March 14, 2019

- Bisection
- RatioCut
- Normalized Cut

Workflow:

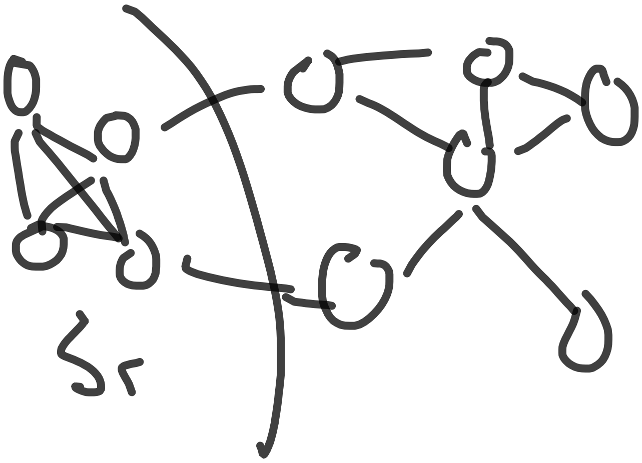
1. Combinatorial problem
2. relax ⟹ tractable (spectral)
3. round from relaxed solution to comb. one

Split graph into two

What about split into k?

Objective function

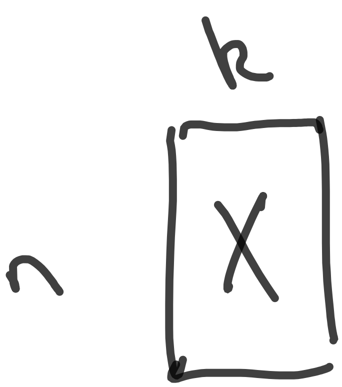
$$\mathrm{NCUT}(S_1, S_2, \ldots, S_k) = \sum_{r=1}^{k} \frac{\mathrm{cut}(S_r)}{\mathrm{vol}(S_r)}$$



cut = # edges leaving the set

$$\mathrm{vol}(S_r) = \sum_{i \in S_r} d_i$$

$$X_{ij} = \begin{cases} 1/\sqrt{\mathrm{vol}(S_j)} & i \in S_j \\ 0 & i \notin S_j \end{cases}$$



Tutorial von Luxburg 2007

Ng, Jordan, Weiss 2001

Claim: $\mathrm{trace}(X^T L X)$

$= NCUT(S_1, \ldots, S_k)$

$\mathrm{trace}(M) = \sum_i M_{ii}$

Proof: $y = X[:, j]$

$$y^T L y = \sum_{\substack{k \in S_j \\ l \notin S_j \\ (k,l) \in E}} (y_k - y_l)^2$$

$$= \sum \left(\frac{1}{\sqrt{vol(S_j)}} - 0 \right)^2$$

$$= \frac{1}{vol(S_j)} \sum 1 = \frac{cut(S_j)}{vol(S_j)}$$

$$\mathrm{trace}(X^T L X) = \sum_j X[:, j]^T L X[:, j]$$

$\square$

$$X_{ij} = \begin{cases} 1/\sqrt{vol(S_j)} & i \in S_j \\ 0 & o/w \end{cases}$$

Claim: $X^T D X = I$

Proof: $e_i^T X^T L X e_j$

$$= 0 \quad \text{if } i \neq j$$

$$x_j^T D x_j = \sum_{k \in S_j} d_k x_k^2$$

$$= \sum_{k \in S_j} d_k \frac{1}{vol(S_j)}$$

$$= \frac{1}{vol(S_j)} \cancel{\sum_{k \in S_j} d_k} \; vol(S_j)$$

$$= 1 \qquad \square$$

Optimization

NCUT

$$\min_{\substack{X \\ S_1, \dots, S_k}} \operatorname{trace}(X^T L X)$$

$$\text{s.t.} \quad X^T D X = I$$

relax ⇒ tractable

$$\cancel{X_{ij} = \begin{cases} 1/\sqrt{vol(S_j)} & i \in S_j \\ 0 & i \notin S_j \end{cases}}$$

NP-hard

$$Z = D^{1/2} X$$

$$\min \operatorname{trace}(Z^T \underbrace{D^{-1/2} L D^{-1/2}}_{\substack{N \\ \text{normalized Laplacian}}} Z)$$

$$\text{s.t.} \quad Z^T Z = I$$

Z

Pick k vectors $z_1, \ldots, z_k$

$$\min_r \left\{ \sum_{j=1}^{k} z_j^T N z_j \right.$$

s.t. $\{z_j\}$ orthonormal

Z consists of the first k eigenvectors of N

$$N V_k = \Lambda_k V_k$$

$$X^* = D^{-1/2} V_k$$

X^* embeds each node into $\mathbb{R}^k$

need to round

most common is k-means

Random walks and NCUT

Graph is conn. & not bipartite

Stationary dist π

1. Pick a starting following π
2. Take one RW step

Claim: $NCUT(S,T)$
= probability that we switch clusters 1 $\to$ 2

Proof: A, B disjoint node sets

z_1 state of random process after step 1

z_2 state after step 2

$$\Pr(z_1 \in A,\ z_2 \in B)$$

$$\pi_i = \frac{d_i}{2m}$$

$$= \sum_{\substack{i \in A \\ j \in B \\ (i,j) \in E}} \pi_i P_{ji}$$

$$P_{ji} = \frac{1}{d_i}$$

$$= \sum \frac{d_i}{2m} \frac{1}{d_i}$$

$$= \frac{1}{2m} \sum 1 = \frac{1}{2m} \operatorname{cut}(A, B)$$

$$\Pr(z_2 \in T \mid z_1 \in S)$$

$$\frac{\operatorname{cut}(S,T)}{2m}$$

$$= \frac{\Pr(z_1 \in S,\ z_2 \in T)}{\Pr(z_1 \in S)}$$

$$\Pr(z_1 \in S) = \sum_{i \in S} \pi_i = \sum \frac{d_i}{2m}$$

$$\Rightarrow \quad \frac{\operatorname{cut}(S,T)}{\sum_{i \in S} d_i} \leftarrow \operatorname{vol}(S)$$

$$= \frac{\operatorname{cut}(S,T)}{\operatorname{vol}(S)}$$

Modularity

$$x = \begin{cases} 1 & i \in S \\ -1 & i \in T = V \setminus S \end{cases}$$

The modularity of the assignment x is

$$Q(x) = \sum_{1 \le i,j \le n} \underbrace{\left[A_{ij} - \frac{d_i d_j}{2m} \right]}_{B} \mathbb{I}(x_i = x_j)$$

B_{ij} is how "surprising" link (i,j) is in the clustering

Want x to <u>maximize</u> the modularity

$$\mathbb{I}(x_i = x_j) = \frac{1}{2}(x_i x_j + 1)$$

$$Q(x) = \sum B_{ij}(x_i x_j + 1)\frac{1}{2}$$

$$\max Q(x)$$

$$\Leftrightarrow \max \sum B_{ij} x_i x_j = x^T B x$$

$$\text{s.t.} \quad \cancel{x_i = \begin{cases} 1 & i \in S \\ -1 & i \notin S \end{cases}}$$

relax $\Rightarrow$ tractable

$$x^T x = n$$

NP-hard

$$\max \quad x^T B x$$
$$\text{s.t.} \quad \|x\|_2^2 = n$$

$$B = V \Lambda V^T$$

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

$$\lambda_1 < \lambda_2 < \cdots < \lambda_n$$

Choose x to be evec of largest positive eigenvalue

round: $+$ entry $\Rightarrow S$

$-$ entry $\Rightarrow T$

$$B_{ij} = A_{ij} - \frac{d_i d_j}{2m}$$

no longer sparse!

$$B = \underbrace{A}_{\text{sparse}} - \underbrace{\frac{1}{2m} d d^T}_{\text{low-rank}}$$

we can apply B to a vector in time nnz(A)

$$NCUT(S,T) = \frac{cut(S)}{vol(S)} + \frac{cut(T)}{vol(T)}$$

Claim:

$$Q(x) = -2\,\mathrm{cut}(S) + \frac{1}{m}\mathrm{vol}(S)\,\mathrm{vol}(T)$$

$$\min 2\,\mathrm{cut}(S) - \frac{1}{m}\mathrm{vol}(S)\,\mathrm{vol}(T)$$

Proof: Gleich & Mahoney

Mining Large Graphs