(1) Graph Matrices

weighted adjacency matrix A

$\rightarrow$ symmetric (undirected)

$\rightarrow$ $a_{ij} \geq 0$

degree matrix $\Delta = \begin{pmatrix} d_1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n \end{pmatrix}$ $\quad \Delta^{-1} = \begin{pmatrix} 1/d_1 & & & 0 \\ & 1/d_2 & & \\ & & \ddots & \\ 0 & & & 1/d_n \end{pmatrix}$

$d_i = \sum_{j=1}^{n} a_{ij}$

normalized A (Markov matrix): M

$M = \underline{\Delta}$

row stochastic matrix

$m_{ij} = P(x_i \rightarrow x_j)$

Probability of transition from x_i to x_j

each row sums to 1.

$\lambda_1 = 1$

$u_1 = \begin{pmatrix} \\ \\ \end{pmatrix}$

Laplacian matrix

$$\underline{L = \Delta - A}$$

$$L = \begin{pmatrix} d_1 - a_{11} & -a_{12} & -a_{13} & \cdots & -a_{1n} \\ -a_{21} & d_2 - a_{22} & -a_{23} & \cdots & \\ \vdots & & \ddots & & \\ & & & & \end{pmatrix}$$

$$\begin{pmatrix} & \vdots & \ddots & & \\ & & & \ddots & d_n - a_{nn} \end{pmatrix}$$

L is positive semi-definite

every row of L sums to 0

L is PSD

eigenvalues $\lambda_i \geq 0$

real

$$\lambda_1 \geq \lambda_2 \ldots \geq \boxed{\lambda_n = 0}$$

normalized Laplacian

1) asymmetric : $L^a = \Delta^{-1} L$

acts like PSD

real, non-negative eigenvalue ~~too~~!

$$L^a = \begin{pmatrix} \frac{d_1 - a_{11}}{d_1} & -\frac{a_{12}}{d_1} & -\frac{a_{13}}{d_1} & \cdots & -\frac{a_{1n}}{d_1} \\ & & & & \end{pmatrix}$$

2) Symmetric :

$$L^s = \Delta^{-1/2} L \Delta^{-1/2}$$

row

$$\Delta^{-1/2} = \begin{pmatrix} & 0 \\ 0 & \end{pmatrix}$$

$$= \begin{pmatrix} \frac{d_1 - a_{11}}{\sqrt{d_1} \cdot \sqrt{d_1}} & -\frac{a_{12}}{\sqrt{d_1}\sqrt{d_2}} & \cdots & -\frac{a_{1n}}{\sqrt{d_1}\sqrt{d_n}} \\ -\frac{a_{21}}{\sqrt{d_1}\sqrt{d_2}} & \frac{d_2 - a_{22}}{\sqrt{d_2}\sqrt{d_2}} & \cdots & -\frac{a_{2n}}{\sqrt{d_2}\sqrt{d_n}} \end{pmatrix}$$

L^s also symmetric

symmetric
positive-semidefinite

---

graph clustering

→ partitioning

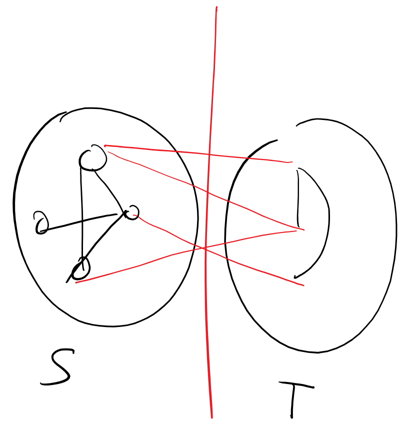
→ k-way

→ balance criteria (clusters should not be too small)

Given A : weighted adjacency matrix

W : cut-weight function

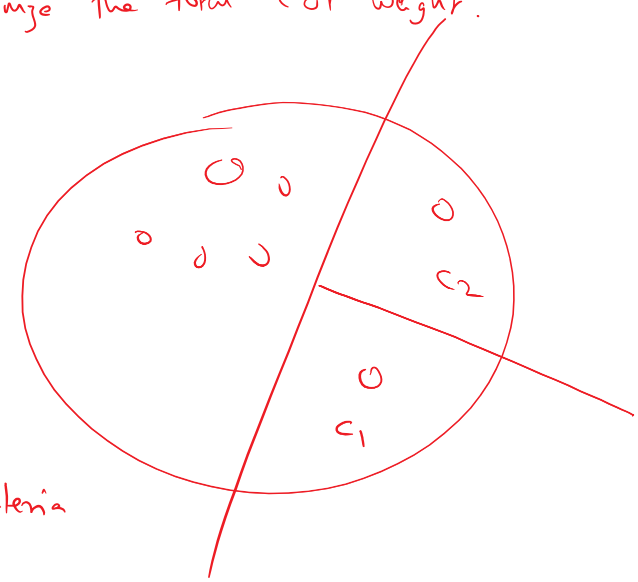
$$\boxed{W(S,T) : \sum_{i \in S} \sum_{j \in T} a_{ij}}$$



$a_{ij} \equiv$ affinity / similarity

k-way cuts to minimize the total cut weight.

trivial clustering



we need balance criteria

① Ratio cut objective : $G=(V,E)$, A : adj matrix, k : # of clusters

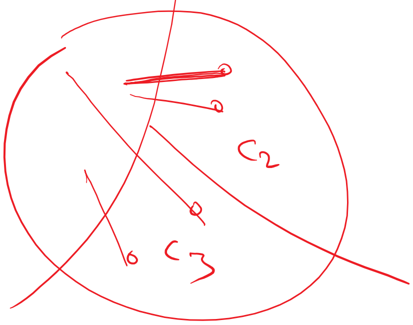
(1) Ratio objective : $G=(V,E)$, A : adj matrix , k : # of clusters

$$\zeta = \{c_1, c_2, \ldots, c_k\}$$ k-way

$$c_i \cap c_j = \phi \,, \quad \bigcup c_i = V$$

$$\boxed{\min_{\{c_1, c_2, \ldots c_k\}} J_{rc} = \sum_{i=1}^{k} \frac{W(c_i, \bar{c}_i)}{|c_i|}}$$

minimize sum of external weights & maximize the size



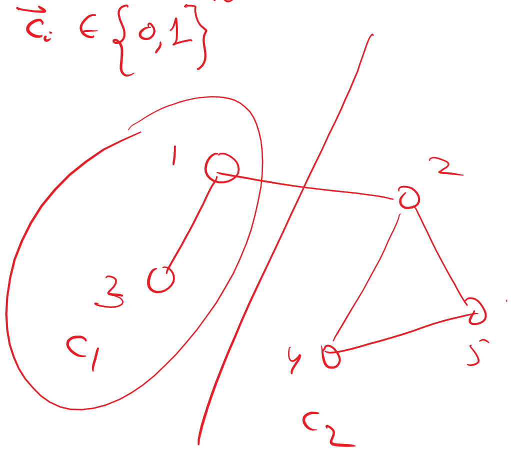
$$\bar{c}_i = V - c_i$$

$|c_i|$ = # of vertices in c_i

define a binary cluster indicator vector

$$\vec{c}_i \equiv c_{ij} = \begin{cases} 1 & \text{if } x_j \in c_i \\ 0 & \text{otherwise} \end{cases}$$

$\vec{c}_i \in \{0,1\}^n$



$$\vec{c}_1 = \begin{pmatrix} \\ \\ \\ \end{pmatrix} \qquad \vec{c}_2 = \begin{pmatrix} \\ \\ \\ \end{pmatrix}$$

$$|c_i| = \vec{c}_i^{\,T} c_i$$

dot product of cluster indicators

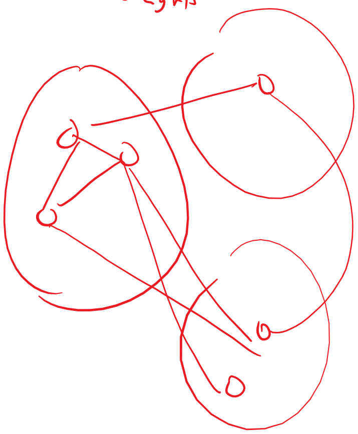
$$\underbrace{W(c_i, \bar{c}_i)}_{\text{total external weights}} = W(c_i, V) - W(c_i, c_i)$$

total external weights

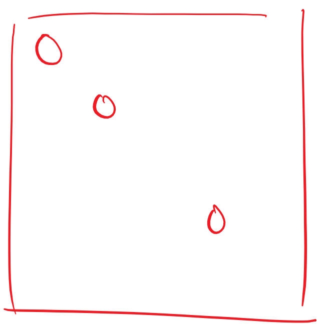
total weight for c_i (internal edges + external edges)

total internal weights

c_1



$$W(c_i, V) = \sum_{j \in c_i} \sum_{k \in V} a_{jk}$$



$$= \sum_{\in c_i} d_j$$

$$= \underbrace{\vec{c}_i^T \Delta c_i}_{\text{left/right multiplication leads to selection of only the cluster elements.}}$$

$$W(c_i, c_i) = \sum_{j \in c_i} \sum_{k \in c_i} a_{jk}$$

$$= \vec{c}_i^T \left(A \right) \vec{c}_i$$

$$W(c_i, \bar{c}_i) = W(c_i, V) - W(c_i, c_i)$$

$$= \vec{c}_i^T \Delta \vec{c}_i - \vec{c}_i^T A \vec{c}_i$$

$$= \vec{c}_i^T \underbrace{\left(\Delta - A \right)}_{} \vec{c}_i$$

$$= \vec{c}_i^T L \vec{c}_i$$

$$= \vec{c}_i^{\,T} L \vec{c}_i$$

$$\min_{\{\vec{c}_1, \vec{c}_2, \dots, \vec{c}_k\}} J_{rc} = \sum_{i=1}^{k} \frac{\vec{c}_i^{\,T} L \vec{c}_i}{\vec{c}_i^{\,T} \vec{c}_i}$$

Ratio-cut

what are the $\vec{c}_i$'s? k of them

$\vec{c}_i \in \{0/1\}^n$: binary vectors

$\|\vec{c}_i\|^2 = \vec{c}_i^{\,T} \vec{c}_i$

NP-hard

relax the condition that $\vec{c}_i$ have to binary

$\vec{c}_i \in \mathbb{R}^n$

$$\sum_{i=1}^{k} \frac{\vec{c}_i^{\,T} L \vec{c}_i}{\|\vec{c}_i\|^2} = \sum_{i=1}^{k} \left(\frac{\vec{c}_i}{\|c_i\|}\right)^T L \left(\frac{\vec{c}_i}{\|c_i\|}\right)$$

relaxed $\min_{\{u_1, u_2, \dots u_k\}} J_{rc} = \sum_{i=1}^{k} u_i^T L u_i$

$u_i \in \mathbb{R}$

s.t. Constraint that $u_i^T u_i = 1$.

solutions are the eigenvectors of L

$$\frac{\partial\, u_i^T L u_i - \lambda_i (u_i^T u_i - 1)}{\partial u_i}$$

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n = 0$$

$$2 L u_i - 2\lambda_i u_i = 0$$

$$\Rightarrow \boxed{L u_i = \lambda_i u_i}$$

k-smallest eigen-value

$u_{n-k+1} \dots u_n$

... smallest eigen-value

$\equiv \boxed{Lu_i = \lambda_i u_i}$

$$\underbrace{u_{n-k+1} \cdots u_n}_{\text{Optimal solution}}$$

$\to$ to the relaxed problem

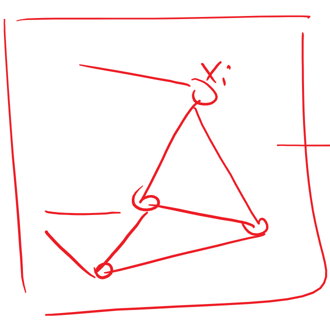
$A \to L \equiv \Delta - A$

$\lambda \qquad \lambda_n = 0$

$$\underbrace{\boxed{\vec{u}_{n-k+1} \quad \vec{u}_{n-1} \quad \tilde{}}}_{k\text{ - eigenvectors}} \in \mathbb{R}^n$$

$\Longrightarrow$ how to convert to binary $\vec{c}_i$ vectors

G



L

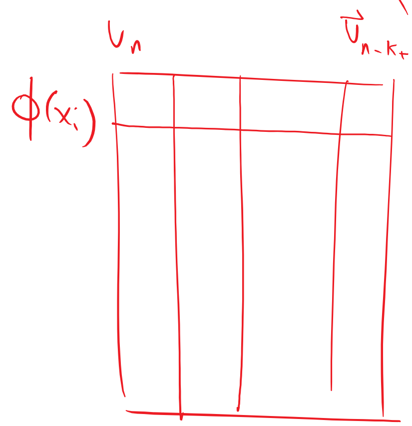
L^a

L^s

$\xrightarrow{\phi}$

graph embedding

k dimensions



orthogonal to each other

feature space "explicit"

Q

$\Downarrow$

k-means

$\Downarrow$

extract the k-clusters

Normalized cut objective

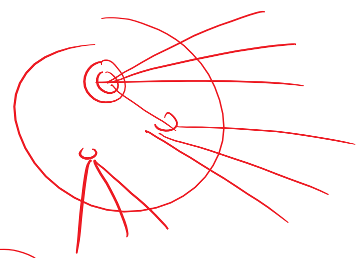
min $\bar{J}_{nc} = \sum^{k} \frac{W(C_i, \bar{C_i})}{vol(C_i)}$

$|C_i|$ = # of vertices

$Vol(C_i) = W(C_i, V) = \sum_{j \in C_i} d_j$

→ taking the weights into account

$$J_{nc} = \sum_{i=1}^{k} \frac{\vec{c}_i L \vec{c}_i}{\vec{c}_i \Delta c_i}$$



$\frac{1}{\Delta}(L)$

$\frac{L}{\Delta} \rightarrow \Delta^{-1} L \rightarrow \sum \vec{u}_i^{T} \Delta^{-1} L \vec{u}_i$

$\frac{L}{\Delta^{1/2} \cdot \Delta^{1/2}} \rightarrow \Delta^{-1/2} L \Delta^{-1/2} \rightarrow \sum \vec{u}_i^{T} (\Delta^{-1/2} L \Delta^{-1/2}) \vec{u}_i$

Solving J_{nc} → k smaller eigen-values/vectors of $\Delta^{-1} L = L^{a}$

→ " " $\Delta^{-1/2} L \Delta^{-1/2} = L^{s}$

Modularity

$\left(\underbrace{\text{observed probability of edges within a cluster}}_{\text{observe}} - \underbrace{\text{expected probability}}_{\text{null model}} \right)$

$$J_m = \left(\frac{w(C_i, C_i)}{w(C_i, V)} - \left(\frac{w(C_i, V)}{w(V, V)} \right)^2 \right)$$

Prob of internal edge weights

$$= \sum_{i=1}^{k} \left(\left(\frac{c_i^T A c_i}{c_i^T \Delta c_i} \right) - \frac{\left(c_i^T \Delta c_i \right)^2}{\underline{tr(\Delta)^2}} \right)$$

$$\underset{=}{\max} \quad J_m = \sum_{i=1}^{k} c_i^T Q \vec{c}_i \longleftarrow \underline{\underline{K \text{ largest}}} \text{ eigenvalue \& eigenvectors of } Q$$

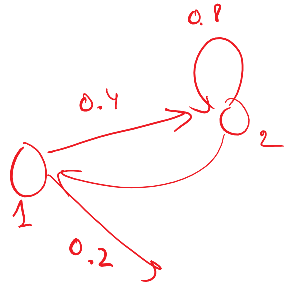
Modularity matrix

extract upto k positive eigenvals

Q : symmetric but not PSD

<u>Markov chain clustering (MCL)</u>

$M = \Delta^{-1} A$: transition probability matrix



m_{1j} = prob of jump from 1 to j
$= P(j \mid 1)$ after 1 step

1) $= P(j|1)$ after 1 step

row stochastic $\equiv$ every row is a prob. vector

$$\boxed{M^2(1,j) = P(j|1) \text{ after 2 steps..}}$$

M^t: t-step markov matrix

$$\equiv M^t_{ij} : P(x_i \to x_j \mid t\text{-steps})$$

$$\pi_0 = \underbrace{(0.2, 0.1, \dots)}_{\text{initial starting vector}}{}^T$$

$P(\text{start at state/node } x_i)$

$\pi_1, \pi_2, \dots \pi_\infty$

$$\pi_\infty = \text{dominant eigenvector of } M^T$$

Page-rank

---

1. walk for one step (markov step) $M = M \times M$
2. Inflation: increase high prob, decrease low prob

$$M' = \Gamma(M, r) = \text{every element to power } r$$

normalize each row

M

$$\begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 0.4 & 0.6 \\ 2 & 0.1 & 0.9 \end{array}$$

1 0.4 0.6

Normalize each row

$\gamma = 2$

$$\begin{pmatrix} 0.16 & 0.36 \\ 0.01 & 0.81 \end{pmatrix} \longrightarrow \begin{pmatrix} 0.16/0.52 & 0.36/0.52 \\ 0.01/0.82 & 0.81/0.82 \end{pmatrix} \longrightarrow \begin{pmatrix} & \\ 0.01 & 0.99 \end{pmatrix}$$

③ after Convergence

$M : \quad M_{ij} \longrightarrow$ edge from $v_i \xrightarrow[M_{ij}]{} v_j$

directed graph.

→ weakly connected components

→ clusters

γ : Inflation parameter

- small γ close to 1 → large / few clusters
- large γ → very many small clusters