

Programming Assignment 2

Spectral Clustering on Star War

Dataset

- Dataset內含有69位Start War角色，分屬於三大政營
- 所給的adjacent_matrix.csv:
 - 描述這69位角色間的關係(0或1)
 - header和index是角色編號
- 令 $k=3$ ，請實做spectral clustering來將這些角色分成三個集團(communities)

作業要求

- **不可以**直接跑任何spectral clustering的API
- 請依照課程投影片：
 - 利用adjacent matrix建置ratio-cut Laplacian L
 - **允許**用eigenvector的API來取出 L 的前三小eigenvectors, 並用其來建置 \hat{X}
 - 最小的eigenvalue會為0!! 其eigenvector會是一個constant vector, 可取, 也可不用取這個eigenvector來建置 \hat{X} (也就是只取倒數第二與第三個eigenvector)
 - 建置 \hat{X} 後, 將每個row視為一個角色的representation vector, 且用k-means來將這些vectors(角色) 分成3群。
 - **K-means請自己刻, 不可跑任何API**

上傳文件

(csv 檔案中每個 column 是一個維度)

- 至多2頁的報告 (40%)
- 程式碼 (30%)
- 前三小eigenvectors (15%)
 - 由小到大，將前三小eigenvalue之eigenvector內容存成 *eigen.csv* 。不需四捨五入。
- Community detection結果 (15%)
 - 請依照角色的number，由小到大，依照spectral clustering結果標註其所屬政營，政營用數字代表，從0開始，且產生如下格式的*result.csv*檔案。我們會計算該community detection的NMI

最小

Vector1 (以逗號隔開)

第二小

Vector2 (以逗號隔開)

第三小

Vector3 (以逗號隔開)

0,2,0,1, ...

角色0的政營編號為0, 角色1的政營編號為2,
角色2的政營編號為0, 角色3的政營編號為1 ...

(csv 檔案中每個 column 是一個角色)

作業注意事項

- **No** late submission will be accepted.
- 上傳的 *eigen.csv* 和 *result.csv* 若格式有誤，則該項次不予計分。但可於一週內重新上傳結果與程式碼，分數會9折計算。（遲交者不適用）
- 除了計算eigenvalue、eigenvector可以使用package之外，其他計算皆不可使用套件，違者作業不算分。