Chapter 6

Random Walks, Mixing Times

6.1 Basic definitions

Given a finite graph (random or not) one can define a random walk by its **transition kernel**

$$K(x,y) = 1/d(x) \qquad \text{if } y \sim x$$

where $y \sim x$ is short for y is a neighor of x and $d(x)$ is the degree of x. $K(x,y)$ is also called the **transition probability**, and is often denoted by p. We will use both notations in this book. If we let $D = \sum_x d(x)$ then $\pi(x) = d(x)/D$ is a stationary distribution that satisfies the **detailed balance condition**

$$\pi(x)K(x,y) = \frac{1_{x\sim y}}{D} = \pi(y)K(y,x).$$

Stationary distributions with this property are also called **reversible**, since when started the chain in state π the process has the same distribution going backwards in time. To explain this note that

$$P_\pi(X_0 = y|X_1 = x) = \frac{P_\pi(X_0 = y, X_1 = x)}{P_\pi(X_1 = x)} = \frac{\pi(y)K(y,x)}{\pi(x)} = \frac{\pi(x)K(x,y)}{\pi(x)} = K(x,y).$$

To avoid the problem of periodicity, we will often make the random walk **lazy**

$$\bar{K}(x,x) = 1/2 \qquad \bar{K}(x,y) = (1/2)K(x,y) \quad \text{when } x \neq y.$$

If the graph is connected then the kernel is irreducible. If lazy it is aperiodic so it converges to equilibrium, i.e. $K^n(x,y) \to \pi(y)$. Here we will be interested in how fast the convergence occurs. To measure the distance from equilibrium we can use the **total variation distance**. If μ and ν are probability distributions then

$$\|\mu - \nu\|_{TV} = \max_A |\mu(A) - \nu(A)| = \frac{1}{2}\sum_x |\mu(x) - \nu(x)|.$$

We will usually drop the subscript TV. The distance from equilibrium at time t can be defined as

$$d(t) = \sup_x \|K^t(x,\cdot) - \pi\| = \sup_\mu \|\mu K^t - \pi\|$$

where μ is a probability distribution and $\mu K^t(y) = \sum_x \mu(x) K^t(x,y)$. It is useful to also define

$$\bar{d}(t) = \sup_{x,y} \|K^t(x,\cdot) - K^t(y,\cdot)\| = \sup_{\mu,\nu} \|\mu K^t - \nu K^t\|$$

$\bar{d}$ has the useful property that it is **submultipliative**

$$\bar{d}(s+t) \le \bar{d}(s) \cdot \bar{d}(t). \tag{6.1.1}$$

It is easy to see that

$$d(t) \le \bar{d}(t) \le 2d(t). \tag{6.1.2}$$

The time to get within ϵ in total variation distance is

$$t_{mix}(\epsilon) = \min\{t : d(t) < \epsilon\}.$$

We define the **mixing time** by

$$t_{mix} = t_{mix}(1/4). \tag{6.1.3}$$

To see why we choose $\epsilon = 1/4$ note that using (6.1.2) and (6.1.1)

$$d(\ell t_{mix}) \le \bar{d}(\ell t_{mix}) \le \bar{d}(t_{mix})^\ell \le (2\epsilon)^\ell = 2^{-\ell}$$

when $\epsilon = 1/4$.

6.1.1 Bounds on convergence

Consider a Markov chain transition kernel $K(i,j)$ on $\{1, 2, \ldots n\}$ with reversible stationary distribution π_i, i.e., $\pi_i K(i,j) = \pi_j K(j,i)$. For the next result we will measure the distance from equilibrium using the **relative pointwise distance**

$$\Delta(t) = \max_{i,j} \left| \frac{K^t(i,j)}{\pi_j} - 1 \right|$$

which is larger than the total variation distance

$$\Delta(t) \ge \max_i \sum_j \left| \frac{K^t(i,j)}{\pi_j} - 1 \right| \pi_j = \max_i \sum_j |K^t(i,j) - \pi_j|.$$

Let D be a diagonal matrix with entries $\pi_1, \pi_2, \ldots \pi_n$ and $a = D^{1/2} K D^{-1/2}$. Since

$$\begin{aligned} a(i,j) &= \pi_i^{1/2} K(i,j) \pi_j^{-1/2} = \pi_i^{-1/2} \cdot \pi(i) K(i,j) \cdot \pi_j^{-1/2} \\ &= \pi_i^{-1/2} \cdot \pi(j) K(j,i) \cdot \pi_j^{-1/2} = \pi_j^{1/2} K(j,i) \pi_i^{-1/2} = a(j,i) \end{aligned}$$

matrix theory tells us that $a(i,j)$ has real eigenvalues $1 = \lambda_0 \geq \lambda_1 \geq \ldots \lambda_{n-1} \geq -1$. Let $\lambda_{max} = \max\{\lambda_1, |\lambda_{n-1}|\}$ be the eigenvalue with largest magnitude. The lazy chain has transition kernel $\bar{K} = (I+K)/2$ so all $\bar{\lambda}_i = (1+\lambda_i)/2 \geq 0$ and we do not have to worry about $|\lambda_{n-1}|$.

To explain the interest in eigenvalues near -1, note that if the chain is periodic with period two then we can find a set S that $p(x, S^c) = 1$ for $x \in S$ and $p(x,S) = 1$ for $x \in S^c$. The function that has $f(x) = 1$ for $x \in S$ and $f(x) = -1$ for $x \in S^c$ has $Pf = -f$, i.e., it is an eigenvector with eigenvalue -1. Thus $|\lambda_{n-1}|$ measure how close the chain is to having period 2. In general, Markov chains can have states with any period $d \geq 3$ but that is impossible in the reversible case.

The next result is from Sinclair and Jerrum (1989), but similar results can be found in many other places.

Theorem 6.1.1. *Let K be the transition matrix of an irreducible reversible Markov chain on $\{1, 2, \ldots n\}$ with stationary distribution π and let $\pi_{min} = \min_j \pi_j$. Then*

$$\Delta(t) \leq \frac{\lambda_{max}^t}{\pi_{min}} \qquad |K^t(i,j) - \pi_j| \leq \lambda_{max}^t \sqrt{\pi_j/\pi_i}.$$

Proof. Since a is symmetric, we can select an orthonormal basis e_m, $0 \leq m < n$ of eigenvectors of a, and a has spectral decomposition:

$$a = \sum_{m=0}^{n-1} \lambda_m e_m e_m^T.$$

The matrix $B_m = e_m e_m^T$ has $B_m^2 = B_m$, and $B_\ell B_m = 0$ if $\ell \neq m$ so

$$a^t(i,j) = \sum_{m=0}^{n-1} \lambda_m^t B_m(i,j) = \sum_{m=0}^{n-1} \lambda_m^t e_m(i) e_m(j)$$

$e_0(i) = \pi_i^{1/2}$ so

$$K^t(i,j) = (D^{-1/2} a^t D^{1/2})_{i,j} = \pi_j + \sqrt{\frac{\pi_j}{\pi_i}} \sum_{m=1}^{n-1} \lambda_m^t e_m(i) e_m(j). \tag{6.1.4}$$

From this it follows that

$$\Delta(t) = \max_{i,j} \frac{\left|\sum_{m=1}^{n-1} \lambda_m^t e_m(i) e_m(j)\right|}{\sqrt{\pi_i \pi_j}} \leq \lambda_{max}^t \frac{\max_{i,j} \sum_{m=1}^{n-1} |e_m(i)||e_m(j)|}{\pi_{min}}.$$

The Cauchy-Schwarz inequality implies

$$\sum_{m=1}^{n-1} |e_m(i)||e_m(j)| \leq \left(\sum_{m=1}^{n-1} |e_m(i)|^2 \sum_{m=1}^{n-1} |e_m(j)|^2 \right)^{1/2} \leq 1. \tag{6.1.5}$$

To see that $\sum_{m=1}^{n-1}|e_m(i)|^2 \le 1$ note that if δ_i is the vector with 1 in the ith place and 0 otherwise then expanding in the orthonormal basis $\delta_i = \sum_{m=0}^{n-1} e_m(i)e_m$, so the desired result follows by taking the L^2 norm of both sides of the equation. To get the second result combine (6.1.4) and (6.1.5). □

6.1.2 Continuous time chains

If jumps occur at rate one then there are a Poisson mean t jumps by time t so the transition probability is

$$H_t(x,y) = e^{-t}\sum_{m=0}^{\infty}\frac{t^m}{m!}K^m(x,y). \tag{6.1.6}$$

If λ_i is an eigenvalue of K then $e^{-t(1-\lambda_i)}$ is an eigenvalue of H_t. Thus there are no negative eigenvalues to worry about and we have

Theorem 6.1.2. *If $\gamma = 1-\lambda_1$ is the spectral gap of K and $\Delta(t) = \max_{i,j}|H_t(i,j)/\pi(j)-1|$ then*

$$\Delta(t) \le \frac{e^{-\gamma t}}{\pi_{min}} \qquad |H_t(i,j)-\pi_j| \le e^{-\gamma t}\sqrt{\pi_j/\pi_i}.$$

Proof. This follows immediately from (6.1.6) and Theorem 6.1.1. □

Given a reversible Markov transition kernel $K(x,y)$ we define an inner prodcut by

$$\mathcal{E}(f,g) = \frac{1}{2}\sum_{x,y}(f(x)-f(y))(g(x)-g(y))\pi(x)K(x,y)$$

and the **Dirichlet form** by

$$\mathcal{E}(f) = \frac{1}{2}\sum_{x,y}(f(x)-f(y))^2\pi(x)K(x,y). \tag{6.1.7}$$

Introducing the inner product $< f,g >_\pi = \sum_x f(x)g(x)\pi(x)$, a little algebra shows

$$\mathcal{E}(f,f) = \langle f,(I-K)f\rangle_\pi.$$

If we define the variance by $\text{var}_\pi(f) = E_\pi(f-E_\pi f)^2$ then the spectral gap can be computed from the **variational formula**

$$1-\lambda_1 = \min\{\mathcal{E}(f,f) : \text{var}_\pi(f) = 1\}. \tag{6.1.8}$$

To see this note that $\mathcal{E}(f,f)$ is not affected by subtracting a constant from f so

$$1-\lambda_1 = \min\{\langle f,f\rangle_\pi - \langle f,Kf\rangle_\pi : E_\pi f = 0, \langle f,f\rangle_\pi = 1\}$$

and the result follows from the usual variational formula for λ_1 for the nonnegative symmetric matrix $a_{i,j} = \pi(i)K(i,j)$, i.e.,

$$\lambda_1 = \max\left\{\sum_{i,j} x_i a_{i,j} x_j : \sum_i x_i^2 = 1\right\}.$$

Up to this point we have obtained the continuous time results from the discrete case. They can also be proved directly without too much effort. The developments here follow Chapter 2 of Saloff-Coste (1996).

Lemma 6.1.3. *Let K be a Markov kernel with spectral gap γ then the semigroup $H_t = e^{-t(I-K)}$ satisfies*

$$\|H_t f - \pi(f)\|_2^2 \le e^{-2\gamma t} var_\pi(f) \qquad \text{for all } f \in \ell^2(\pi).$$

Proof. Set $u(t) = \text{var}_\pi(H_t f) = \|H_t(f - \pi(f))\|_2^2 = \|H_t(f) - \pi(f)\|_2^2$. Then

$$u'(t) = -2\mathcal{E}(H_t(f) - \pi(f), H_t(f) - \pi(f)) \le -2u(t).$$

Integrating gives $u(t) \le e^{-2\gamma t}u(0)$ which gives the desried result. □

This leads to a slightly weaker version of Theorem 6.1.2

Theorem 6.1.4. *Let $h_t^x = H_x^t(\cdot)/\pi(\cdot)$. If $\gamma = 1 - \lambda_1$ is the spectral gap of K then*

$$\|h_t^x - 1\|_2 \le \sqrt{1/\pi(x)}e^{-\gamma t} \qquad |h_t(i,j) - \pi_j| \le e^{-\gamma t}\sqrt{\pi_j/\pi_i}.$$

Proof. Let $g_x(y) = 1/\pi(x)$ if $y = x$, 0 otherwise. Lemma 6.1.3 implies

$$\|h_t^x - 1\|_2 \le \sqrt{\frac{1-\pi(x)}{\pi(x)}}e^{-\gamma t}.$$

To prove the second result note that

$$\begin{aligned} |h_t(i,j) - \pi_j| &= \left|\sum_z (h_{t/2}(x,z) - 1)(h_{t/2}(z,y) - 1)\pi(z)\right| \\ &\le \|h_{t/2}^x\|_2 \|h_{t/2}^y\|_2 \le e^{-\gamma t}\sqrt{\pi_j/\pi_i} \end{aligned}$$

where on the second step we have used a Cauchy-Schwarz inequality. □

Relaxation time

There are many ways to quantify the amount of time needed to reach equilibrium. Earlier we defined the mixing time in (6.1.3) in terms of

$$d(t) = \sup_x \|K^t(x,\cdot) - \pi(\cdot)\|.$$

One can define the **relaxation time** in terms of the absolute spectral gap

$$\gamma_* = \max\{|\lambda_i| : \lambda_i \neq 1\}$$

where the λ_i are the eigenvalues of the transition matrix, by setting

$$t_{rel} = 1/\gamma_*. \tag{6.1.9}$$

Combining Theorems 6.1.1 and 6.1.2 we have that in discrete or continuous time

$$|\Delta(\ell t_{rel})| \leq \frac{e^{-\ell}}{\pi_{min}}.$$

6.2 Markov chains and electrical networks

To lead into the theory developed in the next section, we will describe the connection between reversible Markov chains and electrical networks. The idea goes back to Nash-Williams (1959). One can find a leisurely description accessible to undergraduates in the delightful little book by Doyle and Snell (1984), or a more terse account for professional probabilists in Griffeath and Liggett (1982). Intermediate between these two extremes is Chapter 1 in Grimmett's *Probability on Graphs,* which has the same clarity as his excellent book on percolation.

Doyle and Snell work in diuscrete time, Griffeath and Ligett in continuous time. Our first task is show that for the quantities we want to compute here there is no difference, except in the size of the minimum energy, so it is enough to prove the results in one of the two settings.

Given a Markov chain transition probability $p(x, y)$ on a finite set S and a reversible probability measure π, we can let

$$Q(x, y) = \pi(x)p(x, y) = \pi(y)p(y, x)$$

be the flow of probability mass across the unoriented edge $\{x, y\}$ in equilibrium. Given a continuous time Markov chain that jumps from x to y at rate $\alpha(x, y)$ and a reversible probability measure π, we can let

$$Q(x, y) = \pi(x)\alpha(x, y) = \pi(y)\alpha(y, x).$$

To go from discrete to continuous time we can declare that jumps happen at overall rate 1 and let $\alpha(x, y) = p(x, y)$. In the other direction if we let $\alpha(x) = \sum_y \alpha(x, y)$ then $p(x, y) = \alpha(x, y)/\alpha(x)$ is the transition probaility of the embedded jump chain. In either case, to create the connection with electrical networks, we define the **resistance** of the edge to be $R(x, y) = 1/Q(x, y)$.

Given an initial point a and a set B let $V(x) = P_x(T_a < T_B)$ where

$$\begin{aligned} T_a &= \inf\{n \geq 0 : X_n = a\} \text{ is the time of the first visit to } a. \\ T_B &= \inf\{n \geq 0 : X_n \in B\} \text{ is the time of the first visit to } B. \end{aligned}$$

Note that the hitting probabilities $P_x(T_a < T_B)$ are the same in discrete or continuous time. Since we have used $n \geq 0$ in the definitions, it follows that $V(a) = 1$ and $V(b) = 0$ if $b \in B$. We will stop the Markov chain at time $\tau = T_a \wedge T_B$. Let $C = \{x \in S : x \neq a, x \notin B\}$ be the **continuation region** where the random walk continues to jump. The Markov property implies that for $x \in C$ we have

$$V(x) = \sum_y p(x, y)V(y). \tag{6.2.1}$$

The last equation shows that $V(X_{n\wedge\tau})$ is a martingale so

$$V(x) = P_x(T_a < T_B).$$

Rearranging (6.2.1) we have

$$0 = \sum_y p(x,y)(V(x) - V(y)) = \frac{1}{\pi(x)} \sum_y Q(x,y)(V(x) - V(y)). \tag{6.2.2}$$

If we think of $V(x)$ as the voltage at x, **Ohm's law** says that

$$i(x,y) = Q(x,y)[V(x) - V(y)] \tag{6.2.3}$$

is the current that flows from x to y and (6.2.2) is **Kirkhoff's law**: the net current flowing through any $x \in C$ is 0. Given a function on S, define the energy of f by

$$\mathcal{E}(f) = \sum_{\{x,y\}} (f(x) - f(y))^2 Q(x,y)$$

where $\{x,y\}$ indicates that we are summing over edges = unordered pairs.

The next result is on page 63 of Doyle and Snell (1984), and is Theorem (2.1) in Griffeath and Liggett (1982) with $a = 0$ and $B = \Lambda$. The minimum energies in discrete and continuous time differ by a constant factor.

Theorem 6.2.1. Thomson's Principle. *If we minimize $\mathcal{E}(f)$ over all functions with $f(a) = 1$ and $f(b) = 0$ for all $b \in B$ then the minimum occurs at $V(x) = P_x(T_a < T_B)$. The minimum energy is*

$$i_a = \sum_y Q(a,y)[1 - V(y)], \tag{6.2.4}$$

i.e., the current flow out of a or $P_0(T_B < T_a^+)$ where $T_a^+ = \inf\{n \geq 1 : X_n = a\}$.

Proof. Recall that for a random variable X with $EX^2 < \infty$, $E(X - c)^2$ is minimized when $c = EX$. From this it follows that for fixed $x \in C$, $\sum_y p(x,y)[c - f(y)]^2$ is minimized by $c = \sum_y p(x,y)f(y)$. Thus the minimizer must have $f(x) = \sum_j p(x,y)f(y)$ for all $x \in C$ we have proved the first result.

To prove the second result, we sum over ordered pairs (x,y)

$$\begin{aligned} \mathcal{E}(f) &= \frac{1}{2} \sum_{(x,y)} (V(x) - V(y))^2 Q(x,y) \\ &= \frac{1}{2} \sum_{(x,y)} i(x,y)(V(x) - V(y)) = \sum_{(x,y)} i(x,y)V(x) \end{aligned}$$

since $i(y,x) = -i(x,y)$. If $x \in C$ then $\sum_y i(x,y) = 0$. If $x \in B$ then $V(x) = 0$ so we are only left with the terms with $x = a$. $V(a) = 1$ and $i(a,y) = Q(a,y)(1 - V(y)$ so the desired result follows. $\square$

6.2.1 Transience is finite effective resistance

Given an infinite graph and a point a, consider now the sequence of minimization problems that come from taking $B_n = \{x : \text{dist}(a, x) = n\}$ and let $S_n = \cup_{m=0}^n B_m$. Since a function f on S_n with $f = 0$ on B_n can be extended to be 0 on $S_\infty - S_n$ the minimum energy is decreasing in n. From the formula for the energy in (6.2.4) we see that the minimum energy tends to 0 if and only if the random walk on the full graph is recurrent.

This gives a result that Doyle and Snell (1984) call **Rayleigh's monotonicity law**. (2.7) in Liggett and Griffeath (1982) gives a similar result.

Theorem 6.2.2. *If X_n and $\bar{X}_n$ have flow matrices $Q \le \bar{Q}$ and $\bar{X}_n$ is recurrent then X_n is.*

This comparison solves Exercise 9 on page 425 of the third edition of Feller volume I. We have made some minor edits in the statement to make it fit better into our discussion.

Example 6.2.3. Plane random walk with reflecting barriers. Consider a symmetric random walk in a bounded region $D \subset \mathbb{Z}^2$. The boundary is reflecting in the sense that, whenever the particle in the unrestricted random walk the particle would leave the region, it is not allowed to jump. Show that if every point in the region can be reached from every other point, and the region has $K < \infty$ points there is a stationary distribution with $u(x) = 1/K$, $x \in D$. (If D is unbounded then the states are persistent null states and $u(x) = 1$, $x \in D$ is a stationary measure.)

The bounded case is obvious. In the unbounded case it is clear that $u(x) = 1$ is a stationary measure. However, it is far from obvious how to establish recurrence using elementary computations. Theorem 6.2.4 will provide another proof using the connection with resistance.

To connect with the title of this subsection, we will now change our perspective and look for an energy minimizing flow in which the current flowing from a is 1. The solution to the new problem $\bar{V}(x) = V(x)/i_a$. The value at a, $\bar{V}(a) = 1/i_a$ is called the **effective resistance**, R_{EFF}. In this formulation, we see that the random walk on the infinite graph is recurrent if and only if $R_{EFF} \to \infty$ as $n \to \infty$.

There are a number of applications of this fact to checking recurrence or transience of Markov chains. See Sections 3.4, 3.6, and 6 in Doyle and Snell (1984). We will only prove one such result, the **Nash-Williams recurrence criterion**. This result is (2.10) from Griffeath and Liggett (1982) so we will formulate it in continuous time with $\alpha_{i,j}$ the rate of jumps from i to j and $\alpha_i = \sum_j \alpha_{ij}$. Suppose that

(i) the state space S can be partitioned as $S = \sum_{k=0}^\infty \Lambda_k$ where Σ indicates a disjoint union, and that if $i \in \Lambda_k$ and $\alpha_{i,j} > 0$ then $j \in \Lambda_{k-1} \cup \Lambda_k \cup \Lambda_{k+1}$ where $\Lambda_{-1} = \emptyset$

(ii) $\sum_{i \in \Lambda_k} \alpha_i < \infty$

(iii) S is a graph, $\Lambda_0 = \{0\}$ and Λ_k is the set of sites at distance k from 0.

Theorem 6.2.4. *If $P_0(T_{\Lambda_m} < \infty) = 1$ for all $m \geq 1$ then*

$$P_0(T_{\Lambda_m} < T_0^+) \leq (\alpha_0 \Sigma_m)^{-1} \quad \text{where } \Sigma_m = \sum_{k=1}^m (\alpha_k)^{-1}.$$

so X_t is recurrent provided $\sum_{k=1}^\infty \alpha_k^{-1} = \infty$.

Suppose that that the graph is an unbounded subset of $\mathbb{Z}^2$ that contains 0 and has $\alpha_{i,j} = 1/4$ when $i, j \in G$ are nearest neighbors. If we let $\Lambda_m = \{(i,j) : |i| + (j) = m\}$ then $\alpha \leq Ck$ so the chain is recurrent.

6.2.2 A finite energy flow implies transience

Our final result here takes a different approach to checking transience due to Terry Lyons in a paper he wrote when he was an E.R. Hedrick Assistant Professor at UCLA, where he learned about of the work of Griffeath and Liggett (1982). His proof was motivated a 1952 result of Royden which gave a necessary and sufficient condition for the covering surface of a compact Riemann surface to have a Green's function.

The result he proved will be used in Section 6.9. There we will say that a function θ on the vertices of a tree T is a flow if $\theta \geq 0$ and

$$\theta(x) = \sum_{x \to y} \theta(y)$$

where $x \to y$ indicates that y is descendant of x. To extend this to a general graph define for neighbors x, y, define $u_{x,y}$ Which is the flowed through the oriented edge (x, y) to have the following properties

(i) $u_{x,y} = -u_{y,x}$,
(ii) there is an x_0 with $\sum_y u_{x_0,y} \neq 0$ and $\sum_y u_{x,y} = 0$ for $x \neq x_0$.

On the tree, x_0 is the root. If x is the parent of y then $u(x,y) = \theta(y) - \theta(x)$ and $u(y,x) = \theta(x) - \theta(y)$.

Lyons considered a discrete time reversible Markov chain with transition probability $p(x,y)$ with stationary distribution $\pi(x)$. To fit his work into our notations we let

$$Q(x,y) = \pi(x)p(x,y) = \pi(y)p(y,x).$$

Theorem 6.2.5. *The Markov chain p with state space S is transient if (and only if) we can find $u_{x,y}$ that satisfy (i), (ii), and*

$$(iii) \qquad \sum_{x,y \in S} u_{x,y}^2 / Q(x,y) < \infty.$$

The Q in the denominator may look strange but as in the case of the current, the optimal $u(x,y) = [V(x) - V(y)]Q(x,y)$ so the energy is

$$\sum_{x,y} [V(x) - V(y)]^2 Q(x,y).$$

Another physical interpretation. Suppose the Markov chain is constructed from the flows $a_{i,j} = a_{ji}$ with $\pi_i = \sum_{i,j} a_{i,j}$ and $p_{i,j} = a_{i,j}/\pi_i$. Suppose i and j are connected by a pipe with cross-sectional area $a_{i,j}$, and let u_{ij} be the volume rate at which the fluid flows between them. The mass of the fluid is a_{ij}, its velocity is $u_{i,j}/a_{ij}$ so the total kinetic energy is

$$\sum_{ij} a_{ij} \left(\frac{u_{ij}}{a_{ij}}\right)^2 = \sum_{ij} \frac{u_{ij}^2}{a_{ij}}.$$

Proof. The first step is an elementary Hilbert space argument. Let H be the space of all sequences v_{ij} satisfying (iii), and define an inner product on H by

$$\langle v, w\rangle = \sum_{ij} \frac{v_{ij} w_{ij}}{a_{ij}}.$$

For each m define $\ell_m \in H$ by

$$\ell_{ij}^m = \delta_{mi} a_{mk}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. If we think of a_{ij} as a matrix then ℓ^m is the mth row with all the other entries set equal to 0.

To check that $\langle \ell^m, \ell^m\rangle = \pi_m$ we note that

$$\langle \ell^m, \ell^m\rangle = \sum_j \frac{a_{mj}^2}{a_{mj}} = \pi_m$$

and that $u \in H$ has property (ii) if

$$(ii') \qquad \langle \ell^{i_0} u\rangle \neq 0 \quad \text{and when } m \neq i_0 \quad \langle \ell^m u\rangle = 0.$$

Let E be the affine space of $u \in H$ satisfying (ii'). Let w be the unique vector in E that minimizes (iii). By a standard argument from Hilbert space w exists and and characterized by the property

$$\langle w, w - e\rangle = 0 \quad \text{for all } e \in E. \tag{6.2.5}$$

Our second goal is to construct a function W on S (rather than w on $S \times S$) so that

$$W_j - W_i = \frac{w_{ij}}{a_{ij}}.$$

To do this we need to show that if $j_0, j_1, \ldots j_n = j_0$ is a chain of vertices with $j_m \neq j_n$ and $p(j_k, j_{k+1}) \neq 0$ then

$$\sum_{k=0}^{n-1} \frac{w(j_k, j_{k+1})}{a(j_k, j_{k+1})} = 0.$$

Define $f(j_k, j_{k+1}) = 1$ and $f(j_{k+1}, j_k) = -1$ for all $k < n$, with $f = 0$ otherwise, and notice that reversibility implies $\langle f, e^i \rangle = 0$, and hence $w - f \in E$. Using (6.2.5) and then the definition of f

$$0 = \langle f, w \rangle = \sum_{k=0}^{n-1} \frac{w(j_k, j_{k+1})}{a(j_k, j_{k+1})} - \sum_{k=0}^{n-1} \frac{w(j_{k+1}, j_k)}{a(j_{k+1}, j_k)} = 2 \sum_{k=0}^{n-1} \frac{w(j_k, j_{k+1})}{a(j_k, j_{k+1})}$$

since $w_{ij} = -w_{i,j}$ and $a_{ij} = a_{j,i}$.

W has properties (a) $W(i_0) = 0$, (b) $W \not\equiv 0$ (since $\langle w, \ell^{i_0} \rangle \neq 0$)

$$\text{(c) } W_i = \sum_j p_{ij} W_j \quad \text{and} \quad \text{(d) } \sum_{ij} a_{ij} (W_i - W_j)^2 < \infty.$$

To complete the proof we will show that the existence of W this is incompatible with recurrence. To do this let T be the hitting time of i_0. $W(X_{k \wedge T})$ is a martingale that converges to 0 as $k \to \infty$. If we can show that it is L^2 bounded, i.e.,

$$M_i = \sup_k E_i([W(Y_K \wedge T) - W(Y_0)]^2) < \infty$$

then it follows that $W(i) \equiv 0$ contradicting (b). The orthogonality of martingale increments (Theorem 4.4.7) in PTE5 implies that

$$M_i = \sum_j g(i,j) \sum_k p(j,k)(W_j - W_k)^2 \tag{6.2.6}$$

where $g(i,j)$ is the expected number of visits to j before hitting i_0. Reversibility and the Markov property imply

$$g(i,j) \leq \frac{\pi_j}{\pi_i} g(j,i) \leq \frac{\pi_j}{\pi_i} g(i,i).$$

The probability of hitting i_0 eventually is 1, so the probability of starting from i and hitting i_0 before i must be > 0 and hence $g(i,i) < \infty$. Using (6.2.6) we see that $M_i < \infty$ which completes the proof. □

6.3 Conductance

Let $Q(x,y) = \pi(x)K(x,y)$ be the flow across the edge $\{x,y\}$, let $Q(S,S^c) = \sum_{x\in S, y\in S^c} Q(x,y)$ and define

$$\Phi(S) = \frac{Q(S,S^c)}{\pi(S)} \qquad \Phi_\star = \min_{\pi(S)\le 1/2} \Phi(S).$$

Levin and Peres (2017) call $\Phi_\star$ the **bottleneck ratio**. To see what this has to do with the rate of convergence to equlibirum we will state and prove their Theorem 7.4. The mixing time was defined in (6.1.3).

Theorem 6.3.1. $t_{mix}(1/4) \ge 1/4\Phi_\star)$.

Proof. Suppose that X_t is stationary, i.e., the distribution is π for all t.

$$P_\pi(X_0 \in A, X_t \in A^c) \le \sum_{s=1}^t P_\pi(X_{s-1} \in A, X_s \in A^c) = tQ(A,A^c).$$

This implies $P_\pi(X_t \in A^c | X_0 \in A) \le t\Phi(A)$ so there is an $x \in A$ with

$$P^t(x,A) \ge 1 - t\Phi_\star(A)$$

and we conclude $d(t) \ge 1 - t\Phi(A) - \pi(A)$. If $\pi(A) \le 1/2$ and $t < 1/(4\Phi(A))$ then $d(t) > 1/4$. This shows $t_{mix} \ge 1/\Phi(A))$ and maximizing over A gives the desired result. □

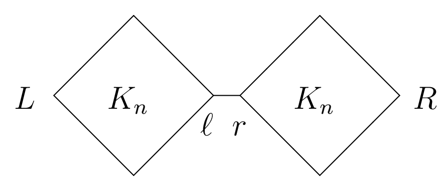

Figure 6.1: Picture of Example 6.3.2

Example 6.3.2. Take two complete graphs on n vertices, call them L and R. Pick $\ell \in L$ and $r \in R$ and connect them by an edge. To simplify the computation we formulate the chain in continuous time with jumps across each edge occur at rate $1/n$. Let P_R denote the law of the Markov chain starting with all states in R having probability $1/n$ and $P_R(X_0 \in L) = 0$. If we let $u(t) = P_R(X_t \in R)$. then

$$\begin{aligned} u'(t) &= \frac{P_R(X_t = r) - P_R(X_t = \ell)}{n} \\ &\approx \frac{P_R(X_t \in R) - P_R(X_t \in L)}{n^2} = \frac{u(t) - (1-u(t))}{n^2} \end{aligned}$$

where in the second step we have used the approximation that all sites in R at time t have the same probability. (6.3.1) provides support for this but we leave the tedious details of writing a rigorous proof to the reader. Letting $v(t) = 2u(t) - 1$ we have $v'(t) = 2u'(y) = 2v(t)/n^2$. Since $v(0) = 1$, we have $v(t) = \exp(-2t/n^2)$ and

$$u(t) = \frac{1 + v(t)}{2} = \frac{1 + e^{-2t/n^2}}{2}.$$

To compute $\Phi(R, L)$ in this example we must first compute $\pi(x)$. To do this we note that ℓ and r have degree n while all the other vertices that we will call x have degree $n - 1$ so

$$\pi(\ell) = \pi(r) = \frac{n}{2(n + (n-1)^2)} \qquad \pi(x) = \frac{n}{2(n + (n-1)^2)} \tag{6.3.1}$$

and the other vertices x in equilibrium we have $\pi(x) = \pi(\ell) \cdot (n-1)/n$. Simplifying the exact formula to $\pi(y) \approx 1/2n$ for all y.

$$\Phi(R, L) = \frac{(1/2n) \cdot (1/n)}{1/2} = n^{-2}.$$

Since we start with all of the mass on R and in equlibirum $\pi(R) = 1/2$ it follows that the time to equilibrium is at least $O(n^2)$.

6.3.1 Cheeger's inequality

Since we learned about Markov chain mixing times from Laurent, our next result is Lemma 3.3.7 in Saloff-Coste (1996). His $I = 2h$ so the constants are different. Saloff-Coste attributes the result to Diaconis and Stroock (1991), who in turn named the result Cheeger's inequality in honor of the eigenvalue bound in differential geometry. Levin and Peres (2017) cite Sincalir and Jerrum (1989) and Lawler and Sokal (1988).

Theorem 6.3.3. *The spectral gap has*

$$\frac{h^2}{2} \le 1 - \lambda_1 \le 2h.$$

Proof. Taking $f = 1_S$ in the variational formula (6.1.8) we have

$$\mathcal{E}(1_S, 1_S) = Q(S, S^c)$$

and $\text{var}_\pi(1_S) = \pi(S)(1 - \pi(S))$, so $1 - \lambda_1 \le Q(S, S^c)/\pi(S)(1 - \pi(S))$. The right-hand side is the same for S and S^c, so we can restrict our attention to $\pi(S) \le 1/2$. Since $1 - \pi(S) \ge 1/2$, we have $1 - \lambda_1 \le 2h$.

For the other direction, let $F_t = \{x : f(x) \ge t\}$ and let f_t be the indicator function of the set F_t. Since only differences $f(x) - f(y)$ appear in $\mathcal{E}(f, f)$, defined in (6.1.7), we can without loss of generality suppose that the median of f is 0, i.e., $\pi(F_t) \le 1/2$ for $t > 0$, and

$\pi(F_t^c) \le 1/2$ for $t < 0$. Our next step is to compute something that would be the Dirichlet form if we had squared the increment.

$$\begin{aligned}
\frac{1}{2}\sum_{x,y} |f(x)-f(y)|Q(x,y) &= \sum_{f(x)>f(y)} (f(x)-f(y))Q(x,y) \\
&= \sum_{x,y} \int_{-\infty}^{\infty} 1_{\{f(y)<t<f(x)\}} Q(x,y)\,dt \\
&= \int_0^{\infty} |\partial F_t|\,dt + \int_{-\infty}^{0} |\partial F_t^c|\,dt \\
&\ge h\left(\int_0^{\infty} \pi(F_t)\,dt + \int_{-\infty}^{0} \pi(F_t^c)\,dt\right) = h\pi(|f|).
\end{aligned}$$

Continuing to suppose that the median of f is 0, let $g = f^2 \operatorname{sgn}(f)$, where $\operatorname{sgn}(x) = 1$ if $x > 0$, $\operatorname{sgn}(x) = -1$ if $x < 0$, and $\operatorname{sgn}(0) = 0$. $|g| = f^2$ so the last inequality implies

$$2h\pi(f^2) \le \sum_{x,y} |g(x)-g(y)|Q(x,y) \le \sum_{x,y} |f(x)-f(y)|(|f(x)|+|f(y)|)Q(x,y).$$

To check the last inequality, we can suppose without loss of generality that $f(x) > 0$ and $f(x) > f(y)$. If $f(y) \ge 0$ we have an inequality, while if $f(y) < 0$ we have $f^2(x) + f^2(y) < (|f(x)| + |f(y|)^2$. Using the Cauchy-Schwarz inequality now the above is

$$\begin{aligned}
&\le \left(\sum_{x,y} (f(x)-f(y))^2 Q(x,y)\right)^{1/2} \cdot \left(\sum_{x,y} (|f(x)|+|f(y)|)^2 Q(x,y)\right)^{1/2} \\
&\le (2\mathcal{E}(f,f))^{1/2}(4\pi(f^2))^{1/2}.
\end{aligned}$$

Rearranging gives $(2\mathcal{E}(f,f))^{1/2} \ge h(\pi(f^2))^{1/2}$. Squaring we have

$$\mathcal{E}(f,f) \ge \frac{h^2}{2}\pi(f^2) \ge \frac{h^2}{2} E_\pi(f - E_\pi f)^2$$

which proves the desired result. □

Example 6.3.4. Markov chains on graphs. Let G be a finite connected graph, $d(x)$ be the degree of x, and define a transition kernel by $K(x,x) = 1/2$, $K(x,y) = 1/2d(x)$ if $x \sim y$ and $K(x,y) = 0$ otherwise. Our K can be written $(I+p)/2$ where p is another transition probability, so all of the eigenvalues of K are in $[0,1]$, and $\lambda_{max} = \lambda_1$. $\pi(x) = d(x)/D$ where $D = \sum_{y \in G} d(y)$, defines a reversible stationary distribution since $\pi(x)K(x,y) = 1/2D = \pi(y)K(y,x)$. Letting $e(S,S^c)$ is the number of edges between S and S^c, and $\mathrm{vol}(S)$ be the sum of the degrees in S, we have

$$h = \frac{1}{2} \min_{\pi(S) \le 1/2} \frac{e(S,S^c)}{\mathrm{vol}(S)}.$$

When $d(x) \equiv d$, $h = \iota/2d$ where

$$\iota = \min_{|S| \le n/2} \frac{e(S, S^c)}{|S|}$$

is the **edge isoperimetric constant**.

The next two examples show that both inequalities iun Theorem 6.3.3 are "sharp."

Example 6.3.5. Random walk on the circle. To illustrate the use of Theorem 6.3.3 and to show that one cannot get rid of the power 2 from the lower bound, consider random walk on the circle $\mathbb{Z}$ mod n in which we stay put with probability 1/2 and jump from x to $x \pm 1$ with probability 1/4 each. Taking $S = \{1, 2, \dots n/2\}$ we see that

$$\iota = \frac{2}{n/2} = 4/n.$$

To bound the spectral gap, we let $f(x) = \sin(\pi x/n)$. Since $\sin(a+b) = \sin a \cos b + \sin b \cos a$ we have

$$(I - K)f(x) = f(x)(1 - \cos(\pi/n))/2$$

and $1 - \lambda_1 \le (1 - \cos(\pi/n))/2 \sim \pi^2/4n^2$ as $n \to \infty$. Using Theorem 6.1.1 gives an upper bound on the convergence time of order $O(n^2 \log n)$. However using the local central limit theorem for random walk on $\mathbb{Z}$ it is easy to see that $\Delta(t) \le \epsilon$ at a time $K_\epsilon n^2$.

Example 6.3.6. n-dimensional hypercube. Consider the lazy random walk on the hypercube $\{0, 1\}^n$ and let $S = \{x : x_1 = 0\}$. Since $\pi(S) = 1/2$.

$$\begin{aligned}\Phi(S) &= 2 \sum_{x \in S, y \in S^c} 2^{-n} p(x, y) \\ &= 2^{-n+1} \cdot 2^{n-1} \cdot \frac{1}{2n} = \frac{1}{2n}.\end{aligned}$$

To determine the rate of convergence to equilibrium note that the lazy chain may be formulated as: ocik a coordinate at random and then flip a fiar coin to determine its state. From this it is clear that when we have touched all the coordinates the system is in equilibrium. To compute the spectral gap note that the coordinates are independent so $\gamma = 1/n$. It is not hard to verify this by hand nut the details can be found in Section 12.4 of Levin and Peres (2017) which concerns product chains.

6.3.2 Mixing times and the conductance profile

In some examples given an initial state i, it is possible to define stopping times T (which are called **stationary times** so that X_T has the the stationary distribution. Define $H(i, \pi)$ the minimum value of ET for all such stopping times and let $\mathcal{H} = \max_i H(i, \pi)$. For more on stationary times (and strong stationary times) see Chapter 6 in Levin and Peres (2017)

To be completely accurate we should redefine the mixing time by

$$T_{mix} = \max_i \min\{t : d_{TV}(K^t(i,\cdot),\pi) < 1/e\}.$$

Earlier, see (6.1.3), we used 1/4 instead of $1/e$. The choice of thrshold is not important as long as it is small enough, but it does affect the C_i in the next result. Aldous (1988) has shown, see also Aldous, Lovász, and Winkler (1997), that

$$C_1\mathcal{H} \le T_{mix} \le C_2\mathcal{H}.$$

Define the **conductance profile** by

$$\Phi(x) = \min_{S:0<\pi(s)\le x} \frac{Q(S,S^c)}{\pi(S)\pi(S^c)}.$$

Lovàsz and Kannan (1999) have shown that

$$\mathcal{H} \le 32\int_{\pi_{min}}^{1/2} \frac{dx}{x\Phi(x)^2}.$$

Morris and Peres (2003) used their notion of evolving sets to sharpen this result to

$$\text{If} \quad n \ge \int_{\pi(i)\wedge\pi(j)}^{4/\epsilon} \frac{4\,dx}{x\Phi(x)^2} \quad \text{then} \quad \left|\frac{K^n(i,j)}{\pi(i)} - 1\right| \le \epsilon.$$

These results are useful for improving rate of convergence results in some examples. However in some of our favorite examples the worst conductance occurs for small sets, so we will instead use a recent result of Fountoulakis and Reed (2007).

Theorem 6.3.7. *If $\Phi_c(x)$ is the minimum $Q(S,S^c)/\pi(S)\pi(S^c)$ over all connected sets S with $x/2 \le \pi(S) \le x$ then*

$$T_{mix} \le 32\int_{\pi_{min}}^{1/2} \frac{dx}{x\Phi_c(x)^2}.$$

6.4 Fixed degree distribution, minimum degree 3

Gkantsidis, Mihail, and Saberi (2003) have proved the following:

Theorem 6.4.1. *Consider a random graph with a fixed degree distribution in which the minimum degree is $r \geq 3$. There is a constant $\alpha_0 > 0$ so that with probability tending to 1 as $n \to \infty$*

$$\min_{\pi(S) \leq 1/2} \frac{e(S, S^c)}{vol(S)} \geq \alpha_0,$$

and hence by Theorem 6.1.1 it follows that the mixing time is $\leq C \log n$.

The diameters of these graphs are of $O(\log n)$ so it cannot occur at a faster rate. The condition $r \geq 3$ is necessary since if there is a positive density of vertices of degree 2 then there will be paths of length $O(\log n)$ in which each vertex has degree 2 and if we start in the middle of the path then the mixing time will be $\geq O(\log^2 n)$. We will consider that case in the next section.

Proof. We say that a set of vertices S is bad if $e(S, S^c)/\text{vol}(S) \leq \alpha$. Our goal is to show that $\bar{P}(\exists \text{ bad } S) \to 0$. There are at most $\binom{D/r}{k/r}$ sets that have volume k.

Let $f(m)$ be the number of ways of dividing m objects into pairs.

$$f(m) = \frac{m!}{(m/2)! 2^{m/2}}. \tag{6.4.1}$$

Let $P(k, \ell)$ that there is a set S with $\text{vol}(S) = k$ has $e(S, S^c) = \ell$.

$$P(k, \ell) \leq \binom{k}{\ell} \binom{D-k}{\ell} \ell! f(k-\ell) f(D-k-\ell) \frac{1}{f(D)}. \tag{6.4.2}$$

To see this recall that in the random configuration model we pair the D half-edges at random, which can be done in $f(D)$ ways. We pick ℓ of k half-edges in S and ℓ of those $D-k$ in S^c. The ℓ which will make up $e(S, S^c)$ can be paired in $\ell!$ ways. Then the remaining $k-\ell$ half-edges in S can be paired in $f(k-\ell)$ ways and the $D-k-\ell$ in S^c in $f(D-k-\ell)$ ways.

To make it easier to compare with the argument in GMS we change values $\ell = \alpha k$. Taking into account the number of choices of S, the probability of a bad set with volume k and $e(S, S^c) = \alpha k$ is

$$\binom{D/r}{k/r} \binom{k}{\alpha k} \binom{D-k}{\alpha k} \frac{(\alpha k)! f(k-\alpha k) f(D-k-\alpha k)}{f(D)}. \tag{6.4.3}$$

Their formula (10) is this with $s! = (\alpha k)!$ replaced by the larger $f(2\alpha k)$. They also have a factor αk to account for $1 \leq s \leq \alpha k$.

To bound the binomial coefficients, the following lemma is useful

Lemma 6.4.2.

$$\binom{n}{m} \leq \frac{n^m}{m!} \leq \frac{n^m}{m^m e^{-m}}.$$

Proof. The first inequality follows from $n(n-1)\cdots(n-m+1) \le n^m$. For the second we note that the series expansion of e^m has only positive terms so $e^m > m^m/m!$. □

From Lemma 6.4.2, we see that the three binomial coefficients in (6.4.3) are

$$\le \left(\frac{De}{k}\right)^{k/r} \left(\frac{e}{\alpha}\right)^{2\alpha k} \left(\frac{D-k}{k}\right)^{\alpha k}. \tag{6.4.4}$$

Here, to prepare for a later step, we have transferred part of the bound for the third term into the second.

To bound the f's in (6.4.3) we use Stirling's formula to conclude

$$f(m) = \frac{m!}{(m/2)!2^{m/2}} \sim C\frac{m^{m+1/2}e^{-m}}{(m/2)^{m/2+1/2}e^{-m/2}2^{m/2}} = C(m/e)^{m/2}.$$

From this we see that the fraction in (6.4.3) is

$$\le Ck^{1/2}\frac{(\alpha k/e)^{\alpha k}(k(1-\alpha)/e)^{k(1-\alpha)/2}((D-(1+\alpha)k)/e)^{(D-(1+\alpha)k)/2}}{(D/e)^{D/2}} \tag{6.4.5}$$

$$= Ck^{1/2}(\alpha k)^{\alpha k}D^{-\alpha k}\left(\frac{k(1-\alpha)}{D}\right)^{k(1-\alpha)/2}\left(1-\frac{(1+\alpha)k}{D}\right)^{(D-(1+\alpha)k)/2} \tag{6.4.6}$$

since the exponents in the numerator sum to $D/2$.

Combining (6.4.4) and (6.4.6) gives an upper bound

$$\le Ck^{1/2}\left(\frac{De}{k}\right)^{k/r}\left(\frac{e^2}{\alpha}\right)^{\alpha k}\left(\frac{D-k}{D}\right)^{\alpha k} \cdot \left(\frac{k(1-\alpha)}{D}\right)^{k(1-\alpha)/2}\left(1-\frac{(1+\alpha)k}{D}\right)^{(D-(1+\alpha)k)/2}.$$

Ignoring the $Ck^{1/2}$'s, the first term is the first term from (6.4.4), the second and third terms come from combining the second and third terms of (6.4.4) and with the first and second terms of (6.4.6), while the remainder of the formula comes from (6.4.6). Using $\alpha > 0$ and $D - k < D$ and rearranging we have

$$\le Ck^{1/2}e^{k/r}\left(\frac{e^2}{\alpha}\right)^{\alpha k}\cdot\left(\frac{k}{D}\right)^{k(1-\alpha)/2-k/r}\left(1-\frac{(1+\alpha)k}{D}\right)^{(D-(1+\alpha)k)/2}.$$

Setting $\beta = e^2/\alpha$ and $\gamma = (1-\alpha)/2 - 1/r$ we have

$$\le Ck^{1/2}e^{k/r}\beta^{\alpha k}\cdot\left(\frac{k}{D}\right)^{\gamma k}\left(1-\frac{(1+\alpha)k}{D}\right)^{(D-(1+\alpha)k)/2}. \tag{6.4.7}$$

Comparing with formula (17) in GMS, we see that apart from the differences that result from our use of $(\alpha k)!$ instead of $f(2\alpha k)$, they are missing the $e^{k/r}$ and we have retained an extra term to compensate for the error. Let

$$G(k) = e^{k/r}\beta^{\alpha k} \cdot \left(\frac{k}{D}\right)^{\gamma k} \left(1 - \frac{(1+\alpha)k}{D}\right)^{(D-(1+\alpha)k)/2}.$$

$Ck^{1/2} \le Cn^{1/2}$ so we can show $h \ge \alpha_0$ by showing that for $0 \le \alpha \le \alpha_0$

$$\sup_{1\le k\le D/2} G(k) = o(n^{-5/2})$$

because then we can sum our estimate over $k \le D/2$ and $s = \alpha k$ with $\alpha \le \alpha_0$ and end up with a result that is $o(1)$.

$\beta^{\alpha k} = \exp(\eta k)$ where $\eta = \alpha \log(e^2/\alpha) \to 0$ as $\alpha \to 0$. Ignoring this term, and setting $k = D/2$, $\alpha = 0$

$$G(D/2) = e^{D/2r}(1/2)^{[1/2-1/r]D/2+D/4} = \left(e^{1/3}(1/2)^{2/3}\right)^{D/2}$$

when $r = 3$, the worst case. Since $4 > e$, the quantity in parentheses is < 1 when $\alpha = 0$ and hence also when $0 \le \alpha \le \alpha_0$, if α_0 is small.

To extend this result to other values of k, let

$$H(k) = \log G(k) = \frac{k}{r} + k\alpha\log\beta + k\gamma\log(k/D) + \frac{D-(1+\alpha)k}{2}\log\left(1 - \frac{(1+\alpha)k}{D}\right).$$

Since $G(k) = \exp(H(k))$, differentiating gives $G'(k) = G(k)H'(k)$ where

$$H'(k) = \frac{1}{r} + \alpha\log\beta + \gamma\log(k/D) + \gamma - \frac{(1+\alpha)}{2}\log\left(\frac{D-(1+\alpha)k}{D}\right) + \frac{D-(1+\alpha)k}{2} \cdot \frac{D}{D-(1+\alpha)k} \cdot \left(\frac{-(1+\alpha)}{D}\right).$$

Differentiating again $G''(k) = G(k)(H'(k)^2 + H''(k))$ where

$$H''(k) = \frac{\gamma}{k} - \frac{(1+\alpha)}{2} \cdot \frac{D}{D-(1+\alpha)k} \cdot \left(\frac{-(1+\alpha)}{D}\right) > 0.$$

From the last calculation we see that $G(k)$ is convex. We have control of the value for $k = D/2$. It remains then to inspect the values for small k. Dropping the last factor which is < 1

$$G(k) \le e^{k/r}\beta^{k\alpha}\left(\frac{1}{\alpha D}\right)^{\gamma k}.$$

When $0 \le \alpha \le \alpha_0 \le 1/24$, $\gamma \ge 7/48$ and hence $G(24) \le Cn^{-7/2}$. Since $e(S, S^c) \ge 1$ there is nothing to prove for $k \le 1/\alpha_0 = 24$ and the proof is complete. □

6.5 Effect of degree 2 vertices

In the previous section we saw that if the minimum degree on a connected graph with n vertices is 3 then the random walk mixes in time $O(\log n)$. This conclusion is false for configuration model graphs if there are vertices of degree 2. It is easy to see that such graphs will have paths of length $\geq \delta \log n$ in which all the vertices have degree 2. If we start in the middle of such a path, it will with positive probability take time $\delta^2 \log^2 n$ to escape from it.

I knew this when I was writing the first edition of this book, which was published in 2007. I wanted to prove an upper bound of the same order, but I could not cope with the complexities of a general graph, so I created a simple example. Start with a random 3-regular graph H with $(1-a)n$ vertices, and hence $3(1-a)n/2$ edges. We produce a new graph G_{23} by replacing each edge by a path with a geometric number of edges with success probability r, i.e., with probability $(1-r)^{j-1}r$ we have j edges. The number of vertices of degree 2 in one of these paths has mean $(1/r) - 1$ so if we pick r so that

$$\frac{3(1-a)}{2} \cdot ((1/r) - 1) = a,$$

then we asymptotically have n vertices and $p_2 = a$.

Our main result is that

Theorem 6.5.1. *The mixing time of the lazy random walk on G_{23} is $\Theta(\log^2 n)$.*

To prove an upper bound, we will use mixing time result in Theorem 6.3.7. I appreciate the fact that on a trip to Montreal (circa 2006) Bruce Reed (a) spent a long time explaining the proof to me, and (b) gave me a tutorial on the wonders of Belgian beer.

Proof. To begin we need a simple combinatorial result.

Lemma 6.5.2. *The number of connected subsets of H of size k containing a fixed vertex v_0 is $\leq 3^{3k}$.*

Proof. Given a connected set V of vertices of H, define the set $W = \{(x, y) : x, y \in V, x \neq y\}$. Note that if $(x, y) \in W$ then $(y, x) \in W$ and think of these as two oriented edges between x and y. We will show that there is a Hamiltonian path starting from v_0 that traverses each oriented edge at most once. The number of edges in W is at most $3k$. At each stage we have at most 3 choices so the number of such paths is $\leq 3^{3k}$ this proves the desired result.

To construct the path start at v_0 and pick an outgoing edge. When we are at a vertex $v \neq v_0$ we have used one more incoming edge than outgoing edge so we have at least one way out. This procedure may terminate by coming back to v_0 at a time when there are no more outgoing edges. If so, and we have not exhausted the graph, then ther is some vertex v_1 on the current path with an outgoing edge. Repeat the construction starting from v_1 using edges not in the current path. We will eventually come back to v_1. We can combine the two paths by using the old path from v_0 to the first visit to v_1, using the new path to go from v_1 to v_1, and then the old path to return from v_1 to v_0. Repeating this construction we will eventually exhaust all of the edges. □

Let B be a connected subset of G, and let $A = B \cap H$. It is easy to see that A is a connected subset of H. By the isoperimetric inequality for random regular graphs, there is an $\alpha > 0$ so that $|\partial A| \geq \alpha |A|$, where ∂A is the set of edges (x, y) with $x \in A$ and $y \notin A$. From the construction of the graph it is easy to see that $|\partial B| = |\partial A|$.

It remains to see how big $|B|/|A|$ can be. When $|A| = 1$ we can have $|B| = O(\log n)$. The key to the proof is to show that the ratio cannot be big when $|A|$ is. Let X_i be i.i.d with $P(X_i = j) = (1-r)^j r$ and let $S_m = X_1 + \cdots X_m$.

Lemma 6.5.3. *There are constants β and γ so that*

$$P(S_m \geq \beta \log n + \gamma m) \leq n^{-2}(2/81)^m.$$

Proof. The moment generating function

$$\psi(\theta) = Ee^{\theta X_i} = \sum_{j=0}^{\infty} (e^\theta(1-r))^j r = \frac{r}{1 - e^\theta(1-r)}.$$

when $e^\theta(1-r) < 1$. If we pick $\theta > 0$ so that $e^\theta(1-r) = 1 - r/2$ then $\psi(\theta) = 2$. Markov's inequality implies

$$P(S_m \geq \beta \log n + \gamma m) \leq \psi(\theta)^m \exp(-\theta[\beta \log n + \gamma m]).$$

Letting $\beta = 2/\theta$ and $\gamma = 81/\theta$ the desired result follows. □

If $|A| = k$ then the number of edges adjacent to some point in A is $\geq k + 2$, the value for a tree and $\leq 3k$. Since the number of connected sets of size k is $\leq n27^k$ it follows that with probability $1 - O(n^{-1})$ we have $|B| \leq \beta \log n + 3\gamma |A|$ for connected sets B. From this it follows that

$$|\partial B| = |\partial A| \geq \alpha |A| \geq \frac{\alpha}{\gamma}(|B| - \beta \log n)$$

if $|B| \geq 2\beta \log n$ then $|\partial B|/|B| \geq c$, while for $|B| \leq 2\beta \log n$, $|\partial B|/|B| \geq 2/|B|$.

To evaluate $\int_{1/3n}^1 dx/(x\Phi(x)^2)$ up to a constant factor we note that

$$\int_{2\beta \log n/n}^{1/2} \frac{dx}{x} = O(\log n)$$

while changing variables $y = nx$, $dy = n\,dx$ shows

$$\int_{1/3n}^{2\beta \log n/n} \frac{dx}{x(2/xn)^2} = \int_{1/3}^{2\beta \log n} (y/2)\,dy = O(\log^2 n)$$

and completes the proof. □

Other related results:

Fountoulakis and Reed (2008) investigated Erdős-Rényi graphs with λ ranging from $(1+\epsilon)/n$ to $(2\log n)/n$ and identified the range of values for which the mixing time is $\Theta((\log n)^2)$. See Section **??**.

Benjamini, Kozma, and Wormald (2014) studied the Erdős-Rényi graph $G(n,m)$ with a fixed number of edges m and $m/n \to c > 1$ and showed that the mixing time was $\Theta(\log^2 n)$. They did this by showing that (i) these graphs are *decorated expanders*: there is an expander subgraph B which they call the *strong core*, which leaves only small components when it is deleted, and (ii) proving a mixing time result for decorated expanders

Ding, Lubetzky, and Peres (2014) gave a description of the giant component of the Erdős-Rényi graph outside the critical window, i.e., $\lambda = (1+\epsilon)/n$ with $\epsilon^3 n \to \infty$. In rough terms there are three steps (i) one builds a graph with minimum degree 3 using the configuration model, (ii) subdivide each edge into a path using i.i.d. geometrics, and (iii) attach a Poisson number of subcritical Galton-Watson trees to each vertex.

This work is a sequel to a paper the trio wrote in 2011 with J.H. Kim on the "anatomy of a young giant component," which concentrates on the situation ϵ is small. In both cases the constructed model is only **contiguous** to the giant component for the Erdős-Rényi graph. When two random graph models are contiguous they have the same qualitative properties: anything that hold a.a.s. (asymptotically almost surely) for one model holds a.a.s for the other, but quantitative properties may differ. See Section 9.6 of Janson, Luczak, and Ruciński's book *Random Graphs* for a brief course in contiguity or see Kim's (2008) paper for more thorough treatment

6.6 Connected Erdös-Rényi graphs

In this section we will consider random walk on Erdős-Rényi$(n, (c\log n)/n)$ with $c > 1$, which Theorem 1.9.1 has shown is connected with high probability for large n. The word "connected" in the title of the section is a bit of a red herring because we make this assumption on the value of λ not to avoid the trivial obstruction to convergence that somes from lack of irreducibility, but to have the stability that occurs when all vertices have large degree. At the end of this section we will describe results of Fountoulakis and Reed (2008) that apply to all values of

$$\lambda \in [(1+\epsilon)/n, (2\log n)/n]$$

and show how the mixing rate evolves as λ is increases.

Let $d(x)$ be the degree of x, write $x \sim y$ if x and y are neighbors. To avoid problems associated with periodicity, we run the lazy random walk with

$$K(x,x) = 1/2 \qquad K(x,y) = 1/2d(x) \quad \text{if } x \sim y$$

and $K(x,y) = 0$ otherwise. Our analysis of the mixing time follows Cooper and Frieze (2003) who were primarily interested in the cover time $\mathcal{C}_G$, i.e., the time to visit all the vertices.

Theorem 6.6.1. *Suppose that $np = c_n \log n$ where $c_n = O(1)$ and $(c_n - 1)\log n \to \infty$. Then whp*

$$\mathcal{C}_G \sim c \log\left(\frac{c}{c-1}\right) n \log n.$$

To prove this result they needed to establish a number of prperties of the Erdős-Rényi graph $G_{n,p}$ including

Theorem 6.6.2. *Consider $ER(n, (c\log n)/n)$ with $c > 1$. The lazy random walk mixes in time $O(\log n)$.*

Proof. We begin by estimating the maximum and minimum degrees of vertices.

Lemma 6.6.3. *There is a constant $\delta > 0$ so that if n is large then*

$$\delta c \log n \le d(x) \le 4c\log n \qquad \textit{for all } x.$$

Proof. By the large deviations result in Lemma **??** if $X = \text{Binomial}(n,p)$ then

$$P(X \ge np(1+y)) \le \exp(-npy^2/2(1+y)).$$

Taking $p = (c\log n)/n$, and $y = 3$

$$P(X \ge 4c\log n) \le \exp(-9(c\log n)/8) = n^{-9c/8}.$$

Since we have assumed $c > 1$ with high probability, the maximum degree in the graph is $\le 4c\log n$.

To get a lower bound, we need the more precise result in Lemma **??**. The function

$$H(a) = a\log(a/p) + (1-a)\log((1-a)/(1-p))$$

defined there has $H(0) = -\log(1-p)$, which is sensible since $P(X=0) = (1-p)^n$. When $p = c\log n$

$$(1 - (c\log n)/n)^n \le n^{-c}.$$

Taking $a = (\delta c \log n)/n$, we have

$$H(a) = \frac{\delta c \log n}{n}\log\delta + \left(1 - \frac{\delta c\log n}{n}\right)\log\left(\frac{1-\delta c\log n/n}{1-c\log n/n}\right).$$

The logarithm in the second term is

$$\log\left(1 + \frac{(1-\delta)c\log n/n}{1-c\log n/n}\right) \sim (1-\delta)c\log n/n$$

as $n\to\infty$. As $\delta\to 0$, $\delta\log\delta\to 0$, so if δ is small enough then $H(\delta)\sim(b\log n)/n$ with $b>1$ as $n\to\infty$ and we conclude that with probability that tends to one, the minimum degree in the graph is $\ge \delta c\log n$. □

To prove the theorem we will estimate the conductance h introduced in Section 6.3. . By considering the number of edges we see that

$$\mathrm{vol}(G) = 2\mathrm{binomial}(\binom{n}{2}, c\log n/n)$$

which has mean $\sim cn\log n$ and variance $\sim cn\log n$, so $\mathrm{vol}(G)\sim cn\log n$. The maximum degree $\le 4c\log n$ with high probability for large n, so if $|S^c|\le n/10$ then

$$\mathrm{vol}(S^c) \le 4c\log n\cdot n/10$$

and hence for large n no set with $|S|\ge 9n/10$ will have $\pi(S)\le 1/2$.

Case 1. Consider $B = \{S : n/(c\log n)\le|S|\le 9n/10\}$ and let $s=|S|$. Using Lemma 6.4.2 we have

$$\binom{n}{s} \le (n/s)^s e^s = \exp(s[\log(n/s)+1]). \tag{6.6.1}$$

Lemma 1.9.5 says that if $X = \mathrm{binomial}(n,p)$ then

$$P(X\le np(1-y)) \le \exp(-npy^2/2).$$

The number of edges from S to S^c is $\mathrm{binomial}(s(n-s), c\log n)/n)$. Using the large deviations result with $n = s(n-s)$, $p=p$, and $y=1/2$ gives that the binomial probability

$$\le \exp(-s(n-s)(c\log n)/8n). \tag{6.6.2}$$

Combining the last two estimates and using $n - s \geq n/10$, $s \geq n/(c\log n)$ we see that there is a $S \in B$ with $e(S, S^c) \leq s(n-s)p/2$ is

$$\leq n \cdot \exp\left(-\frac{n}{c\log n}\left[\frac{c\log n}{80} - \log(c\log n) - 1\right]\right)$$

which goes to 0 exponentially fast as $n \to \infty$. To finish up now we note that

$$s(n-s)p/2 = s(n-s)(c\log n)/2n \geq sc(\log n)/20$$

while $\text{vol}(S) \leq 4sc\log n$, so for sets in B we have $e(S, S^c)/\text{vol}(S) \geq 1/80$.

Case 2. Finally we have to deal with the small sets $A = \{S : 1 \leq |S| \leq n/(c\log n)\}$. In this case we upper bound $e(S, S)$ in order to conclude $e(S, S^c)$ is large. $E|e(S, S)| \leq (s^2/2)p \leq s/2$ so

$$P(\exists S \in A : e(S, S) \geq s\log\log n) \leq C\binom{n}{s}\binom{s^2/2}{s\log\log n}p^{s\log\log n}.$$

The right-hand side is the probability $e(S, S) = s\log\log n$, ignoring the fact that this may not be an integer. However in this part of the tail, the probabilities decay exponentially fast. Bounding the binomial coefficients using Lemma 6.4.2, and filling in the value of p

$$\leq C\left(\frac{ne}{s}\right)^s \frac{(s^2/2)^{s\log\log n}p^{s\log\log n}}{(s\log\log n)^{s\log\log n}e^{-s\log\log n}} = C\left(\frac{ne}{s}\right)^s\left(\frac{s}{2\log\log n}\cdot\frac{ec\log n}{n}\right)^{s\log\log n}.$$

Reorganizing we have

$$= C\exp\left(s[\log(ne) - \log s] + s\log\log n\left[\log s + \log\left(\frac{ec\log n}{2n\log\log n}\right)\right]\right).$$

Differentiating the exponent with respect to s we have

$$\log(ne) - \log s - 1 + \log\log n[\log s + \log(ec\log n) - \log(2n\log\log n)] + \log\log n.$$

When $1 \leq s \leq n/(c\log n)$ this is negative, so the worst case is $s = 1$. In this case the quantity of interest is

$$= \exp\left(\log(ne) + \log\log n[\log(ec\log n) - \log(2n\log\log n)]\right)$$

which tends to 0 as $n \to \infty$.

To bound $e(S, S^c)$ we note that $e(S, S^c) = d(S) - e(S, S) \geq s\delta\log n - s\log\log n \geq s(\delta/2)\log n$ when n is large. $\text{vol}(S) \leq 4s\log n$, so for sets in A we have $e(S, S^c)/\text{vol}(S) \geq \delta/8$. ∎

6.6.1. Mixing time changes from connected to sparse

Fountoulakis and Reed (2008) proved results about the mixing times of random walks on Erdős-Rényi graphs that over the entire range from $(1+\epsilon)/n$ to $(2\log n)/n$. In the first result $p(n)$ is the probability an edge is present and $d(n) = np(n)$ is the average degree.

Theorem 6.6.4. *For every $p = p(n)$ with $d(n) - \log n \to \infty$ we have*

$$\left| T_{mix}(G_{n,p}) - \frac{\log n}{\log d} \right| \leq 3.$$

Letting $H_{n,p}$ be the giant component of $G_{n,p}$ and writing a.a.s. (asyptotically almost surely) instead of w.h.p.

Lemma 6.6.5. *For $p < (\log n)/5n$, $H_{n,p}$ a.a.s. contains paths of length more than $(\log n)/d$ all of whose interior vertices have degree two.*

To state the next two results we need some notation. Let

$$\bar{P}^t_{x_0} = (1/t) \sum_{m=1}^{t} P^m(x_0, \cdot) \qquad T'_{mix}(G) = \sup_{x_0} \min\{t : d_{TV}(\bar{P}^t_{x_0}, \pi) < 1/e).$$

Theorem 6.6.6. *For every $p(n)$ with $\sqrt{(\log n) \log\log n} \leq np(n) \leq 2\log n$ we have*

$$\left| T'_{mix}(G) - \frac{\log n}{\log d} \right| = O((\log n/d)^2)$$

Theorem 6.6.7. *For every $p(n)$ with $1 + \epsilon \leq np(n) \leq \sqrt{(\log n) \log\log n}$ we have*

$$|T'_{mix}(G)| = O((\log n/d)^2).$$

To explain why the transition between the last two results happens at $p = \sqrt{(\log n) \log\log n}$ we note that for this value of p

$$(\log n/d)^2 = \frac{\log n}{\log\log n} \sim \frac{\log n}{\log d}.$$

6.7 The Cutoff Phenomenon

A sequence of finite state Markov chains is said to exhibit **cutoff** if its distance from the stationary distribution π drops from 1 to 0 over a window that is $o(t_n)$ where t_n is the time to converge to equilibrium. To make this notion precise we need some more definitions. Let

$$d_n(t) = \max_{x \in \mathcal{S}} \|P_x(X_t \in \cdot) - \pi\|_{TV}$$

be the distance from equilibrium and let

$$t^n_{mix}(\epsilon) = \min\{t : d_n(t) < \epsilon\}.$$

We say there is **cutoff** if for any $\epsilon > 0$

$$\lim_{n\to\infty} t^n_{mix}(\epsilon)/t^n_{mix}(1-\epsilon) \to 1.$$

We say there is **cutoff at** t_n **with window size** w_n if

$$\lim_{\lambda\to\infty} \liminf_{n\to\infty} d_n(t_n - \lambda w_n) = 1,$$
$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} d_n(t_n + \lambda w_n) = 0.$$

A nice introduction to cutoff is given in Diaconis (1996), an "Inaugural Article" written on the occasion of his election to the National Academy of Science in April 1995. What follows is our rendition of his work, which lacks much of the information presented there and the style of the original. We begin with an example of a Markov chain that does not have cutoff.

Example 6.7.1. Simple random walk on the integers modulo n. To avoid problems with periodicity the transition probability has

$$p(i, i+1) = p(i, i) = p(i, i-1) = 1/3$$

where the arithmetic is done modulo n so that $p(0, n-1) = p(n-1, 0) = 1/3$. The stationary distribution in uniform $\pi(i) = 1/n$. It is easy to show, see e.g., Diaconis (1986).

Theorem 6.7.2. $\|P^k_0 - \pi\| \sim Ck/n^2$.

Here and in the next three examples we have changed notation so that $P^k_x = P_x(X_k \in \cdot)$. It is clear from the central limit theoerm that the total variation distance is ≈ 1 when $k = o(n^2)$, and that we need $k \gg n^2$ for the normal distribution of the position of the walk on $\mathbb{Z}$ to flatten out enough so that when reduced modulo n it is almost uniform.

It is surprising (to me at least) that cutoff occurs in

Example 6.7.3. Ehernfest chain. There are two urns and d balls. In the simplest version on each step one ball is chosen at random and transferred to the other urn. However to avoid periodicity, we do nothing with probability $1/(d+1)$.

$$p(i,i) = \frac{1}{d+1} \qquad p(i,i-1) = \frac{i}{d+1} \qquad p(i,i+1) = \frac{d-i}{d+1}.$$

Theorem 6.7.4. *For the Enrehfest chain started at 0, if* $k = \frac{1}{4}(d+1)(\log d + \theta)$ *then*

$$\|P_0^k - \pi\| \le \frac{1}{\sqrt{2}}(\exp(e^{-\theta}) - 1)^{1/2}$$

while if $k = (1/4)d(\log d - \theta)$ *then the total variation distance tends to 1 as* $d, \theta \to \infty$.

Example 6.7.5. Random transpositions. This example considered by Diaconis and Shahshahani (1981) was perhaps the first example where a sharp cutoff was demonstrated. Picture n cards at random on a table. At each time your left and right hands choose cards independently (so left=right with probability $1/n$) and then the cards positions are switched. The stationary distribution is uniform over all the $n!$ arrangements

Theorem 6.7.6. *Let* $k = (1/2)n(\log n + \theta)$ *with* $\theta > 0$, *For any starting position* x,

$$\|P_x^k - \pi\| \le Ae^{-\theta/2}.$$

To explain the answer, suppose we start with the identity permutation. The expected number of fixed points of a randomly chosen permutation is 1, and the distribution is roughly Poisson. By results for the coupon collectors problem, see Example 6.7.10 for more detail, it takes time $\sim (n/2)\log(n)$ until we have moved every card at least once. If there are a lot of cards that have not moved the chain is not in equilibrium because there are too many fixed points.

Example 6.7.7. Riffle shuffles are an approximation of how humans, that are not magicians by training, shuffle cards. The deck is cut into two pieces according to a symmetric binomial distribution. Then the two halves are riffled together according to the following rule. If the pile in the left hand has A cards and the one in the right hand has B cards then the next card drops from the left pile with probability $A/(A+B)$ and from the right pile with probability $B/(A+B)$. Let $p(x,y)$ be the transition probability of this shuffling method, which is called the **Gilbert-Shannon-Reeds riffle shuffle**. Bayer and Diaconis (1992) proved

Theorem 6.7.8. *Let* $k = (3/2)\log_2 n + \theta$. *Then*

$$\|P_x^k - \pi\| = 1 - 2\Phi(-2^{-\theta}/4\sqrt{3}) + O(1/\sqrt{n})$$

where Φ *is the standard normal distribution function.*

There are by now many examples. We will add two more from Levin and Peres (2017) that are closely related to the first four and then leave it to the reader to explore the subject further. For the next two example references such as the one to Proposition 7.14 are to their book.

Example 6.7.9. Random walk on the hypercube. The state space is $\{0,1\}^n$. The dynamics are simple. We pick a coordinate at random and flip it. To avoid the problems of periodicity we consider the lazy walk that does nothing with probability 1/2.

This example is used throughout the book by Levin and Peres (2017). Their Proposition 7.14 shows that

$$t_{mix}(1-\epsilon) \geq \frac{1}{2} n \log n - c_\ell(\epsilon) n.$$

In their Example 12.19 using the known eigenvalues and eigenfunctions for the chain, it was proved that

$$t_{mix}(\epsilon) \leq \frac{1}{2} n \log n - c_u(\epsilon) n.$$

Thus the lazy random walk on the hypercube has cut off at $(1/2) n \log n$ with a window of size $O(n)$.

Example 6.7.10. Top to random shuffle. Suppose to simplify the discussion that initially the Ace of spades (A♠) is at the bottom of the deck. It will gradually rise to the top as cards are inserted blow it. It is easy to show by induction that when there are k cards below the A♠ they are in random order. Thus if we let τ_n be the time at which the A♠ is inserted into a random position in the deck, the system is in equilibirum.

In the terminology of Chapter 6 of Levin and Peres (2017) τ_n is a **stationary time**. It has the additional property that

$$P_x(\tau_n = t, X_\tau = y) = P_x(\tau_n = t)\pi(y)$$

so it is a **strong stationary time**, i.e., the chain is stationary even if we condition on when τ occurs. Proposition 6.11 says that if τ is a strong stationary times then

$$d_n(t) \equiv \|P^t(x,\cdot) - \pi\|_{TV} \leq P(\tau_n > t).$$

In the case under consideration τ_n is a coupon collectors random variable. It is a sum of independent geometric(k/n) random variables with $k = 1, \ldots n$. It is well known that (see e.g., Example 2.2.7 in PTE5) that

$$E\tau_n \sim n \log n \qquad \text{var}\,(\tau_n) \sim N^2 \sum_{m=1}^{\infty} m^{-2}.$$

Proposition 2.4. says that

$$P(\tau_n > n \log n + cn) \leq e^{-c}. \tag{6.7.1}$$

The proof is so easy that we go ahead and give it

Proof. Let A_i be the probability that card i has not been drawn by time $n \log n + cn$.

$$P(\tau > n \log n + cn) = P\left(\cup_{i=i}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$
$$= n \left(1 - \frac{1}{n}\right)^{n \log n + cn} \leq n \exp(-\log n - c) = e^{-c}$$

which is the desired result. □

In Section 7.4.1 they look at the time it it takes for the jth card from the bottom to reach the top in order to prove (see their Proposition 7.15).

Theorem 6.7.11. *Given $\epsilon > 0$ there is an $\alpha(\epsilon)$ so that for large n*

$$d_n(n \log n - \alpha n) \geq 1 - \epsilon.$$

This shows that there is cut off at $n \log n$ with a window of size $O(n)$.

6.8 Random regular graphs

Let $\mathcal{G}(n,d)$ be the collection of random regular graph in which all vertices have degree d. The case $d = 1$ is boring: each vertex has degree 1, so the graph is a collection of isolated edges. The case $d = 2$ is not much better: the graph is a union of circles of random sizes. In what follows, we restrict our attention to $d \geq 3$ even if we do not state that explicitly. Using Theorem 6.4.1 implies that when $d \geq 3$ the mixing time on a random d-regular graph is $\leq C \log n$.

Lubetzky and Sly (2010) have identified the constant C in the mixing time and shown that there is cutoff with a window of size $O(\sqrt{\log_{d-1} n})$.

Theorem 6.8.1. *(Theorem 1.) Let $G \sim \mathcal{G}(n,d)$ be a random regular graph in which all vertices have degree $d \geq 3$. Then the simple random walk exhibits cutoff at $t_n = (d/(d-2)) \log_{d-1} n$ with a window of size $w_n = O(\sqrt{\log n})$. Furthermore*

$$t_{mix}(s) = \frac{d}{d-2} \log_{d-1} n - \Phi^{-1}(s)(\Lambda + o(1))\sqrt{\log_{d-1} n} \tag{6.8.1}$$

where $\Lambda = 2\sqrt{d(d-1)}/(d-2)^{3/2}$.and Φ is the cumulative distribution function of the normal.

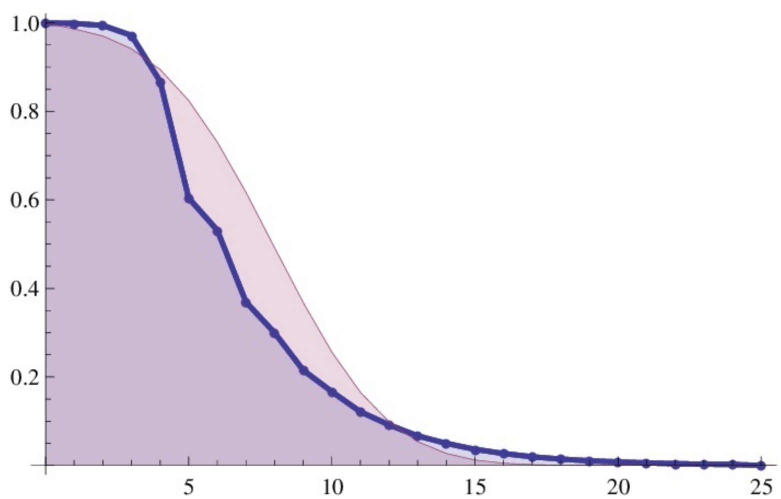

Figure 6.2: Distance from stationarity for a random 6-regular graph with 5000 vertices compared with theoretical result. Picture from Lubetzky and Sly (2010)

It is easy to guess that there is cutoff at time t_n with a window of width w_n. Seen from a fixed vertex that we will call 0, the graph G looks locally like a tree in which the root 0 has degree d while the other vertices have one edge pointing back toward 0 and $d-1$ pointing away. If we assume the approximate picture is exactly correct, then the diameter of the graph is $\log_{d-1} n$ and the distance of the random walk from the root, X_m, when it is positive, changes by $+1$ with probability $(d-1)/d$ and by -1 with probability $1/d$.

The mean of one step of the random walk (ignoring the times it is at 0) is $(d-2)/d$ so by the law of large numbers the time to distance $\log_{d-1} n$ is asymptotically

$$T_n \equiv \frac{d}{d-2} \log_{d-1} n.$$

Clearly we cannot be in equilibrium at time $(1-\epsilon)T_n$ since there are only $O(n^{1-\epsilon})$ vertices reachable in that time and the stationary distribution is uniform.

The more precise result given in (6.8.1) comes by reasoning that we cannot be in equilibrium until the random walk has moved by more that $\log_{d-1} n$ and using the central limit theorem to evaluate that probability. The funky formula for Λ should indicate that there is more to sharp result in the displayed formula, than just computing the variance of the steps of the random walk, so my advice to the reader is to wait until the formula is done for you in (6.8.7).

Main ideas of the proof

The hard part of the proof of Theorem 1 is to show that once the walker has moved a distance larger than $\log_{d-1} n$ it is in equilibrium. The proof is based on two ideas.

I. The first important idea, which is widely known, is that the graph is **locally tree like**. To prove a result to quantify this, let

$$B_t(u) = \{v \in V : \text{dist}(u,v) \le t\} \qquad \partial B_t(u) = B_t(u) - B_{t-1}(u).$$

Let $\text{tx}(B_t)$ be the maximum number of edges that can be deleted from B_t and still keep it connected. Here tx is short for tree excess. Sharpening the conclusion in our Theorem 1.2.3 they show the following. Here and for the rest of the sections Lemmas are numbered as in their paper.

Lemma 2.1. *Let $G \sim \mathcal{G}(n,d)$ for a fixed $d \ge 3$ and let $s = \lfloor (1/5) \log_{d-1} n \rfloor$ then with high probability $tx(B_s(u)) \le 1$ for all $u \in V(G)$.*

"Proof." When we grow the clusters out to distance s there are (ignoring constants) $n^{1/5}$ vertices. In running the exploration process for time $m = n^{1/5}$ the number of collisions will be smaller than binomial$(m, m/n)$. The probability of more than 2 collisions is

$$\le \binom{m}{2} \cdot (m/n)^2 \le m^4/n^2 = O(n^{-6/5})$$

so the expected number of vertices for which this occurs is $n^{-1/5}$ and the probability of the existence of such a vertex $\to 0$.

Theorem 1.2.3 shows that for a given vertex the number of collisions before the cluster grows to size n^α is $n^{2\alpha-1}$. If $\alpha = 2/3$ then this is $O(n^{1/3})$. While there are a large number of these, vertices with this problem, which we will call bad vertices are rare. Let A_k be the probability that the random walk has stepped on a bad vertex by time k. Berestycki and Durrett

(2008) who studied random walk on a three regular graph proved that $P(A_{3(1-\epsilon)\log n}) \to 0)$, so with high probability the walk will get to time $3(1-\epsilon)\log n$ without noticing that it is not on a 3-regular tree. While the last picture is useful it breaks down badly when the walks reaches a distance equal to the diameter.

II. To analyze what happens beyond distance $\log_{d-1} n$ becomes much easier one uses the notion of the **cover tree of a graph**. Let $G = (V, E)$ be a d-regular graph and let $a \in V$. The cover tree is a mapping $\phi : \mathcal{T} \to V$ where $\mathcal{T}$ is a d-regular tree with root ρ so that $\phi(a) = \rho$ which preserves the neighborhool relationship:

$$\mathcal{N}_G(\phi(x)) = \{\phi(y) : y \in \mathcal{N}_{\mathcal{T}}(x)\}$$

where for $H = G, \mathcal{T}$, $\mathcal{N}_H(y)$ are the neighbors of y in H.

To prove Theorem 1 we need several result about the structure of the cover tree. In order to conserve the reader's energy and to make it easier to see the "big picture" we state these without proof, and refer the reader to the paper for details. To state these lemmas we need four definitions

$$K = \lfloor \log_{d-1} \log n \rfloor, \qquad R = \lfloor (4/7) \log_{d-1} n \rfloor, \qquad T = \lfloor (1/2) \log_{d-1} n \rfloor.$$

A vertex $a \in V$ is a **K-root** if and only if the induced subgraph on $B_K(a)$ is a tree.

Lemma 3.2. *Suppose that every $u \in V$ has $tx(B_{5K}(u)) \le 1$. Then for any $v \in V$ the simple random walk of length $4K$ ends in a K-root with probability $1 - o(1)$. In particular, the number of K-roots is $n - o(n)$.*

Lemma 3.3. *With high probability every K-root u satisfies*

$$|\partial B_t(u)| \ge (1 - o(1)) d(d-1)^{t-1} \quad \text{for all } t < R.$$

Let $\partial B_t^*(u)$ denote the set of vertices in $\partial B_t(u)$ with a single (simple) path of length t to u.

Lemma 3.4. *With high probability any two K-roots with $dist(u, v) > 2K$ satisfy*

$$|\partial B_t^*(u) - B_{t+1}(v)| = (1 - o(1)) d(d-1)^{t-1} \quad \text{for all } t < R - 1.$$

Let $\mathcal{S}_k$ be the number of simple paths of length k between u and v.

Lemma 3.5. *With high probability any two K-roots with $dist(u, v) > 2K$ satisfy*

$$\mathcal{S}_{2T+t}(u, v) \ge (1 - o(1)) \frac{1}{n} d(d-1)^{2T+t-1}$$

for all $2K \le \ell \le (1/20) \log_{d-1} n$

Proof of Theorem 6.8.1. By Lemma 3.2 after K steps whp the random walk is at a K-root. Since we are only seeking to establish t_{mix} to an accuracy of $o(\sqrt{\log_{d-1} n})$ and since $K = o(\sqrt{\log_{d-1} n})$ it is enough to suppose we are starting at a K-root.

By Lemma 3.5

$$\mathcal{S}_{2T+t}(u,v) \geq (1-o(1))\frac{1}{n}d(d-1)^{2T+t-1} \qquad \text{for all } 2K \leq \ell \leq (1/20)\log_{d-1} n.$$

Since each simple path in G corresponds to a simple path in the cover tree $\mathcal{T}$

$$\begin{aligned} &|\{w \in \mathcal{T} : \phi(w) = v, \operatorname{dist}(\rho, w) = 2T+\ell\}| \\ &\qquad \geq \mathcal{S}_{2T+\ell}(u,v) \geq \frac{1-o(1)}{n} d(d-1)^{2T+t-1} \end{aligned} \tag{6.8.2}$$

when $2K \leq \ell \leq (1/20)\log_{d-1} n$.

Let X_t be a simple random walk on $\mathcal{T}$ started from ρ and let $W_t = \phi(X_t)$ be the corresponding SRW on G started from u. Note that by symmetry. the random walk conditioned on $\operatorname{dist}(\rho, X_t) = k$ is uniform on the $d(d-1)^{k-1}$ points at distance k from ρ in $\mathcal{T}$

A random walk on $\mathcal{T}$ with $d \geq 3$ is transient so it returns to ρ only finitely many times. If $X_t \neq \rho$ then on the next step the change the distance from ρ increases by 1 with probability $(d-1)/d$ and decreases by 1 with probability $1/d$. Let ξ be a random variable with this distribution. $E\xi = (d-2)/d$ and

$$\operatorname{var}(\xi) = 1 - (E\xi)^2 = 1 - \frac{d^2-4d+4}{d^2} = \frac{4(d-1)}{d^2}.$$

Therefore the central limit theorem gives

$$\operatorname{dist}(X_t,\rho) = \frac{t(d-2)/d}{\sqrt{t}\cdot 2\sqrt{d-1}/d} \Rightarrow \operatorname{normal}(0,1). \tag{6.8.3}$$

Let A be the set of vertices that are K roots and whose distance from u is $\geq 2K$. Since the number of vertices with distance $2K$ is $o(n)$ it follows from by Lemma 3.2 that $|A| \geq n - o(n)$. If $v \in A$ and

$$t = \left\lfloor \frac{d}{d-2}\log_{d-1} n + k\sqrt{\log_{d-1} n} \right\rfloor. \tag{6.8.4}$$

There are many points $w \in \mathcal{T}$ with $\phi(w) = v$ so

$$P(W_t = v) = \sum_{j=0}^{t} P(\operatorname{dist}(\rho, X_t) = j)\frac{|\{w \in \mathcal{T} : \phi(w) = v, \operatorname{dist}(\rho,w) = j\}|}{d(d-1)^{j-1}}.$$

Recalling $2T = 2\lfloor (1/2)\log_{d-1} n\rfloor \approx \log_{d-1} n$ and using (6.8.2) the above

$$\begin{aligned} &= \sum_{\ell=2K}^{(1/20)\log_{d-1} n} P(\operatorname{dist}(\rho, X_t) = 2T+\ell)\cdot\frac{1+o(1)}{n} \\ &= \frac{1+o(1)}{n}\cdot P(\log_{d-1} n + 2K \leq \operatorname{dist}(\rho, X_t) \leq 1.05\log_{d-1} n) \\ &= \frac{1+o(1)}{n}\cdot P(\operatorname{dist}(\rho, X_t) - \log_{d-1} n \geq 0). \end{aligned}$$

Since in the second line $K = o(\sqrt{\log_{d-1} n})$ and the other inequality holds with probability $\to 1$. Substituting in the value of t

$$\text{dist}(\rho, X_t) - t(d-2)/d = \log_{d-1} n + k\sqrt{\log_{d-1} n} \cdot (d-2)/d$$

which implies $P(\text{dist}(\rho, X_t) - \log_{d-1} n \geq 0)$ is

$$= P\left(\frac{\text{dist}(\rho, X_t) - t(d-2)/d}{\sqrt{t} \cdot 2\sqrt{d-1}/d} \geq \frac{-k\sqrt{\log_{d-1} n} \cdot (d-2)/d}{\sqrt{t} \cdot 2\sqrt{d-1}/d}\right). \tag{6.8.5}$$

Using the central limit theorem in (6.8.3) we have

$$P(\text{dist}(\rho, X_t) - \log_{d-1} n \geq 0) \approx 1 - \Phi(-k/\Lambda) \tag{6.8.6}$$

where Φ is the distribution function of the standard normal and

$$\Lambda = \sqrt{\frac{d}{d-2}} \cdot \frac{2\sqrt{d-1}/d}{(d-2)/d} = \frac{2\sqrt{d(d-1)}}{(d-2)^{3/2}}. \tag{6.8.7}$$

to get to this easily start by flipping the right hand side of (6.8.5) and then use the definition of t in (6.8.4).

To complete the proof now recall that A is the set of vertices tha are K-roots and we use (6.8.6)

$$\begin{aligned}
\|P(W_t \in \cdot) - \pi\| &= \sum_{v \in V} \max\left(\frac{1}{n} - P(W_t = v), 0\right) \\
&\leq \frac{n - |A|}{n} + \sum_{v \in A} \max\left(\frac{1}{n} - P(W_t = v), 0\right) \\
&= o(1) + (1 + o(1)) \cdot \frac{1}{n}\Phi(-k/\Lambda).
\end{aligned}$$

At this point the proof of the matching lower bound is not very hard but again we refer the reader to the paper for details.

6.9 Random walk on Galton-Watson trees

In this section we describe results of Lyons, Pemantle, and Peres (1995). Consider a supercitical Galton-Watson branching process in which each individual has k children with probability p_k, the generating function $f(s) = \sum_{k=0}^{\infty} p_k s^k$ and the mean number of children $\mu = f'(1) > 1$. Started with a single progenitor, on the event of nonextinction the process yields a random infinte family tree T called a **Galton-Watson tree**. The focus of the paper by LPP is on the asymptotic properties of simple random walk on T that on each step jumps from a vertex to a randomly chosen neighbor. To simplify things we will assume in the first three parts of this section that $p_0 = 0$. In the fourth on biased random walks, the situation is more interesting when $p_0 > 0$.

6.9.1 Transience

The question of transience of random walks on Galton-Watson trees first arose in work of Grimmett and Kesten (1984) on "random electrical networks." They considered a complete graph on $n + 2$ vertices in which the resistance of an edge has

$$P(R = \infty) = 1 - \frac{\gamma(n)}{n} \qquad P(R \le x) = \frac{\gamma(n)}{n} F(x)$$

where F is a fixed distribution function. If $\gamma(n) \to \gamma \in (0, \infty)$ then the connected set of edges with $R(e) < \infty$ that contain a fixed vertex converges to a Galton-Watson process with a Poisson(γ) offspring distribution.

They proved that if $\gamma(n) \ge n^\beta$ for some $\beta > 0$ (and hence the probability the graph of finite resistance edges is connected tends to 1) then effective reistance $\mathcal{R}_n$ between two random chosen vertices satisfies

$$\gamma(n)\mathcal{R}_n \to 2 \left(\int_0^\infty x^{-1}\, dF(x) \right)^{-1} \quad \text{in probability.}$$

They stated two results for the case $\gamma(n) \to \gamma$.

Theorem 6.9.1. *If $\gamma < 1$ then $P(\mathcal{R}_n = \infty) \to 1$.*

Theorem 6.9.2. *If $\gamma > 1$ then $\mathcal{R}_n$ converges to a limit $\mathcal{R}_\infty$ with*

$$P(\mathcal{R}_\infty = \infty) = 2q(\gamma) - q(\gamma)^2$$

where $q(\gamma)$ is the extinction probability of the Poisson(γ) branching process.

The work was done while Grimmett was visiting Cornell in 1983, but only in 2001 were the missing proofs put on the arXiv.

In 1990 Russ Lyons published a ground breaking paper on random walks and percolation on a class of trees much more general than Galton-Watson trees. To be precise, he proved results for trees with a well-defined **branching number**, a growth rate inspired by a method of Furstenberg used to compute Hausdorff dimensions of sets. In the case of Galton-Watson tree, where the branching number is just the mean, his result gives

Theorem 6.9.3. *If we assume that $\mu = \sum_k kp_k > 1$ and condition on the event that the branching process does not die out, then the random walk on the resulting tree is transient with probability one*

Proof. To prove transience using Terry Lyons' result given in Theorem 6.2.5 we will construct a flow with finite energy. To construct our flow, we will pick K so that $\sum_{k=1}^{\infty} p_k \min\{k, K\} > 1$ and modify the offspring distribution so that $\bar{p}_K = 1 - \sum_{j=1}^{k-1} p_j$ and $\bar{p}_j = p_j$ for $j < K$. From the electrical networks point of view, the truncated branching process yields a tree that is less transient because it has higher resistance. See Theorem 6.2.2.

By arguments in the previous paragraph, it suffices to prove our result when he distribution is bounded or more generally when $\sum_k k^2 p_k < \infty$. In this case if we let B_n be the vertices at distance n from the root, $Z_n = |B_n|$ then $Z_n/\mu^n \to W$ in L^2, so $EW^2 = 1$. We will call this value $W(\rho)$ and set $\theta(\rho) = W(\rho)$. Given a site $x \in B_n$, let x^+ be the tree consisting of x and the descendants of x. We define a flow by $\theta(x) = W(x^+)/\mu^n$ where $W(x^+)$ is the limit random variable $\lim_{m\to\infty} Z_m(x^+)/\mu^m$ for the tree x^+. To check the flow property we note that

$$\theta(x) = \frac{1}{\mu} \sum_{z:x\to z} \theta(z)$$

and there is a positive flow from the root ρ.

To check (iii) in Theorem 6.2.5 note that $Q(x, y) \equiv 1$, so taking expected value

$$E \sum_{(y_x,x)} [W(x^+)/\mu^n]^2 = \sum_{n=1}^{\infty} 1/\mu^n < \infty$$

so for almost every Galton-Watson tree the sum is finite. □

Remark 6.9.4. The proofs above generalize easily to the **homesick random walk** in which jumps back toward the root from vertices with k children have probability $\lambda/(\lambda + k)$ where $\lambda > 1$. In this case $Q(y_x, x) = \lambda^{-n}$ when $x \in B_n$, so the walk is transient if $\lambda < \mu$.

6.9.2 Escape rate

In a Galton-Watson tree each vertex has k descendants with probability p_k. Lyons, Pemantle and Peres (1995) studied the asymptotic behavior of the discrete time random walk X_n that jumps to each of the neighbors with equal probability.

Theorem 6.9.5. *If $\mu = \sum_{k=1}^{\infty} kp_k \in (1, \infty)$ and $p_0 = 0$, then*

$$\frac{|X_n|}{n} \to \ell \equiv \sum_k \frac{k-1}{k+1} p_k \tag{6.9.1}$$

Proof. The limiting speed is remarkably simple. If we are at the root ρ then the distance from the root will be 1 the next time. Other vertices with k children have 1 parent so the drift at these vertices is $(k - 1)/(k + 1)$.

The proof in Lyons, Pemantle, and Peres (1995) is only for a sophisticated reader. For people like me it is fortunate that Lyons and Peres (2017) have a proof that stays close to the intuition. Let X_n be the location on the tree at time n and define the drift at x by

$$\delta(x) = \begin{cases} 1 & x = \rho \\ \frac{k-1}{k+1} & x \neq \rho, \deg(x) = k+1 \end{cases}$$

From this it is clear that if $|X_n|$ is the distance from the root then

$$M_n = |X_n| - \sum_{m=0}^{n-1} \delta(X_m) \quad \text{is a martingale.}$$

Since jumps of the martingale are bounded by $K = 2$, we have a very good bound on the deviations from the mean. See Theorem **??**.

Lemma 6.9.6. Azuma-Hoeffding inequality.

$$P(M_n - M_0 \geq L) \leq \exp\left(-\frac{L^2}{2nK}\right)$$

The last observation is extremely useful but we still need to do some work to show that the fraction of time spent at vertices with $\delta(x) = (k-1)/(k+1)$ children is p_k. In order to do this, we need to find a stationary distribution for the **environment process**, i.e., the tree as seen from the current vertex. As we have defined things the root of the Galton-Watson tree is different from the other vertices since it has stochastically one fewer neighbos. To remedy this defect we consider what LPP call the **augmented Galton-Waston tree measure**, or **AGW**. This is obtained by added another neighbor z of the root and an independent Galton-Watson tree T' rooted at z. The new neighbor is the missing parent of the root, but we will not interrupt the proof to complete that thought.

To define the dynamics of the environment process, we declare that when the random walks jumps to a randomly chosen neighbor x of the root(T) then the environment chain jumps to MoveRoot(T, x), which, as the notation should suggest, is the tree T with the root moved to be at x. Theorem 3.1 in LPP shows:

Theorem 6.9.7. *The environment chain with initial distribution* **AGW** *is stationary and reversible.*

We declare this to be obvious and refer the reader to LPP for a proof. From Theorem 6.9.7 it follows that on **AGW** $\delta(X_n)$ is stationary. A result in Section 8.1 of LPP implies that it is ergodic, so an application of the ergodic theorem completes the proof. □

6.9.3 Dimension drop

The developments here require mathematics outside the author's skill set, so we will only describe what is true and refer the reader to LPP for proofs.

Theorem 6.9.8. *The Hausdorff dimension of harmonic measure on the boundary of a non-degenerate Galton-Watson tree T is almost surely a constant $\mathbf{d} < \log \mu = dim(\partial T)$ there is a Borel subset of ∂T of full harmonic measure and dimension d.*

The next result explains our interest in that conclusion. In it $|\Gamma^n|$ is the number of vertices in Γ on level n.

Theorem 6.9.9. *For any $\epsilon > 0$ and for almost every Galton-Watson tree T, there is a subtree Γ of T of growth*

$$\lim_{n\to\infty} |\Gamma^n|^{1/n} = e^{\mathbf{d}} < \mu$$

such that with probability $1 - \epsilon$ the sample path of the random walk on T is contained in Γ.

The proof of Theorem 6.9.8 gives an abstract integral formula for $\mathbf{d}$

$$\mathbf{d} = \frac{1}{\ell} \int_{s=0}^{\infty} \int_{t=0}^{\infty} \frac{\log(1+s)}{1 + s^{-1} + t^{-1}} \, dF(t) \, dF(s)$$

where F is the distribution function of the effective conductance from the root to infinite of the Galton-Watson tree. For example, this gives that $\mathbf{d} = \log 1.47$ for the tree with generating function $(s+s^2)/2$, and hence $m = 1.5$. Figure 1.2 in LPP graphs the distribution of the conductance in this case, which has a shape that is reminiscent of a stegosaurus.

6.9.4 Biased random walks

In this section we follow Lyons, Pemantle and Peres (1996), or LPP2 for short. For $\lambda > 0$ the biased random walk on a tree $\mathcal{T}$ with root ρ is the time homogeneous Markov chain X_n, $n \geq 0$ on the vertices of $\mathcal{T}$ so that if u is a vertex with children $v_1, \ldots, v_k$ and parent v_0 then the transition probability is

$$p(u,x) = \begin{cases} 1/(\lambda + k) & x = v_1, \ldots v_k \\ \lambda/(\lambda + k) & x = v_0 \end{cases}$$

Suppose that $\mathcal{T}$ is a Galton-Watson tree with offspring generating function $f(z) = \sum_k p_k z^k$, mean offspring $\mu > 1$ and let q be the extinction probability of the branching process. Lyons (1990) has shown that the biased random walk is transient if $1 < \lambda < \mu$. LPP2 show that when $1 < \lambda < \mu$ the random walk escape to infinity at a positive speed. LPP2 have results for the Hausdorff dimension of the limit set of random walk, which is again smaller than the diemnsion of the boundary. Here, we will concentrate here on results that are more surprising and easier to prove.

For example, if $p_0 > 0$ and the bias is away from the root, i.e., $0 < \lambda < 1$, then the speed is positive if and only if $\lambda > f'(q)$. The condition $p_0 > 0$ implies that the tree has dead ends. If $0 < \lambda < f'(q)$ then the probability of jumps back toward the root is not large enough so that the time to escape from a dead end has finite mean.

Speed of the random walk

Theorem 6.9.10. *(Theorem 3.1.) For $1 < \lambda < \mu$ and for almost every Galton-Watson tree conditioned on non-extinction, $\lim_{n\to\infty} |X_n|/n$ exists and is a positive constant depending only on λ and the offspring distribution. A lower bound is*

$$\frac{(1-\lambda^{-1})^3}{12} \cdot (1-q_\lambda)^2$$

where q_λ is the smallest nonneagtive number satisfying

$$f(1-\lambda^{-1}(1-q_\lambda)) = q_\lambda$$

Proof. We say that n is a**fresh epoch** if $X_n \neq X_k$ for all $k < n$ and a **regeneration epoch** if $X_{n-1} \neq X_k$ for all $k \geq n$. Finally let P_{non} be the measure of the Galton-Watson tree conditioned on non-extinction. The first step is a strong Markov property which we declare to be obvious.

Lemma 6.9.11. *(Lemma 3.2) Let A be a measurable set of infinite trees and let $\mathcal{F}_n$ be the σ-field generated by the events $\{X_i \neq X_j\}$ for $0 \leq i < j \leq n$. Let α be a stopping time with respect to $\mathcal{F}_n$ which is a fresh epoch and let $\mathcal{T}^\alpha$ be the descendant subtree of X_α. Then*

$$P_{non}(T^\alpha \in A|\mathcal{F}_\alpha) = P_{non}(T \in A)$$

Lemma 6.9.12. *(Lemma 3.3) Let $1 < \lambda < \mu$. For almost every Galton-Watson tree $\mathcal{T}$ on the set of nonextinction, and for almost every sample path of the random walk there are infinitely many regeneration epochs.*

Proof. Given the σ-field $\mathcal{F}_m$ up to a fixed time m the probability of another regeneration is bounded below by a constant > 0. □

Lemma 6.9.13. *(Proposition 3.4) For $1 < \lambda < \mu$ on the event of nonexitinction the $\{\tau_{n+1} - \tau_n, n \geq 1\}$ are i.i.d. as are the increments $\{|X_{\tau(n+1)}| - |X_{\tau(n)}|, n \geq 1\}$.*

Proof. This is also intuitively clear. However, writing out the proof requires a fair amount of notation. For details, see pages 253–255 of LPP2. □

To complete the proof, LPP2 show that $E(\tau_{n+1} - \tau_n) < \infty$ and is upper bounded by the reciprocal of the lower bound given in Theorem **??**. □

Outward biased random walks

Theorem 6.9.14. *(Theorem 4.1.) Suppose that $p_0 > 0$. Let $\mathcal{T}$ be a Galton-Watson tree conditioned on nonextinction. The speed of the bised random walk exists and is constant almost surely. It is zero if $0 \leq \lambda \leq f'(q)$ and positive $f'(q) < \lambda \leq 1$.*

Proof. The key to the proof is a decomposition of a Galton-Watson tree into the **backbone** which is the collection of individuals with an infinite line of descent, and **shrubs** that consist of the descendants of individuals whose lineage dies out and have a parent on the back bone. Theorem 1.1.9 implies that the backbone is a Galton-Watson process with offspring generating function

$$g(z) = \frac{f((1-q)z+z)-z}{q}$$

where q is the extinction probability. The shrubs are a branching process with generating function $h(z) = q^{-1}h(qz)$.

From Theorem 5.5.11 and 5.5.7 in PTE5, if we let r be the root of a shrub $\mathcal{B}$ and π is the stationary distribution then

$$E_r T_r = 1/\pi(r) = \nu_r(\mathcal{B})$$

where μ_r is the cycle trick measure based at r. Since the random walk is reversible if y is a child of x we have

$$\nu_r(x) = \lambda \nu_r(y)$$

Since $\mu_r(r) = 1$, the reversibility condition implies that if y is a distance m from the root then $\nu(y) = \lambda^{-m}$ and the total mass of the measure is

$$\mu_r(\mathcal{B}) = \sum_{m=0}^{\infty} V_m \lambda^{-m}$$

where V_m is the number of vertices at distance m. In the case of a shrub this sum has expectation

$$\mu_r(\mathcal{B}) = \sum_{m=0}^{\infty} h'(1)^m \lambda^{-m} = \begin{cases} \frac{1}{1-f'(q)/\lambda} & \lambda > f'(q) \\ \infty & \lambda \le f'(q) \end{cases}$$

When $0 < \lambda \le f'(q)$ it follows that the expected time between regeneration epochs is infinite. If $f'(q) < \lambda < 1$, then when the random walk visits x it makes a geometrically distributed number of visits to shrubs before it moves on to another vertex, In this case the random walk on the Galton-Watson tree is just a time change of the random walk on the backbone, which has an offsrping distribution with no mass at 0 and the conclusion follows from previous results. □

6.10 Sparse Erdős-Rényi graphs

In Section 6.4 we saw that if we have a random graph generated by the configuration model in which vertices of degree 2 have positive probability then the mixing time of the random walk on the giant component is at least $C\log^2 n$ since there are paths of length $\log n$ in which all vertices have degree 2. If the walk starts in the middle of such a path then it takes time $O(\log^2 n$ to escape from the set.

Berestycki, Lubetzky, Peres, and Sly (2018) have shown that if we start at a randomly chosen vertex in the giant component then the mixing time becomes $O(\log n)$ and there is cutoff with a window of $(\log n)^{1/2+o(1)}$.

Theorem 6.10.1. *Let $\mathcal{C}_1$ be the giant component of an Erdős-Rényi$(n,\lambda/n)$ random graph with $\lambda > 1$ fixed. For any $0<\epsilon<1$ whp the mixing time of the random walk starting from a randomly chosen vertex $x \in \mathcal{C}_1$ satisfies*

$$|t^v_{mix}(\epsilon) - (\nu\mathbf{d})^{-1}\log n| \le (\log n)^{1/2+o(1)}.$$

Here ν is the speed of a random walk on a Poisson(λ) Galton-Watson tree and $\mathbf{d}$ is the dimension of harmonic measure on the ends of the tree, concepts that were introduced in the previous section. In addition to the result for Erdős-Rényi they have a result for a graph G with general degree distribution p_k. Condition (6.10.1) is strong but weaker than having finite exponential moments.

Theorem 6.10.2. *Suppose that for fixed $\delta > 0$ the random variable Z given by $P(Z = k-1) \propto kp_k$ satisifes*

$$P(Z > \Delta_n) = o(1/n) \qquad \Delta_n \equiv \exp((\log n)^{1/2-\delta}) \tag{6.10.1}$$

and let $t_\star = (\nu\mathbf{d})^{-1}\log n$ where ν and $\mathbf{d}$ are the speed of the random walk and the dimension drop for the Galton-Watson tree with offspring distributon Z.

(i) Set $w_n = (\log n)^{1/2}(\log\log n)^3$. If $p_2 < 1-\delta$ and $1+\delta < EZ < K$ for all n then whp on the event that z is in the largest component of G

$$d^z_{TV}(t_\star - w_n) > 1-\epsilon) \qquad d^z_{TV}(t_\star + w_n) < \epsilon.$$

(ii) Set $w_n = (\log n)^{1/2}$. If $Z > 2$ and $EZ < K$ for all n then for any $\epsilon > 0$ there is a $\gamma < \infty$ sot aht with probability at least $1-\epsilon-o(1)$

$$d^z_{TV}(t_\star - w_n) > 1-\epsilon) \qquad d^z_{TV}(t_\star + w_n) < \epsilon.$$

The structure theorem of Ding, Lubetzky, and Peres (2014) gives an intuitive explanation. From page 3 of BLPS: "A contiguous model for the giant componet is given by (i) choosing a kernel uniformly over graphs on i.i.d. Poisson turncated to be at least

3, (ii) subdividing every edge into i.i.d. geometric variables, and (iii) hanging i.i.d. Poisson Galton-Watson trees on every vertex. Observe that steps (ii) and (iii) introduce i.i.d. delays with an exponential tail for the walk; thus starting from a uniform random vertex would essentially eliminate all but the typical $O(1)$ delays and it should mix on the kernel (whp an expander) in time $O(\log n)$"

Dimension drop results in the slow down. Note that while the kth generation of the Galton-Watson tree has size $\approx \lambda^k$, the random walk concentrates on a set of size $e^{k\mathbf{d}}$. This means that on the cover tree of the random graph the random walk has to get a distance $\mathbf{d}^{-1}\log n$ from the root before it can possibily reach all n vertices. To do this it will take time $t_\star = (\nu\mathbf{d})^{-1}\log n$.

To see what this means back on the graph, we quote Corollary 3.4 from their paper

Theorem 6.10.3. *Consider the random walk $\mathcal{X}_t$ start from a randomly chosen vertex $z \in \mathcal{C}_1$ either in the setting of Theorem 6.10.1 or 6.10.2. Let ν denote the speed of random walk on the corresponding Galton-Watson tree, and let λ be the mean of its offspring distribution. For evey fixed $a > 0$ if $t = a\log_\lambda n$ then*

$$\frac{dist(z, \mathcal{X}_t)}{\log_\lambda n} \to (\nu a \wedge 1) \quad \textit{in probability.}$$

6.11 References

Aldous, D. (1989) *The Poisson Clumping Heuristic.* Springer, New York

Aldous, D., Lovász, L., and Winkler, P. (1997) Mixing times for uniformly ergodic Markov chains. *Stoch. Proc. Appl.* 125, 408–420

Bayer, D., and Diaconis, P. (1992) Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.* 2, 37–50

Benjamini, I., Kozma, G, and Wormald, N (2014) The mixing time of the giant component of a random graph. *Rand. Struct. Alg.* 45 (3), 383–407

Berestycki, N., and Durrett, R. (2008) Limit behavior for the idstance of a random walk. *Electron. J. Probab.* 13, paper 44

Berestycki, N., Lubetzky, E., Peres, Y., and Sly, A. (2018) Random walks on the random graph. *Ann. Probab.* 46, 456-490

Cooper, C. and Frieze, A. (2003) The cover time of sparse random graphs. *Proceedings of 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, 140–147.

Diaconis, P. (1986) *Group representations in probability and statistics.* IMS Lecture Notes-Monographs series.

Diaconis, P. (1996) The cut-off phenomenon in finite Markov chains *Proc. Natl. Acad. Sci. USA* 93, 1659–1664

Diaconis, P., and Shahshahani, M. (1981) Generating a random permutation with random transpositions. *Probab. Theory Rel. Fields.* 57 (1981), 159–179

Diaconis, P., and Stroock, D. (1991) Geometric bounds for the eigenvalues of Markov chains. *Ann. Appl. Probab.* 1, 36-61

Ding, J., Lubetzky, E., and Peres, Y. (2014) Anatomy of a giant component: the strictly supercritical regime. *European J. Combinatorics.* 35, 155–168

Ding, J., Kim, J.H., Lubetzky, E., and Peres, Y. (2011) Anatomy of a young giant component in the random graph. *Rand. Struct. Alg.* 39, 139–178

Doyle, P.G., and Snell, J.L. (1984) *Random Walks and Electrical Networks.* Carus Monographs 22, MAA, Washington, D.C.

Fountoulakis, N., and Reed, B. (2007) Faster mixing and small bottlenecks. *Probab. Theory Rel. Fields.* 137, 475–486

Fountoulakis, N., and Reed, B. (2008) the evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Rand. Struct. Alg.* 33, 68–86

Gkantsis, C., Mihail, M., and Saberi, A. (2003) Conductance and congestion in power law graphs. *Proceedings of the 2003 ACM SIGMETRICS international conference on measurement and modeling of computer systems*, 148–159

Griffeath, D., and Liggett, T.M. (1982) Critical phenomena for Spitzer's reversible nearest particvle systems. *Ann. Probab.* 10, 881-895

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

Grimmett, G., and Kesten, H. (2001) Random electrical networks on complete graphs. II: proofs. arXiv:0107068

Kim, Jeong Han (2008) Poisson cloning model for random graphs. pages 873–897 in International Congress of Mathematics, Zurich European Matheamtical Society

Lawler, G., and Sokal, A. (1988) Bounds on the L^2 spectrum for Markov chainsand Markov processes: A generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.* 309, 557–580

Lovász, L., and Kannan, R. (1999) Faster mixing via average conductance. *Symposium on the Theory of Computing 1999*, 282–287

Lubetzky, E., and Sly, A. (2010) Cutoff phenomena for random walks on random regular graphs. *Duke Math. J.* 153, 475–510.

Lyons, R. (1990) Random walks and percolation on trees. *Ann. Probab.* 18, 931-958

Lyons, R., Pemantle, R., and Peres, Y. (1995) Ergodic theory on Galton-Watson trees: Speed of random walk and dimensions of harmonic measure. *Ergodic Theory Dynamical Systems.* 15, 593–619

Lyons, R., Pemantle, R., and Peres. L. (1996) Biased random walk on Galton-Watson trees *Prob. Theory Relat. Fields.* 106, 249–264

Lyons, T. (1984) A simple criterion for transience of a reversible Markov chain. *Ann. Probab.* 11, 393–402

Morris, B., and Peres, Y. (2003) Evolving sets and mixing. *Symposium on the Theory of Computing 2003*, 279–286

Nash-Williams, C. St. J. (1959) Random walks ans currents in electrical networks. *Proc. Camb. Phil. Soc.* 55, 181-194

Saloff-Coste, L. (1996) Lectures on finite Markov chains. *Ecole d'eté de probabilités de Saint-Flour XXVI.* Springer Lecture Notes in Math, Volume 1665

Sinclair, A., and Jerrum, M. (1989) Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation.* 82, 93–133