Case Studies C
(Eigenvalue Decomposition)

Case I: Find roots of a polynomial.

Let $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} + x^n$

There are n solutions (called roots) of $p(x) = 0$ in $\mathbb{C}$.

Construct an $n \times n$ matrix

$$A_p = \begin{bmatrix} 0 & & & & -a_0 \\ 1 & \ddots & & 0 & -a_1 \\ & \ddots & \ddots & & \vdots \\ 0 & & \ddots & 0 & \vdots \\ & & & 1 & -a_{n-1} \end{bmatrix}$$

Then $\det(\lambda I - A_p) = a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n$

So, the eigenvalues of A_p are roots of $p(x)$.

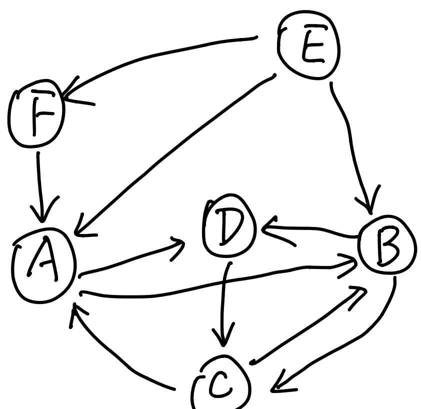
We can use QR algorithm to find eigenvalues of A_p.

This is widely used in available software packages, e.g. Matlab.

Case II: Page Rank (Revisited)

Problem: Rank the webpages?

Data: The linkages of webpages, e.g.



$A, B, \cdots, F$ are webpages.

"E → F" means webpage E has a hyperlink to F, and so on.

We have seen that, in Google's page rank, the ranking problem is formulated as a solution of linear equation:

$$\pi = \frac{1-p}{n} \cdot \mathbb{1} + pA\pi,$$

where $\pi \in \mathbb{R}^n$ is the score vector, $\mathbb{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$, and A is the normalized adjancency matrix with each column sum 1.

The equation is rewritten as

$$(I - pA)\pi = \frac{1-p}{n} \cdot \mathbb{1}$$

Because π is a probability distribution, $\mathbb{1}^T \pi = 1$. Therefore,

$$(I - pA)\pi = \frac{1-p}{n} \cdot \mathbb{1}\mathbb{1}^T \pi$$

Then $\left(pA + \frac{1-p}{n}\mathbb{1}\mathbb{1}^T\right)\pi = \pi$,

which implies that π is an eigenvector of $pA + \frac{1-p}{n}\mathbb{1}\mathbb{1}^T$ with associated eigenvalue 1.

Perron-Frobenius Theorem says that $\lambda = 1$ is the unique largest eigenvalue of $pA + \frac{1-p}{n}\mathbb{1}\mathbb{1}^T$.

Therefore, if we apply power iteration to $pA + \frac{1-p}{n}\mathbb{1}\mathbb{1}^T$, then we will get $\lambda_1 = 1$ and the eigenvector x_1. A renormalization of x_1, will give π, since $\mathbb{1}^T\pi = 1$ and $x_1 \geq 0$ by Perron-Frobenius theorem.

$$\pi = \frac{x_1}{\|x_1\|_1}$$

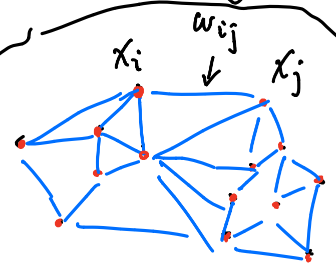
Case III: Spectral Clustering and Graph Cut

Given a data sets $\{x_i\}_{i=1}^m$, $x_i \in \mathbb{R}^n$, how to separate them into clusters?

Graph model: Construct a graph (V, E), where

— each point x_i is a vertex $i \in V$ of the graph,

— the similarity of x_i and x_j is the weight for the edge $(i,j)\in E$.
$\widehat{}$ w_{ij}



We may choose

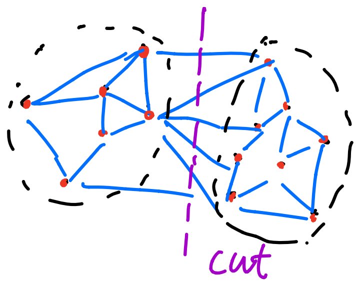
$$w_{ij} = e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}}$$

In this figure, longer edge, smaller weight

To separate points into, say, 2 clusters, we only need to separate the graph to two disconnect graphs.

The cost is that:

we have to cut some edges



A natural criteria for the cutting is to minimize the total weight of cut edges.

Let A, B be two subsets of V.

Define the cut of A and B as

$$\text{cut}(A,B) = \sum_{\substack{i\in A\\ j\in B}} w_{ij}$$

So we solve the problem:

$$\min_{S\in V} \text{cut}(S, \bar{S}), \quad \text{where} \quad \bar{S} = V\setminus S.$$

This is called $\boxed{\text{minimum cut}}$.

The min cut is hard to solve, because it is a combinatorial problem.

We can solve it approximately by reformulating it to an eigenvalue problem.

Let $z \in \mathbb{R}^m$ be an indicator vector satisfying

$$z_i = \begin{cases} 1 & \text{if } i \in S \\ -1 & \text{if } i \in \bar{S} \end{cases}$$

Then,
$$\begin{aligned}
\text{cut}(S, \bar{S}) &= \sum_{\substack{i \in S \\ j \in \bar{S}}} w_{ij} \\
&= \sum_{i,j} \begin{cases} w_{ij} = \frac{1}{4} w_{ij}(z_i - z_j)^2 & \text{if } i \in S,\ j \in \bar{S} \\ 0 = \frac{1}{4} w_{ij}(z_i - z_j)^2 & \text{if } i \in S\ \ j \in S \\ 0 = \frac{1}{4} w_{ij}(z_i - z_j)^2 & \text{if } i \in \bar{S},\ j \in \bar{S} \end{cases} \\
&= \frac{1}{4} \sum_{i,j} w_{ij}(z_i - z_j)^2
\end{aligned}$$

Let $D = \text{diag}(W\mathbb{1})$, where $(W\mathbb{1})_i = \sum_j w_{ij}$

Then,
$$\begin{aligned}
\text{cut}(S, \bar{S}) &= \frac{1}{4} \sum_{i,j} w_{ij}(z_i - z_j)^2 \\
&= \frac{1}{4}\left(\sum_{i,j} w_{ij} z_i^2 - 2\sum_{ij} w_{ij} z_i z_j + \sum_i \sum_j w_{ij} z_j^2 \right) \\
&\overset{\text{Because } W \text{ is symmetric.}}{=} \frac{1}{4} \cdot 2 \cdot \left(\sum_{ij} w_{ij} z_i^2 - \sum_{ij} w_{ij} z_i z_j \right) \\
&= \frac{1}{2} \cdot \left(\sum_i \left(\sum_j w_{ij} \right) z_i^2 - z^T W z \right) \\
&= \frac{1}{2} \left(\sum_i D_{ii} z_i^2 - z^T W z \right) \\
&= \frac{1}{2} \left(z^T D z - z^T W z \right) \\
&= \frac{1}{2} z^T (D - W) z
\end{aligned}$$

Then, min cut becomes

$$\begin{aligned}
\min_{z \in \mathbb{R}^m} \quad & z^T (D - W) z \\
\text{s.t.} \quad & z_i = \pm 1, \ i = 1, \cdots, m.
\end{aligned}$$

The constraint is difficult to handle. We relax it to

$$\{ z \mid z_i = \pm 1, i = 1, \cdots, m \} \xrightarrow{\text{relaxation}} z^T \mathbb{1} = 0, \quad z \neq 0. \text{ and } \|z\|_2 = \sqrt{n}$$

In this relaxation, we don't expect z_i's are exatly ± 1, but some z_i's are positive and some are negative.

So min cut is approximated by

$$\min_{\|z\|_2 = \sqrt{n}} \quad z^T (D - W) z \qquad \text{s.t.} \quad z^T \mathbb{1} = 0 \qquad (P)$$

After we get a solution z, we set $S=\{i \mid z_i>0\}$, $\bar{S}=\{i \mid z_i<0\}$.

It remains to solve (P).

The matrix $L \equiv D-W$ is called the Laplacian matrix of the graph (V,E)

① L is spsd, because: $L=L^T$ and

$$z^T L z = z^T(D-W)z = \sum_{i,j} W_{ij}(z_i-z_j)^2 \geq 0 \quad \forall z \in \mathbb{R}^n.$$

Therefore, all eigenvalues of L are non-negative.

$\boxed{\begin{array}{l}\text{Theorem: Let } A \text{ spsd. Then all eigenvalues are non-negative.} \\ \text{proof. By contradiction. Let } A = X\Lambda X^T \text{ be its eigenvalue decomposition.} \\ \quad \text{Suppose } \lambda_i<0. \text{ Let } x_i \text{ be corresponding eigenvector.} \\ \quad \text{Then, } x_i^T A x_i = x_i^T X \Lambda X^T x_i = \lambda_i x_i^T x_i \underset{\substack{\uparrow \\ \text{Contradiction.}}}{<0}. \end{array}}$

② $L1=(D-W)1 = D1-W1 = 0\,1$ which implies

0 is an eigenvalue of L and 1 is the corresponding eigenvector,

Based on these two observations, we have the following theorem:

Theorem: A solution of (P) is $z=\sqrt{n}\,x_{n-1}$, where x_{n-1} is an eigenvector of the 2nd smallest eigenvalue of $L \equiv D-W$.

Proof. (P) is equivalent to $\min_{\|z\|_2=1} z^T L z$, z

The eigenvalue decomposition of L must be

$$L = [x_1 \cdots x_{n-1} \ \tfrac{1}{\sqrt{n}}] \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_{n-1} & \\ & & & 0 \end{bmatrix} \begin{bmatrix} x_1^T \\ \vdots \\ x_{n-1}^T \\ 1/\sqrt{n} \end{bmatrix} = \sum_{i=1}^{n-1} \lambda_i x_i x_i^T, \text{ where}$$

x_i, $i=1,2,\cdots,n-1$ are orthogonal to 1, and $\lambda_1,\cdots,\lambda_{n-1} \geq 0$.

For any $z \in \mathbb{R}^n$ satisfying $z^T 1 = 0$, we must have

$$z = \sum_{i=1}^{n-1} c_i x_i,$$

and $\sum c_i^2 = n$.

(because $\{z \in \mathbb{R}^n \mid z^T 1 = 0\}$ is $(n-1)$-dimensional and $x_1,\cdots,x_{n-1}$ is an orthonormal basis)

So, $z^T L z = \sum_{i=1}^{n-1} \lambda_i (x_i^T z)^2 = \sum_{i=1}^{n-1} \lambda_i c_i^2$

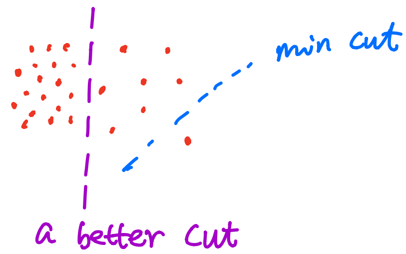
Since $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_{n-1} \geq 0$,

$$z^T L z \geq \lambda_{n-1} \cdot \sum_{i=1}^{n-1} c_i^2 = n \lambda_{n-1}.$$

The lower bound is attend when $c_1 = c_2 = \cdots = c_{n-2} = 0$ and $c_{n-1} = \sqrt{n}$.

So $z = \sqrt{n}\, x_{n-1}$ is the solution of (P). ⊠.

min-cut may not be good:



Normalized Cut:

Min-cut fails because it gives two partitions where one is tiny and the other is huge.

To make the two partitions in similar sizes, define

$$\mathrm{vol}(A) = \sum_{\substack{i \in A \\ j \in V}} w_{ij} \quad \text{—— the size of } A$$

and

$$\mathrm{Ncut}(A,B) = \frac{\mathrm{cut}(A,B)}{\mathrm{vol}(A)} + \frac{\mathrm{cut}(A,B)}{\mathrm{vol}(B)} \quad \text{—— normalized cut}$$

Therefore, we solve

$$\boxed{\min_{S \in V} \mathrm{Ncut}(S, \bar{S}), \qquad \text{where } \bar{S} = V \setminus S.}$$

Again, this normalized cut can be reformulated as an eigenvalue problem.

Let $z \in \mathbb{R}^m$ be

$$z_i = \begin{cases} \frac{1}{\mathrm{vol}(S)} & \text{if } i \in S \\ -\frac{1}{\mathrm{vol}(\bar{S})} & \text{if } i \in \bar{S} \end{cases}$$

Then, $z^T L z = z^T (D - W) z = \sum_{i,j} w_{ij} (z_i - z_j)^2$

$$= \sum_{i \in S} w_{ij} \left(\frac{1}{\mathrm{vol}(S)} + \frac{1}{\mathrm{vol}(\bar{S})} \right)^2$$

$j \in \bar{S}$

and $z^T D z = \sum_i d_i z_i^2 = \sum_{i \in S} \frac{d_i}{(vol(S))^2} + \sum_{j \in \bar{S}} \frac{d_j}{(vol(\bar{S}))^2} = \frac{1}{vol(S)} + \frac{1}{vol(\bar{S})}$

Therefore $Ncut = \frac{z^T L z}{z^T D z}$

So $\min_z \frac{z^T L z}{z^T D z}$

s.t. $z_i \in \left\{ \frac{1}{vol(S)}, -\frac{1}{vol(\bar{S})} \right\}^m$

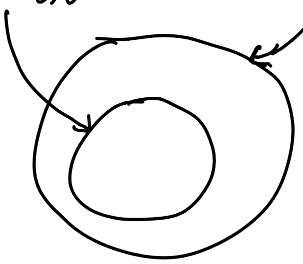
hard to handle

$z_i \in \left\{ \frac{1}{vol(S)}, -\frac{1}{vol(\bar{S})} \right\}^m$

$$\Rightarrow \boxed{\begin{aligned} z^T D 1 &= \sum_i d_i z_i \\ &= \sum_{i \in S} \frac{d_i}{(vol(S))} - \sum_{j \in \bar{S}} \frac{d_j}{(vol(\bar{S}))} \\ &= 1 - 1 = 0 \end{aligned}}$$

We relax the constraint

$$z_i \in \left\{ \frac{1}{vol(S)}, -\frac{1}{vol(\bar{S})} \right\}^m \xrightarrow{\text{relaxation}} z^T D 1 = 0$$



(we solve the minimization in a larger set)

We obtain: $\min_z \frac{z^T L z}{z^T D z}$ s.t. $z^T D 1 = 0$.

Let $y = D^{1/2} z$.

Assume the graph is connected, so that D is invertible. $\Rightarrow z = D^{-1/2} y$

Then $\frac{z^T L z}{z^T D z} = \frac{y^T D^{-1/2} L D^{-1/2} y}{y^T y}$

and $y^T (D^{1/2} 1) = 0$

We solve $\min_y \frac{y^T D^{-1/2} L D^{-1/2} y}{y^T y}$ s.t. $y^T (D^{1/2} 1) = 0$ (NCut-R)

$\tilde{L} = D^{-1/2} L D^{-1/2}$ is called normalized Laplacian

① $\tilde{L}$ is spsd.

② 0 is the smallest eigenvalue of $\tilde{L}$ and the associated eigenvector is $D^{1/2} 1$.

A similar proof reveals:

Theorem: A solution to (Ncut-R) is the eigenvector of the 2nd smallest eigenvalue of $\tilde{L} = D^{-1/2} L D^{-1/2} = D^{-1/2}(D-W)D^{-1/2}$

$$= I - D^{-1/2} W D^{-1/2}.$$

Also, λ_{n-1} is the cost of N-cut.

In summary, in normalized cut.

① Construct w_{ij} (similarity between i and j), hence $W = [w_{ij}]_{i,j}$

② $D = \mathrm{diag}(W\mathbf{1})$

③ $\tilde{L} = I - D^{-1/2} W D^{-1/2}$

④ Compute the eigenvector y corresponding to λ_{n-1} of $\tilde{L}$.

⑤ $z = D^{-1/2} y$.

⑥ Define the two groups by:

$S = \{ i \mid z_i > 0 \}$ $\quad$ $\bar{S} = \{ i \mid z_i < 0 \}$

If we want more groups, then:

① Use N-cut get S_1, S_2.

② Apply N-cut to S_1 and S_2 to get

S_{11}, S_{12} and S_{21}, S_{22}

and the cost of Ncut: $\lambda_{n-1}^{(1)}$ $\quad$ $\lambda_{n-1}^{(2)}$

If $\lambda_{n-1}^{(1)} \le \lambda_{n-1}^{(2)}$, $\quad S_{11}, S_{12}, S_2$

If $\lambda_{n-1}^{(1)} > \lambda_{n-1}^{(2)}$, $\quad S_1, S_{21}, S_{22}$

③ Apply N-cut to the resulting 3-groups and so on.

Results: