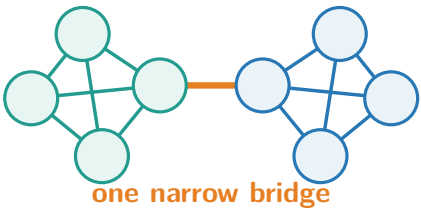
Spectral Graph Theory, Expansion, and Cheeger's Inequality

How eigenvalues reveal cuts, connectivity, and random-walk behavior

Can an eigenvalue detect a bottleneck?



Structural questions

- Is the graph connected?
- Does it contain a sparse cut?
- Does a random walk mix rapidly?
- Can we find a bottleneck efficiently?

We will answer these questions through eigenvalues and eigenvectors of three closely related matrices.

Three views of the same graph



- The Laplacian quadratic form measures variation across edges.
- The second Laplacian eigenvalue measures resistance to separation.
- Cheeger's inequality converts an eigenvector into a sparse cut.
- The same spectral gap controls convergence of random walks.

We focus on D-regular undirected graphs

Throughout the lecture,

$$G = (V, E), \qquad V = \{1, \dots, n\},$$

is an undirected, loopless, D-regular graph.
Parallel edges are allowed and are counted with multiplicity.

Why regular graphs?

Every vertex has the same degree, so

$$\text{degree matrix} = DI,$$

the stationary distribution is uniform, and all relevant matrices share the same eigenvectors.

The theory extends to nonregular graphs using

$$D_{\deg}^{-1}A \quad \text{and} \quad I - D_{\deg}^{-1/2} A D_{\deg}^{-1/2},$$

Part I

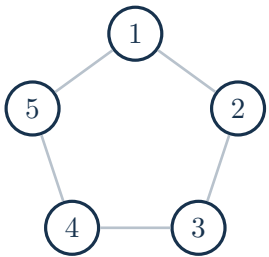
Adjacency and random-walk matrices

The adjacency matrix records the edges

The adjacency matrix $A \in \mathbb{R}^{n \times n}$ is defined by

$$A_{ij} = \#\{\text{edges between } i \text{ and } j\}.$$

For a simple graph, $A_{ij} \in \{0, 1\}$.



$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Regularity produces a graph-independent eigenvector

Since every vertex has degree D, every row sum of A is D:

$$A\mathbf{1} = D\mathbf{1}.$$

Because G is undirected,

$$A = A^{\mathsf{T}},$$

so every column sum is also D.

The trivial adjacency eigenpair

$$\boxed{A\mathbf{1} = D\mathbf{1}.}$$

The vector $\mathbf{1}$ and eigenvalue D occur for every D-regular graph, independently of its finer structure.

Since A is real and symmetric, it has an orthonormal eigenbasis and only real

The random-walk matrix averages over neighbors

Define

$$P := \frac{1}{D} A.$$

Then

$$P_{ji} = \Pr[\text{move from } i \text{ to } j \text{ in one step}].$$

For the standard basis vector e_i,

$$P e_i = \frac{1}{D} \sum_{j \sim_G i} e_j,$$

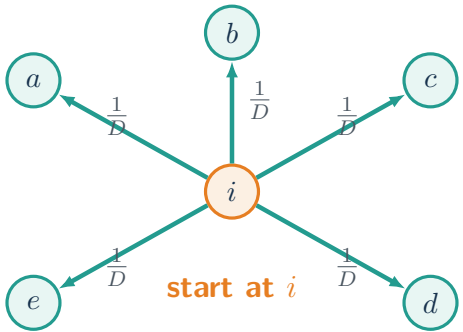
where neighbors are counted with edge multiplicity.

Moreover,

$$P\mathbf{1} = \mathbf{1}, \qquad \mathbf{1}^{\mathsf{T}} P = \mathbf{1}^{\mathsf{T}}.$$

Thus P is symmetric and doubly stochastic.

One multiplication performs one random step



If the initial distribution is deterministic, $p = e_i$, then Pp is the uniform distribution on the D neighbors of i.

Probability distributions are column vectors

Represent a distribution p on V by

$$p_i = \Pr[X = i], \qquad p_i \geq 0, \qquad \mathbf{1}^\mathsf{T} p = 1.$$

Since P has nonnegative entries,

$$Pp \geq 0.$$

Since every column of P sums to one,

$$\mathbf{1}^\mathsf{T} Pp = \mathbf{1}^\mathsf{T} p = 1.$$

Therefore Pp is again a probability distribution.

The uniform distribution is stationary

$$\pi := \frac{1}{n}\mathbf{1}, \qquad P\pi = \pi.$$

Linearity explains the random-walk update

Every distribution is a convex combination of deterministic ones:

$$p = \sum_{i=1}^{n} p_i e_i.$$

Therefore

$$Pp = \sum_{i=1}^{n} p_i P e_i.$$

This is exactly the distribution produced by:

1. sampling X according to p;
2. choosing a uniformly random edge incident to X;
3. moving across that edge to a vertex Y.

In coordinates,

$$(Pp)_j = \sum_{i=1}^{n} \Pr[Y = j \mid X = i] \Pr[X = i].$$

Powers of P encode longer random walks

Let

$$X_0 = j, \qquad X_{t+1} = \text{a uniformly random neighbor of } X_t.$$

Then

$$\boxed{\Pr[X_t = i] = \left\langle e_i, P^t e_j \right\rangle = (P^t)_{ij}.}$$

Proof by induction

The case $t = 0$ is immediate. If the claim holds at time t, then

$$\begin{aligned}
\Pr[X_{t+1} = i] &= \sum_{k=1}^{n} \Pr[X_{t+1} = i \mid X_t = k] \Pr[X_t = k] \\
&= \sum_{k=1}^{n} P_{ik} (P^t)_{kj} \\
&= (P^{t+1})_{ij}.
\end{aligned}$$

The spectral theorem makes repeated walks simple

Since P is symmetric, there are orthonormal eigenvectors $u_1, \dots, u_n$ and real eigenvalues $\mu_1, \dots, \mu_n$ such that

$$P = U \operatorname{diag}(\mu_1, \dots, \mu_n) U^{\mathsf{T}} = \sum_{r=1}^{n} \mu_r u_r u_r^{\mathsf{T}}.$$

Consequently,

$$P^t = \sum_{r=1}^{n} \mu_r^t u_r u_r^{\mathsf{T}}.$$

We may take

$$u_1 = \frac{1}{\sqrt{n}} \mathbf{1}, \qquad \mu_1 = 1.$$

Every component orthogonal to $\mathbf{1}$ is multiplied by μ_r^t after t steps.

The nontrivial eigenvalues contain the graph's information

Order the eigenvalues of P as

$$1 = \mu_1 \geq \mu_2 \geq \cdots \geq \mu_n \geq -1.$$



- The multiplicity of $\mu = 1$ equals the number of connected components.
- Thus G is connected exactly when $\mu_2 < 1$.
- An eigenvalue -1 occurs when a connected graph is bipartite.

Part II

The graph Laplacian

Four spectra express the same information

Define the combinatorial and normalized Laplacians by

$$L := DI - A, \qquad \mathcal{L} := \frac{1}{D}L = I - P.$$

If u_i is a P-eigenvector with eigenvalue μ_i, then

$$Au_i = D\mu_i u_i,$$
$$Lu_i = D(1-\mu_i)u_i,$$

and

$$\mathcal{L}u_i = (1-\mu_i)u_i.$$

Writing

$$\lambda_i := D(1-\mu_i), \qquad \gamma_i := 1-\mu_i = \frac{\lambda_i}{D},$$

we have

$$0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2D.$$

The trivial Laplacian eigenvalue is zero

Since $A\mathbf{1} = D\mathbf{1}$,

$$L\mathbf{1} = (DI - A)\mathbf{1} = 0.$$

Thus

$$\boxed{\lambda_1 = 0, \qquad u_1 = \frac{1}{\sqrt{n}}\mathbf{1}.}$$

The first nontrivial Laplacian eigenvalue is

$$\lambda_2.$$

Its normalized version,

$$\gamma := \frac{\lambda_2}{D} = 1 - \mu_2,$$

is the spectral gap relevant to expansion.

Each edge contributes a rank-one Laplacian

For an edge $e = \{i, j\}$, define its signed incidence vector

$$b_e := e_i - e_j.$$

Its edge Laplacian is

$$L_e := b_e b_e^{\mathsf{T}}.$$

Restricted to coordinates i, j, it has the block

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$



$$b_e = e_i - e_j$$

The graph Laplacian is a sum over edges

Summing one edge Laplacian for each edge occurrence gives

$$\boxed{L_G = \sum_{e \in E} L_e = \sum_{\{i,j\} \in E} (e_i - e_j)(e_i - e_j)^{\mathsf{T}}.}$$

Indeed:

- each edge incident to i contributes 1 to L_{ii};
- since the graph is D-regular, $L_{ii} = D$;
- each edge between i and j contributes -1 to L_{ij} and L_{ji}.

Therefore the sum is exactly

$$DI - A.$$

The edge decomposition proves positive semidefiniteness

For every $f \in \mathbb{R}^n$,

$$\begin{aligned} f^{\mathsf{T}} L f &= \sum_{\{i,j\} \in E} f^{\mathsf{T}} (e_i - e_j)(e_i - e_j)^{\mathsf{T}} f \\ &= \sum_{\{i,j\} \in E} \left((e_i - e_j)^{\mathsf{T}} f \right)^2 \\ &= \boxed{\sum_{\{i,j\} \in E} (f_i - f_j)^2} \\ &\geq 0. \end{aligned}$$

Hence

$$\boxed{L \succeq 0.}$$

The quadratic form measures how much f changes across graph edges.

The Laplacian energy measures lack of smoothness

Define the energy of f by

$$\mathcal{E}_G(f) := f^{\mathsf{T}} L f = \sum_{\{i,j\} \in E} (f_i - f_j)^2.$$

Small energy

Adjacent vertices usually receive similar values.

$$f_i \approx f_j \quad \text{for most } \{i,j\} \in E.$$

Large energy

Many edges connect vertices with very different values.

$|f_i - f_j|$ is often large.

Spectral graph theory studies the smoothest nonconstant functions on the graph.

The kernel records the connected components

Since every term in

$$f^{\mathsf{T}} L f = \sum_{\{i,j\} \in E} (f_i - f_j)^2$$

is nonnegative,

$$f^{\mathsf{T}} L f = 0 \quad \Longleftrightarrow \quad f_i = f_j \text{ across every edge.}$$

Therefore $Lf = 0$ exactly when f is constant on every connected component.

Spectral characterization of connectivity

$$\dim \ker L = \#\{\text{connected components of } G\}.$$

In particular,

$$G \text{ is connected} \quad \Longleftrightarrow \quad \lambda_2 > 0.$$

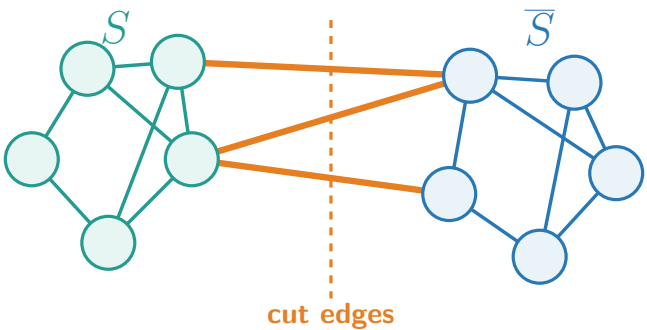
Part III

Cuts, sparsity, and conductance

A cut separates a set from its complement

For $S \subseteq V$, define

$$E(S, \overline{S}) := \Big\{ \{i, j\} \in E : i \in S,\ j \notin S \Big\}.$$



Indicator vectors turn cut size into Laplacian energy

Let $\mathbf{1}_S \in \mathbb{R}^n$ be the indicator vector

$$\mathbf{1}_S(i) = \begin{cases} 1, & i \in S, \\ 0, & i \notin S. \end{cases}$$

For an edge $\{i, j\}$,

$$\big(\mathbf{1}_S(i) - \mathbf{1}_S(j)\big)^2 = \begin{cases} 1, & \{i,j\} \in E(S, \overline{S}), \\ 0, & \text{otherwise.} \end{cases}$$

Therefore

$$\boxed{\mathbf{1}_S^{\mathsf{T}} L \mathbf{1}_S = |E(S, \overline{S})|.}$$

Since $\|\mathbf{1}_S\|_2^2 = |S|$,

$$\frac{\mathbf{1}_S^{\mathsf{T}} L \mathbf{1}_S}{\|\mathbf{1}_S\|_2^2} = \frac{|E(S, \overline{S})|}{|S|}.$$

Conductance normalizes the boundary by available edges

For a nonempty set S with $|S| \leq n/2$, define

$$\boxed{\phi(S) := \frac{|E(S, \overline{S})|}{D|S|}.}$$

Since $D|S|$ is the total number of edge incidences at vertices of S,

$$0 \leq \phi(S) \leq 1.$$

Define the conductance of the graph by

$$\phi_G := \min_{\substack{\varnothing \neq S \subseteq V \\ |S| \leq n/2}} \phi(S).$$

Small ϕ_G means that some set has many internal connections but few edges leading out.

Sparsity uses both sides of the cut

Define the sparsity of a cut by

$$\psi(S) := \frac{|E(S, \overline{S})|}{|S|\,|\overline{S}|}.$$

Since

$$|E(S, \overline{S})| = D|S|\phi(S),$$

we have the exact identity

$$\psi(S) = \frac{D}{|\overline{S}|}\phi(S).$$

When $|S| \leq n/2$,

$$\frac{n}{2} \leq |\overline{S}| \leq n,$$

and hence

$$\boxed{\frac{D}{n}\phi(S) \leq \psi(S) \leq \frac{2D}{n}\phi(S).}$$

Conductance is a one-step escape probability

Start a random walk from a uniformly random vertex of S:

$$p_0 = \frac{\mathbf{1}_S}{|S|}.$$

The probability that the next vertex lies outside S is

$$\begin{aligned}\Pr[X_1 \notin S] &= \left\langle \mathbf{1}_{\overline{S}}, P\frac{\mathbf{1}_S}{|S|} \right\rangle \\ &= \frac{|E(S,\overline{S})|}{D|S|} \\ &= \boxed{\phi(S)}.\end{aligned}$$

Interpretation

Conductance is the probability of *leaving* S in one step when the starting point is uniform on S.

Sparse cuts create metastable regions



If $\phi(S)$ is very small, a random walk beginning in S has difficulty discovering the rest of the graph.
This is the first indication that expansion and mixing should be spectrally related.

Part IV

The spectral gap certifies expansion

The second eigenvalue minimizes a Rayleigh quotient

The Courant–Fischer characterization gives

$$\boxed{\lambda_2 = \min_{\substack{x \neq 0 \\ x \perp \mathbf{1}}} \frac{x^{\mathsf{T}} L x}{\|x\|_2^2}.}$$

Equivalently, every $x \perp \mathbf{1}$ satisfies

$$\frac{x^{\mathsf{T}} L x}{\|x\|_2^2} \geq \lambda_2.$$

Interpretation

λ_2 is the minimum energy of a nonconstant, mean-zero function of unit norm.

To compare λ_2 with a cut, we need a cut vector orthogonal to $\mathbf{1}$.

A raw indicator is not orthogonal to the trivial vector

Although

$$\frac{\mathbf{1}_S^{\mathsf{T}} L \mathbf{1}_S}{\|\mathbf{1}_S\|_2^2} = D\phi(S),$$

we cannot directly compare this quotient with λ_2, because

$$\langle \mathbf{1}, \mathbf{1}_S \rangle = |S| \neq 0.$$

Center the indicator by defining

$$v_S := |\overline{S}|\, \mathbf{1}_S - |S|\, \mathbf{1}_{\overline{S}}.$$

Then

$$\langle \mathbf{1}, v_S \rangle = |S||\overline{S}| - |\overline{S}||S| = 0.$$

The centered cut vector jumps by n across the cut

The coordinates of v_S are

$$(v_S)_i = \begin{cases} |\overline{S}|, & i \in S, \\ -|S|, & i \in \overline{S}. \end{cases}$$

For an edge inside one side,

$$(v_S)_i - (v_S)_j = 0.$$

For an edge crossing the cut,

$$(v_S)_i - (v_S)_j = |\overline{S}| + |S| = n.$$

Hence the Laplacian energy is

$$\boxed{v_S^{\mathsf{T}} L v_S = n^2 |E(S, \overline{S})|.}$$

The norm of the centered cut vector is explicit

We have

$$\begin{aligned}\|v_S\|_2^2 &= |S|\,|\overline{S}|^2 + |\overline{S}|\,|S|^2 \\ &= |S||\overline{S}|\Big(|S| + |\overline{S}|\Big) \\ &= \boxed{n|S||\overline{S}|.}\end{aligned}$$

Therefore

$$\begin{aligned}\frac{v_S^{\mathsf{T}} L v_S}{\|v_S\|_2^2} &= \frac{n^2|E(S,\overline{S})|}{n|S||\overline{S}|} \\ &= n\psi(S).\end{aligned}$$

The spectral gap lower-bounds every cut

Since $v_S \perp \mathbf{1}$, the variational characterization gives

$$\lambda_2 \leq \frac{v_S^{\mathsf{T}} L v_S}{\|v_S\|_2^2} = n\psi(S).$$

Using

$$\psi(S) = \frac{D}{|\overline{S}|}\phi(S)$$

and $|\overline{S}| \geq n/2$,

$$n\psi(S) = \frac{nD}{|\overline{S}|}\phi(S) \leq 2D\phi(S).$$

Thus every nonempty S with $|S| \leq n/2$ satisfies

$$\boxed{\phi(S) \geq \frac{\lambda_2}{2D}.}$$

A spectral expander is an edge expander

Define the normalized spectral gap

$$\gamma := \frac{\lambda_2}{D} = 1 - \mu_2.$$

Taking the minimum over all admissible S,

$$\boxed{\phi_G \geq \frac{\gamma}{2}.}$$

Expansion terminology

- G is an α-edge expander if $\phi_G \geq \alpha$.
- Under the lecture's convention, G is an α-spectral expander if

$$\frac{\lambda_2}{2D} \geq \alpha.$$

Part V

Cheeger's inequality and spectral partitioning

Cheeger's inequality works in both directions

Let

$$\phi_G = \min_{\substack{\varnothing \neq S \subseteq V \\ |S| \leq n/2}} \phi(S), \qquad \gamma = \frac{\lambda_2}{D}.$$

Cheeger's inequality for D-regular graphs

$$\boxed{\frac{\gamma}{2} \leq \phi_G \leq \sqrt{2\gamma}.}$$

In terms of the combinatorial Laplacian eigenvalue,

$$\boxed{\frac{\lambda_2}{2D} \leq \phi_G \leq \sqrt{\frac{2\lambda_2}{D}}.}$$

We have already proved the lower bound. The upper bound is constructive.

The proof rounds an eigenvector into a cut

Let u be an eigenvector satisfying

$$Lu = \lambda_2 u, \qquad u \perp \mathbf{1}.$$

The proof has two steps:

1. **Truncate to one sign.** Construct a nonnegative vector y such that

$$|\operatorname{supp}(y)| \leq \frac{n}{2}, \qquad \frac{y^{\mathsf{T}} L y}{\|y\|_2^2} \leq \lambda_2.$$

2. **Threshold y.** Find a set $S \subseteq \operatorname{supp}(y)$ such that

$$\phi(S) \leq \sqrt{\frac{2y^{\mathsf{T}} L y}{D \|y\|_2^2}}.$$

Combining the two inequalities proves the theorem.

Preprocessing lemma: keep the smaller sign support

Lemma

Let $u \perp \mathbf{1}$ satisfy $Lu = \lambda_2 u$. There exists a nonzero vector $y \geq 0$ such that

$$|\operatorname{supp}(y)| \leq \frac{n}{2}$$

and

$$\frac{y^{\mathsf{T}} L y}{\|y\|_2^2} \leq \lambda_2.$$

Since $u \perp \mathbf{1}$ and $u \neq 0$, it has both positive and negative coordinates.
At least one of the sets

$$\{i : u_i > 0\} \qquad \{i : u_i < 0\}$$

Choose the sign before taking the positive part

Replace u by $-u$, if necessary, so that

$$|\{i : u_i > 0\}| \leq \frac{n}{2}.$$

Define

$$y_i := \max\{u_i, 0\}.$$

Then

$$y \geq 0, \qquad \operatorname{supp}(y) = \{i : u_i > 0\}, \qquad |\operatorname{supp}(y)| \leq \frac{n}{2}.$$

Moreover,

$$y_i = u_i \quad \text{whenever } y_i > 0,$$

and for every j,

$$y_j \geq u_j.$$

Truncation does not increase the Rayleigh quotient

For every i with $y_i > 0$,

$$\begin{aligned}(Ly)_i &= Dy_i - \sum_{j \sim i} y_j \\ &\leq Du_i - \sum_{j \sim i} u_j \\ &= (Lu)_i \\ &= \lambda_2 u_i \\ &= \lambda_2 y_i.\end{aligned}$$

Multiplying by y_i and summing,

$$\begin{aligned}y^{\mathsf{T}} L y &= \sum_{i : y_i > 0} y_i (Ly)_i \\ &\leq \lambda_2 \sum_{i : y_i > 0} y_i^2\end{aligned}$$

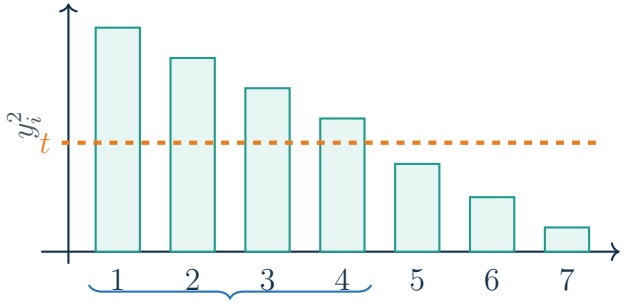
Thresholding converts values into nested cuts

Rescale y so that

$$0 \le y_i \le 1.$$

This does not change its Rayleigh quotient.
For $t \in [0, 1]$, define

$$S_t := \{i : y_i^2 > t\}, \qquad w(t) := \mathbf{1}_{S_t}.$$



The thresholding lemma

Lemma

Let $y \geq 0$ be nonzero and satisfy

$$|\operatorname{supp}(y)| \leq \frac{n}{2}.$$

Then some nonempty threshold set $S_t \subseteq \operatorname{supp}(y)$ satisfies

$$\boxed{\phi(S_t) \leq \sqrt{\frac{2y^{\mathsf{T}}Ly}{D\|y\|_2^2}}.}$$

We prove this by choosing

$$t \sim \operatorname{Uniform}[0,1]$$

and comparing

Expected threshold-set size equals squared mass

Since $w_i(t) \in \{0, 1\}$,

$$\|w(t)\|_2^2 = |S_t|.$$

For each i,

$$\Pr[i \in S_t] = \Pr[t < y_i^2] = y_i^2.$$

Therefore

$$\begin{aligned}\mathbb{E}[|S_t|] &= \sum_{i=1}^{n} \Pr[i \in S_t] \\ &= \sum_{i=1}^{n} y_i^2 \\ &= \|y\|_2^2 .\end{aligned}$$

Multiplying by D,

An edge is cut exactly when the threshold falls between

For an edge $\{i, j\}$,

$$(w_i(t) - w_j(t))^2 = 1$$

exactly when t lies strictly between y_i^2 and y_j^2.
Thus

$$\mathbb{E}[(w_i(t) - w_j(t))^2] = |y_i^2 - y_j^2|.$$

Since $y_i, y_j \geq 0$,

$$|y_i^2 - y_j^2| = |y_i - y_j|(y_i + y_j).$$

Summing over edges,

$$\mathbb{E}[w(t)^\mathsf{T} L w(t)] = \sum_{\{i,j\} \in E} |y_i - y_j|(y_i + y_j).$$

Cauchy–Schwarz separates variation from magnitude

Applying Cauchy–Schwarz,

$$\begin{aligned}\mathbb{E}[w^{\mathsf{T}}Lw] &\le \sqrt{\sum_{\{i,j\}\in E}(y_i - y_j)^2}\sqrt{\sum_{\{i,j\}\in E}(y_i + y_j)^2} \\ &= \sqrt{y^{\mathsf{T}}Ly}\sqrt{\sum_{\{i,j\}\in E}(y_i + y_j)^2}.\end{aligned}$$

The first factor is exactly the Laplacian energy.
It remains to control the second factor using regularity.

Regularity bounds the second Cauchy–Schwarz factor

Using

$$(a+b)^2 \le 2a^2 + 2b^2,$$

we obtain

$$\begin{aligned}\sum_{\{i,j\}\in E} (y_i + y_j)^2 &\le 2 \sum_{\{i,j\}\in E} (y_i^2 + y_j^2) \\ &= 2D \sum_{i=1}^{n} y_i^2 \\ &= 2D \, \|y\|_2^2 .\end{aligned}$$

Therefore

$$\mathbb{E}[w^{\mathsf{T}} L w] \le \sqrt{2D \, y^{\mathsf{T}} L y} \; \|y\|_2$$

$$= \|y\|_2 \sqrt{2 y^{\mathsf{T}} L y}$$

Averaging guarantees one good threshold

Define

$$T := \sqrt{\frac{2y^{\mathsf{T}}Ly}{D\,\|y\|_2^2}}.$$

We proved

$$\mathbb{E}[w^{\mathsf{T}}Lw] \le T\,\mathbb{E}[D\,\|w\|_2^2].$$

Hence some nonempty threshold set satisfies

$$w^{\mathsf{T}}Lw \le TD\,\|w\|_2^2\,.$$

Since $w = \mathbf{1}_{S_t}$,

$$w^{\mathsf{T}}Lw = |E(S_t, \overline{S_t})|, \qquad \|w\|_2^2 = |S_t|.$$

Therefore

$$\phi(S_t) \le T.$$

The support bound makes this a valid conductance cut

Every threshold set satisfies

$$S_t \subseteq \operatorname{supp}(y).$$

Since preprocessing ensured

$$|\operatorname{supp}(y)| \leq \frac{n}{2},$$

we also have

$$|S_t| \leq \frac{n}{2}.$$

Thus its conductance is exactly

$$\phi(S_t) = \frac{w^{\mathsf{T}} L w}{D \|w\|_2^2}.$$

The thresholding lemma therefore produces a valid candidate for the minimum defining ϕ_G.

Combining the lemmas proves Cheeger's upper bound

The preprocessing lemma gives

$$\frac{y^{\mathsf{T}} L y}{\|y\|_2^2} \leq \lambda_2.$$

The thresholding lemma gives a set S_t with

$$\begin{aligned}\phi(S_t) &\leq \sqrt{\frac{2y^{\mathsf{T}} L y}{D\,\|y\|_2^2}} \\ &\leq \sqrt{\frac{2\lambda_2}{D}} \\ &= \sqrt{2\gamma}.\end{aligned}$$

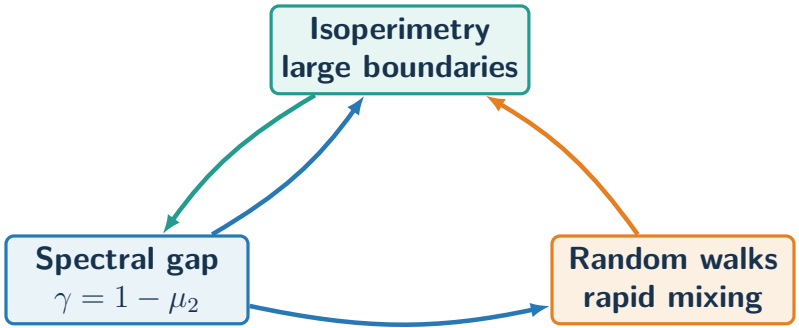
Since $\phi_G \leq \phi(S_t)$,

$$\boxed{\phi_G \leq \sqrt{2\gamma}}$$

Part VI

Expansion, spectral gap, and mixing

Three phenomena describe the same obstruction



A bottleneck simultaneously creates

a sparse cut, a slowly varying eigenvector, a slowly mixing walk.

Cheeger's inequality relates geometry and spectrum

Writing $h := \phi_G$, Cheeger's inequality says

$$\frac{\gamma}{2} \leq h \leq \sqrt{2\gamma}.$$

Equivalently,

$$\boxed{\frac{h^2}{2} \leq \gamma \leq 2h.}$$

Thus:

- a constant spectral gap forces constant edge expansion;
- constant edge expansion forces a constant spectral gap;
- a very small cut forces a small second eigenvalue;
- a very small second eigenvalue yields a sparse sweep cut.

Expansion and spectral gap are quantitatively equivalent up to the quadratic loss in Cheeger's upper bound.

The power-method viewpoint explains mixing

Decompose an initial distribution as

$$p_0 = \pi + \sum_{i=2}^{n} \alpha_i u_i, \qquad \pi = \frac{1}{n}\mathbf{1}.$$

After t steps,

$$P^t p_0 = \pi + \sum_{i=2}^{n} \alpha_i \mu_i^t u_i.$$

This is precisely the phenomenon behind the power method:

- the trivial eigenvector $u_1 = \mathbf{1}/\sqrt{n}$ survives;
- every orthogonal component is multiplied by a smaller eigenvalue power;
- a spectral gap yields exponential convergence toward u_1.

For mixing, negative eigenvalues require one additional precaution.

Bipartite graphs require a lazy random walk

In a connected bipartite graph,

$$\mu_n = -1.$$

The ordinary walk alternates between the two sides, so it cannot converge to the uniform distribution at every time.
Define the lazy walk

$$\widetilde{P} := \frac{1}{2}(I + P).$$

At every step:

- remain at the current vertex with probability $1/2$;
- take an ordinary random-walk step with probability $1/2$.

Its eigenvalues are

$$\widetilde{\mu}_i = \frac{1 + \mu_i}{2} \in [0, 1].$$

The spectral gap gives exponential mixing

For the lazy walk,

$$\widetilde{\mu}_2 = \frac{1+\mu_2}{2} = 1 - \frac{\gamma}{2}.$$

Therefore

$$\begin{aligned}\left\|\widetilde{P}^t p_0 - \pi\right\|_2 &\le \left(1-\frac{\gamma}{2}\right)^t \|p_0 - \pi\|_2 \\ &\le e^{-\gamma t/2} \|p_0 - \pi\|_2 .\end{aligned}$$

Using

$$\|q\|_{\mathrm{TV}} \le \frac{\sqrt{n}}{2} \|q\|_2 ,$$

a constant spectral gap implies mixing in

$$O(\log n)$$

steps, and more generally in

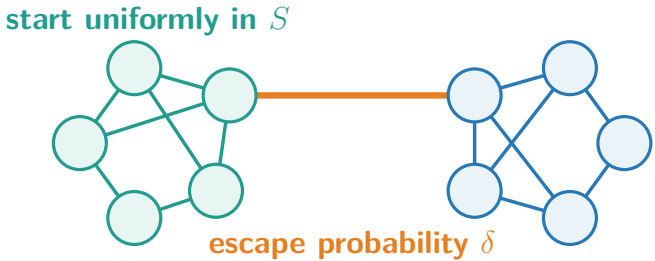
$$O\left(\frac{1}{\gamma}\left(\log n + \log\frac{1}{\varepsilon}\right)\right)$$

A sparse cut prevents rapid mixing

Suppose

$$\phi(S) = \delta, \qquad |S| \leq \frac{n}{2},$$

and start from the uniform distribution on S.



For much less than $1/\delta$ steps, escape is unlikely

Consider a stationary walk $Y_0, Y_1, \ldots$, with Y_0 uniform on V, conditioned on $Y_0 \in S$.
This conditioning makes Y_0 uniform on S.
At any time r, stationarity gives

$$\Pr[Y_r \in S,\ Y_{r+1} \notin S] = \frac{|E(S, \overline{S})|}{Dn} = \frac{|S|}{n}\phi(S).$$

Therefore

$$\Pr[Y_r \in S,\ Y_{r+1} \notin S \mid Y_0 \in S] \leq \phi(S) = \delta.$$

By a union bound,

$$\Pr[\text{the walk leaves } S \text{ during its first } t \text{ steps}] \leq t\delta.$$

Slow escape gives a mixing-time lower bound

If

$$t \le \frac{1}{4\delta},$$

then a walk starting uniformly in S satisfies

$$\Pr[X_t \in S] \ge 1 - t\delta \ge \frac{3}{4}.$$

The uniform distribution on V assigns mass

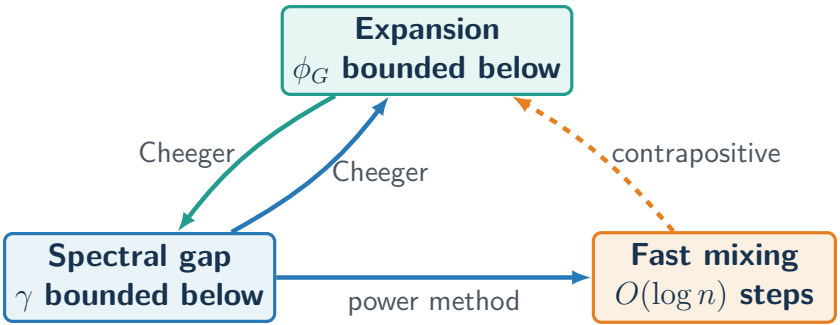
$$\pi(S) = \frac{|S|}{n} \le \frac{1}{2}.$$

Hence

$$\|\mathcal{L}(X_t) - \pi\|_{\mathrm{TV}} \ge \Pr[X_t \in S] - \pi(S) \ge \frac{1}{4}.$$

Thus a cut of conductance δ prevents mixing in substantially fewer than $1/\delta$ steps.

The three implications fit together



Sparse cuts, small spectral gaps, and slow mixing are different manifestations of the same bottleneck.

What the spectrum revealed

1. The random-walk matrix encodes t-step transition probabilities:
$$\Pr[X_t = i \mid X_0 = j] = (P^t)_{ij}.$$
2. The Laplacian quadratic form measures edgewise variation:
$$f^{\mathsf{T}} L f = \sum_{\{i,j\} \in E} (f_i - f_j)^2.$$
3. Indicator vectors turn Laplacian energy into cut size.
4. The second eigenvalue certifies expansion:
$$\phi_G \geq \frac{\lambda_2}{2D}.$$
5. Thresholding a second eigenvector finds a sparse cut:
$$\phi_G \leq \sqrt{\frac{2\lambda_2}{D}}.$$
6. The same spectral gap drives exponential mixing of lazy walks.

Geometry, spectrum, and dynamics are three views of one graph.

Cheeger's inequality is the bridge.