Laplacian matrix

Author and editors

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland (ORCID 0000-0001-7538-0990)
Editors: Konstantina Olioumtsevits (ORCID 0000-0003-2697-8659), Ekkehard Schnoor (ORCID 0000-0001-8252-1499), Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Abstract

The Laplacian matrix of a weighted graph with n nodes is the $n \times n$ matrix $\mathbf{L}^{(\mathcal{G})} = \mathbf{D}^{(\mathcal{G})} - \mathbf{A}^{(\mathcal{G})}$, the degree matrix minus the adjacency matrix of the edge weights. Its quadratic form $\mathbf{x}^\top \mathbf{L}^{(\mathcal{G})} \mathbf{x}$ is the sum, over the edges, of the edge weight times the squared difference of the two entries of $\mathbf{x}$ at the ends of the edge. The Laplacian matrix is therefore a symmetric matrix and positive semi-definite (psd), the constant vector is an eigenvector with eigenvalue 0, and the multiplicity of the eigenvalue 0 equals the number of connected components of the graph. Two normalized variants divide the edge weights by the node degrees; they share their eigenvalues, which lie between 0 and 2. In machine learning (ML), the quadratic form is the penalty term of Laplacian regularization, which makes the values learned at connected nodes similar and predicts the value at an unlabeled node as the weighted average of its neighbors. The eigenvectors of its smallest eigenvalues are the coordinates that spectral clustering feeds to k-means.

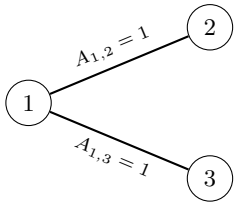
Definition

Twelve weather stations around Krems an der Donau, from Melk and Zwettl in the west to Retz and Langenlebarn in the east, each recorded a mean temperature for the year 2024: 12.6 degrees Celsius at Krems, 9.6 on the Jauerling mountain. Two stations at most 35 km apart are joined by an edge whose weight decays with their distance, so the stations form a weighted graph $\mathcal{G}$ with 27 edges (Fig. 2(a)). How much does the temperature vary along the edges of $\mathcal{G}$? One answer is the sum, over the edges, of the edge weight times the squared difference of the two temperatures at its ends. The Laplacian matrix of $\mathcal{G}$ is the matrix that computes this sum as a quadratic form in the vector of the twelve temperatures.

Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with nodes $\mathcal{V} = \{1, \ldots, n\}$ and nonnegative edge weights $A_{i,i'} = A_{i',i}$, where $A_{i,i'} = 0$ when $\mathcal{G}$ has no edge $\{i, i'\}$ (von Luxburg, 2007, Sect. 2.1). The edge weights are the entries of the adjacency matrix $\mathbf{A}^{(\mathcal{G})} \in \mathbb{R}^{n \times n}$, and the node degree $d^{(i)} := \sum_{i'=1}^{n} A_{i,i'}$ of node i is the sum of the weights of its edges. The Laplacian matrix of $\mathcal{G}$ is the difference

$$\mathbf{L}^{(\mathcal{G})} := \mathbf{D}^{(\mathcal{G})} - \mathbf{A}^{(\mathcal{G})}, \text{ with the degree matrix } \mathbf{D}^{(\mathcal{G})} := \operatorname{diag}\left\{d^{(1)}, \ldots, d^{(n)}\right\}. \quad (1)$$

The diagonal entry $L^{(\mathcal{G})}_{i,i}$ of the Laplacian matrix (1) is the node degree $d^{(i)}$, and the off-diagonal entry $L^{(\mathcal{G})}_{i,i'}$ is the negated edge weight $-A_{i,i'}$, so every row sums to zero (Hastie et al., 2009, Sect. 14.5.3; von Luxburg, 2007, Sect. 3.1). Fig. 1 carries out the construction for a graph with three nodes and two edges. For the station graph, the edge between Krems and Langenlois, 8 km apart, has weight 0.92, the longest edges have weight 0.23, and Krems has degree 3.18 from its six neighbors.



$$\mathbf{L}^{(\mathcal{G})} = \mathbf{D}^{(\mathcal{G})} - \mathbf{A}^{(\mathcal{G})} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

Figure 1: An undirected graph with three nodes and two edges, both with edge weight 1, and its Laplacian matrix (1). Node 1 has degree 2 and nodes 2 and 3 have degree 1, which are the diagonal entries; each edge contributes -1 to the two off-diagonal entries it joins, and the remaining entries are 0

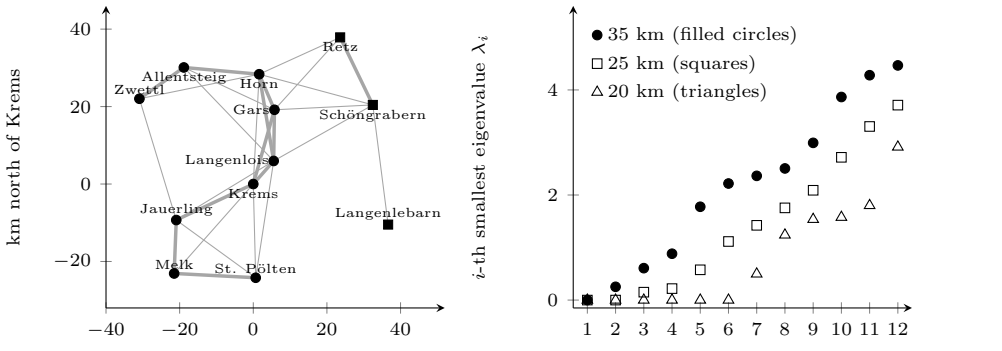
The Laplacian matrix is characterized by its quadratic form. For every vector $\mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$, one entry per node,

$$\mathbf{x}^\top \mathbf{L}^{(\mathcal{G})} \mathbf{x} = \sum_{\{i,i'\} \in \mathcal{E}} A_{i,i'} \big(x_i - x_{i'}\big)^2. \tag{2}$$

The identity follows from expanding $\mathbf{x}^\top \mathbf{D}^{(\mathcal{G})} \mathbf{x} - \mathbf{x}^\top \mathbf{A}^{(\mathcal{G})} \mathbf{x} = \sum_{i=1}^n d^{(i)} x_i^2 - \sum_{i=1}^n \sum_{i'=1}^n A_{i,i'} x_i x_{i'}$ and collecting the two terms of each edge (Hastie et al., 2009, Sect. 14.5.3; von Luxburg, 2007, Prop. 1). For the station graph and the vector of mean temperatures, the quadratic form (2) equals 26.9 squared degrees; the edge between Jauerling and Melk, with weight 0.79 and a temperature difference of 2.8 degrees, contributes 6.0 of it.

Three properties of the Laplacian matrix follow from the quadratic form (2) (von Luxburg, 2007, Prop. 1, Prop. 2; Mohar, 1991). First, $\mathbf{L}^{(\mathcal{G})}$ is a symmetric matrix and positive semi-definite (psd), since every term of the sum is nonnegative, so its eigenvalues are real and can be ordered as $0 \le \lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$. Second, the constant vector $\mathbf{1} = (1, \ldots, 1)^\top$ makes every term vanish, and $\mathbf{L}^{(\mathcal{G})} \mathbf{1} = \mathbf{0}$ since every row sums to zero, so $\lambda_1 = 0$ with the eigenvector $\mathbf{1}$. Third, the multiplicity of the eigenvalue 0 equals the number of connected components of $\mathcal{G}$. A vector $\mathbf{x}$ with $\mathbf{x}^\top \mathbf{L}^{(\mathcal{G})} \mathbf{x} = 0$ takes the same value at the two ends of every edge, so it is constant on each component, and the indicator vectors of the components span the nullspace of $\mathbf{L}^{(\mathcal{G})}$. The station graph is connected, so $\lambda_1 = 0$ is simple and $\lambda_2 = 0.25$ is positive. Joining only stations at most 25 km apart leaves Langenlebarn without an edge and the graph with two components,

and the limit 20 km leaves six components; the eigenvalue 0 then has multiplicity two and six (Fig. 2(b)). The second-smallest eigenvalue λ_2 is the algebraic connectivity of $\mathcal{G}$, and the eigenvector of λ_2 is the Fiedler vector (Fiedler, 1973). The signs of its entries split the nodes into two clusters with few edges between them (Fiedler, 1975). For the station graph, the three stations more than 20 km east of Krems, Langenlebarn, Schöngrabern, and Retz, have negative entries and the other nine have positive entries.



(a) the station graph

(b) eigenvalues of the Laplacian matrix of three graphs

Figure 2: (a) Twelve weather stations around Krems an der Donau, drawn at their positions in km east and north of Krems and joined by an edge when at most 35 km apart; a thick line marks an edge weight above 1/2, a thin line one below. Filled circles mark the stations with a nonnegative entry of the Fiedler vector, filled squares those with a negative entry. (b) The eigenvalues of the Laplacian matrix in increasing order for the graph in (a) (filled circles) and for the graphs that join stations at most 25 km (squares) and 20 km (triangles) apart, the legend naming the largest edge length. The eigenvalue 0 appears once, twice, and six times, the number of connected components of each graph. Data generated by `pythondemos/LapMat.py`

Two normalized variants of the Laplacian matrix divide the edge weights by the node degrees (von Luxburg, 2007, Sect. 3.2),

$$\mathbf{L}^{(\mathrm{sym})} := \big(\mathbf{D}^{(\mathcal{G})}\big)^{-1/2}\mathbf{L}^{(\mathcal{G})}\big(\mathbf{D}^{(\mathcal{G})}\big)^{-1/2} \text{ and } \mathbf{L}^{(\mathrm{rw})} := \big(\mathbf{D}^{(\mathcal{G})}\big)^{-1}\mathbf{L}^{(\mathcal{G})},$$

for a graph without isolated nodes. The first is a symmetric matrix whose quadratic form is the sum (2) with each x_i replaced by $x_i/\sqrt{d^{(i)}}$. The second is $\mathbf{I} - \big(\mathbf{D}^{(\mathcal{G})}\big)^{-1}\mathbf{A}^{(\mathcal{G})}$, the identity matrix minus the transition matrix of the random walk on $\mathcal{G}$ that moves to a neighbor with probability proportional to the edge weight. The two variants have the same eigenvalues, which lie between 0 and 2 (Chung, 1997, Lemma 1.7); the eigenvalue 0 has the eigenvector $\mathbf{1}$ for $\mathbf{L}^{(\mathrm{rw})}$ and $\big(\mathbf{D}^{(\mathcal{G})}\big)^{1/2}\mathbf{1}$ for $\mathbf{L}^{(\mathrm{sym})}$, and its multiplicity again counts the connected components (von Luxburg, 2007, Prop. 3, Prop. 4). For the station graph, the

largest eigenvalue of $\mathbf{L}^{(\mathrm{sym})}$ is 1.57. Which of the three matrices a method uses is a design choice (von Luxburg, 2007, Sect. 8.5): spectral clustering has been formulated with $\mathbf{L}^{(\mathrm{rw})}$ (Shi and Malik, 2000), with $\mathbf{L}^{(\mathrm{sym})}$ (Ng et al., 2001), and with $\mathbf{L}^{(\mathcal{G})}$ (von Luxburg, 2007, Sect. 4).

In machine learning (ML), the Laplacian matrix enters through its quadratic form (2) as a penalty term. Consider data points that are the nodes of $\mathcal{G}$, which is then a similarity graph, with a label y_i known only for the nodes in a subset $\mathcal{A} \subset \mathcal{V}$. Examples are the stations with a record, or the users of a social network who filled in a profile field. A hypothesis is a vector $\mathbf{x} \in \mathbb{R}^n$ that assigns one value to every node, and the hypothesis space is $\mathbb{R}^n$. Laplacian regularization adds the quadratic form to the average squared error loss on the labeled nodes,

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{|\mathcal{A}|} \sum_{i\in\mathcal{A}} \big(x_i - y_i\big)^2 + \alpha\, \mathbf{x}^\top \mathbf{L}^{(\mathcal{G})} \mathbf{x}, \tag{3}$$

with a hyperparameter $\alpha > 0$ (Belkin et al., 2004; Zhou and Schölkopf, 2004). Adding a penalty term is one of the three routes that regularization distinguishes, and this one implements the assumption that nodes joined by a heavy edge carry similar values. The training error alone says nothing about a node outside $\mathcal{A}$; the penalty term is the only term that constrains its value, so the generalization from the labeled to the unlabeled nodes comes from the edges. The objective (3) is convex and its minimizer solves a linear system, since the gradient of the quadratic form is $2\mathbf{L}^{(\mathcal{G})}\mathbf{x}$. As $\alpha \to 0$, the minimizer pins $x_i = y_i$ on the labeled nodes and minimizes the quadratic form over the remaining entries. The entry of $\mathbf{L}^{(\mathcal{G})}\mathbf{x}$ at an unlabeled node is then zero. Such a node takes the weighted average of its neighbors, $x_i = (1/d^{(i)}) \sum_{i'=1}^{n} A_{i,i'} x_{i'}$, which is the harmonic solution of Zhu et al. (2003).

Leaving each station of Fig. 2(a) out in turn and predicting its mean temperature by this average of its neighbors gives a mean absolute error of 0.86 degrees. Predicting the mean of the other eleven stations instead gives 1.12 degrees. The largest error is at Jauerling, 955 m above sea level, predicted at 12.3 degrees against 9.6 recorded. The edge weights encode distance but not altitude, so the penalty asks a mountain station to resemble the valley stations around it. The same quadratic form, with the model parameters of local models in place of the scalars x_i, is the generalized total variation (GTV) that generalized total variation minimization (GTVMin) penalizes in federated learning (FL). The eigenvectors of the smallest eigenvalues of a Laplacian matrix are the coordinates that spectral clustering hands to k-means (von Luxburg, 2007, Sect. 4).

Cross References

graph, weighted graph, edge weight, adjacency matrix, node degree, matrix, psd, eigenvalue, eigenvector, EVD, connected, algebraic connectivity, Fiedler vector, spectral clustering, graph clustering, similarity graph, regularization, penalty term, GTV, GTVMin

References

1. Chung (1997). Spectral Graph Theory. American Mathematical Society, Providence, RI, USA.

2. Fiedler (1973). Algebraic connectivity of graphs. Czechoslovak Mathematical Journal, 23(2), pp. 298–305.

3. Fiedler (1975). A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. Czechoslovak Mathematical Journal, 25(4), pp. 619–633.

4. von Luxburg (2007). A tutorial on spectral clustering. Statist. Comput., 17(4), pp. 395–416.

5. Mohar (1991). The Laplacian Spectrum of Graphs. In: Graph Theory, Combinatorics, and Applications, pp. 871–898. Wiley.

6. Ng et al. (2001). On Spectral Clustering: Analysis and an Algorithm. In: Adv. Neural Inf. Process. Syst., pp. 849–856.

7. Shi and Malik (2000). Normalized Cuts and Image Segmentation. IEEE Trans. Pattern Anal. Mach. Intell., 22(8), pp. 888–905.

8. Belkin et al. (2004). Regularization and semi-supervised learning on large graphs. In: COLT, pp. 624–638.

9. Hastie et al. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed., Springer Science+Business Media, New York, NY, USA.

10. Zhou and Schölkopf (2004). A regularization framework for learning from graph data. In: ICML workshop on statistical relational learning and Its connections to other fields, pp. 67–68.

11. Zhu et al. (2003). Semi-supervised learning using gaussian fields and harmonic functions. In: ICML, pp. 912–919.