

姓名:葉冠宏 學號:r11943113 社群媒體分析 hw2

在 spectral clustering 中，為了要去求出 ratio-cut Laplacian，我們先去讀取 adjacent_matrix-1-1 去得出 adjacency matrix A ，然後去計算各自 node 的 degree，得出矩陣 D 。最後經由 $L=D-A$ ，我們可以求出 ratio-cut Laplacian L 。我們經由 `api numpy.linalg` 去求出 L 的 eigen value 和 eigen vector。我們去排序 eigen value，並取出前三小的 eigen value 所對應的 eigen vector。我們去轉置前三小的 eigen vector 矩陣，得出 X ，把結果存到 eigen.csv 中。

我們建置 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。在 k mean 的演算法中，我們一開始先隨機取 3 個點來當作 centroid(中心點)，然後去計算每個點和該 3 個 centroid 的距離，並取最小的距離來當作那一個點所屬的 cluster。在下一個 iteration 中，我們去把每一個 cluster 內的點的平均位置來當作該 cluster 的新的 centroid 位置。再下一輪的 iteration 中，我們就去計算每個點和新的 centroid 的距離是多少，並選擇最小的距離來當作該點所屬的 cluster 是什麼。就這樣經過多個 iteration 的重複計算，我們可以得到一個 cluster 較精確的中心點的所屬位置，以及每一個點應該所屬的 cluster 是什麼。值得注意的是，如果 cluster 的新的 centroid 位置和舊的 centroid 位置如果距離差別不大的話，那 iteration 就會停止了。最後我們輸出每個點所屬的 cluster 結果於 result.csv。