Chapter 3

Martingale techniques

Martingales are a central tool in probability theory. In this chapter we illustrate their use, as well as some related concepts, on a number of applications in discrete probability. We begin with a quick review.

3.1 Background

To be written. See [Dur10, Sections 4.1, 5.2, 5.7].

3.1.1 Stopping times

3.1.2 ▷ *Markov chains: exponential tail of hitting times, and some cover time bounds*

To be written. See [AF, Sections 2.4.3 and 2.6].

3.1.3 Martingales

3.1.4 ▷ *Percolation on trees: critical regime*

To be written. See [Per09, Sections 2 and 3].

3.2 Concentration for martingales

The Chernoff-Cramér method extends naturally to martingales. This observation leads to powerful concentration inequalities that apply beyond the case of sums of

---

Version: October 27, 2014

independent variables.*

3.2.1 Azuma-Hoeffding inequality

The main result of this section is the following generalization of Hoeffding's inequality, Theorem 2.50.

Theorem 3.1 (Azuma-Hoeffding inequality). *Let $(Z_t)_{t\in\mathbb{Z}_+}$ be a martingale with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}_+}$. Assume that (Z_t) has* bounded increments*, i.e., there are finite, nonnegative constants (c_t) such that for all $t \geq 1$, almost surely,* *bounded increments*

$$|Z_t - Z_{t-1}| \leq c_t.$$

Then for all $\beta > 0$

$$\mathbb{P}[Z_t - \mathbb{E}Z_t \geq \beta] \leq \exp\left(-\frac{\beta^2}{2\sum_{r\leq t} c_r^2}\right).$$

Applying this inequality to $(-Z_t)$ gives a tail bound in the other direction.

Proof of Theorem 3.1. Because $(Z_t - Z_0)_t = (Z_t - \mathbb{E}Z_t)_t$ is also a martingale with bounded increments we assume w.l.o.g. that $Z_0 = 0$. As in the Chernoff-Cramér method, we start by applying (the exponential version of) Markov's inequality

$$\mathbb{P}[Z_t \geq \beta] \leq \frac{\mathbb{E}\left[e^{sZ_t}\right]}{e^{s\beta}} = \frac{\mathbb{E}\left[e^{s\sum_{r=1}^{t}(Z_r - Z_{r-1})}\right]}{e^{s\beta}}. \tag{3.1}$$

This time, however, the terms of the sum in the exponent are not independent. Instead, to exploit the martingale property, we condition on the filtration

$$\mathbb{E}\left[\mathbb{E}\left[e^{s\sum_{r=1}^{t}(Z_r - Z_{r-1})} \,\middle|\, \mathcal{F}_{t-1}\right]\right] = \mathbb{E}\left[e^{s\sum_{r=1}^{t-1}(Z_r - Z_{r-1})}\,\mathbb{E}\left[e^{s(Z_t - Z_{t-1})} \,\middle|\, \mathcal{F}_{t-1}\right]\right].$$

The bounded increments property implies that, conditioned on $\mathcal{F}_{t-1}$, the random variable $Z_t - Z_{t-1}$ lies in an interval of length $2c_t$. Hence by Hoeffding's lemma, Lemma 2.51, it holds almost surely that

$$\mathbb{E}\left[e^{s(Z_t - Z_{t-1})} \,\middle|\, \mathcal{F}_{t-1}\right] \leq \exp\left(\frac{s^2(2c_t)^2/4}{2}\right) = \exp\left(\frac{c_t^2 s^2}{2}\right). \tag{3.2}$$

Arguing by induction, we obtain

$$\mathbb{E}\left[e^{sZ_t}\right] \leq \exp\left(\frac{s^2\sum_{r\leq t} c_r^2}{2}\right),$$

*Requires: Section 2.3.5.

and choosing $s = \beta / \sum_{r \leq t} c_r^2$ in (3.1) gives the result. Put differently, we have proved that Z_t is sub-Gaussian with variance factor $\sum_{r \leq t} c_r^2$. ■

In Theorem 3.1 the *martingale differences* (X_t), where $X_t := Z_t - Z_{t-1}$, are not only pairwise uncorrelated by Lemma **??**, i.e.,

martingale differences

$$\mathbb{E}[X_s X_r] = 0, \qquad \forall r \neq s,$$

but by the same argument they are in fact "mutually uncorrelated,"

$$\mathbb{E}\left[X_{j_1} \cdots X_{j_k}\right] = 0, \qquad \forall\, k \geq 1,\ \forall\, 1 \leq j_1 < \cdots < j_k.$$

This much stronger property perhaps helps explain why $\sum_{r \leq t} X_r$ is highly concentrated. This point is the subject of Exercise 3.1, which guides the reader through a slightly different proof of the Azuma-Hoeffding inequality. Compare with Exercises 2.4 and 2.5.

3.2.2 Method of bounded differences

The power of the Azuma-Hoeffding inequality is that it produces tail inequalities for quantities other than sums of independent variables.

McDiarmid's inequality The following useful corollary, which illustrates this point, is referred to as the *method of bounded differences*. It also goes under the name of *McDiarmid's inequality*. First a definition:

Definition 3.2 (Lipschitz condition). *We say that the function* $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ *is* c-Lipschitz *if for all* $i = 1, \ldots, n$ *and all* $(x_1, \ldots, x_n) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$

Lipschitz condition

$$\sup_{y, y' \in \mathcal{X}_i} \left| f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y', x_{i+1}, \ldots, x_n) \right| \leq c.$$

To see the connection with the usual definition of Lipschitz continuity, note that f above is Lipschitz continuous with Lipschitz constant c if one uses the Euclidean metric on $\mathbb{R}$ and the product (discrete) metric

$$\rho(x, x') := \sum_{i=1}^{n} \mathbb{1}_{\{x_i \neq x'_i\}},$$

for $x, x' \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$.

Corollary 3.3 (Method of bounded differences)**.** *Let $X_1, \dots, X_n$ be independent random variables where X_i is $\mathcal{X}_i$-valued for all i, and let $X = (X_1, \dots, X_n)$. Assume $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ is a measurable function that is c-Lipschitz. Then for all $b > 0$*

$$\mathbb{P}[f(X) - \mathbb{E}f(X) \geq \beta] \leq \exp\left(-\frac{2\beta^2}{c^2 n}\right).$$

Once again, applying the inequality to $-f$ gives a tail bound in the other direction.

Proof of Corollary 3.3. Note that by the Lipschitz condition the function f is bounded, as for any $x, x^* \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$,

$$|f(x)| \leq |f(x^*)| + nc =: B^*,$$

so that $\mathbb{E}|f(X)| \leq B^* < +\infty$. The idea of the proof is to consider the *Doob martingale*

Doob martingale

$$Z_i = \mathbb{E}[f(X) \,|\, \mathcal{F}_i],$$

where $\mathcal{F}_i = \sigma(X_1, \dots, X_i)$, and apply the Azuma-Hoeffding inequality. Indeed, note that $Z_n = \mathbb{E}[f(X) \,|\, \mathcal{F}_n] = f(X)$ and $Z_0 = \mathbb{E}[f(X)]$, and the result will follow once we show that the martingale (Z_i) has bounded increments.

Lemma 3.4 (Lipschitz condition $\implies$ bounded increments)**.** *Let $X_1, \dots, X_n$ be independent random variables where X_i is $\mathcal{X}_i$-valued for all i, and let $X = (X_1, \dots, X_n)$. Let $\mathcal{F}_i = \sigma(X_1, \dots, X_i)$ be the corresponding filtration. Assume $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ is a measurable function that is c-Lipschitz. Then the Doob martingale $Z_i = \mathbb{E}[f(X) \,|\, \mathcal{F}_i]$ has bounded increments with bound c.*

Proof. Let X_i' be an independent copy of X_i, and let

$$X' = (X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n).$$

Then

$$\begin{aligned}
|Z_i - Z_{i-1}| &= |\mathbb{E}[f(X) \,|\, \mathcal{F}_i] - \mathbb{E}[f(X) \,|\, \mathcal{F}_{i-1}]| \\
&= \left|\mathbb{E}[f(X) \,|\, \mathcal{F}_i] - \mathbb{E}[f(X') \,|\, \mathcal{F}_{i-1}]\right| \qquad (3.3) \\
&= \left|\mathbb{E}[f(X) \,|\, \mathcal{F}_i] - \mathbb{E}[f(X') \,|\, \mathcal{F}_i]\right| \qquad (3.4) \\
&= \left|\mathbb{E}[f(X) - f(X') \,|\, \mathcal{F}_i]\right| \\
&\leq \mathbb{E}\left[|f(X) - f(X')| \,|\, \mathcal{F}_i\right] \\
&\leq c,
\end{aligned}$$

where we applied the Lipschitz condition on the last line. Note that we crucially used the independence of the X_ks in (3.3) and (3.4). ■

We return to the proof of the corollary. Applying the Azuma-Hoeffding inequality—almost—gives the result. Note that we have proved a factor of $\frac{1}{2}$ instead of 2 in the exponent. Obtaining the better factor requires an additional observation. Note that, conditioned on $X_1, \ldots, X_{i-1}$ and $X_{i+1}, \ldots, X_n$, the random variable $f(X) - f(X')$ above lies almost surely in the interval

$$\Bigg[\inf_{y \in \mathcal{X}_i} f(X_1, \ldots, X_{i-1}, y, X_{i+1}, \ldots, X_n) - f(X'), \\ \sup_{y' \in \mathcal{X}_i} f(X_1, \ldots, X_{i-1}, y', X_{i+1}, \ldots, X_n) - f(X')\Bigg].$$

Averaging over $X_{i+1}, \ldots, X_n$, we get that $Z_i - Z_{i-1} = \mathbb{E}[f(X) - f(X') \,|\, \mathcal{F}_i]$ lies almost surely in the interval

$$\Bigg[\mathbb{E}\Big[\inf_{y \in \mathcal{X}_i} f(X_1, \ldots, X_{i-1}, y, X_{i+1}, \ldots, X_n) - f(X') \,\Big|\, \mathcal{F}_i\Big], \\ \mathbb{E}\Big[\sup_{y' \in \mathcal{X}_i} f(X_1, \ldots, X_{i-1}, y', X_{i+1}, \ldots, X_n) - f(X') \,\Big|\, \mathcal{F}_i\Big]\Bigg],$$

which, because it does not depend on X_i and the X_js are independent, is equal to

$$\Bigg[\mathbb{E}\Big[\inf_{y \in \mathcal{X}_i} f(X_1, \ldots, X_{i-1}, y, X_{i+1}, \ldots, X_n) - f(X') \,\Big|\, \mathcal{F}_{i-1}\Big], \\ \mathbb{E}\Big[\sup_{y' \in \mathcal{X}_i} f(X_1, \ldots, X_{i-1}, y', X_{i+1}, \ldots, X_n) - f(X') \,\Big|\, \mathcal{F}_{i-1}\Big]\Bigg].$$

Hence, conditioned on $\mathcal{F}_{i-1}$, the increment $Z_i - Z_{i-1}$ lies almost surely in an interval of length

$$\mathbb{E}\Bigg[\sup_{y \in \mathcal{X}_i} f(X_1, \ldots, X_{i-1}, y, X_{i+1}, \ldots, X_n) \\ - \inf_{y' \in \mathcal{X}_i} f(X_1, \ldots, X_{i-1}, y', X_{i+1}, \ldots, X_n) \,\Bigg|\, \mathcal{F}_{i-1}\Bigg] \leq c.$$

Using this fact directly in the application of Hoeffding's lemma in (3.2) proves the claim. ■

The following simple example shows that, without the independence assumption, the conclusion of Lemma 3.4 in general fails to hold.

Example 3.5 (A counterexample)**.** Let $f(x, y) = x + y$ where $x, y \in \{0, 1\}$. Clearly, f is 1-Lipschitz. Let X be a uniform random variable on $\{0, 1\}$ and let Y have conditional distribution

$$Y \,|\, X \sim \begin{cases} X, & \text{w.p. } 1-\varepsilon, \\ 1-X, & \text{w.p. } \varepsilon, \end{cases}$$

for some $\varepsilon > 0$ small. Then

$$\mathbb{E}[f(X,Y)\,|\,X] = X + \mathbb{E}[Y\,|\,X] = X + (1-\varepsilon)X + \varepsilon(1-X) = 2(1-\varepsilon)X + \varepsilon.$$

Hence, note that

$$|\mathbb{E}[f(X,Y)\,|\,X=1] - \mathbb{E}[f(X,Y)\,|\,X=0]| = 2(1-\varepsilon) > 1,$$

for ε small enough. In particular, the corresponding Doob martingale does not have bounded increments with bound 1 despite the fact that f itself is 1-Lipschitz. ◀

Examples The moral of McDiarmid's inequality is that functions of *independent* variables that are *smooth*, in the sense that they do not depend too much on any one of their variables, are concentrated around their mean. Here are some straightforward applications.

Example 3.6 (Balls and bins: empty bins)**.** Suppose we throw m balls into n bins independently, uniformly at random. The number of empty bins, $Z_{n,m}$, is centered at

$$\mathbb{E}Z_{n,m} = n\left(1 - \frac{1}{n}\right)^m.$$

Writing $Z_{n,m}$ as the sum of indicators $\sum_{i=1}^n \mathbb{1}_{B_i}$, where B_i is the event that bin i is empty, is a natural first attempt at proving concentration around the mean. However there is a problem—the B_is are *not independent*. Indeed because there is a fixed number of bins the event B_i intuitively makes the other such events less likely. Instead let X_j be the index of the bin in which ball j lands. The X_js are independent by construction and, moreover, $Z_{n,m} = f(X_1, \ldots, X_m)$ with f satisfying the Lipschitz condition with bound 1. Hence by the method of bounded differences

$$\mathbb{P}\left[\left|Z_{n,m} - n\left(1-\frac{1}{n}\right)^m\right| \geq b\sqrt{m}\right] \leq 2e^{-2b^2}.$$

◀

Example 3.7 (Pattern matching)**.** Let $X = (X_1, X_2, \ldots, X_n)$ be i.i.d. random variables taking values uniformly at random in a finite set S of size $s = |S|$. Let $a = (a_1, \ldots, a_k)$ be a fixed substring of elements of S. We are interested in the number of occurrences of a as a (consecutive) substring in X, which we denote by N_n. Denote by E_i the event that the substring of X starting at i is a. Summing over the starting positions and using the linearity of expectation, the mean of N_n is

$$\mathbb{E}N_n = \mathbb{E}\left[\sum_{i=1}^{n-k+1} \mathbb{1}_{E_i}\right] = (n-k+1)\left(\frac{1}{s}\right)^k.$$

However the $\mathbb{1}_{E_i}$s are not independent. So we cannot use a Chernoff bound for Poisson trials. Instead we use the fact that $N_n = f(X)$ where f satisfies the Lipschitz condition with bound k, as each X_i appears in at most k substrings of length k. By the method of bounded differences, for all $b > 0$,

$$\mathbb{P}\left[|N_n - \mathbb{E}N_n| \geq bk\sqrt{n}\right] \leq 2e^{-2b^2}.$$

◀

The last two examples are perhaps not surprising in that they involve *sums* of *"almost independent"* indicator variables. One might reasonably expect a sub-Gaussian type inequality in that case. The next application is much more striking.

3.2.3 ▷ *Erdös-Rényi graphs: exposure martingales, and application to the chromatic number*

Exposure martingales A common way to apply the Azuma-Hoeffding inequality in the context of Erdös-Rényi graphs is to consider a so-called *exposure martingale*. Let $G \sim \mathbb{G}_{n,p}$ and let F be function on graphs such that $\mathbb{E}_{n,p}|F(G)| < +\infty$ for all n, p. For $i = 1, \ldots, n$, denote by H_i the subgraph of G induced by the first i vertices, where the vertices are ordered arbitrarily. Then the filtration $\mathcal{H}_i = \sigma(H_1, \ldots, H_i)$, $i = 1, \ldots, n$, corresponds to exposing the vertices of G one at a time. The Doob martingale

$$Z_i = \mathbb{E}_{n,p}[F(G) \,|\, \mathcal{H}_i], \qquad i = 1, \ldots, n,$$

is known as a *vertex exposure martingale*. Edge exposure can be defined similarly by exposing the edges one at a time in some arbitrary order.

exposure martingale

As an example, consider the chromatic number $\chi(G)$, i.e., the smallest number of colors needed in a proper coloring of G. We claim that the corresponding vertex exposure martingale (Z_i) has bounded increments with bound 1.

Lemma 3.8. *Changing the edges adjacent to a single vertex can change the chromatic number by at most* 1.

Proof. Changing the edges adjacent to v can increase the chromatic number by at most 1 as one can always use an extra color for v. On the other hand, if the chromatic number were to decrease by more than 1, reversing the change and using the previous observation would give a contradiction. ∎

Write $\chi(G) = f(X_2, \ldots, X_n)$ for some function f where $X_i = (\mathbb{1}_{\{\{i,j\} \in G\}} : j \leq i)$. Because X_i depends only on (a subset of) the edges adjacent to node i, the previous lemma implies that f is 1-Lipschitz. Furthermore, since the X_is are independent as they involve disjoint subsets of edges (which is the reason behind the definition of vertex exposure), Lemma 3.4 implies that (Z_i) has bounded increments with bound 1. Hence, for all $0 < p < 1$ and n, by an immediate application of the Azuma-Hoeffding inequality:

Claim 3.9.

$$\mathbb{P}_{n,p}\left[|\chi(G) - \mathbb{E}_{n,p}[\chi(G)]| \geq b\sqrt{n-1}\right] \leq 2e^{-b^2/2}.$$

In fact applying the method of bounded differences directly to f gives a slightly better bound—but the constants rarely matter. Observe also that *edge exposure* results in a much weaker bound as the $\Theta(n^2)$ steps of the corresponding martingale produce only a *linear in* n deviation for the same tail probability. (The reader may want to ponder the apparent paradox: using a larger number of independent variables seemingly leads to weaker concentration here.)

Remark 3.10. *Note that Claim 3.9 tells us nothing about the expectation of* $\chi(G)$*. It turns out that, up to logarithmic factors,* $\mathbb{E}_{n,p_n}[\chi(G)]$ *is of order* np_n *when* $p_n \sim n^{-\alpha}$ *for some* $0 < \alpha < 1$*. We will not prove this result here. See the "Bibliographic remarks" at the end of this chapter for more on the chromatic number of Erdös-Rényi graphs.*

$\chi(G)$ **is concentrated on few values** Much stronger concentration results can be obtained: when $p_n = n^{-\alpha}$ with $\alpha > \frac{1}{2}$, $\chi(G)$ is in fact concentrated on two values! We give a partial result along those lines which illustrates a less straightforward choice of martingale in the Azuma-Hoeffding inequality.

Claim 3.11. *Let* $p_n = n^{-\alpha}$ *with* $\alpha > \frac{5}{6}$ *and let* $G_n \sim \mathbb{G}_{n,p_n}$*. Then for any* $\varepsilon > 0$ *there is* $\varphi_n := \varphi_n(\alpha, \varepsilon)$ *such that*

$$\mathbb{P}_{n,p_n}\left[\,\varphi_n \leq \chi(G_n) \leq \varphi_n + 3\,\right] \geq 1 - \varepsilon,$$

for all n *large enough.*

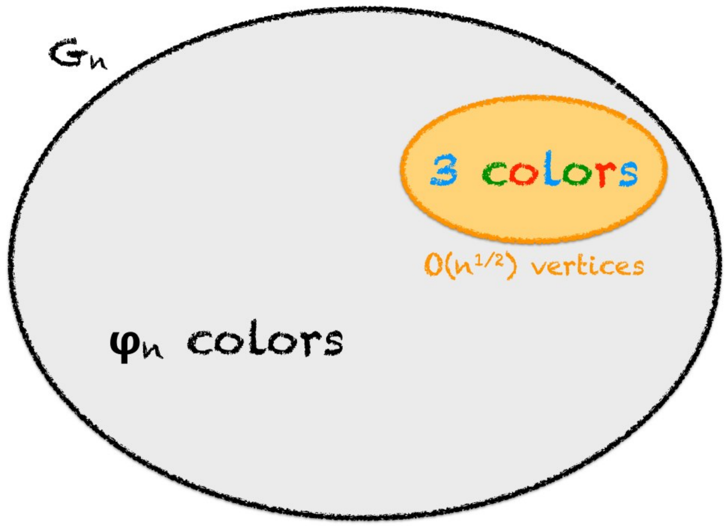

Figure 3.1: All but $O(\sqrt{n})$ vertices are colored using φ_n colors. The remaining vertices are colored using 3 additional colors.

Proof. We consider the following martingale. Let φ_n be the smallest integer such that

$$\mathbb{P}_{n,p_n}[\chi(G_n) \leq \varphi_n] > \frac{\varepsilon}{3}. \tag{3.5}$$

Let $F_n(G_n)$ be the minimal size of a set of vertices, U, in G_n such that $G_n \backslash U$ is φ_n-colorable. Let (Z_i) be the corresponding vertex exposure martingale. The idea of the proof is to show that all but $O(\sqrt{n})$ vertices can be φ_n-colored, and the remaining vertices can be colored using 3 additional colors. See Figure 3.2.3.

We claim that (Z_i) has bounded increments with bound 1.

Lemma 3.12. *Changing the edges adjacent to a single vertex can change F_n by at most* 1.

Proof. Changing the edges adjacent to v can increase F_n by at most 1. Indeed, if F_n increases, it must be that $v \notin U$ and we can add v to U. On the other hand, if F_n were to decrease by more than 1, reversing the change and using the previous observation would give a contradiction. ■

Lemma 3.4 implies that (Z_i) has bounded increments with bound 1.

Choose b_ε such that $e^{-b_\varepsilon^2/2} = \frac{\varepsilon}{3}$. Then, applying the Azuma-Hoeffding inequality to $(-Z_i)$,

$$\mathbb{P}_{n,p_n}\left[F_n(G_n) - \mathbb{E}_{n,p_n}[F_n(G_n)] \leq -b_\varepsilon\sqrt{n-1}\right] \leq \frac{\varepsilon}{3}$$

which, since $\mathbb{P}_{n,p_n}[F_n(G_n) = 0] = \mathbb{P}_{n,p_n}[\chi(G_n) \leq \varphi_n] > \frac{\varepsilon}{3}$, implies that

$$\mathbb{E}_{n,p_n}[F_n(G_n)] \leq b_\varepsilon\sqrt{n-1}.$$

Applying the Azuma-Hoeffding inequality to (Z_i) gives

$$\begin{aligned} &\mathbb{P}_{n,p_n}\left[F_n(G_n) \geq 2b_\varepsilon\sqrt{n-1}\right] &(3.6)\\ &\quad\leq \mathbb{P}_{n,p_n}\left[F_n(G_n) - \mathbb{E}_{n,p_n}[F_n(G_n)] \geq b_\varepsilon\sqrt{n-1}\right]\\ &\quad\leq \frac{\varepsilon}{3}. &(3.7)\end{aligned}$$

So with probability at least $1 - \frac{\varepsilon}{3}$, we can color all vertices but $2b_\varepsilon\sqrt{n-1}$ using φ_n colors. Let U be the remaining uncolored vertices.

We claim that, with high probability, we can color the vertices in U using at most 3 extra colors.

Lemma 3.13. *Fix $c > 0$, $\alpha > \frac{5}{6}$ and $\varepsilon > 0$. Let $G_n \sim \mathbb{G}_{n,p_n}$ with $p_n = n^{-\alpha}$. For all n large enough,*

$$\mathbb{P}_{n,p_n}\left[\textit{every subset of } c\sqrt{n} \textit{ vertices of } G_n \textit{ can be 3-colored}\right] \geq 1 - \frac{\varepsilon}{3}. \quad (3.8)$$

Proof. We use the first moment method, Corollary 2.5. To bound the probability that a subset of vertices is not 3-colorable, we consider a minimal such subset and notice that all of its vertices must have degree at least 3. Indeed, suppose W is not 3-colorable but that all of its subsets are (we call such a subset *minimal, non 3-colorable*), and suppose that $w \in W$ has degree less than 3. Then $W\backslash\{w\}$ is 3-colorable. But, since w has fewer than 3 neighbors, it can also be properly colored without adding a new color—a contradiction. In particular, the subgraph of G_n induced by W must have at least $\frac{3}{2}|W|$ edges.

minimal, non 3-colorable subset

Let Y_n be the number of minimal, non 3-colorable subsets of vertices of G_n of size at most $c\sqrt{n}$. By the argument above, the probability that a subset of vertices of G_n of size ℓ is minimal, non 3-colorable is at most $\binom{\binom{\ell}{2}}{\frac{3\ell}{2}} p_n^{\frac{3\ell}{2}}$ by a union bound over subsets of edges of size $\frac{3\ell}{2}$. Then, by the first moment method,

$$\begin{aligned}
\mathbb{P}_{n,p_n}[Y_n > 0] &\leq \mathbb{E}_{n,p_n} Y_n \\
&\leq \sum_{\ell=4}^{c\sqrt{n}} \binom{n}{\ell} \binom{\binom{\ell}{2}}{\frac{3\ell}{2}} p_n^{\frac{3\ell}{2}} \\
&\leq \sum_{\ell=4}^{c\sqrt{n}} \left(\frac{en}{\ell}\right)^{\ell} \left(\frac{e\ell}{3}\right)^{\frac{3\ell}{2}} n^{-\frac{3\ell\alpha}{2}} \\
&\leq \sum_{\ell=4}^{c\sqrt{n}} \left(\frac{e^{\frac{5}{2}} n^{1-\frac{3\alpha}{2}} \ell^{\frac{1}{2}}}{3^{\frac{3}{2}}}\right)^{\ell} \\
&\leq \sum_{\ell=4}^{c\sqrt{n}} \left(c' n^{\frac{5}{4}-\frac{3\alpha}{2}}\right)^{\ell} \\
&\leq O\left(n^{\frac{5}{4}-\frac{3\alpha}{2}}\right)^{4} \\
&\to 0,
\end{aligned}$$

as $n \to +\infty$, for some $c' > 0$, where we used that $\frac{5}{4} - \frac{3\alpha}{2} < \frac{5}{4} - \frac{5}{4} = 0$ when $\alpha > \frac{5}{6}$. ■

By the choice of φ_n in (3.5),

$$\mathbb{P}_{n,p_n}[\chi(G_n) < \varphi_n] \leq \frac{\varepsilon}{3}.$$

By (3.7) and (3.8),

$$\mathbb{P}_{n,p_n}[\chi(G_n) > \varphi_n + 3] \leq \frac{2\varepsilon}{3}.$$

So, overall,

$$\mathbb{P}_{n,p_n}[\varphi_n \leq \chi(G_n) \leq \varphi_n + 3] \geq 1 - \varepsilon.$$

■

3.2.4 ▷ *Hypercube: concentration of measure*

For $A \subseteq \{0,1\}^n$ a subset of the hypercube and $r > 0$, we let

$$A_r = \left\{ x \in \{0,1\}^n \,:\, \inf_{a \in A} \|x - a\|_1 \leq r \right\},$$

be the points at ℓ^1 distance r from A.

Fix $\varepsilon \in (0, 1/2)$ and assume that $|A| \geq \varepsilon 2^n$. Let λ_ε be such that $e^{-2\lambda_\varepsilon^2} = \varepsilon$. The following application of the method of bounded differences indicates that much of the uniform measure on the high-dimensional hypercube lies in a close neighborhood of A, an example of the *concentration of measure phenomenon*.

Claim 3.14.

$$r > 2\lambda_\varepsilon\sqrt{n} \implies |A_r| \geq (1-\varepsilon)2^n.$$

Proof. Let $X = (X_1, \ldots, X_n)$ be uniformly distributed in $\{0,1\}^n$. Note that the coordinates are in fact independent. The function $f(x) = \inf_{a \in A} \|x - a\|_1$ satisfies the Lipschitz condition with bound 1. Indeed changing one coordinate of x can only increase the ℓ^1 distance to the closest point to x by 1. Hence the method of bounded differences gives

$$\mathbb{P}\left[\mathbb{E}f(X) - f(X) \geq \beta\right] \leq \exp\left(-\frac{2\beta^2}{n}\right).$$

Choosing $\beta = \mathbb{E}f(X)$ and noting that $f(x) \leq 0$ if and only if $x \in A$ gives

$$\mathbb{P}[A] \leq \exp\left(-\frac{2(\mathbb{E}f(X))^2}{n}\right),$$

or, rearranging and using our assumption on A,

$$\mathbb{E}f(X) \leq \sqrt{\frac{1}{2} n \log \frac{1}{\mathbb{P}[A]}} \leq \sqrt{\frac{1}{2} n \log \frac{1}{\varepsilon}} = \lambda_\varepsilon \sqrt{n}.$$

By a second application of the method of bounded differences with $\beta = \lambda_\varepsilon\sqrt{n}$,

$$\mathbb{P}\left[f(X) \geq 2\lambda_\varepsilon\sqrt{n}\right] \leq \mathbb{P}\left[f(X) - \mathbb{E}f(X) \geq b\right] \leq \exp\left(-\frac{2\beta^2}{n}\right) = \varepsilon.$$

The result follows by observing that, with $r > 2\lambda_\varepsilon\sqrt{n}$,

$$\frac{|A_r|}{2^n} \geq \mathbb{P}\left[f(X) < 2\lambda_\varepsilon\sqrt{n}\right] \geq 1 - \varepsilon.$$

∎

Claim 3.14 is striking for two reasons: 1) the radius $2\lambda_\varepsilon\sqrt{n}$ is much smaller than n, the diameter of $\{0,1\}^n$; and 2) it applies to *any* A. The smallest r such that $|A_r| \geq (1-\varepsilon)2^n$ in general depends on A. Here are two extremes:

- For $\gamma > 0$, let

$$B(\gamma) := \left\{ x \in \{0,1\}^n \,:\, \|x\|_1 \leq \frac{n}{2} - \gamma\sqrt{\frac{n}{4}} \right\}.$$

Note that

$$\begin{aligned} \frac{1}{2^n}|B(\gamma)| &= \sum_{\ell=0}^{\frac{n}{2}-\gamma\sqrt{\frac{n}{4}}} \binom{n}{\ell} 2^{-n} \\ &= \mathbb{P}\left[Y_n \leq \frac{n}{2} - \gamma\sqrt{\frac{n}{4}}\right] \\ &= \mathbb{P}\left[\frac{Y_n - n/2}{\sqrt{n/4}} \leq -\gamma\right], \end{aligned} \tag{3.9}$$

where $Y_n \sim B(n, \frac{1}{2})$. By the Berry-Esséen theorem (e.g., [Dur10, Theorem 3.4.9]), there is a $C > 0$ such that

$$\left| \mathbb{P}\left[\frac{Y_n - n/2}{\sqrt{n/4}} \leq -\gamma\right] - \mathbb{P}[Z \leq -\gamma] \right| \leq \frac{C}{\sqrt{n}},$$

where $Z \sim N(0,1)$. Let $\varepsilon < \varepsilon' < 1/2$ and let $\gamma_{\varepsilon'}$ be such that $\mathbb{P}[Z \leq -\gamma_{\varepsilon'}] = \varepsilon'$. Then setting $A := B(\gamma_{\varepsilon'})$, for n large enough,

$$|A| \geq \varepsilon 2^n,$$

by (3.9). On the other hand, setting $r := \gamma_{\varepsilon'}\sqrt{n/4}$, we have

$$A_r \subseteq B(0),$$

so that $|A_r| \leq \frac{1}{2}2^n < (1-\varepsilon)2^n$. We have shown that $r = \Omega(\sqrt{n})$ is in general required for Claim 3.14 to hold.

- Assume for simplicity that $N := \varepsilon 2^n$ is an integer. Let $A \subseteq \{0,1\}^n$ be constructed as follows: starting from the empty set, add points in $\{0,1\}^n$ to A independently, uniformly at random until $|A| = N$. Set $r := 2$. By the first moment method, Corollary 2.5, the probability that $A_r \subset \{0,1\}^n$ is at most

$$\begin{aligned}
\mathbb{P}[|\{0,1\}^n \backslash A_r| > 0] &\leq \mathbb{E}|\{0,1\}^n \backslash A_r| \\
&= \sum_{x \in \{0,1\}^n} \mathbb{P}[x \notin A_r] \\
&\leq 2^n \left(1 - \frac{\binom{n}{2}}{2^n}\right)^{\varepsilon 2^n} \\
&\leq 2^n e^{-\varepsilon \binom{n}{2}} \\
&\to 0,
\end{aligned}$$

where, on the third line, we considered only the first N picks in the construction of A. In particular, as $n \to +\infty$, $\mathbb{P}[|\{0,1\}^n \backslash A_r| > 0] < 1$. So for n large enough there is a set A such that $A_r = \{0,1\}^n$ where $r = 2$.

Remark 3.15. *In fact, it can be shown that sets of the form $\{x : \|x\|_1 \leq s\}$ have the smallest "expansion" among subsets of $\{0,1\}^n$ of the same size, a result known as Harper's vertex isoperimetric theorem. See, e.g., [BLM13, Theorem 7.6 and Exercises 7.11-7.13].*

3.2.5 ▷ *Preferential attachment graphs: degree sequence*

Let $(G_t)_{t \geq 1} \sim \mathrm{PA}_m$ be a preferential attachment graph process with parameter $m \geq 1$. A key feature of preferential attachment graphs is a power-law degree sequence: the fraction of vertices with degree d behaves like $\propto d^{-\alpha}$ for some $\alpha > 0$, i.e., it has a fat tail. We prove this in the case of scale-free trees, $m = 1$. In contrast, we will show in Section 4.1.2 that (sparse) Erdös-Rényi graphs have an asymptotically Poisson-distributed degree sequence.

Power law degree sequence Let $D_i(t)$ be the degree of the i-th vertex, v_i, in G_t, and denote by

$$N_d(t) := \sum_{i=0}^{t} \mathbb{1}_{\{D_i(t)=d\}},$$

the number of vertices of degree d in G_t. Define

$$f_d := \frac{4}{d(d+1)(d+2)}, \qquad d \geq 1. \tag{3.10}$$

Claim 3.16.

$$\frac{1}{t} N_d(t) \to_{\mathrm{p}} f_d, \qquad \forall d \geq 1.$$

Proof. Claim 3.16 follows from the following lemmas. Fix $\delta > 0$.

Lemma 3.17 (Convergence of the mean)**.**

$$\frac{1}{t}\mathbb{E}N_d(t) \to f_d, \qquad \forall d \geq 1.$$

Lemma 3.18 (Concentration around the mean)**.**

$$\mathbb{P}\left[\left|\frac{1}{t}N_d(t) - \frac{1}{t}\mathbb{E}N_d(t)\right| \geq \sqrt{\frac{2\log\delta^{-1}}{t}}\right] \leq 2\delta, \qquad \forall d \geq 1, \forall t.$$

An alternative representation of the process We start with the proof of Lemma 3.18, which follows from an application of the method of bounded differences.

Proof of Lemma 3.18. In our description of the preferential attachment process, the random choices made at each time depend in a seemingly complicated way on previous choices. In order to establish concentration of the process around its mean, we introduce a clever, alternative construction of the $m = 1$ case which has the advantage that it involves *independent* choices.

We start with a single vertex v_0. At time 1, we add a single vertex v_1 and an edge e_1 connecting v_0 and v_1. For bookkeeping we orient edges away from the vertex of lower time index. For all $s \geq 2$, let X_s be an independent, uniformly chosen edge extremity among the edges in G_{s-1}, i.e., pick a uniform element in

$$\mathcal{X}_s := \{(1, \text{tail}), (1, \text{head}), \ldots, (s-1, \text{tail}), (s-1, \text{head})\}.$$

To form G_s, attach a new edge e_s to the vertex of G_{s-1} corresponding to X_s. A vertex of degree d' in G_{s-1} is selected with probability $\frac{d'}{2(s-1)}$, as it should. Note that X_s can be picked in advance independently of the sequence $(G_{s'})_{s'<s}$.

For instance, if $x_2 = (1, \text{head})$, $x_3 = (2, \text{tail})$ and $x_4 = (3, \text{head})$, the graph obtained at time 4 is depicted in Figure 3.2.

We claim that $N_d(t) =: h(X_2, \ldots, X_t)$ seen as a function of $X_2, \ldots, X_t$ is 2-Lipschitz. Indeed let $(x_2, \ldots, x_t)$ be a realization of $(X_2, \ldots, X_t)$ and let $y \in \mathcal{X}_s$ with $y \neq x_s$. Replacing $x_s = (i, \text{end})$ with $y = (j, \text{end}')$ where $i, j \in \{1, \ldots, s-1\}$ and $\text{end}, \text{end}' \in \{\text{tail}, \text{head}\}$ has the effect of redirecting the head of edge e_s from the end of e_i to the end$'$ of e_j. This redirection also brings along with it the heads of all other edges associated with the choice (s, head). But, crucially, those changes only affect the degrees of the vertices corresponding to (i, end) and (j, end') in the original graph. Hence the number of vertices with degree d changes by at most 2.

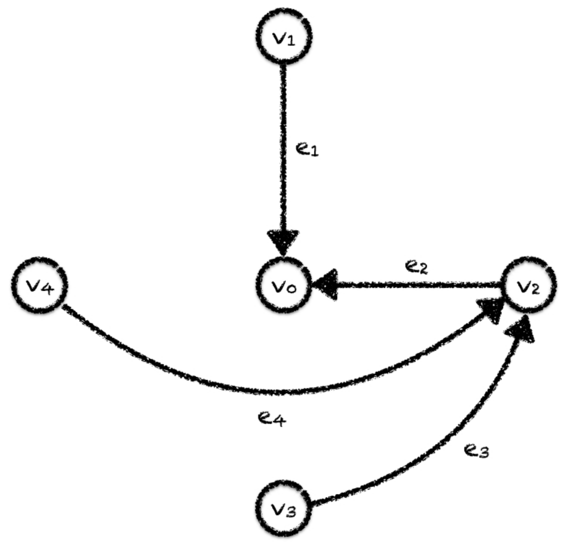

Figure 3.2: Graph obtained when $x_2 = (1, \text{head})$, $x_3 = (2, \text{tail})$ and $x_4 = (3, \text{head})$.

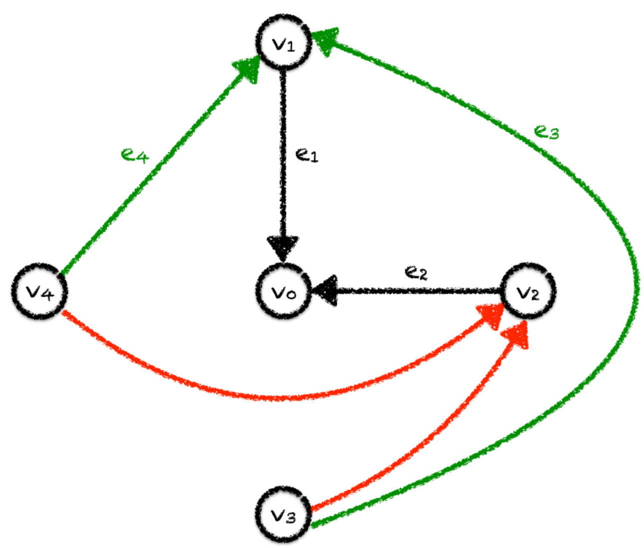

Figure 3.3: Substituting $x_3 = (2, \text{tail})$ with $y = (1, \text{tail})$ in the example of Figure 3.2 has the effect of replacing the red edges with the green edges. Note that only the degrees of vertices v_1 and v_2 are affected by this change.

For instance, returning to the example of Figure 3.2. If we replace $x_3 = (2, \mathrm{tail})$ with $y = (1, \mathrm{tail})$, one obtains the graph in Figure 3.3. Note that only the degrees of vertices v_1 and v_2 are affected by this change.

By the method of bounded differences, for all $\beta > 0$,

$$\mathbb{P}[|N_d(t) - \mathbb{E}N_d(t)| \geq \beta] \leq 2\exp\left(-\frac{2\beta^2}{(2)^2(t-1)}\right),$$

which, choosing $\beta = \sqrt{2t\log\delta^{-1}}$, we can re-write as

$$\mathbb{P}\left[\left|\frac{1}{t}N_d(t) - \frac{1}{t}\mathbb{E}N_d(t)\right| \geq \sqrt{\frac{2\log\delta^{-1}}{t}}\right] \leq 2\delta.$$

■

Dynamics of the mean Once again the method of bounded differences tells us nothing about the mean, which must be analyzed by other means. The proof of Lemma 3.17 does not rely on the Azuma-Hoeffding inequality but is given for completeness (and may be skipped).

Proof of Lemma 3.17. The idea of the proof is to derive a recursion for f_d by considering the evolution of $\mathbb{E}N_d(t)$ and taking a limit as $t \to +\infty$. By the description of the preferential attachment process, the following recursion holds for $t \geq d$

$$\mathbb{E}N_d(t+1) - \mathbb{E}N_d(t) = \underbrace{\frac{d-1}{2t}\mathbb{E}N_{d-1}(t)}_{\text{(a)}} - \underbrace{\frac{d}{2t}\mathbb{E}N_d(t)}_{\text{(b)}} + \underbrace{\mathbb{1}_{\{d=1\}}}_{\text{(c)}}, \tag{3.11}$$

and $\mathbb{E}N_d(d-1) = 0$. Indeed: (a) for $d \geq 2$, $N_d(t)$ increases by 1 if a vertex of degree $d-1$ is picked, an event of probability $\frac{d-1}{2t}N_{d-1}(t)$ because the sum of degrees at time t is twice the number of edges, i.e., $2t$; (b) for $d \geq 1$, $N_d(t)$ decreases by 1 if a vertex of degree d is picked, an event of probability $\frac{d}{2t}N_d(t)$; and (c) the last term comes from the fact that the new vertex always has degree 1. We re-write (3.11) as

$$\begin{aligned}
\mathbb{E}N_d(t+1) &= \mathbb{E}N_d(t) + \frac{d-1}{2t}\mathbb{E}N_{d-1}(t) - \frac{d}{2t}\mathbb{E}N_d(t) + \mathbb{1}_{\{d=1\}} \\
&= \left(1 - \frac{d/2}{t}\right)\mathbb{E}N_d(t) + \left\{\frac{d-1}{2}\left[\frac{1}{t}\mathbb{E}N_{d-1}(t)\right] + \mathbb{1}_{\{d=1\}}\right\} \\
&=: \left(1 - \frac{d/2}{t}\right)\mathbb{E}N_d(t) + g_d(t), \qquad (3.12)
\end{aligned}$$

where $g_d(t)$ is defined as the expression in curly brackets on the second line. We show by induction on d that $\frac{1}{t}\mathbb{E}N_d(t) \to f_d$. Because of the form of the recursion, the following lemma is what we need to proceed.

Lemma 3.19. *Let f be a function of $t \in \mathbb{N}$ satisfying the following recursion*

$$f(t+1) = \left(1 - \frac{\alpha}{t}\right) f(t) + g(t), \qquad \forall t \geq t_0$$

with $g(t) \to g \in (-\infty, +\infty)$ as $t \to +\infty$, and where $\alpha > 0, t_0 \geq 2\alpha, f(t_0) \geq 0$ are constants. Then

$$\frac{1}{t} f(t) \to \frac{g}{1+\alpha},$$

as $t \to +\infty$.

The proof of this lemma is given after the proof of Claim 3.16. We first conclude the proof of Lemma 3.17. First let $d = 1$. In that case, $g_1(t) = g_1 := 1$, $\alpha := 1/2$, and $t_0 := 1$. By Lemma 3.19,

$$\frac{1}{t}\mathbb{E}N_1(t) \to \frac{1}{1+1/2} = \frac{2}{3} = f_1.$$

Assuming by induction that $\frac{1}{t}\mathbb{E}N_{d'}(t) \to f_{d'}$ for all $d' < d$ we get

$$g_d(t) \to g_d := \frac{d-1}{2} f_{d-1},$$

as $t \to +\infty$. Using Lemma 3.19 with $\alpha := d/2$ and $t_0 := d$, we obtain

$$\frac{1}{t}\mathbb{E}N_d(t) \to \frac{1}{1+d/2}\left[\frac{d-1}{2} f_{d-1}\right] = \frac{d-1}{d+2} \cdot \frac{4}{(d-1)d(d+1)} = f_d,$$

where we used (3.10). That concludes the proof of Lemma 3.17. ■

To prove Claim 3.16, we combine Lemmas 3.17 and 3.18. Fix any $\delta, \varepsilon > 0$. Choose t' large enough that for all $t \geq t'$

$$\max\left\{\left|\frac{1}{t}\mathbb{E}N_d(t) - f_d\right|, \sqrt{\frac{2\log\delta^{-1}}{t}}\right\} \leq \varepsilon.$$

Then

$$\mathbb{P}\left[\left|\frac{1}{t}N_d(t) - f_d\right| \geq 2\varepsilon\right] \leq 2\delta,$$

for all $t \geq t'$. That proves convergence in probability. ■

Proof of the technical lemma It remains to prove Lemma 3.19.

Proof of Lemma 3.19. By induction on t, we have

$$\begin{aligned}
f(t+1) &= \left(1-\frac{\alpha}{t}\right) f(t) + g(t) \\
&= \left(1-\frac{\alpha}{t}\right)\left[\left(1-\frac{\alpha}{t-1}\right) f(t-1) + g(t-1)\right] + g(t) \\
&= \left(1-\frac{\alpha}{t}\right) g(t-1) + g(t) + \left(1-\frac{\alpha}{t}\right)\left(1-\frac{\alpha}{t-1}\right) f(t-1) \\
&= \cdots \\
&= \sum_{i=1}^{t-t_0} g(t-i) \prod_{j=0}^{i-1}\left(1-\frac{\alpha}{t-j}\right) + f(t_0)\prod_{j=0}^{t-t_0}\left(1-\frac{\alpha}{t-j}\right),
\end{aligned}$$

or

$$f(t+1) = \sum_{s=t_0}^{t} g(s) \prod_{r=s+1}^{t}\left(1-\frac{\alpha}{r}\right) + f(t_0)\prod_{r=t_0}^{t}\left(1-\frac{\alpha}{r}\right). \tag{3.13}$$

To guess the answer note that, for large s, $g(s)$ is roughly constant and that the product in the first term behaves like

$$\exp\left(-\sum_{r=s+1}^{t}\frac{\alpha}{r}\right) \approx \exp\left(-\alpha(\log t - \log s)\right) \approx \frac{s^\alpha}{t^\alpha}.$$

So approximating the sum by an integral we get that $f(t+1) \approx \frac{gt}{\alpha+1}$.

Formally, we use that there is a constant $\gamma = 0.577\ldots$ such that (see e.g. [LL10, Lemma 12.1.3])

$$\sum_{\ell=1}^{m}\frac{1}{\ell} = \log m + \gamma + \Theta(m^{-1}),$$

and that by a Taylor expansion, for $|z| \le 1/2$,

$$\log(1-z) = -z + \Theta(z^2).$$

Fix $\eta > 0$ small and take t large enough that $\eta t > 2\alpha$ and $|g(s) - g| < \eta$ for all $s \ge \eta t$. Then, for $s+1 \ge t_0$,

$$\begin{aligned}
\sum_{r=s+1}^{t}\log\left(1-\frac{\alpha}{r}\right) &= -\sum_{r=s+1}^{t}\left\{\frac{\alpha}{r} + \Theta(r^{-2})\right\} \\
&= -\alpha\left(\log t - \log s\right) + \Theta(s^{-1}),
\end{aligned}$$

so, taking exponentials,

$$\prod_{r=s+1}^{t}\left(1-\frac{\alpha}{r}\right)=\frac{s^{\alpha}}{t^{\alpha}}(1+\Theta(s^{-1})).$$

Hence

$$\frac{1}{t}f(t_0)\prod_{r=t_0}^{t}\left(1-\frac{\alpha}{r}\right)=\frac{t_0^{\alpha}}{t^{\alpha+1}}(1+\Theta(t_0^{-1}))\to 0.$$

Moreover

$$\begin{aligned}\frac{1}{t}\sum_{s=\eta t}^{t}g(s)\prod_{r=s+1}^{t}\left(1-\frac{\alpha}{r}\right) &\leq \frac{1}{t}\sum_{s=\eta t}^{t}(g+\eta)\frac{s^{\alpha}}{t^{\alpha}}(1+\Theta(s^{-1}))\\ &\leq O(\eta)+(1+\Theta(t^{-1}))\frac{g}{t^{\alpha+1}}\sum_{s=\eta t}^{t}s^{\alpha}\\ &\leq O(\eta)+(1+\Theta(t^{-1}))\frac{g}{t^{\alpha+1}}\frac{(t+1)^{\alpha+1}}{\alpha+1}\\ &\to O(\eta)+\frac{g}{\alpha+1},\end{aligned}$$

and, similarly,

$$\begin{aligned}\frac{1}{t}\sum_{s=t_0}^{\eta t-1}g(s)\prod_{r=s+1}^{t}\left(1-\frac{\alpha}{r}\right) &\leq \frac{1}{t}\sum_{s=t_0}^{\eta t-1}(g+\eta)\frac{s^{\alpha}}{t^{\alpha}}(1+\Theta(s^{-1}))\\ &\leq \frac{\eta t}{t}(g+\delta)\frac{(\eta t)^{\alpha}}{t^{\alpha}}(1+\Theta(t_0^{-1}))\\ &\to O(\eta^{\alpha+1}).\end{aligned}$$

Plugging these inequalities back into (3.13), we get

$$\limsup_{t}\frac{1}{t}f(t+1)\leq\frac{g}{1+\alpha}+O(\eta).$$

A similar inequality holds in the other direction. Taking $\eta\to 0$ concludes the proof. ■

Remark 3.20. *A more quantitative result (uniform in t and d) can be derived. See, e.g., [vdH14, Sections 8.5, 8.6]. See the same reference for the case $m>1$.*

3.3 Electrical networks

In this section we develop a classical link between random walks and electrical networks. The electrical interpretation is merely a useful physical analogy. The mathematical substance of the connection starts with the following well-known observation.

Let (X_t) be a Markov chain with transition matrix P on a finite or countable state space V. For two disjoint subsets A, Z of V, the probability of hitting A before Z

$$h(x) = \mathbb{P}_x[\tau_A < \tau_Z], \tag{3.14}$$

seen as a function of the starting point $x \in V$, is harmonic on $W := (A \cup Z)^c$ in the sense that

$$h(x) = \sum_y P(x,y)h(y), \qquad \forall x \in W, \tag{3.15}$$

where $h \equiv 1$ (respectively $\equiv 0$) on A (respectively Z). Indeed by the Markov property, after one step of the chain, for $x \in W$

$$\begin{aligned} \mathbb{P}_x[\tau_A < \tau_Z] &= \sum_{y \notin A \cup Z} P(x,y)\, \mathbb{P}_y[\tau_A < \tau_Z] \\ &\qquad + \sum_{y \in A} P(x,y) \cdot 1 + \sum_{y \in Z} P(x,y) \cdot 0 \\ &= \sum_y P(x,y)\, \mathbb{P}_y[\tau_A < \tau_Z]. \end{aligned} \tag{3.16}$$

Quantities such as (3.14) arise naturally, for instance in the study of recurrence, and the connection to potential theory, the study of harmonic functions, proves fruitful in that context—and beyond—as we outline in this section.

First we re-write (3.15) to reveal the electrical interpretation. For this we switch to reversible chains. Recall that a reversible Markov chain is equivalent to a random walk on a network $\mathcal{N} = (G, c)$ where the edges of G correspond to transitions of positive probability. If the chain is reversible with respect to a stationary measure π, then the edge weights are $c(x,y) = \pi(x)P(x,y)$. In this notation (3.15) becomes

$$h(x) = \frac{1}{c(x)} \sum_{y \sim x} c(x,y)h(y), \qquad \forall x \in (A \cup Z)^c, \tag{3.17}$$

where $c(x) := \sum_{y \sim x} c(x,y) = \pi(x)$. In words, $h(x)$ is the weighted average of its neighboring values. Now comes the electrical analogy: if one interprets $c(x,y)$ as a conductance, a function satisfying (3.17) is known as a voltage or potential function. The voltages at A and Z are 1 and 0 respectively. We show in the next subsection by a martingale argument that, under appropriate conditions, such a voltage exists and is unique. To see why martingales come in, let $\mathcal{F}_t = \sigma(X_0, \ldots, X_t)$ and $\tau^* := \tau_{A \cup Z}$. Notice that, by a one-step calculation again, (3.15) implies that

$$h(X_{t\wedge\tau^*}) = \mathbb{E}\left[h(X_{(t+1)\wedge\tau^*}) \,|\, \mathcal{F}_t\right], \qquad \forall t \geq 0,$$

i.e., $(h(X_{t\wedge\tau^*}))_t$ is a martingale with respect to $(\mathcal{F}_t)$.

3.3.1 Martingales and the Dirichlet problem

Although the rest of Section 3.3 is concerned with reversible Markov chains, the current subsection applies to the non-reversible case as well. The following definition will be useful below. Let σ be a stopping time for a Markov chain (X_t). The *Green function* of the chain stopped at σ is given by

Green function

$$\mathscr{G}_\sigma(x, y) = \mathbb{E}_x\left[\sum_{0\leq t<\sigma} \mathbb{1}_{\{X_t = y\}}\right], \qquad x, y \in V$$

i.e., it is the expected number of visits to y *before* σ when started at x. We will use the notation $h|_Z$ for the function h restricted to the subset Z.

Existence and uniqueness of a harmonic extension We begin with a general problem.

Theorem 3.21 (Existence and uniqueness)**.** *Let P be an irreducible transition matrix on a finite or countable spate space V. Let W be a finite, proper subset of V and let $h : W^c \to \mathbb{R}$ be a bounded function on $W^c = V\backslash W$. Then there exists a unique extension of h to W that is* harmonic *on W, i.e., it satisfies*

harmonic function

$$h(x) = \sum_y P(x, y)h(y), \qquad \forall x \in W. \tag{3.18}$$

The solution is given by $h(x) = \mathbb{E}_x[h(X_{\tau_{W^c}})]$.

Proof. We first argue about uniqueness. Suppose h is defined over all of V and satisfies (3.18). Let $\tau^* := \tau_{W^c}$. Then the process $(h(X_{t\wedge\tau^*}))_t$ is a martingale: on $\{\tau^* \leq t\}$,

$$\mathbb{E}[h(X_{(t+1)\wedge\tau^*}) \,|\, \mathcal{F}_t] = h(X_{\tau^*}) = h(X_{t\wedge\tau^*}),$$

and on $\{\tau^* > t\}$

$$\mathbb{E}[h(X_{(t+1)\wedge\tau^*}) \,|\, \mathcal{F}_t] = \sum_y P(X_t, y)h(y) = h(X_t) = h(X_{t\wedge\tau^*}).$$

Because W is finite and the chain is irreducible, we have $\tau^* < +\infty$ a.s. See Lemma **??**. Moreover the process is bounded because h is bounded on W^c and W is finite. Hence by the bounded convergence theorem (or the optional stopping theorem)

$$h(x) = \mathbb{E}_x[h(X_0)] = \mathbb{E}_x[h(X_{t\wedge\tau^*})] \to \mathbb{E}_x[h(X_{\tau^*})], \qquad \forall x \in W,$$

which implies that h is unique.

For the existence, simply define

$$h(x) = \mathbb{E}_x[h(X_{\tau^*})], \qquad \forall x \in W,$$

and use the Markov property as in (3.16). ■

For some insights on what happens when the assumptions of Theorem 3.21 are not satisfied, see Exercise 3.3. For an alternative proof of uniqueness based on the maximum principle, see Exercise 3.4.

The previous result is related to the classical Dirichlet problem in partial differential equations. To see the connection, note first that the proof above still works if one only specifies h on the outer boundary of W

$$\partial_{\mathcal{V}} W = \{z \in V \backslash W \,:\, \exists y \in W, P(y,z) > 0\}.$$

Introduce the *Laplacian* operator on $\mathcal{N}$

Laplacian

$$\Delta_{\mathcal{N}} f(x) = \left[\sum_y P(x,y) f(y)\right] - f(x) = \sum_y P(x,y)[f(y) - f(x)].$$

We have proved that, under the assumptions of Theorem 3.21, there exists a unique solution to

$$\begin{cases} \Delta_{\mathcal{N}} f(x) = 0, & \forall x \in W, \\ f(x) = h(x), & \forall x \in \partial_{\mathcal{V}} W, \end{cases} \tag{3.19}$$

and that solution is given by $f(x) = \mathbb{E}_x[h(X_{\tau_{W^c}})]$, for $x \in W \cup \partial_{\mathcal{V}} W$. The system (3.19) is called a *Dirichlet problem*. The Laplacian above can be interpreted as a discretized version of the standard Laplacian. For instance, for simple random walk on $\mathbb{Z}$ (with $\pi \equiv 1$), $\Delta_{\mathcal{N}} f(x) = \frac{1}{2}\{[f(x+1) - f(x)] - [f(x) - f(x-1)]\}$ which is a discretized second derivative.

Dirichlet problem

Applications Before developing the electrical network theory, we point out that Theorem 3.21 has many more applications. One of its consequences is that harmonic functions on finite, connected networks are constant.

Corollary 3.22. *Let P be an irreducible transition matrix on a finite state space V. If h is harmonic on all of V, then it is constant.*

Proof. Fix the value of h at an arbitrary vertex z and set $W = V\backslash\{z\}$. Applying Theorem 3.21, for all $x \in W$, $h(x) = \mathbb{E}_x[h(X_{\tau_{W^c}})] = h(z)$. ∎

As an example of application of this corollary, we prove the following surprising result: in a finite, irreducible Markov chain, the expected time to hit a target chosen at random according to the stationary distribution does not depend on the starting point.

Theorem 3.23 (Random target lemma)**.** *Let (X_t) be an irreducible Markov chain on a finite state space V with transition matrix P and stationary distribution π. Then*

$$h(x) := \sum_{y\in V} \pi(y)\, \mathbb{E}_x[\tau_y]$$

does not in fact depend on x.

Proof. By assumption, $\mathbb{E}_x[\tau_y] < +\infty$ for all x, y. By Corollary 3.22, it suffices to show that $h(x) := \sum_y \pi(y)\, \mathbb{E}_x[\tau_y]$ is harmonic on all of V. As before, it is natural to expand $\mathbb{E}_x[\tau_y]$ according to the first step of the chain,

$$\mathbb{E}_x[\tau_y] = \mathbb{1}_{\{x\neq y\}} \left(1 + \sum_z P(x,z)\, \mathbb{E}_z[\tau_y]\right).$$

Substituting into $h(x)$ gives

$$\begin{aligned} h(x) &= (1-\pi(x)) + \sum_z \sum_{y\neq x} \pi(y) P(x,z)\, \mathbb{E}_z[\tau_y] \\ &= (1-\pi(x)) + \sum_z P(x,z)\left(h(z) - \pi(x)\, \mathbb{E}_z[\tau_x]\right). \end{aligned}$$

Rearranging, we get

$$\Delta_{\mathcal{N}} h(x) = \pi(x)\left(1 + \sum_z P(x,z)\mathbb{E}_z[\tau_x]\right) - 1 = 0,$$

where we used $1/\pi(x) = \mathbb{E}_x[\tau_x^+] = 1 + \sum_z P(x,z)\, \mathbb{E}_z[\tau_x]$. ∎

3.3.2 Basic electrical network theory

We now develop the basic theory of electrical networks and their connections to random walks. We begin with a few definitions.

Definitions Let $\mathcal{N} = (G, c)$ be a finite or countable network. Throughout this section we assume that $\mathcal{N}$ is connected and locally finite. In the context of electrical networks, edge weights are called *conductances*. The reciprocal of the conductances are called *resistances* and are denoted by $r(x, y) = 1/c(x, y)$, for all $x \sim y$. Both c and r are symmetric. For an edge $e = \{x, y\}$ we also write $c(e) := c(x, y)$ and $r(e) := r(x, y)$. Recall that the transition matrix of the random walk on $\mathcal{N}$ satisfies

conductance

resistance

$$P(x, y) = \frac{c(x, y)}{\sum_{y \sim x} c(x, y)}.$$

Let A, Z be disjoint, non-empty subsets of V such that $W := (A \cup Z)^c$ is finite. For our purposes it will suffice to take A to be a singleton, i.e. $A = \{a\}$ for some a. Then a is called the *source* and Z is called the *sink-set*, or *sink* for short.

source, sink

As an immediate corollary of Theorem 3.21, we obtain the existence and uniqueness of a voltage function, defined formally in the next corollary. It will be useful to consider voltages taking an arbitrary value at a, but we always set the voltage on Z to 0. Note in the definition below that if v is a voltage with value v_0 at a, then $\tilde{v}(x) = v(x)/v_0$ is a voltage with value 1 at a.

Corollary 3.24 (Voltage)**.** *Fix $v_0 > 0$. Let $\mathcal{N} = (G, c)$ be a finite or countable, connected network with $G = (V, E)$. Let $A := \{a\}$, Z be disjoint non-empty subsets of V such that $W = (A \cup Z)^c$ is non-empty and finite. Then there exists a unique* voltage*, i.e., a function v on V such that v is harmonic on W*

voltage

$$v(x) = \frac{1}{c(x)} \sum_{y \sim x} c(x, y) v(y), \qquad \forall x \in W, \tag{3.20}$$

where $c(x) = \sum_{y \sim x} c(x, y)$, and

$$v(a) = v_0 \qquad \textit{and} \qquad v|_Z \equiv 0. \tag{3.21}$$

Moreover

$$\frac{v(x)}{v_0} = \mathbb{P}_x[\tau_a < \tau_Z], \tag{3.22}$$

for the corresponding random walk on $\mathcal{N}$.

Proof. Set $h(x) = v(x)$ on $A \cup Z$. Theorem 3.21 gives the result. ■

Let v be a voltage function on $\mathcal{N}$ with source a and sink Z. The Laplacian-based formulation of harmonicity, (3.19), can be interpreted in flow terms as follows. We define the *current* function $i(x, y) := c(x, y)[v(x) - v(y)]$ or, equivalently, $v(x) - v(y) = r(x, y)\, i(x, y)$. The latter definition is usually referred to as *Ohm's "law."* Notice that the current function is defined on ordered pairs of vertices and is anti-symmetric, i.e., $i(x, y) = -i(y, x)$. In terms of the current function, the harmonicity of v is then expressed as

current

Ohm's law

$$\sum_{y \sim x} i(x, y) = 0, \qquad \forall x \in W,$$

i.e., i is a flow on W. This set of equations is known as *Kirchhoff's node law*. We also refer to these constraints as flow-conservation constraints. To be clear, the current function is not just any flow. It is a flow that can be written as a potential difference according to Ohm's law. Such a current also satisfies *Kirchhoff's cycle law*: if $x_1 \sim x_2 \sim \cdots \sim x_k \sim x_{k+1} = x_1$ is a cycle, then

Kirchhoff's node law

Kirchhoff's cycle law

$$\sum_{j=1}^{k} i(x_j, x_{j+1})\, r(x_j, x_{j+1}) = 0,$$

as can be seen by substituting Ohm's law. The *strength* of the current is defined as

strength

$$\|i\| = \sum_{y \sim a} i(a, y).$$

The definition of $i(x, y)$ ensures that the flow out of the source is nonnegative as $\mathbb{P}_y[\tau_a < \tau_Z] \leq 1 = \mathbb{P}_a[\tau_a < \tau_Z]$ for all $y \sim a$. Note that by multiplying the voltage by a constant we obtain a current which is similarly scaled. Up to that scaling, the current function is unique from the uniqueness of the voltage. We will often consider the *unit current* where we scale v and i so as to enforce that $\|i\| = 1$.

unit current

Remark 3.25. *Note that the definition of the current depends crucially on the reversibility of the chain, i.e., on the fact that* $c(x, y) = c(y, x)$. *For non-reversible chains, it is not clear how to interpret the system* (3.19) *as flow conservation as it involves only the outgoing transitions (which in general are not related to the incoming transitions).*

Summing up the previous paragraph, to determine the voltage it suffices to find functions v and i that simultaneously satisfy Ohm's law and Kirchhoff's node law. Here is an example.

Example 3.26 (Network reduction: birth-and-death chain)**.** Let $\mathcal{N}$ be the line on $\{0, 1, \ldots, n\}$ with $j \sim k \iff |j - k| = 1$ and arbitrary (positive) conductances on the edges. Let (X_t) be the corresponding walk. We use the principle above to compute $\mathbb{P}_x[\tau_0 < \tau_n]$ for $1 \leq x \leq n-1$. Consider the voltage function v when $v(0) = 1$ and $v(n) = 0$ with current i. The desired quantity is $v(x)$ by Corollary 3.24. Note that because i is a flow on $\mathcal{N}$, the flow into every vertex equals the flow out of that vertex, and we must have $i(y, y+1) = i(0,1) = \|i\|$ for all y. To compute $v(x)$, we note that it remains the same if we replace the path $0 \sim 1 \sim \cdots \sim x$ with a single edge of resistance $R_{0,x} = r(0,1) + \cdots + r(x-1, x)$. Indeed leave the voltage unchanged on the remaining nodes and define the current on the new edge as $\|i\|$. Kirchhoff's node law is automatically satisfied by the argument above. To check Ohm's law on the new "super-edge," note that on the original network $\mathcal{N}$

$$\begin{aligned} v(0) - v(x) &= (v(0) - v(1)) + \cdots + (v(x-1) - v(x)) \\ &= r(x-1, x)i(x-1, x) + \cdots + r(0,1)i(0,1) \\ &= [r(0,1) + \cdots + r(x-1, x)]\|i\| \\ &= R_{0,x}\|i\|. \end{aligned}$$

Ohm's law is also satisfied on every other edge because nothing has changed there. That proves the claim. We do the same reduction on the other side of x by replacing $x \sim x+1 \sim \cdots \sim n$ with a single edge of resistance $R_{x,n} = r(x, x+1) + \cdots + r(n-1, n)$. See Figure 3.26. Because the voltage at x has not changed, we can compute $v(x) = \mathbb{P}_x[\tau_0 < \tau_n]$ directly on the reduced network, where it is now a straightforward computation. Indeed, starting at x, the reduced walk jumps to 0 with probability proportional to the conductance on the new super-edge $0 \sim x$ (or the reciprocal of the resistance), i.e.,

$$\begin{aligned} \mathbb{P}_x[\tau_0 < \tau_n] &= \frac{R_{0,x}^{-1}}{R_{0,x}^{-1} + R_{x,n}^{-1}} \\ &= \frac{R_{x,n}}{R_{x,n} + R_{0,x}} \\ &= \frac{r(x, x+1) + \cdots + r(n-1, n)}{r(0,1) + \cdots + r(n-1, n)}. \end{aligned}$$

Some special cases:

- *Simple random walk.* In the case of simple random walk, all resistances are equal and we get

$$\mathbb{P}_x[\tau_0 < \tau_n] = \frac{n-x}{n}.$$

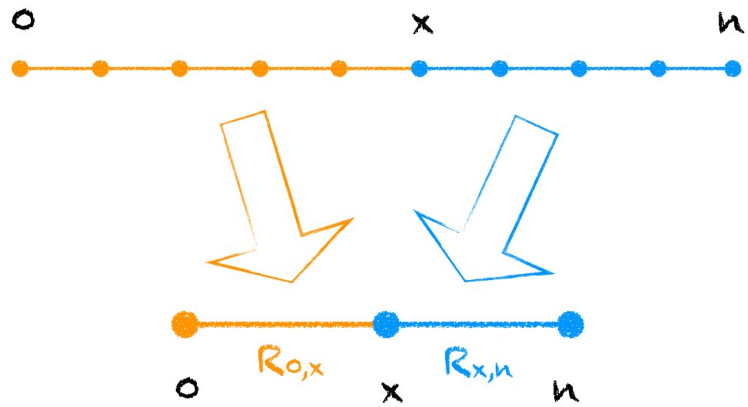

Figure 3.4: Reduced network.

- *Gambler's ruin.* The gambler's ruin example corresponds to taking $c(j, j+1) = (p/q)^j$ or $r(j, j+1) = (q/p)^j$, for some $0 < p < 1$. In this case we obtain

$$\mathbb{P}_x[\tau_0 < \tau_n] = \frac{\sum_{j=x}^{n-1}(q/p)^j}{\sum_{j=0}^{n-1}(q/p)^j} = \frac{(q/p)^x(1-(q/p)^{n-x})}{1-(q/p)^n} = \frac{(p/q)^{n-x}-1}{(p/q)^n-1},$$

when $p \neq q$ (otherwise we get back the simple random walk case).

◀

The above example illustrates the *series law*: resistances in series add up. There is a similar *parallel law*: conductances in parallel add up. To formalize these laws, one needs to introduce multigraphs. This is straightforward, but to avoid complicating the notation further we will not do this here. (See the "Bibliographic remarks" for more.) Another useful network reduction technique is *shorting*, in which we identify, or glue together, vertices with the same voltage while keeping existing edges. Here is an example.

series law, parallel law

shorting

Example 3.27 (Network reduction: binary tree)**.** Let $\mathcal{N}$ be the rooted binary tree with n levels $\widehat{\mathbb{T}}_2^n$ and equal conductances on all edges. Let 0 be the root. Pick an arbitrary leaf and denote it by n. The remaining vertices on the path between 0 and n, which we refer to as the main path, will be denoted by $1, \ldots, n-1$ moving away from the root. We claim that, for all $0 < x < n$, it holds that

$$\mathbb{P}_x[\tau_0 < \tau_n] = (n - x)/n.$$

Indeed let v be the voltage with values 1 and 0 at $a = 0$ and $Z = \{z\}$ with $z = n$ respectively. Let i be the corresponding current. Notice that, for each $0 \leq y < n$, the current—as a flow—has nowhere to go on the subtree T_y hanging from y away from the main path. The leaves of the subtree are dead ends. Hence the current must be 0 on T_y and by Ohm's law the voltage must be constant on it, i.e., every vertex in T_y has voltage $v(y)$. Imagine collapsing all vertices in T_y, including y, into a single vertex (and removing the self-loops so created). Doing this for every vertex on the main path results in a new reduced network which is formed of a single path as in Example 3.26. Note that the voltage and the current can be taken to be the same as they were previously on the main path. Indeed, with this choice, Ohm's law is automatically satisfied. Moreover, because there is no current on the hanging subtrees in the original network, Kirchhoff's node law is also satisfied on the reduced network, as no current is lost. Hence the answer can be obtained from Example 3.26. That proves the claim. (You should convince yourself that this result is obvious from a probabilistic point of view.) ◀

We gave a probabilistic interpretation of the voltage. What about the current? The following result says that, roughly speaking, $i(x, y)$ is the net traffic on the edge $\{x, y\}$ from x to y. We start with an important formula for the voltage at a. For the walk started at a, we use the shorthand

$$\mathbb{P}[a \to Z] := \mathbb{P}_a[\tau_Z < \tau_a^+],$$

for the *escape probability*.

escape probability

Lemma 3.28. *Let v be a voltage on $\mathcal{N}$ with source a and sink Z. Let i be the associated current. Then*

$$\frac{v(a)}{\|i\|} = \frac{1}{c(a)\,\mathbb{P}[a \to Z]}. \tag{3.23}$$

Proof. Using the usual one-step trick,

$$\begin{aligned}
\mathbb{P}[a \to Z] &= \sum_{x\sim a} P(a,x)\mathbb{P}_x[\tau_Z < \tau_a] \\
&= \sum_{x\sim a} \frac{c(a,x)}{c(a)}\left(1 - \frac{v(x)}{v(a)}\right) \\
&= \frac{1}{c(a)v(a)} \sum_{x\sim a} c(a,x)[v(a) - v(x)] \\
&= \frac{1}{c(a)v(a)} \sum_{x\sim a} i(a,x),
\end{aligned}$$

where we used Corollary 3.24 on the second line and Ohm's law on the last line. Rearranging gives the result. ■

Theorem 3.29 (Probabilistic interpretation of the current)**.** *For $x \sim y$, let $N^Z_{x\to y}$ be the number of transitions from x to y up to the time of the first visit to the sink Z for the random walk on $\mathcal{N}$ started at a. Let v be the voltage corresponding to the unit current i. Then the following formulas hold:*

$$v(x) = \frac{\mathscr{G}_{\tau_Z}(a,x)}{c(x)}, \qquad \forall x, \tag{3.24}$$

and

$$i(x,y) = \mathbb{E}_a[N^Z_{x\to y} - N^Z_{y\to x}], \qquad \forall x \sim y.$$

Proof. We prove the formula for the voltage by showing that $v(x)$ as defined above is harmonic on $W = V\backslash\{a\} \cup Z$. Note first that $\mathscr{G}_{\tau_Z}(a,z) = 0$ for all $z \in Z$ by definition, or $0 = v(z) = \frac{\mathscr{G}_{\tau_Z}(a,z)}{c(z)}$. Moreover, to compute $\mathscr{G}_{\tau_Z}(a,a)$, note that the number of visits to a before the first visit to Z is geometric with success probability $\mathbb{P}[a \to Z]$ by the strong Markov property and hence

$$\mathscr{G}_{\tau_Z}(a,a) = \frac{1}{\mathbb{P}[a \to Z]},$$

and by the previous lemma $v(a) = \frac{\mathscr{G}_{\tau_Z}(a,a)}{c(a)}$, as required. To establish the formula for $x \in W$, we compute the quantity $\frac{1}{c(x)}\sum_{y\sim x}\mathbb{E}_a[N^Z_{y\to x}]$ in two ways. First, because each visit to $x \in W$ must enter through one of x's neighbors (including itself in the presence of a self-loop), we get

$$\frac{1}{c(x)}\sum_{y\sim x}\mathbb{E}_a[N^Z_{y\to x}] = \frac{\mathscr{G}_{\tau_Z}(a,x)}{c(x)}. \tag{3.25}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}_a[N^Z_{y\to x}] &= \mathbb{E}_a\left[\sum_{0\le t<\tau_Z} \mathbb{1}_{\{X_t=y, X_{t+1}=x\}}\right] \\
&= \sum_{t\ge 0} \mathbb{P}_a\left[X_t = y, X_{t+1} = x, \tau_Z > t\right] \\
&= \sum_{t\ge 0} \mathbb{P}_a[\tau_Z > t]\mathbb{P}_a[X_t = y \,|\, \tau_Z > t]P(y,x) \\
&= P(y,x)\,\mathbb{E}_a\left[\sum_{0\le t<\tau_Z} \mathbb{1}_{\{X_t=y\}}\right] \\
&= P(y,x)\,\mathscr{G}_{\tau_Z}(a,y), \qquad (3.26)
\end{aligned}$$

so that, summing over y, we obtain this time

$$\frac{1}{c(x)}\sum_{y\sim x}\mathbb{E}_a[N^Z_{y\to x}] = \frac{1}{c(x)}\sum_{y\sim x} P(y,x)\,\mathscr{G}_{\tau_Z}(a,y) = \sum_{y\sim x} P(x,y)\frac{\mathscr{G}_{\tau_Z}(a,y)}{c(y)}, \qquad (3.27)$$

where we used reversibility. Equating (3.25) and (3.27) shows that $\frac{\mathscr{G}_{\tau_Z}(a,x)}{c(x)}$ is harmonic on W and hence must be equal to the voltage function by Corollary 3.24.

Finally, by (3.26),

$$\begin{aligned}
\mathbb{E}_a[N^Z_{x\to y} - N^Z_{y\to x}] &= P(x,y)\,\mathscr{G}_{\tau_Z}(a,x) - P(y,x)\,\mathscr{G}_{\tau_Z}(a,y) \\
&= P(x,y)v(x)c(x) - P(y,x)v(y)c(y) \\
&= c(x,y)[v(x) - v(y)] \\
&= i(x,y).
\end{aligned}$$

That concludes the proof. ■

Remark 3.30. *Formula* (3.24) *relies crucially on reversibility. Indeed assume the chain has stationary distribution* π. *Then, in probabilistic terms,* (3.24) *reads*

$$\pi(x)\mathbb{P}_x[\tau_a < \tau_Z] = \frac{\mathscr{G}_{\tau_Z}(a,x)}{\pi(a)\mathbb{P}[a\to Z]},$$

where we used (3.22) *and* (3.23)*, and the fact that the current has unit strength. This is not true in general for non-reversible chains. Take for instance a deterministic walk on a directed cycle of size* n *with* x *on the directed path from* a *to* $Z = \{z\}$. *In that case the l.h.s. is* 0 *but the r.h.s. is* n.

Example 3.31 (Network reduction: binary tree (continued))**.** Recall the setting of Example 3.27. We argued that the current on side edges, i.e., edges of subtrees hanging from the main path, is 0. This is clear from the probabilistic interpretation of the current: in a walk from a to z, any traversal of a side edge must be undone at a later time. ◀

The network reduction techniques illustrated above are useful. But the power of the electrical network perspective is more apparent in what comes next: the definition of the effective resistance and, especially, its variational characterization.

Effective resistance Before proceeding further, let us recall our original motivation. Let $\mathcal{N} = (G, c)$ be a countable, locally finite, connected network and let (X_t) be the corresponding walk. Recall that a vertex a in G is transient if $\mathbb{P}_a[\tau_a^+ < +\infty] < 1$.

To relate this to our setting, consider an *exhaustive sequence* of induced subgraphs G_n of G which for our purposes is defined as: G_0 contains only a, $G_n \subseteq G_{n+1}$, $G = \bigcup_n G_n$, and every G_n is finite and connected. Such a sequence always exists by iteratively adding the neighbors of the previous vertices and using that G is locally finite and connected. Let Z_n be the set of vertices of G not in G_n. Then, by Lemma **??**, $\mathbb{P}_a[\tau_{Z_n} \wedge \tau_a^+ = +\infty] = 0$ for all n by our assumptions on (G_n). Hence, the remaining possibilities are

exhaustive sequence

$$\begin{aligned} 1 &= \mathbb{P}_a[\exists n,\ \tau_a^+ < \tau_{Z_n}] + \mathbb{P}_a[\forall n,\ \tau_{Z_n} < \tau_a^+] \\ &= \mathbb{P}_a[\tau_a^+ < +\infty] + \lim_n \mathbb{P}[a \to Z_n]. \end{aligned}$$

Therefore a is transient if and only if $\lim_n \mathbb{P}[a \to Z_n] > 0$. Note that the limit exists because the sequence of events $\{\tau_{Z_n} < \tau_a^+\}$ is decreasing by construction. By a sandwiching argument the limit also does not depend on the exhaustive sequence. (Exercise.) Hence we define

$$\mathbb{P}[a \to \infty] := \lim_n \mathbb{P}[a \to Z_n] > 0.$$

We use Lemma 3.28 to characterize this limit using electrical network notions.

But, first, here comes the key definition. In Lemma 3.28, $v(a)$ can be thought of as the potential difference between the source and the sink, and $\|i\|$ can be thought of as the total current flowing through the network from the source to the sink. Hence, viewing the network as a single "super-edge," Equation (3.23) is the analogue of Ohm's law if we interpret $c(a)\,\mathbb{P}[a \to Z]$ as a "conductance."

Definition 3.32 (Effective resistance and conductance)**.** *Let $\mathcal{N} = (G, c)$ be a finite or countable, locally finite, connected network. Let $A = \{a\}$ and Z be disjoint* *non-empty subsets of the vertex set V such that $W := V \backslash (A \cup Z)$ is finite. Let v be a voltage from source a to sink Z and let i be the corresponding current. The* effective resistance *between a and Z is defined as*

effective resistance

$$\mathscr{R}(a \leftrightarrow Z) := \frac{1}{c(a)\, \mathbb{P}[a \to Z]} = \frac{v(a)}{\|i\|},$$

where the rightmost equality holds by Lemma 3.28. The reciprocal is called the effective conductance *and denoted by $\mathscr{C}(a \leftrightarrow Z) := 1/\mathscr{R}(a \leftrightarrow Z)$.*

effective conductance

Going back to recurrence, for an exhaustive sequence (G_n) with (Z_n) as above, it is natural to define

$$\mathscr{R}(a \leftrightarrow \infty) := \lim_n \mathscr{R}(a \leftrightarrow Z_n),$$

where, once again, the limit does not depend on the choice of exhaustive sequence.

Theorem 3.33 (Recurrence and resistance)**.** *Let $\mathcal{N} = (G, c)$ be a countable, locally finite, connected network. Vertex a (and hence all vertices) in $\mathcal{N}$ is transient if and only if $\mathscr{R}(a \leftrightarrow \infty) < +\infty$.*

Proof. This follows immediately from the definition of the effective resistance. Recall that, on a connected network, all states have the same *type* (recurrent or transient). ■

type

Note that the network reduction techniques we discussed previously leave both the voltage and the current strength unchanged on the reduced network. Hence they also leave the effective resistance unchanged.

Example 3.34 (Gambler's ruin chain revisited)**.** Extend the gambler's ruin chain of Example 3.26 to all of $\mathbb{Z}_+$. We determine when this chain is transient. Because it is irreducible, all states have the same type and it suffices to look at 0. Consider the exhaustive sequence obtained by letting G_n be the graph restricted to $[0, n-1]$ and letting $Z_n = [n, +\infty)$. To compute the effective resistance $\mathscr{R}(0 \leftrightarrow Z_n)$, we use the same reduction as in Example 3.26, except that this time we reduce the network all the way to a single edge. That edge has resistance

$$\mathscr{R}(0 \leftrightarrow Z_n) = \sum_{j=0}^{n-1} r(j, j+1) = \sum_{j=0}^{n-1} (q/p)^j = \frac{(q/p)^n - 1}{(q/p) - 1},$$

when $p \neq q$, and similarly it has value n in the $p = q$ case. Hence

$$\mathscr{R}(0 \leftrightarrow \infty) = \begin{cases} +\infty, & p \leq 1/2, \\ \frac{p}{2p-1}, & p > 1/2. \end{cases}$$

So 0 is transient if and only if $p > 1/2$. ◀

Example 3.35 (Biased walk on the b-ary tree)**.** Fix $\lambda \in (0, +\infty)$. Consider the rooted, infinite b-ary tree with conductance λ^j on all edges between level $j-1$ and j, $j \geq 1$. We determine when this chain is transient. Because it is irreducible, all states have the same type and it suffices to look at the root. Denote the root by 0. For an exhaustive sequence, let G_n be the root together with the first $n-1$ levels. Let Z_n be as before. To compute $\mathscr{R}(0 \leftrightarrow Z_n)$: 1) glue together all vertices of Z_n; 2) glue together all vertices on the same level of G_n; 3) replace parallel edges with a single edge whose conductance is the sum of the conductances; 4) let the current on this edge be the sum of the currents; and 5) leave the voltages unchanged. Note that Ohm's law and Kirchhoff's node law are still satisfied. Hence we have not changed the effective resistance. (This is an application of the parallel law.) The reduced network is now a line. Denote the new vertices $0, 1, \ldots, n$. The conductance on the edge between j and $j+1$ is $b^{j+1}\lambda^j = b(b\lambda)^j$. So this is the chain from the previous example with $(p/q) = b\lambda$ where all conductances are scaled by a factor of b. Hence

$$\mathscr{R}(0 \leftrightarrow \infty) = \begin{cases} +\infty, & b\lambda \leq 1, \\ \frac{1}{b(1-(b\lambda)^{-1})}, & b\lambda > 1. \end{cases}$$

So the root is transient if and only if $b\lambda > 1$. ◀

3.3.3 Bounding the effective resistance

The examples we analyzed so far were atypical in that it was possible to reduce the network down to a single edge using simple rules and read off the effective resistance. In general, we need more robust techniques to bound the effective resistance. The following two variational principles provide a powerful approach for this purpose.

Variational principles Recall that *flow θ from source a to sink Z* on a countable, locally finite, connected network $\mathcal{N} = (G, c)$ is a function on pairs of adjacent vertices such that: θ is anti-symmetric, i.e., $\theta(x,y) = -\theta(y,x)$ for all $x \sim y$; and it satisfies the flow-conservation constraint $\sum_{y \sim x} \theta(x,y) = 0$ on all vertices x except those in $\{a\} \cup Z$. The strength of the flow is $\|\theta\| := \sum_{y \sim a} \theta(a,y)$. The current is a special flow—one that can be written as a potential difference according to Ohm's law. As we show next, it can also be characterized as a flow minimizing a certain energy. Specifically, the *energy* of a flow θ is defined as

flow

energy

$$\mathscr{E}(\theta) = \frac{1}{2} \sum_{x,y} r(x,y)[\theta(x,y)]^2.$$

The proof of the variational principle we present here employs a neat trick, convex duality. In particular, it reveals that the voltage and current are dual in the sense of convex analysis.

Theorem 3.36 (Thomson's principle)**.** *Let $\mathcal{N} = (G, c)$ be a finite, connected network. The effective resistance between source a and sink Z is characterized by*

$$\mathscr{R}(a \leftrightarrow Z) = \inf \{\mathscr{E}(\theta) \,:\, \theta \text{ is a unit flow between } a \text{ and } Z\}. \tag{3.28}$$

The unique minimizer is the unit current.

Proof. It will be convenient to work in matrix form. Choose an arbitrary orientation of $\mathcal{N}$, i.e., replace each edge $\{x, y\}$ with either $\langle x, y\rangle$ or $\langle y, x\rangle$. Let $\overrightarrow{G}$ be the corresponding directed graph. Think of the flow $\boldsymbol{\theta}$ as a vector with one component for each oriented edge. Then the flow constraint can be written as a linear system $A\boldsymbol{\theta} = \boldsymbol{b}$. Here the matrix A has a column for each edge and a row for each vertex *except those in* Z. The entries of A are of the form $A_{x,\langle x,y\rangle} = 1$, $A_{y,\langle x,y\rangle} = -1$, and 0 otherwise. The vector $\boldsymbol{b}$ has 0s everywhere except for $b_a = 1$. Let $\boldsymbol{r}$ be the vector of resistances and let R be the diagonal matrix with diagonal $\boldsymbol{r}$. In matrix form the optimization problem (3.28) reads

$$\mathscr{E}^* = \inf\{\boldsymbol{\theta}' R \boldsymbol{\theta} \,:\, A\boldsymbol{\theta} = \boldsymbol{b}\},$$

where $'$ denotes the transpose.

Introduce the *Lagrangian*

Lagrangian

$$\mathscr{L}(\boldsymbol{\theta}; \boldsymbol{h}) := \boldsymbol{\theta}' R \boldsymbol{\theta} - 2\boldsymbol{h}'(A\boldsymbol{\theta} - \boldsymbol{b}),$$

where $\boldsymbol{h}$ has an entry for all vertices except those in Z. (The reason for the factor of 2 will be clear below.) For all $\boldsymbol{h}$,

$$\mathscr{E}^* \geq \inf_{\boldsymbol{\theta}} \mathscr{L}(\boldsymbol{\theta}; \boldsymbol{h}),$$

because those $\boldsymbol{\theta}$s with $A\boldsymbol{\theta} = \boldsymbol{b}$ make the second term vanish in $\mathscr{L}(\boldsymbol{\theta}; \boldsymbol{h})$. Since $\mathscr{L}(\boldsymbol{\theta}; \boldsymbol{h})$ is strictly convex as a function of $\boldsymbol{\theta}$, the solution is characterized by the usual optimality condition which in this case reads $2R\boldsymbol{\theta} - 2A'\boldsymbol{h} = 0$, or

$$\boldsymbol{\theta} = R^{-1} A' \boldsymbol{h}. \tag{3.29}$$

Substituting into the Lagrangian and simplifying, we have proved that

$$\mathscr{E}(\boldsymbol{\theta}) \geq \mathscr{E}^* \geq -\boldsymbol{h}' A R^{-1} A' \boldsymbol{h} + 2\boldsymbol{h}'\boldsymbol{b} =: \mathscr{L}^*(\boldsymbol{h}), \qquad \forall \boldsymbol{h} \text{ and flow } \boldsymbol{\theta}. \tag{3.30}$$

This inequality is a statement of weak duality.

To show that a flow $\boldsymbol{\theta}$ is optimal it suffices to find $\boldsymbol{h}$ such that the l.h.s. in (3.30) equals $\mathscr{E}(\boldsymbol{\theta}) = \boldsymbol{\theta}' R \boldsymbol{\theta}$. Not surprisingly, when $\boldsymbol{\theta}$ is the unit current, the suitable $\boldsymbol{h}$ turns out to be the corresponding voltage. To see this, observe that $A'\boldsymbol{h}$ is the vector of neighboring node differences

$$A'\boldsymbol{h} = (h(x) - h(y))_{\langle x,y\rangle \in \overrightarrow{G}}\,. \tag{3.31}$$

Hence the optimality condition (3.29) is nothing but Ohm's law in matrix form. Therefore, if $\boldsymbol{i}$ is the unit current and $\boldsymbol{v}$ is the associated voltage in vector form, it holds that

$$\mathscr{L}^*(\boldsymbol{v}) = \mathscr{L}(\boldsymbol{i}; \boldsymbol{v}) = \mathscr{E}(\boldsymbol{i}),$$

where the first equality follows from the optimality of $\boldsymbol{v}$ and the second equality follows from the fact that $A\boldsymbol{i} = \boldsymbol{b}$. So we must have $\mathscr{E}(\boldsymbol{i}) = \mathscr{E}^*$. As for uniqueness, note that two minimizers $\boldsymbol{\theta}$, $\boldsymbol{\theta}'$ satisfy

$$\mathscr{E}^* = \frac{\mathscr{E}(\boldsymbol{\theta}) + \mathscr{E}(\boldsymbol{\theta}')}{2} = \mathscr{E}\left(\frac{\boldsymbol{\theta} + \boldsymbol{\theta}'}{2}\right) + \mathscr{E}\left(\frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{2}\right).$$

The first term on the r.h.s. is greater or equal than $\mathscr{E}^*$ because the average of two unit flows is still a unit flow. The second term is nonnegative by definition. Hence the latter must be zero and $\boldsymbol{\theta} = \boldsymbol{\theta}'$.

To conclude the proof, it remains to compute the optimal value. The matrix

$$\Delta_{\mathcal{N}} := AR^{-1}A',$$

can be interpreted as the Laplacian operator of Section 3.3.1 in matrix form, i.e., for each row x it takes a conductance-weighted average of the neighboring values and subtracts the value at x

$$\begin{aligned}
\left(AR^{-1}A'\boldsymbol{v}\right)_x = &\sum_{y:\langle x,y\rangle \in \overrightarrow{G}} \Big[c(x,y)(v(x) - v(y))\Big] \\
&\qquad - \sum_{y:\langle y,x\rangle \in \overrightarrow{G}} \Big[c(y,x)(v(y) - v(x))\Big] \\
= &\sum_{y\sim x} \Big[c(x,y)(v(x) - v(y))\Big],
\end{aligned}$$

where we used (3.31) and the fact that $r(x,y)^{-1} = c(x,y)$ and $c(x,y) = c(y,x)$. So $\Delta_{\mathcal{N}}\boldsymbol{v}$ is zero everywhere except for the row corresponding to a where it is

$$\sum_{y\sim a} c(a,y)[v(a) - v(y)] = \sum_{y\sim a} i(a,y) = 1,$$

where we used Ohm's law and the fact that the current has unit strength. We have finally

$$\mathscr{E}^* = \mathscr{L}^*(\boldsymbol{v}) = -\boldsymbol{v}' A R^{-1} A' \boldsymbol{v} + 2\boldsymbol{v}'\boldsymbol{b} = -v(a) + 2v(a) = v(a) = \mathscr{R}(a \leftrightarrow Z),$$

by (3.28), where used again that $\|i\| = 1$. ■

Observe that the convex combination $\boldsymbol{\alpha}$ minimizing the sum of squares $\sum_j \alpha_j^2$ is uniform. In a similar manner, Thomson's principle stipulates roughly speaking that the more the flow can be spread out over the network, the lower is the effective resistance (penalizing flow on edges with higher resistance). Pólya's theorem below provides a vivid illustration. Here is a simple example suggesting that, in a sense, the current is indeed a well distributed flow.

Example 3.37 (Random walk on the complete graph)**.** Let $\mathcal{N}$ be the complete graph on $\{1, \ldots, n\}$ with unit resistances, and let $a = 1$ and $Z = \{n\}$. Assume $n > 2$. The effective resistance is straightforward to compute in this case. Indeed, the escape probability (with a slight abuse of notation) is

$$\mathbb{P}[1 \to n] = \frac{1}{n-1} + \frac{1}{2}\left(1 - \frac{1}{n-1}\right) = \frac{n}{2(n-1)},$$

as we either jump to n immediately or jump to one of the remaining nodes, in which case we reach n first with probability $1/2$ by symmetry. Hence, since $c(1) = n-1$, we get

$$\mathscr{R}(1 \leftrightarrow n) = \frac{2}{n},$$

from the definition of the effective resistance. We now look for the optimal flow. Putting a flow of 1 on the edge $(1, n)$ gives an upper bound of 1, which is far from the optimal $\frac{2}{n}$. Spreading the flow a bit more by pushing $1/2$ through the edge $(1, n)$ and $1/2$ through the path $1 \sim 2 \sim n$ gives the slightly better bound $1/4 + 2(1/4) = 3/4$. Taking this further, putting a flow of $\frac{1}{n-1}$ on $(1, n)$ as well as on each two-edge path to n through the remaining neighbors of 1 gives the yet improved bound

$$\frac{1}{(n-1)^2} + 2(n-2)\frac{1}{(n-1)^2} = \frac{2n-3}{(n-1)^2} = \frac{2}{n} \cdot \frac{2n^2 - 3n}{2n^2 - 4n + 2} > \frac{2}{n},$$

when $n > 2$. Because the direct path from 1 to n has a somewhat lower resistance, the optimal flow is obtained by increasing the flow on that edge slightly. Namely, for a flow α on $(1, n)$ we get an energy of $\alpha^2 + 2(n-2)[\frac{1-\alpha}{n-2}]^2$ which is minimized at $\alpha = \frac{2}{n}$ where it is indeed

$$\left(\frac{2}{n}\right)^2 + \frac{2}{n-2}\left(\frac{n-2}{n}\right)^2 = \frac{2}{n}\left(\frac{2}{n} + \frac{n-2}{n}\right) = \frac{2}{n}.$$

◀

The matrix $\Delta_{\mathcal{N}} = AR^{-1}A'$ in the proof of Thomson's principle is the *Laplacian matrix*. As we noted above, because $A'\boldsymbol{h}$ is the vector of neighboring node differences, we have

Laplacian matrix

$$\boldsymbol{h}'\Delta_{\mathcal{N}}\boldsymbol{h} = \frac{1}{2}\sum_{x,y} c(x,y)[h(y) - h(x)]^2,$$

where we implicitly fix $h|_Z \equiv 0$, which is called the *Dirichlet energy*. Thinking of $\nabla_{\mathcal{N}} := A'$ as a discrete *gradient* operator, the Dirichlet energy can be interpreted as the weighted norm of the gradient of $\boldsymbol{h}$. The following dual to Thomson's principle is essentially a reformulation of the Dirichlet problem. Exercise 3.6 asks for a proof.

Dirichlet energy, gradient

Theorem 3.38 (Dirichlet's principle)**.** *Let $\mathcal{N} = (G, c)$ be a finite, connected network. The effective conductance between source a and sink Z is characterized by*

$$\mathscr{C}(a \leftrightarrow Z) = \inf\left\{\frac{1}{2}\sum_{x,y} c(x,y)[h(y) - h(x)]^2 \;:\; h(a) = 1,\; h|_Z \equiv 0\right\}.$$

The unique minimizer is the voltage v with $v(a) = 1$.

The following lower bound is a typical application of Thomson's principle. See Pólya's theorem below for an example of its use.

Definition 3.39 (Cutset)**.** *On a finite graph, a* cutset *separating a from Z is a set of edges Π such that any path between a and Z must include at least one edge in Π. Similarly, on a countable network, a cutset separating a from ∞ is a set of edges that must be crossed by any infinite self-avoiding path from a.*

cutset

Corollary 3.40 (Nash-Williams inequality)**.** *Let $\mathcal{N}$ be a finite, connected network and let $\{\Pi_j\}_{j=1}^n$ be a collection of disjoint cutsets separating source a from sink Z. Then*

$$\mathscr{R}(a \leftrightarrow Z) \geq \sum_{j=1}^{n}\left(\sum_{e\in\Pi_j} c(e)\right)^{-1}.$$

Similarly, if $\mathcal{N}$ is a countable, locally finite, connected network, then for any collection $\{\Pi_j\}_j$ of finite, disjoint cutsets separating a from ∞,

$$\mathscr{R}(a \leftrightarrow \infty) \geq \sum_{j}\left(\sum_{e\in\Pi_j} c(e)\right)^{-1}.$$

Proof. We will need the following lemma.

Lemma 3.41. *Let $\mathcal{N}$ be finite. For any flow θ between a and Z and any cutset Π separating a from Z, it holds that*

$$\sum_{e \in \Pi} |\theta(e)| \geq \|\theta\|.$$

Proof. Intuitively, the flow out of a must cross Π to reach Z. Formally, let W_Π be the set of vertices reachable from a without crossing Π, let Z_Π be the set of vertices not in W_Π that are incident with an edge in Π and let $V_\Pi = W_\Pi \cup Z_\Pi$. For $x \in W_\Pi \backslash \{a\}$, note that by definition of a cutset $x \notin Z$. Moreover, all neighbors of x in V are in fact in V_Π: if $y \sim x$ is not in Z_Π then it is reachable from a through x without crossing Π and therefore it is in W_Π. Hence,

$$\sum_{y \in V_\Pi \,:\, y \sim x} \theta(x, y) = \sum_{y \in V \,:\, y \sim x} \theta(x, y) = 0, \tag{3.32}$$

or in other words θ is a flow from a to Z_Π on the graph G_Π induced by V_Π. By the same argument, this flow has strength

$$\sum_{y \in V_\Pi \,:\, y \sim a} \theta(a, y) = \sum_{y \in V \,:\, y \sim a} \theta(a, y) = \|\theta\|. \tag{3.33}$$

By (3.32) and (3.33) and the anti-symmetry of θ,

$$\begin{aligned}
0 &= \sum_{x \in V_\Pi} \sum_{y \in V_\Pi \,:\, y \sim x} \theta(x, y) \\
&= \|\theta\| + \sum_{x \in Z_\Pi} \sum_{y \in V_\Pi \,:\, y \sim x} \theta(x, y) \\
&= \|\theta\| + \sum_{x \in Z_\Pi} \sum_{y \in Z_\Pi \,:\, y \sim x} \theta(x, y) + \sum_{x \in Z_\Pi} \sum_{y \in W_\Pi \,:\, y \sim x} \theta(x, y) \\
&= \|\theta\| + \sum_{x \in Z_\Pi} \sum_{y \in W_\Pi \,:\, y \sim x} \theta(x, y) \\
&\geq \|\theta\| - \sum_{e \in \Pi} |\theta(e)|,
\end{aligned}$$

as $x \in Z_\Pi, y \in W_\Pi, y \sim x$ implies $\{x, y\} \in \Pi$. That concludes the proof. ■

Returning to the proof of the claim, consider the case where $\mathcal{N}$ is finite. For any unit flow from a to Z, by Cauchy-Schwarz and the lemma above

$$\begin{aligned}\sum_{e\in\Pi_j} c(e) \sum_{e\in\Pi_j} r(e)[\theta(e)]^2 &\geq \left(\sum_{e\in\Pi_j} \sqrt{c(e)r(e)}\ |\theta(e)|\right)^2 \\ &= \left(\sum_{e\in\Pi_j} |\theta(e)|\right)^2 \\ &\geq 1.\end{aligned}$$

Summing over j, using the disjointness of the cutsets, and rearranging gives the result in the finite case.

The infinite case follows from a similar argument. Note that, after removing a finite cutset Π separating a from $+\infty$, the connected component containing a must be finite by definition of Π. ■

Another typical application of Thomson's principle is the following monotonicity property (which is not obvious from a probabilistic point of view).

Corollary 3.42. *Adding an edge to a finite, connected network cannot increase the effective resistance between a source a and a sink Z. In particular, if the added edge is not incident to a, then $\mathbb{P}[a \to Z]$ cannot decrease.*

Proof. The additional edge enlarges the space of possible flows, so by Thomson's principle it can only lower the resistance or leave it as is. The second statement follows from the definition of the effective resistance. ■

More generally:

Corollary 3.43 (Rayleigh's principle)**.** *Let $\mathcal{N}$ and $\mathcal{N}'$ be two networks on the same finite, connected graph G such that, for each edge in G, the resistance in $\mathcal{N}'$ is greater than it is in $\mathcal{N}$. Then, for any source a and sink Z,*

$$\mathscr{R}_{\mathcal{N}}(a \leftrightarrow Z) \leq \mathscr{R}_{\mathcal{N}'}(a \leftrightarrow Z).$$

Proof. Compare the energies of an arbitrary flow on $\mathcal{N}$ and $\mathcal{N}'$, and apply Thomson's principle. ■

Application to recurrence Combining Theorem 3.33 and Thomson's principle, we derive a flow-based criterion for recurrence. To state the result, it is convenient to introduce the notion of a *unit flow θ from source a to ∞* on a countable, locally finite network: θ is anti-symmetric, it satisfies the flow-conservation constraint on all vertices but a, and $\|\theta\| := \sum_{y\sim a} \theta(a, y) = 1$. Note that the energy $\mathscr{E}(\theta)$ of such a flow is well defined in $[0, +\infty]$.

flow to ∞

Theorem 3.44 (Recurrence and finite-energy flows)**.** *Let $\mathcal{N} = (G, c)$ be a countable, locally finite, connected network. Vertex a (and hence all vertices) in $\mathcal{N}$ is transient if and only if there is a unit flow from a to ∞ of finite energy.*

Proof. One direction is immediate. Suppose such a flow exists and has energy bounded by $B < +\infty$. Let (G_n) be an exhaustive sequence with associated sinks (Z_n). A unit flow from a to ∞ on $\mathcal{N}$ yields, by projection, a unit flow from a to Z_n. This projected flow also has energy bounded by B. Hence Thomson's principle implies $\mathscr{R}(a \leftrightarrow Z_n) \leq B$ for all n and transience follows from Theorem 3.33.

Proving the other direction involves producing a flow to ∞. Suppose a is transient and let (G_n) be an exhaustive sequence as above. Then Theorem 3.33 implies that $\mathscr{R}(a \leftrightarrow Z_n) \leq \mathscr{R}(a \leftrightarrow \infty) < B$ for some $B < +\infty$ and Thomson's principle guarantees in turn the existence of a flow θ_n from a to Z_n with energy bounded by B. In particular there is a unit current i_n, and associated voltage v_n, of energy bounded by B. So it remains to use the sequence of current flows (i_n) to construct a flow to ∞ on the infinite network. The technical point is to show that the limit of (i_n) exists and is indeed a flow. For this, consider the random walk on $\mathcal{N}$ started at a. Let $Y_n(x)$ be the number of visits to x before hitting Z_n the first time. By the monotone convergence theorem, $\mathbb{E}_a Y_n(x) \to \mathbb{E}_a Y_\infty(x)$ where $Y_\infty(x)$ is the total number of visits to x. By (3.24), $\mathbb{E}_a Y_n(x) = c(x) v_n(x)$. So we can now define

$$v_\infty(x) := \lim_n v_n(x),$$

and then

$$i_\infty(x,y) := c(x,y)[v_\infty(y) - v_\infty(x)] = \lim_n c(x,y)[v_n(y) - v_n(x)] = \lim_n i_n(x,y),$$

by Ohm's law (when n is large enough that both x and y are in G_n). Because i_n is a flow for all n, by taking limits in the flow-conservation constraints we see that so is i_∞. Note that the partial sums

$$\sum_{x,y \in G_n} c(x,y)[i_\infty(x,y)]^2 = \lim_{\ell \geq n} \sum_{x,y \in G_n} c(x,y)[i_\ell(x,y)]^2 \leq \limsup_{\ell \geq n} \mathscr{E}(i_\ell) < B,$$

uniformly in n. Because the l.h.s. converges to the energy of i_∞ by the monotone convergence theorem, we are done. ■

Example 3.45 (Random walk on trees: recurrence[†])**.** To be written. See [Per99, Theorem 13.1]. ◀

We can now prove the following classical result.

[†]Requires: Section 2.2.4.

Theorem 3.46 (Pólya's theorem)**.** *Random walk on $\mathbb{L}^d$ is recurrent for $d \leq 2$ and transient for $d \geq 3$.*

We prove the theorem for $d = 2, 3$ using the tools developed in this section. The other cases follow by Rayleigh's principle. Of course, there are elementary proofs of this result. But we will show below that the electrical network approach has the advantage of being robust to the details of the lattice. For a different argument, see Exercise 2.7.

Proof of Theorem 3.46. The case $d = 2$ follows from the Nash-Williams inequality by letting Π_j be the set of edges connecting vertices of ℓ^∞ norm j and $j+1$. (Recall that the ℓ^∞ norm is defined as $\|x\|_\infty = \max_j |x_j|$.) Using the fact that all conductances are 1, that $|\Pi_j| = O(j)$, and that $\sum_j j^{-1}$ diverges, recurrence is established.

Now consider the case $d = 3$ and let $a = 0$. We construct a finite-energy flow to ∞ using the *method of random paths*. Note that a simple way to produce a unit flow to ∞ is to push a flow of 1 through an infinite self-avoiding path. Taking this a step further, let μ be a probability measure on infinite self-avoiding paths and define the anti-symmetric function

$$\theta(x, y) := \mathbb{E}[\mathbb{1}_{\langle x,y\rangle \in \Gamma} - \mathbb{1}_{\langle y,x\rangle \in \Gamma}] = \mathbb{P}[\langle x, y\rangle \in \Gamma] - \mathbb{P}[\langle y, x\rangle \in \Gamma],$$

where Γ is a random path distributed according to μ, oriented away from 0. Observe that $\sum_{y\sim x}[\mathbb{1}_{\langle x,y\rangle\in\Gamma} - \mathbb{1}_{\langle y,x\rangle\in\Gamma}] = 0$ for any $x \neq 0$ because vertices visited by Γ are entered and exited exactly once. That same sum is 1 at $x = 0$. Hence θ is a unit flow to ∞. Finally, for edge $e = \{x, y\}$, let

$$\mu(e) := \mathbb{P}[\langle x, y\rangle \in \Gamma \text{ or } \langle y, x\rangle \in \Gamma] = \mathbb{P}[\langle x, y\rangle \in \Gamma] + \mathbb{P}[\langle y, x\rangle \in \Gamma] \geq \theta(x, y),$$

where we used that a self-avoiding path Γ cannot visit both $\langle x, y\rangle$ and $\langle y, x\rangle$. Thomson's principle gives the following bound.

Claim 3.47 (Method of random paths)**.**

$$\mathscr{R}(0 \leftrightarrow \infty) \leq \sum_e [\mu(e)]^2. \tag{3.34}$$

For a measure μ concentrated on a single path, the sum above is infinite. To obtain a useful bound, what we need is a large collection of spread out paths. On the lattice $\mathbb{L}^3$, we construct μ as follows. Let $\boldsymbol{U}$ be a uniformly random point on the unit sphere in $\mathbb{R}^3$ and let γ be the ray from 0 to ∞ going through $\boldsymbol{U}$. Imagine centering a unit cube around each point in $\mathbb{Z}^3$ whose edges are aligned with the axes. Then γ traverses an infinite number of such cubes. Let Γ be the corresponding self-avoiding path in the lattice $\mathbb{L}^3$. To see that this procedure indeed produces a path observe that γ, upon exiting a cube around a point $z \in \mathbb{Z}^3$, enters the cube of a neighboring point $z' \in \mathbb{Z}^3$ through a face corresponding to the edge between z and z' on the lattice $\mathbb{L}^3$ (unless it goes through a corner of the cube, but this has probability 0). To argue that μ distributes its mass among sufficiently spread out paths, we bound the probability that a vertex is visited by Γ. Let z be an arbitrary vertex in $\mathbb{Z}^3$. Because the sphere of radius $\|z\|_2$ around the origin in $\mathbb{R}^3$ has area $O(\|z\|_2^2)$ and its intersection with the unit cube centered around z has area $O(1)$, it follows that

$$\mathbb{P}[z \in \Gamma] = O\left(1/\|z\|_2^2\right).$$

That immediately implies a similar bound on the probability that an edge is visited by Γ. Moreover:

Lemma 3.48. *There are $O(j^2)$ edges with an endpoint at ℓ^2 distance within $[j, j+1]$ from the origin.*

Proof. Consider a ball of ℓ^2 radius $1/2$ centered around each vertex of ℓ^2 norm within $[j, j+1]$. These balls are non-intersecting and have total volume $\Omega(N_j)$ where N_j is the number of such vertices. On the other hand, the volume of the shell of ℓ^2 inner and outer radii $j - 1/2$ and $j + 3/2$ centered around the origin is

$$\frac{4}{3}\pi(j+3/2)^3 - \frac{4}{3}\pi(j-1/2)^3 = O(j^2),$$

hence $N_j = O(j^2)$. Finally note that each vertex has 6 incident edges. ■

Plugging these bounds into (3.34), we get

$$\mathscr{R}(0 \leftrightarrow \infty) \leq \sum_j O(j^2) \cdot \left[O(1/j^2)\right]^2 = O\left(\textstyle\sum_j j^{-2}\right) < +\infty.$$

Transience follows from Theorem 3.44. (This argument clearly does not work on $\mathbb{L}$ where there are only two rays. You should convince yourself that it does not work on $\mathbb{L}^2$ either. But see Exercise 3.7.) ■

Finally we derive a useful general result illustrating the robustness reaped from Thomson's principle. At a high level, a rough embedding from $\mathcal{N}$ to $\mathcal{N}'$ is a mapping of the edges of $\mathcal{N}$ to paths of $\mathcal{N}'$ of comparable overall resistance that do not overlap much. The formal definition follows. As we will see, the purpose of a rough embedding is to allow a flow on $\mathcal{N}$ to be morphed into a flow on $\mathcal{N}'$ of comparable energy.

Definition 3.49 (Rough embedding)**.** *Let $\mathcal{N} = (G, c)$ and $\mathcal{N}' = (G', c')$ be networks with resistances r and r' respectively. We say that a map ϕ from the vertices of G to the vertices of G' is a* rough embedding *if there are constants $\alpha, \beta < +\infty$ and a map Φ defined on the edges of G such that:*

rough embedding

1. *for every edge $e = \{x, y\}$ in G, $\Phi(e)$ is a non-empty, self-avoiding path of edges of G' between $\phi(x)$ and $\phi(y)$ such that*

$$\sum_{e' \in \Phi(e)} r'(e') \leq \alpha\, r(e),$$

2. *for every edge e' in G', there are no more than β edges in G whose image under Φ contains e'.*

(The map ϕ need not be a bijection.)

We say that two networks are *roughly equivalent* if there exist rough embeddings between them, one in each direction.

rough equivalence

Example 3.50 (Independent coordinates walk)**.** Let $\mathcal{N} = \mathbb{L}^d$ with unit resistances and let $\mathcal{N}'$ be the network on the subset of $\mathbb{Z}^d$ corresponding to $(Y_t^{(1)}, \ldots, Y_t^{(d)})$, where the $(Y_t^{(i)})$s are independent simple random walks on $\mathbb{Z}$ started at 0. Note that $\mathcal{N}'$ contains only those points of $\mathbb{Z}^d$ with coordinates of identical parities. We claim that the networks $\mathcal{N}$ and $\mathcal{N}'$ are roughly equivalent.

- *$\mathcal{N}$ to $\mathcal{N}'$:* Consider the map ϕ which associates to each $x \in \mathcal{N}$ a closest point in $\mathcal{N}'$ chosen in some arbitrary manner. For Φ, associate to each edge $e = \{x, y\} \in \mathcal{N}$ a shortest path in $\mathcal{N}'$ between $\phi(x)$ and $\phi(y)$, again chosen arbitrarily. If $\phi(x) = \phi(y)$, choose an arbitrary, non-empty, shortest cycle through $\phi(x)$.

- *$\mathcal{N}'$ to $\mathcal{N}$:* Consider the map ϕ which associates to each $x \in \mathcal{N}'$ the corresponding point x in $\mathcal{N}$. Construct Φ similarly to the previous case.

◀

See Exercise 3.9 for an important generalization of the previous example. Our main result about roughly equivalent networks is that they have the same type.

Theorem 3.51 (Recurrence and rough equivalence)**.** *Let $\mathcal{N}$ and $\mathcal{N}'$ be roughly equivalent, locally finite, connected networks. Then $\mathcal{N}$ is transient if and only if $\mathcal{N}'$ is transient.*

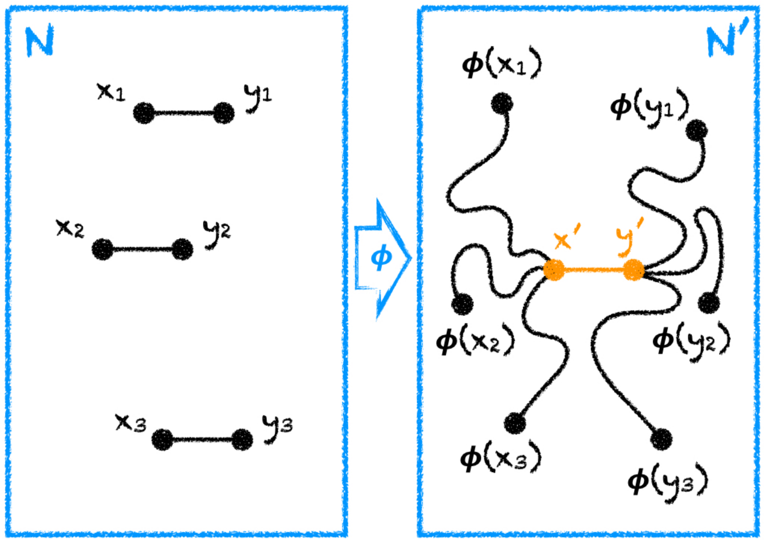

Figure 3.5: The flow on $\langle x', y' \rangle$ is the sum of the flows on $\langle x_1, y_1 \rangle$, $\langle x_2, y_2 \rangle$, and $\langle x_3, y_3 \rangle$.

Proof. Assume $\mathcal{N}$ is transient and let θ be a unit flow from some a to ∞ of finite energy. The existence of this flow is guaranteed by Theorem 3.44. Let ϕ, Φ be a rough embedding with parameters α and β.

The basic idea of the proof is to map the flow θ onto $\mathcal{N}'$ using Φ. Because flows are directional, it will be convenient to think of edges as being directed. Recall that $\langle x, y \rangle$ denotes the directed edge from x to y. For $e = \{x, y\}$ in $\mathcal{N}$, we write $\langle x', y' \rangle \in \overrightarrow{\Phi}(x, y)$ to mean that $\{x', y'\} \in \Phi(e)$ and that x' is visited before y' in the path $\Phi(e)$ from $\phi(x)$ to $\phi(y)$. (If $\phi(x) = \phi(y)$, choose an arbitrary orientation of the cycle $\Phi(e)$ for $\overrightarrow{\Phi}(x, y)$ and the reversed orientation for $\overrightarrow{\Phi}(y, x)$.) Then define, for x', y' with $\{x', y'\}$ in $\mathcal{N}'$,

$$\theta'(x', y') := \sum_{\langle x, y \rangle : \langle x', y' \rangle \in \overrightarrow{\Phi}(x, y)} \theta(x, y). \tag{3.35}$$

See Figure 3.5.

We claim that θ' is a flow to ∞ of finite energy on $\mathcal{N}'$. We first check that θ' is a flow.

1. *(Anti-symmetry)* By construction, $\theta'(y', x') = -\theta'(x', y')$, i.e., θ' is antisymmetric, because θ itself is anti-symmetric.

2. *(Flow conservation)* Next we check the flow-conservation constraints. Fix z' in $\mathcal{N}'$. By Condition 2 in Definition 3.49, there are finitely many edges e in $\mathcal{N}$ such $\Phi(e)$ visits z'. Let $e = \{x, y\}$ be such an edge. There are two cases:

   - Assume first that $\phi(x), \phi(y) \neq z'$ and let $\langle u', z'\rangle$, $\langle z', w'\rangle$ be the directed edges incident with z' on the path $\Phi(e)$ oriented from $\phi(x)$ to $\phi(y)$. Observe that, in the definition of θ', $\langle y, x\rangle$ contributes $\theta(y, x) = -\theta(x, y)$ to $\theta'(z', u')$ and $\langle x, y\rangle$ contributes $\theta(x, y)$ to $\theta'(z', w')$. So these contributions cancel out in the flow-conservation constraint for z', i.e., in the sum $\sum_{v' \sim z'} \theta'(z', v')$.
   - If instead $e = \{x, y\}$ is such that $\phi(x) = z'$, let $\langle z', w'\rangle$ be the first edge on the path $\Phi(e)$ from $\phi(x)$ to $\phi(y)$. Edge $\langle x, y\rangle$ contributes $\theta(x, y)$ to $\theta'(z', w')$. (A similar statement applies to $\phi(y) = z'$.)

   From the two cases above, summing over all paths visiting z' gives

$$\sum_{v' : v' \sim z'} \theta'(z', v') = \sum_{z : \phi(z) = z'} \left(\sum_{v \sim z} \theta(z, v) \right).$$

   Because θ is a flow, the sum on the r.h.s. is 0 unless $a \in \phi^{-1}(\{z'\})$ in which case it is 1. We have shown that θ' is a unit flow from $\phi(a)$ to ∞.

It remains to bound the energy of θ'. By (3.35), Cauchy-Schwarz, and Condition 2 in Definition 3.49,

$$\begin{aligned} \theta'(x', y')^2 &= \left[\sum_{\langle x,y\rangle : \langle x', y'\rangle \in \overrightarrow{\Phi}(x,y)} \theta(x, y) \right]^2 \\ &\leq \left[\sum_{\langle x,y\rangle : \langle x', y'\rangle \in \overrightarrow{\Phi}(x,y)} 1 \right] \left[\sum_{\langle x,y\rangle : \langle x', y'\rangle \in \overrightarrow{\Phi}(x,y)} \theta(x, y)^2 \right] \\ &\leq \beta \sum_{\langle x,y\rangle : \langle x', y'\rangle \in \overrightarrow{\Phi}(x,y)} \theta(x, y)^2. \end{aligned}$$

Summing over all pairs and using Condition 1 in Definition 3.49 gives

$$\begin{aligned}\frac{1}{2}\sum_{x',y'} r'(x',y')\theta'(x',y')^2 &\leq \beta\frac{1}{2}\sum_{x',y'} r'(x',y') \sum_{\langle x,y\rangle:\langle x',y'\rangle\in\overrightarrow{\Phi}(x,y)} \theta(x,y)^2 \\ &= \beta\frac{1}{2}\sum_{x,y}\theta(x,y)^2 \sum_{\langle x',y'\rangle\in\overrightarrow{\Phi}(x,y)} r'(x',y') \\ &\leq \alpha\beta\frac{1}{2}\sum_{x,y} r(x,y)\theta(x,y)^2,\end{aligned}$$

which is finite by assumption. That concludes the proof. ■

Example 3.52 (Independent coordinates walk (continued))**.** Consider again the networks $\mathcal{N}$ and $\mathcal{N}'$ in Example 3.50. Because they are roughly equivalent, they have the same type. This leads to yet another proof of Pólya's theorem. Recall that, because the number of returns to 0 is geometric with success probability equal to the escape probability, random walk on $\mathcal{N}'$ is recurrent if and only if the expected number of visits to 0 is finite. By independence of the coordinates, this expectation can be written as

$$\sum_{t\geq 0}\left(\mathbb{P}_0\left[Y_{2t}^{(1)}=0\right]\right)^d = \sum_{t\geq 0}\left(\binom{2t}{t}2^{-2t}\right)^d = \sum_{t\geq 0}\Theta(t^{-d/2}),$$

where we used Stirling's formula. The r.h.s. is finite if and only if $d \geq 3$. That implies random walk on $\mathcal{N}'$ is transient under the same condition. By rough equivalence, the same is true of $\mathcal{N}$. ◀

Other applications So far we have emphasized applications to recurrence. Here we show that electrical network theory can also be used to bound certain hitting times. In Sections 3.3.5 and 3.3.6, we give further applications beyond random walks on graphs.

An application of Lemma **??** gives another probabilistic interpretation of the effective resistance—and a useful formula.

Theorem 3.53 (Commute time identity)**.** *Let $\mathcal{N} = (G, c)$ be a finite, connected network with vertex set V. For $x \neq y$, let the* commute time *$\tau_{x,y}$ be the time of the first return to x after the first visit to y. Then* commute time

$$\mathbb{E}_x[\tau_{x,y}] = \mathbb{E}_x[\tau_y] + \mathbb{E}_y[\tau_x] = c_{\mathcal{N}}\mathscr{R}(x \leftrightarrow y),$$

where $c_{\mathcal{N}} = 2\sum_{e=\{x,y\}\in\mathcal{N}} c(e)$.

Proof. This follows immediately from Lemma **??** and the definition of the effective resistance. Specifically,

$$\begin{aligned}
\mathbb{E}_x[\tau_y] + \mathbb{E}_y[\tau_x] &= \frac{1}{\pi_x(\mathbb{P}_x[\tau_y < \tau_x^+])} \\
&= \frac{1}{\{c(x)/(2\sum_{e=\{x,y\}\in\mathcal{N}} c(e))\}\mathbb{P}_x[\tau_y < \tau_x^+]} \\
&= c_{\mathcal{N}}\, \mathscr{R}(x \leftrightarrow y).
\end{aligned}$$

■

Example 3.54 (Random walk on the torus)**.** Consider random walk on the d-dimensional torus $\mathbb{L}_n^d$ with unit resistances. We use the commute time identity to lower bound the mean hitting time $\mathbb{E}_x[\tau_y]$ for arbitrary vertices $x \neq y$ at graph distance k on $\mathbb{L}_n^d$. To use Theorem 3.53, note that by symmetry $\mathbb{E}_x[\tau_y] = \mathbb{E}_y[\tau_x]$ so that

$$\mathbb{E}_x[\tau_y] = \frac{1}{2}\, c_{\mathcal{N}}\, \mathscr{R}(x \leftrightarrow y) = n^d\, \mathscr{R}(x \leftrightarrow y). \tag{3.36}$$

To simplify, assume n is odd and identify the vertices of $\mathbb{L}_n^d$ with the box

$$B := \{-(n-1)/2, \ldots, (n-1)/2\}^d,$$

in $\mathbb{L}^d$ centered at $x = 0$. The rest of the argument is essentially identical to the first half of the proof of Theorem 3.46. Let $\partial B_j^\infty = \{z \in \mathbb{L}^d \,:\, \|z\|_\infty = j\}$ and let Π_j be the set of edges between ∂B_j^∞ and ∂B_{j+1}^∞. Note that on B the ℓ^1 norm of y is at most k. Since the ℓ^∞ norm is at least $1/d$ times the ℓ^1 norm on $\mathbb{L}^d$, there exists $J = O(k)$ such that all Π_js, $j \leq J$, are cutsets separating x from y. By the Nash-Williams inequality

$$\mathscr{R}(x \leftrightarrow y) \geq \sum_{0 \leq j \leq J} |\Pi_j|^{-1} = \sum_{0 \leq j \leq J} \Omega\left(j^{-(d-1)}\right) = \begin{cases} \Omega(\log k), & d = 2 \\ \Omega(1), & d \geq 3. \end{cases}$$

From (3.36), we get:

Claim 3.55.

$$\mathbb{E}_x[\tau_y] = \begin{cases} \Omega(n^d \log k), & d = 2 \\ \Omega(n^d), & d \geq 3. \end{cases}$$

◀

Remark 3.56. *The bounds in the previous example are tight up to constants. See [LPW06, Proposition 10.13].*

3.3.4 ▷ *Random walk in a random environment: supercritical percolation clusters*

In this section, we apply the random paths approach to random walk on percolation clusters.[‡]

To be written. See [LP, Section 5.5].

3.3.5 ▷ *Uniform spanning trees: Wilson's method*

In this section and the next one, we describe applications of electrical network theory to topics seemingly unrelated to random walks on networks, namely uniform spanning trees and Ising models on trees.

Uniform spanning trees Let $G = (V, E)$ be a finite connected graph. Recall that a spanning tree is a subtree of G containing all its vertices. A *uniform spanning tree* is a spanning tree T chosen uniformly at random among all spanning trees of G. (The reader interested only in Wilson's method for generating uniform spanning trees may jump ahead to the second half of this section.) *uniform spanning tree*

A fundamental property of uniform spanning trees is the following negative correlation between edges.

Claim 3.57.

$$\mathbb{P}[e \in T \,|\, e' \in T] \leq \mathbb{P}[e \in T], \qquad \forall e \neq e' \in G.$$

This property is perhaps not surprising. For one, the number of edges in a spanning tree is fixed, so the inclusion of e' makes it seemingly less likely for other edges to be present. Yet proving Claim 3.57 is non-trivial. The only known proof relies on the electrical network perspective developed in this section. The key to the proof is a remarkable formula for the inclusion of an edge in a uniform spanning tree.

Theorem 3.58 (Kirchhoff's resistance formula)**.** *Let $G = (V, E)$ be a finite, connected graph and let $\mathcal{N}$ be the network on G with unit resistances. If T is a uniform spanning tree on G, then for all $e = \{x, y\}$*

$$\mathbb{P}[e \in T] = \mathscr{R}(x \leftrightarrow y).$$

Before explaining how this formula arises, we show that it implies Claim 3.57.

[‡]Requires: Section 2.2.

Proof of Claim 3.57. By Bayes' rule and a short calculation, we can instead prove

$$\mathbb{P}[e \in T \mid e' \notin T] \geq \mathbb{P}[e \in T], \tag{3.37}$$

unless $\mathbb{P}[e' \in T] \in \{0, 1\}$ or $\mathbb{P}[e \in T] \in \{0, 1\}$ in which case the claim is vacuous. (In fact these probabilities cannot be 0. Why? Can they be equal to 1?) Picking a uniform spanning tree on $\mathcal{N}$ conditioned on $\{e' \notin T\}$ is the same as picking a uniform spanning tree on the modified network $\mathcal{N}'$ where e' is removed. By Rayleigh's principle,

$$\mathscr{R}_{\mathcal{N}'}(x \leftrightarrow y) \geq \mathscr{R}_{\mathcal{N}}(x \leftrightarrow y),$$

and Kirchhoff's resistance formula gives (3.37). ∎

Remark 3.59. *More generally, thinking of a uniform spanning tree T as a random subset of edges, the law of T has the property of negative associations, defined as follows. An event $\mathcal{A} \subseteq 2^E$ is said to be* increasing *if $\omega \cup \{e\} \in \mathcal{A}$ whenever $\omega \in \mathcal{A}$. The event $\mathcal{A}$ is said to* depend only on *$F \subseteq E$ if for all $\omega_1, \omega_2 \in 2^E$ that agree on F, either both are in $\mathcal{A}$ or neither is. The law, $\mathbb{P}_T$, of T has* negative associations *in the sense that for any two increasing events $\mathcal{A}$ and $\mathcal{B}$ that depend only on disjoint sets of edges, we have $\mathbb{P}_T[\mathcal{A} \cap \mathcal{B}] \leq \mathbb{P}_T[\mathcal{A}]\mathbb{P}_T[\mathcal{B}]$, i.e., $\mathcal{A}$ and $\mathcal{B}$ are* negatively correlated. *See [LP, Exercise 4.6]. (To see why the events considered depend on* disjoint edges*, see for instance what happens when $\mathcal{A} \subseteq \mathcal{B}$.)*

Let $e = \{x, y\}$. To get some insight into Kirchhoff's resistance formula, we first note that, if i is the unit current from x to y and v is the associated voltage, by definition of the effective resistance

$$\mathscr{R}(x \leftrightarrow y) = \frac{v(x)}{\|i\|} = c(e)(v(x) - v(y)) = i(x, y), \tag{3.38}$$

where we used Ohm's law as well as the fact that $c(e) = 1$, $v(y) = 0$, and $\|i\| = 1$. Note the difference between $\|i\|$ and $i(x, y)$. Although $\|i\| = 1$, $i(x, y)$ is only the current along the edge to y. Furthermore by the probabilistic interpretation of the current, with $Z = \{y\}$,

$$i(x, y) = \mathbb{E}_x[N^Z_{x \to y} - N^Z_{y \to x}] = \mathbb{P}_x\left[\langle x, y\rangle \text{ is traversed before } \tau_y\right]. \tag{3.39}$$

Indeed, started at x, $N^Z_{y \to x} = 0$ and $N^Z_{x \to y} \in \{0, 1\}$. Kirchhoff's resistance formula is then established by relating the random walk on $\mathcal{N}$ to the probability that e is present in a uniform spanning tree T. To do this we introduce a random-walk-based algorithm for generating uniform spanning trees. This rather miraculous procedure, known as *Wilson's method*, is of independent interest. For a classical connection between random walks and spanning trees, see Exercise 3.11.

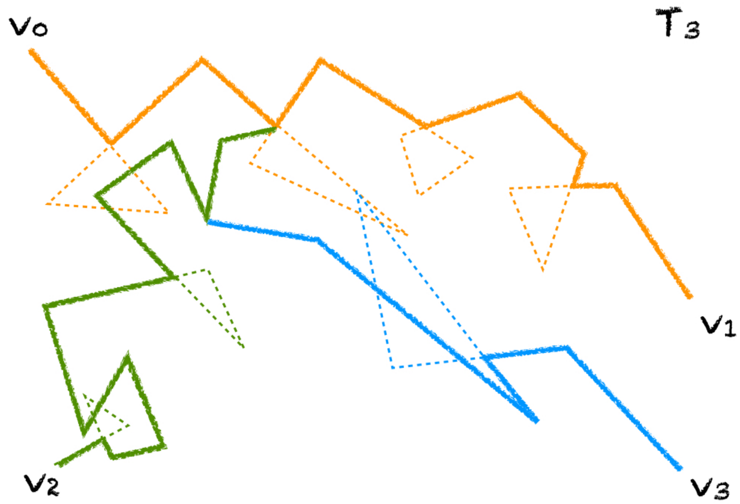

Figure 3.6: An illustration of Wilson's method. The dotted lines indicate erased loops.

Wilson's method It will be somewhat more transparent to work in a more general context. Let $\mathcal{N} = (G, c)$ be a finite, connected network on G with arbitrary conductances and define the *weight* of a spanning tree T on $\mathcal{N}$ as

$$W(T) = \prod_{e \in T} c(e).$$

With a slight abuse, we continue to call a tree T picked at random among all spanning trees of G with probability proportional to $W(T)$ a "uniform" spanning tree on $\mathcal{N}$.

To state Wilson's method, we need the notion of *loop erasure*. Let $\mathcal{P} = x_0 \sim \ldots \sim x_k$ be a path in $\mathcal{N}$. The loop erasure of $\mathcal{P}$ is obtained by removing cycles in the order they appear. That is, let j^* be the smallest j such that $x_j = x_\ell$ for some $\ell < j$. Remove the subpath $x_{\ell+1} \sim \cdots \sim x_j$ from $\mathcal{P}$, and repeat. The resulting self-avoiding path is denoted by $\mathrm{LE}(\mathcal{P})$. *loop erasure*

Let ρ be an arbitrary vertex of G, which refer to as the root, and let T_0 be the subtree made up of ρ alone. Order arbitrarily the vertices $v_0, \ldots, v_{n-1}$ of G, starting with the root $v_0 := \rho$. Wilson's method constructs an increasing sequence of subtrees as follows. See Figure 3.6. Let $T := T_0$.

1. Let v be the vertex of G not in T with lowest index. Perform random walk

on $\mathcal{N}$ started at v until the first visit to a vertex of T. Let $\mathcal{P}$ be the resulting path.

2. Add the loop erasure $\mathrm{LE}(\mathcal{P})$ to T.

3. Repeat until all vertices of G are in T.

Let $T_0, \ldots, T_m$ be the sequence of subtrees produced by Wilson's method.

Claim 3.60. *Forgetting the root, T_m is a uniform spanning tree on $\mathcal{N}$.*

This claim is far from obvious. Before proving it, we finish the proof of Kirchhoff's resistance formula.

Proof of Theorem 3.58. From (3.38) and (3.39), it suffices to prove that, for $e = \{x, y\}$,

$$\mathbb{P}_x [\langle x, y\rangle \text{ is traversed before } \tau_y] = \mathbb{P}[e \in T],$$

where the probability on the l.h.s. refers to random walk on $\mathcal{N}$ started at x and the probability on the r.h.s. refers to a uniform spanning tree T on $\mathcal{N}$. Generate T using Wilson's method started at root $\rho = y$ with the choice $v_1 = x$. If the sample path from x to y during the first iteration of Wilson's method includes $\langle x, y\rangle$, then the loop erasure is simply $x \sim y$ and e is in T. On the other hand, if the sample path from x to y does not include $\langle x, y\rangle$, then e cannot be used at a later stage because it would create a cycle. That immediately proves the claim. ■

It remains to prove the claim.

Proof of Claim 3.60. The idea of the proof is to cast Wilson's method in the more general framework of cycle popping algorithms. We begin by explaining how such algorithms work.

Let P be the transition matrix corresponding to random walk on $\mathcal{N} = (G, c)$ with $G = (V, E)$. To each vertex $x \neq \rho$ in V, we assign an independent stack of directed edges

$$\mathcal{S}_0^x := (\langle x, Y_1^x\rangle_1, \langle x, Y_2^x\rangle_2, \ldots)$$

where each Y_j^x is chosen independently at random from the distribution $P(x, \cdot)$. In particular all Y_j^xs are neighbors of x in $\mathcal{N}$. The index j in $\langle x, Y_j^x\rangle_j$ is usually referred to as the *color* of the edge. It keeps track of the position of the edge in the original stack. (Picture $\mathcal{S}^x$ as a spring-loaded plate dispenser located on vertex x.) *color*

We consider a process which involves popping edges off the stacks. We use the notation $\mathcal{S}^x$ to denote the *current* stack at x. The initial assignment of the stack is $\mathcal{S}^x := \mathcal{S}_0^x$ as above. Given the current stacks $(\mathcal{S}^x)$, we call *visible graph* the directed graph over V with edges $\mathrm{Top}(\mathcal{S}^x)$ for all $x \neq \rho$, where $\mathrm{Top}(\mathcal{S}^x)$ is *visible graph* the first edge in the current stack $\mathcal{S}^x$. The latter are referred to as *visible edges*. We denote the current visible graph by $\overrightarrow{G}_\odot$. Note that $\overrightarrow{G}_\odot$ has out-degree 1 for all $x \neq \rho$ and the root has out-degree 0. In particular all (undirected) cycles in $\overrightarrow{G}_\odot$ are in fact directed cycles. Indeed, a set of edges forming a cycle that is not directed must have a vertex of out-degree 2. Recall the following characterization (see Lemma **??**): a cycle-free undirected graph with n vertices and $n-1$ edges is a spanning tree. Hence, if there is no cycle in $\overrightarrow{G}_\odot$ then it must be a spanning tree where, furthermore, all edges point towards the root. Such a tree is also known as a *spanning arborescence*.

visible edge

spanning arborescence

As the name suggests, a cycle popping algorithm proceeds by popping cycles in $\overrightarrow{G}_\odot$ off the tops of the stacks until a spanning arborescence is produced. That is, at every iteration, if $\overrightarrow{G}_\odot$ contains at least one cycle, then a cycle $\overrightarrow{C}$ is picked according to some rule, the top of each stack in $\overrightarrow{C}$ is popped, and a new visible graph $\overrightarrow{G}_\odot$ is revealed. See Figure 3.7 for an illustration.

With these definitions in place, the proof of the claim involves the following steps.

1. *Wilson's method is a cycle popping algorithm.* We can think of the initial stacks $(\mathcal{S}_0^x)$ as corresponding to picking—ahead of time—all potential transitions in the random walks used by Wilson's method. With this representation, Wilson's method boils down to a recipe for choosing which cycle to pop next. Indeed, at each iteration, we start from a vertex v not in the current tree T. Following the visible edges from v traces a path whose distribution is that of random walk on $\mathcal{N}$. Loop erasure then corresponds to popping cycles. We pop only those visible edges on the removed cycles as they originate from vertices that will be visited again by the algorithm and for which a new transition will then be needed. Those visible edges in the remaining loop-erased path are not popped—they are part of the final arborescence.

2. *The popping order does not matter.* We just argued that Wilson's method is a cycle popping algorithm. In fact we claim that any cycle popping algorithm, i.e., no matter what popping choices are made along the way, produces the same final arborescence. To make this precise, we identify the popped cycles uniquely. This is where the colors come in. A *colored cycle* is a directed cycle over V made of colored edges from the stacks (not necessarily of the same color and not necessarily in the current visible graph). We say that a colored cycle $\overrightarrow{C}$ is *poppable* for a visible graph $\overrightarrow{G}_\odot$ if there exists a sequence of colored cycles $\overrightarrow{C}_1, \ldots, \overrightarrow{C}_r = \overrightarrow{C}$ that can be popped in that order starting from $\overrightarrow{G}_\odot$. Note that, by this definition, $\overrightarrow{C}_1$ is a directed cycle in $\overrightarrow{G}_\odot$. Now

   colored cycle

   poppable cycle

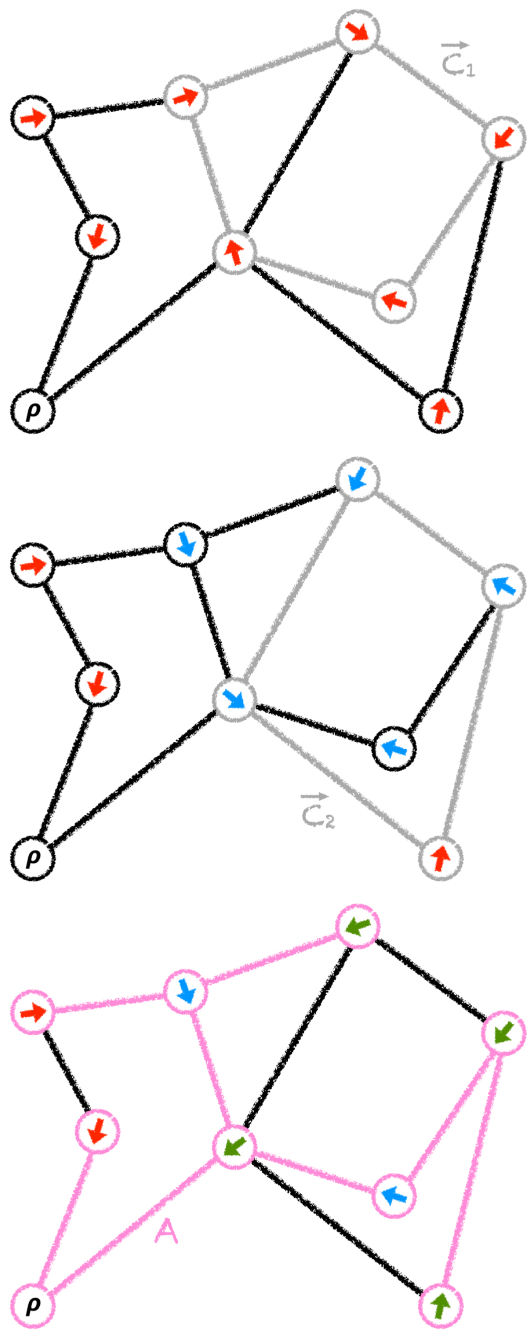

Figure 3.7: A realization of a cycle popping algorithm (from top to bottom). In all three figures, the underlying graph is G while the arrows depict the visible edges.

we claim that if $\overrightarrow{C}'_1$ were popped first instead of $\overrightarrow{C}_1$, producing the new visible graph $\overrightarrow{G}'_\odot$, then $\overrightarrow{C}$ would still be poppable for $\overrightarrow{G}'_\odot$. This claim implies that, in any cycle popping algorithm, either an infinite number of cycles are popped or eventually all poppable cycles are popped—independently of the order—producing the same outcome. To prove the claim, note first that if $\overrightarrow{C}'_1 = \overrightarrow{C}$ or if $\overrightarrow{C}'_1$ does not share a vertex with any of $\overrightarrow{C}_1, \ldots, \overrightarrow{C}_r$ there is nothing to prove. So let $\overrightarrow{C}_j$ be the first cycle in the sequence sharing a vertex with $\overrightarrow{C}'_1$, say x. Let $\langle x, y\rangle_c$ and $\langle x, y'\rangle_{c'}$ be the colored edges emanating from x in $\overrightarrow{C}_j$ and $\overrightarrow{C}'_1$ respectively. By definition, x is not on any of $\overrightarrow{C}_1, \ldots, \overrightarrow{C}_{j-1}$ so the edge originating from x is not popped by that sequence and we must have $\langle x, y\rangle_c = \langle x, y'\rangle_{c'}$ as colored edges. In particular, the vertex y is also a shared vertex of $\overrightarrow{C}_j$ and $\overrightarrow{C}'_1$, and the same argument applies to it. Proceeding by induction leads to the conclusion that $\overrightarrow{C}'_1 = \overrightarrow{C}_j$ as colored cycles. But then $\overrightarrow{C}$ is clearly poppable for the visible graph resulting from popping $\overrightarrow{C}'_1$ first, because it can be popped with the rearranged sequence $\overrightarrow{C}'_1 = \overrightarrow{C}_j, \overrightarrow{C}_1, \ldots, \overrightarrow{C}_{j-1}, \overrightarrow{C}_{j+1}, \ldots, \overrightarrow{C}_r = \overrightarrow{C}$, where we used the fact that $\overrightarrow{C}'_1$ does not share a vertex with $\overrightarrow{C}_1, \ldots, \overrightarrow{C}_{j-1}$.

3. *Termination occurs in finite time almost surely.* We have shown so far that, in any cycle popping algorithm, either an infinite number of cycles are popped or eventually all poppable cycles are popped. But Wilson's method—a cycle popping algorithm as we have shown—stops after a finite amount of time with probability 1. Indeed, because the network is finite and connected, the random walk started at each iteration hits the current T in finite time almost surely (by Lemma **??**). To sum up, all cycle popping algorithms terminate and produce the same spanning arborescence. It remains to compute the distribution of the outcome.

4. *The arborescence has the desired distribution.* Let $\mathcal{A}$ be the spanning arborescence produced by any cycle popping algorithm on the stacks $(\mathcal{S}_0^x)$. To compute the distribution of $\mathcal{A}$, we first compute the distribution of a particular cycle popping realization leading to $\mathcal{A}$. Because the popping order does not matter, by "realization" we mean a collection $\mathcal{C}$ of colored cycles together with a final spanning arborescence $\mathcal{A}$. Notice that what lies in the stacks *under* $\mathcal{A}$ is not relevant to the realization, i.e., the same outcome is produced no matter what is under $\mathcal{A}$. So, from the distribution of the stacks, the probability of observing $(\mathcal{C}, \mathcal{A})$ is simply the product of the transitions

corresponding to the directed edges in $\mathcal{C}$ and $\mathcal{A}$, i.e.,

$$\prod_{\vec{e} \in \mathcal{C} \cup \mathcal{A}} P(\vec{e}) = \Psi(\mathcal{A}) \prod_{\vec{C} \in \mathcal{C}} \Psi\left(\vec{C}\right),$$

where the function Ψ returns the product of the transition probabilities of a set of directed edges. Thanks to the product form on the r.h.s., summing over all possible $\mathcal{C}$s gives that the probability of producing $\mathcal{A}$ is proportional to $\Psi(\mathcal{A})$. For this argument to work though, there are two small details to take care of. First, note that we want the probability of the *uncolored* arborescence. But observe that, in fact, there is no need to keep track of the colors on the edges of $\mathcal{A}$ because these are determined by $\mathcal{C}$. Secondly, we need for the collection of possible $\mathcal{C}$s *not to vary with* $\mathcal{A}$. But it is clear that any arborescence could lie under any $\mathcal{C}$.

To see that we are done, let T be the undirected spanning tree corresponding to the outcome, $\mathcal{A}$, of Wilson's method. Then, because $P(x, y) = \frac{c(x,y)}{c(x)}$, we get

$$\Psi(\mathcal{A}) = \frac{W(T)}{\prod_{x \neq \rho} c(x)},$$

where note that the denominator does not depend on T. So if we forget the orientation of $\mathcal{A}$, which is determined by the root, we get a spanning tree whose distribution is proportional to $W(T)$, as required. ■

3.3.6 ▷ *Ising model on trees: the reconstruction problem*

To be written. See [Per99, Section 16].

Exercises

Exercise 3.1 (Azuma-Hoeffding: a second proof)**.** This exercise leads the reader through an alternative proof of the Azuma-Hoeffding inequality.

a) Show that for all $x \in [-1, 1]$ and $a > 0$

$$e^{ax} \leq \cosh a + x \sinh a.$$

b) Use a Taylor expansion to show that for all x

$$\cosh x \leq e^{x^2/2}.$$

c) Let $X_1, \ldots, X_n$ be (not necessarily independent) random variables such that, for all i, $|X_i| \leq c_i$ for some constant $c_i < +\infty$ and

$$\mathbb{E}\left[X_{i_1} \cdots X_{i_k}\right] = 0, \qquad \forall\, 1 \leq k \leq n,\ \forall\, 1 \leq i_1 < \cdots < i_k \leq n. \tag{3.40}$$

Show, using a) and b), that for all $b > 0$

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geq b\right] \leq \exp\left(-\frac{b^2}{2\sum_{i=1}^{n} c_i^2}\right).$$

d) Prove that c) implies the Azuma-Hoeffding inequality as stated in Theorem 3.1.

e) Show that the random variables in Exercise 2.5 do not satisfy (3.40) (without using the claim in part b) of that exercise).

Exercise 3.2 (Kirchhoff's laws)**.** Consider a finite, connected network with a source and a sink. Show that an anti-symmetric function on the edges satisfying Kirchhoff's two laws is a current function (i.e., it corresponds to a voltage function through Ohm's law).

Exercise 3.3 (Dirichlet problem: non-uniqueness)**.** Let (X_t) be the birth-and-death chain on $\mathbb{Z}_+$ with $P(x, x+1) = p$ and $P(x, x-1) = 1 - p$ for all $x \geq 1$, and $P(0, 1) = 1$, for some $0 < p < 1$. Fix $h(0) = 1$.

a) When $p > 1/2$, show that there is more than one bounded extension of h to $\mathbb{Z}_+\backslash\{0\}$ that is harmonic on $\mathbb{Z}_+\backslash\{0\}$. [Hint: Consider $\mathbb{P}_x[\tau_0 = +\infty]$.]

b) When $p \leq 1/2$, show that there exists a unique bounded extension of h to $\mathbb{Z}_+\backslash\{0\}$ that is harmonic on $\mathbb{Z}_+\backslash\{0\}$.

Exercise 3.4 (Maximum principle)**.** Let $\mathcal{N} = (G, c)$ be a finite or countable, connected network with $G = (V, E)$. Let W be a finite, connected, proper subset of V.

a) Let $h : V \to \mathbb{R}$ be a function on V. Prove the maximum principle: if h is harmonic on W, i.e., it satisfies

$$h(x) = \frac{1}{c(x)} \sum_{y \sim x} c(x, y) h(y), \qquad \forall x \in W,$$

and if h achieves its supremum on W, then h is constant on $W \cup \partial_{\mathcal{V}} W$, where

$$\partial_{\mathcal{V}} W = \{z \in V \backslash W \,:\, \exists y \in W, y \sim z\}.$$

b) Let $h : W^c \to \mathbb{R}$ be a bounded function on $W^c := V \backslash W$. Let h_1 and h_2 be extensions of h to W that are harmonic on W. Use part a) to prove that $h_1 \equiv h_2$.

Exercise 3.5 (Effective resistance: metric)**.** Show that effective resistances between pairs of vertices form a metric.

Exercise 3.6 (Dirichlet principle: proof)**.** Prove Theorem 3.38.

Exercise 3.7 (Random walk on $\mathbb{L}^2$: effective resistance)**.** Consider random walk on $\mathbb{L}^2$, which we showed is recurrent. Let (G_n) be the exhaustive sequence corresponding to vertices at distance at most n from the origin and let Z_n be the corresponding sink-set. Show that $\mathscr{R}(0 \leftrightarrow Z_n) = \Theta(\log n)$. [Hint: Use the Nash-Williams inequality and the method of random paths.]

Exercise 3.8 (Random walk on regular graphs: effective resistance)**.** Let G be a d-regular graph with n vertices and $d > n/2$. Let $\mathcal{N}$ be the network (G, c) with unit conductances. Let a and z be arbitrary distinct vertices.

a) Show that there are at least $2d - n$ vertices $x \neq a, z$ such that $a \sim x \sim z$ is a path.

b) Prove that

$$\mathscr{R}(a \leftrightarrow z) \leq \frac{2rn}{2r - n}.$$

Exercise 3.9 (Rough isometries)**.** Graphs $G = (V, E)$ and $G' = (V', E')$ are *roughly isometric* (or quasi-isometric) if there is a map $\phi : V \to V'$ and constants $0 < \alpha, \beta < +\infty$ such that for all $x, y \in V$ *rough isometry*

$$\alpha^{-1} d(x, y) - \beta \leq d'(\phi(x), \phi(y)) \leq \alpha d(x, y) + \beta,$$

where d and d' are the graph distances on G and G' respectively, and furthermore all vertices in G' are within distance β of the image of V. Let $\mathcal{N} = (G, c)$ and $\mathcal{N}' = (G', c')$ be countable, connected networks with uniformly bounded conductances, resistances and degrees. Prove that if G and G' are roughly isometric then $\mathcal{N}$ and $\mathcal{N}'$ are roughly equivalent. [Hint: Start by proving that being roughly isometric is an equivalence relation.]

Exercise 3.10 (Random walk on the cycle: hitting time)**.** Use the commute time identity to compute $\mathbb{E}_x[\tau_y]$ in Example 3.54 in the case $d = 1$. Give a second proof using a direct martingale argument.

Exercise 3.11 (Markov chain tree theorem)**.** Let P be the transition matrix of a finite, irreducible Markov chain with stationary distribution π. Let G be the directed graph corresponding to the positive transitions of P. For an arborescence $\mathcal{A}$ of G, define its weight as

$$\Psi(\mathcal{A}) = \prod_{\vec{e} \in \mathcal{A}} P(\vec{e}).$$

Consider the following process on spanning arborescences over G. Let ρ be the root of the current spanning arborescence $\mathcal{A}$. Pick an outgoing edge $\vec{e} = \langle \rho, x \rangle$ of ρ according to $P(\rho, \cdot)$. Add $\vec{e}$ to $\mathcal{A}$. This creates a cycle. Remove the edge of this cycle originating from x, producing a new arborescence $\mathcal{A}'$ with root x. Repeat the process.

a) Show that this chain is irreducible.

b) Show that Ψ is a stationary measure for this chain.

c) Prove the *Markov chain tree theorem*: The stationary distribution π of P is proportional to

$$\pi_x = \sum_{\mathcal{A} \,:\, \mathrm{root}(\mathcal{A}) = x} \Psi(\mathcal{A}).$$

Bibliographic remarks

Section 3.1 The textbooks [Dur10] and [Wil91] contain excellent introductions to martingales.

Section 3.2 The Azuma-Hoeffding inequality is due to Hoeffding [Hoe63] and Azuma [Azu67]. The version of the inequality in Exercise 3.1 is from [Ste97]. The method of bounded differences has its origins in the works of Yurinskii [Yur76], Maurey [Mau79], Milman and Schechtman [MS86], Rhee and Talagrand [RT87], and Shamir and Spencer [SS87]. It was popularized by McDiarmid [McD89]. Example 3.7 is taken from [MU05, Section 12.5]. The presentation in Section **??** follows [AS11, Section 7.3]. Claim 3.9 is due to Shamir and Spencer [SS87]. The 2-point concentration result alluded to in Section 3.2.3 is due to Alon and Krivelevich [AK97]. For the full story on the chromatic number of Erdös-Rényi graphs, see [JLR11, Chapter 7]. Claim 3.16 is due to Bollobás, Riordan, Spencer, and Tusnády [BRST01]. It confirmed simulations of Barabási and Albert [BA99]. The expectation was analyzed by Dorogovtsev, Mendes, and Samukhin [DMS00].

For much more on preferential attachment models see [Dur06] and [vdH14]. Section 3.2.4 borrows from [BLM13, Section 7.1] and [Pet, Section 6.3]. General references on the concentration of measure phenomenon and concentration inequalities are [Led01] and [BLM13].

Section 3.3 The material in Sections 3.3.1-3.3.5 borrows heavily from [LPW06, Chapters 9, 10], [AF, Chapters 2, 3] and, especially, [LP, Sections 2.1-2.6, 4.1-4.2, 5.5]. The classical reference on potential theory and its probabilistic counterpart is [Doo01]. For the discrete case and the electrical network point of view, the book of Doyle and Snell is excellent [DS84]. In particular the series and parallel laws are defined and illustrated. See also [LP, Section 2.3] for more examples. For an introduction to convex optimization and duality, see e.g. [BV04]. The Nash-Williams inequality is due to Nash-Williams [NW59]. The result in Example 3.45 is due to R. Lyons [Lyo90]. An elementary proof of Pólya's theorem can be found in [Dur10, Section 4.2]. The flow we used in the proof of Pólya's theorem is essentially due to T. Lyons [Lyo83]. The commute time identity was proved by Chandra, Raghavan, Ruzzo, Smolensky and Tiwari [CRR$^+$89]. Wilson's method is due to Wilson [Wil96]. A related method for generating uniform spanning trees was introduced by Aldous [Ald90] and Broder [Bro89]. A connection between loop-erased random walks and uniform spanning trees had previously been established by Pemantle [Pem91] using the Aldous-Broder method. For more on negative correlation in uniform spanning trees, see e.g. [LP, Section 4.2]. For a proof of the matrix tree theorem using Wilson's method, see [KRS]. For a discussion of the running time of Wilson's method and other spanning tree generation approaches, see [Wil96].