Machine Learning with Graphs: Spectral Graph Theory 1/3 - Intro, Laplacian

Arlei Silva

Spring 2022

Spectral Graph Theory (SGT) provides elegant connections between graphs and matrices—especially the so-called graph Laplacian. While most relevant graph problems are NP-hard, a few of them can be approximated as eigenvector/eigenvalue problems, which are solvable in poly-time. Here, we will introduce the Laplacian matrix and show how we can use its spectrum to solve machine learning problems, such as graph partitioning.

Linear algebra

Spectral graph theory relies on a few fundamental concepts and theorems from Linear Algebra. If you need a refresher on the subject, Gilbert Strang's books [5, 4] are good references.

We say that $(\lambda, \mathbf{x})$ is an *eigenvalue-eigenvector* pair for a matrix M if:

$$M\mathbf{x} = \lambda\mathbf{x}$$

Eigenvalues can also be defined as roots of a matrix polynomial:

$$det(\lambda I - M)$$

We will focus on the case where M is symmetric. Thus, we can use the *spectral theorem* to show that M has real eigenvalues and *orthonormal* eigenvectors.

Theorem 1 *(The spectral theorem) A real symmetric matrix $M \in \mathbb{R}^{n\times n}$ has real eigenvalues $\lambda_1, \dots \lambda_n$ and orthonormal eigenvectors $\mathbf{x}_1, \dots \mathbf{x}_n$.*

The *quadratic form* of a square matrix M is the scalar $\mathbf{x}^T M\mathbf{x} = \sum_{ij} M_{i,j}\mathbf{x}_i\mathbf{x}_j$. We can use the quadratic form to define the *Rayleigh–Ritz quotient* of M, $\mathbf{x}^T M\mathbf{x}/\mathbf{x}^T\mathbf{x}$. The *Courant-Fischer theorem* applies the Rayleigh-Ritz quotient of M to define eigenvalues as the solution of optimization problems.

Theorem 2 *(Courant-Fischer theorem) The eigenvalues $\lambda_1, \dots \lambda_n$ of a symmetric matrix M are such that:*

$$\lambda_k = \max_{\substack{S \subseteq \mathbb{R}^n \\ dim(S)=k}} \min_{\substack{\mathbf{x} \in S \\ \mathbf{x} \neq 0}} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \min_{\substack{T \subseteq \mathbb{R}^n \\ dim(S)=n-k+1}} \max_{\substack{\mathbf{x} \in T \\ \mathbf{x} \neq 0}} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

where S and T are subspaces and $dim(S)$ is the dimension of S.

The following definitions for $\lambda_1, \lambda_2, \lambda_3, \ldots$ follow from Courant-Fischer:

$$\lambda_1 = \min_{\mathbf{x} \neq 0} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

$$\lambda_2 = \min_{\substack{\mathbf{x} \neq 0 \\ \mathbf{x} \perp \mathbf{x}_1}} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

$$\lambda_3 = \min_{\substack{\mathbf{x} \neq 0 \\ \mathbf{x} \perp \{\mathbf{x}_1, \mathbf{x}_2\}}} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

Notice that, for orthonormal eigenvectors, the denominator $\mathbf{x}^T \mathbf{x} = 1$.

The Laplacian matrix

We will assume that G is undirected and unweighted. However, everything discussed here can be easily generalized to the case of weighted graphs.

The Laplacian can be defined in terms of the *adjacency* and *degree* matrices. The adjacency matrix $A \in \mathbb{R}^{n \times n}$ of a graph $G = (V, E)$ is defined as:

$$A_{ij} = \begin{cases} 1 & \text{if } (v_i, v_j) \in E \\ 0 & \text{otherwise} \end{cases}$$

And the degree matrix of G is defined as:

$$D_{ij} = \begin{cases} deg(v_i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

The Laplacian $L \in \mathbb{R}^{n \times n}$ of G is the difference between the degree and adjacency matrices:

$$L = D - A$$

The first interesting property of the Laplacian is that its quadratic form is a measure of the "smoothness" of § in G:

$$\mathbf{x}^T L \mathbf{x} = \sum_{(v_i, v_j) \in E} (x_i - x_j)^2$$

It follows that L is positive semi-definite (PSD), since $\mathbf{x}^T L \mathbf{x} \geq 0$. Moreover, we can use the Courant-Fischer theorem to show that its first eigenvalue $\lambda_1 = 0$

with eigenvector $\mathbf{1}_n/\sqrt{n}$ ($\mathbf{x}^T L\mathbf{x} = 0$). Using a similar argument, we can show that the second eigenvalue $\lambda_2 = 0$ iff G is disconnected—we can again find a vector $\mathbf{x}$ such that $\mathbf{x}^T L\mathbf{x} = 0$ and $\mathbf{x} \perp \mathbf{1}_n/\sqrt{n}$.

So far, our understanding of the spectrum of the Laplacian is limited to $(\lambda_1, \mathbf{x}_1)$ and, if the graph is disconnected, $(\lambda_2, \mathbf{x}_2)$. But what can we say about $\lambda_2, \dots \lambda_n$ and $\mathbf{x}_1, \dots \mathbf{x}_n$ in general? We can gain some insights by looking at specific graphs.

A clique (or complete graph) with n vertices has Laplacian $L = nI - \mathbf{1}_{n\times n} = nI - \mathbf{1}_n^T\mathbf{1}_n$, where I is the identity matrix. By definition, $nI\mathbf{x} = n\mathbf{x}$ for any $\mathbf{x}$. Moreover, eigenvectors of $\mathbf{1}_n^T\mathbf{1}_n$ are such that $\mathbf{1}_n^T \sum_i \mathbf{x}[i] = \lambda\mathbf{x}$. We know that $\mathbf{x}_n = \mathbf{1}_n$ is an eigenvector with eigenvalue $\lambda_n = n$. All other eigenvectors $\mathbf{x}_1, \dots \mathbf{x}_{n-1}$ must be orthogonal to $\mathbf{1}_n$ ($\sum_i \mathbf{x}[i] = 0$) and thus their associated eigenvalue must be $\lambda_1 = \dots \lambda_{n-1} = 0$. Because these eigenvectors are also eigenvectors of nI, they are eigenvectors of L with eigenvalues $\lambda'_1 = n - n = 0$ and $\lambda_2 = \dots \lambda_n = n - 0 = n$.

References

[1] Fan RK Chung and Fan Chung Graham. *Spectral graph theory*. American Mathematical Soc., 1997.

[2] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.

[3] Daniel Spielman. Spectral graph theory. *Combinatorial scientific computing*, 2012.

[4] Gilbert Strang. *Linear algebra and learning from data*. Wellesley-Cambridge Press, 2019.

[5] Gilbert Strang. Introduction to linear algebra. 2021.