Computer Science 2: Advanced Algorithms Fall 2026

Lecture 4 — September 7, 2026

Prof. Jarosław Błasiok *Scribe: Alessandro Marchetti, Giuseppe Nicola Liso*

1 Overview

In the last lecture we studied the spectral theorem and scalar and block power iteration, including Gaussian initialization and eigengap-controlled eigenvector and invariant-subspace recovery.

In this lecture we study adjacency, random-walk, and Laplacian matrices and use them to relate cuts and conductance, the normalized spectral gap, and random-walk mixing through Cheeger's inequality.

The central objective is to show that three descriptions of a regular graph (i.e its cut geometry, the spectrum of its associated matrices, and the dynamics of its random walk) express the same underlying bottleneck phenomenon. Cheeger's inequality connects conductance with the normalized spectral gap, while spectral decomposition relates this gap to the rate at which a lazy random walk approaches the uniform distribution.

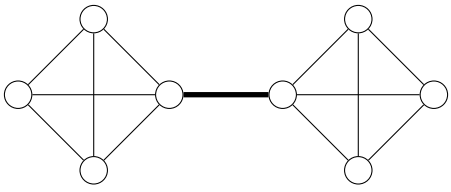
1.1 Spectral Graph Theory

Spectral graph theory studies structural properties of a graph through the eigenvalues and eigenvectors of matrices associated with it. The principal questions are:

- is the graph connected?
- does it contain a sparse cut?
- does a random walk mix rapidly?
- can a bottleneck be found efficiently?

A random walk repeatedly moves from the current vertex across a uniformly chosen incident edge. Mixing measures how rapidly its distribution loses dependence on the initial vertex and approaches stationarity. A narrow bridge between two dense regions obstructs this process and is simultaneously visible as a sparse cut and as a small spectral gap.

The following graph illustrates two dense regions connected by one narrow bridge.



The graph has three complementary descriptions:

- *geometry*: cuts and their edge boundaries;
- *spectrum*: eigenvalues and eigenvectors;
- *dynamics*: random walks and mixing.

Throughout, let $G = (V, E)$ be an undirected, loopless, D-regular graph with $V = \{1, \dots, n\}$. Parallel edges are permitted and counted with multiplicity. Regularity implies that

- the degree matrix is DI;
- the stationary distribution is uniform;
- the adjacency, random-walk, and Laplacian matrices share the same eigenvectors.

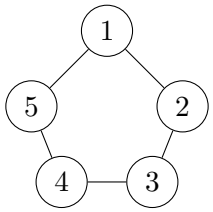
For a nonregular graph with degree matrix D_{deg}, the row-vector transition convention uses $D_{\text{deg}}^{-1}A$, whereas the present column-vector convention uses AD_{deg}^{-1}. The symmetric normalized Laplacian is $I - D_{\text{deg}}^{-1/2}AD_{\text{deg}}^{-1/2}$.

1.2 Adjacency and random walk matrices

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

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

Thus $A_{ij} \in \{0, 1\}$ for a simple graph, while parallel edge occurrences produce larger entries. Since the graph is undirected and loopless, $A = A^\top$ and $A_{ii} = 0$. For the five-cycle shown below,



its adjacency matrix is

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

Every row sum of A is D[1], because row i counts the edge occurrences incident to vertex i. Symmetry then gives the same statement for columns. Therefore

$$A\mathbf{1} = D\mathbf{1}, \qquad \mathbf{1}^\top A = D\mathbf{1}^\top.$$

[1]Throughout the notes we will avoid referring every time to D_{deg}, but for ease of the notation just D.

Consequently, $\mathbf{1}$ is an eigenvector of A with eigenvalue D. Since A is real and symmetric, the spectral theorem provides a real orthonormal eigenbasis.

A **random walk** is a stochastic process $X_0, X_1, \dots$ on V. Conditionally on $X_t = i$, one of the D incident edge occurrences is selected uniformly, with probability $1/D$, and the walk crosses that edge. Its **random-walk matrix** is

$$P := \frac{1}{D}A, \qquad P_{ji} = \frac{A_{ji}}{D} = \mathbb{P}[X_{t+1} = j \mid X_t = i].$$

The second index is the current vertex and the first is the destination, in accordance with the column-vector convention. If e_i denotes the ith standard basis vector, then e_i is also the deterministic distribution concentrated at i, and multiplication by e_i selects column i:

$$(Pe_i)_j = P_{ji} = \frac{A_{ji}}{D}.$$

Thus Pe_i is the distribution obtained after one step from i.

Consider a probability distribution as a column vector p, where $p_i = \mathbb{P}[X = i]$, $p_i \geq 0$, and $\mathbf{1}^\top p = 1$. Every such vector is the convex combination $p = \sum_i p_i e_i$. If X has distribution p and Y is obtained by one random-walk step, linearity and the law of total probability give

$$Pp = \sum_{i=1}^{n} p_i P e_i, \qquad (Pp)_j = \sum_{i=1}^{n} \mathbb{P}[Y = j \mid X = i]\,\mathbb{P}[X = i] = \mathbb{P}[Y = j].$$

Thus multiplication by P performs one random-walk step. The entries of P are nonnegative. Moreover, its row and column sums equal one:

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

Hence P is doubly stochastic. In particular, if p is a probability distribution, then $Pp \geq 0$ and

$$\mathbf{1}^\top Pp = (\mathbf{1}^\top P)p = \mathbf{1}^\top p = 1.$$

Therefore P maps probability distributions to probability distributions.

Theorem 1. *The uniform distribution is stationary.*

Proof. Let $\pi := n^{-1}\mathbf{1}$. Since each coordinate is $1/n$, π is the uniform probability distribution. Moreover,

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

Thus a uniformly distributed position remains uniformly distributed after one step.

■

The factors $1/n$ and $1/D$ have different roles: $1/n$ is the probability of a specified current vertex under π, whereas $1/D$ is the conditional probability of a specified incident edge. Their product $1/(nD)$ is the joint probability of being at that vertex and selecting that edge in stationarity.

Powers of P encode longer walks. If $X_0 = j$, then

$$\mathbb{P}[X_t = i \mid X_0 = j] = (P^t)_{ij}.$$

For $t = 0$, both sides equal 1 when $i = j$ and 0 otherwise. If the identity holds at time t, conditioning on $X_t = k$ gives

$$\begin{aligned}\mathbb{P}[X_{t+1} = i \mid X_0 = j] &= \sum_{k=1}^{n} \mathbb{P}[X_{t+1} = i \mid X_t = k]\, \mathbb{P}[X_t = k \mid X_0 = j] \\ &= \sum_{k=1}^{n} P_{ik}(P^t)_{kj} = (P^{t+1})_{ij}.\end{aligned}$$

This proves the identity by induction. Consequently, an arbitrary initial distribution p_0 evolves according to $p_t = P^t p_0$.

Since P is symmetric, the *spectral theorem* provides orthonormal eigenvectors $u_1, \ldots, u_n$ and real eigenvalues $\mu_1, \ldots, \mu_n$ such that

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

Since $U^\top U = I$, repeated multiplication gives

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

The identity $P\mathbf{1} = \mathbf{1}$ supplies the normalized eigenpair

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

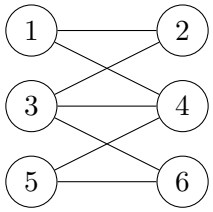
Every eigenvalue of P belongs to $[-1, 1]$. Indeed, let $Pv = \mu v$, and choose i such that $|v_i| = \max_j |v_j|$. Since a nonzero eigenvector has $|v_i| > 0$,

$$\begin{aligned}|\mu|\,|v_i| = |(Pv)_i| &= \left|\sum_j P_{ij} v_j\right| \le \sum_j P_{ij}|v_j| \\ &\le |v_i| \sum_j P_{ij} = |v_i|.\end{aligned}$$

Thus $|\mu| \le 1$. Order the eigenvalues by numerical value,

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

The following graph is bipartite.



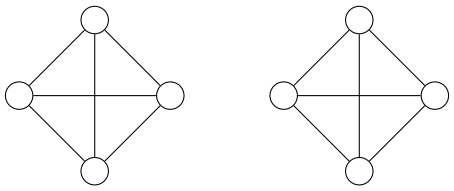
For a connected graph, -1 is an eigenvalue of P if and only if the graph is bipartite. If $V = L \cup R$ is a bipartition, define $s_i = 1$ on L and $s_i = -1$ on R. Every neighbor of a vertex lies on the opposite side, hence $Ps = -s$. Conversely, suppose $Pv = -v$ and choose i with $|v_i| = M := \max_j |v_j|$. The equality

$$-v_i = (Pv)_i = \frac{1}{D} \sum_{j \sim i} v_j$$

expresses an extreme value of magnitude M as an average of values in $[-M, M]$. Equality is possible only if every neighbor has value $-v_i$. Connectedness propagates this relation to every vertex, so all values have magnitude M and every edge joins opposite signs. The positive and negative vertices therefore form a bipartition.

An ordinary walk on a bipartite graph alternates between the two sides: after an even number of steps it is on the starting side, and after an odd number it is on the opposite side. The **lazy random walk**, which stays at the current vertex with probability $1/2$ and makes an ordinary step with probability $1/2$, removes this periodicity.

The multiplicity of the eigenvalue 1 records the number of connected components. The following graph has two components.



For the illustrated graph, the two component indicators are

$$v_1 = (1, 1, 1, 1, 0, 0, 0, 0)^\top, \qquad v_2 = (0, 0, 0, 0, 1, 1, 1, 1)^\top,$$

and $Pv_1 = v_1$, $Pv_2 = v_2$.

More generally, let $C_1, \ldots, C_c$ be the connected components. Each indicator $\mathbf{1}_{C_r}$ satisfies $P\mathbf{1}_{C_r} = \mathbf{1}_{C_r}$, and the indicators are linearly independent because their supports are disjoint. Hence the multiplicity of 1 is at least c. Conversely, if $Pv = v$, choose a vertex at which v is maximal within a fixed component. Its value equals the average of its neighbors' values, all of which are no larger. Every neighbor must therefore have the same value. Repeating this argument along paths shows that v is constant on each component. The space of such vectors has dimension c, proving

$$\mathrm{mult}_P(1) = c.$$

Consequently,

$$G \text{ is connected} \quad \Longleftrightarrow \quad \mu_2 < 1.$$

1.3 Graph Laplacian

The **combinatorial** and **normalized Laplacians** are

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

If $Pu_i = \mu_i u_i$, then $Au_i = D\mu_i u_i$ and

$$L_G u_i = D(1-\mu_i)u_i =: \lambda_i u_i, \qquad \mathcal{L}_G u_i = (1-\mu_i)u_i =: \gamma_i u_i.$$

Thus all four matrices have the same eigenvectors. Since the μ_i are ordered decreasingly by value, the $\lambda_i = D(1-\mu_i)$ are ordered increasingly:

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

The zero eigenvalue also follows directly from $L_G\mathbf{1} = (DI - A)\mathbf{1} = 0$. The first nontrivial Laplacian eigenvalue and its normalized version are

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

The quantity γ is the normalized spectral gap between the trivial eigenvalue $\mu_1 = 1$ and the second-largest eigenvalue μ_2.

For an edge occurrence $e = \{i, j\}$, choose an arbitrary orientation and define the signed incidence vector $b_e = e_i - e_j$. The **edge Laplacian** $L_e = b_e b_e^\top$ has the block

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

on coordinates i, j and is zero elsewhere. Reversing the orientation changes b_e to $-b_e$ but leaves $b_e b_e^\top$ unchanged. Summing over edge occurrences gives

$$\sum_{e \in E} b_e b_e^\top = DI - A = L_G.$$

Indeed, every edge incident to i contributes 1 to the diagonal entry (i, i), producing D, and every edge between distinct vertices i, j contributes -1 to entries (i, j) and (j, i).

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

$$\begin{aligned} f^\top L_G f &= \sum_{e \in E} f^\top b_e b_e^\top f = \sum_{\{i,j\} \in E} \left((e_i - e_j)^\top f\right)^2 \\ &= \sum_{\{i,j\} \in E} (f_i - f_j)^2 \ge 0. \end{aligned} \tag{1}$$

Hence $L_G \succeq 0$: the Laplacian is positive semidefinite. Its quadratic form is the Laplacian energy and measures the variation of f across edges. It is small when adjacent vertices usually receive similar values.

The same identity characterizes the kernel. Since every summand is nonnegative,

$$f^\top L_G f = 0 \quad \iff \quad f_i = f_j \text{ across every edge.}$$

Equality along edges propagates along paths, so f is constant on each connected component. Conversely, every componentwise-constant vector has zero energy and belongs to the kernel. The component indicators form a basis, and therefore

$$\dim \ker L_G = \#\{\text{connected components of } G\}, \qquad G \text{ connected} \iff \lambda_2 > 0.$$

1.4 Cuts, sparsity and conductance

For $S \subseteq V$, let $\overline{S} = V \setminus S$ and define its edge boundary by

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

The indicator vector of S is

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

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

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

Applying the energy identity (1) therefore gives

$$\mathbf{1}_S^\top L_G \mathbf{1}_S = |E(S, \overline{S})|.$$

Since $\|\mathbf{1}_S\|_2^2 = |S|$, its Rayleigh quotient is

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

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

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

The denominator $D|S|$ is the total number of incident edge choices at vertices of S, while the numerator counts those choices that leave S.

Definition 2. *The **conductance** of the graph $G = (V, E)$ is*

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

Observation 3. *The restriction $|S| \le n/2$ selects the smaller side of each nontrivial cut. Thus no cut is omitted, and small ϕ_G means that some set has few outgoing edges relative to its available incident edges.*

Definition 4. *The **sparsity** of a cut is*

$$\psi(S) := \frac{|E(S, \overline{S})|}{|S|\,|\overline{S}|}. \tag{3}$$

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

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

For $|S| \le n/2$, $n/2 \le |\overline{S}| \le n$, and hence

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

Observation 5. *Conductance is the probability of leaving S in one step when the starting point is uniform on S.*

Let X_0 be uniform on S, so its distribution is $p_0 = \mathbf{1}_S/|S|$, and let X_1 be the vertex reached after one transition. Then

$$\begin{aligned}\mathbb{P}[X_1 \notin S] = \mathbf{1}_{\overline{S}}^{\top} P \frac{\mathbf{1}_S}{|S|} &= \frac{1}{D|S|} \sum_{j \in \overline{S}} \sum_{i \in S} A_{ji} \\ &= \frac{|E(S, \overline{S})|}{D|S|} = \phi(S).\end{aligned}$$

1.5 Spectral gaps

The *Courant–Fischer characterization* gives

$$\lambda_2 = \min_{0 \neq x \perp \mathbf{1}} \frac{x^{\top} L_G x}{\|x\|_2^2}.$$

Thus λ_2 is the minimum Rayleigh quotient of the Laplacian energy among nonzero mean-zero vectors. Equivalently, every nonzero $x \perp \mathbf{1}$ satisfies

$$\lambda_2 \leq \frac{x^{\top} L_G x}{\|x\|_2^2}.$$

The comparison with cuts has two directions:

1. a large λ_2 lower-bounds the conductance of every admissible set;
2. a small λ_2 produces an admissible set of small conductance.

The first direction uses a centered cut vector; the second rounds an eigenvector into a cut.

1. The raw indicator is not admissible because

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

   Define the centered cut vector

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

   Its coordinates sum to zero, so $v_S \perp \mathbf{1}$. The vector is constant on each side of the cut and changes by $|S| + |\overline{S}| = n$ across every boundary edge. Hence

   $$v_S^{\top} L_G v_S = n^2 |E(S, \overline{S})|.$$

   Its squared norm is

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

Since v_S is admissible in the variational characterization,

$$\begin{aligned}\lambda_2 &\le \frac{v_S^\top L_G v_S}{\|v_S\|_2^2} = \frac{n|E(S,\overline{S})|}{|S|\,|\overline{S}|} \\ &= n\psi(S) = \frac{nD}{|\overline{S}|}\phi(S) \le 2D\phi(S),\end{aligned}$$

where the final inequality uses $|\overline{S}| \ge n/2$. Thus every nonempty S with $|S| \le n/2$ satisfies

$$\phi(S) \ge \frac{\lambda_2}{2D}. \tag{4}$$

Taking the minimum over admissible sets gives $\phi_G \ge \gamma/2$.

2. Let $u \ne 0$ satisfy $L_G u = \lambda_2 u$ and $u \perp \mathbf{1}$. The constructive direction follows from two lemmas.

Lemma 6 (Preprocessing)**.** *Let $u \perp \mathbf{1}$ satisfy $L_G u = \lambda_2 u$. There exists a nonzero vector $y \ge 0$ such that $|\operatorname{supp}(y)| \le n/2$ and*

$$\frac{y^\top L_G y}{\|y\|_2^2} \le \lambda_2.$$

Proof. Since $u \perp \mathbf{1}$ and $u \ne 0$, the vector u has both positive and negative coordinates. At least one of the two sign supports has size at most $n/2$. Replace u by $-u$ if necessary so that $|\{i : u_i > 0\}| \le n/2$, and define $y_i := \max\{u_i, 0\}$. Then

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

For $y_i > 0$, $y_i = u_i$, while $y_j \ge u_j$ for every j. Since all adjacency entries are nonnegative,

$$\begin{aligned}(L_G y)_i &= D y_i - \sum_j A_{ij} y_j \le D u_i - \sum_j A_{ij} u_j \\ &= (L_G u)_i = \lambda_2 u_i = \lambda_2 y_i.\end{aligned}$$

Multiplying by $y_i > 0$ and summing over the positive support gives

$$y^\top L_G y = \sum_{i:y_i>0} y_i (L_G y)_i \le \lambda_2 \sum_{i:y_i>0} y_i^2 = \lambda_2 \|y\|_2^2.$$

■

Rescale y by its positive maximum coordinate so that $0 \le y_i \le 1$. This does not change its Rayleigh quotient because both numerator and denominator scale quadratically.

For $t \in [0,1]$, define $S_t := \{i : y_i^2 > t\}$ and $w(t) := \mathbf{1}_{S_t}$.

Lemma 7 (Thresholding)**.** *Let $y \ge 0$ be nonzero and satisfy $|\operatorname{supp}(y)| \le n/2$. Then some nonempty threshold set $S_t \subseteq \operatorname{supp}(y)$ satisfies*

$$\phi(S_t) \le \sqrt{\frac{2y^\top L_G y}{D\|y\|_2^2}}.$$

Proof. Choose t uniformly in $[0,1]$. Since $0 \le y_i^2 \le 1$, the event $i \in S_t$ is the interval $0 \le t < y_i^2$, whose probability is y_i^2. Hence, by linearity of expectation,

$$\mathbb{E}\,|S_t| = \sum_{i=1}^{n} \mathbb{P}[i \in S_t] = \sum_{i=1}^{n} y_i^2 = \|y\|_2^2.$$

Since $w(t)$ is an indicator vector, $\|w(t)\|_2^2 = |S_t|$. Therefore

$$\mathbb{E}[D\|w(t)\|_2^2] = D\|y\|_2^2.$$

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

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

exactly when t lies between y_i^2 and y_j^2. The length of this interval is $|y_i^2 - y_j^2|$, so

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

where the last equality uses $y_i, y_j \ge 0$. Summing over edges and applying Cauchy–Schwarz gives

$$\begin{aligned}
\mathbb{E}[w(t)^\top L_G w(t)] &= \sum_{\{i,j\}\in E} |y_i - y_j|(y_i + y_j) \\
&\le \left(\sum_{\{i,j\}\in E} (y_i - y_j)^2\right)^{1/2} \left(\sum_{\{i,j\}\in E} (y_i + y_j)^2\right)^{1/2} \\
&\le \sqrt{y^\top L_G y}\left(2\sum_{\{i,j\}\in E} (y_i^2 + y_j^2)\right)^{1/2} \\
&= \sqrt{2D\, y^\top L_G y}\,\|y\|_2.
\end{aligned}$$

The second inequality uses $(a+b)^2 \le 2a^2 + 2b^2$. The last equality uses D-regularity: each y_i^2 is counted once for every incident edge, hence exactly D times. Define

$$T := \sqrt{\frac{2y^\top L_G y}{D\|y\|_2^2}}.$$

The two expectation bounds now yield

$$\mathbb{E}[w(t)^\top L_G w(t)] \le TD\|y\|_2^2 = T\,\mathbb{E}[D\|w(t)\|_2^2].$$

Suppose, for contradiction, that every nonempty threshold set satisfies

$$w(t)^\top L_G w(t) > TD\|w(t)\|_2^2.$$

Empty threshold sets contribute zero to both sides. Since $y \ne 0$, S_t is nonempty throughout the interval $0 \le t < \max_i y_i^2$, which has positive length. Averaging the strict inequality over this interval would contradict the preceding expectation bound. Hence some nonempty threshold satisfies

$$w(t)^\top L_G w(t) \le TD\|w(t)\|_2^2.$$

For this threshold, $S_t \subseteq \mathrm{supp}(y)$, so $|S_t| \le n/2$. Moreover, $w(t) = \mathbf{1}_{S_t}$ implies $w(t)^\top L_G w(t) = |E(S_t, \overline{S_t})|$ and $\|w(t)\|_2^2 = |S_t|$. Therefore

$$\phi(S_t) = \frac{w(t)^\top L_G w(t)}{D\|w(t)\|_2^2} \le T.$$

■

Observation 8. *Since S_t is nonempty and has size at most $n/2$, the thresholding lemma produces a valid candidate for the minimum defining ϕ_G.*

The following inequality was first established by Cheeger [1] in the setting of Riemannian manifolds. Its discrete counterpart for graphs was subsequently developed by Dodziuk [2] and Alon and Milman [3]. Since then, Cheeger-type inequalities have been extensively studied in theoretical computer science; see, in particular, [4] for an improved version of the inequality and [5] for a higher-order generalization.

Theorem 9 (Cheeger's Inequality for D-regular graphs)**.** *Let $G = (V, E)$ and let L_G be its combinatorial Laplacian. Then*

$$\frac{\lambda_2}{2D} \le \phi_G \le \sqrt{\frac{2\lambda_2}{D}}.$$

Proof. The lower bound follows from (4) after minimizing over all admissible sets. For the upper bound, the Preprocessing Lemma (6) gives

$$\frac{y^\top L_G y}{\|y\|_2^2} \le \lambda_2,$$

and the Thresholding Lemma (7) gives an admissible set S_t such that

$$\phi(S_t) \le \sqrt{\frac{2y^\top L_G y}{D\|y\|_2^2}} \le \sqrt{\frac{2\lambda_2}{D}}.$$

Since $\phi_G \le \phi(S_t)$, the upper bound follows.

■

In terms of the normalized gap $\gamma = \lambda_2/D = 1 - \mu_2$, Cheeger's inequality is

$$\frac{\gamma}{2} \le \phi_G \le \sqrt{2\gamma}.$$

1.6 Three phenomena describe the same obstruction

A bottleneck simultaneously creates

- a sparse cut;
- a slowly varying eigenvector;

- a slowly mixing walk.

For the normalized spectral gap $\gamma = \lambda_2/D$, Cheeger's inequality is equivalently written as

$$\frac{\phi_G^2}{2} \leq \gamma \leq 2\phi_G.$$

- **Constant spectral gap ⇒ constant edge expansion.** If $\gamma \geq c > 0$, then $\phi_G \geq \gamma/2 \geq c/2$. Since ϕ_G is the minimum conductance, every nonempty S with $|S| \leq n/2$ satisfies $\phi(S) \geq c/2$.
- **Constant edge expansion ⇒ constant spectral gap.** If $\phi_G \geq c > 0$, then $\gamma \geq \phi_G^2/2 \geq c^2/2$. This implication incurs a quadratic loss.
- **A very sparse cut forces a small spectral gap.** If some admissible set S has small conductance, then $\phi_G \leq \phi(S)$ is small and $\gamma \leq 2\phi_G$ is also small.
- **A very small spectral gap yields a sparse threshold cut.** If γ is small, then $\phi_G \leq \sqrt{2\gamma}$. The thresholding argument in Cheeger's inequality constructs such a cut from a nonnegative vector obtained from a second Laplacian eigenvector.

Consider a D-regular graph formed by two dense regions A and B, with b edge occurrences crossing between them, and assume $|A| \leq n/2$.

- There are $D|A|$ incident edge choices at vertices of A, and exactly b of them cross to B. Hence

  $$\phi(A) = \frac{b}{D|A|}.$$

  If b is small relative to $D|A|$, then A defines a bottleneck and $\gamma \leq 2\phi_G \leq 2b/(D|A|)$.
- A walk whose initial vertex is uniform on A crosses to B in one step with probability $b/(D|A|) = \phi(A)$. It therefore escapes slowly when this ratio is small.
- For the centered cut vector $x = \mathbf{1}_A - (|A|/n)\mathbf{1}$,

  $$x^\top L_G x = b, \qquad \|x\|_2^2 = |A|\left(1 - \frac{|A|}{n}\right).$$

  Thus the cut produces a low-energy vector orthogonal to $\mathbf{1}$, while thresholding a suitable low-energy vector produces a sparse cut.

Observation 10. *The inequality $\gamma/2 \leq \phi_G$ converts a spectral-gap lower bound into an expansion lower bound and, contrapositively, converts a sparse cut into a small spectral gap. The inequality $\phi_G \leq \sqrt{2\gamma}$ converts an expansion lower bound into a spectral-gap lower bound and, through thresholding, converts a small spectral gap into a sparse cut. Expansion and spectral gap are quantitatively related, but they are not numerically identical.*

The evolution of a random walk has the same spectral mechanism as power iteration. Let $u_1 = n^{-1/2}\mathbf{1}$ and $\pi = n^{-1}\mathbf{1}$. Since every probability distribution p_0 satisfies $u_1^\top p_0 = 1/\sqrt{n}$, it has a unique decomposition

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

Consequently,

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

For the ordinary walk, convergence is controlled by $\max_{i\geq 2} |\mu_i|$, not only by μ_2. In particular, a connected bipartite graph has $\mu_n = -1$, and the ordinary walk is periodic.

The *lazy* transition matrix is

$$\widetilde{P} = \frac{I + P}{2}.$$

At each step, the walk remains at its current vertex with probability $1/2$ and performs an ordinary random-walk step with probability $1/2$. The vectors u_i remain eigenvectors, and the corresponding eigenvalues are

$$\widetilde{\mu}_i = \frac{1 + \mu_i}{2} \in [0, 1], \qquad \widetilde{\mu}_2 = 1 - \frac{\gamma}{2}.$$

Claim 11. *For the lazy walk, a positive spectral gap yields exponential convergence to the uniform distribution.*

Because the eigenvectors are orthonormal and $0 \leq \widetilde{\mu}_i \leq \widetilde{\mu}_2$ for every $i \geq 2$,

$$\begin{aligned} \|\widetilde{P}^t p_0 - \pi\|_2^2 &= \sum_{i=2}^{n} \alpha_i^2 \widetilde{\mu}_i^{2t} \\ &\leq \left(1 - \frac{\gamma}{2}\right)^{2t} \sum_{i=2}^{n} \alpha_i^2. \end{aligned}$$

Taking square roots and using $1 - x \leq e^{-x}$ gives

$$\|\widetilde{P}^t p_0 - \pi\|_2 \leq e^{-\gamma t} \|p_0 - \pi\|_2.$$

For probability distributions p and q, total variation distance is

$$\|p - q\|_{\mathrm{TV}} = \frac{1}{2} \|p - q\|_1.$$

The Cauchy–Schwarz inequality gives $\|z\|_1 \leq \sqrt{n} \|z\|_2$. Moreover, for every probability distribution p_0,

$$\|p_0 - \pi\|_2^2 = \sum_{i=1}^{n} p_{0,i}^2 - \frac{1}{n} \leq 1 - \frac{1}{n} < 1.$$

Therefore,

$$\|\widetilde{P}^t p_0 - \pi\|_{\mathrm{TV}} \leq \frac{\sqrt{n}}{2} e^{-\gamma t}.$$

If

$$t_{\mathrm{mix}}(\varepsilon) := \min \left\{ t : \max_{j \in V} \|\widetilde{P}^t e_j - \pi\|_{\mathrm{TV}} \leq \varepsilon \right\},$$

then, for $\gamma > 0$ and $0 < \varepsilon < 1/2$,

$$t_{\mathrm{mix}}(\varepsilon) \leq \left\lceil \frac{2}{\gamma} \log\left(\frac{\sqrt{n}}{2\varepsilon} \right) \right\rceil = O\left(\frac{\log n + \log(1/\varepsilon)}{\gamma} \right).$$

If $\gamma = 0$, the graph is disconnected and convergence from every initial vertex to the uniform distribution does not hold.

■

Claim 12. *A sparse cut prevents rapid mixing.*

Let S satisfy $|S| \le n/2$ and $\phi(S) = \delta$, and let the ordinary random walk $X_0, X_1, \dots$ start from the uniform distribution on S. If $\delta = 0$, the walk cannot leave S and therefore cannot converge to the uniform distribution on V. Assume henceforth that $\delta > 0$. To control the escape probability, first consider an unconditioned stationary ordinary walk $Y_0, Y_1, \dots$ with Y_0 uniform on V. At every time r, stationarity gives

$$\begin{aligned}\mathbb{P}[Y_r \in S, Y_{r+1} \notin S] &= \sum_{i \in S} \mathbb{P}[Y_r = i]\, \mathbb{P}[Y_{r+1} \notin S \mid Y_r = i] \\ &= \sum_{i \in S} \frac{1}{n} \frac{|\{\text{edges from } i \text{ to } \overline{S}\}|}{D} \\ &= \frac{|E(S, \overline{S})|}{Dn} = \frac{|S|}{n}\phi(S) = \frac{|S|}{n}\delta\end{aligned}$$

Conditioning the entire process on $Y_0 \in S$ gives the same law as the process X. Hence, for $F_r = \{X_r \in S,\, X_{r+1} \notin S\}$,

$$\mathbb{P}[F_r] \le \frac{\mathbb{P}[Y_r \in S,\, Y_{r+1} \notin S]}{\mathbb{P}[Y_0 \in S]} = \frac{(|S|/n)\delta}{|S|/n} = \delta.$$

If the walk leaves S during its first t steps, then at least one of the events $F_0, \dots, F_{t-1}$ occurs. The union bound yields

$$\mathbb{P}[X \text{ leaves } S \text{ during its first } t \text{ steps}] \le \sum_{r=0}^{t-1} \mathbb{P}[F_r] \le t\delta.$$

If $t \le 1/(4\delta)$, then $\mathbb{P}[X_t \in S] \ge 1 - t\delta \ge 3/4$, whereas $\pi(S) = |S|/n \le 1/2$. Consequently,

$$\| \operatorname{Law}(X_t) - \pi \|_{\mathrm{TV}} \ge \mathbb{P}[X_t \in S] - \pi(S) \ge \frac{1}{4}.$$

The law of X_t is the average of the laws obtained from deterministic starting vertices in S. By convexity of total variation distance, at least one such starting vertex is at total variation distance at least $1/4$ from π. Thus the ordinary walk cannot mix in fewer than a constant multiple of $1/\delta$ steps. For the lazy walk, a crossing is attempted with probability $1/2$, so its one-step escape probability is $\delta/2$. Applying the same argument to a set attaining ϕ_G gives, when $\phi_G > 0$,

$$t_{\mathrm{mix}}(1/4) = \Omega\left(\frac{1}{\phi_G}\right).$$

Connectedness, positive conductance, a positive normalized spectral gap, and convergence of the lazy walk to the uniform distribution are qualitatively equivalent. Quantitatively, Cheeger's inequality and the mixing bounds show that sparse cuts, small spectral gaps, and slow mixing are manifestations of the same bottleneck, although ϕ_G, γ, and t_{mix} are distinct parameters.

AI use disclosure LLM were used to provide additional insights on both our personal notes and slides, to polish the english grammar and to generate Tikz pictures.

References

[1] Jeff Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. *Problems in Analysis*, 195–199, 1970.

[2] Józef Dodziuk. Difference equations, isoperimetric inequality and transience of certain random walks. *Transactions of the American Mathematical Society*, 284(2):787–794, 1984.

[3] Noga Alon, Vitali Davidovich Milman, λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.

[4] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, Luca Trevisan. Improved Cheeger's Inequality: Analysis of Spectral Partitioning Algorithms through Higher Order Spectral Gap. *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, 11–20, 2013.

[5] James R Lee, Shayan Oveis Gharan, Luca Trevisan, Multi-way Spectral Partitioning and Higher-Order Cheeger Inequalities *Journal of the ACM*, 61(6), Article 37, 2014.