Nonnegative and spectral matrix theory with applications to network analysis

Dario Fasino[*]

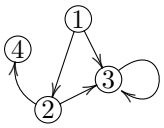
Lecture notes (extended version) for the course held at the Rome-Moscow School on Matrix Methods and Applied Linear Algebra 2018. This document is intended for internal circulation only.

1 Basic definitions

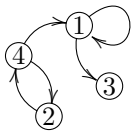
1.1 Nonnegative matrices and graphs

A *(directed) graph* is a pair $\mathcal{G} = (V, E)$ where V is a finite set of *nodes* (or *vertices*) and $E \subseteq V \times V$ is a set of *(oriented) edges*. In pictures, nodes can be visualized as points and edges as arrows or lines joining them. Hereafter, I generally assume $V = \{1, \dots, n\}$ and write $i \to j$ to indicate that $(i, j) \in E$. Edges of the form (i, i) are called *loops*.

A graph $\mathcal{G} = (V, E)$ can be completely described by its *adjacency matrix*, which is the $n \times n$ matrix A such that $A_{ij} = 1$ if $j \to i$ and $A_{ij} = 0$ otherwise.[1] The notation $A = A_{\mathcal{G}}$ indicates that A is the adjacency matrix of $\mathcal{G}$.

$\mathcal{G}:$  $\rightsquigarrow \qquad A_{\mathcal{G}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$

Conversely, for any given matrix $A \in \mathbb{R}^{n \times n}$ the *graph associated to* A is the graph $\mathcal{G}_A = (V, E)$ such that $V = \{1, \dots, n\}$ and $j \to i \iff A_{ij} \neq 0$. Thus, if the entries of A belong to the set $\{0, 1\}$ then A is the adjacency matrix of $\mathcal{G}_A$.

$A = \begin{pmatrix} * & 0 & 0 & * \\ 0 & 0 & 0 & * \\ * & 0 & 0 & 0 \\ 0 & * & 0 & 0 \end{pmatrix} \qquad \rightsquigarrow \qquad \mathcal{G}_A:$ 

A graph is *undirected* (or *non-oriented*) when $i \to j \iff j \to i$, that is, when its adjacency matrix is symmetric. In that case, edges are depicted as lines instead of arrows, and the notation $i \sim j$ replaces both $i \to j$ and $j \to i$.

Let $A = A_{\mathcal{G}}$ and let $v \in \mathbb{R}^n$. If we consider v_i as a score placed on node i then it is useful to look at the matrix-vector product $w = Av$ as propagating the scores along the edges of $\mathcal{G}$. In fact, direct inspection shows that the numbers w_i are obtained by propagating the v_i's along the edges of the graph and summing up the contributions arriving at each node:

$$w_i = \sum_{j=1}^{n} A_{ij} v_j = \sum_{j : j \to i} v_j.$$

Definition 1.1. *A* walk of length $k \geq 1$ *in* $\mathcal{G}$ *is any sequence of nodes* $i_0, i_1, \dots, i_k$ *such that* $i_{j-1} \to i_j$ *(or* $i_{j-1} \sim i_j$ *in the undirected case) for* $j = 1, \dots, k$*. We say that the walk* $i_0, i_1, \dots, i_k$ *starts at* i_0 *and terminates at* i_k*.*

[*]Dept. Mathematics, Computer Science and Physics, University of Udine, Italy. Email: dario.fasino@uniud.it.
The author's work has been partially supported by Istituto Nazionale di Alta Matematica (INdAM, Italy).

[1] Various authors define the adjacency matrix as $A_{ij} = 1$ if $i \to j$ and 0 otherwise. I prefer the other definition for simplicity of subsequent notations.

An useful result is included here below; the simple proof proceeds by induction and is omitted for brevity.

Lemma 1.2. *Let $A = A_{\mathcal{G}}$. For any $k \in \mathbb{N}$ and $i, j = 1, \dots, n$ the value of $(A^k)_{ij}$ is equal to the number of different walks of length k starting at j and terminating at i.*

Remark 1.3. *Many applications require the treatment of* weighted graphs*. These are graphs where every edge $(i, j) \in E$ is associated to a real (usually positive) number, say w_{ij}. For example, w_{ij} may indicate the strenght of the tie or the length of a physical link existing between nodes i and j in a network found in the real world. In this case the adjacency matrix is defined as $A_{ij} = w_{ji}$ if $j \to i$ and $A_{ij} = 0$ otherwise. In this document graphs are not weighted. By the way, virtually all results presented here apply also to weighted graphs, with due (almost obvious) modifications.*

Other useful notations are the following:

- An all-zeros matrix is denoted by O. An all-ones vector is denoted by $\mathbf{1}$. The i-th canonical vector is denoted by e_i and I denotes an identity matrix.
- Inequality operators like $\geq$ or $>$ are extended to matrices and vectors in the entrywise sense; for example, $A \geq O$ means that all elements of A are nonnegative. Analogously, the absolute value is extended to vectors entrywise: If $v = (v_1, \dots, v_n)^T$ then $|v| = (|v_1|, \dots, |v_n|)^T$.
- The spectral radius of a square matrix A is denoted by $\rho(A)$:

$$\rho(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

1.2 Irreducible matrices and strongly connected graphs

Definition 1.4. *The matrix $A \in \mathbb{R}^{n\times n}$ is* reducible *if there is a permutation matrix P such that the matrix $B = PAP^T$ is in (lower) block triangular form:*

$$B = PAP^T = \begin{pmatrix} B_{11} & B_{12} \\ O & B_{22} \end{pmatrix},$$

where the diagonal blocks B_{11}, B_{22} are square matrices. An irreducible *matrix is a matrix that is not reducible.*

Exercise 1.5. Prove this: A is irreducible $\iff A^T$ is irreducible.

Definition 1.6. *A graph is* strongly connected *if any two nodes are connected by a walk.*

The two preceding definitions are connected by the following important result:

Theorem 1.7. *A matrix $A \in \mathbb{R}^{n\times n}$ is irreducible if and only if $\mathcal{G}_A$ is strongly connected.*

Proof. Suppose that A is reducible. Apart of a permutation (which corresponds to a renumbering of the nodes of $\mathcal{G}_A$) we can assume that A is already in reduced block triangular form:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ O & A_{22} \end{pmatrix}, \qquad A_{11} \in \mathbb{R}^{n_1\times n_1}, \qquad A_{22} \in \mathbb{R}^{n_2\times n_2}, \tag{1}$$

with $n_1+n_2 = n$. Hence, in $\mathcal{G}_A$ there are no edges connecting nodes $1, \dots, n_1$ to nodes $n_1+1, \dots, n$. As a consequence, there are no walks going from nodes $n_1 + 1, \dots, n$ to nodes $1, \dots, n_1$, and the graph is not strongly connected.

Conversely, if $\mathcal{G}_A$ is not strongly connected then there are two distinct nodes, say i and j, such that there is no walk from i to j. Let $\mathcal{I}$ be the set of all nodes that can be reached by a walk starting from i, and let $\mathcal{J}$ be its complementary set. Without loss of generality, we can suppose that $\mathcal{I} = \{1, \dots, n_1\}$ and $\mathcal{J} = \{n_1 + 1, \dots, n\}$. It is not difficult to realize that A has the form (1), hence it is reducible. ∎

2 An introduction to Perron–Frobenius theory

At the beginning of 20th century, Oskar Perron investigated spectral properties of matrices having all positive entries. Shortly after, Ferdinand G. Frobenius extended most of Perron's results to some matrices with nonnegative entries. Today we call Perron–Frobenius theory a wealth of results on spectral properties of nonnegative matrices and operators, originated by those studies.[2]

[2] Besides C. Meyer's book [12], a good reference for Perron–Frobenius theory is the book by R. S. Varga, *Matrix Iterative Analysis* (1962).

Quoting from [12, p. 662]: *"The Perron–Frobenius theory is elegant. It is a testament to the fact that beautiful mathematics eventually tends to be useful, and useful mathematics eventually tends to be beautiful."*

2.1 Perron's theorem

Theorem 2.1. *Let $A \in \mathbb{R}^{n\times n}$, $A > O$. Then, (1) A has a positive eigenvalue equal to $\rho(A)$. (2) To $\rho(A)$ it is associated an eigenvector $x > 0$. (3) $\rho(A)$ is a simple eigenvalue of A, that is, it corresponds to a single Jordan block of order* 1. *(4) $\rho(A)$ is a* dominant *eigenvalue of A, that is, if μ is any other eigenvalue of A then $|\mu| < \rho(A)$.*

Proof. In this proof,[3] the following simple fact will be used repeatedly.

Lemma 2.2. *If $A > O$ then for any vector $y \geq 0$ with $y \neq 0$ we have $Ay > 0$. In particular, for any real vector z, there exists a real number $\varepsilon > 0$ such that $Ay > \varepsilon z$.*

For any nonzero vector $v \geq 0$ let

$$\Lambda_v = \{\varepsilon > 0 : Av \geq \varepsilon v\}$$

and set $\Lambda = \cup_{v\gneq 0}\Lambda_v$. It is easy to see that $\Lambda \neq \varnothing$ (for example, $A_{ii} \in \Lambda_{e_i}$) and that it is bounded. Let $\lambda = \sup\Lambda$. We will show that there exists a vector $x > 0$ such that $Ax = \lambda x$. First, by the definition of λ, we can pick a sequence $\{\lambda^{(j)}\}$ converging to λ. By the definition of Λ we also have vectors $x^{(j)} \geq 0$ such that $Ax^{(j)} \geq \lambda^{(j)}x^{(j)}$. Without loss of generality, we can choose $x^{(j)}$ with $\|x^{(j)}\| = 1$. Using a compactness argument, after possibly passing to a subsequence we can also assume that $x^{(j)}$ converges to a vector x. By the way x is obtained we have $x \geq 0$, $Ax \geq \lambda x$, and $\|x\| = 1$.

To show that $Ax = \lambda x$, proceed by contradiction. Assume that $Ax \neq \lambda x$ and let $y = Ax - \lambda x$. Thus $y \neq 0$ by assumption, and since $y = \lim_{j\to\infty} Ax^{(j)} - \lambda^{(j)}x^{(j)}$, we also have $y \geq 0$. Let $z = Ax$. By Lemma 2.2 there is $\varepsilon > 0$ so that $Ay \geq \varepsilon z$. Moreover,

$$Az = A(Ax) = Ay + \lambda Ax \geq (\lambda_1 + \varepsilon)z,$$

hence $\lambda + \varepsilon \in \Lambda_z$, contradicting the fact that $\lambda = \sup\Lambda$. So $Ax = \lambda x$. In addition, Lemma 2.2 also implies that $\lambda x = Ax > 0$ and therefore $x > 0$.

Now we show that $\lambda = \rho(A)$. If μ is an eigenvalue (which in general is a complex number) of A with an eigenvector v, then by triangle inequality we have $A|v| \geq |Av| = |\mu||v|$, hence $|\mu| \in \Lambda_{|v|}$. Since $\lambda = \sup\Lambda$ we obtain $\lambda \geq |\mu|$, hence $\lambda = \rho(A)$.

To prove claim (3) we use the following lemma whose proof relies on the Jordan normal form of A and is not included.[4]

Lemma 2.3. *Let $Ax = \lambda x$. Then, λ is not simple if and only if there exists $y \neq 0$ such that $y^TA = \lambda y^T$ and $y^Tx = 0$.*

If λ is not simple then by Lemma 2.3 there exists a vector $0 \neq y \in \mathbb{R}^n$ such that $y^TA = \rho(A)y^T$ and $y^Tx = 0$. Since $x > 0$, y must have both positive and negative entries. Let z be the vector defined as

$$z_j = \begin{cases} y_j & \text{if } y_j > 0 \\ 0 & \text{else.} \end{cases}$$

It is not difficult to see that $\lambda z^T \leq z^TA$ with strict inequality in at least one entry. Owing to the inequality $x > 0$ we deduce

$$\lambda z^Tx < z^TAx = \lambda z^Tx,$$

a contradiction.

To prove that $\rho(A)$ is dominant let $\varepsilon > 0$ be such that $B = A - \varepsilon I > 0$. Note that μ is an eigenvalue of A iff $\mu - \varepsilon$ is an eigenvalue of B. In particular, $\rho(B) = \rho(A) - \varepsilon$. Now, if $|\mu - \varepsilon| \leq \rho(B)$ then the identity $|\mu| = \rho(A)$ is fulfilled iff $\mu = \rho(A)$. So if $\mu \neq \rho(A)$ then $|\mu| < \rho(A)$, that is, $\rho(A)$ is dominant. ∎

[3] This proof of Perron's theorem is borrowed from L. Ni, *A Perron-type theorem on the principal eigenvalue of nonsymmetric elliptic operators*. Amer. Math. Monthly, 121 (2014), 903–908.

[4] To prove Lemma 2.3 it is sufficient to consider the case when A is diagonal or it consists of a single nontrivial Jordan block. A complete proof can be established on the basis of the discussion of these two fundamental cases.

2.2 Extension to irreducible nonnegative matrices

Theorem 2.1 cannot be immediately extended to nonnegative matrices, as shown by simple counterexamples.

Exercise 2.4. For any of the following matrices, what claim of Perron's theorem is not fulfilled?

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

If $A \geq O$ then we can only say that there exists a vector $x \geq 0$ such that $Ax = \rho(A)x$ with $\rho(A) \geq 0$ (use a continuity argument). However, if we add the irreducibility hypothesis then we can say much more.

Theorem 2.5 (Perron–Frobenius)**.** *Let $A \in \mathbb{R}^{n \times n}$ be an irreducible, nonnegative matrix. Then, all claims in Theorem 2.1 hold true except the last one.*

Proof. (sketch) First of all, observe that if $A \geq O$ is irreducible then $\rho(A) > 0$. In fact, if $\rho(A) = 0$ then A is *nilpotent*, that is, there exists a positive integer m such that $A^m = O$. In particular, $A^m e_1 = 0$. According to the "score propagation" interpretation of matrix-vector products with powers of A, we conclude that all walks in $\mathcal{G}_A$ starting from node 1 sooner or later arrive to nodes without outgoing edges. But, owing to Theorem 1.7, $\mathcal{G}_A$ is strongly connected, a contradiction.

For any $0 < \alpha < 1/\rho(A)$ the matrix $I - \alpha A$ is invertible and $B = (I - \alpha A)^{-1} > O$. Moreover, A and B have the same eigenvectors, and λ is an eigenvalue of A iff $\mu = 1/(1 - \alpha\lambda)$ is an eigenvalue of B. From this, one can deduce that also claims (2) and (3) of Theorem 2.1 hold true. ∎

It is usual to call $\rho(A)$ the *Perron eigenvalue* of A. Any associated positive eigenvector is a *Perron eigenvector*. The forthcoming lemma shows easily computable lower and upper bounds for the Perron eigenvalue of a nonnegative irreducible matrix.

Lemma 2.6. *Let $A \geq O$ be irreducible. Suppose that for some vector $w \geq 0$ and scalars $0 \leq \alpha < \beta$ we have $\alpha w \leq Aw \leq \beta w$, with strict inequalities in at least one entry. Then $\alpha < \rho(A) < \beta$.*

Proof. Since A^T is nonnegative and irreducible (why?), there exists a vector $y > 0$ such that $y^T A = \rho(A) y^T$. We have $y^T w > 0$ and moreover,

$$\alpha y^T w < y^T A w < \beta y^T w,$$

and the claim follows from the identity $y^T A w = \rho(A) y^T w$. ∎

The Perron eigenvalue of a nonnegative, irreducible matrix is a monotonic function of its entries, as shown in the following result.

Theorem 2.7. *If A and B are two nonnegative, irreducible matrices with $O \leq A \leq B$ and $A \neq B$ then $\rho(A) < \rho(B)$.*

Proof. Let z be a Perron eigenvector of B. We have $Az \lneqq Bz = \rho(B)z$, and the claim follows from Lemma 2.6. ∎

2.3 Primitive matrices

Definition 2.8. *A matrix $A \geq O$ is* primitive *if there exists a positive integer k such that $A^k > O$.*

Note that a positive matrix is primitive (set $k = 1$). Moreover, a primitive matrix is necessarily irreducible (since the powers of a block triangular matrix are in the same block triangular form) but the converse is not true (the matrix $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ is a counterexample). The following result says that the Perron eigenvalue of a primitive matrix $A \geq O$ is dominant; all other eigenvalues are smaller in modulus.

Theorem 2.9. *If $A \geq O$ is primitive then $\rho(A)$ is a* dominant *eigenvalue, that is, if λ is any eigenvalue of A different from $\rho(A)$ then $|\lambda| < \rho(A)$.*

PROOF. By contradiction, let λ be an eigenvalue of A such that $\lambda \neq \rho(A)$ and $|\lambda| = \rho(A)$. Let k be an integer such that $A^k > O$. Then both λ^k and $\rho(A)^k$ are eigenvalues of A^k. However $|\lambda|^k = \rho(A)^k$ thus violating Theorem 2.1. ■

Remark 2.10. *Theorem 2.1 allows us to conclude that if we apply the power method (with normalization) to a nonnegative, primitive matrix starting from a positive vector then the method will converge to a Perron vector:*

$$x^{(0)} > 0,\ x^{(k+1)} = Ax^{(k)}/\|Ax^{(k)}\| \quad \Longrightarrow \quad \lim_{k\to\infty} x^{(k)} = x,\ Ax = \rho(A)x.$$

This fact is not true for a generic irreducible $A \geq O$. For example, examine the behaviour of the power method applied to the matrix $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ which is nonnegative and irreducible but not primitive.

2.4 Bounding perturbations on Perron vectors

The forthcoming theorem, borrowed from [5], provides a bound on the relative change in a Perron eigenvector when some matrix rows are changed. The result states that when a few rows of a nonnegative irreducible matrix are modified then the relative changes in the corresponding elements of the Perron vector bound the relative changes in the other elements.

Theorem 2.11. *Let $A, \hat{A}$ be irreducible, nonegative matrices, let $Ax = \rho x$ and $\hat{A}\hat{x} = \hat{\rho}\hat{x}$ be the corresponding Perron eigenpairs. Suppose that $\hat{A}$ is obtained by changing a few entries of A. Let $\mathcal{I}$ be the index set of unchanged rows:*

$$\mathcal{I} = \{i : A_{i,:} = \hat{A}_{i,:}\}.$$

Hence,

$$\forall i \in \mathcal{I}, \qquad \frac{\rho}{\hat{\rho}} \min_{j=1\ldots n} \frac{\hat{x}_j}{x_j} \leq \frac{\hat{x}_i}{x_i} \leq \frac{\rho}{\hat{\rho}} \max_{j=1\ldots n} \frac{\hat{x}_j}{x_j}.$$

In particular, if $\hat{\rho} > \rho$ then $\max_{i\in\mathcal{I}} \frac{\hat{x}_i}{x_i} < \max_{j\notin\mathcal{I}} \frac{\hat{x}_j}{x_j}$ while if $\hat{\rho} < \rho$ then $\min_{i\in\mathcal{I}} \frac{\hat{x}_i}{x_i} < \min_{j\notin\mathcal{I}} \frac{\hat{x}_j}{x_j}$.

PROOF. Firstly, note that by hypotheses we have $x, \hat{x} > 0$. For any $i \in \mathcal{I}$ we have

$$\begin{aligned} \frac{\hat{x}_i}{x_i} = \frac{\hat{\rho}\hat{x}_i}{\hat{\rho}x_i} &= \frac{1}{\hat{\rho}x_i} \sum_j \hat{A}_{ij}\hat{x}_j \\ &= \frac{1}{\hat{\rho}x_i} \sum_j A_{ij}x_j \frac{\hat{x}_j}{x_j} \\ &\leq \frac{1}{\hat{\rho}x_i} \left(\max_j \frac{\hat{x}_j}{x_j}\right) \rho x_i = \frac{\rho}{\hat{\rho}} \left(\max_j \frac{\hat{x}_j}{x_j}\right). \end{aligned}$$

The opposite inequality is obtained analogously. Furthermore, if $\rho/\hat{\rho} < 1$ then $\max_{i\in\mathcal{I}} \frac{\hat{x}_i}{x_i} < \max_j \frac{\hat{x}_j}{x_j}$, whence $\max_j \frac{\hat{x}_j}{x_j} = \max_{j\notin\mathcal{I}} \frac{\hat{x}_j}{x_j}$, and analogously for the other inequality when $\rho/\hat{\rho} > 1$. ■

Exercise 2.12. Prove the following result:[5] Let v be a nonnegative vector, let $B = A + e_i v^T$. If x, y are positive Perron vectors of A and B, respectively, then $y_i/x_i > y_j/x_j$ for $j \neq i$.

2.5 Applications: A simple epidemic model

If the graph $\mathcal{G}$ represents a computer network, or a social network, and $A = A_{\mathcal{G}}$, then the number $\rho(A)$ plays an important role in modelling (computer or biologic, respectively) virus propagation in $\mathcal{G}$. The smaller $\rho(A)$ the better the robustness of the network against the spread of viruses. Hereafter, I present a simple virus propagation model which has been discussed e.g., in [4, §2.7] and [3].[6]

[5] Found in L. Elsner, C. Johnson, M. Neumann; *Czech. Math. J.* 32 (1982), 99–109.

[6] A nonlinear, more sophisticated model of virus propagation in computer networks has been developed in: Van Mieghem P., Omic J., Kooij R., Virus spread in networks, *IEEE/ACM Transactions on Networking* 17 (2009), 1–14, with the same conclusion concerning $\rho(A)$.

Consider a virus spreading on $\mathcal{G}$; at each time step, a contagious node may infect its neighbors with probability δ (virus birth rate). At the same time, an infected node may also be cured with probability β (virus curing rate). If the number $p_i(t)$ measures the amount of infection of node i at time t, then the model is

$$p_i(t) = (1-\beta)p_i(t-1) + \delta \sum_{j:j\to i} p_j(t-1),$$

where $p(t) = (p_1(t), \dots, p_n(t))^T$ and $p(0)$ is the initial state of infection. With simple passages, the model can be rewritten in matrix notation as

$$p(t) = Mp(t-1), \qquad M = (1-\delta)I + \beta A.$$

Hence, if the initial vector $p(0)$ is known then the preceding equation allows to simulate the evolution of the epidemics for $t = 1, 2\dots$ In fact, the simulation reduces to the iterations of the power method for the matrix M. As a consequence, we have the following cases:

- If $\rho(M) > 1$ then the epidemic will affect the whole network, independently on the initial state of the network.
- If $\rho(M) = 1$ then the network will approach a stationary state that may depend on the initial state.
- If $\rho(M) < 1$ then the epidemic will fade out, independently on the initial state.

Hence, the asymptotic behaviour of the epidemic is governed by $\rho(M)$, and the entries of the Perron eigenvector of M indicate the long-term infection level of each node. However, $\rho(M)$ depends on $\rho(A)$, as shown hereafter.

Theorem 2.13. *$\rho(A) = (\rho(M) - 1 + \delta)/\beta$.*

Proof. Let x be a Perron vector of M, $Mx = \rho(M)x$. Then,

$$\rho(M)x = ((1-\delta)I + \beta A)x = (1-\delta)x + \beta Ax.$$

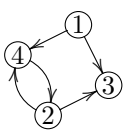
Then $Ax = \lambda x$ with $\lambda = (\rho(M) - 1 + \delta)/\beta$. Since $x \geq 0$, λ must be the Perron eigenvalue of A, that is, $\rho(A) = \lambda$, and the claim is proved. ∎

In conclusion, $\rho(M) < 1$ if and only if $\rho(A) < \delta/\beta$. Hence, the spectral radius $\rho(A)$ plays an important role in modelling virus propagation. The critical threshold that separates global network infection from global network health is determined by $\rho(A)$. The epidemic will fade out if and only if the ratio δ/β is larger than the threshold. This fact leads also to the conclusion that, if we want to contrast the epidemic by immunizing some of the nodes, then the best policy is to immunize those nodes whose removal from the network yields the largest decrease in the spectral radius of the adjacency matrix [3].

2.6 More exercises and problems

1. Let $A \geq O$ be irreducible. Prove this: If $(\lambda I - A)^{-1}$ exists and is nonnegative then $\lambda > \rho(A)$.
   Hint: Note that $(\lambda I - A)^{-1}$ must be irreducible. Let (μ, x) be a Perron eigenpair of $(\lambda I - A)^{-1}$. Deduce $Ax = (\lambda - 1/\mu)x$ and prove that x is a Perron vector of A.

2. A graph $\mathcal{G} = (V, E)$ is called *bipartite* if $V = V_1 \cup V_2$, $V_1 \cap V_2 = \varnothing$, and every edge belongs to either $V_1 \times V_2$ or $V_2 \times V_1$. In other words, a graph is bipartite if and only if its nodes can be colored black or white in such a way that no edge connects two nodes with the same color. For example, the graph on page 1 is not bipartite while the following one is:

   $\mathcal{G}:$  $\rightsquigarrow \qquad A_G = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$

   Let $\mathcal{G}$ be a strongly connected bipartite graph.

(a) Prove that $A_{\mathcal{G}}$ is not primitive.

(b) Assume that $\mathcal{G}$ is undirected. Find all eigenpairs (λ, x) of $A_{\mathcal{G}}$ with $|\lambda| = \rho(A_{\mathcal{G}})$.

Hint: Assume $V_1 = \{1, \dots, m\}$ *and* $V_2 = \{m+1, \dots, n\}$. *What is the block structure of* A^k*?*

3. Let $O \leq A \lneqq B$ and let B be irreducible. Prove that $\rho(A) < \rho(B)$.

3 A brief introduction to the analysis of complex networks

Complex networks is a common name for various real networks which are usually presented by graphs with a large number of nodes: Internet graphs, collaboration graphs, e-mail graphs, social networks, transport networks, and many other. Roughly speaking, a complex network is a graph found in the real world. The term *network analysis* refers to a wealth of mathematical techniques aiming at describing the structure, function, and evolution of complex networks.

- One of the main tasks in network analysis is the localization of nodes that, in some sense, are the "most important" in a given graph. The main tool to quantify the relevance of nodes in a graph is through the computation of suitably defined *centrality indices*. Many centrality indices have been invented during time. Each one of them refers to a particular definition of "importance" or "relevance" that is most useful in a given context.
- Graphs and networks can be considered as a whole, rather than as sets of connected nodes. A *graph invariant* (or *topological index*) is a single number associated to a graph which quantifies some macroscopic feature or topological property of that graph.
- *Graph partitioning* is the problem of dividing the vertices of a graph into a given number of disjoint subsets such that the total weight of edges between such sets is minimized. The best known example of a graph partitioning problem is the problem of dividing a graph into two subsets of comparable size, such that the number of edges between them is minimized.
- *Community detection* differs from graph partitioning in that the number and size of the subsets into which the network is divided are generally not apriori specified. Moreover, the set of edges between different subsets is not necessarily "small". Instead it is assumed that the graph is intrinsically structured into communities or groups of vertices which are more or less evidently delimited, the aim being to reveal the presence and the consistency of such groups.

3.1 Main notations and definitions

Two graphs $\mathcal{G} = (V, E)$ and $\mathcal{G}' = (V, E')$ are called *isomorphic* if there exists a permutation matrix P such that $A_{\mathcal{G}'} = P A_{\mathcal{G}} P^T$. Hence, two graphs are isomorphic if and only if one of them can be obtained from the other by simply renumbering the nodes. Furthermore, if there exists a permutation matrix $P \neq I$ such that $A_{\mathcal{G}} = P A_{\mathcal{G}} P^T$ then $\mathcal{G}$ has a nontrivial *automorphism*.

Example 3.1. *The two graphs here below are isomorphic:*

Moreover, $\mathcal{G}_1$ *owns a nontrivial isomorphism with itself (that is, an* automorphism*) since the roles of nodes* 2 *and* 3 *can be interchanged.*

Let $A = A_{\mathcal{G}}$. The *in-degree* and the *out-degree* of node i are respectively the numbers

$$d_i^{\text{in}} = \sum_{j=1}^{n} A_{ij}, \qquad d_i^{\text{out}} = \sum_{j=1}^{n} A_{ji}.$$

They represent the number (or overall weight, in the weighted case) of edges that arrive to or depart from node i, respectively. If $\mathcal{G}$ is not oriented the two numbers are the same and their common value is the *degree* d_i.

Let $V = \{1, \dots, n\}$ and let Γ_n be the set of all graphs whose node set is V. A *graph invariant* (or *topological index*) is any function $f : \Gamma_n \mapsto \mathbb{R}$ which is invariant under graph isomorphisms: If $A_{\mathcal{G}'} = PA_{\mathcal{G}}P^T$ then $f(\mathcal{G}) = f(\mathcal{G}')$. Thus if two graphs are isomorphic then they have the same graph invariants.

A *centrality index* is any function $c : \Gamma_n \mapsto \mathbb{R}^n$ such that if $A_{\mathcal{G}'} = PA_{\mathcal{G}}P^T$ then $Pc(\mathcal{G}) = c(\mathcal{G}')$. Thus if two graphs are isomorphic then corresponding nodes have the same centrality indices.

The degree vectors $d^{\text{in}} = A\mathbf{1}$ and $d^{\text{out}} = A^T\mathbf{1}$ are the most simple centrality indices. Clearly, $\mathbf{1}^T d^{\text{in}} = \mathbf{1}^T d^{\text{out}}$, and the sum is equal to the total edge weight of $\mathcal{G}$, which is a graph invariant called *volume*. Many interesting graph invariants and centrality indices are based on spectral properties of $A_{\mathcal{G}}$ and variations thereof.

Remark 3.2. *Let $\mathcal{G} \in \Gamma_n$ and let $A = A_{\mathcal{G}}$. Suppose that there exists a permutation matrix $P \neq I$ such that $A = PAP^T$ (that is, $\mathcal{G}$ owns a nontrivial automorphism). Hence, if c is a centrality index computed on $\mathcal{G}$ then $c = Pc$. In particular, if $Pe_i = e_j$ then $c_i = c_j$, that is, nodes that are related by graph automorphisms get the same centrality indices. It is interesting to note that if $\mathcal{G}$ is strongly connected then this condition is fulfilled by the Perron vector of A, since $\rho(A)$ is simple: If A is irreducible, $Ax = \rho(A)x$ and $A = PAP^T$ for some permutation matrix then $Px = x$.*

Example 3.3. *A* star graph *with n nodes is the undirected graph whose adjacency matrix is*

$$A = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}.$$

Note that $\operatorname{rank}(A) = 2$ *and the trace is zero, so A has only two nonzero eigenvalues and the spectrum of A is $\{-\rho, 0, \rho\}$. Moreover, A is irreducible. Using graph automorphisms we can assume that the Perron vector of A has the form $x = (\alpha, 1, \dots, 1)^T$. In this way it is possible to reduce the problem $Ax = \rho x$ to the solution of two scalar equations in two unknowns. The solution is $x = (\rho, 1, \dots, 1)^T$ with $\rho = \sqrt{n-1}$.*

4 Spectral centralities

The purpose of this section is to describe some of the most important centrality indices, whose definition is largely based on tools and concepts borrowed from linear algebra. The common feature shared by virtually all these indices is that they are Perron eigenvectors of suitably defined nonnegative matrices.

4.1 The Bonacich index

Let $\mathcal{G}$ be a directed graph and let $A = A_{\mathcal{G}}$. In mid '90s, the american sociologist Phillip Bonacich proposed to use the Perron vector of A as a centrality index for social networks [1]. The original idea is that *a node is important if it is linked by other important nodes.* This sort of circular definition can be formalized rigorously by assuming that the centrality value of node i is proportional to the sum of centrality values of all nodes j such that $j \to i$:

$$\lambda b_i = \sum_{j:j\to i} b_j = \sum_{j=1}^{n} A_{ij} b_j.$$

Hence, the vector $b = (b_1, \dots, b_n)^T$ fulfills the eigenvalue equation $Ab = \lambda b$. Among the possible solutions of the previous equation, the Bonacich index[7] is the one which corresponds to the Perron eigenpair of A. In fact, if $\mathcal{G}$ is strongly connected then the Bonacich index obtains immediately a number of useful properties from Perron–Frobenius theory:

- It is uniquely defined, apart of a scaling factor, and its entries are positive (every node gets a nonzero score).

[7] This attribution is not standard in network analysis. In fact, the Perron vector of $A_{\mathcal{G}}$ is usually called the *eigenvector centrality* of $\mathcal{G}$ in modern literature. By the way, its use as a centrality index for social network analysis has been popularized by P. Bonacich.

- If A is primitive then it can be computed by means of the power method.
- If we add a new edge $p \to q$ to $\mathcal{G}$ then the node whose Bonacich index receives the largest relative increase is q (by Theorem 2.11), consistently to the intuition that its relevance is increasing most.

4.2 PageRank

One of the best known centrality indices for arbitrary graphs is PageRank, whose fortune started from its introduction in the Google search engine [11]. The recent account by D. Gleich [9] includes an impressive list of more than 20 PageRank-related centrality indices currently used within different domains including bibliometry, social networks, literature, biology... The original formula by S. Brin and L. Page [2] defines the PageRank vector $\pi = (\pi_1, \ldots, \pi_n)^T$ of a graph $\mathcal{G}$ as the solution of the following linear system:

$$\pi_i = (1-\alpha) + \alpha \sum_{j:j\to i} \frac{\pi_j}{d_j^{\text{out}}},$$

where $\alpha \in (0,1)$ is a fixed constant called the *damping factor*, originally set to $\alpha = 0.85$. In matrix form,

$$(I - \alpha M)\pi = (1-\alpha)\mathbf{1}, \tag{2}$$

where $M \geq O$ is the so-called *link matrix* which is defined as

$$M_{ij} = \begin{cases} A_{ij}/d_j^{\text{out}} & \text{if } d_j^{\text{out}} > 0 \\ 0 & \text{otherwise} \end{cases}$$

and $A = A_{\mathcal{G}}$. For simplicity of exposition, hereafter I assume that all nodes in $\mathcal{G}$ have at least one outgoing edge, that is, $d^{\text{out}} > 0$. In this case, the sum of all entries in any column of M is 1 (check this). Unfortunately, M is seldom irreducible. Nevertheless, we can say something about $\rho(M)$:

Lemma 4.1. $\rho(M) = 1$.

Proof. Since $M \geq O$, there exists $x \geq 0$ such that $Mx = \rho(M)x$. Moreover, we can rewrite $\sum_i M_{ij} = 1$ as $M^T\mathbf{1} = \mathbf{1}$. Hence, $\mathbf{1}^T x = \mathbf{1}^T Mx = \rho(M)\mathbf{1}^T x$, and the proof is complete, by observing that $\mathbf{1}^T x > 0$. ∎

As a consequence, we obtain that $I - \alpha M$ is nonsingular, so that π is well defined from (2). Indeed, by Lemma 4.1 all eigenvalues of $I - \alpha M$ are contained in the circle $\{z \in \mathbb{C} : |1 - z| \leq \alpha\}$, which excludes 0 since $\alpha < 1$ by hypothesis. The surprise here is that there exists a matrix $\Gamma > O$ (which is called *Google matrix*) such that π is a Perron eigenvector of Γ.

Theorem 4.2. *Let $\Gamma \in \mathbb{R}^{n\times n}$ be the positive matrix defined as*

$$\Gamma = \alpha M + \frac{1-\alpha}{n}\mathbf{1}\mathbf{1}^T.$$

for $0 < \alpha < 1$. Then, $\rho(\Gamma) = 1$ and the vector π defined in (2) is a Perron eigenvector. Moreover, if $\lambda \neq 1$ is another eigenvalue of Γ then $|\lambda| \leq \alpha$.

Proof. Observe that $\Gamma > O$ by construction. Simple computations show that $\Gamma^T\mathbf{1} = \mathbf{1}$, so that $\rho(\Gamma) = 1$ is the Perron eigenvalue. Let x be a Perron eigenvector of Γ normalized so that $\mathbf{1}^T x = n$. Then,

$$x = \Gamma x = \alpha Mx + \frac{1-\alpha}{n}\mathbf{1}\mathbf{1}^T x = \alpha Mx + (1-\alpha)\mathbf{1}.$$

Rearranging terms, $(I - \alpha M)x = (1-\alpha)\mathbf{1}$, which is (2). Finally, if $\Gamma v = \lambda v$ with $\lambda \neq 1$ then we must have $\mathbf{1}^T v = \mathbf{1}^T\Gamma v = \lambda\mathbf{1}^T v = 0$, whence either $\lambda = 0$ or $\mathbf{1}^T v = 0$. In the latter case,

$$\lambda v = \Gamma v = \alpha Mv + \frac{1-\alpha}{n}\mathbf{1}\mathbf{1}^T v = \alpha Mv.$$

Consequently, $|\lambda| \leq \rho(\alpha M) = \alpha$. ∎

4.3 Hubs and Authorities

Almost in the same year S. Brin and L. Page invented PageRank, J. Kleinberg introduced another algorithm to evaluate the relevance of documents in a large hypertext, such as the Internet [10, 11]. This algorithm (HITS, *Hypertext Induced Topic Search*) quantifies the importance of nodes in a graph according to two centrality indices: the *hub score* and the *authority score*.

Very informally, the hub score of a node is a measure of how good it is as "access point" or "portal", while the authority score is a measure of how good a node is as "informative document". Kleinberg's original idea is that *a node is a good hub if it points to good authorities; and a node is a good authority if it is pointed by good hubs.* This "mutual reinforcement" concept has been formalized by the following equations. Let h_i and a_i be the hub score and authority score of node i, respectively. Then,

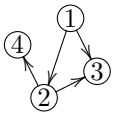
$$\lambda h_i = \sum_{j:i\to j} a_j \qquad \lambda a_i = \sum_{j:j\to i} h_j, \qquad i = 1\ldots, n, \tag{3}$$

where λ is a proportionality constant, to be defined. In matrix notations, $\lambda h = A^T a$ and $\lambda a = Ah$. The two equations can be uncoupled as follows:

$$\lambda^2 h = A^T A h, \qquad \lambda^2 a = A A^T a.$$

Let $M_{\text{hub}} = A^T A$ and $M_{\text{auth}} = AA^T$ be the hub matrix and the authority matrix, respectively. These two matrices are symmetric, nonnegative, positive semidefinite, and have exactly the same eigenvalues (why?). In particular, $\rho(M_{\text{hub}}) = \rho(M_{\text{auth}})$. The preferred solution to (3) corresponds to Perron eigenvectors, with $\lambda = \sqrt{\rho(M_{\text{hub}})}$.

Example 4.3. *Let's compute HITS scores for the following graph:*



The adjacency and hub matrix are

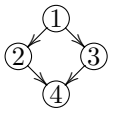
$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \qquad M_{\text{hub}} = A^T A = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The eigenvalues of M_{hub} *are* $3, 1, 0, 0$. *An eigenvector associated to* $\rho(M_{\text{hub}}) = 3$ *is* $h = (1, 1, 0, 0)^T$. *We can compute authority scores from the formula* $\lambda a = Ah$. *We obtain* $a = (0, 1, 2, 1)^T/\sqrt{3}$. *We conclude that nodes* 1 *and* 2 *are good hubs, nodes* 3 *and* 4 *are not (indeed, they have no outgoing links). The best authority node is* 3, *which is pointed by both best hubs. Node* 1 *is not an authority, because it has no ingoing links.*

Exercise 4.4. Suppose that in a given graph $\mathcal{G}$ there are two nodes, say i and j such that for every $k \in V$ it holds $k \to i \Rightarrow k \to j$. Prove that $a_i \le a_j$.

The HITS algorithm as proposed in [10] is essentially the power method with normalization applied to M_{hub} (or, equivalently, to M_{auth}) starting from the initial vector $\mathbf{1}$. Unfortunately, M_{hub} and M_{auth} are usually not irreducible, even if the original graph is strongly connected. Hence, the largest eigenvalue of these matrices may be not simple, and this fact implies that HITS scores may be not uniquely defined (apart of the scaling factor), since the convergence of the power method can be affected by the choice of the starting vector.

Example 4.5. *Consider the following graph and its hub matrix,*



$$\rightsquigarrow \qquad M_{\text{hub}} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The eigenvalues of M_{hub} *are* $2, 2, 0, 0$. *Any vector of the form* $h = (\alpha\ \beta\ \beta\ 0)^T$ *is an eigenvector corresponding to* $\rho(M_{\text{hub}}) = 2$. *If we apply the power method to* M_{hub} *starting from* $(1\ 1\ 1\ 1)^T$ *we obtain* $h \propto (1\ 1\ 1\ 0)^T$, *while if the starting vector is* $(1\ 0\ 0\ 0)^T$ *then we obtain* $h \propto (1\ 0\ 0\ 0)^T$.

Various modifications of the basic HITS algorithm have been devised in order to make hub-authority scores well defined under rather general hypotheses, see e.g., [6]. One of these tricks is described in the following exercise:

Exercise 4.6. Let $\widehat{A} = A + \varepsilon I$ where $\varepsilon > 0$ (note: this modification corresponds to adding a loop with weight ε to every node in the graph) and let $\widehat{M}_{\text{hub}} = \widehat{A}^T \widehat{A}$. Prove that if $\mathcal{G}$ is not disconnected[8] then $\widehat{M}_{\text{hub}}$ is irreducible.

Hint: $\widehat{M}_{\text{hub}} = \varepsilon(A + A^T) +$ *other nonnegative matrices.*

4.4 More exercises and problems

1. [The Q-indices] Let $\mathcal{G} = (V, E)$ be an undirected, connected graph. Consider the following centrality indices (called Q-indices) for both nodes and edges: To any $i \in V$ and $(i, j) \in E$ associate nonnegative variables q_i and e_{ij} respectively, by means of these equations:

$$\lambda q_i = \sum_{j : i \sim j} e_{ij}, \qquad e_{ij} = q_i + q_j,$$

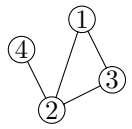
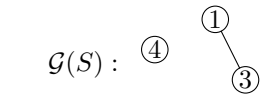
where λ is a constant to be determined. According to these equations, the importance of an edge depends on the importance of its nodes, and the importance of a node is proportional to the sum of importance of its edges. Discuss existence, uniqueness, and positivity of the solution to the previous equations.

Hint: Rearranging equations, the vector $q = (q_1, \ldots, q_n)^T$ *can be written as the Perron vector of a nonnegative matrix.*

5 Nodal domains

In what follows, the graph $\mathcal{G} = (V, E)$ is assumed to be undirected (that is, $A_{\mathcal{G}}$ is symmetric). Moreover, the *subgraph induced by* $S \subseteq V$ is the graph $\mathcal{G}(S)$ whose adjacency matrix is $[A]_{i,j \in S}$.

Let $0 \neq v \in \mathbb{R}^n$ and consider the set $S = \{i : v_i \geq 0\}$. The subgraph $\mathcal{G}(S)$ may result in a collection of subgraphs which are disconnected one from the other. These components are called *nodal domains* of v. For example, for the following graph $\mathcal{G}$ and vector v,

$\mathcal{G}$:  $\quad v = \begin{pmatrix} 1 \\ -2 \\ 0 \\ 0.1 \end{pmatrix} \quad \rightsquigarrow \quad \mathcal{G}(S)$: 

the set $S = \{i : v_i \geq 0\}$ induces the nodal domains $\mathcal{G}(\{1, 3\})$ and $\mathcal{G}(\{4\})$.

Nodal domains of eigenvectors of various graph-related matrices (not only adjacency but also Laplacian and modularity matrices) are useful tools to provide approximate solutions of graph partitioning and community detection problems. In fact, various problems that require to partition a complex network into cohesive regions, or to locate "clusters" or "communities", require the solution of combinatorial optimization problems whose computational cost grows exponentially fast in n. By means of relaxation techniques, solutions of these problems can be approximated efficiently by nodal domains of eigenvectors of suitably defined matrices, see e.g., [13] or [14] for a survey.

Let $A = A_{\mathcal{G}}$. A Perron vector v has positive entries, so that v has only one nodal domain which is $\mathcal{G}$ itself. Obviously, we cannot say the same for other eigenvectors (why?). The goal of this section is to show an interesting result by Fiedler [8] concerning the nodal domains of eigenvectors associated to non-dominant eigenvalues of A. Before going further, I recall a basic fact in matrix theory, see e.g., [15, §5.7]:

Lemma 5.1. *Let $M \in \mathbb{R}^{p \times p}$ be a symmetric matrix, and let $N \in \mathbb{R}^{q \times q}$ be one of its principal submatrices. Let $\lambda_1(M) \geq \lambda_2(M) \geq \ldots \geq \lambda_p(M)$ and $\lambda_1(N) \geq \lambda_2(N) \geq \ldots \geq \lambda_q(N)$ denote the eigenvalues of M and N counted with their multiplicity, respectively. Then, $\lambda_i(M) \geq \lambda_i(N)$ for $i = 1, \ldots, q$.*

[8] A graph is *disconnected* if its vertex set can be partitioned into two subsets, $V = V_1 \cup V_2$ and $V_1 \cap V_2 = \varnothing$, so that no edge belongs to $(V_1 \times V_2) \cup (V_2 \times V_1)$.

Theorem 5.2. *Let $A \geq O$ be irreducible and symmetric. Let $\rho(A) = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n$ be its eigenvalues, let v be an eigenvector associated to λ_2, and let $S = \{i \in V : v \geq 0\}$. Then $\mathcal{G}(S)$ is connected.*

Proof. Proceed by contradiction. Assume that $S = S_1 \cup S_2$ with $S_1 \cap S_2 = \varnothing$, both $\mathcal{G}(S_1)$ and $\mathcal{G}(S_2)$ are connected but there is no edge joining V_1 with V_2. By a suitable permutation of rows and columns, we can assume that $v = (v_1, v_2, v_3)^T$ where $v_1 \geq 0$ and $v_2 \geq 0$ are the entries with indices in S_1 and S_2, respectively, and $v_3 < 0$ are the entries with indices in $\bar{S}$. Accordingly, the structure of A is

$$A = \begin{pmatrix} A_{11} & O & A_{13} \\ O & A_{22} & A_{23} \\ * & * & * \end{pmatrix}$$

where A_{11} and A_{22} are irreducible and both A_{13} and A_{23} are nonzero (because A is irreducible). Then, the equation $Av = \lambda_2 v$ yields

$$\begin{aligned} A_{11}v_1 + A_{13}v_3 &= \lambda_2 v_1 \\ A_{22}v_2 + A_{23}v_3 &= \lambda_2 v_2. \end{aligned}$$

Let y_1 and y_2 be Perron eigenvectors of A_{11} and A_{22}, respectively: $A_{ii}y_i = \rho(A_{ii})y_i$. Remind that $A_{ii} = A_{ii}^T$. Then,

$$\underbrace{y_i^T A_{ii} v_i}_{=\rho(A_{ii})y_i^T v_i} + \underbrace{y_i^T A_{i3} v_3}_{<0} = \lambda_2 y_i^T v_i, \qquad i = 1, 2.$$

Since $y_i^T v_i > 0$ we get $\rho(A_{ii}) > \lambda_2$ for $i = 1, 2$. Hence, the submatrix $\left(\begin{smallmatrix} A_{11} & O \\ O & A_{22} \end{smallmatrix}\right)$ has at least 2 eigenvalues that are $> \lambda_2$, thus contradicting Lemma 5.1. ∎

Remarks:

- By applying Theorem 5.2 to $-v$ in place of v, you can deduce easily that also the set $\{i : v_i \leq 0\}$ induces a connected subgraph.
- The argument of the proof of Theorem 5.2 can be extended naturally to eigenvalues λ_i with $i \geq 2$. The result is that, if $Av = \lambda_i v$ and $S = \{i : v_i \geq 0\}$ then $\mathcal{G}(S)$ is composed by no more than $i - 1$ connected components, see e.g., [8].

5.1 Applications: Spectral graph bisection

A graph partitioning problem requires to partition the nodes of a given graph $\mathcal{G} = (V, E)$ into pairwise disjoint sets (also called clusters) so that the number of edges running across different sets is minimized, in some sense. Hereafter, the following notations will be used in correspondence with an arbitrary set $S \subseteq V$:

- Denote by $|S|$ its cardinality (that is, the number of its elements), by $\bar{S}$ its complement (that is, $\bar{S} = V \setminus S$) and by $\mathbf{1}_S$ its characteristic vector, that is $(\mathbf{1}_S)_i = 1$ if $i \in S$ and 0 otherwise.
- Let $\operatorname{vol} S = \sum_{i \in S} d_i$ be the *volume* of S (recall that d_i is the degree of node i). Note: $\operatorname{vol} S = d^T \mathbf{1}_S$ and $d = A\mathbf{1}$.
- Let $e_{\text{in}}(S) = \mathbf{1}_S^T A \mathbf{1}_S$ and $e_{\text{out}}(S) = \mathbf{1}_S^T A \mathbf{1}_{\bar{S}} = \operatorname{vol} S - e_{\text{in}}(S)$. Note: $e_{\text{out}}(S)$ is the number of edges joining S with $\bar{S}$ while $e_{\text{in}}(S)$ is twice the number of edges whose endpoints are both in S (but loops are counted only once).

Let's consider the special graph partitioning problem where we want to split V into two subsets S and $\bar{S}$, with $S \cup \bar{S} = V$ and $S \cap \bar{S} = \varnothing$. The pair $\{S, \bar{S}\}$ is a *cut* in $\mathcal{G}$. For any $S \subseteq V$ consider the number

$$H(S) = e_{\text{out}}(S)/|S|,$$

which is sometimes called the *conductance of* S. A set with high conductance has a relatively large amount of edges connecting it to its complement, with respect to the number of nodes. Conversely, a set having low conductance is a set that can be easily separated from the rest of the graph, by removing a quite small number of edges. In the framework of graph partitoning preblems, a useful

merit function of the graph cut $\{S, \bar{S}\}$ (which is easily generalized to more than two sets) is the following:

$$h(S, \bar{S}) = H(S) + H(\bar{S}) = \ldots = \frac{n}{|S||\bar{S}|} e_{\text{out}}(S),$$

since $e_{\text{out}}(S) = e_{\text{out}}(\bar{S})$. One of the main graph partitioning problems consists in computing

$$h_{\mathcal{G}} = \min_{S \subseteq V} h(S, \bar{S}) \tag{4}$$

which is an important graph invariant. Indeed, a set S attaining the minimum splits the graphs into two parts that are comparable in size and are connected by relatively few edges. The task of finding that set S is very hard (it's an NP-complete problem). To help its solution, there exists an heuristic technique known as *spectral bisection*, which is based on nodal domains and often goes very close to the exact solution.

5.1.1 The Laplacian matrix

Let $D = \text{Diag}(d_1, \ldots, d_n)$. The matrix $L = D - A$ is called *Laplacian matrix of* $\mathcal{G}$. This is one of the most useful matrices associated to a graph. The study of its spectral properties and applications has been pioneered by M. Fiedler, see e.g., [7]. For every $v \in \mathbb{R}^n$ we have

$$v^T L v = \sum_{i \sim j} (v_i - v_j)^2.$$

Thus, L is positive semidefinite; the vector $\mathbf{1}$ is in the kernel of L, that is $L\mathbf{1} = 0$; and the dimension of $\ker(L)$ is 1 if and only if $\mathcal{G}$ is connected.[9] For any given $S \subseteq V$ we have

$$\mathbf{1}_S^T L \mathbf{1}_S = \mathbf{1}_S^T D \mathbf{1}_S - \mathbf{1}_S^T A \mathbf{1}_S = \text{vol}\, S - e_{\text{in}}(S) = e_{\text{out}}(S).$$

Define $v \in \mathbb{R}^n$ as $v = \mathbf{1}_S - (|S|/n)\mathbf{1}$, that is

$$v_i = \begin{cases} |\bar{S}|/n & i \in S \\ -|S|/n & i \notin S. \end{cases}$$

You can easily verify the following identitites:

$$\mathbf{1}^T v = 0, \qquad v^T v = \frac{|S||\bar{S}|}{n}, \qquad v^T L v = e_{\text{out}}(S), \qquad h(S, \bar{S}) = \frac{v^T L v}{v^T v}. \tag{5}$$

We obtain a nontrivial lower bound for the number $h_{\mathcal{G}}$ defined in (4):

Theorem 5.3. *Let $\mathcal{G}$ be connected, and let $0 = \lambda_1 < \lambda_2 \leq \ldots \lambda_n$ be the eigenvalues of L. Then, $\lambda_2 \leq h_{\mathcal{G}}$.*

Proof. Owing to the variational characterization of the eigenvalues of a symmetric matrix (see e.g., [15, §5.6]) we have exactly

$$\lambda_2 = \min_{v : \mathbf{1}^T v = 0} \frac{v^T L v}{v^T v}.$$

Moreover, by (5), $\lambda_2 \leq h(S, \bar{S})$ for all S. ∎

Hence, the eigenvalue λ_2, which is named the *algebraic connectivity of* $\mathcal{G}$ after [7], tells us how easy is to split the graph into two (roughly balanced) pieces. Indeed, if $\lambda_2 \approx 0$ then $\mathcal{G}$ can be easily disconnected by removing a few edges (in particular, if $\lambda_2 = 0$ then $\mathcal{G}$ is already disconnected) while if λ_2 is large then also $h_{\mathcal{G}}$ must be large.

[9] More precisely, the dimension of $\ker(L)$ is equal to the number of connected components of $\mathcal{G}$.

5.1.2 Spectral bisection via the Fiedler vector

The nodal domains of an eigenvector associated to λ_2 often provide good approximations to the cut $\{S, \bar{S}\}$ which minimizes $h(S, \bar{S})$. Their connectedness is considered in the following result:

Theorem 5.4. *Let $\mathcal{G}$ be a connected, undirected graph. Let $0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_n$ be the eigenvalues of the Laplacian matrix L, let f be an eigenvector associated to λ_2 and let $S = \{i : f_i \geq 0\}$. Then $\mathcal{G}(S)$ is connected.*

Proof. By choosing a sufficiently large positive constant α, the matrix $M = \alpha I - L = \alpha I - D + A$ is nonnegative and irreducible. Moreover, any eigenvector of M is also an eigenvector of L, and conversely. Indeed, $Mv = \mu v \iff Lv = (\alpha - \mu)v$. Hence, the eigenvalues of M are the numbers $\alpha > \alpha - \lambda_2 \geq \ldots \geq \lambda_n$. The claim follows immediately from Theorem 5.2. ∎

The vector f appearing in the preceding theorem is usually referred to as a *Fiedler vector*. By replacing f with $-f$, it is also possible to prove that the subgraph induced by $\{i : f_i \leq 0\}$ is connected.

References

[1] P. Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2 (1972), 113–120.

[2] S. Brin, L. Page. *The anatomy of a large-scale hypertextual web search engine.* Proceedings of the 7th international conference on World Wide Web (WWW). Brisbane, Australia (1998), 107–117.

[3] D. Chackrabarti, Y. Wang, C. Wang, J. Leskovec, C. Faloutsos. Epidemic Thresholds in Real Networks. *ACM Transactions on Information and System Security*, 10 (2008), n. 13.

[4] D. Cvetković, S. Simić. Graph spectra in Computer Science. *Lin. Algebra Appl.*, 434 (2011), 1545–1562.

[5] E. Dietzenbacher. Perturbations of matrices: a theorem on the Perron vector and its applications to input-output models. *J. of Economics*, 48 (1988), 389–412.

[6] A. Farahat, T. Lofaro, J. Miller, G. C. Rae, L. A. Ward. Authority rankings from HITS, PageRank, and SALSA: existence, uniqueness, and effect of initialization. *SIAM J. Sci. Comput.* 27 (2006), 1181–1201.

[7] M. Fiedler. Algebraic connectivity of graphs. *Czech. Math. J.*, 23 (1973), 298–305.

[8] M. Fiedler. A property of nonnegative symetric matrices and its application to graph theory. *Czech. Math. J.*, 25 (1975), 619–633.

[9] D. F. Gleich. PageRank beyond the web. *SIAM Rev.*, 57 (2015), 321–363.

[10] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46 (1999), 604–632.

[11] A. N. Langville, C. D. Meyer. A survey of eigenvector methods for Web information retrieval. *SIAM Rev.*, 47(2005), 135–161.

[12] C. Meyer. *Matrix Analysis and Applied Linear Algebra.* SIAM, 2000.

[13] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74 (2006) 036104.

[14] M. Nascimento, A. de Carvalho. Spectral methods for graph clustering — A survey. *European Journal of Operational Research*, 211 (2011), 221–231.

[15] E. E. Tyrtyshnikov. *A Brief Introduction to Numerical Analysis.* Birkhäuser, 1997.