A MEASURE OF SIMILARITY BETWEEN GRAPH VERTICES. WITH APPLICATIONS TO SYNONYM EXTRACTION AND WEB SEARCHING

VINCENT D. BLONDEL AND PAUL VAN DOOREN

ABSTRACT. We introduce a concept of similarity between vertices of directed graphs. Let G_A and G_B be two directed graphs with respectively n_A and n_B vertices. We define a $n_A \times n_B$ *similarity matrix* $\mathbf{S}$ whose real entry s_{ij} expresses how similar vertex i (in G_A) is to vertex j (in G_B) : we say that s_{ij} is their *similarity score*. In the special case where $G_A = G_B = G$, the score s_{ij} is the similarity score between the vertices i and j of G and the square similarity matrix $\mathbf{S}$ is the *self-similarity matrix* of the graph G. We point out that Kleinberg's "hub and authority" method to identify web-pages relevant to a given query can be viewed as a special case of our definition in the case where one of the graphs has two vertices and a unique directed edge between them. In analogy to Kleinberg, we show that our similarity scores are given by the components of a dominant vector of a non-negative matrix and we propose a simple iterative method to compute them. Potential applications of our similarity concept are manifold and we illustrate one application for the automatic extraction of synonyms in a monolingual dictionary.

Keywords: Algorithms, graph algorithms, graph theory, eigenvalues of graphs.

AMS Mathematics Subject Classification (2000): 05C50, 05C85, 15A18, 68R10.

Preprint available from:
`http://www.inma.ucl.ac.be/~blondel/publications/`

Date: September 2002.

1. Generalizing hubs and authorities

Efficient web search engines such as Google are often based on the idea of characterizing the most important vertices in a graph representing the connections or links between pages on the web. One such method, proposed by Kleinberg [16], identifies in a set of pages relevant to a query search those that are good *hubs* or good *authorities*. For example, for the query "automobile makers", the home-pages of Ford, Toyota and other car makers are good authorities, whereas web pages that list these home-pages are good hubs. Good hubs are those that point to good authorities, and good authorities are those that are pointed to by good hubs. From these implicit relations, Kleinberg derives an iterative method that assigns an "authority score" and a "hub score" to every vertex of a given graph. These scores can be obtained as the limit of a converging iterative process which we now describe.

Let G be a graph with edge set E and let h_j and a_j be the hub and authority scores of the vertex j. We let these scores be initialized by some positive values and then update them simultaneously for all vertices according to the following *mutually reinforcing relation* : the hub score of vertex j is set equal to the sum of the authority scores of all vertices pointed to by j and, similarly, the authority score of vertex j is set equal to the sum of the hub scores of all vertices pointing to j :

$$\begin{cases} h_j & \leftarrow \quad \sum_{i:(j,i)\in E} a_i \\ a_j & \leftarrow \quad \sum_{i:(i,j)\in E} h_i \end{cases}$$

Let B be the adjacency matrix of G and let h and a be the vectors of hub and authority scores. The above updating equations take the simple form

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

which we denote in compact form by

$$x_{k+1} = M\, x_k, \qquad k = 0, 1, \ldots$$

where

$$x_k = \begin{bmatrix} h \\ a \end{bmatrix}_k, \quad M = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix}.$$

We are only interested in the relative scores and we will therefore consider the *normalized* vector sequence

$$z_0 = x_0, \quad z_{k+1} = \frac{Mz_k}{\|Mz_k\|_2}, \quad k = 0, 1, \ldots$$

Ideally, we would like to take the limit of the sequence z_k as a definition for the hub and authority scores. There are two difficulties with such a definition. Firstly, the sequence does not always converge. In fact, non-negative matrices M with the above block structure always have *two* real eigenvalue of largest magnitude and the resulting sequence z_k almost never converges. Notice however that the matrix M^2 is symmetric and so, even though the sequence z_k may not converge, the even and odd sub-sequences do converge. Let us define

$$z_{even} = \lim_{k\to\infty} z_{2k} \quad \text{and} \quad z_{odd} = \lim_{k\to\infty} z_{2k+1}.$$

and let us consider both limits for the moment. The second difficulty is that the limit vectors z_{even} and z_{odd} do in general depend on the initial vector z_0 and there is no apparent natural choice for z_0. In Theorem 2.2, we define the set of all limit vectors obtained when starting from a positive initial vector

$$Z = \{z_{even}(z_0), z_{odd}(z_0) : z_0 > 0\}.$$

and prove that the vector z_{even} obtained for $z_0 = \mathbf{1}$ is the vector of largest possible 1-norm among all vectors in Z (throughout this paper we denote by $\mathbf{1}$ the vector, or matrix, whose entries are all equal to 1; the appropriate dimension of $\mathbf{1}$ is always clear from the context). Because of this extremal property, we take the two sub-vectors of $z_{even}(\mathbf{1})$ as definitions for the hub and authority scores. In the case of the above matrix M, we have

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

and from this it follows that, if the dominant invariant subspaces associated to B^TB and BB^T have dimension one, then the normalized hub and authority scores are simply given by the normalized dominant eigenvectors of B^TB and BB^T, respectively. This is the definition used in [16] for the authority and hub scores of the vertices of G. The arbitrary choice of $z_0 = \mathbf{1}$ made in [16] is given here an extremal norm justification. Notice that when the invariant subspace has dimension one, then there is nothing particular about the starting vector $\mathbf{1}$ since any other positive vector z_0 would give the same result.

We now generalize this construction. The authority score of the vertex j of G can be seen as a similarity score between j and the vertex *authority* in the graph

$$hub \longrightarrow authority$$

and, similarly, the hub score of j can be seen as a similarity score between j and the vertex *hub*. The mutually reinforcing updating iteration used above can be generalized to graphs that are different from the hub-authority structure graph. The idea of this generalization is quite simple; we illustrate it in this introduction on the path graph with three vertices and provide a general definition for arbitrary graphs in Section 3. Let G be a graph with edge set E and adjacency matrix B and consider the *structure graph*

$$1 \longrightarrow 2 \longrightarrow 3.$$

To the vertex j of G we associate three scores x_{j1}, x_{j2} and x_{j3}; one for each vertex of the structure graph. We initialize these scores at some positive value and then update them according to the following mutually reinforcing relation

$$\begin{cases} x_{j1} & \leftarrow & & \sum_{i:(j,i)\in E} x_{i2} & \\ x_{j2} & \leftarrow & \sum_{i:(i,j)\in E} x_{i1} & & + \sum_{i:(j,i)\in E} x_{i3} \\ x_{j3} & \leftarrow & & \sum_{i:(i,j)\in E} x_{i2} & \end{cases}$$

or, in matrix form (we denote by x_i the column vector with entries x_{ji}),

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

which we again denote $x_{k+1} = Mx_k$. The situation is now identical to that of the previous example and all convergence arguments given there apply here as well. The matrix M^2 is symmetric and non-negative, the normalized even and odd iterates converge and the limit $z_{even}(\mathbf{1})$ is among all possible limits one that has largest possible 1-norm. We take the three components of this extremal limit $z_{even}(\mathbf{1})$ as definition for the similarity scores s_1, s_2 and s_3 and define the similarity matrix by

$$\mathbf{S} = [s_1\ s_2\ s_3].$$

(In Section 4, we prove that the "central similarity score" s_2 can be obtained more directly from B by computing the dominating eigenvector of the matrix $BB^T + B^TB$.)

The rest of this paper is organized as follows. In Section 2, we describe some standard Perron-Frobenius results for non-negative matrices that will be useful in the rest of the paper. In Section 3, we give a precise definition of the similarity matrix together with different alternative definitions. The definition immediately translates into an approximation algorithm and we discuss in that section some complexity aspects of this algorithm. In Section 4 we describe similarity matrices for the situation where one of the two graphs is a path graph; path graphs of lengths 2 and 3 are those that are discussed in this introduction. In Section 5, we consider the special case $G_A = G_B = G$ for which the score s_{ij} is the similarity between the vertices i and j in the graph G. Section 6 deals with vertex-transitive graphs. Vertex-transitive graphs are graphs for which all vertices play the same rôle; cycle graphs, for example, are vertex-transitive. We prove that when one of the graphs is vertex-transitive, then the similarity matrix has rank one. We also consider in that section the case where one of the two graphs is undirected and prove that the similarity matrix has then rank one. In a final section we report results obtained for the automatic synonym extraction in a dictionary by using the central score of a graph.

2. Graphs and non-negative matrices

With any directed graph $G = (V, E)$ one can associate a non-negative matrix via an indexation of its vertices. The so-called *adjacency matrix* of G is the matrix $B \in \mathbb{N}^{n\times n}$ whose entry d_{ij} equals the number of edges from vertex i to vertex j. Conversely, a square matrix B whose entries are non-negative integer numbers, defines a directed graph G with d_{ij} edges between i and j. Let B be the adjacency matrix of some graph G; the entry $(B^k)_{ij}$ is equal to the number of paths of length k from vertex i to vertex j. From this it follows 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 $(B^k)_{ij} > 0$. Matrices that satisfy this property are said to be *irreducible*.

The Perron-Frobenius theory [13] establishes interesting properties about the eigenvectors and eigenvalues for non-negative and irreducible matrices. Let us denote the spectral radius of the matrix C – i.e. the largest magnitude of its eigenvalues – by $\rho(C)$. The following results follow from [13, 3].

Theorem 2.1. *Let C be a non-negative matrix. Then*
(i) the spectral radius $\rho(C)$ is an eigenvalue of C – called the Perron root – and there exists an associated non-negative vector $x \geqslant 0$ ($x \neq 0$) – called the Perron vector – such that $Cx = \rho x$
(ii) if C is irreducible, then the algebraic multiplicity of the Perron root ρ is equal to one and there is a positive vector $x > 0$ such that $Cx = \rho x$
(iii) if C is symmetric, then the algebraic and geometric multiplicity

of the Perron root ρ are equal and there is a non-negative basis $X \geqslant 0$ associated to the invariant subspace associated to ρ, such that $CX = \rho X$.

Proof : The proofs of (i) and (ii) are given in e.g. [13, 3]. For (iii) we use the fact that every symmetric non-negative matrix C can be permuted to a block-diagonal matrix with irreducible blocks C_i on the diagonal [13, 3]. Let ρ be the spectral radius of C. The invariant subspace of C is obtained from the normalized Perron vectors of the C_i blocks, appropriately padded with zeros. The basis X one obtains that way is then non-negative and orthogonal. □

In the sequel, we shall also need the notion of orthogonal projection on vector subspaces. Let $\mathcal{V}$ be a linear subspace of $\mathbb{R}^n$ and let $v \in \mathbb{R}^n$. The *orthogonal projection* of v on $\mathcal{V}$ is the vector in $\mathcal{V}$ with smallest distance to v. The matrix representation of this projection is obtained as follows. Let $\{v_1, \ldots, v_m\}$ be an orthogonal basis for $\mathcal{V}$ and arrange these column vectors in a matrix V. The projection of v on $\mathcal{V}$ is then given by $\Pi v = VV^T v$ and the matrix $\Pi = VV^T$ is the *orthogonal projector* on $\mathcal{V}$. From the previous theorem it follows that, if the matrix C is non-negative and symmetric, then the elements of the orthogonal projector Π on the vector space associated to the Perron root of C are all non-negative.

The next theorem will be used to justify our definition of similarity matrix between two graphs. The result describes the limits points of sequences associated to symmetric non-negative linear transformations.

Theorem 2.2. *Let M be a symmetric non-negative matrix of spectral radius ρ. Let $z_0 > 0$ and consider the sequence*

$$z_{k+1} = Mz_k / \|Mz_k\|_2, \quad k = 0, \ldots$$

Then the subsequences z_{2k} and z_{2k+1} converge to the limits

$$z_{even}(z_0) = \lim_{k\to\infty} z_{2k} = \frac{\Pi z_0}{\|\Pi z_0\|_2},$$

and

$$z_{odd}(z_0) = \lim_{k\to\infty} z_{2k+1} = \frac{\Pi M z_0}{\|\Pi M z_0\|_2},$$

where Π is the orthogonal projector on the invariant subspace of M^2 associated to its Perron root ρ^2. In addition to this, the set of all possible limits is given by:

$$Z = \{z_{even}(z_0), z_{odd}(z_0) : z_0 > 0\} = \{\Pi z / \|\Pi z\|_2 : z > 0\}$$

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

Proof :
Let us denote the invariant subspaces of M corresponding to ρ, to $-\rho$ and to the rest of the spectrum, respectively by $\mathcal{V}_\rho$, $\mathcal{V}_{-\rho}$ and $\mathcal{V}_\mu$. Assume that these spaces are non-trivial, and that we have orthonormal bases for them :

$$MV_\rho = \rho V_\rho, \quad MV_{-\rho} = -\rho V_{-\rho}, \quad MV_\mu = V_\mu M_\mu, \tag{2.1}$$

where M_μ has a spectral radius μ strictly less than ρ. The eigenvalue decomposition can then be rewritten in block form :
(2.2)

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

It then follows that

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

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

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

and since $\rho(M_\mu) = \mu < \rho$ it follows from multiplying this by z_0 and Mz_0 that

$$z_{2k} - \frac{\Pi z_0}{\|\Pi z_0\|_2} = O(\mu/\rho)^{2k} \tag{2.3}$$

and

$$z_{2k+1} - \frac{\Pi M z_0}{\|\Pi M z_0\|_2} = O(\mu/\rho)^{2k} \tag{2.4}$$

provided the initial vectors z_0 and Mz_0 have a non-zero component in the relevant subspaces, i.e. provided Πz_0 and $\Pi M z_0$ are non-zero. But the 2-norm of these vectors equal $z_0^T \Pi z_0$ and $z_0^T M \Pi M z_0$ since $\Pi^2 = \Pi$. These norms are both non-zero since $z_0 > 0$ and both Π and $M\Pi M$ are non-negative.

It follows from the non-negativity of M and the formula for $z_{even}(z_0)$ and $z_{odd}(z_0)$ that both limits lie in $\{\Pi z/\|\Pi z\| : z > 0\}$. Let us now show that every element $\hat{z}_0 \in \{\Pi z/\|\Pi z\| : z > 0\}$ can be obtained as $z_{even}(z_0)$ for some $z_0 > 0$. Since the entries of Π are non-negative, so are those of $\hat{z}_0$. This vector may however have some of its entries equal to zero. From $\hat{z}_0$ we construct z_0 by adding ϵ to all the zero entries of $\hat{z}_0$. The vector $z_0 - \hat{z}_0$ is clearly orthogonal to $\mathcal{V}_\rho \oplus \mathcal{V}_{-\rho}$ and will therefore vanish in the iteration of M^2. Thus we have $z_{even}(z_0) = \hat{z}_0$ for $z_0 > 0$, as requested.

We now prove the last statement. The matrix Π and all vectors are non-negative 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

$$\left\| \frac{\Pi z_0}{\|\Pi z_0\|_2} \right\|_1 = \frac{\mathbf{1}^T \Pi^2 z_0}{\sqrt{{z_0}^T \Pi^2 z_0}}$$

Applying the Schwarz inequality to Πz_0 and $\Pi \mathbf{1}$ yields

$$|\mathbf{1}^T \Pi^2 z_0| \leqslant \sqrt{z_0^T \Pi^2 z_0}.\sqrt{\mathbf{1}^T \Pi^2 \mathbf{1}}.$$

with equality only when $\Pi z_0 = \lambda \Pi \mathbf{1}$ for some $\lambda \in \mathbb{R}$. From this the proof easily follows. $\square$

Corollary 2.3. *When $-\rho$ is not an eigenvalue of M, then the sequence z_k simply converges to*

$$z_{even}(z_0) = z_{odd}(z_0) = \frac{\Pi z_0}{\|\Pi z_0\|_2}.$$

Proof :
In this case $\mathcal{V}_{-\rho} = \{0\}$ and the invariant subspace of M^2 corresponding to its Perron root ρ^2 is also the invariant subspace of M corresponding to its Perron root ρ. For the rest, the proof follows the same lines as in the previous theorem. $\square$

3. Similarity between vertices in graphs

We now introduce our definition of graph similarity for arbitrary graphs. Let G_A and G_B be two directed graphs with respectively n_A and n_B vertices. We think of G_A as a "structure graph" that plays the role of the graphs *hub* $\longrightarrow$ *authority* and $1 \longrightarrow 2 \longrightarrow 3$ in the introductory examples. Let pre(v) (respectively post(v)) denote the set of ancestors (respectively descendants) of the vertex v. We consider real scores x_{ij} for $i = 1, \ldots, n_B$ and $j = 1, \ldots, n_A$ and simultaneously update all scores according to the following updating equations

$$[x_{ij}]_{k+1} = \sum_{r \in \mathrm{pre}(i),\, s \in \mathrm{pre}(j)} [x_{rs}]_k + \sum_{r \in \mathrm{post}(i),\, s \in \mathrm{post}(j)} [x_{rs}]_k \tag{3.5}$$

These equations coincide with those given in the introduction. The equations can be written in more compact matrix form. Let X_k be the $n_B \times n_A$ matrix of entries $[x_{ij}]_k$. Then (3.5) takes the form

$$X_{k+1} = BX_kA^T + B^TX_kA, \qquad k = 0, 1, \dots \tag{3.6}$$

where A and B are the adjacency matrices of G_A and G_B. In this updating equation, the entries of X_{k+1} depend linearly on those of X_k. We can make this dependance more explicit by using the matrix-to-vector operator that develops a matrix into a vector by taking its columns one by one. This operator, denoted vec, satisfies the elementary property $\operatorname{vec}(CXD) = (D^T \otimes C) \operatorname{vec}(X)$ in which $\otimes$ denotes the Kronecker tensorial product (also denoted tensorial, direct or categorial product). For a proof of this property, see Lemma 4.3.1 in [14]. Applying this property to (3.6) we immediately obtain

$$x_{k+1} = (A \otimes B + A^T \otimes B^T)\, x_k \tag{3.7}$$

where $x_k = \operatorname{vec}(X_k)$. This is the format used in the introduction. Combining this observation with Theorem 2.2 we deduce the following property for the normalized sequence Z_k.

Corollary 3.1. *Let G_A and G_B be two graphs with adjacency matrices A and B, fix some initial positive matrix $Z_0 > 0$ and define*

$$Z_{k+1} = \frac{BZ_kA^T + B^TZ_kA}{\|BZ_kA^T + B^TZ_kA\|_2} \qquad k = 0, 1, \dots .$$

Then, the matrix subsequences Z_{2k} and Z_{2k+1} converge to Z_{even} and Z_{odd}. Moreover, among all the matrices in the set

$$\{Z_{even}(Z_0), Z_{odd}(Z_0) : Z_0 > 0\}$$

the matrix $Z_{even}(\mathbf{1})$ is the unique matrix of largest 1-norm.

In order to be consistent with the vector norm appearing in Theorem 2.2, the matrix norm $\|.\|_2$ we use here is the square root of the sum of all squared entries (this norm is known as the Euclidean or Frobenius norm), and the 1-norm $\|.\|_1$ is the sum of all entries magnitudes. In view of this result, the next definition is now justified.

Definition 3.2. *Let G_A and G_B be two graphs with adjacency matrices A and B. The similarity matrix between G_A and G_B is the limit matrix*

$$\mathbf{S} = \lim_{k \to +\infty} Z_{2k}$$

obtained for $Z_0 = \mathbf{1}$ and

$$Z_{k+1} = \frac{BZ_kA^T + B^TZ_kA}{\|BZ_kA^T + B^TZ_kA\|_2}, \qquad k = 0, 1, \dots$$

Notice that it follows from this definition that the similarity matrix between G_B and G_A is the transpose of the similarity matrix between G_A and G_B. Similarity matrices can also be obtained directly by projecting the matrix $\mathbf{1}$ on an invariant subspace associated to the graphs. From Theorem 2.2 we easily obtain :

Corollary 3.3. *Let G_A and G_B be two graphs with adjacency matrices A and B and define*

$$M = A \otimes B + A^T \otimes B^T. \tag{3.8}$$

Consider the vector

$$\mathbf{s} = \frac{\Pi \mathbf{1}}{\|\Pi \mathbf{1}\|_2}$$

where Π is the orthogonal projector on the invariant subspace associated to the Perron root ρ^2 of M^2. Let this vector be partitioned as

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_{n_A} \end{bmatrix}$$

Then, the $n_B \times n_A$ similarity matrix between G_A and G_B is given by $\mathbf{S} = [\mathbf{s}_1 \ldots \mathbf{s}_{n_A}]$.

This theorem can also be stated in a compact matrix language using the matrix equivalent of the vector mapping $x \to Mx$ defined in (3.7) :

$$X \to \hat{M}(X) := BXA^T + B^T XA.$$

The correspondence $x = \mathrm{vec}(X)$ clearly preserves fixed points, and those of $\rho^2 X = \hat{M}^2(X)$ form a linear subspace, just as those of $\rho^2 x = M^2 x$. Let us denote the orthogonal projector onto this linear subspace by $\hat{\Pi}$, then the similarity matrix $\mathbf{S}$ is the normalized projection onto this subspace :

$$\mathbf{S} = \frac{\hat{\Pi}\mathbf{1}}{\|\hat{\Pi}\mathbf{1}\|_2}.$$

The similarity matrix can also be defined by its extremal property.

Corollary 3.4. *The similarity matrix of the graphs G_A and G_B of adjacency matrices A and B is the unique matrix of largest 1-norm among all matrices X that maximize the expression*

$$\frac{\|BXA^T + B^T XA\|_2}{\|X\|_2}. \tag{3.9}$$

Proof :
The above expression can also be written as $\|\hat{M}(X)\|_2/\|X\|_2$ which is the induced 2-norm of the linear transformation $\hat{M}(.)$. It is well known [13] that each dominant eigenvector X of $\hat{M}^2$ is a maximizer of this expression. It was shown above that $\mathbf{S}$ is the unique matrix of largest 1-norm in that set. $\square$

A direct algorithmic transcription of the definition leads to an approximation algorithm for computing similarity matrices of graphs. The complexity of this algorithm is easy to estimate. Let G_A, G_B be two graphs with n_A, n_B vertices and e_A, e_B edges. Then the products BZ_k and $B^T Z_k$ require less than $2n_A.e_B$ additions and multiplications each, while the subsequent products $(BZ_k)A^T$ and $(B^T Z_k)A$ require less than $2n_B.e_A$ additions and multiplications each. The sum and the calculation of the Frobenius norm requires $2n_A.n_B$ additions and multiplications, while the scaling requires 1 division and $n_A.n_B$ multiplications. Let us define

$$\alpha_A := e_A/n_A, \quad \alpha_B := e_B/n_B$$

as the average number of non-zero elements per row of A and B, respectively, then the total complexity per step of recurrence is of the order of

$$4(\alpha_A + \alpha_B)n_A n_B \text{ additions and multiplications.}$$

As was shown in Theorem 2.2, the convergence of the even iterates of the above recurrence is linear with ratio $(\mu/\rho)^2$. The number of floating point operations needed to compute $\mathbf{S}$ to ϵ accuracy is therefore of the order of

$$4n_A n_B \frac{(\alpha_A + \alpha_B)\log \epsilon}{(\log \mu - \log \rho)}.$$

For particular classes of adjacency matrices, one can compute the similarity matrix $\mathbf{S}$ directly from the dominant invariant subspaces of matrices of the size of A or B. We provide explicit expressions for a few classes in the next sections.

4. Hubs, authorities, central scores and path graphs

As explained in the introduction, the hub and authority scores of a graph G_B can be expressed in terms of the adjacency matrix of G_B.

Theorem 4.1. *Let B be the adjacency matrix of the graph G_B. The normalized hub and authority scores of the vertices of G_B are given by the normalized dominant eigenvectors of the matrices $B^T B$ and BB^T,* *provided the corresponding Perron root is of multiplicity 1. Otherwise, it is the normalized projection of the vector* $\mathbf{1}$ *on the respective dominant invariant subspaces.*

The condition on the multiplicity of the Perron root is not superfluous. Indeed, even for connected graphs, BB^T and B^TB may have multiple dominant roots: for cycle graph for example, both BB^T and B^TB are the identity matrix.

Another interesting structure graph is the path graph of length three:

$$1 \longrightarrow 2 \longrightarrow 3$$

Similarly to the hub and authority scores, the resulting similarity score with vertex 2, a score that we call *central score*, can be given an explicit expression. This central score has been successfully used for the purpose of automatic extraction of synonyms in a dictionary. This application is described in more details in Section 7.

Theorem 4.2. *Let* B *be the adjacency matrix of the graph* G_B*. The normalized central scores of the vertices of* G_B *are given by the normalized dominant eigenvector of the matrix*

$$B^TB + BB^T,$$

provided the corresponding Perron root is of multiplicity 1. Otherwise, it is the normalized projection of the vector $\mathbf{1}$ *on the dominant invariant subspace.*

Proof :
The corresponding matrix M is as follows :

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

and so

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

and the result then follows from the definition of the similarity scores, provided the central matrix $B^TB + BB^T$ has a dominant root ρ^2 of M^2. This can be seen as follows. The matrix M can be permuted to

$$PMP^T = \begin{bmatrix} 0 & E \\ E^T & 0 \end{bmatrix}, \quad \text{where } \; E := \begin{bmatrix} B \\ B^T \end{bmatrix}.$$

Let now V and U be orthonormal bases for the right and left singular subspaces of E [13] :

$$EV = \rho U, \quad E^T U = \rho V, \tag{4.10}$$

then clearly V and U are also the dominant invariant subspaces of $E^T E$ and EE^T, respectively, since

$$E^T E V = \rho^2 V, \quad E E^T U = \rho^2 U. \tag{4.11}$$

Moreover, $\Pi_v := VV^T$ and $\Pi_u := UU^T$ are the respective projectors of the block diagonal matrices $E^T E$ and EE^T of M^2 corresponding to its Perron root ρ^2. The projector Π is nothing but $diag\{\Pi_v, \Pi_u\}$ and hence the sub-vectors of $\Pi\mathbf{1}$ are the vectors $\Pi_v\mathbf{1}$ and $\Pi_u\mathbf{1}$, which can be computed from the smaller matrices $E^T E$ or EE^T. Since $E^T E = B^T B + BB^T$ the central vector $\Pi_v\mathbf{1}$ is the middle vector of $\Pi\mathbf{1}$. It is worth pointing out that (4.10) also yields a relation between the two smaller projectors :

$$\rho^2 \Pi_v = E^T \Pi_u E, \quad \rho^2 \Pi_u = E \Pi_v E^T.$$

□

In order to illustrate an interest of this structure graph over the *hub-authority* structure graph we consider here the special case of the "bow-tie graph" G_B represented in Figure 1. If we label the center vertex first, then label the m left vertices and finally the n right vertices, the adjacency matrix for this graph is given by :

$$B := \left[\begin{array}{c|c|c} 0 & 0 \cdots 0 & 1 \cdots 1 \\ \hline 1 & & \\ \vdots & \mathbf{0}_n & \mathbf{0} \\ 1 & & \\ \hline 0 & & \\ \vdots & \mathbf{0} & \mathbf{0}_m \\ 0 & & \end{array}\right] \tag{4.12}$$

The matrix $B^T B + BB^T$ is equal to

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

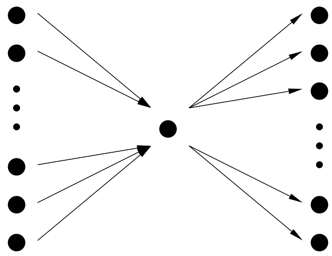


FIGURE 1. Bow-tie graph

and, following Theorem 4.2, the Perron root of M is equal to $\rho = \sqrt{n+m}$ and the similarity score matrix is given by

$$\mathbf{S} = \frac{1}{\sqrt{2}\rho} \left[\begin{array}{c|c|c} 0 & \rho & 0 \\ \hline 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ \hline 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{array} \right]. \tag{4.13}$$

This result holds irrespective of the relative value of m and n. Let us call the three vertices of G_A, 1, center and 3, respectively. One could view a *center* as a node through which much information is passed on. This similarity matrix $\mathbf{S}$ indicates that vertex 1 of G_B looks very much like a center, the left vertices of G_B look like 1's, and the right vertices of G_B look like 3's. If on the other hand we analyze the graph G_B with the hub-authority structure graph G_A of Kleinberg, then the similarity scores $\mathbf{S}$ differ for $m < n$ and $m > n$:

$$\mathbf{S}_{m>n} = \frac{1}{\sqrt{2}\rho} \left[\begin{array}{c|c} \rho & 0 \\ \hline 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ \hline 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{array} \right], \rho = \sqrt{m}, \quad \mathbf{S}_{m<n} = \frac{1}{\sqrt{2}\rho} \left[\begin{array}{c|c} 0 & \rho \\ \hline 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ \hline 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{array} \right], \rho = \sqrt{n}. \tag{4.14}$$

This shows a weakness of this structure graph, since the vertices of G_B that deserve the label of hub or authority completely change between $m > n$ and $m < n$.

The above structure graphs are path graphs of length 2 and 3. We now consider path graphs of arbitrary length ℓ.

Corollary 4.3. *Let B be the adjacency matrix of the graph G_B. Let G_A be the path graph of length ℓ:*

$$G_A \;:\quad 1 \longrightarrow 2 \longrightarrow \cdots \longrightarrow \ell.$$

Then the odd and even columns of the similarity matrix $\mathbf{S}$ can be computed independently as the projection of $\mathbf{1}$ on the dominant invariant subspaces of EE^T and E^TE where

$$E = \begin{bmatrix} B & & & \\ B^T & \ddots & & \\ & \ddots & B & \\ & & B^T & B \end{bmatrix} \quad \text{or} \quad E = \begin{bmatrix} B & & \\ B^T & \ddots & \\ & \ddots & B \\ & & B^T \end{bmatrix}$$

for ℓ even and ℓ odd, respectively.

Proof :
The matrix M corresponding to this configuration is as follows :

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

which can be permuted to $PMP^T = \begin{bmatrix} 0 & E \\ E^T & 0 \end{bmatrix}$, where

$$E = \begin{bmatrix} B & & & \\ B^T & \ddots & & \\ & \ddots & B & \\ & & B^T & B \end{bmatrix} \quad \text{or} \quad E = \begin{bmatrix} B & & \\ B^T & \ddots & \\ & \ddots & B \\ & & B^T \end{bmatrix}$$

for ℓ even and ℓ odd, respectively. We point out that this is the so-called "odd-even" ordering of the bipartite graph M. The rest of the proof is as in Theorem 4.2 □

5. Self-similarity matrix of a graph

When we compare two equal graphs $G_A = G_B = G$, the similarity matrix $\mathbf{S}$ is a square matrix whose entries are similarity scores between vertices of G; this matrix is the *self-similarity matrix* of G. Various graphs and their corresponding self-similarity matrices are represented in Figure 2. In general, we expect vertices to have a high similarity score with themselves; that is, we expect the diagonal entries of self-similarity matrices to be large. We prove in the next theorem that the largest entry of a self-similarity matrix always appear on the diagonal and that, except for trivial cases (see Example 2 on Figure 2), the diagonal elements of a self-similarity matrix are non-zero. As is shown with the Example 6 of Figure 2, it is however not true that diagonal elements dominate all elements on the same line and column.

Theorem 5.1. *The self-similarity matrix of a graph is always positive semi-definite. In particular, the largest element of the matrix always appears on diagonal, and if a diagonal entry is equal to zero, then the corresponding line and column are equal to zero.*

Proof
Since $A = B$, the iteration of the normalized matrices Z_k now becomes

$$Z_{k+1} = AZ_kA^T + A^T Z_k A/\|AZ_kA^T + A^T Z_k A\|_2, \quad Z_0 = \mathbf{1}.$$

Since the scaled sum of two positive semi-definite matrices is also positive semi-definite, it is clear that all matrices Z_k will be positive semi-definite. Moreover, positive semi-definite matrices are a closed set and hence the limit $\mathbf{S}$ will also be positive semi-definite. The properties mentioned in the Theorem are well known properties of positive semi-definite matrices. □

When vertices of a graph are similar to each other, such as in cycle graphs, we expect to have a self-similarity matrix whose entries are all equal. This is indeed the case. Let us recall here that a graph is said to be *vertex-transitive* (or *vertex symmetric*) if all vertices play the same rôle in the graph. More formally, a graph G of adjacency matrix A is vertex-transitive if associated to any pair of vertices i, j, there is a permutation matrix T that satisfies $T(i) = j$ and $T^{-1}AT = A$.

Theorem 5.2. *All entries of the self-similarity matrix of a vertex-transitive graph are equal to* $1/n$.

Proof
This is a direct consequence of Corollary 6.3 which will be proved in the next section. □

This theorem includes the case of cycle graphs. We can also derive explicit expressions for the self-similarity matrices of path graphs.

Theorem 5.3. *The self-similarity matrix of the path graph of length* ℓ *is a diagonal matrix with diagonal elements equal to* $\sin(j\pi/(\ell+1))$, $j = 1, \ldots, \ell$.

Proof
The product of two path graphs is a disjoint union of path graphs and so the matrix M corresponding to this graph can be permuted to a block diagonal arrangement of Jacobi matrices

$$J_j := \begin{bmatrix} 0 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & 0 & 1 \\ & & 1 & 0 \end{bmatrix}$$

of dimension $j = 1, \ldots \ell$. The largest of these blocks corresponds to the Perron root of M and eigenvalues on eigenvectors of these matrices are well-known [19]. The vertices of this block correspond to the diagonal elements of $\mathbf{S}$. $\square$

Example 3 in Figure 2 gives explicit numerical values for the case $\ell = 3$.

6. Graphs whose vertices are symmetric to each other

In this section we analyze properties of the similarity matrix when one of the two graphs (say G_A) has all its vertices symmetric to each other, or when the adjacency matrix of one of the graphs is normal. Vertex-transitive graphs have been introduced in the previous section. We refer the reader to [4] for the following lemma.

Lemma 6.1. *A vertex-transitive graph G_A has an adjacency matrix A with Perron root of algebraic multiplicity 1. The vector $\mathbf{1}$ is the corresponding Perron vector of both A and A^T.*

When G_A is vertex-transitive, the similarity matrix is a rank one matrix.

Theorem 6.2. *Let G_A, G_B be two graphs and assume that G_A is vertex-transitive. Then the similarity matrix between G_A and G_B is a rank one matrix of the form*

$$\mathbf{S} = \alpha\, \mathbf{1} v^T$$

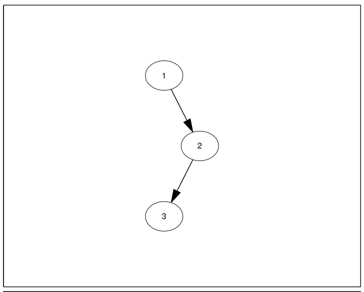
where $v = \Pi\mathbf{1}$ is the projection of $\mathbf{1}$ on the dominant invariant subspace of $(B + B^T)^2$ and α is the scaling factor $\alpha = 1/\|\mathbf{1}v^T\|$.

Proof
It easily follows from Lemma 6.1 that each matrix Z_k of the iteration 3.2 is of rank one and of the type $\mathbf{1}v_k^T/\sqrt{n_A}$, where

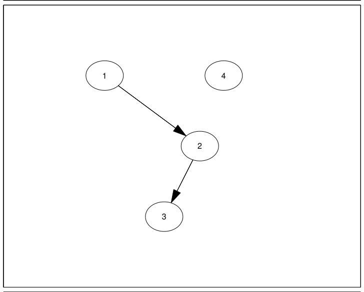
$$v_{k+1} = (B + B^T)v_k/\|(B + B^T)v_k\|_2, \quad v_0 = \mathbf{1}.$$

This clearly converges to $\Pi_\beta\mathbf{1}/\|\Pi_\beta\mathbf{1}\|_2$ where Π_β is the projector on the dominant invariant subspace of $(B + B^T)^2$. $\square$

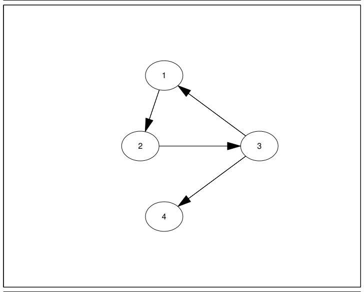
Corollary 6.3. *If G_A and G_B are vertex symmetric then the entries of their similarity matrix are all equal to $1/\sqrt{n_A n_B}$.*



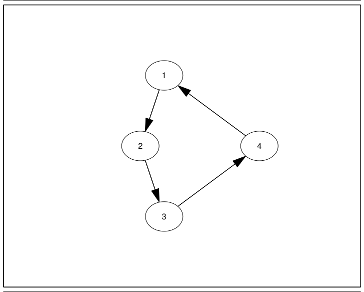
$$\begin{bmatrix} 0.408 & 0 & 0 \\ 0 & 0.816 & 0 \\ 0 & 0 & 0.408 \end{bmatrix}$$



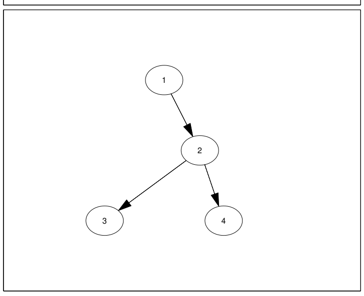
$$\begin{bmatrix} 0.408 & 0 & 0 & 0 \\ 0 & 0.816 & 0 & 0 \\ 0 & 0 & 0.408 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$



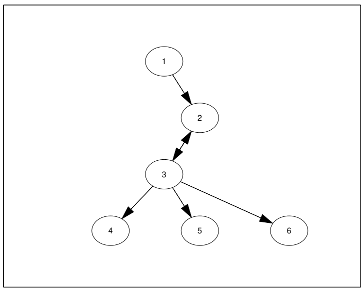
$$\begin{bmatrix} 0.431 & 0 & 0 & 0.259 \\ 0 & 0.431 & 0 & 0 \\ 0 & 0 & 0.652 & 0 \\ 0.259 & 0 & 0 & 0.259 \end{bmatrix}$$



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



$$\begin{bmatrix} 0.182 & 0 & 0 & 0 \\ 0 & 0.912 & 0 & 0 \\ 0 & 0 & 0.182 & 0.182 \\ 0 & 0 & 0.182 & 0.182 \end{bmatrix}$$



$$\begin{bmatrix} 0.103 & 0 & 0.251 & 0 & 0 & 0 \\ 0 & 0.202 & 0 & 0.096 & 0.096 & 0.096 \\ 0.251 & 0 & 0.845 & 0 & 0 & 0 \\ 0 & 0.096 & 0 & 0.074 & 0.074 & 0.074 \\ 0 & 0.096 & 0 & 0.074 & 0.074 & 0.074 \\ 0 & 0.096 & 0 & 0.074 & 0.074 & 0.074 \end{bmatrix}$$

FIGURE 2. Graphs and their corresponding self-similarity matrix.

Cycle graphs have an adjacency matrix A that satisfies $AA^T = I$. This property corresponds to the fact that, in a cycle graph, all forward-backward paths from a vertex return to that vertex. More generally, we consider in the next theorem graphs that have an adjacency matrix A that is normal, i.e., that have an adjacency matrix A such that $AA^T = A^T A$. In particular, graphs that have a symmetric adjacency matrix satisfy this property. We prove below that when one of the graphs has a normal adjacency matrix, then the similarity matrix has rank one; and we provide an explicit expression for this matrix.

Lemma 6.4. *Let A be normal and non-negative and let α be its Perron root. Then the projectors $\Pi_{+\alpha}$ and $\Pi_{-\alpha}$ on its invariant subspaces corresponding to the eigenvalues $+\alpha$ and $-\alpha$ are also the corresponding projectors for A^T.*

Proof
This follows from the fact that every real eigenvector of a real eigenvalue of a normal matrix is also an eigenvector of A^T. □

Theorem 6.5. *Let A be a normal non-negative matrix with Perron root α and let $\Pi_{+\alpha}$, $\Pi_{-\alpha}$ be the projectors on its invariant subspaces corresponding to the eigenvalues $+\alpha$ and $-\alpha$. Let β be the Perron root of $(B + B^T)$, and let Π_β be the projector on the invariant subspace of $(B + B^T)^2$ corresponding to the eigenvalue β^2. Then $M = (A \otimes B + A^T \otimes B^T)$ has spectral radius $\rho = \alpha \cdot \beta$ and $\Pi = (\Pi_{+\alpha} + \Pi_{-\alpha}) \otimes \Pi_\beta$ is the projector on the invariant subspace of M^2 corresponding to its Perron root ρ^2.*

Proof :
Since A is normal there exists a unitary matrix U which diagonalizes

both A and A^T :

$$A = U\Lambda U^*, \quad A^T = U\overline{\Lambda}U^*$$

and the columns $u_i, i = 1, \ldots, n_A$ of U are their common eigenvectors (notice that u_i is real only if λ_i is real as well). Therefore

$$(U^* \otimes I)M(U \otimes I) = (U^* \otimes I)(A \otimes B + A^T \otimes B^T)(U \otimes I) = \Lambda \otimes B + \overline{\Lambda} \otimes B^T$$

and the eigenvalues of M are those of the Hermitian matrices

$$H_i := \lambda_i B + \overline{\lambda}_i B^T,$$

which obviously are bounded by $|\lambda_i|\beta$ where β is the Perron root of $(B + B^T)$. Moreover, if $v_j^{(i)}, j = 1, \ldots, n_B$ are the eigenvectors of H_i then those of M are given by

$$u_i \otimes v_j^{(i)}, \quad i = 1, \ldots n_a, \quad j = 1, \ldots, n_B$$

and they can again only be real if λ_i is real. Since we want *real* eigenvectors corresponding to extremal eigenvalues of M we only need to consider the largest real eigenvalues of A, i.e. $\pm\alpha$ where α is the Perron root of A. Since A is also normal we have

$$A\Pi_\alpha = A^T\Pi_\alpha = \alpha\Pi_\alpha, \quad A\Pi_{-\alpha} = A^T\Pi_{-\alpha} = -\alpha\Pi_{-\alpha}.$$

It then follows that

$$(A \otimes B + A^T \otimes B^T)^2((\Pi_{+\alpha} + \Pi_{-\alpha}) \otimes \Pi_\beta) = \alpha^2(\Pi_{+\alpha} + \Pi_{-\alpha}) \otimes \beta^2\Pi_\beta,$$

which completes the proof. $\square$

It follows from this theorem that applying the projector Π to the vector $\mathbf{1}$ yields the vector

$$(\Pi_{+\alpha} + \Pi_{-\alpha})\mathbf{1} \otimes \Pi_\beta\mathbf{1},$$

which decouples the problem into two smaller ones.

Corollary 6.6. *Let G_A and G_B be two graphs and assume that A is a normal matrix. Then the similarity matrix between G_A and G_B is a rank one matrix $\mathbf{S} = uv^T$ where*

$$u = \frac{(\Pi_{+\alpha} + \Pi_{-\alpha})\mathbf{1}}{\|(\Pi_{+\alpha} + \Pi_{-\alpha})\mathbf{1}\|_2}, \quad v = \frac{\Pi_\beta\mathbf{1}}{\|\Pi_\beta\mathbf{1}\|_2}.$$

When one of the graphs G_A or G_B is vertex-transitive or has a normal adjacency matrix, the resulting similarity matrix $\mathbf{S}$ has rank one. Adjacency matrices of vertex-transitive graphs and normal matrices have the property that the projector $\Pi_{+\alpha}$ on the invariant subspace corresponding to the Perron root of A is also the projector on the subspace of A^T (and similarly for $-\alpha$). We conjecture here that the similarity matrix can only be of rank one if either A or B have this property.

7. Application of the central score to automatic extraction of synonyms

We illustrate in this last section the use of the central similarity score introduced in Section 4 for the automatic extraction of synonyms from a monolingual dictionary. Our method uses a graph constructed from the dictionary and is based on the assumption that synonyms have many words in common in their definitions and are used in the definition of many common words. We briefly outline our method below and then discuss the results obtained for four query words with the Webster dictionary. For a complete description of this application we refer the interested reader to [5] from which this section is extracted.

The method is fairly simple. Starting from a dictionary, we first construct the associated *dictionary graph* G; each word of the dictionary is a vertex of the graph and there is an edge from u to v if v appears in the definition of u. Then, associated to a given query word w, we construct a *neighborhood graph* G_w which is the subgraph of G whose vertices are those pointed by w or pointing to w. Finally, we compute the similarity score of the vertices of the graph G_w with the vertex 2 in the structure graph

$$1 \longrightarrow 2 \longrightarrow 3$$

and rank the words obtained by decreasing score. This score is the central score for which a simple computation method is presented in Section 4. Because of the way the neighborhood graph is constructed, we expect the word with highest central score to be a good candidate for synonymy.

Before proceeding to the description of the results obtained, we briefly describe the dictionary graph. We used the Online Plain Text English Dictionary [2] which is based on the "Project Gutenberg Etext of Webster's Unabridged Dictionary" which is in turn based on the 1913 US Webster's Unabridged Dictionary. The dictionary consists of 27 HTML files (one for each letter of the alphabet, and one for several additions). These files are freely available from the web site `http://www.gutenberg.net/`. The resulting graph has $112,169$ vertices and $1,398,424$ edges. It can be downloaded from the web-page `http://www.eleves.ens.fr/home/senellar/`. We have analyzed several features of the graph: connectivity and strong connectivity, number of connected components, distribution of connected components, degree distributions, graph diameter, etc. Some of our findings are reported in [17].

In order to be able to evaluate the quality of our synonym extraction method, we have compared the results produced with three other lists of synonyms. Two of these (Distance and ArcRank) were compiled automatically by two other synonym extraction methods (see [5] for details), and one of them lists synonyms obtained from the hand-made resource WordNet freely available on the WWW, [1]. The order of appearance of the words for this last source is arbitrary, whereas it is well defined for the three other methods. We have not kept the query word in the list of synonyms, since this has not much sense except for our method, where it is interesting to note that in every example we have experimented, the original word appears as the first word of the list; a point that tends to give credit to our method. We have examined the first ten results obtained on four query words chosen for their variety:

(1) **disappear**: a word with various synonyms such as **vanish**.
(2) **parallelogram**: a very specific word with no true synonyms but with some similar words: **quadrilateral**, **square**, **rectangle**, **rhomb**...
(3) **sugar**: a common word with different meanings (in chemistry, cooking, dietetics...). One can expect **glucose** as a candidate.
(4) **science**: a common and vague word. It is hard to say what to expect as synonym. Perhaps **knowledge** is the best option.

In order to have an objective evaluation of the different methods, we have asked a sample of 21 persons to give a mark (from 0 to 10) to the lists of synonyms, according to their relevance to synonymy. The lists were of course presented in random order for each word. The results obtained are given in the Tables 1, 2, 3 and 4. The last two lines of each of these tables gives the average mark and the standard deviation.

Concerning **disappear**, the distance method and our method do pretty well; **vanish**, **cease**, **fade**, **die**, **pass**, **dissipate**, **faint** are very relevant (one must not forget that verbs necessarily appear without their postposition). **dissipate** or **faint** are relevant too. However, some words like **light** or **port** are completely irrelevant, but they appear only in 6th, 7th or 8th position. If we compare these two methods, we observe that our method is better: an important synonym like **pass** takes a good ranking, whereas **port** or **appear** go out of the top ten words. It is hard to explain this phenomenon, but we can say that

	Distance	Our method	ArcRank	Wordnet
1	**vanish**	**vanish**	**epidemic**	**vanish**
2	**wear**	**pass**	**disappearing**	**go away**
3	**die**	**die**	**port**	**end**
4	**sail**	**wear**	**dissipate**	**finish**
5	**faint**	**faint**	**cease**	**terminate**
6	**light**	**fade**	**eat**	**cease**
7	**port**	**sail**	**gradually**	
8	**absorb**	**light**	**instrumental**	
9	**appear**	**dissipate**	**darkness**	
10	**cease**	**cease**	**efface**	
Mark	3.6	6.3	1.2	7.5
Std dev.	1.8	1.7	1.2	1.4

TABLE 1. Proposed synonyms for **disappear**

	Distance	Our method	ArcRank	Wordnet
1	**square**	**square**	**quadrilateral**	**quadrilateral**
2	**parallel**	**rhomb**	**gnomon**	**quadrangle**
3	**rhomb**	**parallel**	**right-lined**	**tetragon**
4	**prism**	**figure**	**rectangle**	
5	**figure**	**prism**	**consequently**	
6	**equal**	**equal**	**parallelepiped**	
7	**quadrilateral**	**opposite**	**parallel**	
8	**opposite**	**angles**	**cylinder**	
9	**altitude**	**quadrilateral**	**popular**	
10	**parallelepiped**	**rectangle**	**prism**	
Mark	4.6	4.8	3.3	6.3
Std dev.	2.7	2.5	2.2	2.5

TABLE 2. Proposed synonyms for **parallelogram**

the mutually reinforcing aspect of our method is apparently a positive point. On the contrary, ArcRank gives rather poor results with words such as **eat**, **instrumental** or **epidemic** that are out of the point.

Because the neighborhood graph of **parallelogram** is rather small (30 vertices), the first two algorithms give similar results, which are not absurd: **square**, **rhomb**, **quadrilateral**, **rectangle**, **figure** are rather interesting. Other words are less relevant but still are in the semantic domain of **parallelogram**. ArcRank which also works on the same subgraph does not give as interesting words, although **gnomon** makes its appearance, since **consequently** or **popular** are irrelevant. It is interesting to note that Wordnet is here less rich because it focuses on

	Distance	Our method	ArcRank	Wordnet
1	**juice**	**cane**	**granulation**	**sweetening**
2	**starch**	**starch**	**shrub**	**sweetener**
2	**cane**	**sucrose**	**sucrose**	**carbohydrate**
4	**milk**	**milk**	**preserve**	**saccharide**
5	**molasses**	**sweet**	**honeyed**	**organic compound**
6	**sucrose**	**dextrose**	**property**	**saccarify**
7	**wax**	**molasses**	**sorghum**	**sweeten**
8	**root**	**juice**	**grocer**	**dulcify**
9	**crystalline**	**glucose**	**acetate**	**edulcorate**
10	**confection**	**lactose**	**saccharine**	**dulcorate**
Mark	3.9	6.3	4.3	6.2
Std dev.	2.0	2.4	2.3	2.9

TABLE 3. Proposed synonyms for **sugar**

	Distance	Our method	ArcRank	Wordnet
1	**art**	**art**	**formulate**	**knowledge domain**
2	**branch**	**branch**	**arithmetic**	**knowledge base**
3	**nature**	**law**	**systematize**	**discipline**
4	**law**	**study**	**scientific**	**subject**
5	**knowledge**	**practice**	**knowledge**	**subject area**
6	**principle**	**natural**	**geometry**	**subject field**
7	**life**	**knowledge**	**philosophical**	**field**
8	**natural**	**learning**	**learning**	**field of study**
9	**electricity**	**theory**	**expertness**	**ability**
10	**biology**	**principle**	**mathematics**	**power**
Mark	3.6	4.4	3.2	7.1
Std dev.	2.0	2.5	2.9	2.6

TABLE 4. Proposed synonyms for **science**

a particular aspect (**quadrilateral**).

Once more, the results given by ArcRank for **sugar** are mainly irrelevant (**property**, **grocer**, ...). Our method is again better than the distance method: **starch**, **sucrose**, **sweet**, **dextrose**, **glucose**, **lactose** are highly relevant words, even if the first given near-synonym (**cane**) is not as good. Its given mark is even better than for Wordnet.

The results for **science** are perhaps the most difficult to analyze. The distance method and ours are comparable. ArcRank gives perhaps better results than for other words but is still poorer than the two other methods.

As a conclusion, the first two algorithms give interesting and relevant words, whereas it is clear that ArcRank is not adapted to the search for synonyms. The use of the central score and its mutually reinforcing relationship demonstrates its superiority on the basic distance method, even if the difference is not obvious for all words. The quality of the results obtained with these different methods is still quite different to that of hand-made dictionaries such as Wordnet. Still, these automatic techniques show their interest, since they present more complete aspects of a word than hand-made dictionaries. They can profitably be used to broaden a topic (see the example of **parallelogram**) and to help with the compilation of synonyms dictionaries.

8. Concluding remarks

In this paper, we introduce a new concept of similarity matrix and explain how to associate a *score* for the similarity of the vertices of two graphs. We show how this score can be computed and indicate how it extends the concept of hub and authority scores introduced by Kleinberg. We prove several properties and illustrate the strength and weakness of this new concept. Investigations of properties and applications of the similarity matrix of graphs can be pursued in several directions. We outline here two possible research directions.
One natural extension of our concept is to consider networks rather than graphs; this amounts to consider adjacency matrices with arbitrary real entries and not just integers. The definitions and results presented in this paper use only the property that the adjacency matrices involved have non-negative entries, and so all results remain valid for networks with non-negative weights. The extension to networks makes a sensitivity analysis possible: How sensitive is the similarity matrix to the weights in the network? Experiments and qualitative arguments show that, for most networks, similarity scores are almost everywhere continuous functions of the network entries. Perhaps this can be analyzed for models for random graphs such as those that appear in [6]? These questions can probably also be related to the large literature on eigenvalues and eigenspaces of graphs; see, e.g., [7], [8] and [9].

More specific questions on the similarity matrix also arise. One open problem is to characterize the pairs of matrices that give rise to a rank one similarity matrix. The structure of these pairs is conjectured at the end of Section 6. Is this conjecture correct? A long-standing graph question also arise when trying to characterize the graphs whose similarity matrices have only positive entries. The positive entries of the similarity matrix between the graphs G_A and G_B can be obtained as follows. One construct the product graph, symmetrize it, and then identify in the resulting graph the connected component(s) of largest possible Perron root. The indices of the vertices in that graph correspond exactly to the nonzero entries in the similarity matrix of G_A and G_B. The entries of the similarity matrix will thus be all positive if and only if the symmetrized product graph is connected; that is, if and only if, the product graph of G_A and G_B is weakly connected. The problem of characterizing all pairs of graphs that have a weakly connected product was introduced and analyzed in 1966 in [10]. That reference provides sufficient conditions for the product to be weakly connected. Despite several subsequent contributions on this question (see, e.g. [11]), the problem of efficiently characterizing all pairs of graphs that have a weakly connected product is a problem that, to our knowledge, is still open.

Another topic of interest is to investigate how the concepts proposed here can be used, possibly in modified form, for evaluating the similarity between two graphs, for clustering vertices or graphs, for pattern recognition in graphs and for data mining purposes.

Acknowledgment

Three of our students, Maureen Heymans, Anahí Gajardo and Pierre Senellart have provided inputs on several ideas developed in this paper. Their respective contributions are as follows.

A first definition of similarity score appears in Maureen Heymans Master's thesis performed at the University of Louvain in 2000-2001 under the supervision of the authors [12]. The definition appearing in the thesis differs from the one we use in this paper but the essential ideas are the same. In the spirit of Kleinberg, Maureen Heymans Master's thesis also describes applications of the similarity score to information retrieval in the graph of the internet.
Anahí Gajardo has worked with us during a postdoctoral visit at the University of Louvain in 2001. Her input has been helpful to derive the definition of the similarity score given here. Also, the bow-tie example appearing in Section 4 to ilustrate the interest of the central score is due to her.
Finally, the application of the central score to the automatic search of synonyms was developed by Pierre Senellart during his Master Thesis in 2002 at the University of Louvain under the supervision of VB. A

complete description of that application appears in [5] and a survey on the automatic discovery of synonyms is given in [18].
We are pleased to acknowledge the inputs of all these students.

This paper presents research supported by National Science Foundation under Grant No. CCR 99-12415, by the University of Louvain under a grant FSR2000, and by the Belgian Programme on Inter-university Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with its authors.

References

1. Wordnet 1.6, `http://www.cogsci.princeton.edu/~wn/`, 1998.
2. The Online Plain Text English Dictionary, `http://msowww.anu.edu.au/~ralph/OPTED/`, 2000.
3. A. Berman and R. J. Plemmons, *Non-negative matrices in the mathematical sciences*, Classics in Applied Mathematics 9, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.
4. Norman Biggs, *Algebraic Graph Theory (Second Edition)*, Cambridge Mathematical Library, Cambridge University Press, 1993.
5. Vincent D. Blondel, Pierre P. Senellart, Automatic extraction of synonyms in a dictionary, Technical report 2001-89, Universit catholique de Louvain, Louvain-la-Neuve, Belgium. Also: Proceedings of the SIAM Text Mining Workshop, Arlington (Virginia, USA) April 11, 2002.
6. B. Bollobas, Random Graphs, Academic Press, 1985.
7. Fan R. K. Chung, Spectral Graph Theory, American Mathematical Society, 1997.
8. Dragoš Cvetković, Peter Rowlinson, Slobodan Simić, Eigenspaces of graphs, Cambridge University Press, 1997.
9. Dragoš Cvetković, M. Doob, H. Sachs, Spectra of graphs – Theory and applications (Third edition), Johann Ambrosius Barth Verlag, 1995.
10. Frank Harary, C. Trauth, Connectedness of products of two directed graphs, J. SIAM Appl. Math., 14, pp. 150-154, 1966.
11. Hartfiel, D. J. and Maxson, C. J., The chainable matrix, a special combinatorial matrix, Discrete Math., Volume 12, pp. 245–256, 1975.
12. Maureen Heymans, Extraction d'information dans les graphes, et application aux moteurs de recherche sur internet. Mémoire de fin d'études. Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2001.
13. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, London, 1985.
14. R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, London, 1991.
15. Jan Jannink and Gio Wiederhold, Thesaurus Entry Extraction from an On-line Dictionary, Proceedings of Fusion 1999, Sunnyvale CA, 1999.
16. Jon M. Kleinberg, Authoritative sources in a hyperlinked environment, Journal of the ACM, 46:5, pp. 604–632, 1999.

17. Pierre Senellart. Extraction of information in large graphs. Automatic search for synonyms. Masters Internship Report. Technical report 2001-90, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2001.
18. Pierre P. Senellart, Vincent D. Blondel, Automatic discovery of similar words. To appear in "A Comprehensive Survey of Text Mining", Springer-Verlag, 2003.
19. G. Strang, The discrete cosine, SIAM Review, Volume 41, Number 1, pp. 135-147, 1999.

Division of Applied Mathematics, Université catholique de Louvain, 4 avenue Georges Lemaitre, B-1348 Louvain-la-Neuve, Belgium
E-mail address: `blondel@inma.ucl.ac.be`

Division of Applied Mathematics, Université catholique de Louvain, 4 avenue Georges Lemaitre, B-1348 Louvain-la-Neuve, Belgium
E-mail address: `vdooren@csam.ucl.ac.be`