



# 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:



# Acknowledgements

First of all, I would like to thank my promotor, Philippe Cara for his incredible attention to detail, for the patience he had with me and the many reminders to continue and to work hard. He also showed me the way to this beautiful topic for a Master thesis. Besides his role as promotor, I will never forget him as professor of the Linear Algebra course in the very first term of the BSc in Mathematics, that without any doubt was the most memorable course during my college years: the hard study and the stress I had before entering the exam room are memories that I still will be telling in my old age.

Secondly, I would like to thank Roger Van Nieuwenhuyze, my math lecturer during the teacher training I followed for one year at the Hogeschool-Universiteit Brussel. Without his encouragements, I had never dared to leave and study Mathematics at university level.

I also want to thank all professors and teaching assistant who have learned me a lot during the past five years. A special thanks goes to Reen Tallon, who always managed to overcome any administrative burden caused by the combination of programs.

...

## Abstract

## Samenvatting

# Contents

<b>1</b>	<b>Preliminaries and notations</b>	<b>5</b>
1.1	Some families of matrices . . . . .	5
1.1.1	Permutation matrices . . . . .	5
1.1.2	Nonnegative and primitive matrices . . . . .	6
1.1.3	Irreducible nonnegative matrices . . . . .	7
1.2	Perron-Frobenius Theorem . . . . .	9
1.2.1	Spectral radii of nonnegative matrices . . . . .	9
1.2.2	Proof . . . . .	13
1.2.3	Example . . . . .	14
1.3	Numerical analysis . . . . .	14
1.3.1	Bachmann-Landau notations . . . . .	14
	Big Oh . . . . .	14
	Small Oh . . . . .	15
	Asymptotical Equality . . . . .	15
1.3.2	The Power Method . . . . .	15
	The algorithm . . . . .	17
	Computational cost & Usage . . . . .	19
	Example . . . . .	20
1.4	Graphs . . . . .	20
1.4.1	General definitions . . . . .	21
	Product graphs . . . . .	22
	Colored graphs . . . . .	23
	Adjacency matrices . . . . .	23
1.4.2	Strong connectivity . . . . .	24
<b>2</b>	<b>Similarity on graphs</b>	<b>25</b>
2.1	The HITS algorithm of Kleinberg . . . . .	25
2.1.1	History . . . . .	25
2.1.2	Motivation . . . . .	26
2.1.3	Constructing relevant graphs of webpages . . . . .	27
2.1.4	Hubs and Authorities . . . . .	28
2.1.5	Convergence of the algorithm . . . . .	32
2.1.6	Examples . . . . .	35
	Searching for math professors at the VUB . . . . .	35
	Predictors in the Eurovision Song Contest 2009-2014 . . . . .	37
2.1.7	Final reflection . . . . .	42

2.2	Node similarity . . . . .	42
	From Hubs and Authorities to structure graphs . . . . .	42
2.2.1	Convergence of the sequence $\mathbf{z}^{(k)}$ . . . . .	46
2.2.2	Compact form & Similarity matrices . . . . .	52
2.2.3	Algorithm & Computational cost . . . . .	55
2.2.4	Special cases . . . . .	56
	HITS-algorithm . . . . .	56
	Central scores . . . . .	57
	Self-similarity of a graph . . . . .	59
2.3	Node-edge similarity . . . . .	60
2.3.1	Coupled node-edge similarity scores . . . . .	61
2.3.2	The algorithm . . . . .	64
2.3.3	Difference with node similarity . . . . .	65
2.3.4	Example . . . . .	66
2.4	Colored graphs . . . . .	66
2.4.1	Colored nodes . . . . .	67
<b>Appendix A Listings</b>		<b>72</b>
<b>Appendix B Results of the Eurovision Song Contest 2009-2014</b>		<b>76</b>

# 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  $\pi$  of  $n$  elements:*

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

*with:*

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

*the associated **permutation matrix**  $P_\pi$  is the  $n \times n$ -matrix obtained by permuting the rows of the identity matrix  $I_n$  according to  $\pi$ . 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  $P_\pi$  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  $k' \in \{1, \dots, n\}$  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.  $\square$

**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 irreducible 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  $1 \times 1$ -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  $\{1, \dots, n\}$  with  $a_{ij} = 0$  for  $i \in S$ ,  $j \notin S$ . Then  $A$  is reducible.

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

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

$\sigma$  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  $k + 1 \leq d \leq n$ :

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

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  $i \in \{i_1, i_2, \dots, i_k\}$ ; similarly, since  $k + 1 \leq d \leq n$ , we have  $j \in \{j_1, j_2, \dots, j_{n-k}\}$ . By assumption, for such a pair  $i, j$ , we have  $a_{ij} = 0$ . That completes the proof.  $\square$

We now 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  $A, A^2, \dots, A^n$  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  $PAP^T PAP^T = PA^2P^T$ , 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  $t = 1, 2, \dots$  and for any  $r, s$  corresponding to an entry of the zero submatrix in  $PAP^T$ . Now, for  $t = 1, \dots, n$ :

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

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

**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

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

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

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

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  $\lambda_i$ ,  $1 \leq i \leq n$ . 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  $\lambda_i$  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

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

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

$$A\mathbf{x} \geq p\mathbf{x} \quad (1.3)$$

We now consider the nonnegative quantity  $r$  defined by

$$r = \sup_{\substack{\mathbf{x} \geq 0 \\ \mathbf{x} \neq \mathbf{0}}} \{r_{\mathbf{x}}\} \quad (1.4)$$

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:

$$r = \sup_{\mathbf{y} \in Q} \{r_{\mathbf{y}}\} \quad (1.5)$$

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:

$$A\mathbf{z} \geq r\mathbf{z}, \quad (1.6)$$

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  $r > 0$ .  $\square$

**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  $\mathbf{w} = (I + A)^{n-1}\mathbf{z}$ , then we have  $\mathbf{z} > 0$ , completing the proof.  $\square$

**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  $\beta$  is any eigenvalue of  $B$ , then

$$|\beta| \leq r, \quad (1.7)$$

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:

$$B = e^{i\phi} D A D^{-1}, \quad (1.8)$$

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

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

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

$$|\beta| |\mathbf{y}| \leq |B| |\mathbf{y}| \leq A |\mathbf{y}|, \quad (1.9)$$

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,

$$r |\mathbf{y}| = |B| |\mathbf{y}| = A |\mathbf{y}|, \quad (1.10)$$

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

$$|B| = A \quad (1.11)$$

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

$$D = \text{diag} \left\{ \frac{y_1}{|y_1|}, \dots, \frac{y_n}{|y_n|} \right\}.$$

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

$$\mathbf{y} = D |\mathbf{y}|. \quad (1.12)$$

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

$$C |\mathbf{y}| = r |\mathbf{y}|, \quad (1.13)$$

where

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

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

$$C |\mathbf{y}| = |B| |\mathbf{y}| = A |\mathbf{y}|. \quad (1.15)$$

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

$$|C| = |B| = A. \quad (1.16)$$

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  $\beta$  with  $|\beta| = r$ , which completes the proof.  $\square$

**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.  $\square$

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

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

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.  $\square$

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  $M(a_1, \dots, a_k)$ . Suppose  $x_1, \dots, x_k$  are differentiable vector functions, then  $M(x_1, \dots, x_k)$  is differentiable and:

$$\frac{d}{dt}M(x_1, \dots, x_k) = M\left(\frac{d}{dt}x_1, x_2, \dots, x_k\right) + M\left(x_1, \frac{d}{dt}x_2, \dots, x_k\right) + \dots + M\left(x_1, x_2, \dots, \frac{d}{dt}x_k\right)$$

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

$$\frac{d}{dt} \det(x_1, \dots, x_k) = \det\left(\frac{d}{dt}x_1, x_2, \dots, x_k\right) + \det\left(x_1, \frac{d}{dt}x_2, \dots, x_k\right) + \dots + \det\left(x_1, x_2, \dots, \frac{d}{dt}x_k\right)$$

**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:

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

*Proof.* The proof is immediately done by direct computation:

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

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

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

□

### 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 a 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  $\|\mathbf{x}\| = 1$  ( $\|\mathbf{x}\|$  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 whether 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.

#### Big 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  $g \Leftrightarrow \lim_{x \rightarrow +\infty} \frac{f(x)}{g(x)} = 1$ . 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 adaptations 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  $\lambda$  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  $\lambda_1, \lambda_2, \dots, \lambda_n$ . When

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

$\lambda_1$  is called the **dominant eigenvector**.

**Corollary 1.3.6.** The dominant eigenvector  $\lambda_1$  of a  $n \times n$ -matrix  $A$  is real.

*Proof.* This is trivial, if  $\lambda_1$  would be complex, the complex conjugate of  $\lambda_1$  would also be an eigenvalue with the same modulus.  $\square$

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

*Proof.*  $\Leftarrow$  Consider  $A \in \mathbb{R}^{n \times n}$  with (not necessarily different) eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$  and the eigenbasis  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ . Let:

$$X = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$$

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  $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ , 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 linear independent. This means that  $A$  has  $n$  linear independent eigenvectors, forming an eigenbasis.  $\square$

**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  $\lambda_1$ . We know that  $A$  has eigenbasis  $V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ . Every vector  $\mathbf{x}^{(0)} \in \mathbb{R}^n$  can be written as a linear 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:

$$\begin{aligned} A^k \mathbf{x}^{(0)} &= \sum_{i=1}^n \xi_i A^k \mathbf{v}_i \\ &= \sum_{i=1}^n \xi_i \lambda_i^k \mathbf{v}_i \\ &= \lambda_1^k \left\{ \xi_1 \mathbf{v}_1 + \xi_2 \left( \frac{\lambda_2}{\lambda_1} \right)^k \mathbf{v}_2 + \dots + \xi_n \left( \frac{\lambda_n}{\lambda_1} \right)^k \mathbf{v}_n \right\} \end{aligned}$$

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

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

so:

$$\mathbf{x}^{(k)} = \lambda_1^k \xi_1 \mathbf{v}_1 + o(1) \text{ for } k \rightarrow \infty \quad (1.18)$$

This means that for  $k$  large enough  $\mathbf{x}^{(k+1)}$  is equal to  $\lambda_1$  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  $\lambda_1$ .  $\mathbf{x}^{(k)}$  will be a corresponding eigenvector because it's 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  $\lambda_1$  in the power method by:

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

To decide on the sign of  $\lambda_1$ , just divide two non-zero components of  $\mathbf{x}^{(k+1)}$  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 underflow 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  $k = 0, 1, \dots$ :

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

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:

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

So:

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

Because  $\mu_{k+1}$  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 5. Just to give an understandable algorithm, we used the Euclidean norm in this algorithm<sup>1</sup>.

---

<sup>1</sup> $\|\cdot\|_2$  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,

$\mu_k$ : an estimation of the dominant eigenvalue.

**begin** power\_method( $\mathbf{y}^{(0)}$ ,  $k$ )

$k = 1$  ;

**repeat**

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

$\mu_k = \|\mathbf{z}^{(k)}\|_2$ ;

$\mathbf{y}^{(k)} = \frac{\mathbf{z}^{(k)}}{\mu_k}$ ;

$k = k + 1$

**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  $\mu_k$  of  $\lambda_1$ :

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

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:

$$\text{power\_method} \in O\left(\frac{-1}{\log \lambda_1 - \lambda_2}\right)$$

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

$k$	$\mu_k$	$k$	$\mu_k$
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  $\mu_k$  of Example 1.3.9.

of  $A$ , we can predict the number of steps  $n = -5/\log \lambda_2/\lambda_1 = -5/(\log 2/3) \approx 28$ . 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.

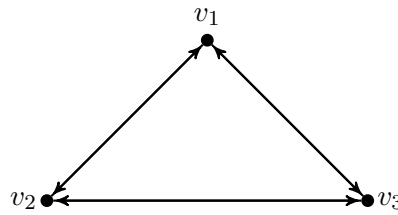
The definitions and results in this section are mainly based on the course ‘Discrete Mathematics’ by P. Cara [CARA].

### 1.4.1 General definitions

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

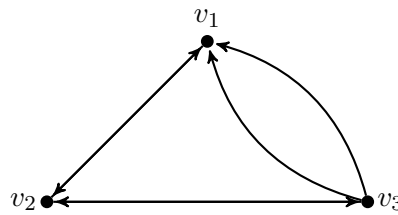
When the relation  $\rightarrow$  is symmetric, we call the graph **undirected**, in this case we often write  $\sim$  instead of  $\rightarrow$ .

**Example 1.4.2.** The graph



is an undirected graph with vertices  $v_1, v_2, v_3$ . The adjacency relation  $\rightarrow$  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  $(V, \mu)$  with  $V$  a set and  $\mu : V \times V \rightarrow \mathbb{N}$  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  $\mu$  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  $\mathcal{G} = (V, \mu)$  is the induced subgraph  $\mathcal{G}_v$  with vertex set  $V'$  consisting of all vertices adjacent to  $v$  without  $v$  itself and with the multiplicity function  $\mu'$ , which is the restriction of  $\mu$  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.** Let  $\mathcal{G} = (V, \rightarrow)$  be a graph with vertices  $v_1, v_2, \dots, v_n \in V$  and edges (defined as ordered pairs)  $(v_{x_1}, v_{y_1})_1, (v_{x_1}, v_{y_1})_2, \dots, (v_{x_m}, v_{y_m})_m \in \rightarrow$  (with  $(v_{x_i}, v_{y_i}) \in V$ ). Let  $u \rightarrow v$  be an edge between  $u, v \in V$ . We call  $u$  the **source node** and  $v$  the **terminal node** of the edge.  $s_{\mathcal{G}}(i)$  denotes the source node  $u_i$  of edge  $i$ ,  $t_{\mathcal{G}}(i)$  denotes the terminal node  $w_i$  of edge  $i$ .

**Definition 1.4.6.** The **order of a finite graph**  $\mathcal{G}$  is the number of vertices of  $\mathcal{G}$  and is denoted by  $|\mathcal{G}|$ .

**Definition 1.4.7.** The **indegree of a vertex**  $v$  in a graph  $\mathcal{G}$  is the number of times  $v$  is a terminal node of an edge.

**Definition 1.4.8.** The **outdegree of a vertex**  $v$  in a graph  $\mathcal{G}$  is the number of times  $v$  is a source node of an edge.

**Definition 1.4.9.** The **degree of a vertex**  $v$  in a graph  $\mathcal{G}$  is the sum of the indegree and outdegree of vertex  $v$ .

**Definition 1.4.10.** A **walk** in a graph  $\mathcal{G}$  is a sequence of vertices

$$v_0, v_1, \dots, v_k$$

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

**Definition 1.4.11.** If all edges are distinct in a walk in a graph  $\mathcal{G}$ , we call the walk a **path**.

**Definition 1.4.12.** A **cycle** is a walk from  $v_0$  to  $v_0$  in which all vertices except  $v_0$  are distinct.

**Definition 1.4.13.** A **simple graph** is an undirected graph  $\mathcal{G} = (V, \mu)$  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 1.4.14.** Take two graphs  $\mathcal{G} = (U, \rightarrow)$ ,  $\mathcal{H} = (V, \rightarrow')$ , the **product graph**  $\mathcal{G} \times \mathcal{H}$  is the graph with  $|\mathcal{G}| \cdot |\mathcal{H}|$  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  $\mathcal{G}$  and there is an edge between  $v_j$  and  $v_l$  in  $\mathcal{H}$ .

### Colored graphs

**Definition 1.4.15.** A **node colored graph**  $\mathcal{G}$  is quadruple  $(V, \rightarrow, C, a)$  with  $V$  a set of vertices,  $\rightarrow$  an adjacency relation,  $C$  a set of colors and  $a$  a surjective function  $a : V \rightarrow C$  that assigns to each vertex one color.

**Definition 1.4.16.** In a node colored graph  $\mathcal{G} = (V, \rightarrow, C, a)$ ,  $c_{\mathcal{G}}(i)$  denotes the number of vertices of color  $i$ . So:

$$c_{\mathcal{G}}(i) = |\{(i, v_j) \in C \times V | a(v_j) = i\}|$$

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

**Corollary 1.4.18.** The adjacency matrix of an undirected graph  $\mathcal{G} = (V, \sim)$  is a symmetric matrix.

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

**Theorem 1.4.19.** Let  $k > 0$ . The element on place  $(i, j)$  in  $A_{\mathcal{G}}^k$  contains the number of walks of length  $k$  from  $i$  to  $j$  in the graph  $\mathcal{G} = (V, \mu)$ .

*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  $A_{\mathcal{G}}$ .

Let  $v_l$  be a vertex of  $\mathcal{G}$ . If there are  $b_{il}$  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  $A_{\mathcal{G}}^k$  so  $c_{ij}$  is exactly the element on place  $(i, j)$  in the matrix product

$$A_{\mathcal{G}}^k A_{\mathcal{G}} = A_{\mathcal{G}}^{k+1}.$$

□

**Example 1.4.20.** 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* and introduce the concept of connectivity.

**Definition 1.4.21.** An undirected graph  $\mathcal{G}$  is **connected** if it possible to establish a path from any vertex to any other vertex.

**Definition 1.4.22.** A directed graph  $\mathcal{G}(V, \rightarrow)$  is **connected** if the underlying undirected graph (remove all arrows on the edges) is connected, the directed graph  $\mathcal{G}$  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 1.4.23.** Let  $\mathcal{G}$  be a (directed) graph with adjacency matrix  $A$ . Then  $\mathcal{G}$  is strongly connected if and only if  $A$  is irreducible.

*Proof.* From Theorem 1.4.19 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.

□

## Chapter 2

# Similarity on graphs

In the previous chapter all the basic terminology and results were introduced, now we 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. Next, we extend the notion of similarity on directed graphs: the method of Blondel only returns the concept of node similarity which is in fact a measurement of how similar two nodes of two graphs are to each other.

## 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 were 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 Jon 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](http://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 into further detail about the PageRank-algorithm. In the following paragraphs, the HITS algorithm is extensively explained.

### 2.1.2 Motivation

Kleinberg's work originates in the problems that arise 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 occurrence. 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](http://www.vdab.be). Unfortunately, the query 'job offers' is used in over a million pages on the internet and [www.vdab.be](http://www.vdab.be) is not the one using the term most often. Therefore, there is no way to favor [www.vdab.be](http://www.vdab.be) in a text-based ranking function. This is 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 manufacturers such as "[www.apple.com](http://www.apple.com)", "[www.samsung.com](http://www.samsung.com)",... would be good authorities for this search query and these are the results a user expects 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:** $\sigma$ : 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  $S_\sigma$  satisfying all the properties of our wish list.

```

begin create_graph( $\sigma, \mathcal{E}, t, d$ )
  Let  $R_\sigma$  denote the top  $t$  results of  $\mathcal{E}$  on  $\sigma$ ;
  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  $S_\sigma$ ;
    if  $|\Gamma^-(p)| \leq d$  then
      | Add all pages in  $\Gamma^-(p)$  to  $S_\sigma$ ;
    else
      | Add an arbitrary set of  $d$  pages from  $\Gamma^-(p)$  to  $S_\sigma$ ;
    end
  end
  return  $S_\sigma$ ;
end

```

**Algorithm 2:** Algorithm to construct  $S_\sigma$ .

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  $\mathcal{G} = (V, \rightarrow)$ : the nodes correspond to the pages, and if there is a link from page  $p$  to page  $q$ , there is an arc  $p \rightarrow q$ . Suppose a search query is performed, specified by a query  $\sigma$ . 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  $Q_\sigma$  of all pages where the query  $\sigma$  occurs. For this, we could use any already existing text-based search engine. But, for our algorithm  $Q_\sigma$  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  $Q_\sigma$  does not contain some of the most important authorities because they never use the query string  $\sigma$  on their website.

Therefore, we wish to transform the set  $Q_\sigma$  to a set  $S_\sigma$  of pages following this ‘wish list’ of properties:

1.  $S_\sigma$  is relatively small,
2.  $S_\sigma$  is rich in relevant pages,
3.  $S_\sigma$  contains most of the strongest authorities.

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

Thus, we obtain  $S_\sigma$  by expanding  $R_\sigma$  to include any page pointed to by a page in  $R_\sigma$ . We also add  $d$  pages that point to a page in  $R_\sigma$ .  $d$  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  $S_\sigma$  relatively small. Some experiments in [KLEINBERG] showed that this algorithm resulted in a  $S_\sigma$  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  $R_\sigma$  to be added to  $S_\sigma$ .

Denote the resulting graph of the page set  $S_\sigma$  by  $\mathcal{G}[S_\sigma]$ . Note that  $\mathcal{G}[S_\sigma]$  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  $\mathcal{G}[S_\sigma]$  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  $\mathcal{G}[S_\sigma]$ , 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  $\mathcal{G}[S_\sigma]$ , we get a new graph  $\mathcal{G}'_\sigma$  which is exactly what we need to perform our link analysis.

#### 2.1.4 Hubs and Authorities

A very simple approach would now be to order the pages in  $\mathcal{G}'_\sigma$  by their indegree. Although this approach can sometimes return good search results, this heuristic is often too simple because  $S_\sigma$  will probably contain some web pages with a lot of incoming links without being very relevant to the search query  $\sigma$  (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  $\mathcal{G}'_\sigma$  can tell us a lot more than it may seem at first glance. Authoritative pages relevant to query  $\sigma$  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  $\sigma$ . Figure 2.1.1 shows what this means conceptually.

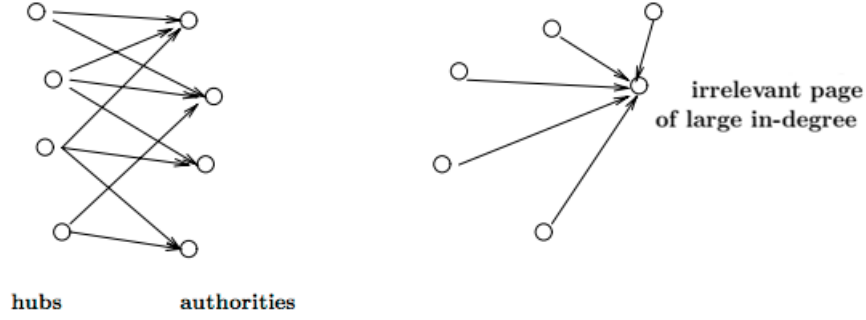


Figure 2.1.1: The concept of hubs and authorities (Source: [KLEINBERG])

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  $\mathcal{G}'_\sigma = (V, \rightarrow)$  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 sum of 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  $v_j$ .

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

The basic operations in which hubs and authorities reinforce one another are depicted in Figure 2.1.2. deze tekening wordt nog herwerkt

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

$$\begin{pmatrix} \mathbf{h} \\ \mathbf{a} \end{pmatrix}^{(k+1)} = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{h} \\ \mathbf{a} \end{pmatrix}^{(k)}, \quad k = 0, 1, \dots,$$

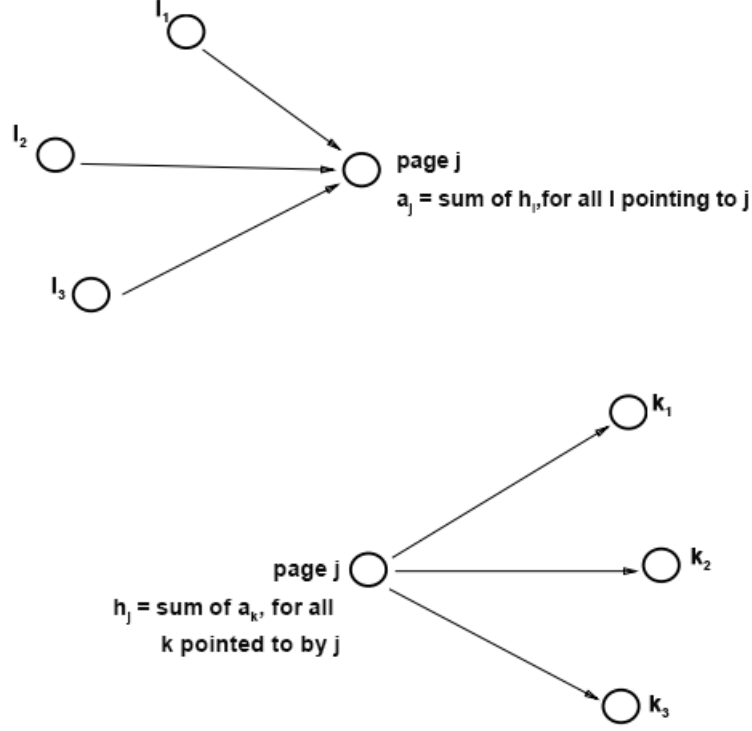


Figure 2.1.2: The basic operations in the reinforcing relation between hubs and authorities (Source: [KLEINBERG])

In compact form, we denote

$$\mathbf{x}^{(k+1)} = M\mathbf{x}^{(k)}, \quad k = 0, 1, \dots, \quad (2.1)$$

where

$$\mathbf{x}^{(k)} = \begin{pmatrix} \mathbf{h} \\ \mathbf{a} \end{pmatrix}^{(k)}, \quad M = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$$

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:

$$\mathbf{z}^{(0)} = \mathbf{x}^{(0)} > 0, \quad \mathbf{z}^{(k+1)} = \frac{M\mathbf{z}^{(k)}}{\|M\mathbf{z}^{(k)}\|_2}, \quad k = 0, 1, \dots, \quad (2.2)$$

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 sake of simplicity, we make the natural choice<sup>1</sup>  $\mathbf{1} \in \mathbb{R}^{2n}$ . The limit to which the

<sup>1</sup> $\mathbf{1}$  is a matrix, or vector, whose entries are all equal to 1.

**Data:**

$\mathcal{G}$ : 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( $\mathcal{G}, k$ )

    Set  $\mathbf{a}^{(0)} = (1, 1, \dots, 1) \in \mathbb{R}^n$ ;

    Set  $\mathbf{h}^{(0)} = (1, 1, \dots, 1) \in \mathbb{R}^n$ ;

**for**  $i = 1, 2, \dots, k$  **do**

        Calculate

$\mathbf{h}'^{(i)} = \left( \sum_{m:(v_1, v_m) \in \rightarrow} \mathbf{a}_m^{(i-1)}, \sum_{m:(v_2, v_m) \in \rightarrow} \mathbf{a}_m^{(i-1)}, \dots, \sum_{m:(v_n, v_m) \in \rightarrow} \mathbf{a}_m^{(i-1)} \right)$ ;

        Normalize  $\mathbf{h}'^{(i)}$  obtaining  $\mathbf{h}^{(i)}$ ;

        Calculate  $\mathbf{a}'^{(i)} = \left( \sum_{m:(v_m, v_1) \in \rightarrow} \mathbf{h}_m^{(i)}, \sum_{m:(v_m, v_2) \in \rightarrow} \mathbf{h}_m^{(i)}, \dots, \sum_{m:(v_m, v_n) \in \rightarrow} \mathbf{h}_m^{(i)} \right)$ ;

        Normalize  $\mathbf{a}'^{(i)}$  obtaining  $\mathbf{a}^{(i)}$ ;

**end**

**return**  $(\mathbf{h}^{(k)}, \mathbf{a}^{(k)})$ ;

**end**

**Algorithm 3:** The iterative HITS-algorithm.

**Data:**

$\mathcal{G}$ : a graph of  $n$  linked pages.

$k$ : natural number.

$c$ : natural number.

**Result:** A vector  $((\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_c), (\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_c))$  containing exactly the nodes of the  $c$  top hubs and  $c$  top authorities.

**begin** filter( $\mathcal{G}, k, c$ )

$(\mathbf{h}, \mathbf{a}) = \text{hits}(\mathcal{G}, k)$ ;

    Sort the pages with the  $c$  largest values in  $\mathbf{h}$ , resulting in a vector of nodes

$(\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_c)$ ;

    Sort the pages with the  $c$  largest values in  $\mathbf{a}$ , resulting in a vector of nodes

$(\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_c)$ ;

**return**  $((\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_c), (\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_c))$ ;

**end**

**Algorithm 4:** Returning the top  $c$  hubs and authorities

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

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 prove that the sequence converges, theoretically we can keep on iterating until a fixed point is approximated. 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 of 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 get 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



proportional: 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 now want to prove that for arbitrarily large values of  $k$ , the sequence  $Z^{(k)}$  converge to a limit  $(\mathbf{h}', \mathbf{a}')$ . Before 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  $M$ . This is the first condition of the power method we introduced in section 1.3.2. 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 (a unique dominant eigenvalue means the largest eigenvalue with multiplicity 1), simply because this is not true in general.<sup>2</sup> 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 (See for example Theorem 2.2.3 in [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  $S_\sigma$  will be linked to each other, the graph  $\mathcal{G}'_\sigma$  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  $M$ , the matrix 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*

---

<sup>2</sup>The matrix  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  is a simple counterexample of a symmetric, nonnegative, real matrix that has no unique dominant eigenvalue.

different) real eigenvalues corresponding to real eigenvectors.

*Proof.* First, treat  $A$  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  $\lambda$  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  $\lambda_1$  by the previous theorem. Let  $\mathbf{x}_1$  be a corresponding eigenvector 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  $V_1 = \{\mathbf{x}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  of  $\mathbb{R}^n$ . Let:

$$S_1 = [\mathbf{x}_1, \mathbf{v}_2, \dots, \mathbf{v}_n],$$

since  $S_1$  is orthonormal, we get  $S_1^t = S_1^{-1}$ . 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:

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

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:

$$\begin{aligned} (S'_2)^{-1}S_1^{-1}AS_1S'_2 &= \left( \begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0}^t & S_2^t \end{array} \right) (S_1^{-1}AS_1) \left( \begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0}^t & S_2 \end{array} \right) \\ &= \left( \begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0}^t & S_2^t \end{array} \right) \left( \begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0}^t & A_1 \end{array} \right) \left( \begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0}^t & S_2 \end{array} \right) \\ &= \left( \begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0}^t & S_2^t A_1 S_2 \end{array} \right) \\ &= \left( \begin{array}{c|c} \lambda_1 & \mathbf{0} \\ \hline \mathbf{0}^t & D' \end{array} \right) \end{aligned}$$

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.  $\square$

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

$$\mathbf{z}^{(0)} = \mathbf{1} \in \mathbb{R}^n, \mathbf{z}^{(k+1)} = \frac{M\mathbf{z}^{(k)}}{\|M\mathbf{z}^{(k)}\|_2}, \quad k = 0, 1, \dots,$$

*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.  $\square$

We conclude with a nice corollary.

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

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

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

*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  $\mathbf{a}_k \leftarrow (B^T B)^k \mathbf{a}_0$ . Let  $\mathbf{a}_0$  be  $\mathbf{1} \in \mathbb{R}^n$ . From the previous theorem we also know that:

$$\lim_{k \rightarrow \infty} \mathbf{h}_k = \mathbf{h} \quad \text{and} \quad \lim_{k \rightarrow \infty} \mathbf{a}_k = \mathbf{a},$$

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  $B^T B$ .  $\square$

## 2.1.6 Examples

### Searching for math professors at the VUB

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

- **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 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:

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

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.

### Predictors in the Eurovision Song Contest 2009-2014

The Eurovision Song Contest is an annual competition between countries whose public broadcaster is part of the EBU-network. The contest is the biggest music competition in the world, reaching about 200 million annually.

The contest consists of three shows: 2 semi-finals and 1 grand final. From each semi-final, 10 countries proceed to the grand final. Each country, also those who dropped out during the semi-finals, gives points during the voting of the grand final. The voting during the grand final takes place after all the countries have performed their song. Each country is called and awards 12 points to their favorite song, 10 points to their second favorite, and then points from 8 down to 1 to other eight songs. Countries can not vote for themselves.

The voting system is in fact a positional voting system that is very similar to the Borda count method (see [SAARI] for a scientific explanation of Borda count): the list of points of a country represents the ranking of the 10 best countries in the voting of that country. So the points are values on an ordinal scale.

The complete voting procedure during the Eurovision Song Contest can be seen as a directed graph: all the participating countries are the nodes and the edges represent the points between the countries (when country  $a$  assigns 3 points to country  $b$ , then there are 3 edges from  $a$  to  $b$ ).

Let  $A$  be the adjacency matrix of the voting during a song contest ( $A$  will in fact be just a points table). If we take  $A$  as input for the HITS-algorithm, we expect that the country with the highest authority score will be the winner of the competition. Actually we expect a lot more: when we order the countries based on their authority score, we expect that this ordering will be practically equal to the final ranking of the contest. This is based on the simple fact that Borda count just sums up points, and we only expect very small differences when the difference in points is low between two countries, because then the hub scores can influence the result of the HITS-algorithm a bit.

But what is very interesting now, is the role that the hub scores are playing. In fact these scores can be seen as a kind of ‘predictive value’ of a country: a country with a high hub score will have assigned points in such way that it is seen as a reliable source, meaning that the points of that country will match well with the final result of the contest.

Let’s take the Eurovision Song Contest 2014 as an example.

The final results of the contest where (the complete result table can be found in Appendix B):

1. Austria (290 points)
2. The Netherlands (238 points)
3. Sweden (218 points)
4. Armenia (174 points)
5. Hungary (143 points)
6. Ukraine (113 points)
7. Russia (89 points)
8. Norway (88 points)
9. Denmark (74 points)
10. Spain (74 points)
11. Finland (72 points)
12. Romania (72 points)
13. Switzerland (64 points)
14. Poland (62 points)
15. Iceland (58 points)
16. Belarus (43 points)
17. United Kingdom (40 points)
18. Germany (39 points)
19. Montenegro (37 points)
20. Greece (35 points)
21. Italy (33 points)
22. Azerbaijan (33 points)
23. Malta (32 points)
24. San Marino (14 points)

25. Slovenia (9 points)

26. France (2 points)

Now we calculate the hub and authority scores, based on the full scoreboard and the results are presented in Table 2.1.6.

Participant	Authority Score	Participant	Hub Score
Austria	0.285110029	Portugal	0.173610946
The Netherlands	0.235720093	Finland	0.172574335
Sweden	0.212949392	Belgium	0.169584694
Armenia	0.156091783	Latvia	0.166295198
Hungary	0.124453384	Spain	0.165764986
Ukraine	0.095410297	Hungary	0.165699760
Norway	0.086504385	Iceland	0.165062483
Denmark	0.074474911	Estonia	0.162056401
Finland	0.070675128	Denmark	0.160549539
Spain	0.068432379	Lithuania	0.159920120
Russia	0.065352465	Greece	0.159278464
Romania	0.062560768	Norway	0.156676045
Switzerland	0.055868228	Slovenia	0.156533823
Iceland	0.053973263	Sweden	0.155725338
Poland	0.052246035	Romania	0.155212094
United Kingdom	0.037978889	France	0.153993935
Germany	0.029958317	Switzerland	0.153864143
Belarus	0.027945852	Israel	0.153059334
Malta	0.026911408	Ireland	0.147584917
Italy	0.023817428	The Netherlands	0.145211066
Montenegro	0.023425296	United Kingdom	0.144342619
Azerbaijan	0.022613716	Austria	0.137956312
Greece	0.021816155	Germany	0.136869443
San Marino	0.009315756	Ukraine	0.135769452
Slovenia	0.005843162	Italy	0.121638547
France	0.002167387	Malta	0.119063784
Albania	0	Georgia	0.117423087
Belgium	0	Moldova	0.115874904
Estonia	0	Poland	0.112296570
FYR Macedonia	0	FYR Macedonia	0.107531180
Georgia	0	Russia	0.102149330
Ireland	0	San Marino	0.098204303
Israel	0	Montenegro	0.097193444
Latvia	0	Albania	0.091897617
Lithuania	0	Belarus	0.089227164
Moldova	0	Azerbaijan	0.068005452
Portugal	0	Armenia	0.050899422

Table 2.1: The authority and hub scores of the countries during the Final of the Eurovision Song Contest 2014



Notice that ordering the countries by their authority scores indeed returns the final ranking of the contest. The countries who have an authority score equal to 0 are the countries that didn't make it to the final and couldn't therefore receive any points. When looking at the hub scores, it appears that Portugal 'predicted' the final ranking the best. When we look at the points awarded by Portugal during the final, this is indeed true:

- **12 points:** Austria
- **10 points:** The Netherlands
- **8 points:** Sweden
- **7 points:** Switzerland
- **6 points:** Hungary
- **5 points:** Denmark
- **4 points:** Armenia
- **3 points:** Norway
- **2 points:** Russia
- **1 point:** Romania

No less than 7 countries in the points of Portugal have achieved the final top 10, the top 3 results of Portugal are even equal to the final top 3! Russia, Romania and Switzerland are the 3 countries where Portugal awarded points to but didn't achieve the top 10, but still they are placed 11, 12 and 13 in the final ranking. So it is absolutely not surprising that Portugal is the country with the highest hub score.

Note that countries who scored very well during the final, usually have a moderate hub score: this is due to the fact that countries can not vote for themselves. The average hub score is, in general, also lower than the average authority score, this is because countries can award points to only 10 countries, but (qualified) countries can receive points from every country, except themselves. The 'predictive value' of a country is therefore limited to only 10 countries, while the voting produces a complete final ranking of 26 countries.

Of course, it's tempting to research the predictive value of countries during a couple of years. We opted for the period 2009-2014 because during the last five years the voting procedure remained unchanged: the points awarded by each country are based on 50% televoting and 50% jury vote. So we take the average of the hub scores of the last five years of each country. We only put one condition on these countries: they should have participated at 4 out of 5 times, more than 1 absence would possibly make the average misleading. The result is presented in Table 2.1.6. The complete results of each year can be found in Appendix B.

We see that Hungary, Cyprus and Belgium are the best predictors during the last five years. The bottom is also not very surprising: in 2010, 2012 and 2013 Albania did not receive enough televotes so the jury decided their points for 100% (see [EUROVISION] for more information), making their judging process vary from one year to another. Azerbaijan and Armenia scored very high in the last five years, (e.g. Azerbaijan reached the top 5 every year except 2014, won in 2011 and became second in 2013) but due to their dispute about the Nagorno-Karabakh region, they never exchanged a single point during the last five years.

Participant	2014	2013	2012	2011	2010	2009	Average
Hungary	0.165700	0.150320	0.151291	0.130650	-	0.158260	<b>0.151244</b>
Cyprus	-	0.161413	0.140385	0.145691	0.144175	0.146567	<b>0.147646</b>
Belgium	0.169585	0.159189	0.153462	0.118487	0.147362	0.122234	<b>0.145053</b>
Lithuania	0.159920	0.141552	0.144121	0.127425	0.144634	0.152601	<b>0.145042</b>
The Netherlands	0.145211	0.133926	0.147638	0.137637	0.139085	0.163395	<b>0.144482</b>
Israel	0.153059	0.157744	0.139948	0.131677	0.118970	0.157031	<b>0.143072</b>
Spain	0.165765	0.153609	0.133174	0.114255	0.162626	0.122739	<b>0.142028</b>
Latvia	0.166295	0.134627	0.138436	0.121329	0.142837	0.143788	<b>0.141219</b>
Estonia	0.162056	0.141096	0.131495	0.143254	0.136098	0.131401	<b>0.140900</b>
Slovenia	0.156534	0.141375	0.148665	0.118682	0.137277	0.133178	<b>0.139285</b>
Malta	0.119064	0.146482	0.122804	0.162192	0.138073	0.139682	<b>0.138049</b>
Iceland	0.165062	0.149906	0.129060	0.130358	0.137492	0.113493	<b>0.137562</b>
Denmark	0.160550	0.112496	0.139704	0.104782	0.144174	0.156365	<b>0.136345</b>
Greece	0.159278	0.144943	0.124054	0.145912	0.095629	0.148054	<b>0.136312</b>
Austria	0.137956	0.127058	0.153796	0.129429	-	0.132217	<b>0.136091</b>
Croatia	-	0.169310	0.132023	0.129980	0.114014	0.134540	<b>0.135974</b>
France	0.153994	0.135187	0.151435	0.137268	0.118194	0.109034	<b>0.134185</b>
Russia	0.102149	0.130277	0.129907	0.143360	0.141861	0.152992	<b>0.133424</b>
Sweden	0.155725	0.122816	0.099063	0.106864	0.153858	0.160329	<b>0.133109</b>
Ireland	0.147585	0.137790	0.132864	0.103704	0.147411	0.128126	<b>0.132913</b>
Romania	0.155212	0.141742	0.119829	0.131675	0.128084	0.119466	<b>0.132668</b>
Germany	0.136869	0.126650	0.157167	0.103942	0.115909	0.155134	<b>0.132612</b>
Finland	0.172574	0.108008	0.141472	0.101891	0.133330	0.132773	<b>0.131675</b>
Bulgaria	-	0.136342	0.150772	0.110776	0.136208	0.122794	<b>0.131378</b>
Norway	0.156676	0.098933	0.154850	0.099012	0.155455	0.107546	<b>0.128745</b>
United Kingdom	0.144343	0.143615	0.121922	0.096917	0.136172	0.133630	<b>0.128594</b>
Portugal	0.173611	-	0.116836	0.130723	0.107674	0.105899	<b>0.126949</b>
Serbia	-	0.165969	0.113907	0.095025	0.132455	0.123080	<b>0.126087</b>
Ukraine	0.135769	0.106438	0.123998	0.120992	0.139278	0.142826	<b>0.125295</b>
FYR Macedonia	0.107531	0.138521	0.132612	0.116197	0.127626	0.124228	<b>0.124453</b>
Belarus	0.089227	0.143351	0.120914	0.138934	0.104108	0.149588	<b>0.124354</b>
Moldova	0.115875	0.139519	0.119439	0.126125	0.119696	0.119161	<b>0.123302</b>
Switzerland	0.153864	0.117585	0.125128	0.092621	0.127077	0.117733	<b>0.122335</b>
Georgia	0.117423	0.147386	0.123067	0.132744	0.086281	-	<b>0.121380</b>
Albania	0.091898	0.102282	0.098185	0.142108	0.140538	0.141493	<b>0.119417</b>
Azerbaijan	0.068005	0.109553	0.113396	0.119173	0.120818	0.109357	<b>0.106717</b>
Armenia	0.050899	0.125792	-	0.128531	0.101509	0.109782	<b>0.103303</b>

Table 2.2: The average of the hub scores between 2009-2014. Countries must have participated at least 4 times.

### 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  $\mathcal{G}'_o$  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  $\mathcal{G}'_o$  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 Node similarity

This section provides a detailed overview of 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 rigorously 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 greater than 1 and therefore we will develop a method to work around this situation.

### From Hubs and Authorities to structure graphs

We start with an introduction on how we generalize the construction of the previous section. We don’t present any proofs in this subsection, the results are shown in the next subsections.

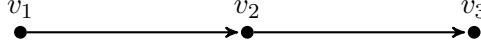
Remember from the previous section we constructed a graph  $\mathcal{G}'_o$  and calculated hub and authorities scores for each vertex. Now, for any directed graph  $\mathcal{G}$ , the authority score of a vertex  $v_j$  of  $\mathcal{G}$  can be thought of as a *score* between  $v_j$  of  $\mathcal{G}$  and the vertex denoted as *authority* of the graph:

$$hub \bullet \longrightarrow \bullet authority$$

and, conversely, the hub score of vertex  $v_j$  of  $\mathcal{G}$  can be thought of as a score between  $v_j$  and the vertex denoted as *hub*. We call the hub-authority graph a *structure graph* and we

already know the resulting iterative method from the previous section. We will call this scores *similarity scores*.

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  $\mathcal{G}(W, \rightarrow)$  be a graph. With each vertex  $w_i$  of  $\mathcal{G}$  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  $\mathcal{G}$ ),

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

which we, again, can denote by

$$\mathbf{x}^{(k+1)} = M\mathbf{x}^{(k)}. \quad (2.3)$$

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 result is the limit of the normalized vector sequence:

$$\mathbf{z}^{(0)} = \mathbf{x}^{(0)} > 0, \mathbf{z}^{(k+1)} = \frac{M\mathbf{z}^{(k)}}{\|M\mathbf{z}^{(k)}\|_2}, \quad k = 0, 1, \dots, \quad (2.4)$$

Remember that the HITS-algorithm assumed that  $M$  has 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 kinds of directed graphs. We will see that without this assumption, the sequence 2.4 does not always converge but oscillates between the limits:

$$\mathbf{z}_{\text{even}} = \lim_{k \rightarrow \infty} \mathbf{z}^{(2k)} \quad \text{and} \quad \mathbf{z}_{\text{odd}} = \lim_{k \rightarrow \infty} \mathbf{z}^{(2k+1)}$$

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:

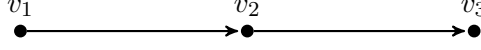
$$Z = \{\mathbf{z}_{\text{even}}(\mathbf{z}^{(0)}), \mathbf{z}_{\text{odd}}(\mathbf{z}^{(0)}) : \mathbf{z}^{(0)} > 0\},$$

and we would like to select a particular vector in that set. We will prove later on that the vector  $\mathbf{z}^{(0)} = \mathbf{1}$  is a good choice because it's the unique vector with the largest possible Manhattan<sup>3</sup> (we will prove this in Theorem 2.2.10) norm. We denote  $\mathbf{z}_{\text{even}}(\mathbf{1})$ .

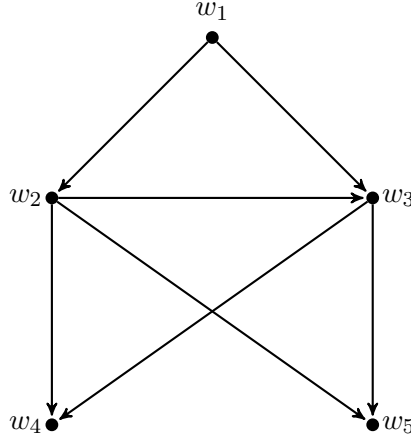
<sup>3</sup> $\|\mathbf{x}\|_1 := \sum_{i=1}^n |x_i|$  for  $\mathbf{x} \in \mathbb{R}^n$  is the Manhattan norm.

The extremal limit  $\mathbf{z}_{\text{even}}(\mathbf{1})$  will be defined as the *similarity matrix*.  
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  $\mathcal{G}(W, \rightarrow)$  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 get the following similarity matrix (a numerical algorithm to calculate this is presented later on in this section):

$$S = \begin{pmatrix} 0.3557 & 0.1265 & 0 \\ 0.3102 & 0.3451 & 0.0557 \\ 0.2732 & 0.4619 & 0.4115 \\ 0 & 0.1579 & 0.3557 \\ 0 & 0.0840 & 0.1521 \end{pmatrix}$$

The similarity score of  $w_4$  with  $v_2$  of the structure graph is equal to 0.1579.

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

$$x_{ij}^{(k+1)} := \sum_{r:(v_r, v_i) \in \rightarrow', s:(u_s, u_j) \in \rightarrow} x_{rs}^{(k)} + \sum_{r:(v_i, v_r) \in \rightarrow', s:(u_j, u_s) \in \rightarrow} x_{rs}^{(k)} \quad (2.5)$$

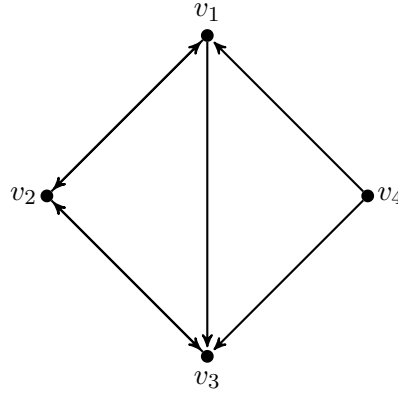
Consider the product graph  $\mathcal{G} \times \mathcal{H}$  (see Definition 1.4.14). The above updating equation is equivalent to replacing the scores of all vertices of the product graph by the sum of the scores of the vertices linked by an incoming or outgoing edge.

Equation (2.4) can be rewritten in a more compact matrix form. Let  $X_k$  be the  $n_{\mathcal{H}} \times n_{\mathcal{G}}$  matrix of entries  $x_{ij}$  at iteration  $k$ , and  $A$  and  $B$  are the adjacency matrices of  $\mathcal{G}$  and  $\mathcal{H}$ . Then the updating equations can be written as:

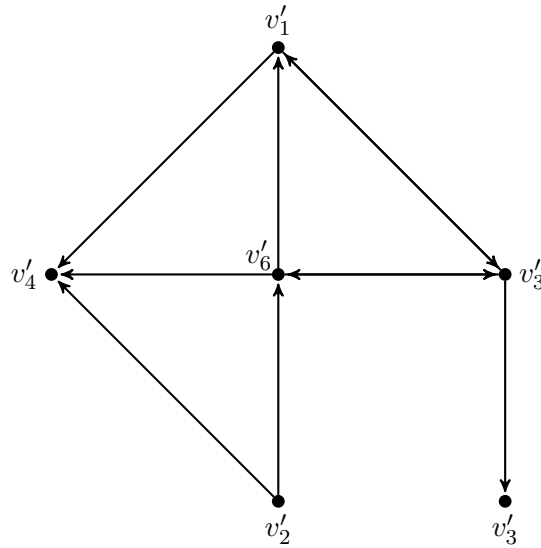
$$X^{(k+1)} = BX^{(k)}A^T + B^T X^{(k)}A, \quad k = 0, 1, \dots, \quad (2.6)$$

We'll prove that the normalized even and odd iterates of this updating equation converge and that the limit  $\mathbf{z}_{\text{even}}(\mathbf{1})$  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  $\mathcal{G}_A(V, \rightarrow)$  be the following graph:



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



We get 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  $\mathcal{G}_A$  is most similar to vertex  $v'_3$  in  $\mathcal{G}_B$  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 (2.4) converges for even and odd iterates. We will prove this at 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 2.2.7 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 by the Power Method. This approach is far too naive when introducing similarity on graph vertices, because graphs can lead to matrices where there is no unique dominant eigenvalue. 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. Next, two lemmas are presented which will also contribute to the proof of the spectral radius formula.

Before we prove the following lemma, remember the well known theorem about the Jordan canonical form of a square matrix  $A$  (see Theorem in [KIEBOOM2011]).

**Theorem 2.2.3.** *A square complex matrix  $A$  is similar to a block diagonal matrix*

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{pmatrix}$$

where each block  $J_i$  is a square matrix of the form:

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}$$

So there exists an (invertible) matrix  $P$  such that

$$P^{-1}AP = J$$

$J$  is called the **Jordan normal form** of  $A$ .

**Lemma 2.2.4.** *Let  $A$  be an  $n \times n$  matrix and  $\epsilon > 0$ , there exist a matrix norm  $\|\cdot\|$  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:

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

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

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

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:

$$\|M\| = \max_{\|\mathbf{x}\|_1=1} \|D(1/\epsilon)S^{-1}MSD(\epsilon)\mathbf{x}\|_1 \quad (2.7)$$

$$= \max_{l \in [1:n]} \sum_{k=1}^n |(D(1/\epsilon)S^{-1}MSD(\epsilon))_{k,l}|. \quad (2.8)$$

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  $\mathbb{C}^{n \times n}$ .  $\square$

**Theorem 2.2.5. Spectral radius formula** *Let  $A$  be an  $n \times n$  matrix and let  $\|\cdot\|$  be a matrix norm then:*

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



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

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

so:

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

Taking the limit as  $k \rightarrow \infty$  gives  $\rho(A) \leq \lim_{k \rightarrow \infty} \|A^k\|^{1/k}$ . 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 2.2.4, we know that there exists a matrix norm  $\|\cdot\|$  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  $k \geq 0$ ,

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

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

Now we are ready for our first big result: since one is confronted in practice with non-negative matrices that are not necessarily irreducible, we extend the Perron-Frobenius and see what remains without this assumption. We first start with a lemma and next we prove 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 2.2.6.** *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  $|\cdot|$ ).*

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

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

by using some trivial properties of the absolute value function. Let  $\|\cdot\|_2$  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  $m = 1, 2, \dots$ . If we now let  $m \rightarrow \infty$  and apply the spectral radius formula from Theorem 2.2.5 we get:

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

$\square$

**Theorem 2.2.7.** *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  $\epsilon_1 > \epsilon_2 > \dots$  with  $\lim_{k \rightarrow \infty} \epsilon_k = 0$  such that  $\lim_{k \rightarrow \infty} \mathbf{x}(\epsilon_k) = \mathbf{x}$  exists. Since  $\mathbf{x}(\epsilon_k) > 0$  for all  $k = 1, 2, \dots$ , it must be that  $\mathbf{x} = \lim_{k \rightarrow \infty} \mathbf{x}(\epsilon_k) \geq 0$ ;  $\mathbf{x} = 0$  is impossible because:

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

By Lemma 2.2.6,  $\rho(A(\epsilon_k)) \geq \rho(A(\epsilon_{k+1})) \geq \dots \geq \rho(A)$  for all  $k = 1, 2, \dots$ , so the sequence  $\{\rho(A(\epsilon_k))\}_{k=1,2,\dots}$  is a monotone decreasing sequence. Thus,  $\rho = \lim_{k \rightarrow \infty} \rho(A(\epsilon_k))$  exists and  $\rho \geq \rho(A)$ . From the fact that

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

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

Now that we know that any nonnegative matrix  $M$  has its 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 2.2.8.** *Let  $M$  be a symmetric nonnegative matrix with spectral radius  $\rho$ . Then the algebraic and geometric multiplicity of the Perron root  $\rho$  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  $\Pi$  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  $\rho$  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.  $\square$

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 2.2.9.** *Let  $\mathcal{V}$  be a linear subspace of  $\mathbb{R}^n$  with orthonormal basis  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ . 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.* <sup>4</sup> We use the connection between transposes and the standard inner product to find

---

<sup>4</sup> $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$  with  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  is the standard inner product of two vectors in  $\mathbb{R}^n$ .

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:

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

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

$$\begin{aligned} \Pi^2 &= (V V^T)^2 \\ &= V V^T V V^T \\ &= V V^{-1} V V^T \\ &= I_n V V^T \\ &= V V^T \\ &= \Pi \end{aligned}$$

□

**Theorem 2.2.10.** *Let  $M$  be a symmetric nonnegative, real matrix of spectral radius  $\rho$ . Let  $Z^{(0)} > 0$  and consider the sequence*

$$\mathbf{z}^{(k+1)} = \frac{M \mathbf{z}^{(k)}}{\|M \mathbf{z}^{(k)}\|_2}, k = 0, 1, \dots$$

*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  $\Pi$  is the orthogonal projector on the eigenspace associated with the Perron root  $\rho$ . When  $-\rho$  is an eigenvalue of  $M$ , then the subsequences  $\mathbf{z}^{(2k)}$  and  $\mathbf{z}^{(2k+1)}$  converge to the limits*

$$\mathbf{z}_{\text{even}}(Z^{(0)}) = \lim_{k \rightarrow \infty} \mathbf{z}^{(2k)} = \frac{\Pi \mathbf{z}^{(0)}}{\|\Pi \mathbf{z}^{(0)}\|_2} \quad \text{and} \quad \mathbf{z}_{\text{odd}}(\mathbf{z}^{(0)}) = \lim_{k \rightarrow \infty} \mathbf{z}^{(2k+1)} = \frac{\Pi M \mathbf{z}^{(0)}}{\|\Pi M \mathbf{z}^{(0)}\|_2}.$$

*where  $\Pi$  is the orthogonal projector on the sums of the invariant subspaces associated with  $\rho$  and  $-\rho$ . In both cases the set of all possible limits is given by:*

$$\{\mathbf{z}_{\text{even}}(Z^{(0)}), \mathbf{z}_{\text{odd}}(Z^{(0)}) : \mathbf{z}^{(0)} > 0\} = \left\{ \frac{\Pi \mathbf{z}}{\|\Pi \mathbf{z}\|_2} : \mathbf{z} > 0 \right\}$$

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

*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  $\rho$ ,  $-\rho$  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 ( $\rho$  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  $M_\mu$  is a square matrix (diagonal if  $V_\mu$  is a basis of eigenvectors of the rest of the spectrum, see Theorem 2.1.2) with spectral radius  $\mu$  strictly less than  $\rho$ .

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 *eigendecomposition* 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  $\rho^2$ . 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  $\Pi$  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  $\Pi$  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  $\Pi$  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  $\Pi$  and all vectors are nonnegative and  $\Pi^2 = \Pi$ :

$$\begin{aligned} |\langle \Pi \mathbf{1}, \Pi \mathbf{z}^{(0)} \rangle| &\leq \|\Pi \mathbf{1}\| \cdot \|\Pi \mathbf{z}^{(0)}\| \\ \Leftrightarrow |(\Pi \mathbf{1})^T \Pi \mathbf{z}^{(0)}| &\leq \sqrt{\mathbf{1}^T \Pi^2 \mathbf{1}} \cdot \sqrt{\mathbf{z}^{(0)T} \Pi^2 \mathbf{z}^{(0)}} \\ \Leftrightarrow |\mathbf{1}^T \Pi^T \Pi \mathbf{z}^{(0)}| &\leq \sqrt{\mathbf{1}^T \Pi^2 \mathbf{1}} \cdot \sqrt{\mathbf{z}^{(0)T} \Pi^2 \mathbf{z}^{(0)}} \\ \Leftrightarrow |\mathbf{1}^T \Pi^2 \mathbf{z}^{(0)}| &\leq \sqrt{\mathbf{1}^T \Pi^2 \mathbf{1}} \cdot \sqrt{\mathbf{z}^{(0)T} \Pi^2 \mathbf{z}^{(0)}}, \end{aligned}$$

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.  $\square$

## 2.2.2 Compact form & Similarity matrices

We now come to the formal definition of the similarity matrix of two directed graphs  $\mathcal{G}(U, \rightarrow)$  and  $\mathcal{H}(V, \rightarrow')$ , by proving that the mutually reinforcing relation is given by ( $A$  and  $B$  are the adjacency matrices of  $\mathcal{G}$  and  $\mathcal{H}$ ):

$$X^{(k+1)} = BX^{(k)}A^T + B^T X^{(k)}A, \quad k = 0, 1, \dots, \quad (2.9)$$

which we already introduced in 2.6. To prove this relation, we take a detour via a nice property of the Kronecker product and the vec-operator. The vec-operator is a convenient way of stacking columns of a matrix left to right. With this operator we can rewrite 2.9 in  $\mathbf{x}^{(k+1)} = M\mathbf{x}^{(k)}$  with  $M$  equal to a sum of Kronecker products. This will allow us to apply Theorem 2.2.10.

**Definition 2.2.11.** With each  $m \times n$ -matrix  $A = [a_{ij}]$  we can associate the vector  $\text{vec}(A) \in \mathbb{R}^{mn}$  defined by:

$$\text{vec}(A) = \begin{pmatrix} a_{11} & \dots & a_{m1} & a_{12} & \dots & a_{m2} & a_{1n} & \dots & a_{mn} \end{pmatrix}^T$$

**Definition 2.2.12.** The **Kronecker product** (or **tensor product**) of an  $m \times n$ -matrix  $A = (a_{ij})$  and a  $p \times q$ -matrix  $B = (b_{ij})$  is denoted by  $A \otimes B$  and is defined by the matrix:

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix} \in \mathbb{R}^{mp \times nq}$$

**Lemma 2.2.13.** Consider a  $m \times n$ -matrix  $A = (a_{ij})$  and a  $p \times q$ -matrix  $B = (b_{ij})$ , we have:

$$A^T \otimes B^T = (A \otimes B)^T$$

*Proof.* By direct computation, we get:

$$\begin{aligned} (A \otimes B)^T &= \left( \begin{array}{ccc} a_{11} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ \vdots & \ddots & \vdots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pq} \end{pmatrix} & \dots & a_{1n} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ \vdots & \ddots & \vdots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pq} \end{pmatrix} \\ \vdots & \ddots & \vdots \\ a_{m1} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ \vdots & \ddots & \vdots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pq} \end{pmatrix} & \dots & a_{mn} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ \vdots & \ddots & \vdots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pq} \end{pmatrix} \end{array} \right)^T \\ &= \left( \begin{array}{ccc} a_{11} \begin{pmatrix} b_{11} & b_{21} & \dots & b_{p1} \\ \vdots & \ddots & \vdots & \vdots \\ b_{1q} & b_{2q} & \dots & b_{pq} \end{pmatrix} & \dots & a_{m1} \begin{pmatrix} b_{11} & b_{21} & \dots & b_{p1} \\ \vdots & \ddots & \vdots & \vdots \\ b_{1q} & b_{2q} & \dots & b_{pq} \end{pmatrix} \\ \vdots & \ddots & \vdots \\ a_{1n} \begin{pmatrix} b_{11} & b_{21} & \dots & b_{p1} \\ \vdots & \ddots & \vdots & \vdots \\ b_{1q} & b_{2q} & \dots & b_{pq} \end{pmatrix} & \dots & a_{mn} \begin{pmatrix} b_{11} & b_{21} & \dots & b_{p1} \\ \vdots & \ddots & \vdots & \vdots \\ b_{1q} & b_{2q} & \dots & b_{pq} \end{pmatrix} \end{array} \right)^T \\ &= \begin{pmatrix} a_{11}B^T & \dots & a_{m1}B^T \\ \vdots & \ddots & \vdots \\ a_{1n}B^T & \dots & a_{mn}B^T \end{pmatrix} \\ &= A^T \otimes B^T \end{aligned}$$

□

**Theorem 2.2.14.** Let  $A$  a  $m \times n$ -matrix,  $B$  a  $p \times q$ -matrix and  $C$  a  $m \times p$  matrix, the matrix equation:

$$AXB = C$$

with  $X$  an unknown  $n \times p$ -matrix, is equivalent tot the system of  $qm$  equations and  $np$  unkonws given by:

$$(B^T \otimes A) \text{vec}(X) = \text{vec}(C)$$

so:

$$\text{vec}(AXB) = (B^T \otimes A) \text{vec}(X)$$

*Proof.* Let  $Q_k$  be the  $k$ th column of a given matrix  $Q$ , we get:

$$\begin{aligned} (AXB)_k &= A(XB)_k \\ &= AXB_k \\ &= A \sum_{i=1}^p b_{ik} X_i \\ &= Ab_{1k}X_1 + Ab_{2k}X_2 + \dots + Ab_{pk}X_p \\ &= \begin{pmatrix} b_{1k}A & b_{2k}A & \dots & b_{pk}A \end{pmatrix} \text{vec}(X) \\ &= (B_k^T \otimes A) \text{vec}(X) \end{aligned}$$

We get:

$$\text{vec}(AXB) = \begin{pmatrix} B_1^T \otimes A \\ \vdots \\ B_q^T \otimes A \end{pmatrix} \text{vec}(X)$$

But it follows immediately that:

$$\text{vec}(AXB) = (B^T \otimes A) \text{vec}(X)$$

because the transpose of a column of  $B$  is a row of  $B^T$ .  $\square$

**Theorem 2.2.15.** *Let  $\mathcal{G}$  and  $\mathcal{H}$  be two directed graphs with adjacency matrices  $A = [a_{ij}] \in \mathbb{R}^{n \times n}$  and  $B = [b_{ij}] \in \mathbb{R}^{m \times m}$  and let  $Z^{(0)} > 0$  be an initial positive matrix, and define<sup>5</sup>:*

$$Z^{(k+1)} = \frac{BZ^{(k)}A^T + B^T Z^{(k)}A}{\|BZ^{(k)}A^T + B^T Z^{(k)}A\|_F}, \quad k = 0, 1, \dots$$

*Then the matrix subsequences  $Z^{(2k)}$  and  $Z^{(2k+1)}$  converge to  $Z_{\text{even}}$  and  $Z_{\text{odd}}$  and among all the matrices in the set of all possible limits:*

$$\{Z_{\text{even}}(Z^{(0)}), Z_{\text{odd}}(Z^{(0)}) : Z^{(0)} > 0\}$$

*the matrix  $Z_{\text{even}}(\mathbf{1})$  is the unique matrix of largest 1-norm<sup>6</sup>.*

*Proof.* We first rewrite the compact form of 2.9:

$$\begin{aligned} X^{(k+1)} &= BX^{(k)}A^T + B^T X^{(k)}A \\ \Leftrightarrow \text{vec}(X^{(k+1)}) &= \text{vec}(BX^{(k)}A^T + B^T X^{(k)}A) \\ \Leftrightarrow \text{vec}(X^{(k+1)}) &= \text{vec}(BX^{(k)}A^T) + \text{vec}(B^T X^{(k)}A) \\ \Leftrightarrow \text{vec}(X^{(k+1)}) &= ((A^T)^T \otimes B) \text{vec}(X^{(k)}) + (A^T \otimes B^T) \text{vec}(X^{(k)}) \\ \Leftrightarrow \text{vec}(X^{(k+1)}) &= (A \otimes B + A^T \otimes B^T) \text{vec}(X^{(k)}) \end{aligned}$$

If we define  $\mathbf{x}_k = \text{vec}(X^{(k)})$  and  $M = A \otimes B + A^T \otimes B^T$  we get:

$$\mathbf{x}^{(k+1)} = M\mathbf{x}^{(k)},$$

exactly the same compact form as introduced in the beginning of this section in 2.3. If we can prove that  $M$  is symmetric and nonnegative, then we can apply Theorem 2.2.10 and the result follows.  $M$  is, of course, nonnegative because the adjacency matrices  $A$  and  $B$  are always nonnegative. To see that  $M$  is symmetric, first notice that  $M$  can be rewritten using Lemma 2.2.13 to:

$$M = A \otimes B + (A \otimes B)^T,$$

because  $A \otimes B$  is a  $nm \times nm$  square matrix, we know that  $M$  is symmetric because it's the sum of a square matrix and its transpose. In order to stay consistent with the vector norm appearing in Theorem 2.2.10, the matrix norm  $\|\cdot\|_F$  or the *Frobenius* norm is the square root of the sum of all squared entries and the 1-norm  $\|\cdot\|_1$  is the sum of all magnitudes of the entries. We can now apply Theorem 2.2.10 and the result follows.  $\square$

<sup>5</sup>  $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n (a_{ij})^2}$  is the *Frobenius* norm for a  $m \times n$ -matrix  $A$ .

<sup>6</sup>  $\|A\|_{1,1} = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|$  is the 1-norm for a  $m \times n$ -matrix  $A$

**Definition 2.2.16.** Let  $\mathcal{G}, \mathcal{H}$  be two directed graphs, then the unique matrix

$$S = Z_{\text{even}}(\mathbf{1}) = \lim_{k \rightarrow +\infty} Z^{(2k)}$$

(with the notations of the previous theorem) is the **similarity matrix** between  $\mathcal{G}$  and  $\mathcal{H}$ .

Notice that it follows from 2.9 that the similarity matrix between  $\mathcal{H}$  and  $\mathcal{G}$  is the transpose of the similarity matrix between  $\mathcal{G}$  and  $\mathcal{H}$ .

### 2.2.3 Algorithm & Computational cost

**Data:**

$A$ : the  $n \times n$  adjacency matrix of a directed graph  $\mathcal{G}$

$B$ : the  $m \times m$  adjacency matrix of a directed graph  $\mathcal{H}$ .

TOL: tolerance for the estimation error.

**Result:**

$S$ : the similarity matrix between  $\mathcal{G}$  and  $\mathcal{H}$ .

**begin** similarity\_matrix( $A, B, TOL$ )

$k = 1$  ;

$Z^{(0)} = \mathbf{1}$  ( $n \times m$ -matrix with all entries equal to 1);

$\mu = n \times m$ -matrix with all entries equal to TOL;

**repeat**

$$Z^{(k)} = \frac{BZ^{(k-1)}A^T + B^T Z^{(k-1)}A}{\|BZ^{(k-1)}A^T + B^T Z^{(k-1)}A\|_F};$$

$k = k + 1$ ;

**until**  $k$  is even and  $|Z - Z^{(k-2)}| < \mu$ ;

**return**  $Z^{(k)}$ ;

**end**

**Algorithm 5:** Algorithm for calculating the similarity matrix between  $\mathcal{G}$  and  $\mathcal{H}$

The compact form introduced in the previous section leads directly to the approximation algorithm 5. Note that the algorithm must iterate an even number of times and to check if the tolerance limit is reached, only the even iteration steps are considered. This is of course an obvious consequence of the definition of the similarity matrix. A Matlab implementation of the algorithm can be found in Listing A.2 in Appendix A.

To estimate the computational cost, we first notice that the algorithm is a matrix analogue to the power method introduced in section 1.3.2. We already showed in Theorem 2.2.10 that the convergence of the even iterates in the above recurrence relation is linear with ratio  $(\mu/\rho)^{2n}$ . With the accuracy level TOL, we become: **Ik weet dat ik  $\approx$  nog moet definiëren, maar ik weet niet goed hoe.**

$$\left| \frac{\mu}{\rho} \right|^{2n} \approx \text{TOL}$$

So, for example, for  $\text{TOL} = 10^{-5}$  we get  $n = -5/(2 \log \mu/\lambda)$ , we become:

$$\text{similarity\_matrix} \in O\left(\frac{-1}{\log \mu - \log \rho}\right)$$

Note that this  $O$  holds for any estimation error TOL of the form  $10^{-e}$  with  $e \in \mathbb{N}$ . Also note that this is completely analogous to the computational cost of the power method, which



comes as no surprise since the algorithms are similar, but remember from Theorem 2.2.7 that we cannot encounter a situation in which  $\mu = \rho$ , since we know that  $\rho$  is the unique dominant eigenvalue (in contrast to the power method, where we don't know in general whether a matrix has a unique dominant eigenvalue or not).

### 2.2.4 Special cases

We now consider some special cases of similarity scores between two vertices of graphs.

#### HITS-algorithm

Remember that the HITS-algorithm assumed that

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

has a unique dominant eigenvalue to calculate the hub and authority scores of nodes in a graph  $\mathcal{G}_\sigma$  with adjacency matrix  $B$ . This assumption is a direct result of the power method, because it's one of the conditions to apply the algorithm. One of our goals of this section was to drop this assumption, making the concept of similarity more robust. We finally arrived at this result.

**Theorem 2.2.17.** *Let  $\mathcal{G}$  be a graph with adjacency matrix  $B$ . The normalized hub and authority scores of the vertices are given by the normalized dominant eigenvectors of the matrices  $BB^T$  and  $B^TB$ , provided the corresponding Perron root is of multiplicity 1. Otherwise, it is the normalized projection of the vector  $\mathbf{1}$  on the eigenspace of the Perron root.*

*Proof.* The corresponding matrix  $M$  is:

$$M = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$$

so:

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

and the result follows from Theorem 2.2.10 under the condition that the matrix  $M^2$  has a dominant root  $\rho^2$ . This can be seen as follows: let  $V$  and  $U$  be orthonormal bases for the dominant eigenspaces of  $B$ :

$$BV = \rho U, \quad B^T U = \rho V$$

then clearly  $V$  and  $U$  are also bases for the dominant eigenspaces of  $B^TB$  and  $BB^T$  so:

$$B^T B V = \rho^2 V, \quad B B^T U = \rho^2 U.$$

The projectors associated with the dominant eigenvalues of  $BB^T$  and  $B^TB$  are, respectively:

$$\Pi_v = VV^T \quad \text{and} \quad \Pi_u = UU^T,$$

the projector of  $\Pi$  of  $M^2$  is then:

$$\Pi = \text{diag}(\Pi_v, \Pi_u),$$

and hence the subvectors of  $\Pi \mathbf{1}$  are the vectors  $\Pi_v \mathbf{1}$  and  $\Pi_u \mathbf{1}$ , which can be computed with  $B^TB$  and  $BB^T$ .  $\square$

### Central scores

As for the hub and authority scores we can give an explicit expression for the similarity score with vertex 2, which we will call the *central score*.

**Theorem 2.2.18.** *Let  $\mathcal{G}$  be a graph with adjacency matrix  $B$ . The normalized similarity scores of vertex 2 of the path graph  $1 \rightarrow 2 \rightarrow 3$  with the vertices of graph  $\mathcal{G}$  are called the central scores and are given by the normalized dominant eigenvector of the matrix  $B^T B + B B^T$ , provided the corresponding Perron root is of multiplicity 1. Otherwise, it is the normalized projection of the vector 1 on the dominant invariant subspace.*

*Proof.* Let

$$M = \begin{pmatrix} 0 & B & 0 \\ B^T & 0 & B \\ 0 & B^T & 0 \end{pmatrix}$$

so:

$$M^2 = \begin{pmatrix} B B^T & 0 & B B \\ 0 & B^T B + B B^T & 0 \\ B^T B^T & 0 & B^T B \end{pmatrix}$$

because  $M$  is nonnegative and symmetric, the result follows from Theorem 2.2.10 under the condition that the central matrix  $B^T B + B B^T$  has a dominant root  $\rho^2$  of  $M^2$ . We can state this condition otherwise, because  $M$  can be permuted to:

$$M = P^T \begin{pmatrix} 0 & E \\ E^T & 0 \end{pmatrix} P, \quad \text{where } E = \begin{pmatrix} B \\ B^T \end{pmatrix}$$

Now let  $V$  and  $U$  be orthonormal bases for the dominant eigenspaces of  $E^T E$  and  $E E^T$ , since:

$$E^T E V = \rho^2 V, \quad E E^T U = \rho^2 U.$$

Moreover,

$$P M^2 P^T = \begin{pmatrix} E E^T & 0 \\ 0 & E^T E \end{pmatrix},$$

and let

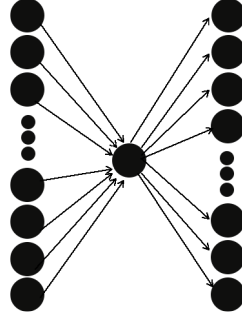
$$\Pi_v = V V^T \quad \text{and} \quad \Pi_u = U U^T$$

be the projectors associated with the dominant eigenvalues of  $E E^T$  and  $E^T E$ . The project  $\Pi$  of  $M^2$  is then equal to:

$$\Pi = P^T \text{diag}(\Pi_v, \Pi_u) P,$$

and it follows that the subvectors of  $\Pi \mathbf{1}$  are the vectors  $\Pi_v \mathbf{1}$  and  $\Pi_u \mathbf{1}$ , which can be computed from  $E^T E$  or  $E E^T$ . Since  $E^T E = B^T B + B B^T$ , the central vector  $\Pi_v \mathbf{1}$  is the central vector of  $\Pi \mathbf{1}$ .  $\square$

**Example 2.2.19.** In order to illustrate the intuitive meaning of calculating a similarity matrix where the path graph  $1 \rightarrow 2 \rightarrow 3$  is the structure graph, consider the following directed bow-tie graph:



Label the center vertex first, then the  $m$  left vertices and finally the  $n$  right vertices, then the adjacency matrix of the bow-tie graph becomes:

$$B = \left( \begin{array}{c|ccc|ccc} 0 & 0 & \dots & 0 & 1 & \dots & 1 \\ \hline 1 & & & & & & \\ \vdots & & \mathbf{0}_n & & & & \mathbf{0} \\ \hline 1 & & & & & & \\ \hline 0 & & & & & & \\ \vdots & & \mathbf{0} & & & \mathbf{0}_m & \\ \hline 0 & & & & & & \end{array} \right)$$

By direct computation, we get that the matrix  $B^T B + B B^T$  is equal to:

$$B^T B + B B^T = \begin{pmatrix} m+n & 0 & 0 \\ 0 & \mathbf{1}_n & 0 \\ 0 & 0 & \mathbf{1}_m \end{pmatrix}$$

By Theorem 2.2.18, the Perron root of  $M$  is equal to  $\rho = \sqrt{n+m}$  and the similarity matrix is the  $(1+m+n) \times 3$ -matrix:

$$S = \frac{1}{\sqrt{m+n+1}} \left( \begin{array}{c|c|c} 0 & 1 & 0 \\ \hline 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ \hline 1 & 0 & 0 \\ \hline 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ \hline 0 & 0 & 1 \end{array} \right)$$

If we see vertex 2 of the path graph  $1 \rightarrow 2 \rightarrow 3$  as the *center*, which can be seen as a vertex through which much information is passed, then it is not surprising that  $S$  indicates that vertex 1 of the directed bow-tie graph is the only one that looks like a center. The left vertices of the bow-tie graph look like vertex 1 of the path graph and the right vertices of the bow-tie graph look like vertex 3. This is a beautiful example because it confirms our intuition.

### Self-similarity of a graph

When we compute the similarity matrix of two equal graphs  $\mathcal{G} = \mathcal{H}$ , the similarity matrix  $S$  is square matrix with as entries the similarity scores between vertices of  $\mathcal{G}$ , we call  $S$  the *self-similarity matrix* of  $\mathcal{G}$  in this case.

Intuitively, we expect that vertices have the highest similarity scores with themselves, which means that the largest entries are located on the diagonal of  $S$ . We prove in the next theorem that the largest entry of a self-similarity matrix appears always on the diagonal and that, except for some special cases, the diagonal elements of a self-similarity matrix are nonzero. This doesn't mean that the diagonal elements are always larger than the other elements on the same row or column and we conclude this paragraph with some easy examples the show this.

**Theorem 2.2.20.** *The self-similarity matrix of a graph  $\mathcal{G}$  is positive semidefinite<sup>7</sup>. The largest entry of the matrix appears on the diagonal and if a diagonal entry is equal to zero, all the entries of the corresponding row and column are also equal to zero.*

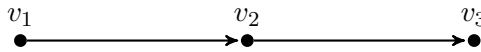
**Hier ga ik nog enkele relevante eigenschappen van positieve semidefinitie matrices toevoegen, dit bewijs toch net iets te kort door de bocht me dunkt.**

*Proof.* Since  $A = B$ , the compact form of Theorem 2.2.15 becomes:

$$Z^{(k+1)} = \frac{BZ^{(k)}A^T + B^T Z^{(k)}A}{\|BZ^{(k)}A^T + B^T Z^{(k)}A\|_F}, \quad Z^{(0)} = \mathbf{1},$$

It is clear that all matrices  $Z^{(k)}$  are positive semidefinite because the scaled sum of two positive semidefinite matrices is also positive semidefinite. Since the positive semidefinite matrices form a closed set (in the usual topology), the limit  $S$  will also be positive semidefinite. The rest of the theorem are well known properties of positive semidefinite matrices.  $\square$

**Example 2.2.21.** The self-similarity matrix of the graph:



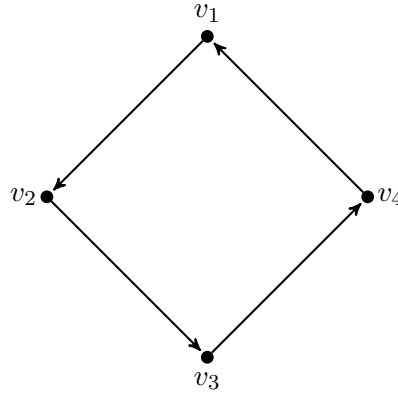
is equal to:

$$\begin{pmatrix} 0.5774 & 0 & 0 \\ 0 & 0.5774 & 0 \\ 0 & 0 & 0.5774 \end{pmatrix}$$

**Example 2.2.22.** The self-similarity matrix of the graph:

---

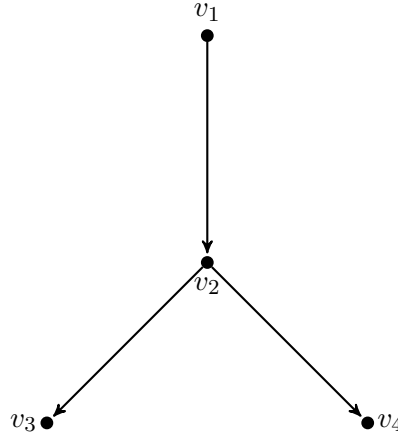
<sup>7</sup>A  $n \times n$ -matrix  $A$  is called positive-semidefinite if  $\mathbf{x}^T A \mathbf{x} \geq 0$  for all  $\mathbf{x}$  in  $\mathbb{R}^n$ .



is equal to:

$$\begin{pmatrix} 0.250 & 0.250 & 0.250 & 0.250 \\ 0.250 & 0.250 & 0.250 & 0.250 \\ 0.250 & 0.250 & 0.250 & 0.250 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{pmatrix}$$

**Example 2.2.23.** The self-similarity matrix of the graph:



is equal to:

$$\begin{pmatrix} 0.4082 & 0 & 0 & 0 \\ 0 & 0.4082 & 0 & 0 \\ 0 & 0 & 0.4082 & 0.4082 \\ 0 & 0 & 0.4082 & 0.4082 \end{pmatrix}$$

## 2.3 Node-edge similarity

The paper of Blondel et al. [BLONDEL] caused a flow of successive papers which build upon the concept of similarity between two graphs. For instance, the paper ‘Graph similarity scoring and matching’ of Laura A Zager and Goerge C. Verghese [ZAGER] expands the notion of node similarity presented in the previous section to similarity between the edges of two graphs. The paper was presented in 2006. Intuitively an edge of a graph  $\mathcal{G}$  is similar to an edge of graph  $\mathcal{H}$  if their *source* and *terminal nodes* are similar. As a consequence, the notion of similarity between edges introduces a coupling between edge and node similarity scores.

The algorithm presented in this paper is therefore an extension of the algorithm presented in the previous section.

### 2.3.1 Coupled node-edge similarity scores

We now present the extended algorithm allowing us to calculate not only a node similarity scores, but also edge similarity scores. The algorithm will use a new sort of matrices that represent a graph. Recall Definition 1.4.5 of source and terminal nodes.

**Definition 2.3.1.** Let be  $\mathcal{G} = (V, \rightarrow)$  be a graph with adjacency matrix  $A$ , vertices  $v_1, v_2, \dots, v_n \in V$  and edges  $(v_{x_1}, v_{y_1})_1, (v_{x_1}, v_{y_1})_2, \dots, (v_{x_m}, v_{y_m})_m \in \rightarrow$  (with  $(v_{x_i}, v_{y_i}) \in V$ ). We define the **source-edge matrix**  $A_S$  as a  $n \times m$  matrix with entries:

$$(A_S)_{ij} = \begin{cases} 1 & \text{if } s_{\mathcal{G}}(j) = v_i \\ 0 & \text{otherwise} \end{cases},$$

the notation  $A_S$  is derived from the adjacency matrix  $A$ , when the graph would have another adjacency matrix, for instance  $B$ , we would write  $B_S$ .

**Definition 2.3.2.** Let be  $\mathcal{G} = (V, \rightarrow)$  be a graph with adjacency matrix  $A$ , vertices  $v_1, v_2, \dots, v_n \in V$  and edges  $(v_{x_1}, v_{y_1})_1, (v_{x_1}, v_{y_1})_2, \dots, (v_{x_m}, v_{y_m})_m \in \rightarrow$  (with  $(v_{x_i}, v_{y_i}) \in V$ ). We define the **terminus-edge matrix**  $A_T$  as a  $n \times m$  matrix with entries:

$$(A_T)_{ij} = \begin{cases} 1 & \text{if } t_{\mathcal{G}}(j) = v_i \\ 0 & \text{otherwise} \end{cases},$$

the notation  $A_T$  is derived from the adjacency matrix  $A$ , when the graph would have another adjacency matrix, for instance  $B$ , we would write  $B_T$ .

**Property 2.3.3.** Let  $\mathcal{G} = (V, \rightarrow)$  be a graph, then  $A_S A_S^T$  is a diagonal matrix where the  $i$ th diagonal entry is equal the outdegree of vertex  $v_i$ .

*Proof.* By direct computation, we get:

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

Assume  $i \neq j$ . Then for each  $k$ ,  $(A_S)_{ik} (A_S)_{jk} = 0$  since vertex  $v_i$  and vertex  $v_j$  can't be both the source node of edge  $k$ .

When  $i = j$ , then for each  $k$ ,  $(A_S)_{ik} (A_S)_{jk} = ((A_S)_{ik})^2$  equals 0 or 1 depending on whether  $v_i$  is a starting point or not, so 1 is added to  $(A_S A_S^T)_{ii}$  each time an edge 'departs' from  $v_i$ , this is exactly the outdegree of vertex  $v_i$ .  $\square$

In an analogous way, we prove:

**Property 2.3.4.** Let  $\mathcal{G} = (V, \rightarrow)$  be a graph, then  $A_T A_T^T$  is a diagonal matrix where the  $i$ th diagonal entry is equal the indegree of vertex  $v_i$ .

**Property 2.3.5.** Let  $\mathcal{G} = (V, \rightarrow)$  be a graph, then the adjacency matrix  $A$  is equal to  $A_S A_T^T$ .

*Proof.* By direct computation, we get:

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

The terms  $(A_S)_{ik} (A_T)_{jk}$  equal 1 if edge  $k$  goes from  $v_i$  to  $v_j$  and since the sum is taken over all edges we conclude:

$$(A_S A_T^T)_{ij} = A_{ij}.$$

□

Let  $\mathcal{G}(V, \rightarrow)$ ,  $\mathcal{H}(U, \rightarrow')$  be two (directed) graphs,  $\mathcal{G}$  has  $n_{\mathcal{G}}$  vertices and  $m_{\mathcal{G}}$  edges and  $\mathcal{H}$  has  $n_{\mathcal{H}}$  vertices and  $m_{\mathcal{H}}$ . Remember the following updating equation from (2.5), which returns the (node) similarity score between vertices  $u_i$  from  $\mathcal{H}$  and  $v_j$  from  $\mathcal{G}$ :  $\mathcal{G} = (U, \rightarrow)$ :

$$x_{ij}^{(k+1)} = \sum_{r: (u_r, u_i) \in \rightarrow', w: (v_w, v_j) \in \rightarrow} x_{rw}^{(k)} + \sum_{r: (u_i, u_r) \in \rightarrow', w: (v_j, v_w) \in \rightarrow} x_{rw}^{(k)},$$

if we number the edges of  $\mathcal{G}$  and  $\mathcal{H}$  as  $(v_{x'_1}, v_{y'_1})_1, (v_{x'_1}, v_{y'_1})_2, \dots, (v_{x'_{m_{\mathcal{G}}}}, v_{y'_{m_{\mathcal{G}}}})_{m_{\mathcal{G}}} \in \rightarrow$  (with  $(v_{x'_i}, v_{y'_i}) \in V$ ) and  $(u_{x'_1}, v_{y'_1})_1, (u_{x'_1}, v_{y'_1})_2, \dots, (u_{x'_{m_{\mathcal{H}}}}, v_{y'_{m_{\mathcal{H}}}})_{m_{\mathcal{H}}} \in \rightarrow'$  (with  $(u_{x'_i}, u_{y'_i}) \in U$ ), this can be rewritten as:

$$x_{ij}^{(k+1)} = \sum_{t_{\mathcal{H}}(p)=u_i, t_{\mathcal{G}}(q)=v_j} x_{s_{\mathcal{H}}(p)s_{\mathcal{G}}(q)}^{(k)} + \sum_{s_{\mathcal{H}}(p)=u_i, s_{\mathcal{G}}(q)=v_j} x_{t_{\mathcal{H}}(p)t_{\mathcal{G}}(q)}^{(k)}.$$

We now extend this mutually reinforcing relation with the notion of edge similarity.  $x_{ij}$  denotes again the node similarity score between vertex  $u_i$  from  $\mathcal{H}$  and vertex  $v_j$  from  $\mathcal{G}$  and  $y_{pq}$  denotes the edge similarity score between edge  $p$  from  $\mathcal{H}$  and edge  $q$  in  $\mathcal{G}$ , the update equations for edge and node similarity scores take now the following form:

$$y_{pq}^{(k+1)} = x_{s_{\mathcal{H}}(p)s_{\mathcal{G}}(q)}^{(k)} + x_{t_{\mathcal{H}}(p)t_{\mathcal{G}}(q)}^{(k)} \quad (2.10)$$

$$x_{ij}^{(k+1)} = \sum_{t_{\mathcal{H}}(r)=u_i, t_{\mathcal{G}}(w)=v_j} y_{rw}^{(k)} + \sum_{s_{\mathcal{H}}(t)=u_i, s_{\mathcal{G}}(w)=v_j} y_{rw}^{(k)} \quad (2.11)$$

With the same reasoning as presented in the previous section, these scores can be assembled into matrices  $Y^{(k)}$  and  $X^{(k)}$  by using the source-edge matrices  $A_S$  and the terminus-edge matrices  $A_T$ . Let  $A$  be the adjacency matrix of  $\mathcal{G}$  and  $B$  be the adjacency matrix of  $\mathcal{H}$ , let  $X^{(k)}$  be the  $n_{\mathcal{H}} \times n_{\mathcal{G}}$ -matrix with entries  $x_{ij}$ , the node similarity scores at iteration step  $k$  and let  $Y^{(k)}$  be the  $m_{\mathcal{H}} \times m_{\mathcal{G}}$ -matrix with entries  $y_{pq}$ , the edge similarity scores at step  $k$ . The equations (2.10) and (2.11) can be rewritten as:

$$Y'^{(k+1)} = B_S^T X^{(k)} A_S + B_T^T X^{(k)} A_T \quad (2.12)$$

$$X'^{(k+1)} = B_S Y^{(k)} A_S^T + B_T Y^{(k)} A_T^T \quad (2.13)$$

for  $k = 0, 1, \dots$

Of course we want to customize this equations in a way that we can apply Theorem 2.2.10 to prove convergence. This will be completely analogous to Theorem 2.2.15, but we can achieve a slightly better result: not only the even and odd iterates will converge, the iteration converges as a whole. Lastly, remember that one of the conditions of Theorem 2.2.10 is to normalize the results at each iteration step. Therefore, the following theorem comes as no surprise: **Bij dit bewijs moet nog een stuk bijgevoegd worden dat bewijst dat de uiteindelijke matrix M enkel positieve eigenwaarden heeft.**

**Theorem 2.3.6.** *Let  $\mathcal{G}$  and  $\mathcal{H}$  be two graphs with adjacency matrices  $A$  and  $B$ ,  $\mathcal{G}$  has  $n_{\mathcal{G}}$  vertices and  $m_{\mathcal{G}} \text{raf}$  edges and  $\mathcal{H}$  has  $n_{\mathcal{H}}$  vertices and  $m_{\mathcal{H}} \text{rafeen}$  edges, define:*

$$Y^{(k+1)} = \frac{B_S^T X^{(k)} A_S + B_T^T X^{(k)} A_T}{\|B_S^T X^{(k)} A_S + B_T^T X^{(k)} A_T\|_F} \quad (2.14)$$

$$X^{(k+1)} = \frac{B_S Y^{(k)} A_S^T + B_T Y^{(k)} A_T^T}{\|B_S Y^{(k)} A_S^T + B_T Y^{(k)} A_T^T\|_F} \quad (2.15)$$

for  $k = 0, 1, \dots$

Then the matrix subsequences  $X^{(2k)}$ ,  $Y^{(2k)}$  and  $X^{(k+1)}$ ,  $Y^{(k+1)}$  converge to  $X_{\text{even}}$ ,  $Y_{\text{even}}$  and  $X_{\text{odd}}$ ,  $Y_{\text{odd}}$ . If we take:

$$\begin{aligned} \mathbf{X}^{(0)} &= \mathbf{1} \in \mathbb{R}^{n_{\mathcal{H}} \times n_{\mathcal{G}}} \\ \mathbf{Y}^{(0)} &= \mathbf{1} \in \mathbb{R}^{m_{\mathcal{H}} \times m_{\mathcal{G}}} \end{aligned}$$

as initial matrices, then  $X_{\text{even}}(\mathbf{1})$ ,  $Y_{\text{even}}(\mathbf{1})$  are the unique matrices of largest 1-norm among all possible limits with positive initial matrices.

*Proof.* By Theorem 2.2.14 we can rewrite (2.12) as follows:

$$\begin{aligned} Y'^{(k+1)} &= B_S^T X'^{(k)} A_S + B_T^T X'^{(k)} A_T \\ \Leftrightarrow \text{vec}(Y'^{(k+1)}) &= \text{vec}(B_S^T X'^{(k)} A_S + B_T^T X'^{(k)} A_T) \\ \Leftrightarrow \text{vec}(Y'^{(k+1)}) &= \text{vec}(B_S^T X'^{(k)} A_S) + \text{vec}(B_T^T X'^{(k)} A_T) \\ \Leftrightarrow \text{vec}(Y'^{(k+1)}) &= (A_S^T \otimes B_S^T) \text{vec}(X'^{(k)}) + (A_T^T \otimes B_T^T) \text{vec}(X'^{(k)}) \\ \Leftrightarrow \text{vec}(Y'^{(k+1)}) &= (A_S^T \otimes B_S^T + A_T^T \otimes B_T^T) \text{vec}(X'^{(k)}) \end{aligned}$$

Completely analogous we can also rewrite (2.13):

$$\text{vec}(X'^{(k+1)}) = (A_S \otimes B_S + A_T \otimes B_T) \text{vec}(Y^k),$$

define  $\mathbf{y}^{(k)} = \text{vec}(Y'^{(k+1)})$  and  $\mathbf{x}^{(k)} = \text{vec}(X'^{(k+1)})$ , we get:

$$\begin{aligned} \mathbf{y}^{(k+1)} &= (A_S^T \otimes B_S^T + A_T^T \otimes B_T^T) \mathbf{x}^{(k)} \\ \mathbf{x}^{(k+1)} &= (A_S \otimes B_S + A_T \otimes B_T) \mathbf{y}^{(k)} \end{aligned}$$

If we define  $G = A_S^T \otimes B_S^T + A_T^T \otimes B_T^T$ , then with Lemma 2.2.13 and the well known properties of transpose matrices:

$$\begin{aligned} G^T &= (A_S^T \otimes B_S^T + A_T^T \otimes B_T^T)^T \\ &= ((A_S \otimes B_S)^T + (A_T \otimes B_T)^T)^T \\ &= ((A_S \otimes B_S)^T)^T + ((A_T \otimes B_T)^T)^T \\ &= A_S \otimes B_S + A_T \otimes B_T \end{aligned}$$

So we get:

$$\begin{aligned} \mathbf{y}^{(k+1)} &= G \mathbf{x}^{(k)} \\ \mathbf{x}^{(k+1)} &= G^T \mathbf{y}^{(k)}, \end{aligned}$$



$G$  is a  $n_{\mathcal{G}}n_{\mathcal{H}} \times m_{\mathcal{G}}m_{\mathcal{H}}$ -matrix, the previous expressions can be concatenated to a single matrix update equation (we define matrix  $M$  and  $\mathbf{z}^{(k+1)}$ ):

$$\mathbf{z}^{(k+1)} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^{(k+1)} = \begin{pmatrix} \mathbf{0}_{m_{\mathcal{G}}m_{\mathcal{H}}} & G^T \\ G & \mathbf{0}_{n_{\mathcal{G}}n_{\mathcal{H}}} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}^{(k)} = M\mathbf{z}^{(k)},$$

$M$  is clearly nonnegative because  $G$  and  $G^T$  consists of sums of Kronecker products of  $A_S, B_S, A^T$  and/or  $B^T$ , all matrices with entries equal to zero or one,  $M$  is as a  $(n_{\mathcal{G}}n_{\mathcal{H}} + m_{\mathcal{G}}m_{\mathcal{H}}) \times (n_{\mathcal{G}}n_{\mathcal{H}} + m_{\mathcal{G}}m_{\mathcal{H}})$ -matrix clearly symmetric, so the result follows immediately from Theorem 2.2.10. The appearance of the Perron norm can be explained in the same way as in Theorem 2.2.15, note that we write  $Y^{(k)}$  and  $X^{(k)}$  for the normalized results at iteration step  $k$ .  $\square$

We define now  $X_{\text{even}}(\mathbf{1})$  as the node similarity matrix and  $Y_{\text{even}}(\mathbf{1})$  as the edge similarity matrix.

### 2.3.2 The algorithm

**Data:**

$A_S$ : the  $n_{\mathcal{G}} \times m_{\mathcal{G}}$  source-edge matrix of a graph  $\mathcal{G}$

$A_T$ : the  $n_{\mathcal{G}} \times m_{\mathcal{G}}$  terminal-edge matrix of a graph  $\mathcal{G}$

$B_S$ : the  $n_{\mathcal{H}} \times m_{\mathcal{H}}$  source-edge matrix of a graph  $\mathcal{H}$

$B_T$ : the  $n_{\mathcal{H}} \times m_{\mathcal{H}}$  terminal-edge matrix of a graph  $\mathcal{H}$

TOL: tolerance for the estimation error.

**Result:**

$X$ : the node similarity matrix between  $\mathcal{G}$  and  $\mathcal{H}$

$Y$ : the edge similarity matrix between  $\mathcal{G}$  and  $\mathcal{H}$

**begin** node\_edge\_similarity\_matrix( $A_S, A_T, B_S, B_T, TOL$ )

$k = 1$  ;

$X^{(0)} = \mathbf{1}$  ( $n_{\mathcal{H}} \times n_{\mathcal{G}}$ -matrix with all entries equal to 1);

$Y^{(0)} = \mathbf{1}$  ( $m_{\mathcal{H}} \times m_{\mathcal{G}}$ -matrix with all entries equal to 1);

$\mu_X = n_{\mathcal{H}} \times n_{\mathcal{G}}$ -matrix with all entries equal to TOL;

$\mu_Y = m_{\mathcal{H}} \times m_{\mathcal{G}}$ -matrix with all entries equal to TOL;

**repeat**

$$Y^{(k)} = \frac{B_S^T X^{(k-1)} A_S + B_T^T X^{(k-1)} A_T}{\|B_S^T X^{(k-1)} A_S + B_T^T X^{(k-1)} A_T\|_F};$$

$$X^{(k)} = \frac{B_S Y^{(k)} A_S^T + B_T Y^{(k)} A_T^T}{\|B_S Y^{(k)} A_S^T + B_T Y^{(k)} A_T^T\|_F};$$

$k = k + 1$ ;

**until**  $|X^{(k)} - X^{(k-1)}| < \mu_X$  and  $|Y^{(k)} - Y^{(k-1)}| < \mu_Y$ ;

**return**  $X^{(k)}, Y^{(k)}$ ;

**end**

**Algorithm 6:** Algorithm for calculating the node and edge similarity matrix  $X$  and  $Y$  between  $\mathcal{G}$  and  $\mathcal{H}$ .

The algorithm to calculate the node and edge similarity scores is presented in pseudocode in Algorithm 6. Remember that the absolute value that is mentioned is the same as in Notation 1.1.8. Besides the difference calculations, the main difference with the algorithm of the previous section (Algorithm 5) is that the whole sequence converges, not only the even (or odd) iterates, making this algorithm twice as fast. Also notice that we implement a *sequential* update: when  $Y^{(k)}$  is calculated, it's result is immediately used in  $X^{(k)}$ . This is not according to Theorem 2.3.6, which suggest *simultaneous* updating equations: in that case  $X^{(k)}$  uses  $Y^{(k-1)}$  in it's calculations. It is not difficult to see, by an analogous argument of Theorem 2.3.6, that both approaches yield the same set of similarity scores. Numerical analysts, however, will always prefer the sequential update procedure because it performs slightly better.

A Matlab implementation can be found in Listing A.3 in Appendix A. Because giving source-edge matrices and terminal-edge matrices as input is quite unnatural, in Listing A.4 and Listing A.5 some Matlab code is also presented to transform an adjacency matrix into a source-edge matrix and a terminal-edge matrix. The resulting matrices represent an edge numbering left-to-right, based on the entries of the adjacency matrix. The algorithm to calculate the source-edge matrix based on the adjacency matrix is also written in pseudocode in Algorithm 7, the calculation of the terminal-edge matrix is completely analogous. A Matlab implementation of Algorithm 6 that takes two adjacency matrices as input can be found in Listing A.6.

**Data:**

$A$ : the  $n \times n$  adjacency matrix of a graph  $\mathcal{G}$

**Result:**

$A_S$ : the source-edge matrix of graph  $\mathcal{G}$

**begin** source\_edge\_matrix( $A$ )

```

     $m$  = sum of all entries of  $A$  ;
     $A_S$  = initialize a  $n \times m$ -matrix wit all entries equal to 0;
    current_edge = 0 ;
    for  $i : 1$  to  $n$  do
        for  $j : 1$  to  $n$  do
            if  $(A)_{ij} > 0$  then
                for  $e : 1$  to  $(A)_{ij}$  do
                     $(A_S)_{i, \text{current\_edge}}$  = 1;
                    current_edge = current_edge + 1 ;
                end
            end
        end
    end

```

**end**

**return**  $A_S$ ;

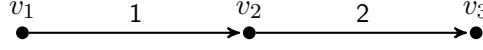
**Algorithm 7:** Algorithm for calculating the source-edge matrix  $A_S$  based on the adjacency matrix  $A$  of a graph  $\mathcal{G}$ .

### 2.3.3 Difference with node similarity

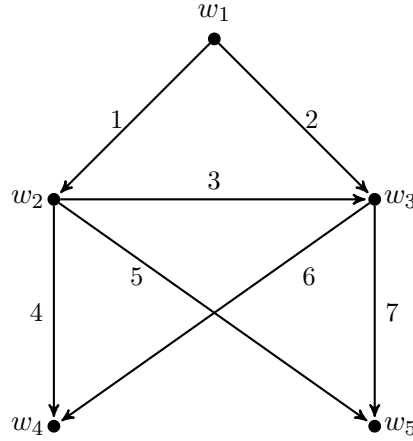
Dit wordt nog uitgewerkt

### 2.3.4 Example

**Example 2.3.7.** We retake Example 2.2.2 and number the edges as follows, let  $\mathcal{G} = (V, \rightarrow)$  be :



Let  $\mathcal{H} = (W, \rightarrow)$  be the following graph:



Then the node similarity matrix is:

$$X = \begin{pmatrix} 0.2338 & 0.0718 & 0 \\ 0.2472 & 0.3230 & 0.0128 \\ 0.1841 & 0.7553 & 0.3185 \\ 0 & 0.0935 & 0.2338 \\ 0 & 0.0441 & 0.0576 \end{pmatrix},$$

and the edge similarity matrix equals:

$$Y = \begin{pmatrix} 0.2166 & 0.0329 \\ 0.3847 & 0.1518 \\ 0.3899 & 0.2495 \\ 0.1325 & 0.2166 \\ 0.1133 & 0.1480 \\ 0.3653 & 0.4176 \\ 0.1080 & 0.3847 \end{pmatrix}$$

If you look at the structure of  $\mathcal{H}$  and compare it with the structure of  $\mathcal{G}$ , then intuitively it is not surprising that edge 3 of  $\mathcal{H}$  is most similar to edge 1 of  $\mathcal{G}$  and edge 6 of  $\mathcal{H}$  is most similar to edge 2 of  $\mathcal{G}$ .

## 2.4 Colored graphs

In this last section, we extend the notion of similarity to colored graphs. The paper [VANDOREN] is written by two Belgian professors Paul Van Dooren and Catherin Fraikin from the French speaking part of the Catholic University of Louvain in 2009.

Graph coloring is already introduced in paragraph 1.4.1 and we will construct a method that returns similarity matrices for two graphs where the coloring is on the nodes or on the edges of both graphs.

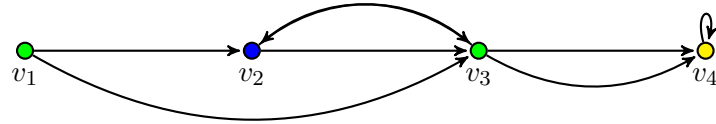
### 2.4.1 Colored nodes

We first extend the node similarity introduced in section 2.2 to node colored graphs. Take two graphs  $\mathcal{G} = (V, \rightarrow, C, a)$  and  $\mathcal{H} = (U, \rightarrow', C, a')$  with  $|C|$  (remember that  $a, a'$  are surjective) different colors and assume that the nodes in both graphs are renumbered such that those of color 1 come first, then those of color 2, ... The adjacency matrices  $A$  (of graph  $\mathcal{G}$ ) and  $B$  (of graph  $\mathcal{H}$ ) can then be partitioned as follows:

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1|C|} \\ A_{21} & A_{22} & \dots & A_{2|C|} \\ \vdots & \vdots & \ddots & \vdots \\ A_{|C|1} & A_{|C|2} & \dots & A_{|C||C|} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} B_{11} & B_{12} & \dots & B_{1|C|} \\ B_{21} & B_{22} & \dots & B_{2|C|} \\ \vdots & \vdots & \ddots & \vdots \\ B_{|C|1} & B_{|C|2} & \dots & B_{|C||C|} \end{pmatrix}$$

Remember that  $c_{\mathcal{G}}(i)$  denotes the number of vertices of color  $i$  in graph  $\mathcal{G}$ , then each block  $A_{ij} \in \mathbb{R}^{c_{\mathcal{G}}(i) \times c_{\mathcal{G}}(j)}$  and  $B_{ij} \in \mathbb{R}^{c_{\mathcal{H}}(i) \times c_{\mathcal{H}}(j)}$  describes the adjacency relations between the nodes of color  $i$  to the vertices of color  $j$  in both  $A$  and  $B$ . In fact, by just renumbering the vertices such that the nodes with the same color succeed each other, you immediately get such partitioning. If you see this renumbering as a permutation on the nodes, then performing on the original adjacency matrix a left and right multiplication with the corresponding permutation matrix (see Definition 1.1.1) leads to this adjusted adjacency matrix. To make the idea more comprehensible, we give an example.

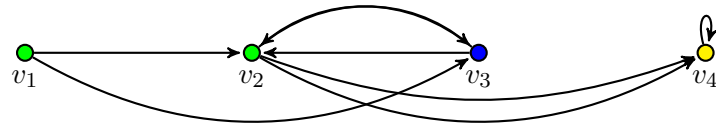
**Example 2.4.1.** Let  $\mathcal{G}$  be following graph:



The adjacency matrix equals:

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

If we order the colors as: {green, blue, yellow} (so color 1 = green, color 2 = blue, color 3 = yellow), we can renumber the vertices and get the following graph:



The adjacency matrix equals:

$$A' = PAP = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

which can be partitioned in blocks as follows (we have 3 colors: 1 = green, 2 = blue, 3 = yellow):

$$A' = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \begin{pmatrix} 0 \\ 2 \end{pmatrix} \\ \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix}.$$

Now, we only will compare the nodes of the same in both graphs, so that we define similarity matrices  $S_{ii} \in \mathbb{R}^{c_{\mathcal{G}}(i) \times c_{\mathcal{H}}(i)}$ , with  $i = 1, \dots, |C|$ , which we can put in a block-diagonal similarity matrix:

$$S = \begin{pmatrix} S_{11} & 0 & \dots & 0 \\ 0 & S_{12} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & S_{|C||C|} \end{pmatrix}$$

**Theorem 2.4.2.** *Let  $\mathcal{G} = (V, \rightarrow, C, a)$  and  $\mathcal{H} = (V, \rightarrow, C, a')$  be two node colored graphs and define:*

$$\begin{aligned} Z_1^{(k+1)} &= \sum_{i \in \{1, \dots, |C|\}} B_{1i} S_{ii}^{(k)} A_{1i}^T + B_{i1}^T S_{ii}^{(k)} A_{i1} \\ Z_2^{(k+1)} &= \sum_{i \in \{1, \dots, |C|\}} B_{2i} S_{ii}^{(k)} A_{2i}^T + B_{i2}^T S_{ii}^{(k)} A_{i2} \\ &\vdots \\ Z_{|C|}^{(k+1)} &= \sum_{i \in \{1, \dots, |C|\}} B_{|C|i} S_{ii}^{(k)} A_{|C|i}^T + B_{i|C|}^T S_{ii}^{(k)} A_{i|C|} \\ (S_{11}, S_{22}, \dots, S_{|C||C|})^{(k+1)} &= \frac{(Z_1^{(k+1)}, Z_2^{(k+1)}, \dots, Z_{|C|}^{(k+1)})}{\left\| (Z_1^{(k+1)}, Z_2^{(k+1)}, \dots, Z_{|C|}^{(k+1)}) \right\|_F} \end{aligned}$$

for  $k = 0, 1, \dots$

Then the the matrix subsequences  $Z_j^{(2k)}$  for every  $j \in \{1, \dots, |C|\}$  and  $(S_{11}, S_{22}, \dots, S_{|C||C|})^{(2k)}$  converge to  $Z_j^{even}$  and  $(S_{11}, S_{22}, \dots, S_{|C||C|})^{even}$ . Also the odd subsequences converge. If we take:

$$S_{jj}^{(0)} = \mathbf{1} \in \mathbb{R}^{c_{\mathcal{H}}(j) \times c_{\mathcal{G}}(j)}$$

as initial matrices, then the resulting  $S_{jj}^{even}(\mathbf{1})$  are the unique matrices of largest 1-norm among all possible limits with positive start vector.

*Proof.* By induction on  $|C|$ . Remember from Definition 1.4.15 that the function  $c$  is surjective, meaning that  $C$  only consists of colors that are actually used.

For  $|C| = 1$ , we have a graph with all vertices having the same color, which can be seen as an uncoloured graph. This just the normal case as described in section 2.2.

Although redundant, it is instructive to prove the case  $|C| = 2$ , because the generalization in the induction step is then immediately clear, so consider the partitioned adjacency matrices:

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

the equations of the theorem are in this case:

$$\begin{aligned} Z_1^{(k+1)} &= B_{11} S_{11}^{(k)} A_{11}^T + B_{11}^T S_{11}^{(k)} A_{11} + B_{12} S_{22}^{(k)} A_{12}^T + B_{21}^T S_{22}^{(k)} A_{21} \\ Z_2^{(k+1)} &= B_{21} S_{11}^{(k)} A_{21}^T + B_{12}^T S_{11}^{(k)} A_{12} + B_{22} S_{22}^{(k)} A_{22}^T + B_{22}^T S_{22}^{(k)} A_{22} \\ (S_{11}, S_{22})^{(k+1)} &= \frac{(Z_1^{(k+1)}, Z_2^{(k+1)})}{\|(Z_1^{(k+1)}, Z_2^{(k+1)})\|_F} \end{aligned}$$

Let  $Z_1'^{(k+1)}$  be the not normalized version  $Z^{(k+1)}$ , then we can write (see Theorem 2.2.14):

$$\begin{aligned} Z_1'^{(k+1)} &= B_{11} Z_1'^{(k)} A_{11}^T + B_{11}^T Z_1'^{(k)} A_{11} + B_{12} Z_2'^{(k)} A_{12}^T + B_{21}^T Z_2'^{(k)} A_{21} \\ \Leftrightarrow \text{vec}(Z_1'^{(k+1)}) &= \text{vec}(B_{11} Z_1'^{(k)} A_{11}^T + B_{11}^T Z_1'^{(k)} A_{11} + B_{12} Z_2'^{(k)} A_{12}^T + B_{21}^T Z_2'^{(k)} A_{21}) \\ \Leftrightarrow \text{vec}(Z_1'^{(k+1)}) &= (A_{11} \otimes B_{11}) \text{vec} Z_1'^{(k)} + (A_{11}^T \otimes B_{11}^T) \text{vec}(Z_1'^{(k)}) \\ &\quad + (A_{12} \otimes B_{12}) \text{vec}(Z_2'^{(k)}) + (A_{21}^T \otimes B_{21}^T) \text{vec}(Z_2'^{(k)}) \\ \Leftrightarrow \text{vec}(Z_1'^{(k+1)}) &= (A_{11} \otimes B_{11} + A_{11}^T \otimes B_{11}^T) \text{vec}(Z_1'^{(k)}) \\ &\quad + (A_{12} \otimes B_{12} + A_{21}^T \otimes B_{21}^T) \text{vec}(Z_2'^{(k)}) \end{aligned}$$

In a similar way, we can write  $Z_2'^{(k+1)}$ , which is the not normalized version of  $Z_2^{(k+1)}$ , as:

$$\text{vec}(Z_2'^{(k+1)}) = (A_{21} \otimes B_{21} + A_{12}^T \otimes B_{12}^T) \text{vec} Z_1'^{(k)} + (A_{22} \otimes B_{22} + A_{22}^T \otimes B_{22}^T) \text{vec} Z_2'^{(k)}$$

If we define  $M, \mathbf{z}^{k+1}$  as follows, the previous expressions concatenate to a single matrix update equation:

$$\begin{aligned} \mathbf{z}^{k+1} &= \begin{pmatrix} \text{vec}(Z_1') \\ \text{vec}(Z_2') \end{pmatrix}^{(k+1)} \\ &= \begin{pmatrix} A_{11} \otimes B_{11} + A_{11}^T \otimes B_{11}^T & A_{12} \otimes B_{12} + A_{21}^T \otimes B_{21}^T \\ A_{21} \otimes B_{21} + A_{12}^T \otimes B_{12}^T & A_{22} \otimes B_{22} + A_{22}^T \otimes B_{22}^T \end{pmatrix} \begin{pmatrix} \text{vec}(Z_1') \\ \text{vec}(Z_2') \end{pmatrix}^{(k)} \\ &= M \mathbf{z}^{(k)} \end{aligned}$$

Notice that the diagonal blocks in  $M$  are related to links between the nodes with the same color, while the off diagonal blocks refer to links between nodes of another color. As always, we want to use Theorem 2.2.10 to get the result.  $M$  is clearly nonnegative, because every block  $A_{ij}$  and  $B_{ij}$  is nonnegative (it's derived from the nonnegative adjacency matrices  $A$  and  $B$ ). Proving that  $M$  is symmetric a bit trickier, but notice that  $A_{11} \otimes B_{11} + (A_{11} \otimes B_{11})^T$

is a symmetric  $c_{\mathcal{G}}(1)c_{\mathcal{H}}(1) \times c_{\mathcal{G}}(1)c_{\mathcal{H}}(1)$  matrix (because it's the sum of a matrix with his transpose), and  $A_{22} \otimes B_{22} + (A_{22} \otimes B_{22})^T$  is a symmetric  $c_{\mathcal{G}}(2)c_{\mathcal{H}}(2) \times c_{\mathcal{G}}(2)c_{\mathcal{H}}(2)$ . If we define  $G = A_{21} \otimes B_{21} + A_{12}^T \otimes B_{12}^T$ ,  $G$  is a  $c_{\mathcal{G}}(2)c_{\mathcal{H}}(2) \times c_{\mathcal{G}}(1)c_{\mathcal{H}}(1)$ -matrix. Now, notice the following relation between the off diagonal blocks:

$$\begin{aligned} G^T &= (A_{21} \otimes B_{21} + A_{12}^T \otimes B_{12}^T)^T \\ &= (A_{21} \otimes B_{21})^T + ((A_{12}^T \otimes B_{12}^T)^T)^T \\ &= A_{12} \otimes B_{12} + A_{21}^T \otimes B_{21}^T \end{aligned}$$

$G^T$  is a  $c_{\mathcal{G}}(1)c_{\mathcal{H}}(1) \times c_{\mathcal{G}}(2)c_{\mathcal{H}}(2)$ -matrix. So we can rewrite  $M$  as:

$$M = \begin{pmatrix} A_{11} \otimes B_{11} + A_{11}^T \otimes B_{11}^T & G^T \\ G & A_{22} \otimes B_{22} + A_{22}^T \otimes B_{22}^T \end{pmatrix}$$

$M$  is a  $(c_{\mathcal{G}}(1)c_{\mathcal{H}}(1) + c_{\mathcal{G}}(2)c_{\mathcal{H}}(2)) \times (c_{\mathcal{G}}(1)c_{\mathcal{H}}(1) + c_{\mathcal{G}}(2)c_{\mathcal{H}}(2))$ -matrix, we want to prove that  $(M)_{ij} = (M)_{ji}$  and to do so we distinguish all possible cases:

- If  $i \leq c_{\mathcal{G}}(1)c_{\mathcal{H}}(1), j \leq c_{\mathcal{G}}(1)c_{\mathcal{H}}(1)$ , then  $(M)_{ij}$  and  $(M)_{ji}$  will both be entries of  $A_{11} \otimes B_{11} + A_{11}^T \otimes B_{11}^T$  and this submatrix was symmetric.
- If  $i \leq c_{\mathcal{G}}(1)c_{\mathcal{H}}(1), j > c_{\mathcal{G}}(1)c_{\mathcal{H}}(1)$ , then  $(M)_{ij}$  will be an entry of  $G^T$  and  $(M)_{ji}$  will be an entry of  $G$ , so they are equal.
- If  $i > c_{\mathcal{G}}(1)c_{\mathcal{H}}(1), j > c_{\mathcal{G}}(1)c_{\mathcal{H}}(1)$ , then  $(M)_{ij}$  and  $(M)_{ji}$  will both be entries of  $A_{22} \otimes B_{22} + A_{22}^T \otimes B_{22}^T$  and this submatrix was symmetric.
- If  $i > c_{\mathcal{G}}(1)c_{\mathcal{H}}(1), j \leq c_{\mathcal{G}}(1)c_{\mathcal{H}}(1)$ , then  $(M)_{ij}$  will be an entry of  $G$  and  $(M)_{ji}$  will be an entry of  $G^T$ , so they are equal.

The result immediately follows from Theorem 2.2.10. We motivated the usage of the Frobenius norm already in the proof of Theorem 2.2.15. Notice that the normalization after each iteration step happens ‘together’ by dividing by  $\|(Z_1^{(k+1)}, Z_2^{(k+1)})\|_F$ , because this is in accordance to the conditions of Theorem 2.2.10. Normalizing  $Z_1^{(k+1)}, Z_2^{(k+1)}$  separately after the expressions is calculated, gives an iterative process that is different from the one described in Theorem 2.2.10.

We now prove the induction step  $|C| = n - 1 \Rightarrow |C| = n$ . The only crucial thing to prove is that  $M$  is again symmetric, the rest of the steps consist of an easy expansion of the case  $|C| = 2$ .  $M$  is in this case equal to:

$$M = \begin{pmatrix} A_{11} \otimes B_{11} + A_{11}^T \otimes B_{11}^T & \cdots & A_{1(n-1)} \otimes B_{1(n-1)} + A_{(n-1)1}^T \otimes B_{(n-1)1}^T & A_{1n} \otimes B_{1n} + A_{n1}^T \otimes B_{n1}^T \\ \vdots & \ddots & \vdots & \vdots \\ A_{(n-1)1} \otimes B_{(n-1)1} + A_{1(n-1)}^T \otimes B_{1(n-1)}^T & \cdots & A_{(n-1)(n-1)} \otimes B_{(n-1)(n-1)} + A_{(n-1)(n-1)}^T \otimes B_{(n-1)(n-1)}^T & A_{(n-1)n} \otimes B_{(n-1)n} + A_{n(n-1)}^T \otimes B_{n(n-1)}^T \\ A_{n1} \otimes B_{n1} + A_{1n}^T \otimes B_{1n}^T & \cdots & A_{n(n-1)} \otimes B_{n(n-1)} + A_{(n-1)n}^T \otimes B_{(n-1)n}^T & A_{nn} \otimes B_{nn} + A_{nn}^T \otimes B_{nn}^T \end{pmatrix}$$

which can be seen as:

$$M = \begin{pmatrix} & & & A_{1n} \otimes B_{1n} \\ & & & +A_{n1}^T \otimes B_{n1}^T \\ & & M' & \vdots \\ & & & A_{(n-1)n} \otimes B_{(n-1)n} \\ & & & +A_{n(n-1)}^T \otimes B_{n(n-1)}^T \\ A_{n1} \otimes B_{n1} & & A_{n(n-1)} \otimes B_{n(n-1)} & A_{nn} \otimes B_{nn} \\ +A_{1n}^T \otimes B_{1n}^T & \cdots & +A_{(n-1)n}^T \otimes B_{(n-1)n}^T & +A_{nn}^T \otimes B_{nn}^T \end{pmatrix}$$

From the induction hypothesis we now that  $M'$  is symmetric. It's clear that the entries in the last column of  $M$  are the transpose of the entries in the last row. It follows that  $M$  is again symmetric.

□



# Appendix A

## Listings

Listing A.1: The MatLab code for the Power Method described in algorithm 1

```
function [ y, lambda, it_num ] = power_method ( n, a, y, it_max, tol )

% Force Y to be a column vector.
y = y ( : );
lambda = 0;

% Force Y to be a vector of unit norm.

y = y / norm ( y );

it_num = 0;

y_old = y;

ay = a * y;
lambda = y' * ay;
y = ay / norm ( ay );
if ( lambda < 0.0 )
    y = - y;
end

val_dif = 0.0;

cos_yly2 = y' * y_old;
sin_yly2 = sqrt ( ( 1.0 - cos_yly2 ) * ( 1.0 + cos_yly2 ) );

for it_num = 1 : it_max

    lambda_old = lambda;
    y_old = y;

    ay = a * y;
    lambda = y' * ay;
    y = ay / norm ( ay );
    if ( lambda < 0.0 )
        y = - y;
    end
```

```

    val_dif = abs ( lambda - lambda_old );

    sin_yly2 = 0;
    cos_yly2 = y' * y_old;
    sin_yly2 = sqrt ( ( 1.0 - cos_yly2 ) * ( 1.0 + cos_yly2 ) );

    if ( val_dif <= tol )
        break
    end

end

y = ay / lambda;

return
end

```

Listing A.2: The MatLab code for algorithm 5.

```

function [Z] = similarity_matrix(A,B, TOL)
    Z_0 = ones(size(B,1),size(A,1));
    mu(1:size(B,1),1:size(A,1)) = TOL;
    Z = Z_0;
    Z_previouslyeven = Z_0;
    k=1;
    while true;
        Y = norm((B*Z*transpose(A)+transpose(B)*Z*A), 'fro');
        X = B*Z*transpose(A)+transpose(B)*Z*A;
        Z = X/Y;
        if mod(k,2) == 0
            difference = abs(Z-Z_previouslyeven);
            disp(difference);
            Z_previouslyeven = Z;
            if (difference < mu)
                break;
            end
        end
        k = k + 1;
    end
    return;
end

```

Listing A.3: The MatLab code for algorithm 6.

```

function [X, Y] = node_edge_similarity_matrices(AS, AT, BS, BT, TOL)
    X = ones(size(BS,1),size(AS,1));
    Y = ones(size(BS,2),size(AS,2));
    mu_X(1:size(BS,1),1:size(AS,1)) = TOL;
    mu_Y(1:size(BS,2),1:size(AS,2)) = TOL;
    X_previouslyeven = X;
    Y_previouslyeven = Y;
    k=1;
    while true;
        Y = (transpose(BS)*X*AS+transpose(BT)*X*AT)
            /norm(transpose(BS)*X*AS+transpose(BT)*X*AT, 'fro');
    end

```

```

X = (BS*Y*transpose(AS)+BT*Y*transpose(AT))
    /norm(BS*Y*transpose(AS)+BT*Y*transpose(AT), 'fro');

difference_X = abs(X-X_previous);
difference_Y = abs(Y-Y_previous);
X_previous = X;
Y_previous = Y;
    if difference_X < mu_X
        if difference_Y < mu_Y
            break;
        end
    end
end
return;
end

```

Listing A.4: The MatLab code to converse an adjacency matrix to a source-edge matrix (edges are numbered left-to-right based on the adjacency matrix).

```

function [AS] = source_edge_matrix(A)
    number_of_edges = sum(A(:));
    AS = zeros(size(A,1), number_of_edges); %initialize the souce-edge matrix A_S
    current_edge = 1;
    for i=1:size(A,1)
        for j=1:size(A,2)
            if A(i,j) > 0
                for e=1:A(i,j)
                    AS(i,current_edge) = 1;
                    current_edge = current_edge + 1;
                end
            end
        end
    end
end
end

```

Listing A.5: The MatLab code to converse an adjacency matrix to a terminal-edge matrix (edges are numbered left-to-right based on the adjacency matrix).

```

function [AS] = terminal_edge_matrix(A)
    number_of_edges = sum(A(:));
    AS = zeros(size(A,1), number_of_edges); %initialize the terminal-edge matrix A_T
    current_edge = 1;
    for i=1:size(A,1)
        for j=1:size(A,2)
            if A(i,j) > 0
                for e=1:A(i,j)
                    AS(j,current_edge) = 1;
                    current_edge = current_edge + 1;
                end
            end
        end
    end
end
end

```

Listing A.6: The MatLab code for algorithm 6 but with adjacency matrices as input (edges are numbered left-to-right based on the adjacency matrices).

```
function [X,Y] = node_edge_similarity_matrices_with_adjacency_matrix(A, B, TOL)
    AS = source_edge_matrix(A);
    AT = terminal_edge_matrix(A);
    BS = source_edge_matrix(B);
    BT = terminal_edge_matrix(B);
    [X,Y] = node_edge_similarity_matrix(AS, AT, BS, BT, TOL);
end
```

# Results of the Eurovision Song Contest 2009-2014

76

# 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.
- [BRUALDI] . A. Brualdi, H. J. Ryser, *Combinatorial Matrix Theory*, Cambridge University Press, 1991
- [LANCASTER] . 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 University 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 Institute 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.
- [EUROVISION] *Official website of the Eurovision Song Contest*, [www.eurovision.tv](http://www.eurovision.tv), May, 2015.
- [SAARI] D. G. Saari, *Mathematical Structure of Voting Paradoxes: II. Positional Voting*, Journal of Economic Theory, 2000.
- [FOUC] S. Foucart, *Linear Algebra and Matrix Analysis*, Drexel University, 2010.
- [HORN] R.A. Horn, C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [HORN2] R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, 1991.
- [ZAGER] L. A. Zager, George C. Verghese, *Graph similarity scoring and matching*, Applied Mathematics Letters, January, 2007. i Analysis  
, Cambridge University Press, 1991.
- [VANDOREN] P. Van Doren, C. Fraikin, *Similarity matrices for colored graphs*, Bulletin of the Belgian Mathematical Society Simon Stevin, 2009.