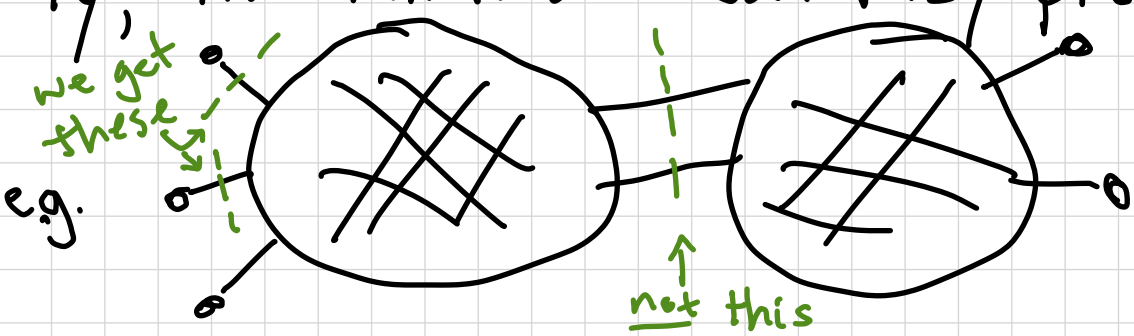
CSE5401 Lec. 8

1. Sparse cuts in graphs
2. Relating eigenvalues to sparsest cut
3. Algorithm for finding sparse cuts

Sometimes it is useful to partition a graph into well-connected pieces where the graph is well-connected, towards the design of a "divide-and-conquer" style algorithm that has a base case specialized to such components. We might recall algorithms for the Min Cut problem from undergraduate algorithms: recall that a cut is a non-trivial subset $S \subsetneq V$, $S \neq \emptyset$. The size of the cut is the number of edges crossing: $E(S) = \{(u,v) \in E : u \in S, v \notin S\}$.

Unfortunately, the minimum cut may produce a highly imbalanced partition, eg.



One possible fix is to penalize imbalanced cuts by examining the fraction of edges that cross the cut:

$$\Phi(S) = \frac{|E(S)|}{\min\{|\{(u,v) \in E : u \in S\}|, |\{(u,v) \in E : v \notin S\}|}$$ the "conductance".

Recalling that in a random walk on an undirected graph the stationary distribution is uniform on edges, this is the probability that the walk crosses from the smaller side in a step from the stationary distribution.

We recall that (at least when eigenvectors exist) the second-largest eigenvector gives a bound on the number

of steps to reach a stationary distribution. So, it is natural to conjecture that there is a connection.

Theorem (Cheeger) $\frac{1}{2}(1-\lambda_2) \le \min_S \Phi(S) \le \sqrt{2(1-\lambda_2)}$

Both quantities govern the "mixing time" of a random walk.

First we'll establish the lower inequality. Recall we had

Lemma $\frac{1}{2}\sum_{i,j} P_{ij}(x_i - x_j)^2 = \|\vec{x}\|_2^2 - \vec{x}^T P \vec{x}$ for any symmetric random walk transition matrix P.

Where $1-\lambda_2 = \min_{\substack{\vec{x}:\vec{x}\perp\vec{1}\\ \vec{x}\neq 0}} \left\{1 - \frac{\vec{x}^T P \vec{x}}{\|\vec{x}\|_2^2}\right\} = \min_{\substack{\vec{x}:\vec{x}\perp\vec{1}\\ \vec{x}\neq 0}} \left\{\frac{\frac{1}{2}\sum_{ij} P_{ij}(x_i-x_j)^2}{\|\vec{x}\|_2^2}\right\}$

In particular, if the graph is d-regular, the transition matrix $P = \frac{1}{d}E$ where E is the adjacency matrix, and for $\{0,1\}$-valued $\vec{x}$, how does the expression relate to the number of edges? We have $|E(\{i: x_i = 1\})| = \frac{1}{2}\sum_{i,j} E_{ij}(x_i - x_j)^2$ and $|\{(i,j)\in E : x_i = 1\}| = d\sum_{i=1}^{n} x_i = d\|\vec{x}\|_2^2$. Thus: $\frac{\frac{1}{2}\sum_{ij} P_{ij}(x_i-x_j)^2}{\|\vec{x}\|_2^2} = \frac{|E(\{i:x_i=1\})|}{|\{(i,j)\in E : x_i = 1\}|}$

This is almost $\min_{S\neq V, S\neq\emptyset} \Phi(S)$, if we minimize between $|\{(i,j)\in E : x_i = 1\}|$ and $|\{(i,j)\in E : x_j = 0\}|$. But that only makes the expression larger; and, we know that the larger of the two is at least $d\cdot\frac{n}{2}$

so, $\min\{|\{(i,j)\in E : x_i = 1\}|, |\{(i,j)\in E : x_i = 0\}|\} \le \frac{2}{dn}\cdot|\{(i,j)\in E : x_i = 1\}||\{(i,j)\in E : x_i = 0\}|$

Claim $\min_{\substack{\vec{x}:\vec{x}\perp\vec{1}\\ \vec{x}\neq 0}} \left\{\frac{\frac{1}{2}\sum_{ij} P_{ij}(x_i-x_j)^2}{2\|\vec{x}\|_2^2}\right\} \le \min_{\vec{x}\in\{0,1\}^n\setminus\{\vec{0},\vec{1}\}} \left\{\frac{\frac{1}{2}\sum_{ij} E_{ij}(x_i-x_j)^2}{\frac{2}{dn}|\{(i,j)\in E : x_i = 1\}||\{(i,j)\in E : x_j = 0\}|}\right\}$

Proof: For any $\vec{y}\in\{0,1\}^n\setminus\{\vec{0},\vec{1}\}$, put $\rho = \frac{|\{i: y_i = 1\}|}{n}$ and $\vec{x} = \vec{y} - \rho\vec{1}$.

Now $\langle\vec{x},\vec{1}\rangle = \sum_{i:y_i=1}(1-\rho) + \sum_{i:y_i=0}-\rho = |\{i: y_i = 1\}| - n\cdot\rho = 0$ and $(x_i-x_j)^2 = (y_i-y_j)^2$ for all i,j. $\|\vec{x}\|_2^2 = \sum_{i:y_i=1}(1-\rho)^2 + \sum_{i:y_i=0}\rho^2 = n\rho^2 + |\{i:y_i=1\}|(1-2\rho) = |\{i:y_i=1\}|(\cancel{\rho}+1-\cancel{2\rho})$

so $\frac{\frac{1}{2}\sum_{ij} P_{ij}(x_i-x_j)^2}{\|\vec{x}\|_2^2} = \frac{\frac{1}{2}\sum_{ij} E_{ij}(y_i-y_j)^2}{d|\{i:y_i=1\}||\{i:y_i=0\}|\cdot\frac{1}{n}} = \frac{\frac{1}{2}\sum_{ij} E_{ij}(y_i-y_j)^2}{\frac{1}{dn}|\{(i,j)\in E : y_i = 1\}||\{(i,j)\in E : y_i = 0\}|}$.

Thus, for every $\vec{y}$ on the RHS, there is an $\vec{x}$ that gives the same value in the minimum on the LHS. Hence, the LHS can be no larger. ∎

This completes the lower inequality. For the upper inequality, we will give an algorithm to construct a cut demonstrating $\min_{S \subsetneq V, S \neq \emptyset} \Phi(S) \leq \sqrt{2(1-\lambda_2)}$. (A "square root" from optimal)

Given an undirected graph $G=(V,E)$, put $\vec{x} \leftarrow$ 2nd eigenvector of E.
Sort the pairs $(i, x_i) : i = 1, \ldots, n$ by the x_i-component. (Get $i_1, i_2, \ldots, i_n$)
Put $j^* \leftarrow \arg\min_{j \in \{1,\ldots,n-1\}} \Phi(\{i_1, i_2, \ldots, i_j\})$
Return $\{i_j : j \leq j^*\}$.

Note that given $\sum_{kl} E_{kl}(y_k^{(j)} - y_l^{(j)})^2 + |\{(k,l) \in E : y_k^{(j)} = 1\}|$ for $y_i^{(j)} = \begin{cases} 1 & i = i_k, k \leq j \\ 0 & \text{o.w.} \end{cases}$ we can compute these for $\vec{y}^{(j+1)}$ in time $O(\deg(i_{j+1}))$.

Thus, the running time is dominated by the time to compute the eigenvector $\vec{x}$; the rest is $O(|E|)$. But, we'll see that an approximate eigenvector is enough, $\vec{x}$ s.t. $\|P\vec{x}\| \geq (1-\varepsilon)\lambda_2 \|\vec{x}\|$.

So we can also use the power method to obtain $\vec{x}$. (review later)

For the analysis, we'll define a distribution on thresholds τ such that for $\vec{y}(\tau) = \begin{cases} 1 & \text{if } x_i \leq \tau \\ 0 & \text{o.w.} \end{cases}$

$\mathbb{E}\left[\sum_{ij} E_{ij}(y_i(\tau) - y_j(\tau))^2 - \sqrt{2(1-\lambda_2)} \min\{|\{(i,j) \in E : y_i(\tau) = 1\}|, |\{(i,j) \in E : y_j(\tau) = 0\}|\}\right] \leq 0$ so for some τ^*

$\sum_{ij} E_{ij}(y_i(\tau^*) - y_j(\tau^*))^2 - \sqrt{2(1-\lambda_2)} \min\{|\{(i,j) \in E : y_i(\tau^*) = 1\}|, |\{(i,j) \in E : y_j(\tau^*) = 0\}|\} \leq 0$ giving a cut $S^* = \{i : y_i(\tau^*) = 1\}$ with $\Phi(S^*) \leq \sqrt{2(1-\lambda_2)}$. Recall that $2(1-\lambda_2) = \frac{\sum_{ij} E_{ij}(x_i - x_j)^2}{d\|\vec{x}\|_2^2}$ for our eigenvector $\vec{x}$. We'd like τ to cut the edge (i,j) with probability $(x_i - x_j)^2$ so that $\mathbb{E}\left[\sum_{ij} E_{ij}(y_i(\tau) - y_j(\tau))^2\right]$ $= \sum_{ij} E_{ij}(x_i - x_j)^2$. We won't quite manage this, but we can use a pdf $f(t) \propto |t|$ so that $\Pr[i,j \text{ cut}] = \Pr[x_i \leq \tau \leq x_j] \propto \int_{x_i}^{x_j} |t|\, dt = \begin{cases} x_j^2 - x_i^2 & \text{if } \operatorname{sgn}(x_i) = \operatorname{sgn}(x_j) \\ x_j^2 + x_i^2 & \text{o.w.} \end{cases}$

We can get a convenient upper bound on both cases as follows. When $\operatorname{sgn}(x_i) = \operatorname{sgn}(x_j)$, $|x_j^2 - x_i^2| = |(x_j - x_i)(x_j + x_i)|$ and when $\operatorname{sgn}(x_i) \neq \operatorname{sgn}(x_j)$ $x_j^2 + x_i^2 \leq (x_j - x_i)^2 \leq |x_j - x_i|(|x_j| + |x_i|)$ by Δ-ineq.

Recall that the Rayleigh quotient does not change when we rescale $\vec{x}$, so we can rescale it so that $\int_{x_{i_1}}^{x_{i_n}} 2|t|\,dt = 1$ and thus $\Pr[(i,j) \text{ cut}] \le |x_j - x_i|(|x_j| + |x_i|)$ for all $(i,j) \in E$.

We can now bound the expected number of edges cut:

$$\mathbb{E}\left[\sum_{ij} E_{ij}(y_i(\tau) - y_j(\tau))^2\right] \le \sum_{ij} E_{ij}|x_j - x_i|(|x_j| + |x_i|), \text{ by Cauchy-Schwarz:}$$

$$\le \sqrt{\sum_{ij} E_{ij}(x_j - x_i)^2}\ \sqrt{\sum_{ij} E_{ij}(|x_i| + |x_j|)^2}$$

$2x_i^2 + 2x_j^2 - (|x_i| + |x_j|)^2 = (|x_i| - |x_j|)^2 \ge 0 \Rightarrow (|x_i| + |x_j|)^2 \le 2(x_i^2 + x_j^2)$ so

$$\sqrt{2\sum_{ij} E_{ij}(x_i^2 + x_j^2)} = 2\sqrt{\sum_{i=1}^n d x_i^2}$$

$$\le \underbrace{\sqrt{\frac{\sum_{ij} E_{ij}(x_j - x_i)^2}{d\sum_{i=1}^n x_i^2}}}_{\sqrt{2(1-\lambda_2)}}\ 2\, d\sum_{i=1}^n x_i^2$$

Thus, it will suffice to also bound $\mathbb{E}\left[\min\{\|y(\tau)\|_2^2, \|\vec{1} - \vec{y}(\tau)\|_2^2\}\right] \ge \|\vec{x}\|_2^2$.

Note that if we add $\mu \cdot \vec{1}$ to $\vec{x}$, $\vec{z} = \vec{x} - \mu \cdot \vec{1}$ has $(z_j - z_i)^2 = (x_j - x_i)^2$ for all i,j and if we define $y_i(\tau) = 1$ if $\tau \ge z_i$ instead, we get a bound $\sqrt{2\frac{\sum_{ij} E_{ij}(z_j - z_i)^2}{d\sum z_i^2}}\, d\sum_{i=1}^n z_i^2$. The advantage is that we can put $\mu = x_{i_{n/2}}$ so that whichever side of the cut is smaller, they all have the same sign. Note that since $\vec{x} \perp \vec{1}$, $\langle z, z\rangle = \langle x, x\rangle - 2\mu\langle x, 1\rangle + \mu^2 n$

Therefore $\sqrt{\frac{\sum_{ij} E_{ij}(z_j - z_i)^2}{d\sum z_i^2}} \le \sqrt{1-\lambda_2}$ still. To conclude,

Claim $\mathbb{E}\left[\min\{\|y(\tau)\|_2^2, \|\vec{1} - \vec{y}(\tau)\|_2^2\}\right] = \sum_{i=1}^n z_i^2$

Proof? For each i_j, $j \le n/2$, i_j is in the smaller set when $z_{i_j} \le \tau \le 0$, which occurs with prob. $\int_{z_{i_j}}^0 2|t|\,dt = z_{i_j}^2$. For i_j, $j > n/2$, i_j is in the smaller set when $0 \le \tau < z_{i_j}$, i.e., with prob. $\int_0^{z_{i_j}} 2|t|\,dt = z_{i_j}^2$ again. So, by linearity of expectation, the expected size of the smaller side of the cut is $\sum_{i=1}^n z_i^2$ ▨

We now turn to finding an approximate eigenvector. For the proof, we only needed that $\vec{x} \perp \vec{1}$ and that the Rayleigh quotient is approximately λ_2; if it was $(1-\varepsilon)\lambda_2$, we'd get a cut of conductance $\sqrt{2(1-(1-\varepsilon)\lambda_2)}$. The power method can give us this efficiently:

Initialize $\vec{x}^{(0)}$: $x_i^{(0)} = \pm 1$ unif. at random
Put $\vec{x}^{(1)} = \vec{x}^{(0)} - \langle \vec{x}^{(0)}, \vec{1} \rangle \cdot \vec{1}$ (project to $\vec{x}^{(1)} \perp \vec{1}$)
For $t = 1, \ldots, T$ put $\vec{x}^{(t+1)} = P \cdot \vec{x}^{(t)}$
Return $\vec{x}^{(T+1)}$

The analysis is very similar to the analysis of mixing. We'll suppose the graph is not bipartite. Then if we consider writing $\vec{x}^{(0)}$ in the basis of eigenvectors of P, $\vec{x}^{(0)} = a_1 \vec{v}^{(1)} + a_2 \vec{v}^{(2)} + \cdots + a_n \vec{v}^{(n)}$, $\vec{x}^{(1)}$ simply zeros out a_1 and then $\vec{x}^{(t+1)} = \sum_{i \geq 2}^{n} a_i \lambda^{(i)t} \vec{v}^{(i)}$. Let's suppose k of the eigenvalues have magnitude at least $(1-\frac{\varepsilon}{2})|\lambda_2|$, so the other $n-k$ have magnitude at most $(1-\frac{\varepsilon}{2})|\lambda_2|$. After $t = O(\frac{1}{\varepsilon} \log \frac{n}{a_2})$ iterations, the projection of $\vec{x}^{(t+1)}$ on the span of these small eigenvalues is at most $(1-\frac{\varepsilon}{2})^t |\lambda_2|^t \cdot \|\vec{x}^{(0)}\|_2 < |a_2| \frac{\varepsilon}{3} |\lambda_2|^t$. The projection on the span of the large eigenvalues is at least the projection on $\vec{v}^{(2)}$, i.e., at least $|\lambda_2|^t \cdot |a_2|$. Thus, the projection of $\vec{x}^{(t+1)}$ on the small eigenvalue subspace has norm less than a $\frac{\varepsilon}{3}$-fraction of $\|\vec{x}^{(t+1)}\|_2$; thus, $\frac{\vec{x}^{(t+1)T}}{\|\vec{x}^{(t+1)}\|_2} P \frac{\vec{x}^{(t+1)}}{\|\vec{x}^{(t+1)}\|_2} \geq (1-\frac{\varepsilon}{2})\lambda_2(1-\frac{\varepsilon}{3})$
$> (1-\varepsilon)\lambda_2$.

Also observe that $\vec{x}^{(t+1)} \perp \vec{1}$ for all t.

We now only need a lower bound on $|a_2|$ to bound the number

of iterations needed. $a_2 = \langle \vec{x}^{(0)}, \vec{v}^{(2)} \rangle$. a_2 may be negative and has expected value 0, but a_2^2 has expectation

$$\mathbb{E}\left[\langle \vec{x}^{(0)}, \vec{v}^{(2)} \rangle^2\right] = \sum_{i,j} v_i^{(2)} v_j^{(2)} \underbrace{\mathbb{E}\left[x_i^{(0)} x_j^{(0)}\right]}_{0 \text{ if } i \neq j,\ 1 \text{ o.w.}} = \sum_{i=1}^{n} v_i^{(2)2} = \|\vec{v}^{(2)}\|_2^2.$$

Likewise, $\mathbb{E}\left[\langle \vec{x}^{(0)}, \vec{v}^{(2)} \rangle^4\right] = \sum_{i,j,k,l} v_i^{(2)} v_j^{(2)} v_k^{(2)} v_l^{(2)} \mathbb{E}\left[x_i^{(0)} x_j^{(0)} x_k^{(0)} x_l^{(0)}\right]$

0 unless $i=j$ & $k=l$ or $i=k$ & $j=l$ or $i=l$ & $j=k$. Note that these triple count $i=j=k=l$, so we must subtract this:

$$= 3\sum_{i,j} v_i^{(2)2} v_j^{(2)2} - 2\sum_{i=1}^{n} v_i^{(2)4} < 3 \cdot \|\vec{v}^{(2)}\|_2^4.$$

Theorem (Paley-Zygmund) $\Pr\left[\underbrace{Z \geq \delta \mathbb{E}[Z]}_{\text{"A"}}\right] \geq (1-\delta)^2 \frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}$ for $Z \geq 0$.

Proof: $\mathbb{E}[Z] = \mathbb{E}[Z \cdot 1_A] + \underbrace{\mathbb{E}[Z \mid \neg A]}_{\leq \delta \mathbb{E}[Z]} (1 - \Pr[A])$

$$\leq \sqrt{\mathbb{E}[Z^2]\,\mathbb{E}[1_A]}$$

$$(1-\delta)\mathbb{E}[Z] \leq \sqrt{\mathbb{E}[Z^2]}\sqrt{\Pr[A]} - \underbrace{\delta \mathbb{E}[Z] \Pr[A]}_{\leq 0}$$

$$\frac{(1-\delta)^2 \mathbb{E}[Z]^2}{\mathbb{E}[Z^2]} \leq \Pr[A] \quad \blacksquare$$

Here, $Z = \langle \vec{x}^{(0)}, \vec{v}^{(2)} \rangle^2$, and so we get $\Pr[a_2^2 \geq \delta] \geq (1-\delta)^2 \cdot \frac{1}{3}$

A constant # of trials hence suffices to get any $\delta > 0$ lower bound. Each iteration takes time $O(|E|)$, so overall we get a $(1-\varepsilon)$-approximate eigenvector in time $O(\frac{1}{\varepsilon}|E| \log n)$. As we mentioned earlier, this dominates the time to find a cut, and gives our overall running time.