Course Notes: Probability on Trees and Networks, Fall 2004

BY YUVAL PERES

§1. Random Graphs.

Let n be a positive integer, and $0 \le p \le 1$. The Random Graph $G(n,p)$ is a probability space over the set of graphs on n vertices in which each of the possible $\binom{n}{2}$ edges appear with probability p independently of all other edges. For a graph G, denote by $C_1(G)$ the size of its largest connected component. We will show that $C_1(G)$ experiences a phase transition when $p = \frac{\lambda}{n}$ for fixed λ. More precisely, when $\lambda < 1$ we have $|C_1(G(n,p))| = O(\log(n))$, when $\lambda > 1$ we have $|C_1(G(n,p))| = \Theta(n)$ and when $\lambda = 1$ we have that $|C_1(G(n,p))|n^{-2/3}$ has a non-trivial limiting distribution. Before proving this, we will need some preparations.

We first prove a large deviation inequality due to Hoeffding (1963), also known as the Hoeffding-Azuma inequality. We follow the exposition by Steele (1997).

THEOREM 1.1. *Let $\{X_1, \ldots, X_n\}$ be bounded random variables such that*

$$\mathbf{E}\Big[X_{i_1} \cdots X_{i_k}\Big] = 0 \qquad \forall \quad 1 \le i_1 < \ldots < i_k \,,$$

(for instance, independent variables with zero mean or martingale differences). Then

$$\mathbf{P}\left[\sum_{i=1}^n X_i \ge L\right] \le e^{-L^2/\left(2\sum_{i=1}^n \|X_i\|_\infty^2\right)} \,.$$

Proof. For any sequences of constants $\{a_i\}$ and $\{b_i\}$, we have

$$\mathbf{E}\Big[\prod_{i=1}^n (a_i + b_i X_i)\Big] = \prod_{i=1}^n a_i \,. \tag{1.1}$$

Since the function $f(x) = e^{ax}$ is convex on the interval $[-1,1]$, it follows that for any $x \in [-1,1]$

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

If we now let $x = X_i/\|X_i\|_\infty$ and $a = t\|X_i\|_\infty$ we find

$$\exp\left(t\sum_{i=1}^n X_i\right) \le \prod_{i=1}^n \left(\cosh(t\|X_i\|_\infty) + \frac{X_i}{\|X_i\|_\infty}\sinh(t\|X_i\|_\infty)\right) \,.$$

When we take expectations and use (1.1), we find

$$\mathbf{E}\exp\left(t\sum_{i=1}^{n}X_i\right)\le\prod_{i=1}^{n}\cosh(t\|X_i\|_\infty)\,,$$

so, by the elementary bound $\cosh x=\sum_{k=0}^{\infty}\frac{x^{2k}}{(2k)!}\le\sum_{k=0}^{\infty}\frac{x^{2k}}{2^k k!}=e^{x^2/2}$, we have

$$\mathbf{E}\exp\left(t\sum_{i=1}^{n}X_i\right)\le\exp\left(\frac{1}{2}t^2\sum_{i=1}^{n}\|X_i\|_\infty^2\right)\,.$$

By Markov's inequality and the above we have that for any $t>0$

$$\mathbf{P}\left[\sum_{i=1}^{n}X_i\ge L\right]=\mathbf{P}\left[\exp\left(t\sum_{i=1}^{n}X_i\right)\ge e^{Lt}\right]\le e^{-Lt}\exp\left(\frac{t^2}{2}\sum_{i=1}^{n}\|X_i\|_\infty^2\right)\,,$$

so letting $t=L(\sum_{i=1}^{n}\|X_i\|_\infty^2)^{-1}$ we obtain the required result. ∎

We will also need the following:

LEMMA 1.2. *Let* $0<\lambda<1$, $X\sim\mathrm{Bin}(n-1,\frac{\lambda}{n})$ *and* $X^*\sim\mathrm{Poisson}(\lambda)$, *then for any* $k\ge 0$,

$$\mathbf{P}[X\ge k]\le\mathbf{P}[X^*\ge k]\,.$$

REMARK 1.3. In this case we say that X^* *stochastically dominates* X.

Proof. For $1\le j\le n-1$ define I_j to be an indicator random variable taking 1 with probability $\frac{\lambda}{n}$, and define I_j^* to be a Poisson random variable with mean $\frac{\lambda}{n-1}$, all independent. So,

$$X\sim\sum_{j=1}^{n-1}I_j,\qquad X^*\sim\sum_{j=1}^{n-1}I_j^*\,.$$

Thus, it suffices to prove that I_j^* stochastically dominates I_j, which we leave as a simple calculus exercise for the reader (see Exercise 1.1). ∎

For any vertex v, denote by $C(v)$ the connected component of $G=G(n,p)$ that contains v. We define a procedure which finds $C(v)$ for a given v. In this procedure, vertices will be either dead, live or neutral. At each time unit t, let Y_t be the number of live vertices and N_t be the number of neutral vertices. At time $t=0$, $Y_0=1$, v is live and all other vertices are neutral. At each time unit t we take a live vertex w and check all pairs $\{w,w'\}$, w' neutral, for membership in G. If $\{w,w'\}\in E(G)$ we make w' live, otherwise it stays neutral. After searching all neutral w' we set w dead and let Y_t equal the new number of live vertices. When there are no more live vertices the process ends and $C(v)$ is the set of dead vertices. After this process ends, we choose an arbitrary vertex and continue in the same fashion. This motivates the definition of the following process $\{Y_t\}_{t=0}^{\infty}$. For convenience define $N_t = n - t - Y_t$ and let

$$Y_0 = 1\,, \qquad Y_t = Y_{t-1} - 1 + \begin{cases} \mathrm{Bin}(N_{t-1}, p), & N_{t-1} > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Under this definition, for each t in which $N_t > 0$ we have that Y_t counts the number of currently active vertices minus the number of components discovered up to time t. Observe that by induction, Y_t is distributed as $\mathrm{Bin}(n-1, 1-(1-p)^t) + 1 - t$. Moreover, let τ_0 be the stopping time

$$\tau_0 = \min_t \{Y_t = 0\}\,,$$

then τ_0 is $|C(v)|$. Similarly, if we define by induction for $k > 0$ the stopping times τ_k by

$$\tau_k = \min\{Y_t = -k : t > \tau_{k-1}, N_t \geq 0\}$$

then if $\tau_k < \infty$ then $\tau_k - \tau_{k-1}$ is the size of the $(k+1)$th component discovered in the procedure.

THEOREM 1.4. *Let $\lambda < 1$, then a.a.s. $|C_1(G(n,p))| = O(\log(n))$. More precisely, for any $\epsilon > 0$ we have*

$$\mathbf{P}\Big[|C_1(G(n,p))| > \frac{(1+\epsilon)\log n}{|\log\lambda - \lambda + 1|}\Big] \to 0\,,$$

as $n \to \infty$.

Proof. For any $1 \leq t \leq n$ define i.i.d. random variables $X_t^* \sim \mathrm{Poisson}(\lambda)$, $Y_0^* = 1$ and $Y_t^* = Y_{t-1}^* + X_t^* - 1$. Clearly $Y_t^* \sim \mathrm{Poisson}(\lambda t) - (t-1)$. Also, since $N_t \leq n-1$ for all t, the variable Y_t^* stochastically dominates Y_t, so

$$\mathbf{P}[|C(v)| > t] \leq \mathbf{P}[Y_t > 0] \leq \mathbf{P}[Y_t^* > 0] = \mathbf{P}[\mathrm{Poisson}(\lambda t) > t-1] = \sum_{k \geq t} \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

$$\leq e^{-\lambda t}\frac{(\lambda t)^t}{t!} \sum_{k \geq 0} \frac{(\lambda t)^k}{t^k} \leq e^{-\lambda t}\frac{(\lambda t)^t}{t!}\frac{1}{1-\lambda}\,.$$

Since $t! \geq (t/e)^t$ for all t, it follows that

$$\mathbf{P}[|C(v)| > t] \leq \frac{1}{1-\lambda} e^{(\log\lambda - \lambda + 1)t}\,.$$

Thus, by the union bound, for any $\epsilon > 0$,

$$\mathbf{P}\left[\exists v : \quad |C(v)| > \frac{(1+\epsilon)\log n}{|\log\lambda - \lambda + 1|}\right] \leq n^{-\epsilon} \to 0\,,$$

which concludes the proof. ∎

Theorem 1.5. *Let $\lambda > 1$, then a.a.s. $|C_1(G(n,p))| = \Theta(n)$. In particular, if β is the unique positive solution of $1 - e^{-\lambda\beta} = \beta$, then for any $\epsilon > 0$*

$$\mathbf{P}\Big[\Big|\frac{|C_1(G(n,p))|}{\beta n} - 1\Big| > \epsilon\Big] \to 0\,,$$

as $n \to \infty$.

Proof. Fix $\epsilon, \epsilon' > 0$. We will show that

$$\mathbf{P}\Big(\bigcup_{t > (\beta+\epsilon)n} Y_t > -\epsilon' n\Big) < e^{-cn}\,, \tag{1.2}$$

for some $c > 0$, and that

$$\mathbf{P}\Big(\bigcup_{\log n < t < (\beta-\epsilon)n} Y_t < 0\Big) < n^{-c}\,. \tag{1.3}$$

Hence, with high probability we find a component with at least $(\beta-\epsilon)n - \log n$ vertices and no more than $(\beta+\epsilon)n$ vertices. Moreover, since after finding this component we are left with $m < (1-\beta+\epsilon)n$ unexplored vertices, the remaining graph is distributed as $G(m,p)$ and $mp < \lambda(1-\beta+\epsilon)$. Since the generating function $f(s) = e^{\lambda(s-1)}$ is convex and $1-\beta$ is a fixed point, it follows by Rolle's theorem that the derivative of f at $1-\beta$ is smaller than 1. This yields $\lambda(1-\beta+\epsilon) < 1$, for small enough ϵ, and so by Theorem 1.4, the remaining graph has components of size at most $\Theta(\log n)$, and indeed the analysis gives the maximal component.

The proof of (1.2) and (1.3) are based on Theorem 1.1 and on the fact that Y_t is distributed as $\mathrm{Bin}(n-1, 1-(1-p)^t) + 1 - t$. To prove (1.2) let $t = \alpha n$ where $\alpha \geq \beta + \epsilon$, estimate $1-(1-p)^t \geq 1 - e^{-\alpha\lambda}$ and observe that for $\alpha > \beta$ we have that $1 - e^{-\alpha\lambda} < \alpha$, so for small enough ϵ' (which depends only on ϵ),

$$\mathbf{P}(Y_t > -\epsilon' n) \leq \mathbf{P}\Big(\mathrm{Bin}(n-1, 1-e^{-\alpha\lambda}) > (\alpha-\epsilon')n - 1\Big) \leq e^{-ct}\,,$$

for some fixed $c > 0$, by Theorem 1.1.

To prove (1.3) again write $t = \alpha n$ (note that here α can depend on n) and estimate $1-(1-p)^t \geq 1 - e^{-\alpha\lambda}$. Observe that for $\alpha < \beta$ we have $1 - e^{-\alpha\lambda} > \alpha$ so

$$\mathbf{P}(Y_t < 0) \leq e^{-c\alpha n}$$

for some fixed $c > 0$, by Theorem 1.1. By summing over t, and recalling that $\alpha > \log n/n$ we get (1.3). ∎

THEOREM 1.6. *If $p = \frac{1}{n}$, then for any $b > 0$ we have*

$$\lim_{a\to 0}\lim_{n\to\infty} \mathbf{P}\Big[\exists \text{ component of size } \in (an^{2/3}, bn^{2/3})\Big] = 1\,,$$

as $n \to \infty$.

THEOREM 1.7. *If $p = \frac{1}{n}$, then for any $A > 0$ and $n > 7$ we have*

$$\mathbf{P}[|C_1(G(n,p))| > An^{2/3}] \le \frac{4}{A^2}.$$

Before proving the theorems we first give a proof of Cayley's formula for the number of labelled trees which is due to Joyal (see Gessel and Stanley 1995). Denote $[n] = \{1, \dots, n\}$.

THEOREM 1.8. *The number of distinct labelled trees on $\{1, \dots, n\}$ is n^{n-2}.*

Proof. First note that because every permutation can be represented as a product of disjoint directed cycles, it follows that for any finite set S the number of sets of cycles of elements of S (each element appearing exactly once in some cycle) is equal to the number of linear arrangements of the elements of S. The number of functions from $[n]$ to $[n]$ is clearly n^n. To each such function f we may associate its functional digraph, which has an arc from i to $f(i)$ for each i in $[n]$. Every weakly connected component of the functional digraph can be represented by a cycle of rooted trees. So n^n is also the number of linear arrangements of rooted trees on $[n]$. We claim now that $n^n = n^2 t_n$, where t_n is the number of trees on $[n]$.

It is clear that $n^2 t_n$ is the number of triples (x, y, T), where $x, y \in [n]$ and T is a tree on $[n]$. Given such a triple, we obtain a linear arrangement of rooted trees by removing all arcs on the unique path from x to y and taking the nodes on this path to be the roots of the trees that remain. This correspondence is bijective, and thus $t_n = n^{n-2}$. ∎

Proof of Theorem 1.6. We use a static approach. Our aim is to find the asymptotic behavior of the number of trees of size between $an^{2/3}$ and $bn^{2/3}$ for any fixed $0 < a < b$, and to conclude that for any $\epsilon > 0$ there exists $N > 0$ and $a > 0$ such that for any $n > N$ there exists a tree component of size $> an^{2/3}$ with probability $> 1 - \epsilon$.

Assume $p = \frac{1}{n}$, and for any k let T_k be the random variable counting the number of tree components of size k in $G(n,p)$. Recall that by Theorem 1.8 there are k^{k-2} possible trees on k labelled vertices, so

$$\mathbf{E}T_k = \binom{n}{k} k^{k-2} \left(\frac{1}{n}\right)^{k-1} \left(1 - \frac{1}{n}\right)^{k(n-k)+\binom{k}{2}-(k-1)}.$$

This is approximately

$$\frac{n(n-1)\dots(n-k+1)}{\sqrt{2\pi}k^{5/2}}\left(\frac{1}{n}\right)^{k-1} e^{\frac{k^2-3k}{2n}}\,.$$

Hence,

$$\log \mathbf{E}T_k = \sum_{j=1}^{k-1} \log\left(1-\frac{j}{n}\right) + \frac{k^2-3k}{2n} - \log k^{5/2} + \log n + \log(2\pi)^{-1/2} + o(1)\,.$$

Recall that the Taylor expansion of $\log(1-x)$ is $-x-\frac{x^2}{2}+O(x^3)$ as $x\to 0$, so

$$\log \mathbf{E}T_k = \sum_{j=1}^{k-1}\left(-\frac{j}{n}-\frac{j^2}{2n^2}+O(\frac{j^3}{n^3})\right) + \frac{k^2-3k}{2n} + \log n - \log k^{5/2} + \log(2\pi)^{-1/2} + o(1)\,,$$

which adds up to

$$\log \mathbf{E}T_k = -\frac{k^3}{6n^2} - \frac{3k}{2n} - \log k^{5/2} + \log n + O\left(\frac{k^4}{n^3}\right) + \log(2\pi)^{-1/2} + o(1)\,.$$

This implies that if $k/n^{2/3}\to\infty$ then $\mathbf{E}T_k \to 0$ and thus a.a.s. there are no tree components of size k in $G(n,\frac{1}{n})$. Fix $0<a<b$ and let $S=\sum_{k=an^{2/3}}^{bn^{2/3}} T_k$. So,

$$\begin{aligned}
\mathbf{E}S &= \frac{(1+o(1))n}{\sqrt{2\pi}} \sum_{k=an^{2/3}}^{bn^{2/3}} e^{-\frac{k^3}{6n^2}+O\left(\frac{k^4}{n^3}+\frac{k}{n}\right)} k^{-5/2} \\
&= \frac{1+o(1)}{\sqrt{2\pi}} \sum_{k=an^{2/3}}^{bn^{2/3}} e^{-\frac{1}{6}\left(\frac{k}{n^{2/3}}\right)^3} \left(\frac{k}{n^{2/3}}\right)^{-5/2} \frac{1}{n^{2/3}} \\
&\to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^3}{6}} x^{-5/2} dx\,.
\end{aligned}$$

For the second moment of S, similar computations (see Exercise 5) yield

$$\mathbf{E}S^2 \to \int_a^b \int_a^b x^{-5/2} y^{-5/2} e^{-\frac{1}{6}(x+y)^3} dxdy + \int_a^b e^{-\frac{x^3}{6}} x^{-5/2} dx\,.$$

Recall that if $X\ge 0$ is random variable then, by Cauchy Schwarz,

$$\mathbf{P}[X>0] \ge \frac{(\mathbf{E}X)^2}{\mathbf{E}[X^2]}\,.$$

So to prove the second claim of the Theorem it remains to prove that

$$\lim_{a\to 0}\lim_{n\to\infty} \frac{(\mathbf{E}S)^2}{\mathbf{E}[S^2]} = 1\,,$$

which is indeed true (see Exercise 5). ∎

Proof of Theorem 1.7. For a vertex $v \in V$, let $\mathcal{C}(v)$ denote the connected component which contains v. We will prove that for any fixed $v \in V$

$$\mathbf{E}|\mathcal{C}(v)| \leq 3n^{1/3} + 5\,. \tag{1.4}$$

Assuming (1.4) the theorem follows since if we let $\{v_1, \ldots, v_n\}$ be the vertices, then

$$\mathbf{E}|\mathcal{C}(v)| = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}|\mathcal{C}(v_i)| = \frac{1}{n}\mathbf{E}\sum_{j} |\mathcal{C}_j|^2\,.$$

Symmetry implies the first equality, and the second equality follows from the observation that in the middle term we sum $|\mathcal{C}_i|$ exactly $|\mathcal{C}_i|$ times, for every component $\mathcal{C}_i$. Thus by (1.4),

$$\mathbf{E}|\mathcal{C}_1|^2 \leq \mathbf{E}\sum_{j} |\mathcal{C}_j|^2 \leq 3n^{4/3} + 5n\,,$$

and Markov's inequality concludes the proof.

To prove (1.4) we define the following process. Let $\alpha \in (0, 1/2)$ be chosen later, let $\{\xi_t\}$ be a sequence of i.i.d. $\mathrm{Bin}(n, 1/n)$ random variables, and let $\{\eta_t\}$ be a sequence of i.i.d. $\mathrm{Bin}(n - n^{2\alpha}, 1/n)$ random variables (independent of $\{\xi_t\}$). Let $W_0 = 1$ and

$$W_t = W_{t-1} - 1 + \begin{cases} \xi_t, & t \leq \tau_\alpha + n^{2\alpha} \\ \eta_t, & t > \tau_\alpha + n^{2\alpha} \end{cases}$$

for $t > 0$, where

$$\tau_\alpha = \min\{t : W_t = 0 \text{ or } W_t \geq n^\alpha\}\,.$$

It is clear that we can couple Y_t and W_t so that $Y_t \leq W_t$, hence to prove (1.4) it suffices to bound the expectation of $\tilde{T} = \min\{t : W_t = 0\}$ by $3n^{1/3} + 5$ for some choice of α (in fact, we will choose $\alpha = 1/3$). In order to show this, we will prove the following lemmas:

LEMMA 1.9.

$$\mathbf{E}[W_{\tau_\alpha + n^{2\alpha}} \mid W_{\tau_\alpha} > 0] \leq n^\alpha + 1\,.$$

LEMMA 1.10.

$$\mathbf{P}[W_{\tau_\alpha} > 0] \leq n^{-\alpha}\,.$$

LEMMA 1.11.

$$\mathbf{E}\tau_\alpha \leq n^\alpha + 4\,.$$

Assume the lemmas. Given $W_{\tau_\alpha} > 0$, the process $Z_t = W_{\tau_\alpha + n^{2\alpha} + t}$ is a random walk with drift $-n^{2\alpha - 1}$, and the only negative increment it can have is -1. Thus Wald's lemma (or optional sampling applied to the martingale $Z_t + tn^{2\alpha-1}$) implies that the stopping time $\gamma = \min\{t \ge 0 : Z_t = 0\}$ satisfies

$$\mathbf{E}[\gamma \mid W_{\tau_\alpha} > 0] = \mathbf{E}[W_{\tau_\alpha + n^{2\alpha}} \mid W_{\tau_\alpha} > 0] n^{1-2\alpha} \le (n^\alpha + 1) n^{1-2\alpha}, \tag{1.5}$$

where the last inequality is by Lemma 1.9.

Observe that if $W_{\tau_\alpha} = 0$ we have $\tilde{T} = \tau_\alpha$, so

$$\tilde{T} \le \tau_\alpha + n^{2\alpha} \mathbf{1}_{(W_{\tau_\alpha} > 0)} + \gamma \cdot \mathbf{1}_{(W_{\tau_\alpha} > 0)} .$$

The expectations of the first two terms are bounded by Lemma 1.11 and Lemma 1.10, and the expectation of the third term is controlled by (1.5) and Lemma 1.10. Summing these contributions,

$$\mathbf{E}\tilde{T} \le 2n^\alpha + n^{1-2\alpha} + n^{1-3\alpha} + 4 .$$

Taking $\alpha = 1/3$ gives $\mathbf{E}\tilde{T} \le 3n^{1/3} + 5$. ∎

We proceed to the proofs of the lemmas. We will use the following simple fact.

Lemma 1.12. *Let X be distributed as $Bin(n,p)$, and let $r \in [0,n]$ be an integer. Then*

$$\mathbf{E}[X - r \mid X \ge r] \le (n-r)p, \tag{1.6}$$

and

$$\mathbf{E}\Big[(X-r)^2 \mid X \ge r\Big] \le [(n-r)p]^2 + (n-r)p(1-p) .$$

Proof. Write X as a sum of i.i.d. indicator variables and condition on the first τ such that the partial sum of the initial τ indicators equals r. Then given τ, the difference $X - \tau$ has $\mathrm{Bin}(n-\tau, p)$ distribution. Since $\tau \ge r$, the lemma follows. ∎

Proof of Lemma 1.9. Since at times $t \in [\tau_\alpha, \tau_\alpha + n^{2\alpha}]$ the process $\{W_t\}$ has mean 0 increments, by the strong Markov property it suffices to prove that

$$\mathbf{E}[W_{\tau_\alpha} \mid W_{\tau_\alpha} > 0] \le n^\alpha + 1 .$$

Observe that conditioned on $W_{\tau_\alpha - 1} = h$ and $W_{\tau_\alpha} > 0$, the variable $W_{\tau_\alpha} - h + 1$ is distributed as a $\mathrm{Bin}(n, 1/n)$ random variable X conditioned on the event that $X \ge r = n^\alpha - h + 1$, so $X - r \overset{d}{=} W_{\tau_\alpha} - n^\alpha$. Hence by (1.6),

$$\mathbf{E}[W_{\tau_\alpha} - n^\alpha \mid W_{\tau_\alpha - 1} = h,\ W_{\tau_\alpha} > 0] \le 1. \tag{1.7}$$

Averaging over the values of $W_{\tau_\alpha - 1}$, we conclude that

$$\mathbf{E}[W_{\tau_\alpha} - n^\alpha \mid W_{\tau_\alpha} > 0] \le 1 .$$

∎

Proof of Lemma 1.10. By the Optional Sampling Theorem,

$$1 = \mathbf{E}W_0 = \mathbf{E}W_{\tau_\alpha} = \mathbf{P}[W_{\tau_\alpha} > 0]\mathbf{E}[W_{\tau_\alpha} \mid W_{\tau_\alpha} > 0]\,,$$

and clearly $\mathbf{E}[W_{\tau_\alpha} \mid W_{\tau_\alpha} > 0] \geq n^\alpha$, which finishes the proof of the lemma. ∎

Proof of Lemma 1.11. Let $\sigma^2 = 1 - 1/n$. It is easy to check that $W^2_{t\wedge\tau_\alpha} - \sigma^2(t \wedge \tau_\alpha)$ is a martingale with expectation 1. By the Optional Sampling Theorem,

$$\mathbf{E}W^2_{\tau_\alpha} = 1 + \sigma^2\mathbf{E}\tau_\alpha\,.$$

Squaring $W_{\tau_\alpha} = n^\alpha + (W_{\tau_\alpha} - n^\alpha)$ gives

$$W^2_{\tau_\alpha} = n^{2\alpha} + 2n^\alpha(W_{\tau_\alpha} - n^\alpha) + (W_{\tau_\alpha} - n^\alpha)^2\,.$$

To bound the conditional expectation of the right hand side given $W_{\tau_\alpha} > 0$, use Lemma 1.12 as in (1.7). Assuming $n^\alpha \geq 2$, this gives

$$\mathbf{E}\Big[W^2_{\tau_\alpha} \mid W_{\tau_\alpha} > 0\Big] \leq n^{2\alpha} + 2n^\alpha + 2 \leq n^{2\alpha} + 3n^\alpha\,,$$

and so by Lemma 1.10

$$\sigma^2\mathbf{E}\tau_\alpha \leq \mathbf{P}[W_{\tau_\alpha} > 0]\mathbf{E}\Big[W^2_{\tau_\alpha} \mid W_{\tau_\alpha} > 0\Big] \leq n^\alpha + 3\,,$$

and hence $\mathbf{E}\tau_\alpha \leq n^\alpha + 4$. ∎

Exercises

1. Complete the proof of Lemma 1.2.
2. Show that the a.a.s. upper bound for the largest component size of the random graph in the subcritical phase is a tight bound. I.e., for any $\epsilon > 0$ show that a.a.s.

$$C_1(G(n,p)) \geq \frac{(1-\epsilon)\log n}{|\log\lambda - \lambda + 1|}\,.$$

3.
   i. Let τ be the random variable counting the total population size of a branching process and denote by $G(s)$ its generating function and by $F(s)$ the generating function of the offspring distribution. Prove that

$$G(s) = sF(G(s))\,.$$

ii. (Spitzer's lemma) Suppose $a_1, \ldots, a_k \in \mathbb{Z}$ and $\sum_{i=1}^k a_i = -1$, then there is precisely one cyclic shift on $(1, \ldots, k)$ for which all partial sums are no less than 0.

iii. Let τ denote the total size of the population of a branching process with offspring distribution X and let $X_1, X_2, \ldots$ be i.i.d. random variables distributed as X. Prove that

$$\mathbf{P}[\tau = k] = \frac{1}{k}\mathbf{P}\left[\sum_{j=1}^{k} X_j = k-1\right].$$

iv. For fixed v denote by $C(v)$ the connected component v belongs to. Let A_k be the event that $C(v)$ is a tree component of size k. Prove that

$$\lim_{n\to\infty} \mathbf{P}[A_k] = \frac{(\lambda k)^{k-1}e^{\lambda k}}{k!}.$$

Note that this implies that for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\Big(\Big|\frac{|T_k|}{n} - \frac{(\lambda k)^{k-1}e^{\lambda k}}{k!}\Big| > \epsilon\Big) = 0.$$

4. Let $G(n,p)$ be the random graph with $p = \frac{\lambda}{n}$ such that $\lambda < 1$. Let T denote the number of vertices belonging to trees. Show that $T = (1-o(1))n$ a.a.s.

5.

i. Let S denote the number of trees of size not larger than $an^{2/3}$ and not smaller than $bn^{2/3}$ in $G(n, \frac{1}{n})$ for fixed $0 < a < b$. Prove that

$$\lim_{n\to\infty} \mathbf{E}S^2 = \int_a^b \int_a^b x^{-5/2}y^{-5/2}e^{-\frac{1}{6}(x+y)^3}dxdy + \int_a^b e^{-\frac{x^3}{6}}x^{-5/2}dx.$$

ii. Show that

$$\lim_{a\to 0}\lim_{n\to\infty} \frac{(\mathbf{E}S)^2}{\mathbf{E}[S^2]} = 1.$$

<u>*Solutions*</u>

1. We need to prove $\mathbf{P}[I_j^* > 0] \geq \mathbf{P}[I_j > 0]$, i.e. that $1 - e^{-\lambda/(n-1)} \geq \frac{\lambda}{n}$. This amounts to proving $e^{\frac{\lambda}{n-1}} \geq \frac{n}{n-\lambda}$. Indeed, recall that $e^x > 1 + x$ for any $x > 0$, so $e^{\frac{\lambda}{n-1}} \geq 1 + \frac{\lambda}{n-1}$ which is indeed no less than $\frac{n}{n-\lambda}$ when $\lambda < 1$.

3.

i. Condition on the number of children of the first individual, and recall that each new family has the same distribution as the original one. If $\tau_1, \tau_2 \ldots$ are i.i.d. random variables distributed as τ, we get

$$G(s) = \mathbf{E}[s^\tau] = \sum_{k=0}^{\infty} \mathbf{P}[X = k]\mathbf{E}[s^{1+\tau_1+\ldots+\tau_k}] = s\sum_{k=0}^{\infty} \mathbf{P}[X = k]G(s)^k = sF(G(s)).$$

ii. Continue the sequence periodically such that $a_{k+l} = a_l$ for any natural number $l > 0$. Take $j \in \{0, \ldots, k-1\}$ such that j is the first global minimum of $a_1 + \ldots + a_j$. It is easy to see that for that unique j,

$$a_{j+1} + \ldots + a_{j+l} \geq 0 \quad \forall l < k\,.$$

iii. By the definition of a branching process, $\tau = \min\left\{k : 1 + \sum_{i=1}^{k} X_i = k\right\}$. For any k, let $Y_1, \ldots, Y_k$ be i.i.d. random variables distributed as $X - 1$. Do a uniform random cyclic shift of the variables to get $Y_1', \ldots, Y_k'$. Clearly $(Y_1', \ldots, Y_k')$ is distributed as $(X_1 - 1, \ldots, X_k - 1)$. If $Y_1 + \ldots + Y_k = -1$ then by Spitzer's lemma, there is precisely one cyclic shift for which all partial sums are non-negative. We conclude that

$$\begin{aligned}\mathbf{P}[\tau = k] &= \mathbf{P}\left[\sum_{i=1}^{k} X_i = k-1, \quad \sum_{i=1}^{l} X_i \geq l \quad \forall l < k\right] \\ &= \frac{1}{k}\mathbf{P}\Big[\sum_{i=1}^{k} Y_i' = -1\Big] = \frac{1}{k}\mathbf{P}\Big[\sum_{i=1}^{k} X_i = k-1\Big]\,.\end{aligned}$$

iv. Let T be the total population of a branching process with offspring distribution Poisson(λ). Because the Poisson distribution is the limiting distribution of Binomials, it is clear that for any fixed k

$$\lim_{n\to\infty} \mathbf{P}[A_k] = \mathbf{P}[T = k]\,.$$

By our previous arguments we can now compute this probability:

$$\mathbf{P}[T = k] = \frac{1}{k}\mathbf{P}[\text{Poisson}(\lambda k) = k-1] = \frac{(\lambda k)^{k-1}e^{\lambda k}}{k!}\,.$$

5.

i. For any k and ℓ denote by $T_{k,\ell}$ the random variable counting the pairs of disjoint tree components of size k and ℓ, respectively. We have

$$\mathbf{E}S^2 = \sum_{an^{2/3} \leq k,\ell \leq bn^{2/3}} \mathbf{E}T_{k,\ell} + \sum_{an^{2/3} \leq k \leq bn^{2/3}} \mathbf{E}T_k\,.$$

This leads to a computation similar to the one in the proof of Theorem 1.6:

$$\begin{aligned}\mathbf{E}T_{k,\ell} &= \binom{n}{k}\binom{n-k}{\ell} k^{k-2}\ell^{\ell-2}\left(\frac{1}{n}\right)^{k+\ell-2}\left(1 - \frac{1}{n}\right)^{\binom{k}{2}+\binom{\ell}{2}+k(n-k)+\ell(n-\ell)-(k-1)-(\ell-1)} \\ &= \frac{(1+o(1))n^2}{2\pi}\prod_{j=1}^{k-1}\left(1 - \frac{j}{n}\right)\prod_{j=1}^{\ell-1}\left(1 - \left(\frac{j}{n} + \frac{k}{n}\right)\right)(k\ell)^{-5/2}e^{\frac{k^2}{2n}+\frac{\ell^2}{2n}+\frac{\ell k}{n}}\,.\end{aligned}$$

Again, taking logs and using $\log(1-x) = -x - \frac{x^2}{2} + O(x^3)$ as $x \to 0$ yields

$$\begin{aligned}
\log \mathbf{E}T_{k,\ell} &= 2\log n - \sum_{j=1}^{k-1}\left(\frac{j}{n} + \frac{j^2}{2n^2}\right) - \sum_{j=1}^{\ell-1}\left(\frac{j+k}{n} + \frac{j^2+2jk+k^2}{2n^2}\right) \\
&\quad + \frac{k^2+\ell^2+2\ell k}{2n} + \log(k\ell)^{-5/2} + o(1) \\
&= 2\log n + \log(k\ell)^{-5/2} - \frac{1}{6}\left(\frac{k}{n^{2/3}} + \frac{\ell}{n^{2/3}}\right)^3 + o(1)\,.
\end{aligned}$$

As before, the sum of $\mathbf{E}T_{k,\ell}$ converges to an integral,

$$\sum_{k,\ell} \mathbf{E}T_{k,\ell} \to \int_a^b \int_a^b x^{-5/2} y^{-5/2} e^{\frac{1}{6}(x+y)^3}\, dxdy\,,$$

so finally, as n tends to infinity,

$$\mathbf{E}S^2 \to \int_a^b \int_a^b x^{-5/2} y^{-5/2} e^{-\frac{1}{6}(x+y)^3}\, dxdy + \int_a^b e^{-\frac{x^3}{6}} x^{-5/2} dx\,.$$

ii. Fix $b > 0$ and denote $F(a) = \int_a^b \int_a^b x^{-5/2} y^{-5/2} e^{\frac{1}{6}(x+y)^3} dxdy$ and $G(a) = \int_a^b e^{-\frac{x^3}{6}} x^{-5/2} dx$. Clearly $\lim_{a\to 0} G(a) = \infty$, so $\lim_{a\to 0} G(a)/G(a)^2 = 0$ and it is enough to prove $\lim_{a\to 0} G(a)^2/F(a) = 1$. First observe that for any $a > 0$, $G(a)^2 > F(a)$. Secondly, for any $0 < \delta < b - a$ we have

$$\begin{aligned}
1 \le \frac{G(a)^2}{F(a)} &\le \frac{\left(\int_a^{a+\delta} e^{-\frac{x^3}{6}} x^{-5/2} dx + \int_{a+\delta}^b e^{-\frac{x^3}{6}} x^{-5/2} dx\right)^2}{\int_a^{a+\delta} \int_a^{a+\delta} x^{-5/2} y^{-5/2} e^{-\frac{1}{6}(x+y)^3} dxdy} \\
&\le \frac{\left(\int_a^{a+\delta} x^{-5/2} dx\right)^2 + 2\int_{a+\delta}^b x^{-5/2} dx \int_a^{a+\delta} x^{-5/2} dx + \left(\int_{a+\delta}^b x^{-5/2} dx\right)^2}{e^{-\frac{8}{6}\delta^3}\left(\int_a^{a+\delta} x^{-5/2} dx\right)^2}\,.
\end{aligned}$$

Since for any such δ the integral $\int_{a+\delta}^b x^{-5/2} dx$ is uniformly bounded in a, we get that for any $0 < \delta < b - a$

$$1 \le \limsup_{a\to 0} \frac{G(a)^2}{F(a)} \le e^{\frac{8}{6}\delta^3}\,,$$

which of course implies $\lim_{a\to 0} \frac{G(a)^2}{F(a)} = 1$, and this concludes our proof.

§2. Isoperimetric Inequalities.

For a graph $G = (V, E)$, define the *boundary* ∂A of a set $A \subset V$ as the set of edges with one end in A and the other in $V - A$.

THEOREM 2.1. *(Discrete Isoperimetric Inequality) Let $A \subset \mathbb{Z}^d$ be a finite set, then*

$$|\partial A| \geq 2d|A|^{\frac{d-1}{d}} .$$

REMARK 2.2. Observe that the $2d$ constant in the inequality is the best possible as the example of the d-dimensional cube shows: If $A = [0, n)^d \cap \mathbb{Z}^d$, then $|A| = n^d$ and $|\partial A| = 2dn^{d-1}$.

For every $1 \leq i \leq d$, define the projection $\mathcal{P}_i : \mathbb{Z}^d \to \mathbb{Z}^{d-1}$ simply as the function dropping the ith coordinate, i.e., $\mathcal{P}_i(x_1, \ldots, x_d) = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$. Theorem 2.1 follows easily from the following lemma.

LEMMA 2.3. *(Discrete Loomis and Whitney inequality, 1949) For any finite $A \subset \mathbb{Z}^d$,*

$$|A|^{d-1} \leq \prod_{i=1}^{d} |\mathcal{P}_i(A)| .$$

Proof of Theorem 2.1. The important observation is that $|\partial A| \geq 2\sum_{i=1}^{d} |\mathcal{P}_i A|$. To see this, observe that any vertex in $\mathcal{P}_i(A)$ matches to a straight line in the ith coordinate direction which "stabs" A. Thus, since A is finite, to any vertex in $\mathcal{P}_i(A)$ you can always match two distinct edges in ∂A: the first and last edges on the straight line which intersects A. Using this and the arithmetic-geometric mean inequality we get

$$|A|^{d-1} \leq \prod_{i=1}^{d} |\mathcal{P}_i(A)| \leq \left(\frac{1}{d}\sum_{i=1}^{d} |\mathcal{P}_i(A)|\right)^d \leq \left(\frac{|\partial A|}{2d}\right)^d ,$$

as required. ∎

To prove Lemma 2.3, we introduce some notions of entropy. Let X be a random variable taking values $x_1, \ldots, x_n$. Denote $p(x) := \mathbf{P}[X = x]$, and define the *entropy* of X to be

$$H(X) = \sum_{i=1}^{n} p(x_i) \log(1/p(x_i)) .$$

Given another random variable, Y, taking values $y_1, \ldots, y_m$, denote $p(x, y) := \mathbf{P}[X = x, Y = y]$ and define the *conditional entropy* $H(X \mid Y)$ *of* X *given* Y as

$$H(X \mid Y) = H(X, Y) - H(Y) = \sum_{x_i, y_j} p(x_i, y_j) \log(1/p(x_i, y_j)) - \sum_{y_j} p(y_j) \log(1/p(y_j)) .$$

PROPOSITION 2.4.

(i) *If X takes n values, then $H(X) \le \log n$.*

(ii) $H(X \mid Y) \le H(X)$ *and* $H(X \mid Y, Z) \le H(X \mid Z)$.

Proof.

Note that $H(X) - \log n = \sum_{i=1}^{n} p(x_i) \log \frac{1}{np(x_i)}$. Plugging in the inequality $\log t \le t - 1$, valid for all $t > 0$, we get

$$H(X) - \log n \le \sum_{i=1}^{n} p(x_i) \left(\frac{1}{np(x_i)} - 1 \right) = 0\,.$$

Part (ii) of the Proposition is proved similarly and is left as an exercise for the reader. ∎

Our next step in proving Lemma 2.3 is a theorem of J. Shearer (see Han 1978 and Chung, Frankl, Graham and Shearer 1986).

THEOREM 2.5. *(Shearer's inequality) Let $X_1, \ldots, X_d$ be random variables taking finitely many values, and let $S_1, \ldots, S_l \subset \{1, \ldots, d\}$ be sets with the property that any $j = 1, \ldots, d$ is a member of precisely r of the S_i's. Then*

$$rH(X_1, \ldots, X_d) \le \sum_{i=1}^{l} H(\{X_j : j \in S_i\})\,.$$

Proof. For any i, by definition, we have the telescoping sum

$$H(X_j : j \in S_i) = \sum_{j \in S_i} H(X_j \mid X_u : u \in S_i, u < j) \ge \sum_{j \in S_i} H(X_j \mid X_u : u < j)\,,$$

where the last inequality follows from part (ii) of Proposition 2.4. Sum over i and recall that each j appears in precisely r of the S_i's to get

$$\sum_{i=1}^{l} H(X_j : j \in S_i) \ge \sum_{i=1}^{l} \sum_{j \in S_i} H(X_j \mid X_u : u < j) = \sum_{j=1}^{d} rH(X_j \mid X_u : u < j) = rH(X_1, \ldots, X_d)\,,$$

where the last equality follows from the definition of conditional entropy. ∎

Proof of Lemma 2.3.

Take random variables $X_1, \ldots, X_d$ such that $(X_1, \ldots, X_d)$ is distributed uniformly on A. Clearly $H(X_1, \ldots, X_d) = \log |A|$, and by Proposition 2.4, $H(X_1, \ldots X_{i-1}, X_{i+1}, \ldots, X_d) \le \log |\mathcal{P}_i(A)|$. Now use Theorem 2.5 on $X_1, \ldots, X_d$ and $S_i = \{1, \ldots, d\} - \{i\}$ to find that

$$(d-1) \log |A| \le \sum_{i=1}^{d} \log |\mathcal{P}_i(A)|\,,$$

as required. ∎

We now present an argument which gives a lower bound to the isoperimetric profile of a Cayley graph based only on the rate of growth of the graph. This proof is attributed to Coulhon and Saloff-Coste. Denote by $B(n)$ the ball of radius n surrounding the identity and by $V(n)$ the number of vertices in $B(n)$.

THEOREM 2.6. *Let $G = (V, E)$ be a Cayley graph of degree d. Let $\phi(\ell) = \inf\{n : V(n) > \ell\}$, then for all finite $K \subset V$ such that $|K| \le |V|/2$, we have*

$$\frac{|\partial K|}{|K|} \ge \frac{1}{2\phi(2|K|)}.$$

Proof. Let $n = \phi(2|K|)$ and take the ball such that $V(n) \ge 2|K|$. Fix $x \in K$, then a uniform random $g \in B(n)$ has probability at least $1/2$ to have $gx \notin K$. So the expected value of $|\{x \in K : gx \notin K\}|$ is at least $|K|/2$, hence there is some $g \in B(n)$ that achieves this. Now notice that if s is a generator, the number of vertices that leave K due to s acting on them is at most $|\partial K|$, i.e., $|\{x \in K : sx \notin K\}| \le |\partial K|$. By the same token, if g is at distance at most n from the origin, then $|\{x \in K : gx \notin K\}| \le n|\partial K|$, so we get

$$|K|/2 \le n|\partial K|,$$

which yields the result. ∎

<u>*Exercises*</u>

1. Prove part (ii) of Proposition 2.4.
2. Show that there exists a constant $C_d > 0$ such that if A is a subgraph of the box $\{0, \dots, n-1\}^d$ and $|A| < \frac{n^d}{2}$, then

$$|\partial A| \ge C_d |A|^{\frac{d-1}{d}}.$$

<u>*Solutions*</u>

1. Note that $H(X|Y) \le H(X)$ is a special case of $H(X|Y,Z) \le H(X|Y)$ by taking Y to be constant. Using the fact that $\log t \le t - 1$ for any $t > 0$, we get

$$\begin{aligned} H(X|Y,Z) - H(X|Y) &= H(X,Y,Z) - H(Y,Z) - H(X,Y) + H(Y) \\ &= \sum_{x,y,z} p(x,y,z) \log \frac{p(y,z)p(x,y)}{p(x,y,z)p(y)} \\ &\le \sum_{x,y,z} p(x,y,z) \left(\frac{p(y,z)p(x,y)}{p(x,y,z)p(y)} - 1 \right) \\ &= \sum_{x,y,z} \frac{p(y,z)p(x,y)}{p(y)} - 1 = 0. \end{aligned}$$

2. Let $A \subset \{0, \ldots, n-1\}^d$ with $|A| < \frac{n^d}{2}$. Let m be such that $|\mathcal{P}_m(A)|$ is maximal over all projections, and let

$$F = \{a \in \mathcal{P}_m(A) : |\mathcal{P}_m^{-1}(a)| = n\} .$$

Notice that for any $a \notin F$ there is at least one edge in ∂A, and for different a's we get disjoint edges, i.e., $|\partial A| \geq |\mathcal{P}_m(A) \setminus F|$. By Theorem 2.1 we get

$$|A|^{d-1} \leq \prod_{j=1}^{d} |\mathcal{P}_j(A)| \leq |\mathcal{P}_m(A)|^d ,$$

and so $|\mathcal{P}_m(A)| \geq |A|/|A|^{1/d} \geq 2^{1/d}|A|/n \geq 2^{1/d}|F|$, which together yield

$$|\partial A| \geq |\mathcal{P}_m(A) \setminus F| \geq (1 - 2^{-1/d})|\mathcal{P}_m(A)| \geq (1 - 2^{-1/d})|A|^{\frac{d-1}{d}} .$$

§3. Markov Type of Metric Spaces.

Recall that a Markov chain $\{Z_t\}_{t=0}^{\infty}$ with transition probabilities $p_{ij} := \mathbf{P}[Z_{t+1} = j \mid Z_t = i]$ on the state space $\{1, \ldots, n\}$ is *stationary* if $\pi_i := \mathbf{P}[Z_t = i]$ does not depend on t, and it is (*time*) *reversible* if $\pi_i p_{ij} = \pi_j p_{ji}$ for every $i, j \in \{1, \ldots, n\}$.

Definition (Ball 1992): Given a metric space (X, d) we say that X has Markov type 2 if there exists a constant $M > 0$ such that for every stationary reversible Markov chain $\{Z_t\}_{t=0}^{\infty}$ on $\{1, \ldots, n\}$, every mapping $f : \{1, \ldots, n\} \to X$ and every time $t \in \mathbb{N}$,

$$\mathbf{E}d(f(Z_t), f(Z_0))^2 \leq M^2 t \mathbf{E}d(f(Z_1), f(Z_0))^2 .$$

We will prove that the real line and weighted trees have Markov type 2 (see Exercise 1 for a space which does not have Markov type 2). The result for trees is due to Naor, Peres, Schramm and Sheffield (2004).

THEOREM 3.1. *$\mathbb{R}$ has Markov type* 2.

Recall that any tree with edge weights defines a metric on the vertices: the distance between two vertices in a weighted tree is the sum of the weights along the unique path between the vertices.

THEOREM 3.2. *Weighted trees have uniform Markov type* 2.

Let $P = (p_{ij})$ be the transition matrix of the Markov chain. Observe that time reversibility is equivalent to the assertion that P is a self-adjoint operator in $L^2(\pi)$. This is because

$$\langle Pf, g\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}\pi(i)f(j)g(i) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ji}\pi(j)f(j)g(i) = \langle f, Pg\rangle\,.$$

This in turn implies that $L^2(\pi)$ has an orthogonal basis of eigenfunctions of P with real eigenvalues. Since P is a stochastic matrix,

$$\max_i \sum_{j=1}^{n} p_{ij}|f(j)| \le \max_i |f(i)|\,,$$

which implies $\|Pf\|_\infty \le \|f\|_\infty$, and thus if λ is an eigenvalue of P then $|\lambda| \le 1$. We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. We prove the statement with constant $M = 1$. Note that

$$\mathbf{E}d(f(Z_t), f(Z_0))^2 = \sum_{i,j} \pi_i p_{ij}^{(t)}[f(i) - f(j)]^2 = 2\langle (I - P^t)f, f\rangle\,,$$

and similarly

$$\mathbf{E}d(f(Z_1), f(Z_0))^2 = 2\langle (I - P)f, f\rangle\,.$$

Thus it suffices to prove that

$$\langle (I - P^t)f, f\rangle \le t\langle (I - P)f, f\rangle\,.$$

Indeed, if f is an eigenfunction with eigenvalue λ, this reduces to proving $(1-\lambda^t) \le t(1-\lambda)$. Since $|\lambda| \le 1$, this reduces to

$$1 + \lambda + \cdots + \lambda^{t-1} \le t\,,$$

which is obviously true.

The claim follows for any other f by taking $f = \sum_{j=1}^{n} a_j f_j$ where $\{f_j\}$ is an orthonormal basis of eigenfunctions:

$$\langle (I - P^t)f, f\rangle = \sum_{j=1}^{n} a_j^2 \langle (I - P^t)f_j, f_j\rangle \le \sum_{j=1}^{n} a_j^2 t\langle (I - P)f_j, f_j\rangle = t\langle (I - P)f, f\rangle\,.$$

∎

To prove Theorem 3.2 we first prove a lemma.

LEMMA 3.3. *Let $\{Z_t\}_{t=0}^{\infty}$ be a stationary time reversible Markov chain on $\{1,\ldots,n\}$ and $f:\{1,\ldots,n\}\to\mathbb{R}$. Then, for every time $t>0$,*

$$\mathbf{E}\max_{0\le s\le t}[f(Z_s)-f(Z_0)]^2 \le 15t\mathbf{E}[f(Z_1)-f(Z_0)]^2\,.$$

Proof. Let π be the stationary distribution. Define $P: L^2(\pi)\to L^2(\pi)$ by $(Pf)(i) = \mathbf{E}[f(Z_{s+1})|Z_s=i] = \sum_{j=1}^n p_{ij}f(j)$. For any $s\in\{0,\ldots,t-1\}$, let

$$D_s = f(Z_{s+1}) - (Pf)(Z_s)\,.$$

Since $\mathbf{E}[D_s|Z_1,\ldots,Z_s] = \mathbf{E}[D_s|Z_s] = 0$, the D_s are martingale differences with respect to the natural filtration of $Z_1,\ldots,Z_t$. Also, because of time-reversibility,

$$\tilde{D}_s = f(Z_{s-1}) - (Pf)(Z_s)$$

are martingale differences with respect to the natural filtration on $Z_t,\ldots,Z_1$. Note that $D_s - \tilde{D}_s = f(Z_{s+1}) - f(Z_{s-1})$, which implies that for any m,

$$f(Z_{2m}) - f(Z_0) = \sum_{k=1}^m D_{2k-1} - \sum_{k=1}^m \tilde{D}_{2k-1}\,.$$

So,

$$\max_{0\le s\le t} f(Z_s) - f(Z_0) \le \max_{m\le t/2}\sum_{k=1}^m D_{2k-1} + \max_{m\le t/2}\sum_{k=1}^m -\tilde{D}_{2k-1} + \max_{\ell\le t/2}|f(Z_{2\ell+1}) - f(Z_{2\ell})|\,.$$

Take squares and use the fact $(a+b+c)^2 \le 3(a^2+b^2+c^2)$, which is implied by the Cauchy-Schwarz inequality, to get

$$\begin{aligned}\max_{0\le s\le t}|f(Z_s)-f(Z_0)|^2 \le& 3\max_{m\le t/2}\left|\sum_{k=1}^m D_{2k-1}\right|^2 + 3\max_{m\le t/2}\left|\sum_{k=1}^m \tilde{D}_{2k-1}\right|^2 \\ &+ 3\sum_{\ell\le t/2}\left|f(Z_{2\ell+1}) - f(Z_{2\ell})\right|^2\,.\end{aligned}$$

We will use Doob's L^2 maximum inequality for martingales (see, e.g., Durrett 1996)

$$\mathbf{E}\max_{0\le s\le t} M_s^2 \le 4\,\mathbf{E}|M_t|^2\,.$$

Consider

$$M_{s+1} = \sum_{j \le s, j \text{ odd}} D_j \,.$$

Since M_s is still a martingale, we have

$$\begin{aligned}\mathbf{E} \max_{0 \le s \le t} |f(Z_s) - f(Z_0)|^2 \le & 12\mathbf{E}\Big| \sum_{k=1}^{\lfloor t/2 \rfloor} D_{2k-1} \Big|^2 + 12\mathbf{E}\Big| \sum_{k=1}^{\lfloor t/2 \rfloor} \tilde{D}_{2k-1} \Big|^2 \\ & + 3 \sum_{0 \le \ell \le t/2} \mathbf{E}\Big| f(Z_{2\ell+1}) - f(Z_{2\ell}) \Big|^2 \,.\end{aligned}$$

Denote $V = \mathbf{E}[|f(Z_1) - f(Z_0)|^2]$, and notice that

$$D_0 = f(Z_1) - f(Z_0) - \mathbf{E}[f(Z_1) - f(Z_0) \mid Z_0] \,,$$

which implies that D_0 is orthogonal to $\mathbf{E}[f(Z_1) - f(Z_0) \mid Z_0]$ in $L^2(\pi)$. So, by the Pythagorian law, for any s we have $\mathbf{E}[D_s^2] = \mathbf{E}[D_0^2] \le V$. Summing everything up gives

$$\mathbf{E} \max_{0 \le s \le t} |f(Z_s) - f(Z_0)|^2 \le 6tV + 6tV + 3(t/2+1)V \le 15tV \,,$$

which concludes our proof. ∎

Proof of Theorem 3.2.

Let T be a weighted tree, $\{Z_j\}$ be a reversible Markov chain on $\{1, \ldots, n\}$ and $F : \{1, \ldots, n\} \to T$. Choose an arbitrary root and set for any vertex v, $\psi(v) = d(\mathit{root}, v)$. If $v_0, \ldots, v_t$ is a path in the tree, then

$$d(v_0, v_t) \le \max_{0 \le j \le t} \left(|\psi(v_0) - \psi(v_j)| + |\psi(v_t) - \psi(v_j)| \right) \,,$$

since choosing the closest vertex to the root on the path yields equality.

Let $X_j = F(Z_j)$. Connect X_i to X_{i+1} by the shortest path for any $0 \le i \le t-1$ to get a path between X_0 and X_t. Since now the closest vertex to the root can be on any of the shortest paths between X_j and X_{j+1}, we get

$$d(X_0, X_t) \le \max_{0 \le j < t} \Big(|\psi(X_0) - \psi(X_j)| + |\psi(X_t) - \psi(X_j)| + 2d(X_j, X_{j+1}) \Big) \,.$$

Square, and use Cauchy-Schwarz again,

$$d(X_0, X_t)^2 \le 3 \max_{0 \le j \le t} \Big(|\psi(X_0) - \psi(X_j)|^2 + |\psi(X_t) - \psi(X_j)|^2 \Big) + 12 \sum_{0 \le j < t} d^2(X_j, X_{j+1}) \,.$$

By Lemma 3.3 with $f = \psi F$ we get,

$$\mathbf{E}d(X_0, X_t)^2 \le 90t\mathbf{E}|\psi(X_0) - \psi(X_1)|^2 + 6 \sum_{0 \le j \le t} \mathbf{E}d^2(X_j, X_{j+1})\,.$$

Since in any metric space $|\psi(X_1) - \psi(X_0)| \le d(X_0, X_1)$ and since the Markov chain is stationary we have $\mathbf{E}d(X_0, X_1) = \mathbf{E}d(X_j, X_{j+1})$ for any j. So

$$\mathbf{E}d(X_0, X_t)^2 \le 96t\mathbf{E}d(X_0, X_1)^2\,,$$

which concludes our proof. ∎

<u>*Exercises*</u>

1. Prove that the n-dimensional hypercubes do not have uniform Markov type 2.

<u>*Solutions*</u>

1. We show that a simple random walk on the hypercube satisfies

$$\mathbf{E}d(Y_t, Y_0) \ge \frac{t}{2} \qquad \forall t < \frac{n}{4}\,.$$

It follows from Jensen's inequality that $\mathbf{E}d^2(Y_t, Y_0) > t^2/4$ for $t < n/4$, implying that the hypercubes do not have uniform Markov type 2. Indeed, at each step at time $t < \frac{n}{4}$ we have probability at least $\frac{3}{4}$ to increase the distance by 1 and probability at most $\frac{1}{4}$ of decreasing it by 1, so

$$\mathbf{E}d(Y_t, Y_0) \ge \frac{3}{4}t - \frac{1}{4}t = \frac{t}{2}\,.$$

§4. Random Walks and Electrical Networks.

THEOREM 4.1. *Consider the* 2*-dimensional square, i.e. the spanning subgraph of* $\mathbb{Z}^2$ *on vertices* $\{1, \ldots, n\}^2$*. Let* $\mathbf{a} = (1, 1)$ *and* $\mathbf{z} = (n, n)$*. The probability that a simple random walk, starting at* $\mathbf{a}$ *will reach* $\mathbf{z}$ *before returning to* $\mathbf{a}$ *is* $\Theta(1/\log n)$*.*

Proof. Recall that this is equivalent to showing that the effective resistance between $\mathbf{a}$ and $\mathbf{z}$ is $\Theta(\log n)$. By considering the natural disjoint diagonal cutsets, the Nash-Williams inequality easily gives

$$R_{\text{eff}}(\mathbf{a} \leftrightarrow \mathbf{z}) \ge 2 \sum_{k=1}^{n-1} \frac{1}{2k} \ge \log(n-1)\,.$$

To get a lower bound, recall that the effective resistance is the minimal energy of a unit flow from a to z. So, to get a matching upper bound, we build a unit flow on the square. Recall that Pólya's Urn is a Markov chain on the square where the probability of moving from (a, b) to $(a+1, b)$ is $\frac{a}{a+b}$ and to $(a, b+1)$ is $\frac{b}{a+b}$. Traditionally, a is the number of black balls in a bin and b is the number of white balls, and on each step we draw at random a ball from the bin, return it and add a new ball to the bin with the same color of the drawn ball. Run the process on the square and stop when you reach the main diagonal ($\{(x, x)\}$). Direct all horizontal edges east, vertical edges north and define for all edges in the bottom left half of the square

$$f(e) = \mathbf{P}[\text{the process went through e}] \, .$$

To finish the construction, reflect the bottom left half on the main diagonal to get the flow values for the upper right half of the square.

It is a well-known fact that in Pólya's Urn process, for any k, all pairs (a, b) of which $a + b = k$ are equally probable to be reached. To see this, observe that the number of black balls in the process after k steps is distributed like the number of real numbers to the left of X_0 among $X_1, \ldots, X_k$, where $\{X_i\}_{i=0}^{k}$ are independent random variables distributed uniformly on $[0, 1]$. Thus, if (a, b) is a point in the square having $a + b = k$, the sum of the flow on the two edges pointing at it is $\frac{1}{k}$. Thus, the energy of this flow is

$$\mathcal{E}(f) \leq \sum_{k=1}^{n} 2 \left(\frac{1}{k} \right)^2 k \leq 2 \log(n + 1) \, .$$

∎

Since unit current flows have probabilistic interpretations, constructions of these flows can lead to useful observations. Let G be a connected graph with unit resistances for all edges (the simple random walk case) and marked vertices a and z. Choose uniformly at random a spanning tree of G, and for any edge $e = \vec{xy}$ take

$$f(e) = \mathbf{E}[\text{net number of crossings of } e \text{ in the unique path } a \to z] \, ,$$

where the net number of crossings of e is $+1$ if the path crosses e from x to y, -1 if it crosses from y to x, and 0 if it does not cross e.

Proposition 4.2. *f is the unit current flow.*

Proof. We first observe that f is a unit flow. Note that

$$\sum_{x} f(\vec{ax}) = \mathbf{E} \sum_{a \sim x} \text{net crossings of } \vec{ax} \, .$$

But since any flow from a to z goes through precisely one neighbor of a (in the right direction) this is always 1. Similarly, $\sum_y f(\vec{yz}) = 1$. Also, if $w \notin \{a, z\}$, any path from a to z passing through w has precisely one edge in the positive direction and one edge in the negative direction, thus $\sum_x f(\vec{wx}) = 0$. To deduce that f is unit current flow, recall that it is enough to check that the cycle law holds, i.e., if $x_0, x_1, \ldots, x_\ell = x_0$ is a cycle in G then

$$\sum_{i=0}^{l-1} f(x_i, x_{i+1}) = 0\,.$$

Let $e_1, \ldots, e_\ell$ be a cycle in G. To show that the cycle law holds, for any i we build an injection from trees in which the $a \to z$ path uses e_i in the positive direction to trees in which the $a \to z$ path uses e_j in the negative direction for some $1 \leq j \leq \ell$. Note that for any i, the union of trees of which the $a \to z$ path uses e_i in the positive direction and their image under the injection contribute precisely 0 to the sum of the flow over the cycle. Summing up on i gives that the cycle law holds.

Given a tree where the $a \to z$ path uses e_i in the positive direction, delete e_i and let $C(a)$ and $C(z)$ be the connected components containing a and z, respectively. Choose e_j to be the first edge of the cycle, starting from e_i, which connects $C(z)$ to $C(a)$. Add e_j to connect the two components. This yields a new tree of which the $a \to z$ path goes through e_j in the negative direction. ∎

DEFINITION. Given a Markov chain, define the *Green function* $g_\tau(a, x)$ to be the expected number of visits to x before time τ, starting at a. That is,

$$g_\tau(a, x) = \mathbf{E}_a \sum_{t=0}^{\tau-1} \mathbf{1}_{\{Y_t = x\}}\,.$$

LEMMA 4.3. *(Aldous, Fill 2005) If τ is a stopping time for a finite and irreducible Markov chain, and $\mathbf{P}_a[Y_\tau = a] = 1$, then*

$$\frac{g_\tau(a, x)}{\mathbf{E}_a \tau} = \pi(x) \qquad \forall x\,.$$

Proof. Let V denote the state space of the chain. Note that for any index t and $y \in V$ we have

$$\sum_{x \in V} \mathbf{1}_{\{Y_t = x\}} \mathbf{1}_{\{Y_{t+1} = y\}} = \mathbf{1}_{\{Y_{t+1} = y\}}\,.$$

Since $\mathbf{P}_a[Y_\tau = a] = 1$, we have $\mathbf{1}_{\{Y_0 = y\}} = \mathbf{1}_{\{Y_\tau = y\}}$ with probability 1. This implies

$$\sum_{x \in V} g_\tau(a, x) p(x, y) = g_\tau(a, y)\,,$$

meaning that the vector $(g_\tau(a,x))_{x\in V}$ is invariant, and clearly

$$\sum_{x\in V} g_\tau(a,x) = \mathbf{E}_a \sum_{t=0}^{\tau-1}\sum_{x\in V} \mathbf{1}_{\{Y_t=x\}} = \mathbf{E}_a\tau\,,$$

since $\sum_{x\in V}\mathbf{1}_{\{Y_t=x\}} = 1$ with probability 1. Thus, normalizing $(g_\tau(a,x))_{x\in V}$ gives the stationary distribution. ∎

Corollary 4.4. *Let τ_a^+ be the first time returning to a when starting from a. Clearly $g_{\tau_a^+}(a,a) = 1$. Then, by the lemma,*

$$\mathbf{E}\tau_a^+ = \frac{1}{\pi(a)}\,.$$

Define the *commute time* between a and b as the first time to visit a after visiting b. That is,

$$\tau = \min\{j : Y_j = a, \exists i < j \quad Y_i = b\}\,.$$

We will use the lemma to compute the expected commute time between any two points.

Proposition 4.5. *Let C_v denote the conductances in a finite, irreducible and reversible Markov chain. For any states a and b in the chain, let τ be their commute time. Then,*

$$\mathbf{E}_a\tau = \left(\sum_{v\in V} C_v\right) R_{\text{eff}}(a\leftrightarrow b)\,.$$

Proof. Denote $C = \sum_v C_v$. By the lemma,

$$\frac{g_\tau(a,a)}{\mathbf{E}_a\tau} = \pi(a) = \frac{C_a}{C}\,.$$

By definition, after visiting b the chain does not visit a until time τ, so $g_\tau(a,a) = g_{\tau_b}(a,a)$. Notice that, because of lack of memory, $g_{\tau_b}(a,a)$ is the expected value of a geometric random variable with success probability $\mathbf{P}_a(\tau_b < \tau_a^+)$, so

$$g_\tau(a,a) = \frac{1}{\mathbf{P}_a(\tau_b<\tau_a^+)} = C_a R_{\text{eff}}(a\leftrightarrow b)\,,$$

and we conclude that

$$\mathbf{E}_a\tau = C R_{\text{eff}}(a\leftrightarrow b)\,.$$

∎

Corollary 4.6. *In the 2-dimensional square $\{1,\ldots,n\}^2$, the commute time between $a = (1,1)$ and $z = (n,n)$ is $\Theta(n^2 \log n)$.*

Proof. Follows from Proposition 4.5 and Theorem 4.1 immediately. ∎

Matthews' cover time inequality

The *Cover Time* $\mathcal{C}$ of a finite Markov chain is the first time the process visits all states, i.e.,

$$\mathcal{C} = \min\{t : \forall x \in V \quad \exists j \le t \quad Y_j = x\}\,.$$

Theorem 4.7. *(Matthews 1988) For any irreducible finite Markov chain on n states,*

$$\mathbf{E}[\mathcal{C}] \le \left(\max_{a,b\in V} \mathbf{E}_a \tau_b\right)\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right)\,.$$

Proof. Order the states according to a random permutation, $(j_1,\ldots,j_n)$. Let t_k be the first time all $(j_1,\ldots,j_k)$ were visited, and let $L_k = Y_{t_k}$ be the state we are in at time t_k. Then,

$$\mathbf{E}[t_k - t_{k-1} | Y_0,\ldots,Y_{t_{k-1}}, j_1,\ldots,j_k] = \mathbf{E}_{L_{k-1}}(\tau_{j_k})\mathbf{1}_{\{L_k = j_k\}}\,.$$

Now, since all permutations are equally probable, it is clear that $P(L_k = j_k) = 1/k$, so we conclude that

$$\mathbf{E}[t_k - t_{k-1}] \le \left(\max_{a,b\in V} \mathbf{E}_a \tau_b\right)\frac{1}{k}\,,$$

and summing over k yields the result. ∎

The same technique can lead to computing lower bounds, as well. For $A \subset \{1,\ldots,n\}$, consider the the cover time of A denoted by $\mathcal{C}_A$. Clearly $\mathbf{E}[\mathcal{C}] \ge \mathbf{E}[\mathcal{C}_A]$. Let $t^A_{\min} = \min_{a,b\in A, a\ne b} \mathbf{E}_a \tau_b$, then similarly to the last proof, we have

$$\mathbf{E}[\mathcal{C}_A] \ge t^A_{\min}\left(1 + \frac{1}{2} + \cdots + \frac{1}{|A|}\right)\,,$$

which makes the following true:

Theorem 4.8. *For any irreducible finite Markov chain on n states,*

$$\mathbf{E}[\mathcal{C}] \ge \max_{A\subset V} \quad t^A_{\min}\left(1 + \frac{1}{2} + \cdots + \frac{1}{|A|}\right)\,.$$

Lemma 4.9. *(The Cycle Lemma) Let a, b, c be distinct states in a reversible Markov chain, and denote by τ_{bca} the first time the trajectory of the chain contains the states b, c, a in that order. Then,*

$$\mathbf{E}_a[\tau_{bca}] = \mathbf{E}_a[\tau_{cba}]\,.$$

Remark. This lemma does not follow from an easy path-reversal argument. Take, for instance, the path *acabca*, then $\tau_{bca} = 6$ but τ_{cba} for the reversed path is 4.

Proof. By adding $\mathbf{E}_\pi[\tau_a]$ to both sides of the equation, where π is the stationary distribution, it becomes equivalent to

$$\mathbf{E}_\pi[\tau_{abca}] = \mathbf{E}_\pi[\tau_{acba}] .$$

But, actually, much more is true, because of reversibility we have that for any $k > 0$

$$\mathbf{P}_\pi[\tau_{abca} > k] = \mathbf{P}_\pi[\tau_{acba} > k] ,$$

which concludes the proof. ∎

REMARK 4.10. See Exercise 7,8 to see how Lemma 4.9 generalizes to cycles of any length, and to the non-reversible case.

Exercises

1. Show that for any finite and reversible Markov chain, the function

$$f(e) = \mathbf{E}_a[\text{net number of crossings of } e \text{ before visiting } z] ,$$

defined on all edges, is the unit current flow from a to z. Deduce that in the loop erased walk from a to z the expected net number of crossings of an edge e is also a unit current flow.

2. Find the effective resistance between two neighboring vertices in the 2-dimensional torus with unit edge resistances. Show it tends to $\frac{1}{2}$ as $n \to \infty$. Deduce that in a simple random walk on $\mathbb{Z}^2$ starting from the origin, the probability of visiting $(1,0)$ before returning to the origin is $\frac{1}{2}$.
3. Consider the 2-dimensional torus with unit edge resistances. Show that there exists a constant γ such that for any $0 < \alpha < \frac{1}{2}$ and for any two points a and z of distance αn, we have $R_{\text{eff}}(a \leftrightarrow z) \sim \gamma \log n$. Find γ explicitly.
4. Find the asymptotic value of the expected cover time of the hypercube $\{0,1\}^n$.
5. Show that for any irreducible Markov chain, the expected hitting time of a random, π-distributed target is independent of the starting state. That is, show that

$$f(x) = \sum_y \pi(y) \mathbf{E}_x \tau_y$$

is constant.

6. Show that if the Markov chain is reversible, the constant from previous question is equal to

$$\sum_{\lambda_i < 1} \frac{1}{1 - \lambda_i} ,$$

where λ_i are the eigenvalues of the transition matrix, and the sum includes the multiplicity of the eigenvalues.

7. Let $\hat{P}$ be the time-reversed chain of a Markov chain with transition matrix P, i.e., $\hat{p}(x,y) = \pi(x)p(x,y)/\pi(y)$.
   i. Prove that $\mathbf{E}_\pi[\tau_a] = \hat{\mathbf{E}}_\pi[\tau_a]$, where $\hat{\mathbf{E}}$ represents the expectation operator of the time-reversed chain.
   ii. Use (i) to prove the following generalization of Lemma 4.9

$$\mathbf{E}_a[\tau_{bca}] = \hat{\mathbf{E}}_a[\tau_{cba}] .$$

8. Generalize Lemma 4.9 for cycles of any length, i.e., for distinct states $a_1, \ldots, a_k$ of a reversible Markov chain, prove that

$$\mathbf{E}_{a_1}[\tau_{a_2,\ldots,a_k,a_1}] = \mathbf{E}_{a_1}[\tau_{a_k,\ldots,a_2,a_1}] ,$$

and similarly for non-reversible chains.

<u>*Solutions*</u>

5. Let x, y be states in the chain and $M > 0$ an integer. Let τ be the following stopping time: start from y, run the chain until you hit x, run the chain further M steps, and after finishing, run the chain until you hit y. Clearly, τ satisfy the conditions of Lemma 4.3, and $\mathbf{E}_y\tau = E_y\tau_x + M + \mathbf{E}_{X_{\tau_x+M}}\tau_y$. So by Lemma 4.3 we get

$$g_\tau(y,y) = \pi(y)\left[\mathbf{E}_y\tau_x + M + \mathbf{E}_{X_{\tau_x+M}}\tau_y\right] ,$$

and clearly

$$g_\tau(y,y) = g_{\tau_x}(y,y) + \sum_{t=0}^{M-1} p_{xy}^t .$$

Now, $g_{\tau_x}(y,y) = g_{\tilde{\tau}}(y,y)$ where $\tilde{\tau}$ is the commute time between y and x, so again by Lemma 4.3, we have

$$g_{\tau_x}(y,y) = \pi(y)\left[\mathbf{E}_x\tau_y + \mathbf{E}_y\tau_x\right] .$$

Rearranging gives

$$\sum_{t=0}^{M-1} (p_{xy}^t - \pi(y)) = \pi(y)\left[\mathbf{E}_{X_{\tau_x+M}}\tau_y - \mathbf{E}_x\tau_y\right] .$$

Note that summing over y the LHS gives 0, so

$$\sum_y \pi(y)\mathbf{E}_x\tau_y = \sum_y \pi(y)\mathbf{E}_{X_{\tau_x+M}}\tau_y .$$

Because of the Markov property, X_{τ_x+M} has $P_x(X_M)$ distribution which converges to the stationary distribution π as $M \to \infty$ and, since the chain is finite, we get

$$\sum_y \pi(y)\mathbf{E}_x\tau_y = \sum_y \pi(y)\mathbf{E}_\pi\tau_y \,,$$

which does not depend on x, as required.

§5. The Cheeger Constant and Mixing Time.

In this section we investigate the speed with which a finite, reversible Markov chain converges to the stationary distribution. Let P be the transition matrix of the chain, and let π be the stationary distribution. For any two states x and y, consider the distance

$$\frac{P^t(x,y)-\pi(y)}{\pi(y)} \,.$$

Since P is reversible, as we have seen before, it has only real eigenvalues, all in $[-1,1]$, denoted $\lambda_n \leq \ldots \leq \lambda_1 = 1$. We will show that under some conditions, when P exhibits what is known as a *spectral gap*, i.e., the second largest eigenvalue of P is strictly smaller than 1, the chain converges to the stationary distribution in an exponential rate in t.

Note that in this section all inner products are with respect to the stationary measure π, i.e.,

$$\langle f,g\rangle = \sum_{i=1}^{n} \pi(i)f(i)g(i) \,.$$

Definition We call P *irreducible* if

$$\forall x,y \quad \exists k \quad P^k(x,y) > 0 \,.$$

We call P *irreducible a-periodic* if

$$\exists k \quad \forall x,y \quad P^k(x,y) > 0 \,.$$

LEMMA 5.1. *Let $\lambda_* = \max_{i\geq 2}|\lambda_i|$ be the second largest (in absolute value) eigenvalue of a finite, irreducible, a-periodic, reversible Markov chain with transition matrix P, then $\lambda_* < 1$.*

For a proof of this lemma see exercise 2.

Theorem 5.2. *Under the conditions of Lemma 5.1, let π be the stationary distribution of P and $\pi_* = \min_x \pi(x)$ and denote $g = 1 - \lambda_*$. Then, for any states x and y,*

$$\left| \frac{P^t(x,y) - \pi(y)}{\pi(y)} \right| \le \frac{e^{-gt}}{\pi_*} .$$

Proof. For any x let ψ_x the following function on the state space

$$\psi_x(z) = \frac{\mathbf{1}_{\{z=x\}}}{\pi(x)} .$$

Since for any x and y we have $(P^t\psi_y)(x) = \frac{P^t(x,y)}{\pi(y)}$ we get that

$$\frac{P^t(x,y) - \pi(y)}{\pi(y)} = \langle \psi_x, P^t\psi_y - \mathbf{1} \rangle .$$

Since $P\mathbf{1} = \mathbf{1}$, we have

$$\langle \psi_x, P^t\psi_y - \mathbf{1} \rangle = \langle \psi_x, P^t(\psi_y - \mathbf{1}) \rangle \le \|\psi_x\|_2 \|P^t(\psi_y - \mathbf{1})\|_2 ,$$

where the last inequality is Cauchy-Schwarz. Let $(f_i)_{i=1}^n$ be a basis of orthogonal eigenvectors with respect to P, where f_i is an eigenvector of eigenvalue λ_i. Since $\psi_y - \mathbf{1}$ is orthogonal to $\mathbf{1}$ there exist constants $(a_i)_{i=2}^n$ such that $\psi_y - \mathbf{1} = \sum_{i=2}^n a_i f_i$, so

$$\|P^t(\psi_y - \mathbf{1})\|_2^2 = \|\sum_{i=2}^n \lambda_i^t a_i f_i\|_2^2 \le \lambda_*^{2t} \|\psi_y - \mathbf{1}\|_2^2 .$$

It is easy to check that $\|\psi_y - \mathbf{1}\|_2 \le \|\psi_y\|_2$, and so

$$\left| \frac{P^t(x,y) - \pi(y)}{\pi(y)} \right| \le \lambda_*^t \|\psi_x\|_2 \|\psi_y\|_2 = \frac{\lambda_*^t}{\sqrt{\pi(x)\pi(y)}} \le \frac{e^{-gt}}{\pi_*} .$$

∎

Intuitively, if a Markov chain has ”bottlenecks”, it will take it more time to mix. To formulate this intuition, we define what is known as the *Cheeger constant*.

For any two states in a stationary Markov chain a and b, let $q(a,b) = \pi(a)p(a,b)$, and for any two subsets of states A and B, let $Q(A,B) = \sum_{a \in A, b \in B} q(a,b)$.

Definition The Cheeger constant of a stationary Markov chain is

$$\Phi_* = \min_{S : \pi(S) \le 1/2} \Phi_S ,$$

where

$$\Phi_S = \frac{Q(S, S^c)}{\pi(S)} .$$

The following theorem (see Alon, 1986, Jerrum and Sinclair, 1989 and Lawler and Sokal, 1988), together with the previous one, connects the isoperimetric properties of a Markov chain with its mixing time. We assume that the chain is “lazy”, i.e., for any state x we have that $p(x,x) \ge \frac{1}{2}$. In that case, $P = \frac{I + \tilde{P}}{2}$ where $\tilde{P}$ is another stochastic matrix, and hence all the eigenvalues of P are in $[0,1]$, and $\lambda_* = \lambda_2$.

Theorem 5.3. *Let λ_2 be the second eigenvalue of a reversible and lazy Markov chain, and $g = 1 - \lambda_2$. Then,*

$$\frac{\Phi_*^2}{2} \le g \le 2\Phi_* \,.$$

Proof. The upper bound is easy. Recall that

$$\lambda_2 = \max_{f \perp \mathbf{1}} \frac{\langle Pf, f\rangle}{\langle f, f\rangle} \,,$$

so

$$g = \min_{f \perp \mathbf{1}} \frac{\langle (I-P)f, f\rangle}{\langle f, f\rangle} \,. \tag{5.1}$$

As we have seen before, expanding the nominator gives

$$\langle (I-P)f, f\rangle = \frac{1}{2}\sum_{x,y} \pi(x)p(x,y)[f(y)-f(x)]^2 \,,$$

thus

$$g = \min_{f\perp\mathbf{1}} \frac{\sum_{x,y}\pi(x)p(x,y)[f(y)-f(x)]^2}{\sum_{x,y}\pi(x)\pi(y)[f(x)-f(y)]^2} \,.$$

To get $g \le 2\Phi_*$, for any S with $\pi(S) \le 1/2$ define a function f by $f(x) = \pi(S^c)$ for $x \in S$ and $f(x) = \pi(S)$ for $x \notin S$. $\mathbf{E}f = 0$, so $f \perp \mathbf{1}$. Using this f we get that

$$g \le \frac{2Q(S,S^c)}{2\pi(S)\pi(S^c)} \le \frac{2Q(S,S^c)}{\pi(S)} \le 2\Phi_S \,,$$

and so $g \le 2\Phi_*$.

To prove the lower bound we first prove a lemma.

Lemma 5.4. *Given a function $\psi \ge 0$ on a state space V of a stationary Markov chain, order V such that ψ is monotone decreasing and assume $\pi\{\psi > 0\} \le 1/2$, then*

$$\|\psi\|_{L^1(\pi)} \le \Phi_*^{-1} \sum_{x<y} [\psi(x) - \psi(y)]q(x,y) \,.$$

Proof. Let $t > 0$, then $\Phi_* \le \Phi_S$ with $S = \{\psi > t\}$, so we have

$$\pi\{\psi > t\} \le \Phi_*^{-1} \sum_{x,y} q(x,y)\mathbf{1}_{\{\psi(x) > t \ge \psi(y)\}} \,.$$

Now, integrating on t gives

$$\|\psi\|_{L^1(\pi)} \le \Phi_*^{-1} \sum_{x<y} [\psi(x) - \psi(y)]q(x,y) \,,$$

since $\int_0^\infty \pi\{\psi > t\}dt = \|\psi\|_{L^1(\pi)}$ and $\int_0^\infty \mathbf{1}_{\{\psi(x)>t\ge\psi(y)\}} = \psi(x) - \psi(y)$. ∎

Now take an eigenfunction f_2 such that $Pf_2 = \lambda_2 f_2$ and $\pi\{f_2 > 0\} \le 1/2$ (if this does not hold, take $-f_2$). Define a new function $f = \max(f_2, 0)$. Observe that,

$$[(I-P)f](x) \le gf(x) \qquad \forall x\,.$$

This is because if $f(x) = 0$, this translates to $-(Pf)(x) \le 0$ which is true since $f \ge 0$, and if $f(x) > 0$, then, since $f \ge f_2$, we have $[(I-P)f](x) \le [(I-P)f_2](x) \le gf_2(x) = gf(x)$. Since $f \ge 0$, we get

$$\langle (I-P)f, f\rangle \le g\langle f, f\rangle\,,$$

or equivalently,

$$g \ge \frac{\langle (I-P)f, f\rangle}{\langle f, f\rangle}\,.$$

Note that this looks like a contradiction to (5.1), but it is not since f is not orthogonal to $\mathbf{1}$. Denote $R = \frac{\langle (I-P)f,f\rangle}{\langle f,f\rangle}$. Now, apply the lemma with $\psi = f^2$ to get

$$\langle f, f\rangle^2 \le \Phi_*^{-2}\left[\sum_{x<y}(f^2(x) - f^2(y))q(x,y)\right]^2\,.$$

By the Cauchy-Schwarz inequality we get

$$\langle f, f\rangle^2 \le \Phi_*^{-2}\left[\sum_{x<y}(f(x) - f(y))^2 q(x,y)\right]\left[\sum_{x<y}(f(x) + f(y))^2 q(x,y)\right]\,.$$

Recall that $\langle (I-P)f, f\rangle = \sum_{x<y}(f(x)-f(y))^2 q(x,y)$ and use the fact that $(f(x)+f(y))^2 = 2f^2(x) + 2f^2(y) - (f(x)-f(y))^2$ to get that

$$\langle f, f\rangle^2 \le \Phi_*^{-2}\langle (I-P)f, f\rangle[2\langle f, f\rangle - \langle (I-P)f, f\rangle]\,.$$

Divide by $\langle f, f\rangle^2$ to get

$$\Phi_*^2 \le R(2-R)\,,$$

so

$$1 - \Phi_*^2 \ge 1 - 2R + R^2 = (1-R)^2 \ge (1-g)^2\,.$$

One additional computation,

$$\left(1 - \frac{\Phi_*^2}{2}\right)^2 \ge 1 - \Phi_*^2 \ge (1-g)^2\,,$$

yields that $g \ge \frac{\Phi_*^2}{2}$, as required. ∎

A family of graphs $\{G_n\}$ is said to be an (n, d, λ) *expander* family if all of the following three conditions hold for all n:

(i) $|V(G_n)| = n$.

(ii) G_n is d-regular.

(iii) The Cheeger constant of the simple random walk on the graph satisfies $\Phi_*(G_n) \geq \lambda$.

We now construct a a family of 3-regular expander graphs. This is the first construction of an expander family and it is due to Pinsker (1973). Let $G = (V, E)$ be a bipartite graph with equal sides, A and B, each with n vertices. Denote $A, B = \{1, \ldots, n\}$. Draw uniformly at random two permutations $\pi_1, \pi_2 \in S_n$, and set the edge set to be $E = \{(i, i), (i, \pi_1(i)), (i, \pi_2(i)) : 1 \leq i \leq n\}$.

THEOREM 5.5. *With positive probability, G has a positive Cheeger constant, i.e., there exists $\delta > 0$ such that for any $S \subset V$ with $|S| \leq n$ we have*

$$\frac{\#\{\textit{edges between } S \textit{ and } S^c\}}{|S|} > \delta \,.$$

Proof.

It is enough to prove that any $S \subset A$ of size $k \leq \frac{n}{2}$ has at least $(1+\delta)k$ neighbors in B. This is because for any $S \subset V$ simply consider the side in which S has more vertices, and if this side has more than $\frac{n}{2}$ vertices, just look at an arbitrary subset of size exactly $\frac{n}{2}$ vertices. Let $S \subset A$ be a set of size $k \leq \frac{n}{2}$, and denote by $N(S)$ the neighborhood of S. We wish to bound the probability that $|N(S)| \leq (1+\delta)k$. Since (i, i) is an edge for any $1 \leq i \leq k$, we get immediately that $|N(S)| \geq k$. So all we have to enumerate is the surplus δk vertices that a set which contains $N(S)$ will have, and to make sure both $\pi_1(S)$ and $\pi_2(S)$ fall within that set. This argument gives

$$\mathbf{P}\Big[|N(S)| \leq (1+\delta)k\Big] \leq \frac{\binom{n}{\delta k}\binom{(1+\delta)k}{k}^2}{\binom{n}{k}^2} \,,$$

so

$$\mathbf{P}\Big[\exists S \,, |S| \leq \frac{n}{2} \,, |N(S)| \leq (1+\delta)k\Big] \leq \sum_{k=1}^{\frac{n}{2}} \binom{n}{k} \frac{\binom{n}{\delta k}\binom{(1+\delta)k}{\delta k}^2}{\binom{n}{k}^2} \,,$$

which is strictly less than 1 for $\delta > 0$ small enough by Exercise 1. ∎

<u>*Exercises*</u>

1. To complete the proof of Theorem 5.5, prove that there exists $\delta > 0$ such that

$$\sum_{k=1}^{\frac{n}{2}} \frac{\binom{n}{\delta k}\binom{(1+\delta)k}{\delta k}^2}{\binom{n}{k}} < 1 \,.$$

2. Let $\lambda_* = \max_{i\geq 2}|\lambda_i|$ be the second largest (in absolute value) eigenvalue of a finite, irreducible, a-periodic, reversible Markov chain with transition matrix P, and stationary distribution π. Prove that $\lambda_* < 1$.

<u>*Solutions*</u>

1. We bound $\binom{n}{\delta k} \leq \frac{n^{\delta k}}{(\delta k)!}$, similarly $\binom{(1+\delta)k}{\delta k}$ and $\binom{n}{k} \geq \frac{n^k}{k^k}$. This gives

$$\sum_{k=1}^{\frac{n}{2}} \frac{\binom{n}{\delta k}\binom{(1+\delta)k}{\delta k}^2}{\binom{n}{k}} \leq \sum_{k=1}^{\frac{n}{2}} \frac{n^{\delta k}((1+\delta)k)^{2\delta k}k^k}{(\delta k)!^3 n^k}.$$

Recall that for any integer ℓ we have $\ell! > (\ell/e)^\ell$, and bound $(\delta k)!$ by this. We get

$$\sum_{k=1}^{\frac{n}{2}} \frac{\binom{n}{\delta k}\binom{(1+\delta)k}{\delta k}^2}{\binom{n}{k}} \leq \sum_{k=1}^{\log n} \left(\frac{\log n}{n}\right)^{(1-\delta)k} \left[\frac{e^3(1+\delta)^2}{\delta^3}\right]^{\delta k} + \sum_{k=\log n}^{\frac{n}{2}} \left(\frac{k}{n}\right)^{(1-\delta)k} \left[\frac{e^3(1+\delta)^2}{\delta^3}\right]^{\delta k}.$$

The first sum clearly tends to 0 as n tends to ∞, for any $\delta \in (0,1)$, and since $\frac{k}{n} \leq \frac{1}{2}$ and $\frac{1}{2}^{(1-\delta)}\left[\frac{e^3(1+\delta)^2}{\delta^3}\right]^\delta < 0.9$ for $\delta < 0.01$, for any such delta the second sum tends to 0 as n tends to ∞.

2. Define Π to be an $n \times n$ matrix with $\Pi(i,j) = \pi_j$. Fix k such that for all states x, y we have $P^k(x,y) > 0$. Since $\pi_j > 0$ for all j, we can find $\epsilon > 0$ and a stochastic matrix M such that

$$P^k = \epsilon\Pi + (1-\epsilon)M\,. \tag{5.2}$$

Since M is a linear combination of two self-adjoint matrices in $L^2(\pi)$, it is self-adjoint and hence has an orthogonal basis of eigenvectors $\{f_1, \dots, f_n\}$ where $f_1 = (1, \dots, 1)$ and $f_1 \perp f_j$ for all $j > 1$. Observe that if $f_1 \perp f_j$ then $\Pi f_j = 0$. Recall that since M is self-adjoint and stochastic all its eigenvalues are real and are at most 1 in absolute value. By (5.2), it follows that for any $j > 1$ we have that f_j is an eigenvector of P^k with eigenvalue which has absolute value at most $1-\epsilon$, hence $\lambda_* < (1-\epsilon)^{1/k} < 1$.

§6. Harmonic Functions and Random Walks.

Definition Let P be a Markov operator and let f be a real function on the state space. f is called *harmonic* if $Pf = f$, i.e. if

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

Definition Let $\mathcal{S}$ be a measure space. A *coupling* of two $\mathcal{S}$-valued random variables X and X', or of their distributions μ and μ', is a probability measure on $\mathcal{S} \times \mathcal{S}$ having marginals μ and μ', i.e., for every event $A \subset \mathcal{S}$

$$\mathbf{P}\Big[\{(x,x') : x \in A\}\Big] = \mu(A)$$

and

$$\mathbf{P}\Big[\{(x,x') : x' \in A\}\Big] = \mu'(A)\,.$$

Theorem 6.1. *Let P be a transition matrix for a Markov chain on a countable state space V. If for any $x, y \in V$ exists a coupling (X_n, Y_n) such that $\{X_n\}$ and $\{Y_n\}$ are distributed according to the chain, $(X_0, Y_0) = (x, y)$ and*

$$\lim_{n \to \infty} \mathbf{P}[X_n \neq Y_n] = 0\,,$$

then any bounded harmonic function $f : V \to \mathbb{R}$ is constant.

Proof. Let $x, y \in V$ and let (X_n, Y_n) be such a coupling. Let $f : V \to \mathbb{R}$ be a bounded harmonic function. Observe that since f is harmonic, for any n we have

$$\mathbf{E}[f(X_{n+1}) - f(X_n) \mid X_n = x] = \sum_{y \sim x} P(x,y)(f(y) - f(x)) = 0\,,$$

and by taking expectation over x we get that $\mathbf{E}f(X_{n+1}) = \mathbf{E}f(X_n)$. So for any n we have that $|f(x) - f(y)| = |\mathbf{E}f(X_n) - \mathbf{E}f(Y_n)|$, and if $|f|$ is bounded by M, then

$$|f(x) - f(y)| = |\mathbf{E}f(X_n) - \mathbf{E}f(Y_n)| \leq 2M\mathbf{P}[X_n \neq Y_n] \to 0\,,$$

and so $f(x) = f(y)$ for every $x, y \in V$.

∎

Theorem 6.2. *(Blackwell, 1955) All bounded harmonic functions on $\mathbb{Z}^d$ are constant.*

Proof. By Theorem 6.1 it is enough to find a Markov chain and a coupling (X_n, Y_n) such that $(X_0, Y_0) = (x, y)$ and $P(X_n \neq Y_n) \to 0$. We take $\{X_n\}$ and $\{Y_n\}$ to be lazy simple random walks on $\mathbb{Z}^d$ starting at x and y respectively, i.e. for any $x \in \mathbb{Z}^d$ we have $p(x,x) = 1/2$, and the transition probability to any one of the $2d$ neighbors is $1/(4d)$.

We couple $\{X_n\}$ and $\{Y_n\}$ coordinate-wise. Draw a direction $i \in \{1, \dots, d\}$ at random. If $X_n(i) = Y_n(i)$, then with probability $1/2$ leave both at the same position and with probability $1/2$ move them together in direction i. If $X_n(i) \neq Y_n(i)$, choose either X_n or Y_n with probability $1/2$ and move it in direction i. Observe that both walks are distributed as lazy simple random walks. Also, in any coordinate the difference between the walks is a lazy simple one-dimensional random walk with 0 as an absorbing state. This implies that with probability 1 there exists some N such that for all $n > N$ we have $X_n = Y_n$, and thus $P(X_n \neq Y_n) \to 0$, as required. ∎

On infinite regular trees the situation is quite different.

PROPOSITION 6.3. *For any $d \geq 3$, let G be the infinite d-regular tree, and let ρ be its root. Choose one neighbor of ρ and let $A \subset G$ be all the descendants of that vertex. Let f be a real function on G defined by*

$$f(x) = P_x[X_n \in A \text{ for all but finitely many } n]\,.$$

Then f is a non-constant bounded harmonic function.

Proof. Denote by B the event $\{X_n \in A$ for all but finitely many $n\}$. Observe that for any x,

$$f(x) = P_x(B) = \sum_y P_x(X_1 = y) P_y(B) = \sum_y P_x(X_1 = y) f(y)\,,$$

implying that f is a bounded harmonic function.

To show f is not constant, consider $x, y \in A$ such that y is a child of x. Note that the probability of the simple random walk starting at y never to hit x is the same as the probability that a random walk on $\mathbb{Z}$, with probability $1/d$ going left and $1 - 1/d$ going right, starting at 1, never hits 0. If p denotes the above probability, then clearly $f(y) = f(x) + p$, and since $d > 2$, it is known that $p > 0$, and so $f(y) > f(x)$. ∎

Our aim is to formulate a connection between the rate of escape of a Markov chain on a graph, and the existence of non-constant bounded harmonic functions on the graph.

PROPOSITION 6.4. *If $\{X_n\}$ is a simple random walk on a vertex transitive graph then*

$$\lim_{n\to\infty} \frac{\mathbf{E} d(X_0, X_n)}{n}$$

exists.

Proof. The sequence $\mathbf{E}d(X_0, X_n)$ is sub-additive, since using the triangle inequality and the transitivity, we get

$$\mathbf{E}d(X_0, X_{n+m}) \leq \mathbf{E}d(X_0, X_n) + \mathbf{E}d(X_n, X_{n+m}) = \mathbf{E}d(X_0, X_n) + \mathbf{E}d(X_0, X_m),$$

and so by Lemma 6.5 we conclude the proof. ∎

In the proof we used the well-known lemma due to Fekete:

LEMMA 6.5. *If $\{a_n\}$ sub-additive, i.e. $a_{n+m} \leq a_n + a_m$ for all n, m, then*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_n \frac{a_n}{n}.$$

Definition The *lamplighter* group over $\mathbb{Z}$ is a group with elements being all pairs (f, x), where $x \in Z$ and $f : \mathbb{Z} \to \mathbb{Z}_2$, with $f(i) = 0$ for all but finitely many i. Traditionally, x represents the lamplighter's location and f the states of the lamps. The identity element of the group is (f, x) where $x = 0$ and $f(i) = 0$ for all i. We take three generators of the group which yield the following moves:

(i) Lamplighter moves one step left.
(ii) Lamplighter moves one step right.
(iii) Lamplighter changes the current state of the lamp at her position.

Equivalently, the multiplication rule of the group is $(f, x)(g, y) = (h, x + y)$ where $h(i) = f(i) + g(i - x)$.

A natural generalization of this group is the lamplighter group over $\mathbb{Z}^d$, in which the lamplighter moves in d dimensions.

We are interested in the speed of the simple random walk on the Cayley graph of the lamplighter group over $\mathbb{Z}^d$ with the generators specified above.

THEOREM 6.6. *Let $d > 2$ and let $\{X_n\}$ be a random walk on the lamplighter group with the specified generators. Assume that X_n starts at the identity, and that it has the "lazy" property (i.e., with probability $1/4$, we have $X_n = X_{n-1}$, with probability $1/4$, the lamplighter changes the lamp's state, and with probability $1/(4d)$ the lamplighter walks to one of the $2d$ neighbors). Then*

$$\lim_{n\to\infty} \frac{\mathbf{E}d(X_n, id)}{n} > 0.$$

In other words, the walk has positive speed.

To prove Theorem 6.6, we first provide a lemma:

LEMMA 6.7. *Let p denote the probability that a simple random walk on a vertex-transitive graph, starting at the origin, never returns to the origin. Let R_n denote the number of distinct sites visited by the walk up to time n. Then*

$$\lim_{n\to\infty} \frac{\mathbf{E}R_n}{n} = p\,.$$

Proof. Let I_i^n be the indicator random variable indicating whether X_i is a state that is not visited again between time $i+1$ and n, or equivalently

$$I_i^n = \begin{cases} 1, & X_i \notin \{X_{i+1}, \ldots, X_n\} \\ 0, & \text{otherwise.} \end{cases}$$

Clearly $R_n = \sum_{i=1}^n I_i^n$, and for any fixed i we have $\mathbf{E}I_i^n \to p$ as $n \to \infty$, since the graph is vertex-transitive. $\mathbf{E}I_i^n$ is decreasing in n, and increasing in i, hence, for any $\epsilon > 0$, we have m such that for all i and $n > i + m$ we have $p \le \mathbf{E}I_i^n \le p + \epsilon$. So

$$p \le \frac{\mathbf{E}R_n}{n} = \frac{\sum_{i=1}^n \mathbf{E}I_i^n}{n} \le \frac{(p+\epsilon)(n-m)+m}{n} \to p + \epsilon\,.$$

∎

Proof of Theorem 6.6. Let p be the probability that the lazy simple random walk on $\mathbb{Z}^d$, starting at the origin, never returns to the origin. Since $d > 2$, we have $p > 0$. Notice that since all the lamps in the identity element are turned off, $d(X_n, id)$ is at least the number of lit lamps. Note that each lamp that the lamplighter visited has probability at least $1/4$ of being lit after the lamplighter leaves, so by Lemma 6.7,

$$\frac{\mathbf{E}d(X_n, id)}{n} \ge \frac{\mathbf{E}\#\text{ lamps lit after } n \text{ steps}}{n} \ge \frac{\mathbf{E}\#\text{ lamps visited}}{4n} \to p/4 > 0\,.$$

∎

THEOREM 6.8. *All bounded harmonic functions on the lamplighter group on $\mathbb{Z}^d$ are constant for $d = 1, 2$. This is not true for $d > 2$.*

Proof. Assume $d = 1, 2$. As in Theorem 6.2, we couple two copies of a random walk on the group, (f_n, X_n) and (g_n, Y_n), starting from different elements of the group, (f_0, X_0) and (g_0, Y_0). Exactly as in Theorem 6.2, we couple X_n and Y_n (which are random walks on $\mathbb{Z}^d$), so after a finite number of steps, the lamplighter in both random walks is in the same position; however, the states of the lamps can be different. Now couple the lamplighters' moves such that in every step the current lamp will have the same state in both random walks. Since in $d = 1, 2$ the random walk on $\mathbb{Z}^d$ is recurrent, in finite time we will go over all previously visited lamps almost surely, and thus, as before, this implies that the function is constant.

For $d > 2$, define φ on the group by

$$\varphi(f, x) = \mathbf{P}_{(f,x)}[\text{lamp at the origin is on only for a finite amount of time}]\,.$$

It is easy to check that φ is harmonic. To see that φ is not constant, recall that the walk on $\mathbb{Z}^d$, for $d > 2$, is transient, and so we have

$$\lim_{N\to\infty} \mathbf{P}[\exists n > N : X_n = 0] = 0\,.$$

Take N such that $\mathbf{P}[\exists n > N \quad X_n = 0] < 1/3$ and let $x = X_N$. The probability of never returning to the origin from x is less than $1/3$. Now take f such that $f(0) = 0$ and g such that $g(0) = 1$. Clearly $\varphi(f, x) \geq 2/3$ and $\varphi(g, x) \leq 1/3$. ∎

Suppose μ is a probability measure on an Abelian group G such that the set $\{g : \mu(g) > 0\}$ generates G. A function f on G is μ-harmonic if $\int_G f(x + y)d\mu(y) = f(x)$ for all $x \in G$. We now extend Theorem 6.2. The following theorem is due to Choquet and Deny (1960). We present a proof by Raugi (2004).

THEOREM 6.9. *Let h be a bounded μ-harmonic function on G, then for all $x \in G$, for μ-almost every $y \in G$, we have $h(x + y) = h(x)$ (i.e., h is constant).*

Proof. Consider the sequence of functions defined by

$$u_1(x) = \int_G \Big(h(x + y_1) - h(x)\Big)^2 \mu(dy_1)\,,$$

and for all $n \geq 2$

$$u_n(x) = \int_G \Big(h(x + y_1 + \cdots + y_n) - h(x + y_2 + \cdots + y_n)\Big)^2 \mu(dy_1)\ldots\mu(dy_n)\,.$$

Notice that, by Cauchy-Schwarz and since h is μ-harmonic, we get that for all $x \in G$

$$\begin{aligned} u_n(x) &= \int \mu(dy_1)\cdots\mu(dy_{n-1}) \int \Big(h(x + y_1 + \cdots + y_n) - h(x + y_2 + \cdots + y_n)\Big)^2 \mu(dy_n) \\ &\geq \int \mu(dy_1)\cdots\mu(dy_{n-1}) \Big(\int h(x + y_1 + \cdots + y_n) - h(x + y_2 + \cdots + y_n)\mu(dy_n)\Big)^2 \\ &= u_{n-1}(x)\,. \end{aligned}$$

Also, since h is μ-harmonic, it is easy to observe that

$$u_n(x) = \int h^2(x{+}y_1{+}\cdots{+}y_n)\mu(dy_1)\ldots\mu(dy_n) - \int h^2(x{+}y_1{+}\cdots{+}y_{n-1})\mu(dy_1)\ldots\mu(dy_{n-1})\,.$$

So the series $\sum_{n\geq 1} u_n(x)$ is a telescoping series, and since h is bounded, it is also a converging series of increasing non-negative terms, hence $u_1(x) = 0$, which concludes the proof. ∎

Exercises

1. Suppose μ is a probability measure on an Abelian group G such that $\{g : \mu(g) > 0\}$ generates G. Show that any bounded μ-harmonic function on G is constant, using a coupling argument.
2. Given a transition matrix P and a positive harmonic function h on the state space (i.e. $Ph = h$), check that the matrix $\tilde{P}$ defined by

$$\tilde{P}(x, y) = P(x, y)h(y)/h(x)$$

is a transition matrix (this is known as the Doob h-transform).
3. Show that on $\mathbb{Z}^d$, any harmonic function with sublinear growth is constant.
4. Show that on $\mathbb{Z}^d$, any positive harmonic function is constant.

Solutions

1. We may assume without loss of generality that $\mu(e) \geq 1/2$, where e is the unit of G. Indeed, if this is not the case, replace μ with $\tilde{\mu} = \frac{1}{2}(\mu + \nu)$ where ν is the probability measure assigning mass 1 to the identity element e, and note that f is μ-harmonic iff it is $\tilde{\mu}$-harmonic. Let X_n and Y_n be Markov chains on G starting from h_1 and h_2 respectively, with transition rules $p(g, gh) = \mu(h)$. By Theorem 6.1 it suffices to exhibit a coupling between X_n and Y_n so that $\mathbf{P}[X_n \neq Y_n] \to 0$ as $n \to \infty$. For $\mu(g_i) > 0$ and $g_i \neq e$ write:

$$h_1 = g_1^{\alpha_1} g_2^{\alpha_2} \dots g_k^{\alpha_k}$$
$$h_2 = g_1^{\beta_1} g_2^{\beta_2} \dots g_k^{\beta_k}$$

where $g_i \neq g_j$ for $i \neq j$ and $\alpha_i, \beta_i \geq 0$ for all i. Define $\alpha_i(0) = \alpha_i$ and $\beta_i(0) = \beta_i$, and set

$$\alpha_i(n+1) = \begin{cases} \alpha_i(n) + 1, & \text{if } X_{n+1} = X_n g_i \\ \alpha_i(n), & \text{otherwise;} \end{cases}$$
$$\beta_i(n+1) = \begin{cases} \beta_i(n) + 1, & \text{if } X_{n+1} = X_n g_i \\ \beta_i(n), & \text{otherwise.} \end{cases}$$

Also, define $S_n = \{g_i : \alpha_i(n) \neq \beta_i(n)\}$. The coupling is as follows:

i. $X_{n+1} = X_n g$ and $Y_{n+1} = Y_n g$ with probability $\mu(g)$ for $g \notin S_n \cup \{e\}$.
ii. $X_{n+1} = X_n g$ and $Y_{n+1} = Y_n$ with probabilty $\mu(g)/2$ for $g \in S_n$.
iii. $X_{n+1} = X_n$ and $Y_{n+1} = Y_n g$ with probabilty $\mu(g)/2$ for $g \in S_n$.
iv. $X_{n+1} = X_n$ and $Y_{n+1} = Y_n$ with probability $\mu(e) - \mu(S_n)$.

Since $\mu(g_i) > 0$, we see that with probability one $\alpha_i(n) \to \infty$ and $\beta_i(n) \to \infty$ as $n \to \infty$. Restricting to those n_k for which either $\alpha_i(n_k + 1) \neq \alpha_i(n_k)$ or $\beta_i(n_k + 1) \neq \beta_i(n_k)$, we see that $\alpha_i(n_k) - \beta_i(n_k)$ is a simple random walk on $\mathbb{Z}$ with an absorbing state at 0. Recurrence of simple random walk on $\mathbb{Z}$ implies that $\lim_{n\to\infty} \alpha_i(n) - \beta_i(n) = 0$ a.s. Since this is true for all i, we conclude that $\mathbf{P}[X_n \neq Y_n] \to 0$ as $n \to \infty$. Finally, since h_1 and h_2 are arbitrary, f must be constant.

§7. Embeddings of Finite Metric Spaces.

Definition An invertible mapping $f : X \to Y$, where (X, d_X) and (Y, d_Y) are metric spaces, is a C-embedding if there exists a number $r > 0$ such that for all $x, y \in X$

$$r d_X(x,y) \leq d_Y(f(x), f(y)) \leq C r d_X(x,y)\,.$$

The infimum of numbers C such that f is a C-embedding is called the distortion of f and is denoted by $dist(f)$. Equivalently, $dist(f) = \|f\|_{\mathrm{Lip}} \|f^{-1}\|_{\mathrm{Lip}}$, where

$$\|f\|_{\mathrm{Lip}} = \sup\left\{ \frac{d_Y(f(x), f(y))}{d_X(x,y)} : x, y \in X, x \neq y \right\}.$$

We will be interested in embeddings of finite metric spaces and in application of Markov type 2 results to prove lower bounds on distortions of embeddings of certain spaces. We will see that the any embedding of the hypercube $\{0,1\}^k$ in Hilbert space has distortion at least $c\sqrt{k}$, for some $c > 0$ (Enflo, 1969). In Exercise 2 we will show by Markov type arguments that any embedding of an expander family into Hilbert space has distortion at least $\Omega(\log n)$. This was originally shown by Linial, London and Rabinovich (1995) to prove that a theorem of Bourgain (1985), stating that any metric on n points can be embedded in $\ell_p^{\log n}$ with distortion $O(\log n)$, is tight.

We first prove a dimension reduction lemma due to Johnson and Lindenstrauss (1984).

Lemma 7.1. *For any $0 < \epsilon < 1/2$ and $v_1, \ldots, v_n \in \mathbb{R}^n$ with Euclidean metric, there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^k$ where $k = O(\log n/\epsilon^2)$, with distortion at most $1+\epsilon$ on the n point space $\{v_1, \ldots, v_n\}$.*

Proof. Let $A = \frac{1}{\sqrt{k}}(X_i^{(j)})_{1\leq i\leq n, 1\leq j\leq k}$ be an $n \times k$ matrix where the entries $X_i^{(j)}$ are independent standard normal $N(0,1)$ random variables. We prove that with positive probability this map has distortion at most $1+\epsilon$. For any $i \neq j$, let $u = \frac{v_i - v_j}{\|v_i - v_j\|} \in S^{n-1}$, and denote $u = (u_1, \ldots, u_n)$. Clearly,

$$uA = \frac{1}{\sqrt{k}} \left(\sum_{i=1}^{n} u_i X_i^{(1)}, \cdots, \sum_{i=1}^{n} u_i X_i^{(k)} \right).$$

So

$$\|uA\|^2 = \frac{1}{k}\sum_{j=1}^{k}\left(\sum_{i=1}^{n} u_i X_i^{(j)}\right)^2 .$$

Note that for any j the sum $\sum_{i=1}^{n} u_i X_i^{(j)}$ is distributed as a standard normal random variable with mean 0, and since $\sum_{i=1}^{n} u_i^2 = 1$, the variance is 1. So $\|uA\|^2$ is distributed as $\frac{1}{k}\sum_{j=1}^{k} Y_j^2$, where $Y_1, \ldots, Y_k$ are independent standard normal $N(0,1)$ random variables. We wish to show that uA is concentrated around its mean. To achieve that we compute the moment generating function of Y^2 where $Y \sim N(0,1)$. For any real $\lambda < 1/2$ we have

$$\mathbf{E}e^{\lambda Y^2} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\lambda y^2} e^{-y^2/2} dy = \frac{1}{\sqrt{1-2\lambda}} ,$$

and using Taylor expansion we get

$$\begin{aligned}\varphi(\lambda) &= |\log \mathbf{E}e^{\lambda(Y^2-1)}| = |-\frac{1}{2}\log(1-2\lambda) - \lambda| \\ &= \sum_{k=2}^{\infty}\frac{2^{k-1}\lambda^k}{k} \le 2\lambda^2(1 + 2\lambda + (2\lambda)^2 + \cdots) = \frac{2\lambda^2}{1-2\lambda} .\end{aligned}$$

Now,

$$\mathbf{P}[\|uA\|^2 > 1+\epsilon] = \mathbf{P}\left[e^{\lambda\sum_{i=1}^{k}(Y_j^2-1)} > e^{\lambda\epsilon k}\right] \le e^{-\lambda\epsilon k}e^{k\varphi(\lambda)} \le \exp\left(-\lambda\epsilon k + \frac{2\lambda^2 k}{1-2\lambda}\right) .$$

Taking $\lambda = \epsilon/4$ and $k \ge 24\log n/\epsilon^2$, and recalling that $\epsilon < 1/2$, yields

$$\mathbf{P}[\|uA\|^2 > 1+\epsilon] \le \exp(-\epsilon^2 k/12) \le n^{-2} .$$

One can prove similarly that

$$\mathbf{P}[\|uA\|^2 < 1-\epsilon] \le n^{-2} .$$

Since we have $\binom{n}{2}$ pairs of vectors v_i, v_j we showed that with positive probability, for all $i \ne j$,

$$(1-\epsilon)\|v_i - v_j\| \le \|v_i A - v_j A\| \le (1+\epsilon)\|v_i - v_j\| ,$$

which implies that the distortion of A is no more than $1+\epsilon$. ∎

Remark 7.2. From algorithmic perspective it is important to achieve Lemma 7.1 using i.i.d., ± 1 with probability $1/2$, random variables as our $X_i^{(j)}$. This is in fact possible for any random variable X for which there exists a constant $C > 0$ such that $\mathbf{E}e^{\lambda X} \le e^{C\lambda^2}$

(for $X = \pm 1$ with probability 1/2 we have $\mathbf{E}e^{\lambda X} = \cosh(\lambda) \leq e^{\lambda^2/2}$) by the following argument:

Let $Y = \sum_{i=1}^k u_i X_i$ with $\sum_{j=1}^k u_j^2 = 1$ and let Z be distributed $N(0,1)$ independently of $\{X_i\}$. Recall that for all real α we have $\mathbf{E}e^{\alpha Z} = e^{\alpha^2/2}$. Since Y and Z are independent, using Fubini's Theorem we get that for any $\lambda < \frac{C}{2}$

$$\begin{aligned}\mathbf{E}e^{\lambda Y^2} &= \mathbf{E}e^{\frac{(\sqrt{2\lambda}Y)^2}{2}} = \mathbf{E}e^{\sqrt{2\lambda}YZ} = \mathbf{E}e^{\sum_{i=1}^k \sqrt{2\lambda}u_i X_i Z} = \mathbf{E}\,\mathbf{E}\Big[e^{\sum_{i=1}^k \sqrt{2\lambda}u_i X_i Z} \mid Z\Big] \\ &\leq \mathbf{E}e^{C\sum_{i=1}^k \lambda u_i^2 Z^2} = \mathbf{E}e^{C\lambda Z^2} = \frac{1}{\sqrt{1-2C\lambda}}\,,\end{aligned}$$

and the rest of the argument is the same as Lemma 7.1.

Theorem 7.3. *(Bourgain, 1985) Every n-point metric space (X,d) can be embedded in an $O(\log n)$ -dimensional Euclidean space with an $O(\log n)$ distortion.*

Proof. We follow Linial, London and Rabinovich (1995). Let $\alpha > 0$ be determined later. For each cardinality $k < n$ which is a power of 2, randomly pick $\alpha \log n$ sets $A \subset X$ independently, by including each $x \in X$ with probability $1/k$. We have drawn $O(\log^2 n)$ sets $A_1, \ldots, A_{O(\log^2 n)}$. Map every vertex $x \in X$ to the vector

$$\frac{1}{\log n}(d(x,A_1), d(x,A_2), \ldots)\,.$$

Denote this mapping by f. We will show this mapping to $\ell_2^{O(\log^2 n)}$ has almost surely $O(\log n)$ distortion, and using Lemma 7.1 this yields the required result.

It is easy to observe that this map is not expanding. By the triangle inequality, for any $x, y \in X$ and any $A_i \subset X$ we have $|d(x,A_i) - d(y,A_i)| \leq d(x,y)$, so

$$\|f(x) - f(y)\|_2^2 \leq \frac{1}{\log^2 n} \sum_{i=1}^{\alpha \log^2 n} |d(x,A_i) - d(y,A_i)|^2 \leq \alpha d(x,y)^2\,.$$

For the lower bound, let $B(x,\rho) = \{y \in X | d(x,y) \leq \rho\}$ and $B^o(x,\rho) = \{y \in X | d(x,y) < \rho\}$ denote the closed and open balls of radius ρ centered at x. Consider two points $x \neq y \in X$. Let $\rho_0 = 0$, and let ρ_t be the least radius ρ for which both $|B(x,\rho)| \geq 2^t$ and $|B(y,\rho)| \geq 2^t$. We define ρ_t as long as $\rho_t < \frac{1}{4}d(x,y)$, and let $\hat{t}$ be the largest such index. Also let $\rho_{\hat{t}+1} = \frac{d(x,y)}{4}$. Observe that $B(y,\rho_j)$ and $B(x,\rho_i)$ are always disjoint.

Notice that $A \cap B^o(x,\rho_t) = \varnothing \iff d(x,A) \geq \rho_t$, and $A \cap B(y,\rho_{t-1}) \neq \varnothing \iff d(y,A) \leq \rho_{t-1}$. Therefore, if both conditions hold, then $|d(y,A) - d(x,A)| \geq \rho_t - \rho_{t-1}$.

Let us assume that $|B^o(x,\rho_t)| < 2^t$ (otherwise we argue for y). On the other hand, $|B(y,\rho_{t-1})| \geq 2^{t-1}$. Let $k = 2^t$ and let $A \subset X$ be chosen randomly by including each $x \in X$ with probability $1/k$. We have

$$\mathbf{P}[A \text{ misses } B^o(x,\rho_t)] \geq (1-2^{-t})^{2^t} \geq \frac{1}{4}\,,$$

and

$$\mathbf{P}[A \text{ hits } B(y,\rho_{t-1})] \geq 1-(1-2^{-t})^{2^{t-1}} \geq 1-e^{-1/2} \geq \frac{1}{2}\,.$$

Since these events are independent, such an A has probability at least $\frac{1}{8}$ to both intersect $B(y,\rho_{t-1})$ and miss $B^o(x,\rho_t)$. Since for each k we choose $\alpha \log n$ such sets, by Theorem 1.1, the probability that less than $\frac{\alpha \log n}{16}$ of them have the previous property is less than

$$e^{-2(\alpha \log n/16)^2/(\alpha \log n)} \leq n^{-\alpha/128} \leq n^{-5}\,,$$

by choosing α such that $\alpha/128 > 5$. So with probability tending to 1, for any $x, y \in X$ and k we have at least $\alpha \log n/16$ sets which satisfy the condition. Summing it up gives

$$\|f(x)-f(y)\|_2^2 \geq \frac{1}{\log^2 n} \sum_{i=1}^{\hat{t}+1} \frac{\alpha \log n}{16} (\rho_i - \rho_{i-1})^2\,.$$

Since $\sum_{i=1}^{\hat{t}+1}(\rho_i - \rho_{i-1}) = \rho_{\hat{t}+1} = \frac{d(x,y)}{4}$, we have

$$\|f(x)-f(y)\|_2^2 \geq \frac{\alpha}{16 \log n}\left(\frac{d(x,y)}{4(\hat{t}+1)}\right)^2 (\hat{t}+1) \geq \frac{\alpha d(x,y)^2}{256(\hat{t}+1)\log n} \geq \frac{\alpha d(x,y)^2}{256 \log^2 n}\,,$$

hence the distortion of f is $O(\log n)$ with probability tending to 1. ∎

Proposition 7.4. *(Enflo, 1969) There exists $c > 0$ such that any embedding of the hypercube $\{0,1\}^k$ in Hilbert space has distortion at least $c\sqrt{k}$.*

Proof. Recall that in Exercise 1 of Chapter 3 we proved that if $\{X_j\}$ is a simple random walk in the hypercube, then

$$\mathbf{E}d(X_0, X_j) \geq \frac{j}{2} \qquad \forall j \leq k/4\,.$$

Take $j = \frac{k}{4}$. By Jensen's inequality, $\mathbf{E}d^2(X_0, X_{k/4}) \geq k^2/64$. Now let $f : \{0,1\}^k \to L^2$ be a map. Assume without loss of generality that f is a non-expanding, i.e., $\|f\|_{\text{Lip}} = 1$ (otherwise, take $f/\|f\|_{\text{Lip}}$). By Theorem 3.1 it follows that L^2 has Markov type 2 with constant $M = 1$, so,

$$\mathbf{E}d^2\Big(f(X_0), f(X_{k/4})\Big) \leq k\,.$$

We conclude

$$\|f^{-1}\|_{\text{Lip}}^2 k \geq \|f^{-1}\|_{\text{Lip}}^2 \mathbf{E}d^2(f(X_0), f(X_{k/4})) \geq \mathbf{E}d^2(X_0, X_{k/4}) \geq k^2/64\,,$$

hence $\|f^{-1}\|_{\text{Lip}} \geq \frac{\sqrt{k}}{8}$, which implies the result. ∎

Remark 7.5. Enflo's original proof gives $c = 1$. See Exercise 1 for the proof of this fact.

We now prove a theorem of Bourgain (1986).

Theorem 7.6. *Any embedding of a binary tree of depth M and $n = 2^{M+1} - 1$ vertices into a Hilbert space has distortion $\Omega(\sqrt{\log M}) = \Omega(\sqrt{\log \log n})$.*

Remark 7.7. See Exercise 3 for an embedding with distortion $O(\sqrt{\log M})$.

We first prove two lemmas.

Lemma 7.8. *Let $M = 2^m$ for $m \geq 1$ and $y_0, \ldots, y_n \in \mathbb{R}$, then*

$$\sum_{i=1}^{M} (y_i - y_{i-1})^2 = \frac{(y_M - y_0)^2}{M} + \sum_{k=1}^{m} \frac{1}{2^k} \sum_{j=1}^{2^{m-k}} \left(y_{j2^k} - 2y_{(2j-1)2^{k-1}} + y_{(j-1)2^k}\right)^2 .$$

Proof. This can be proved by induction on m, however, we will prove it using Parseval's identity. Consider the Haar orthonormal basis of $\mathbb{R}^M$ which is defined by the following vectors: for any $1 \leq k \leq m$ and any $1 \leq j \leq 2^{m-k}$ let $I(k;j)$ denote the set of indices $\{(j-1)2^k + 1, \ldots, j2^k\}$ and define

$$\psi_{I(k;j)}(i) = \begin{cases} \frac{1}{2^{k/2}}, & (j-1)2^k < i \leq (2j-1)2^{k-1}\,; \\ -\frac{1}{2^{k/2}}, & (2j-1)2^{k-1} < i \leq j2^k. \end{cases}$$

Together with the vector $\psi_1 = \frac{1}{\sqrt{M}}(1, \ldots, 1)$ this gives 2^m orthonormal vectors in $\mathbb{R}^M$. Now define $z \in \mathbb{R}^M$ by $z_i = y_i - y_{i-1}$, so the LHS of the lemma becomes $\sum_{i=1}^{M} z_i^2$, which, by Parseval's identity, is

$$\langle z, z\rangle = \langle z, \psi_1\rangle^2 + \sum_{k=1}^{m} \sum_{j=1}^{2^{m-k}} \langle z, \psi_{I(k;j)}\rangle^2 ,$$

which can easily be seen to be the RHS of the lemma. ∎

Lemma 7.9. *Let $M = 2^m$, and suppose that $Y_0, Y_1, \ldots$ is a function of a Markov chain taking values in Hilbert space. For any $1 \leq k \leq m$ and $1 \leq j \leq 2^{m-k}$ let $r = (2j-1)2^{k-1}$ and let $\tilde{Y}(k;j)$ denote the random process which is equal to $\{Y_t\}$ for time $t \leq r$ and evolves independently for time $t > r$. Write $\mathcal{A}_{i=1}^{M}(\cdot) = \frac{1}{M}\sum_{i=1}^{M}(\cdot)$ for the averaging operator. Then*

$$\mathbf{E}\left[\mathcal{A}_{i=1}^{M} \|Y_i - Y_{i-1}\|^2\right] \geq \mathbf{E}\left[\frac{1}{2}\sum_{k=1}^{m} \mathcal{A}_{j=1}^{2^{m-k}} \frac{\|Y_{j2^k} - \tilde{Y}_{j2^k}(k;j)\|^2}{2^{2k}}\right] .$$

Proof. Since all the distances in the lemma are squared, we can assume without loss of generality that $\{Y_t\}$ is real valued. Let k, j be as in the lemma. Write $\mathbf{E}_r(\cdot) = \mathbf{E}[\cdot \mid Y_r]$

and $\tilde{Y} = \tilde{Y}(k;j)$. Let $t > r$ and denote $\mu_r = \mathbf{E}_r[Y_t]$. Note that by the definition of $\tilde{Y}$, we have that Y_t and $\tilde{Y}_t$ are independent given Y_r, and so $\mathbf{E}_r[Y_t\tilde{Y}_t] = \mathbf{E}_r[Y_t]\mathbf{E}_r[\tilde{Y}_t]$. Also, since Y_t has the same distribution as $\tilde{Y}_t$, we have

$$\mathbf{E}_r|Y_t - \tilde{Y}_t|^2 = \mathbf{E}_r|(Y_t - \mu_r) - (\tilde{Y}_t - \mu_r)|^2 = 2\mathbf{E}_r(Y_t - \mu_r)^2 \leq 2\mathbf{E}_r(Y_t - \lambda_r)^2\,,$$

for any λ_r which is Y_r-measurable. The last inequality follows from the fact that $\mathbf{E}_r(Y_t - \mu_r)^2$ is the squared length of the projection of Y_t on the space of Y_r-measurable functions. Taking expectation w.r.t to Y_r on the last inequality gives

$$\mathbf{E}(Y_t - \lambda_r)^2 \geq \frac{1}{2}\mathbf{E}(Y_t - \tilde{Y}_t)^2\,. \tag{7.1}$$

Now apply Lemma 7.8 with $y_i = Y_i$:

$$\mathcal{A}_{i=1}^{M}(Y_i - Y_{i-1})^2 \geq \sum_{k=1}^{m} \mathcal{A}_{j=1}^{2^{m-k}} \frac{(Y_{j2^k} - 2Y_{(2j-1)2^{k-1}} + Y_{(j-1)2^k})^2}{2^{2k}}\,.$$

Take expectations and apply (7.1) with $\lambda_r = -2Y_{(2j-1)2^{k-1}} + Y_{(j-1)2^k}$ to get

$$\mathbf{E}\mathcal{A}_{i=1}^{M}(Y_i - Y_{i-1})^2 \geq \frac{1}{2}\sum_{k=1}^{m} \mathbf{E}\mathcal{A}_{j=1}^{2^{m-k}} \frac{(Y_{j2^k} - \tilde{Y}_{j2^k}(k;j))^2}{2^{2k}}\,,$$

as required. ∎

Proof of Theorem 7.6. Let T denote the full binary tree with depth $M = 2^m$ (for general depths, consider the tree up to depth which a power of 2) . Let $\{Z_i\}$ be the forward random walk on it starting from the root (i.e., at each vertex it goes right/left with probability 1/2). Clearly $d(Z_i, Z_{i+1})^2 = 1$ a.s., so $\mathbf{E}\mathcal{A}_{i=1}^{M} d(Z_i, Z_{i-1})^2 = 1$. Also, in the forward random walk, after the splitting at time r, with probability 1/2 the two independent walks will accumulate distance which is twice the number of steps. Thus, $\mathbf{E}d^2(Z_{j2^k}, \tilde{Z}_{j2^k}(k;j)) \geq 2^{2k-1}$, and we get that

$$\mathbf{E}\mathcal{A}_{i=1}^{M} d^2(Z_i, Z_{i-1}) = 1 \leq \frac{2}{m}\sum_{k=1}^{m} \mathbf{E}\mathcal{A}_{j=1}^{2^{m-k}} \frac{d^2(Z_{j2^k}, \tilde{Z}_{j2^k}(k;j))}{2^{2k}}\,.$$

Now let $F : T \to H$ be an embedding with $\|F\|_{\mathrm{Lip}} = 1$, then the previous inequality holds for $F(Z_i)$ up to a factor of $\|F^{-1}\|_{\mathrm{Lip}}$, i.e.

$$\mathbf{E}\mathcal{A}_{i=1}^{M} d^2(F(Z_i), F(Z_{i-1})) \leq \frac{2\|F^{-1}\|_{\mathrm{Lip}}^2}{m}\sum_{k=1}^{m} \mathbf{E}\mathcal{A}_{j=1}^{2^{m-k}} \frac{d^2(F(Z_{j2^k}), F(\tilde{Z}_{j2^k}(k;j)))}{2^{2k}}\,.$$

By Lemma 7.9 we have

$$\mathbf{E}\mathcal{A}_{i=1}^{M} d^2(F(Z_i), F(Z_{i-1})) \geq \mathbf{E}\left[\frac{1}{2}\sum_{k=1}^{m} \mathcal{A}_{j=1}^{2^{m-k}} \frac{d^2(F(Z_{j2^k}) - F(\tilde{Y}_{j2^k}(k;j)))}{2^{2k}}\right]\,,$$

which, combined with previous inequality, yields $\|F^{-1}\|_{\mathrm{Lip}}^2 \geq \frac{m}{4}$, as required. ∎

Exercises

1. (Enflo, 1969) Let $\Omega_d = \{-1, 1\}^d$ be the d-dimensional hypercube with ℓ_1 metric. Show that any $f : \Omega_d \to L^2$ has distortion at least $\sqrt{d}$ and give an example for an f with $\sqrt{d}$ distortion.
2. Use the proof that $\mathbb{R}$ has Markov type 2 to show that for an (n, d, λ)-expander family, any invertible mapping f of the vertices to a Hilbert space have

$$\|f\|_{\mathrm{Lip}} \|f^{-1}\|_{\mathrm{Lip}} \geq C_{d,\lambda} \log n \,,$$

where $C_{d,\lambda}$ is constant.

Solutions

1. As an example, take f to be the identity embedding $\Omega_d \to \ell_2^d$. Clearly $d_2(f(x), f(y)) \leq d(x, y)$, and by Cauchy-Schwarz we have

$$d(x, y) = 2 \sum_{i=1}^{d} \mathbf{1}_{(x_i \neq y_i)} \leq d_2(x, y)\sqrt{d} \,.$$

For the lower bound, a simple computation gives that for any $x_1, x_2, x_3, x_4 \in L^2$ we have that the sum of squares of the diagonals is smaller than the sum of squared edges lengths

$$d(x_1, x_4)^2 + d(x_2, x_3)^2 \leq d(x_1, x_2)^2 + d(x_2, x_4)^2 + d(x_3, x_4)^2 + d(x_1, x_3)^2 \,.$$

This can be easily extended by induction on d to

$$\sum_{x \in \Omega_d} \|f(x) - f(-x)\|^2 \leq \sum_{x \sim y} \|f(x) - f(y)\|^2 \,.$$

Assume now that $L > 0$ satisfies

$$d_1(x, y) \leq \|f(x) - f(y)\|_2 \leq L d_1(x, y) \,.$$

Then

$$\sum_{x \sim y} \|f(x) - f(y)\|^2 \leq 4L^2 d 2^d \,,$$

and also

$$\sum_{x \in \Omega_d} \|f(x) - f(-x)\|^2 \geq 4d^2 2^d \,,$$

which, together with the previous argument, imply that $L \geq \sqrt{d}$, as required.

2. Let $\{X_j\}$ be the simple random walk on the expander, with transition matrix P. Since the family is d-regular, the random walk has uniform stationary distribution. By Theorem 5.2 it follows that for any $x, y \in V$

$$P^t(x,y) \le \pi(y) + e^{-gt}\,,$$

where $g = 1-\lambda$. Take $\alpha > 0$ such that $g\alpha < 1$, and take $t = \alpha \log n$. Then

$$P^t(x,y) \le \frac{1}{n} + e^{-g\alpha \log n} \le 2e^{-g\alpha \log n}\,.$$

Fix $\gamma > 0$ small enough such that $d^\gamma e^{-g\alpha} < 1$. We wish to show that up to time $t = \alpha \log n$, the random walk on the expander has positive speed. Indeed, for any $x \in V$, since the ball $B(x, \gamma \log n)$ of radius $\gamma \log n$ around x has at most $d^{\gamma \log n}$ vertices, it follows that

$$\mathbf{P}_x[X_t \in B(x, \gamma \log n)] \le d^{\gamma \log n} 2e^{-g\alpha \log n} \to 0\,.$$

This in turn implies that for large enough n

$$\mathbf{E}d^2(X_0, X_t) > \frac{\gamma^2 \log^2 n}{2}\,.$$

Let $f : V \to L^2$, and assume without loss of generality that $\|f\|_{\mathrm{Lip}} = 1$ (otherwise take $f/\|f\|_{\mathrm{Lip}}$). Recall that in the proof of Theorem 3.1 we actually proved that

$$\mathbf{E}d(f(X_t), f(X_0))^2 \le (1 + \lambda + \lambda^2 + \cdots + \lambda^{t-1})\mathbf{E}d(f(Z_1), f(Z_0))^2\,.$$

This immediately implies that

$$\mathbf{E}d^2(f(X_0), f(X_t)) \le \frac{1}{g}\,.$$

Now, similarly to the proof of Proposition 7.4 we conclude that $\|f^{-1}\|_{\mathrm{Lip}} \ge \sqrt{g}\gamma \log n/\sqrt{2}$.

§8. The Ising Model, Potts Model and Random Cluster Model.

Let $G = (V, E)$ be a finite graph and let β be a real positive number. The *Ferromagnetic Ising model* at temperature $t = \frac{1}{\beta}$ is a probability space over all $\sigma : V \to \{-1, 1\}$ with measure μ_β defined by

$$\mu_\beta(\sigma) = \frac{1}{Z} e^{\beta \sum_{x\sim y} \sigma(x)\sigma(y)}, \tag{8.1}$$

where $Z = \sum_\sigma e^{\beta \sum_{x\sim y} \sigma(x)\sigma(y)}$ is the normalizing factor. This measure is called the *Gibbs measure.*

Note that (8.1) can also be written as

$$\mu_\beta(\sigma) = \frac{1}{Z} e^{\beta \left[|E| - 2\sum_{x\sim y} \mathbf{1}_{(\sigma(x)\neq\sigma(y))}\right]} = \frac{1}{\tilde{Z}} e^{-2\beta \sum_{x\sim y} \mathbf{1}_{(\sigma(x)\neq\sigma(y))}}, \tag{8.2}$$

where $\tilde{Z}$ is the new normalizing factor. The *Potts model* is a generalization of the Ising model in which the probability space is over all $\sigma : V \to \{1, \ldots, q\}$ for some integer $q \geq 2$ with measure as in (8.2)

$$\mu_{\beta,q}(\sigma) = \frac{1}{\tilde{Z}} e^{-2\beta \sum_{x\sim y} \mathbf{1}_{(\sigma(x)\neq\sigma(y))}}.$$

This measure is called the Gibbs *free* measure. An important variation is imposing boundary conditions on the model. Fix $B \subset V$ and let $\gamma : B \to \{1, \ldots, q\}$. The Potts model with boundary condition γ is a probability space on all $\sigma : V \to \{1, \ldots, q\}$ such that $\sigma_{|B} = \gamma$ again with measure as in (8.2)

$$\mu^\gamma_{\beta,q}(\sigma) = \frac{1}{Z} e^{-2\beta \sum_{x\sim y} \mathbf{1}_{(\sigma(x)\neq\sigma(y))}}, \tag{8.3}$$

where Z is the normalizing factor.

We will consider now as the underlying graph the square, i.e. the induced subgraph of $\mathbb{Z}^2$ on the vertex set $\{-n, \ldots, n\}^2$. Since the Gibbs free measure is invariant under permutations of $\{1, \ldots, q\}$, it is clear that $\sigma(v)$ is uniform on $\{1, \ldots, q\}$ for any vertex v. Denote by μ^+_β the Gibbs measure of the Potts model on the square with boundary conditions $B = \{(i, j) : \max(|i|, |j|) = n\}$ and $\gamma(v) = 1$ for all $v \in B$ (i.e., all 1's on the boundary of the square).

Under these boundary conditions, the distribution of the center $(0, 0)$, as n tends to infinity, experiences a phase transition.

Theorem 8.1. *There exists* $0 < \beta_0 < \beta_1 < \infty$ *such that for all* $\beta < \beta_0$ *and all* $j \in \{1, \ldots, q\}$

$$\lim_{n\to\infty} \left|\mu^+_\beta(\sigma((0,0)) = j) - \frac{1}{q}\right| = 0\,,$$

and for all $\beta > \beta_1$

$$\limsup_{n\to\infty} \left|\mu^+_\beta(\sigma((0,0)) = 1) - \frac{1}{q}\right| > 0\,.$$

In order to prove Theorem 8.1 we detour to Percolation theory on $\mathbb{Z}^2$.

Let $G = (V, E)$ be a finite or infinite graph and let $p \in [0, 1]$. Bernoulli Percolation on G is a probability measure on subgraphs of G such that each edge is deleted from the graph with probability $1-p$ and remains in it with probability p independently of all other edges (Kolmogorov's consistency theorem is used in the case G is infinite). The deleted edges are called *closed* edges, while the edges remaining are called *open*.

On an infinite graph, by Kolmogorov's 0-1 law, the event "there exists an infinite open component" has always probability 0 or 1. On the other hand, the probability of the event "the origin is contained in an infinite open component", $\Theta(p)$, is strictly smaller than 1 for any $p < 1$ and it experiences a phase transition as p goes from 0 to 1.

Theorem 8.2. *In* $\mathbb{Z}^2$ *we have* $\Theta(p) = 0$ *for all* $p < \frac{1}{3}$ *and* $\Theta(p) > 0$ *for all* $p > \frac{2}{3}$.

Proof. Denote by V_n the square $[-n, n]^2$ and by $o \leftrightarrow \partial V_n$ the event that the origin is connected by an open path to the boundary of V_n. Clearly

$$\Theta(p) = \lim_{n\to\infty} \mathbf{P}_p[o \leftrightarrow \partial V_n]\,.$$

Assume now $p < \frac{1}{3}$ and let Γ be the set of all self-avoiding paths from the origin to the boundary of V_n. Since the paths Γ can never backtrack there are at most $43^{\ell-1}$ such paths of length ℓ (this bound is obviously crude, but it suffices), thus

$$\mathbf{P}_p[o \leftrightarrow \partial V_n] \le \sum_{\psi\in\Gamma} \mathbf{P}_p[\psi \text{ is open}] = \sum_{\ell\ge n} \#\{\psi \in \Gamma \text{ with length } \ell\} p^l \le \sum_{\ell\ge n} 43^{\ell-1} p^\ell \to 0\,.$$

Now assume $p > \frac{2}{3}$. For a subgraph G of $\mathbb{Z}^2$ define the dual graph to the graph where the set of vertices are the faces of $\mathbb{Z}^2$ and an edge exists in the dual graph iff the edge connecting the two faces is nonexistent in G. A moment of reflection gives that the origin is not connected to the boundary of V_n if and only if there exists an open closed path (i.e., a path of open edges beginning and ending at the same vertex) in the dual graph which separates the origin from V_n. The important observation is that in $\mathbb{Z}^2$ the dual of the percolation model is distributed as a percolation model on $\mathbb{Z}^2$ with edge probability $1-p$ instead of p. Let k be a natural number such that

$$\sum_{\ell \geq k} 4\ell 3^{\ell-1}(1-p)^{\ell-1} < 1\,.$$

Since $p > \frac{2}{3}$, such k exists. Now, for any $n > k$, if all the edges in V_k are open, and ∂V_k is connected to ∂V_n by an open path, then clearly the origin is connected to ∂V_n, so we have

$$\mathbf{P}_p[o \leftrightarrow \partial V_n] \geq p^{4(k+1)^2}\mathbf{P}_p[\partial V_k \leftrightarrow \partial V_n]\,.$$

If ∂V_k is not connected to ∂V_n by an open path then there exists an open closed path in the dual graph surrounding V_k of length $\ell \geq k$. This path must cross the ray $[0,\infty)$ in distance at most ℓ and we consider this to be the starting point of the path. Summing this up, we have at most $4\ell 3^{\ell-1}$ possible paths, which gives

$$\mathbf{P}_p[\partial V_k \not\leftrightarrow \partial V_n] \leq \sum_{\ell \geq k} 4\ell 3^{\ell-1}(1-p)^{\ell-1} < 1\,,$$

and so we get that $\lim_{n\to\infty} \mathbf{P}_p[\partial V_k \leftrightarrow \partial V_n] > 0$ which in turn implies that $\lim_{n\to\infty} \mathbf{P}_p[o \leftrightarrow \partial V_n] > 0$, as required. ∎

Remark 8.3. In fact, in $\mathbb{Z}^2$, we have $\inf\{p : \Theta(p) > 0\} = \frac{1}{2}$ (see Kesten, 1980).

We continue with preparations for the proof of Theorem 8.1 by introducing the *Random Cluster* model (see Fortuin and Kasteleyn, 1972). Let $G = (V,E)$ be a graph. For an edge configuration $\eta : E \to \{0,1\}$ we write $k(\eta)$ for the number of connected components in the subgraph of G containing all vertices and all edges e such that $\eta(e) = 1$. The Random Cluster model with parameters $p \in [0,1]$ and $q > 0$ is a probability space over all $\eta : E \to \{0,1\}$, with measure

$$\phi_{p,q}(\eta) = \frac{1}{Z_{p,q}} p^{\sum_e \eta(e)}(1-p)^{\sum_e (1-\eta(e))} q^{k(\eta)}\,.$$

For integer $q > 0$ we define the model with boundary conditions. Let $B \subset V$ and $\gamma : B \to \{1,\ldots,q\}$, define $\tilde{k}(\eta)$ to be the number of components in η that do not contain any vertex of B. We define the random cluster model with the restriction that vertices in B which are connected by an open path of η have the same value under γ; in other words we define the measure

$$\phi^{\gamma}_{p,q}(\eta) = \frac{1}{Z^{\gamma}_{p,q}} p^{\sum_e \eta(e)}(1-p)^{\sum_e (1-\eta(e))} q^{\tilde{k}(\eta)} \prod_{v,u \in B, \gamma(v) \neq \gamma(u)} \Big(1 - \mathbf{1}_{(u \leftrightarrow v)}\Big)\,. \qquad (8.4)$$

We now present a coupling argument which binds the Potts model and the Random Cluster model for integer q. Let $G = (V, E)$ be a graph, and let $B \subset V$ and $\gamma : B \to \{1, \ldots, q\}$ be boundary conditions. We define a measure on $\{1, \ldots, q\}^V \times \{0,1\}^E$ by assigning each vertex in $V - B$ a value chosen from $\{1, \ldots, q\}$ according to uniform distribution, then assigning each edge value 1 or 0 independently with probability p and $1-p$ respectively and do all this conditioned on the event that no two vertices with different values (values of vertices of B are defined by γ) have an edge with value 1 connecting them. In other words, for all $\sigma \in \{1, \ldots, q\}^V$ such that $\sigma_{|B} = \gamma$ and $\eta \in \{0,1\}^E$, we have

$$\mathbf{P}^{\gamma}_{p,q}[\sigma, \eta] = \frac{1}{\tilde{Z}^{\gamma}_{p,q}} p^{\sum_e \eta(e)} (1-p)^{\sum_e (1-\eta(e))} \mathbf{1}_{\{(\sigma(x)-\sigma(y))\eta(e)=0\}} ,$$

where again $\tilde{Z}^{\gamma}_{p,q}$ is the appropriate normalizing factor, and the probability of any (σ, η) such that $\sigma_{|B} \neq \gamma$ is 0. The measure $\mathbf{P}^{\gamma}_{p,q}$ was introduced by Swendsen and Wang (1987) and made more explicit by Edwards and Sokal (1988).

The following theorem states that the vertex marginal of $\mathbf{P}^{\gamma}_{p,q}$ is the Gibbs measure for the Potts model with the same boundary conditions, and the edge marginal is the Random Cluster model with the same boundary conditions.

THEOREM 8.4. *Let π_V and π_E be the projection measures of $\mathbf{P}^{\gamma}_{p,q}$ on $\{1, \ldots, q\}^V$ and $\{0,1\}^E$ respectively. Let $\beta = -\frac{1}{2}\log(1-p)$, then*

$$\pi_V = \mu^{\gamma}_{\beta,q} ,$$

and

$$\pi_E = \phi^{\gamma}_{p,q} ,$$

where $\mu^{\gamma}_{\beta,q}$ is the Gibbs measure for the Potts model (8.3), and $\phi^{\gamma}_{p,q}$ is the Random Cluster model (8.4).

Proof. We simply sum up the marginals. Let Z be the normalizing factor in $\mathbf{P}^{\gamma}_{p,q}$. Fix $\sigma \in \{1, \ldots, q\}^V$ such that $\sigma_{|B} = \gamma$, and sum up over all $\eta \in \{0,1\}^E$:

$$\pi_V(\sigma) = \sum_{\eta \in \{0,1\}^E} \mathbf{P}^{\gamma}_{p,q}[\sigma, \eta] = \frac{1}{Z} \sum_{\eta \in \{0,1\}^E} p^{\sum_e \eta(e)} (1-p)^{\sum_e (1-\eta(e))} \mathbf{1}_{\{(\sigma(x)-\sigma(y))\eta(e)=0\}} .$$

Now note that in each of the non-zero items in the summation, for any edge $e = (x, y) \in E$ such that $\sigma(x) \neq \sigma(y)$, we must have $\eta(e) = 0$, so all terms have in common the product of

$(1-p)$ over all such edges. What is left is clearly the expansion of $\prod_{(x,y)\in E,\sigma(x)=\sigma(y)}(p+(1-p))=1$. So we get

$$\pi_V(\sigma)=\frac{1}{Z}\prod_{(x,y)\in E,\sigma(x)\neq\sigma(y)}(1-p)=\frac{1}{Z}e^{-2\beta\sum_{x\sim y}\mathbf{1}_{(\sigma(x)\neq\sigma(y))}}=\mu^{\gamma}_{\beta,q}(\sigma)\,,$$

where the last equality is true since Z must be the same normalization factor as in (8.3).

Now we fix $\eta\in\{0,1\}^E$. If there are two vertices in the boundary B with different γ values and an open η path connecting them, then for any σ we have $\mathbf{P}^{\gamma}_{p,q}[\sigma,\eta]=0$. This is because the open path has different values on its two endpoints, thus, for σ complying with γ, somewhere on this path we have an edge connecting two vertices with different σ values.

So it is safe to assume that η has no open path between any two vertices in B with different boundary values. We sum over all $\sigma\in\{1,\dots,q\}^V$ such that $\sigma_{|B}=\gamma$. Note that given such η there are exactly $q^{\tilde{k}(\eta)}$ configurations σ that are allowed (since all the values of σ on components containing a boundary vertex have one possible value, the γ value of that vertex). We get

$$\pi_E(\eta)=\sum_{\sigma}\mathbf{P}^{\gamma}_{p,q}=q^{\tilde{k}(\eta)}\frac{1}{Z}p^{\sum_e\eta(e)}(1-p)^{\sum_e(1-\eta(e))}\prod_{v,u\in B,\gamma(v)\neq\gamma(u)}\Big(1-\mathbf{1}_{(u\leftrightarrow v)}\Big)=\phi^{\gamma}_{p,q}(\eta)\,.$$

∎

Corollary 8.5. *Let $p=1-e^{-2\beta}$ and suppose we pick a random configuration $\sigma\in\{1,\dots,q\}^V$ as follows:*

1. *Pick a random edge configuration $\eta\in\{0,1\}^E$ according to the Random Cluster measure $\phi^{\gamma}_{p,q}$.*
2. *For each connected component C of η which does not contain any boundary vertex, pick a value uniformly at random from $\{1,\dots,q\}$, assign this value to every vertex of C and do this independently for all such components. To any component containing a boundary vertex assign the boundary value to each vertex of the component.*

Then σ is distributed according to the Gibbs measure $\mu^{\gamma}_{\beta,q}$ of the Potts model with the same boundary conditions.

We consider now the case when G is the square, i.e. the induced subgraph of $\mathbb{Z}^2$ on the vertex set $\{-n,\dots,n\}^2$. We take the boundary to be $B=\{(i,j):\max(|i|,|j|)=n\}$ and $\gamma(v)=1$ for all $v\in B$. Let $\phi^{\gamma}_{p,q}$ be the Random Cluster measure, then for any edge $e=(x,y)\in E$ and any configuration $\eta\in\{0,1\}^{E\setminus\{e\}}$, if x and y are connected via open edges in η, or both are connected by η to the boundary, we have

$$\phi^{\gamma}_{p,q}(e\text{ is open}|\eta)=p\,,$$

and otherwise

$$\phi^\gamma_{p,q}(e \text{ is open}|\eta) = \frac{p}{p+(1-p)q}\,.$$

This implies that for any partial configuration $\eta \in \{0,1\}^{E'}$ where $E' \subset E$ and $e \notin E'$ we have

$$\frac{p}{p+(1-p)q} \leq \phi^\gamma_{p,q}(e \text{ is open}|\eta) \leq p\,, \tag{8.5}$$

since $\phi^\gamma_{p,q}(e \text{ is open}|\eta)$ is just a weighted average of all configurations that miss exactly e and contain η.

Let $\mathbf{P}_p$ be the Bernoulli Percolation measure on G with edge probability p. (8.5) allows us to prove easily that the Random Cluster model is stochastically dominated above and below by Bernoulli Percolation.

LEMMA 8.6. *Let $p \in [0,1]$, $q \geq 1$ and $\tilde{p} = \frac{p}{p+(1-p)q}$, then*

$$\mathbf{P}_{\tilde{p}} \preceq \phi^\gamma_{p,q} \preceq \mathbf{P}_p\,.$$

Proof. Denote the edges E by $e_1,\ldots,e_m$. Let $U_1,\ldots,U_m$ be independent uniform $[0,1]$ random variables. We define two edge configurations $w, w' : E \to \{0,1\}$. For any edge $e_i \in E$ we have $w(e_i) = 1$ if and only if $U_i \leq \tilde{p}$. For any $e_i \in E$ let $\eta_i : \{e_1,\ldots,e_{i-1}\} \to \{0,1\}$ be a partial edge configuration determined by w, i.e. for any $j < i$ we have $\eta_i(e_j) = w(e_j)$. Now for any $e_i \in E$ let $w'(e_i) = 1$ if and only if $U_i \leq \phi^\gamma_{p,q}(e_i \text{ is open}|\eta_i)$. It is easy to verify that w and w' are distributed as $\mathbf{P}_{\tilde{p}}$ and $\phi^\gamma_{p,q}$ respectively, and that for any $e \in E$, because of (8.5), we have $w(e) \leq w'(e)$ with probability 1. This coupling shows $\mathbf{P}_{\tilde{p}} \preceq \phi^\gamma_{p,q}$, and the other side of the inequality is proven similarly. ∎

We are finally ready to prove Theorem 8.1.

Proof of Theorem 8.1. Take $\beta_0 = -\frac{1}{2}\log(1-\frac{1}{3})$, let $\beta < \beta_0$ and set $p = 1 - e^{-2\beta} < \frac{1}{3}$. Draw a random configuration $\sigma \in \{1,\ldots,q\}^V$ according to $\mu^+_{\beta,q}$ and to Corollary 8.5. It is clear that if the origin is not connected to the boundary, $\sigma((0,0))$ is uniformly distributed on $\{1,\ldots,q\}$, so for any $j \in \{1,\ldots,q\}$

$$\left|\mu^+_{\beta,q}(\sigma((0,0)) = j) - \frac{1}{q}\right| \leq \phi^\gamma_{p,q}\Big(o \leftrightarrow \partial V_n\Big)\,.$$

But, since the event $o \leftrightarrow \partial V_n$ is increasing, by Lemma 8.6 and Theorem 8.2 we have

$$\phi^\gamma_{p,q}\Big(o \leftrightarrow \partial V_n\Big) \leq \mathbf{P}_p[o \leftrightarrow \partial V_n] \to 0\,,$$

concluding the first part of the theorem.

Similarly, since for any q, we have $\tilde{p} = \frac{p}{p+(1-p)q} \to 1$ as $p \to 1$, we may choose p_0 such that $\tilde{p}_0 > \frac{2}{3}$. Take $\beta_1 = -\frac{1}{2}\log(1-p_0)$, let $\beta > \beta_1$ and set $p' = 1 - e^{-2\beta} > p_0$. Again, draw a random configuration $\sigma \in \{1,\ldots,q\}^V$ according to $\mu^+_{\beta,q}$ and to Corollary 8.5. If the origin is connected to the boundary, its σ value is 1 with probability 1, otherwise it is uniformly distributed on $\{1,\ldots,q\}$, so we get

$$\Big|\mu^+_{\beta,q}(\sigma((0,0)) = 1) - \frac{1}{q}\Big| = \Big(1 - \frac{1}{q}\Big)\phi^\gamma_{p',q}\Big(o \leftrightarrow \partial V_n\Big)\,.$$

Since $\tilde{p'} > 2/3$, by Lemma 8.6 and Theorem 8.2 we have

$$\limsup_{n\to\infty} \phi^\gamma_{p',q}\Big(o \leftrightarrow \partial V_n\Big) \geq \limsup_{n\to\infty} \mathbf{P}_{\tilde{p}'}\Big[o \leftrightarrow \partial V_n\Big] > 0\,,$$

concluding the second part of the theorem. ∎

REFERENCES

Aldous D. and Fill J.A. (2005). Reversible markov chains and random walks on graphs, available at http://www.stat.berkeley.edu/ aldous/RWG/book.html.

Alon, N. (1986). Eigenvalues and expanders. Combinatorica 6, no. 2, 83–96.

Ball K. (1992). Markov Chains, Riesz Transforms and Lipschitz Maps, Geom. Funct. Anal. 2, 137–172.

Blackwell D. (1955). On transient Markov processes with a countable number of states and stationary transition probabilities. Ann. Math. Statist. 26, 654–658.

Bourgain J. (1985). On Lipschitz embedding of finite metric spaces in Hilbert space. Israel J. Math. 52 (1985), no. 1-2, 46–52.

Bourgain J. (1986). The metrical interpretation of superreflexivity in Banach spaces. Israel J. Math. 56, no. 2, 222–230.

Choquet G., Deny J. (1960). Sur l'quation de convolution $= *\sigma$. (French) C. R. Acad. Sci. Paris 250 1960 799–801.

Chung F.R.K., Frankl P., Graham R. and Shearer J.B. (1986). Some intersection theorems for ordered sets and graphs, J. Combin. Theory Ser. A. 48, 23-37.

Durrett R. (1996). Probability: Theory and Examples, second edition. Duxbury Press, Belmont CA.

Edwards R. G., Sokal A. D. (1988). Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. Phys. Rev. D (3) 38 (1988), no. 6, 2009–2012.

Fortuin, C. M., Kasteleyn, P. W. (1972). On the random-cluster model. I. Introduction and relation to other models. Physica 57 (1972), 536–564.

Gessel I.M. and Stanley R.P. (1995). Algebraic enumeration in Handbook of Combinatorics, vol. 2, Elsevier, Amsterdam, and MIT Press, Cambridge, pp. 1021-1061.

Han T. S. (1978). Nonnegative entropy measures of multivariate symmetric correlations. Information and Control 36, 2 133–156.

Hoeffding W. (1963). Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58 13–30.

Jerrum M. and Sinclair A. (1989). Approximating the permanent. SIAM J. Comput. 18, no. 6, 1149–1178.

Johnson W. B., Lindenstrauss J. (1984). Extensions of Lipschitz mappings into a Hilbert space. Conference in modern analysis and probability (New Haven, Conn., 1982), 189–206, Contemp. Math., 26, Amer. Math. Soc., Providence, RI.

Kesten H. (1980). The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. Comm. Math. Phys. 74 (1980), no. 1, 41–59.

Lawler G.F. and Sokal A.D. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. Trans. Amer. Math. Soc. 309, no. 2, 557–580.

Linial N., London E. and Rabinovich Y. (1995). The geometry of graphs and some of its algorithmic applications, , Combinatorica 15, no. 2, 215-245.

Loomis L.H. and Whitney, H. (1949). An inequality related to the isoperimetric inequality. Bull. Amer. Math. Soc 55, 961–962.

Matthews P. (1988). Covering problems for Markov chains. Ann. Probab. 16, no. 3, 1215–1228.

Naor A., Peres Y., Schramm O. and Sheffield S. (2004). Markov chains in smooth Banach spaces and Gromov hyperbolic metric spaces. http://front.math.ucdavis.edu /math.FA/0410422

Pinsker M. (1973). On the complexity of a concentrator. Proc 7th Int. Teletraffic Congr., Stockholm, pp. 318/1-318/4.

Raugi A. (2004). A general Choquet-Deny theorem for nilpotent groups. Ann. Inst. H. Poincar Probab. Statist. 40 (2004), no. 6, 677–683.

Steele, M.J. (1997). Probability Theory and Combinatorial Optimization, Society for Industrial and Applied Mathematics.

Swendsen R. H., Wang J. S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. Phys Rev Lett. 1987; 58(2):86-88.