A Note on Spectral Clustering

by Langtian Ma

1 Similarity Graphs

Definition 1. (Similarity graphs) *Given a set of data points $\{x_i\}_{i=1}^n$, and some notion of similarity s_{ij}, which measures the similarity between x_i and x_j. Then we can representing the data in form of the similarity graph $G=(V,E)$, where each vertex v_i is in this graph represents a data point x_i. x_i and x_j are connected if $s_{ij}>\tau$, where τ is a threshhold, usually set to 0.*

Then the problem of clustering can be reformulated using the similarity graph: we want to find a partition of similarity graph.

2 Graph Notation

- Let $G=(V,E)$ be an undirected weighted graph with vertex set $V=\{v_1,\dots,v_n\}$. Each edge between two vertices v_i and v_j carries a non-negative weight $w_{ij}>0$.
- The weighted **adjacency matrix** is the matrix $W=(w_{ij})_{i,j=1,\dots n}$. $w_{ij}=0$ means v_i and v_j are not connected. We require $w_{ij}=w_{ji}$ since the graph is undirected.
- The degree of a vertex is defined as
$$d_i=\sum_{j=1}^{n} w_{ij}.$$
The degree matrix is defined as $D=\mathrm{diag}(d_1,\dots d_n)$.
- Given a subset of vertices $A\subset V$, we denote its complement $V\setminus A$ by $\bar{A}$. We define the indicator vector $\mathbb{1}_A=(f_1,\dots f_n)^T\in\mathbb{R}^n$ where $f_i=\mathbb{1}\{v_i\in A\}$. For convenience we use $i\in A$ to represent $v_i\in A$.
- For $A,B\in V$ we define
$$W(A,B)=\sum_{i\in A,j\in B} w_{ij}$$
- Two different ways to measure size of A:
$$|A|:=\text{the number of vertices in } A$$
$$\mathrm{vol}(A):=\sum_{i\in A} d_i$$
- $A\in V$ is **connected** if any two vertices of A can be joined by a path such that all intermediate points also lie in A. A is called a **connected component** if it is connected and if there are no connections between vertices in A and $\bar{A}$.
- The nonempty sets A_1,...A_k form a partitioin of the graph if $A_i\wedge A_j=\varnothing$ and $A_i\cup\dots\cup A_k=V$.
- Eigenvalues will always be ordered increasingly, respecting multiplicities.

- "The first k eigenvectors" refers to the eigenvectors corresponding to the k smallest eigenvalues.

3 Graph Laplacians

3.1 The unnormalized graph Laplacian

Definition 2. *The Unnormalized Graph Laplacian is defined as*

$$L = D - W.$$

Proposition 3. (Properties of L) *The matrix L have the following properties:*

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

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

2. *L is symmetric and positive semi-definite.*

3. *The smallest eigenvalue of L is 0, the corresponding eigenvector is $\mathbf{1} = (1, \dots 1)^T \in \mathbb{R}^n$.*

4. *L has n non-negative, real eigenvalue $0 = \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$.*

Remark 4. Unnormalized graph Laplacian does not depdend on diagonal elements of the adjacency matrix W.

Proposition 5. *Let G be an undirected graph with non-negative weights. Then the multiplicity k of the eigenvalue 0 of L equals the number of connected components $A_1, \dots, A_k$ in the graph. The eigenspace of eigenvalue 0 is spanned by the indicator vectors $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_k}$.*

3.2 The normalized graph Laplacians

Definition 6. *There are two matrices which are called normalized graph Laplacians:*

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

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

Remark 7. L_{sym} is symmetric and L_{rw} is closely related to random walk.

Proposition 8. (Properties of L_{sym} and L_{rw})

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

$$f^T L_{\text{sym}} f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} \left(\frac{f_i}{\sqrt{d_i}} - \frac{f_i}{\sqrt{d_j}} \right)^2$$

2. *λ is an eigenvalue of L_{rw} with eigenvector u if and only if λ is an eigenvalue of L_{sym} with eigenvector $w = D^{-1/2} u$.*

3. *λ is an eigenvalue of L_{rw} with eigenvector u if and only if λ and u solve the generalized eigen-problem $Lu = \lambda Du$.*

4. *0 is an eigenvalue of L_{rw} with $\mathbf{1}$ as eigenvector. 0 is an eigenvalue of L_{sym} with eigenvector $D^{1/2}\mathbf{1}$.*

5. *L_{sym} and L_{rw} are positive semi-definite.*

Proposition 9. *Let G be an undirected graph with non-negative weights. Then the multiplicity k of the eigenvalue 0 of both L_{rw} and L_{sym} equals the number of connected components $A_1, \ldots, A_k$ in the graph. For L_{rw} the eigenspace of eigenvalue 0 is spanned by $\mathbb{1}_{A_i}$ of those components. For L_{sym}, the eigenspace of 0 is spanned by the vectors $D^{-1/2}\mathbb{1}_{A_i}$.*

4 Spectral Clustering Algorithms

Algorithm 1

(Unnormalized spectral clustering)
Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of the clusters to construct.

1. Construct a similarity graph based on S, let W be its weighted adjacency matrix.
2. Compute the unnormalized Laplacian L.
3. **Compute the first k eigenvectors $u_1, \ldots, u_k$ of L.**
4. Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing $u_1, \ldots u_k$ as columns.
5. For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of U.
6. 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 $A_i = \{j \,|\, y_j \in C_i\}$.

Algorithm 2

(Normalized spectral clustering with L_{rw})
Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of the clusters to construct.

1. Construct a similarity graph based on S, let W be its weighted adjacency matrix.
2. Compute the unnormalized Laplacian L.
3. **Compute the first k eigenvectors $u_1, \ldots, u_k$ of the generalized eigenproblem $Lu = \lambda Du$, i.e. eigenvectors of L_{rw}.**
4. Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing $u_1, \ldots u_k$ as columns.
5. For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of U.
6. 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 $A_i = \{j \,|\, y_j \in C_i\}$.

Algorithm 3

(Normalized spectral clustering with L_{sym})
Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of the clusters to construct.

1. Construct a similarity graph based on S, let W be its weighted adjacency matrix.
2. Compute the unnormalized Laplacian L.
3. **Compute the first k eigenvectors $u_1, \ldots, u_k$ of L_{sym}.**
4. Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing $u_1, \ldots u_k$ as columns.
5. **Form the matrix T by normalizing the rows of U to 1.**

6. For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of U.
7. 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 $A_i = \{j \,|\, y_j \in C_i\}$.