A Measure of Similarity between Graph Vertices: Applications to Synonym Extraction and Web Searching*

Vincent D. Blondel[†]
Anahí Gajardo[‡]
Maureen Heymans[§]
Pierre Senellart[¶]
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 an $n_B \times n_A$ *similarity matrix* $\mathbf{S}$ whose real entry s_{ij} expresses how similar vertex j (in G_A) is to vertex i (in G_B): we say that s_{ij} is their *similarity score*. The similarity matrix can be obtained as the limit of the normalized even iterates of $S_{k+1} = BS_kA^T + B^TS_kA$, where A and B are adjacency matrices of the graphs and S_0 is a matrix whose entries are all equal to 1. In the special case where $G_A = G_B = G$, the matrix $\mathbf{S}$ is square and the score s_{ij} is the similarity score between the vertices i and j of 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 eigenvector of a nonnegative matrix. Potential applications of our similarity concept are numerous. We illustrate an application for the automatic extraction of synonyms in a monolingual dictionary.

Key words. algorithms, graph algorithms, graph theory, eigenvalues of graphs

AMS subject classifications. 05C50, 05C85, 15A18, 68R10

DOI. 10.1137/S0036144502415960

1. Generalizing Hubs and Authorities. Efficient web search engines such as Google are often based on the idea of characterizing the most important vertices

[†]Division of Applied Mathematics, Université catholique de Louvain, 4 Ave. G. Lemaitre, B-1348 Louvain-la-Neuve, Belgium (blondel@inma.ucl.ac.be, http://www.inma.ucl.ac.be/~blondel; vdooren@csam.ucl.ac.be, http://www.inma.ucl.ac.be/~vdooren).

[‡]Departamento de Ingeniera Matemática, Universidad de Concepción, Casilla 160-C, Concepción, Chile (anahi@ing-mat.udec.cl).

[§]Google Inc., 2400 Bayshore Parkway, Mountain View, CA 94043 (maureen@google.com).

[¶]Computer Science Department, École Normale Supérieure, 45 rue d'Ulm, F-75230 Paris cedex 05, France (Pierre.Senellart@ens.fr).

in a graph representing the connections or links between pages on the web. One such method, proposed by Kleinberg [18], identifies in a set of pages relevant to a query search the subset of pages that are good *hubs* or the subset of pages that are good *authorities*. For example, for the query "university," the home-pages of Oxford, Harvard, and other universities are good authorities, whereas web-pages that point to these home-pages are good hubs. Good hubs are pages that point to good authorities, and good authorities are pages 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 $G = (V, E)$ be a graph with vertex set V and with edge set E and let h_j and a_j be the hub and authority scores of 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:

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

Let B be the matrix whose entry (i, j) is equal to the number of edges between the vertices i and j in G (the *adjacency matrix* of G), and let h and a be the vectors of hub and authority scores. The above updating equations then take the simple form

$$\left[\begin{array}{c} h \\ a \end{array}\right]_{k+1} = \left[\begin{array}{cc} 0 & B \\ B^T & 0 \end{array}\right]\left[\begin{array}{c} h \\ a \end{array}\right]_k, \qquad k = 0, 1, \dots,$$

which we denote in compact form by

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

where

$$x_k = \left[\begin{array}{c} h \\ a \end{array}\right]_k, \quad M = \left[\begin{array}{cc} 0 & B \\ B^T & 0 \end{array}\right].$$

Notice that the matrix M is symmetric and nonnegative.[1] We are interested only in the relative scores and we will therefore consider the *normalized* vector sequence

(1.1) $$z_0 = x_0 > 0, \quad z_{k+1} = \frac{Mz_k}{\|Mz_k\|_2}, \quad k = 0, 1, \dots,$$

where $\|\cdot\|_2$ is the Euclidean vector norm. 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.

A first difficulty is that the sequence z_k does not always converge. In fact, sequences associated with nonnegative matrices M with the above block structure almost never converge, but rather oscillate between the limits

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

---

[1] A matrix or a vector Z will be said to be nonnegative (positive) if all its components are nonnegative (positive); we write $Z \geq 0$ ($Z > 0$) to denote this.

We prove in Theorem 2 that this is true in general for symmetric nonnegative matrices, and that either the sequence resulting from (1.1) converges or it doesn't, and then the even and odd subsequences do converge. 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 apparently natural choice for z_0. The set of all limit vectors obtained when starting from a positive initial vector is given by

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

and we would like to select one particular vector in that set. The vector z_{even} obtained for $z_0 = \mathbf{1}$ (we denote by $\mathbf{1}$ the vector, or matrix, whose entries are all equal to 1) has several interesting features that qualify it as a good choice: it is particularly easy to compute, it possesses several nice properties (see in particular section 4), and it has the extremal property, proved in Theorem 2, of being the unique vector in Z of the largest possible 1-norm (the 1-norm of a vector is the sum of all the magnitudes of its entries). Because of these features, we take the two subvectors of $z_{even}(\mathbf{1})$ as definitions for the hub and authority scores. Notice that the above matrix M has the property that

$$M^2 = \left[\begin{array}{cc} BB^T & 0 \\ 0 & B^T B \end{array} \right],$$

and from this equality it follows that, if the dominant invariant subspaces associated with BB^T and $B^T B$ have dimension 1, then the normalized hub and authority scores are simply given by the normalized dominant eigenvectors of BB^T and $B^T B$. This is the definition used in [18] for the authority and hub scores of the vertices of G. The arbitrary choice of $z_0 = \mathbf{1}$ made in [18] is shown here to have an extremal norm justification. Notice that when the invariant subspace has dimension 1, 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 vertex j of G can be thought of as a similarity score between vertex j of G and vertex *authority* of the graph

$$hub \longrightarrow authority$$

and, similarly, the hub score of vertex j of G can be seen as a similarity score between vertex j and 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 easier to grasp with an example; we illustrate it first on the path graph with three vertices and then provide a definition for arbitrary graphs. Let G be a graph with edge set E and adjacency matrix B and consider the *structure graph*

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

With each vertex j of 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 some positive value and then update them according to the following mutually reinforcing relation:

$$\left\{ \begin{array}{lll} x_{i1} & \leftarrow & \sum_{j:(i,j)\in E} x_{i2}, \\ x_{i2} & \leftarrow \sum_{j:(j,i)\in E} x_{i1} & + \sum_{j:(i,j)\in E} x_{i3}, \\ x_{i3} & \leftarrow & \sum_{j:(j,i)\in E} x_{i2}, \end{array} \right.$$
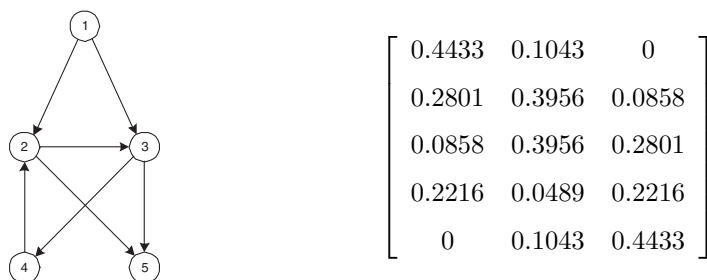
$$\begin{bmatrix} 0.4433 & 0.1043 & 0 \\ 0.2801 & 0.3956 & 0.0858 \\ 0.0858 & 0.3956 & 0.2801 \\ 0.2216 & 0.0489 & 0.2216 \\ 0 & 0.1043 & 0.4433 \end{bmatrix}$$

Fig. 1.1   *A graph and its similarity matrix with the structure graph* $1 \longrightarrow 2 \longrightarrow 3$. *The similarity score of vertex* 3 *with vertex* 2 *of the structure graph is equal to* 0.3956.

or, in matrix form (we denote by $\mathbf{x}_j$ the column vector with entries x_{ij}),

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix}_{k+1} = \begin{bmatrix} 0 & B & 0 \\ B^T & 0 & B \\ 0 & B^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix}_k, \qquad k = 0, 1, \ldots,$$

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 is symmetric and nonnegative, the normalized even and odd iterates converge, and the limit $z_{even}(\mathbf{1})$ is, among all possible limits, the unique vector with largest possible 1-norm. We take the three components of this extremal limit $z_{even}(\mathbf{1})$ as the definition of the similarity scores s_1, s_2, and s_3 and define the similarity matrix by $\mathbf{S} = [s_1 \ s_2 \ s_3]$. A numerical example of such a similarity matrix is shown in Figure 1.1. Note that we shall prove in Theorem 6 that the score s_2 can be obtained more directly from B by computing the dominating eigenvector of the matrix $BB^T + B^TB$.

We now come to a description of the general case. Assume that we have two directed graphs G_A and G_B with n_A and n_B vertices and edge sets E_A and E_B. 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 above examples. 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:

$$x_{ij} \leftarrow \sum_{r:(r,i)\in E_B, \ s:(s,j)\in E_A} x_{rs} + \sum_{r:(i,r)\in E_B, \ s:(j,s)\in E_A} x_{rs}.$$

This equation can be given an interpretation in terms of the product graph of G_A and G_B. The *product graph* of G_A and G_B is a graph that has $n_A.n_B$ vertices and that has an edge between vertices (i_1, j_1) and (i_2, j_2) if there is an edge between i_1 and i_2 in G_A and there is an edge between j_1 and j_2 in G_B. The above updating equation is then equivalent to replacing the scores of all vertices of the product graph by the sum of the scores of the vertices linked by an outgoing or incoming edge.

Equation (1.1) can also be written in more compact matrix form. Let X_k be the $n_B \times n_A$ matrix of entries x_{ij} at iteration k. Then the updating equations take the simple form

$$(1.2) \qquad\qquad X_{k+1} = BX_kA^T + B^TX_kA, \qquad k = 0, 1, \ldots,$$

where A and B are the adjacency matrices of G_A and G_B. We prove in section 3 that, as for the above examples, the normalized even and odd iterates of this updating
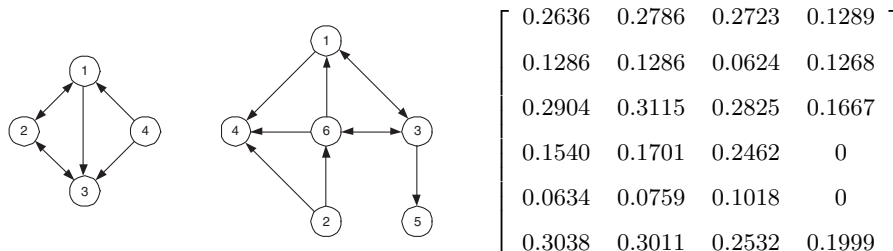
$$\begin{bmatrix} 0.2636 & 0.2786 & 0.2723 & 0.1289 \\ 0.1286 & 0.1286 & 0.0624 & 0.1268 \\ 0.2904 & 0.3115 & 0.2825 & 0.1667 \\ 0.1540 & 0.1701 & 0.2462 & 0 \\ 0.0634 & 0.0759 & 0.1018 & 0 \\ 0.3038 & 0.3011 & 0.2532 & 0.1999 \end{bmatrix}$$

Fig. 1.2 *Two graphs G_A, G_B and their similarity matrix. The vertex of G_A which is most similar to vertex 5 in G_B is vertex 3.*

equation converge and that the limit $Z_{even}(\mathbf{1})$ is among all possible limits the only one with largest 1-norm. We take this limit as the definition of the similarity matrix. An example of two graphs and their similarity matrix is shown in Figure 1.2.

It is interesting to note that in the database literature similar ideas have been proposed [20], [17]. The application there is information retrieval in large databases with which a particular graph structure can be associated. The ideas presented in these conference papers are obviously linked to those of the present paper, but in order to guarantee convergence of the proposed iteration to a unique fixed point, the iteration has to be slightly modified using, e.g., certain weighting coefficients. Therefore, the hub and authority score of Kleinberg is not a special case of their definition. We thank the authors of [17] for drawing our attention to those references.

The rest of this paper is organized as follows. In section 2, we describe some standard Perron–Frobenius results for nonnegative matrices that are 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 the algorithm. In section 4 we describe similarity matrices for the situation where one of the two graphs is a path graph of length 2 or 3. 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 a single graph G. Section 6 deals with graphs whose similarity matrix have rank 1. We prove there that if one of the graphs is regular or if one of the graphs is undirected, then the similarity matrix has rank 1. Regular graphs are graphs whose vertices have the same in-degrees and the same out-degrees; cycle graphs, for example, are regular. In a final section we report results obtained for the automatic synonym extraction in a dictionary by using the central score of a graph. A short version of this paper appears as a conference contribution [6].

2. Graphs and Nonnegative Matrices. With any directed graph $G = (V, E)$ one can associate a nonnegative 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 b_{ij} equals the number of edges from vertex i to vertex 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*.

In what follows, we shall 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 unique vector in $\mathcal{V}$ with smallest Euclidean distance to v. A matrix

representation of this projection can be obtained as follows. Let $\{v_1, \ldots, v_m\}$ be an orthonormal basis for $\mathcal{V}$ and arrange the column vectors v_i 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 $\mathcal{V}$. Projectors have the property that $\Pi^2 = \Pi$.

The Perron–Frobenius theory [14] establishes interesting properties about the eigenvectors and eigenvalues of nonnegative matrices. Let the largest magnitude of the eigenvalues of the matrix M (the *spectral radius* of M) be denoted by $\rho(M)$. According to the Perron–Frobenius theorem, the spectral radius of a nonnegative matrix M is an eigenvalue of M (called the *Perron root*), and there exists an associated nonnegative vector $x \geq 0$ ($x \neq 0$) such that $Mx = \rho x$ (called the *Perron vector*). In the case of symmetric matrices, more specific results can be obtained.

THEOREM 1. *Let M be a symmetric nonnegative matrix of spectral radius ρ. Then the algebraic and geometric multiplicity of the Perron root ρ are equal; there is a nonnegative matrix $X \geq 0$ whose columns span the invariant subspace associated with the Perron root; and the elements of the orthogonal projector Π on the vector space associated with the Perron root of M are all nonnegative.*

Proof. We use the facts that any symmetric nonnegative matrix M can be permuted to a block-diagonal matrix with irreducible blocks M_i on the diagonal [14, 3] and that the algebraic multiplicity of the Perron root of an irreducible nonnegative matrix is equal to 1. From these combined facts it follows that the algebraic and geometric multiplicities of the Perron root ρ of M are equal. Moreover, the corresponding invariant subspace of M is obtained from the normalized Perron vectors of the M_i blocks, appropriately padded with zeros. The basis X one obtains that way is then nonnegative and orthonormal. ☐

The next theorem will be used to justify our definition of similarity matrix between two graphs. The result describes the limit vectors of sequences associated with symmetric nonnegative linear transformations.

THEOREM 2. *Let M be a symmetric nonnegative matrix of spectral radius ρ. Let $z_0 > 0$ and consider the sequence*

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

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 z_k simply converges to $\Pi z_0 / \|\Pi z_0\|_2$, where Π is the orthogonal projector on the invariant subspace associated with the Perron root ρ. When $-\rho$ is an eigenvalue of M, then the subsequences z_{2k} and z_{2k+1} converge to the limits

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

where Π is the orthogonal projector on the sums of the invariant subspaces associated with ρ and $-\rho$. In both cases 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. We prove only the case where $-\rho$ is an eigenvalue; the other case is a trivial modification. 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 nontrivial and that we have orthonormal bases for them:

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

where M_μ is a square matrix (diagonal if V_μ is the basis of eigenvectors) with spectral radius μ strictly less than ρ. The eigenvalue decomposition can then be rewritten in block-diagonal form:

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

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 $M z_0$ that

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

and

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

provided the initial vectors z_0 and $M z_0$ have a nonzero component in the relevant subspaces, i.e., provided Πz_0 and $\Pi M z_0$ are nonzero. But the Euclidean norms 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 nonzero since $z_0 > 0$ and both Π and $M \Pi M$ are nonnegative and nonzero.

It follows from the nonnegativity of M and the formula for $z_{even}(z_0)$ and $z_{odd}(z_0)$ that both limits lie in $\{\Pi z / \|\Pi z\|_2 : z > 0\}$. Let us now show that every element $\hat{z}_0 \in \{\Pi z / \|\Pi z\|_2 : z > 0\}$ can be obtained as $z_{even}(z_0)$ for some $z_0 > 0$. Since the entries of Π are nonnegative, 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 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

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

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

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

with equality only when $\Pi z_0 = \lambda \Pi \mathbf{1}$ for some $\lambda \in \mathbb{C}$. But since Πz_0 and $\Pi \mathbf{1}$ are both real nonnegative, the proof easily follows. $\qquad \square$

3. Similarity between Vertices in Graphs. We now come to a formal definition of the similarity matrix of two directed graphs G_A and G_B. The updating equation for the similarity matrix is motivated in the introduction and is given by the linear mapping

$$(3.1) \qquad X_{k+1} = BX_kA^T + B^TX_kA, \qquad k = 0, 1, \ldots,$$

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 $\text{vec}(CXD) = (D^T \otimes C)\,\text{vec}(X)$ in which $\otimes$ denotes the Kronecker product (also denoted the tensorial or direct product). For a proof of this property, see Lemma 4.3.1 in [15]. Applying this property to (3.1) we immediately obtain

$$(3.2) \qquad x_{k+1} = (A \otimes B + A^T \otimes B^T)\,x_k,$$

where $x_k = \text{vec}(X_k)$. This is the format used in the introduction. Combining this observation with Theorem 2, we deduce the following theorem.

THEOREM 3. *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*

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

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 the largest 1-norm.

In order to be consistent with the vector norm appearing in Theorem 2, the matrix norm $\|.\|_F$ 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 the magnitudes of all entries. One can also provide a definition of the set Z in terms of one of its extremal properties.

THEOREM 4. *Let G_A and G_B be two graphs with adjacency matrices A and B and consider the notation of Theorem 3. The set*

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

and the set of all positive matrices that maximize the expression

$$\frac{\|BXA^T + B^TXA\|_F}{\|X\|_F}$$

are equal. Moreover, among all matrices in this set, there is a unique matrix $\mathbf{S}$ whose 1-norm is maximal. This matrix can be obtained as

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

for $Z_0 = \mathbf{1}$.

Proof. The above expression can also be written as $\|L(X)\|_2/\|X\|_2$, which is the induced 2-norm of the linear mapping L defined by $L(X) = BXA^T + B^TXA$. It is well known [14] that each dominant eigenvector X of L^2 is a maximizer of this expression. It was shown above that **S** is the unique matrix of largest 1-norm in that set. $\quad\square$

We take the matrix **S** appearing in this theorem as the definition of the similarity matrix between G_A and G_B. 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.

A direct algorithmic transcription of the definition leads to an approximation algorithm for computing similarity matrices of graphs:

1. Set $Z_0 = \mathbf{1}$.
2. Iterate an even number of times

$$Z_{k+1} = \frac{BZ_kA^T + B^TZ_kA}{\|BZ_kA^T + B^TZ_kA\|_F}$$

   and stop upon convergence.
3. Output **S** is the last value of Z_k.

This algorithm is a matrix analogue to the classical power method (see [14]) to compute a dominant eigenvector of a matrix. 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, respectively. Then the products BZ_k and B^TZ_k require less than $2n_A.e_B$ additions and multiplications each, while the subsequent products $(BZ_k)A^T$ and $(B^TZ_k)A$ 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 one 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 nonzero elements per row of A and B, respectively; then the total complexity per iteration step is of the order of $4(\alpha_A + \alpha_B)n_An_B$ additions and multiplications. As was shown in Theorem 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 **S** to ϵ accuracy with the power method is therefore of the order of

$$8n_An_B\frac{(\alpha_A + \alpha_B)\log\epsilon}{(\log\mu - \log\rho)}.$$

Other sparse matrix methods could be used here, but we do not address such algorithmic aspects in this paper. For particular classes of adjacency matrices, one can compute the similarity matrix **S** directly from the dominant invariant subspaces of matrices of the size of A or B. We provide explicit expressions for a few such classes in the next section.

4. Hubs, Authorities, and Central Scores. As explained in the introduction, the hub and authority scores of a graph can be expressed in terms of its adjacency matrix.

THEOREM 5. *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 BB^T and B^TB, provided the corresponding Perron root*

is of multiplicity 1. *Otherwise, it is the normalized projection of the vector* **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 $B^T B$ may have multiple dominant roots: for cycle graphs, for example, both BB^T and $B^T B$ are the identity matrix.

Another interesting structure graph is the path graph of length 3:

$$1 \longrightarrow 2 \longrightarrow 3$$

As for the hub and authority scores, we can give an explicit expression for the similarity score with vertex 2, a score that we will call the *central score*. This central score has been successfully used for the purpose of automatic extraction of synonyms in a dictionary. This application is described in more detail in section 7.

THEOREM 6. *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 $B^T B + BB^T$, 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. The corresponding matrix M is as follows:

$$M = \left[\begin{array}{ccc} 0 & B & 0 \\ B^T & 0 & B \\ 0 & B^T & 0 \end{array} \right]$$

and so

$$M^2 = \left[\begin{array}{ccc} BB^T & 0 & BB \\ 0 & B^T B + BB^T & 0 \\ B^T B^T & 0 & B^T B \end{array} \right],$$

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

$$M = P^T \left[\begin{array}{cc} 0 & E \\ E^T & 0 \end{array} \right] P, \quad \text{where } E := \left[\begin{array}{c} B \\ B^T \end{array} \right].$$

Now let V and U be orthonormal bases for the dominant right and left singular subspaces of E [14]:

(4.1) $$EV = \rho U, \quad E^T U = \rho V;$$

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

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

Moreover,

$$PM^2 P^T = \left[\begin{array}{cc} EE^T & 0 \\ 0 & E^T E \end{array} \right],$$

and the projectors associated with the dominant eigenvalues of EE^T and $E^T E$ are, respectively, $\Pi_v := VV^T$ and $\Pi_u := UU^T$. The projector Π of M^2 is then nothing
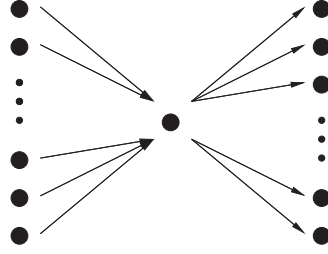
Fig. 4.1   *A directed bow-tie graph. Kleinberg's hub score of the center vertex is equal to $1/\sqrt{m+1}$ if $m > n$ and to 0 if $m < n$. The central score of this vertex is equal to $1/\sqrt{m+n+1}$ independently of the relative values of m and n.*

but $P^T \text{diag}\{\Pi_v, \Pi_u\}P$, and hence the subvectors 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 (4.1) also yields a relation between the two smaller projectors:

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

In order to illustrate that path graphs of length 3 may have an advantage over the *hub–authority* structure graph we consider here the special case of the "directed bow-tie graph" G_B represented in Figure 4.1. 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

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

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

$$
B^T B + BB^T = \left[
\begin{array}{ccc}
m+n & 0 & 0 \\
0 & \mathbf{1}_n & 0 \\
0 & 0 & \mathbf{1}_m
\end{array}
\right],
$$

and, following Theorem 6, the Perron root of M is equal to $\rho = \sqrt{n+m}$ and the similarity matrix is given by the $(1+m+n) \times 3$ matrix

$$
\mathbf{S} = \frac{1}{\sqrt{m+n+1}}
\left[
\begin{array}{c|c|c}
0 & 1 & 0 \\
\hline
1 & 0 & 0 \\
\vdots & \vdots & \vdots \\
1 & 0 & 0 \\
\hline
0 & 0 & 1 \\
\vdots & \vdots & \vdots \\
0 & 0 & 1
\end{array}
\right].
$$

This result holds irrespective of the relative value of m and n. Let us call the three vertices of the path graph, 1, center, and 3, respectively. One could view a *center* as a vertex through which much information is passed on. This similarity matrix $\mathbf{S}$ indicates that vertex 1 of G_B looks 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 of Kleinberg, then the similarity scores $\mathbf{S}$ differ for $m < n$ and $m > n$:

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

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 $m < n$.

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 5.1. 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 appears on the diagonal and that, except for trivial cases, the diagonal elements of a self-similarity matrix are nonzero. As can be seen from elementary examples, it is, however, not true that diagonal elements always dominate all elements on the same row and column.

THEOREM 7. *The self-similarity matrix of a graph is positive semidefinite. In particular, the largest element of the matrix appears on the diagonal, and if a diagonal entry is equal to zero the corresponding row and column are equal to zero.*

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

$$Z_{k+1} = \frac{A Z_k A^T + A^T Z_k A}{\|A Z_k A^T + A^T Z_k A\|_F}, \quad Z_0 = \mathbf{1}.$$

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

When vertices of a graph are similar to each other, such as in cycle graphs, we expect to have a self-similarity matrix with all entries equal. This is indeed the case, as will be proved in the next section. We can also derive explicit expressions for the self-similarity matrices of path graphs.

THEOREM 8. *The self-similarity matrix of a path graph is a diagonal matrix.*

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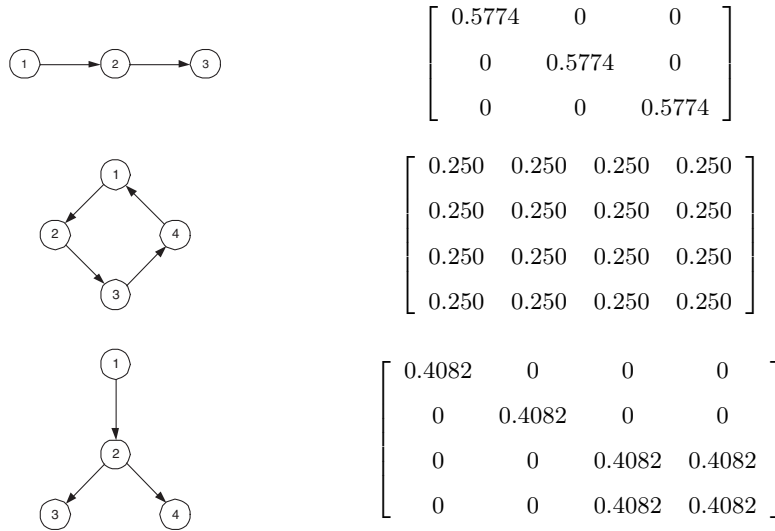
$$
\begin{bmatrix}
0.5774 & 0 & 0 \\
0 & 0.5774 & 0 \\
0 & 0 & 0.5774
\end{bmatrix}
$$

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

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

Fig. 5.1 *Graphs and their corresponding self-similarity matrix. The self-similarity matrix of a graph gives a measure of how similar vertices are to each other.*

arrangement of Jacobi matrices

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

of dimension $j = 1, \ldots, \ell$, where ℓ is the dimension of the given path graph. The largest of these blocks corresponds to the Perron root ρ of M. There is only one largest block and its vertices correspond to the diagonal elements of **S**. As shown in [19], $\rho = 2\cos(\pi/(\ell+1))$, but M has both eigenvalues $\pm\rho$ and the corresponding vectors have the elements $(\pm)^j \sin(j\pi/(\ell+1))$, $j = 1, \ldots, \ell$, from which $\Pi\mathbf{1}$ can easily be computed. □

6. Similarity Matrices of Rank 1. In this section we describe two classes of graphs that lead to similarity matrices that have rank 1. We consider the case when one of the two graphs is regular (a graph is regular if the in-degrees of its vertices are all equal and the out-degrees are also equal), and the case when the adjacency matrix of one of the graphs is normal (a matrix A is normal if it satisfies $AA^T = A^T A$). In both cases we prove that the similarity matrix has rank 1. Graphs that are not directed have a symmetric adjacency matrix, and symmetric matrices are normal, therefore graphs that are not directed always generate similarity matrices that have rank 1.

THEOREM 9. *Let G_A, G_B be two graphs of adjacency matrices A and B and assume that G_A is regular. Then the similarity matrix between G_A and G_B is a rank 1 matrix of the form*

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

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

Proof. It is known (see, e.g., [4]) that a regular graph G_A has an adjacency matrix A with Perron root of algebraic multiplicity 1 and that the vector $\mathbf{1}$ is the corresponding Perron vector of both A and A^T. It easily follows from this that each matrix Z_k of the iteration defining the similarity matrix is of rank 1 and of the type $v_k \mathbf{1}^T / \sqrt{n_A}$, where

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

This clearly converges to $\Pi \mathbf{1} / \|\Pi \mathbf{1}\|_2$, where Π is the projector on the dominant invariant subspace of $(B + B^T)^2$.   □

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 $AA^T = A^T A$.

THEOREM 10. *Let G_A and G_B be two graphs of adjacency matrices A and B and assume that one of the adjacency matrices is normal. Then the similarity matrix between G_A and G_B has rank 1.*

Proof. Let A be the normal matrix and let α be its Perron root. Then 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 normal we also have that its real eigenvectors are also eigenvectors of A^T. Therefore,

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

It then follows that

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

and hence $\Pi := (\Pi_{+\alpha} + \Pi_{-\alpha}) \otimes \Pi_\beta$ is the projector of the dominant root $\alpha^2 \beta^2$ of M^2. Applying this projector to the vector $\mathbf{1}$ yields the vector

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

which corresponds to the rank 1 matrix

$$\mathbf{S} = (\Pi_{+\alpha} + \Pi_{-\alpha})\mathbf{1}\Pi_{\beta}. \qquad \square$$

When one of the graphs G_A or G_B is regular or has a normal adjacency matrix, the resulting similarity matrix $\mathbf{S}$ has rank 1. Adjacency matrices of regular graphs and normal matrices have the property that the projector Π on the invariant subspace corresponding to the Perron root of A is also the projector on the subspace of A^T. As a consequence, $\rho(A + A^T) = 2\rho(A)$. In this context we formulate the following conjecture.

CONJECTURE 11. *The similarity matrix of two graphs has rank* 1 *if and only if one of the graphs has the property that its adjacency matrix D is such that $\rho(D+D^T) = 2\rho(D)$.*

7. Application 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 appear together in the definition of many words. We briefly outline our method below and then discuss the results obtained with the Webster dictionary on four query words. The application given in this section is based on [5], to which we refer the interested reader for a complete description.

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 with a given query word w, we construct a *neighborhood graph G_w*, which is the subgraph of G whose vertices are pointed to by w or are pointing to w (see, e.g., Figure 7.1). Finally, we compute the similarity score of the vertices of the graph G_w with the central vertex in the structure graph

$$1 \longrightarrow 2 \longrightarrow 3$$

and rank the words by decreasing score. Because of the way the neighborhood graph is constructed, we expect the words with highest central score to be good candidates 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 U.S. 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/.

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; the method ArcRank is described in [16]), and one of them lists synonyms obtained from the handmade resource WordNet freely available on the web [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 does not make much sense except
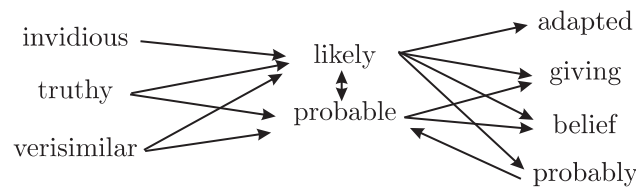
Fig. 7.1   *Part of the neighborhood graph associated with the word "likely." The graph contains all words used in the definition of "likely" and all words using "likely" in their definition. Synonyms are identified by selecting in this graph those vertices of highest central score.*

Table 7.1   *Proposed synonyms for **disappear**.*

	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

for our method, where it is interesting to note that in every example with which 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 a synonym. Perhaps **knowledge** is the best candidate.

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 Tables 7.1, 7.2, 7.3, and 7.4. The last two lines of each of these tables gives the average mark and its standard deviation.

Concerning **disappear**, the *distance method* and *our method* do pretty well; **vanish**, **cease**, **fade**, **die**, **pass**, **dissipate**, and **faint** are very relevant (one must not forget that verbs necessarily appear without their postposition). Some words like **light** or **port** are completely irrelevant, but they appear only in sixth, seventh, or eighth position. If we compare these two methods, we observe that our method is better: an important synonym like **pass** gets a good ranking, whereas **port** or **appear** are not in the top ten words. It is hard to explain this phenomenon, but we can say that the mutually reinforcing aspect of our method apparently has a positive

Table 7.2 *Proposed synonyms for **parallelogram**.*

	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 7.3 *Proposed synonyms for **sugar**.*

	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 7.4 *Proposed synonyms for **science**.*

	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

effect. In contrast to this, ArcRank gives rather poor results including words such as
eat, **instrumental**, or **epidemic** that are irrelevant.

Because the neighborhood graph of **parallelogram** is rather small (30 vertices),
the first two algorithms give similar results, which are reasonable: **square**, **rhomb**,
quadrilateral, **rectangle**, and **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 results of the same quality: **consequently**

and **popular** are clearly irrelevant, but **gnomon** is an interesting addition. It is interesting to note that WordNet is here less rich because it focuses on 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**, and **lactose** are highly relevant words, even if the first given near-synonym (**cane**) is not as good. Note that our method has marks that are even better than those of 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.

In 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 to 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 from that of handmade dictionaries such as WordNet. Still, these automatic techniques are interesting, since they present more complete aspects of a word than handmade dictionaries. They can profitably be used to broaden a topic (see the example of **parallelogram**) and to help with the compilation of synonym dictionaries.

8. Concluding Remarks. In this paper, we introduce a new concept, similarity matrices, and explain how to associate a *score* with 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 some possible research directions.

One natural extension of our concept is to consider networks rather than graphs; this amounts to considering 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 nonnegative entries, and so all results remain valid for networks with nonnegative 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 [7]? These questions can probably also be related to the large literature on eigenvalues and invariant subspaces of graphs; see, e.g., [8], [9], and [10].

It appears natural to investigate the possible use of the similarity matrix of two graphs to detect whether the graphs are isomorphic. (The membership of the graph isomorphism problem in the complexity classes P or NP-complete is so far unsettled.) If two graphs are isomorphic, then their similarity matrix can be made symmetric by column (or row) permutation. It is easy to check in polynomial time if such a permutation is possible and if it is unique. (When all entries of the similarity matrix are distinct, it can only be unique.) In the case where no such permutation exists or where only one permutation is possible, one can immediately conclude by answering negatively or by checking the proposed permutation. In the case where many permutations render the similarity matrix symmetric, all of them have to be checked, and this leads to a possibly exponential number of permutations to verify. It

would be interesting to see how this heuristic compares to other heuristics for graph isomorphism and to investigate whether other features of the similarity matrix can be used to limit the number of permutations to consider.

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 1 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 arises 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. First 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 [11]. That reference provides sufficient conditions for the product to be weakly connected. Despite several subsequent contributions on this question (see, e.g., [12]), the problem of efficiently characterizing all pairs of graphs that have a weakly connected product is, to our knowledge, 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.

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, *Nonnegative Matrices in the Mathematical Sciences*, Classics Appl. Math. 9, SIAM, Philadelphia, 1994.

[4] N. Biggs, *Algebraic Graph Theory*, 2nd ed., Cambridge Mathematical Library, Cambridge University Press, Cambridge, UK, 1993.

[5] V. D. BLONDEL AND P. P. SENELLART, *Automatic extraction of synonyms in a dictionary*, in Proceedings of the SIAM Text Mining Workshop, Arlington, VA, 2002.

[6] V. D. BLONDEL AND P. VAN DOOREN, *Similarity matrices for pairs of graphs*, in Proceedings of the ICALP 2003 Conference, J. C. M. Baeten et al., eds., Lecture Notes in Comput. Sci. 2719, Springer-Verlag, New York, 2003, pp. 739–550.

[7] B. BOLLOBAS, *Random Graphs*, Academic Press, New York, 1985.

[8] F. R. K. CHUNG, *Spectral Graph Theory*, AMS, Providence, RI, 1997.

[9] D. CVETKOVIĆ, P. ROWLINSON, AND S. SIMIĆ, *Eigenspaces of Graphs*, Cambridge University Press, Cambridge, UK, 1997.

[10] D. CVETKOVIĆ, M. DOOB, AND H. SACHS, *Spectra of Graphs—Theory and Applications*, 3rd ed., Johann Ambrosius Barth Verlag, Leipzig, 1995.

[11] F. HARARY AND C. A. TRAUTH, JR., *Connectedness of products of two directed graphs*, SIAM J. Appl. Math., 14 (1966), pp. 250–254.

[12] D. J. HARTFIEL AND C. J. MAXSON, *The chainable matrix, a special combinatorial matrix*, Discrete Math., 12 (1975), pp. 245–256.

[13] M. HEYMANS, *Extraction d'information dans les graphes, et application aux moteurs de recherche sur internet*, Engineer's thesis, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, 2001.

[14] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, London, 1985.

[15] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, London, 1991.

[16] J. JANNINK AND G. WIEDERHOLD, *Thesaurus entry extraction from an on-line dictionary*, in Proceedings of Fusion 1999, Sunnyvale, CA, 1999.

[17] G. JEH AND J. WIDOM, *SimRank: A Measure of Structural-Context Similarity*, in Proceedings of the KDD2002 Conference, Edmonton, AB, Canada, 2002.

[18] J. M. KLEINBERG, *Authoritative sources in a hyperlinked environment*, J. ACM, 46 (1999), pp. 604–632.

[19] M. MARCUS AND H. MINC, *A Survey of Matrix Theory and Matrix Inequalities*, Allyn and Bacon, Boston, MA, 1964.

[20] S. MELNIK, H. GARCIA-MOLINA, AND E. RAHM, *Similarity flooding: A versatile graph matching algorithm and its application to schema matching*, in Proceedings of the 18th ICDE, San Jose, CA, 2002.

[21] P. P. SENELLART AND V. D. BLONDEL, *Automatic discovery of similar words*, in A Comprehensive Survey of Text Mining, M. Berry, ed., Springer-Verlag, New York, 2003.