Graph Laplacian and the Rankability of Data

Thomas R. Cameron and Heather Smith

April 9, 2019

Abstract

This note compiles information on the graph Laplacian for weighted directed graphs and its application to the rankability of data. Much of development follows the work of Bauer [2]. However, given the application to rankability of data, our interest lies in non-normalized graph Laplacians with the focus on the out degree rather than the in degree of the vertices.

1 Introduction

Throughout this note, we consider finite simple loopless graphs. Let $\Gamma = (V, E, w)$ be a weighted directed graph on n vertices, where V denotes the vertex set, E denotes the edge set, and $w\colon V \times V \to \mathbb{R}$ is the associated weight function of the graph. For a directed edge $e = (i, j) \in E$, we say that there is an edge from i to j. The weight of $e = (i, j)$ is given by w_{ij} and we use the convention that $w_{ij} = 0$ if and only if $e(i, j) \notin E$.

Let $\mathbb{G}$ denote the set of weighted directed graphs, and let $\mathbb{G}^+$ denote the subset of weighted directed graphs with non-negative weights. The out degree of vertex i is defined by $d_i^{out} = \sum_j w_{ij}$. A graph Γ is said to have a *spanning tree* if there exists a vertex from which all other vertices can be reached following directed edges. A directed graph Γ is *weakly connected* if replacing all of its edges with undirected edges produces a connected undirected graph. A directed graph Γ is *strongly connected* if for any pair of distinct vertices i and j, there is a path from i to j and a path from j to i.

Let $C(V)$ denote the space of complex valued functions on V. The graph *Laplace operator* for $\Gamma \in G$ is defined as $\Delta\colon C(V) \to C(V)$, where

$$\Delta f(i) = \begin{cases} f(i)d_i^{out} - \sum_j w_{ij} f(j) & \text{if } d_i^{out} \neq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

for all $f \in C(V)$. An equivalent definition of the Laplace operator for non-weighted graphs is given in [6].

We say that the vertex i is *isolated*, if $w_{ij} = 0$ for all $j \in V$. Similarly, the vertex i is said to be *quasi-isolated*, if $d_i^{out} = \sum_j w_{ij} = 0$. Note that isolated implies quasi-isolated, but not vice versa, except for graphs $\Gamma \in G^+$, where the definitions of isolated and quasi-isolated are equivalent.

Let $\Gamma = (V, E, w) \in G$ be a graph and $\Gamma' = (V', E', w')$ be an induced subgraph of Γ, i.e., $V' \subseteq V$, $E' = E \cap (V' \times V')$, and $w' = w \restriction_{E'}$. We say that Γ' is isolated if $w_{ij} = 0$ for all $i \in V'$ and $j \notin V'$. Similarly, Γ' is said to be quasi-isolated if $\sum_{j \in V \setminus V'} w_{ij} = 0$ for all $i \in V'$.

Let $V_R \subseteq V$ be the subset of vertices that are not quasi-isolated. The *reduced Laplace operator* $\Delta_R\colon C(V_R) \to C(V_R)$ is defined as

$$\Delta_R f(i) = f(i)d_i^{out} - \sum_{j \in V_R} w_{ij} f(j), \qquad i \in V_R, \tag{2}$$

where $f \in C(V_R)$ and d_i^{out} is the out degree of vertex in i in Γ.

We are particularly interested in the eigenvalues of the graph Laplace operator, which we denote by $\sigma(\Delta)$. For this reason, we often make use of the Frobenius normal form [4]:

$$\Delta = \begin{pmatrix} \Delta_1 & \Delta_{12} & \cdots & \Delta_{1z} \\ 0 & \Delta_2 & \cdots & \Delta_{2z} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Delta_z \end{pmatrix}, \tag{3}$$

where $\Delta_1, \ldots, \Delta_z$ are square matrices corresponding to the Laplace operator restricted to the strongly connected components $\Gamma_1, \ldots, \Gamma_z$ of Γ. Let V_k denote the vertex set of the strongly connected component Γ_k. Then, the off-diagonal elements of Δ_k are equal to $-w_{ij}$ for all $i, j \in V_k$ if $d_i^{out} \neq 0$ and zero otherwise; the diagonal elements are equal to d_i^{out} if $d_i^{out} \neq 0$ and zero otherwise. If V_k does not contain a quasi-isolated vertex, i.e., $d_i^{out} \neq 0$ for all $i \in V_k$, then Δ_k is irreducible. Finally, the submatrices Δ_{kl}, $1 \leq k < l \leq z$, are determined by the connectivity of different strongly connected components. In particular, Δ_{kl} contains all elements of the form $-w_{ij}$ for all $i \in V_k$ and all $j \in V_l$.

2 Spectral Properties

In what follows, we discuss the properties of the graph Laplace operator. We begin with some basic properties which we use to derive necessary and sufficient conditions for isolated subgraphs and zero eigenvalues. Finally, we prove necessary and sufficient conditions for perfect dominance graphs and their spectrum.

Proposition 2.1. *Let $\Gamma \in \mathbb{G}$ be a graph on n vertices, and let Δ be the Laplace operator of Γ. Then, the following properties hold:*

i. Zero is always an eigenvalue of Δ.

ii. The spectrum of Δ is symmetric with respect to the real axis.

iii. Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of Δ. Then,

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} d_i^{out}.$$

iv. If the weights of Γ are multiplied by a non-zero constant c, then the eigenvalues of Δ are scaled by c.

v. Let $V_R \subseteq V$ denote the subset of vertices that are not quasi-isolated. Then,

$$\sigma(\Delta) = \sigma(\Delta_R) \cup \{0, \textit{repeated } |V \setminus V_R| \textit{ times}\}.$$

vi. Let Δ be represented in its Frobenius normal form as in (3). Then,

$$\sigma(\Delta) = \bigcup_{i=1}^{z} \sigma(\Delta_i).$$

vii. The spectrum of Δ is the union of the spectra of the reduced Laplace operator over the weakly connected components of Γ.

Proof.

i. Let $\mathbf{e}$ denote the vector of all ones and $\mathbf{0}$ denote the vector of all zeros. Since the row sums of Δ are equal to zero, it follows that $\Delta\mathbf{e} = \mathbf{0}$ and, therefore, $0 \in \sigma(\Delta)$.

ii. Since the weights of Γ are real, it follows that the matrix representation of Δ is a $n \times n$ real matrix. Therefore, the eigenvalues of Δ come in complex conjugate pairs.

iii. Follows from the fact that the sum of the eigenvalues of a matrix is equal to the trace of that matrix.

iv. Suppose the weights of Γ are multiplied by a non-zero constant c. Then, the Laplace operator Δ of Γ must also be scaled by c. Therefore, the eigenvalues of Δ are scaled by c.

v. Note that Δ can be decomposed into its action over V_R and its action over $V \setminus V_R$. Suppose that $f_R \in C(V_R)$ is an eigenfunction of Δ_R. Then, $f \in C(V)$ defined by $f(i) = f_R(i)$ for $i \in V_R$ and $f(i) = 0$ for $i \in V \setminus V_R$ is an eigenfunction of Δ. Furthermore, the restriction of Δ to $V \setminus V_R$ is an $|V \setminus V_R|$ dimension operator that maps everything to zero.

vi. Follows from the Frobenius normal form of Δ and the fact that the spectrum of a block upper triangular matrix is equal to the union of the spectra of main diagonal blocks.

vii. Follows from the Frobenius normal form of Δ and the fact that the spectrum of a block upper triangular matrix is equal to the union of the spectra of main diagonal blocks.

□

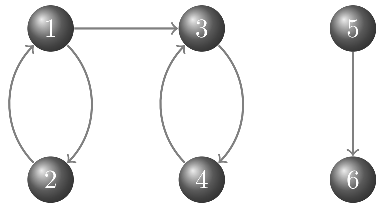

Figure 1: Directed graph with binary weights on 6 vertices.

Example 2.2. *Let Γ denote the directed graph with binary weights shown in Figure 1. Note that Γ is made up of two weakly connected components, with corresponding vertex sets $\{1,2,3,4\}$ and $\{5,6\}$. Furthermore, Γ is made up of four strongly connected components, with corresponding vertex sets $\{1,2\}$, $\{3,4\}$, $\{5\}$, and $\{6\}$. The only isolated (quasi-isolated) vertex is 6; therefore, $V_R = \{1,2,3,4,5\}$. The Laplace operator of Γ has the following Frobenius normal form:*

$$\Delta = \begin{pmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The Laplace operator restricted to V_R is the submatrix of Δ formed by rows and columns 1 through 5, which we denote by $\Delta_{1:5,1:5}$. Also, the Laplace operator restricted to the strongly connected components of Γ can be seen in the submatrices $\Delta_{1:2,1:2}$, $\Delta_{3:4,3:4}$, $\Delta_{5,5}$, and $\Delta_{6,6}$. Finally, the Laplace operator restricted to the weakly connected components of Γ can be seen in the submatrices $\Delta_{1:4,1:4}$ and $\Delta_{5:6,5:6}$.

Therefore, we can view the spectrum of Δ as the union in Proposition 2.1(v):

$$\begin{aligned} \sigma(\Delta) &= \sigma(\Delta_{1:5,1:5}) \cup \{0\} \\ &= \left\{ \frac{1}{2}(3+\sqrt{5}), 2, 1, \frac{1}{2}(3-\sqrt{5}), 0 \right\} \cup \{0\}, \end{aligned}$$

or as the union in Proposition 2.1(vi):

$$\begin{aligned} \sigma(\Delta) &= \sigma(\Delta_{1:2,1:2}) \cup \sigma(\Delta_{3:4,3:4}) \cup \sigma(\Delta_{5,5}) \cup \sigma(\Delta_{6,6}) \\ &= \left\{ \frac{1}{2}(3+\sqrt{5}), \frac{1}{2}(3-\sqrt{5}) \right\} \cup \{2,0\} \cup \{1\} \cup \{0\}, \end{aligned}$$

or as the union in Proposition 2.1(vii):

$$\begin{aligned} \sigma(\Delta) &= \sigma(\Delta_{1:4,1:4}) \cup \sigma(\Delta_{1:2,1:2}) \\ &= \left\{ \frac{1}{2}(3+\sqrt{5}), 2, \frac{1}{2}(3-\sqrt{5}), 0 \right\} \cup \{1,0\}. \end{aligned}$$

□

2.1 Isolated Components

Proposition 2.1 (vi–vii) implies that it sufficient to study the spectral properties of strongly and weakly connected components of a graph Γ. To that end, we begin with the following simple observation.

Lemma 2.3. *Let $\Gamma \in \mathbb{G}$ and Δ be the Laplace operator Δ of Γ be represented in its Frobenius normal form as in (3). Then,*

i. If Δ_i is isolated, then $\Delta_{ij} = 0$ for all $j > i$.

ii. If Δ_i is quasi-isolated, then the row sums of $\Delta_{i,(i+1)}, \ldots, \Delta_{i,z}$ add up to zero.

Moreover, if $\Gamma \in \mathbb{G}^+$, then

iii. Δ_i is isolated if and only if $\Delta_{ij} = 0$ for all $j > i$.

iv. Δ_i *is quasi-isolated if and only if the row sums of* $\Delta_{i,(i+1)}, \ldots, \Delta_{i,z}$ *add up to zero.*

While the spectrum of the Laplacian over weakly and strongly connected subgraphs of Γ is a subset of the spectrum of the Laplacian of Γ, this is not, in general, true for all induced subgraphs Γ' of Γ. The following result establishes sufficient conditions for when $\sigma(\Delta(\Gamma')) \subseteq \sigma(\Delta(\Gamma))$.

Lemma 2.4. *Let* $\Gamma \in \mathbb{G}$ *and* Γ' *be an induced subgraph of* Γ. *If one of the following conditions is satisfied*

i. Γ' *consists of* p, $1 \leq p \leq z$, *strongly connected components of* Γ *and is quasi-isolated,*

ii. Γ' *is isolated,*

then $\sigma(\Delta(\Gamma')) \subseteq \sigma(\Delta(\Gamma))$.

Proof.

i. Suppose that Γ' is quasi-isolated and consists of p strongly connected components of Γ. Without loss of generality, we assume that Γ' consists of the strongly connected components $\Gamma_1, \ldots, \Gamma_p$. Furthermore, we note that any $i \in V'$ such that $i \notin \bigcup_{j=1}^{p} \Gamma_j$ must be a quasi-isolated vertex and, therefore, corresponds to a zero eigenvalue of Δ. Therefore, it suffices to assume that $\Gamma' = \bigcup_{j=1}^{p} \Gamma_j$. Finally, since Γ' is quasi-isolated, for all $i \in V'$ we have

$$d_i^{out} = \sum_{j \in V} w_{ij} = \sum_{j \in V'} w_{ij} + \sum_{j \in V \setminus V'} w_{ij} = \sum_{j \in V'} w_{ij},$$

i.e., the out-degree of each vertex $i \in V'$ is not affected by the vertices in $V \setminus V'$. Thus,

$$\sigma(\Delta(\Gamma')) = \bigcup_{i=1}^{p} \sigma(\Delta_i) \subseteq \bigcup_{i=1}^{z} \sigma(\Delta_i) = \sigma(\Delta(\Gamma)).$$

ii. If Γ' is isolated, then Γ' must consist of p, $1 \leq p \leq z$, strongly connected components of Γ. Therefore, the second assertion follows from the first.

□

We will make use of the following famous result from Olga Taussky in our proof of Lemma 2.6.

Theorem 2.5 (Taussky)**.** *A complex* $n \times n$ *matrix* A *is non-singular if* A *is irreducible and* $|A_{ii}| \geq \sum_{j \neq i} |A_{ij}|$ *with equality in at most* $(n-1)$ *cases.*

Proof. See Theorem II in [5]. □

Lemma 2.6. *Let* $\Gamma \in \mathbb{G}^+$ *and let the Laplace operator* Δ *of* Γ *be represented in its Frobenius normal form as in* (3)*. Then, zero is an eigenvalue (in fact a simple eigenvalue) of* Δ_i *if and only if* Γ_i *is isolated.*

Proof. We observe that the result is trivial for Γ_i consisting of a single vertex since $\Delta_i = 0$ in that case. Therefore, we assume that Γ_i consists of more than one vertex and note that since $\Gamma \in \mathbb{G}^+$ it follows that Δ_i is irreducible.

Suppose that Γ_i is not isolated. Then, there exists a vertex $k \in V_i$ such that $w_{kl} \neq 0$ for some $l \notin V_i$. Therefore, since the weights are non-negative, we have

$$|\Delta_{kk}| = d_k^{out} > \sum_{j \in V_i} w_{kj} = \sum_{j \in V_i} |\Delta_{kj}|.$$

For all other $j \in V_i$, we have

$$|\Delta_{jj}| = d_j^{out} \geq \sum_{l \in V_i} w_{jl} = \sum_{l \in V_i} |\Delta_{jl}|.$$

Therefore, by Theorem 2.5, it follows that Δ_i is non-singular, i.e., zero is not an eigenvalue of Δ_i.

Now, suppose that Γ_i is isolated. Note that multiplying a row of Δ_i by a non-zero scalar will not affect the potential zero eigenvalue of Δ_i. Therefore, we consider Δ_i in its normalized form, i.e., we divide each nonzero row of Δ_i by the corresponding diagonal entry in that row. Next, we define $P = I - \Delta_i$, where I is the identity matrix of the same size as Δ_i. Note that P is a non-negative, irreducible, row-stochastic matrix. Therefore, by the classical Perron-Frobenius theorem [3], it follows that one is a simple eigenvalue of P and, hence, zero is a simple eigenvalue of Δ_i. □

Theorem 2.7. *Let $\Gamma \in \mathbb{G}^+$. Then, the following statements are equivalent:*

i. The algebraic multiplicity of the zero eigenvalue of Δ is equal to $k \in \mathbb{N}$.

ii. There exist k isolated strongly connected components in Γ.

iii. The minimum number of directed trees needed to span the whole graph is equal to $k \in \mathbb{N}$.

Proof. Note that $(i) \Leftrightarrow (ii)$ follows from Lemma 2.6 and $(ii) \Leftrightarrow (iii)$ follows from the Frobenius normal form and Lemma 2.3. □

2.2 Directed Acyclic Graphs

Let $\Gamma \in \mathbb{G}$. We defined a *directed cycle* in Γ as a cycle with all edges oriented in the same direction. Furthermore, a vertex of Γ that is contained in at least one directed cycle is called a *cyclic vertex*. We say that Γ is a *directed acyclic graph* if none of its vertex are cyclic. The set of all directed acyclic graphs is denoted by $\mathbb{G}^{ac}$. We begin this section with the following observation that follows immediately from the Frobenius normal from Δ.

Lemma 2.8. *The following statements are equivalent:*

i. $\Gamma \in \mathbb{G}^{ac}$ is a directed acyclic graph.

ii. Every strongly connected component of Γ consists of exactly one vertex.

iii. Δ represented in its Frobenius normal form is upper triangular.

We are now ready to provide a spectral characterization for directed acyclic graphs with non-negative weights.

Theorem 2.9. *The following results hold for directed acyclic graphs.*

i. If $\Gamma \in \mathbb{G}^{ac}$, then $\sigma(\Delta) = \{d_1^{out}, \dots, d_{n-1}^{out}, d_n^{out}\}$. Furthermore, the algebraic multiplicity of the zero eigenvalue is equal to $|V \setminus V_R|$, where $V_R \subseteq V$ is the set of vertices that are not quasi-isolated.

ii. Let $\Gamma \in \mathbb{G}^+$. Then, $\sigma(\Delta) = \{d_1^{out}, \dots, d_{n-1}^{out}, d_n^{out}\}$ if and only if $\Gamma \in \mathbb{G}^{ac,+}$.

Proof.

i. If $\Gamma \in \mathbb{G}^{ac}$, then it follows from Lemma 2.8 that the Frobenius normal form of Δ is upper triangular. Since the eigenvalues of an upper triangular matrix are its main diagonal entries, it follows that $\sigma(\Delta) = \{d_1^{out}, \dots, d_{n-1}^{out}, d_n^{out}\}$. Furthermore, every zero eigenvalue corresponds to a vertex that is quasi-isolated, and the results follows.

ii. All that remains is to show that if $\Gamma \in \mathbb{G}^+$ and $\sigma(\Delta) = \{d_1^{out}, \dots, d_{n-1}^{out}, d_n^{out}\}$, then $\Gamma \in \mathbb{G}^{ac,+}$. For the sake of contradiction, we suppose that $\Gamma \in \mathbb{G}^+$ and $\sigma(\Delta) = \{d_1^{out}, \dots, d_{n-1}^{out}, d_n^{out}\}$, but $\Gamma \notin \mathbb{G}^{ac,+}$. Then, by Lemma 2.8, there exists a strongly connected component Γ_i in Γ consisting of at least two vertices.

   First, assume that Γ_i is isolated. Then, by Lemma 2.6, exactly one eigenvalue of Δ_i is equal to zero.

   Now, suppose that Γ_i is not isolated.

□

As an immediate corollary of Theorem 2.9, we have a spectral characterization for perfect dominance graphs.

Corollary 2.10. *Let Γ be a directed graph with binary weights on n vertices. Then, Γ is a perfect dominance graph if and only if $\sigma(\Delta) = \{d_1^{out}, \dots, d_{n-1}^{out}, d_n^{out}\}$ and there is a re-ordering of the vertices such that $d_i^{out} = n - i$ for $i = 1, \dots, n$.*

Proof. Let $\Gamma \in \mathbb{G}^{ac,+}$, then it follows from Theorem 2.9 that $\sigma(\Delta) = \{d_1^{out}, \dots, d_{n-1}^{out}\}$. Furthermore, if Γ is a perfect dominance graph, then there is a re-ordering of the vertices such that $d_i^{out} = n - i$ for $i = 1, \dots, n$.

Conversely, suppose that $\sigma(\Delta) = \{d_1^{out}, \dots, d_n^{out}\}$ and there exists a re-ordering of the vertices such that $d_i^{out} = n - i$ for $i = 1, \dots, n$. Then, it follows from Theorem 2.9 that Γ must be an acyclic graph. Therefore, vertex 1 must point to all other vertices, and there is no vertex that points to vertex 1. Similarly, vertex 2 must point to vertices $i = 3, \dots, n$, and there is not vertex $i \neq 1$ that points to vertex 2. Repeating this argument for all vertices, we conclude that Γ is a perfect dominance graph. □

3 Rankability Measure

Recently a new problem was proposed in [1], the *rankability problem*, which refers to a dataset's inherent ability to produce a meaningful ranking of its items. In their paper, Anderson et al. introduce a rankability measure based on the distance of a dataset's adjacency matrix and the closest perfect dominance graph(s). This distance is defined by the minimum number of changes k, i.e., edges added or removed, that are needed to obtain a perfect dominance graph, and the number p of perfect dominance graphs that can be obtained with k changes. Computing this measure is expensive, but in doing so one obtains an abundance of information about the dataset. In this section, we propose another strategy for computing a rankability measure. While our measure does not include the additional information gained by the measure in [1], it is far more efficient to compute and has the potential to be used on large datasets.

Algorithm 1 Rankability measure of dataset with given adjacency matrix A.

function $[r] = \text{rank}(A)$:
  $n = \text{size}(A)$
  $x = [\text{sum}(A[i,:]) \text{ for } i \text{ in range}(n)]$
  $D = \text{diag}(x)$
  $L = D - A$
  $s = [n - k \text{ for } k \text{ in range}(1, n+1)]$
  $e = \text{eigvals}(L)$
  $r = \frac{d(e,s)+d(x,s)}{2(n-1)}$

Example 3.1. *For the graphs in Figure 3 of [1], we compute our rankability measure with respect to the Hausdorff and Matching distance. The results are shown in the table below.*

Graph	Hausdorff	Matching
Dominance Graph	0.0	0.0
Perturbed Dominance Graph	0.062	0.062
Perturbed Random Graph	0.223	0.270
Nearly Disconnected	0.400	0.400
Random	0.565	0.565
Cyclic	0.70	0.70
Completely Connected	0.80	1.0
Empty Graph	1.0	1.0

References

[1] P. E. Anderson, T. P. Chartier, and A. N. Langville, *The rankability of data*, SIMODS, (2018).

[2] F. Bauer, *Normalized graph laplacians for directed graphs*, Linear Algebra Appl., 436 (2012), pp. 4193–4222.

[3] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, SIAM, Philadelphia, PA, 1994.

[4] R. Brualdi and H. Ryser, *Combinatorial Matrix Theory*, Cambridge University Press, United Kingdom, 1991.

[5] O. Taussky, *A recurring theorem on determinants*, Amer. Math. Monthly, 56 (1949), pp. 672–676.

[6] C. W. Wu, *On Rayleigh-Ritz ratios of a generalized Laplacian matrix of drected graphs*, Linear Algebra Appl., (2005), pp. 207–227.