Issue: $\frac{cut(S)}{vol(S)}$ is not necessarily conductance — it only is if

$vol(S) \le m = |E|$ $\quad [vol(V) = 2m]$

Fixing the issue

Choose y so that $\lambda_2 = R(y) = \frac{y^T L y}{y^T D y}$

and scale it so that $-1 \le y_i \le 1$

for $i \in [n] = \{1, 2, \ldots, n\}$.

Note that $y^T d = 0$. $\quad [x = D^{1/2} y]$

$\mathcal{L} x = \lambda_2 x$

(1) Let $z = y - ce$ where $c \in \mathbb{R}$ will be chosen later and $e = [1\ 1 \cdots 1]^T$

Then

$$\boxed{R(z) = \frac{\sum_{ij \in E} (z_i - z_j)^2}{\sum_{i \in V} d_i z_i^2}} = \frac{\sum_{ij \in E} (y_i - y_j)^2}{\sum_{i \in V} (y_i - c)^2 d_i}$$

$$= \frac{\sum_{ij \in E} (y_i - y_j)^2}{\sum_{i \in V} y_i^2 d_i + c^2 d_i - 2 y_i c d_i}$$

$$= \frac{\sum_{ij \in E} (y_i - y_j)^2}{\left[\sum_{i \in V} y_i^2 d_i\right] + \underbrace{c^2 \left[\sum_{i \in V} d_i\right]}_{>0} - \cancel{2c \sum_{i \in V} y_i d_i}^{\,0}}$$

$y^T d = 0$

$$\leq \underline{R(y)} = \lambda_2$$

(2) For the chosen c, let

$$S_c = \{ i \in V : y_i < c \}$$

$$\hat{S}_c = \{ i \in V : y_i \geq c \}$$

We will choose c so that

$vol(S_c) \leq m$ and $vol(\hat{S}_c) \leq m$

How? Order entries of y by size and consider their corresponding degrees.

$$y_1 \leq y_2 \leq \cdots \leq y_r \leq y_{r+1} \leq y_{r+2} \leq \cdots \leq y_{n-1} \leq y_n$$

$$d_1 \quad d_2 \quad \cdots \quad d_r \quad d_{r+1} \quad d_{r+2} \quad \quad d_{n-1} \quad d_n$$

Let r be the integer satisfying

$$\boxed{\sum_{i=1}^{r} d_i \leq m} \quad \text{and} \quad \sum_{i=1}^{r+1} d_i > m$$

This means that $\displaystyle\sum_{i=r+2}^{n} d_i \leq m.$

Choose $c = y_{r+1}$. Then

$$\boxed{\begin{aligned} vol(S_c) &= \sum_{i=1}^{r} d_i \le m \\ vol(\hat{S}_c) &= \sum_{i=r+2}^{n} d_i \le m \end{aligned}} \quad (B) \quad \checkmark$$

Setting $z = y - (y_{r+1})e$, define z^+ and z^- by

$$z_i^+ = \begin{cases} z_i & \text{if } z_i > 0 \quad y_i > c \\ 0 & \text{otherwise} \end{cases}$$

$$z_i^- = \begin{cases} |z_i| & \text{if } z_i < 0 \\ 0 & \text{otherwise} \end{cases}$$

so that $z = z^+ - z^-$

By (B), we know that

$$\boxed{\begin{aligned} &\underline{vol(supp(z^+)}) \leq m \\ &vol(supp(z^-)) \leq m \end{aligned}}$$

$$supp(x) = \{ i \in V : x_i \neq 0 \}$$

(3) Claim: Either $R(z^+) \leq R(z)$

or $R(z^-) \leq R(z)$

Proof: For $ij \in E$ if z_i and z_j have the same sign, then

$$(z_i - z_j)^2 = (z_i^+ - z_j^+)^2 + (z_i^- - z_j^-)^2$$

If z_i and z_j have opposite signs, then

$$(z_i - z_j)^2 = z_i^2 + z_j^2 \underline{- 2 z_i z_j}$$

$$\geq z_i^2 + z_j^2$$

$$= \boxed{(z_i^+ - z_j^+)^2 + (z_i^- - z_i^-)^2}$$

Therefore,

$$\underline{z^T L z} = \sum_{ij \in E} (z_i - z_j)^2$$

$$\geq \sum_{ij \in E} (z_i^+ - z_j^+)^2 + \sum_{ij \in E} (z_i^- - z_j^-)^2$$

$$= z_+^T L z_+ + z_-^T L z_-$$

Furthermore, for each $i \in V$

$$z_i^2 = (z_i^+ + z_i^-)^2 = (z_i^+)^2 + (z_i^-)^2 + 2 z_i^+ z_i^-$$

$$= (z_i^+)^2 + (z_i^-)^2$$

$$\sum_{i \in V} d_i z_i^2 = \sum_{i \in V} d_i (z_i^+)^2 + \sum_{i \in V} d_i (z_i^-)^2$$

$$\underline{z^T D z} = z_+^T D z + z_-^T D z$$

So

$$\lambda_2 = R(y) \geq R(z) = \frac{z^T L z}{z^T D z} \geq \frac{z_+^T L z_+ + z_-^T L z_-}{z_+^T D z_+ + z_-^T D z_-}$$

$$\geq \min \left\{ \frac{z_+^T L z_+}{z_+^T D z_+}, \frac{z_-^T L z_-}{z_-^T D z_-} \right\}$$

$$= \min \{ R(z^+), R(z^-) \}$$

Claim proven. $\blacksquare$

(4) Now, we separately apply Lemma 1 to both z^+ and z^- to find

$$S^+ \subseteq \operatorname{supp}(z^+) \qquad \operatorname{vol}(S^+) \leq m$$

$$S^- \subseteq \operatorname{supp}(z^-) \qquad \operatorname{vol}(S^-) \leq m$$

with $\phi(S^+) = \dfrac{\operatorname{cut}(S^+)}{\operatorname{vol}(S^+)} \leq \sqrt{2R(z^+)}$

$$\phi(S^-) = \frac{cut(S^-)}{vol(S^-)} \leq \sqrt{2R(z^-)}$$

one of these is $\leq \sqrt{2R(z)} \leq \sqrt{2R(y)} = \sqrt{2\lambda_2}$

So I have some set S such that

$$\phi(S) \leq \sqrt{2\lambda_2}$$

$$y_1 \leq y_2 \leq y_3 \cdots \leq y_n$$

The spectral clustering algorithm checks all sets of the form

$$\boxed{S_t = \{i \in V : y_i < t\}}$$

In the lemma, for z^+, we check all cuts of the form

$$\rightarrow S_a = \{i \in V : (z_i^+)^2 > a\}$$

but since all values of z^+ are nonnegative, this is equivalent to checking all sets of the form

$$S_b = \{ i \in V : z_i^+ > b \}$$

(Similar for z^-)

Final observation: running a full sweep cut procedure on y will check all of the sets we would consider by applying Lemma 1 to z^+ and z^-.

Example

$$y = \begin{bmatrix} -0.8 & -0.7 & 0.0 & \underbrace{0.1}_{y_{r+1}} & 0.8 & 0.9 \end{bmatrix}$$

$$\underset{=y-ce}{z} = \begin{bmatrix} -0.9 & -0.8 & -0.1 & 0 & 0.7 & 0.8 \end{bmatrix}$$

$$z^+ = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.7 & 0.8 \end{bmatrix}$$

$$z^- = \begin{bmatrix} 0.9 & 0.8 & 0.1 & 0 & 0 & 0 \end{bmatrix}$$

How good of an approximation algorithm is spectral clustering?

A ρ-approximation is a set S satisfying

$$\phi^* = \phi(S^*) \leq \phi(S) \leq \rho\, \phi(S^*)$$

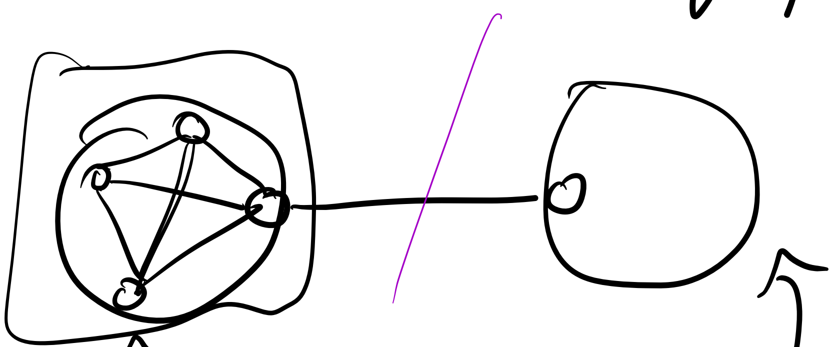
We have a set S satisfying

$$\phi^* \leq \phi(S) \leq \sqrt{2\lambda_2} \leq \sqrt{4\phi^*} = \frac{\sqrt{4}}{\sqrt{\phi^*}}\, \phi^*$$

$$\left(\frac{\lambda_2}{2} \leq \phi^* \quad \text{so} \quad 2\lambda_2 \leq 4\phi^* \right)$$

What is the worst case value of $\frac{\sqrt{4}}{\sqrt{\phi^*}}$?

Consider a dumbell graph



cliques of size $\frac{n}{2}$

$$\phi(S^*) = \frac{1}{1 + \frac{n}{2}\left(\frac{n}{2}-1\right)} \approx \frac{c_1}{n^2}$$

in this case $\frac{1}{\sqrt{\phi^*}} = \Theta(n)$

$$\frac{1}{\sqrt{\frac{c}{n^2}}} = \frac{n}{\sqrt{c}}$$

In the worst case, this is

only a $O(n)$-approximation to the optimal conductance.

We will use other techniques to get an $O(\log n)$ approximation.