Vrije Universiteit Brussel

Faculty of Science
Master in Mathematics - specialization: Education

Similarity on Combinatorial Structures

Graduation thesis submitted in partial fulfillment of the requirements for
the degree of Master of Science in Mathematics - specialization: Education

Filip Moons

Promotor: Prof. Dr. P. Cara
Stagebegeleider: Promotor

Stagementor:

June 2015

Contents

Chapter 1

Preliminaries and notations

The Perron-Frobenius theorem states that a real square matrix with nonnegative entries has a unique largest real eigenvalue with an eigenvector that has only positive entries. The theorem was proved by Oskar Perron (1880-1975) in 1907 for strictly positive entries and extended by Ferdinand Georg Frobenius (1849-1917) to irreducible matrices with nonnegative entries.

1.1 Some families of matrices

In this section, we first introduce different kinds of matrices. Note that all matrices in this master thesis have real entries, unless otherwise stated. We start with permutation matrices and their uses. With permutation matrices, we can introduce irreducible matrices. Also nonnegative and primitive square matrices are presented. After defining those, we look at the Perron-Frobenius theorem.

1.1.1 Permutation matrices

Definition 1.1.1. *Given a permutation π of n elements:*

$$\pi : \{1, \dots, n\} \to \{1, \dots, n\},$$

with:

$$\pi = \begin{pmatrix} 1 & 2 & \cdots & n \\ \pi(1) & \pi(2) & \cdots & \pi(n) \end{pmatrix}$$

*the associated **permutation matrix** P_π is the $n \times n$-matrix obtained by permuting the rows of the identity matrix I_n according to π. So:*

$$P_\pi = \begin{bmatrix} \mathbf{e}_{\pi(1)} \\ \mathbf{e}_{\pi(2)} \\ \vdots \\ \mathbf{e}_{\pi(n)} \end{bmatrix}.$$

where $\mathbf{e}_j$ is the j-th row of I_n.

Example 1.1.2. The permutation matrix P_π corresponding to the permutation

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 3 \end{pmatrix}$$

is:

$$P_\pi = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Note that $p_{ij} = 1$ if and only if $\pi(i) = j$.

Property 1.1.3. *A permutation matrix P satisfies:*

$$PP^T = I_n,$$

where P^T is the transpose and I_n is the identity matrix.

Proof. By direct computation, we get:

$$(PP^T)_{ij} = \sum_{k=1}^{n} P_{ik} P_{kj}^T = \sum_{k=1}^{n} P_{ik} P_{jk}$$

Assume $i \neq j$. Then for each k, $P_{ik}P_{jk} = 0$ since there is only one nonzero entry in the k-th row and $i \neq j$, P_{ik} and P_{jk} can't be both the nonzero entry. So, $(PP^T)_{ij} = 0$ when $i \neq j$.

When $i = j$, then there exists a $k' \in \{1, \ldots, n\}$ with $P_{ik'}P_{jk'} = 1$, since there is only one nonzero entry in the k-th row, this k' is unique, which results in $\sum_{k=1}^{n} P_{ik}P_{jk} = (PP^T)_{ij} = 1$.

In other words,

$$(PP^T)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases},$$

this is exactly the formula for the entries of the identity matrix. □

Corollary 1.1.4. *The transpose of a permutation matrix P is its inverse:*

$$P^T = P^{-1}.$$

This can also more easily be concluded by the fact that a permutation matrix is clearly an orthogonal matrix (a real $n \times n$-matrix with orthonormal entries).

1.1.2 Nonnegative and primitive matrices

Definition 1.1.5. *Let A and B be two real $n \times r$-matrices. Then, $A \geq B$ (respectively $A > B$) if $a_{ij} \geq b_{ij}$ (respectively $a_{ij} > b_{ij}$) for all $1 \leq i \leq n, 1 \leq j \leq r$.*

Definition 1.1.6. *A real $n \times r$-matrix A is **nonnegative** if $A \geq 0$, with 0 the $n \times r$-null matrix.*

Definition 1.1.7. *A real $n \times r$-matrix A is **positive** if $A > 0$, with 0 the $n \times r$-null matrix.*

Since row vectors are $1 \times n$-matrices, we shall use the terms nonnegative and positive vector throughout.

Notation 1.1.8. *Let B be an arbitrary complex $n \times r$-matrix, then $|B|$ denotes the matrix with entries $|b_{ij}|$. This is not to be confused with the determinant of a square matrix B, which we denote by* $\det(B)$.

Definition 1.1.9. *A nonnegative square matrix A is called **primitive** if there is a $k \in \mathbb{N}_0$ such that all entries of A^k are positive.*

1.1.3 Irreducible nonnegative matrices

In developing the Perron-Frobenius theory, we shall first establish a series of theorems and lemmas on nonnegative irredicuble square matrices.

Definition 1.1.10. *A square matrix A is called **reducible** if there is a permutation matrix P such that*

$$PAP^T = \begin{pmatrix} B & 0 \\ C & D \end{pmatrix}$$

*where B, and D are square matrices, each of size at least one and 0 is a zero matrix. A square matrix A is called **irreducible** if it is not reducible.*

It follows immediately that a 1×1-matrix is always irreducible by definition. We now show a useful property to identify a reducible matrix.

Property 1.1.11. *Let A be an $n \times n$-matrix with $n \geq 2$. Consider a nonempty, proper subset S of $\{1, \ldots, n\}$ with $a_{ij} = 0$ for $i \in S$, $j \notin S$. Then A is reducible.*

Proof. Let $S = \{i_1, i_2, \ldots, i_k\}$, where we assume, without loss of generality, that $i_1 < i_2 < \cdots < i_{k-1} < i_k$. Let S^c be the complement of S, consisting of the ordered set of elements $j_1 < j_3 < \cdots < j_{n-k}$. Consider the permutation σ of $\{1, 2, \ldots, n\}$ given by

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & k & k+1 & k+2 & \ldots & n \\ i_1 & i_2 & \ldots & i_k & j_1 & j_2 & \cdots & j_{n-k} \end{pmatrix}$$

σ can be represented by the permutation matrix $P_\sigma = (p_{ij})$, where $p_{rs} = 1$ if $\sigma(r) = s$. We prove that

$$PAP^T = \begin{pmatrix} B & 0 \\ C & D \end{pmatrix}$$

where B and D are square matrices and 0 is a $k \times (n-k)$ zero matrix. Consider row c and column d, where $1 \leq c \leq k$ and $k+1 \leq d \leq n$:

$$(PAP^T)_{cd} = \sum_i \sum_j p_{ci} a_{ij} p_{dj}. \tag{1.1}$$

It is enough to show that each term in the summation is zero. Suppose $p_{ci} = p_{dj} = 1$. Thus $\sigma(c) = i$ and $\sigma(d) = j$. Since $1 \leq c \leq k$, then $i \in \{i_1, i_2, \ldots, i_k\}$; similarly, since $k+1 \leq d \leq n$, we have $j \in \{j_1, j_2, \ldots, j_{n-k}\}$. By assumption, for such a pair i, j, we have $a_{ij} = 0$. That completes the proof. □

We know prove some equivalent definitions for a nonnegative, irreducible square matrix.

Theorem 1.1.12. *Let $A \geq 0$ be a nonnegative $n \times n$-matrix. Then the following conditions are equivalent:*

(1) A is irreducible.

(2) $(I + A)^{n-1} > 0$

(3) For any pair (i, j), with $1 \leq i, j \leq n$, there is a positive integer $k = k(i, j) \leq n$ such that $(A^k)_{ij} > 0$.

Proof. (1) $\Rightarrow$ (2): Let $\mathbf{x} \geq 0, \mathbf{x} \neq \mathbf{0}$ be an arbitrary vector in $\mathbb{R}^n$. If a coordinate of $\mathbf{x}$ is positive, the same coordinate is positive in $\mathbf{x} + A\mathbf{x} = (I + A)\mathbf{x}$ as well. We claim that $(I + A)\mathbf{x}$ has fewer zero coordinates than $\mathbf{x}$ as long as $\mathbf{x}$ has a zero coordinate. If this claim is not true, then the number of zero coordinates must be at least equal, this means that for each coordinate j with $x_j = 0$ we would have that $x_j + (A\mathbf{x})_j = 0$. Let $J = \{j : x_j > 0\}$. For any $j \notin J, r \in J$, we have $(A\mathbf{x})_j = \sum_k a_{jk}x_k = 0$ and $x_r > 0$. It must be that $a_{jr} = 0$. It follows from Property 1.1.11 that A is reducible, which is a contradiction and the claim is proved. Thus $(I + A)\mathbf{x}$ has at most $n - 2$ zero coordinates. Continuing in this manner we conclude that $(I + A)^{n-1}\mathbf{x} > 0$. Let $\mathbf{x} = \mathbf{e}_i$, then the corresponding column of $(I + A)^{n-1}$ must be positive. Thus (2) holds.

(2) $\Rightarrow$ (3): We have $(I + A)^{n-1} > 0$, $A \geq 0$, so $A \neq 0$ and

$$A(I + A)^{n-1} = \sum_{k=1}^{n} \binom{n-1}{k-1} A^k > 0.$$

Thus for any i, j at least one of the matrices $A, A^2, \ldots, A^n$ has its (i, j)-th element entry positive.

(3) $\Rightarrow$ (1): Suppose A is reducible. Then for some permutation matrix P,

$$PAP^T = \begin{pmatrix} B_1 & 0 \\ C_1 & D_1 \end{pmatrix}$$

where B_1 and D_1 are square matrices. Furthermore, we know from Property 1.1.3 that $PAP^TPAP^T = PA^2P^T$, whence for some square matrices B_2, C_2 we have:

$$PA^2P^T = \begin{pmatrix} B_2 & 0 \\ C_2 & D_2 \end{pmatrix}$$

More generally, for some matrix C_t and square matrices B_t and D_t,

$$PA^tP^T = \begin{pmatrix} B_t & 0 \\ C_t & D_t \end{pmatrix}$$

Thus $(PA^tP^T)_{rs} = 0$ for $t = 1, 2, \ldots$ and for any r, s corresponding to an entry of the zero submatrix in PAP^T. Now, for $t = 1, \ldots, n$:

$$0 = (PA^tP^T)_{rs} = \sum_k \sum_l p_{rk} a_{kl}^{(t)} p_{st}$$

By using the same reasoning as in 1.1, choose k, l so that $p_{rk} = p_{sl} = 1$. Then $a_{kl}^{(t)} = 0$ for all t, contradicting the hypothesis. This completes the proof. □

Corollary 1.1.13. *If A is irreducible then $I + A$ is primitive.*

Corollary 1.1.14. *A^T is irreducible whenever A is irreducible.*

Property 1.1.15. *No row or column of an irreducible matrix A can vanish. This means that A cannot have a row or a column of zeros.*

Proof. Suppose that A has a zero row, then it could be permuted to

$$PAP^T = \begin{pmatrix} 0 & 0\dots 0 \\ \mathrm{c}_1 & \\ \vdots & D \\ \mathrm{c}_n & \end{pmatrix}$$

by some permutation matrix P. It follows from Definition 1.1.10 that A is reducible. Similarly, if A has zero column, it can be permuted to

$$PAP^T = \begin{pmatrix} & 0 \\ B & \vdots \\ & 0 \\ c_1 \dots c_n & 0 \end{pmatrix}$$

again from Definition 1.1.10 we conclude that A is reducible.

□

1.2 Perron-Frobenius Theorem

1.2.1 Spectral radii of nonnegative matrices

Definition 1.2.1. *Let A be an $n \times n$-matrix with complex entries and eigenvalues λ_i, $1 \leq i \leq n$. Then:*

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|$$

*is called the **spectral radius** of the matrix A.*

Geometrically, if all the eigenvalues λ_i of A are plotted in the complex plane, then $\rho(A)$ is the radius of the smallest disk $|z| \leq R$, with center at the origin, which includes all the eigenvalues of the matrix A.

We now establish a series of lemmas on nonnegative irreducible square matrices. These lemmas will allow us to prove the Perron-Frobenius at the end of this section.

If $A \geq 0$ is an irreducible $n \times n$-matrix and $\mathbf{x}$, a vector of size n with $\mathbf{0} \neq \mathbf{x} \geq 0$, let

$$r_{\mathbf{x}} = \min \left\{ \frac{\sum_{j=1}^{n} a_{ij} x_j}{x_i} \right\} \tag{1.2}$$

where the minimum is taken over all i for which $x_i > 0$. Clearly, $r_{\mathbf{x}}$ is a nonnegative real number and is the supremum of all $p \geq 0$ for which

$$A\mathbf{x} \geq p\mathbf{x} \tag{1.3}$$

We now consider the nonnegative quantity r defined by

$$r = \sup_{\substack{\mathbf{x} \geq 0 \\ \mathbf{x} \neq \mathbf{0}}} \{r_{\mathbf{x}}\} \tag{1.4}$$

As $r_{\mathbf{x}}$ and $r_{\alpha\mathbf{x}}$ have the same value for any scalar $\alpha > 0$, we need consider only the set B of vectors $\mathbf{x} \geq 0$ with $||\mathbf{x}|| = 1$, and we correspondingly let Q be the set of all vectors $\mathbf{y} = (I + A)^{n-1}\mathbf{x}$ where $\mathbf{x} \in B$. From Theorem 1.1.12 , Q consists only of positive vectors. Multiplying both sides of the inequality $A\mathbf{x} \geq r_{\mathbf{x}}\mathbf{x}$ by $(I + A)^{n-1}$, we obtain:

$$\forall \mathbf{y} \in Q : A\mathbf{y} \geq r_{\mathbf{x}}\mathbf{y},$$

and we conclude from (1.3) that $r_{\mathbf{y}} \geq r_{\mathbf{x}}$. Therefore, the quantity r of (1.4) can be defined equivalently as:

$$r = \sup_{\mathbf{y} \in Q}\{r_{\mathbf{y}}\} \tag{1.5}$$

As B is a compact set (in the usual topology) of vectors, so is Q, and as $r_{\mathbf{y}}$ is a continuous function on Q, we know from the extreme value theorem that there necessarily exists a positive vector $\mathbf{z}$ for which:

$$A\mathbf{z} \geq r\mathbf{z}, \tag{1.6}$$

and no vector $\mathbf{w} \geq 0$ exists for which $A\mathbf{w} > r\mathbf{w}$.

Definition 1.2.2. *We call all nonnegative, nonzero vectors* $\mathbf{z}$ *satisfying (1.6)* ***extremal vectors*** *of the matrix* A.

Lemma 1.2.3. *If* $A \geq 0$ *is an irreducible* $n \times n$*-matrix, the quantity* r *of (1.4) is positive.*

Proof. If $\mathbf{x}$ is the positive vector whose coordinates are all unity, then since the matrix A is irreducible, we know from Property 1.1.15 that no row of A can vanish, and consequently no component of $A\mathbf{x}$ can vanish. Thus, $r_{\mathbf{x}} > 0$, proving that $r > 0$. □

Lemma 1.2.4. *If* $A \geq 0$ *is an irreducible* $n \times n$*-matrix, each extremal vector* $\mathbf{z}$ *is a positive eigenvector of* A *with corresponding eigenvalue* r *of (1.4), i.e.,* $A\mathbf{z} = r\mathbf{z}$ *and* $\mathbf{z} > 0$.

Proof. Let $\mathbf{z}$ be an extremal vector with $A\mathbf{z} - r\mathbf{z} = \mathbf{t}$. If $\mathbf{t} \neq \mathbf{0}$, then some coordinate of $\mathbf{t}$ is positive; multiplying through by the matrix $(I + A)^{n-1}$, we have:

$$A\mathbf{w} - r\mathbf{w} > 0, \text{ with } \mathbf{w} = (I + A)^{n-1}\mathbf{z}$$

from Theorem 1.1.12 we know that $\mathbf{w} > 0$. It would then follow that $r_{\mathbf{w}} > r$, contradicting the definition of r in (1.5). Thus $A\mathbf{z} = r\mathbf{z}$, and since $\mathbf{w} > 0$ and $\mathbf{w} = (1 + r)^{n-1}\mathbf{z}$, then we have $\mathbf{z} > 0$, completing the proof. □

Lemma 1.2.5. *Let* $A \geq 0$ *be an irreducible* $n \times n$*-matrix, and let* B *be an* $n \times n$*- complex matrix with* $|B| \leq A$. *If* β *is any eigenvalue of* B, *then*

$$|\beta| \leq r, \tag{1.7}$$

where r *is the positive quantity of (1.4). Moreover, equality is valid in (1.7), i.e.,* $\beta = re^{i\phi}$, *if and only if* $|B| = A$, *and where* B *has the form:*

$$B = e^{i\phi}DAD^{-1}, \tag{1.8}$$

and D *is a diagonal matrix whose diagonal entries have modulus unity.*

Proof. If $\beta\mathbf{y} = B\mathbf{y}$ where $\mathbf{y} \neq \mathbf{0}$, then

$$\beta y_i = \sum_{j=1}^{n} b_{ij} y_i, \text{ with } 1 \leq i \leq n.$$

Using the hypotheses of the lemma and the notation of Definition 1.1.8, it follows that:

$$|\beta||\mathbf{y}| \leq |B||\mathbf{y}| \leq A|\mathbf{y}|, \tag{1.9}$$

which implies that $|\beta| \leq r_{|\mathbf{y}|} \leq r$, proving (1.7). If $|\beta| = r$, then $|\mathbf{y}|$ is an extremal vector of A. Therefore, from Lemma 1.2.4, $|\mathbf{y}|$ is a positive eigenvector of A corresponding to the positive eigenvalue r. Thus,

$$r|\mathbf{y}| = |B||\mathbf{y}| = A|\mathbf{y}|, \tag{1.10}$$

and since $|\mathbf{y}| > 0$, we conclude from (1.10) and the hypothesis $|B| \leq A$ that

$$|B| = A \tag{1.11}$$

For the vector $\mathbf{y}$, where $|\mathbf{y}| > 0$, let

$$D = \operatorname{diag}\left\{\frac{y_1}{|y_1|}, \ldots, \frac{y_n}{|y_n|}\right\}.$$

It is clear that the diagonal entries of D have modulus unity, and

$$\mathbf{y} = D|\mathbf{y}|. \tag{1.12}$$

Setting $\beta = re^{i\phi}$, then $B\mathbf{y} = \beta\mathbf{y}$ can be written as:

$$C|\mathbf{y}| = r|\mathbf{y}|, \tag{1.13}$$

where

$$C = e^{-i\phi} D^{-1} B D. \tag{1.14}$$

From (1.10) and (1.13), equiting terms equal to $r|\mathbf{y}|$ we have

$$C|\mathbf{y}| = |B||\mathbf{y}| = A|\mathbf{y}|. \tag{1.15}$$

From the definition of the matrix C in (1.14), $|C| = |B|$. Combining with (1.11), we have:

$$|C| = |B| = A. \tag{1.16}$$

Thus, from (1.15) we conclude that $C|\mathbf{y}| = |C||\mathbf{y}|$, and as $|\mathbf{y}| > 0$, it follows that $C = |C|$ and thus $C = A$ from (1.16). Combining this result with (1.14), gives the desired result that $B = e^{i\phi} D A D^{-1}$. Conversely, it is obvious that if B has the form in (1.8), then $|B| = A$, and B has an eigenvalue β with $|\beta| = r$, which completes the proof. □

Corollary 1.2.6. *If $A \geq 0$ is an irreducible $n \times n$-matrix, then the positive eigenvalue r of Lemma 1.2.4 equals the spectral radius $\rho(A)$ of A*

Proof. Setting $B = A$ in Lemma 1.2.5 immediately gives us this result. □

In other words, if $A \geq 0$ is an irreducible $n \times n$-matrix, its spectral radius $\rho(A)$ is positive, and the intersection in the complex plane of the circle $|z| = \rho(A)$ with the positive real axis is an eigenvalue of A.

Definition 1.2.7. *A **principal square submatrix** of an $n \times n$-matrix A is any matrix obtained by crossing out any j rows and the corresponding j columns of A, with $1 \leq j \leq n$.*

Lemma 1.2.8. *If $A \geq 0$ is an irreducible $n \times n$-matrix, and B is any principal square submatrix of A, then $\rho(B) < \rho(A)$.*

Proof. If B is any principal submatrix of A, then there is an $n \times n$-permutation matrix P such that $B = A_{11}$ where

$$C = \begin{pmatrix} A_{11} & 0 \\ 0 & 0 \end{pmatrix}; PAP^T = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{1.17}$$

Here, A_{11} and A_{22} are, respectively, $m \times m$ and $(n-m) \times (n-m)$ principal square submatrices of PAP^T, $1 \leq m \leq n$. Clearly, $0 \leq C \leq PAP^T$, and $\rho(C) = \rho(B) = \rho(A_{11})$, but as $C = |C| \neq PAP^T$, the conclusion follows immediately from Lemma 1.2.5 and Corollary 1.2.6. □

The following lemma is used to prove that $\rho(A)$ is a simple eigenvalue of A in the Perron-Frobenius theorem. The proof uses the extension of the product rule of derivation for multilinear functions $M(a_1, \ldots, a_k)$. Suppose $x_1, .., x_k$ are differentiable vector functions, then $M(x_1, \ldots, x_k)$ is differentiable and:

$$\frac{\mathrm{d}}{\mathrm{d}t} M(x_{1,k}) = M(\frac{\mathrm{d}}{\mathrm{d}t} x_1, x_2, \ldots, x_k) + M(x_1, \frac{\mathrm{d}}{\mathrm{d}t} x_2, \ldots, x_k) + \ldots + M(x_1, x_2, \ldots, \frac{\mathrm{d}}{\mathrm{d}t} x_k)$$

The most important application of this rule is for the derivative of the determinant:

$$\frac{\mathrm{d}}{\mathrm{d}t} \det(x_1, \ldots, x_k) = \det(\frac{\mathrm{d}}{\mathrm{d}t} x_1, x_2, \ldots, x_k) + \det(x_1, \frac{\mathrm{d}}{\mathrm{d}t} x_2, \ldots, x_k) + \ldots + \det(x_1, x_2, \ldots, \frac{\mathrm{d}}{\mathrm{d}t} x_k)$$

Lemma 1.2.9. *Let A be an $n \times n$-matrix over the complex numbers and let $\phi(A, \lambda) = \det(\lambda I_n - A)$ be the characteristic polynomial of A. Let B_i be the principal submatrix of A formed by deleting the i-th row and column of A and let $\phi(B_i, \lambda)$ be the characteristic polynomial of B_i. Then:*

$$\phi'(A, \lambda) = \frac{\mathrm{d}\phi(A, \lambda)}{\mathrm{d}\lambda} = \sum_i \phi(B_i, \lambda)$$

Proof. The proof is immediately done by direct computation:

$$\phi(A, \lambda) = \det \begin{bmatrix} \lambda - a_{1,1} & -a_{1,2} & \ldots & -a_{1,n} \\ -a_{2,1} & \lambda - a_{2,2} & \ldots & -a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n,1} & -a_{n,2} & \ldots & \lambda - a_{n,n} \end{bmatrix}.$$

Using the extension of the product rule of derivation for multilinear functions

$$\phi'(A,\lambda) = \det\begin{bmatrix} 1 & 0 & \dots & 0 \\ -a_{2,1} & \lambda - a_{2,2} & \dots & -a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n,1} & -a_{n,2} & \dots & \lambda - a_{n,n} \end{bmatrix} + \det\begin{bmatrix} \lambda - a_{1,1} & -a_{1,2} & \dots & -a_{1,n} \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n,1} & -a_{n,2} & \dots & \lambda - a_{n,n} \end{bmatrix} + \dots$$
$$+ \det\begin{bmatrix} \lambda - a_{1,1} & -a_{1,2} & \dots & -a_{1,n} \\ -a_{2,1} & \lambda - a_{2,2} & \dots & -a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \sum_i \phi(B_i,\lambda).$$

□

1.2.2 Proof

We now collect the above results into the following main theorem: we finally arrived at the Perron-Frobenius Theorem:

Theorem 1.2.10. *(**Perron-Frobenius theorem**) Let $A \geq 0$ be an irreducible $n \times n$-matrix. Then,*

1. *A has a positive real eigenvalue equal to its spectral radius.*
2. *To $\rho(A)$ there corresponds an eigenvector $\mathbf{x} > 0$.*
3. *$\rho(A)$ increases when any entry of A increases.*
4. *$\rho(A)$ is a simple eigenvalue of A.*
5. *If $A\mathbf{x} = \rho(A)\mathbf{x}$ where $\mathbf{x} > 0$ and $\mathbf{x}$ is a normalized vector, then $\mathbf{x}$ is unique.*

Proof. (1) and (2) follow immediately from Lemma 1.2.4 and Corollary 1.2.6.

(3) Suppose we increase some entry of the matrix A, giving us a new irreducible matrix $\tilde{A}$ where $\tilde{A} \geq A$ and $\tilde{A} \neq A$. Applying Lemma 1.2.5, we conclude that $\rho(\tilde{A}) > \rho(A)$.

(4) $\rho(A)$ is a simple eigenvalue of A, i.e., $\rho(A)$ is a zero of multiplicity one of the characteristic polynomial $\phi(\lambda) = \det(\lambda I_n - A)$, we make use of Lemma 1.2.9 by using the fact that $\phi'(\lambda)$ is the sum of the determinants of the principal $(n-1) \times (n-1)$ submatrices of $\lambda I - A$. If A_i is any principal submatrix of A, then from Lemma 1.2.8, $\det(\lambda I - A_i)$ (with I the identity matrix with the same size as the principal submatrix A_i) cannot vanish for any $\lambda \geq \rho(A)$. From this it follows that:

$$\det(\rho(A)I - A_i) > 0,$$

and thus

$$\phi'(\rho(A)) > 0.$$

Consequently, $\rho(A)$ cannot be z zero of $\phi(\lambda)$ of multiplicity greater than one and thus $\rho(A)$ is a simple eigenvalue of A.

(5) If $A\mathbf{x} = \rho(A)\mathbf{x}$ where $\mathbf{x} > 0$ and $||x|| = 1$ ($||x||$ denotes the standard Euclidean norm), we cannot find another eigenvector $\mathbf{y} \neq s\mathbf{x}$, with s a scalar, of A with $A\mathbf{y} = \rho(A)\mathbf{y}$, so that the eigenvector $\mathbf{x}$, meaning that the normalized eigenvector $\mathbf{x}$ is uniquely determined. □

With the previous proof in mind, the following definition comes not unexpected:

Definition 1.2.11. *If a matrix A has an eigenvalue equal to the spectral radius $\rho(A)$, this eigenvalue is called the* ***Perron root****, the corresponding eigenvector is called the* ***Perron root****.*

1.2.3 Example

To check wether a matrix with nonnegative entries is primitive, irreducible or neither, we just have to replace all nonzero entries by 1 since this does not affect the classification. The matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

is strictly positive and thus primitive. The matrices

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

both have 1 as a double eigenvalue hence can not be irreducible. The matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ satisfies:

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

and hence is primitive. The same goes for

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

this matrix is irreducible but not primitive. Its eigenvalues are 1 and -1.

1.3 Numerical analysis

1.3.1 Bachmann-Landau notations

For comparing the computational cost of algorithms, it's important to know the *Bachmann-Landau notations*. These notations are used to describe the limiting behavior of a function in terms of simpler functions. These notations are used a lot in computer science to classify algorithms by how their number of steps depends on changes in input size. We are only interested in the effects on the number of steps for really large input sizes, so constants don't play any role in the classification.

Bigh Oh

Definition 1.3.1. ***(Big Oh)***
Big Oh is the set of all functions f that are bounded above by g asymptotically (up to constant factor).

$$O(g(n)) = \{f | \exists c, n_0 \geq 0 : \forall n \geq n_0 : 0 \leq f(n) \leq cg(n)\}$$

We now proof a very simple lemma to show that indeed constant factors doesn't matter for Big Oh:

Lemma 1.3.2. $\forall k > 0 : O(k.g(n)) = O(g(n))$

Proof.

$$\begin{aligned} O(k.g(n)) &= \{f | \exists c, n_0 \geq 0 : \forall n \geq n_0 : 0 \leq f(n) \leq k.c.g(n)\} \\ &= \{f | \exists c, n_0 \geq 0 : \forall n \geq n_0 : 0 \leq f(n) \leq (k.c).g(n)\} \\ \text{let c' = k.c} & \\ &= \{f | \exists c', n_0 \geq 0 : \forall n \geq n_0 : 0 \leq f(n) \leq c'.g(n)\} \\ &= O(g(n)) \end{aligned}$$

□

Small Oh

Definition 1.3.3. ***(Small Oh)***
Small Oh is the set of all functions f that are dominated by g asymptotically.

$$o(g(n)) = \{f | \forall \varepsilon > 0, \exists n_0 \forall n \geq n_0 : f(n) \leq \varepsilon.g(n)\}$$

Note that the small oh-notation is a much stronger statement than the corresponding big oh-notation: every function that is in the small oh of g is also in big oh, but the inverse isn't necessarily true. Intuitively, $f(x) \in o(g(x))$ means that $g(x)$ grows much faster than $f(x)$.

Asymptotical Equality

Definition 1.3.4. ***(Asymptotically Equal)***
Let f and g real functions, then f is asymptotically equal to $g \Leftrightarrow \lim_{x \to +\infty} \frac{f(x)}{g(x)} = 1$. Notation: $f \sim g$.

In fact asymptotical equality, can also be defined as an equivalency relation: $f \sim g \Leftrightarrow (f - g) \in o(g)$. It's trivially clear that as $f \sim g \Rightarrow f \in O(g)$.

1.3.2 The Power Method

We now introduce the *classical power method*, also called the *Von Mises iteration* ([GOLUB]) because adaptions of this iterative method will appear in the following chapters of this master thesis. The power method is an eigenvalue algorithm that, given a diagonalizable matrix A, finds the eigenvalue λ with the greatest magnitude and a corresponding eigenvector $\mathbf{v}$ such that:

$$A\mathbf{v} = \lambda\mathbf{v}.$$

The power method is special because it doesn't use any matrix decomposition technique for obtaining results, making it suitable for very large matrices. At the other hand, it only finds one eigenvalue with a corresponding eigenvector and the the iterative process might converge very slowly.

There are plenty of variations of the power method available that overcome all these difficulties (finding only 1 eigenvalue/eigenvector, slow convergence, A must be diagonalizable, ...) but we limit our discussion here to the very basic method, therefore we call it the *classical* power method.

We first introduce some needed definitions, theorems and notations.

Definition 1.3.5. *Consider a real $n \times n$-matrix A with (not necessarily different) eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. When*

$$|\lambda_1| > |\lambda_2| \leq |\lambda_3| \leq \ldots \leq |\lambda_n|,$$

*λ_1 is called the **dominant eigenvector**.*

Corollary 1.3.6. *The dominant eigenvector λ_1 of a $n \times n$-matrix A is real.*

Proof. This is trivial, if λ_1 would be complex, the complex conjugate of λ_1 would also be an eigenvalue with the same modulus. □

Theorem 1.3.7. *A $n \times n$ matrix A is diagonalizable if and only if it has an eigenbasis (a basis containing only lineair independent eigenvectors).*

Proof. $\Leftarrow$ Consider $A \in \mathbb{R}^{n \times n}$ with (not necessarily different) eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and the eigenbasis $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$. Let:

$$X = [\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}]$$

Because the columns of X are linearly independent, X is invertible. Since:

$$AX_i = \lambda_i X_i$$

we also have:

$$AX = X\Lambda$$

with $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$, thus:

$$X^{-1}AX = \Lambda.$$

$\Rightarrow$ When $A \in \mathbb{R}^{n \times n}$ is diagonalizable, then there exist an invertible matrix T so:

$$TAT^{-1} = D$$

with D diagonal. Let $Y = T^{-1}$, then:

$$AY = YD.$$

This means that the columns of Y are eigenvectors where the corresponding eigenvalues are noted in the diagonal matrix D. Because T and Y are invertible, the columns of Y are lineair independent. This means that A has n lineair independent eigenvectors, forming an eigenbasis. □

Notation 1.3.8. *Let $\mathbf{x}^{(i)}$ denote vector $\mathbf{x}$ at iteration step i.*

The algorithm

Let $A \in \mathbb{R}^{n\times n}$ be a diagonalizable matrix with dominant eigenvalue λ_1 . We know that A has eigenbasis $V = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$. Every vector $\mathbf{x^{(0)}} \in \mathbb{R}^n$ can be written as a lineair combination of elements in X, because X spans the space $\mathbb{R}^n$. So:

$$\mathbf{x}^{(0)} = \sum_{i=1}^{n} \xi_i \mathbf{v_i}.$$

Now construct the sequence of vectors $\mathbf{x^{(k)}}$:

$$\mathbf{x^{(k)}} = A\mathbf{x^{(k-1)}} = A^k\mathbf{x^{(0)}}$$

Now:

$$\begin{aligned} A^k\mathbf{x^{(0)}} &= \sum_{i=1}^{n} \xi_i A^k \mathbf{v_i} \\ &= \sum_{i=1}^{n} \xi_i \lambda_i^k \mathbf{v_i} \\ &= \lambda_1^k \left\{ \xi_1 \mathbf{v_1} + \xi_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k \mathbf{v_2} + \ldots + \xi_n \left(\frac{\lambda_n}{\lambda_1}\right)^k \mathbf{v_n} \right\} \end{aligned}$$

Because $|\lambda_i| < |\lambda_1|$ for $i > 1$, we have that

$$\left(\frac{\lambda_i}{\lambda_1}\right)^k \to 0 \text{ for } k \to \infty$$

so:

$$\mathbf{x^{(k)}} = \lambda_1^k \xi_1 \mathbf{v_1} + o(1) \text{ for } k \to \infty \tag{1.18}$$

This means that for k large enough $\mathbf{x^{(k+1)}}$ is equal to λ_1 times $\mathbf{x^{(k)}}$. So when, the ratio between the corresponding vector entries in $\mathbf{x^{(k+1)}}$ and $\mathbf{x^{(k)}}$ becomes constant after k iteration steps, then this ratio will be equal to the dominant eigenvalue λ_1. $\mathbf{x^{(k)}}$ will be a corresponding eigenvector because it's is proportional to $\mathbf{v_1}$.

The start value $\mathbf{x^{(0)}}$ must only satisfy the condition that $\xi_1 \neq 0$, in other words: $\mathbf{x^{(0)}}$ must have a non-zero component belonging to the dominant eigenvector. In general, each randomly chosen option for $\mathbf{x^{(0)}}$ will normally fulfill this requirement. Even if we are so unlucky to pick a starting vector which doesn't, subsequent $\mathbf{x^{(k)}}$ will again fulfill the requirement because rounding errors sustained during the iteration will have a component in this direction.

A practical problem arises now when one of the components of $\mathbf{x^{(k)}}$ is equal to zero. If we want to take the ratio between the corresponding components of $\mathbf{x^{(k)}}$ and $\mathbf{x^{(k+1)}}$ we get a division by zero. We can solve this by a property that any norm function has:

$$\|\lambda_1 \mathbf{x^{(k)}}\| = |\lambda_1| \|\mathbf{x^{(k+1)}}\|.$$

Because $\mathbf{x^{(k)}} \neq 0$ we have that $\|\mathbf{x^{(k+1)}}\| \approx \|\lambda_1 \mathbf{x^{(k+1)}}\|$, so we calculate λ_1 in the power method by:

$$|\lambda_1| = \lim_{k\to\infty} \frac{\|\mathbf{x^{(k+1)}}\|}{\|\mathbf{x^{(k)}}\|}$$

To decide on the sign of λ_1, just divide two non-zero components of f $\mathbf{x^{(k+1)}}$ and $\mathbf{x^{(k)}}$.

Another issue to address is that the components of $\mathbf{x^{(k)}} = A^k\mathbf{x^{(0)}}$ can become very high or very low, which can cause an overflow or onderflow in the real number representation of computers. To avoid this, we use normed versions of the $\mathbf{x^{(k)}}$-vectors: we start with a vector $\mathbf{y^{(0)}}$ with $\|\mathbf{y^{(0)}}\| = 1$. Subsequently, we calculate for $k = 0, 1, \ldots$:

$$\mathbf{z^{(k+1)}} = A\mathbf{y^{(k)}}, \quad \mu_{k+1} = \|\mathbf{z^{(k+1)}}\|, \quad \mathbf{y^{(k+1)}} = \frac{\mathbf{z^{(k+1)}}}{\mu_{k+1}}.$$

The vectors $\mathbf{y^{(k)}}$ have all magnitude 1 and the components of $\mathbf{z^{(k)}}$ are restricted because:

$$\|\mathbf{z^{(k)}}\| = \|A\mathbf{y^{(k-1)}}\| \leq \|A\|\|\mathbf{y^{(k-1)}}\| = \|A\|$$

and when A is invertible we have $\|\mathbf{y^{(k-1)}}\| = 1 \leq \|A^{-1}\|\|\mathbf{z^{(k)}}\|$. Thus $\|\mathbf{z^{(k)}}\| \geq \frac{1}{\|A^{-1}\|}$. If we want to calculate the eigenvalue we can use:

$$\begin{aligned} A^k\mathbf{y^{(0)}} &= A^{k-1}A\mathbf{y^{(0)}} = A^{k-1}\mathbf{z^{(1)}} = A^{k-1}\mu_1\mathbf{y^{(1)}} \\ &= A^{k-2}\mu_1 A\mathbf{y^{(1)}} = A^{k-2}\mu_1\mathbf{z^{(2)}} = A^{k-2}\mu_1\mu_2\mathbf{y^{(2)}} \\ &= \ldots \\ &= \mu_1\mu_2\ldots\mu_k\mathbf{y^{(k)}} \end{aligned}$$

So:

$$|\lambda_1| = \lim_{k\to\infty}\frac{\|A^{k+1}\mathbf{y^{(0)}}\|}{\|A^k\mathbf{y^{(0)}}\|} = \lim_{k\to\infty}\frac{\mu_1\mu_2\ldots\mu_{k+1}\|\mathbf{y^{(k+1)}}\|}{\mu_1\mu_2\ldots\mu_k\|\mathbf{y^{(k)}}\|} = \lim_{k\to\infty}\mu_{k+1}$$

Because μ_{k+1} converges, a good choice for a stop condition for our numerical algorithm could be

$$|\mu_k - \mu_{k-1}| < Tol,$$

which guarantees an estimation error of at most Tol (usually $Tol = 10^{-5}$) for the approximation of the dominant eigenvalue. With all this information, we construct algorithm 1. Just to give an understandable algorithm, we used the Euclidean norm in this algorithm[1].

[1] $||.||_2$ is the Euclidean vector norm.

Data:
$\mathbf{y^{(0)}}$: a start vector with $\|\mathbf{y^{(0)}}\|_2 = 1$,
Tol: Tolerance for the estimation error.
Result:
$\mathbf{y^{(k)}}$: an estimation of a dominant eigenvector,
μ_k: an estimation of the dominant eigenvalue.
begin `power_method(`$\mathbf{y^{(0)}}$`,` k`)`
  $k = 1$;
  repeat
    $\mathbf{z^{(k)}} = A\mathbf{y^{(k-1)}}$;
    $\mu_k = \|\mathbf{z^{(k)}}\|_2$;
    $\mathbf{y^{(k)}} = \dfrac{\mathbf{z^{(k)}}}{\mu_k}$;
    $k = k + 1$
  until $k > 2$ *and* $|\mu_k - \mu_{k-1}| < Tol$,;
  if *the components* $\mathbf{y^{(k)}}$ *and* $\mathbf{y^{(k-1)}}$ *have a different sign* **then**
    $\mu_k = -\mu_k$;
  end
  return $\mathbf{y^{(k)}}, \mu_k$;
end

Algorithm 1: The Power method

Computational cost & Usage

The computational cost of the algorithm is determined by the speed at which the $o(1)$ terms in 1.18 go to zero. This is indicated by the slowest converging term $(\lambda_2/\lambda_1)^k$. This means that the algorithm converges slowly when there is an eigenvalue close in magnitude to the dominant eigenvalue. We get following expression for approximation μ_k of λ_1:

$$|\mu^k - \lambda_1| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$$

In our algorithm we accepted an estimation error of 10^{-5}, so the number of steps n can be computed as:

$$\left|\frac{\lambda_2}{\lambda_1}\right|^n \approx Tol$$

So, for example, for $Tol = 10^{-5}$ we get $n = -5/\log \lambda_1/\lambda_2$, we become:

$$\texttt{power\_method} \in O\left(\frac{-1}{\log \lambda_2/\lambda_2}\right)$$

Note that this O holds for any estimation error Tol of the form 10^{-e} with $e \in \mathbb{N}$.

When $\lambda_2 \approx \lambda_1$ we see that the power method (almost) has infinitely many steps. Since we do not know the eigenvalues of A, this means we cannot know in advance whether the power method will work or not. Recall that the eigenvalues of a real matrix A are in general complex, and occur in conjugate pairs. This means, when the dominant eigenvalue of A is not real, the power method will certainly fail. Therefore, it is a good idea to apply the power method only to matrix whose eigenvalues are known to be real. The only thing that can

go wrong with those matrices is that the dominant eigenvalue has an algebraic multiplicity larger than 1.

From the Perron-Frobenius theorem in 1.2.10, we also get another good choice: namely the irreducible matrices or any matrix with strictly positive entries. Indeed, The Perron-Frobenius theorem tells us that they have a unique dominant eigenvalue.

Example

Example 1.3.9. Consider the matrix:

$$A = \begin{pmatrix} 1 & -3 & 5 \\ -1 & -7 & 11 \\ -1 & -9 & 13 \end{pmatrix}$$

A has a dominant eigenvalue 3 and a double eigenvalue 2. A corresponding dominant eigenvector is $(1, 1, 1)$. Now we use the classical power method with start vector $\mathbf{y^{(0)}} = (1, 0, 0)$, $Tol = 10^{-5}$ and we become the values in Table 1.3.9. Because we know the eigenvalues

k	μ_k		k	μ_k
0	1.00000		15	3.00459
1	1.73205		16	3.00305
2	4.12311		17	3.00204
3	4.06564		18	3.00136
4	3.58774		19	3.00090
5	3.34047		20	3.00060
6	3.20743		21	3.00040
7	3.13055		22	3.00026
8	3.08385		23	3.00017
9	3.05457		24	3.00011
10	3.03582		25	3.00008
11	3.02363		26	3.00005
12	3.01564		27	3.00003
13	3.01037		28	3.00002
14	3.00689		29	3.00001

Table 1.1: The iteration values μ_k of Example 1.3.9.

of A, we can predict the number of steps $n = -5/\log \lambda_2/\lambda_2 = -5/(\log 2/3) \approx 28$. Because $\lambda_2/\lambda_1 = 2/3$ is not that small, the convergence here is also not that fast. We have for the approximation of a corresponding eigenvector

$$\mathbf{y^{(29)}} = (-0.577348, -0.577352, -0.577351)$$

which is (more or less) in proportion with $(1, 1, 1)$.

1.4 Graphs

After introducing different kinds of matrices and proving the Perron-Frobenius theorem, we now take a closer look at graphs. Here too, we'll look at different families of graphs and prove

some relevant properties about them. We also link the concept of graphs with different kinds of matrices, deepening our insight of some theorems of the previous section.

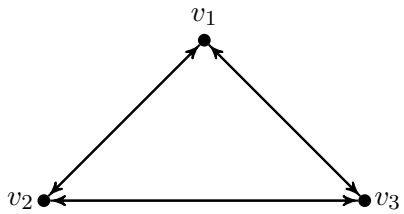
The definitions and results in this section are mainly based on the course 'Discrete Mathematics' by P. Cara [CARA].

1.4.1 General definitions

Definition 1.4.1. *A **graph** is a an ordered pair $(V, \to)$ where V is a set and $\to$ is a relation. The elements of V are called **vertices** and $\to$ is called the **adjacency relation**. Let $u, v \in V$, then the ordered pair (u, v) belonging to $\to$ is called an **arc** or **edge** and we write $u \to v$. We also say that u is **adjacent** to v. When $v \to v$ (with $v \in V$) we say that the graph has a **loop** in v. A graph $(V, \to)$ is most of the time denoted by calligraphic letters $\mathscr{G}, \mathscr{H}, ...$*

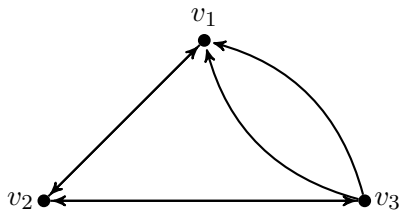
When the relation $\to$ is symmetric, we call the graph **undirected**, in this case we often write $\sim$ instead of $\to$.

Example 1.4.2. The graph



is an undirected graph with vertices v_1, v_2, v_3. The adjacency relation $\to$ equals $\{(v_1, v_2), (v_2, v_1), (v_2, v_3), (v_3, v_2), (v_3, v_1), (v_1, v_3)\}$.

There is a small problem with our definition, because not all graphs are taken into account, for example the graph below is not a graph following our definition because you can not define multiple edges between vertices in a relation.



Therefore we introduce a more general definition and introduce the concept of multiplicity of an edge:

Definition 1.4.3. *A graph is an ordered pair (V, μ) with V a set and $\mu : V \times V \to \mathbb{N}$ a function that gives the **multiplicity** of an edge. The function is defined as follows:*

- *when $\mu(u, v) = 0$ we say that u and v are not adjacent;*
- *when $\mu(u, v) = k > 0$ we say that there are k edges from u to v.*

It is clear that our previous definition fits perfectly in this more general definition, by constructing μ in this case as follows:

$$\mu(u,v) = \begin{cases} 1 & \text{if } (u,v) \in \rightarrow \\ 0 & \text{otherwise.} \end{cases}$$

Definition 1.4.4. *The **neighbourhood** of a vertex v of a graph $\mathscr{G} = (V, \mu)$ is the induced subgraph $\mathscr{G}_v$ with vertex set V' consisting of all vertices adjacent to v without v itself and with the multiplicity function μ', which is the restriction of μ to the vertices in V'. A vertex with a neighbourhood equal to the empty graph (a graph with an empty set of vertices) is called* isolated.

Definition 1.4.5. *The **order of a finite graph** $\mathscr{G}$ is the number of vertices of $\mathscr{G}$ and is denoted by $|\mathscr{G}|$.*

Definition 1.4.6. *The **degree of a vertex** v **in a graph** $\mathscr{G}$ is the number of edges containing v and is denoted by* $\deg(v)$*, so:*

$$\deg v = |\mathscr{G}_v|$$

Definition 1.4.7. *A **walk** in a graph $\mathscr{G}$ is a sequence of vertices*

$$v_0, v_1, \ldots, v_k$$

*such that $v_{i-1} \rightarrow v_i$ for each $i \in \{1, \ldots, k\}$. The **length** of the walk is k, one less than the number of vertices.*

Definition 1.4.8. *If all edges are distinct in a walk in a graph $\mathscr{G}$, we call the walk a **path**.*

Definition 1.4.9. *A **cycle** is a walk from v_0 to v_0 in which all vertices except v_0 are distinct.*

Definition 1.4.10. *A **simple graph** is an undirected graph $\mathscr{G}(V, \mu)$ containing no loops and for all vertices $v_i, v_j \in V$, we have that the multiplicity $\mu(v_i, v_j)$ is at most 1.*

Product graphs

Definition 1.4.11. *Take two graphs $\mathscr{G}(U, \rightarrow), \mathscr{H}(V, \rightarrow')$, the **product graph** $\mathscr{G} \times \mathscr{H}$ is the graph with $|\mathscr{G}|.|\mathscr{H}|$ vertices and that has an edge between vertices (u_i, v_j) and (u_k, v_l) if there is an edge between u_i and u_k in $\mathscr{G}$ and there is an edge between v_j and v_l in $\mathscr{H}$.*

Adjacency matrices

We now represent a finite graph in the form of an adjacency matrix. This matrix gives a lot of useful information about the graph and vice versa.

Definition 1.4.12. *Let $\mathscr{G} = (V, \mu)$ be a graph of order n and define a numbering on the vertices $v_1, \ldots, v_n$. Then the **adjacency matrix** $A_{\mathscr{G}}$ of $\mathscr{G}$ is the real $n \times n$-matrix with a_{ij} equal to $\mu(v_i, v_j)$.*

Corollary 1.4.13. *The adjacency matrix of an undirected graph $\mathscr{G} = (V, \sim)$ is a symmetric matrix.*

Proof. This is trivial by the definition of an undirected graph. □

Theorem 1.4.14. *Let $k > 0$. The element on place (i, j) in $A^k_{\mathscr{G}}$ contains the number of walks of length k from i to j in the graph $\mathscr{G} = (V, \mu)$.*

Proof. By induction on k.

For $k = 1$ we count the walks of length 1. These are edges and the result follows immediately from the definition of $A_{\mathscr{G}}$.

Let v_l be a vertex of $\mathscr{G}$. If there are b_{ij} walks of length k from i to l and a_{lj} walks of length 1 (edges) from v_l to v_j, then there are $b_{il}a_{lj}$ walks of length $k+1$ from v_i to v_j passing vertex v_l. Therefore, the number of walks of length $k + 1$ between v_i and v_j is equal to:

$$\sum_{l \in V} b_{il}a_{lj} =: c_{ij}.$$

By the induction hypothesis we now that b_{il} equals the element on place (i, l) in $A^k_{\mathscr{G}}$ so c_{ij} is exactly the element on place (i, j) in the matrix product

$$A^k_{\mathscr{G}} A_{\mathscr{G}} = A^{k+1}_{\mathscr{G}}.$$

□

Example 1.4.15. The adjacency matrix of the graph in Example 1.4.2 is:

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

1.4.2 Strong connectivity

In this section, we take a closer look at *directed graphs*. We already showed that the adjacency matrix of a simple graph $\mathscr{G}$ is irreducible if and only if $\mathscr{G}$ is connected. But this is not appropriate for checking the irreducibility of matrices using graphs, because the only matrices we can check are the ones containing entries equal to 0 or 1. Can we find a more general method for checking irreducibility of a matrix using graphs? Luckily, we can and that's where directed graphs come in: we will introduce a method for turning a matrix in to a directed graph and vice versa and link the irreducibility of this matrix to a property of the directed graph.

Definition 1.4.16. *An undirected graph $\mathscr{G}$ is **connected** if it possible to establish a path from any vertex to any other vertex.*

Definition 1.4.17. *A directed graph $\mathscr{G}(V, \rightarrow)$ is **connected** if the underlying undirected graph (remove all arrows on the edges) is connected, the directed graph $\mathscr{G}$ is **strongly connected** if there is a path in each direction between each pair of vertices of the graph.*

In the next proof, we study the equivalence of the matrix property of irreducibility of Definition 1.1.10 with the concept of the strongly connected directed graphs of a matrix:

Theorem 1.4.18. *Let $\mathscr{G}$ be a (directed) graph with adjacency matrix A. Then $\mathscr{G}$ is strongly connected if and only if A is irreducible.*

Proof. From Theorem 1.4.14 we know that a graph is strongly connected if and only if for every pair of indices i and j there is an integer k such that $(A^k)_{ij} > 0$, from Theorem 1.1.12 we know this means that A is irreducible and vice versa.

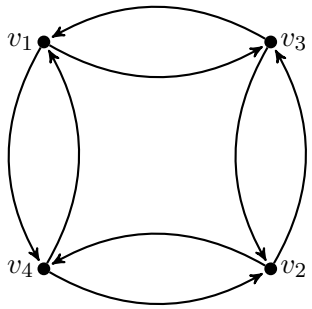
□

We now link directed graphs with $n \times n$-matrices and vice versa. Consider any $n \times n$-matrix A, and consider any n distinct points $v_1, v_2, \ldots, v_n$ in the plane, which will be the nodes of the directed graph. For every nonzero entry a_{ij} of the matrix, we connect the node v_i to the node v_j by means of an arc $v_i \to v_j$ directed from v_i to v_j. In this way, with every $n \times n$-matrix A can be associated a **directed** graph $\mathscr{G}(A)$.

Example 1.4.19. Consider the matrix:

$$B = \begin{pmatrix} 0 & 0 & 89 & 7 \\ 0 & 0 & 2 & 2 \\ 123 & 9 & 0 & 0 \\ 14 & 89 & 0 & 0 \end{pmatrix}.$$

We get the directed graph:



Notice that this graph is strongly connected.

Corollary 1.4.20. *An $n \times n$-matrix A is irreducible if and only if its directed graph $\mathscr{G}(A)$ is strongly connected.*

Proof. This follows immediately from the previous proof. □

Chapter 2

Similarity on graphs

In the previous chapter all the basic terminology and results were introduced, we now take an extensive look at the concept of similarity on graphs. Similarity on graphs is a fairly new concept to compare the nodes of two graphs. The concept arose from the research on algorithms for web searching engines (like *Google*, *Yahoo*,...) in the late nineties. More specifically, Jon M. Kleinberg introduced in his paper 'Authoritative Sources in a Hyperlinked Environment' ([KLEINBERG]) the famous 'HITS algorithm' for extracting information from the link structure of websites. The method leads to an iterative algorithm where graphs represent the link structure of a collection of websites on a specific topic. Because this paper formed the basis of later research on similarity on graphs, the whole idea and algorithm of Kleinberg is introduced in the first section of this chapter. In 2004, V.D. Blondel et al. ([BLONDEL]) generalized the algorithm of Kleinberg, introducing the notion of similarity on directed graphs. This similarity is covered in the second section. With this similarity on directed graphs, there is a much wider scope of applications than just search algorithms. Finally, we conclude by looking at similarity on colored graphs in the last section. Both sections on similarity on directed and colored graph conclude with an application on the Eurovision Song Contest results between 1980-1990.

2.1 The HITS algorithm of Kleinberg

2.1.1 History

Back in the nineties, internet became more and more popular by the public. The popular search engines back then where Altavista and Yahoo, but they weren't as advanced as search engines today. The main pitfall of the first search engines was that the search results were purely based on the number of occurrences of a word in a webpage. This was a pitfall for many reasons. The first reason was the growing popularity of the internet: as more and more webpages were put online, simply getting the relevant pages to a search query in this text-based manner, was a process that could possibly return millions of relevant pages. Also *content similarity* was an issue: a website owner can easily cheat in a text-based search system by just adding and repeating some very popular search words, making his website appear in the results of a large number of search queries. Two possible solutions were simultaneously invented in 1997 and 1998. The first one was the *PageRank*-system developed by Larry Page and Sergey Brin ([PAGE]). The PageRank system led to the foundation of the immensely popular Google search engine. Meanwhile, also Jhon Kleinberg came up with his own solution,

the *HITS algorithm* (hyperlink-induced topic search). At that time, he was both a professor in the Computer Science Department at the Cornell University and researcher for IBM. The algorithm is used inter alia today by the Ask search engine (www.ask.com). Both these algorithms use the hyperlinks between webpages to rank search results. Because this master thesis is about similarity and this concept is introduced on graphs as a generalization of the HITS algorithm, we don't go to into further detail about the PageRank-algorithm. In the following paragraphs, the HITS algorithm is extensively explained.

2.1.2 Motivation

Kleinbergs work originates in the problems that arises with text-based searching the WWW. Text-based searching just counts all the occurrences of a given search query on webpages and returns a set of webpages ordered by decreasing occurency. When a user supplies a search query, we probably face an *abundance problem* with this method: the number of pages that could reasonably be returned as relevant is far too large for a human user to digest. To provide effective search results under these conditions, we need to filter the 'authoritative' ones. We face some complications when we want to filter the 'authoritative' webpages in a text-based system. For example, if we search for '`job offers in Flanders` ' the most authoritative page and expected first result in a search engine would be `www.vdab.be`. Unfortunately, the query '`job offers`' is used in over a million pages on the internet and `www.vdab.be` is not the one using the term most often. Therefore, there is no way to favor `www.vdab.be` in a text-based ranking function. This a recurring phenomenon, as another example if you search for the query '`computer brands`', there is no reason at all to be sure that the website of Apple or Toshiba even contain this search term.

The HITS algorithm solves these difficulties by analyzing the hyperlink structure among webpages. The idea is that hyperlinks encode a sort of human judgment and that this judgement is crucial to formulate a notion of authority. Specifically, when a page p includes a link to page q, it means that p gives a *conferred authority* on q. Again we face difficulties, because this conferred authority doesn't hold for every link. Links are created for a wide variety of reasons, for example, a large number of links are created for navigation within a website (e.g. "Return to homepage") and these have of course nothing to do with a notion of authority.

The HITS method is based on the relationship between the *authorities* for a topic and those pages that link to many related authorities, called *hubs*. Page p is called an *authority* for the query "`smartphone brand`" if it contains valuable information on the subject. In our example websites of smartphone manifacturers such as "`www.apple.com`", "`www.samsung.com`",... would be good authorities for this search query and these are the results a user expect from a search engine.

A hub is a second category of pages needed to find good authorities. Their role is to advertise authoritative pages. Hubs contain useful links toward these authorities. In our example, consumer websites with reviews on smartphones, websites of smartphone shops,... would be good hubs. In fact, hubs point the search process in the 'right direction'.

To really grasp the idea, we make an analogy with everyday life. If you tell a friend that you think of buying a new smartphone, he might tell you his experiences with smartphones and he will probably share some opinions he got from other friends. He might suggest you some good models and good brands. Now, you are more inclined to buy a smartphone that your friend suggested. Well, this idea is used in the HITS-method: your friend served as hub,

Data:
σ: a query string.
$\mathcal{E}$: a text-based search engine.
t: natural number (usually initiated to 200)
d: natural number (usually initiated to 50).
Result: A page set S_σ satisfying all the properties of our wish list.
begin `create_graph(`σ, $\mathcal{E}$, t, d`)`
  Let R_σ denote the top t results of $\mathcal{E}$ on σ;
  Set $S_\sigma := R_\sigma$;
  for *each page* $p \in R_\sigma$ **do**
    Let $\Gamma^+(p)$ denote the set of all pages p points to;
    Let $\Gamma^-(p)$ denote the set of all pages pointing to p;
    Add all pages in $\Gamma^+(p)$ to S_σ;
    if $|\Gamma^-(p)| \leq d$ **then**
      Add all pages in $\Gamma^-(p)$ to S_σ;
    else
      Add an arbitrary set of d pages from $\Gamma^-(p)$ to S_σ;
    end
  end
  return S_σ;
end

Algorithm 2: Algorithm to construct S_σ.

the brands and models he suggested are good authorities.

2.1.3 Constructing relevant graphs of webpages

Any collection of hyperlinked pages can be transformed to a directed graph $\mathscr{G} = (V, \to)$: the nodes correspond to the pages, and if there is a link from page p to page q, there is an arc $p \to q$. Suppose a search query is preformed, specified by a query σ. We wish to determine the authoritative pages by an analysis of the link structure. But first we have to construct a subgraph of the internet on which our algorithm will operate. We want to make the computational effort as efficient as possible, so we restrict the subgraph to the set Q_σ of all pages where the query σ occurs. For this, we could use any already existing text-based search engine. But, for our algorithm Q_σ is possibly much too big: it may contain millions of pages making it impossible for any computer to preform the algorithm. Moreover it is, as explained in the motivation in 2.1.2, possible that Q_σ does not contain some of the most important authorities because they never use the query string σ on their website.

Therefore, we wish to transform the set Q_σ to a set S_σ of pages following this 'wish list' of properties:

1. S_σ is relatively small,
2. S_σ is rich in relevant pages,
3. S_σ contains most of the strongest authorities.

By keeping S_σ small, the computational cost of preforming non-trivial algorithms can be kept under control. By the property of being rich in relevant pages, it will be easier to find good authorities.

To construct S_σ, we first construct a *root set* R_σ with the t highest-ranked pages for σ using a text-based search engine (they sort results based on the occurence of σ). Typically, t is set about 200. R_σ complies with properties 1 and 2 of our wish list, but because $R_\sigma \subset Q_\sigma$, it fails from satisfying property 3. Now we use the root set R_σ to create the set S_σ satisfying our complete wish list. When a strong authority is not in R_σ, it is very likely that at least one of the pages in R_σ points to this authority. Hence, by using the pages in R_σ, we can expand it to S_σ by looking at the links that enter and leave R_σ. We get algorithm 2.

Thus, we obtain S_σ by expanding R_σ to include any page pointed to by a page in R_σ. We also add d pages that point to a page in R_σ. d is usually initiated to 50. The parameter d is crucial to stay in accordance with property 1 of our wish list. Indeed, a webpage can be pointed to by several thousands and thousands of other webpages, and we don't want to include them all if we want to keep S_σ relatively small. Some experiments in [KLEINBERG] showed that this algorithm resulted in a S_σ with a size in the range of 1000 to 5000 web pages. Property 3 of our wish list is usually met because a strong authority need only be reference once in the t pages of the root set R_σ to be added to S_σ.

Denote the resulting graph of the page set S_σ by $\mathscr{G}[S_\sigma]$. Note that $\mathscr{G}[S_\sigma]$ will contain a lot of links serving only navigational purposes within a website. As mentioned before, these links have nothing to do with the the notion of authority and they must be removed from our final graph if we want a good determination of the authoritative pages by an analysis of the link structure. A very simple heuristic can be used to derive a subgraph of $\mathscr{G}[S_\sigma]$ leaving out all the navigational links: we make a distinction between *transverse* links and *intrinsic* links. Transverse links are links between different domain names (e.g. a link between `www.vub.ac.be` and `www.ua.ac.be`) and intrinsic links are links between the same domain name (e.g. a link between `www.vub.ac.be` and `dwis.vub.ac.be`). Intrinsic links exist to allow navigation within a website and they tell us very little about the authority of the pages they point to. Therefore, we delete all intrinsic links from $\mathscr{G}[S_\sigma]$, keeping only the arcs corresponding to transverse links.

Our graph still contains some meaningless links in the context of page authority. Suppose a large number of pages from the same domain name have a transverse link to the same page p. Most of the time, this means a form of advertisement (by example 'Website created by. . . .' at the bottom of each page). It is useful to only allow m pages (m is usually initiated to 6) from the same domain name to have a transverse link to the same page. If m is exceeded, all the transverse links must be deleted from the graph. Note, however, that not all links to advertisements will be erased because on most web pages, advertisements change on every page which avoids the exceedance of m.

Applying the two described heuristics above on $\mathscr{G}[S_\sigma]$, we get a new graph $\mathscr{G}'_\sigma$ which is exactly what we need to preform our link analysis.

2.1.4 Hubs and Authorities

A very simple approach would now be to order the pages in $\mathscr{G}'_\sigma$ by their in-degree. Although this approach can sometimes return good search results, this heuristic is often too simple because S_σ will probably contain some web pages with a lot of incoming links without being very relevant to the search query σ (e.g. advertisements). With these incoming links, those

web pages are ranked high in the final search result, which we want to avoid.

Do we have to return to a text-based approach to avoid irrelevant web pages being on top of the search results? No, the link structure of $\mathscr{G}'_\sigma$ can tell us a lot more than it may seem at first glance. Authoritative pages relevant to query σ should indeed have a large in-degree, but there should also be a considerable overlap in the sets of pages that point to authoritative pages. This set of pages that point to authoritative pages are called *hubs*. Hubs have links to several authoritative pages and they sort of "concentrate" all the authorities on query σ. Figure 2.1.1 shows what this means conceptually.

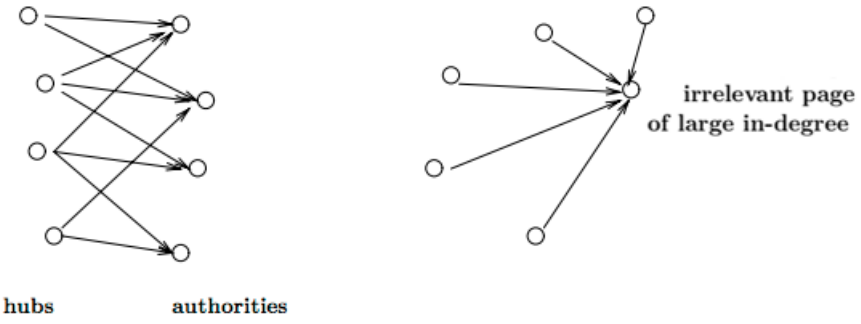


Figure 2.1.1: The concept of hubs and authorities (Source: [KLEINBERG])

So, for each page j we assign two scores, an *authority score* which estimates the value of the content of the page and a *hub score* which estimates the value of the outgoing links to other pages. We get now a *mutually reinforcing relation*: a good hub is a page pointing to many good authorities, a good authority is a page that is pointed to by many good hubs. This leads us to a *mutually reinforcing relation* resulting in an iterative method to break this circularity.

So let $\mathscr{G}'_\sigma = (V, \to)$ and let h_j and a_j be the hub and authority scores of vertex v_j (corresponding with page j). These scores must be initialized by some positive start values and then updated simultaneously for all vertices. This leads to a *mutually reinforcing relation* in which the hub score of v_j is set equal to the authority scores of all vertices pointed to by v_j and in an equal manner the authority score of v_j is set equal to the sum of the hub scores of all vertices pointing to v_j.

$$\begin{cases} h_j := \sum_{i:(v_j,v_i)\in\to} a_i, \\ a_j := \sum_{i:(v_i,v_j)\in\to} h_i. \end{cases}$$

The basic operations in which hubs and authorities reinforce one another are depicted in Figure 2.1.2.

Let B be the adjacency matrix of $\mathscr{G}'_\sigma$ and denote $\mathbf{a}$ as the authority vector with coordinates $(a_1, a_2, \ldots, a_n)$ (with $n = |\mathscr{G}'_\sigma|$, the number of pages) and $\mathbf{h}$ as the hub vector. The mutually reinforcing relation can now be rewritten as:

$$\begin{pmatrix} \mathbf{h} \\ \mathbf{a} \end{pmatrix}_{k+1} = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{h} \\ \mathbf{a} \end{pmatrix}_k, \quad k = 0, 1, \ldots,$$

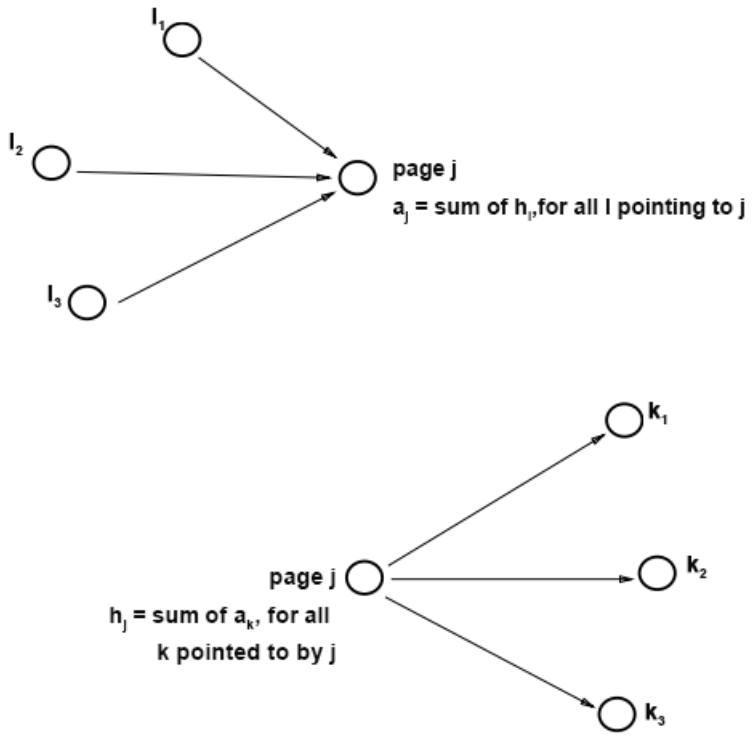


Figure 2.1.2: The basic operations in the reinforcing relation between hubs and authorities (Source: [KLEINBERG])

In compact form, we denote

$$\mathbf{x_{k+1}} = M\mathbf{x_k}, \quad k = 0, 1, \ldots, \tag{2.1}$$

where

$$\mathbf{x_k} = \begin{pmatrix} \mathbf{h} \\ \mathbf{a} \end{pmatrix}_k, M = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$$

After each iteration, we have to normalize h_j and a_j. Indeed, we want to get the authority and hub weights for each page and in order to compare these after each iteration step, they must be normalized because only the relative differences do matter, otherwise the whole procedure would be meaningless. Pages with larger a_j-scores are viewed as being better authorities, pages with larger h_j-scores are better hubs.

We get the following sequence (with z_0 some positive start value) of normalized vectors:

$$\mathbf{z}_0 = \mathbf{x}^{(0)} > 0, \mathbf{z}_{k+1} = \frac{M\mathbf{z}_k}{||M\mathbf{z}_k||_2}, \quad k = 0, 1, \ldots, \tag{2.2}$$

How do we decide on $\mathbf{x}^{(0)}$? We will see that any positive vector in $\mathbb{R}^{2n}$ is a good choice, but for the shake of simplicity, we choose the natural choice[1] $\mathbf{1} \in \mathbb{R}^{2n}$. The limit to which

[1] $\mathbf{1}$ is a matrix, or vector, whose entries are all equal to 1.

Data:
$\mathscr{G}$: a graph of n linked pages.
k: natural number.
Result: A vector $(\mathbf{h}, \mathbf{a})$ containing the hub and authority scores after k steps.
begin `hits(`$\mathscr{G}$`, k)`
  Set $\mathbf{a}^{(0)} = (1, 1, \ldots, 1) \in \mathbb{R}^n$;
  Set $\mathbf{h}^{(0)} = (1, 1, \ldots, 1) \in \mathbb{R}^n$;
  for $i = 1, 2, \ldots, k$ **do**
    Calculate $\mathbf{h}'^{(i)} = \left(\sum_{m:(v_1,v_m)\in\to} \mathbf{a}_m^{(i-1)}, \sum_{m:(v_2,v_m)\in\to} \mathbf{a}_m^{(i-1)}, \ldots, \sum_{m:(v_n,v_m)\in\to} \mathbf{a}_m^{(i-1)}\right)$;
    Normalize $\mathbf{h}'^{(i)}$ obtaining $\mathbf{h}^{(i)}$;
    Calculate $\mathbf{a}'^{(i)} = \left(\sum_{m:(v_m,v_1)\in\to} \mathbf{h}_m^{(i)}, \sum_{m:(v_m,v_2)\in\to} \mathbf{h}_m^{(i)}, \ldots, \sum_{m:(v_m,v_n)\in\to} \mathbf{h}_m^{(i)}\right)$;
    Normalize $\mathbf{a}'^{(i)}$ obtaining $\mathbf{a}^{(i)}$;
  end
  return $(\mathbf{h}^{(k)}, \mathbf{a}^{(k)})$;
end

Algorithm 3: The iterative HITS-algorithm.

Data:
$\mathscr{G}$: a graph of n linked pages.
k: natural number.
c: natural number.
Result: A vector $((\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_c), (\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_c))$ containing exactly the nodes of the c top hubs and c top authorities.
begin `filter(`$\mathscr{G}$`, k, c)`
  $(\mathbf{h}, \mathbf{a})$ = `hits(`$\mathscr{G}$`, k)`;
  Sort the pages with the c largest values in $\mathbf{h}$, resulting in a vector of nodes $(\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_c)$;
  Sort the pages with the c largest values in $\mathbf{a}$, resulting in a vector of nodes $(\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_c)$;
  return $((\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_c), (\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_c))$;
end

Algorithm 4: Returning the top c hubs and authorities

the sequence converges results in 'definitive' hub and authority scores for each page in the graph $\mathscr{G}'_\sigma$.

To compute the iterative algorithm, we update the hub and authority scores in an alternating form (by each step we have to normalize the scores). Because we will proof that te sequence converges, theoretically we can keep on iterating until a fixed point is reached. But in most practical settings, we choose a fixed number of steps k to reduce the computational cost because we can not know beforehand how large k has to be to reach the limit. But off course, it is extremely important to know that method converges anyway. Let $\mathbf{x}^{(i)}$ denote vector $\mathbf{x}$ at iteration step i as in Notation 1.3.8, and we become Algorithm 3.

To filter the top c hubs and the top c authorities, you can use the trivial algorithm 4.

How do we decide on the values of k and c? It's immediately clear that c and k must be

propertional: for low c values, a lower value for the number of iteration steps k is appropriate and vice versa. Experiments in [KLEINBERG] showed that k set to 20 is sufficient to become stable for finding the 5 best hubs and authorities, thus for $c = 5$.

2.1.5 Convergence of the algorithm

We know want to proof that for arbitrarily large values of k, the sequence z_k converge to a limit $(\mathbf{h}', \mathbf{a}')$. Before we want to prove the convergence, note that adjacency matrices are nonnegative by definition, and thus the matrix M is nonnegative too. M is also clearly a symmetric $n' \times n'$-matrix with nonnegative, real entries. We prove that such matrices have n' (not necessarily different) real eigenvalues and that we can diagonalize M. Remember that we introduced the power method in 1.3.2 and with this proofs about symmetric, real matrices, and with these theorems we showed that the first condition of the power method namely being a diagonalizable matrix) is met. If we can also prove the second condition (having a unique dominant eigenvalue), convergence is immediately shown by the power method.

However, there is a problem here: we can not prove that nonnegative symmetric matrices have a unique dominant eigenvalue, simply because this is not true in general.[2] In the original paper of Kleinberg ([KLEINBERG]) he solves this issue by simply imposing that the matrix M has a unique dominant eigenvalue and he doesn't pay any further attention to this problem. He presents it as 'a small, technical assumption for the sake of simplicity'.

Is this justified in practice? Actually it is, because you can prove with probability theory that a random matrix C_n, with a probability tending to 1, has no repeated eigenvalues as the size of the matrix goes to infinity ([DEIFT]). You can also defend this differently: the only reason why we can't use the Perron-Frobenius theorem (see 1.2.10) here, is because M will have zero entries (not all pages in S_σ will be linked to each other, the graph $\mathscr{G}'_\sigma$ is not strongly connected in general). But, it is intuitively clear that by adding 1 to each entry of M, the final results of the algorithm (a sorted vector with the best hubs and authorities) will not be changed at all, because pages with larger indegrees and outdegrees will continue to get better hub and authority scores (note, however, that the relative hub and authority scores can fluctuate a bit and the algorithm will converge slower because of the lack of zero entries). So, by adding 1 to each entry of M, M becomes a positive, real matrix and we know from the Perron-Frobenius that these matrices have a unique dominant eigenvalue. So yes, the 'small, technical assumption' in the paper of Kleinberg is justified.

Now that this problem is solved, we present the relevant theorems below. We also impose on the matrix M that it has a unique dominant eigenvalue with the preceding explanations in mind. Remember that we will generalize the idea of the HITS algorithm to introduce similarity on graphs. Therefore, we will reconsider the convergence of the (generalized) algorithm again in the following section, and there we prove that there exists also a limit even when the matrix M has no unique dominant eigenvalue. The reason why we don't present this result immediately, is because we want to present the results as authentic possible and we want to show the evolution of the ideas in the successive papers.

Theorem 2.1.1. *If A is a symmetric, real $n \times n$-matrix, then it has n (not necessarily different) real eigenvalues corresponding to real eigenvectors.*

[2] The matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is a simple counterexample for a symmetric, nonnegatieve, real matrix that has no unique dominant eigenvalue.

Proof. First, threat A as complex matrix. The characteristic polynomial $\det(A - \lambda I)$ has n roots in $\mathbb{C}$ and each root is an eigenvalue for A. Let $\lambda \in \mathbb{C}$ be any eigenvalue and $\mathbf{v} \in \mathbb{C}^n$ be a corresponding eigenvector for A. We have:

$$A\mathbf{v} = \lambda\mathbf{v}.$$

As $A = A^t$, we also get:

$$\mathbf{v}^t A = \lambda\mathbf{v}^t.$$

Taking the complex conjugate of both sides we get (A is a real matrix):

$$\bar{\mathbf{v}}^t A = \bar{\lambda}\bar{\mathbf{v}}^t$$

We get:

$$\bar{\mathbf{v}}^t A\mathbf{v} = (\bar{\mathbf{v}}^t A)\mathbf{v} = (\bar{\lambda}\bar{\mathbf{v}}^t)\mathbf{v} = \bar{\lambda}\bar{\mathbf{v}}^t\mathbf{v}.$$

We also have:

$$\bar{\mathbf{v}}^t A\mathbf{v} = \bar{\mathbf{v}}^t(A\mathbf{v}) = \lambda\bar{\mathbf{v}}^t\mathbf{v}.$$

Hence:

$$\bar{\lambda}\bar{\mathbf{v}}^t\mathbf{v} = \lambda\bar{\mathbf{v}}^t\mathbf{v}.$$

We conclude that $\lambda = \bar{\lambda}$ for $\mathbf{v} \neq 0$. We proved that every eigenvalue of A is real. If λ is an eigenvalue of A, then the matrix $(A - \lambda I)$ is not invertible so a vector $\mathbf{s} \in \mathbb{R}^n$ exists with

$$(A - \lambda I)\mathbf{s} = 0,$$

proving that also the corresponding eigenvector is real.

□

Theorem 2.1.2. ***(Symmetric Schur Decomposition)*** *Let A be a real symmetric matrix, then there exist an orthogonal matrix P such that:*

(i) $P^{-1}AP = D$, *a diagonal matrix,*

(ii) The diagonal entries of D are the eigenvalues of A,

(iii) The column vectors of P are the eigenvectors of the eigenvalues of A.

Proof. By induction on the order of the matrix. For $n = 1$ the theorem is trivial. Let A be a symmetric $n \times n$-matrix. A has at least one eigenvalue λ_1 by the previous theorem. Let $\mathbf{x_1}$ be a corresponding eigenvalue with $\|\mathbf{x_1}\| = 1$ and $A\mathbf{x_1} = \lambda_1\mathbf{x_1}$. By the Gram-Schmidt procedure, we construct an orthonormal basis $V_1 = \{\mathbf{x_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ of $\mathbb{R}^n$. Let:

$$S_1 = [\mathbf{x_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}],$$

since S_1 is orthonormal, we get $S_1^t = S^{-1}$. Consider the matrix: $S_1^{-1}AS_1$. We have:

$$(S_1^{-1}AS_1)^t = (S_1^tAS_1)^t = S_1^tA^tS_1 = S_1^{-1}AS_1$$

Thus $S_1^{-1}AS_1$ is a symmetric matrix. Since $S_1\mathbf{e_1} = \mathbf{x_1}$, we get:

$$\begin{aligned} S_1^{-1}AS_1\mathbf{e_1} &= (S_1^{-1}A)(\mathbf{x_1}) \\ &= S_1^{-1}(\lambda_1\mathbf{x_1}) \\ &= \lambda_1(S_1^{-1}\mathbf{x_1}) \\ &= \lambda_1\mathbf{e_1} \end{aligned}$$

So we get:

$$S_1^{-1}AS_1 = \left(\begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0}^t & \mathbf{A_1} \end{array}\right),$$

with $\mathbf{0}$ a vector of zero entries of size $n-1$ and A_1 an $(n-1)\times(n-1)$ symmetric matrix. We know by induction that there exist a $(n-1)\times(n-1)$ orthogonal matrix S_2 such that $S_2^{-1}A_1S_2 = D'$ with D' an $(n-1)\times(n-1)$ diagonal matrix. Let:

$$S_2' = \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0}^t & S_2 \end{array}\right),$$

and also S_2' is an orthogonal matrix, we get:

$$\begin{aligned} \left(S_2'\right)^{-1} S_1^{-1}AS_1S_2' &= \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0}^t & S_2^t \end{array}\right)\left(S_1^{-1}AS_1\right)\left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0}^t & S_2 \end{array}\right) \\ &= \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0}^t & S_2^t \end{array}\right)\left(\begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0}^t & A_1 \end{array}\right)\left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0}^t & S_2 \end{array}\right) \\ &= \left(\begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0}^t & S_2^tA_1S_2 \end{array}\right) \\ &= \left(\begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0}^t & D' \end{array}\right) \end{aligned}$$

Thus, if we put

$$\begin{aligned} P &= S_1S_2' \\ D &= \left(\begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0}^t & D' \end{array}\right), \end{aligned}$$

we have proved (1). From the definition of diagonalizable matrices and Theorem 1.3.7 (ii) and (iii) immediately follow. □

Theorem 2.1.3. *Giving a graph $\mathscr{G}$ with n linked pages, the sequence as defined in the previous paragraph:*

$$\mathbf{z}_0 = \mathbf{1} \in \mathbb{R}^n, \mathbf{z}_{k+1} = \frac{M\mathbf{z}_k}{||M\mathbf{z}_k||_2}, \quad k = 0, 1, \ldots,$$

converges when M has a unique dominant eigenvalue.

Proof. Since 1) M is diagonalizable as symmetric matrix by Theorem 2.1.2 and 2) M has a unique dominant eigenvalue, it follows from the power method that the sequence will converge to a corresponding dominating eigenvector $(\mathbf{h}, \mathbf{a})$. This eigenvector contains the hub and authority scores. □

We conclude with a nice corollary.

Corollary 2.1.4. *The second power of the matrix M has the form:*

$$M^2 = \begin{pmatrix} BB^T & 0 \\ 0 & B^TB \end{pmatrix},$$

and the normalized hub and authority scores are given by the dominant eigenvectors of BB^T and B^TB.

Proof. By the compact form given in equation 2.1, we see that $\mathbf{h_k} \leftarrow (BB^T)^{k-1}B\mathbf{a_0}$ and $\mathbf{a_k} \leftarrow (B^TB)^k\mathbf{a_0}$. Let $\mathbf{a_0}$ be $\mathbf{1} \in \mathbb{R}^n$. From the previous theorem we also know that:

$$\lim_{k\to\infty} \mathbf{h_k} = \mathbf{h} \quad \text{and} \quad \lim_{k\to\infty} \mathbf{a_k} = \mathbf{a},$$

and also from the previous proof we know that $(\mathbf{h}, \mathbf{a})$ is the dominant eigenvector of M. It follows immediately that also $\mathbf{h}$ is the dominant eigenvector of BB^T and $\mathbf{a}$ is the dominant eigenvector of B^TB. □

2.1.6 Example

Example 2.1.5. We conclude this section with a fictitious example of the HITS-algorithm. Suppose you are looking for `math professors vub` with a text-based search engine and you get the following results:

- The website of the mathematics department of the VUB,
- The website of the faculty of science of the VUB,
- The websites of 4 math professors,
- The website of 10 PhD students at the the mathematics department of the VUB.

Lets take a look at the link structure of these web pages (remember that it is a fictitious example):

- The website of the mathematics department at the VUB links to the websites of all the 4 professors, the 10 PhD students and the faculty of science,
- The website of the faculty of science of the VUB links to the websites of all the 4 math professors and the mathematics department,
- The websites of the 4 math professors link to the website of the Mathematics department and the faculty of science,
- The websites of the 10 PhD students at the the VUB link to the the website of their promotor. 1 professor has 4 PhD students, the other 3 professors have 2 PhD students.

We can now construct the graph $\mathscr{G}_\sigma$ (of course this graph is not completely made according to Algorithm 2) and we have the following adjacency matrix of $\mathscr{G}_\sigma$:

- **Row 1**: website of the mathematics department,
- **Row 2**: website of the faculty of science,
- **Row 3**: website of the professor with the 4 PhD students,
- **Row 4, 5, 6**: websites of the professors with the 2 PhD students,
- **Row 7, 8, 9, 10**: websites of the 4 Phd students of the professor on row 3,
- **Row 11, 12**: websites of the 2 Phd students of the professor on row 4,

- **Row 12, 14**: websites of the 2 Phd students of the professor on row 5,
- **Row 14, 16**: websites of the 2 Phd students of the professor on row 6,

Leads to:

$$B = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

Intuitively, we expect - when applying the HITS-method - that the professor with his 4 PhD student will have the largest authority score, immediately followed by the other 3 professors. The website of the mathematics department is clearly the best hub in this example and should get the largest hub score. Also the website of the faculty of science should get a high hub score.

We now apply the HITS-method by calculating the dominant eigenvector of BB^T (this returns the hub scores) and the dominant eigenvector of B^TB (this returns the authority scores) with the power method (see 1.3.2). We get:

$$\mathbf{a} = \begin{pmatrix} 0.1979 \\ 0.3162 \\ \mathbf{0.3688} \\ 0.3231 \\ 0.3231 \\ 0.3231 \\ 0.2029 \\ 0.2029 \\ 0.2029 \\ 0.2029 \\ 0.2029 \\ 0.2029 \\ 0.2029 \\ 0.2029 \\ 0.2029 \\ 0.2029 \end{pmatrix} \quad \text{and} \quad \mathbf{h} = \begin{pmatrix} \mathbf{0.8645} \\ 0.3605 \\ 0.1207 \\ 0.1207 \\ 0.1207 \\ 0.1207 \\ 0.0866 \\ 0.0866 \\ 0.0866 \\ 0.0866 \\ 0.0758 \\ 0.0758 \\ 0.0758 \\ 0.0758 \\ 0.0758 \\ 0.0758 \end{pmatrix}$$

We see that the websites of the 4 math professors are indeed the best authorities for the search query `math professors vub` and that the website of the mathematics department is an extremely good hub (this is very logic because it links to all the other relevant websites). The professor with his 4 PhD students would be ranked first in the search results (he has the highest authority score), the other professors would appear just underneath him. Obviously, a hub score of 0.8645 is so high that it would be quite exceptional in a graph containing a lot more websites (it's very unlikely that you find a website containing links to all the other pages/nodes in the graph). Nevertheless, we conclude that the HITS-algorithm returns the results we wanted intuitively.

2.1.7 Final reflection

The HITS-algorithm is one of the few algorithms that has the ability to rank pages according to a specific search query. Also the computational cost of the HITS-algorithm, which equals the cost of the power method (see 1.3.2), is not excessive and feasible for most servers. The result of the HITS-algorithm for popular queries will also be cached by most search engines, which reduces the computational cost even more because the saved results can be served directly to the user without any new calculations.

The biggest disadvantage of the HITS-algorithm is that it suffers from *topic drift*: the graph $\mathscr{G}_\sigma$ could contain nodes which have high authority scores for the query but are completely irrelevant. E.g. `Facebook` is nowadays a universally popular website, almost every website contains a 'like' or 'share' button linking to Facebook, and Facebook itself contains tons of posts linking to other webpages. This means that Facebook has a great chance to appear in almost any $\mathscr{G}_\sigma$ and receive a high authority score because the original HITS-algorithm as presented here cannot detect such 'universally popular' websites. The same goes for other social media websites and some advertisements.

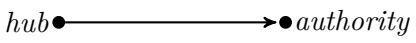
Nowadays, we know that `Ask.com` uses this algorithm. In fact, most search engines are very secretive about their search algorithm (e.g. `Google`) to make profit and avoid cheating by webmasters. Still, the chances are that other search engines use some variant of the algorithm as well, in combination with a lot of other procedures.

2.2 Similarity on directed graphs

This section summarizes the paper '*A Measure of Similarity between Graph Vertices: Applications to Synonym Extraction and Web Searching*'[BLONDEL] of V. D. Blondel and others. The paper generalizes the HITS-algorithm leading to the concept of similarity on directed graphs. This concept is explained in detail and far more mathematically rigorous than in the previous section. Recall our assumption in the previous section stating that the matrix M has to have a unique dominant eigenvalue. Although this is satisfactory in practical examples, we want to construct a concept that works for all types of directed graphs, also the ones leading to matrices with dominant eigenvalues with multiplicity more than 1 and therefore we will develop a method to work around this situation. The section concludes with a selfmade example.

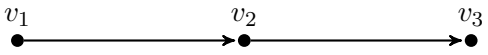
From Hubs and Authorities to structure graphs

We generalize the construction of the previous section. Remember that we constructed a graph $\mathscr{G}'_\sigma$ and calculated hub and authorities scores for each vertex. Now, for any directed graph $\mathscr{G}$, the authority score of a vertex v_j of $\mathscr{G}$ can be thought of as a similarity score between v_j of $\mathscr{G}$ and vertex *authority* of the graph



and, conversely, the hub score of vertex v_j of $\mathscr{G}$ can be thought of as a similarity score between v_j and vertex *hub*. We call the hub-authority graph a *structure graph* and we already know the resulting iterative method from the previous section.

The central question is now: which *mutually reinforcing relation* (iterative method) do we get when using another structure graph, different from the hub-authority structure graph? We start with an example. In our example, we use as structure graph a path graph with three vertices v_1, v_2, v_3.



Let $\mathscr{G}(W, \rightarrow)$ be a graph. With each vertex w_i of $\mathscr{G}$ we now associate three scores x_{i1}, x_{i2} and x_{i3}, one for each vertex of the structure graph. We initialize these scores with a positive value and then update them according to the mutually reinforcing relation:

$$\begin{cases} x_{i1} := & \sum_{j:(w_i,w_j)\in\rightarrow} x_{j2}, \\ x_{i2} := \sum_{j:(w_j,w_i)\in\rightarrow} x_{j1} & + \sum_{j:(w_i,w_j)\in\rightarrow} x_{j3}, \\ x_{i3} := & \sum_{j:(w_j,w_i)\in\rightarrow} x_{j2}, \end{cases}$$

or, in matrix form ($\mathbf{x_j}$ denotes the column vector with entries x_{ij}, B is the adjacency matrix of graph $\mathscr{G}$),

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}_{k+1} = \begin{pmatrix} 0 & B & 0 \\ B^T & 0 & B \\ 0 & B^T & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}_k$$

which we, again, can denote by $\mathbf{x_{k+1}} = M\mathbf{x_k}$. The principle is now exactly the same as the previous example with hubs and authorities. The matrix M is symmetric and nonnegative, and again the results is the limit of the normalized vector sequence:

$$\mathbf{z}_0 = \mathbf{x}^{(0)} > 0, \mathbf{z}_{k+1} = \frac{M\mathbf{z}_k}{||M\mathbf{z}_k||_2}, \quad k = 0, 1, \ldots, \tag{2.3}$$

Remember that the HITS-algorithm assumed that M as a unique dominant eigenvalue but we don't want to make this assumption in this section because we want a concept that can be applied to all kind of directed graphs. We will see that without this assumption, the sequence 2.3 does not always converge but oscillates between the limits:

$$\mathbf{z}_{\text{even}} = \lim_{k\to\infty} \mathbf{z}_{2k} \quad \text{and} \quad \mathbf{z}_{\text{odd}} = \lim_{k\to\infty} \mathbf{z}_{2k+1}$$

The limit vectors $\mathbf{z}_{\text{even}}$ and $\mathbf{z}_{\text{odd}}$ do in general depend on the initial vector $\mathbf{z}_0$. The set of all limit vectors obtained when starting from a positive initial vector is given by:

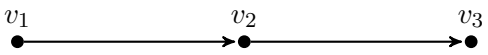
$$Z = \{\mathbf{z}_{\text{even}}(\mathbf{z}_0), \mathbf{z}_{\text{odd}}(\mathbf{z}_0) : \mathbf{z}_0 > 0\},$$

and we would like to select a particular vector in that set. The vector $\mathbf{z}_0 = \mathbf{1}$ is a good choice: it's the unique vector with the largest possible Manhattan[3] (we will prove this in Theorem 2.2.12) norm and makes computations easy. We denote $z_{\text{even}}(\mathbf{1})$.

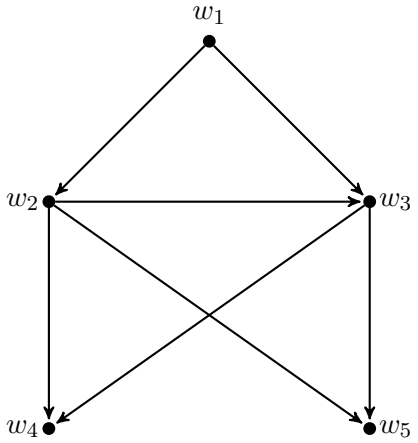
The three components of the extremal limit $z_{\text{even}}(\mathbf{1})$ are now defined as the *similarity scores* $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$ and the *similarity matrix* is defined by $S = [\mathbf{s}_1\ \mathbf{s}_2\ \mathbf{s}_3]$. We will also prove in **??** that $\mathbf{s}_2$ can be obtained from B by calculating the dominant eigenvector of the matrix $BB^T + B^TB$.

We now give a numerical example.

Example 2.2.1. Take as structure graph again the path graph with three vertices v_1, v_2, v_3:



Let $\mathscr{G}(W, \to)$ be the following graph:



Then the adjacency matrix B is:

$$B = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

By using the described mutually reinforcing updating iteration we become the following similarity matrix (a numerical algorithm to calculate this is presented later on in this section together with some proofs that facilitate the calculation):

$$S = \begin{pmatrix} 0.4433 & 0.1043 & 0 \\ 0.2801 & 0.3956 & 0.0858 \\ 0.0858 & 0.3956 & 0.2801 \\ 0.2216 & 0.0489 & 0.2216 \\ 0 & 0.1043 & 0.4433 \end{pmatrix}$$

[3] $\|\boldsymbol{x}\|_1 := \sum_{i=1}^{n} |x_i|$ for $\mathbf{x} \in \mathbb{R}^n$is the Manhattan norm.

The similarity score of w_4 with v_2 of the structure graph is equal to 0.0489.

We now construct the general case. Take two directed graphs $\mathscr{G}(U, \to)$ and $\mathscr{H}(V, \to')$ with $n_{\mathscr{G}}$ and $n_{\mathscr{H}}$ the order of the graphs. We think of $\mathscr{G}$ as the structure graph (such as the graphs hub $\to$ authority and the graph $1 \to 2 \to 3$ in the previous paragraphs). We get the following mutually reinforcing updating iteration with the following updating equations:

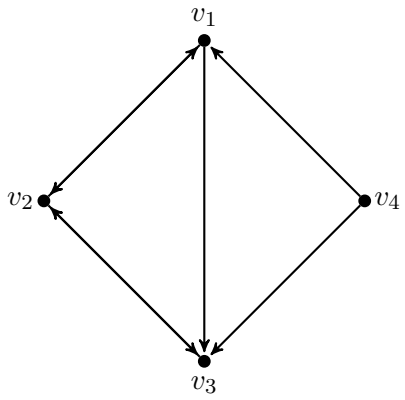
$$x_{ij} := \sum_{r:(v_r,v_i)\in\to', s:(u_s,u_j)\in\to} x_{rs} + \sum_{r:(v_i,v_r)\in\to', s:(u_j,u_s)\in\to} x_{rs}$$

Consider the product graph $\mathscr{G} \times \mathscr{H}$ (see Definition 1.4.11). The above updating equation is equivalent to replacing the scores of all vertices of the product graph by the sum of the scores of the vertices linked by an incoming or outgoing edge. Equation 2.3 can also be rewritten in a a more compact matrix form. Let X_k be the $n_{\mathscr{H}} \times n_{\mathscr{G}}$ matrix of entries x_{ij} at iteration k, and A and B are the adjacency matrices of $\mathscr{G}$ and $\mathscr{H}$. Then the updating equations can be written as:

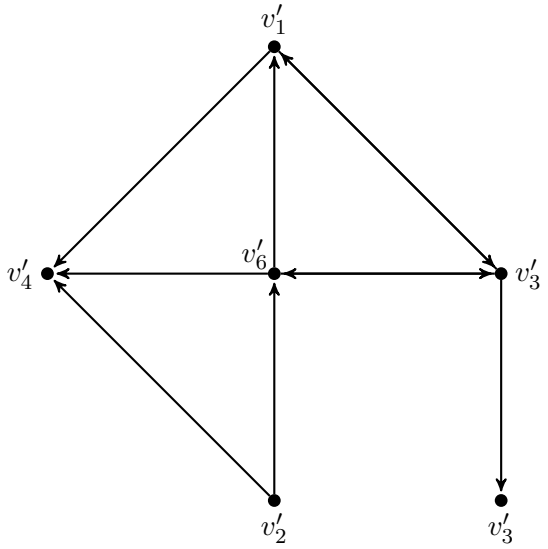
$$X_{k+1} = BX_kA^T + B^TX_kA, \quad k = 0, 1, \ldots, \tag{2.4}$$

We'll prove that the normalized even and odd iterates of this updating equation converge and that the limit $z_{\text{even}}(\mathbf{1})$ is the limit with the largest Manhattan norm. This limit is the definition of the similarity matrix. The following example shows a calculated similarity matrix of two directed graphs.

Example 2.2.2. Let $\mathscr{G}_A(V, \to)$ be the following graph:



Let $\mathscr{G}_B(V', \to')$ be the following graph:



We become the following similarity matrix (a numerical algorithm to calculate this matrix is introduced later in this section):

$$S = \begin{pmatrix} 0.2636 & 0.2786 & 0.2723 & 0.1289 \\ 0.1286 & 0.1286 & 0.0624 & 0.1268 \\ 0.2904 & 0.3115 & 0.2825 & 0.1667 \\ 0.1540 & 0.1701 & 0.2462 & 0 \\ 0.0634 & 0.0759 & 0.1018 & 0 \\ 0.3038 & 0.3011 & 0.2532 & 0.1999 \end{pmatrix}$$

We see for example, that vertex v_2 of $\mathscr{G}_A$ is most similar to vertex v_3' in $\mathscr{G}_B$ because the similarity score s_{32} is the highest among the similarity scores in s_2.

2.2.1 Convergence of the sequence $\mathbf{z}_k$

In the introduction, we mentioned already that the sequence in Equation 2.3 converges for even and odd iterates. We will prove this add the end of this subsection. But before we arrive there, we first need some results on the eigenvectors and eigenvalues of nonnegative matrices. The Perron-Frobenius applies only to nonnegative, *irreducible* matrices, but we will prove in Theorem 2.2.9 that also nonnegative matrices M have a Perron root (see Definition 1.2.11) that has an associated Perron vector. We will also investigate more specific results in the case M is not only nonnegative, but also symmetric.

The reason why we prove all this is clear: remember that in the previous chapter we assumed that M has a unique dominant eigenvalue and showed convergence on the basis of the Power Method. This approach is far to naive when introducing similarity on graph vertices, because graphs can lead to matrices where there is no unique dominant eigenvalue (for instance cycle graphs will lead to such an M in combination with the hub→authority structure graph). To solve this, a more profound mathematical analysis of the concept is needed, which we present here.

We first want to prove the spectral radius formula, but before we can handle this we have to introduce matrix norms and show that on finite dimensional vector spaces they are all

equivalent. Next, two lemmas are presented which will also contribute to the proof of the spectral radius formula.

Definition 2.2.3. *A function $\|.\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ is called a **matrix norm** if the following conditions hold ($\|A\|$ denotes the norm of the matrix A):*

1. $\|A\| \geq 0$,
2. $\|A\| = 0$ *if and only if* $A = 0$,
3. $\|A\| = |\alpha|\|A\|$ *for all* $\alpha \in \mathbb{C}$,
4. $\|A + B\| \leq \|A\| + \|B\|$ *for all matrices* $A, B \in \mathbb{C}^{m\times n}$

Theorem 2.2.4. ***(Equivalence of norms)*** *For any two matrix norms $\|.\|_\alpha$ and $\|.\|_\beta$ we have:*

$$s\|A\|_\alpha \leq \|A\|_\beta \leq t\|A\|_\alpha \quad \text{with } A \in \mathbb{R}^{m\times n}, r, s \in \mathbb{R}^+$$

Proof. This theorem holds for any finite dimensional vector space, but we focus only on the real case. It is straightforward to see that the equivalence of norms is an equivalence relation. Thus, it is sufficient to show that every norm on $\mathbb{R}^{m\times n}$ is equivalent to the Euclidean matrix norm: **nog onvolledig** □

Lemma 2.2.5. *Let A be an $n \times n$ matrix and let $\|.\|$ be a matrix norm then:*

$$\rho(A) \leq \|A\|$$

Proof. Let λ be an eigenvalue of A and let $\mathbf{x} \neq 0$ be a corresponding eigenvector. From $Ax = \lambda x$, we have:

$$AX = \lambda X, \quad \text{where } X = \begin{bmatrix} \mathbf{x} & \dots & \mathbf{x} \end{bmatrix} \in \mathbb{R}^{n\times n} \setminus \{0\}.$$

It follows that:

$$|\lambda|\|X\| = \|\lambda X\| = \|AX\| \leq \|A\|\|X\|,$$

and simplifying by $\|X\|(> 0)$ gives:

$$|\lambda| \leq \|A\|.$$

This holds for all eigenvalues, taking the maxim over all these eigenvalues gives the result. □

VRAAG: De Jordannormaalvorm wordt gezien als voorkennis, dit werd ook (weliswaar zeer kort) behandeld bij professor Kieboom in het eerste jaar, mag ik ervan uitgaan dat dit voorkennis is of moet dit toch een sectie worden bij het eerste hoofdstuk?

Lemma 2.2.6. *Let A be an $n \times n$ matrix and $\epsilon > 0$, there exist a matix norm $\|.\|$ such that:*

$$\|A\| \leq \rho(A) + \epsilon$$

Proof. The Jordan canonical form of A is:

$$A = S \begin{bmatrix} J_{n_1}(\lambda_1) & 0 & \dots & 0 \\ 0 & J_{n_2}(\lambda_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & J_{n_k}(\lambda_k) \end{bmatrix} S^{-1},$$

where $S \in \mathbb{R}^{n\times n}$ is an invertible matrix, $\lambda_1, \dots, \lambda_k$ are the eigenvalues of A and $n_1 + \dots + n_k = n$. Let:

$$D(\eta) = \begin{bmatrix} D_{n_1}(\eta) & 0 & \dots & 0 \\ 0 & D_{n_2}(\eta) & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ 0 & \dots & 0 & D_{n_k}(\eta) \end{bmatrix} \quad \text{with } D_m(\eta) = \begin{bmatrix} \eta & 0 & \dots & 0 \\ 0 & \eta^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \eta^m \end{bmatrix}$$

Since the left multiplication by $D_m(1/\epsilon)$ multiplies the ith row by $1/\epsilon^i$ and the right multiplication on the right by $D_m(\eta)$ multiplies the jth colum by ϵ^j, we calculate:

$$D(1/\epsilon)S^{-1}ASD(\epsilon) = \begin{bmatrix} B_{n_1}(\lambda_1, \epsilon) & 0 & \dots & 0 \\ 0 & B_{n_2}(\lambda_2, \epsilon) & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ 0 & \dots & 0 & B_{n_k}(\lambda_k, \epsilon) \end{bmatrix}$$

with

$$B_m(\lambda, \epsilon) = D_m(1/\epsilon)J_m(\lambda)D_m(\epsilon) = \begin{bmatrix} \lambda & \epsilon & 0 & \dots & 0 \\ 0 & \lambda & \epsilon & 0 & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \lambda & \epsilon \\ 0 & \dots & 0 & 0 & \lambda \end{bmatrix}$$

We now define the matrix norm for $M \in \mathbb{R}^{n\times n}$ by:

$$\|M\| = \max_{\|\mathbf{x}\|_1=1} \|D(1/\epsilon)S^{-1}MSD(\epsilon)\mathbf{x}\|_1 \tag{2.5}$$

$$= \max_{l\in[1:n]} \sum_{k=1}^{n} |(D(1/\epsilon)S^{-1}MSD(\epsilon))_{k,l}|. \tag{2.6}$$

The conditions for being a matrix norm are trivially met because we know that $\max_{\|\mathbf{x}\|_1=1} \|A\mathbf{x}\|$ is a matrix norm for any A in $\mathbb{C}^{n\times n}$. □

Theorem 2.2.7. ***Spectral radius formula*** *Let a be an $n \times n$ matrix and let $\|.\|$ be a matrix norm then:*

$$\rho(A) = \lim_{k\to\infty} \|A^k\|^{1/k}$$

Proof. Given $k \geq 0$, we use Lemma 2.2.5 to write:

$$\rho(A)^k = \rho(A^k) \leq \|A^k\|,$$

so:

$$\rho(A) \leq \|A^k\|^{1/k}.$$

Taking the limit as $k \to \infty$ gives $\rho(A) \leq \lim_{k\to\infty} \|A^k\|^{1/k}$. To establish the reverse inequality, we need to prove that, for any $\epsilon > 0$, there exists a $K \geq 0$ such that $\|A^k\|^{1/k} \leq \rho(A)+\epsilon$ for all $k \geq K$. From Lemma 2.2.6, we know that there exists a matrix norm $\|.\|$ so $\|A| \leq \rho(A)+\epsilon/2$. Moreover, by the equivalence of the norms on $\mathbb{R}^{n\times n}$, we know that there exists some constant $C > 0$ such that $\|M\| \leq C\|M\|$ for all $M \in \mathbb{R}^{n\times n}$. Then, for any $k \geq 0$,

$$\begin{aligned} \|A^k\| &\leq C\|A^k\| \leq C\|A\|^k \leq C(\rho(A) + \epsilon/2)^k, \\ \|A^k\|^{1/k} &\leq C^{1/k}(\rho(A) + \epsilon/2) \to_{k\to\infty} \rho(A) + \epsilon/2 \end{aligned}$$

This implies the existence of $K \geq 0$ such that $\|A^k\|^{1/k} \leq \rho(A) + \epsilon$ for $k \geq K$, as desired. □

Now we are ready for are first big result: since one is confronted in practice with nonnegative matrices that are not necessary irreducible, we extend the Perron-Frobenius and see what remains without this assumption. We first start with a lemma and next we proof that the spectral radius $\rho(M)$ of a nonnegative matrix M is an eigenvalue of M, the Perron root. Moreover, there exists an associated nonnegative eigenvector $\mathbf{x} \geq 0(\mathbf{x} \neq 0)$, the Perron vector, such that $M\mathbf{x} = \rho\mathbf{x}$.

Lemma 2.2.8. *Let A, B be $n \times n$-matrices, if $|A| \leq B$, then $\rho(A) \leq \rho(|A|) \leq \rho(B)$. (See Definition 1.1.8 for the definition of $|.|$).*

Proof. For every $m = 1, 2, \ldots$ we have

$$|A^m| \leq |A|^m \leq B^m$$

by using some trivial properties of the absolute value function. Let $||.||_2$ be the matrix 2-norm induced by the Euclidean vector norm: for any matrix M, we have $\|M\|_2 = \max_{\|\mathbf{x}\|_2=1} \|M\mathbf{x}\|_2$. For this matrix norm it is trivial to see that if $|M| \leq |M'|$ (see Definition 1.1.5) it follows that $\|M\|_2 \leq \|M'\|_2$ and also $\|M\|_2 = \| \, |M| \, \|_2$, we get:

$$\|A^m\|_2 \leq \| \, |A|^m \, \|_2 \leq \|B^m\|_2$$

and

$$\|A^m\|_2^{1/m} \leq \| \, |A|^m \, \|_2^{1/m} \leq \|B^m\|_2^{1/m}$$

for all $m = 1, 2, \ldots$ If we now let $m \to \infty$ and apply the spectral radius formula from Theorem 2.2.7 we get:

$$\rho(A) \leq \rho(\|A\|) \leq \rho(B).$$

□

Theorem 2.2.9. *If $A \geq 0$ is an $n \times n$-matrix, then $\rho(A)$ is an eigenvalue of A and there is a nonnegative vector $\mathbf{x} \geq 0, \mathbf{x} \neq 0$, such that $A\mathbf{x} = \rho(A)\mathbf{x}$.*

Proof. For any $\epsilon > 0$ define $A(\epsilon) = [a_{ij} + \epsilon] > 0$. Denote by $\mathbf{x}(\epsilon)$ the Perron vector of $A(\epsilon)$, so $\mathbf{x}(\epsilon) > 0$ and $\sum_{i=1}^n \mathbf{x}(\epsilon)_i = 1$. Since the set of vectors $\{\mathbf{x}(\epsilon) : \epsilon > 0\}$ is contained in the compact set $\{x : x \in \mathbb{C}^n, \|\mathbf{x}\|_1 \leq 1\}$, there is a monotone decreasing sequence $\epsilon_1, \epsilon_2, \ldots$ with

$\lim_{k\to\infty} \epsilon_k = 0$ such that $\lim_{k\to\infty} \mathbf{x}(\epsilon_k) = \mathbf{x}$ exists. Since $\mathbf{x}(\epsilon_k) > 0$ for all $k = 1, 2, \ldots$, it must be that $\mathbf{x} = \lim_{k\to\infty} \mathbf{x}(\epsilon_k) \geq 0$; $\mathbf{x} = 0$ is impossible because:

$$\sum_{i=1}^{n} \mathbf{x}_i = \lim_{k\to\infty} \sum_{i=1}^{n} \mathbf{x}(\epsilon_k)_i = 1$$

By Lemma 2.2.8, $\rho(A(\epsilon_k)) \geq \rho(A(\epsilon_{k+1}) \geq \ldots \geq \rho(A)$ for all $k = 1, 2, \ldots$, so the sequence $\{\rho(A(\epsilon_k))\}_{k=1,2,\ldots}$ is a monotone decreasing sequence. Thus, $\rho = \lim_{k\to\infty} \rho(A(\epsilon_k))$ exists and $\rho \geq \rho(A)$. From the fact that

$$\begin{aligned} A\mathbf{x} &= \lim_{k\to\infty} A(\epsilon_k)\mathbf{x}(\epsilon_k) \\ &= \lim_{k\to\infty} \rho(A(\epsilon_k))\mathbf{x}(\epsilon_k) \\ &= \lim_{k\to\infty} \rho(A(\epsilon_k)) \lim_{k\to\infty} \mathbf{x}(\epsilon_k) \\ &= \rho\mathbf{x}, \end{aligned}$$

and the fact that $\mathbf{x} \neq 0$ we conclude that ρ is an eigenvalue of A. But then $\rho \leq \rho(A)$ so it must be that $\rho = \rho(A)$. □

Now that we know that any nonnegative matrix M has it's spectral radius as an eigenvalue and there exists an associated nonnegative eigenvector, we will see if we can get more specific results when handling nonnegative, symmetric matrices.

Theorem 2.2.10. *Let M be a symmetric nonnegative matrix with spectral radius ρ. Then the algebraic and geometric multiplicity of the Perron root ρ are equal; there is a nonnegative matrix X whose columns span the invariant subspace associated with the Perron root; and the elements of the orthogonal projector Π on the vector space associated with the Perron root of M are all nonnegative.*

Proof. We know that any symmetric nonnegative matrix M can be permuted to a Jordan cononical form with irreducible blocks M_i on the diagonal. We also know from the Perron-Frobenius theorem (see 1.2.10) that the algebraic multiplicity of the Perron root of an irreducible nonnegative matrix is equal to 1. It follows from both facts that the algebraic multiplicities and the geometric multiplicities of the Perron root ρ of M are equal. The corresponding eigenspace of M is obtained from the normalized Perron vectors of the M_i blocks padded with zeros. This normalized Perron vectors form a basis X which is nonnegative and orthonormal. □

We are almost ready for the theorem about the convergence of $\mathbf{z}_k$, but we first need an easy result about the orthogonal projection.

Theorem 2.2.11. *Let $\mathcal{V}$ be a linear subspace of $\mathbb{R}^n$ with orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$. Arrange the column vectors $\mathbf{v}_i$ in a matrix V and let $\mathbf{x} \in \mathbb{R}^n$. The **orthogonal projection** of $\mathbf{x}$ on $\mathcal{V}$ is then given by:*

$$\Pi\mathbf{x} = VV^T\mathbf{x},$$

*the matrix $\Pi = VV^T$ is the **orthogonal projector**. Projectors have the property that $\Pi^2 = \Pi$*

Proof. [4] We use the connection between transposes and the standard inner product to find the matrix orthogonal projection on the subspace $\mathcal{V}$:

$$\begin{aligned}\text{proj}_{\mathcal{V}}\mathbf{x} &= \sum_{i=1}^{m}\frac{\langle \mathbf{v}_i, \mathbf{x}\rangle}{\|\mathbf{v}_i\|}\mathbf{v}_i \\ &= \sum_{i=1}^{m}\langle \mathbf{v}_i, \mathbf{x}\rangle \mathbf{v}_1 \\ &= \sum_{i=1}^{m}\mathbf{v}_i\langle \mathbf{v}_i, \mathbf{x}\rangle\end{aligned}$$

Remember that $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^T\mathbf{y}$ so:

$$\begin{aligned}&= \sum_{i=1}^{m}\mathbf{v}_i(\mathbf{v}_i^T\mathbf{x}) \\ &= \sum_{i=1}^{m}(\mathbf{v}_i\mathbf{v}_i^T)\mathbf{x} \\ &= VV^T\mathbf{x}\end{aligned}$$

Proving that $\Pi^2 = \Pi$ is also trivial, remember that for an orthogonal matrix A it holds that $A^T = A^{-1}$:

$$\begin{aligned}\Pi^2 &= (VV^T)^2 \\ &= VV^TVV^T \\ &= VV^{-1}VV^T \\ &= I_nVV^T \\ &= VV^T \\ &= \Pi\end{aligned}$$

□

Theorem 2.2.12. *Let M be a symmetric nonnegative, real matrix of spectral radius ρ. Let $z_0 > 0$ and consider the sequence*

$$\mathbf{z}_{k+1} = \frac{M\mathbf{z}_k}{||M\mathbf{z}_k||_2}, k = 0, 1, \ldots$$

Two convergence cases can occur depending on whether or not $-\rho$ is an eigenvalue of M. When $-\rho$ is not an eigenvalue of M, then the sequence of z_k simply converges to $\frac{\Pi\mathbf{z}_0}{||\Pi\mathbf{z}_0||_2}$, where Π is the orthogonal projector on the invariant subspace associated with the Perron root ρ. When $-\rho$ is an eigenvalue of M, then the subsequences $\mathbf{z}_{2k}$ and $\mathbf{z}_{2k+1}$ converge to the limits

$$\mathbf{z}_{even}(z_0) = \lim_{k\to\infty}\mathbf{z}_{2k} = \frac{\Pi\mathbf{z}_0}{||\Pi\mathbf{z}_0||_2} \quad \textit{and} \quad \mathbf{z}_{odd}(\mathbf{z}_0) = \lim_{k\to\infty}\mathbf{z}_{2k+1} = \frac{\Pi M\mathbf{z}_0}{||\Pi M\mathbf{z}_0||_2}.$$

and the vector $\mathbf{z}_{even}(\mathbf{1})$ is the unique vector of largest possible Manhattan norm in that set.

[4] $\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{i=1}^{n} x_iy_i$ with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is the standard inner product of two vectors in $\mathbb{R}^n$.

Proof. We prove only the case where $-\rho$ is an eigenvalue, the other case is an easy adaption. Denote the invariant subspaces of M corresponding to ρ, $-\rho$ and to the rest of the spectrum, respectively by $\mathcal{V}_\rho, \mathcal{V}_{-\rho}$ and $\mathcal{V}_\mu$. From the previous theorem, we know that $\mathcal{V}_\rho, \mathcal{V}_{-\rho}$ are certainly nontrivial (ρ and $-\rho$ have at least multiplicity 1, so the eigenspace contains also at least 1 vector), also assume that $\mathcal{V}_\mu$ is nontrivial (if $\mathcal{V}_\mu$ would be trivial, the rest of the proof gets only easier). We have:

$$MV_\rho = \rho V_\rho, \quad MV_{-\rho} = -\rho V_{-\rho}, \quad MV_\mu = V_\mu M_\mu,$$

where M_μ is a square matrix (diagonal if V_μ is a basis of eigenvectors of the rest of the spectrum, see Theorem 2.1.2) with spectral radius μ strictly less than ρ.

Remember from Theorem 2.1.2 that $P^{-1}MP = D$, with D a diagonal matrix, this can be rewritten as $M = PDP^{-1}$ or $M = PDP^T$ (P is an orthogonal matrix), we can rewrite this in this case as (this is the so called *eigen decomposition* for symmetric matrices):

$$\begin{aligned} M &= \begin{bmatrix} V_\rho & V_{-\rho} & V_\mu \end{bmatrix} \begin{bmatrix} \rho I & & \\ & -\rho I & \\ & & M_\mu \end{bmatrix} \begin{bmatrix} V_\rho & V_{-\rho} & V_\mu \end{bmatrix}^T \\ &= \rho V_\rho V_\rho^T - \rho V_{-\rho} V_{-\rho}^T + V_\mu M_\mu V_\mu^T \end{aligned}$$

It then follows that:

$$M^2 = \rho^2 \Pi + V_\mu M_\mu^2 V_\mu^T,$$

where

$$\Pi = V_\rho V_\rho^T + V_{-\rho} V_{-\rho}^T$$

is the orthogonal projector onto the invariant subspace $\mathcal{V}_\rho \oplus \mathcal{V}_{-\rho}$ of M^2 corresponding to ρ^2. We also have:

$$M^{2k} = \rho^{2k} \Pi + V_\mu M_\mu^{2k} V_\mu^T,$$

and since $\rho M_\mu = \mu < \rho$, it follows from multiplying this by $\mathbf{z}_0$ and $M\mathbf{z}_0$ that:

$$\mathbf{z}_{2k} = \frac{\Pi \mathbf{z}_0}{\|\Pi \mathbf{z}_0\|_2} + O\left(\frac{\mu}{\rho}\right)^{2k}$$

and

$$\mathbf{z}_{2k+1} = \frac{\Pi M \mathbf{z}_0}{\|\Pi M \mathbf{z}_0\|_2} + O\left(\frac{\mu}{\rho}\right)^{2k},$$

provided $\Pi\mathbf{z}_0$ and $\Pi M\mathbf{z}_0$ are nonzero, so the initial vectors $\mathbf{z}_0$ and $M\mathbf{z}_0$ must have a nonzero component. This is no problem, because the Euclidean norms of these vectors equal $\mathbf{z}_0^T \Pi \mathbf{z}_0$ and $\mathbf{z}_0^T M \Pi M \mathbf{z}_0$ since $\Pi^2 = \Pi$. These norms are both nonzero since $\mathbf{z}_0 > 0$ and both Π and $M\Pi$ are nonnegative and nonzero.

From the fact that M is nonnegative and the formula for $\mathbf{z}_{\text{even}}(\mathbf{z}_0)$ and $\mathbf{z}_{\text{odd}}(\mathbf{z}_0)$ we conclude that both limits lie in $\{\Pi\mathbf{z}/\|\Pi\mathbf{z}\|_2 : \mathbf{z} > 0\}$. We now show that every element $\tilde{\mathbf{z}}_0 \in \{\Pi\mathbf{z}/\|\Pi\mathbf{z}\|_2 : \mathbf{z} > 0\}$ can be obtained as $\mathbf{z}_{\text{even}}(\mathbf{z}_0)$ for some $\mathbf{z}_0 > 0$. Since the entries of Π are nonnegative, so are those of $\tilde{\mathbf{z}}_0$. This vector may have some zero entries. From $\tilde{\mathbf{z}}_0$ we construct$\mathbf{z}_0$ by adding $\epsilon > 0$ to all the zero entries of $\tilde{\mathbf{z}}_0$. The vector $\mathbf{z}_0 - \tilde{\mathbf{z}}_0$ is clearly orthogonal to $\mathcal{V}_\rho \oplus \mathcal{V}_{-\rho}$ and will vanish in the iteration of M^2. Thus we have $\mathbf{z}_{\text{even}}(\mathbf{z}_0) = \tilde{\mathbf{z}}_0$ for $\mathbf{z}_0 > 0$, proving our statement.

We now prove the last statement. The matrix Π and all vectors are nonnegative and $\Pi^2 = \Pi$ and so:

$$\left\| \frac{\Pi \mathbf{1}}{\|\Pi \mathbf{1}\|_2} \right\|_1 = \sqrt{\mathbf{1}^T \Pi^2 \mathbf{1}}$$

and also:

$$\left\| \frac{\Pi \mathbf{z}_0}{\|\Pi \mathbf{z}_0\|_2} \right\|_1 = \frac{\mathbf{1}^T \Pi^2 \mathbf{z}_0}{\sqrt{\mathbf{z}_0^T \Pi^2 \mathbf{z}_0}}$$

We apply the Cauchy-Schwarz inequality to $\Pi\mathbf{1}$ and $\Pi\mathbf{z}_0$, remember that $\Pi = WW^T$ for some W, so $\Pi^T = (WW^T)^T = (W^T)^T W^T = WW^T = \Pi$, that the matrix Π and all vectors are nonnegative and $\Pi^2 = \Pi$:

$$\begin{aligned}
|\langle \Pi\mathbf{1}, \Pi\mathbf{z}_0 \rangle| &\leq \|\Pi\mathbf{1}\|.\|\Pi\mathbf{z}_0\| \\
|(\Pi\mathbf{1})^T \Pi\mathbf{z}_0| &\leq \sqrt{\mathbf{1}^T\Pi^2\mathbf{1}}.\sqrt{\mathbf{z}_0^T\Pi^2\mathbf{z}_0} \\
|\mathbf{1}^T\Pi^T\Pi\mathbf{z}_0| &\leq \sqrt{\mathbf{1}^T\Pi^2\mathbf{1}}.\sqrt{\mathbf{z}_0^T\Pi^2\mathbf{z}_0} \\
|\mathbf{1}^T\Pi^2\mathbf{z}_0| &\leq \sqrt{\mathbf{1}^T\Pi^2\mathbf{1}}.\sqrt{\mathbf{z}_0^T\Pi^2\mathbf{z}_0},
\end{aligned}$$

equality arises only when $\Pi\mathbf{z}_0 = \lambda\Pi\mathbf{1}$ for some $\lambda \in \mathbb{C}$. But since $\Pi\mathbf{z}_0$ and $\Pi\mathbf{1}$ are both real nonnegative, the last statement is proved.

□

De rest van de thesis laat ik even weg door enerzijds een paar compilatieproblemen en anderzijds nog onvolkomendheden.

Bibliography

[EVES] H. Eves, *Elementary Matrix Theory*, Dover Publications, 2012.

[BAPAT] . B. Bapat, T.E.. Raghavan, *Nonnegative Matrices and Applications*, Encyclopedia of Mathematics and its Applications 64, Cambridge University Press, 1997.

[STERN2010] S. Sternberg, *The Perron-Frobenius theorem*, Chapter 9 in Dynamical Systems, Dover Publications, 2010.

[NOUT2008] D. Noutsos, *Perron-Frobenius theory and some extensions* (lecture notes), Department of Mathematics, University of Ioannina, May 2008.

[VARGA] . S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, 1962.

[1] . A. Brualdi, H. J. Ryser, *Combinatorial Matrix Theory*, Cambridge University Press, 1991

[2] . Lancaster, M. Tismenetsky, *The Theory of Matrices: With Applications*, Academic Press, 1985

[MEU2011] W. De Meuter, *Algoritmen en Datastructuren I*, Vrije Universiteit Brussel, 2011.

[KIEBOOM2011] R. Kieboom, *Aanvulling Lineaire Algebra*, Vrije Universiteit Brussel, 2011.

[ADHEMAR] . Bultheel, *Inleiding tot de numerieke wiskunde*, Acco, 2006.

[BIN1977] K.G. Binmore, *Mathematical Analysis: A Straightforward Approach*, Cambridge Universitiy Press, 1977.

[GOLUB] G. Golub, C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1996.

[CARA] P. Cara, *Discrete Wiskunde*, Vrije Universiteit Brussel - Dienst Uitgaven, 2011

[LIFSHITS] Y. Lifshits, *Four Results of Jon Kleinberg: A Talk for St. Petersburg Mathematical Society*, Steklov Insitute of Mathematics at St. Petersburg, 2007.

[PAGE] S. Brin, L. Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Seventh International World-Wide Web Conference, 1998.

[BLONDEL] V. D Blondel, A. Gajardo, M. Heymans, P. Senellart, P. Van Dooren, *A Measure of Similarity between Graph Vertices: Applications to Synonym Extraction and Web Searching*, Society for Industrial and Applied Mathematics, 2004.

[KLEINBERG] J. M. Kleinberg, *Authoritative sources in a hyperlinked environment*, J. ACM, 1999.

[DEIFT] P. Deift, D. Gioev, *Random Matrix Theory: Invariant Ensembles and Universality*, American Mathematical Society, January, 2009.

[FOUC] S. Foucart, *Linear Algebra and Matrix Analysis*, Drexel University, 2010.

[HORN] R.A. Horn, C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.