Chapter 5

Spectral methods

In this chapter, we develop spectral techniques. We highlight some applications to Markov chain mixing and network analysis. The main tools are the *spectral theorem* and the *variational characterization of eigenvalues*, which we review in Section 5.1 together with some related results. We also give a brief introduction to *spectral graph theory* and detail an application to community recovery. In Section 5.2 we apply the spectral theorem to reversible Markov chains. In particular we define the *spectral gap* and establish its close relationship to the mixing time. Roughly speaking, we show through an eigendecomposition of the transition matrix that the gap between the eigenvalue 1 (which is the largest in absolute value) and the rest of the spectrum drives how fast P^t converges to the stationary distribution. We give several examples. We then show in Section 5.3 that the spectral gap can be bounded using certain isoperimetric properties of the underlying network. We prove *Cheeger's inequality*, which quantifies this relationship, and introduce expander graphs, an important family of graphs with good "expansion." Applications to mixing times are also discussed. One specific technique is the "canonical paths method," which bounds the spectral graph by formalizing a notion of congestion in the network.

5.1 Background

We first review some important concepts from linear algebra. In particular, we recall the spectral theorem as well as the variational characterization of eigenvalues. We also derive a few perturbation results. We end this section with an application

Version: December 20, 2023
Modern Discrete Probability: An Essential Toolkit

to community recovery in network analysis.

5.1.1 Eigenvalues and their variational characterization

When a matrix $A \in \mathbb{R}^{d\times d}$ is *symmetric*, that is, $a_{ij} = a_{ji}$ for all i, j, a remarkable result is that A is similar to a diagonal matrix by an orthogonal transformation. Put differently, there exists an orthonormal basis of $\mathbb{R}^d$ made of eigenvectors of A. Recall that a matrix $Q \in \mathbb{R}^{d\times d}$ is *orthogonal* if $QQ^T = I_{d\times d}$ and $Q^T Q = I_{d\times d}$, where $I_{d\times d}$ is the $d \times d$ identity matrix. In words, its columns form an orthonormal basis of $\mathbb{R}^d$. For a vector $\mathbf{z} = (z_1, \ldots, z_d)$, we let $\mathrm{diag}(\mathbf{z}) = \mathrm{diag}(z_1, \ldots, z_d)$ be the diagonal matrix with diagonal entries $z_1, \ldots, z_d$. Unless specified otherwise, a vector is by default a "column vector" and its transpose is a "row vector."

symmetric matrix

orthogonal matrix

Theorem 5.1.1 (Spectral theorem)**.** *Let $A \in \mathbb{R}^{d\times d}$ be a symmetric matrix, that is, $A^T = A$. Then A has d orthonormal eigenvectors $\mathbf{q}_1, \ldots, \mathbf{q}_d$ with corresponding (not necessarily distinct) real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. In matrix form, this is written as the matrix factorization*

spectral theorem

$$A = Q\Lambda Q^T = \sum_{i=1}^{d} \lambda_i \mathbf{q}_i \mathbf{q}_i^T,$$

where Q has columns $\mathbf{q}_1, \ldots, \mathbf{q}_d$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$. We refer to this factorization as a spectral decomposition of A.

The proof uses a greedy sequence maximizing the quadratic form $\langle \mathbf{v}, A\mathbf{v}\rangle$. For a hint as to why that might come about, note that for a unit eigenvector $\mathbf{v}$ with eigenvalue λ we have $\langle \mathbf{v}, A\mathbf{v}\rangle = \langle \mathbf{v}, \lambda\mathbf{v}\rangle = \lambda$.

We will need the following formula. Consider the block matrices

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where $\mathbf{y} \in \mathbb{R}^{d_1}$, $\mathbf{z} \in \mathbb{R}^{d_2}$, $A \in \mathbb{R}^{d_1\times d_1}$, $B \in \mathbb{R}^{d_1\times d_2}$, $C \in \mathbb{R}^{d_2\times d_1}$, and $D \in \mathbb{R}^{d_2\times d_2}$. Then it follows by direct calculation that

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix}^T \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} = \mathbf{y}^T A\mathbf{y} + \mathbf{y}^T B\mathbf{z} + \mathbf{z}^T C\mathbf{y} + \mathbf{z}^T D\mathbf{z}. \tag{5.1.1}$$

We will also need the following linear algebra fact. Let $\mathbf{v}_1, \ldots, \mathbf{v}_j$ be orthonormal vectors in $\mathbb{R}^d$, with $j < d$. Then they can be completed into an orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_d$ of $\mathbb{R}^d$.

Proof of Theorem 5.1.1. We proceed by induction.

A first eigenvector Let $A_1 = A$. Maximizing over the objective function $\langle \mathbf{v}, A_1 \mathbf{v}\rangle$, we let

$$\mathbf{v}_1 \in \arg\max\{\langle \mathbf{v}, A_1 \mathbf{v}\rangle : \|\mathbf{v}\|_2 = 1\},$$

and

$$\lambda_1 = \max\{\langle \mathbf{v}, A_1 \mathbf{v}\rangle : \|\mathbf{v}\|_2 = 1\}.$$

Complete $\mathbf{v}_1$ into an orthonormal basis of $\mathbb{R}^d$, $\mathbf{v}_1$, $\hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_d$, and form the block matrix

$$\hat{W}_1 := \begin{pmatrix} \mathbf{v}_1 & \hat{V}_1 \end{pmatrix}$$

where the columns of $\hat{V}_1$ are $\hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_d$. Note that $\hat{W}_1$ is orthogonal by construction.

Getting one step closer to diagonalization We show next that $\hat{W}_1$ gets us one step closer to a diagonal matrix by similarity transformation. Note first that

$$\hat{W}_1^T A_1 \hat{W}_1 = \begin{pmatrix} \lambda_1 & \mathbf{w}_1^T \\ \mathbf{w}_1 & A_2 \end{pmatrix}$$

where $\mathbf{w}_1 := \hat{V}_1^T A_1 \mathbf{v}_1$ and $A_2 := \hat{V}_1^T A_1 \hat{V}_1$. The key claim is that $\mathbf{w}_1 = \mathbf{0}$. This follows from an argument by contradiction.

Suppose $\mathbf{w}_1 \neq \mathbf{0}$ and consider the unit vector

$$\mathbf{z} := \hat{W}_1 \times \frac{1}{\sqrt{1 + \delta^2 \|\mathbf{w}_1\|_2^2}} \begin{pmatrix} 1 \\ \delta \mathbf{w}_1 \end{pmatrix}$$

which achieves objective value

$$\begin{aligned}
\mathbf{z}^T A_1 \mathbf{z} &= \frac{1}{1 + \delta^2 \|\mathbf{w}_1\|_2^2} \begin{pmatrix} 1 \\ \delta \mathbf{w}_1 \end{pmatrix}^T \begin{pmatrix} \lambda_1 & \mathbf{w}_1^T \\ \mathbf{w}_1 & A_2 \end{pmatrix} \begin{pmatrix} 1 \\ \delta \mathbf{w}_1 \end{pmatrix} \\
&= \frac{1}{1 + \delta^2 \|\mathbf{w}_1\|_2^2} \left(\lambda_1 + 2\delta \|\mathbf{w}_1\|_2^2 + \delta^2 \mathbf{w}_1^T A_2 \mathbf{w}_1\right),
\end{aligned}$$

where we used (5.1.1). By the Taylor expansion,

$$\frac{1}{1 + \epsilon^2} = 1 - \epsilon^2 + O(\epsilon^4),$$

for δ small enough,

$$\begin{aligned}
\mathbf{z}^T A_1 \mathbf{z} &= (\lambda_1 + 2\delta \|\mathbf{w}_1\|_2^2 + \delta^2 \mathbf{w}_1^T A_2 \mathbf{w}_1)(1 - \delta^2 \|\mathbf{w}_1\|_2^2 + O(\delta^4)) \\
&= \lambda_1 + 2\delta \|\mathbf{w}_1\|_2^2 + O(\delta^2) \\
&> \lambda_1.
\end{aligned}$$

That gives the desired contradiction.

So, letting $W_1 := \hat{W}_1$,

$$W_1^T A_1 W_1 = \begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{pmatrix}.$$

Finally note that $A_2 = \hat{V}_1^T A_1 \hat{V}_1$ is symmetric since

$$A_2^T = (\hat{V}_1^T A_1 \hat{V}_1)^T = \hat{V}_1^T A_1^T \hat{V}_1 = \hat{V}_1^T A_1 \hat{V}_1 = A_2,$$

by the symmetry of A_1 itself.

Next step of the induction Apply the same argument to the symmetric submatrix $A_2 \in \mathbb{R}^{(d-1)\times(d-1)}$, let $\hat{W}_2 \in \mathbb{R}^{(d-1)\times(d-1)}$ be the corresponding orthogonal matrix, and define λ_2 and A_3 through the equation

$$\hat{W}_2^T A_2 \hat{W}_2 = \begin{pmatrix} \lambda_2 & \mathbf{0} \\ \mathbf{0} & A_3 \end{pmatrix}.$$

Define the block matrix

$$W_2 = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \hat{W}_2 \end{pmatrix}$$

and observe that

$$\begin{aligned} W_2^T W_1^T A_1 W_1 W_2 &= W_2^T \begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{pmatrix} W_2 \\ &= \begin{pmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & \hat{W}_2^T A_2 \hat{W}_2 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & 0 & \mathbf{0} \\ 0 & \lambda_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_3 \end{pmatrix}. \end{aligned}$$

Proceeding similarly by induction gives the claim, with the final Q being the product of the W_is (which is orthogonal as the product of orthogonal matrices). ■

We derive an important variational characterization inspired by the proof of the spectral theorem. We will need the following quantity.

Definition 5.1.2 (Rayleigh quotient). *Let $A \in \mathbb{R}^{d\times d}$ be a symmetric matrix. The* Rayleigh quotient *of A is defined as*

Rayleigh quotient

$$\mathcal{R}_A(\mathbf{u}) = \frac{\langle \mathbf{u}, A\mathbf{u}\rangle}{\langle \mathbf{u}, \mathbf{u}\rangle}$$

which is defined for any $\mathbf{u} \neq \mathbf{0}$ in $\mathbb{R}^d$.

We let the *span* of a collection of vectors be defined as

span

$$\mathrm{span}(\boldsymbol{u}_{\mathbf{1}},\ldots,\boldsymbol{u}_n) := \left\{\sum_{i=1}^n \alpha_i \boldsymbol{u}_i \,:\, \alpha_1,\ldots,\alpha_n \in \mathbb{R}\right\}.$$

Courant-Fischer

Theorem 5.1.3 (Courant-Fischer theorem)**.** *Let $A \in \mathbb{R}^{d\times d}$ be a symmetric matrix with spectral decomposition $A = \sum_{i=1}^d \lambda_i \mathbf{v}_i \mathbf{v}_i^T$ where $\lambda_1 \geq \cdots \geq \lambda_d$. For each $k = 1,\ldots,d$, define the subspace*

$$\mathcal{V}_k = \mathrm{span}(\mathbf{v}_1,\ldots,\mathbf{v}_k) \quad \textit{and} \quad \mathcal{W}_{d-k+1} = \mathrm{span}(\mathbf{v}_k,\ldots,\mathbf{v}_d).$$

Then, for all $k = 1,\ldots,d$,

$$\lambda_k = \min_{\mathbf{u}\in\mathcal{V}_k} \mathcal{R}_A(\mathbf{u}) = \max_{\mathbf{u}\in\mathcal{W}_{d-k+1}} \mathcal{R}_A(\mathbf{u}).$$

Furthermore we have the following min-max formulas, which do not depend on the choice of spectral decomposition, for all $k = 1,\ldots,d$

$$\lambda_k = \max_{\dim(\mathcal{V})=k} \min_{\mathbf{u}\in\mathcal{V}} \mathcal{R}_A(\mathbf{u}) = \min_{\dim(\mathcal{W})=d-k+1} \max_{\mathbf{u}\in\mathcal{W}} \mathcal{R}_A(\mathbf{u}).$$

Note that, in all these formulas, the vector $\mathbf{u} = \mathbf{v}_k$ is optimal. To derive the "local" formula, the first ones above, we expand a vector in $\mathcal{V}_k$ into the basis $\mathbf{v}_1,\ldots,\mathbf{v}_k$ and use the fact that $\mathcal{R}_A(\mathbf{v}_i) = \lambda_i$ and that eigenvalues are in nonincreasing order. The "global" formulas then follow from a dimension argument.

We will need the following dimension-based fact. Let $\mathcal{U},\mathcal{V} \subseteq \mathbb{R}^d$ be linear subspaces such that $\dim(\mathcal{U}) + \dim(\mathcal{V}) > d$, where $\dim(\mathcal{U})$ denotes the dimension of $\mathcal{U}$. Then there exists a nonzero vector in the intersection $\mathcal{U}\cap\mathcal{V}$. That is,

$$\dim(\mathcal{U}) + \dim(\mathcal{V}) > d \implies (\mathcal{U}\cap\mathcal{V})\setminus\{\mathbf{0}\} \neq \emptyset. \tag{5.1.2}$$

Proof of Theorem 5.1.3. We first prove the local formulas, that is, the ones involving a *specific* decomposition.

Local formulas Since $\mathbf{v}_1,\ldots,\mathbf{v}_k$ form an orthonormal basis of $\mathcal{V}_k$, any nonzero vector $\mathbf{u}\in\mathcal{V}_k$ can be written as $\mathbf{u} = \sum_{i=1}^k \langle \mathbf{u},\mathbf{v}_i\rangle \mathbf{v}_i$ and it follows that

$$\langle \mathbf{u},\mathbf{u}\rangle = \sum_{i=1}^k \langle \mathbf{u},\mathbf{v}_i\rangle^2$$

$$\langle \mathbf{u},A\mathbf{u}\rangle = \left\langle \mathbf{u}, \sum_{i=1}^k \langle \mathbf{u},\mathbf{v}_i\rangle \lambda_i \mathbf{v}_i\right\rangle = \sum_{i=1}^k \lambda_i \langle \mathbf{u},\mathbf{v}_i\rangle^2.$$

Thus,

$$\mathcal{R}_A(\mathbf{u}) = \frac{\langle \mathbf{u}, A\mathbf{u}\rangle}{\langle \mathbf{u}, \mathbf{u}\rangle} = \frac{\sum_{i=1}^k \lambda_i \langle \mathbf{u}, \mathbf{v}_i\rangle^2}{\sum_{i=1}^k \langle \mathbf{u}, \mathbf{v}_i\rangle^2} \geq \lambda_k \frac{\sum_{i=1}^k \langle \mathbf{u}, \mathbf{v}_i\rangle^2}{\sum_{i=1}^k \langle \mathbf{u}, \mathbf{v}_i\rangle^2} = \lambda_k$$

where we used $\lambda_1 \geq \cdots \geq \lambda_k$ and the fact that $\langle \mathbf{u}, \mathbf{v}_i\rangle^2 \geq 0$. Moreover $\mathcal{R}_A(\mathbf{v}_k) = \lambda_k$. So we have established

$$\lambda_k = \min_{\mathbf{u} \in \mathcal{V}_k} \mathcal{R}_A(\mathbf{u}).$$

The expression in terms of $\mathcal{W}_{d-k+1}$ is proved similarly.

Global formulas Since $\mathcal{V}_k$ has dimension k, it follows from the local formula that

$$\lambda_k = \min_{\mathbf{u} \in \mathcal{V}_k} \mathcal{R}_A(\mathbf{u}) \leq \max_{\dim(\mathcal{V})=k} \min_{\mathbf{u} \in \mathcal{V}} \mathcal{R}_A(\mathbf{u}).$$

Let $\mathcal{V}$ be any subspace with dimension k. Because $\mathcal{W}_{d-k+1}$ has dimension $d-k+1$, we have that $\dim(\mathcal{V}) + \dim(\mathcal{W}_{d-k+1}) > d$ and there must be nonzero vector $\mathbf{u}_0$ in the intersection $\mathcal{V} \cap \mathcal{W}_{d-k+1}$ by the dimension-based fact above. We then have by the other local formula that

$$\lambda_k = \max_{\mathbf{u} \in \mathcal{W}_{d-k+1}} \mathcal{R}_A(\mathbf{u}) \geq \mathcal{R}_A(\mathbf{u}_0) \geq \min_{\mathbf{u} \in \mathcal{V}} \mathcal{R}_A(\mathbf{u}).$$

Since this inequality holds for any subspace of dimension k, we have

$$\lambda_k \geq \max_{\dim(\mathcal{V})=k} \min_{\mathbf{u} \in \mathcal{V}} \mathcal{R}_A(\mathbf{u}).$$

Combining with the inequality in the other direction above gives the claim. The other global formula is proved similarly. ■

5.1.2 Elements of spectral graph theory

We apply the variational characterization of eigenvalues to matrices arising in graph theory. In this section, graphs have no self-loop.

Unweighted graphs As we have previously seen, a convenient way of specifying a graph is through a matrix representation. Assume the undirected graph $G = (V, E)$ has $n = |V|$ vertices. Recall that the adjacency matrix A of G is the $n \times n$ symmetric matrix defined as

$$A_{xy} = \begin{cases} 1 & \text{if } \{x, y\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Another matrix of interest is the Laplacian matrix. It is related to the Laplace operator we encountered previously. We will show in particular that it contains useful information about the connectedness of the graph. Recall that, given a graph $G = (V, E)$, the quantity $\delta(v)$ denotes the degree of $v \in V$.

Definition 5.1.4 (Graph Laplacian)**.** *Let $G = (V, E)$ be a graph with vertices $V = \{1, \ldots, n\}$ and adjacency matrix $A \in \mathbb{R}^{n\times n}$. Let $D = \mathrm{diag}(\delta(1), \ldots, \delta(n))$ be the degree matrix. The* graph Laplacian *(or* Laplacian matrix*, or* Laplacian *for short) associated to G is defined as $L = D - A$. Its entries are*

graph Laplacian

$$l_{ij} = \begin{cases} \delta(i) & \text{if } i = j, \\ -1 & \text{if } \{i, j\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Observe that the Laplacian L of a graph G is a symmetric matrix:

$$L^T = (D - A)^T = D^T - A^T = D - A,$$

where we used that both D and A are themselves symmetric. The associated quadratic form is particularly simple and will play an important role.

Lemma 5.1.5 (Laplacian quadratic form)**.** *Let $G = (V, E)$ be a graph with $n = |V|$ vertices. Its Laplacian L is a positive semi-definite matrix and furthermore we have the following formula for the* Laplacian quadratic form *(or* Dirichlet energy*)*

Laplacian quadratic form

$$\mathbf{x}^T L\mathbf{x} = \sum_{e=\{i,j\}\in E} (x_i - x_j)^2,$$

for any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Proof of Lemma 5.1.5. Let B be an oriented incidence matrix of G (see Definition 1.1.16). We claim that $BB^T = L$. Indeed, for $i \neq j$, entry (i, j) of BB^T is a sum over all edges containing i and j as endvertices, of which there is at most one. When $e = \{i, j\} \in E$, that entry is -1, since one of i or j has a 1 in the column of B corresponding to e and the other one has a -1. For $i = j$, letting b_{xy} be entry (x, y) of B,

$$(BB^T)_{ii} = \sum_{e=\{x,y\}\in E: i\in e} b_{xy}^2 = \delta(i).$$

That shows that $BB^T = L$ entry-by-entry.

For any $\mathbf{x}$, we have $(B^T\mathbf{x})_k = x_v - x_u$ if the edge $e_k = \{u, v\}$ is oriented as (u, v) under B. That implies

$$\mathbf{x}^T L\mathbf{x} = \mathbf{x}^T BB^T\mathbf{x} = \|B^T\mathbf{x}\|_2^2 = \sum_{e=\{i,j\}\in E} (x_i - x_j)^2.$$

Since the latter quantity is always nonnegative, it also implies that L is positive semidefinite. ■

As a convention, we denote the eigenvalues of a Laplacian matrix L by

$$0 \leq \mu_1 \leq \mu_2 \leq \cdots \leq \mu_n,$$

and we will refer to them as *Laplacian eigenvalues*. Here is a simple observation. For any $G = (V, E)$, the constant unit vector

Laplacian eigenvalues

$$\mathbf{y}_1 = \frac{1}{\sqrt{n}}(1, \ldots, 1),$$

is an eigenvector of the Laplacian with eigenvalue 0. Indeed, let B be an oriented incidence matrix of G and recall from the proof of Lemma 5.1.5 that $L = BB^T$. By construction $B^T\mathbf{y}_1 = \mathbf{0}$ since each column of B has exactly one 1 and one -1. So $L\mathbf{y}_1 = BB^T\mathbf{y}_1 = \mathbf{0}$ as claimed. In general, the constant vector may not be the only eigenvector with eigenvalue one.

We are now ready to derive connectivity consequences. Recall that, for any graph G, the Laplacian eigenvalue $\mu_1 = 0$.

Lemma 5.1.6 (Laplacian and connectivity)**.** *If G is connected, then the Laplacian eigenvalue $\mu_2 > 0$.*

Proof. Let $G = (V, E)$ with $n = |V|$ and let $L = \sum_{i=1}^n \mu_i \mathbf{y}_i \mathbf{y}_i^T$ be a spectral decomposition of its Laplacian L with $0 = \mu_1 \leq \cdots \leq \mu_n$. Suppose by way of contradiction that $\mu_2 = 0$. Any eigenvector $\mathbf{y} = (y_1, \ldots, y_n)$ with 0 eigenvalue satisfies $L\mathbf{y} = \mathbf{0}$ by definition. By Lemma 5.1.5 then

$$0 = \mathbf{y}^T L \mathbf{y} = \sum_{e=\{i,j\} \in E} (y_i - y_j)^2.$$

In order for this to hold, it must be that any two adjacent vertices i and j have $y_i = y_j$. That is, $\{i, j\} \in E$ implies $y_i = y_j$. Furthermore, because G is connected, between any two of its vertices u and v (adjacent or not) there is a path $u = w_0 \sim \cdots \sim w_k = v$ along which the y_ws must be the same. Thus $\mathbf{y}$ is a constant vector.

But that is a contradiction since the eigenvectors $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are in fact linearly independent, so that $\mathbf{y}_1$ and $\mathbf{y}_2$ cannot both be a constant vector. ■

The quantity μ_2 is sometimes referred to as the *algebraic connectivity* of the graph. The corresponding eigenvector, $\mathbf{y}_2$, is known as the *Fiedler vector*.

Fiedler vector

We will be interested in more quantitative results of this type. Before proceeding, we start with a simple observation. By our proof of Theorem 5.1.1, the largest eigenvalue μ_n of the Laplacian L is the solution to the optimization problem

$$\mu_n = \max\{\langle \mathbf{x}, L\mathbf{x}\rangle : \|\mathbf{x}\|_2 = 1\}.$$

Such extremal characterization is useful in order to bound the eigenvalue μ_n, since any choice of $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$ gives a lower bound through the quantity $\langle \mathbf{x}, L\mathbf{x}\rangle$. We give a simple consequence.

Lemma 5.1.7 (Laplacian and degree)**.** *Let $G = (V, E)$ be a graph with maximum degree $\bar{\delta}$. Let μ_n be the largest Laplacian eigenvalue. Then*

$$\mu_n \geq \bar{\delta} + 1.$$

Proof. Let $u \in V$ be a vertex with degree $\bar{\delta}$. Let $\mathbf{z}$ be the vector with entries

$$z_i = \begin{cases} \bar{\delta} & \text{if } i = u, \\ -1 & \text{if } \{i, u\} \in E, \\ 0 & \text{otherwise,} \end{cases}$$

and let $\mathbf{x}$ be the unit vector $\mathbf{z}/\|\mathbf{z}\|_2$. By definition of the degree of u, $\|\mathbf{z}\|_2^2 = \bar{\delta}^2 + \bar{\delta}(-1)^2 = \bar{\delta}(\bar{\delta} + 1)$.

Using the Lemma 5.1.5,

$$\begin{aligned} \langle \mathbf{z}, L\mathbf{z}\rangle &= \sum_{e=\{i,j\}\in E} (z_i - z_j)^2 \\ &\geq \sum_{i:\{i,u\}\in E} (z_i - z_u)^2 \\ &= \sum_{i:\{i,u\}\in E} (-1 - \bar{\delta})^2 \\ &= \bar{\delta}(\bar{\delta} + 1)^2, \end{aligned}$$

where we restricted the sum to those edges incident with u and used the fact that all terms in the sum are nonnegative. Finally

$$\langle \mathbf{x}, L\mathbf{x}\rangle = \left\langle \frac{\mathbf{z}}{\|\mathbf{z}\|_2}, L\frac{\mathbf{z}}{\|\mathbf{z}\|_2} \right\rangle = \frac{1}{\|\mathbf{z}\|_2^2}\langle \mathbf{z}, L\mathbf{z}\rangle = \frac{\bar{\delta}(\bar{\delta} + 1)^2}{\bar{\delta}(\bar{\delta} + 1)} = \bar{\delta} + 1,$$

so that

$$\mu_n = \max\{\langle \mathbf{x}', L\mathbf{x}'\rangle : \|\mathbf{x}'\|_2 = 1\} \geq \langle \mathbf{x}, L\mathbf{x}\rangle = \bar{\delta} + 1,$$

as claimed. ∎

A special case of Courant-Fischer (Theorem 5.1.3) for the Laplacian matrix is the following.

Corollary 5.1.8 (Variational characterization of μ_2)**.** *Let $G = (V, E)$ be a graph with $n = |V|$ vertices. Assume the Laplacian L of G has spectral decomposition $L = \sum_{i=1}^n \mu_i \mathbf{y}_i \mathbf{y}_i^T$ with $0 = \mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$ and $\mathbf{y}_1 = \frac{1}{\sqrt{n}}(1, \ldots, 1)$. Then*

$$\mu_2 = \min\left\{\frac{\sum_{\{u,v\}\in E}(x_u - x_v)^2}{\sum_{u=1}^n x_u^2} : \mathbf{x} = (x_1, \ldots, x_n) \neq \mathbf{0}, \sum_{u=1}^n x_u = 0\right\}.$$

Proof. By Theorem 5.1.3,

$$\mu_2 = \min_{\mathbf{x}\in\mathcal{W}_{d-1}} \mathcal{R}_L(\mathbf{x}).$$

Since $\mathbf{y}_1$ is constant and $\mathcal{W}_{d-1}$ is the subspace orthogonal to it, this is equivalent to restrictring the minimization to those nonzero $\mathbf{x}$s such that

$$0 = \langle \mathbf{x}, \mathbf{y}_1 \rangle = \frac{1}{\sqrt{n}} \sum_{u=1}^m x_u.$$

Moreover, by Lemma 5.1.5,

$$\langle \mathbf{x}, L\mathbf{x} \rangle = \sum_{\{u,v\}\in E} (x_u - x_v)^2$$

so the Rayleigh quotient is

$$\mathcal{R}_L(\mathbf{x}) = \frac{\langle \mathbf{x}, L\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \frac{\sum_{\{u,v\}\in E}(x_u - x_v)^2}{\sum_{u=1}^n x_u^2}.$$

That proves the claim. ■

One application of this extremal characterization is a graph drawing heuristic. Consider the entries of the second Laplacian eigenvector $\mathbf{y}_2$ normalized to have unit norm. The entries are centered around 0 by the condition $\sum_{u=1}^n x_u = 0$. Because it minimizes the quantity

$$\frac{\sum_{\{u,v\}\in E}(x_u - x_v)^2}{\sum_{u=1}^n x_u^2},$$

over all centered unit vectors, $\mathbf{y}_2$ tends to assign similar coordinates to adjacent vertices. A similar reasoning applies to the third Laplacian eigenvector, which in addition is orthogonal to the second one. See Figure 5.1 for an illustration.

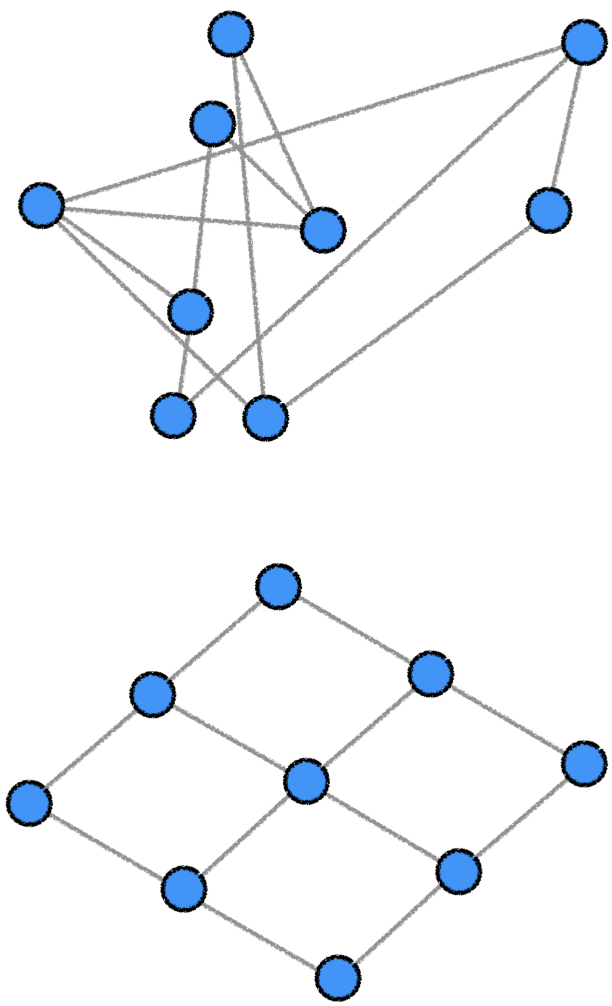

Figure 5.1: Top: A 3-by-3 grid graph with vertices located at independent uniformly random points in a square. Bottom: The same 3-by-3 grid graph with vertices located at the coordinates corresponding to the second and third eigenvectors of the Laplacian matrix. That is, vertex i is located at position $(y_{2,i}, y_{3,i})$.

Example 5.1.9 (Two-component graph)**.** Let $G = (V, E)$ be a graph with two connected components $\emptyset \neq V_1, V_2 \subseteq V$. By the properties of connected components, we have $V_1 \cap V_2 = \emptyset$ and $V_1 \cup V_2 = V$. Assume the Laplacian L of G has spectral decomposition $L = \sum_{i=1}^n \mu_i \mathbf{y}_i \mathbf{y}_i^T$ with $0 = \mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$ and $\mathbf{y}_1 = \frac{1}{\sqrt{n}}(1, \ldots, 1)$. We claimed earlier that for such a graph $\mu_2 = 0$. We prove this here using Corollary 5.1.8:

$$\mu_2 = \min \left\{ \sum_{\{u,v\} \in E} (x_u - x_v)^2 : \right.$$
$$\left. \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n, \ \sum_{u=1}^n x_u = 0, \ \sum_{u=1}^n x_u^2 = 1 \right\}.$$

Based on this characterization, it suffices to find a vector $\mathbf{x}$ satisfying $\sum_{u=1}^n x_u = 0$ and $\sum_{u=1}^n x_u^2 = 1$ such that $\sum_{\{u,v\} \in E}(x_u - x_v)^2 = 0$. Indeed, since $\mu_2 \geq 0$ and any such $\mathbf{x}$ gives an upper bound on μ_2, we then necessarily have that $\mu_2 = 0$.

For $\sum_{\{u,v\} \in E}(x_u - x_v)^2$ to be 0, one might be tempted to take a constant vector $\mathbf{x}$. But then we could not satisfy $\sum_{u=1}^n x_u = 0$ and $\sum_{u=1}^n x_u^2 = 1$. Instead, we modify this guess slightly. Because the graph has two connected components, there is no edge between V_1 and V_2. Hence we can assign a different value to each component and still get $\sum_{\{u,v\} \in E}(x_u - x_v)^2 = 0$. So we look for a vector $\mathbf{x} = (x_1, \ldots, x_n)$ of the form

$$x_u = \begin{cases} \alpha & \text{if } u \in V_1, \\ \beta & \text{if } u \in V_2. \end{cases}$$

To satisfy the constraints on $\mathbf{x}$, we require

$$\sum_{u=1}^n x_u = \sum_{u \in V_1} \alpha + \sum_{u \in V_2} \beta = |V_1|\alpha + |V_2|\beta = 0,$$

and

$$\sum_{u=1}^n x_u^2 = \sum_{u \in V_1} \alpha^2 + \sum_{u \in V_2} \beta^2 = |V_1|\alpha^2 + |V_2|\beta^2 = 1.$$

Replacing the first equation in the second one, we get

$$|V_1| \left(\frac{-|V_2|\beta}{|V_1|} \right)^2 + |V_2|\beta^2 = \frac{|V_2|^2\beta^2}{|V_1|} + |V_2|\beta^2 = 1,$$

or

$$\beta^2 = \frac{|V_1|}{|V_2|(|V_2| + |V_1|)} = \frac{|V_1|}{n|V_2|}.$$

Take

$$\beta = -\sqrt{\frac{|V_1|}{n|V_2|}}, \qquad \alpha = \frac{-|V_2|\beta}{|V_1|} = \sqrt{\frac{|V_2|}{n|V_1|}}.$$

The vector $\mathbf{x}$ we constructed is in fact an eigenvector of L. Indeed, let B be an oriented incidence matrix of G. Then, for $e_k = \{u, v\}$, $(B^T\mathbf{x})_k$ is either $x_u - x_v$ or $x_v - x_u$. In both cases, that is 0. So $L\mathbf{x} = BB^T\mathbf{x} = \mathbf{0}$, that is, $\mathbf{x}$ is an eigenvector of L with eigenvalue 0.

We have shown that $\mu_2 = 0$ when G has two connected components. A slight modification of this argument shows that $\mu_2 = 0$ whenever G is not connected. ◀

Networks In the case of a network (i.e., edge-weighted graph) $G = (V, E, w)$, the Laplacian can be defined as follows. As usual, we assume that $w : E \to \mathbb{R}_+$ is a function that assigns positive real weights to the edges. We write $w_e = w_{ij}$ for the weight of edge $e = \{i, j\}$. Recall that the degree of a vertex i is,

$$\delta(i) = \sum_{j:\{i,j\}\in E} w_{ij},$$

the adjacency matrix A of G is the $n \times n$ symmetric matrix defined as

$$A_{ij} = \begin{cases} w_{ij} & \text{if } \{i,j\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 5.1.10 (Network Laplacian)**.** *Let $G = (V, E, w)$ be a network with $n = |V|$ vertices and adjacency matrix A. Let $D = \mathrm{diag}(\delta(1), \ldots, \delta(n))$ be the degree matrix. The* network Laplacian *(or* Laplacian matrix*, or* Laplacian *for short) associated to G is defined as $L = D - A$.*

It can be shown (see Exercise 5.2) that the Laplacian quadratic form satisfies in the edge-weighted case

$$\langle \mathbf{x}, L\mathbf{x} \rangle = \sum_{\{i,j\}\in E} w_{ij}(x_i - x_j)^2, \tag{5.1.3}$$

for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. (The keen observer will have noticed that we already encountered this quantity as the "Dirichlet energy" in Section 3.3.3; more on this in Section 5.3.) As a positive semidefinite matrix (see again Exercise 5.2), the network Laplacian has an orthonormal basis of eigenvectors with nonnegative eigenvalues that satisfy the variational characterization we derived above. In particular, if we denote the eigenvalues $0 = \mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$, it follows from

Courant-Fischer (Theorem 5.1.3) that

$$\mu_2 = \min\left\{ \sum_{\{u,v\}\in E} w_{uv}(x_u - x_v)^2 \; : \right.$$
$$\left. \mathbf{x} = (x_1,\ldots,x_n)^T \in \mathbb{R}^n, \sum_{u=1}^n x_u = 0, \sum_{u=1}^n x_u^2 = 1 \right\}.$$

Other variants of the Laplacian are useful. We introduce the normalized Laplacian next.

Definition 5.1.11 (Normalized Laplacian)**.** *The* normalized Laplacian *of* $G = (V, E, w)$ *with adjacency matrix* A *and degree matrix* D *is defined as*

normalized Laplacian

$$\mathcal{L} = I - D^{-1/2}AD^{-1/2}.$$

The entries of $\mathcal{L}$ are

$$\mathcal{L}_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ -\frac{w_{ij}}{\sqrt{\delta(i)\delta(j)}} & \text{otherwise.} \end{cases}$$

We also note the following relation to the (unnormalized) Laplacian:

$$\mathcal{L} = D^{-1/2}LD^{-1/2}. \tag{5.1.4}$$

We check that the normalized Laplacian is symmetric:

$$\begin{aligned} \mathcal{L}^T &= I^T - (D^{-1/2}AD^{-1/2})^T \\ &= I - (D^{-1/2})^T A^T (D^{-1/2})^T \\ &= I - D^{-1/2}AD^{-1/2} \\ &= \mathcal{L}. \end{aligned}$$

It is also positive semidefinite. Indeed,

$$\mathbf{x}^T\mathcal{L}\mathbf{x} = \mathbf{x}^T D^{-1/2}LD^{-1/2}\mathbf{x} = (D^{-1/2}\mathbf{x})^T L(D^{-1/2}\mathbf{x}) \geq 0,$$

by the properties of the Laplacian. Hence by the spectral theorem (Theorem 5.1.1), we can write

$$\mathcal{L} = \sum_{i=1}^n \eta_i \mathbf{z}_i \mathbf{z}_i^T,$$

where the $\mathbf{z}_i$s are orthonormal eigenvectors of $\mathcal{L}$ and the eigenvalues satisfy

$$0 \leq \eta_1 \leq \eta_2 \leq \cdots \leq \eta_n.$$

One more observation: because the constant vector is an eigenvector of L with eigenvalue 0, we get from (5.1.4) that $D^{1/2}\mathbf{1}$ is an eigenvector of $\mathcal{L}$ with eigenvalue 0. So $\eta_1 = 0$ and we set

$$(\mathbf{z}_1)_i = \left(\frac{D^{1/2}\mathbf{1}}{\|D^{1/2}\mathbf{1}\|_2}\right)_i = \sqrt{\frac{\delta(i)}{\sum_{i\in V}\delta(i)}}, \quad \forall i \in [n],$$

which makes $\mathbf{z}_1$ into a unit norm vector. The relationship to the Laplacian implies (see Exercise 5.3) that

$$\mathbf{x}^T\mathcal{L}\mathbf{x} = \sum_{\{i,j\}\in E} w_{ij}\left(\frac{x_i}{\sqrt{\delta(i)}} - \frac{x_j}{\sqrt{\delta(j)}}\right)^2,$$

for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Through the change of variables

$$y_i = \frac{x_i}{\sqrt{\delta(i)}},$$

Courant-Fischer (Theorem 5.1.3) gives this time

$$\eta_2 = \min\left\{\sum_{\{u,v\}\in E} w_{uv}(y_u - y_v)^2 \ : \ \mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n, \sum_{u=1}^n \delta(u) y_u = 0, \sum_{u=1}^n \delta(u) y_u^2 = 1\right\}. \tag{5.1.5}$$

5.1.3 Perturbation results

We will need some perturbation results for eigenvalues and eigenvectors. Recall the following definition. Define $\mathbb{S}^{m-1} = \{\mathbf{x} \in \mathbb{R}^m \ : \ \|\mathbf{x}\|_2 = 1\}$. The spectral norm (or induced 2-norm or 2-norm) of a matrix $A \in \mathbb{R}^{n\times m}$ is

$$\|A\|_2 := \max_{\mathbf{0}\neq\mathbf{x}\in\mathbb{R}^m} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\mathbf{x}\in\mathbb{S}^{m-1}} \|A\mathbf{x}\|.$$

The induced 2-norm of a matrix has many other useful properties.

Lemma 5.1.12 (Properties of the induced norm)**.** *Let $A, B \in \mathbb{R}^{n\times m}$ and $\alpha \in \mathbb{R}$. The following hold:*

(i) $\|A\mathbf{x}\|_2 \leq \|A\|_2\|\mathbf{x}\|_2$, $\forall \mathbf{x} \in \mathbb{R}^m$

(ii) $\|A\|_2 \geq 0$

(iii) $\|A\|_2 = 0$ *if and only if* $A = 0$

(iv) $\|\alpha A\|_2 = |\alpha|\|A\|_2$

(v) $\|A + B\|_2 \leq \|A\|_2 + \|B\|_2$

(vi) $\|AB\|_2 \leq \|A\|_2\|B\|_2$.

Proof. These properties all follow from the definition of the induced norm and the corresponding properties for the vector norm:

- Claims (i) and (ii) are immediate from the definition.
- For (ii) note that $\|A\|_2 = 0$ implies $\|A\mathbf{x}\|_2 = 0, \forall \mathbf{x} \in \mathbb{S}^{m-1}$, so that $A\mathbf{x} = \mathbf{0}, \forall \mathbf{x} \in \mathbb{S}^{m-1}$. In particular, $A_{ij} = \mathbf{e}_i^T A\mathbf{e}_j = 0, \forall i, j$.
- For (iv), (v), (vi), observe that for all $\mathbf{x} \in \mathbb{S}^{m-1}$

$$\|\alpha A\mathbf{x}\|_2 = |\alpha|\|A\mathbf{x}\|_2$$

$$\|(A + B)\mathbf{x}\|_2 = \|A\mathbf{x} + B\mathbf{x}\|_2 \leq \|A\mathbf{x}\|_2 + \|B\mathbf{x}\|_2 \leq \|A\|_2 + \|B\|_2$$
$$\|(AB)\mathbf{x}\|_2 = \|A(B\mathbf{x})\|_2 \leq \|A\|_2\|B\mathbf{x}\|_2 \leq \|A\|_2\|B\|_2.$$

∎

Perturbations of eigenvalues For a symmetric matrix $C \in \mathbb{R}^{d\times d}$, we let $\lambda_j(C)$, $j = 1, \ldots, d$, be the eigenvalues of C in nonincreasing order with corresponding orthonormal eigenvectors $\mathbf{v}_j(C)$, $j = 1, \ldots, d$. As in the Courant-Fischer theorem (Theorem 5.1.3), define the subspaces

$$\mathcal{V}_k(C) = \mathrm{span}(\mathbf{v}_1(C), \ldots, \mathbf{v}_k(C))$$

and

$$\mathcal{W}_{d-k+1}(C) = \mathrm{span}(\mathbf{v}_k(C), \ldots, \mathbf{v}_d(C)).$$

The following lemma is one version of what is known as *Weyl's inequality*. *Weyl's inequality*

Lemma 5.1.13 (Weyl's inequality)**.** *Let $A \in \mathbb{R}^{d\times d}$ and $B \in \mathbb{R}^{d\times d}$ be symmetric matrices. Then, for all $j = 1, \ldots, d$,*

$$\max_{j\in[d]} |\lambda_j(B) - \lambda_j(A)| \leq \|B - A\|_2.$$

Proof. Let $H = B - A$. We prove only one upper bound. The other one follows from interchanging the roles of A and B. Because

$$\dim(\mathcal{V}_j(B)) + \dim(\mathcal{W}_{d-j+1}(A)) = j + (d - j + 1) = d + 1 > d,$$

it follows from (5.1.2) that $\mathcal{V}_j(B) \cap \mathcal{W}_{d-j+1}(A)$ contains a nonzero vector. Let $\mathbf{v}$ be a unit vector in that intersection.

By Theorem 5.1.3,

$$\lambda_j(B) \leq \langle \mathbf{v}, (A + H)\mathbf{v}\rangle = \langle \mathbf{v}, A\mathbf{v}\rangle + \langle \mathbf{v}, H\mathbf{v}\rangle \leq \lambda_j(A) + \langle \mathbf{v}, H\mathbf{v}\rangle.$$

Moreover, by Cauchy-Schwarz (Theorem B.4.8), since $\|\mathbf{v}\|_2 = 1$

$$\langle \mathbf{v}, H\mathbf{v}\rangle \leq \|\mathbf{v}\|_2 \|H\mathbf{v}\|_2 \leq \|H\|_2,$$

which proves the claim after rearranging. ∎

Perturbations of eigenvectors While Weyl's inequality (Lemma 5.1.13) indicates that the eigenvalues of A and B are close when $\|A - B\|_2$ is small, it says nothing about the eigenvectors. The following theorem remediates that. It is traditionally stated in terms of the angle between the eigenvectors (whereby the name). Here we give a version that is more suited to the applications we will encounter. We do not optimize the constants. We use the same notation as in the previous paragraph. Recall Parseval's identity: if $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d$ is an orthonormal basis of $\mathbb{R}^d$, then $\|\boldsymbol{x}\|^2 = \sum_{i=1}^d \langle \boldsymbol{x}, \boldsymbol{u}_i\rangle^2$.

Theorem 5.1.14 (Davis-Kahan $\sin\theta$ theorem)**.** *Let $A \in \mathbb{R}^{d\times d}$ and $B \in \mathbb{R}^{d\times d}$ be symmetric matrices. For an $i \in \{1, \ldots, d\}$, assume that*

$$\delta := \min_{j\neq i} |\lambda_i(A) - \lambda_j(A)| > 0.$$

Then

$$\min_{s\in\{+1,-1\}} \|\mathbf{v}_i(A) - s\mathbf{v}_i(B)\|_2^2 \leq \frac{8\|A - B\|_2^2}{\delta^2}.$$

Proof. Expand $\mathbf{v}_i(B)$ in the basis formed by the eigenvectors of A, that is,

$$\mathbf{v}_i(B) = \sum_{j=1}^{d} \langle \mathbf{v}_i(B), \mathbf{v}_j(A) \rangle \, \mathbf{v}_j(A),$$

where we used the orthonormality of the $\mathbf{v}_j(A)$s. On the one hand,

$$\begin{aligned}
&\|(A - \lambda_i(A) I_{d\times d}) \, \mathbf{v}_i(B)\|_2^2 \\
&= \left\| \sum_{j=1}^{d} \langle \mathbf{v}_i(B), \mathbf{v}_j(A) \rangle (A - \lambda_i(A) I_{d\times d}) \, \mathbf{v}_j(A) \right\|_2^2 \\
&= \left\| \sum_{j=1, j\neq i}^{d} \langle \mathbf{v}_i(B), \mathbf{v}_j(A) \rangle (\lambda_j(A) - \lambda_i(A)) \, \mathbf{v}_j(A) \right\|_2^2 \\
&= \sum_{j=1, j\neq i}^{d} \langle \mathbf{v}_i(B), \mathbf{v}_j(A) \rangle^2 (\lambda_j(A) - \lambda_i(A))^2 \\
&\geq \delta^2 (1 - \langle \mathbf{v}_i(B), \mathbf{v}_i(A) \rangle^2),
\end{aligned}$$

where, on the last two lines, we used the orthonormality of the $\mathbf{v}_j(A)$s and $\mathbf{v}_j(B)$s through Parseval's identity, as well as the definition of δ.

On the other hand, letting $E = A - B$, by the triangle inequality

$$\begin{aligned}
\|(A - \lambda_i(A) I) \, \mathbf{v}_i(B)\|_2 &= \|(B + E - \lambda_i(A) I) \, \mathbf{v}_i(B)\|_2 \\
&\leq \|(B - \lambda_i(A) I) \, \mathbf{v}_i(B)\|_2 + \|E \, \mathbf{v}_i(B)\|_2 \\
&\leq |\lambda_i(B) - \lambda_i(A)| \|\mathbf{v}_i(B)\|_2 + \|E\|_2 \|\mathbf{v}_i(B)\|_2 \\
&= 2\|E\|_2,
\end{aligned}$$

where we used Lemma 5.1.12 and Weyl's inequality.

Combining the last two inequalities gives

$$(1 - \langle \mathbf{v}_i(B), \mathbf{v}_i(A) \rangle^2) \leq \frac{4\|E\|_2^2}{\delta^2}.$$

The result follows by noting that, since $|\langle \mathbf{v}_i(B), \mathbf{v}_i(A) \rangle| \leq 1$ by Cauchy-Schwarz (Theorem B.4.8), we have

$$\begin{aligned}
\min_{s\in\{+1,-1\}} \|\mathbf{v}_i(A) - s\mathbf{v}_i(B)\|^2 &= 2 - 2|\langle \mathbf{v}_i(B), \mathbf{v}_i(A) \rangle| \\
&\leq 2(1 - \langle \mathbf{v}_i(B), \mathbf{v}_i(A) \rangle^2) \\
&\leq \frac{8\|E\|_2^2}{\delta^2}.
\end{aligned}$$

■

5.1.4 ▷ *Data science: community recovery*

A common task in network analysis is to recover hidden community structure. Informally, we seek groups of vertices with more edges within the groups then to the rest of the graph. More rigorously, providing statistical guarantees on the output of a community recovery algorithm requires some underlying random graph model. The standard model for this purpose is the *stochastic blockmodel*, a generalization of the Erdös-Rényi graph model with a "planted partition."

Stochastic blockmodel and recovery requirement We restrict ourselves to the simple case of two strictly balanced communities. Consider a random graph on n (even) nodes where there are two communities, labeled $+1$ and -1, consisting of $n/2$ nodes. Each vertex $i \in V$ is assigned a community label $X_i \in \{1, -1\}$ as follows: a subset of $n/2$ vertices is chosen uniformly at random among all such subsets to form community $+1$, and the rest of the vertices form community -1. For two nodes i, j, the edge $\{i, j\}$ is present with probability p if they belong to the same community, and with probability q otherwise. All edges are independent. The following 2×2 matrix describes the edge density within and across the two communities:

$$W = \begin{array}{c c} & \begin{array}{cc} +1 & -1 \end{array} \\ \begin{array}{c} +1 \\ -1 \end{array} & \begin{bmatrix} p & q \\ q & p \end{bmatrix} \end{array}.$$

We assume that $p \geq q$, encoding the fact that vertices belonging to the same community are more likely to share an edge. To summarize, we say that $(X, G) \sim \mathrm{SBM}_{n,p,q}$ if:

stochastic blockmodel

1. *(Communities)* The assignment $X = (X_1, \ldots, X_n)$ is uniformly random over

$$\Pi_2^n := \{\mathbf{x} \in \{+1, -1\}^n : \mathbf{x}^T \mathbf{1} = 0\},$$

where $\mathbf{1} = (1, \cdots, 1)$ is the all-one vector.

2. *(Graph)* Conditioned on X, the graph $G = ([n], E)$ has independent edges where $\{i, j\}$ is present with probability W_{X_i, X_j} for $\forall i < j$.

We denote the corresponding measure by $\mathbb{P}_{n,p,q}$. We allow p and q to depend on n (although we do not make that dependence explicit).

Roughly speaking, the *community recovery problem* is the following: given G, output X. There are different notions of recovery.

Definition 5.1.15 (Agreement)**.** *The agreement between two community assignment vectors $\mathbf{x}, \mathbf{y} \in \{+1, -1\}^n$ is the largest fraction of common assignments between $\mathbf{x}$ and $\pm\mathbf{y}$, that is,*

$$\alpha(\mathbf{x}, \mathbf{y}) = \max_{s \in \{+1,-1\}} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{x_i = s y_i\}.$$

The role of s in this formula is to account for the fact that the community names are not meaningful.

Now consider the following recovery requirements. These are asymptotic notions, as $n \to +\infty$.

recovery

Definition 5.1.16 (Recovery requirement)**.** *Let $(X, G) \sim \mathrm{SBM}_{n,p,q}$. For any estimator $\hat{X} := \hat{X}(G) \in \Pi_2^n$, we say that it achieves:*

- *exact recovery* if $\mathbb{P}_{n,p,q}[\alpha(X, \hat{X}) = 1] = 1 - o(1)$; or
- *almost exact recovery* if $\mathbb{P}_{n,p,q}[\alpha(X, \hat{X}) = 1 - o(1)] = 1 - o(1)$.

Next we establish sufficient conditions for almost exact recovery. First we describe a natural estimator $\hat{X}$.

MAP estimator and spectral clustering A natural starting point is the maximum a posteriori (MAP) estimator. Let $\Omega(X)$ be the balanced partition of $[n]$ corresponding to X and $\hat{\Omega}(G)$ be the one corresponding to $\hat{X}(G)$. The probability of error, that is, the probability of not recovering the true partition, is given by

$$\mathbb{P}[\Omega(X) \neq \hat{\Omega}(G)] = \sum_{g} \mathbb{P}[\hat{\Omega}(g) \neq \Omega(X) \,|\, G = g]\, \mathbb{P}[G = g], \tag{5.1.6}$$

where the sum is over all graphs on n vertices (i.e., all possible subsets of edges present) and we dropped the subscript n, p, q to simplify the notation. The MAP estimator $\hat{\Omega}^{\mathrm{MAP}}(G)$ is obtained by minimizing each term $\mathbb{P}[\hat{\Omega}(g) \neq \Omega(X) \,|\, G = g]$ individually (note that $\mathbb{P}[G = g] > 0$ for all g by definition of the $\mathrm{SBM}_{n,p,q}$, a probability which does not depend on the estimator). Equivalently we choose for each g a partition γ that maximizes the posterior probability

$$\begin{aligned}\mathbb{P}[\Omega(X) = \gamma \,|\, G = g] &= \frac{\mathbb{P}[G = g \,|\, \Omega(X) = \gamma]\, \mathbb{P}[\Omega(X) = \gamma]}{\mathbb{P}[G = g]} \\ &= \mathbb{P}[G = g \,|\, \Omega(X) = \gamma] \cdot \frac{1}{|\Pi_2^n|\, \mathbb{P}[G = g]},\end{aligned} \tag{5.1.7}$$

where we applied Bayes' rule on the first line and the uniformity of the partition X on the second line.

Based on (5.1.7), we seek a partition that maximizes $\mathbb{P}[G = g \,|\, \Omega(X) = \gamma]$. We compute this last probability explicitly. For fixed g, let $M := M(g)$ be the number of edges in g. For any γ, denote by $M_{\text{in}} := M_{\text{in}}(g, \gamma)$ and $M_{\text{out}} := M_{\text{out}}(g, \gamma)$ the number of edges within and across communities respectively, and note that $M_{\text{in}} = M - M_{\text{out}}$. By definition of the $\mathrm{SBM}_{n,p,q}$ model, the probability of a graph g given a partition γ is expressed simply as

$$\begin{aligned}
&\mathbb{P}[G = g \,|\, \Omega(X) = \gamma] \\
&= q^{M_{\text{out}}}(1-q)^{(\frac{n}{2})^2 - M_{\text{out}}} p^{M_{\text{in}}}(1-p)^{\{\binom{n}{2} - (\frac{n}{2})^2\} - M_{\text{in}}} \\
&= q^{M_{\text{out}}}(1-q)^{(\frac{n}{2})^2 - M_{\text{out}}} p^{M - M_{\text{out}}}(1-p)^{\{\binom{n}{2} - (\frac{n}{2})^2\} - \{M - M_{\text{out}}\}} \\
&= \left[\frac{q}{1-q} \cdot \frac{1-p}{p}\right]^{M_{\text{out}}} \left\{(1-q)^{(\frac{n}{2})^2} p^M (1-p)^{\{\binom{n}{2} - (\frac{n}{2})^2\} - M}\right\}.
\end{aligned}$$

The expression in curly brackets does not depend on the partition γ. Moreover, since we assume that $p \geq q$, we have that $\left[\frac{q}{1-q} \cdot \frac{1-p}{p}\right] \leq 1$ (which can be checked directly by rearranging and cancelling). Therefore, to maximize $\mathbb{P}[G = g \,|\, \Omega(X) = \gamma]$ over γ for a fixed g, we need to choose a partition that results in the smallest possible value of M_{out}, the number of edges across the two communities. This problem is well-known in combinatorial optimization, where it is referred to as the *minimum bisection problem*. It is unfortunately NP-hard and we consider a relaxation that admits a polynomial-time algorithmic solution. *minimum bisection problem*

To see how this comes about, observe that the minimum bisection problem can be reformulated as

$$\max_{\mathbf{x} \in \{+1,-1\}^n,\ \mathbf{x}^T \mathbf{1} = 0} \mathbf{x}^T A \mathbf{x}$$

where A is the $n \times n$ adjacency matrix. Replacing the combinatorial constraint $\mathbf{x} \in \{+1, -1\}^n$ by $\mathbf{x} \in \mathbb{R}^n$ with $\|\mathbf{x}\|_2 = n$ leads to the relaxation

$$\begin{aligned}
&\max_{\mathbf{z} \in \mathbb{R}^n,\ \mathbf{z}^T \mathbf{1} = 0,\ \|\mathbf{z}\|_2 = n} \mathbf{z}^T A \mathbf{z} \\
&= \max_{\mathbf{0} \neq \mathbf{z} \in \mathbb{R}^n,\ \mathbf{z}^T \mathbf{1} = 0} \left(n \frac{\mathbf{z}}{\|\mathbf{z}\|_2}\right)^T A \left(n \frac{\mathbf{z}}{\|\mathbf{z}\|_2}\right) \\
&= n^2 \max_{\mathbf{0} \neq \mathbf{z} \in \mathbb{R}^n,\ \mathbf{z}^T \mathbf{1} = 0} \frac{\mathbf{z}^T A \mathbf{z}}{\mathbf{z}^T \mathbf{z}},
\end{aligned}$$

where we changed the notation from $\mathbf{x}$ to $\mathbf{z}$ to emphasize that the solution no longer encodes a partition. We recognize the Rayleigh quotient of A as the objective function in the final formulation. At this point, it is tempting to use Courant-Fischer

(Theorem 5.1.3) and conclude that the maximum above is achieved at the second eigenvalue of A. Note however that the vector $\mathbf{1}$ (appearing in the orthogonality constraint $\mathbf{z}^T\mathbf{1} = 0$) is not in general an eigenvector of A (unless the graph happens to be regular). To leverage the variational characterization of eigenvalues in a statistically justified way, we instead turn to the expected adjacency matrix and then establish concentration.

Lemma 5.1.17 (Expected adjacency). *Let $(X, G) \sim \mathrm{SBM}_{n,p,q}$, let A be the adjacency matrix of G and let $\mathcal{A}_X = \mathbb{E}_{n,p,q}[A \,|\, X]$. Then*

$$\mathcal{A}_X = n\,\frac{p+q}{2}\,\mathbf{u}_1\mathbf{u}_1^T + n\,\frac{p-q}{2}\,\mathbf{u}_2\mathbf{u}_2^T - p\,I,$$

where

$$\mathbf{u}_1 = \frac{1}{\sqrt{n}}\mathbf{1}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{n}}X.$$

Proof. For any distinct pair i, j, the term

$$\left(n\,\frac{p+q}{2}\,\mathbf{u}_1\mathbf{u}_1^T\right)_{i,j} = n\frac{p+q}{2}\left(\frac{1}{\sqrt{n}}\right)^2 = \frac{p+q}{2}$$

while the term

$$\left(n\,\frac{p-q}{2}\,\mathbf{u}_2\mathbf{u}_2^T\right)_{i,j} = n\frac{p-q}{2}\left(\frac{1}{\sqrt{n}}\right)^2 X_iX_j = \frac{p-q}{2}X_iX_j.$$

The product X_iX_j is 1 when i and j belong to the same community and is -1 otherwise. In the former case, summing the two terms indeed gives p, while in the latter case it gives q. Finally, the term $-pI$ accounts for the fact that A has zeros on the diagonal. ■

Now condition on X and observe that $\mathbf{u}_1$ and $\mathbf{u}_2$ in Lemma 5.1.17 are orthogonal by our assumption that X corresponds to a balanced partition (i.e., with two communities of equal size). Hence we deduce that an eigenvector decomposition of $\mathcal{A}_X$ is formed of $\mathbf{u}_1$, $\mathbf{u}_2$ and any orthonormal basis of the orthogonal complement of the span of $\mathbf{u}_1$ and $\mathbf{u}_2$, with respective eigenvalues

$$n\,\frac{p+q}{2} - p, \qquad n\,\frac{p-q}{2} - p, \qquad -p.$$

So the second largest eigenvalue of $\mathcal{A}_X$ is $\lambda_2(\mathcal{A}_X) = n\,\frac{p-q}{2} - p$ (independently of X), and Courant-Fischer implies

$$\max_{\mathbf{0} \neq \mathbf{z} \in \mathbb{R}^n, \mathbf{z}^T\mathbf{1} = 0} \frac{\mathbf{z}^T\mathcal{A}_X\mathbf{z}}{\mathbf{z}^T\mathbf{z}} = \lambda_2(\mathcal{A}_X).$$

The corresponding eigenvector, up to scaling and sign, is precisely what we are trying to recover, namely, the community assignment X.

These observations motivate the following *spectral clustering* approach. *spectral clustering*

1. *Input:* graph G with adjacency matrix A.
2. Compute an eigenvector decomposition of A.
3. Let $\hat{\mathbf{u}}_2$ be the eigenvector corresponding to the second largest eigenvalue.
4. *Output:* $\hat{X}(G) = \mathrm{sgn}\left(\hat{\mathbf{u}}_2\right)$.

Here we used the notation

$$(\mathrm{sgn}(\mathbf{z}))_i = \begin{cases} +1 & \text{if } z_i \geq 0, \\ -1 & \text{otherwise.} \end{cases}$$

Because we used A rather than $\mathcal{A}_X$ (which we do not know), it is not immediate that this approach will work. Below, we use Davis-Kahan (Theorem 5.1.14) to show that, under some conditions, the second eigenvector of A is concentrated around that of $\mathcal{A}_X$—and therefore almost exact recovery holds.

Before getting to the analysis, we make a final algorithmic remark. The "clustering" above, specifically taking the sign of the second eigenvector, works in this toy model but is perhaps somewhat naive. More generally, in a spectral clustering method, one uses the top eigenvectors (deciding how many is a bit of an art) of the adjacency matrix (or of another matrix associated to the graph such as the Laplacian or normalized Laplacian) to obtain a low-dimensional representation of the input. Then in a second step, one uses a clustering algorithm, for example, k-means clustering, to extract communities in the low-dimensional space.

Almost exact recovery We prove the following. We restrict ourselves to the case where p and q are constants not depending on n.

Theorem 5.1.18. *Let $(X, G) \sim \mathrm{SBM}_{n,p,q}$ and let A be the adjacency matrix of G. Let $\mu := \min\left\{q, \frac{p-q}{2}\right\} > 0$. Clustering according to the sign of the second eigenvector of A identifies the two communities of G with probability at least $1 - e^{-n}$, except for C/μ^2 misclassified nodes for some constant $C > 0$.*

There are two key ingredients to the proof: concentration of the adjacency matrix and perturbation arguments.

We start with the former.

Lemma 5.1.19 (Norm of the centered adjacency matrix)**.** *Let* $(X, G) \sim \mathrm{SBM}_{n,p,q}$*, let* A *be the adjacency matrix of* G *and let* $\mathcal{A}_X = \mathbb{E}_{n,p,q}[A \,|\, X]$*. There is a constant* $C' > 0$ *such that, conditioned on* X*,*

$$\|A - \mathcal{A}_X\|_2 \leq C'\sqrt{n},$$

with probability at least $1 - e^{-n}$*.*

Proof. Condition on X. We use Theorem 2.4.28 on the random matrix $R := A - \mathcal{A}_X$. The entries of R are centered and independent (conditionally on X). Moreover they are bounded. Indeed, for $i \neq j$, $A_{ij} \in \{0, 1\}$ while $(\mathcal{A}_X)_{ij} \in \{q, p\}$. So $R_{ij} \leq [-p, 1-q]$. On the diagonal, $R_{ii} = 0$. Hence, by Hoeffding's lemma (Lemma 2.4.12), the entries are sub-Gaussian with variance factor

$$\frac{1}{4}(1 - q - (-p))^2 \leq 1.$$

Taking $t = \sqrt{n}$ in Theorem 2.4.28, there is a constant $C > 0$ such that with probability $1 - e^{-n}$

$$\|A - \mathcal{A}_X\|_2 \leq C\sqrt{1}(\sqrt{n} + \sqrt{n} + \sqrt{n}).$$

Adjusting the constant gives the claim. ■

We are ready to prove the theorem.

Proof of Theorem 5.1.18. Condition on X. To apply the Davis-Kahan theorem (Theorem 5.1.14), we need to bound the smallest gap δ between the second largest eigenvalue of $\mathcal{A}_X$ and its other eigenvalues. Recall that the eigenvalues are

$$n\,\frac{p+q}{2} - p, \qquad n\,\frac{p-q}{2} - p, \qquad -p,$$

so

$$\delta = \min\left\{n\,\frac{p-q}{2}, n\,q\right\} = n\,\mu > 0.$$

By Davis-Kahan and Lemma 5.1.19, with probability at least $1 - e^{-n}$, there is $\theta \in \{+1, -1\}$ such that

$$\|\mathbf{u}_2 - \theta\,\hat{\mathbf{u}}_2\|_2^2 \leq \frac{8\|A - \mathcal{A}_X\|_2^2}{\delta^2} \leq \frac{8(C'\sqrt{n})^2}{(n\mu)^2} = \frac{C}{n\,\mu^2},$$

by adjusting the constant. Note that this bound holds for *any* X.

Rearranging and expanding the norm, we get

$$\sum_i \left|\sqrt{n}\,(\mathbf{u}_2)_i - \sqrt{n}\,\theta\,(\hat{\mathbf{u}}_2)_i\right|^2 \leq \frac{C}{\mu^2}.$$

If the signs of $(\mathbf{u}_2)_i$ and $\theta\,(\hat{\mathbf{u}}_2)_i$ disagree, then the i-th term in the sum above is ≥ 1. So there can be at most C/μ^2 such disagreements. That establishes the desired bound on the number of misclassified nodes. ■

Remark 5.1.20. *It was shown in [YP14, MNS15a, AS15] that almost exact recovery in the balanced two-community model* SBM_{n,p_n,q_n} *with* $p_n = a_n/n$ *and* $q_n = b_n/n$ *is achievable (and computationally efficiently so) if and only if*

$$\frac{(a_n - b_n)^2}{(a_n + b_n)} = \omega(1).$$

On the other hand, it was shown in [ABH16, MNS15a] that exact recovery in the SBM_{n,p_n,q_n} *with* $p_n = \alpha \log n/n$ *and* $q_n = \beta \log n/n$ *is achievable and computationally efficiently so if* $\sqrt{\alpha} - \sqrt{\beta} > 2$ *and not achievable if* $\sqrt{\alpha} - \sqrt{\beta} < 2$.

5.2 Spectral techniques for reversible Markov chains

In this section, we apply the spectral theorem to reversible Markov chains. Throughout (X_t) is an irreducible Markov chain on a state space V with transition matrix P reversible with respect to a positive stationary measure $\pi > 0$. Recall that this means that $\pi(x)P(x,y) = \pi(y)P(y,x)$ for all $x, y \in V$. We also assume that P is irreducible.

A Hilbert space It will be convenient to introduce a Hilbert space of functions over V. Let $\ell^2(V,\pi)$ be the space of functions $f : V \to \mathbb{R}$ such that $\sum_{x\in V} \pi(x) f(x)^2 < +\infty$. Equipped with the following inner product, it forms a Hilbert space (i.e., a real inner product space that is also a complete metric space (see Theorem B.4.10) with respect to the induced metric; we will work mostly in finite dimension where it is merely a slight generalization of Euclidean space). For $f, g \in \ell^2(V,\pi)$, define

$$\langle f, g\rangle_\pi := \sum_{x\in V} \pi(x) f(x) g(x),$$

and

$$\|f\|_\pi^2 := \langle f, f\rangle_\pi.$$

The inner product is well-defined since the series is summable by Hölder's inequality (Theorem B.4.8), which implies the Cauchy-Schwarz inequality

$$\langle f, g\rangle_\pi \leq \|f\|_\pi \|g\|_\pi.$$

Minkowski's inequality (Theorem B.4.9) implies the triangle inequality

$$\|f + g\|_\pi \leq \|f\|_\pi + \|g\|_\pi.$$

The integral with respect to π (see Appendix B) reduces in this case to a sum

$$\pi(f) := \sum_{x \in V} \pi(x) f(x),$$

provided $\pi(|f|) < +\infty$ or $f \geq 0$. Here $|f|$ is defined as $|f|(x) := |f(x)|$ for all $x \in V$. We also write $\pi f = \pi(f)$ to simplify the notation.

We recall some standard Hilbert space facts. The countable collection of functions $\{f_i\}_{i=1}^{\infty}$ in $\ell^2(V, \pi)$ is an orthonormal basis if: (i) $\langle f_i, f_j\rangle_\pi = 0$ if $i \neq j$ and $=1$ if $i = j$; and (ii) any $f \in \ell^2(V, \pi)$ can be written as $\lim_{n\to+\infty} \sum_{i=1}^{n} \langle f_i, f\rangle_\pi f_i = f$ where the limit is in the norm. We then have ***Parseval's identity***: for any $g \in \ell^2(V, \pi)$ *Parseval's identity*

$$\|g\|_\pi^2 = \sum_{j=1}^{\infty} \langle g, f_j\rangle_\pi^2. \tag{5.2.1}$$

Think of P as an operator on $\ell^2(V, \pi)$. That is, let $Pf : V \to \mathbb{R}$ be defined as

$$(Pf)(x) := \sum_{y \in V} P(x, y) f(y),$$

for $x \in V$. For any $f \in \ell^2(V, \pi)$, Pf is well-defined and further we have

$$\|Pf\|_\pi \leq \|f\|_\pi. \tag{5.2.2}$$

Indeed by Cauchy-Schwarz, stochasticity, Fubini and stationarity

$$\begin{aligned}
\|P|f|\|_\pi^2 &= \sum_x \pi(x) \left[\sum_y P(x, y)|f(y)|\right]^2 \\
&\leq \sum_x \pi(x) \left[\sum_y P(x, y)|f(y)|^2 \sum_z P(x, z)\right] \\
&= \sum_y \sum_x \pi(x) P(x, y) f(y)^2 \\
&= \sum_y \pi(y) f(y)^2 \\
&= \|f\|_\pi^2 < +\infty.
\end{aligned} \tag{5.2.3}$$

This shows that Pf is well-defined since $\pi > 0$ and hence the series in square brackets on the first line is finite for all x. Applying the same argument to $\|Pf\|_\pi^2$ gives the inequality above.

Everything above holds whether or not P is reversible, so long as π is a stationary measure. Now we use reversibility. We claim that, when P reversible, then it is *self-adjoint*, that is,

$$\langle f, Pg\rangle_\pi = \langle Pf, g\rangle_\pi \qquad \forall f, g \in \ell^2(V, \pi). \tag{5.2.4}$$

This follows immediately by reversibility

$$\begin{aligned}
\langle f, Pg\rangle_\pi &= \sum_{x\in V} \pi(x) f(x) \sum_{y \in V} P(x,y) g(y)\\
&= \sum_{x\in V}\sum_{y\in V} \pi(y) P(y,x) f(x) g(y)\\
&= \sum_{y\in V} \pi(y) g(y) \sum_{x\in V} P(y,x) f(x)\\
&= \langle Pf, g\rangle_\pi,
\end{aligned}$$

where we argue as in (5.2.3) to justify using Fubini.

Throughout this section, we denote by 0 and 1 the all-zero and all-one functions respectively.

5.2.1 Spectral gap

In this subsection, we restrict ourselves to a finite state space V. Our goal is to bound the mixing time of (X_t) in terms of the eigenvalues of the transition matrix P. We assume that π is now the stationary distribution, that is, $\sum_{x\in V} \pi(x) = 1$ (which is unique by Theorem 1.1.24 and irreducibility). We also let $n := |V| < +\infty$.

Spectral decomposition Self-adjointness generalizes the notion of a symmetric matrix, with one consequence being that a version of the spectral theorem applies to P (at least in this finite-dimensional case; see Section 5.2.5 for more discussion on this). For completeness, we derive it from Theorem 5.1.1. It will be convenient to assume without loss of generality that $V = [n]$ and identify functions in $\ell^2(V, \pi)$ with vectors in $\mathbb{R}^n$.

Theorem 5.2.1 (Reversibility: spectral theorem)**.** *There is an orthonormal basis of $\ell^2(V, \pi)$ formed of real eigenfunctions $\{f_j\}_{j=1}^n$ of P with real eigenvalues $\{\lambda_j\}_{j=1}^n$.*

Proof. Let D_π be the diagonal matrix with π on the diagonal. By reversibility,

$$\begin{aligned}
M(x,y) &:= (D_\pi^{1/2} P D_\pi^{-1/2})_{x,y} \\
&= \sqrt{\frac{\pi(x)}{\pi(y)}} P(x,y) \\
&= \sqrt{\frac{\pi(y)}{\pi(x)}} P(y,x) \\
&= (D_\pi^{1/2} P D_\pi^{-1/2})_{y,x} \\
&= M(y,x).
\end{aligned}$$

So $M = (M(x,y))_{x,y} = D_\pi^{1/2} P D_\pi^{-1/2}$ is a symmetric matrix. By the spectral theorem (Theorem 5.1.1), it has real eigenvectors $\{\phi_j\}_{j=1}^n$ forming an orthonormal basis of $\mathbb{R}^n$ with corresponding real eigenvalues $\{\lambda_j\}_{j=1}^n$. Define $f_j := D_\pi^{-1/2}\phi_j$. Then

$$\begin{aligned}
P f_j &= P D_\pi^{-1/2} \phi_j \\
&= D_\pi^{-1/2} D_\pi^{1/2} P D_\pi^{-1/2} \phi_j \\
&= D_\pi^{-1/2} M \phi_j \\
&= \lambda_j D_\pi^{-1/2} \phi_j \\
&= \lambda_j f_j,
\end{aligned}$$

and

$$\begin{aligned}
\langle f_i, f_j \rangle_\pi &= \langle D_\pi^{-1/2}\phi_i, D_\pi^{-1/2}\phi_j \rangle_\pi \\
&= \sum_{x \in V} \pi(x)[\pi(x)^{-1/2}\phi_i(x)][\pi(x)^{-1/2}\phi_j(x)] \\
&= \sum_{x \in V} \phi_i(x)\phi_j(x).
\end{aligned}$$

Because $\{\phi_j\}_{j=1}^n$ is an orthonormal basis of $\mathbb{R}^n$, we have that $\{f_j\}_{j=1}^n$ is an orthonormal basis of $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_\pi)$. ■

We collect a few more facts about the eigenbasis. Recall that

$$\|f\|_\infty = \max_{x \in V} |f(x)|.$$

Lemma 5.2.2. *Any eigenvalue λ of P satisfies $|\lambda| \le 1$.*

Proof. It holds that

$$Pf = \lambda f \implies |\lambda|\|f\|_\infty = \|Pf\|_\infty = \max_x \left|\sum_y P(x,y)f(y)\right| \leq \|f\|_\infty.$$

Rearranging gives the claim. ■

We order the eigenvalues $1 \geq \lambda_1 \geq \cdots \geq \lambda_n \geq -1$. The second eigenvalue will play an important role below.

Lemma 5.2.3. *We have $\lambda_1 = 1$ and $\lambda_2 < 1$. Also we can take $f_1 = 1$.*

Proof. Because P is stochastic, the all-one vector is a right eigenvector with eigenvalue 1. Any eigenfunction with eigenvalue 1 is harmonic with respect to P on V (see (3.3.2)). By Corollary 3.3.3, for a finite, irreducible chain the only harmonic functions are the constant functions. So the eigenspace corresponding to 1 is one-dimensional. We must have $\lambda_2 < 1$ by Lemma 5.2.2. ■

When the chain is aperiodic, it cannot have an eigenvalue -1. Exercise 5.9 asks for a proof.

Lemma 5.2.4. *If P has an eigenvalue equal to -1, then P is not aperiodic.*

Lemma 5.2.5. *For all $j \neq 1$, $\pi f_j = 0$.*

Proof. By orthonormality, $\langle f_1, f_j \rangle_\pi = 0$. Now use the fact that $f_1 = 1$. ■

Let $\delta_x(y) := \mathbf{1}_{\{x=y\}}$.

Lemma 5.2.6. *For all x, y,*

$$\sum_{j=1}^n f_j(x) f_j(y) = \pi(x)^{-1} \delta_x(y).$$

Proof. Using the notation of Theorem 5.2.1, the matrix Φ whose columns are the ϕ_js is orthogonal so $\Phi\Phi^T = I$. That is,

$$\sum_{j=1}^n \phi_j(x)\phi_j(y) = \delta_x(y),$$

or

$$\sum_{j=1}^n \sqrt{\pi(x)\pi(y)} f_j(x) f_j(y) = \delta_x(y).$$

Rearranging gives the result. ■

Using the eigendecomposition of P, we get the following expression for its t-th power P^t.

Theorem 5.2.7 (Spectral decomposition of P^t)**.** *Let $\{f_j\}_{j=1}^n$ be the eigenfunctions of a reversible and irreducible transition matrix P with corresponding eigenvalues $\{\lambda_j\}_{j=1}^n$, as defined previously. Assume $\lambda_1 \geq \cdots \geq \lambda_n$. We have the decomposition*

$$\frac{P^t(x,y)}{\pi(y)} = 1 + \sum_{j=2}^n f_j(x) f_j(y) \lambda_j^t.$$

Proof. Let F be the matrix whose columns are the eigenvectors $\{f_j\}_{j=1}^n$ and let D_λ be the diagonal matrix with $\{\lambda_j\}_{j=1}^n$ on the diagonal. Using the notation in the proof of Theorem 5.2.1,

$$D_\pi^{1/2} P^t D_\pi^{-1/2} = M^t = (D_\pi^{1/2} F) D_\lambda^t (D_\pi^{1/2} F)^T,$$

which after rearranging becomes

$$P^t D_\pi^{-1} = F D_\lambda^t F^T.$$

Expanding and using Lemma 5.2.3 gives the result. ∎

Example 5.2.8 (Two-state chain)**.** Let $V := \{0,1\}$ and

$$P := \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

for $\alpha, \beta \in (0,1)$. Observe that P is reversible with respect to the stationary distribution

$$\pi := \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta} \right).$$

We know that $f_1 = 1$ is an eigenfunction with eigenvalue 1. As can be checked by direct computation, the other eigenfunction (in vector form) is

$$f_2 := \left(\sqrt{\frac{\alpha}{\beta}}, -\sqrt{\frac{\beta}{\alpha}} \right),$$

with eigenvalue $\lambda_2 := 1 - \alpha - \beta$. We normalized f_2 so that $\|f_2\|_\pi^2 = 1$.

By Theorem 5.2.7, the spectral decomposition at time t is therefore

$$P^t D_\pi^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + (1-\alpha-\beta)^t \begin{pmatrix} \frac{\alpha}{\beta} & -1 \\ -1 & \frac{\beta}{\alpha} \end{pmatrix}.$$

Or, rearranging,

$$P^t = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix} + (1-\alpha-\beta)^t \begin{pmatrix} \frac{\alpha}{\alpha+\beta} & -\frac{\alpha}{\alpha+\beta} \\ -\frac{\beta}{\alpha+\beta} & \frac{\beta}{\alpha+\beta} \end{pmatrix}.$$

Note for instance that the case $\alpha + \beta = 1$ corresponds to a rank-one P, which immediately converges to stationarity.

Assume $\beta \geq \alpha$. Then, by (1.1.6) and Lemma 4.1.9,

$$d(t) = \max_x \frac{1}{2} \sum_y |P^t(x,y) - \pi(y)| = \frac{\beta}{\alpha+\beta} |1-\alpha-\beta|^t.$$

As a result,

$$\mathrm{t_{mix}}(\varepsilon) = \left\lceil \frac{\log\left(\varepsilon \frac{\alpha+\beta}{\beta}\right)}{\log|1-\alpha-\beta|} \right\rceil = \left\lceil \frac{\log \varepsilon^{-1} - \log\left(\frac{\alpha+\beta}{\beta}\right)}{\log|1-\alpha-\beta|^{-1}} \right\rceil.$$

◀

Spectral gap and mixing Assume further that P is aperiodic. Recall that by the convergence theorem (Theorem 1.1.33), for all x, y, $P^t(x,y) \to \pi(y)$ as $t \to +\infty$, and that the mixing time (Definition 1.1.35) is

$$\mathrm{t_{mix}}(\varepsilon) := \min\{t \geq 0 : d(t) \leq \varepsilon\},$$

where $d(t) := \max_{x \in V} \|P^t(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}}$. It will be convenient to work with a different notion of distance.

Definition 5.2.9 (Separation distance)**.** *The* separation distance *is defined as*

separation distance

$$s_x(t) := \max_{y \in V} \left[1 - \frac{P^t(x,y)}{\pi(y)}\right],$$

and we let $s(t) := \max_{x \in V} s_x(t)$.

Lemma 5.2.10 (Separation distance and total variation distance)**.**

$$d(t) \leq s(t).$$

Proof. By Lemma 4.1.15,

$$\begin{aligned} \|P^t(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} &= \sum_{y : P^t(x,y) < \pi(y)} \left[\pi(y) - P^t(x,y)\right] \\ &= \sum_{y : P^t(x,y) < \pi(y)} \pi(y) \left[1 - \frac{P^t(x,y)}{\pi(y)}\right] \\ &\leq s_x(t). \end{aligned}$$

Since this holds for any x, the claim follows. ■

It follows that, from the spectral decomposition (Theorem 5.2.7), the speed of convergence of $P^t(x, y)$ to $\pi(y)$ is dominated by the largest eigenvalue of P not equal to 1.

Definition 5.2.11 (Spectral gap)**.** *The* absolute spectral gap *is* $\gamma_* := 1 - \lambda_*$ *where* $\lambda_* := |\lambda_2| \vee |\lambda_n|$. *The* spectral gap *is* $\gamma := 1 - \lambda_2$.

absolute spectral gap

By Lemmas 5.2.3 and 5.2.4, we have $\gamma_* > 0$ when P is irreducible and aperiodic. Note that the eigenvalues of the lazy version $\frac{1}{2}P + \frac{1}{2}I$ of P are $\{\frac{1}{2}(\lambda_j + 1)\}_{j=1}^n$ which are all nonnegative. So, there, $\gamma_* = \gamma$.

Definition 5.2.12 (Relaxation time)**.** *The* relaxation time *is defined as*

relaxation time

$$\mathrm{t_{rel}} := \gamma_*^{-1}.$$

Example 5.2.13 (Two-state chain (continued))**.** Returning to Example 5.2.8, there are two cases:

- $\alpha + \beta \leq 1$: In that case the (absolute) spectral gap is $\gamma_* = \gamma = \alpha + \beta$ and the relaxation time is $\mathrm{t_{rel}} = 1/(\alpha + \beta)$.
- $\alpha + \beta > 1$: In that case the absolute spectral gap is $\gamma_* = 2 - \alpha - \beta$ and the relaxation time is $\mathrm{t_{rel}} = 1/(2 - \alpha - \beta)$.

◀

The following result clarifies the relationship between the mixing and relaxation times. Let $\pi_{\min} = \min_x \pi(x)$.

Theorem 5.2.14 (Mixing time and relaxation time)**.** *Let P be reversible, irreducible, and aperiodic with positive stationary distribution π. For all $\varepsilon > 0$,*

$$(\mathrm{t_{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right) \leq \mathrm{t_{mix}}(\varepsilon) \leq \log\left(\frac{1}{\varepsilon \pi_{\min}}\right) \mathrm{t_{rel}}.$$

Proof. We start with the upper bound. By Lemma 5.2.10, it suffices to find t such that $s(t) \leq \varepsilon$. By the spectral decomposition and Cauchy-Schwarz,

$$\left|\frac{P^t(x,y)}{\pi(y)} - 1\right| \leq \lambda_*^t \sum_{j=2}^n |f_j(x) f_j(y)| \leq \lambda_*^t \sqrt{\sum_{j=2}^n f_j(x)^2 \sum_{j=2}^n f_j(y)^2}.$$

By Lemma 5.2.6, $\sum_{j=2}^n f_j(x)^2 \leq \pi(x)^{-1}$. Plugging this back above, we get

$$\left|\frac{P^t(x,y)}{\pi(y)} - 1\right| \leq \lambda_*^t \sqrt{\pi(x)^{-1}\pi(y)^{-1}} \leq \frac{\lambda_*^t}{\pi_{\min}} = \frac{(1-\gamma_*)^t}{\pi_{\min}} \leq \frac{e^{-\gamma_* t}}{\pi_{\min}}, \quad (5.2.5)$$

where we used that $1 - z \leq e^{-z}$ for all $z \in \mathbb{R}$ (see Exercise 1.16). Observe that the right-hand side is less than ε when $t \geq \log\left(\frac{1}{\varepsilon \pi_{\min}}\right) t_{\mathrm{rel}}$.

For the lower bound, let f_* be an eigenfunction associated with an eigenvalue achieving $\lambda_* := |\lambda_2| \vee |\lambda_n|$. Let z be such that $|f_*(z)| = \|f_*\|_\infty$. By Lemma 5.2.5, $\pi f_* = 0$. Hence

$$\begin{aligned}
\lambda_*^t |f_*(z)| &= |P^t f_*(z)| \\
&= \left| \sum_y [P^t(z,y) f_*(y) - \pi(y) f_*(y)] \right| \\
&\leq \|f_*\|_\infty \sum_y |P^t(z,y) - \pi(y)| \leq \|f_*\|_\infty 2d(t),
\end{aligned}$$

so $d(t) \geq \frac{1}{2}\lambda_*^t$. When $t = t_{\mathrm{mix}}(\varepsilon)$, $\varepsilon \geq d(t_{\mathrm{mix}}(\varepsilon)) \geq \frac{1}{2}\lambda_*^{t_{\mathrm{mix}}(\varepsilon)}$. Therefore, rearranging and taking a logarithm, we get

$$t_{\mathrm{mix}}(\varepsilon)\left(\frac{1}{\lambda_*} - 1\right) \geq t_{\mathrm{mix}}(\varepsilon) \log\left(\frac{1}{\lambda_*}\right) \geq \log\left(\frac{1}{2\varepsilon}\right),$$

where we used $z = 1 - \lambda_*^{-1}$ in $1 - z \leq e^{-z}$ to get the first inequality. The result follows from $\left(\frac{1}{\lambda_*} - 1\right)^{-1} = \left(\frac{1-\lambda_*}{\lambda_*}\right)^{-1} = \left(\frac{\gamma_*}{1-\gamma_*}\right)^{-1} = t_{\mathrm{rel}} - 1$. ■

5.2.2 ▷ *Random walks: a spectral look at cycles and hypercubes*

We illustrate the results in the previous subsection to random walk on cycles and hypercubes.

Random walk on a cycle

Consider simple random walk on the n-cycle (see Example 1.1.17). That is, $V := \{0, 1, \ldots, n-1\}$ and $P(x,y) = 1/2$ if and only if $|x - y| = 1 \mod n$. We assume that n is odd to avoid periodicity issues. Let $\pi \equiv n^{-1}$ be the stationary distribution (by symmetry and $|V| = n$). We showed in Section 4.3.2 that (for the lazy version of the chain) the mixing time is $t_{\mathrm{mix}}(\varepsilon) = \Theta(n^2)$.

Here we use spectral techniques. We first compute the eigendecomposition, which in this case can be determined explicitly.

Lemma 5.2.15 (Cycle: eigenbasis)**.** *For $j = 1, \ldots, n-1$, the function*

$$g_j(x) := \sqrt{2} \cos\left(\frac{2\pi j x}{n}\right), \qquad x = 0, 1, \ldots, n-1,$$

is an eigenfunction of P with eigenvalue

$$\mu_j := \cos\left(\frac{2\pi j}{n}\right),$$

and $g_0 = 1$ is an eigenfunction with eigenvalue 1. Moreover the g_js are orthonormal in $\ell^2(V, \pi)$.

Proof. We know from Lemma 5.2.3 that 1 is an eigenfunction with eigenvalue 1. Let $j \in \{1, \ldots, n-1\}$. Note that, for all x, switching momentarily to the complex representation (where we use i for the imaginary unit)

$$\begin{aligned}
\sum_y P(x,y) g_j(y) &= \frac{1}{2}\left[\sqrt{2}\cos\left(\frac{2\pi j(x-1)}{n}\right) + \sqrt{2}\cos\left(\frac{2\pi j(x+1)}{n}\right)\right] \\
&= \frac{\sqrt{2}}{2}\left[\frac{e^{i\frac{2\pi j(x-1)}{n}} + e^{-i\frac{2\pi j(x-1)}{n}}}{2} + \frac{e^{i\frac{2\pi j(x+1)}{n}} + e^{-i\frac{2\pi j(x+1)}{n}}}{2}\right] \\
&= \sqrt{2}\left[\frac{e^{i\frac{2\pi jx}{n}} + e^{-i\frac{2\pi jx}{n}}}{2}\right]\left[\frac{e^{i\frac{2\pi j}{n}} + e^{-i\frac{2\pi j}{n}}}{2}\right] \\
&= \left[\sqrt{2}\cos\left(\frac{2\pi jx}{n}\right)\right]\left[\cos\left(\frac{2\pi j}{n}\right)\right] \\
&= \cos\left(\frac{2\pi j}{n}\right) g_j(x).
\end{aligned}$$

The orthonormality follows from standard trigonometric identities. We prove only that the g_js have unit norm. We use the Dirichtlet kernel (see Exercise 5.8)

$$1 + 2\sum_{k=1}^{n} \cos k\theta = \frac{\sin((n+1/2)\theta)}{\sin(\theta/2)},$$

for $\theta \neq 0$, and the identity $\cos^2(\theta) = \frac{1}{2}(1 + \cos(2\theta))$. For $j = 0$, $g_j = 1$ and the norm squared is $\sum_x \pi(x) = 1$. For $j \neq 0$, we have $\|g_j\|_\pi^2$ is

$$\begin{aligned}
\sum_{x \in V} \pi(x) g_j(x)^2 &= \frac{1}{n}\sum_{x=0}^{n-1} 2\cos^2\left(\frac{2\pi jx}{n}\right) \\
&= \frac{1}{n}\sum_{x=0}^{n-1}\left(1 + \cos\left(\frac{4\pi jx}{n}\right)\right) \\
&= 1 + \frac{1}{n}\sum_{k=1}^{n}\cos\left(k\frac{4\pi j}{n}\right) \\
&= 1 + \frac{1}{2}\left[\frac{\sin((n+1/2)(4\pi j/n))}{\sin((4\pi j/n)/2)} - 1\right],
\end{aligned}$$

which is indeed 1. ■

From the eigenvalues, we derive the relaxation time (Definition 5.2.12) analytically.

Theorem 5.2.16 (Cycle: relaxation time)**.** *The relaxation time for lazy simple random walk on the n-cycle is*

$$\mathrm{t}_{\mathrm{rel}} = \frac{1}{1 - \cos\left(\frac{2\pi}{n}\right)} = \Theta(n^2).$$

Proof. By Lemma 5.2.15, the absolute spectral gap (Definition 5.2.11) is $1 - \cos\left(\frac{2\pi}{n}\right)$, using that n is odd. By a Taylor expansion,

$$1 - \cos\left(\frac{2\pi}{n}\right) = \frac{4\pi^2}{n^2} + O(n^{-4}). \quad ■$$

Since $\pi_{\min} = 1/n$, we get $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = O(n^2 \log n)$ and $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = \Omega(n^2)$ by Theorem 5.2.14.

It turns out our upper bound is off by a logarithmic factor. A sharper bound on the mixing time can be obtained by working directly with the spectral decomposition. By Lemma 4.1.9 and Cauchy-Schwarz (Theorem B.4.8), for any $x \in V$,

$$\begin{aligned}
4\|P^t(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}}^2 &= \left\{\sum_y \pi(y)\left|\frac{P^t(x,y)}{\pi(y)} - 1\right|\right\}^2 \\
&\leq \sum_y \pi(y)\left(\frac{P^t(x,y)}{\pi(y)} - 1\right)^2 \\
&= \left\|\sum_{j=1}^{n-1} \mu_j^t g_j(x) g_j\right\|_\pi^2 \\
&= \sum_{j=1}^{n-1} \mu_j^{2t} g_j(x)^2,
\end{aligned}$$

where we used the spectral decomposition of P^t (Theorem 5.2.7) on the third line and Parseval's identity (i.e., (5.2.1)) on the fourth line.

Here comes the trick: the total variation distance does not depend on the starting point x by symmetry. Multiplying by $\pi(x)$ and summing over x—on the right-hand side only—gives

$$\begin{aligned}
4\|P^t(x,\cdot)-\pi(\cdot)\|_{\mathrm{TV}}^2 &\leq \sum_x \pi(x)\sum_{j=1}^{n-1}\mu_j^{2t} g_j(x)^2 \\
&= \sum_{j=1}^{n-1}\mu_j^{2t}\sum_x \pi(x) g_j(x)^2 \\
&= \sum_{j=1}^{n-1}\mu_j^{2t},
\end{aligned}$$

where we used that $\|g_j\|_\pi^2 = 1$.

We get

$$4d(t)^2 \leq \sum_{j=1}^{n-1}\cos^{2t}\left(\frac{2\pi j}{n}\right) = 2\sum_{j=1}^{(n-1)/2}\cos^{2t}\left(\frac{2\pi j}{n}\right).$$

For $x \in [0,\pi/2)$, $\cos x \leq e^{-x^2/2}$ (see Exercise 1.16). Then

$$\begin{aligned}
4d(t)^2 &\leq 2\sum_{j=1}^{(n-1)/2}\exp\left(-\frac{4\pi^2 j^2}{n^2}t\right) \\
&\leq 2\exp\left(-\frac{4\pi^2}{n^2}t\right)\sum_{j=1}^{\infty}\exp\left(-\frac{4\pi^2(j^2-1)}{n^2}t\right) \\
&\leq 2\exp\left(-\frac{4\pi^2}{n^2}t\right)\sum_{\ell=0}^{\infty}\exp\left(-\frac{4\pi^2 t}{n^2}\ell\right) \\
&= \frac{2\exp\left(-\frac{4\pi^2}{n^2}t\right)}{1-\exp\left(-\frac{4\pi^2}{n^2}t\right)}.
\end{aligned}$$

So $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = O(n^2)$.

Random walk on the hypercube

Consider simple random walk on the hypercube $V := \{-1,+1\}^n$ where $x \sim y$ if they differ at exactly one coordinate. We consider the lazy version to avoid issues of periodicity (see Example 1.1.31). Let P be the transition matrix and let $\pi \equiv 2^{-n}$ be the stationary distribution (by symmetry and $|V| = 2^n$). We showed in Section 4.3.2 that $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = \Theta(n\log n)$. Here we use spectral techniques.

For $J \subseteq [n]$, we let

$$\chi_J(x) = \prod_{j \in J} x_j, \qquad x \in V.$$

These are called *parity functions*. We show that the parity functions form an eigenbasis of the transition matrix.

parity function

Lemma 5.2.17 (Hypercube: eigenbasis). *For all $J \subseteq [n]$, the function χ_J is an eigenfunction of P with eigenvalue*

$$\mu_J := \frac{n - |J|}{n}.$$

Moreover the χ_Js are orthonormal in $\ell^2(V, \pi)$.

Proof. For $x \in V$ and $i \in [n]$, let $x^{[i]}$ be x where coordinate i is flipped. Note that, for all J, x,

$$\begin{aligned}
\sum_y P(x,y)\chi_J(y) &= \frac{1}{2}\chi_J(x) + \frac{1}{2}\sum_{i=1}^n \frac{1}{n}\chi_J(x^{[i]}) \\
&= \left\{\frac{1}{2} + \frac{1}{2}\frac{n-|J|}{n}\right\}\chi_J(x) - \frac{1}{2}\frac{|J|}{n}\chi_J(x) \\
&= \frac{n-|J|}{n}\chi_J(x).
\end{aligned}$$

For the orthonormality, note that

$$\sum_{x \in V} \pi(x)\chi_J(x)^2 = \sum_{x \in V} \frac{1}{2^n}\prod_{j \in J} x_j^2 = 1.$$

For $J \neq J' \subseteq [n]$,

$$\begin{aligned}
&\sum_{x \in V} \pi(x)\chi_J(x)\chi_{J'}(x) \\
&= \sum_{x \in V} \frac{1}{2^n} \prod_{j \in J \cap J'} x_j^2 \prod_{j \in J \setminus J'} x_j \prod_{j \in J' \setminus J} x_j \\
&= \frac{2^{|J \cap J'|}}{2^n} \prod_{j \in J \setminus J'} \left(\sum_{x_j \in \{-1,+1\}} x_j \right) \prod_{j \in J' \setminus J} \left(\sum_{x_j \in \{-1,+1\}} x_j \right) \\
&= 0,
\end{aligned}$$

since at least one of $J \setminus J'$ or $J' \setminus J$ is nonempty. ■

From the eigenvalues, we obtain the relaxation time.

Theorem 5.2.18 (Hypercube: relaxation time)**.** *The relaxation time for lazy simple random walk on the n-dimensional hypercube is*

$$t_{\mathrm{rel}} = n.$$

Proof. From Lemma 5.2.17, the absolute spectral gap is

$$\gamma_* = \gamma = 1 - \frac{n-1}{n} = \frac{1}{n}. \quad \blacksquare$$

Note that $\pi_{\min} = 1/2^n$. Hence, by Theorem 5.2.14, we have $t_{\mathrm{mix}}(\varepsilon) = O(n^2)$ and $t_{\mathrm{mix}}(\varepsilon) = \Omega(n)$. Those bounds, it turns out, are both off.

As we did for the cycle, we obtain a sharper upper bound by working directly with the spectral decomposition. By the same argument we used there,

$$4d(t)^2 \leq \sum_{J \neq \emptyset} \mu_J^{2t}.$$

Then

$$\begin{aligned}
4d(t)^2 &\leq \sum_{J \neq \emptyset} \left(\frac{n - |J|}{n}\right)^{2t} \\
&= \sum_{\ell=1}^{n} \binom{n}{\ell} \left(1 - \frac{\ell}{n}\right)^{2t} \\
&\leq \sum_{\ell=1}^{n} \binom{n}{\ell} \exp\left(-\frac{2t\ell}{n}\right) \\
&= \left(1 + \exp\left(-\frac{2t}{n}\right)\right)^{n} - 1,
\end{aligned}$$

where we used that $1 - x \leq e^{-x}$ for all x (see Exercise 1.16). So, by definition, $t_{\mathrm{mix}}(\varepsilon) \leq \frac{1}{2} n \log n + O(n)$.

Remark 5.2.19. *In fact, lazy simple random walk on the n-dimensional hypercube has a "cutoff" at $(1/2)n \log n$. Roughly speaking, within a time window of size $O(n)$, the total variation distance to the stationary distribution goes from near 1 to near 0. See, e.g., [LPW06, Section 18.2.2].*

5.2.3 ▷ *Markov chains: Varopoulos-Carne and diameter-based bounds on the mixing time*

If (S_t) is simple random walk on $\mathbb{Z}$, then Lemma 2.4.3 guarantees that for any $x, y \in \mathbb{Z}$

$$P^t(x,y) \leq e^{-|x-y|^2/2t}, \tag{5.2.6}$$

where P is the transition matrix of (S_t). Interestingly a similar bound holds for *any* reversible Markov chain—and Lemma 2.4.3 plays an unexpected role in its proof. An application to mixing times is discussed below.

Varopoulos-Carne bound

Our main bound is the following. Recall that a reversible Markov chain is equivalent to a random walk on the network corresponding to its positive transition probabilities (see Definition 1.2.7 and the discussion following it).

Theorem 5.2.20 (Varopoulos-Carne bound)**.** *Let P be the transition matrix of an irreducible Markov chain (X_t) on the countable state space V. Assume further that P is reversible with respect to the stationary measure π and that the corresponding network $\mathcal{N}$ is locally finite. Then the following hold*

$$\forall x, y \in V, \forall t \in \mathbb{N}, \qquad P^t(x,y) \leq 2\sqrt{\frac{\pi(y)}{\pi(x)}}e^{-\rho(x,y)^2/2t},$$

where $\rho(x,y)$ is the graph distance between x and y on $\mathcal{N}$.

As a sanity check before proving the theorem, note that if the chain is aperiodic and π is the stationary distribution then by the convergence theorem (Theorem 1.1.33)

$$P^t(x,y) \to \pi(y) \leq 2\sqrt{\frac{\pi(y)}{\pi(x)}}, \qquad \text{as } t \to +\infty,$$

since $\pi(x), \pi(y) \leq 1$.

Proof of Theorem 5.2.20. The idea of the proof is to show that

$$P^t(x,y) \leq 2\sqrt{\frac{\pi(y)}{\pi(x)}}\mathbb{P}[S_t \geq \rho(x,y)],$$

where again (S_t) is simple random walk on $\mathbb{Z}$ started at 0, and then use the Chernoff bound (Lemma 2.4.3).

By the local finiteness assumption, only a finite number of states can be reached by time t. Hence we can reduce the problem to a finite state space. More precisely, let $\tilde{V} = \{z \in V \,:\, \rho(x,z) \leq t\}$ and for $z, w \in \tilde{V}$

$$\tilde{P}(z,w) = \begin{cases} P(z,w) & \text{if } z \neq w, \\ P(z,z) + P(z, V \setminus \tilde{V}) & \text{otherwise.} \end{cases}$$

By construction $\tilde{P}$ is reversible with respect to $\tilde{\pi} = \pi/\pi(\tilde{V})$ on $\tilde{V}$. Because within time t one never reaches a state z where $P(z, V \setminus \tilde{V}) > 0$, by Chapman-Kolmogorov (Theorem 1.1.20) and using the fact that $\tilde{\pi}(y)/\tilde{\pi}(x) = \pi(y)/\pi(x)$, it suffices to prove the result for $\tilde{P}$. Hence we assume without loss of generality that V is finite with $|V| = n$.

To relate (X_t) to simple random walk on $\mathbb{Z}$, we use a special representation of P^t based on Chebyshev polynomials. For $\xi = \cos\theta \in [-1,1]$,

$$T_k(\xi) = \cos k\theta,$$

is a *Chebyshev polynomial of the first kind*. Note that $|T_k(\xi)| \leq 1$ on $[-1,1]$ by definition. The classical trigonometric identity (to see this, write it in complex form) *Chebyshev polynomials*

$$\cos((k+1)\theta) + \cos((k-1)\theta) = 2\cos\theta\cos(k\theta),$$

implies the recursion

$$T_{k+1}(\xi) + T_{k-1}(\xi) = 2\xi\, T_k(\xi),$$

which in turn implies that T_k is indeed a polynomial. It has degree k from induction and the fact that $T_0(\xi) = 1$ and $T_1(\xi) = \xi$. The connection to simple random walk on $\mathbb{Z}$ comes from the following somewhat miraculous representation (which does not rely on reversibility). Let $T_k(P)$ denote the polynomial T_k evaluated at P as a matrix polynomial.

Lemma 5.2.21.

$$P^t = \sum_{k=-t}^{t} \mathbb{P}[S_t = k]\, T_{|k|}(P).$$

Proof. It suffices to prove

$$\xi^t = \sum_{k=-t}^{t} \mathbb{P}[S_t = k]\, T_{|k|}(\xi),$$

as an identity of polynomials. By the binomial theorem (Appendix A),

$$\xi^t = \left(\frac{e^{i\theta}+e^{-i\theta}}{2}\right)^t = \sum_{\ell=0}^{t} 2^{-t}\binom{t}{\ell}(e^{i\theta})^\ell(e^{-i\theta})^{t-\ell} = \sum_{k=-t}^{t}\mathbb{P}[S_t = k]e^{ik\theta},$$

where we used that the probability that

$$S_t = -t + 2\ell = (+1)\ell + (-1)(t-\ell),$$

is the event of making ℓ steps to the right and $t-\ell$ steps to the left. Now take real parts on both sides and use that $\cos(k\theta) = \cos(-k\theta)$ to get the claim. (Put differently, $(\cos\theta)^t$ is the characteristic function $\mathbb{E}[e^{i\theta S_t}]$ of S_t.) ■

We bound $T_k(P)(x,y)$ as follows.

Lemma 5.2.22. *It holds that*

$$T_k(P)(x,y) = 0, \qquad \forall k < \rho(x,y),$$

and

$$T_k(P)(x,y) \leq \sqrt{\frac{\pi(y)}{\pi(x)}}, \qquad \forall k \geq \rho(x,y).$$

Proof. Note that $T_k(P)(x,y) = 0$ when $k < \rho(x,y)$ because $T_k(P)(x,y)$ is a function of the entries $P^\ell(x,y)$ for $\ell \leq k$, all of which are 0.

We work on $\ell^2(V,\pi)$. Let $f_1,\ldots,f_n$ be an eigendecomposition of P orthonormal with respect to the inner product $\langle\cdot,\cdot\rangle_\pi$ with eigenvalues $\lambda_1,\ldots,\lambda_n \in [-1,1]$. Such a decomposition exists by Theorem 5.2.1. Then $f_1,\ldots,f_n$ is also an eigendecomposition of the polynomial $T_k(P)$ with eigenvalues

$$T_k(\lambda_1),\ldots,T_k(\lambda_n) \in [-1,1],$$

by the definition of the Chebyshev polynomials. By decomposing any function $f = \sum_{i=1}^n \alpha_i f_i$ over this eigenbasis, that implies that

$$\begin{aligned}
\|T_k(P)f\|_\pi^2 &= \left\|\sum_{i=1}^n \alpha_i T_k(\lambda_i) f_i\right\|_\pi^2 \\
&= \sum_{i=1}^n \alpha_i^2\, T_k(\lambda_i)^2 \\
&\leq \sum_{i=1}^n \alpha_i^2 \\
&= \|f\|_\pi^2,
\end{aligned} \tag{5.2.7}$$

where we used Parseval's identity (5.2.1) twice and the fact that $T_k(\lambda_i)^2 \in [0,1]$.

Let δ_z denote the point mass at z. By Cauchy-Schwarz (Theorem B.4.8) and (5.2.7),

$$T_k(P)(x,y) = \frac{\langle \delta_x, T_k(P)\delta_y\rangle_\pi}{\pi(x)} \leq \frac{\|\delta_x\|_\pi \|\delta_y\|_\pi}{\pi(x)} = \frac{\sqrt{\pi(x)}\sqrt{\pi(y)}}{\pi(x)} = \sqrt{\frac{\pi(y)}{\pi(x)}},$$

for any k (in particular for $k \geq \rho(x,y)$) and we have proved the claim. ■

Combining the two lemmas gives the result. ■

Remark 5.2.23. *The local finiteness assumption is made for simplicity only. The result holds for any countable-space, reversible chain. See [LP16, Section 13.2].*

Lower bound on mixing Let (X_t) be an irreducible aperiodic (for now not necessarily reversible) Markov chain with finite state space V and stationary distribution π. Recall that, for a fixed $0 < \varepsilon < 1/2$, the mixing time is

$$\mathrm{t_{mix}}(\varepsilon) = \min\{t \,:\, d(t) \leq \varepsilon\},$$

where

$$d(t) = \max_{x \in V} \|P^t(x,\cdot\,) - \pi\|_{\mathrm{TV}}.$$

It is intuitively clear that $\mathrm{t_{mix}}(\varepsilon)$ is at least of the order of the "diameter" of the transition graph of P. For $x,y \in V$, let $\rho(x,y)$ be the graph distance between x and y on the undirected version of the transition graph, that is, ignoring the orientation of the edges. With this definition, a shortest directed path from x to y contains at least $\rho(x,y)$ edges. Here we define the *diameter* of the transition graph as $\Delta := \max_{x,y\in V} \rho(x,y)$. Let x_0, y_0 be a pair of vertices achieving the diameter. Then we claim that $P^{\lfloor(\Delta-1)/2\rfloor}(x_0,\cdot\,)$ and $P^{\lfloor(\Delta-1)/2\rfloor}(y_0,\cdot\,)$ are supported on disjoint sets. To see this let *diameter*

$$A = \{z \in V \,:\, \rho(x_0,z) < \rho(y_0,z)\},$$

be the set of states closer to x_0 than y_0. See Figure 5.2. By the triangle inequality for ρ, any z such that $\rho(x_0,z) \leq \lfloor(\Delta-1)/2\rfloor$ is in A, otherwise we would have $\rho(y_0,z) \leq \rho(x_0,z) \leq \lfloor(\Delta-1)/2\rfloor$ and hence $\rho(x_0,y_0) \leq \rho(x_0,z) + \rho(y_0,z) \leq 2\lfloor(\Delta-1)/2\rfloor < \Delta$, a contradiction. Similarly, if $\rho(y_0,z) \leq \lfloor(\Delta-1)/2\rfloor$, then $z \in A^c$. By the triangle inequality for the total variation distance,

$$\begin{aligned} d(\lfloor(\Delta-1)/2\rfloor) &\geq \frac{1}{2}\left\|P^{\lfloor(\Delta-1)/2\rfloor}(x_0,\cdot\,) - P^{\lfloor(\Delta-1)/2\rfloor}(y_0,\cdot\,)\right\|_{\mathrm{TV}} \\ &\geq \frac{1}{2}\left\{P^{\lfloor(\Delta-1)/2\rfloor}(x_0,A) - P^{\lfloor(\Delta-1)/2\rfloor}(y_0,A)\right\} \\ &= \frac{1}{2}\{1-0\} = \frac{1}{2}, \end{aligned} \tag{5.2.8}$$

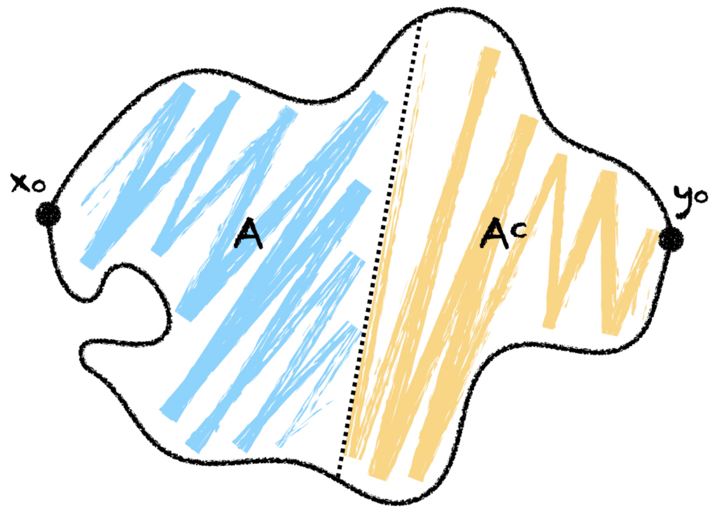

Figure 5.2: The supports of $P^{\lfloor(\Delta-1)/2\rfloor}(x_0, \cdot\,)$ and $P^{\lfloor(\Delta-1)/2\rfloor}(y_0, \cdot\,)$ are contained in A and A^c respectively.

where we used (1.1.4) on the second line, so that:

Claim 5.2.24.

$$\mathrm{t_{mix}}(\varepsilon) \geq \frac{\Delta}{2}.$$

This bound is often far from the truth. Consider for instance simple random walk on a cycle of size n. The diameter is $\Delta = n/2$. But Lemma 2.4.3 suggests that it takes time of order Δ^2 to even reach the antipode of the starting point, let alone achieve stationarity. More generally, when P is *reversible*, the "diffusive behavior" captured by the Varopoulos-Carne bound (Theorem 5.2.20) implies that the mixing time does indeed scale at least as the *square* of the diameter.

Assume that P is reversible with respect to π and has diameter Δ. Letting $n = |V|$ and $\pi_{\min} = \min_{x\in V}\pi(x)$, we then have the following.

Claim 5.2.25. *The following lower bound holds*

$$\mathrm{t_{mix}}(\varepsilon) \geq \frac{\Delta^2}{12\log n + 4|\log \pi_{\min}|},$$

provided $n \geq \frac{16}{(1-2\varepsilon)^2}$.

Proof. The proof is based on the same argument we used to derive our first diameter-based bound, except that the Varopoulos-Carne bound gives a better dependence on the diameter. Namely, let x_0, y_0, and A be as above. By the Varopoulos-Carne bound,

$$P^t(x_0, A^c) = \sum_{z \in A^c} P^t(x_0, z) \leq \sum_{z \in A^c} 2\sqrt{\frac{\pi(z)}{\pi(x_0)}} e^{-\frac{\rho^2(x_0,z)}{2t}} \leq 2n\pi_{\min}^{-1/2} e^{-\frac{\Delta^2}{8t}},$$

where we used that $|A^c| \leq n$ and $\rho(x_0, z) \geq \frac{\Delta}{2}$ for $z \in A^c$. For any

$$t < \frac{\Delta^2}{12 \log n + 4|\log \pi_{\min}|}, \tag{5.2.9}$$

we get that

$$P^t(x_0, A^c) \leq 2n\pi_{\min}^{-1/2} \exp\left(-\frac{3\log n + |\log \pi_{\min}|}{2}\right) = \frac{2}{\sqrt{n}},$$

or $P^t(x_0, A) \geq 1 - \frac{2}{\sqrt{n}}$. Similarly, $P^t(y_0, A) \leq \frac{2}{\sqrt{n}}$ so that arguing as in (5.2.8)

$$d(t) \geq \frac{1}{2}\left\{1 - \frac{2}{\sqrt{n}} - \frac{2}{\sqrt{n}}\right\} = \frac{1}{2} - \frac{2}{\sqrt{n}} \geq \varepsilon,$$

for t as in (5.2.9) and n as in the statement. ∎

Remark 5.2.26. *The dependence on Δ and $\pi_{\min}$ in Claim 5.2.25 cannot be improved. See [LP16, Section 13.3].*

5.2.4 ▷ *Randomized algorithms: Markov chain Monte Carlo and a quantitative ergodic theorem*

In Markov chain Monte Carlo methods, one generates samples from a probability distribution of interest π over some state space V in order to estimate some of its properties, for example, its mean, by designing and then running a Markov chain with stationary distribution π. The Metropolis algorithm from Example 1.1.30 is a standard way of constructing such a chain. These techniques play a central role in Bayesian statistics in particular where π is the so-called posterior distribution given the data.

We restrict ourselves here to finite V and, without loss of generality, we assume that $V = [n]$. Let P be an irreducible chain reversible with respect to a stationary distribution $\pi = (\pi_x)_{x \in V}$. As previously, we work on $\ell^2(V, \pi)$. Let $f : V \to \mathbb{R}$ be a function in $\ell^2(V, \pi)$. Recall that

$$\pi f = \sum_{x \in V} \pi_x f(x).$$

Our goal is to estimate πf from the sample path of the Markov chain $(X_t)_{t \geq 0}$ with transition matrix P. Indeed the ergodic theorem guarantees that

$$\frac{1}{T} \sum_{t=1}^{T} f(X_t) \to \pi f,$$

almost surely as $T \to +\infty$ for any starting point. We derive a simple, quantitative version of this statement that provides insights into how long the chain needs to be run to get an accurate estimate in terms of the spectral gap.

Theorem 5.2.27 (Ergodic theorem: reversible case)**.** *Let $P = (P_{x,y})_{x,y \in V}$ be an irreducible aperiodic transition matrix over a finite state space V reversible with respect to the stationary distribution $\pi = (\pi_x)_{x \in V}$. Let $f : V \to \mathbb{R}$ be a function in $\ell^2(V, \pi)$. Then for any initial distribution $\mu = (\mu_x)_{x \in V}$*

$$\frac{1}{T} \sum_{t=1}^{T} f(X_t) \to \pi f,$$

in probability as $T \to +\infty$. Moreover, for any $\varepsilon > 0$,

$$\mathbb{P}\left[\left|\frac{1}{T} \sum_{t=1}^{T} f(X_t) - \pi f\right| \geq \varepsilon\right] \leq \frac{9\pi_{\min}^{-1} \|f\|_\infty^2 \gamma_*^{-1} \frac{1}{T}}{(\varepsilon - \pi_{\min}^{-1} \|f\|_\infty \gamma_*^{-1} \frac{1}{T})^2},$$

as $T \to +\infty$, where $\gamma_* > 0$ is the absolute spectral gap of P.

Recall that, by Lemmas 5.2.3 and 5.2.4, we have $\gamma_* > 0$ since P is irreducible and aperiodic. We will first need the following lemma.

Lemma 5.2.28 (Convergence of the expectation)**.** *For any initial distribution $\mu = (\mu_x)_{x \in V}$ and any t*

$$|\,\mathbb{E}[f(X_t)] - \pi f\,| \leq (1 - \gamma_*)^t \pi_{\min}^{-1} \|f\|_\infty.$$

Proof. We have

$$|\,\mathbb{E}[f(X_t)] - \pi f\,| = \left| \sum_x \sum_y \mu_x P_{x,y}^t f(y) - \sum_y \pi_y f(y) \right|.$$

Because $\sum_x \mu_x = 1$, the right-hand side is

$$\begin{aligned}
&= \left| \sum_x \sum_y \mu_x P^t_{x,y} f(y) - \sum_x \sum_y \mu_x \pi_y f(y) \right| \\
&\leq \sum_x \mu_x \sum_y \left| P^t_{x,y} - \pi_y \right| |f(y)|,
\end{aligned}$$

by the triangle inequality.

Now by (5.2.5) this is

$$\begin{aligned}
&\leq \sum_x \mu_x \sum_y (1-\gamma_*)^t \frac{\pi_y}{\pi_{\min}} |f(y)| \\
&= (1-\gamma_*)^t \frac{1}{\pi_{\min}} \sum_x \mu_x \sum_y \pi_y |f(y)| \\
&\leq (1-\gamma_*)^t \pi_{\min}^{-1} \|f\|_\infty.
\end{aligned}$$

That proves the claim. ■

Proof of Theorem 5.2.27. We use Chebyshev's inequality (Theorem 2.1.2), similarly to the proof of the L^2 weak law of large numbers (Theorem 2.1.6). In particular, we note that the X_ts are *not* independent.

By Lemma 5.2.28, the expectation of the time average can be bounded as follows

$$\begin{aligned}
\left| \mathbb{E}\left[\frac{1}{T} \sum_{t=1}^T f(X_t) \right] - \pi f \right| &\leq \frac{1}{T} \sum_{t=1}^T |\mathbb{E}[f(X_t)] - \pi f| \\
&\leq \frac{1}{T} \sum_{t=1}^T (1-\gamma_*)^t \pi_{\min}^{-1} \|f\|_\infty \\
&\leq \pi_{\min}^{-1} \|f\|_\infty \frac{1}{T} \sum_{t=0}^{+\infty} (1-\gamma_*)^t \\
&= \pi_{\min}^{-1} \|f\|_\infty \gamma_*^{-1} \frac{1}{T} \to 0,
\end{aligned}$$

as $T \to +\infty$, since $\gamma_* > 0$.

Next we bound the variance of the sum. We have

$$\mathrm{Var}\left[\frac{1}{T} \sum_{t=1}^T f(X_t) \right] = \frac{1}{T^2} \sum_{t=1}^T \mathrm{Var}[f(X_t)] + \frac{2}{T^2} \sum_{1 \leq s < t \leq T} \mathrm{Cov}[f(X_s), f(X_t)].$$

We bound the variance and covariance terms separately.

To obtain convergence, a trivial bound on the variance suffices

$$0 \leq \mathrm{Var}[f(X_t)] \leq \mathbb{E}[f(X_t)^2] \leq \|f\|_\infty^2.$$

Hence,

$$0 \leq \frac{1}{T^2} \sum_{t=1}^{T} \mathrm{Var}[f(X_t)] \leq \frac{T\|f\|_\infty^2}{T^2} \to 0,$$

as $T \to +\infty$.

Bounding the covariance requires a more delicate argument. Fix $1 \leq s < t \leq T$. The trick is to condition on X_s and use the Markov Property (Theorem 1.1.18). By definition of the covariance, the tower property (Lemma B.6.16) and taking out what is known (Lemma B.6.13),

$$\begin{aligned}
&\mathrm{Cov}[f(X_s), f(X_t)] \\
&= \mathbb{E}\left[(f(X_s) - \mathbb{E}[f(X_s)])(f(X_t) - \mathbb{E}[f(X_t)])\right] \\
&= \sum_x \mathbb{E}\left[(f(X_s) - \mathbb{E}[f(X_s)])(f(X_t) - \mathbb{E}[f(X_t)]) \,|\, X_s = x\right] \mathbb{P}[X_s = x] \\
&= \sum_x \mathbb{E}\left[f(X_t) - \mathbb{E}[f(X_t)] \,|\, X_s = x\right] (f(x) - \mathbb{E}[f(X_s)])\, \mathbb{P}[X_s = x].
\end{aligned}$$

We now use the time homogeneity of the chain to note that

$$\begin{aligned}
&\mathbb{E}\left[f(X_t) - \mathbb{E}[f(X_t)] \,|\, X_s = x\right] \\
&= \mathbb{E}\left[f(X_t) \,|\, X_s = x\right] - \mathbb{E}[f(X_t)] \\
&= \mathbb{E}\left[f(X_{t-s}) \,|\, X_0 = x\right] - \mathbb{E}[f(X_t)].
\end{aligned}$$

By Lemma 5.2.28,

$$\begin{aligned}
&|\mathbb{E}\left[f(X_t) - \mathbb{E}[f(X_t)] \,|\, X_s = x\right]| \\
&= |\mathbb{E}\left[f(X_{t-s}) \,|\, X_0 = x\right] - \mathbb{E}[f(X_t)]| \\
&= |(\mathbb{E}\left[f(X_{t-s}) \,|\, X_0 = x\right] - \pi f) - (\mathbb{E}[f(X_t)] - \pi f)| \\
&\leq |\mathbb{E}\left[f(X_{t-s}) \,|\, X_0 = x\right] - \pi f| + |\mathbb{E}[f(X_t)] - \pi f| \\
&\leq (1 - \gamma_*)^{t-s} \pi_{\min}^{-1} \|f\|_\infty + (1 - \gamma_*)^{t} \pi_{\min}^{-1} \|f\|_\infty \\
&\leq 2(1 - \gamma_*)^{t-s} \pi_{\min}^{-1} \|f\|_\infty,
\end{aligned}$$

which does not depend on x. Plugging back above,

$$\begin{aligned}
&|\mathrm{Cov}[f(X_s), f(X_t)]| \\
&\leq \sum_x |\mathbb{E}\left[f(X_t) - \mathbb{E}[f(X_t)] \,|\, X_s = x\right]| \, |f(x) - \mathbb{E}[f(X_s)]| \, \mathbb{P}[X_s = x] \\
&\leq 2(1-\gamma_*)^{t-s}\pi_{\min}^{-1}\|f\|_\infty \sum_x |f(x) - \mathbb{E}[f(X_s)]| \, \mathbb{P}[X_s = x] \\
&\leq 2(1-\gamma_*)^{t-s}\pi_{\min}^{-1}\|f\|_\infty \sum_x 2\|f\|_\infty \mathbb{P}[X_s = x] \\
&\leq 4(1-\gamma_*)^{t-s}\pi_{\min}^{-1}\|f\|_\infty^2.
\end{aligned}$$

Returning to the sum over the covariances, the previous bound gives

$$\begin{aligned}
&\left|\frac{2}{T^2}\sum_{1\leq s<t\leq T} \mathrm{Cov}[f(X_s), f(X_t)]\right| \\
&\leq \frac{2}{T^2}\sum_{1\leq s<t\leq T} |\mathrm{Cov}[f(X_s), f(X_t)]| \\
&\leq \frac{2}{T^2}\sum_{1\leq s<t\leq T} 4(1-\gamma_\star)^{t-s}\pi_{\min}^{-1}\|f\|_\infty^2.
\end{aligned}$$

To evaluate the sum we make the change of variable $h = t - s$ to get that the previous expression is

$$\begin{aligned}
&\leq 4\pi_{\min}^{-1}\|f\|_\infty^2 \frac{2}{T^2}\sum_{1\leq s\leq T}\sum_{h=1}^{T-s}(1-\gamma_*)^h \\
&\leq 4\pi_{\min}^{-1}\|f\|_\infty^2 \frac{2}{T^2}\sum_{1\leq s\leq T}\sum_{h=0}^{+\infty}(1-\gamma_*)^h \\
&= 4\pi_{\min}^{-1}\|f\|_\infty^2 \frac{2}{T^2}\sum_{1\leq s\leq T}\frac{1}{\gamma_*} \\
&= 8\pi_{\min}^{-1}\|f\|_\infty^2 \gamma_*^{-1}\frac{1}{T} \to 0,
\end{aligned}$$

as $T \to +\infty$.

Combining the variance and covariance bounds, we have shown that

$$\mathrm{Var}\left[\frac{1}{T}\sum_{t=1}^T f(X_t)\right] \leq \|f\|_\infty^2 \frac{1}{T} + 8\pi_{\min}^{-1}\|f\|_\infty^2\gamma_*^{-1}\frac{1}{T} \leq 9\pi_{\min}^{-1}\|f\|_\infty^2\gamma_*^{-1}\frac{1}{T}.$$

For any $\varepsilon > 0$

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{1}{T}\sum_{t=1}^{T} f(X_t) - \pi f\right| \geq \varepsilon\right] \\
&= \mathbb{P}\left[\left|\frac{1}{T}\sum_{t=1}^{T} f(X_t) - \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} f(X_t)\right] + \left(\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} f(X_t)\right] - \pi f\right)\right| \geq \varepsilon\right] \\
&\leq \mathbb{P}\left[\left|\frac{1}{T}\sum_{t=1}^{T} f(X_t) - \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} f(X_t)\right]\right| + \left|\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} f(X_t)\right] - \pi f\right| \geq \varepsilon\right] \\
&\leq \mathbb{P}\left[\left|\frac{1}{T}\sum_{t=1}^{T} f(X_t) - \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} f(X_t)\right]\right| \geq \varepsilon - \pi_{\min}^{-1}\|f\|_\infty \gamma_*^{-1}\frac{1}{T}\right].
\end{aligned}$$

We can now apply Chebyshev's inequality to get

$$\mathbb{P}\left[\left|\frac{1}{T}\sum_{t=1}^{T} f(X_t) - \pi f\right| \geq \varepsilon\right] \leq \frac{9\pi_{\min}^{-1}\|f\|_\infty^2 \gamma_*^{-1}\frac{1}{T}}{(\varepsilon - \pi_{\min}^{-1}\|f\|_\infty \gamma_*^{-1}\frac{1}{T})^2} \to 0,$$

as $T \to +\infty$. ■

5.2.5 Spectral radius

The results in this section have so far concerned finite state spaces. The countably infinite case presents a number of complications. We start with a few observations:

- Suppose P is irreducible, aperiodic and positive recurrent. Then we know from the convergence theorem (Theorem 1.1.33) that, if π is the stationary distribution, then for all x

  $$\|P^t(x, \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \to 0,$$

  as $t \to +\infty$. The convergence rate depends on the starting point x. In the infinite state space case, one typically needs to make that dependence explicit to get meaningful results. In particular the mixing time—as we have defined it—may not be a useful concept.

- In the transient and null recurrent cases, there is no stationary distribution to converge to by Theorem 3.1.20. Instead, we have the following by Theorem 3.1.21: if P is an irreducible chain which is either transient or null recurrent, then we have that

  $$\lim_t P^t(x, y) = 0,$$

  for all $x, y \in V$.

- Conditions stronger than reversibility are needed for the spectral theorem—in a form similar to what we used—to apply. Specifically, one needs that P is a *compact operator*: whenever $(f_n)_n \in \ell^2(V, \pi)$ is a bounded sequence, then there exists a subsequence $(f_{n_k})_k$ such that (Pf_{n_k}) converges in the norm. Unfortunately that is often not the case, as the next example illustrates, even in the reversible, positive recurrent setting. *compact operator*

Example 5.2.29 (A positive recurrent chain whose P is not compact)**.** For $p < 1/2$, let (X_t) be the birth-death chain with $V := \{0, 1, 2, \ldots\}$, $P(0,0) := 1-p$, $P(0,1) = p$, $P(x, x+1) := p$ and $P(x, x-1) := 1-p$ for all $x \geq 1$, and $P(x, y) := 0$ if $|x-y| > 1$. As can be checked by direct computation, P is reversible with respect to the stationary distribution $\pi(x) = (1-\gamma)\gamma^x$ for $x \geq 0$ where $\gamma := \frac{p}{1-p}$. For $j \geq 1$, define $g_j(x) := \pi(j)^{-1/2}\mathbf{1}_{\{x=j\}}$. Then $\|g_j\|_\pi^2 = 1$ for all j so $\{g_j\}_j$ is bounded in $\ell^2(V, \pi)$. On the other hand,

$$Pg_j(x) = p\pi(j)^{-1/2}\mathbf{1}_{\{x=j-1\}} + (1-p)\pi(j)^{-1/2}\mathbf{1}_{\{x=j+1\}}.$$

So

$$\begin{aligned}
\|Pg_j\|_\pi^2 &= p^2\pi(j)^{-1}\pi(j-1) + (1-p)^2\pi(j)^{-1}\pi(j+1) \\
&= p^2\frac{1-p}{p} + (1-p)^2\frac{p}{1-p} \\
&= 2p(1-p).
\end{aligned}$$

Hence $\{Pg_j\}_j$ is also bounded. However, for $j > \ell$

$$\begin{aligned}
\|Pg_j - Pg_\ell\|_\pi^2 &\geq (1-p)^2\pi(j)^{-1}\pi(j+1) + p^2\pi(\ell)^{-1}\pi(\ell-1) \\
&= 2p(1-p).
\end{aligned}$$

So $\{Pg_j\}_j$ does not have a converging subsequence. ◀

We will not say much about the spectral theory of infinite networks. In this subsection, we establish a relationship between the operator norm of P—which is related to its spectrum—and the decay of $P^t(x, y)$.

Let $\ell_0(V)$ be the set of real-valued functions on V with finite support. It is dense in $\ell^2(V, \pi)$. Indeed let $v_1, v_2, \ldots$ be an enumeration of V and, for $f \in \ell^2(V, \pi)$, define $f|_n(v_i) := f(v_i)\mathbf{1}_{i \leq n}$ to be f restricted to $v_1, \ldots, v_n$. Then

$$\|f - f|_n\|_\pi^2 = \sum_{i=n+1}^{\infty} \pi(v_i)f(v_i)^2 \to 0, \tag{5.2.10}$$

as $n \to \infty$, since $\|f\|_\pi^2 = \sum_x \pi(x) f(x)^2 < +\infty$. We will also need the following

$$\|Pf - P(f|_n)\|_\pi^2 = \|P(f - f|_n)\|_\pi^2 \leq \|f - f|_n\|_\pi^2 \to 0, \tag{5.2.11}$$

where we used (5.2.2).

Definition 5.2.30 (Operator norm)**.** *The* operator norm *of P is*

operator norm

$$\|P\|_\pi = \sup \left\{ \frac{\|Pf\|_\pi}{\|f\|_\pi} \,:\, f \in \ell_0(V),\ f \neq 0 \right\}.$$

By definition, for any $f \in \ell_0(V)$,

$$\|Pf\|_\pi \leq \|P\|_\pi \|f\|_\pi. \tag{5.2.12}$$

The same can be seen to hold for any $f \in \ell^2(V, \pi)$ by considering the sequence $(f|_n)_n$ and noting that $\|f|_n\|_\pi \to \|f\|_\pi$ and $\|P(f|_n)\|_\pi \to \|Pf\|_\pi$ as $n \to \infty$ by (5.2.10), (5.2.11) and the triangle inequality. This latter observation explains why it suffices to restrict the supremum to ℓ_0 in the definition of the norm.

Note that, by (5.2.2), $\|P\|_\pi \leq 1$. Note further that, if V is finite or more generally if π is summable, then we have in fact $\|P\|_\pi = 1$ by taking $f \equiv 1$ above. When P is self-adjoint, the norm $\|P\|_\pi$ is also equal to what is known as the *spectral radius*, that is, the radius of the smallest disk centered at 0 in the complex plane that contains the spectrum of P. We will not need to define what that means formally here. (But Exercise 5.5 asks for a proof in the setting of symmetric matrices.)

spectral radius

Our main result is the following.

Theorem 5.2.31 (Spectral radius)**.** *Let P be irreducible and reversible with respect to $\pi > 0$. Then*

$$\rho(P) := \limsup_t P^t(x, y)^{1/t} = \|P\|_\pi.$$

In particular, the limit does not depend on x, y. Moreover, for all t,

$$P^t(x, y) \leq \sqrt{\frac{\pi(y)}{\pi(x)}} \|P\|_\pi^t.$$

In the positive recurrent case (for instance if the chain is finite), we have $P^t(x, y) \to \pi(y) > 0$ and so $\rho(P) = 1 = \|P\|_\pi$. The theorem says that the equality between $\rho(P)$ and $\|P\|_\pi$ holds in general for reversible chains.

Proof of Theorem 5.2.31. To see that the limit does not depend on x, y, let $u, v, x, y \in V$ and $k, m \geq 0$ such that $P^m(u, x) > 0$ and $P^k(y, v) > 0$. Then

$$\begin{aligned} &P^{t+m+k}(u,v)^{1/(t+m+k)} \\ &\quad \geq (P^m(u,x)P^t(x,y)P^k(y,v))^{1/(t+m+k)} \\ &\quad \geq P^m(u,x)^{1/(t+m+k)}P^t(x,y)^{1/t}P^k(y,v)^{1/(t+m+k)}, \end{aligned}$$

which shows that $\limsup_t P^t(u,v)^{1/t} \geq \limsup_t P^t(x,y)^{1/t}$ for all u, v, x, y.

We first show that $\rho(P) \leq \|P\|_\pi$. Observe that applying (5.2.4) and (5.2.12) repeatedly gives that P^t is self-adjoint and satisfies the inequality $\|P^t\|_\pi \leq \|P\|_\pi^t$. Because $\|\delta_z\|_\pi^2 = \pi(z) \leq 1$, by Cauchy-Schwarz

$$\pi(x)P^t(x,y) = \langle \delta_x, P^t\delta_y\rangle_\pi \leq \|P\|_\pi^t \|\delta_x\|_\pi \|\delta_y\|_\pi = \|P\|_\pi^t \sqrt{\pi(x)\pi(y)}.$$

Hence $P^t(x,y) \leq \sqrt{\frac{\pi(y)}{\pi(x)}}\|P\|_\pi^t$ and

$$\begin{aligned} \rho(P) &= \limsup_t P^t(x,y)^{1/t} \\ &\leq \limsup_t \left(\sqrt{\frac{\pi(y)}{\pi(x)}}\|P\|_\pi^t\right)^{1/t} \\ &= \|P\|_\pi. \end{aligned}$$

To establish the inequality in the other direction, we make a series of observations. Fix a nonzero $f \in \ell_0(V)$.

- By self-adjointness and Cauchy-Schwarz,

  $$\|P^{t+1}f\|_\pi^2 = \langle P^{t+1}f, P^{t+1}f\rangle_\pi = \langle P^{t+2}f, P^t f\rangle_\pi \leq \|P^{t+2}f\|_\pi \|P^t f\|_\pi,$$

  or

  $$\frac{\|P^{t+1}f\|_\pi}{\|P^t f\|_\pi} \leq \frac{\|P^{t+2}f\|_\pi}{\|P^{t+1}f\|_\pi}.$$

  So $\frac{\|P^{t+1}f\|_\pi}{\|P^t f\|_\pi}$ is non-decreasing and therefore has a limit $L \leq +\infty$. Moreover, for $t = 0$, we get

  $$\frac{\|Pf\|_\pi}{\|f\|_\pi} \leq L, \tag{5.2.13}$$

  so it suffices to prove $L \leq \rho(P)$.

- Observe that

$$\left(\frac{\|P^t f\|_\pi}{\|f\|_\pi}\right)^{1/t} = \left(\frac{\|Pf\|_\pi}{\|f\|_\pi} \times \cdots \times \frac{\|P^t f\|_\pi}{\|P^{t-1} f\|_\pi}\right)^{1/t} \to L,$$

so in fact

$$L = \lim_t \|P^t f\|_\pi^{1/t}.$$

- By self-adjointness again

$$\|P^t f\|_\pi^2 = \langle f, P^{2t} f\rangle_\pi = \sum_x \pi(x) f(x) \sum_y f(y) P^{2t}(x,y).$$

By definition of $\rho(P)$, for any $\varepsilon > 0$, there is a t large enough that

$$P^{2t}(x,y) \leq (\rho(P) + \varepsilon)^{2t},$$

for all x, y in the support of f. For such a t, plugging back into the previous display

$$\|P^t f\|_\pi^{1/t} \leq (\rho(P) + \varepsilon) \left(\sum_x \pi(x)|f(x)| \sum_y |f(y)|\right)^{1/2t}.$$

The expression in parentheses on the right-hand side is finite because f has finite support. Since ε is arbitrary, we get

$$L = \lim_t \|P^t f\|_\pi^{1/t} \leq \rho(P). \tag{5.2.14}$$

So, combining (5.2.13) and (5.2.14), we have shown that $\|P\|_\pi \leq \rho(P)$ and that concludes the proof. ∎

Corollary 5.2.32. *Let P be irreducible and reversible with respect to π. If $\|P\|_\pi < 1$, then P is transient.*

Proof. By Theorem 5.2.31, $P^t(x,x) \leq \|P\|_\pi^t$ so

$$\sum_t P^t(x,x) \leq \sum_t \|P\|_\pi^t < +\infty.$$

Let (X_t) be a chain with transition matrix P. Because

$$\sum_t P^t(x,x) = \mathbb{E}_x\left[\sum_t \mathbf{1}_{\{X_t = x\}}\right],$$

we have that $\sum_t \mathbf{1}_{\{X_t=x\}} < +\infty$, $\mathbb{P}_x$-a.s., and (X_t) is transient. ∎

This is not an if and only if. Random walk on $\mathbb{Z}^3$ is transient, yet $P^{2t}(0,0) = \Theta(t^{-3/2})$ so there $\|P\|_\pi = \rho(P) = 1$.

In the non-reversible case, our definition of $\|P\|_\pi$ still makes sense with respect to any stationary measure π (although P is not self-adjoint). But the equality in Theorem 5.2.31 no longer holds in general.

Example 5.2.33 (Counter-example)**.** Let (X_t) be asymmetric random walk on $\mathbb{Z}$ with probability $p \in (1/2, 1)$ of going to the right. Then both $\pi_0(x) := \left(\frac{p}{1-p}\right)^x$ and $\pi_1(x) := 1$ define stationary measures, but the transition matrix P is only reversible with respect to π_0.

Under π_1, we have $\|P\|_{\pi_1} = 1$. Indeed, let $g_n(x) := \mathbf{1}_{\{|x|\leq n\}}$ and note that

$$(Pg_n)(x) = \mathbf{1}_{\{|x|\leq n-1\}} + p\mathbf{1}_{\{x\,=\,-n\,-\,1\text{ or }-n\}} + (1-p)\mathbf{1}_{\{x\,=\,n\text{ or }n\,+\,1\}},$$

so $\|g_n\|_{\pi_1}^2 = 2n+1$ and $\|Pg_n\|_{\pi_1}^2 \geq 2(n-1)+1$. Hence

$$\limsup_n \frac{\|Pg_n\|_{\pi_1}}{\|g_n\|_{\pi_1}} \geq 1,$$

and $\|P\|_{\pi_1} \geq 1$. But we already showed that $\|P\|_{\pi_1} \leq 1$ in (5.2.2), so the claim follows.

On the other hand, $\mathbb{E}_0[X_t] = (2p-1)t$. So the martingale $Z_t := X_t - (2p-1)t$ (see Example 3.1.29), as a sum of t independent centered random variables in $\{-1-(2p-1), 1-(2p-1)\}$, satisfies the assumptions of the Azuma-Hoeffding inequality (Theorem 3.2.1) with increment bound $c_t := 2$. So

$$\begin{aligned}
P^t(0,0)^{1/t} &\leq \mathbb{P}_0[X_t \leq 0]^{1/t} \\
&= \mathbb{P}_0[X_t - (2p-1)t \leq -(2p-1)t]^{1/t} \\
&\leq e^{-\frac{2(2p-1)^2t^2}{2^2 t}\frac{1}{t}}.
\end{aligned}$$

Therefore

$$\limsup_t P^t(0,0)^{1/t} \leq e^{-(2p-1)^2/2} < 1.$$

◀

5.3 Geometric bounds

The goal of this section is to relate the spectral gap to certain geometric properties of the underlying network, more specifically isoperimetric properties, that is, relationships between the "volume" of sets and their "circumference." The classical *isoperimetric inequality* states that the area enclosed by any rectifiable simple closed curve in the plane is at most the length of the curve squared divided by 4π. Moreover equality is achieved if and only if the curve is a circle.

isoperimetric inequality

Remark 5.3.1. *Here is an easy proof in the smooth case. Suppose $\boldsymbol{r}(s) = (x(s), y(s))$, $s \in [0, 2\pi]$, is the parametrization of a positively oriented, smooth, simple closed curve in the plane centered at the origin with arc-length 2π, where $\|\boldsymbol{r}'(s)\|_2 = 1$ for all s, $\int_0^{2\pi} \boldsymbol{r}(s)\,\mathrm{d}s = \mathbf{0}$ and $x(0) = x(2\pi) = 0$. By Green's theorem, the area enclosed by the curve is*

$$A = \int_0^{2\pi} x(s)y'(s)\,\mathrm{d}s = \frac{1}{2}\int_0^{2\pi} [x(s)^2 + y'(s)^2 - (x(s) - y'(s))^2]\,\mathrm{d}s,$$

where we used that $2ab = a^2 + b^2 - (a-b)^2$. By the one-dimensional Poincaré inequality (Remark 3.2.7),

$$A \leq \frac{1}{2}\int_0^{2\pi} [x(s)^2 + y'(s)^2]\,\mathrm{d}s \leq \frac{1}{2}\int_0^{2\pi} [x'(s)^2 + y'(s)^2]\,\mathrm{d}s = \pi,$$

which is indeed the area of a circle of circumference 2π.

Edge expansion We define our isoperimetric quantity of interest. Let (X_t) be a finite, irreducible Markov chain on V reversible with respect to a stationary measure $\pi > 0$. (In this section, we do not necessarily assume that π is a probability distribution.) Let P be its transition matrix. We think of (X_t) as a random walk on the network $\mathcal{N} = (G, c)$ where G is the transition graph and $c(x, y) := \pi(x)P(x, y) = \pi(y)P(y, x)$.

For a subset $S \subseteq V$, we let the *edge boundary* of S be

edge boundary

$$\partial_{\mathrm{E}} S := \{e = \{x, y\} \in E : x \in S, y \in S^c\}.$$

Let $g : E \to \mathbb{R}_+$ be an edge weight function. For $F \subseteq E$ and $W \subseteq V$, we define

$$|F|_g := \sum_{e \in F} g(e).$$

and

$$|W|_h := \sum_{v \in W} h(v).$$

Finally, for $S \subseteq V$, we let

$$\Phi_{\mathrm{E}}(S; g, h) := \frac{|\partial_{\mathrm{E}} S|_g}{|S|_h}.$$

Roughly speaking, this is the ratio of the "size of the boundary" of a set to its "volume."

Our main definition, the edge expansion constant, quantifies the worst such ratio. First, one last piece of notation: for disjoint subsets $S_0, S_1 \subseteq V$, we let

$$c(S_0, S_1) := \sum_{x_0 \in S_0} \sum_{x_1 \in S_1} c(x_0, x_1).$$

Definition 5.3.2 (Edge expansion)**.** *For a subset of states* $S \subseteq V$*, the* edge expansion constant *(or* bottleneck ratio*) of* S *is*

edge expansion constant

$$\Phi_{\mathrm{E}}(S; c, \pi) = \frac{|\partial_{\mathrm{E}} S|_c}{|S|_\pi} = \frac{c(S, S^c)}{\pi(S)}.$$

We refer to (S, S^c) *as a* cut*. The* edge expansion constant *(or* bottleneck ratio *or* Cheeger number *or* isoperimetric constant[*]*) of* $\mathcal{N}$ *is*

$$\Phi_* := \min\left\{\Phi_{\mathrm{E}}(S; c, \pi)\ :\ S \subseteq V,\ 0 < \pi(S) \leq \frac{1}{2}\right\}.$$

Intuitively, a small value of Φ_* suggests the existence of a "bottleneck" in $\mathcal{N}$. Conversely, a large value of Φ_* indicates that all sets "expand out." See Figure 5.3. Note that the quantity $\Phi_{\mathrm{E}}(S; c, \pi)$ has a natural probabilistic interpretation: pick a stationary state and make one step according to the transition matrix; then $\Phi_{\mathrm{E}}(S; c, \pi)$ is the conditional probability that, given that the first state is in S, the next one is in S^c.

Equivalently, the edge expansion constant can be expressed as

$$\Phi_* := \min\left\{\frac{c(S, S^c)}{\pi(S) \wedge \pi(S^c)}\ :\ S \subseteq V,\ 0 < \pi(S) < 1\right\}.$$

Example 5.3.3 (Edge expansion: complete graph)**.** Let $G = K_n$ be the complete graph on n vertices. Let $c(x, y) = 1/n^2$ for all x, y (corresponding to picking any vertex uniformly at random at the next step) and $\pi(x) = 1/n$ for all x. For simplicity, take n even. Then for a subset S of size $|S| = k$,

$$\Phi_{\mathrm{E}}(S; c, \pi) = \frac{|\partial_{\mathrm{E}} S|_c}{|S|_\pi} = \frac{k(n-k)/n^2}{k/n} = \frac{n-k}{n}.$$

Thus, the minimum is achieved for $k = n/2$ and

$$\Phi_* = \frac{n - n/2}{n} = \frac{1}{2}.$$

◀

[*]It is also called "conductance," but that terminology clashes with our use of the term.

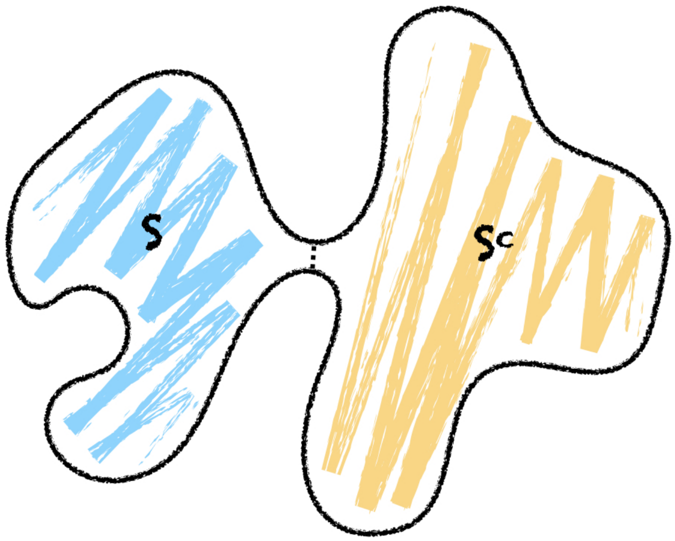

Figure 5.3: A bottleneck.

Dirichlet form, Rayleigh quotient, and normalized Laplacian We relate the edge expansion constant of $\mathcal{N}$ to the spectral gap of P. Recall that we denote by $\lambda_1, \ldots, \lambda_n$ the eigenvalues of P in decreasing order. First, we adapt the variational characterization of Theorem 5.1.3 to the network setting.

The *Dirichlet form* is defined over $\ell^2(V, \pi)$ as the bilinear form

Dirichlet form

$$\mathscr{D}(f, g) := \langle f, (I - P)g \rangle_\pi.$$

The associated quadratic form, also known as *Dirichlet energy*, is $\mathscr{D}(f) := \mathscr{D}(f, f)$.

Dirichlet energy

Note that, using stochasticity and reversibility,

$$
\begin{aligned}
\langle f,(I-P)f\rangle_\pi &= \langle f,f\rangle_\pi - \langle f,Pf\rangle_\pi \\
&= \frac{1}{2}\sum_x \pi(x)f(x)^2 + \frac{1}{2}\sum_y \pi(y)f(y)^2 - \sum_{x,y}\pi(x)f(x)f(y)P(x,y) \\
&= \frac{1}{2}\sum_x f(x)^2\pi(x)\sum_y P(x,y) \\
&\qquad + \frac{1}{2}\sum_y f(y)^2\pi(y)\sum_x P(y,x) - \sum_{x,y}\pi(x)f(x)f(y)P(x,y) \\
&= \frac{1}{2}\sum_{x,y}\pi(x)P(x,y)f(x)^2 \\
&\qquad + \frac{1}{2}\sum_{x,y}\pi(x)P(x,y)f(y)^2 - \sum_{x,y}\pi(x)P(x,y)f(x)f(y) \\
&= \frac{1}{2}\sum_{x,y} c(x,y)[f(x)-f(y)]^2,
\end{aligned}
$$

which is indeed consistent with the expression we encountered previously in Theorem 3.3.25.

The *Rayleigh quotient* for $I-P$ over $\ell^2(V,\pi)$ is then

Rayleigh quotient

$$
\begin{aligned}
\frac{\langle f,(I-P)f\rangle_\pi}{\langle f,f\rangle_\pi} &= \frac{\frac{1}{2}\sum_{x,y}c(x,y)[f(x)-f(y)]^2}{\sum_x \pi(x)f(x)^2} \\
&= \frac{\mathbf{z}^T\mathcal{L}\mathbf{z}}{\mathbf{z}^T\mathbf{z}},
\end{aligned}
$$

where $\mathcal{L}$ is the normalized Laplacian of the network $\mathcal{N}$ and we defined the vector $\boldsymbol{z} = (z_x)_{x\in V}$ with $z_x := \sqrt{\pi(x)}f(x)$. Consequently, the Courant-Fischer theorem (Theorem 5.1.3) in the form (5.1.5) gives the following. Here $\eta_2 = 1-\lambda_2 = \gamma$ is the spectral gap of P, which can also be seen as the second smallest eigenvalue of $I-P$ (which has the same eigenfunctions as P itself). We have

$$
\gamma = \inf\left\{\frac{\langle f,(I-P)f\rangle_\pi}{\langle f,f\rangle_\pi}\ :\ \pi f = 0,\ f\neq 0\right\}.
$$

The infimum is achieved by the eigenfunction f_2 of P corresponding to its second largest eigenvalue λ_2. (Recall from Lemma 5.2.5 that $\pi f_2 = 0$.)

We note further that if $\pi f = 0$ then

$$
\langle f,f\rangle_\pi = \langle f-\pi f, f-\pi f\rangle_\pi = \mathrm{Var}_\pi[f],
$$

where the last expression denotes the variance under π. So the variational characterization of γ implies that

$$\mathrm{Var}_\pi[f] \leq \gamma^{-1}\mathscr{D}(f),$$

for all f such that $\pi f = 0$. In fact, it holds for *any* f by considering $f - \pi f$ and noticing that both sides are unaffected by subtracting a constant to f.

We have shown:

Theorem 5.3.4 (Poincaré inequality for $\mathcal{N}$)**.** *Let P be finite, irreducible and reversible with respect to π. Then*

$$\mathrm{Var}_\pi[f] \leq \gamma^{-1}\mathscr{D}(f), \tag{5.3.1}$$

for all $f \in \ell^2(V\,\pi)$. Equality is achieved by the eigenfunction f_2 of P corresponding to the second largest eigenvalue λ_2.

An inequality of the type

$$\mathrm{Var}_\pi[f] \leq C\mathscr{D}(f), \qquad \forall f, \tag{5.3.2}$$

is known as a *Poincaré inequality*, a simple version of which we encountered previously in Remark 3.2.7. To see the connection with that one-dimensional version, it will be convenient to work with directed edges. Let $\vec{E}$ be an orientation of E, that is, for each $e \in \{x, y\}$, $\vec{E}$ includes either (x, y) or (y, x) with associated weight $c(\vec{e}) := c(e) > 0$ For a function $f : V \to \mathbb{R}$ and an edge $\vec{e} = (x, y) \in \vec{E}$, we define the "discrete gradient"

Poincaré inequality

$$\nabla f(\vec{e}) = f(y) - f(x).$$

With this notation, we can rewrite the Dirichlet energy as

$$\mathscr{D}(f) = \frac{1}{2}\sum_{x,y} c(x,y)[f(x) - f(y)]^2 = \sum_{\vec{e}} c(\vec{e})[\nabla f(\vec{e})]^2, \tag{5.3.3}$$

hence (5.3.1) is a network analogue of (3.2.7).

5.3.1 Cheeger's inequality

The edge expansion constant and the spectral gap are related through the following isoperimetric inequalities. The lower bound is known as *Cheeger's inequality*.

Cheeger's inequality

Theorem 5.3.5. *Let P be a finite, irreducible, reversible Markov transition matrix and let $\gamma = 1 - \lambda_2$ be the spectral gap of P. Then*

$$\frac{\Phi_*^2}{2} \leq \gamma \leq 2\Phi_*.$$

In terms of the relaxation time $\mathrm{t_{rel}} = \gamma^{-1}$, these inequalities have an intuitive meaning: the presence or absence of a bottleneck in the state space leads to slow or fast mixing respectively. We detail some applications to mixing times in the next subsections.

Before giving a proof of the theorem, we start with a trivial—yet insightful—example.

Example 5.3.6 (Two-state chain)**.** Let $V := \{0, 1\}$ and, for $\alpha, \beta \in (0, 1)$,

$$P := \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

which has stationary distribution

$$\pi := \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right).$$

Recall from Example 5.2.8 that the second right eigenvector is

$$f_2 := \left(\sqrt{\frac{\alpha}{\beta}}, -\sqrt{\frac{\beta}{\alpha}}\right) = \left(\sqrt{\frac{\pi_1}{\pi_0}}, -\sqrt{\frac{\pi_0}{\pi_1}}\right),$$

with eigenvalue $\lambda_2 := 1 - \alpha - \beta$, so the spectral gap is $\alpha + \beta$. Assume that $\beta \leq \alpha$. Then the bottleneck ratio is

$$\Phi_* = \frac{c(0,1)}{\pi(0)} = P(0,1) = \alpha.$$

Then Theorem 5.3.5 reads

$$\frac{\alpha^2}{2} \leq \alpha + \beta \leq 2\alpha,$$

which is indeed satisfied for all $0 < \beta \leq \alpha < 1$. Note that the upper bound is tight when $\alpha = \beta$. ◀

Proof of Theorem 5.3.5. We start with the upper bound. In view of the Poincaré inequality for $\mathcal{N}$ (Theorem 5.3.4), to get an upper bound on the spectral gap, it suffices to plug in a well-chosen function f in (5.3.1). Taking a hint from Example 5.3.6, for $S \subseteq V$ with $\pi(S) \in (0, 1/2]$, we let

$$f_S(x) := \begin{cases} -\sqrt{\frac{\pi(S^c)}{\pi(S)}} & x \in S, \\ \sqrt{\frac{\pi(S)}{\pi(S^c)}} & x \in S^c. \end{cases}$$

Then

$$\sum_x \pi(x) f_S(x) = \pi(S)\left[-\sqrt{\frac{\pi(S^c)}{\pi(S)}}\right] + \pi(S^c)\left[\sqrt{\frac{\pi(S)}{\pi(S^c)}}\right] = 0,$$

and

$$\sum_x \pi(x) f_S(x)^2 = \pi(S)\left[-\sqrt{\frac{\pi(S^c)}{\pi(S)}}\right]^2 + \pi(S^c)\left[\sqrt{\frac{\pi(S)}{\pi(S^c)}}\right]^2 = 1.$$

So $\mathrm{Var}_\pi[f_S] = 1$. Hence, from Theorem 5.3.4,

$$\begin{aligned}
\gamma &\leq \frac{\mathscr{D}(f_S)}{\mathrm{Var}_\pi[f_S]} \\
&= \frac{1}{2}\sum_{x,y} c(x,y)[f_S(x) - f_S(y)]^2 \\
&= \sum_{x \in S, y \in S^c} c(x,y)\left[-\sqrt{\frac{\pi(S^c)}{\pi(S)}} - \sqrt{\frac{\pi(S)}{\pi(S^c)}}\right]^2 \\
&= \sum_{x \in S, y \in S^c} c(x,y)\left[-\frac{\pi(S^c) + \pi(S)}{\sqrt{\pi(S)\pi(S^c)}}\right]^2 \\
&= \frac{c(S, S^c)}{\pi(S)\pi(S^c)} \\
&\leq 2\frac{c(S, S^c)}{\pi(S)},
\end{aligned}$$

as claimed.

The other direction is trickier. Because we seek an upper bound on the edge expansion constant Φ_*, our goal is to find a cut (S, S^c) such that

$$\frac{c(S, S^c)}{\pi(S) \wedge \pi(S^c)} \leq \sqrt{2\gamma}. \tag{5.3.4}$$

Because the eigenfunction f_2 achieves γ in Theorem 5.3.4, it is natural to look to it for "good cuts." Thinking of f_2 as a one-dimensional embedding of the network, it turns out to be enough to consider only "sweep cuts" of the form $S := \{v : f_2(v) \leq \theta\}$ for a threshold θ. How to pick the right threshold is less obvious.

Here we use a probabilistic argument, that is, we construct a *random* cut (Z, Z^c). Observe that it suffices that

$$\mathbb{E}\left[c(Z, Z^c)\right] \leq \sqrt{2\gamma}\mathbb{E}\left[\pi(Z) \wedge \pi(Z^c)\right], \tag{5.3.5}$$

since then $\mathbb{E}\left[\sqrt{2\gamma}\pi(Z) \wedge \pi(Z^c) - c(Z, Z^c)\right] \geq 0$, which in turn implies that we have $\mathbb{P}\left[\sqrt{2\gamma}\pi(Z) \wedge \pi(Z^c) - c(Z, Z^c) \geq 0\right] > 0$ by the first moment principle (Theorem 2.2.1); in other words, there exists a cut satisfying (5.3.4).

We now describe the random cut (Z, Z^c):

1. *(Cuts from f_2)* Let again f_2 be an eigenfunction corresponding to the eigenvalue λ_2 of P with $\|f_2\|_\pi^2 = 1$. Order the vertices $V := \{v_1, \ldots, v_n\}$ in such a way that
$$f_2(v_i) \leq f_2(v_{i+1}), \qquad \forall i = 1, \ldots, n-1.$$
As we described above, the function f_2 naturally produces a series of cuts (S_i, S_i^c) where $S_i := \{v_1, \ldots, v_i\}$. By definition of the bottleneck ratio,
$$\Phi_* \leq \frac{c(S_i, S_i^c)}{\pi(S_i) \wedge \pi(S_i^c)}. \tag{5.3.6}$$

2. *(Normalization)* Let
$$m := \min\{i : \pi(S_i) > 1/2\},$$
and define the translated function
$$f := f_2 - f_2(v_m).$$
We further set $g := \alpha f$ where $\alpha > 0$ is chosen so that
$$g(v_1)^2 + g(v_n)^2 = 1.$$
Note that, by construction, $g(v_m) = 0$ and $g(v_1) \leq \cdots \leq g(v_m) = 0 \leq g(v_{m+1}) \leq \cdots \leq g(v_n)$. The function g is related to γ as follows:

Lemma 5.3.7.
$$\frac{1}{2}\sum_{x,y} c(x,y)(g(x) - g(y))^2 \leq \gamma \sum_x \pi(x) g(x)^2.$$

Proof. By Theorem 5.3.4,

$$\gamma = \frac{\mathscr{D}(f_2)}{\mathrm{Var}_\pi[f_2]}.$$

Because neither the numerator nor the denominator is affected by adding a constant, we have also

$$\gamma = \frac{\mathscr{D}(f)}{\mathrm{Var}_\pi[f]}.$$

Furthermore, notice that a constant multiplying f cancels out in the ratio so

$$\gamma = \frac{\mathscr{D}(g)}{\mathrm{Var}_\pi[g]}.$$

Now use the fact that $\mathrm{Var}_\pi[g] \leq \sum_x \pi(x) g(x)^2$. ■

3. *(Random cut)* Pick Θ in $[g(v_1), g(v_n)]$ with density $2|\theta|$. Note that

$$\int_{g(v_1)}^{g(v_n)} 2|\theta| \,\mathrm{d}\theta = g(v_1)^2 + g(v_n)^2 = 1.$$

Finally define

$$Z := \{v_i \,:\, g(v_i) < \Theta\}.$$

The rest of the proof is calculations. We bound the expectations on both sides of (5.3.5).

Lemma 5.3.8. *The following hold:*

(i)

$$\mathbb{E}[\pi(Z) \wedge \pi(Z^c)] = \sum_x \pi(x) g(x)^2.$$

(ii)

$$\mathbb{E}[c(Z, Z^c)] \leq \left(\frac{1}{2}\sum_{x,y} c(x,y)(g(x) - g(y))^2\right)^{1/2} \left(2\sum_x \pi(x) g(x)^2\right)^{1/2}.$$

Lemmas 5.3.7 and 5.3.8 immediately imply (5.3.5) and that concludes the proof of Theorem 5.3.5. So it remains to prove this last lemma.

Proof of Lemma 5.3.8. We start with (i). By definition of g, $\Theta \leq 0$ implies that $\pi(Z) \wedge \pi(Z^c) = \pi(Z)$ and vice versa. Thus

$$\begin{aligned}
\mathbb{E}[\pi(Z) \wedge \pi(Z^c)] &= \mathbb{E}\left[\sum_{\ell<m} \pi(v_\ell)\mathbf{1}_{\{v_\ell \in Z\}}\mathbf{1}_{\{\Theta \leq 0\}} + \sum_{\ell \geq m} \pi(v_\ell)\mathbf{1}_{\{v_\ell \in Z^c\}}\mathbf{1}_{\{\Theta > 0\}}\right] \\
&= \mathbb{E}\left[\sum_{\ell<m} \pi(v_\ell)\mathbf{1}_{\{g(v_\ell) < \Theta \leq 0\}} + \sum_{\ell \geq m} \pi(v_\ell)\mathbf{1}_{\{0 < \Theta \leq g(v_\ell)\}}\right] \\
&= \sum_{\ell<m} \pi(v_\ell)\mathbb{P}\left[g(v_\ell) < \Theta \leq 0\right] + \sum_{\ell \geq m} \pi(v_\ell)\mathbb{P}\left[0 < \Theta \leq g(v_\ell)\right] \\
&= \sum_{\ell<m} \pi(v_\ell)g(v_\ell)^2 + \sum_{\ell \geq m} \pi(v_\ell)g(v_\ell)^2 \\
&= \sum_x \pi(x)g(x)^2,
\end{aligned} \tag{5.3.7}$$

where we integrated over the density of Θ to obtain the fourth line.

We move on to (ii). To compute $\mathbb{E}[c(Z, Z^c)]$, we note that $x_k \in Z$ and $x_\ell \in Z^c$ if and only if $g(v_k) < \Theta \leq g(v_\ell)$. The probability of that event depends on the signs of $g(v_k)$ and $g(v_\ell)$. If $g(v_k)g(v_\ell) \geq 0$,

$$\begin{aligned}
\mathbb{P}[g(v_k) < \Theta \leq g(v_\ell)] &= |g(v_k)^2 - g(v_\ell)^2| \\
&= |g(v_k) - g(v_\ell)||g(v_k) + g(v_\ell)| \\
&= |g(v_k) - g(v_\ell)|(|g(v_k)| + |g(v_\ell)|).
\end{aligned}$$

If $g(v_k)g(v_\ell) < 0$,

$$\begin{aligned}
\mathbb{P}[g(v_k) < \Theta \leq g(v_\ell)] &= g(v_k)^2 + g(v_\ell)^2 \\
&\leq g(v_k)^2 + g(v_\ell)^2 - 2g(v_k)g(v_\ell) \\
&= (g(v_k) - g(v_\ell))^2 \\
&= |g(v_k) - g(v_\ell)|(|g(v_k)| + |g(v_\ell)|).
\end{aligned}$$

We apply Cauchy-Schwarz to get

$$\begin{aligned}\mathbb{E}[c(Z,Z^c)] &= \sum_{k<\ell} c(v_k,v_\ell)\mathbb{P}[g(v_k) < \Theta \le g(v_\ell)]\\ &\le \sum_{k<\ell} c(v_k,v_\ell)|g(v_k)-g(v_\ell)|(|g(v_k)|+|g(v_\ell)|)\\ &\le \left(\sum_{k<\ell} c(v_k,v_\ell)(g(v_k)-g(v_\ell))^2\right)^{1/2}\\ &\qquad \times \left(\sum_{k<\ell} c(v_k,v_\ell)(|g(v_k)|+|g(v_\ell)|)^2\right)^{1/2}.\end{aligned}$$

The expression in the first parentheses is equal to $\frac{1}{2}\sum_{x,y} c(x,y)(g(x)-g(y))^2$. So it remain to bound the expression in the second parentheses.

Note that

$$(|g(x)|+|g(y)|)^2 = 2g(x)^2 + 2g(y)^2 - (|g(x)|-|g(y)|)^2 \le 2g(x)^2 + 2g(y)^2.$$

Therefore, since $\sum_y c(x,y) = \sum_y c(y,x) = \pi(x)$,

$$\begin{aligned}\sum_{k<\ell} c(v_k,v_\ell)(|g(v_k)|+|g(v_\ell)|)^2 &\le \frac{1}{2}\sum_{x,y} c(x,y)(|g(x)|+|g(y)|)^2\\ &\le \sum_x \pi(x)g(x)^2 + \sum_y \pi(y)g(y)^2\\ &= 2\sum_x \pi(x)g(x)^2.\end{aligned}$$

That concludes the proof. ■

■

5.3.2 ▷ *Random walks: trees, cycles, and hypercubes revisited*

We use the techniques of the previous subsection to bound the mixing time of random walk on some simple graphs. In particular we revisit the examples of Section 4.3.2.

b-ary tree Let (Z_t) be lazy simple random walk on the ℓ-level rooted b-ary tree, $\widehat{\mathbb{T}}_b^\ell$. The root, 0, is on level 0 and the leaves, L, are on level ℓ. All vertices have degree $b+1$, except for the root which has degree b and the leaves which have degree 1. Recall that the stationary distribution is

$$\pi(x) := \frac{\delta(x)}{2(n-1)}, \tag{5.3.8}$$

where n is the number of vertices and $\delta(x)$ is the degree of x. We take $b = 2$ to simplify.

It is intuitively clear that each edge of this graph constitutes a bottleneck, with the root being the most "balanced" one. Let x_0 be a leaf of $\widehat{\mathbb{T}}_b^\ell$ and let A be the set of vertices "on the other side of the root (inclusively)," that is, vertices whose graph distance from x_0 is at least ℓ. See Figure 4.7. Let S be the remaining vertices. Then by symmetry $\pi(S) \leq 1/2$. Note that there is a single edge connecting S and $S^c = A$, namely, the edge linking 0 and the root of the subtree T_S formed by the vertices in S. More precisely, let v_S be the root of T_S. From (5.3.8), $P(v_S, 0) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$ (where the $1/2$ accounts for the laziness), $\pi(v_S) = \frac{3}{2n-2}$, and, by symmetry,

$$\pi(S) = \frac{(2n-2-2)/2}{2n-2} = \frac{n-2}{2n-2},$$

where in the numerator we subtracted the degree of the root before dividing the sum of the remaining degrees by 2. Hence

$$\Phi_* \leq \frac{\frac{1}{6}\left(\frac{3}{2n-2}\right)}{\frac{n-2}{2n-2}} = \frac{1}{2(n-2)},$$

By Theorem 5.3.5,

$$\gamma \leq 2\Phi_* \leq \frac{1}{n-2} \qquad \text{and} \qquad \mathrm{t_{rel}} = \gamma^{-1} \geq n-2.$$

Thus by Theorem 5.2.14 and the fact that the chain is lazy

$$\mathrm{t_{mix}}(\varepsilon) \geq (\mathrm{t_{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right) = \Omega(n).$$

We showed in Section 4.3.2, using other techniques, that $\mathrm{t_{mix}}(\varepsilon) = \Theta(n)$.

Cycle Let (Z_t) be lazy simple random walk on the cycle of size n, $\mathbb{Z}_n := \{0, 1 \ldots, n-1\}$, where $i \sim j$ if $|j - i| = 1 \pmod{n}$. Assume n is even.

Consider a subset of vertices S. Note that by symmetry $\pi(S) = \frac{|S|}{n}$. Moreover, for all $i \sim j$, $c(i,j) = \pi(i)P(i,j) = \frac{1}{n} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4n}$. Among all sets of size $|S|$, consecutive vertices minimize the size of the boundary. So

$$\Phi_* \leq \frac{2\frac{1}{4n}}{\frac{\ell}{n}} = \frac{1}{2\ell},$$

for all $\ell \leq n/2$. This expression is minimized for $\ell = n/2$ so

$$\Phi_* = \frac{1}{n}.$$

By Theorem 5.3.5,

$$\frac{1}{2n^2} = \frac{\Phi_*^2}{2} \leq \gamma \leq 2\Phi_* = \frac{2}{n}$$

and

$$\frac{n}{2} \leq \mathrm{t}_{\mathrm{rel}} = \gamma^{-1} \leq 2n^2.$$

Thus by Theorem 5.2.14

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \geq (\mathrm{t}_{\mathrm{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right) = \Omega(n),$$

and

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \log\left(\frac{1}{\varepsilon \pi_{\min}}\right) \mathrm{t}_{\mathrm{rel}} = O(n^2 \log n).$$

We know from exact eigenvalue computations (see Section 5.2.2 where technically we considered the non-lazy chain; laziness only affects the relaxation time by a factor of 2) that in fact $\gamma = \frac{2\pi^2}{n^2} + O(n^{-4})$. We also showed in that section that $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = O(n^2)$. (Exercise 4.15 shows this is tight up to a constant factor.)

Hypercube Let (Z_t) be lazy simple random walk on the n-dimensional hypercube $\mathbb{Z}_2^n := \{0,1\}^n$ where $i \sim j$ if $\|i - j\|_1 = 1$.

To get a bound on the edge expansion constant, consider the set $S = \{x \in \mathbb{Z}_2^n : x_1 = 0\}$. By symmetry $\pi(S) = \frac{1}{2}$. For each $i \sim j$, $c(i,j) = \frac{1}{2^n} \cdot \frac{1}{2} \cdot \frac{1}{n} = \frac{1}{n2^{n+1}}$. Hence

$$\Phi_* \leq \frac{2^{n-1}\frac{1}{n2^{n+1}}}{\frac{1}{2}} = \frac{1}{2n},$$

where in the numerator we used that $|S| = 2^{n-1}$. By Theorem 5.3.5,

$$\gamma \leq 2\Phi_* \leq \frac{1}{n}.$$

Thus by Theorem 5.2.14

$$\mathrm{t_{mix}}(\varepsilon) \geq (\mathrm{t_{rel}} - 1) \log\left(\frac{1}{2\varepsilon}\right) = \Omega(n).$$

We know from exact eigenvalue computations (Section 5.2.2) that in fact $\gamma = \frac{1}{n}$.

We also showed in Section 4.3.2 that $\mathrm{t_{mix}}(\varepsilon) = \Theta(n \log n)$.

5.3.3 ▷ *Random graphs: existence of an expander family and application to mixing*

In many applications, it is useful to construct "bottleneck-free" graphs. In particular, random walks mix rapidly on such graphs. Formally:

Definition 5.3.9 (Expander family). *Let $\{G_n\}_n$ be a collection of finite d-regular graphs with $\lim_n |V(G_n)| = +\infty$, where $V(G_n)$ is the vertex set of G_n. Let*

$$\Phi_*(G_n) := \min\left\{\frac{|\partial_{\mathrm{E}} S|}{d|S|}\ :\ S \subseteq V(G_n),\ 0 < |S| \leq \frac{|V(G_n)|}{2}\right\}$$

denote the edge expansion constant of G_n with unit conductances.[†] *Let $\alpha > 0$. We say that $\{G_n\}_n$ is a (d, α)*-expander family *if for all n*

$$\Phi_*(G_n) \geq \alpha.$$

The key point of the definition is that the edge expansion constant of all graphs in an expander family is bounded away from 0 *uniformly in n*. Note that it is trivial to construct such a family *if we drop the bounded degree assumption*: the edge expansion constant of the complete graph K_n is $1/2$ by Example 5.3.3. On the other hand, it is far from obvious that one can construct a family of *sparse* graphs (i.e., such that $|E(G_n)| = O(|V(G_n)|)$) with an edge expansion constant uniformly bounded away from 0. It turns out that a simple probabilistic construction does the trick.

We will need the following definition. For a subset $S \subseteq V$, we let the *vertex boundary* of S be

vertex boundary

$$\partial_{\mathrm{V}} S := \{y \in S^c\ :\ \exists x \in S \text{ s.t. } x \sim y\}.$$

[†]In terms of random walk, this corresponds to choosing a neighbor uniformly at random and taking the stationary measure equal to the degree. Note that scaling the stationary measure does not affect the edge expansion constant.

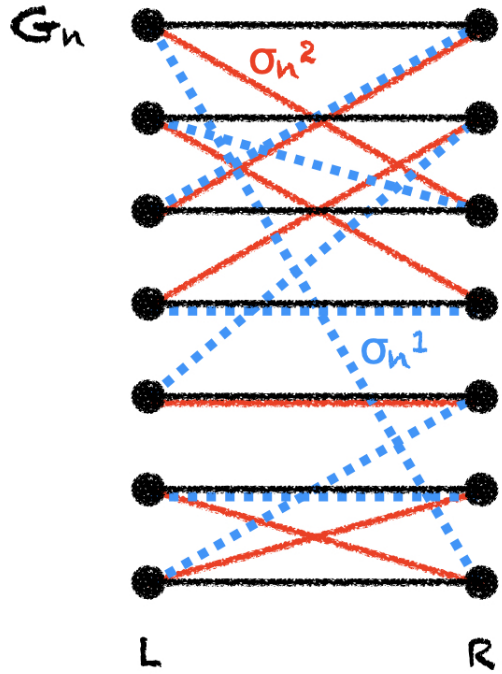

Figure 5.4: A draw from Pinsker's model.

Existence of expander graphs For simplicity, we allow multigraphs (i.e., E is a multiset; or, put differently, there can be multiple edges between the same two vertices) and consider the case $d = 3$. We construct a random bipartite multigraph $G_n = (L_n, R_n, E_n)$ on $2n$ vertices known as *Pinsker's model*. Denote the vertices by $L_n = \{\ell_1, \ldots, \ell_n\}$ and $R_n = \{r_1, \ldots, r_n\}$. Let σ_n^1 and σ_n^2 be independent uniform random permutations of $[n]$. The edge set of G_n is given by

$$E_n := \{(\ell_i, r_i) : i \in [n]\} \cup \left\{(\ell_i, r_{\sigma_n^1(i)}) : i \in [n]\right\} \cup \left\{(\ell_i, r_{\sigma_n^2(i)}) : i \in [n]\right\}.$$

In words, G_n is a union of three independent uniform perfect matchings (and its vertices are labeled so that one of the matchings is $\{(\ell_i, r_i)\}_i$). See Figure 5.4. Observe that, as a multigraph, all vertices of G_n have degree 3. We show that there exists $\alpha > 0$ such that, for all n large enough, with positive (in fact, high) probability G_n has an edge expansion constant bounded below by α. In particular, such a G_n exists for all n large enough and, thus, there exists a $(3, \alpha)$-expander family.

Claim 5.3.10 (Pinsker's model: edge expansion constant)**.** *There exists $\alpha > 0$ such that*

$$\lim_n \mathbb{P}[\Phi_*(G_n) \geq \alpha] = 1.$$

Proof. For convenience, assume n is even. We need to show that with probability going to 1, for any S with $|S| \leq |V(G_n)|/2 = n$, we have $|\partial_{\mathrm{E}} S| \geq \alpha d|S|$ for some $\alpha > 0$. We first reduce the proof to a statement about sets of vertices lying *on one side* of G_n.

Lemma 5.3.11. *There is $\beta > 0$ such that*

$$\lim_n \mathbb{P}\left[|\partial_{\mathrm{V}} K| \geq (1+\beta)|K|,\ \forall K \subseteq L,\ |K| \leq n/2\right] = 1.$$

The same holds for R.

Before proving Lemma 5.3.11, we argue that it implies Claim 5.3.10. Note that the lemma concerns the *vertex* boundary of K. To relate the latter to the edge boundary, let S with $|S| \leq n$, and let $S_L := S \cap L$ and $S_R := S \cap R$. For any subset $K \subseteq S_L$, the size of the edge boundary of S can be bounded below as follows

$$|\partial_{\mathrm{E}} S| \geq |\partial_{\mathrm{V}} K| - |S_R|, \tag{5.3.9}$$

where we took into account that the vertices of $\partial_{\mathrm{V}} K$ in S_R do not contribute to the edge boundary, but the others do as they are incident with at least one edge in $\partial_{\mathrm{E}} S$. It remains to find a good K.

Assume without loss of generality that $|S_L| \geq |S_R|$ (in the other case, just interchange the roles of L and R), and suppose that the event in the lemma holds. In particular, $|S_R| \leq |S|/2$. We claim that there is a subset K of S_L such that

$$|S_R| \leq |S|/2 \leq |K| \leq n/2. \tag{5.3.10}$$

There are two cases:

- If $|S_L| < n/2$, then take $K = S_L$. It follows that $|K| = |S_L| \geq |S|/2$.
- If $|S_L| \geq n/2$, then let K be any subset of S_L of size $n/2$. Since $|S| \leq n$, it follows that $|K| = n/2 \geq |S|/2$.

Under the event in the lemma, $|\partial_{\mathrm{V}} K| \geq (1+\beta)|K|$.

Going back to (5.3.9), using the lower bound on $|\partial_{\mathrm{V}} K|$ and (5.3.10), we get

$$|\partial_{\mathrm{E}} S| \geq (1+\beta)|K| - |K| = \beta|K| \geq \frac{\beta}{2}|S| = \alpha|S|,$$

where we set $\alpha = \beta/2$. Since this holds for any set S with $|S| \leq n$, we have proved Claim 5.3.10.

It remains to prove the lemma.

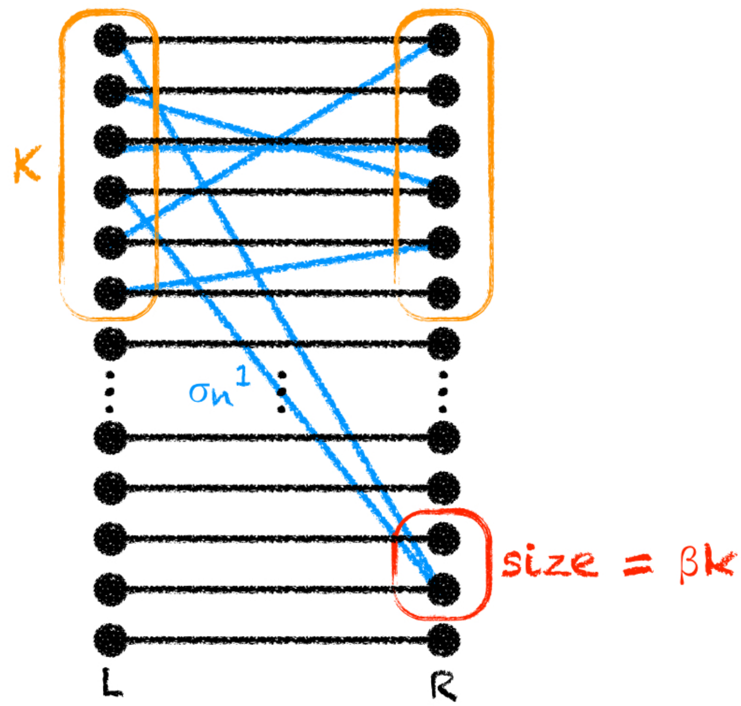

Figure 5.5: Illustration of the main step in proof of the lemma.

Proof of Lemma 5.3.11. Let $K \subseteq L$ with $k := |K| \leq n/2$. Without loss of generality assume $K = \{\ell_1, \ldots, \ell_k\}$. Observe that, by construction, $\partial_{\mathrm{V}} K \supseteq K'$ where $K' = \{r_1, \ldots, r_k\}$. We analyze the "bad event"

$$\mathcal{B}_K := \{|\partial_{\mathrm{V}} K| \leq k + \lfloor \beta k \rfloor\},$$

by considering all subsets of $\{r_{k+1}, \ldots, r_n\}$ of size $\lfloor \beta k \rfloor$ and bounding the probability that *all edges* out of K fall into *one of them and* K'. Note that there are $\binom{n-k}{\lfloor \beta k \rfloor}$ such subsets. See Figure 5.5.

Since σ_n^1 and σ_n^2 are uniform and independent, they each match K to a uniformly chosen subset of the same size in R and we have by a union bound

$$\mathbb{P}[\mathcal{B}_K] \leq \binom{n-k}{\lfloor \beta k \rfloor} \left[\frac{\binom{k+\lfloor \beta k \rfloor}{k}}{\binom{n}{k}} \right]^2 \leq \binom{n}{\lfloor \beta k \rfloor} \frac{\binom{k+\lfloor \beta k \rfloor}{\lfloor \beta k \rfloor}^2}{\binom{n}{k}^2},$$

where we used that $\binom{n}{s} = \binom{n}{n-s}$.

Taking a union bound again, this time over Ks, we have

$$\begin{aligned}\mathbb{P}[\exists K \subseteq L,\ |K| \leq n/2,\ |\partial_{\mathrm{V}} K| \leq& (1+\beta)|K|] \\ &\leq \sum_{K \subseteq L,\ |K| \leq n/2} \mathbb{P}[\mathcal{B}_K] \\ &\leq \sum_{k=1}^{n/2} \binom{n}{k} \binom{n}{\lfloor \beta k \rfloor} \frac{\binom{k+\lfloor \beta k \rfloor}{\lfloor \beta k \rfloor}^2}{\binom{n}{k}^2}. \qquad (5.3.11)\end{aligned}$$

We use the bound $\frac{n^s}{s^s} \leq \binom{n}{s} \leq \frac{e^s n^s}{s^s} \leq \frac{e^t n^t}{t^t}$ for $s \leq t < n$ (see Appendix A; to see the last inequality, note that $\frac{\mathrm{d}}{\mathrm{d}t} \log(\frac{e^t n^t}{t^t}) = \log(\frac{n}{t}) > 0$ for $0 < t < n$). We obtain that the sum in the last display is bounded as

$$\begin{aligned}\sum_{k=1}^{n/2} \binom{n}{k} \binom{n}{\lfloor \beta k \rfloor} \frac{\binom{k+\lfloor \beta k \rfloor}{\lfloor \beta k \rfloor}^2}{\binom{n}{k}^2} &= \sum_{k=1}^{n/2} \binom{n}{\lfloor \beta k \rfloor} \frac{\binom{k+\lfloor \beta k \rfloor}{\lfloor \beta k \rfloor}^2}{\binom{n}{k}} \\ &\leq \sum_{k=1}^{n/2} \frac{e^{\beta k} n^{\beta k}}{(\beta k)^{\beta k}} \frac{\left(\frac{e^{\beta k}(k+\beta k)^{\beta k}}{(\beta k)^{\beta k}}\right)^2}{\frac{n^k}{k^k}} \\ &\leq \sum_{k=1}^{n/2} \left(\frac{k}{n}\right)^{k(1-\beta)} \left(\frac{e^3(1+\beta)^2}{\beta^3}\right)^{\beta k} \\ &= \sum_{k=1}^{\infty} f_n(k), \qquad (5.3.12)\end{aligned}$$

where we defined

$$f_n(k) := \mathbf{1}_{\{k \leq n/2\}} \left(\frac{k}{n}\right)^{k(1-\beta)} \left(\frac{e^3(1+\beta)^2}{\beta^3}\right)^{\beta k}.$$

Let also

$$g(k) := \left[\left(\frac{1}{2}\right)^{1-\beta} \left(\frac{e^3(1+\beta)^2}{\beta^3}\right)^{\beta}\right]^k,$$

and notice that for β small enough

$$|f_n(k)| \leq g(k), \qquad \forall k$$

since $\frac{k}{n} \leq \frac{1}{2}$ for $k \leq \frac{n}{2}$ and

$$\gamma_\beta := \left(\frac{1}{2}\right)^{1-\beta} \left(\frac{e^3(1+\beta)^2}{\beta^3}\right)^{\beta} < 1,$$

using that $\beta^\beta \to 1$ as $\beta \to 0$. Moreover, for each k,

$$f_n(k) \to 0,$$

as $n \to +\infty$, and

$$\sum_{k=1}^{\infty} g(k) \le \frac{1}{1-\gamma_\beta} < +\infty.$$

Hence, by the dominated convergence theorem (Theorem B.4.7), combining (5.3.11) and (5.3.12) we get

$$\mathbb{P}[\exists K \subseteq L,\ |K| \le n/2,\ |\partial_{\mathrm{V}} K| \le (1+\beta)|K|] = \sum_{k=1}^{\infty} f_n(k) \to 0.$$

That concludes the proof. ■

That concludes the proof of Claim 5.3.10. ■

Claim 5.3.10 implies:

Theorem 5.3.12 (Existence of expander family)**.** *For $\alpha > 0$ small enough, there exists a $(3,\alpha)$-expander (multigraph) family.*

Proof. By Claim 5.3.10, for all n large enough, there exists G_n with $\Phi_*(G_n) \ge \alpha$ for some fixed $\alpha > 0$. ■

Fast mixing on expander graphs As we mentioned above, an important property of an expander graph is that random walk on such a graph mixes rapidly. We make this precise.

Claim 5.3.13 (Mixing on expanders)**.** *Let $\{G_n\}$ be a (d,α)-expander family. Then $t_{\mathrm{mix}}(\varepsilon) = \Theta(\log|V(G_n)|)$, where the constant depends on ε and α.*

Proof. Because of the degree assumption, random walk on G_n is reversible with respect to the uniform distribution (see Example 1.1.29). So, by Theorems 5.2.14 and 5.3.5, the mixing time is upper bounded by

$$t_{\mathrm{mix}}(\varepsilon) \le \log\left(\frac{1}{\varepsilon\pi_{\min}}\right) t_{\mathrm{rel}} \le \log\left(\frac{|V(G_n)|}{\varepsilon}\right) 2\alpha^{-2} = O(\log|V(G_n)|).$$

By the diameter-based lower bound on the mixing time for reversible chains (Claim 5.2.25), for n large enough

$$t_{\mathrm{mix}}(\varepsilon) \ge \frac{\Delta^2}{12\log|V(G_n)| + 4|\log \pi_{\min}|},$$

where Δ is the diameter of G_n. For a d-regular graph G_n, the diameter is at least $\log |V(G_n)|$. Indeed, by induction, the number of vertices within graph distance k of any vertex is at most d^k. For d^k to be greater than $|V(G_n)|$, we need $k \geq \log_d |V(G_n)|$. Finally,

$$t_{\mathrm{mix}}(\varepsilon) \geq \frac{(\log_d |V(G_n)|)^2}{16 \log |V(G_n)|} = \Omega(\log |V(G_n)|).$$

That concludes the proof. ■

5.3.4 ▷ *Ising model: Glauber dynamics on complete graphs and expanders*

Let $G = (V, E)$ be a finite, connected graph with maximal degree $\bar{\delta}$. Define $\mathcal{X} := \{-1, +1\}^V$. Recall from Example 1.2.5 that the (ferromagnetic) Ising model on V with *inverse temperature* β is the probability distribution over *spin configurations* $\sigma \in \mathcal{X}$ given by

$$\mu_\beta(\sigma) := \frac{1}{\mathcal{Z}(\beta)} e^{-\beta \mathcal{H}(\sigma)},$$

where

$$\mathcal{H}(\sigma) := -\sum_{i \sim j} \sigma_i \sigma_j,$$

is the *Hamiltonian* and

$$\mathcal{Z}(\beta) := \sum_{\sigma \in \mathcal{X}} e^{-\beta \mathcal{H}(\sigma)},$$

is the *partition function*. In this context, recall that vertices are often referred to as *sites*. The single-site Glauber dynamics of the Ising model (Definition 1.2.8) is the Markov chain on $\mathcal{X}$ which, at each time, selects a site $i \in V$ uniformly at random and updates the spin σ_i according to $\mu_\beta(\sigma)$ conditioned on agreeing with σ at all sites in $V \backslash \{i\}$. Specifically, for $\gamma \in \{-1, +1\}$, $i \in V$, and $\sigma \in \mathcal{X}$, let $\sigma^{i,\gamma}$ be the configuration σ with the state at i being set to γ. Then, letting $n = |V|$, the transition matrix of the Glauber dynamics is

$$Q_\beta(\sigma, \sigma^{i,\gamma}) := \frac{1}{n} \cdot \frac{e^{\gamma \beta S_i(\sigma)}}{e^{-\beta S_i(\sigma)} + e^{\beta S_i(\sigma)}} = \frac{1}{n} \left\{ \frac{1}{2} + \frac{1}{2} \tanh(\gamma \beta S_i(\sigma)) \right\},$$

where

$$S_i(\sigma) := \sum_{j \sim i} \sigma_j.$$

All other transitions have probability 0. Recall that this chain is irreducible and reversible with respect to μ_β. In particular μ_β is the stationary distribution of Q_β.

We showed in Claim 4.3.15 that the Glauber dynamics is fast mixing at high temperature. More precisely we proved that $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = O(n \log n)$ when $\beta < \bar{\delta}^{-1}$. Here we prove a converse: at low temperature, graphs with good enough expansion properties produce exponentially slow mixing of the Glauber dynamics.

Curie-Weiss model

Let $G = K_n$ be the complete graph on n vertices. In this case, the Ising model is often referred to as the *Curie-Weiss model*. It is natural to scale β with n. We define $\alpha := \beta(n-1)$. Since $\bar{\delta} = n-1$, we have that, when $\alpha < 1$, $\beta = \frac{\alpha}{n-1} < \bar{\delta}^{-1}$ so $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = O(n \log n)$. In the other direction, we prove:

Curie-Weiss model

Claim 5.3.14 (Curie-Weiss model: slow mixing at low temperature)**.** *For $\alpha > 1$, $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = \Omega(\exp(r(\alpha)n))$ for some function $r(\alpha) > 0$ not depending on n.*

Proof. We first prove exponential mixing when α is large enough, an argument which will be useful in the generalization to expander graphs below.

The idea of the proof is to bound the edge expansion constant and use Theorem 5.3.5. To simplify the proof, assume n is odd. We denote the edge expansion constant of the chain by $\Phi_*^{\mathcal{X}}$ to avoid confusion with that of the base graph G. Intuitively, because the spins tend to align strongly at low temperature, it takes a considerable amount of time to travel from a configuration with a majority of -1s to a configuration with a majority of $+1$s. Because the model tends to prefer agreeing spins but does not favor any particular spin, a natural place to look for a bottleneck is the set

$$\mathcal{M} := \left\{ \sigma \in \mathcal{X} \,:\, \sum_i \sigma_i < 0 \right\},$$

where the quantity $m(\sigma) := \sum_i \sigma_i$ is called the *magnetization*. Note that the magnetization is positive if and only if a majority of spins are $+1$ and that it forms a Markov chain by itself. So the boundary of the set $\mathcal{M}$ must be crossed to travel from configurations with mostly -1 spins to configurations with mostly -1 spins.

magnetization

Observe further that $\mu_\beta(\mathcal{M}) = 1/2$. The edge expansion constant is hence bounded by

$$\Phi_*^{\mathcal{X}} \leq \frac{\sum_{\sigma \in \mathcal{M}, \sigma' \notin \mathcal{M}} \mu_\beta(\sigma) Q_\beta(\sigma, \sigma')}{\mu_\beta(\mathcal{M})} = 2 \sum_{\sigma \in \mathcal{M}, \sigma' \notin \mathcal{M}} \mu_\beta(\sigma) Q_\beta(\sigma, \sigma'). \quad (5.3.13)$$

Because the Glauber dynamics changes a single spin at a time, in order for $\sigma \in \mathcal{M}$ to be adjacent to a configuration $\sigma' \notin \mathcal{M}$, it must be that

$$\sigma \in \mathcal{M}_{-1} := \{ \sigma \in \mathcal{X} \,:\, m(\sigma) = -1 \},$$

and that $\sigma' = \sigma^{j,+}$ for some site j such that

$$j \in \mathcal{J}_\sigma := \{j \in V \,:\, \sigma_j = -1\}.$$

Because the number of such sites is $(n+1)/2$ on $\mathcal{M}_{-1}$, that is, $|\mathcal{J}_\sigma| = (n+1)/2$ for all $\sigma \in \mathcal{M}_{-1}$, and the Glauber dynamics picks a site uniformly at random, it follows that for $\sigma \in \mathcal{M}_{-1}$

$$\begin{aligned}
\sum_{\sigma\in\mathcal{M},\sigma'\notin\mathcal{M}} \mu_\beta(\sigma) Q_\beta(\sigma,\sigma') &= \sum_{\sigma\in\mathcal{M}_{-1}} \mu_\beta(\sigma) \sum_{j\in\mathcal{J}_\sigma} Q_\beta(\sigma,\sigma^{j,+}) \\
&\leq \sum_{\sigma\in\mathcal{M}_{-1}} \mu_\beta(\sigma) \frac{(n+1)/2}{n} \qquad (5.3.14)\\
&= \frac{1}{2}\left(1+\frac{1}{n}\right)\mu_\beta(\mathcal{M}_{-1}). \qquad (5.3.15)
\end{aligned}$$

Thus plugging this back in (5.3.13) gives

$$\begin{aligned}
\Phi_*^{\mathcal{X}} &\leq \left(1+\frac{1}{n}\right)\mu_\beta(\mathcal{M}_{-1}) \\
&= (1+o(1)) \sum_{\sigma\in\mathcal{M}_{-1}} \frac{e^{-\beta\mathcal{H}(\sigma)}}{\mathcal{Z}(\beta)} \qquad (5.3.16)\\
&= (1+o(1)) \sum_{\sigma\in\mathcal{M}_{-1}} \frac{\exp\left(\frac{\alpha}{n-1}\left[\binom{|\mathcal{J}_\sigma|}{2} + \binom{|\mathcal{J}_\sigma^c|}{2} - |\mathcal{J}_\sigma||\mathcal{J}_\sigma^c|\right]\right)}{\mathcal{Z}(\beta)}.
\end{aligned}$$

We bound the partition function $\mathcal{Z}(\beta) = \sum_{\sigma\in\mathcal{X}} e^{-\beta\mathcal{H}(\sigma)}$ with the term for the all-(-1) configuration, leading to

$$\begin{aligned}
\Phi_*^{\mathcal{X}} &\leq (1+o(1)) \sum_{\sigma\in\mathcal{M}_{-1}} \frac{\exp\left(\frac{\alpha}{n-1}\left[\binom{|\mathcal{J}_\sigma|}{2} + \binom{|\mathcal{J}_\sigma^c|}{2} - |\mathcal{J}_\sigma||\mathcal{J}_\sigma^c|\right]\right)}{\exp\left(\frac{\alpha}{n-1}\left[\binom{|\mathcal{J}_\sigma|}{2} + \binom{|\mathcal{J}_\sigma^c|}{2} + |\mathcal{J}_\sigma||\mathcal{J}_\sigma^c|\right]\right)} \qquad (5.3.17)\\
&= (1+o(1)) \sum_{\sigma\in\mathcal{M}_{-1}} \exp\left(-\frac{2\alpha}{n-1}|\mathcal{J}_\sigma||\mathcal{J}_\sigma^c|\right) \\
&= (1+o(1)) \binom{n}{(n+1)/2} \exp\left(-\frac{2\alpha}{n-1}\left[\frac{n+1}{2}\right]\left[\frac{n-1}{2}\right]\right) \\
&= (1+o(1)) \sqrt{\frac{2}{\pi n}}\, 2^n (1+o(1)) \exp\left(-\frac{\alpha(n+1)}{2}\right) \\
&\leq C_\alpha \sqrt{\frac{2}{\pi n}} \exp\left(-n\left[\frac{\alpha}{2} - \log 2\right]\right),
\end{aligned}$$

for some constant $C_\alpha > 0$ depending on α, where we used Stirling's formula (see Appendix A). Hence, by Theorems 5.2.14 and 5.3.5, for $\alpha > 2\log 2$ there is $r(\alpha) > 0$

$$\mathrm{t_{mix}}(\varepsilon) \geq (\mathrm{t_{rel}} - 1)\log\left(\frac{1}{2\varepsilon}\right) \geq \exp(r(\alpha)n)\log\left(\frac{1}{2\varepsilon}\right).$$

That proves the weaker result.

We now show that $\alpha > 1$ in fact suffices. For this, we need to improve our bound on the partition function in (5.3.17). Writing

$$\begin{aligned}
\mathcal{Z}(\beta) &= \sum_{\sigma\in\mathcal{X}} e^{-\beta\mathcal{H}(\sigma)} \\
&= \sum_{k=0}^{n}\binom{n}{k}\exp\left(\frac{\alpha}{n-1}\left[\binom{k}{2}+\binom{n-k}{2}-k(n-k)\right]\right) \\
&= 2\sum_{k=0}^{(n-1)/2}\binom{n}{k}\exp\left(\frac{\alpha}{n-1}\left[\binom{k}{2}+\binom{n-k}{2}-k(n-k)\right]\right) \\
&=: 2\sum_{k=0}^{(n-1)/2}\mathcal{Y}_{\alpha,k},
\end{aligned}$$

we see that the partition function is a sum of $O(n)$ exponentially large terms and is therefore dominated by the term corresponding to the largest exponent. Using Stirling's formula,

$$\log\binom{n}{k} = (1+o(1))nH(k/n),$$

where $H(p) = -p\log p - (1-p)\log(1-p)$ is the entropy, and therefore

$$\log\mathcal{Y}_{\alpha,k} = (1+o(1))n\underbrace{\left[H(k/n) + \alpha\frac{(k/n)^2+(1-k/n)^2-2(k/n)(1-k/n)}{2}\right]}_{\mathcal{K}_\alpha(k/n)}.$$

where, for $p \in [0,1]$, we let

$$\mathcal{K}_\alpha(p) := H(p) + \alpha\frac{(1-2p)^2}{2}.$$

Note that the first term in $\mathcal{K}_\alpha(p)$ is increasing on $[0,1/2]$ while the second term is decreasing on $[0,1/2]$. In a sense, we are looking at the tradeoff between the contribution from the entropy (i.e., how many ways are there to have k spins with value -1) and that from the Hamiltonian (i.e., how much such a configuration is favored). We seek to maximize $\mathcal{K}_\alpha(p)$ to determine the leading term in the partition function.

By a straightforward computation,

$$\mathcal{K}'_\alpha(p) = \log\left(\frac{1-p}{p}\right) - 2\alpha(1-2p),$$

and

$$\mathcal{K}''_\alpha(p) = -\frac{1}{p(1-p)} + 4\alpha.$$

Observe first that, when $\alpha < 1$ (i.e., at high temperature), $\mathcal{K}'_\alpha(1/2) = 0$ and $\mathcal{K}''_\alpha(p) < 0$ for all $p \in [0,1]$ since $p(1-p) \leq 1/4$. Hence, in that case, $\mathcal{K}_\alpha$ is maximized at $p = 1/2$.

In our case of interest, on the other hand, that is, when $\alpha > 1$, $\mathcal{K}''_\alpha(p) > 0$ in an interval around $1/2$ so there is $p_* < 1/2$ with $\mathcal{K}_\alpha(p_*) > \mathcal{K}_\alpha(1/2) = 1$. So the distribution significantly favors "unbalanced" configurations and crossing $\mathcal{M}_{-1}$ becomes a bottleneck for the Glauber dynamics. Going back to (5.3.17) and bounding $\mathcal{Z}(\beta) \geq 2\mathcal{Y}_{\alpha,\lfloor p_* n\rfloor}$, we get

$$\Phi^{\mathcal{X}}_* = O\left(\exp(-n[\mathcal{K}_\alpha(p_*) - \mathcal{K}_\alpha(1/2)])\right).$$

Applying Theorems 5.2.14 and 5.3.5 concludes the proof. ■

Expander graphs

In the proof of Claim 5.3.14, the bottleneck slowing down the chain arises as a result of the fact that, when $m(\sigma) = -1$, there is a large number of edges in the base graph K_n connecting $\mathcal{J}_\sigma$ and $\mathcal{J}^c_\sigma$. That produces a low probability for such configurations under the ferromagnetic Ising model, where agreeing spins are favored. The same argument easily extends to expander graphs. In words, we prove something that—at first—may seem a bit counter-intuitive: good expansion properties in the base graph produces a bottleneck in the Glauber dynamics at low temperature.

Claim 5.3.15 (Ising model on expander graphs: slow mixing of the Glauber dynamics)**.** *Let $\{G_n\}_n$ be a (d,γ)-expander family. For large enough inverse temperature $\beta > 0$, the Glauber dynamics of the Ising model on G_n satisfies $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = \Omega(\exp(r(\beta)|V(G_n)|))$ for some function $r(\beta) > 0$ not depending on n.*

Proof. Let μ_β be the probability distribution over spin configurations under the Ising model over $G_n = (V,E)$ with inverse temperature β. Let Q_β be the transition matrix of the Glauber dynamics. For not necessarily disjoint subsets of vertices $W_0, W_1 \subseteq V$ in the base graph G_n, let

$$E(W_0, W_1) := \{\{u,v\} : u \in W_0,\ v \in W_1,\ \{u,v\} \in E\},$$

be the set of edges with one endpoint in W_0 and one endpoint in W_1. Let $N = |V(G_n)|$ and assume it is odd for simplicity. We use the notation in the proof of Claim 5.3.14. Following the argument in that proof, we observe that (5.3.15) and (5.3.16) still hold. Thus

$$\Phi_*^{\mathcal{X}} \leq (1+o(1)) \sum_{\sigma \in \mathcal{M}_{-1}} \frac{\exp\left(\beta\left[|E(\mathcal{J}_\sigma, \mathcal{J}_\sigma)| + |E(\mathcal{J}_\sigma^c, \mathcal{J}_\sigma^c)| - |E(\mathcal{J}_\sigma, \mathcal{J}_\sigma^c)|\right]\right)}{\mathcal{Z}(\beta)}.$$

As we did in (5.3.17), we bound the partition function $\mathcal{Z}(\beta) = \sum_{\sigma \in \mathcal{X}} e^{-\beta \mathcal{H}(\sigma)}$ with the term for the all-(-1) configuration, leading to

$$\begin{aligned}
\Phi_*^{\mathcal{X}} &\leq (1+o(1)) \sum_{\sigma \in \mathcal{M}_{-1}} \frac{\exp\left(\beta\left[|E(\mathcal{J}_\sigma, \mathcal{J}_\sigma)| + |E(\mathcal{J}_\sigma^c, \mathcal{J}_\sigma^c)| - |E(\mathcal{J}_\sigma, \mathcal{J}_\sigma^c)|\right]\right)}{\exp\left(\beta\left[|E(\mathcal{J}_\sigma, \mathcal{J}_\sigma)| + |E(\mathcal{J}_\sigma^c, \mathcal{J}_\sigma^c)| + |E(\mathcal{J}_\sigma, \mathcal{J}_\sigma^c)|\right]\right)} \\
&= (1+o(1)) \sum_{\sigma \in \mathcal{M}_{-1}} \exp\left(-2\beta |E(\mathcal{J}_\sigma, \mathcal{J}_\sigma^c)|\right) \\
&= (1+o(1)) \sum_{\sigma \in \mathcal{M}_{-1}} \exp\left(-2\beta |\partial_{\mathrm{E}} \mathcal{J}_\sigma^c|\right) \\
&\leq (1+o(1)) \binom{N}{(N+1)/2} \exp\left(-2\beta\gamma d |\mathcal{J}_\sigma^c|\right) \\
&= (1+o(1)) \sqrt{\frac{2}{\pi N}}\, 2^N (1+o(1)) \exp\left(-\beta\gamma d (N-1)\right) \\
&\leq C_{\beta,\gamma,d} \sqrt{\frac{2}{\pi N}} \exp\left(-N\left[\beta\gamma d - \log 2\right]\right),
\end{aligned}$$

for some constant $C_{\beta,\gamma,d} > 0$. We used the definition of an expander family (Definition 5.3.9) on the fourth line above. Taking $\beta > 0$ large enough gives the result. ■

5.3.5 Congestion ratio

Recall from (5.3.2) that an inequality of the type

$$\mathrm{Var}_\pi[f] \leq C \mathscr{D}(f), \tag{5.3.18}$$

holding for all f is known as a Poincaré inequality. By Theorem 5.3.4, it implies the lower bound $\gamma \geq C^{-1}$ on the spectral gap $\gamma = 1 - \lambda_2$. In this section, we derive such an inequality using a formal measure of "congestion" in the network.

Let $\mathcal{N} = (G, c)$ be a finite, connected network with $G = (V, E)$. We assume that $c(x, y) = \pi(x)P(x, y)$ and therefore $c(x) = \sum_{y\sim x} c(x, y) = \pi(x)$, where π is the stationary distribution of random walk on $\mathcal{N}$. To state the bound, it will be convenient to work with directed edges—this time in both directions. Let $\widetilde{E}$ contain all edges from E with both orientations, that is, for each $e \in \{x, y\}$, $\widetilde{E}$ includes (x, y) and (y, x) with associated weight $c(x, y) = c(y, x) = c(e) > 0$. For a function $f \in \ell^2(V, \pi)$ and an edge $\vec{e} = (x, y) \in \widetilde{E}$, we define as before

$$\nabla f(\vec{e}) = f(y) - f(x).$$

With this notation, we can rewrite the Dirichlet energy as

$$\mathscr{D}(f) = \frac{1}{2}\sum_{x,y} c(x, y)[f(x) - f(y)]^2 = \frac{1}{2}\sum_{\vec{e}\in\widetilde{E}} c(\vec{e})[\nabla f(\vec{e})]^2. \tag{5.3.19}$$

For each pair of vertices x, y, let $\nu_{x,y}$ be a directed path between x and y in the digraph $\widetilde{G} = (V, \widetilde{E})$, as a collection of directed edges. Let $|\nu_{x,y}|$ be the number of edges in the path. The *congestion ratio* associated with the paths $\boldsymbol{\nu} = \{\nu_{x,y}\}_{x,y\in V}$ is

congestion ratio

$$C_{\boldsymbol{\nu}} = \max_{\vec{e}\in\widetilde{E}} \frac{1}{c(\vec{e})} \sum_{x,y:\vec{e}\in\nu_{x,y}} |\nu_{x,y}|\pi(x)\pi(y).$$

Note that $C_{\boldsymbol{\nu}}$ tends to be large when many selected paths, called *canonical paths*, go through the same "congested" edge. To get a good bound in the theorem below, one must choose canonical paths that are well "spread out."

canonical paths

Theorem 5.3.16 (Canonical paths method)**.** *For any choice of paths $\boldsymbol{\nu}$ as above, we have the following bound on the spectral gap*

$$\gamma \geq \frac{1}{C_{\boldsymbol{\nu}}}.$$

Proof. We establish a Poincaré inequality (5.3.18) with $C := C_{\boldsymbol{\nu}}$. The proof strategy is to start with the variance and manipulate it to bring out canonical paths.

For any $f \in \ell^2(V, \pi)$, it can be checked by expanding that

$$\mathrm{Var}_\pi[f] = \frac{1}{2}\sum_{x,y} \pi(x)\pi(y)(f(x) - f(y))^2. \tag{5.3.20}$$

To bring out terms similar to those in (5.3.19), we write $f(x)-f(y)$ as a telescoping sum over the canonical path between x and y. That is, letting $\vec{e}_1, \ldots, \vec{e}_{|\nu_{x,y}|}$ be the edges in $\nu_{x,y}$, observe that

$$f(y) - f(x) = \sum_{i=1}^{|\nu_{x,y}|} \nabla f(\vec{e}_i).$$

By Cauchy-Schwarz (Theorem B.4.8),

$$\begin{aligned}
(f(y) - f(x))^2 &= \left(\sum_{i=1}^{|\nu_{x,y}|} \nabla f(\vec{e}_i) \right)^2 \\
&\leq \left(\sum_{i=1}^{|\nu_{x,y}|} 1^2 \right) \left(\sum_{i=1}^{|\nu_{x,y}|} \nabla f(\vec{e}_i)^2 \right) \\
&= |\nu_{x,y}| \sum_{\vec{e} \in \nu_{x,y}} \nabla f(\vec{e})^2.
\end{aligned}$$

Combining the last display with (5.3.20) and rearranging, we arrive at

$$\begin{aligned}
\mathrm{Var}_\pi[f] &\leq \frac{1}{2} \sum_{x,y} \pi(x)\pi(y)|\nu_{x,y}| \sum_{\vec{e} \in \nu_{x,y}} \nabla f(\vec{e})^2 \\
&= \frac{1}{2} \sum_{\vec{e} \in \widetilde{E}} \nabla f(\vec{e})^2 \sum_{x,y:\vec{e} \in \nu_{x,y}} |\nu_{x,y}|\pi(x)\pi(y) \\
&= \frac{1}{2} \sum_{\vec{e} \in \widetilde{E}} c(\vec{e}) \nabla f(\vec{e})^2 \left(\frac{1}{c(\vec{e})} \sum_{x,y:\vec{e} \in \nu_{x,y}} |\nu_{x,y}|\pi(x)\pi(y) \right) \\
&\leq C_{\boldsymbol{\nu}}\, \mathscr{D}(f).
\end{aligned}$$

That concludes the proof. ■

We give an example next.

Example 5.3.17 (Random walk inside a box)**.** Consider random walk on the following d-dimensional box with sides of length n:

$$V = [n]^d = \{1, \ldots, n\}^d,$$

$$E = \{x, y \in [n]^d \;:\; \|x - y\|_1 = 1\},$$

$$P(x,y) = \frac{1}{|\{z\,:\,z\sim x\}|}, \quad \forall x,y\in[n]^d,\ x\sim y,$$
$$\pi(x) = \frac{|\{z\,:\,z\sim x\}|}{2|E|},$$

and

$$c(e) = \frac{1}{2|E|}, \forall e \in E.$$

We define $\widetilde{E}$ as before.

We use Theorem 5.3.16 to bound the spectral gap. For $x = (x_1,\ldots,x_d), y = (y_1,\ldots,y_d) \in [n]^d$, we construct $\nu_{x,y}$ by matching each coordinate in turn. That is, for two vertices $w, z \in [n]^d$ with a single distinct coordinate, let $[w,z]$ be the directed path from w to z in $\widetilde{G} = (V, \widetilde{E})$ corresponding to a straight line (or the empty path if $w = z$). Then

$$\nu_{x,y} = \bigcup_{i=1}^{d} \left[(y_1,\ldots,y_{i-1},x_i,x_{i+1},\ldots,x_d),(y_1,\ldots,y_{i-1},y_i,x_{i+1},\ldots,x_d)\right]. \tag{5.3.21}$$

It remains to bound

$$C_{\boldsymbol{\nu}} = \max_{\vec{e}\in\widetilde{E}} \frac{1}{c(\vec{e})} \sum_{x,y:\vec{e}\in\nu_{x,y}} |\nu_{x,y}|\pi(x)\pi(y),$$

from above.

Each term in the union defining $\nu_{x,y}$ contains at most n edges, and therefore

$$|\nu_{x,y}| \le dn, \forall x,y.$$

Not attempting to get the best constant factors, the edge weights (i.e., conductances) satisfy

$$c(\vec{e}) = \frac{1}{2|E|} \ge \frac{1}{2\cdot 2dn^d} = \frac{1}{4dn^d},$$

for all $\vec{e}$, since there are n^d vertices and each has at most $2d$ incident edges. Likewise, for any x,

$$\pi(x) = \frac{|\{z\,:\,z\sim x\}|}{2|E|} \le \frac{2d}{2\cdot(dn^d)/2} = \frac{2}{n^d},$$

where we divided by two in the denominator to account for the double-counting of edges. Hence we get

$$\begin{aligned} C_{\boldsymbol{\nu}} &\leq \max_{\vec{e}\in\widetilde{E}} \frac{1}{1/(4dn^d)} \sum_{x,y:\vec{e}\in\nu_{x,y}} (dn)(2/n^d)(2/n^d) \\ &= \frac{16d^2}{n^{d-1}} \max_{\vec{e}\in\widetilde{E}} \left|\{x,y : \vec{e} \in \nu_{x,y}\}\right|. \end{aligned}$$

To bound the cardinality of the set on the last line, we note that any edge $\vec{e} \in \widetilde{E}$ is of the form

$$\vec{e} = ((z_1,\ldots,z_{i-1},z_i,z_{i+1},\ldots,z_d),(z_1,\ldots,z_{i-1},z_i \pm 1,z_{i+1},\ldots,z_d))$$

that is, the endvertices differ by exactly one unit along a single coordinate. By the construction of the path $\nu_{x,y}$ in (5.3.21), if $\vec{e} \in \nu_{x,y}$ then it must lie in the subpath

$$\begin{aligned} &((z_1,\ldots,z_{i-1},z_i,z_{i+1},\ldots,z_d),(z_1,\ldots,z_{i-1},z_i \pm 1,z_{i+1},\ldots,z_d)) \\ &\quad \in [(y_1,\ldots,y_{i-1},x_i,x_{i+1},\ldots,x_d),(y_1,\ldots,y_{i-1},y_i,x_{i+1},\ldots,x_d)]. \end{aligned}$$

But that imposes constraints on x and y. Namely, we must have

$$y_1 = z_1,\ldots,y_{i-1} = z_{i-1}, x_{i+1} = z_{i+1},\ldots,x_d = z_d.$$

The remaining components of x and y (of which there are i of the former and $d-(i-1)$ of the latter) each has at most n possible values (although not all of them are allowed), so that

$$\left|\{x,y : \vec{e} \in \nu_{x,y}\}\right| \leq n^i n^{d-(i-1)} = n^{d+1}.$$

This upper bound is valid for any $\vec{e}$.

Putting everything together, we get the bound

$$C_{\boldsymbol{\nu}} \leq \frac{16d^2}{n^{d-1}} n^{d+1} = 16d^2n^2,$$

so that

$$\gamma \geq \frac{1}{16d^2n^2}.$$

Observe that this lower bound on the spectral gap depends only mildly (i.e., polynomially) in the dimension. ◀

One advantage of the canonical paths method is that it is somewhat robust to modifying the underlying network through comparison arguments. See Exercise 5.17 for a simple illustration.

Exercises

Exercise 5.1. Let A be an $n \times n$ symmetric random matrix. We assume that the entries on and above the diagonal, $A_{i,j}$, $i \leq j$, are independent uniform in $\{+1, -1\}$ (and each entry below the diagonal is equal to the corresponding entry above). Use Talagrand's inequality (Theorem 3.2.32) to prove concentration of the largest eigenvalue of A around its mean (which you do not need to compute).

Exercise 5.2. Let $G = (V, E, w)$ be a network.

(i) Prove formula (5.1.3) for the Laplacian quadratic form. (Hint: For an orientation $G^\sigma = (V, E^\sigma)$ of G (that is, give an arbitrary direction to each edge to turn it into a digraph), consider the matrix $B^\sigma \in \mathbb{R}^{n \times m}$ where the column corresponding to arc (i, j) has $-\sqrt{w_{ij}}$ in row i and $\sqrt{w_{ij}}$ in row j, and every other entry is 0.)

(i) Show that the network Laplacian is positive semidefinite.

Exercise 5.3. Let $G = (V, E, w)$ be a weighted graph with normalized Laplacian $\mathcal{L}$. Show that

$$\mathbf{x}^T \mathcal{L} \mathbf{x} = \sum_{\{i,j\} \in E} w_{ij} \left(\frac{x_i}{\sqrt{\delta(i)}} - \frac{x_j}{\sqrt{\delta(j)}} \right)^2,$$

for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

Exercise 5.4 (2-norm)**.** Prove that

$$\sup_{\mathbf{x} \in \mathbb{S}^{n-1}} \|A\mathbf{x}\|_2 = \sup_{\substack{\mathbf{x} \in \mathbb{S}^{n-1} \\ \mathbf{y} \in \mathbb{S}^{m-1}}} \langle A\mathbf{x}, \mathbf{y} \rangle.$$

[Hint: Use Cauchy-Schwarz (Theorem B.4.8) for one direction, and set $\mathbf{y} = A\mathbf{x}/\|A\mathbf{x}\|_2$ for the other one.]

Exercise 5.5 (Spectral radius of a symmetric matrix)**.** Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. The set $\sigma(A)$ of eigenvalues of A is called the *spectrum* of A and *spectrum*

$$\rho(A) = \max\{|\lambda| \, : \, \lambda \in \sigma(A)\},$$

is its *spectral radius*. Prove that *spectral radius*

$$\rho(A) = \|A\|_2,$$

where recall that

$$\|A\|_2 = \max_{\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^m} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}.$$

Exercise 5.6 (Community recovery in sparse networks)**.** Assume without proof the following theorem.

Theorem 5.3.18 (Remark 3.13 of [BH16])**.** *Consider a symmetric matrix* $\mathbf{Z} = [Z_{i,j}] \in \mathbb{R}^{n \times n}$ *whose entries are independent and obey,* $\mathbb{E}Z_{i,j} = 0$ *and* $Z_{i,j} \le B$*,* $\forall 1 \le i,j \le n$*,* $\mathbb{E}Z_{i,j}^2 \le \sigma^2$ *then with high probability we have* $||\mathbf{Z}|| \lesssim \sigma\sqrt{n} + B\sqrt{\log n}$*.*

Let $(X, G) \sim \mathrm{SBM}_{n,p_n,q_n}$. Show that, under the conditions $p_n \gtrsim \frac{logn}{n}$ and $\sqrt{\frac{p_n}{n}} = o(p_n - q_n)$, spectral clustering achieves almost exact recovery.

Exercise 5.7 (Parseval's identity)**.** Prove Parseval's identity (i.e., (5.2.1)) in the finite-dimensional case.

Exercise 5.8 (Dirichlet kernel)**.** Prove that for $\theta \neq 0$

$$1 + 2\sum_{k=1}^{n} \cos k\theta = \frac{\sin((n+1/2)\theta)}{\sin(\theta/2)}.$$

[Hint: Switch to the complex representation and use the formula for a geometric series.]

Exercise 5.9 (Eigenvalues and periodicity)**.** Let P be a finite irreducible transition matrix reversible with respect to π over V. Show that if P has a nonzero eigenfunction f with eigenvalue -1, then P is not aperiodic. [Hint: Look at x achieving $\|f\|_\infty$.]

Exercise 5.10 (Mixing time: necessary condition for cutoff)**.** Consider a sequence of Markov chains indexed by $n = 1, 2, \ldots$. Assume that each chain has a finite state space and is irreducible, aperiodic, and reversible. Let $\mathrm{t}_{\mathrm{mix}}^{(n)}(\varepsilon)$ and $\mathrm{t}_{\mathrm{rel}}^{(n)}$ be respectively the mixing time and relaxation time of the n-th chain. The sequence is said to have pre-cutoff if

$$\sup_{0<\varepsilon<1/2} \limsup_{n\to+\infty} \frac{\mathrm{t}_{\mathrm{mix}}^{(n)}(\varepsilon)}{\mathrm{t}_{\mathrm{mix}}^{(n)}(1-\varepsilon)} < +\infty.$$

Show that if for some $\varepsilon > 0$

$$\sup_{n\ge 1} \frac{\mathrm{t}_{\mathrm{mix}}^{(n)}(\varepsilon)}{\mathrm{t}_{\mathrm{rel}}^{(n)}} < +\infty,$$

then there is no pre-cutoff. In particular, there is no cutoff, as defined in Remark 4.3.8.

Exercise 5.11 (Relaxation time and variance)**.** Let P be a finite irreducible transition matrix reversible with respect to π over V. Define

$$\mathrm{Var}_\pi[g] = \sum_{x \in V} \pi(x)[g(x) - \pi g]^2.$$

Let γ_* be the absolute spectral gap of P. Show that

$$\mathrm{Var}_\pi[P^t f] \leq (1 - \gamma_*)^{2t} \mathrm{Var}_\pi[f].$$

Exercise 5.12 (Lumping)**.** Let (X_t) be a Markov chain on a finite state space V with transition P. Suppose there is an equivalence relation $\sim$ on V with equivalence classes $V^\sharp$, denoting by $[x]$ the equivalence class of x, such that $[X_t]$ is a Markov chain with transition matrix $P^\sharp([x],[y]) = P(x,[y])$.

(i) Let $f : V \to \mathbb{R}$ be an eigenfunction of P with eigenvalue λ and assume that f is constant on each equivalence class. Prove that $f^\sharp([x]) := f(x)$ defines an eigenfunction of $P^\sharp$. What is its eigenvalue?

(ii) Suppose $g : V^\sharp \to \mathbb{R}$ is eigenfunction of $P^\sharp$ with eigenvalue λ. Prove that $g^\flat : V \to \mathbb{R}$ defined by $g^\flat(x) := g([x])$ is eigenfunction of P. What is its eigenvalue?

Exercise 5.13 (Random walk on path with reflecting boundaries)**.** Let n be an even positive integer. Let (X_t) be simple random walk on the path $\{1, \ldots, n\}$ with reflecting boundaries, that is, the transition matrix P is defined by $P(x, x-1) = P(x, x+1) = 1/2$ for $x \in \{2, \ldots, n-1\}$, and $P(1, 2) = P(n, n-1) = 1$. Use Exercise 5.12 to compute the eigenfunctions of P. [Hint: Use the results of Section 5.2.2.]

Exercise 5.14 (Product chain)**.** For $j = 1, \ldots, d$, let P_j be a transition matrix on the finite state space V_j reversible with respect to the stationary distribution π_j. Let $w = (w_j)_{j \in [d]}$ be a probability distribution over $[d]$. Consider the following Markov chain (X_t) on $V := V_1 \times \cdots \times V_d$: at each step, pick j according to w, then take one step along the j-th coordinate according to P_j.

(i) Compute the transition matrix P and stationary distribution π of the chain (X_t). Show that P is reversible with respect to π.

(ii) Construct an orthonormal basis of $\ell^2(V, \pi)$ made of eigenfunctions of P in terms of eigenfunctions of the P_js. What are the corresponding eigenvalues?

(iii) Compute the spectral gap γ of P in terms of the spectral gaps γ_j of the P_js.

Exercise 5.15 (Hypercube revisited)**.** Use Exercise 5.14 to recover Lemma 5.2.17.

Exercise 5.16 (Norm and Rayleigh quotient)**.** Let P be irreducible and reversible with respect to $\pi > 0$.

(i) Prove the polarization identity

$$\langle Pf, g\rangle_\pi = \frac{1}{4}\left[\langle P(f+g), f+g\rangle_\pi - \langle P(f-g), f-g\rangle_\pi\right].$$

(ii) Show that

$$\|P\|_\pi = \sup\left\{\frac{\langle f, Pf\rangle_\pi}{\langle f, f\rangle_\pi} \,:\, f \in \ell_0(V),\ f \neq 0\right\}.$$

Exercise 5.17 (Random walk on a box with holes)**.** Consider the random walk in Example 5.3.17 with $d = 2$. Suppose we remove from the network an arbitrary collection of horizontal edges at even heights. Use the canonical paths method to derive a lower bound on the spectral gap of the form $\gamma \geq 1/(Cn^2)$. [Hint: Modify the argument in Example 5.3.17 and relate the congestion ratio before and after the removal.]

Bibliographic Remarks

Section 5.1 General references on the spectral theorem, the Courant-Fischer and perturbation results include the classics [HJ13, Ste98]. Much more on spectral graph theory can be gleaned from [Chu97, Nic18]. Section 5.1.4 is based largely on [Abb18], which gives a broad survey of theoretical results for community recovery, and [Ver18, Section 4.5] as well as on scribe notes by Joowon Lee, Aidan Howells, Govind Gopakumar, and Shuyao Li for "MATH 888: Topics in Mathematical Data Science" taught at the University of Wisconsin–Madison in Fall 2021.

Section 5.2 For a great introduction to Hilbert space theory and its applications (including to the Dirichlet problem), consult [SS05, Chapters 4,5]. Section 5.2.1 borrows from [LP17, Chapter 12]. A representation-theoretic approach to computing eigenvalues and eigenfunctions, greatly generalizing the calculations in 5.2.2, is presented in [Dia88]. The presentation in Section 5.2.3 follows [KP, Section 3] and [LP16, Section 13.3]. The Varopoulos-Carne bound is due to Carne [Car85] and Varopoulos [Var85]. For a probabilistic approach to the Varopoulos-Carne bound see Peyre's proof [Pey08]. The application to mixing times is from [LP16]. There are many textbooks dedicated to Markov chain Monte Carlo (MCMC) and its uses in data analysis, for example, [RC04, GL06, GCS$^+$14]. See also [Dia09]. A good overview of the techniques developed in the statistics literature to bound the rate of convergence of MCMC methods (a combination of coupling and Lyapounov arguments) is [JH01]. A deeper treatment of these ideas is developed in [MT09]. A formal definition of the spectral radius and its relationship to the operator norm can be found, for instance, in [Rud73, Part III].

Section 5.3 This section follows partly the presentation in [LP16, Section 6.4], [LPW06, Section 13.6], and [Spi12]. Various proofs of the isoperimetric inequality can be found in [SS03, SS05]. Theorem 5.3.5 is due to [SJ89, LS88]. The approach to its proof used here is due to Luca Trevisan The original Cheeger inequality was proved, in the context of manifolds, in [Che70]. For a fascinating introduction to expander graphs and their applications, see [HLW06]. A detailed account of the Curie-Weiss model can be found in [FV18]. Section 5.3.5 is based partly on [Ber14, Sections 3 and 4]. The method of canonical paths, and some related comparison techniques, were developed in [JS89, DS91, DSC93b, DSC93a]. For more advanced functional techniques for bounding the mixing time, see e.g. [MT06].