot)| \le 2d$. By a series expansion of $\exp(\cdot)$, each of the at most $2d$ summands decays as $\beta(1 + o(1))$ as $\beta \downarrow 0$. Therefore, for each $\delta > 0$, there exists $\beta' > 0$ such that the sum on the right-hand side is at most δ for all $\beta \in (0, \beta']$. This concludes that $\mathbb{E}(d(X,Y)) \le 1 - \alpha$, where $\alpha = (1 - \delta)/n > 0$. Therefore, by Theorem 4.8,

$$t_{\mathrm{mix}} \le O\big((\log n)/\alpha\big) = O(n \log n)$$

as claimed. □