Chapter 13
Continuous State, Infinite State and Random Environment

David Aldous
Department of Statistics
University of California
Berkeley, CA 94720

James Allen Fill
Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, MD 21218-2692

June 23, 2001

Reference to other Chapters are the following versions:

Chap 2: 7/20/99

Chap 3: 9/2/94

Chap 4: 10/11/94

Chap 4-3: 10/11/00

Chap 5: 4/23/96

Chap 6: 10/31/94

Chap 7: 1/31/94

Chap 7-1: undated

Chap 9: 9/1/99

Chap MCMC: 1/8/01

1 Continuous state space

We have said several times that the theory in this book is fundamentally a theory of *inequalities*. "Universal" or "a priori" inequalities for reversible chains on finite state space, such as those in Chapter 4, should extend unchanged to the continuous space setting. Giving proofs of this, or giving the rigorous setup for continuous-space chains, is outside the scope of our intermediate-level treatment. Instead we just mention a few specific processes which parallel or give insight into topics treated earlier.

1.1 One-dimensional Brownian motion and variants

Let $(B_t, 0 \le t < \infty)$ be one-dimensional *standard Brownian motion* (BM). Mentally picture a particle moving along an erratic continuous random trajectory. Briefly, for $s < t$ the increment $B_t - B_s$ has Normal$(0, t-s)$ distribution, and for non-overlapping intervals (s_i, t_i) the increments $B_{t_i} - B_{s_i}$ are independent. See Norris [46] section 4.4, Karlin and Taylor [31] Chapter 7, or Durrett [20] Chapter 7 for successively more detailed introductions. One can do explicit calculations, directly in the continuous setting, of distributions of many random quantities associated with BM. A particular calculation we need ([20] equation 7.8.12) is

$$G(t) := P\left(\sup_{0 \le s \le t} |B_s| < 1\right) = \frac{4}{\pi} \sum_{m=0}^{\infty} \frac{(-1)^m}{2m+1} \exp(-(2m+1)^2 \pi^2 t/8) \quad (1)$$

where

$$G^{-1}(1/e) = 1.006. \quad (2)$$

One can also regard BM as a limit of rescaled random walk, a result which generalizes the classical central limit theorem. If $(X_m, m = 0, 1, 2, \ldots)$ is simple symmetric random walk on Z, then the central limit theorem implies $m^{-1/2} X_m \stackrel{d}{\to} B_1$ and the generalized result is

$$(m^{-1/2} X_{\lfloor mt \rfloor}, 0 \le t < \infty) \stackrel{d}{\to} (B_t, 0 \le t < \infty) \quad (3)$$

where the convergence here is *weak convergence of processes* (see e.g. Ethier and Kurtz [22] for detailed treatment). For more general random flights on Z, that is $X_m = \sum_{j=1}^m \xi_j$ with $\xi_1, \xi_2, \ldots$ independent and $E\xi = 0$ and $\operatorname{var} \xi = \sigma^2 < \infty$, we have *Donsker's theorem* ([20] Theorem 7.6.6)

$$(m^{-1/2} X_{\lfloor mt \rfloor}, 0 \le t < \infty) \stackrel{d}{\to} (\sigma B_t, 0 \le t < \infty). \quad (4)$$

Many asymptotic results for random walk on the integers or on the n-cycle or on the n-path, and their d-dimensional counterparts, can be explained in terms of Brownian motion or its variants. The variants of interest to us take values in compact sets and have uniform stationary distributions.

Brownian motion on the circle can be defined by

$$B_t^\circ := B_t \bmod 1$$

and then random walk $(X^{(n)}_m, m = 0, 1, 2, \ldots)$ on the n-cycle $\{0, 1, 2, \ldots, n-1\}$ satisfies, by (3),

$$n^{-1}(X^{(n)}_{\lfloor n^2 t \rfloor}, 0 \le t < \infty) \ \stackrel{d}{\to} \ (B^\circ_t, 0 \le t < \infty) \text{ as } n \to \infty. \tag{5}$$

The process B° has eigenvalues $\{2\pi^2 j^2, \quad 0 \le j < \infty\}$ with eigenfunction $\equiv 1$ for $j = 0$ and two eigenfunctions $\cos(2\pi j x)$ and $\sin(2\pi j x)$ for $j \ge 1$. In particular the relaxation time is

$$\tau_2 = \tfrac{1}{2\pi^2}.$$

The result for random walk on the n-cycle (Chapter 5 Example 7)

$$\tau_2 \sim \tfrac{n^2}{2\pi^2} \text{ as } n \to \infty$$

can therefore be viewed as a consequence of the n^2 time-rescaling in (5) which takes random walk on the n-cycle to Brownian motion on the circle. This argument is a prototype for the *weak convergence paradigm*: proving size-asymptotic results for discrete structures in terms of some limiting continuous structure.

Variation distance can be studied via coupling. Construct two Brownian motions on R started from 0 and $x > 0$ as follows. Let $B^{(1)}$ be standard Brownian motion, and let

$$T_{x/2} := \inf\{t : B^{(1)}_t = x/2\}.$$

Then $T_{x/2} < \infty$ a.s. and we can define $B^{(2)}$ by

$$\begin{aligned} B^{(2)}_t &= x - B^{(1)}_t, \quad 0 \le t \le T_{x/2} \\ &= B^{(1)}_t, \quad T_{x/2} \le t < \infty. \end{aligned}$$

That is, the segment of $B^{(2)}$ over $0 \le t \le T_{x/2}$ is the image of the corresponding segment of $B^{(1)}$ under the reflection which takes 0 to x. It is easy to see that $B^{(2)}$ is indeed Brownian motion started at x. This is the *reflection coupling* for Brownian motion. We shall study analogous couplings for variant processes. Given Brownian motion on the circle $B^{\circ 1}$ started at 0, we can construct another Brownian motion on the circle $B^{\circ 2}$ started at $0 < x \le 1/2$ via

$$\begin{aligned} B^{\circ 2}_t &= x - B^{\circ 1}_t \bmod 1, \quad 0 \le t \le T_{\{\frac{x}{2}, \frac{x}{2}+\frac{1}{2}\}} \\ &= B^{\circ 1}_t, \quad T_{\{\frac{x}{2}, \frac{x}{2}+\frac{1}{2}\}} \le t < \infty \end{aligned}$$

where

$$T_{\{\frac{x}{2},\frac{x}{2}+\frac{1}{2}\}} := \inf\{t : B^{\circ 1}_t = \tfrac{x}{2} \text{ or } \tfrac{x}{2}+\tfrac{1}{2}\}.$$

Again, the segment of $B^{\circ 2}$ over $0 \leq t \leq T_{\{\frac{x}{2},\frac{x}{2}+\frac{1}{2}\}}$ is the image of the corresponding segment of $B^{\circ 1}$ under the reflection of the circle which takes 0 to x, so we call it the reflection coupling for Brownian motion on the circle. Because sample paths cannot cross without meeting, it is easy to see that the general coupling inequality (Chapter 4-3 section 1.1) becomes an equality:

$$||P_0(B^\circ_t \in \cdot) - P_x(B^\circ_t \in \cdot)|| = P(T_{\{\frac{x}{2},\frac{x}{2}+\frac{1}{2}\}} > t).$$

The worst starting point is $x = 1/2$, and the hitting time in question can be written as the hitting time $T_{\{-1/4,1/4\}}$ for standard Brownian motion, so

$$\bar{d}(t) = P(T_{\{-1/4,1/4\}} > t) = G(16t) \tag{6}$$

by *Brownian scaling*, that is the property

$$(B_{c^2t}, 0 \leq t < \infty) \ \stackrel{d}{=}\ (cB(t), 0 \leq t < \infty). \tag{7}$$

See the Notes for an alternative formula. Thus for Brownian motion on the circle

$$\tau_1 = \tfrac{1}{16}G^{-1}(1/e) = 0.063. \tag{8}$$

If simple random walk is replaced by aperiodic random flight with step variance σ^2 then the asymptotic values of τ_2 and τ_1 are replaced by τ_2/σ^2 and τ_1/σ^2; this may be deduced using the local central limit theorem ([20] Theorem 2.5.2).

Reflecting Brownian motion $\bar{B}$ on the interval $[0,1]$ is very similar. Intuitively, imagine that upon hitting an endpoint 0 or 1 the particle is instantaneously inserted an infinitesimal distance into the interval. Formally one can construct $\bar{B}_t$ as $\bar{B}_t := \phi(B_t)$ for the concertina map

$$\phi(2j+x) = x, \quad \phi(2j+1+x) = 1-x; \quad 0 \leq x \leq 1,\ j = \ldots -2,-1,0,1,2,\ldots.$$

The process $\bar{B}$ has eigenvalues $\{\pi^2 j^2/2, \quad 0 \leq j < \infty\}$ with eigenfunctions $\cos(\pi j x)$. In particular the relaxation time is

$$\tau_2 = \tfrac{2}{\pi^2}.$$

The result for random walk on the n-path (Chapter 5 Example 8)

$$\tau_2 \sim \tfrac{2n^2}{\pi^2} \text{ as } n \to \infty$$

is another instance of the weak convergence paradigm, a consequence of the n^2 time-rescaling which takes random walk on the n-path to reflecting Brownian motion on the interval. The variation distance function $\bar{d}(t)$ for $\bar{B}$ can be expressed in terms of the corresponding quantity (write as $d^\circ(t)$) for B°. Briefly, it is easy to check

$$(\bar{B}_t, 0 \leq t < \infty) \stackrel{d}{=} (2\min(B^\circ_{t/4}, 1 - B^\circ_{t/4}), 0 \leq t < \infty)$$

and then to deduce $\bar{d}(t) = d^\circ(t/4)$. Then using (8)

$$\tau_1 = \tfrac{1}{4} G^{-1}(1/e) = 0.252. \qquad (9)$$

1.2 d-dimensional Brownian motion

Standard d-dimensional Brownian motion can be written as

$$\mathbf{B}_t = (B^{(1)}_t, \ldots, B^{(d)}_t)$$

where the component processes $(B^{(i)}_t, i = 1, \ldots, d)$ are independent one-dimensional standard Brownian motions. A useful property of $\mathbf{B}$ is *isotropy*: its distribution is invariant under rotations of R^d. In approximating simple random walk $(X_m, m = 0, 1, 2, \ldots)$ on Z^d one needs to be a little careful with scaling constants. The analog of (3) is

$$(m^{-1/2} X_{\lfloor mt \rfloor}, 0 \leq t < \infty) \stackrel{d}{\to} (d^{-1/2} \mathbf{B}_t, 0 \leq t < \infty) \qquad (10)$$

where the factor $d^{-1/2}$ arises because the components of the random walk have variance $1/d$ — see (4). Analogous to (5), random walk $(X^{(n)}_m, m = 0, 1, 2, \ldots)$ on the discrete torus Z^d_n converges to Brownian motion $\mathbf{B}^\circ$ on the continuous torus $[0, 1)^d$:

$$n^{-1}(X^{(n)}_{\lfloor n^2 t \rfloor}, 0 \leq t < \infty) \stackrel{d}{\to} (d^{-1/2} \mathbf{B}^\circ_t, 0 \leq t < \infty) \text{ as } n \to \infty. \qquad (11)$$

1.3 Brownian motion in a convex set

Fix a convex polyhedron $K \subset R^d$. One can define reflecting Brownian motion in K; heuristically, when the particle hits a face it is replaced an infinitesimal distance inside K, orthogonal to the face. As in the previous examples, the stationary distribution is uniform on K. We will outline a proof of

Proposition 1 *For Brownian motion* $\mathbf{B}$ *in a convex polyhedron* K *which is a subset of the ball of radius* r,
(i) $\tau_1 \leq G^{-1}(1/e)\ r^2$
(ii) $\tau_2 \leq 8\pi^{-2}r^2$.

Proof. By the d-dimensional version of Brownian scaling (7) we can reduce to the case $r = 1$. The essential fact is

Lemma 2 *Let* $\bar{B}_t$ *be reflecting Brownian motion on* $[0, 1]$ *started at* 1, *and let* T_0 *be its hitting time on* 0. *Versions* $\mathbf{B}^{(1)}, \mathbf{B}^{(2)}$ *of Brownian motion in* K *started from arbitrary points of* K *can be constructed jointly with* $\bar{B}$ *such that*

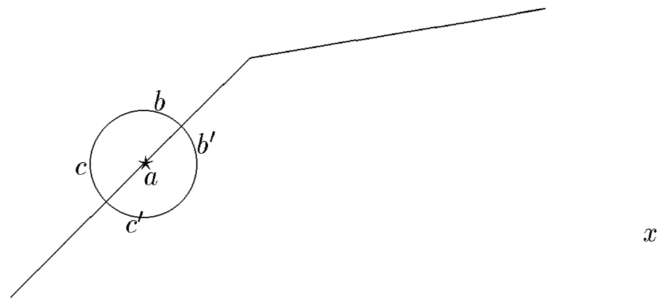
$$|\mathbf{B}_t^{(1)} - \mathbf{B}_t^{(2)}| \leq 2\bar{B}_{\min(t,T_0)}, \quad 0 \leq t < \infty. \tag{12}$$

Granted this fact, $\bar{d}(t)$ for Brownian motion on K satisfies

$$\bar{d}(t) = \max_{\text{startingpoints}} P(\mathbf{B}_t^{(1)} \neq \mathbf{B}_t^{(2)}) \leq P(T_0 > t) = G(t)$$

where the final equality holds because T_0 has the same distribution as the time for Brownian motion stared at 1 to exit the interval $(0, 2)$. This establishes (i) for $r = 1$. Then from the $t \to \infty$ asymptotics of $G(t)$ in (1) we have $\bar{d}(t) = O(\exp(-\pi^2 t/8))$, implying $\tau_2 \leq 8/\pi^2$ by Lemma **??** and establishing (ii).

Sketch proof of Lemma. Details require familiarity with stochastic calculus, but this outline provides the idea. For two Brownian motions in R^d started from $(0, 0, \ldots, 0)$ and from $(x, 0, \ldots, 0)$, one can define the reflection coupling by making the first coordinates evolve as the one-dimensional reflection coupling, and making the other coordinate processes be identical in the two motions. Use isotropy to entend the definition of reflection coupling to arbitrary starting points. Note that the distance between the processes evolves as 2 times one-dimensional Brownian motion, until they meet. The desired joint distribution of $((\mathbf{B}_t^{(1)}, \mathbf{B}_t^{(2)}), \quad 0 \leq t < \infty)$ is obtained by specifying that while both processes are in the interior of K, they evolve as the reflection coupling (and each process reflects orthogonally at faces). As the figure illustrates, the effect of reflection can only be to decrease distance between the two Brownian particles.



For a motion hitting the boundary at a, if the unreflected process is at b or c an infinitesimal time later then the reflected process is at b' or c'. By convexity, for any $x \in K$ we have $|b' - x| \leq |b - x|$; so reflection can only decrease distance between coupled particles. To argue the inequality carefully, let α be the vector normal to the face. The projection $\mathcal{P}_\alpha$ satisfies $|\mathcal{P}_\alpha(b' - x)| \leq |\mathcal{P}_\alpha(b - x)|$. Further, $b - b' \perp \alpha$, implying $P_{\alpha^\perp}(b' - x) = P_{\alpha^\perp}(b - x)$. Therefore by Pythagoras $|b' - x| \leq |b - x|$.

We can therefore write, in stochastic calculus notation,

$$d|\mathbf{B}_t^{(1)} - \mathbf{B}_t^{(2)}| = d(2B_t) - dA_t$$

where B_t is a one-dimensional Brownian motion and A_t is an increasing process (representing the contribution from reflections off faces) which increases only when one process is at a face. But we can construct reflecting Brownian motion $\bar{B}$ in terms of the same underlying B_t by

$$d(2\bar{B}_t) = d(2B_t) - dC_t$$

where C_t (representing the contribution from reflections off the endpoint 1) is increasing until T_0. At time 0 we have (because $r = 1$)

$$|\mathbf{B}_0^{(1)} - \mathbf{B}_0^{(2)}| \leq 2 = 2\bar{B}_0.$$

We have shown

$$d(|\mathbf{B}_t^{(1)} - \mathbf{B}_t^{(2)}| - 2\bar{B}_t) = -dA_t + dC_t.$$

If the desired inequality (12 fails then it fails at some first time t, which can only be a time when dC_t is increasing, that is when $\bar{B}_t = 1$, at which times the inequality holds a priori. $\blacksquare$.

Proposition 1 suggests an approach to the algorithmic question of simulating a uniform random point in a convex set $K \subset R^d$ where d is large, discussed in Chapter 9 section 5.1. If we could simulate the discrete-time chain defined as reflecting Brownian motion $\mathbf{B}$ on K examined at time intervals of h^2/d for some small h (so that the length of a typical step is of order $\sqrt{(h^2/d) \times d} = h$), then Proposition 1 implies that $O(d/h^2)$ steps are enough to approach the stationary distribution. Since the convex set is available only via an oracle, one can attempt to do the simulation via acceptance/rejection. That is, from $\mathbf{x}$ we propose a move to $\mathbf{x}' = \mathbf{x} + \sqrt{h^2/d}\ Z$ where Z has standard d-variate Normal distribution, and accept the move iff $\mathbf{x}' \in K$. While this leads to a plausible heuristic argument, the rigorous difficulty is that it is not clear how close an acceptance/rejection step is to the true step of reflecting Brownian motion. No rigorous argument based directly on Brownian motion has yet been found, though the work of Bubley et al [12] on coupling of random walks has elements in common with reflection coupling.

1.4 Discrete-time chains: an example on the simplex

Discrete-time, continuous-space chains arise in many settings, in particular (Chapter MCMC) in Markov Chain Monte Carlo sampling from a target distribution on R^d. As discussed in that chapter, estimating mixing times for such chains with general target distributions is extremely difficult. The techniques in this book are more directly applicable to chains with (roughly) uniform stationary distribution. The next example is intended to give the flavor of how techniques might be adapted to the continuous setting: we will work through the details of a coupling argument.

Example 3 *A random walk on the simplex.*

Fix d and consider the simplex $\Delta = \{\mathbf{x} = (x_1, \ldots, x_d) : x_i \geq 0, \sum_i x_i = 1\}$. Consider the discrete-time Markov chain $(\mathbf{X}(t), t = 0, 1, 2, \ldots)$ on Δ with steps:

from state $\mathbf{x}$, pick 2 distinct coordinates $\{i, j\}$ uniformly at random, and replace the 2 entries $\{x_i, x_j\}$ by $\{U, x_i + x_j - U\}$ where U is uniform on $(0, x_i + x_j)$.

The stationary distribution π is the uniform distribution on Δ. We will show that the mixing time τ_1 satisfies

$$\tau_1 = O(d^2 \log d) \text{ as } d \to \infty. \tag{13}$$

The process is somewhat reminiscent of card shuffling by random transpositions (Chapter 7 Example 18), so by analogy with that example we expect that in fact $\tau_1 = \Theta(d \log d)$. What we show here is that the coupling analysis of that example (Chapter 4-3 section 1.7) extends fairly easily to the present example.

As a preliminary, let us specify two distinct couplings (A, B) of the uniform$(0, a)$ and the uniform$(0, b)$ distributions. In the *scaling coupling* we take $(A, B) = (aU, bU)$ for U with uniform$(0, 1)$ distribution. In the *greedy coupling* we make $P(A = B)$ have its maximal value, which is $\min(a, b)/\max(a, b)$, and we say the coupling *works* if $A = B$.

Fix $\mathbf{x}(0) \in \Delta$. We now specify a coupling $(\mathbf{X}(t), \mathbf{Y}(t))$ of the chains started with $\mathbf{X}(0) = \mathbf{x}(0)$ and with $\mathbf{Y}(0)$ having the uniform distribution. (This is an atypical couplig argument, in that it matters that one version is the stationary version).

> From state $(\mathbf{x}, \mathbf{y})$, choose the same random pair $\{i, j\}$ for each process, and link the new values x_i' and y_i' (which are uniform on different intervals) via the scaling coupling for the first $t_1 = 3d^2 \log d$ steps, then via the greedy coupling for the next $t_2 = Cd^2$ steps.

We shall show that, for any fixed constant $C > 0$,

$$P(\mathbf{X}(t_1 + t_2) = \mathbf{Y}(t_1 + t_2)) \geq 1 - C^{-1} - o(1) \text{ as } d \to \infty \tag{14}$$

establishing (13).

Consider the effect on l_1 distance $||\mathbf{x} - \mathbf{y}|| := \sum_i |x_i - y_i|$ of a step of the scaling coupling using coordinates $\{i, j\}$. The change is

$$|U(x_i+x_j)-U(y_i+y_j)|+|(1-U)(x_i+x_j)-(1-U)(y_i+y_j)|-|x_i-y_i|-|x_j-y_j|$$

$$= |(x_i + x_j) - (y_i + y_j)| - |x_i - y_i| - |x_j - y_j|$$

$$= \begin{cases} 0 & \text{if sgn } (x_i - y_i) = \text{sgn } (x_j - y_j) \\ -2\min(|x_i - y_i|, |x_j - y_j|) & \text{if not .} \end{cases}$$

Thus

$$E_{(\mathbf{x},\mathbf{y})}\left(||\mathbf{X}(1) - \mathbf{Y}(1)|| - ||\mathbf{x} - \mathbf{y}||\right)$$

$$= \frac{-2}{d(d-1)} \sum_i \sum_{j \neq i} \min(|x_i - y_i|, |x_j - y_j|) 1_{(\text{sgn } (x_i - y_i) \neq \text{sgn } (x_j - y_j))}$$

$$= \frac{-4}{d(d-1)} \sum_{i \in A} \sum_{j \in B} \min(c_i, d_j)$$

(where $c_i := x_i - y_i$ on $A := \{i : x_i > y_i\}$; $d_j := y_j - x_j$ on $B := \{j : y_j > x_j\}$)

$$\begin{aligned} &= \frac{-4}{d(d-1)} \sum_{i\in A} \sum_{j \in B} \frac{c_i d_j}{\max(c_i, d_j)} \\ &\le \frac{-4}{d(d-1)} \sum_{i\in A} \sum_{j \in B} \frac{c_i d_j}{||\mathbf{x}-\mathbf{y}||/2} \\ &= \frac{-2}{d(d-1)} ||\mathbf{x}-\mathbf{y}|| \end{aligned}$$

because $\sum_{i\in A} c_i = \sum_{j\in B} d_j = ||\mathbf{x}-\mathbf{y}||/2$. So

$$E_{(\mathbf{x},\mathbf{y})}||\mathbf{X}(1)-\mathbf{Y}(1)|| \le \left(1 - \frac{2}{d(d-1)}\right)||\mathbf{x}-\mathbf{y}||.$$

Because $||\mathbf{X}(0)-\mathbf{Y}(0)|| \le 2$, it follows that after t steps using the scaling coupling,

$$E||\mathbf{X}(t)-\mathbf{Y}(t)|| \le 2\left(1 - \frac{2}{d(d-1)}\right)^t.$$

So by taking $t_1 \sim 3d^2 \log d$, after t_1 steps we have

$$P(||\mathbf{X}(t_1)-\mathbf{Y}(t_1)|| \le d^{-5}) = 1 - o(1). \tag{15}$$

Now consider the greedy coupling. If a step works, the l_1 distance $||\mathbf{X}(t)-\mathbf{Y}(t)||$ cannot increase. The chance that a step from $(\mathbf{x},\mathbf{y})$ involving coordinates $\{i,j\}$ works is

$$\begin{aligned} \frac{\min(x_i+x_j, y_i+y_j)}{\max(x_i+x_j, y_i+y_j)} &\ge \frac{y_i+y_j-||\mathbf{x}-\mathbf{y}||}{\max(x_i+x_j, y_i+y_j)} \\ &\ge \frac{y_i+y_j-||\mathbf{x}-\mathbf{y}||}{y_i+y_j+||\mathbf{x}-\mathbf{y}||} \\ &\ge \frac{y_i+y_j-2||\mathbf{x}-\mathbf{y}||}{y_i+y_j} \\ &\ge 1 - \frac{||\mathbf{x}-\mathbf{y}||}{\min(y_i,y_j)}. \end{aligned}$$

So unconditionally

$$P_{(\mathbf{x},\mathbf{y})}(\text{greedy coupling works on first step}) \ge 1 - \frac{||\mathbf{x}-\mathbf{y}||}{\min_k y_k}. \tag{16}$$

Now the uniform distribution $(Y_1^{(d)}, \ldots, Y_d^{(d)})$ on the simplex has the property (use [20] Exercise 2.6.10 and the fact that the uniform distribution on the simplex is the joint distribution of spacings between $d-1$ uniform$(0,1)$ variables and the endpoint 1)

$$\text{if constants } a_d > 0 \text{ satisfy } da_d \to 0 \text{ then } P(Y_1^{(d)} \le a_d) \sim da_d.$$

Since $(\mathbf{Y}(t))$ is the stationary chain and $Y_i^{(d)} \stackrel{d}{=} Y_1^{(d)}$,

$$P(\min_{1\le k\le d} Y_k(t) \le d^{-4.5} \text{ for some } t_1 < t \le t_1 + t_2) \le t_2 d P(Y_1^{(d)} < d^{-4.5})$$

and since $t_2 = O(d^2)$ this bound is $o(1)$. In other words

$$P(\min_{1\le k\le d} Y_k(t) \ge d^{-4.5} \text{ for all } t_1 < t \le t_1 + t_2) = 1 - o(1) \text{ as } d \to \infty.$$

Combining this with (15,16) and the non-increase of l_1 distance, we deduce

$$P(\text{greedy coupling works for all} t_1 < t \le t_1 + t_2) = 1 - o(1). \tag{17}$$

Now consider the number $M(t)$ of unmatched coordinates i at time $t \ge t_1$, that is, the number of i with $X_i(t) \ne Y_i(t)$. Provided the greedy coupling works, this number $M(t)$ cannot increase, and decreases by at least 1 each time two unmatched coordinates are chosen. So we can compare $(M(t_1 + t), t \ge 0)$ with the chain $(N(t), t \ge 0)$ with $N(0) = d$ and

$$P(N(t+1) = m-1 | N(t) = m) = \frac{m(m-1)}{d(d-1)} = 1 - P(N(t+1) = m | N(t) = m).$$

As in the analysis of the shuffling example, the time $T = \min\{t : N(t) = 1\}$ has $ET = \sum_{m=2}^{d} \frac{d(d-1)}{m(m-1)} \le d^2$. When the number $M(t)$ goes strictly below 2 it must become 0, and so

$$\begin{aligned} & P(\text{greedy coupling works for all } t_1 < t \le t_1 + t_2, \mathbf{X}(t_1 + t_2) \ne \mathbf{Y}(t_1 + t_2)) \\ = \;& P(\text{greedy coupling works for all } t_1 < t \le t_1 + t_2, M(t_1 + t_2) > 1) \\ \le \;& P(T > t_2) \le 1/C_2. \end{aligned}$$

This and (17) establish (14).

1.5 Compact groups

Parallel to random flights on finite groups one can discuss discrete-time random flights on classical (continuous) compact groups such as the *orthogonal group* $O(d)$ of $d \times d$ real orthogonal matrices. For instance, specify a *reflection* to be an automorphism which fixes the points in some hyperplane, so that a reflection matrix can be written as

$$A = I - 2xx^T$$

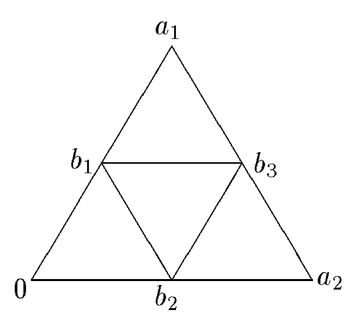
where I is the $d \times d$ identity matrix and x is a unit-length vector in R^d. Assigning to x the Haar measure on the $(d-1)$-sphere creates a *uniform random reflection*, and a sequence of uniform random reflections define a random flight on $O(d)$. Porod [50] shows that the variation threshold satisfies

$$\tau_1 \sim \tfrac{1}{2} d \log d$$

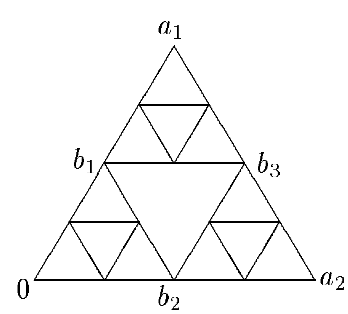
and that the cut-off phenomenon occurs. The result, and its proof via group representation theory, are reminiscent of card-shuffling via random transpositions (Chapter 7 Example 18).

1.6 Brownian motion on a fractal set

Constructions and properties of analogs of Brownian motion taking values in fractal subsets of R^d have been studied in great detail over the last 15 years. Since these processes are most easily viewed as limits of random walks on graphs, we shall say a little about the simplest example. The figure illustrates the first two stages of the construction of the well-known *Sierpinski gasket.*



Graph G_1



Graph G_2

In the topology setting one may regard G_d as a closed subset of R^2, that is as a set of line-segments, and then the closure of $\cup_{d=1}^{\infty} G_d$ is the Sierpinski gasket G (this is equivalent to the usual construction by "cutting out middle triangles"). In the graph setting, regard G_d as a graph and write $(X_t^{(d)}, t = 0, 1, 2, \ldots)$ for discrete-time random walk on G_d started at point 0. Let M_d be the number of steps of $X^{(d)}$ until first hitting point a_1 or a_2. Using symmetry properties of the graphs, there is a simple relationship between the distributions of M_1 and M_2. For the walk on G_2, the length of the time segment until first hitting b_1 or b_2 is distributed as M_1; successive segments (periods until next hitting one of $\{0, a_1, a_2, b_1, b_2, b_3\}$ other than the current one) are like successive steps of the walk on G_1, so the number of segments is distributed as M_1. Using the same argument for general d gives

> M_d is distributed as the d'th generation size in a Galton-Watson branching process with 1 individual in generation 0 and offspring distributed as M_1.

It is easy to calculate $EM_1 = 5$; indeed the distribution of M_1 is determined by its generating function, which can be calculated to be $Ez^{M_1} = z^2/(4-3z)$. So $EM_d = 5^d$. This suggests the existence of a limit process on G after rescaling time, that is a limit

$$(X^{(d)}_{\lfloor 5^{-d} t \rfloor}, 0 \leq t < \infty) \ \xrightarrow{d}\ (X_t^{(\infty)}, 0 \leq t < \infty).$$

In fact we can be more constructive. Branching process theory ([20] Example 4.4.1) shows that $M_d/5^d \ \xrightarrow{d}\ W$ where $EW = 1$ and where W has the self-consistency property

$$\sum_{i=1}^{M} W_i \ \stackrel{d}{=}\ 5W \tag{18}$$

where $(M; W_1, W_2, \ldots)$ are independent, $M \ \stackrel{d}{=}\ M_1$ and $W_i \ \stackrel{d}{=}\ W$. Now in the topological setting, the vertices of G_d are a subset of G. Let $\tilde{X}_t^{(d)}$ be the process on $G_d \subset G$ whose sequence of jumps is as the jumps of the discrete-time walk $X^{(d)}$ but where the times between jumps are independent with distribution $5^{-d}W$. Using (18) we can construct the processes $\tilde{X}^{(d)}$ jointly for all d such that the process $\tilde{X}^{(d)}$, watched only at the times of hitting (successively distinct) points of G_{d-1}, is exactly the process $\tilde{X}^{(d-1)}$. These coupled processes specify a process $X_t^{(\infty)}$ on G at a random subset of times t. It can be shown that this random subset is dense and that sample paths extend continuously to all t, and it is natural to call $X^{(\infty)}$ *Brownian motion on the Sierpinski gasket.*

2 Infinite graphs

There is a huge research literature concerning random walks on infinite discrete groups, and more generally on infinite graphs, and the recent monograph of Woess [59] provides an in-depth treatment. This section focuses narrowly on two aspects of an issue not emphasized in [59]: what does study of random walk on infinite graphs tell us about random walks on finite graphs? One aspect of this issue is that random walks on certain specific infinite graphs may be used to get approximations or inequalities for random walks on specific finite graphs. We treat three examples.

- The infinite lattice Z^d as an approximation to the discrete torus Z^d_N for large N (section 2.4).
- The infinite degree-r tree $\mathbf{T}^r$ and bounds for r-regular expander graphs of large size (section 2.6).
- The hierarchical tree $\mathbf{T}^r_{\mathrm{hier}}$ as an approximation to balanced $(r-1)$-ary trees (section 2.9).

The second aspect concerns properties such as transience, non-trivial boundary, and "spectral radius < 1", which have been well-studied as qualitative properties which an infinite-state chain either possesses or does not possess. What are the quantitative finite-state analogs of such properties? Here actual theorems are scarce; we present conceptual discussion in sections 2.3 and 2.10 as a spur to future research.

2.1 Set-up

We assume the reader has some acquaintance with classical theory (e.g., [20] Chapter 5) for a countable-state irreducible Markov chain, which emphasizes the trichotomy *transient* or *null-recurrent* or *positive-recurrent.* We use the phrase *general chain* to refer to the case of an arbitrary irreducible transition matrix $\mathbf{P}$, without any reversibility assumption.

Recall from Chapter 3 section 2 the identification, in the finite-state setting, of reversible chains and random walks on weighted graphs. Given a reversible chain we defined edge-weights $w_{ij} = \pi_i p_{ij} = \pi_j p_{ji}$; conversely, given edge-weights we defined random walk as the reversible chain

$$p_{vx} = w_{vx}/w_v;\ w_v = \sum_x w_{vx}. \tag{19}$$

In the infinite setting it is convenient (for reasons explained below) to take the "weighted graph" viewpoint. Thus the setting of this section is that we are given a connected weighted graph satisfying

$$w_v \equiv \sum_x w_{vx} < \infty\ \forall x, \quad \sum_v w_v = \infty \tag{20}$$

and we study the associated *random walk* (X_t), i.e., the discrete-time chain with $p_{vx} = w_{vx}/w_v$. So in the unweighted setting ($w_e \equiv 1$), we have nearest-neighbor random walk on a locally finite, infinite graph.

To explain why we adopt this set-up, say π is *invariant* for $\mathbf{P}$ if

$$\sum_i \pi_i p_{ij} = \pi_j\ \forall j; \quad \pi_j > 0\ \forall j.$$

Consider asymmetric random walk on Z, say

$$p_{i,i+1} = 2/3,\ p_{i,i-1} = 1/3;\ -\infty < i < \infty. \tag{21}$$

One easily verifies that each of the two measures $\pi_i = 1$ and $\pi_i = 2^i$ is invariant. Such nonuniqueness makes it awkward to seek to define reversibility of $\mathbf{P}$ via the detailed balance equations

$$\pi_i p_{ij} = \pi_j p_{ji}\ \forall i, j \tag{22}$$

without a prior definition of π. Stating definitions via weighted graphs avoids this difficulty.

The second assumption in (20), that $\sum_v w_v = \infty$, excludes the positive-recurrent case (see Theorem 4 below); because in that case the questions one asks, such as whether the relaxation time τ_2 is finite, can be analyzed by the same techniques as in the finite-state setting.

Our intuitive interpretation of "reversible" in Chapter 3 was "a movie of the chain looks the same run forwards or run backwards". But the chain corresponding to the weighted graph with weights $w_{i,i+1} = 2^i$, which is the chain (21) with $\pi_i = 2^i$, has a particle moving towards $+\infty$ and so certainly doesn't satisfy this intuitive notion. On the other hand, a probabilistic interpretation of an infinite invariant measure π is that if we start at time 0 with independent Poisson(π_v) numbers of particles at vertices v, and let the particles move independently according to $\mathbf{P}$, then the particle process is stationary in time. So the detailed balance equations (22) correspond to the intuitive "movie" notion of reversible *for the infinite particle process*, rather than for a single chain.

2.2 Recurrence and Transience

The next Theorem summarizes parts of the standard theory of general chains (e.g., [20] Chapter 5). Write $\rho_v := P_v(T_v^+ < \infty)$ and let $N_v(\infty)$ be the total number of visits (including time 0) to v.

Theorem 4 *For a general chain, one of the following alternatives holds.*
Recurrent. $\rho_v = 1$ *and* $E_v N_v(\infty) = \infty$ *and* $P_v(N_w(\infty) = \infty) = 1$ *for all* v, w.
Transient. $\rho_v < 1$ *and* $E_v N_v(\infty) < \infty$ *and* $P_v(N_w(\infty) < \infty) = 1$ *for all* v, w.
In the recurrent case there exists an invariant measure π*, unique up to constant multiples, and the chain is either*
positive-recurrent: $E_v T_v^+ < \infty \ \forall v$ *and* $\sum_v \pi_v < \infty$*; or*
null-recurrent: $E_v T_v^+ = \infty \ \forall v$ *and* $\sum_v \pi_v = \infty$.
In the transient and null-recurrent cases, $P_v(X_t = w) \to 0$ *as* $t \to \infty$ *for all* v, w.

Specializing to random walk on a weighted graph, the measure (w_v) is invariant, and the second assumption in (20) implies that the walk cannot be positive-recurrent. By a natural abuse of language we call the weighted graph *recurrent* or *transient*. Because $E_v N_v(\infty) = \sum_t p_{vv}^{(t)}$, Theorem 4 contains the "classical" method to establish transience or recurrence by considering the $t \to \infty$ behavior of $p_{vv}^{(t)}$. This method works easily for random walk on Z^d (section 2.4).

Some of the "electrical network" story from Chapter 3 extends immediately to the infinite setting. Recall the notion of a flow $\mathbf{f}$, and the net flow $f_{(x)}$ out of a vertex x. Say $\mathbf{f}$ is a *unit flow from x to infinity* if $f_{(x)} = 1$ and $f_{(v)} = 0 \ \forall v \neq x$. Thompson's principle (Chapter 3 Proposition 35) extends to the infinite setting, by considering subsets $A_n \downarrow \phi$ (the empty set) with A_n^c finite.

Theorem 5 *Consider a weighted graph satisfying (20). For each v,*

$$\inf \left\{ \tfrac{1}{2} \sum_e f_e^2/w_e : \mathbf{f} \text{ a unit flow from } v \text{ to infinity} \right\} = \frac{1}{w_v(1-\rho_v)}.$$

In particular, the random walk is transient iff for some (all) v there exists a unit flow $\mathbf{f}$ from v to infinity such that $\sum_e f_e^2/w_e < \infty$.

By analogy with the finite setting, we can regard the *inf* as the effective resistance between v and infinity, although (see section **??**) we shall not attempt an axiomatic treatment of infinite electrical networks.

Theorem 5 has the following immediate corollary: of course (a) and (b) are logically equivalent.

Corollary 6 *(a) If a weighted graph is recurrent, then so is any subgraph. (b) To show that a weighted graph is transient, it suffices to find a transient subgraph.*

Thus the classical fact that Z^2 is recurrent implies that a subgraph of Z^2 is recurrent, a fact which is hard to prove by bounding t-step transition probabilities. In the other direction, it is possible (but not trivial) to prove that Z^3 is transient by exhibiting a flow: indeed Doyle and Snell [19] construct a transient tree-like subgraph of Z^3.

Here is a different formulation of the same idea.

Corollary 7 *The return probability $\rho_v = P_v(T_v^+ < \infty)$ cannot increase if a new edge (not incident at v) is added, or the weight of an existing edge (not incident at v) is increased.*

2.3 The finite analog of transience

Recall the mean hitting time parameter τ_0 from Chapter 4. For a sequence of n-state reversible chains, consider the property

$$n^{-1}\tau_0(n) \text{ is bounded as } n \to \infty. \tag{23}$$

We assert, as a conceptual paradigm, that property (23) is the analog of the "transient" property for a single infinite-state chain. The connection is easy to see algebraically for symmetric chains (Chapter 7), where $\tau_0 = E_\pi T_v$ for each v, so that by Chapter 2 Lemma 10

$$n^{-1}\tau_0 = z_{vv} = \sum_{t=0}^{\infty} (p_v v^{(t)} - n^{-1}).$$

The boundedness (in n) of this sum is a natural analog of the transience condition

$$\sum_{t=0}^{\infty} p_{vv}^{(t)} < \infty$$

for a single infinite-state chain. So in principle the methods used to determine transience or recurrence in the infinite-state case ([59] Chapter 1) should be usable to determine whether property (23) holds for finite families, and indeed Proposition 37 of Chapter 3 provides a tool for this purpose. In practice these extremal methods haven't yet proved very successful; early papers [14] proved (23) for expanders in this way, but other methods are easier (see our proof of Chapter 9 Theorem 1). There is well-developed theory ([59] section 6) which establishes recurrence for infinite *planar* graphs under mild assumptions. It is natural to conjecture that under similar assumptions, a planar n-vertex graph has $\tau_0 = \Theta(n \log n)$, as in the case of Z^2 in Proposition 8 below.

2.4 Random walk on Z^d

We consider the lattice Z^d as an infinite $2d$-regular unweighted graph. Write X_t for simple random walk on Z^d, and write $\widetilde{X}_t$ for the continuized random walk. Of course, general random flights (i.e. "random walks", in everyone's terminology except ours) and their numerous variations comprise a well-studied classical topic in probability theory. See Hughes [29] for a wide-ranging intermediate-level treatment, emphasizing physics applications. Our discussion here is very narrow, relating to topics treated elsewhere in this book.

To start some calculations, for $d = 1$ consider

$$\begin{aligned}
\tilde{p}(t) &\equiv P_0(\widetilde{X}_t = 0) \\
&= P(J_t^+ = J_t^-), \text{ where } J_t^+ \text{ and } J_t^- \text{ are the} \\
&\quad \text{independent Poisson}(t/2) \text{ numbers of } +1 \text{ and } -1 \text{ jumps} \\
&= \sum_{n=0}^{\infty} \left(\frac{e^{-t/2}(t/2)^n}{n!} \right)^2 \\
&= e^{-t} I_0(t)
\end{aligned}$$

where $I_0(t) := \sum_{n=0}^{\infty} \frac{(t/2)^{2n}}{(n!)^2}$ is the *modified Bessel function of the first kind of order* 0. Now var $\widetilde{X}_t = t$, and as a consequence of the local CLT (or by quoting asymptotics of the Bessel function I_0) we have

$$\tilde{p}(t) \sim (2\pi t)^{-1/2} \text{ as } t \to \infty. \tag{24}$$

As discussed in Chapter 4 section 6.2 and Chapter 5 Example 17, a great advantage of working in continuous time in dimensions $d \geq 2$ is that the coordinate processes are independent copies of slowed-down one-dimensional processes, so that $\tilde{p}^{(d)}(t) \equiv P_0(\widetilde{X}_t = 0)$ in dimension d satisfies

$$\tilde{p}^{(d)}(t) = (\tilde{p}(t/d))^d = e^{-t}(I_0(t/d))^d. \tag{25}$$

In particular, from (24),

$$\tilde{p}^{(d)}(t) \sim (\tfrac{d}{2\pi})^{d/2} t^{-d/2} \text{ as } t \to \infty. \tag{26}$$

One can do a similar analysis in the discrete time case. In dimension $d = 1$,

$$\begin{aligned} p(t) &\equiv P_0(X_t = 0) \\ &= 2^{-t}\binom{t}{t/2}, t \text{ even} \\ &\sim 2\ (2\pi t)^{-1/2} \text{ as } t \to \infty,\ t \text{ even}. \end{aligned} \tag{27}$$

This agrees with (26) but with an extra "artificial" factor of 2 arising from periodicity. A more tedious argument gives the analog of (26) in discrete time for general d:

$$p^{(d)}(t) \sim 2(\tfrac{d}{2\pi})^{d/2} t^{-d/2} \text{ as } t \to \infty,\ t \text{ even}. \tag{28}$$

From the viewpoint of classical probability, one can regard (26,28) as the special case $j = 0$ of the local CLT: in continuous time in dimension d,

$$\sup_j \left| P_0(\widetilde{X}_t = j) - (\tfrac{d}{2\pi})^{d/2} t^{-d/2} \exp(-d|j|^2/(2t)) \right| = o(t^{-d/2}) \text{ as } t \to \infty$$

where $|j|$ denotes Euclidean norm.

The occupation time $N_0(t)$ satisfies $E_0 N_0(t) = \int_0^t \tilde{p}(s)\ ds$ (continuous time) and $= \sum_{s=0}^{t-1} p(s)$ (discrete time). In either case, as $t \to \infty$,

$$(d = 1)\quad E_0 N_0(t) \sim \sqrt{\tfrac{2}{\pi}}\, t^{1/2} \tag{29}$$

$$(d = 2)\quad E_0 N_0(t) \sim \tfrac{1}{\pi} \log t \tag{30}$$

$$\begin{aligned} (d \geq 3)\quad E_0 N_0(t) \to R_d &\equiv \int_0^\infty \tilde{p}^{(d)}(t) dt \\ &= \int_0^\infty e^{-t}(I_0(t/d))^d\ dt \end{aligned} \tag{31}$$

where $R_d < \infty$ for $d \geq 3$ by (26). This is the classical argument for establishing transience in $d \geq 3$ and recurrence in $d \leq 2$, by applying Theorem 4. Note that the return probability $\rho^{(d)} := P_0(T_0^+ < \infty)$ is related to $E_0 N_0(\infty)$ by $E_0 N_0(\infty) = \frac{1}{1-\rho^{(d)}}$; in other words

$$\rho^{(d)} = \frac{R_d - 1}{R_d}, \ d \geq 3.$$

Textbooks sometimes give the impression that calculating $\rho^{(d)}$ is hard, but one can just calculate numerically the integral (31). Or see [26] for a table.

The quantity $\rho^{(d)}$ has the following sample path interpretation. Let V_t be the number of distinct vertices visited by the walk before time t. Then

$$t^{-1} V_t \to 1 - \rho^{(d)} \text{ a.s. }, \ d \geq 3. \tag{32}$$

The proof of this result is a textbook application of the ergodic theorem for stationary processes: see [20] Theorem 6.3.1.

2.5 The torus Z_m^d

We now discuss how random walk on Z^d relates to $m \to \infty$ asymptotics for random walk on the finite torus Z_m^d, discussed in Chapter 5. We now use superscript $\cdot^{(m)}$ to denote the length parameter. From Chapter 5 Example 17 we have

$$\tau_2^{(m)} = \frac{d}{1 - \cos(2\pi/m)} \sim \frac{dm^2}{2\pi^2}$$

$$\tau_1^{(m)} = \Theta(m^2) \tag{33}$$

where asymptotics are as $m \to \infty$ for fixed d. One can interpret this as a consequence of the dN^2 time rescaling in the wweak convergence of rescaled random walk to Brownian motion of the d-dimensional torus, for which (cf. sections 1.1 and 1.2) $\tau_2 = 2\pi^{-2}$. At (74)–(75) of Chapter 5 we saw that the eigentime identity gave an exact formula for the mean hitting time parameter $\tau_0^{(m)}$, whose asymptotics are, for $d \geq 3$,

$$m^{-d} \tau_0^{(m)} \to \hat{R}_d \equiv \int_0^1 \cdots \int_0^1 \frac{1}{\frac{1}{d}\sum_{u=1}^d (1 - \cos(2\pi x_u))} \, dx_1 \ldots dx_d < \infty. \tag{34}$$

Here we give an independent analysis of this result, and the case $d = 2$.

Proposition 8

$$(d=1) \qquad \tau_0^{(n)} \sim \tfrac{1}{6}n^2 \tag{35}$$

$$(d=2) \qquad \tau_0^{(m)} \sim 2\pi^{-1}m^2\log m \tag{36}$$

$$(d\geq 3) \qquad \tau_0^{(m)} \sim R_d m^d \tag{37}$$

for R_d defined by (31). In particular, the expressions for R_d and $\hat{R}_d$ at (31) and (34) are equal, for $d \geq 3$.

The $d=1$ result is from Chapter 5 (26). We now prove the other cases.

Proof. We may construct continuized random walk $\widetilde{X}_t^{(m)}$ on Z_m^d from continuized random walk $\widetilde{X}_t$ on Z^d by

$$\widetilde{X}_t^{(m)} = \widetilde{X}_t \bmod m \tag{38}$$

and then $P_0(\widetilde{X}_t^{(m)}=0) \geq P_0(\widetilde{X}_t=0)$. So

$$\begin{aligned}
m^{-d}\tau_0^{(m)} &= \int_0^\infty \left(P_0(\widetilde{X}_t^{(m)}=0) - m^{-d}\right)\, dt \\
&\quad \text{(Chapter 2, Corollary 12 and (8))} \\
&= \int_0^\infty \left(P_0(\widetilde{X}_t^{(m)}=0) - m^{-d}\right)^+ dt \text{ by complete monotonicity} \\
&\geq \int_0^\infty \left(P_0(\widetilde{X}_t=0) - m^{-d}\right)^+\, dt \qquad (39) \\
&\to \int_0^\infty P_0(\widetilde{X}_t=0)\, dt \quad = R_d.
\end{aligned}$$

Consider the case $d \geq 3$. To complete the proof, we need the corresponding upper bound, for which it is sufficient to show

$$\int_0^\infty \left(P_0(\widetilde{X}_t^{(m)}=0) - m^{-d} - P_0(\widetilde{X}_t=0)\right)^+ dt \to 0 \text{ as } m\to\infty. \tag{40}$$

To verify (40) without detailed calculations, we first establish a 1-dimensional bound

$$(d=1) \qquad \tilde{p}^{(m)}(t) \leq \tfrac{1}{m} + \tilde{p}(t). \tag{41}$$

To obtain (41) we appeal to a coupling construction (the *reflection coupling*, described in continuous-space in section 1.3 – the discrete-space setting here is similar) which shows that continuized random walks $\widetilde{X}^{(m)}, \widetilde{Y}^{(m)}$ on Z_m with $\widetilde{X}_0^{(m)}=0$ and $\widetilde{Y}_0^{(m)}$ distributed uniformly can be coupled so that

$$\widetilde{Y}_t^{(m)} = 0 \text{ on the event } \{\widetilde{X}_t^{(m)}=0, T\leq t\}$$

where T is the first time that $\widetilde{X}^{(m)}$ goes distance $\lfloor m/2 \rfloor$ from 0. And by considering the construction (38)

$$P(\widetilde{X}^{(m)}_t = 0) \leq P(\widetilde{X}_t = 0) + P(\widetilde{X}^{(m)}_t = 0, T \leq t)$$

and (41) follows, since $P(\widetilde{Y}^{(m)}_t = 0) = 1/m$.

Since the d-dimensional probabilities relate to the 1-dimensional probabilities via $P_0(\widetilde{X}^{(m)}_t = 0) = \left(\tilde{p}^{(m)}(t/d)\right)^d$ and similarly on the infinite lattice, we can use inequality (41) to bound the integrand in (40) as follows.

$$\begin{aligned}
P_0(\widetilde{X}^{(m)}_t = 0) - m^{-d} - P_0(\widetilde{X}_t = 0)& \\
&\leq \left(\tfrac{1}{m} + \tilde{p}(t/d)\right)^d - m^{-d} - (\tilde{p}(t/d))^d \\
&= \sum_{j=1}^{d-1} \binom{d}{j} (\tilde{p}(t/d))^j (\tfrac{1}{m})^{d-j} \\
&= \frac{\tilde{p}(t/d)}{m} \sum_{j=1}^{d-1} \binom{d}{j} (\tilde{p}(t/d))^{j-1} (\tfrac{1}{m})^{d-1-j} \\
&\leq \frac{\tilde{p}(t/d)}{m} \sum_{j=1} d-1 \binom{d}{j} \left[\max((\tilde{p}(t/d))^{d-2}, \tfrac{1}{m})\right]^{d-2} \\
&= (2^d - 2) \frac{\tilde{p}(t/d)}{m} \left[\max((\tilde{p}(t/d))^{d-2}, (\tfrac{1}{m})^{d-2})\right] \\
&\leq (2^d - 2) \frac{\tilde{p}(t/d)}{m} \left[(\tilde{p}(t/d))^{d-2} + \tfrac{1}{m})^{d-2}\right] \\
&= (2^d - 2) \left[\frac{(\tilde{p}(t/d))^{d-1}}{m} + \frac{\tilde{p}(t/d)}{m^{d-1}}\right].
\end{aligned}$$

The fact (24) that $\tilde{p}(t) = \Theta(t^{-1/2})$ for large t easily implies that the integral in (40) over $0 \leq t \leq m^3$ tends to zero. But by (33) and submultiplicativity of $\bar{d}(t)$,

$$0 \leq P_0(\widetilde{X}^{(m)}_t = 0) - m^{-d} \leq d(t) \leq \bar{d}(t) \leq B_1 \exp(-\tfrac{t}{B_2 m^2}) \qquad (42)$$

where B_1, B_2 depend only on d. This easily implies that the integral in (40) over $m^3 \leq t < \infty$ tends to zero, completing the proof of (37).

In the case $d = 2$, we fix $b > 0$ and truncate the integral in (39) at bm^2 to get

$$m^{-2}\tau^{(m)}_0 \quad \geq \quad -b + \int_0^{bm^2} P_0(\widetilde{X}_t = 0)\ dt$$

$$
\begin{aligned}
&= -b + (1+o(1))\tfrac{2}{\pi}\log(bm^2) \text{ by } (30)\\
&= (1+o(1))\tfrac{2}{\pi}\log m.
\end{aligned}
$$

Therefore

$$\tau_0^{(m)} \geq (1+o(1))\tfrac{2}{\pi}m^2\log m.$$

For the corresponding upper bound, since $\int_0^{m^2} P_0(\widetilde{X}_t = 0)dt \sim \frac{2}{\pi}\log m$ by (30), and $m^{-2}\tau_0^{(m)} = \int_0^\infty \left(P_0(\widetilde{X}_t^{(m)} = 0) - m^{-2}\right)\, dt$, it suffices to show that

$$\int_0^\infty \left(P_0(\widetilde{X}_t^{(m)} = 0) - m^{-2} - P_0(\widetilde{X}_t = 0)\right)^+ \, dt$$

$$+\int_{m^2}^\infty \left(P_0(\widetilde{X}_t^{(m)} = 0) - m^{-2}\right)^+ dt = o(\log N). \tag{43}$$

To bound the first of these two integrals, we observe from (41) that $P_0(\widetilde{X}_t^{(m)} = 0) \leq (m^{-1} + \tilde{p}(t/2))^2$, and so the integrand is bounded by $\frac{2}{m}\tilde{p}(t/2)$. Using (24), the first integral is $O(1) = o(\log m)$. To analyze the second integral in (43) we consider separately the ranges $m^2 \leq t \leq m^2\log^{3/2} m$ and $m^2\log^{3/2} m \leq t < \infty$. Over the first range, we again use (41) to bound the integrand by $\frac{2}{m}\tilde{p}(t/2) + (\tilde{p}(t/2))^2$. Again using (24), the integral is bounded by

$$(1+o(1))\frac{2}{\pi^{1/2}m}\int_{m^2}^{m^2\log^{3/2} m} t^{-1/2}dt \;+\; (1+o(1))\pi^{-1}\int_{m^2}^{m^2\log^{3/2} m} t^{-1}dt$$

$$= \Theta(\log^{3/4} m) + \Theta(\log\log m) = o(\log m).$$

To bound the integral over the second range, we use (42) and find

$$
\begin{aligned}
\int_{m^2\log^{3/2} m}^\infty \left(P_0(\widetilde{X}_t^{(m)} = 0) - m^{-2}\right) dt &\leq B_1B_2m^2\exp\left(-\tfrac{\log^{3/2} m}{B_2}\right)\\
&= o(1) = o(\log m).
\end{aligned}
$$

□

2.6 The infinite degree-r tree

Fix $r \geq 3$ and write $\mathbf{T}^r$ for the infinite tree of degree r. We picture $\mathbf{T}^r$ as a "family tree", where the root ϕ has r children, and each other vertex has one parent and $r-1$ children. Being a vertex-transitive graph (recall Chapter 7 section 1.1; for r even, $\mathbf{T}^r$ is the Cayley graph of the free group on $r/2$ generators), one can study many more general "random flights" on $\mathbf{T}^r$ (see Notes), but we shall consider only the simple random walk (X_t).

We can get some information about the walk without resorting to calculations. The "depth" process $d(X_t, \phi)$ is clearly the "reflecting asymmetric random walk" on $Z^+ := \{0, 1, 2, \ldots\}$ with

$$p_{0,1} = 1; \; p_{i,i-1} = 1/r; \; p_{i,i+1} = (r-1)/r, \; i \geq 1.$$

By comparison with asymmetric random walk on all Z, which has drift $(r-2)/r$, we see that

$$t^{-1} d(X_t, \phi) \to \frac{r-2}{r} \text{ a.s. as } t \to \infty. \tag{44}$$

In particular, the number of returns to ϕ is finite and so the walk is transient. Now consider the return probability $\rho = P_\phi(T_\phi^+ < \infty)$ and note that (by considering the first step) $\rho = P_\phi(T_c < \infty)$ where c is a child of ϕ. Considering the first two steps, we obtain the equation $\rho = \frac{1}{r} + \frac{r-1}{r}\rho^2$, and since by transience $\rho < 1$, we see that

$$\rho := P_\phi(T_\phi^+ < \infty) = P_\phi(T_c < \infty) = \frac{1}{r-1}. \tag{45}$$

So

$$E_\phi N_\phi(\infty) = \frac{1}{1-\rho} = \frac{r-1}{r-2}. \tag{46}$$

As at (32), ρ has a sample path interpretation: the number V_t of distinct vertices visited by the walk before time t satisfies

$$t^{-1} V_t \to 1 - \rho = \tfrac{r-2}{r-1} \text{ a.s. as } t \to \infty.$$

By transience, amongst the children of ϕ there is some vertex L_1 which is visited last by the walk; then amongst the children of L_1 there is some vertex L_2 which is visited last by the walk; and so on, to define a "path to infinity" $\phi = L_0, L_1, L_2, \ldots$. By symmetry, given $L_1, L_2, \ldots, L_{i-1}$ the conditional distribution of L_i is uniform over the children of L_{i-1}, so in the natural sense we can describe (L_i) as the uniform random path to infinity.

2.7 Generating function arguments

While the general qualitative behavior of random walk on $\mathbf{T}^r$ is clear from the arguments above, more precise quantitative estimates are most naturally obtained via generating function arguments. For any state i of a Markov chain, the generating functions $G_i(z) := \sum_{t=0}^\infty P_i(X_t = i)z^t$ and $F_i(z) := \sum_{t=1}^\infty P_i(T_i^+ = t)z^t$ are related by

$$G_i = 1 + F_i G_i \tag{47}$$

(this is a small variation on Chapter 2 Lemma 19). Consider simple *symmetric* reflecting random walk on Z^+. Clearly

$$G_0(z) = \sum_{t=0}^\infty \binom{2t}{t} 2^{-2t} z^{2t} = (1-z^2)^{-1/2},$$

the latter identity being the series expansion of $(1-x)^{-1/2}$. So by (47)

$$F_0(z) := \sum_{t=0 \text{ or } 1}^\infty P_0(T_0^+ = 2t)z^{2t} = 1 - (1-z^2)^{1/2}.$$

Consider an excursion of length $2t$, that is, a path $(0 = i_0, i_1, \ldots, i_{2t-1}, i_{2t} = 0)$ with $i_j > 0, 1 \le j \le 2t-1$. This excursion has chance 2^{1-2t} for the symmetric walk on Z^+, and has chance $((r-1)/r)^{t-1}(1/r)^t$ for the asymmetric walk $d(X_t, \phi)$. So

$$\frac{P_\phi(T_\phi^+ = 2t)}{P_0(T_0^+ = 2t)} = \frac{r}{2(r-1)} \left(\frac{4(r-1)}{r^2}\right)^t$$

where the numerator refers to simple RW on the tree, and the denominator refers to simple symmetric reflecting RW on Z^+. So on the tree,

$$F_\phi(z) = \frac{r}{2(r-1)} F_0\left(z\sqrt{\tfrac{4(r-1)}{r^2}}\right) = \frac{r}{2(r-1)}\left(1 - \left(1 - \frac{4(r-1)z^2}{r^2}\right)^{1/2}\right).$$

Then (47) gives an expression for $G_\phi(z)$ which simplifies to

$$G_\phi(z) = \frac{2(r-1)}{r - 2 + \sqrt{r^2 - 4(r-1)z^2}}. \tag{48}$$

In particular, G_ϕ has radius of convergence $1/\beta$, where

$$\beta = 2r^{-1}\sqrt{r-1} < 1. \tag{49}$$

Without going into details, one can now use standard Tauberian arguments to show

$$P_\phi(X_t = \phi) \sim \alpha t^{-3/2}\beta^t, \quad t \text{ even} \tag{50}$$

for a computable constant α, and this format (for different values of α and β) remains true for more general radially symmetric random flights on $\mathbf{T}^r$ ([59] Theorem 19.30). One can also in principle expand (48) as a power series to obtain $P_\phi(X_t = \phi)$. Again we shall not give details, but according to Giacometti [25] one obtains

$$\begin{aligned} P_\phi(X_t = \phi) &= \frac{r-1}{r}\left(\frac{\sqrt{r-1}}{r}\right)^t \frac{\Gamma(1+t)}{\Gamma(2+t/2)\Gamma(1+t/2)} \\ &\quad \times {}_2F_1(\tfrac{t+1}{2}, 1, 2+\tfrac{t}{2}, \tfrac{4(r-1)}{r^2}),\ t \text{ even} \end{aligned} \tag{51}$$

where ${}_2F_1$ is the generalized hypergeometric function.

Finally, the β at (49) can be interpreted as an eigenvalue for the infinite transition matrix (p_{ij}), so we anticipate a corresponding eigenfunction f_2 with

$$\sum_j p_{ij} f_2(j) = \beta f_2(i)\ \forall i, \tag{52}$$

and one can verify this holds for

$$f_2(i) := (1 + \tfrac{r-2}{r} i)(r-1)^{-i/2}. \tag{53}$$

2.8 Comparison arguments

Fix $r \geq 3$ and consider a sequence (G_n) of n-vertex r-regular graphs with $n \to \infty$. Write (X_t^n) for the random walk on G_n. We can compare these random walks with the random walk (X_t^∞) on $\mathbf{T}^r$ via the obvious inequality

$$P_v(X_t^n = v) \geq P_\phi(X_t^\infty = \phi),\ t \geq 0. \tag{54}$$

To spell this out, there is a *universal cover* map $\gamma : \mathbf{T}^r \to G_n$ with $\gamma(\phi) = v$ and such that for each vertex w of $\mathbf{T}^r$ the r edges at w are mapped to the r edges of G_n at $\gamma(w)$. Given the random walk X^∞ on $\mathbf{T}^r$, the definition $X_t^n = \gamma(X_t^\infty)$ constructs random walk on G_n, and (54) holds because $\{X_t^n = v\} \supseteq \{X_t^\infty = \phi\}$.

It is easy to use (54) to obtain asymptotic lower bounds on the fundamental parameters discussed in Chapter 4. Instead of the relaxation time τ_2, it is more natural here to deal directly with the second eigenvalue λ_2.

Lemma 9 *For random walk on n-vertex r-regular graphs, with $r \geq 3$ fixed and $n \to \infty$*
(a) $\liminf n^{-1}\tau_0(n) \geq \frac{r-1}{r-2}$;

(b) $\liminf \frac{\tau_1(n)}{\log n} \geq \frac{r}{(r-2)\log(r-1)}$;
(c) $\liminf \lambda_2(n) \geq \beta := 2r^{-1}\sqrt{r-1}$.

Theory concerning *expanders* (Chapter 9 section 1) shows there exist graphs where the limits above are finite constants (depending on r), so Lemma 9 gives the optimal order of magnitude bound.

Proof. For (a), switch to the continuous-time walk, consider an arbitrary vertex v in G_n, and take $t_0(n) \to \infty$ with $t_0(n)/n \to 0$. Then we repeat the argument around (39) in the torus setting:

$$\begin{aligned} n^{-1} E_\pi T_v &= \int_0^\infty (P_v(X_t^n = v) - \tfrac{1}{n})\ dt \\ &\geq \int_0^{t_0} (P_v(X_t^n = v) - \tfrac{1}{n})\ dt \\ &\geq -\frac{t_0}{n} + \int_0^{t_0} P_\phi(X_t^\infty = \phi)\ dt \text{ by (54)} \\ &\to \int_0^\infty P_\phi(X_t^\infty = \phi)\ dt \\ &= E_\phi N_\phi(\infty) = \frac{r-1}{r-2}, \end{aligned}$$

which is somewhat stronger than assertion (a). Next, the discrete-time spectral representation implies

$$P_v(X_t^n = v) \leq \tfrac{1}{n} + n\beta^t(n).$$

Using (54) and (50), for any $n \to \infty, t \to \infty$ with t even,

$$t^{-3/2}\beta^t(\alpha - o(1)) \leq \tfrac{1}{n} + n\beta^t(n). \tag{55}$$

For (b), the argument for (54) gives a coupling between the process X^n started at v and the process X^∞ started at ϕ such that

$$d^n(X_t^n, v) \leq d^\infty(X_t^\infty, \phi)$$

where d^n and d^∞ denote graph distance. Fix $\varepsilon > 0$ and write $\gamma = \frac{r-2}{r} + \varepsilon$. By the coupling and (44), $P(d^n(X_t^n, v) \geq \gamma t) \to 0$ as $n, t \to \infty$. This remains true in continuous time. Clearly $\tau_1(n) \to \infty$, and so by definition of τ_1 we have

$$\limsup \pi\{w : d^n(w, v) \geq \gamma\tau_1(n)\} \leq e^{-1}.$$

But by counting vertices,

$$\begin{aligned}
\pi\{w : d^n(v,w) \leq d\} &\leq \frac{1+r+r(r-1)+\cdots+r(r-1)^{d-1}}{n} \\
&\to 0 \text{ if } d \sim (1-\varepsilon)\frac{\log n}{\log(r-1)}.
\end{aligned}$$

For these two limit results to be consistent we must have $\gamma\tau_1(n) \geq (1-\varepsilon)\frac{\log n}{\log(r-1)}$ for all large n, establishing (b).

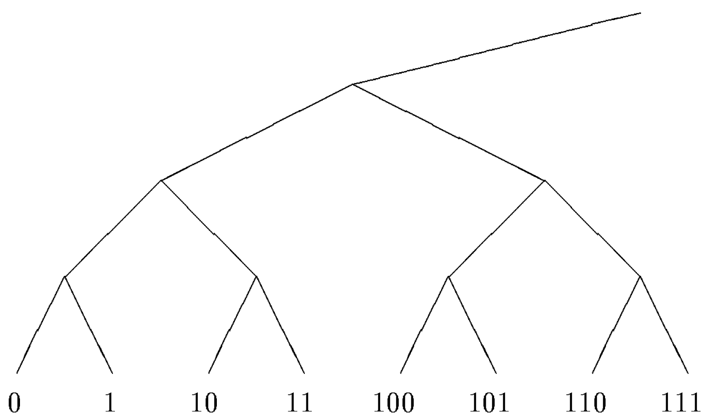
For (c), fix a vertex v_0 of G_n and use the function f_2 at (53) to define $f(v) := f_2(d(v,v_0))$ for all vertices v of G_n. The equality (52) for f_2 on the infinite tree easily implies the inequality $\mathbf{P}f \geq \beta f$ on G_n. Set $\bar{f} := n^{-1}\sum_v f(v)$ and write 1 for the unit function. By the Rayleigh-Ritz characterization (Chapter 4 eq. (73)), writing $\langle g,h\rangle := \sum_{ij}\pi_i g_i p_{ij} h_j$,

$$\begin{aligned}
\lambda_2(n) &\geq \frac{\langle f-\bar{f}1, \mathbf{P}(f-\bar{f}1)\rangle}{||f-\bar{f}1||_2^2} \\
&= \frac{\langle f, \mathbf{P}f\rangle - \bar{f}^2}{||f||_2^2 - \bar{f}^2} \\
&\geq \frac{\beta||f||_2^2 - \bar{f}^2}{||f||_2^2 - \bar{f}^2}.
\end{aligned}$$

As $n \to \infty$ we have $\bar{f} \to 0$ while $||f||_2$ tends to a non-zero limit, establishing (c).

2.9 The hierarchical tree

Fix $r \geq 2$. There is an infinite tree (illustrated for $r = 2$ in the figure) specified as follows. Each vertex is at some height $0,1,2,\ldots$. A vertex at height h has one parent vertex at height $h+1$ and (if $h \geq 1$) r child vertices at height $h-1$. The height-0 vertices are leaves, and the set L of leaves has a natural labeling by finite r-ary strings. The figure illustrates the binary ($r=2$) case, where $L = \{0,1,10,11,100,101,\ldots\}$. L forms an Abelian group under entrywise addition modulo r, e.g. for $r=2$ we have $1101+110 = 1101+0110 = 1011$. Adopting a name used for generalizations of this construction in statistical physics, we call L the *hierarchical lattice* and the tree $\mathbf{T}^r_{\mathrm{hier}}$ the *hierarchical tree.*



Fix a parameter $0 < \lambda < r$. Consider biased random walk X_t on the tree $\mathbf{T}^r_{\text{hier}}$, where from each non-leaf vertex the transition goes to the parent with probability $\lambda/(\lambda + r)$ and to each child with probability $1/(\lambda + r)$. Then consider $Y =$ "X watched only on L", that is the sequence of (not-necessarily distinct) successive leaves visited by X. The group L is distance-transitive (for Hamming distance on L) and Y is a certain isotropic random flight on L. A nice feature of this example is that *without calculation* we can see that Y is recurrent if and only if $\lambda \leq 1$. For consider the path of ancestors of 0. The chain X must spend an infinite time on that path (side-branches are finite); on that path X behaves as asymmetric simple random walk on Z^+, which is recurrent if and only if $\lambda \leq 1$; so X and thence Y visits 0 infinitely often if and only if $\lambda \leq 1$.

Another nice feature is that we can give a fairly explicit expression for the t-step transition probabilities of Y. Writing H for the maximum height reached by X in an excursion from the leaves, then

$$P(H \geq h) = P_1(\hat{T}_h < \hat{T}_0) = \frac{\frac{r}{\lambda} - 1}{(\frac{r}{\lambda})^h - 1}, \quad h \geq 1$$

where $\hat{T}$ denotes hitting time for the height process. Writing M_t for the maximum height reached in t excursions,

$$P(M_t < h) = (P(H < h))^t = \left(1 - \frac{\frac{r}{\lambda} - 1}{(\frac{r}{\lambda})^h - 1}\right)^t.$$

It is clear by symmetry that the distribution of Y_t is conditionally uniform on the leaves which are descendants of the maximal-height vertex previously visited by X. So for leaves v, x with branchpoint at height d,

$$P_v(Y_t = x) = \sum_{h \geq d} r^{-h} P(M_t = h).$$

Since $P(M_t = h) = P(M_t < h+1) - P(M_t < h)$, we have found the "fairly explicit expression" promised above. A brief calculation gives the following time-asymptotics. Fix $s > 0$ and consider $t \sim s(\frac{r}{\lambda})^j$ with $j \to \infty$; then

$$P_v(Y_t = v) \sim r^{-j} f(s); \quad \text{where}$$

$$f(s) = \sum_{i=-\infty}^{\infty} r^{-i} \left(\exp(-s(\tfrac{r}{\lambda} - 1)(\tfrac{r}{\lambda})^{-i-1}) - \exp(-s(\tfrac{r}{\lambda} - 1)(\tfrac{r}{\lambda})^{-i}) \right).$$

In particular,

$$P_v(Y_t = v) = \Theta(t^{-d/2}) \text{ as } t \to \infty, \ d = \frac{2 \log r}{\log r - \log \lambda}. \tag{56}$$

Comparing with (26), this gives a sense in which Y mimics simple random walk on Z^d, for d defined above. Note that d increases continuously from 0 to ∞ as λ increases from 0 to r, and that Y is recurrent if and only if $d \leq 2$.

Though we don't go into details, random walk on the hierarchical lattice is a natural infinite-state analog of biased random walk on the balanced finite tree (Chapter 5 section 2.1). In particular, results in the latter context showed that, writing n for number of vertices, $\tau_0(n) = O(n)$ if and only if $\lambda/r > 1/r$, that is if and only if $d > 2$. This is the condition for transience of the infinite-state walk, confirming the paradigm of section 2.3.

2.10 Towards a classification theory for sequences of finite chains

Three chapters of Woess [59] treat in detail three properties that random walk on an infinite graph may or may not possess:

- transience
- spectral radius < 1
- non-trivial boundary.

Can these be related to properties for sequences of finite chains? We already mentioned (section 2.3) that the property $\tau_0(n) = O(n)$ seems to be the analog of transience. In this speculative section we propose definitions of three other properties for sequences of finite chains, which we name

- compactness
- infinite-dimensionality
- expander-like.

Future research will show whether these are useful definitions! Intuitively we expect that every reasonably "natural" sequence should fall into one of these three classes.

For simplicity we consider reversible random walks on Cayley graphs. It is also convenient to continuize. The resulting chains are special cases of (reversible) *Lévy processes*. We define the general Lévy process to be a continuous-time process with stationary independent increments on a (continuous or discrete) group. Thus the setting for the rest of this section is a sequence $(X_t^{(n)})$ of reversible Lé—vy processes on finite groups $G^{(n)}$ of size $n \to \infty$ through some subsequence. Because we work in continuous time, the eigenvalues satisfy $0 = \lambda_1^{(n)} < \lambda_2^{(n)} \leq \cdots$.

(A): Compactness. Say the sequence $(X_t^{(n)})$ is *compact* if there exists a (discrete or continuous) compact set S and a reversible Lévy process $\widetilde{X}_t$ on S such that

(i) $\tilde{d}(t) \equiv ||P_\pi(\widetilde{X}_t \in \cdot) \to \pi|| \to 0$ as $t \to \infty$;

(ii) $\frac{\lambda_j(n)}{\lambda_2(n)} \to \tilde{\lambda}_j$ as $n \to \infty$, $j \geq 2$; where $1 = \tilde{\lambda}_2 \leq \tilde{\lambda}_3 \leq \cdots$ are the eigenvalues of $(\widetilde{X}_t)$;

(iii) $d^{(n)}(t\ \tau_2(n)) \to \tilde{d}(t)$ as $n \to \infty$; $t > 0$.

These properties formalize the idea that the sequence of random walks form discrete approximations to a limit Lévy process on a compact group, at least as far as mixing times are concerned. Simple random walk on Z_m^d, and the limit Brownian motion on R^d (section 1.2) form the obvious example. Properties (i) and (iii) imply, in particular, that

$$\tau_1(n)/\tau_2(n) \text{ is bounded as } n \to \infty. \tag{57}$$

One might hope that a converse is true:

Does every sequence satisfying (57) have a compact subsequence?

Unfortunately, we are convinced that the answer is "no", for the following reason. Take (X_t^n) which is compact, where the limit Lévy process has function $\tilde{d}(t)$ as at (i). Now consider a product chain $(X_t^{(n)}, Y_t^{(n)})$, where components run independently, and where $Y^{(n)}$ has the cut-off property

(Chapter 7) and $\tau_1^Y(n) \sim \tau_2^X(n)$. Note that by Chapter 7-1 Lemma 1 we have $\tau_2^Y(n) = o(\tau_1^Y(n))$. If the product chain had a subsequential limit, then its total variation function at (i), say $d'(t)$, must satisfy

$$\begin{aligned} d'(t) &= \tilde{d}(t), \quad t > 1 \\ &= 1. \quad t < 1. \end{aligned}$$

But it seems intuitively clear (though we do not know a proof) that every Lévy process on a compact set has continuous $d(\cdot)$. This suggests the following conjecture.

Conjecture 10 *For any sequence of reversible Lévy processes satisfying (57), there exists a subsequence satisfying the definition of* compact *except that condition (ii) is replaced by*
(iv): $\exists t_0 \geq 0$:

$$\begin{aligned} d^{(n)}(t\ \tau_2(n)) &\to 1; \quad t < t_0 \\ &\to \tilde{d}(t); \quad t > t_0. \end{aligned}$$

Before describing the other two classes of chains, we need a definition and some motivating background. In the present setting, the property "trivial boundary" is equivalent (see Notes) to the property

$$\lim_{t\to\infty} ||P_v(X_t \in \cdot) - P_w(X_t \in \cdot)|| = 0, \quad \forall v, w. \tag{58}$$

This suggests that an analogous finite-state property might involve whether the variation distance for nearby starts becomes small before time τ_1. Say that a sequence $(L_n(\varepsilon))$ of subsets is an *asymptotic ε-neighborhood* if

$$||P_\phi(X_{\varepsilon\tau_1} \in \cdot) - P_v(X_{\varepsilon\tau_1} \in \cdot)|| \to 0 \text{ as } n \to \infty$$

uniformly over $v \in L_n(\varepsilon)$; here ϕ is an arbitrary reference vertex. From Chapter 7-1 Lemma 1(b) we can deduce that, if the cut-off property holds, such a neighborhood must have size $|L_n(\varepsilon)| = o(n)$.

(B): Infinite dimensional. Say the sequence $(X_t^{(n)})$ is *infinite-dimensional* if the following three properties hold.

(i) $\tau_1(n) = \Theta(\tau_2 \log\log n)$
(ii) The cut-off property holds

(iii) there exists some $\delta(\varepsilon)$, increasing from 0 to 1 as ε increases from 0 to 1, such that a maximal-size asymptotic ε-neighborhood $(L_n(\varepsilon))$ has

$$\log |L_n(\varepsilon)| = (\log n)^{\delta(\varepsilon)+o(1)} \text{ as } n \to \infty.$$

This definition is an attempt to abstract the essential properties of random walk on the d-cube (Chapter 5 Example 15), where properties (i) and (ii) were already shown. We outline below a proof of property (iii) in that example. Another fundamental example where (i) and (ii) hold is card-shuffling by random transpositions (Chapter 7 Example 18)), and we conjecture that property (iii) also holds there. Conceptually, this class *infinite-dimensional* of sequences is intended (cf. (58)) as the analog of a single random walk with trivial boundary on an infinite-dimensional graph.

Property (iii) for the d-cube. Let $(X(t))$ be continuous-time random walk on the d-cube, and $(X^*(t))$ continuous-time random walk on the b-cube, where $b \leq d$. The natural coupling shows

if $d(v,w) = b$ then

$$||P_v(X(t) \in \cdot) - P_w(X(t) \in \cdot)|| = ||P_{\mathbf{0}}(X^*(tb/d) \in \cdot) - P_{\mathbf{1}}(X^*(tb/d) \in \cdot)||.$$

Take $d \to \infty$ with

$$b(d) \sim d^\alpha, \quad t(d) \sim \tfrac{1}{4}\varepsilon d \log d$$

for some $0 < \alpha, \varepsilon < 1$, so that

$$\lim_d \frac{t(d)b(d)/d}{\frac{1}{4}b(d)\log b(d)} \to \tfrac{\varepsilon}{\alpha}.$$

Since the variation cut-off for the b-cube is at $\frac{1}{4}b\log b$, we see that for vertices v and w at distance $b(d)$,

$$\begin{aligned} ||P_v(X(t(d)) \in \cdot) - P_w(X(t(d)) \in \cdot)|| &\to 1, \quad \varepsilon > \alpha \\ &\to 0, \quad \varepsilon < \alpha. \end{aligned}$$

So a maximal-size asymptotic ε-neighborhood $(L_n(\varepsilon))$ of $\mathbf{0}$ must be of the form $\{w : d(w,\mathbf{0}) \leq d^{\varepsilon+o(1)}\}$. So

$$\log |L_n(\varepsilon)| = \log \binom{d}{d^{\varepsilon+o(1)}} = d^{\varepsilon+o(1)} = (\log n)^{\varepsilon+o(1)}$$

as required.

Finally, we want an analog of a random walk with non-trivial boundary, expressed using property (ii) below.

(C): Expander-like. Say the sequence $(X_t^{(n)})$ is *expander-like* if
(i) $\tau_1 = \Theta(\tau_2 \log n)$
(ii) every asymptotic ε-neighborhood $(L_n(\varepsilon))$ has

$$\log |L_n(\varepsilon)| = (\log n)^{o(1)} \text{ as } n \to \infty$$

(iii) The cut-off property holds.
Recall from Chapter 9 section 1 that, for symmetric graphs which are r-regular expanders for fixed r, we have $\tau_2(n) = \Theta(1)$ and $\tau_1(n) = \Theta(\log n)$. But it is not known whether properties (ii) and (iii) always hold in this setting.

3 Random Walks in Random Environments

In talking about random walk on a weighted graph, we have been assuming the graph is fixed. It is conceptually only a minor modification to consider the case where the "environment" (the graph or the edge-weights) is itself first given in some specified random manner. This has been studied in several rather different contexts, and we will give a brief description of known results without going into many details.

Quantities like our mixing time parameters τ from Chapter 4 are now random quantities $\boldsymbol{\tau}$. In general we shall use **boldface** for quantities depending on the realization of the environment but not depending on a realization of the walk.

3.1 Mixing times for some random regular graphs

There is a body of work on estimating mixing times for various models of random regular graph. We shall prove two simple results which illustrate two basic techniques, and record some of the history in the Notes.

The first result is Proposition 1.2.1 of Lubotzky [37]. This illustrates the technique of proving expansion (i.e., upper-bounding the Cheeger time constant τ_c) by direct counting arguments in the random graph.

Proposition 11 *Let $G_{k,n}$ be the $2k$-regular random graph on vertices $\{1, 2, \ldots, n\}$ with edges $\{(i, \pi_j(i)) : 1 \le i \le n, 1 \le j \le k\}$, where $(\pi_j, 1 \le i \le k)$ are independent uniform random permutations of $\{1, 2, \ldots, n\}$. Write $\boldsymbol{\tau}_c(k, n)$ for* *the Cheeger time constant for random walk on $G_{k,n}$. Then for fixed $k \geq 7$,*

$$P(\boldsymbol{\tau}_c(k,n) > 2k) \to 0 \text{ as } n \to \infty.$$

Note that a realization of $G_{k,n}$ may be disconnected (in which case $\tau_c = \infty$) and have self-loops and multiple edges.

Outline of proof. Suppose a realization of the graph has the property

$$|A| \leq n/2 \Rightarrow |\partial A| \geq |A|/2 \tag{59}$$

where $\partial A := \{\text{edges}(i,j) : i \in A, j \in A^c\}$. Then

$$\tau_c = \sup_{A:1\leq|A|\leq n/2} \frac{k|A|(n-|A|)}{n|\partial A|} \leq \sup_{A:1\leq|A|\leq n/2} \frac{k|A|(n-|A|)}{n|A|/2} \leq 2k.$$

So we want to show that (59) holds with probability $\to 1$ as $n \to \infty$. If (59) fails for some A with $|A| = a$, then there exists B with $|B| = \lfloor \frac{3}{2}a \rfloor = b$ such that

$$\pi_j(A) \subseteq B,\ 1 \leq j \leq k \tag{60}$$

(just take $B = \cup_j \pi_j(A_j)$ plus, if necessary, arbitrary extra vertices). For given A and B, the chance of (60) equals $((b)_a/(n)_a)^k$, where $(n)_a := \prod_{r=0}^{a-1}(n-r)$. So the chance that (59) fails is at most

$$\sum_{1\leq a\leq n/2} q(a), \text{ where } q(a) = \binom{n}{a}\binom{n}{b}\ ((b)_a/(n)_a)^k.$$

So it suffices to verify $\sum_{1\leq a\leq n/2} q(a) \to 0$. And this is a routine but tedious verification (see Notes). □

Of course the bound on τ_c gives, via Cheeger's inequality, a bound on τ_2, and thence a bound on τ_1 via $\tau_1 = O(\tau_2 \log n)$. But Proposition 11 is unsatisfactory in that these bounds get *worse* as k increases, whereas intuitively they should get better. For bounds on τ_1 which improve with k we turn to the second technique, which uses the "$L^1 \leq L^2$" inequality to bound the variation threshold time τ_1. Specifically, recall (Chapter 3 Lemma 8b) that for an n-state reversible chain with uniform stationary distribution, the variation distance $d(t)$ satisfies

$$d(t) \leq 2 \max_i (np_{ii}(2t) - 1)^{1/2}. \tag{61}$$

This is simplest to use for random walk on a group, as illustrated by the following result of Roichman [52].

Proposition 12 *Fix $\alpha > 1$. Given a group G, let S be a random set of $k = \lfloor \log^\alpha |G| \rfloor$ distinct elements of G, and consider random walk on the associated Cayley graph with edges $\{(g, gs) : g \in G, s \in S \cup S^{-1}\}$. For any sequence of groups with $|G| \to \infty$,*

$$P(\boldsymbol{\tau}_1 > t_1) \to 0, \ \ \textit{where } t_1 = \lceil \tfrac{\alpha}{\alpha-1} \tfrac{\log |G|}{\log k} \rceil.$$

Proof. We first give a construction of the random walk jointly with the random set S. Write $A = \{a, b, \ldots\}$ for a set of k symbols, and write $\bar{A} = \{a, a^{-1}, b, b^{-1}, \ldots\}$. Fix $t \geq 1$ and let $(\xi_s, 1 \leq s \leq t)$ be independent uniform on $\bar{A}$. Choose $(g(a), a \in A)$ by uniform sampling without replacement from G, and set $g(a^{-1}) = (g(a))^{-1}$. Then the process $(X_s; 1 \leq s \leq t)$ constructed via $X_s = g(\xi_1)g(\xi_2)\ldots g(\xi_s)$ is distributed as the random walk on the random Cayley graph, started at the identity ι. So $P(X_t = \iota) = E\mathbf{p}_{\iota\iota}(t)$ where $\mathbf{p}_{\iota\iota}(t)$ is the t-step transition probability in the random environment, and by (61) it suffices to take $t = t_1$ (for t_1 defined in the statement of the Proposition) and show

$$|G|P(X_{2t} = \iota) - 1 \to 0. \tag{62}$$

To start the argument, let $J(2t)$ be the number of distinct values taken by $(\langle \xi_s \rangle, 1 \leq s \leq 2t)$, where we define $\langle a \rangle = \langle a^{-1} \rangle = a$. Fix $j \leq t$ and $1 \leq s_1 < s_2 < \ldots < s_j \leq 2t$. Then

$$P(J(2t) = j | \langle \xi_{s_i} \rangle \text{ distinct for } 1 \leq i \leq j) = (j/k)^{2t-j} \leq (t/k)^t.$$

By considering the possible choices of (s_i),

$$P(J(2t) = j) \leq \binom{2t}{j} (t/k)^t.$$

Since $\sum_j \binom{2t}{j} = 2^{2t}$ we deduce

$$P(J(2t) \leq t) \leq (4t/k)^t. \tag{63}$$

Now consider the construction of X_{2t} given above. We claim

$$P(X_{2t} = \iota | \xi_s, 1 \leq s \leq 2t) \leq \frac{1}{|G| - 2t} \text{ on } \{J(2t) > t\}. \tag{64}$$

For if $J(2t) > t$ then there exists some $b \in A$ such that $\langle \xi_s \rangle = b$ for exactly one value of s in $1 \leq s \leq 2t$. So if we condition also on $\{g(a); a \in A, a \neq b\}$, then $X_{2t} = g_1 g(b) g_2$ or $g_1 g(b)^{-1} g_2$ where g_1 and g_2 are determined by the conditioning, and then the conditional probability that $P(X_{2t} = \iota)$ is the conditional probability of $g(b)$ taking a particular value, which is at most $1/(|G| - 2t)$.

Combining (64) and (63),

$$P(X_{2t} = \iota) \leq (4t/k)^t + \tfrac{1}{|G|-2t} \leq (4t/k)^t + \tfrac{1}{|G|} + O(\tfrac{t}{|G|^2}).$$

So proving (62) reduces to proving

$$|G|(4t/k)^t + t/|G| \to 0$$

and the definition of t was made to ensure this.

3.2 Randomizing infinite trees

Simple random walk on the infinite regular tree is a fundamental process, already discussed in section 2.6. There are several natural ways to randomize the environment; we could take an infinite regular tree and attach random edge-weights; or we could consider a Galton–Watson tree, in which numbers of children are random. Let us start by considering these possibilities simultaneously. Fix a distribution $(\xi; W_1, W_2, \ldots, W_\xi)$ where

$$\xi \geq 1; \;\; P(\xi \geq 2) > 0; \;\; W_i > 0, i \leq \xi. \tag{65}$$

Note the (W_i) may be dependent. Construct a tree via:

> the root ϕ has ξ children, and the edge (ϕ, i) to the ith child has weight W_i; repeat recursively for each child, taking independent realizations of the distribution (65).

So the case $\xi_i \equiv r - 1$ gives the randomly-weighted r-ary tree (precisely, the modification where the root has degree $r - 1$ instead of r), and the case $W_i \equiv 1$ gives a Galton–Watson tree. As in Chapter 3 section 2 to each realization of a weighted graph we associate a random walk with transition probabilities proportional to edge-weights. Since random walk on the unweighted r-ary tree is transient, a natural first issue is prove transience in this "random environment" setting. In terms of the electrical network analogy (see comment below Theorem 5), interpreting W as conductance, we want to know whether the (random) resistance $\mathbf{R}$ between ϕ and ∞ is a.s.

finite. By considering the children of ϕ, it is clear that the distribution of $\mathbf{R}$ satisfies

$$\mathbf{R} \stackrel{d}{=} \left(\sum_{i=1}^{\xi}(\mathbf{R}_i + W_i^{-1})^{-1}\right)^{-1} \tag{66}$$

where the $(\mathbf{R}_i)$ are independent of each other and of $(\xi; W_1, W_2, \ldots, W_\xi)$, and $\mathbf{R}_i \stackrel{d}{=} \mathbf{R}$. But $\hat{\mathbf{R}} \equiv \infty$ is a solution of (66), so we need some work to actually prove that $\mathbf{R} < \infty$.

Proposition 13 *The resistance* $\mathbf{R}$ *between* ϕ *and* ∞ *satisfies* $\mathbf{R} < \infty$ *a.s..*

Proof. Write $\mathbf{R}^{(k)}$ for the resistance between ϕ and height k (i.e. the height-k vertices, all shorted together). Clearly $\mathbf{R}^{(k)} \uparrow \mathbf{R}$ as $k \to \infty$, and analogously to (66)

$$\mathbf{R}^{(k+1)} \stackrel{d}{=} \left(\sum_{i=1}^{\xi}(\mathbf{R}_i^{(k)} + W_i^{-1})^{-1}\right)^{-1}$$

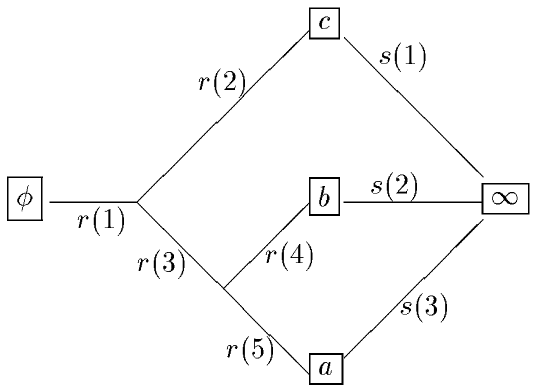
where the $(\mathbf{R}_i^{(k)})$ are independent of each other and of $(\xi; W_1, W_2, \ldots, W_\xi)$, and $\mathbf{R}_i^{(k)} \stackrel{d}{=} \mathbf{R}^{(k)}$.

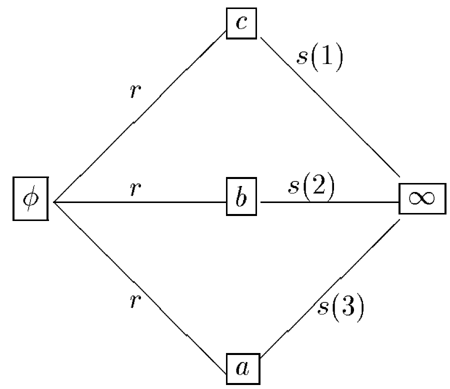
Consider first the special case $\xi \equiv 3$. Choose x such that $P(W_i^{-1} > x \text{ for some } i) \leq 1/16$. Suppose inductively that $P(\mathbf{R}^{(k)} > x) \leq 1/4$ (which holds for $k = 0$ since $\mathbf{R}^{(0)} = 0$). Then

$$P(\mathbf{R}_i^{(k)} + W_i^{-1} > 2x \text{ for at least 2 } i\text{'s }) \leq \tfrac{1}{16} + 3(\tfrac{1}{4})^2 \leq \tfrac{1}{4}.$$

This implies $P(\mathbf{R}^{(k+1)} > x) \leq 1/4$, and the induction goes through. Thus $P(\mathbf{R} > x) \leq 1/4$. By (66) $p := P(\mathbf{R} = \infty)$ satisfies $p = p^3$, so $p = 0$ or 1, and we just eliminated the possibility $p = 1$. So $\mathbf{R} < \infty$ a.s..

Reducing the general case to the special case involves a comparison idea, illustrated by the figure.





Here the edge-weights are *resistances*. In the left network, ϕ is linked to $\{a, b, c\}$ via an arbitrary tree, and in the right network, this tree is replaced by three direct edges, each with resistance $r = 3(r(1) + r(2) + \ldots + r(5))$. We claim that this replacement can only *increase* the resistance between ϕ and ∞. This is a nice illustration of Thompson's principle (Chapter 3 section 7.1) which says that in a realization of either graph, writing $r^*(e)$ for resistance and suming over undirected edges e,

$$R_{\phi\infty} = \inf_{\mathbf{f}} \sum_e r^*(e) f^2(e)$$

where $\mathbf{f} = (f(e))$ is a unit flow from ϕ to ∞. Let $\mathbf{f}$ be the minimizing flow in the right network; use $\mathbf{f}$ to define a flow $\mathbf{g}$ in the left network by specifying that the flow through a (resp. b, c) is the same in the left network and the right network. It is easy to check

$$\text{(left network)} \sum_e r(e) g^2(e) \leq \text{(right network)} \sum_e r(e) f^2(e)$$

and hence the resistance $R_{\phi\infty}$ can indeed only be less in the left network.

In the general case, the fact $P(\xi \geq 2) > 0$ implies that the number of individuals in generation g tends to ∞ a.s. as $g \to \infty$. So in particular we can find 3 distinct individuals $\{A, B, C\}$ in some generation G. Retain the edges linking ϕ with these 3 individuals, and cut all other edges within the first G generations. Repeat recursively for descendants of $\{A, B, C\}$. This procedure constructs an infinite subtree, and it suffices to show that the resistance between ϕ and ∞ in the subtree is a.s. finite. By the comparison argument above, we may replace the network linking ϕ to $\{A, B, C\}$ by three direct edges with the same (random) resistance, and similarly for each stage of the construction of the subtree; this gives another tree $\mathcal{T}$, and it suffices to shows its resistance is finite a.s.. But $\mathcal{T}$ fits the special case $\xi \equiv 3$. ■

It is not difficult (we won't give details) to show that the distribution of $\mathbf{R}$ is the *unique* distribution on $(0, \infty)$ satisfying (66). It does seem difficult to say anything explicit about the distribution of $\mathbf{R}$ in Proposition 13. One can get a little from comparison arguments. On the binary tree ($\xi \equiv 2$), by using the exact potential function and the exact flows from the unweighted case as "test functions" in the Dirichlet principle and Thompson's principle, one obtains

$$E\mathbf{R} \leq EW^{-1}; \quad E\mathbf{R}^{-1} \leq EW.$$

3.3 Bias and speed

Lyons et al [38, 39, 40], summarized in [42] Chapter 10, have studied in detail questions concerning a certain model of biased random walk on deterministic and random infinite trees. Much of their focus is on topics too sophisticated (boundary theory, dimension) to recount here, but let us give one simple result.

Consider the unweighted Galton–Watson tree with offspring distribution $\mu = \text{dist}\ (\xi)$, i.e., the case $W_i \equiv 1$ of (65). Fix a parameter $0 \leq \lambda < \infty$. In the *biased random walk* X_t, from a vertex with r children the walker goes to any particular child with probability $1/(\lambda + r)$, and to the parent with probability $\lambda/(\lambda + r)$. It turns out [40] that the biased random walk is recurrent for $\lambda \geq E\xi$ and transient for $\lambda < E\xi$. We will just prove one half of that result.

Proposition 14 *The biased random walk is a.s. recurrent for $\lambda \geq E\xi$.*

Proof. We use a "method of fictitious roots". That is, to the root ϕ of the Galton-Watson tree we append an extra edge to a "fictitious" root ϕ^*, and we consider random walk on this extended tree (rooted at ϕ^*). Write $\mathbf{q}$ for the probability (conditional on the realization of the tree) that the walk started at ϕ never hits ϕ^*. It will suffice to prove $P(\mathbf{q} = 0) = 1$. Fix a realization of the tree, in which ϕ has z children. Then

$$q = \sum_{i=1}^{z} \frac{1}{\lambda + z}(q_i + (1 - q_i)q)$$

where q_i is the probability (on this realization) that the walk started at the i'th child of ϕ never hits ϕ. Rearrange to see $q = (\sum_i q_i)/(\lambda + \sum_i q_i)$. So on the random tree we have

$$\mathbf{q} \stackrel{d}{=} \frac{\sum_{i=1}^{\xi} \mathbf{q}_i}{\lambda + \sum_{i=1}^{\xi} \mathbf{q}_i}$$

where the $(\mathbf{q}_i)$ are independent of each other and ξ, and $\mathbf{q}_i \stackrel{d}{=} \mathbf{q}$. Applying Jensen's inequality to the concave function $x \to \frac{x}{x+\lambda}$ shows

$$E\mathbf{q} \leq \frac{(E\xi)(E\mathbf{q})}{\lambda + (E\xi)(E\mathbf{q})}.$$

By considering the relevant quadratic equation, one sees that for $\lambda \geq E\xi$ this inequality has no solution with $E\mathbf{q} > 0$. So $E\mathbf{q} = 0$, as required. ■

In the transient case, we expect there to exist a non-random *speed* $s(\lambda, \mu) \leq 1$ such that

$$t^{-1} d(X_t, \phi) \to s(\lambda, \mu) \text{ a.s. as } t \to \infty. \tag{67}$$

Lyons et al [40] show that, when $E\xi < \infty$, (67) is indeed true and that $s(\lambda, \mu) > 0$ for all $1 \leq \lambda < E\xi$. Moreover in the unbiased ($\lambda = 1$) case there is a simple formula [39]

$$s(1, \mu) = E\tfrac{\xi - 1}{\xi + 1}.$$

There is apparently no such simple formula for $s(\lambda, \mu)$ in general. See Lyons et al [41] for several open problems in this area.

3.4 Finite random trees

Cayley's formula ([55] p. 25) says there are n^{n-2} different trees on $n \geq 2$ labeled vertices $\{1, 2, \ldots, n\}$. Assuming each such tree to be equally likely gives one tractable definition (there are others) of *random n-tree* $\mathbf{T}_n$. One can combine the formulas from Chapter 5 section 3 for random walks on general trees with known distributional properties of $\mathbf{T}_n$ to get a variety of formulas for random walk on $\mathbf{T}_n$, an idea going back to Moon [45].

As an illustration it is known [45] that the distance $\mathbf{d}(1, 2)$ between vertex 1 and vertex 2 in $\mathbf{T}_n$ has distribution

$$P(\mathbf{d}(1, 2) = k) = (k + 1) n^{-k} (n - 2)_{k-1}, \ 1 \leq k \leq n - 1$$

where $(m)_s = m(m - 1) \cdots (m - s + 1)$. Routine calculus gives

$$E\mathbf{d}(1, 2) \sim \sqrt{\pi/2}\, n^{1/2}. \tag{68}$$

Now on any n-vertex tree, the mean hitting time $t(i,j) = E_i T_j$ satisfies

$$t(i,j) + t(j,i) = 2(n-1)d(i,j) \tag{69}$$

(Chapter 5 (84)), and so

$$E\mathbf{t}(1,2) = (n-1)E\mathbf{d}(1,2).$$

Combining with (68),

$$E\mathbf{t}(1,2) \sim \sqrt{\pi/2}\; n^{3/2}. \tag{70}$$

Instead of deriving more formulas of this type for random walk on $\mathbf{T}_n$, let's jump to the bottom line. It turns out that all the mixing and hitting time parameters $\boldsymbol{\tau}_u^{(n)}$ of Chapter 4, and the analogous "mean cover time" parameters of Chapter 6, are of order $n^{3/2}$ but are *random* to first order: that is,

$$n^{-3/2}\boldsymbol{\tau}_u^{(n)} \stackrel{d}{\to} \boldsymbol{\tau}_u^{(\infty)} \text{ as } n \to \infty \tag{71}$$

for non-deterministic limits $\boldsymbol{\tau}_u^{(\infty)}$. The fact that all these parameters have the same order is of course reminiscent of the cases of the n-cycle and n-path (Chapter 5 Examples 7 and 8), where all the parameters are $\Theta(n^2)$. And the sophisticated explanation is the same: one can use the weak convergence paradigm (section 1.1). In the present context, the random tree $\mathbf{T}_n$ rescales to a limit *continuum random tree* $\mathbf{T}_\infty$, and the random walk converges (with time rescaled by $n^{3/2}$ and space rescaled by $n^{1/2}$) to Brownian motion on $\mathbf{T}_\infty$, and (analogously to section 1.1) the rescaled limits of the parameters are just the corresponding parameters for the Brownian motion. See the Notes for further comments.

3.5 Randomly-weighted random graphs

Fix a distribution W on $(0,\infty)$ with $EW < \infty$. For each n consider the random graph $G(n, p(n))$, that is the graph on n vertices where each possible edge has chance $p(n)$ to be present. Attach independent random conductances, distributed as W, to the edges. Aspects of this model were studied by Grimmett and Kesten [28]. As they observe, much of the behavior is intuitively rather clear, but technically difficult to prove: we shall just give the intuition.

Case (i): $p(n) = \mu/n$ for fixed $1 < \mu < \infty$. Here the number of edges at vertex 1 is asymptotically Poisson(μ), and the part of the graph within a fixed distance d of vertex 1 is asymptotically like the first d generations

in the random family tree $\mathcal{T}^\infty$ of a Galton–Watson branching process with Poisson(μ) offspring distribution, with independent edge-weights attached. This tree essentially fits the setting of Proposition 13, except that the number of offspring may be zero and so the tree may be finite, but it is not hard to show (modifying the proof of Proposition 13) that the resistance $\mathbf{R}$ in $\mathcal{T}^\infty$ between the root and ∞ satisfies $\{\mathbf{R} < \infty\} = \{\mathcal{T}^\infty \text{ is infinite}\}$ and its distribution is characterized by the analog of (
refRdef). The asymptotic approximation implies that, for $d(n) \to \infty$ slowly, the resistance $\mathbf{R}_{n,d(n)}$ between vertex 1 and the depth-$d(n)$ vertices of $G(n,p(n))$ satisfies $\mathbf{R}_{n,d(n)} \stackrel{d}{\to} \mathbf{R}^{(1)} \stackrel{d}{=} \mathbf{R}$. We claim that the resistance $\mathbf{R}^{(n)}_{1,2}$ between vertices 1 and 2 of $G(n,p(n))$ satisfies

$$\mathbf{R}^{(n)}_{1,2} \stackrel{d}{\to} \mathbf{R}^{(1)} + \mathbf{R}^{(2)}; \text{ where } \mathbf{R}^1 \text{ and } \mathbf{R}^2 \text{ are independent copies of } \mathbf{R}\ .$$

The lower bound is clear by shorting, but the upper bound requires a complicated construction to connect the two sets of vertices at distances $d(n)$ from vertices 1 and 2 in such a way that the effective resistance of this connecting network tends to zero.

The number of edges of the random graph is asymptotic to $n\mu/2$. So the total edge weight $\sum_i \sum_j W_{ij}$ is asymptotic to $n\mu EW$, and by the commute interpretation of resistance the mean commute time $\mathbf{C}^{(n)}_{1,2}$ for random walk on a realization of the graph satisfies

$$n^{-1}\mathbf{C}^{(n)}_{1,2} \stackrel{d}{\to} \mu EW(\mathbf{R}^{(1)} + \mathbf{R}^{(2)}).$$

Case (ii): $p(n) = o(1) = \Omega(n^{\varepsilon-1})$, some $\varepsilon > 0$. Here the degree of vertex 1 tends to ∞, and it is easy to see that the (random) stationary probability π_1 and the (random) transition probabilities and stationary distribution the random walk satisfy

$$\max_j \mathbf{p}_{1,j} \stackrel{p}{\to} 0,\ n\boldsymbol{\pi}_1 \stackrel{p}{\to} 1 \text{ as } n \to \infty.$$

So for fixed $k \geq 1$, the k-step transition probabilities satisfy $p^{(k)}_{11} \stackrel{p}{\to} 0$ as $n \to \infty$. This suggests, but it is technically hard to *prove*, that the (random) fundamental matrix $\mathbf{Z}$ satisfies

$$\mathbf{Z}_{11} \stackrel{p}{\to} 1 \text{ as } n \to \infty. \tag{72}$$

Granted (72), we can apply Lemma 11 of Chapter 2 and deduce that the mean hitting times $\mathbf{t}(\pi,1) = E_\pi T_1$ on a realization of the random graph satisfies

$$n^{-1}\mathbf{t}(\pi,1) = \frac{\mathbf{Z}_{11}}{n\boldsymbol{\pi}_1} \stackrel{p}{\to} 1, \text{ as } n \to \infty. \tag{73}$$

3.6 Random environments in d dimensions

The phrase *random walk in random environment* (RWRE) is mostly used to denote variations of the classical "random flight in d dimensions" model. Such variations have been studied extensively in mathematical physics as well as theoretical probability, and the monograph of Hughes [29] provides thorough coverage. To give the flavor of the subject we quote one result, due to Boivin [8].

Theorem 15 *Assign random conductances* (w_e) *to the edges of the two-dimensional lattice* $\mathbf{Z}^2$*, where*
(i) the process (w_e) *is stationary ergodic.*
(ii) $c_1 \leq w_e \leq c_2$ *a.s., for some constants* $0 < c_1 < c_2 < \infty$.
Let $(X_t; t \geq 0)$ *be the associated random walk on this weighted graph,* $X_0 = 0$. *Then* $t^{-1/2} X_t \stackrel{d}{\to} Z$ *where* Z *is a certain two-dimensional Normal distribution, and moreover this convergence holds for the conditional distribution of* X_t *given the environment, for almost all environments.*

4 Notes on Chapter 13

Section 1. Rigorous setup for discrete-time continuous-space Markov chains is given concisely in Durrett [20] section 5.6 and in detail in Meyn and Tweedie [44]. For the more sophisticated continuous-time setting see e.g. Rogers and Williams [51]. Aldous et al [4] prove some of the Chapter 4 mixing time inequalities in the discrete-time continuous-space setting.

The central limit theorem (for sums of functions of a Markov chain) does not automatically extend from the finite-space setting (Chapter 2 Theorem 17) to the continuous-space setting: regularity conditions are required. See [44] Chapter 17. But a remarkable result of Kipnis - Varadhan [33] shows that for *stationary* reversible chains the central limit theorem remains true under very weak hypotheses.

Sections 1.1 - 1.3. The eigenvalue analysis is classical. The reflection coupling goes back to folklore; see e.g. Lindvall [36] Chapter 6 for applications to multidimensional diffusions and Matthews [43] for Brownian motion in a polyhedron. Burdzy and Kendall [13] give a careful study of coupling for Brownian motion in a triangle. Chen [15] surveys use of coupling to estimate spectral gaps for diffusions on manifolds.

Here is a more concise though less explicit expression for $\bar{d}(t)$ at (6) (and hence for $G(t)$at (1)). Consider Brownian motions B° on the circle started at 0 and at 1/2. At any time t, the former distribution dominates the latter on the interval $(-1/4, 1/4)$ only, and so

$$\begin{aligned}
\bar{d}(t) &= P_0(B_t^\circ \in (-1/4,1/4)) - P_{1/2}(B_t^\circ \in (-1/4,1/4)) \\
&= P_0(B_t^\circ \in (-1/4,1/4)) - P_0(B_t^\circ \in (1/4,3/4)) \\
&= 2P_0(B_t^\circ \in (-1/4,1/4)) \ - 1 \\
&= 2P((t^{1/2}Z) \bmod 1 \in (-1/4,1/4)) - 1
\end{aligned}$$

where Z has Normal$(0,1)$ law. We quoted this expression in the analysis of Chapter 5 Example 7

Section 1.5. Janvresse [30], Porod [50] and Rosenthal [53] study mixing times for other flights on matrix groups involving rotations and reflections; Porod [49] also discusses more general Lie groups.

Section 1.6. The mathematical theory has mostly been developed for classes of nested fractals, of which the Sierpinski gasket is the simplest. See Barlow [6], Lindstrøm [35], Barlow [7] for successively more detailed treatments. Closely related is Brownian motion on the continuum random tree, mentioned in section 3.4.

One-dimensional diffusions. The continuous-space analog of a birth-and-death process is a one-dimensional *diffusion* (X_t), described by a stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$$

where B_t is standard Brownian motion and $\mu(\cdot)$ and $\sigma(\cdot)$ are suitably regular specified functions. See Karlin and Taylor [32] for non-technical introduction. Theoretical treatments standardize (via a one-to-one transformation $R \to R$) to the case $\mu(\cdot) = 0$, though for our purposes the standardization to $\sigma(\cdot) = 1$ is perhaps more natural. In this case, if the formula

$$f(x) \propto \exp\left(\int^x 2\mu(y)dy\right)$$

can give a *density* function $f(x)$ then f is the stationary density. Such diffusions relate to two of our topics.

(i) For MCMC, to estimate a density $f(x) \propto \exp(-H(x))$, one can in principle simulate the diffusion with $\sigma(x) = 1$ and $\mu(x) = -H'(x)/2$. This idea was used in Chapter MCMC section 5.

(ii) Techniques for bounding the relaxation time for one-dimensional diffusions parallel techniques for birth-and-death chains [16].

Section 2. We again refer to Woess [59] for systematic treatment of random walks on infinite graphs.

Our general theme of using the infinite case to obtain limits for finite chains goes back at least to [1], in the case of Z^d; similar ideas occur in the study of interacting particle systems, relating properties of finite and infinite-site models.

Section 2.2. There is a remarkable connection between recurrence of reversible chains and a topic in Bayesian statistics: see Eaton [21]. Properties of random walk on fractal-like infinite subsets of Z^d are studied by Telcs [56, 57].

Section 2.9. One view of (Y_t) is as one of several "toy models" for the notion of random walk on fractional-dimensional lattice. Also, when we seek to study complicated variations of random walk, it is often simpler to use the hierarchical lattice than Z^d itself. See for instance the sophisticated study of self-avoiding walks by Brydges et al [11]; it would be interesting to see whether direct combinatorial methods could reproduce their results.

Section 2.10. Another class of sequences could be defined as follows. There are certain continuous-time, continuous-space reversible processes on

compact spaces which "hit points" and for which $\tau_0 < \infty$; for example
(i) Brownian motion on the circle (section 1.1)
(ii) Brownian motion on certain fractals (section 1.6)
(iii) Brownian motion on the continuum random tree (section 3.4).
So for a sequence of finite-state chains one can define the property

$$\tau_0(n)/\tau_2(n) \text{ is bounded}$$

as the finite analog of "diffusions which hit points". This property holds for the discrete approximations to the examples above: (i) random walk on the n-cycle
(i) random walk on graphs approximating fractals (section 1.6)
(iii) random walk on random n-vertex trees (section 3.4).

Equivalence (58) is hard to find in textbooks. The property "trivial boundary" is equivalent to "no non-constant bounded harmonic functions" ([59] Corollary 24.13), which is equivalent ([58] Theorem 6.5.1) to existence of successful *shift-coupling* of two versions of the chain started at arbitrary points. The property (58) is equivalent ([58] Theorem 4.9.4) to existence of successful *couplings*. In the setting of interest to us (continuized chains on countable space), existence of a shift-coupling (a priori weaker than existence of a coupling) for the discrete-time chain implies existence of a coupling for the continuous-time chain, by using independence of jump chain and hold times.

Section 3. Grimmett [27] surveys "random graphical networks" from a somewhat different viewpoint, emphasising connections with statistical physics models.

Section 3.1. More precise variants of Proposition 11 were developed in the 1970s, e.g. [48, 17]. Lubotzky [37], who attributes this method of proof of Proposition 11 to Sarnak [54], asserts the result for $k \geq 5$ but our own calculations give only $k \geq 7$. Note that Proposition 11 uses the *permutation model* of a $2k$-regular random graph. In the alternative *uniform* model we put $2k$ balls labeled 1, $2k$ balls labeled 2, ... and $2k$ balls labeled n into a box; then draw without replacement two balls at a time, and put an edge between the two vertices. In both models the graphs may be improper (multiple edges or self-loops) and unconnected, but are in fact proper with probability $\Omega(1)$ and connected with probability $1-o(1)$ as $n \to \infty$ for fixed k. Behavior of τ_c in the uniform model is implicitly studied in Bollobás [9]. The L^2 ideas underlying the proof of Proposition 12 were used by Broder and Shamir [10], Friedman [23] and Kahn and Szemerédi [24] in the setting of the permutation model of random r-regular graphs. One result is that $\beta \equiv \max(\lambda_2, -\lambda_n) = O(\frac{2\sqrt{2r-1}}{r})$ with probability $1 - o(1)$. Further results in the "random Cayley graph" spirit of Proposition 12 can be found in [5, 18, 47].

Section 3.2. The monograph of Lyons and Peres [42] contains many more results concerning random walks on infinite deterministic and Galton–Watson trees. A challenging open problem noted in [41] is to prove that $\mathbf{R}$ has absolutely continuous distribution when ξ is non-constant. The method of fictitious roots used in Proposition 14 is also an ingredient in the analysis of cover times on trees [3].

Section 3.4. Moon [45] gives further results in the spirit of (70), e.g. for variances of hitting times. The fact that random walk on $\mathbf{T}_n$ rescales to Brownian motion on a "continuum random tree" $\mathbf{T}_\infty$ was outlined in Aldous [2] section 5 and proved in Krebs [34]. While this makes the "order $n^{3/2}$" property (71) of the parameters essentially obvious, it is still difficult to get explicit information about the limit distributions $\boldsymbol{\tau}^{(\infty)}$. What's known [2] is

(a) $E\boldsymbol{\tau}_0^{(\infty)} = \sqrt{\pi/2}$, as suggested by (70);

(b) $\boldsymbol{\tau}^{(\infty)*} = \frac{8}{3}\sqrt{2\pi}$, from (69) and the known asymptotics for the diameter of $\mathbf{T}_n$;

(c) The "cover and return" time C_n^+ appearing in Chapter 6 satisfies $n^{-3/2}EC_n^+ \to 6\sqrt{2\pi}$, modulo some technical issues.

Section 3.5. Grimmett and Kesten [28] present their results in terms of resistances, without explicitly mentioning random walk, so that results like (73) are only implicit in their work.

References

[1] D.J. Aldous. On the time taken by random walks on finite groups to visit every state. *Z. Wahrsch. Verw. Gebiete*, 62:361–374, 1983.

[2] D.J. Aldous. The continuum random tree II: an overview. In M.T. Barlow and N.H. Bingham, editors, *Stochastic Analysis*, pages 23–70. Cambridge University Press, 1991.

[3] D.J. Aldous. Random walk covering of some special trees. *J. Math. Analysis Appl.*, 157:271–283, 1991.

[4] D.J. Aldous, L. Lovász, and P. Winkler. Mixing times for uniformly ergodic Markov chains. *Stochastic Process. Appl.*, 71:165–185, 1997.

[5] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Struct. Alg.*, 5:271–284, 1994.

[6] M.T. Barlow. Random walks and diffusions on fractals. In *Proc. ICM Kyoto 1990*, pages 1025–1035. Springer–Verlag, 1991.

[7] M.T. Barlow. Diffusions on fractals. In P. Bernard, editor, *Lectures on Probability and Statistics*, volume 1690 of *Lecture Notes in Math.* Springer–Verlag, 1998.

[8] D. Boivin. Weak convergence for reversible random walks in a random environment. *Ann. Probab.*, 21:1427–1440, 1993.

[9] B. Bollobás. The isoperimetric number of random regular graphs. *European J. Combin.*, 9:241–244, 1988.

[10] A. Broder and E. Shamir. On the second eigenvalue of random regular graphs. In *Proc. 28th Symp. Foundations of Computer Sci.*, pages 286–294, 1987.

[11] D. Brydges, S.N. Evans, and J.Z. Imbrie. Self-avoiding walk on a hierarchical lattice in four dimensions. *Ann. Probab.*, 20:82–124, 1992.

[12] R. Bubley, M. Dyer, and M. Jerrum. An elementary analysis of a procedure for sampling points in a convex body. *Random Struct. Alg.*, 12:213–235, 1998.

[13] K. Burdzy and W.S. Kendall. Efficient Markovian couplings: Examples and counterexamples. *Ann. Appl. Probab.*, 10:362–409, 2000.

[14] A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky, and P. Tiwari. The electrical resistance of a graph captures its commute and cover times. *Comput. Complexity*, 6:312–340, 1996/7. Extended abstract originally published in *Proc. 21st ACM Symp. Theory of Computing* (1989) 574-586.

[15] M. F. Chen. Trilogy of couplings and general formulas for lower bound of spectral gap. In L. Accardi and C. Heyde, editors, *Probability Towards 2000*, number 128 in Lecture Notes in Statistics, pages 123–136. Springer–Verlag, 1996.

[16] M. F. Chen. Eigenvalues, inequalities and ergodic theory II. *Adv. Math. (China)*, 28:481–505, 1999.

[17] F.R.K. Chung. On concentrators, superconcentrators, and nonblocking networks. *Bell System Tech. J.*, 58:1765–1777, 1979.

[18] C. Dou and M. Hildebrand. Enumeration and random random walks on finite groups. *Ann. Probab.*, 24:987–1000, 1996.

[19] P.G. Doyle and J.L. Snell. *Random Walks and Electrical Networks*. Mathematical Association of America, Washington DC, 1984.

[20] R. Durrett. *Probability: Theory and Examples*. Wadsworth, Pacific Grove CA, 1991.

[21] M.L. Eaton. Admissability in quadratically regular problems and recurrence of symmetric Markov chains: Why the connection? *J. Statist. Plann. Inference*, 64:231–247, 1997.

[22] S.N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986.

[23] J. Friedman. On the second eigenvalue and random walk in random regular graphs. *Combinatorica*, 11:331–362, 1991.

[24] J. Friedman, J. Kahn, and E. Szemeredi. On the second eigenvalue in random regular graphs. In *Proc. 21st ACM Symp. Theory of Computing*, pages 587–598, 1989.

[25] A. Giacometti. Exact closed form of the return probability on the Bethe lattice. *J. Phys. A: Math. Gen.*, 28:L13–L17, 1995.

[26] P. Griffin. Accelerating beyond the third dimension: Returning to the origin in a simple random walk. *Mathematical Scientist*, 15:24–35, 1990.

[27] G. Grimmett. Random graphical networks. In *Networks and Chaos — Statistical and Probabilistic Aspects*, pages 288–301. Chapman and Hall, London, 1993. Monogr. Statist. Appl. Prob. 50.

[28] G. Grimmett and H. Kesten. Random electrical networks on complete graphs. *J. London Math. Soc. (2)*, 30:171–192, 1984.

[29] B.D. Hughes. *Random Walks and Random Environments.* Oxford University Press, 1995.

[30] E. Janvresse. Spectral gap for Kac's model of the Boltzmann equation. *Ann. Probab.*, 29:288–304, 2001.

[31] S. Karlin and H.M. Taylor. *A First Course in Stochastic Processes.* Academic Press, second edition, 1975.

[32] S. Karlin and H.M. Taylor. *A Second Course in Stochastic Processes.* Academic Press, 1981.

[33] C. Kipnis and S.R.S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104:1–19, 1986.

[34] W.B. Krebs. Brownian motion on the continuum tree. *Probab. Th. Rel. Fields*, 101:421–433, 1995.

[35] T. Lindstrøm. Brownian motion on nested fractals. *Memoirs of the A.M.S.*, 420, 1989.

[36] T. Lindvall. *Lectures on the Coupling Method.* Wiley, 1992.

[37] A. Lubotzky. *Discrete Groups, Expanding Graphs and Invariant Measures.* Birkhauser, 1994. Progress in Mathematics, vol. 125.

[38] R. Lyons and R. Pemantle. Random walk in a random environment and first-passage percolation on trees. *Ann. Probab.*, 20:125–136, 1992.

[39] R. Lyons, R. Pemantle, and Y. Peres. Ergodic theory on Galton–Watson trees: Speed of random walk and dimension of harmonic measure. *Ergodic Theory Dynam. Systems*, 15:593–619, 1995.

[40] R. Lyons, R. Pemantle, and Y. Peres. Biased random walks on Galton–Watson trees. *Probab. Th. Rel. Fields*, 106:249–264, 1996.

[41] R. Lyons, R. Pemantle, and Y. Peres. Unsolved problems concerning random walks on trees. In K. Athreya and P. Jagers, editors, *Classical and Modern Branching Processes*, pages 223–238. Springer–Verlag, 1996.

[42] R. Lyons and Y. Peres. *Probability on Trees and Networks*. Cambridge University Press, 2002. In preparation.

[43] P.C. Matthews. Mixing rates for Brownian motion in a convex polyhedron. *J. Appl. Probab.*, 27:259–268, 1990.

[44] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer–Verlag, 1993.

[45] J.W. Moon. Random walks on random trees. *J. Austral. Math. Soc.*, 15:42–53, 1973.

[46] J.R. Norris. *Markov Chains*. Cambridge University Press, 1997.

[47] I. Pak. Random walk on finite groups with few random generators. *Electon. J. Probab.*, 4(1):1–11, 1999.

[48] N. Pippenger. Superconcentrators. *SIAM J. Comput.*, 6:298–304, 1977.

[49] U. Porod. L_2 lower bounds for a special class of random walks. *Probab. Th. Rel. Fields*, 101:277–289, 1995.

[50] U. Porod. The cut-off phenomenon for random reflections. *Ann. Probab.*, 24:74–96, 1996.

[51] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales: Foundations*, volume 1. Wiley, second edition, 1994.

[52] Y. Roichman. On random random walks. *Ann. Probab.*, 24:1001–1011, 1996.

[53] J. S. Rosenthal. Random rotations: Characters and random walks on $SO(N)$. *Ann. Probab.*, 22:398–423, 1994.

[54] P. Sarnak. *Some Applications of Modular Forms*. Cambridge University Press, 1990. Cambridge Tracts in Math. 99.

[55] R. Stanley. *Enumerative Combinatorics, Vol. 2.* Cambridge University Press, 1999.

[56] A. Telcs. Spectra of graphs and fractal dimensions I. *Probab. Th. Rel. Fields*, 85:489–497, 1990.

[57] A. Telcs. Spectra of graphs and fractal dimensions II. *J. Theoretical Probab.*, 8:77–96, 1995.

[58] H. Thorisson. *Coupling, Stationarity and Regeneration.* Springer–Verlag, 2000.

[59] W. Woess. *Random Walks on Infinite Graphs and Groups.* Number 138 in Cambridge Tracts Math. Cambridge Univ. Press, 2000.