Linear Algebraic Approach to Spectral Clustering

Adam Schaffroth*

April 24, 2023

Abstract

Spectral clustering is a modern class of methods that utilizes eigenvectors, eigenvalues, and similarity measures to identify clusters within unlabelled data. Whereas many traditional clustering algorithms (e.g. k-means) are based on compactness measures, spectral clustering looks at connectivity between data points. The differing approaches can result in radical improvements over the traditional such as k-means clustering. Applications of spectral clustering are displayed across the qualitative and quantitative sciences.

1 Introduction

Spectral clustering's strength is that it does not make assumptions about the form of the clusters in the data. Some traditional techniques like k-means assume that data clusters are spherical and of similar size. Spectral clustering relies on measures of connectivity to decide cluster boundaries whereas k-means chooses clusters with regard to compactness (i.e. distance between data points). This can lead to radically different clustering outputs in the same data set.

Figure 1 demonstrates the difference in output from a clustering algorithm based on connectivity (spectral clustering) and compactness (k-means). Clearly, spectral clustering computes more reasonable partitions. In fact, there are many instances in which spectral clustering outperforms compactness-based algorithms.

Spectral clustering involves variations of a similar process:

1. Create a matrix representation of the graph
    - Create a adjacency matrix, degree matrix, and ultimately, the Laplacian matrix
        - Note that There are multiple ways to construct the Laplacian using adjacency, affinity, or incidence matrices
2. Analyze the spectral decomposition

*Johns Hopkins University Whiting School of Engineering

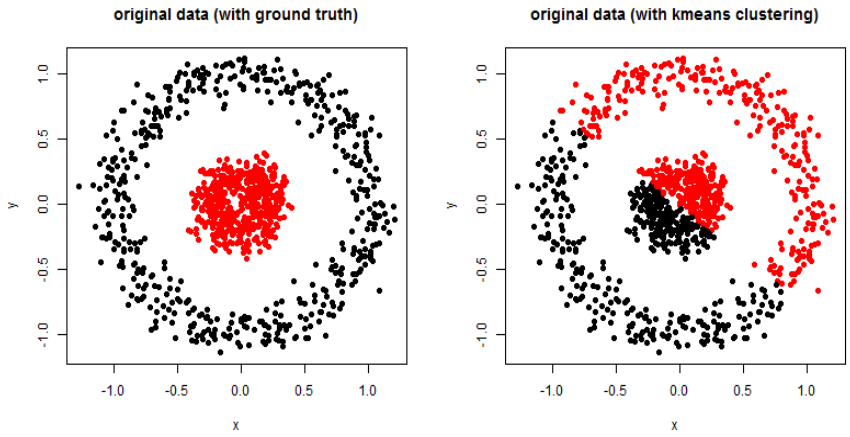


Figure 1: Source: https://rpubs.com/sandipan/199446

- Find the eigenpairs (λ, x) of the Laplacian matrix
- Map vertices into a lower dimesional representation according to the second smallest eigenvector λ_2
  - Note that from analyzing the spectrum (set of eigenvalues) we get the name *spectral* clustering
  - The lower-dimensional mapping is referred to as a *spectral embedding* and is useful beyond spectral partitioning

3. Assigning clusters

- Decide k clusters
- Sort components of a 1-dimensional vector
- Identify split points based on algorithm of your choosing (e.g. k-means)

2 Graph Laplacians

2.1 Graph Background

We want to partition our data set $X = \{x_1, \ldots, x_n\} \in \mathbb{R}$ into k clusters. Our undirected similarity graph, denoted $G = (V, E)$, where vertices $v_i \in V$ and E, represents the edges of G. Let $n \times n$ matrix $W = (w_{ij})_{i,j=1,\ldots,n}$ where $w_{ij} \geq 0$ represents the weight of the edge from v_i to v_j. Note that $w_{ij} = 0$ when there is no edge connecting v_i and v_j. Also, $w_{ij} = w_{ji}$ because G is undirected.

Definition 1 *The* **degree** *of a vertex $v_i \in V$ is defined by*

$$d_i = \sum_{j=1}^{n} w_{ij}.$$

Definition 2 *The* **degree matrix** D *is an* $n \times n$ *diagonal matrix represented by* $D = diag\{d_1, \ldots, d_n\}$.

2.2 Popular Similarity Graphs

There are obviously many ways to create a graph from our data set $X = \{x_1, \ldots, x_n\}$. Three popular choices for spectral clustering are [5][6]:

1. **K-Nearest Neighbors Graphs:** Connect vertices v_i and v_j if:

   - **A:** v_i is amongst the k-nearest neighbors of v_j.
   - **B:** v_j is amongst the k-nearest neighbors of v_i.

   We call the graph where $A \vee B$ the *k-nearest neighbor graph.* We call the graph where $A \wedge B$ the *mutual k-nearest neighbor graph.*

2. ε**-neighborhood graph:** Connect all data points with pairwise distances less than ε.

3. **Fully connected graph:** Connect all data points with a positive similarity metric to each other. A commonly used similarity function is

$$s(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right),$$

   where σ controls the width of neighborhoods.

2.3 Unnormalized Graph Laplacian

Definition 3 *The* **unnormalized graph Laplacian** L *is defined by*

$$L = D - W$$

Definition 4 *The indicator vector* $\mathbb{1}_A$ *is defined by*

$$\mathbb{1}_A = (f_1, \ldots, f_n)^T,$$

where $f_i = 1$ *if* f_i *belongs to subset* $A \in V$*, and* $f_i = 0$ *otherwise.*

Theorem 1 *Matrix* L *satisfies the following:[7][8][6]*

1. $\forall f \in \mathbb{R}^n$,

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

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

3. *The smallest eigenvalue of* L *is* 0 *with eigenvector* $\mathbb{1}$.

4. L *has* n *nonnegative, real-valued eigenvalues* $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$.

3 Analysis of the Spectrum

Let graph G have k connected components $A_1, \ldots, A_k$ with corresponding Laplacian matrices $L_1, \ldots, L_k$. We can write L as a diagonal block matrix:

$$L = \begin{bmatrix} L_1 & & & \\ & L_2 & & \\ & & \ddots & \\ & & & L_k \end{bmatrix}$$

Lemma 1 *The eigenpairs of block matrix L are the union of the eigenpairs of its blocks.*

PROOF:

$$\det(L - \lambda I) = \det(L_1 - \lambda_1 I)\det(L_2 - \lambda I)\ldots\det(L_k - \lambda I)$$

Thus, the eigenvalues λ and corresponding eigenvectors of each block L_i will also be in L. QED

Theorem 2 *The algebraic multiplicity of eigenvalue* $\mathbf{0}$ *for the Laplacian matrix of* $L(G)$ *is exactly 1 iff* G *is connected.[9]*

Theorem 3 *[6] The multiplicity* k *of eigenvalue* 0 *of* L *equals the number of connected components* $A_1, A_2, \ldots, A_k$ *in the graph. The eigenspace of eigenvalue* 0 *is spanned by the indicator vectors* $\mathbb{1}_{A_1}, \mathbb{1}_{A_2}, \ldots, \mathbb{1}_{A_k}$ *of those components.*

PROOF:
$\rightarrow$Graph G has k connected components iff its corresponding adjacency is block diagonoal with k blocks iff $L(G)$ has k blocks. Let

$$L(G) = \begin{bmatrix} L_1 & & & \\ & L_2 & & \\ & & \ddots & \\ & & & L_k \end{bmatrix}.$$

Beacuse each block matrix is symmetric, it is also diagonalizable. By Theorem 2, each block matrix has eigenvalue 0 with algebraic and geometric multiplicity equal to one. We know that $\det(xI - L) = \det(xI - L_1)\det(xI - L_2)\ldots\det(xI - L_k)$, so the algebraic multiplicity of eigenvector $\mathbf{0}$ for L is the sum of the algebraic multiplicity of eigenvector $\mathbf{0}$ for each L_i. If G has k connected components then this sum is $1 + \cdots + 1$ k times, which equals k.
$\leftarrow$ If the algebraic multiplicity of L is k, then the sum of the multiplicities is k. Thus, there must be k blocks in the the Laplacian matrix L. Then G must have k connected components.

QED

4 Unnormalized Spectral Clustering Algorithm

Let $F \in \mathbb{R}^{n \times k}$ be a matrix with orthonormal vectors $f_1, \ldots, f_k$ as columns. We want to find F such that:

$$\min_F \{Tr(F^T LF) \mid F^T F = I\}$$

In other words, we want to find $f_1, \ldots, f_k$ that minimize such that F is an orthogonal matrix. These vectors are the k smallest eigenvectors of unnormalized Laplacian matrix L.

Algorithm 1 Unnormalized Spectral Clustering

[6]

1: Construct a similarity graph (e.g. ε-neighborhood).
2: Compute unnormalized graph Laplacian $L = D - W$.
3: Find the k smallest eigenvectors of L (i.e. $f_1, \ldots, f_k$).
4: Construct $F \in \mathbb{R}^{n \times k}$ from $f_1, \ldots, f_k$.
5: Let rows of F represent vertices in $mathbbmR^k$. Partition these vertices into k clusters using any method (k-means is a popular one).

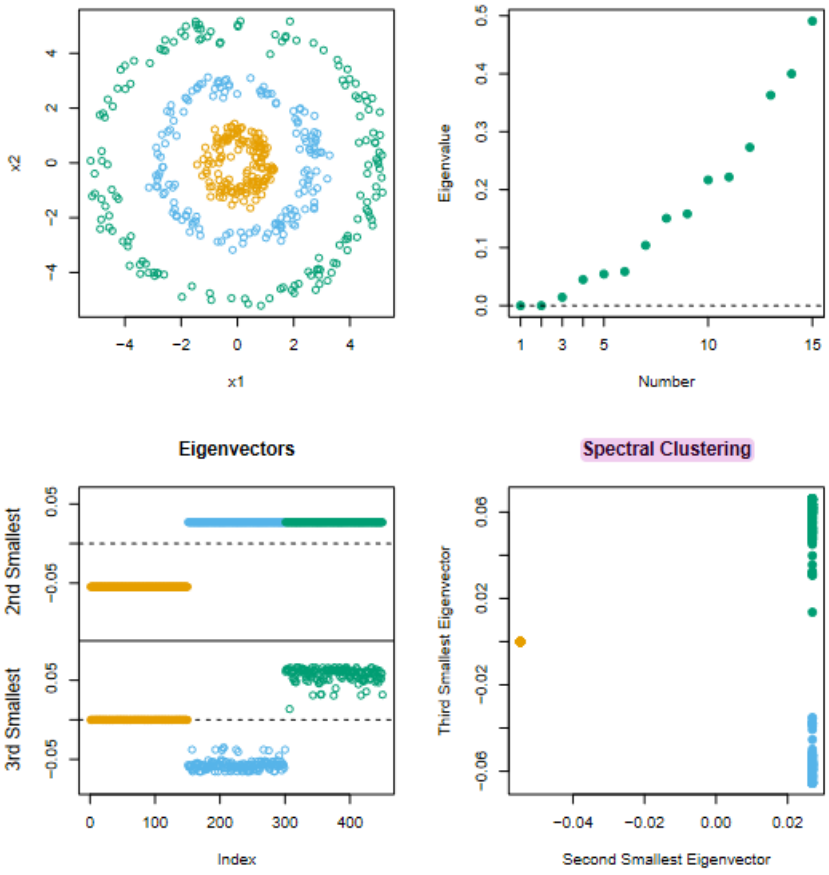
Below are some useful figures to help visualize the steps of spectral clustering up through the spectral decomposition.

Figure 2 shows the results of spectral clustering performed using a k=10 nearest neighbors similarity graph. The bottom left plot shows the eigenvectors of L corresponding to the second and third smallest eigenvalues (the eigenvector corresponding to the smallest eigenvalue is constant and thus, not of interest). The coloring is consistent which how the algorithm partition the data. In the upper-right, the 15 smallest eigenvalues in the spectrum of L are shown. The lower-right plot shows the coordinate of the second and third eigenvectors. The partitioning use here was k-means.

Figure 3 shows the graph view and the algorithm's constructed matrices. The parameter k can usually be determined from the plot of eigenvalues. Here, I would choose k = 3.

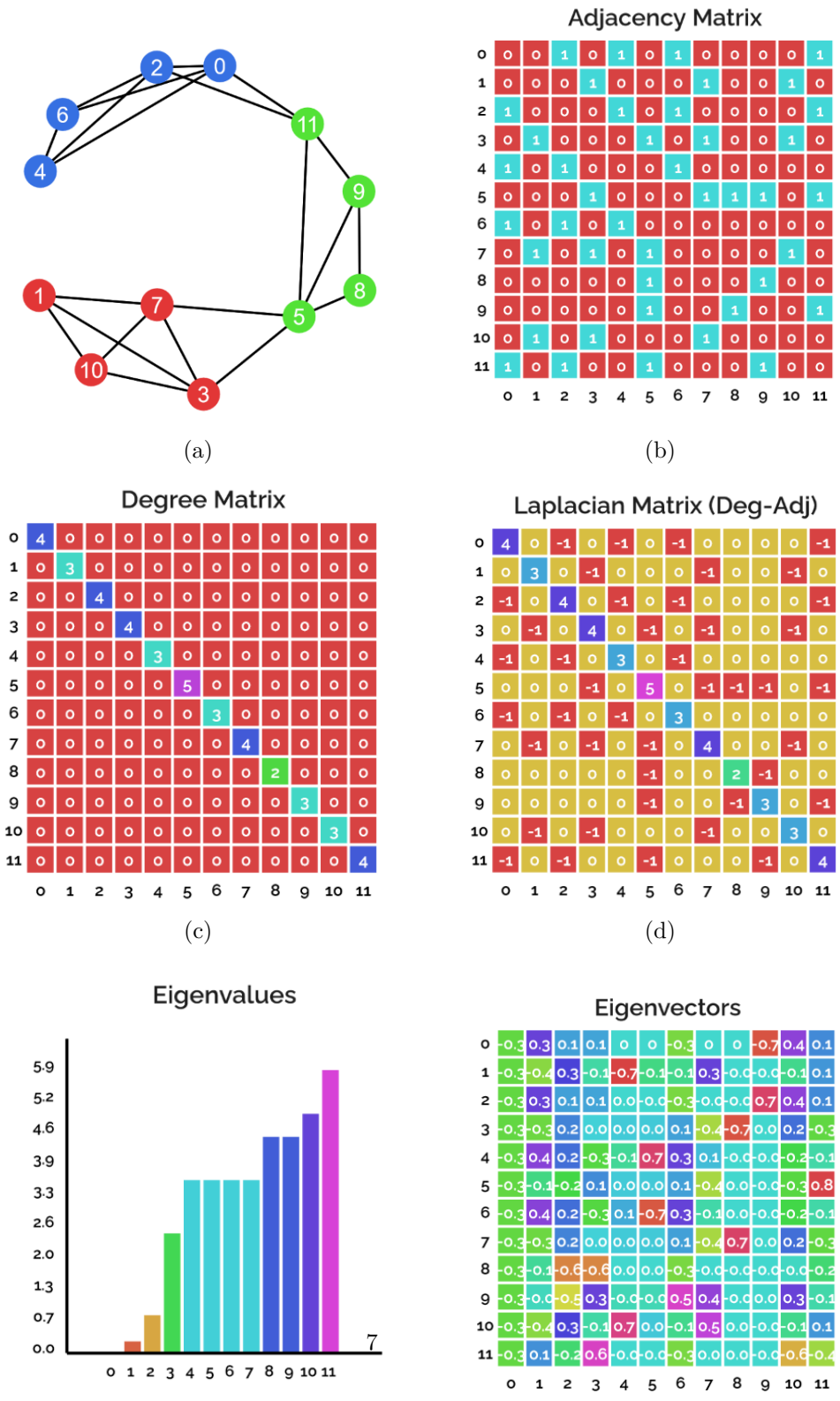
5 Conclusion

There are two versions of normalized graph Laplacians (i.e. the symmetric Laplacian L_{sym} and the random walk Laplacian L_{rw}) which are not covered in this paper. Some of the intuition of spectral clustering can be provided with a *graph-cut* approach, but I have omitted this as well. The "magic" of spectral cluster lies in being able to extract information about the overall connectivity of a graph just by looking at the ascending eigenvalues of the Laplacian matrix.



[10]

Figure 2: Spectral clustering applied to a data set of concentric rings.



(a) (b) (c) (d) (e) (f)

Figure 3: Matrices and eigendecomposition of a simple graph [4].

References

[1] Pothen, A., Simon, H. D., Liou, K. P. (1990). Partitioning Sparse Matrices with Eigenvectors of Graphs. SIAM Journal on Matrix Analysis and Applications, 11(3), 430–452. https://doi.org/10.1137/0611030

[2] Aggarwal, C. C., Reddy, C. K. (2013). Data Clustering: Algorithms and Applications (Chapman Hall/CRC Data Mining and Knowledge Discovery Series) (1st ed.). Chapman and Hall/CRC.

[3] Fiedler, M. (1973). Algebraic connectivity of graphs. Czechoslovak Mathematical Journal, 23(2), 298–305. https://doi.org/10.21136/cmj.1973.101168

[4] Schmidt, D. (2020). Spectral Clustering on Graphs. Spectral Clustering on Graphs Experiment. https://dominikschmidt.xyz/spectral-clustering-exp/

[5] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for em-bedding and clustering. InAdvances in Neural Information Processing Systems 14, pages585–591, 2001.

[6] Ulrike Luxburg. A tutorial on spectral clustering.Statistics and Computing, 17(4):395–416,2007.

[7] Bojan Mohar. The Laplacian spectrum of graphs.Graph Theory, Combinatorics, and Appli-cations, 2:871–898, 1991.

[8] Bojan Mohar. Some applications of Laplace eigenvalues of graphs.Graph Symmetry: Alge-braic Methods and Applications, 497:227–275, 1997.

[9] EIGENVALUES OF THE LAPLACIAN AND THEIR RELATIONSHIP TO THE CONNECTEDNESS OF A GRAPH. (2013). ANNE MARSDEN. https://math.uchicago.edu/ may/REU2013/REUPapers/Marsden.pdf

[10] Hastie, T., Tibshirani, R., Friedman, J. (2021). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics) (2nd ed.). Springer.

Adam Schaffroth

Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218.