Scan the QR to record your attendance

Quick review – Graph Clustering

- Data clustering → Graph (KNN) → graph partitioning (balanced $\min \frac{Cut}{Vol}$)
- $Cut(S, S^c)$, $Vol(S)$, $Assoc(S, S^c)$ … Cheeger Cut, Normalized Cut, Normalized Association

Normalized Association

Indicator F

$$\max_{F\in\{0,1\}^{n\times k}} \sum_{q=1}^{k} \frac{F_{\cdot,q}^T A F_{\cdot,q}}{F_{\cdot,q}^T D F_{\cdot,q}} \quad \text{s.t.} \quad \sum_{q=1}^{k} F_{i,q} = 1 \ \forall i \in V$$

$$\max_{Y\in \text{ binary}^{n\times k}} \operatorname{tr}(Y^T B Y) \ \text{ s.t. } \ Y^T Y = \mathrm{I}_k, \ B = D^{-1/2} A D^{-1/2}$$

$$\text{with } Y_{\cdot,q} = \frac{D^{1/2} F_{\cdot,q}}{\|D^{1/2} F_{\cdot,q}\|_2} \quad \text{(vectorial representation)}$$

Same as Spectral Clustering

$$\text{with } \ B = \Theta^{1/2} A \Theta^{1/2} \overset{\text{EVD}}{=} U \Lambda U^T \in \mathbb{R}^{n\times n}$$

$$\text{and solution } \ Y^\star = U_{\cdot,1:k} \in \mathbb{R}^{n\times k} \quad (k \text{ largest eigenvectors})$$

Normalized Cut

$$\min_{F\in\{0,1\}^{n\times k}} \sum_{q=1}^{k} \frac{F_{\cdot,q}^T L F_{\cdot,q}}{F_{\cdot,q}^T D F_{\cdot,q}}, \quad \text{with } L = D - A \ \text{ and } \ \sum_{q=1}^{k} F_{i,q} = 1 \ \forall i \in V$$

$$\min_{Y\in \text{ binary}^{n\times k}} \operatorname{tr}(Y^T B Y) \ \text{ s.t. } \ Y^T Y = \mathrm{I}_k, \ B = \mathrm{I} - D^{-1/2} A D^{-1/2}$$

$$\text{with } Y_{\cdot,q} = \frac{D^{1/2} F_{\cdot,q}}{\|D^{1/2} F_{\cdot,q}\|_2} \quad \text{(vectorial representation)}$$

k smallest eigenvectors

$$Y^\star = \arg \min_{Y\in\mathbb{R}^{n\times k}} \operatorname{tr}(Y^T B Y)$$

Rotate, binary constraint

$$Z^\star = \arg\min_{Z,R} \|Z - Y^\star R\|_F^2 \quad \text{s.t.} \quad R^T R = \mathrm{I}_k, \ Z \in \{0,1\}^{n\times k}$$

R: rotation matrix

Quick review – Graph Clustering

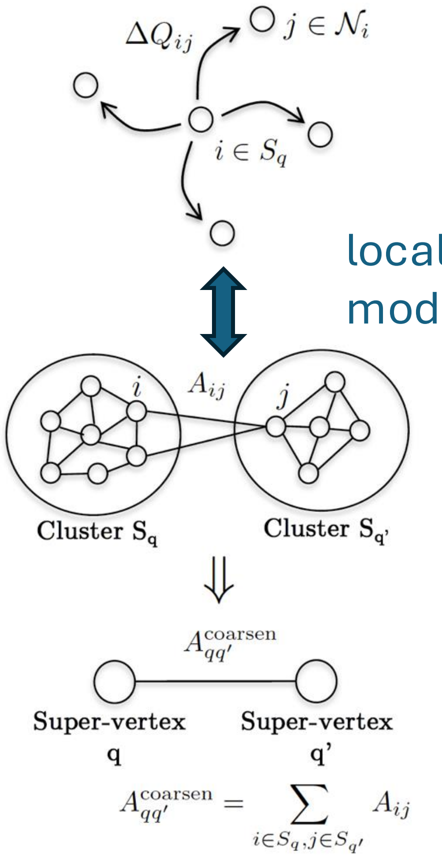
- Metis: coarsen, uncoarsen, graph partitioning & data clustering
- Product cut: robust to cluster outliers
- Louvain algorithm: unknown k, optimize modularity
  - Modularity: actual partition & random edges

$$\max_{k, C: V \to \{1,2\ldots,k\}} \sum_{ij} \left(A_{ij} - \gamma \frac{d_i d_j}{\sum_{i'j'} A_{i'j'}} \right) \delta(C_i, C_j)$$

$$\text{with} \quad \delta(C_i, C_j) = \begin{cases} 1 & \text{if } C_i = C_j \ (i \text{ and } j \text{ belong to the same cluster}) \\ 0 & \text{otherwise} \end{cases}$$

$$\min_{\{S_q\}_{q=1}^k \text{ s.t. } \cup_q S_q = V,\ \cap_q S_q = \emptyset} \sum_{q=1}^{k} \mathrm{Cut}(S_q, S_q^c) - \gamma \mathrm{Vol}(S_q) \mathrm{Vol}(S_q^c)$$

$$\min_{\{S_q\}_{q=1}^k \text{ s.t. } \cup_q S_q = V,\ \cap_q S_q = \emptyset} \sum_{q=1}^{k} \frac{\mathrm{Cut}(S_q, S_q^c)}{\mathrm{Vol}(S_q)\mathrm{Vol}(S_q^c)}, \quad \text{where } \gamma = \min_{\{S_q\}} \mathrm{LCut}(\{S_q\})$$



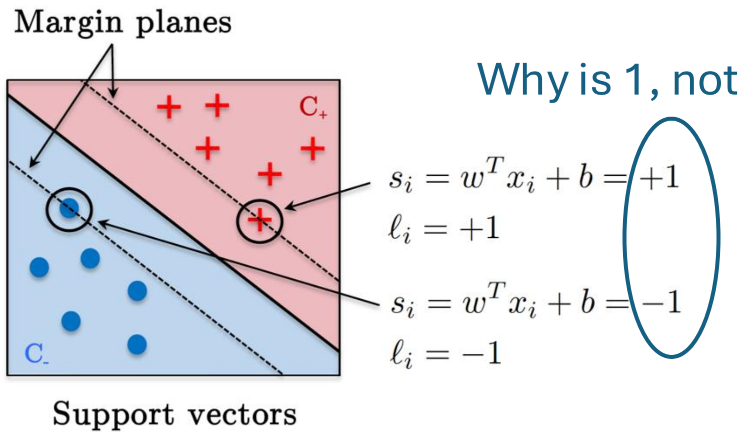
locally maximize modularity

$\gamma \uparrow$ -- Cut $\uparrow$, Vol $\downarrow$ -- $N_{cluster} \uparrow$

$\gamma \downarrow$ -- Cut $\downarrow$, Vol $\uparrow$-- $N_{cluster} \downarrow$

γ: a loose control on $N_{cluster}$

Quick review – Graph SVM



Why is 1, not 2?

- Classification: $s_i = w^\top x + b \to l_i \in \{-1, 1\}$
- Maximize class margins ←→ minimize $||w||^2$ & correct classification

$$\min_w \|w\|_2^2 \text{ s.t. } \ell_i . s_i \geq 1, \; \forall i \in V$$

b*: use support vectors & w*

Dual variable
data & model → only data

$$\text{Given } w = \sum_i \alpha_i \ell_i x_i \in \mathbb{R}^d, \alpha_i \in \mathbb{R}$$

$$\begin{aligned} \text{we have } w^T x &= \sum_i \alpha_i \ell_i x_i^T x \in \mathbb{R} \\ &= \sum_i \alpha_i \ell_i K(x_i, x) \text{ with } K(x_i, x) = x_i^T x \\ &= \alpha^T L K(x), \;\; \alpha, K(x) \in \mathbb{R}^n, L \in \mathbb{R}^{n\times n} \end{aligned}$$

$$\begin{aligned} f_{\text{SVM}}(x) &= \text{sign}(w^T x + b) \in \pm 1 \quad \text{(with primal variable)} \\ &= \text{sign}(\alpha^T L K(x) + b) \in \pm 1 \quad \text{(with dual variable)} \end{aligned}$$

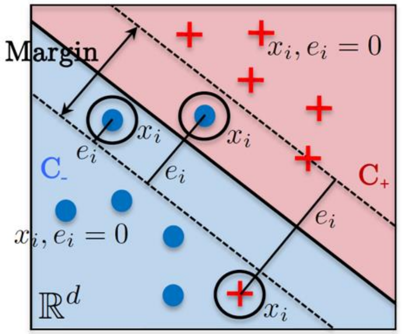
$$\min_{\alpha \geq 0} \frac{1}{2} \alpha^T Q \alpha - \alpha^T 1_n \;\; \text{s.t. } \alpha^T \ell = 0$$

$$\begin{aligned} \text{with } Q &= LKL \in \mathbb{R}^{n\times n} \\ L &= \text{diag}(\ell) \in \mathbb{R}^{n\times n} \\ \ell &= (\ell_1, ..., \ell_n) \in \mathbb{R}^n \\ K &\in \mathbb{R}^{n\times n}, K_{ij} = x_i^T x_j \in \mathbb{R} \end{aligned}$$

Dual optimization problem
Primal-dual iterative scheme

Quick review – Graph SVM

- Soft-margin SVM – non-linearly separable
  - Slack variable e_i, min error & max margin
  - Penalize
    - Misclassification
    - Correct but not confident



Error e_i

$$\min_{w,e} \; \|w\|_2^2 + \lambda \sum_{i=1}^{n} \underline{\max\left(0, 1-\ell_i s_i\right)},$$

$$\text{where} \quad s_i = w^T x_i + b \quad \text{(score function)}$$

$$L_{\text{Hin}}(d_i) = \max\left(0, 1-d_i\right), \; d_i = \ell_i s_i \quad \text{(Hinge loss)}$$

$$\min_{0 \le \alpha \le \lambda} \frac{1}{2}\alpha^T Q \alpha - \alpha^T 1_n \quad \text{s.t.} \;\; \alpha^T \ell = 0 \quad \text{(dual QP problem)}$$

$$\text{with } Q = LKL \in \mathbb{R}^{n \times n}$$

$$L = \text{diag}(\ell) \in \mathbb{R}^{n \times n}$$

$$\ell = (\ell_1, \ldots, \ell_n) \in \mathbb{R}^n$$

$$K \in \mathbb{R}^{n \times n}, K_{ij} = x_i^T x_j \in \mathbb{R} \quad \text{(linear kernel)}$$

- Non-linear / kernel SVM
  - K (linear) → Gaussian, Polynomial
  - Same as Kernel K-means

Core API

- DGL https://www.dgl.ai
  - g=dgl.graph((src, dst), num_nodes): construct graph
  - g.nodes(), g.edges(), g.num_nodes()
  - dgl.add_reverse_edges(g): make undirected graph
- Numpy
  - np.zeros(n), np.eye(n). A.dot(B), A.T
  - np.linalg.norm(M, order, axis)
- Scipy.sparse.linalg.cg(A, b, x_0, tol, maxiter): solve A x = b

Appendix

$$\max_{F \in \{0,1\}^{n \times k}} \sum_{q=1}^{k} \frac{F_{\cdot,q}^T A F_{\cdot,q}}{F_{\cdot,q}^T D F_{\cdot,q}} \quad \text{s.t.} \quad \sum_{q=1}^{k} F_{i,q} = 1 \ \forall i \in V$$

$$\max_{Y \in \text{ binary}^{n \times k}} \operatorname{tr}(Y^T B Y) \quad \text{s.t.} \quad Y^T Y = \mathrm{I}_k, \ B = D^{-1/2} A D^{-1/2}$$

$$\text{with } Y_{\cdot,q} = \frac{D^{1/2} F_{\cdot,q}}{\|D^{1/2} F_{\cdot,q}\|_2} \quad \text{(vectorial representation)}$$

$$\begin{aligned}
Y_{\cdot,q}^{\top} B Y_{\cdot,q} &= \frac{F_{\cdot,q}^{\top} D^{1/2} D^{-1/2} A D^{-1/2} D^{1/2} F_{\cdot,q}}{||D^{1/2} F_{\cdot,q}||_2^2} \\
&= \frac{F_{\cdot,q}^{\top} A F_{\cdot,q}}{F_{\cdot,q}^{\top} D^{1/2} D^{1/2} F_{\cdot,q}} \\
&= \frac{F_{\cdot,q}^{\top} A F_{\cdot,q}}{F_{\cdot,q}^{\top} D F_{\cdot,q}}
\end{aligned}$$

Assume $N_{edge} = m$

$$\begin{aligned}
&\sum_{ij} \left(A_{ij} - \gamma \frac{d_i d_j}{\sum_{i'j'} A_{i'j'}} \right) \delta(C_i, C_j) \\
&= \sum_{ij} \left(A_{ij} - \gamma \frac{d_i d_j}{2m} \right) \delta(C_i, C_j)
\end{aligned}$$

$$\sum_{ij} A_{ij} \delta(C_i, C_j) = Assoc(S_q, S_q^c)$$

$$Cut(S_q, S_q^c) = 2m - Assoc(S_q, S_q^c)$$

$$\begin{aligned}
\sum_q \sum_{ij} \frac{d_i d_j}{2m} \delta(C_i, C_j) &= \frac{Vol(s_q) Vol(s_q)}{2m} \\
&= \sum_q \frac{Vol(s_q)\left(2m - Vol(s_q^c)\right)}{2m} \\
&= \sum_q [Vol(s_q) - \frac{Vol(s_q) Vol(s_q^c)}{2m}] \\
&= 2m - \sum_q \frac{Vol(s_q) Vol(s_q^c)}{2m}
\end{aligned}$$

$$\max \sum_{ij} \left(A_{ij} - \gamma \frac{d_i d_j}{\sum_{i'j'} A_{i'j'}} \right) \delta(C_i, C_j)$$

$$=> \min \sum_q Cut(s_q, s_q^c) - \gamma Vol(s_q) Vol(s_q^c)$$