Summary

常博愛　　資工三　408410086

學習內容簡介：

1. modularity
2. random walk
3. spectral clustering

一． modularity：

1. 定義：為了度量實際網路中社區探索方法的好壞，Newman[3]於 2004 年提 出了一個用於測度複雜網路中社區劃分品質指標——模組度(Modularity)。
2. 公式的涵義與解釋

Modularity第一版:

$$Q==\sum_i (e_{ii}-a_i^2)=Tr(e)-\| e^2 \|$$

Modularity第二版:

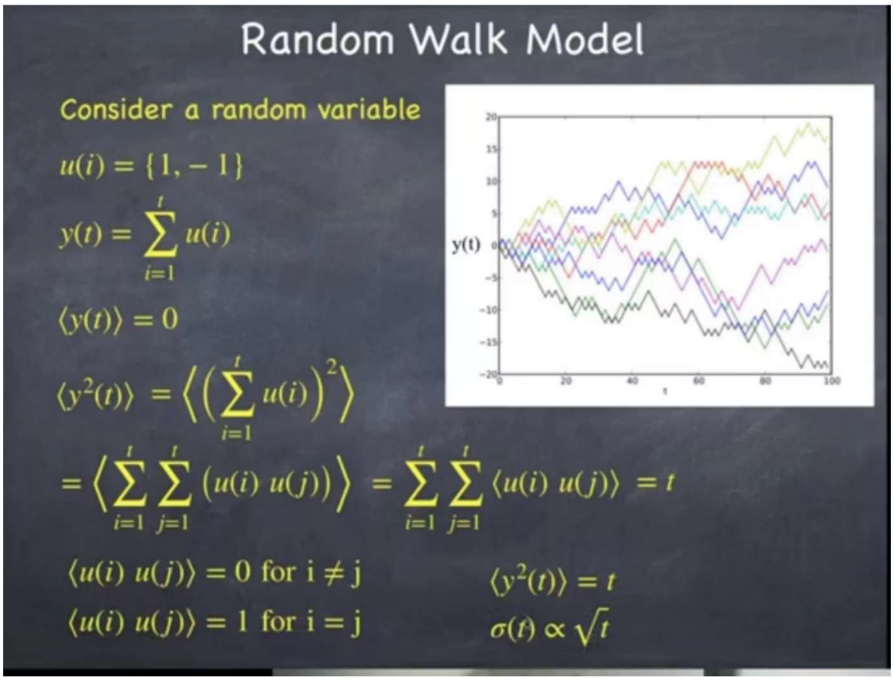
$$Q=\frac{1}{2m}\sum_{ij}(A_{ij}-\frac{k_i k_j}{2m})\delta(i,j)$$

3. resolution limit.

二．**Random walk**

1.Definition:隨機移動的物體從它們開始的地方遊走的過程

2.



4. Random walks in more than one dimension：
   1.Keep same rule as 1-D.
   2.More commonly.
   Note :more memory will be changed.(biased)
5. Biased random walk
6. Random walk in graph network：node embedding

三．Spectral clustering

1. 基于未標準化的拉普拉斯矩陣：

Unnormalized Graph Laplacian

L = D - W

Proposition 1 (Properties of L) The matrix L satisfies the following properties:

1. For every $f \in \mathbb{R}n$ we have

   1. For every $f \in \mathbb{R}^n$ we have

$$f'Lf = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}^2 (f_i - f_j)^2$$

2. L is symmetric and positive semi-definite.

3. The smallest eigenvalue of L is 0, the corresponding eigenvector is the constant one vector $\mathbb{1}$

4. L has n non-negative, real-valued eigenvalues $0 = \lambda i \leqslant \lambda 2 \leqslant \cdots \leqslant \lambda$

5.The multiplicity k of the eigenvalue 0 of L equals the number of connected components $A1, \cdots, Ak$ in the graph.

2. 基于标准化的拉普拉斯矩阵:

Normalized Graph Laplacian

$$L_{sym} := D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$$

$$L_{rw} := D^{-1} L = I - D^{-1} W$$

We denote the first matrix by $Lsym$ as it is a symmetric matrix, and the second one by Lrw as it is closely related to a random walk.

The normalized Laplacians statisfy the following properties:

1. For every $f \in \mathbb{R}n$ we have
2. λ is an eigenvalue of Lrw with eigenvector u if and only if λ is an eigenvalue of $Lsym$ with eigenvector $w = D$ 1/2u.
3. λ is an eigenvalue of Lrw with eigenvector u if and only if λ and u solve the generalized eigen problem $Lu = \lambda Du$.
4. 0 is an eigenvalue of Lrw with the constant one vector $\mathbb{1}$ as eigenvector. 0 is an eigenvalue of $Lsym$ with eigenvector D 1/2 $\mathbb{1}$.
5. $Lsym$ and Lrw are positive semi-definite and have n non-negative real-valued eigenvalues $0 = \lambda i \leqslant \lambda 2 \leqslant \cdots \leqslant \lambda n$.

Algorithms

Unnormalized spectral clustering:

L=D-S

Unnormalized spectral clustering

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct.

- Construct a similarity graph by one of the ways described in Section 2. Let W be its weighted adjacency matrix.
- Compute the unnormalized Laplacian L.
- **Compute the first k eigenvectors $u_1, \ldots, u_k$ of L.**
- Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns.
- For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of U.
- Cluster the points $(y_i)_{i=1,\ldots,n}$ in $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1, \ldots, C_k$.

Output: Clusters $A_1, \ldots, A_k$ with $A_i = \{j \mid y_j \in C_i\}$.

Normalized spectral clustering according to Shi and Malik:

$L_{rw}=D^{-1}L$

Normalized spectral clustering according to Shi and Malik (2000)

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct.

- Construct a similarity graph by one of the ways described in Section 2. Let W be its weighted adjacency matrix.
- Compute the unnormalized Laplacian L.
- **Compute the first k generalized eigenvectors $u_1, \ldots, u_k$ of the generalized eigenproblem $Lu = \lambda Du$.**
- Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns.
- For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of U.
- Cluster the points $(y_i)_{i=1,\ldots,n}$ in $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1, \ldots, C_k$.

Output: Clusters $A_1, \ldots, A_k$ with $A_i = \{j \mid y_j \in C_i\}$.

Normalized spectral clustering according to Ng, Jordan, and Weiss:

$$L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$$

Normalized spectral clustering according to Ng, Jordan, and Weiss (2002)

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct.

- Construct a similarity graph by one of the ways described in Section 2. Let W be its weighted adjacency matrix.
- Compute the normalized Laplacian L_{sym}.
- **Compute the first k eigenvectors $u_1, \ldots, u_k$ of L_{sym}.**
- Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns.
- **Form the matrix $T \in \mathbb{R}^{n \times k}$ from U by normalizing the rows to norm 1,** that is set $t_{ij} = u_{ij} / (\sum_k u_{ik}^2)^{1/2}$.
- For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of T.
- Cluster the points $(y_i)_{i=1,\ldots,n}$ with the k-means algorithm into clusters $C_1, \ldots, C_k$.

Output: Clusters $A_1, \ldots, A_k$ with $A_i = \{j \mid y_j \in C_i\}$.

2.CUT

Ratiocut:

$$RatioCut(G_1, G_2, \ldots, G_k) = \frac{1}{2} \sum_{i=1}^{k} \frac{W(G_i, G_i^C)}{|G_i|}$$

Ncut:

$$NormalizedCut(G_1, G_2, \ldots, G_k) = \frac{1}{2} \sum_{i=1}^{k} \frac{W(G_i, G_i^C)}{\sum_{v \in G_i} d_v}$$

Random walk：

A random walk on a graph is a stochastic process which randomly jumps from vertex to vertex.

• Random walk stays long within the same cluster and seldom jumps between clusters.

• A balanced partition with a low cut will also have the property that the random walk does not have many opportunities to jump between clusters.

• Transition probability pij of jumping from vi to vj

$$p_{ij} = w_{ij}/d_i$$

• The transition matrix P = (pij) i,j = 1,⋯,n of random walk is defined by

$$P = D^{-1}W$$

• If the graph is connected and non-bipartite, the random walk always processes a unique stationary distribution π = (π1, ⋯ , πn)′ , where πi = $di/vol(V)$.

• Relationship between Lrw and P.

$$L_{rw} = I - P$$

• λ is an eigenvalue of Lrw with eigenvector u if and only if 1 − λ is an eigenvalue of P with eigenvector u.

• The largest eigenvectors of P and the smallest eigenvectors of Lrw can be used to describe cluster properties of the graph.

• Random walks and Ncut

Proposition 5 (Ncut **via transition probabilities**) Let G be connected and non bi-partite. Assume that we run the random walk $(X_t)_{t\in N}$ starting with X_0 in the stationary distribution π. For disjoint subsets $A, B \subset V$, denote by $P(B|A) := P(X_1 \in B\,|X_0 \in A)$. Then:

$$Ncut(A, \bar{A}) = P(\bar{A}|A) + P(A|\bar{A}).$$

• A loose relation between spectral clustering and commute distance.

Spectral Clustering

1. Map the vertices of the graph on the rows yi of the matrix U

2. Only take the first k columns of the matrix

Commute Distance

1. Map the vertices on the rows zi of the matrix (Λ †) 1/2U

2. Commute time embedding takes all columns Several authors justify that spectral clustering constructs clusters based on the Euclidean distances between the yi can be interpreted as building clusters of the vertices in the graph based on the commute distance.

演算法的比较:

Which graph Laplacian should be used?

ANS：Look at the degree distribution. There are several arguments which advocate for using normalized rather than unnormalized spectral clustering, and in the normalized case to use the eigenvectors of Lrw rather than those of Lsy

Why normalized is better than unnormalized spectral clustering?

ANS:

Objective1:

Both RatioCut and Ncut directly implement Only Ncut implements Normalized spectral clustering implements both clustering objectives mentioned above, while unnormalized spectral clustering only implements the first obejctive.

Objective2:

1. We want to find a partition such that points in different clusters are dissimilar to each other, that is we want to minimize the between-cluster similarity. In the graph setting, this means to minimize $cut(A, A)$.
2. We want to find a partition such that points in the same cluster are similar to each other, that is we want to maximize the within-cluster similarities $W(A, A)$, and $W(A, A)$.

Only Ncut implements

Normalized spectral clustering implements both clustering objectives mentioned above, while unnormalized spectral clustering only implements the first obejctive.

Why the eigenvectors of Lrw are better than those of $Lsym$?

ANS:

1. Eigenvectors of Lrw are cluster indicator vectors $\mathbb{I}Ai$, while the eigenvectors of $Lsym$ are additionally multiplied with D 1/2 , which might lead to undesired artifacts.
2. Using $Lsym$ also does not have any computational advantages.

譜聚類演算法的**主要優點**有：

1）譜聚類只需要數據之間的相似度矩陣，因此對於處理稀疏數據的聚類很有效。這點傳統聚類演算法比如 K-Means 很難做到。

2）由於使用了降維，因此在處理高維數據聚類時的複雜度比傳統聚類演算法好。

譜聚類演算法的**主要缺點**有：

1）如果最終聚類的維度非常高，則由於降維的幅度不夠，譜聚類的運行速度和最後的聚類效果均不好。

2）聚類效果依賴於相似矩陣，不同的相似矩陣得到的最終聚類效果可能很不同。