Based on lecture notes by Dan Speilman, Madhur Tulsiani, etc.

Gershgorin Circle Theorem

Let $M \in \mathbb{C}^{n\times n}$, $R_i \stackrel{\text{def}}{=} \sum_{j\neq i}\left|M_{ij}\right|$ and $\mathrm{Disc}(a,b) \stackrel{\text{def}}{=} \{z\colon |z-a| \leq b\}$. Then the eigenvalues of M belong to $\cup_{i\in[n]} \mathrm{Disc}(M_{ii}, R_i)$.

- Let $x \in \mathbb{C}^n$ be an eigenvector of M with eigenvalue λ. Let $i \stackrel{\text{def}}{=} \mathrm{argmax}_{\mathrm{j}\in[n]}\left|x_j\right|$.
- $Mx = \lambda x$ implies $\sum_j M_{ij}x_j = \lambda x_i$. Therefore,

$$\lambda = \sum_j M_{ij}x_j/x_i = M_{ii} + \sum_{j\neq i} M_{ij}x_j/x_i$$

- Thus,

$$|\lambda - M_{ii}| = \left|\sum_{j\neq i} M_{ij}x_j/x_i\right| \leq \sum_{j\neq i}\left|M_{ij}\right|\left|\frac{x_j}{x_i}\right| \leq \sum_{j\neq i}\left|M_{ij}\right| = R_i$$

Applications of Gershgorin Circle Theorem

- Graphs of maximum degree d: all eigenvalues of adjacency matrix lie in $[-d, d]$.
- Graphs with self loops: Let G be a graph where vertex i has degree d_i, and also has a self-loop of weight d_i. All eigenvalues of its adjacency matrix lie in $\left[0, 2 \max_i d_i\right]$.

- Diagonally dominant matrices: A matrix M is diagonally dominant if $|M_{ii}| \geq \sum_{j \neq i} \left|M_{ij}\right|$. If M is symmetric and diagonally dominant, then $M \succcurlyeq 0$, i.e. M is positive semidefinite.

Cholesky Decomposition

- Theorem: $M \succcurlyeq 0$ iff there exists a matrix A such that $M = A^T A$.

- $M = V\Lambda V^T = V\Lambda^{1/2}\Lambda^{1/2}V^T = \left(\Lambda^{1/2}V^T\right)^T\left(\Lambda^{1/2}V^T\right)$.

- This decomposition is not unique: Let R be any rotation matrix, and let a, b be any vectors. Then $\langle Ra, Rb\rangle = \langle a, b\rangle$.

- Used in semidefinite programming (SDP) based approximation algorithms, etc.

Courant-Fischer Theorem

Rayleigh quotient $R_M(x) \stackrel{\text{def}}{=} \frac{x^T M x}{x^T x}$

$$R_M(v_i) = \sigma_i$$

- Theorem: For a symmetric matrix M with eigenvalues $\sigma_1 \geq \cdots \geq \sigma_n$ and corresponding eigenvectors $v_1, \ldots, v_n$ respectively,

$$v_1 = \operatorname{argmax}_x \frac{x^T M x}{x^T x} \quad \text{and} \quad v_k = \operatorname{argmax}_{x \perp v_1, \ldots v_{k-1}} \frac{x^T M x}{x^T x}$$

- Theorem:

$$\sigma_k = \max_{S:\text{rank}(S)=k} \min_{x \in S} \frac{x^T M x}{x^T x} = \min_{T:\text{rank}(T)=n-k+1} \max_{x \in T} \frac{x^T M x}{x^T x}$$

- Moreover, the optima is achieved when S is $\text{span}\{v_1, \ldots, v_k\}$ and T is $\text{span}\{v_k, \ldots, v_n\}$. (H.W.)

- Choosing $S = \text{span}\{v_1, \ldots, v_k\}$, for an $x \in S$ write $x = c_1 v_1 + \cdots + c_k v_k$

$$\frac{x^T M x}{x^T x} = \cdots = \frac{\sum_{i \in [k]} c_i^2 \sigma_i}{\sum_{i \in [k]} c_i^2} \geq \sigma_k$$

- Therefore, $\sigma_k \leq \max_{S:\text{rank}(S)=k} \min_{x \in S} \frac{x^T M x}{x^T x}$.

Courant-Fischer Theorem

- Let S be the optimal subspace of rank k and and let $T_k = \text{span}\{v_k, v_{k+1}, \dots, v_n\}$.
- Since $\text{rank}(S) + \text{rank}(T_k) = k + (n - k + 1) > n$, we have $S \cap T_k \neq \emptyset$

$$\min_{x \in \mathrm{S}} \frac{x^T M x}{x^T x} \leq \min_{x \in \mathrm{S} \cap T_k} \frac{x^T M x}{x^T x}$$

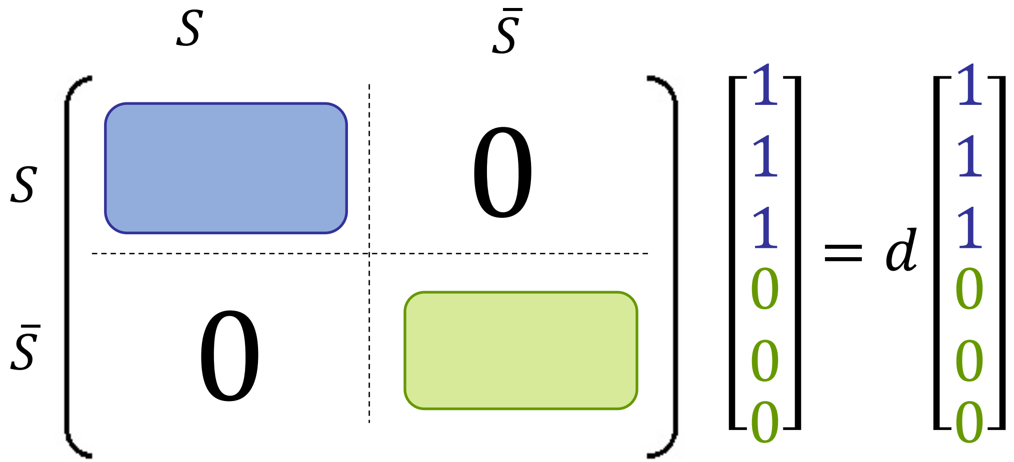
- Let $x \in S \cap T_k$ be the optimal vector above, write $x = c_k v_k + \cdots + c_n v_n$.

$$\frac{x^T M x}{x^T x} = \frac{\sum_{i=k}^{n} c_i^2 \sigma_i}{\sum_{i=k}^{n} c_i^2} \leq \sigma_k$$

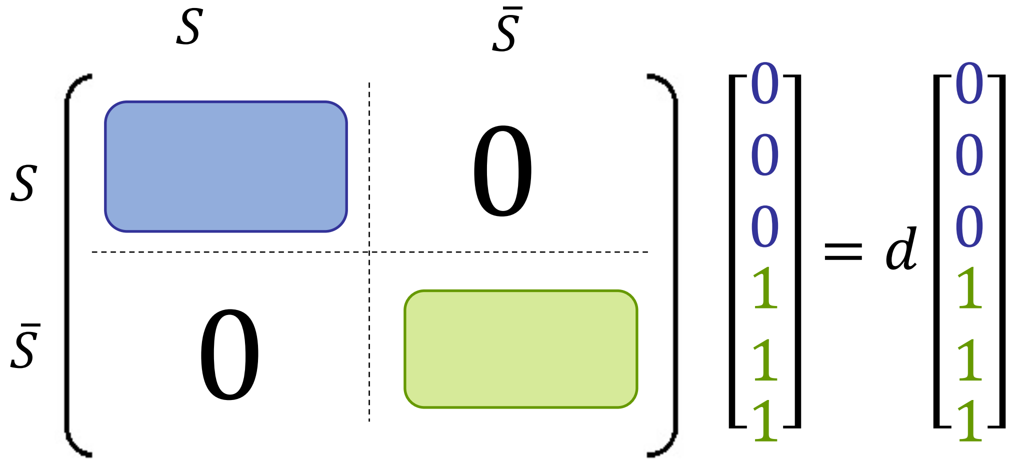
- Therefore, $\max_{S:\text{rank}(S)=k} \min_{x \in S} \frac{x^T M x}{x^T x} \leq \sigma_k$ and hence $\sigma_k = \max_{S:\text{rank}(S)=k} \min_{x \in S} \frac{x^T M x}{x^T x}$.

Graph Matrices

- For a d-regular graph, every row and column of A sums to d. Therefore, $A\mathbf{1} = d\mathbf{1}$.
- Therefore d is an eigenvalue of A with eigenvector $\mathbf{1}$.
- Since all its eigenvalues line in $[-d, d]$, d is its largest eigenvalue.



- Second largest eigenvalue = d iff graph is disconnected.
- Multiplicity of eigenvalue d = number of components in graph.



Non-regular graphs

- A_{ij}= weight of edge $\{i, j\}$. $d_i \stackrel{\text{def}}{=} \sum_{j \neq i} A_{ij}$.

- Normalized adjacency matrix: $D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$
- Eigenvalues lie in $[-1,1]$.

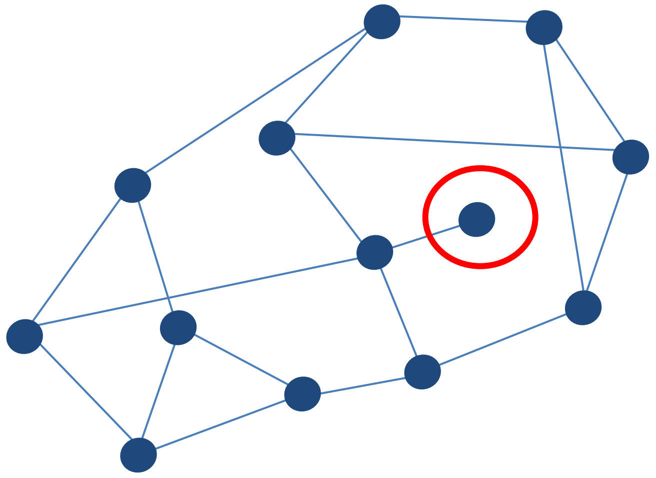
- Laplacian Matrix $L \stackrel{\text{def}}{=} D - A$. Since L is diagonally dominant, $L \succcurlyeq 0$.
- Normalized Laplacian matrix $\mathcal{L} \stackrel{\text{def}}{=} D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$.

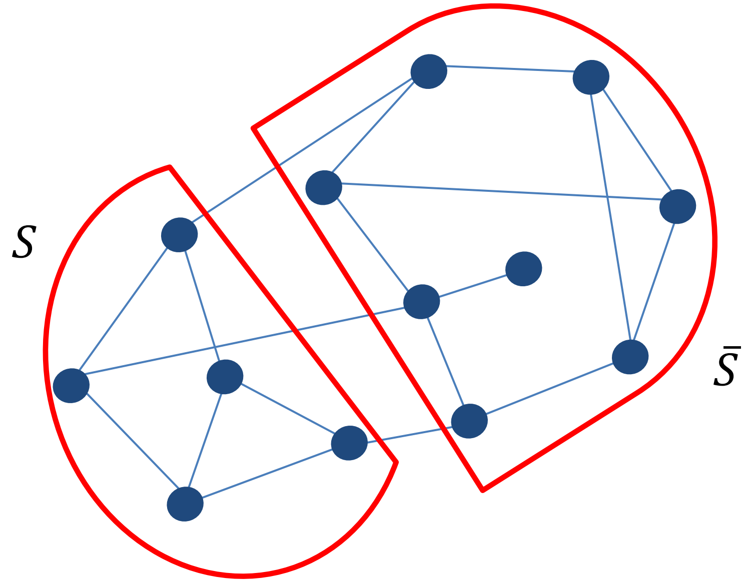
$$\min_x \frac{x^T \mathcal{L} x}{x^T x} = \frac{\left(D^{-1/2} x\right)^T L \left(D^{-1/2} x\right)}{x^T x} \geq 0$$

- Therefore, $\mathcal{L} \succcurlyeq 0$

Laplacian matrix

- λ is an eigenvalue of the normalized adjacency matrix iff $1 - \lambda$ is an eigenvalue of $\mathcal{L}$.
- Let $0 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$ be the eigenvalues of $\mathcal{L}$. Graph is disconnected iff $\lambda_2 = 0$.
- Graph is “close to disconnected” iff λ_2 is “close” to 0?





Graph Cuts

- Expansion of a set S

$$\phi(S) \stackrel{\text{def}}{=} \frac{\sum_{i\in S, j\in V\setminus S} w_{ij}}{\sum_{i\in S} d_i}$$

$\phi(S) = \frac{|E(S,\bar{S})|}{d|S|}$ for d-regular unweighted graphs

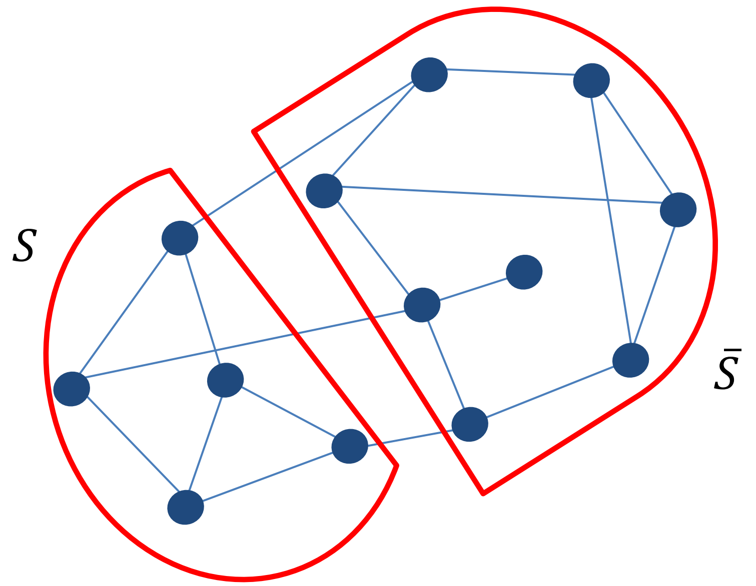
- Other definitions ($\text{vol}(S) \stackrel{\text{def}}{=} \sum_{i\in S} d_i$)

$$\phi(S) \stackrel{\text{def}}{=} \frac{\sum_{i\in S, j\in V\setminus S} w_{ij}}{\min\{\text{vol}(S), \text{vol}(V \setminus S)\}}$$

- Expansion of the graph

$$\phi_G \stackrel{\text{def}}{=} \min_{S:\text{vol}(S)\le \text{vol}(V)/2} \phi(S)$$

$\phi_G = \min_{S:|S|\le|V|/2} \phi(S)$ for d-regular graphs.



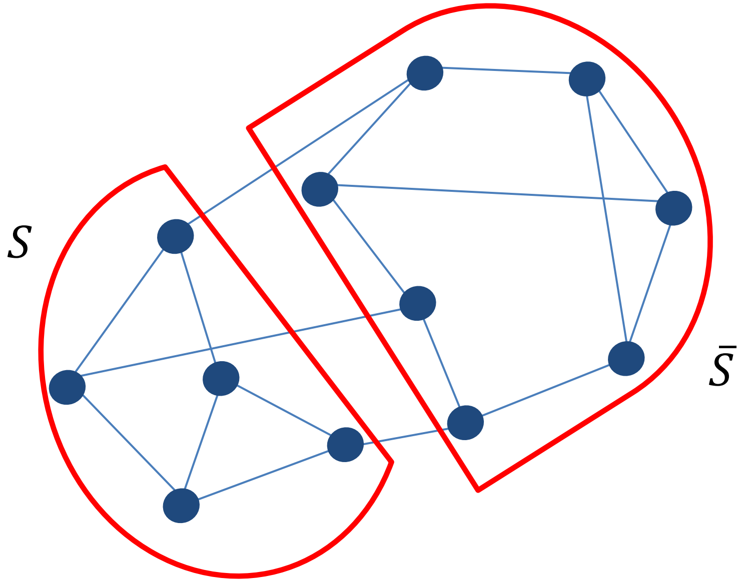
Cheeger's Inequality

- Cheeger's Inequality [Alon, Milman - 85, Alon - 86]

$$\frac{\lambda_2}{2} \leq \phi_G \leq \sqrt{2\lambda_2}$$

- Can be used to find a "sparse cut" in a graph.

- Can be used to certify expansion (approximately).



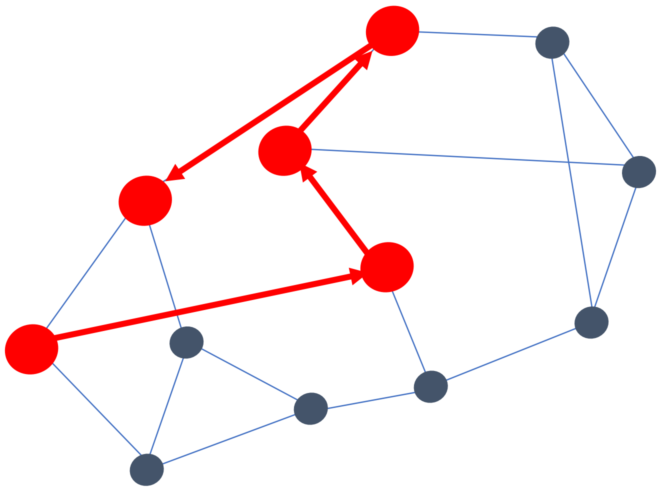
Expander Graphs

- A graph G is said to be an α–expander, if $\phi_G \geq \alpha$. Informally, G is an expander if $\phi_G = \Omega(1)$.
- Examples of expanders: complete graphs, random graphs (w.h.p.), many explicit constructions, etc.
- Examples of non-expanders: cycles, paths, etc.

- A d-regular graph is called a β-spectral expander if $\max_{i \in \{2,\dots,n\}} |\sigma_i| \leq \beta$, where σ_is are the eigenvalues of the adjacency matrix.
- Using Cheeger's inequality, such graphs have expansion at least $(d - \beta)/(2d)$.

- Many applications of expander graphs [Expander Graphs and their Applications -Hoory, Linial and Wigderson].

Random Walks on Graphs



- Start with a probability distribution μ^0
- random walks on regular graphs

$$\mu^{t+1} = \frac{1}{d} A \cdot \mu^t$$

- Stationary distribution $\mu^* = (A/d)\mu^*$
- If graph is connected, then $\mathbf{1}/n$ is the unique stationary distribution.

- δ-mixing time: smallest time t such that $d_{TV}(\mu^t, \mu^*) \leq \delta$ for any starting distribution μ^0.

$$d_{TV}(\mu^t, \mu^*) = \frac{1}{2} \|\mu^t - \mu^*\|_1$$

Applications of Random Walks on expanders

- Let B be a set of strings, Alg a randomized algorithm to decide whether an input string x belongs to B or not. Suppose Alg uses R bits of randomness and satisfies the following properties.
- If $x \in B$, $\Pr_r[\text{Alg}(x,r) = yes] \geq \frac{1}{2}$.
- If $x \notin B$, $\Pr_r[\text{Alg}(x,r) = no] = 1$.

- Alg2: Run Alg with k independent $r_1, \ldots, r_k$, output yes if any of the runs returns yes.
- If $x \in B$, $\Pr_{r_1,\ldots,r_k}[\text{Alg2}(x,r) = yes] \geq 1 - \frac{1}{2^k}$.
- If $x \notin B$, $\Pr_{r_1,\ldots,r_k}[\text{Alg2}(x,r) = no] = 1$.
- Number of bits of randomness needed $=kR$.

- Construct a d-regular expander graph on 2^R vertices for some $d = O(1)$ and $\beta \leq 0.8\, d$.
- Sample r_1 to be uniform random vertex in G. Sample r_{i+1} to be a uniform random neighbor of r_i for $i \geq 1$.
- Number of random bits used = $R + \lceil \log_2 d \rceil (l - 1)$.

- Idea: Fix an $x \in B$. Let $S = \{r: \mathrm{Alg}(x, r) = yes\}$. Since $|S|$ is "large", probability of a random walk avoiding S is small.

- Lemma: Total number of walks of length l between i and j = $\left(A^l\right)_{ij}$. (why?)
- Total number of walks of length l = $\mathbf{1}^T A^l \mathbf{1} = d^l n$

- Define

$$\bar{A}_{ij} = \begin{cases} 0\ if\ i\ or\ j \in S \\ A_{ij}\ otherwise \end{cases}$$

- Total number of walks of length l that avoid S = $\mathbf{1}^T \bar{A}^l \mathbf{1} \leq \lambda_{max}(\bar{A})^l n$
- Idea: Show that $\lambda_{max}(\bar{A}) < d$.
- Probability of avoiding $S \leq \frac{\lambda_{max}(\bar{A})^l n}{d^l n} = \left(\frac{\lambda_{max}(\bar{A})}{d}\right)^l$
- Example:

$$\frac{\mathbf{1}^T \bar{A} \mathbf{1}}{\mathbf{1}^T \mathbf{1}} = \frac{\sum_i \sum_j \bar{A}_{ij}}{n} = \frac{\sum_{i \in V \backslash S} \sum_{j \in V \backslash S} A_{ij}}{n} \leq \frac{\sum_{i \in V \backslash S} d}{n} = \frac{n - |S|}{n} d \leq \frac{d}{2}$$

- For any $x \in \mathbb{R}^n$, define $z_i = \begin{cases} 0\ if\ i \in S \\ x_i\ if\ i \notin S \end{cases}$

$$x^T \bar{A} x = \sum_{ij} \bar{A}_{ij} x_i x_j = \sum_{ij} A_{ij} z_i z_j = z^T A z$$

- Write $z = c_1\mathbf{1} + z_\perp$.

$$z^TAz = (c_1\mathbf{1} + z_\perp)^T A (c_1\mathbf{1} + z_\perp) = (c_1\mathbf{1} + z_\perp)^T (dc_1\mathbf{1} + Az_\perp)$$
$$= d\|c_1\mathbf{1}\|^2 + c_1\mathbf{1}^T Az_\perp + 0 + z_\perp^T Az_\perp = d\|c_1\mathbf{1}\|^2 + z_\perp^T Az_\perp$$

- Then $\langle z, \mathbf{1}\rangle = c_1\langle \mathbf{1}, \mathbf{1}\rangle + 0$. Therefore, $c_1 = \langle z, \mathbf{1}\rangle / n$.

$$\|c_1\mathbf{1}\|^2 = c_1^2 n = n\frac{1}{n^2}\langle z, \mathbf{1}\rangle^2 = \frac{1}{n}\left(\sum_i z_i\right)^2 \leq \frac{1}{n}(n - |S|)\sum_i z_i^2 \leq \frac{1}{2}z^Tz$$

- Since G is a β-spectral expander

$$\frac{z_\perp^T Az_\perp}{z_\perp^T z_\perp} \leq \max_{y \perp \mathbf{1}} \frac{y^TAy}{y^Ty} \leq \beta$$

- Therefore,

$$z^TAz \leq d\|c_1\mathbf{1}\|^2 + \beta\|z_\perp\|^2 = d\|c_1\mathbf{1}\|^2 + \beta(\|z\|^2 - \|c_1\mathbf{1}\|^2)$$
$$= (d - \beta)\|c_1\mathbf{1}\|^2 + \beta(\|z\|^2) \leq (d - \beta)\left(\frac{1}{2}\|z\|^2\right) + \beta\|z\|^2 = \frac{d + \beta}{2}\|z\|^2$$

$$x^T \bar{A} x = z^T A z \leq \frac{d+\beta}{2} \|z\|^2 \leq \frac{d+\beta}{2} \|x\|^2$$

- Therefore, $\lambda_{max}(\bar{A}) \leq \frac{d+\beta}{2}$
- Probability of avoiding $S \leq \left(\frac{\lambda_{max}(\bar{A})}{d}\right)^l \leq (0.9)^l$ (using $\beta \leq 0.8d$).

- Therefore, using $R + O(k)$ bits of randomness, can reduce error probability to $\frac{1}{2^k}$

Mixing time of random walks on graphs

- δ-mixing time: smallest time t such that $d_{TV}(\mu^t, \mu^*) \leq \delta$ for any starting distribution μ^0.
- $\mu^0 = \left\langle \mu^0, \frac{\mathbf{1}}{\sqrt{n}} \right\rangle \frac{\mathbf{1}}{\sqrt{n}} + \mu_\perp^0 = \left(\sum_i \mu_i^0\right) \frac{\mathbf{1}}{n} + \mu_\perp^0 = \frac{\mathbf{1}}{n} + \mu_\perp^0$
- $\mu^t = (A/d)^t \mu^0 = \frac{\mathbf{1}}{n} + (A/d)^t \mu_\perp^0$
- If graph is a β-spectral expander, then

$$\left\| \mu^t - \frac{\mathbf{1}}{n} \right\|_2^2 = \left\| \left(\frac{A}{d}\right)^t \mu_\perp^0 \right\|^2 = (\mu_\perp^0)^T \left(\frac{A}{d}\right)^{2t} (\mu_\perp^0) \leq \left(\frac{\beta}{d}\right)^{2t} \left\| \mu_\perp^0 \right\|^2 \leq \left(\frac{\beta}{d}\right)^{2t}$$

$$\left\| \mu^t - \frac{\mathbf{1}}{n} \right\|_1 \leq \sqrt{n} \left\| \mu^t - \frac{\mathbf{1}}{n} \right\|_2 \leq \sqrt{n} \left(\frac{\beta}{d}\right)^t$$

- Theorem: for $t = O\left(\frac{\log(n/\delta)}{\log(d/\beta)}\right)$, we have $d_{TV}(\mu^t, \mu^*) \leq \delta$.

Mixing Time

- (H.W.) Give an upper bound on mixing time as a function of ϕ_G.

- Levin, Peres, Wilmer – “Markov Chains and Mixing Times”
- Tetali, Montenegro - “Mathematical Aspects of Mixing Times in Markov Chains”.