1 Spectral Graph Theory

We assume the graph has n vertices and m edges. The vertex set is *V* and the edge set is *E*.

1.1 Courant-Fischer Theorem

1. **Eigenvalue version**: Let A be a symmetric matrix with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$, then

$$\lambda_i = \min_{\substack{\text{subspace } W \subseteq \mathbb{R}^n \\ \dim(W) = i}} \max_{\substack{x \in W \\ x \neq 0}} \frac{x^\top A x}{x^\top x} = \max_{\substack{\text{subspace } W \subseteq \mathbb{R}^n \\ \dim(W) = n+1-i}} \min_{\substack{x \in W \\ x \neq 0}} \frac{x^\top A x}{x^\top x}.$$

2. **Eigenbasis version**: Let A be a symmetric matrix with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ and corresponding orthonormal eigenvectors x_1, \ldots, x_n , then

$$\lambda_i = \min_{\substack{x \perp x_1, \dots x_{i-1} \\ x \neq \mathbf{0}}} \frac{x^\top A x}{x^\top x} = \max_{\substack{x \perp x_{i+1}, \dots x_n \\ x \neq \mathbf{0}}} \frac{x^\top A x}{x^\top x}.$$

Note that we also have $\lambda_i = \frac{x_i^\top A x_i}{x_i^\top x_i}$.

Applying Courant-Fischer theorem, we have $\lambda_2 x^{\top} x \leq x^{\top} L x \leq x^{\top} L$ $\lambda_n x^{\top} x$ for all $x \perp 1$, as 1 is the eigenvector of L corresponding to eigenvalue 0. For connected graphs, $\lambda_2 > 0$.

1.2 PSD Order (Loewner Order)

Defined only for symmetric matrices: $A \leq B$ iff for all $x \in \mathbb{R}^n$, we have $x^{\top}Ax \leq x^{\top}Bx$. We also define $G \leq H$ for two graphs G and H iff $L_G \leq L_H$. We always have $G \geq H$ if H is a subgraph

Properties:

- 1. If $A \leq B$ and $B \leq C$, then $A \leq C$.
- 2. If $A \leq B$, then $A + C \leq B + C$ for any symmetric C.
- 3. If $A \leq B$ and $C \leq D$, then $A + C \leq B + D$.
- 4. If A > 0 and $\alpha \ge 1$, then $\frac{1}{\alpha}A \le A \le \alpha A$.
- 5. If $A \leq B$, then $\lambda_i(A) \leq \lambda_i(B)$ for all i. Proof by Courant-Fischer theorem. The converse is not true.
- 6. For any matrix C, if $A \leq B$, then $C^{\top}AC \leq C^{\top}BC$.
- 7. If $0 \le A \le B$, then $B^{-1} \le A^{-1}$.

1.3 Bounding the λ_2 and λ_n

1.3.1 Test Vector Method

Since $\lambda_2 \leq \frac{y^+ L y}{v^+ v}$ for any $y \perp 1$, we can upper bound the λ_2 by any **test vector** y. Similarly, we can lower bound the λ_n by test vectors by $\lambda_n \geq \frac{y^+ L y}{v^\top v}$

- 1. For a complete graph K_n , $L = nI 11^{\top}$ and for any $x \perp 1$ we have $x^{\top}Lx = nx^{\top}x$. Therefore, $\lambda_2(K_n) = \cdots = \lambda_n(K_n) = n$ and any $x \perp 1$ is an eigenvector.
- 2. For a path graph P_n , let x(i) = n + 1 2i be the test vector which satisfies $x \perp 1$, we get $\lambda_2(P_n) \leq \frac{12}{n^2}$. Let x(1) = -1, x(n) = 1 and x(i) = 0 for other i to be the test vector, we get $\lambda_n(P_n) \geq 1$.
- 3. For a complete binary tree T_n (depth equals zero for a single root), let x(i) = 0 for all non-leaf nodes, x(i) = -1 for

even-numbered leaf nodes and x(i) = 1 for odd-numbered leaf nodes be the test vector, we get $\lambda_n(T_n) \ge 1$. Let x(1) = 0, x(i) = 1 for the left subtree of the root and x(i) = -1 for the right subtree of the root be the test vector, we get 2. Conductance of a graph: The conductance $\phi(G) :=$ $\lambda_2(T_n) \leq \frac{2}{n-1}$.

1.3.2 Consequences of PSD Order

Since $x^{\top}(D-A)x = \sum_{(u,v)} w(u,v)(x(u)-x(v))^2 \ge 0$ and $x^{\top}(D+V)$ $A(x) = \sum_{(u,v)} w(u,v)(x(u) + x(v))^2 \ge 0$, we have $D \ge A$ and $D \geq -A$. In addition, we have $D \leq (\max D_{i,i})I$. Therefore, we have $L = D - A \le 2D \le (2 \max D_{i,i})I$, which implies $\lambda_n \leq 2 \max D_{i,i}$ for any graph. For unit-weight graphs, this means $\lambda_n \leq 2 \max \text{degree}(v)$. The bound is tight for a singleedge graph.

To get lower bounds of $\lambda_2(H)$, we first establish $f(n)H \geq G$ for some *G* with known lower bounds on $\lambda_2(G)$. Usually $G = K_n$ because $\lambda_2(K_n) = n$. Then it follows that $\lambda_2(H) \ge \lambda_2(G)/f(n)$.

- 1. **Path Graph** P_n : Let $G_{i,j}$ denote a unit-weight graph consisting of one edge (i, j) and P_n be the path graph connecting 1 and n. Then $(n-1)P_n \geq G_{1,n}$. Proof follows from applying Cauchy-Schwartz for $\delta_i := x(i+1) - x(i)$. For weighted paths, we have $G_{1,n} \leq \left(\sum_{i=1}^{n-1} \frac{1}{w_i}\right) \sum_{i=1}^{n-1} w_i G_{i,i+1}$.
 - Applying path inequality, we have $K_n = \sum_{i < j} G_{i,j} \leq \sum_{i < j} (j 1)$ $i)P_{i,j} \leq \sum_{i < j} (j-i)P_n \leq n^3 P_n$, which implies $\lambda_2(P_n) \geq$ $\lambda_2(K_n)/n^3 = 1/n^2$.
- 2. Any unit-weight graph G: Define the diameter D of a that $\frac{1_S^T L I_S}{1_S^T D I_S} \le \sqrt{2 \frac{z^T L Z}{z^T D Z}}$ graph *G* to be the maximum length of the shortest paths between any two nodes. Let $G_{i,j}^s$ be the shortest path from *i* to *j*. Applying path inequality, we have $K_n = \sum_{i < j} G_{i,j} \le 1$ $\sum_{i < j} DG_{i,j}^s \leq \sum_{i < j} DG \leq n^2 DG$, which implies $\lambda_2(G) \geq \frac{1}{nD}$.
- 3. Complete Binary Tree T_n : Define G_e be the single-edge graph with edge e, and $T_{i,j}$ be the unique path between i and j. Applying the weighted path inequality, we have $K_n = \sum_{i < j} G_{i,j} \le \sum_{i < j} \left(\left(\sum_{e \in T^{i,j}} \frac{1}{w(e)} \right) \left(\sum_{e \in T^{i,j}} w(e) G_e \right) \right) \le$ $(\max_{i < j} \sum_{e \in T^{i,j}} \frac{1}{w(e)}) (\sum_{i < j} \sum_{e \in T^{i,j}} w(e) G_e)$. For *e* connecting level i and i + 1 for $i \in [d - 1]$, we set $w(e) = 2^i$. Then $\max_{i < j} \sum_{e \in T^{i,j}} \frac{1}{w(e)} \le 4$. Since the number of occurrence of e in $T^{i,j}$ for any i < j is upper bounded by $n^2 2^{-i}$, we have $\sum_{i < i} \sum_{e \in T^{i,j}} w(e) G_e \le \sum_e w(e) n^2 2^{-i} G_e = \sum_e n^2 G_e = n^2 T_n$. Therefore, $K_n \leq 4n^2T_n$, which implies $\lambda_2(T_n) \geq \frac{1}{4n}$.

2 Conductance

Definitions:

1. Conductance of a vertex subset: Given $\emptyset \subset S \subset V$, the conductance $\phi(S) := \phi(S) = \frac{|E(S, V \setminus S)|}{\min\{\mathbf{vol}(S), \mathbf{vol}(V \setminus S)\}}$, where $\mathbf{vol}(S) :=$ $\sum_{v \in S} \text{degree}(v)$. Define 1_S to be the *n*-dimensional vector with only 1 for the vertices of S and 0 for the vertices of $V \setminus S$. Assuming $vol(S) \le vol(V)/2$, thus $|E(V, V \setminus S)| = 1$

$$\sum_{(u,v)\in E} (\mathbf{1}_{S}(u) - \mathbf{1}_{S}(v))^{2} = \mathbf{1}_{S}^{\top} L \mathbf{1}_{S} \text{ and } \mathbf{vol}(S) = \mathbf{1}_{S}^{\top} D \mathbf{1}_{S}. \text{ Then } \phi(S) = \frac{\mathbf{1}_{S}^{\top} L \mathbf{1}_{S}}{\mathbf{1}_{S}^{\top} D \mathbf{1}_{S}}.$$

- $\min_{\emptyset \subset S \subset V} \phi(S) = \min_{\emptyset \subset S \subset V}$
- 3. ϕ -expander: For any $\phi \in (0,1]$, we call a graph G to be a ϕ -expander if $\phi(G) \geq \phi$.
- 4. ϕ -expander decomposition of quality q: A partition $\{X_i\}$ of the vertex set V is called a ϕ -expander decomposition of quality q if (1) each induced graph $G[X_i]$ is a ϕ -expander, and (2.i) #edges not contained in any $G[X_i]$ is at most $q \cdot \phi \cdot m$. The second condition is equivalent to (2.ii) The partition removes at most $q \cdot \phi \cdot m$ edges.
- 5. **Normalized Laplacian**: We define the *normalized Laplacian* to be $N := D^{-1/2}LD^{-1/2}$. N is still PSD, with first eigenvalue equals 0 associated with eigenvector $D^{1/2}$ 1. By Courant-Fischer theorem, $\lambda_2(N) = \min_{x+D^{1/2}1} \frac{x^\top N x}{x^\top x} = \min_{z \perp d} \frac{z^\top L z}{z^\top D z}$.

2.1 Cheeger's Inequality

Notice that the $\lambda_2(N)$ has similar forms to $\phi(G)$. Cheeger's inequality aims to bound $\phi(G)$ by $\lambda_2(N)$.

Cheeger's Inequality: $\frac{\lambda_2(N)}{2} \le \phi(G) \le \sqrt{2\lambda_2(N)}$.

The lower bound is proved by restricting the minimum in $\lambda_2(N)$ to be $z_S = 1_S - \alpha 1$ for some α such that $z_S \perp d$. The upper bound is proved by constructing S for any $z \perp d$ such

3 Random Walks on a Graph

A random walk on a graph G is a Markov Chain with transition probability $\mathbb{P}(v_{t+1} = v \mid v_t = u) = w(u, v)/d(u)$ iff $(u, v) \in E$ and 0 otherwise. The transition matrix is thus $W = AD^{-1} =$ $I - D^{1/2}ND^{-1/2}$ and $p_t = W^t p_0$. Define $\pi = \frac{d}{\mathbf{1}^T d}$, thus $\pi = W \pi$ for any *G*, so every *G* has a stationary distribution.

3.1 Lazy Random Walks

A **lazy random walk on a graph** *G* is a random walk, but has half probability to not move for every step. Assuming that G is connected, the lazy random walk guarantees ergodicity of the Markov Chain, and thus convergence to the stationary distribution. The transition matrix is $\tilde{W} = \frac{1}{2}(I + W) =$ $I - \frac{1}{2}D^{1/2}ND^{-1/2}$.

Relation between lazy random walk and normalized Laplacian: For the *i*-th eigenvalue v_i of N associated with eigenvector ψ_i , the \tilde{W} has an eigenvalue $1 - \frac{1}{2}\nu_i$ associated with eigenvector $D^{1/2}\psi_i$. Since $0 \le L \le 2D$, we have $0 \le N \le 2I$ and thus $0 \le \lambda_i(N) \le 2$. Therefore, we conclude that all eigenvalues of $\tilde{W} \in [0,1]$.

Dynamics of lazy random walk: Expanding the starting distribution p_0 by the eigenvectors of \tilde{W} , we have for some $\{\alpha_i\}$ that $p_0 = \sum_{i=1}^n \alpha_i D^{1/2} \psi_i$. Therefore, we have $p_t =$ $\tilde{W}^t p_0 = \sum_{i=1}^n \alpha_i (1 - \frac{1}{2} \nu_i)^t D^{1/2} \psi_i \rightarrow \alpha_1 D^{1/2} \psi_1$ as $\nu_1 = 0$ and

 $v_i > 0$ for $i \neq 1$. Since $\psi_1 \propto D^{1/2} \mathbf{1}$, we have $\psi_1 = \frac{d^{1/2}}{(\mathbf{1}^T d)^{1/2}}$, thus $\alpha_1 = \psi_1^{\top} D^{-1/2} p_0 = \frac{1^{\top} p_0}{(1^{\top} d)^{1/2}} = \frac{1}{(1^{\top} d)^{1/2}}$ and $\alpha_1 D^{1/2} \psi_1 = \pi$, which implies $p_t \to \pi$, the stationary distribution.

Rate of Convergence: For any unit-weight connected graph G and any starting distribution p_0 , we have $||p_t - \pi||_{\infty} \le e^{-\nu_2 t/2} \sqrt{n}$. Therefore, a larger v_2 and smaller vertex set means faster convergence, and the convergence rate is exponential.

3.2 Hitting Time

The expected hitting time from a to s is defined by $\mathbb{E}H_{a,s}$, where $H_{a,s} = \operatorname{argmin}_t \{ v_t = s \mid v_0 = a \}$. We want $\mathbb{E}H_{a,s}$ for all vertices a and denote the vector as h, e.g., h(s) = 0.

By one-step analysis, we have $h(a) = 1 + \sum_{(a,b) \in E} \frac{w(a,b)}{d(a)} h(b) =$ $1 + \mathbf{1}_{a}^{\top} W^{\top} h$, and thus $1 = \mathbf{1}_{a}^{\top} (I - W^{\top}) h$. Combining the equation for all vertices except s, we have $1 - \alpha \mathbf{1}_s = (I - W^\top)h$, where α represents the extra freedom from the n-1 equations. Multiplying both side by *D*, we get $d - \alpha d(s)\mathbf{1}_s = (D - A)h = Lh$, which only have solution when $d - \alpha d(s) \mathbf{1}_s \perp \mathbf{1}$. Therefore, $\alpha = ||d||_1/d(s).$

ted hitting time from all vertices to s. Note that the solution has one extra freedom because dim(ker(L)) = 1, and the correct expected hitting time is h - h(s)1 to enforce the constraint that h(s) = 1. The equation can be solved in $\tilde{O}(m)$.

4 Pseudo-Inverse and Effective Resistance

Given a Laplacian L, its (Moore-Penrose) pseudo inverse is defined to be either of the two equivalents:

- 1. A matrix L^+ that is (1) symmetric, (2) $L^+v = 0$ for $v \in \ker(L)$, and (3) $L^+Lv = LL^+v = v$ for $v \in \ker(L)$.
- 2. Let λ_i, v_i be the *i*-th eigenvalue and eigenvector. Then $L^+ = \sum_{\lambda_i \neq 0} \lambda_i^{-1} v_i v_i^{\top}$.

Property:

- Assume $M = XYX^{\top}$, where X is real and invertible, and Y is real and symmetric. Let Π_M be the orthogonal projection to the image of *M*. Then $M^{+} = \Pi_{M}(X^{\top})^{-1}Y^{+}X^{-1}\Pi_{M}$.
- For symmetric L, $\Pi_L := \sum_{\lambda_i \neq 0} v_i v_i^{\top} = L^{+/2} L L^{+/2} = L^+ L = L L^+$ is the orthogonal projection to the image of L, i.e., $\Pi_L v = 0$ for any $v \in \ker(L)$ and $\Pi_v = v$ for any $v \in \operatorname{im}(L)$. For connected *G*, $\Pi_L = I - \frac{1}{n} \mathbf{1} \mathbf{1}^{\top}$.

The effective resistance between vertex *a* and *b* is defined to be the cost (energy lost) to routing one unit (of positive electric charge) from a to b: $R_{\text{eff}}(a,b) = \min_{Bf=1,-1,a} f^{\top}Rf = \tilde{f}^{\top}R\tilde{f}$, where f is the electric flow. Let \tilde{x} be the electric voltages, we also have $L\tilde{x} = \mathbf{1}_b - \mathbf{1}_a$, and thus $R_{\text{eff}}(a,b) = \tilde{x}^{\top} L\tilde{x} =$ $(\mathbf{1}_b - \mathbf{1}_a)^{\mathsf{T}} L^+ (\mathbf{1}_b - \mathbf{1}_a) = ||L^{+/2} (\mathbf{1}_b - \mathbf{1}_a)||_2^2.$

Effective Resistance is a distance defined on the vertex pairs, i.e. $R_{\text{eff}}(a,c) \leq R_{\text{eff}}(a,b) + R_{\text{eff}}(b,c)$.

Gaussian Elimination for Laplacian

5.1 Optimization View

Solving Lx = d is equivalent to solving $\underset{\sim}{\operatorname{argmin}}_{x} - d^{\top}x + \frac{1}{2}x^{\top}Lx$. By iteratively optimize over x_i , we get a series of similar optimizations. The final optimization is straightforward, then we can back substitute to get x.

5.2 Additive View

Given an invertible square lower/upper triangular matrix M, we can solve Mx = d by back substitution in $O(\mathbf{nnz}(M))$, where nnz(M) means the number of non-zeros in M. Therefore, if we know the **Cholesky decomposition** $L = MM^{T}$ (requires $O(n^3)$), then we can solve $Lx = d = M(M^Tx)$ by (1) solving My = d then (2) solving $M^{\top}x = y$ in $O(\mathbf{nnz}(M))$.

However, the Laplacian is non-invertible, leading to one diagonal of M equals 0. Therefore, we need to play a trick. Define \hat{M} equals M but has value 1 for the zero diagonal, and \hat{D} be a diagonal matrix that has value 0 at the zero diagonal of M and 1 otherwise. Then $L = \hat{M}\hat{D}\hat{M}^{T}$, and each \hat{M} is now invertible. Since $\hat{D}^+ = \hat{D}$, we can find a special solution of Lx = dby (1) solving $\hat{M}z = d$, (2) computing $y = \hat{D}z$, and (3) solving To summarize, by solving $Lh = d - ||d||_1 1$, we can get the expec- $\hat{M}^{\top} x = y$. The solution space is obtained by adding a subspace spanned by 1.

Concentration of Random Matrices

- 1. Chernoff Bound for Bounded independent variables: Suppose $\{X_i \in \mathbb{R}\}$ are independent random variables and $0 \le X_i \le R$. Let $X = \sum_i X_i$ and $\mu = \mathbb{E}X$. Then for any $0 < \epsilon \le 1$, we have $\mathbb{P}(X \ge (1 + \epsilon)\mu) \le \exp(-\frac{\epsilon^2 \mu}{4R})$ and $\mathbb{P}(X \le (1 - \epsilon)\mu) \le \exp(-\frac{\epsilon \mu}{4R}).$
- Bernstein Bound for independent, zero-mean and boun**ded variables:** Suppose $\{\bar{X}_i \in \mathbb{R}\}$ are independent, zeromean random variables and $|X_i| \le R$. Let $X = \sum_i X_i$, and $\sigma^2 = \mathbf{Var}(X)$. Then for any t > 0, we have $\mathbb{P}(|X| \ge t) \le$ $2\exp(\frac{-t^2}{2Rt+4\sigma^2})$.

The proof is similar to Chernoff bound. (1) $\mathbb{P}(X \geq t) =$ $\mathbb{P}(\exp(\theta X) \ge \exp(\theta t)) \le \exp(-\theta t)\mathbb{E}(\exp(\theta X))$, (2) upper bound $\mathbb{E}(\exp(\theta X)) \leq \exp(\theta^2 \sigma^2)$ given $\theta \in (0, \frac{1}{R}]$, which allows $\exp(\theta X_i) \le 1 + \theta X_i + (\theta X_i)^2$, and (3) take the minimum among $\theta \in (0, \frac{1}{R}]$.

Bernstein Bound for independent, zero-mean and boun**ded symmetric matrices**: Suppose $\{X_i \in \mathbb{R}^{n \times n}\}$ are independent, zero-mean, symmetric random matrices and $||X_i|| \le R$, where $||\cdot||$ is the spectral norm (the largest singular value). Let $X = \sum_i X_i$, and $\sigma^2 = \|\sum_{i=1}^n \mathbb{E} X_i^2\|$. Then $\mathbb{P}(||X|| \ge t) \le 2n \exp(\frac{-t^2}{2Rt + 4\sigma^2}).$

6.1 Matrix Functions

 $f(A) = V f(\Lambda) V^{\top}$. This is compatible to the Taylor expan- on of the constructed random graph.

sion $f(x) = \sum_i \alpha_i x^i$, as $f(A) = \sum_i \alpha A^i = V(\sum_i \alpha_i f(\Lambda)) V^{\top} =$ $V f(\Lambda) V^{\top}$.

Monotonicity: Given $f: \mathcal{D} \to \mathcal{C}$ and partial orders $\leq_{\mathcal{C}}$ and $\leq_{\mathcal{D}}$, we call f is monotonically increasing w.r.t. these orders iff for all $d_1 \leq_{\mathcal{D}} d_2 \in \mathcal{D}$ we have $f(d_1) \leq f(d_2)$. For matrix functions, we use the PSD order as the ordering.

Property:

- If the scalar function f is monotonically increasing, then the matrix function $X \to \text{Tr}(f(X))$ is monotonically increa-
- log(·) is monotonically increasing.
- The matrix function $(\cdot)^2$ and $\exp(\cdot)$ is **not** monotone.
- $\exp(A) \le I + A + A^2$ for $||A|| \le 1$.
- $\log(I + A) \leq A$ for A > -I.

6.2 Approximating a Dense Graph by Sparse Graphs

Given PD matrices A, B and $\epsilon > 0$, we say $A \approx_{\epsilon} B$ iff $\frac{1}{1+\epsilon}A \leq$ $B \leq (1+\epsilon)A$. If $L_G \approx_{\epsilon} L_{\tilde{G}}$ and $|\tilde{E}| \ll |E|$, we call \tilde{G} a spectral sparsifier of G.

Properties:

- Define $c_G(T) := \sum_{e \in E \cap (T \times V \setminus T)} w(e)$ to be the value of the cut $(T, V \setminus T)$. If $L_G \approx_{\epsilon} L_{\tilde{G}}$, then for all $T \subset V$, we have $\frac{1}{1+\epsilon}c_G(T) \le c_{\tilde{G}}(T) \le (1+\epsilon)c_G(T)$. The proof is by noticing $c_G(T) = 1_T^{\top} L_G 1_T$.
- $L \approx_{\epsilon} \tilde{L} \Leftrightarrow \Pi_{L} \approx_{\epsilon} L^{+/2} \tilde{L} L^{+/2}$, as $A \leq B$ implies $C^{\top}AC \leq C^{\top}BC$ for any $C \in \mathbb{R}^{n \times n}$.
- For $\epsilon \le 1$, if $\|\Pi_I L^{+/2} \tilde{L} L^{+/2}\| \le \epsilon/2$, then $\Pi_I \approx_{\epsilon} L^{+/2} \tilde{L} L^{+/2}$.

Theorem: Consider a connected graph G = (V, E, w), with n = |V|. For any $0 < \epsilon < 1$ and $0 < \delta < 1$, there exist sampling probabilities p_e for each edge $e \in E$ s.t. if we include each edge e in \tilde{E} independently with probabilty p_e and set its weight $\tilde{w}(e) = \frac{1}{n} w(e)$, then with probability at least $1 - \delta$ the graph Given a real-valued function $f: \mathbb{R} \to \mathbb{R}$ and a symmetric $\tilde{G} = (V, \tilde{E}, \tilde{w})$ satisfies $L_G \approx_{\varepsilon} L_{\tilde{G}}$ and $|\tilde{E}| \leq O(n\varepsilon^{-2}\log(n/\delta))$. matrix A with eigen-decomposition $A = V\Lambda V^{T}$, we define The proof uses Bernstein bounds to prove the concentrati-