The power method, Perron-Frobenius theory, Page-Rank computation

Preliminary considerations, $A \in \mathbb{C}^{n\times n}$

Given a $n \times n$ matrix A with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, the following result holds.

Theorem (inverse power). If λ_i^* is an approximation of the eigenvalue λ_i of a $n\times n$ matrix A, i.e. $|\lambda_i - \lambda_i^*|$ is smaller than $|\lambda_j - \lambda_i^*|$, $\forall\, \lambda_j \neq \lambda_i$, if $m_a(\lambda_i) = m_g(\lambda_i)$, and if $\mathbf{v}_0 \in \mathbb{C}^n$ is not orthogonal to the space spanned by the eigenvectors of A corresponding to λ_i, then the sequence $\{\mathbf{v}_k\}$ generated by the algorithm

$$(A - \lambda_i^* I)\mathbf{a}_k = \mathbf{v}_{k-1},\ \ \mathbf{v}_k = \frac{\mathbf{a}_k}{\|\mathbf{a}_k\|}, \quad k = 1, 2, \dots$$

converges to an eigenvector of A corresponding to the eigenvalue λ_i. If A is diagonalizable, then the rate of convergence is $O((\max_{j:\,\lambda_j\neq\lambda_i} |\frac{\lambda_i-\lambda_i^*}{\lambda_j-\lambda_i^*}|)^k)$.

Remark. In the general case the rate of convergence is

$$\max_{j:\,\lambda_j\neq\lambda_i} \max_{s_{\lambda_j}} O(|p_{s_{\lambda_j}-1}(k)||\frac{\lambda_i - \lambda_i^*}{\lambda_j - \lambda_i^*}|^k),$$

where, given the block diagonal Jordan form of A, $J = X^{-1}AX$, for each $\lambda_j \neq \lambda_i$, the number s_{λ_j} indicates the dimension of the generic Jordan block associated with λ_j, and $p_{s_{\lambda_j}-1}(k)$ is a polynomial of degree $s_{\lambda_j} - 1$, whose coefficients depend on $\frac{1}{(\lambda_j-\lambda_i^*)^r}$, $r = 1, \dots, s_{\lambda_j} - 1$, and on the coefficients of $\mathbf{v}_0$ with respect to the column vectors of X corresponding to the Jordan block under consideration.

proof: See the Appendix. $\square$

Let λ_1 be such that $|\lambda_1| = \rho(A)$ and assume that all λ_i such that $|\lambda_i| = \rho(A)$ are equal to λ_1 (in such case we say that λ_1 dominates the eigenvalues of A). Assume, moreover, that the algebraic and geometric multiplicity of λ_1 are equal. Then the *power method* (see the Theorem below) can be used to compute λ_1 and an eigenvector corresponding to λ_1.

Theorem (power). If λ_1 dominates the eigenvalues of a $n \times n$ matrix A, if $m_a(\lambda_1) = m_g(\lambda_1)$, and if $\mathbf{v}_0 \in \mathbb{C}^n$ is not orthogonal to the space spanned by the eigenvectors of A corresponding to λ_1, then the sequence $\{\mathbf{v}_k\}$ generated by the algorithm

$$\mathbf{a}_k = A\mathbf{v}_{k-1},\ \ \mathbf{v}_k = \frac{\mathbf{a}_k}{\|\mathbf{a}_k\|}, \quad k = 1, 2, \dots$$

converges to an eigenvector of A corresponding to the eigenvalue λ_1. Moreover,

$$\frac{\mathbf{u}^H\mathbf{v}_{k+1}}{\mathbf{u}^H\mathbf{v}_k} \ \ \rightarrow \ \ \lambda_1, \quad k \rightarrow +\infty$$

for any $\mathbf{u}$ for which $\mathbf{u}^H\mathbf{v}_k \neq 0$. If A is diagonalizable, then the rate of convergence is $O((\max_{j:\,\lambda_j\neq\lambda_1} |\frac{\lambda_j}{\lambda_1}|)^k)$.

Remark. In the general case the rate of convergence is

$$\max_{j:\,\lambda_j\neq\lambda_1} \max_{s_{\lambda_j}} O(|p_{s_{\lambda_j}-1}(k)||\frac{\lambda_j}{\lambda_1}|^k),$$

where, given the block diagonal Jordan form of A, $J = X^{-1}AX$, for each $\lambda_j \neq \lambda_1$, the number s_{λ_j} indicates the dimension of the generic Jordan block associated with λ_j, and $p_{s_{\lambda_j}-1}(k)$ is a polynomial of degree $s_{\lambda_j}-1$, whose coefficients depend on $\frac{1}{\lambda_j^r}$, $r=1,\dots,s_{\lambda_j}-1$, and on the coefficients of $\mathbf{v}_0$ with respect to the column vectors of X corresponding to the Jordan block under consideration.

proof: See the Appendix. □

For our purposes it is useful to recall also the following classic result on matrix deflation: how to introduce a matrix whose eigenvalues are all equal to the eigenvalues of A except one, which, instead, is zero.

Theorem. Let A be a $n\times n$ matrix. Let λ_1 be a nonzero eigenvalue of A and $\mathbf{y}_1$ a corresponding eigenvector, i.e. $A\mathbf{y}_1=\lambda_1\mathbf{y}_1$. Call $\lambda_2,\lambda_3,\dots,\lambda_n$ the remaining eigenvalues of A. Then the matrix $W=A-\frac{\lambda_1}{\mathbf{w}^*\mathbf{y}_1}\mathbf{y}_1\mathbf{w}^*$, $\mathbf{w}^*\mathbf{y}_1\neq 0$, has eigenvalues $0,\lambda_2,\lambda_3,\dots,\lambda_n$.

proof. Introduce $S=[\mathbf{y}_1\,\mathbf{z}_2\cdots\mathbf{z}_n]$ non singular, and observe that $p_A(\lambda)=p_{S^{-1}AS}(\lambda)=(\lambda-\lambda_1)q(\lambda)$, $p_W(\lambda)=p_{S^{-1}WS}(\lambda)=\lambda q(\lambda)$. □

Let G be the following matrix

$$G=\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{11}{16} & \frac{1}{4} & \frac{1}{16} \end{bmatrix}.$$

Since $\rho(G)\le\|G\|_\infty=1$, we can say that the spectrum of G lies in the circle $\{z\in\mathbb{C}:\ |z|\le 1\}$.

Note that $G\mathbf{e}=1\cdot\mathbf{e}$, i.e. one eigenvalue of G, $\lambda_1=1$, and its corresponding eigenvector, $\mathbf{y}_1=\mathbf{e}=[1\ 1\ \cdots\ 1]^T$, are known. If λ_2,λ_3 denote the remaining eigenvalues of G, then we can define a matrix W, in terms of G, λ_1, $\mathbf{y}_1$, whose eigenvalues are $0,\lambda_2,\lambda_3$:

$$W=G-\frac{\lambda_1}{\mathbf{w}^*\mathbf{y}_1}\mathbf{y}_1\mathbf{w}^*=G-\frac{1}{\mathbf{w}^*\mathbf{e}}\begin{bmatrix}\mathbf{w}^*\\ \mathbf{w}^*\\ \mathbf{w}^*\end{bmatrix},\quad \forall\mathbf{w},\ \mathbf{w}^*\mathbf{e}\neq 0.$$

Since $(\mathbf{e}_i^T G)\mathbf{e}=\mathbf{e}_i^T(1\cdot\mathbf{e})=1\neq 0$, we choose $\mathbf{w}^*=\mathbf{e}_i^T G$:

$$W=G-\begin{bmatrix}\mathbf{e}_i^T G\\ \mathbf{e}_i^T G\\ \mathbf{e}_i^T G\end{bmatrix}.$$

For $i=1$ the matrix W becomes

$$W=\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{11}{16} & \frac{1}{4} & \frac{1}{16} \end{bmatrix}-\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}=\begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{8} & -\frac{3}{8} \\ \frac{7}{16} & 0 & -\frac{7}{16} \end{bmatrix},$$

so, the eigenvalues of A different from 1 are the eigenvalues of

$$\tilde{W}=\begin{bmatrix} -\frac{1}{8} & -\frac{3}{8} \\ 0 & -\frac{7}{16} \end{bmatrix},$$

i.e. $-\frac{1}{8}$ and $-\frac{7}{16}$.

Thus, 1, $-\frac{1}{8}$ and $-\frac{7}{16}$ are the eigenvalues of G, and also, of course, of G^T. In particular, 1 is eigenvalue of G^T, but note that the eigenvector $\mathbf{p}$ of G^T

corresponding to 1 is not obvious; it must be computed. For example, it can be computed as the limit of the inverse power sequence $\{\mathbf{v}_k\}$ defined as follows (ε small positive number):

$$\mathbf{v}_0 \in \mathbb{R}^3,\ (G^T - (1+\varepsilon)I)\mathbf{a}_k = \mathbf{v}_{k-1},\ \mathbf{v}_k = \frac{\mathbf{a}_k}{\|\mathbf{a}_k\|},\ k = 0, 1, 2, \ldots$$

(rate of convergence: $O(\left|\frac{1-(1+\varepsilon)}{-\frac{1}{8}-(1+\varepsilon)}\right|^k)$). [Here a reference for the inverse power iterations]. We shall see that the vector $\mathbf{p}$ can be also obtained as the limit of the sequence

$$\mathbf{p}_0 \in \mathbb{R}^3,\ \mathbf{p}_0 \text{ positive},\ \|\mathbf{p}_0\|_1 = 1,\ \mathbf{p}_{k+1} = G^T\mathbf{p}_k,\ k = 0, 1, 2, \ldots$$

(rate of convergence: $O(\left|\frac{-\frac{7}{16}}{1}\right|^k)$) [this result is in fact a particular case of the Theorem at the beginning of this section (set $\|\cdot\| = \|\cdot\|_1$, $\mathbf{u} = \mathbf{e}$)]. Even if the rate of convergence of the $\mathbf{p}_k$ is not as good as the rate of convergence of the $\mathbf{v}_k$, the computation of $\mathbf{p}_{k+1}$ from $\mathbf{p}_k$ is much cheaper than the computation of $\mathbf{v}_{k+1}$ from $\mathbf{v}_k$. In fact, for analogous problems, but of high dimension, even one step of the inverse power iterations is prohibitive.

The Perron-Frobenius theory: $A \in \mathbb{R}^{n\times n}$, $A \geq 0$ irreducible [Varga]

Lemma. Let A be a $n \times n$ non negative matrix, i.e. $a_{ij} \geq 0$, $\forall\, i, j$. Assume that A is not reducible. Then $(I + A)^{n-1}$ is a positive matrix, i.e. its entries are all positive.

proof. We shall prove that the vector $(I + A)^{n-1}\mathbf{x}$ is positive whenever $\mathbf{x}$ is a non negative non null vector (prove that this is equivalent to the thesis!).

Let $\mathbf{x}$ be a non negative non null vector. Set $\mathbf{x}_0 = \mathbf{x}$, $\mathbf{x}_1 = (I + A)\mathbf{x}_0 = \mathbf{x}_0 + A\mathbf{x}_0$, ..., $\mathbf{x}_{k+1} = (I + A)\mathbf{x}_k = \mathbf{x}_k + A\mathbf{x}_k$, $k = 1, \ldots, n-2$. Note that $\mathbf{x}_k = (I+A)^k\mathbf{x}$, in particular $\mathbf{x}_{n-1} = (I+A)^{n-1}\mathbf{x}$. So, our aim is to prove that $\mathbf{x}_{n-1}$ is a positive vector. First observe by induction that all $\mathbf{x}_k$ are non negative vectors ($\mathbf{x}_k$ non negative and A non negative imply $A\mathbf{x}_k$ non negative and $\mathbf{x}_{k+1} = \mathbf{x}_k + A\mathbf{x}_k$ non negative). Then the thesis ($\mathbf{x}_{n-1}$ positive) is now proved by showing that $\mathbf{x}_{k+1}$ must have less zeros than $\mathbf{x}_k$ for each $k \in \{0, \ldots, n-2\}$. Note that $\mathbf{x}_{k+1}$ cannot have more zeros than $\mathbf{x}_k$, since $A\mathbf{x}_k$, in the definition $\mathbf{x}_k + A\mathbf{x}_k$ of $\mathbf{x}_{k+1}$, is non negative. Assume $\mathbf{x}_{k+1}$ has the same number of zero entries as $\mathbf{x}_k$. Of course such zeros must be in the same places, i.e. there exists a permutation matrix P such that $P\mathbf{x}_k = \begin{bmatrix} \alpha \\ 0 \end{bmatrix}$, $P\mathbf{x}_{k+1} = \begin{bmatrix} \beta \\ 0 \end{bmatrix}$, with α, β positive vectors of the same dimension m, $1 \leq m \leq n-1$ (why such bounds for m ?). Thus, $P\mathbf{x}_{k+1} = P\mathbf{x}_k + PA\mathbf{x}_k = P\mathbf{x}_k + PAP^TP\mathbf{x}_k$. Consider the following partition of the matrix PAP^T

$$PAP^T = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

where M_{11} is $m \times m$. Then

$$\begin{bmatrix} \beta \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix} + \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} \alpha \\ 0 \end{bmatrix},$$

and, in particular, $M_{21}\alpha = 0$. The latter condition implies $M_{21} = 0$, being α a positive vector and M_{21} a non negative matrix. But this is equivalent to say that A is reducible, against the hypothesis! $\square$

Example. Set

$$I + A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ a & 1 & 0 & 0 \\ 0 & b & 1 & 0 \\ 0 & 0 & c & 1 \end{bmatrix}, \; a, b, c \text{ positive.}$$

Note that A is a non-negative irreducible matrix, and in fact $(I + A)^3$ is a positive matrix, i.e. its entries are positive. Moreover, 3 is the minimum j for which $(I + A)^j$ is positive.

Let A be a $n \times n$ non negative irreducible matrix. Let $\mathbf{x}$ be a non negative non null vector, and associate to $\mathbf{x}$ the number

$$r_{\mathbf{x}} := \min_{i: x_i > 0} \frac{\sum_j a_{ij} x_j}{x_i} = \min_{i: x_i > 0} \frac{(A\mathbf{x})_i}{x_i}.$$

Proposition. $r_{\mathbf{x}}$ is a non negative real number; $r_{\mathbf{x}} = r_{\alpha\mathbf{x}}$ if $\alpha > 0$; $A\mathbf{x} \geq r_{\mathbf{x}}\mathbf{x}$; $r_{\mathbf{x}} = \sup\{\rho \in \mathbb{R} : A\mathbf{x} \geq \rho\mathbf{x}\}$.

proof: easy, left to the reader.

Now associate to A the following number:

$$r = \sup_{\mathbf{x} \geq \mathbf{0}, \, \mathbf{x} \neq \mathbf{0}} r_{\mathbf{x}} = \sup_{\mathbf{x} \geq \mathbf{0}, \, \mathbf{x} \neq \mathbf{0}} \min_{i: x_i > 0} \frac{\sum_j a_{ij} x_j}{x_i}.$$

Proposition. r is a positive real number; if $\mathbf{w} \geq \mathbf{0}$, $\mathbf{w} \neq \mathbf{0}$, is such that $A\mathbf{w} \geq r\mathbf{w}$, then $A\mathbf{w} = r\mathbf{w}$ and $\mathbf{w} > \mathbf{0}$.

proof. If $\mathbf{e} = [1\,1\,\cdots\,1]^T$, then $r_{\mathbf{e}} = \min_{i:\,(\mathbf{e})_i > 0} \frac{\sum_j a_{ij} (\mathbf{e})_j}{(\mathbf{e})_i} = \min_i \sum_j a_{ij} \geq 0$. Assume $r_{\mathbf{e}} = 0$. Then for some k we would have $\sum_j a_{kj} = 0$, so the kth row of A would be null, and thus A would be reducible (exchange the k and n rows), against the hypothesis. It follows that $r \geq r_{\mathbf{e}} > 0$.

proof. Set $\eta = A\mathbf{w} - r\mathbf{w}$. We know that $\eta \geq \mathbf{0}$. Assume $\eta \neq \mathbf{0}$. Then, by the Lemma,

$$\begin{aligned} \mathbf{0} < (I + A)^{n-1}\eta &= (I + A)^{n-1}A\mathbf{w} - (I + A)^{n-1}r\mathbf{w} \\ &= A(I + A)^{n-1}\mathbf{w} - r(I + A)^{n-1}\mathbf{w} = A\mathbf{y} - r\mathbf{y}, \; \mathbf{y} > \mathbf{0}. \end{aligned}$$

i.e. $r < (A\mathbf{y})_i / y_i \; \forall\, i$. Thus $r < r_{\mathbf{y}}$, which is absurd. It follows that $\eta = \mathbf{0}$, that is, $A\mathbf{w} = r\mathbf{w}$. Then, we also have $\mathbf{w} > \mathbf{0}$ since $\mathbf{0} < (I + A)^{n-1}\mathbf{w} = (1 + r)^{n-1}\mathbf{w}$. $\square$

In the following, given $\mathbf{v} \in \mathbb{C}^n$ and $M \in \mathbb{C}^{n \times n}$ we denote by $|\mathbf{v}|$ and $|M|$, respectively, the column vector $(|v_k|)_{k=1}^n$ and the matrix $(|m_{ij}|)_{i,j=1}^n$.

Theorem. There exists a positive vector $\mathbf{z}$ for which $A\mathbf{z} = r\mathbf{z}$; $r = \rho(A)$; if $B \in \mathbb{C}^{n \times n}$, $|B| \leq A$, then $\rho(B) \leq \rho(A)$; if $B \in \mathbb{C}^{n \times n}$, $|B| \leq A$, $|B| \neq A$ then $\rho(B) < \rho(A)$.

proof: We now show that there exists $\mathbf{z} \geq \mathbf{0}$ such that $r = r_{\mathbf{z}}$. Once this is proved we will have the inequality $A\mathbf{z} \geq r_{\mathbf{z}}\mathbf{z} = r\mathbf{z}$ which implies, by the Proposition, $A\mathbf{z} = r\mathbf{z}$ and $\mathbf{z} > \mathbf{0}$.

$$r = \sup_{\mathbf{x}\geq\mathbf{0},\,\mathbf{x}\neq\mathbf{0}} r_{\mathbf{x}} = \sup_{\mathbf{x}\geq\mathbf{0},\,\mathbf{x}\neq\mathbf{0}} r_{\frac{\mathbf{x}}{\|\mathbf{x}\|_2}} = \sup_{\mathbf{x}\geq\mathbf{0},\,\|\mathbf{x}\|_2=1} r_{\mathbf{x}} = (*)$$

We have $r_{\mathbf{x}} \leq r_{(I+A)^{n-1}\mathbf{x}}$ since $r_{(I+A)^{n-1}\mathbf{x}} = \sup\{\rho \in \mathbb{R} : A(I+A)^{n-1}\mathbf{x} \geq \rho(I+A)^{n-1}\mathbf{x}\}$ and $A\mathbf{x} \geq r_{\mathbf{x}}\mathbf{x} \Rightarrow A(I+A)^{n-1}\mathbf{x} = (I+A)^{n-1}A\mathbf{x} \geq r_{\mathbf{x}}(I+A)^{n-1}\mathbf{x}$. Thus,

$$(*) \leq \sup_{\mathbf{x}\geq\mathbf{0},\,\|\mathbf{x}\|_2=1} r_{(I+A)^{n-1}\mathbf{x}} = \sup_{\mathbf{y}\in Q} r_{\mathbf{y}} = \max_{\mathbf{y}\in Q} r_{\mathbf{y}} = r_{\mathbf{z}} \leq r$$

for some $\mathbf{z} \in Q = \{(I+A)^{n-1}\mathbf{x} : \mathbf{x} \geq \mathbf{0}, \|\mathbf{x}\|_2 = 1\}$ ($r_{\mathbf{y}}$ is continuous in $\mathbf{y}$ (why?) and Q is a compact). Note that $\mathbf{z} > \mathbf{0}$.

proof: Let λ be an eigenvalue of A, i.e. $\exists\, \mathbf{y} \neq \mathbf{0} \mid A\mathbf{y} = \lambda\mathbf{y}$. Then $|\lambda||\mathbf{y}| = |\lambda\mathbf{y}| = |A\mathbf{y}| \leq A|\mathbf{y}|$, $|\mathbf{y}| \geq \mathbf{0}$, $|\mathbf{y}| \neq \mathbf{0}$. Thus, by definition of $r_{|\mathbf{y}|}$, $|\lambda| \leq r_{|\mathbf{y}|} \leq r$, and we have the inequality $\rho(A) \leq r$. But r is an eigenvalue of A, thus $r = \rho(A)$.

proof: Let λ be an eigenvalue of B, i.e. $\exists\, \mathbf{y} \neq \mathbf{0} \mid B\mathbf{y} = \lambda\mathbf{y}$. Then $|\lambda||\mathbf{y}| = |\lambda\mathbf{y}| = |B\mathbf{y}| \leq |B||\mathbf{y}| \leq A|\mathbf{y}|$, $|\mathbf{y}| \geq \mathbf{0}$, $|\mathbf{y}| \neq \mathbf{0}$. Thus, by definition of $r_{|\mathbf{y}|}$, $|\lambda| \leq r_{|\mathbf{y}|} \leq r$, and we have the inequality $\rho(B) \leq r = \rho(A)$.

proof: Assume $\rho(B) = \rho(A)$, i.e. there exists λ eigenvalue of B ($B\mathbf{y} = \lambda\mathbf{y}$, $\mathbf{y} \neq \mathbf{0}$) such that $|\lambda| = r$. Then we can add an equality in the above arguments,

$$r|\mathbf{y}| = |\lambda||\mathbf{y}| = |\lambda\mathbf{y}| = |B\mathbf{y}| \leq |B||\mathbf{y}| \leq A|\mathbf{y}|, \quad |\mathbf{y}| \geq \mathbf{0}, |\mathbf{y}| \neq \mathbf{0},$$

obtaining the inequality $r|\mathbf{y}| \leq A|\mathbf{y}|$, $|\mathbf{y}| \geq \mathbf{0}$, $|\mathbf{y}| \neq \mathbf{0}$. But by the above Proposition, such inequality implies $A|\mathbf{y}| = r|\mathbf{y}|$ with $|\mathbf{y}| > \mathbf{0}$. So we have

$$r|\mathbf{y}| = |\lambda||\mathbf{y}| = |\lambda\mathbf{y}| = |B\mathbf{y}| = |B||\mathbf{y}| = A|\mathbf{y}| = r|\mathbf{y}|, \quad |\mathbf{y}| > \mathbf{0},$$

from which it follows $|B| = A$, against the hypothesis! Thus $\rho(B) < \rho(A)$. □

We conclude this section with the following result of the Perron-Frobenius theory, stated without proof (actually we shall give a proof of such result in the case A is positive):

Result. If A is a non-negative irreducible $n \times n$ matrix, then $r = \rho(A)$ is a *simple* eigenvalue of A, i.e. $\lambda - r$ divides $p_A(\lambda)$ but $(\lambda - r)^2$ does not divide $p_A(\lambda)$.

Computing the Perron-pair of A irreducible non-negative by the power method

Let A be an irreducible non-negative $n \times n$ matrix. Then we know that $\rho(A)$ is positive and is a simple eigenvalue of A, and the corresponding eigenvector can be chosen positive. Of course, such eigenvector is uniquely defined if we require that its measure in the 1-norm is one. We also know that $\rho(A) = r := \sup_{\mathbf{x}\geq\mathbf{0},\,\mathbf{x}\neq\mathbf{0}} \min_{i:\, x_i>0} \frac{(A\mathbf{x})_i}{x_i}$.

So, it is uniquely defined the Perron-pair $(\rho(A), \mathbf{z})$. such that $A\mathbf{z} = \rho(A)\mathbf{z}$, $\rho(A)$ positive, $\mathbf{z}$ positive, $\|\mathbf{z}\|_1 = 1$. The power method is a way to compute such Perron-pair.

Theorem(power). Let A be an irreducible non-negative $n \times n$ matrix. Let $\mathbf{a}_0$ be any positive vector, and set $\mathbf{v}_0 = \mathbf{a}_0/\|\mathbf{a}_0\|_1$. Then set

$$\mathbf{a}_{k+1} = A\mathbf{v}_k, \quad \mathbf{v}_{k+1} = \frac{\mathbf{a}_{k+1}}{\|\mathbf{a}_{k+1}\|_1}, \quad k = 0, 1, 2, \dots .$$

Note that the sequences $\{\mathbf{a}_k\}$, $\{\mathbf{v}_k\}$ are well defined sequences of positive vectors, and $\|\mathbf{v}_k\|_1 = 1$, $\forall\, k$. Let X be the non singular matrix defining, by similarity, the Jordan block-diagonal form of A, i.e.

$$X^{-1}AX = J = \begin{bmatrix} r & \mathbf{0}^T \\ \mathbf{0} & B \end{bmatrix}, \quad X = \begin{bmatrix} \mathbf{z} & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix},$$

and let $r = \rho(A)$, λ_j, $j = 2, \ldots, n$, be the eigenvalues of A ($\lambda_j \neq r$, $\forall\, j$). If α in the expression $\mathbf{a}_0 = \alpha\mathbf{z} + \sum_{j=2}^n \alpha_j \mathbf{x}_j$ is nonzero, then, for $k \to +\infty$, we have

$$\mathbf{v}_k - \mathbf{z} \;\to\; \mathbf{0}, \quad \|\mathbf{a}_k\|_1 - \rho(A) \;\to\; 0,$$

provided that $|\lambda_j|$ is smaller than r for all $j = 2, \ldots, n$. In the particular case where A is diagonalizable, the rate of convergence is

$$\big(\max_{j=2\ldots n} \frac{|\lambda_j|}{r}\big)^k$$

[for the general case, use the Remark in Theorem(power) of the section on preliminary considerations].

proof: We prove the Theorem only in the case A diagonalizable, where $A\mathbf{x}_j = \lambda_j \mathbf{x}_j$, $j = 2, \ldots, n$. It is easy to observe that $A^k\mathbf{a}_0 = \alpha r^k \mathbf{z} + \sum_{j=2}^n \alpha_j \lambda_j^k \mathbf{x}_j$ is a positive vector, and that

$$\begin{aligned} \mathbf{v}_k &= \frac{\alpha r^k\mathbf{z} + \sum_{j=2}^n \alpha_j\lambda_j^k\mathbf{x}_j}{\|\alpha r^k\mathbf{z} + \sum_{j=2}^n \alpha_j\lambda_j^k\mathbf{x}_j\|_1} = \frac{\alpha r^k\mathbf{z} + \sum_{j=2}^n \alpha_j\lambda_j^k\mathbf{x}_j}{\mathbf{e}^T(\alpha r^k\mathbf{z} + \sum_{j=2}^n \alpha_j\lambda_j^k\mathbf{x}_j)} \\ &= \frac{\alpha r^k\mathbf{z} + \sum_{j=2}^n \alpha_j\lambda_j^k\mathbf{x}_j}{\alpha r^k\mathbf{e}^T\mathbf{z} + \sum_{j=2}^n \alpha_j\lambda_j^k\mathbf{e}^T\mathbf{x}_j} = \frac{\mathbf{z} + \sum_{j=2}^n \frac{\alpha_j}{\alpha}\left(\frac{\lambda_j}{r}\right)^k \mathbf{x}_j}{1 + \sum_{j=2}^n \frac{\alpha_j}{\alpha}\left(\frac{\lambda_j}{r}\right)^k \mathbf{e}^T\mathbf{x}_j}. \end{aligned}$$

Moreover,

$$\mathbf{a}_{k+1} = A\mathbf{v}_k = \frac{r\mathbf{z} + \sum_{j=2}^n \frac{\alpha_j}{\alpha}\left(\frac{\lambda_j}{r}\right)^k \lambda_j\mathbf{x}_j}{1 + \sum_{j=2}^n \frac{\alpha_j}{\alpha}\left(\frac{\lambda_j}{r}\right)^k \mathbf{e}^T\mathbf{x}_j}$$

and, since $\mathbf{a}_{k+1}$ is positive,

$$\begin{aligned} \|\mathbf{a}_{k+1}\|_1 = \mathbf{e}^T\mathbf{a}_{k+1} \quad &= \quad \frac{r + \sum_{j=2}^n \frac{\alpha_j}{\alpha}\left(\frac{\lambda_j}{r}\right)^k \lambda_j\mathbf{e}^T\mathbf{x}_j}{1 + \sum_{j=2}^n \frac{\alpha_j}{\alpha}\left(\frac{\lambda_j}{r}\right)^k \mathbf{e}^T\mathbf{x}_j} \\ &= \quad \frac{r\left(1 + \sum_{j=2}^n \frac{\alpha_j}{\alpha}\left(\frac{\lambda_j}{r}\right)^{k+1} \mathbf{e}^T\mathbf{x}_j\right)}{1 + \sum_{j=2}^n \frac{\alpha_j}{\alpha}\left(\frac{\lambda_j}{r}\right)^k \mathbf{e}^T\mathbf{x}_j}. \end{aligned}$$

Exercise. Discuss the convergence of the power method when applied to compute the Perron-pair $(1, \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix})$ of the following two 2×2 matrices:

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1/4 & 3/4 \\ 3/4 & 1/4 \end{bmatrix}.$$

In the second case, find a constant c such that $\|\mathbf{v}_k - \mathbf{z}\| \leq c(\frac{1}{2})^k$.

Exercise. Prove the Theorem in case A (non negative, irreducible) is 4×4, and there exists X non singular for which

$$X^{-1}AX = \begin{bmatrix} r & & & \\ & \lambda_2 & 1 & \\ & & \lambda_2 & 1 \\ & & & \lambda_2 \end{bmatrix},$$

with $|\lambda_2|$ smaller than r. Prove that in such case the rate of convergence is

$$|p_2(k)|\left(\left|\frac{\lambda_2}{r}\right|\right)^k,$$

where p_2 is a degree-two polynomial.

Exercise. Set

$$A = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{4} & 0 & 0 \\ \frac{3}{4} & 1 & 0 \end{bmatrix}.$$

Prove that the eigenvalues of A are $\{1, -\frac{1}{2}, -\frac{1}{2}\}$ and that A is not diagonalizable. Find $X = [\mathbf{z}\ \mathbf{x}_2\ \mathbf{x}_3]$ such that

$$X^{-1}AX = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 1 \\ 0 & 0 & -\frac{1}{2} \end{bmatrix}.$$

The only hypothesis A irreducible, non-negative, does not assure that $r = \rho(A)$ is the unique eigenvalue of A whose absolute value is equal to r. We only know that if $|\lambda_j| = r$, $j \in \{2, \ldots, n\}$, then $\lambda_j \neq r$. Let us see examples:

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \text{ Eigenvalues: } -1, 1; \quad \text{Perron-pair: } (1, \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}).$$

$$A = \begin{bmatrix} 0 & a & 0 \\ 1 & 0 & 1 \\ 0 & 1-a & 0 \end{bmatrix}, \ a \in (0,1); \text{ Eig: } -1, 0, 1; \quad \text{Perron-pair: } (1, \begin{bmatrix} \frac{a}{2} \\ \frac{1}{2} \\ \frac{1-a}{2} \end{bmatrix}).$$

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}; \text{ Eigenvalues: } 1, 1, 4; \quad \text{Perron-pair: } (4, \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}).$$

Note that if in the second example A is replaced by A^T, then there is no need of computation in obtaining the perron-pair, since it is clear that $A^T\mathbf{e} = \mathbf{e}$, $\mathbf{e} = [1\ 1\ 1]^T$. Moreover, in the third example $r = \rho(A)$ (which is 4) dominates the remaining eigevalues of A (which are $1, 1$). We note that this fact is true for any *positive* matrix A (see Theorem(positive) below), even if, as the following example shows, it is not a peculiarity of positive matrices:

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}; \text{ Eig: } 2 \mp \sqrt{2}, 2; \quad \text{Perron-pair: } (2+\sqrt{2}, \frac{1}{2+\sqrt{2}} \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix})$$

(here computation is required to obtain the Perron-pair).

Theorem(positive). If A is a $n \times n$ positive matrix and $r = \rho(A), \lambda_2, \dots, \lambda_n$ are its eigenvalues, then $|\lambda_j|$ is smaller than r for all $j = 2, \dots, n$.

proof: Set $W = A - s\mathbf{z}\mathbf{e}^T$. The eigenvalues of W are $r - s, \lambda_2, \dots, \lambda_n$ (a sketch of the proof:

$$Y := \begin{bmatrix} \mathbf{z} & \mathbf{y}_2 & \cdots & \mathbf{y}_n \end{bmatrix} \text{ non singular},$$
$$Y^{-1}AY = \begin{bmatrix} r & \cdots \\ \mathbf{0} & M \end{bmatrix}, \; Y^{-1}WY = \begin{bmatrix} r-s & \cdots \\ \mathbf{0} & M \end{bmatrix}).$$

If there is a value of s for which $|W| \leq A$, $|W| \neq A$, then $\rho(W)$ is smaller than $\rho(A)$, and thus $|\lambda_2|, \dots, |\lambda_n|$ are smaller than $r = \rho(A)$. If A is positive, then such s exists, $s = \min_{i,j} a_{ij}$. $\square$

Irreducible non-negative stochastic-by-columns A and power method

Let A be an irreducible non-negative stochastic by columns $n \times n$ matrix. Then we know that $1 = \rho(A)$ ($A^T\mathbf{e} = \mathbf{e}$, $\rho(A) \leq \|A\|_1 = 1$), that $r = 1 = \rho(A)$ is a simple eigenvalue of A, and that the corresponding eigenvector can be chosen positive. Of course, such eigenvector is uniquely defined if we require that its measure in the 1-norm is one.

So, it is uniquely defined the Perron-pair $(1 = \rho(A), \mathbf{z})$. such that $A\mathbf{z} = \mathbf{z}$, $\mathbf{z}$ positive, $\|\mathbf{z}\|_1 = 1$. The computation of such Perron-pair, i.e. the computation of the vector $\mathbf{z}$, can be performed via the power method. Actually, we now see that the power method in the particular case where A is stochastic-by-columns (besides non-negative and irreducible) can be rewritten in a simpler form and converges independently from the choice of $\mathbf{a}_0$ (provided that 1 dominates the other eigenvalues). These results follow from some remarks, reported in the following Proposition.

Proposition. Let A be an irreducible non-negative stochastic-by-columns $n \times n$ matrix. Then

i)
$$\mathbf{v} \in \mathbb{C}^n \quad \Rightarrow \quad \sum_i (A\mathbf{v})_i = \sum_i v_i$$
($\sum_i (A\mathbf{v})_i = \sum_i \sum_j a_{ij} v_j = \sum_j v_j \sum_i a_{ij} = \sum_j v_j$),

ii)
$$\mathbf{v} \text{ positive}, \; \|\mathbf{v}\|_1 = 1 \quad \Rightarrow \quad A\mathbf{v} \text{ positive}, \; \|A\mathbf{v}\|_1 = 1$$
(use the irreducibility of A and assertion i)),

iii)
$$A\mathbf{v} = \lambda\mathbf{v}, \; \lambda \neq 1, \quad \Rightarrow \quad \sum_i v_i = 0$$
($\sum_i v_i = \sum_i (A\mathbf{v})_i = \lambda \sum_i v_i$, thus $(\lambda - 1)\sum_i v_i = 0$).

Exercise. Assume that $X^{-1}AX = J$ where J is the Jordan block diagonal form of A, where A is an irreducible non-negative stochastic-by-columns $n \times n$ matrix. Assume that $[J]_{11} = 1$. Prove that $\sum_i [X]_{ij} = 0$, $j = 2, \dots, n$.

Corollary(power). Let A be an irreducible non-negative stochastic-by-columns $n \times n$ matrix. Let $\mathbf{a}_0$ be any positive vector such that $\|\mathbf{a}_0\|_1 = 1$. Then set

$$\mathbf{a}_{k+1} = A\mathbf{a}_k, \quad k = 0, 1, 2, \dots \; .$$

The sequence $\{\mathbf{a}_k\}$ is a well defined sequence of positive vectors, and $\|\mathbf{a}_k\|_1 = 1 = \rho(A)$, $\forall\, k$. If $k \to +\infty$, then

$$\mathbf{a}_k - \mathbf{z} \;\to\; \mathbf{0},$$

provided that $|\lambda_j|$ is smaller than 1 for all $j = 2, \ldots, n$. The latter event is assured by Theorem(positive) when A is positive. In the particular case where A is diagonalizable, the rate of convergence is

$$(\max_{j=2\ldots n} |\lambda_j|)^k$$

[for the general case, use the Remark in Theorem(power) of the section on preliminary considerations].

proof: The sequences $\mathbf{a}_k$ and $\mathbf{v}_k$ defined in Theorem(power) coincide, because, by Proposition ii), we have $\|A\mathbf{v}_k\|_1 = \|\mathbf{v}_k\|_1$. It remains to show that α in the expression $\mathbf{a}_0 = \alpha\mathbf{z} + \sum_j \alpha_j \mathbf{x}_j$ is nonzero. Note that $\mathbf{e}^T\mathbf{a}_0 = \alpha\mathbf{e}^T\mathbf{z} + \sum_j \alpha_j \mathbf{e}^T\mathbf{x}_j = \alpha + \sum_j \alpha_j \mathbf{e}^T\mathbf{x}_j$. In the case that A is diagonalizable the thesis follows by Proposition iii) which asserts that $\mathbf{e}^T\mathbf{x}_j$, $j = 2, \ldots, n$, must be zero. In the generic case the proof is left to the reader .... $\square$

Exercise. Discuss existence, unicity, and computation of the Perron-pair for

$$A = \begin{bmatrix} 0 & 1-a & a \\ a & 0 & 1-a \\ 1-a & a & 0 \end{bmatrix},\ a \in [0,1],\ A = \begin{bmatrix} 0 & 0 & 1 \\ a & 0 & 0 \\ 1-a & 1 & 0 \end{bmatrix},\ a \in [0,1],$$

$$A = \begin{bmatrix} 0 & b_1 & & & & \\ 1 & 0 & b_2 & & & \\ & 1-b_1 & 0 & \ddots & & \\ & & 1-b_2 & \ddots & b_{n-2} & \\ & & & \ddots & 0 & 1 \\ & & & & 1-b_{n-2} & 0 \end{bmatrix},\ b_i \in (0,1).$$

Page-rank computation [Berkhin, et al]

Consider an oriented graph with set of verteces $V = \{1, 2, \ldots, n\}$, and set of edges E, $i, j \in E$ if there is a link from i to j.

Associate to such graph the adiacency matrix:

$$L = \begin{cases} 1 & ij \in E \\ 0 & ij \notin E \end{cases}$$

Call $\deg(i)$ the number of edges starting from i. Of course, $\deg(i) = 0$ (no edge starts from i) iff $L_{ij} = 0$, $\forall\, j$. Moreover, $\deg(i) = \sum_j L_{ij}$.

Associate to the graph the transition matrix:

$$P = \begin{cases} \frac{1}{\deg(i)} & ij \in E \\ 0 & ij \notin E \end{cases}$$

Note that $P_{ij} = L_{ij}/\deg(i)$ if i is such that $\deg(i) > 0$, and $P_{ij} = L_{ij} = 0$ otherwise. Moreover, $\sum_j P_{ij} = 1$ if $\deg(i) > 0$ and $\sum_j P_{ij} = 0$ otherwise. So, the matrix P is a non negative matrix *quasi*-stochastic by rows.

Remark. Row i of P is null iff no edge starts from i; column j of P is null iff no edge points to j.

Let $\mathbf{p} \in \mathbb{R}^n$ be the vector whose entry j, p_j, is the importance (authority) of the vertex j. Then

$$p_j = \sum_{i:\, i\to j} \frac{p_i}{\deg(i)} = \sum_{i=1}^{n} P_{ij} p_i = \sum_{i=1}^{n} P^T_{ji} p_i = (P^T \mathbf{p})_j, \quad \mathbf{p} = P^T \mathbf{p}$$

(note that such fixed point $\mathbf{p}$ may not exist, or, if exists, may be not unique or with zero entries; see below).

Let $p_j^{(k+1)}$ be the probability that at step $k+1$ of my visit of the graph (navigation on the web) I am on the vertex (page) j. Then

$$\begin{aligned} p_j^{(k+1)} &= \textstyle\sum_{i:\, i\to j} \frac{p_i^{(k)}}{\deg(i)} = \sum_{i=1}^{n} P_{ij} p_i^{(k)} \\ &= \textstyle\sum_{i=1}^{n} P^T_{ji} p_i^{(k)} = (P^T \mathbf{p}^{(k)})_j, \quad \mathbf{p}^{(k+1)} = P^T \mathbf{p}^{(k)} \end{aligned}$$

(note that such sequence of vector probabilities exists and is uniquely defined, once $\mathbf{p}^{(0)}$ is given, but the $\mathbf{p}^{(k)}$ may loss the possible ddp property of $\mathbf{p}^{(0)}$; see below). [A vector $\mathbf{w}$ is said ddp (discrete distribution of probability) if $\mathbf{w}$ is positive and $\|\mathbf{w}\|_1 = 1$].

Note that it is natural to require that: $p_i^{(k)} > 0$ (at step k there is a probability that I am on vertex i), $\sum_j p_j^{(k)} = 1$ (at step k I am on some vertex); $p_i > 0$ (any vertex has a portion of importance ...), $\sum_i p_i = 1$ (... of the total 1).

So, the following facts must be true:
a) $\mathbf{p}$ such that $\mathbf{p} = P^T\mathbf{p}$, $\mathbf{p} > \mathbf{0}$, $\|\mathbf{p}\|_1 = 1$, exists and is uniquely defined,
b) the method $\mathbf{p}^{(0)} = ddp$, $\mathbf{p}^{(k+1)} = P^T \mathbf{p}^{(k)}$ converges to $\mathbf{p}$.

We now show that in order to have a) and b), the matrix P^T must be both stochastic by columns and irreducible.

Theorem(stoc). If the above facts a) and b) are true, then P^T must be stochastic by columns.

proof: We know that $\exists!\ \mathbf{p}$ such that $\mathbf{p} = P^T\mathbf{p}$, $\mathbf{p} > \mathbf{0}$, $\|\mathbf{p}\|_1 = 1$. Assume that P^T is quasi-stochastic but not stochastic by columns. Then

$$\begin{aligned} \|P^T\mathbf{p}\|_1 &= \textstyle\sum_i (P^T\mathbf{p})_i = \sum_i \sum_j (P^T)_{ij} p_j = \sum_j \sum_i P_{ji} p_j \\ &= \textstyle\sum_{j:\,\deg(j)>0} 1\cdot p_j + \sum_{j:\,\deg(j)=0} 0\cdot p_j < \sum_j p_j = \|\mathbf{p}\|_1, \end{aligned}$$

analogously,

$$\|\mathbf{p}^{(k+1)}\|_1 = \|P^T\mathbf{p}^{(k)}\|_1 \le \|\mathbf{p}^{(k)}\|_1 \le \|\mathbf{p}^{(1)}\|_1 < \|\mathbf{p}^{(0)}\|_1 = 1.$$

Thus $\mathbf{p}$ cannot be equal to $P^T\mathbf{p}$, and $\mathbf{p}^{(k)}$ cannot converge to a ddp. □

Theorem(irred). If the above facts a) and b) are true, then P^T must be irreducible.

proof: assume P^T reducible. Then there exists a permutation matrix Q such that

$$Q^T P^T Q = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}, \qquad (*)$$

with A_{11} and A_{22} at least 1×1 square matrices. Note that $Q^T P^T Q$ is stochastic by columns, like P^T. We know that $\exists!\ \mathbf{p}$ such that $\mathbf{p} = P^T\mathbf{p}$, $\mathbf{p} > \mathbf{0}$, $\|\mathbf{p}\|_1 = 1$, but this is equivalent to say that $\exists!\ Q^T\mathbf{p}$ such that

$$Q^T\mathbf{p} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} Q^T\mathbf{p}, \quad Q^T\mathbf{p} > \mathbf{0}, \ \|Q^T\mathbf{p}\|_1 = 1. \qquad (**)$$

Case 1: assume $A_{12} = 0$. Then A_{11} and A_{22} are non negative stochastic by columns matrices. We can assume they are also irreducible (why?). Then, by the Perron-Frobenius theory,

$$\exists!\ \mathbf{y}_1, \mathbf{y}_2 > \mathbf{0}, \ \|\mathbf{y}_1\|_1 = \|\mathbf{y}_2\|_1 = 1, \ \mathbf{y}_1 = A_{11}\mathbf{y}_1, \ \mathbf{y}_2 = A_{22}\mathbf{y}_2.$$

and, as a consequence, for all $\alpha \in (0,1)$ we have

$$\begin{bmatrix} \alpha\mathbf{y}_1 \\ (1-\alpha)\mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} \alpha\mathbf{y}_1 \\ (1-\alpha)\mathbf{y}_2 \end{bmatrix},$$
$$\begin{bmatrix} \alpha\mathbf{y}_1 \\ (1-\alpha)\mathbf{y}_2 \end{bmatrix} > \mathbf{0}, \ \| \begin{bmatrix} \alpha\mathbf{y}_1 \\ (1-\alpha)\mathbf{y}_2 \end{bmatrix} \|_1 = 1.$$

So, all vectors $\mathbf{p}$ such that $Q^T\mathbf{p} = \begin{bmatrix} \alpha\mathbf{y}_1 \\ (1-\alpha)\mathbf{y}_2 \end{bmatrix}$ satisfy the properties $\mathbf{p} > \mathbf{0}$, $\|\mathbf{p}\|_1 = 1$, $\mathbf{p} = P^T\mathbf{p}$, which is against the hypothesis of unicity.

Case 2: assume $A_{12} \neq 0$. Then A_{22} in (*) is not stochastic by columns, thus A_{22} may have no eigenvalue equal to 1, i.e. the equations in (**) involving A_{22} may be verified only if part of $\mathbf{p}$ is null, against the hypothesis of positiveness of $\mathbf{p}$. $\square$

Viceversa, we know that if the non negative matrix P^T is irreducible and stochastic by columns, then $1 = \rho(P^T)$ is a simple eigenvalue of P^T and there exists a unique vector $\mathbf{p}$ such that $\mathbf{p} = P^T\mathbf{p}$, $\mathbf{p} > \mathbf{0}$, $\|\mathbf{p}\|_1 = 1$. We also know that such hypotheses are not sufficient to assure the convergence (to $\mathbf{p}$) of the sequence $\mathbf{p}^{(k+1)} = P^T\mathbf{p}^{(k)}$, $\mathbf{p}^{(0)} > \mathbf{0}$, $\|\mathbf{p}_0\|_1 = 1$, or, equivalently, to assure that the remaining eigenvalues of P^T have absolute value smaller than 1. We can only say that the sequence $\{\mathbf{p}^{(k)}\}$ is a well defined sequence of ddp.

We have to modify P (the graph) so to make well posed ($\exists!$) the mathematical problem and to make convergent the algorithm for solving it.

Make P^T stochastic:

$$P' = P + \mathbf{d}\mathbf{v}^T, \quad \mathbf{d} = \begin{bmatrix} \delta_{\deg(1),0} \\ \vdots \\ \delta_{\deg(n),0} \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{bmatrix}$$

i.e, where P has null rows P' has the row vector $\mathbf{v}^T$. The vertex i with $\deg(i) = 0$ now links to all verteces of the graph. [We discuss the uniform case, but what follows can be repeated for the more general case $\mathbf{v}$ =ddp].

Observe that $(P')^T$ is stochastic by columns, $(P')^T \geq 0$, thus 1 is eigenvalue of $(P')^T$ and the other eigenvalues of $(P')^T$, $\lambda'_2, \ldots, \lambda'_n$, are such that $|\lambda'_j| \leq 1$.

Make $(P')^T$ irreducible:

$$P'' = cP' + (1-c)\mathbf{e}\mathbf{v}^T, \quad \mathbf{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \ c \in (0,1).$$

Since

$$\mathbf{e}_i^T P'' = \begin{cases} c\mathbf{v}^T + (1-c)\mathbf{v}^T = \mathbf{v}^T & \deg(i) = 0 \\ c[\cdots 0 \ \frac{1}{\deg(i)} \ 0 \cdots] + (1-c)\mathbf{v}^T & \deg(i) > 0 \end{cases}$$

we are assuming that a visitor of the graph can go from the vertex i to one of its neighborhoods with probability $c/\deg(i) + (1-c)/n$, and with probability $(1-c)/n$ to an arbitrary vertex of the graph. Of course the parameter c must be chosen near to 1, in order to maintain our model faithful to the way the graph (web) is visited.

Observe that $(P'')^T$ is stochastic by columns and positive, therefore, in particular, it is non negative and irreducible. So, we have all we need.

Theorem(page-rank). $1 = \rho((P'')^T)$ is a simple eigenvalue of $(P'')^T$, there exists a unique vector $\mathbf{p}$ such that $\mathbf{p} = (P'')^T\mathbf{p}$, $\mathbf{p} > \mathbf{0}$, $\|\mathbf{p}\|_1 = 1$ (i.e. we have fact (a)), and the other eigenvalues of $(P'')^T$, $\lambda''_2, \ldots, \lambda''_n$, are such that $|\lambda''_j| < 1$ (by Theorem(positive)). Thus, $\mathbf{p}^{(k+1)} = (P'')^T\mathbf{p}^{(k)}$, $\mathbf{p}^{(0)} > \mathbf{0}$, $\|\mathbf{p}^{(0)}\|_1 = 1$, is a sequence of ddp convergent to $\mathbf{p}$ and, in case P'' is diagonalizable,

$$\|\mathbf{p}^{(k)} - \mathbf{p}\| = O((\max_{j=2,\ldots,n} |\lambda''_j|)^k)$$

[for the general case see the Remark in Theorem(power) of Section 1] (i.e. we have fact (b)). Moreover, for the particular choice of $(P'')^T$, the cost of each step of the power method is $O(n)$ and is dominated by the cost of the matrix-vector multiplication $P^T\mathbf{z}$, and if A is diagonalizable, then the rate of convergence is

$$\|\mathbf{p}^{(k)} - \mathbf{p}\| = O(c^k)$$

[for the general case see the Remark in Theorem(power) of Section 1] (Google-search engine sets $c = 0.85$ [Berkhin]).

proof: We have to prove only the final assertions.

$O(n)$ arithmetic operations are sufficient to perform each step of the power method. We have

$$(P'')^T = c(P^T + \mathbf{v}\mathbf{d}^T) + (1-c)\mathbf{v}\mathbf{e}^T,$$

and, if $\mathbf{x} \geq \mathbf{0}$, then

$$\begin{aligned} &(P'')^T\mathbf{x} = cP^T\mathbf{x} + \gamma\mathbf{v}, \\ &\gamma = c\mathbf{d}^T\mathbf{x} + (1-c)\mathbf{e}^T\mathbf{x} = \mathbf{e}^T\mathbf{x} - c[\mathbf{e}^T\mathbf{x} - \mathbf{d}^T\mathbf{x}] = \|\mathbf{x}\|_1 - c\|P^T\mathbf{x}\|_1 \end{aligned}$$

(why the latter equality holds?). Thus, in order to compute $\mathbf{p}^{(k+1)}$ from $\mathbf{p}^{(k)}$ one can use the following function

$$\begin{aligned} &\mathbf{y} = cP^T\mathbf{x}, \\ &\gamma = \|\mathbf{x}\|_1 - \|\mathbf{y}\|_1, \\ &(P'')^T\mathbf{x} = \mathbf{y} + \gamma\mathbf{v} \end{aligned}$$

where the dominant operation is the matrix-vector multiplication $P^T\mathbf{x}$. Note that each row j of P^T has (on the average) a very small number of nonzero entries, i.e. exactly the number of vertices pointing to j, so $P^T\mathbf{x}$ can be computed with $O(n)$ arithmetic operations. Note also that in order to implement the above function one needs only $2n$ memory allocations.

Rate of convergence $\|\mathbf{p}^{(k)} - \mathbf{p}\| = O(c^k)$. We first prove that if $\mathbf{e}^T\mathbf{v} = 1$ then

$$p_{P'}(\lambda) = (\lambda - 1)p_{n-1}(\lambda) \quad \Rightarrow \quad p_{P'+\frac{1-c}{c}\mathbf{e}\mathbf{v}^T}(\lambda) = (\lambda - \frac{1}{c})p_{n-1}(\lambda). \qquad (***)$$

In fact,

$$S = \begin{bmatrix} \mathbf{e} & \mathbf{y}_2 & \cdots & \mathbf{y}_n \end{bmatrix}, \ \det(S) \neq 0,$$

$$S^{-1}P'S = \begin{bmatrix} 1 & \mathbf{u}^T \\ \mathbf{0} & M \end{bmatrix}, \ p_{P'}(\lambda) = (\lambda - 1)p_M(\lambda),$$

$$S^{-1}(P' + \tfrac{1-c}{c}\mathbf{e}\mathbf{v}^T)S = \begin{bmatrix} 1 & \mathbf{u}^T \\ \mathbf{0} & M \end{bmatrix} + \tfrac{1-c}{c}S^{-1}\mathbf{e}\mathbf{v}^T S$$
$$= \begin{bmatrix} 1 & \mathbf{u}^T \\ \mathbf{0} & M \end{bmatrix} + \tfrac{1-c}{c}\begin{bmatrix} 1 & \cdots & * \\ 0 & \cdots & 0 \\ & & \end{bmatrix} = \begin{bmatrix} \frac{1}{c} & \tilde{\mathbf{u}}^T \\ \mathbf{0} & M \end{bmatrix}.$$

As a consequence of (***), if $1, \lambda'_2, \ldots, \lambda'_n$ are the eigenvalues of P' (recall that $|\lambda'_j| \leq 1$, $j = 2, \ldots, n$), then the eigenvalues of $P' + \frac{1-c}{c}\mathbf{e}\mathbf{v}^T$ are $\frac{1}{c}, \lambda'_2, \ldots, \lambda'_n$, and thus the eigenvalues of $cP' + (1-c)\mathbf{e}\mathbf{v}^T$ are $1, c\lambda'_2, \ldots, c\lambda'_n$.

It follows that if A is diagonalizable, then $\|\mathbf{p}^{(k)} - \mathbf{p}\| = O((\max_{j=2,\ldots,n}|c\lambda'_j|)^k) = O(c^k)$. $\square$

Why to compute $\mathbf{p}$? [Berkhin].

QUERY: Berkhin survey.

Go in the inverted terms document file, which is a table containing a row for each term of a collection's dictionary. In such file, for each term there is a list of all documents that contain such term

$$\begin{array}{l} \vdots \\ \text{term} \ \rightarrow \ \text{LISTA}_{term} = \{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\} \\ \vdots \\ \text{Berkhin} \ \rightarrow \ \text{LISTA}_{Berkhin} = \{1, 4, 6\} \\ \vdots \\ \text{survey} \ \rightarrow \ \text{LISTA}_{survey} = \{1, 3\} \\ \vdots \end{array}$$

Define the set of relevance of the query

$$\cup_{term \in \text{QUERY}} \text{LISTA}_{term} = \{1, 3, 4, 6\}$$

Reading $\mathbf{p}$ ($\mathbf{p}$ is updated once each month) considers and orders the corresponding set of authorities $\{p_1, p_3, p_4, p_6\}$, for example $p_4 \geq p_6 \geq p_3 \geq p_1$

Finally, show the titles of the documents $1, 3, 4, 6$ in the order $4, 6, 3, 1$, from the one with greatest authority to the one with smallest authority.

Main criticism: This procedure, being independent from the query allows a fast answer, but does not make distinction between pages with authority from pages with authority on a specific subject.

Exercise. Draw the graph whose transition matrix is

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Note that P is non negative, reducible and quasi-stochastic, but not stochastic, by rows. Prove that there is no positive vector $\mathbf{p}$ such that $\mathbf{p} = P^T\mathbf{p}$. Starting from P and proceeding as indicated in the theory, introduce a non negative matrix P' stochastic by rows. Note that P' is reducible, like P. Prove that there is no positive vector $\mathbf{p}$ such that $\mathbf{p} = (P')^T\mathbf{p}$. Starting from P' and proceeding as indicated in the theory, introduce a non negative matrix P'' irreducible and stochastic by rows. Prove that there exists a unique positive vector $\mathbf{p}$ such that $\mathbf{p} = (P'')^T\mathbf{p}$, $\|\mathbf{p}\|_1 = 1$, and describe an algorithm for the computation of $\mathbf{p}$. Here below is an approximation of such vector $\mathbf{p}$:

$$\mathbf{p}^T = [0.03721\ 0.05396\ 0.04151\ 0.3751\ 0.206\ 0.2862].$$

Assume, for example, that the set of relevance of a query is $\{1, 3, 4, 6\}$. Then the documents $1, 3, 4, 6$ are listed in the order 4, 6, 3, 1, being $p_4 \geq p_6 \geq p_3 \geq p_1$.

APPENDIX

Proof (Theorem(inverse power)):
Assume A diagonalizable, $A\mathbf{x}_j = \lambda_j\mathbf{x}_j$, $\{\mathbf{x}_j\}_{j=1}^n$ linearly independent. Then

$$A\mathbf{x}_j - \lambda_i^*\mathbf{x}_j = (\lambda_j - \lambda_i^*)\mathbf{x}_j \quad \Rightarrow \quad (A - \lambda_i^* I)^{-m}\mathbf{x}_j = \frac{1}{(\lambda_j - \lambda_i^*)^m}\mathbf{x}_j.$$

Call α_j the numbers for which $\mathbf{v}_0 = \sum_{j:\,\lambda_j=\lambda_i} \alpha_j\mathbf{x}_j + \sum_{j:\,\lambda_j\neq\lambda_i} \alpha_j\mathbf{x}_j$. Note that the first sum is a non null vector (by the assumption on $\mathbf{v}_0$). Then the following equality holds

$$(\lambda_i - \lambda_i^*)^m (A - \lambda_i^* I)^{-m}\mathbf{v}_0 = \sum_{j:\,\lambda_j=\lambda_i} \alpha_j\mathbf{x}_j + \sum_{j:\,\lambda_j\neq\lambda_i} \alpha_j \Big(\frac{\lambda_i - \lambda_i^*}{\lambda_j - \lambda_i^*}\Big)^m \mathbf{x}_j$$

which, for $m \to +\infty$, yields the thesis.

Assume that A is not diagonalizable. In this case call $\mathbf{x}_j$, $j = 1, \ldots, n$, the columns of the non singular matrix X for which $X^{-1}AX = J$ where J is the block diagonal Jordan form of A. Assume that the upper-left diagonal block of J is diagonal and its diagonal contains all λ_j such that $\lambda_j = \lambda := \lambda_i$. Assume that there are t of such eigenvalues. Then consider any other diagonal block of J; of course it corresponds to an eigenvalue λ_j different from λ, call such eigenvalue μ. The block under consideration has an order, say $s = s_\mu$.

$$J = \begin{bmatrix} \lambda & & & & & & & & & \\ & \lambda & & & & & & & & \\ & & \ddots & & & & & & & \\ & & & \lambda & & & & & & \\ & & & & \bullet & & & & & \\ & & & & & \mu & 1 & & & \\ & & & & & & \mu & \ddots & & \\ & & & & & & & \ddots & 1 & \\ & & & & & & & & \mu & \\ & & & & & & & & & \bullet \end{bmatrix}.$$

Restrict the matrix equation $X^{-1}AX = J$ to this block, thus, for some $r \geq t$ we have

$$A[\mathbf{x}_{r+1}\ \mathbf{x}_{r+2}\ \mathbf{x}_{r+3}\ \cdots\ \mathbf{x}_{r+s}] = [\mathbf{x}_{r+1}\ \mathbf{x}_{r+2}\ \mathbf{x}_{r+3}\ \cdots\ \mathbf{x}_{r+s}] \begin{bmatrix} \mu & 1 & & \\ & \mu & \ddots & \\ & & \ddots & 1 \\ & & & \mu \end{bmatrix}.$$

It follows that

$$\begin{aligned} &A\mathbf{x}_{r+1} = \mu\mathbf{x}_{r+1},\ (A - \lambda^* I)\mathbf{x}_{r+1} = (\mu - \lambda^*)\mathbf{x}_{r+1}, \\ &(A - \lambda^* I)^{-1}\mathbf{x}_{r+1} = \tfrac{1}{\mu-\lambda^*}\mathbf{x}_{r+1},\ (A - \lambda^* I)^{-m}\mathbf{x}_{r+1} = \tfrac{1}{(\mu-\lambda^*)^m}\mathbf{x}_{r+1}. \end{aligned}$$

$$\begin{aligned} &A\mathbf{x}_{r+2} = \mu\mathbf{x}_{r+2} + \mathbf{x}_{r+1},\ (A - \lambda^* I)\mathbf{x}_{r+2} = (\mu - \lambda^*)\mathbf{x}_{r+2} + \mathbf{x}_{r+1}, \\ &(A - \lambda^* I)^{-1}\mathbf{x}_{r+2} = \tfrac{1}{\mu-\lambda^*}\mathbf{x}_{r+2} - \tfrac{1}{(\mu-\lambda^*)^2}\mathbf{x}_{r+1}, \\ &(A - \lambda^* I)^{-m}\mathbf{x}_{r+2} = \tfrac{1}{(\mu-\lambda^*)^m}\mathbf{x}_{r+2} - \tfrac{m}{(\mu-\lambda^*)^{m+1}}\mathbf{x}_{r+1}. \end{aligned}$$

$$\begin{aligned} &A\mathbf{x}_{r+3} = \mu\mathbf{x}_{r+3} + \mathbf{x}_{r+2},\ (A - \lambda^* I)\mathbf{x}_{r+3} = (\mu - \lambda^*)\mathbf{x}_{r+3} + \mathbf{x}_{r+2}, \\ &(A - \lambda^* I)^{-1}\mathbf{x}_{r+3} = \tfrac{1}{\mu-\lambda^*}\mathbf{x}_{r+3} - \tfrac{1}{(\mu-\lambda^*)^2}\mathbf{x}_{r+2} + \tfrac{1}{(\mu-\lambda^*)^3}\mathbf{x}_{r+1}, \\ &(A - \lambda^* I)^{-m}\mathbf{x}_{r+3} = \tfrac{1}{(\mu-\lambda^*)^m}\mathbf{x}_{r+3} - \tfrac{m}{(\mu-\lambda^*)^{m+1}}\mathbf{x}_{r+2} + \tfrac{\frac{1}{2}(m^2+m)}{(\mu-\lambda^*)^{m+2}}\mathbf{x}_{r+1}. \end{aligned}$$

$$\begin{aligned} &A\mathbf{x}_{r+4} = \mu\mathbf{x}_{r+4} + \mathbf{x}_{r+3},\ (A - \lambda^* I)\mathbf{x}_{r+4} = (\mu - \lambda^*)\mathbf{x}_{r+4} + \mathbf{x}_{r+3}, \\ &(A - \lambda^* I)^{-1}\mathbf{x}_{r+4} = \tfrac{1}{\mu-\lambda^*}\mathbf{x}_{r+4} - \tfrac{1}{(\mu-\lambda^*)^2}\mathbf{x}_{r+3} \\ &\qquad + \tfrac{1}{(\mu-\lambda^*)^3}\mathbf{x}_{r+2} - \tfrac{1}{(\mu-\lambda^*)^4}\mathbf{x}_{r+1}, \\ &(A - \lambda^* I)^{-m}\mathbf{x}_{r+4} = \tfrac{1}{(\mu-\lambda^*)^m}\mathbf{x}_{r+4} - \tfrac{m}{(\mu-\lambda^*)^{m+1}}\mathbf{x}_{r+3} \\ &\qquad + \tfrac{\frac{1}{2}(m^2+m)}{(\mu-\lambda^*)^{m+2}}\mathbf{x}_{r+2} - \tfrac{\frac{1}{6}m^3+\frac{1}{2}m^2+\frac{1}{3}m}{(\mu-\lambda^*)^{m+3}}\mathbf{x}_{r+1}, \end{aligned}$$

$$(A - \lambda^* I)^{-m}\mathbf{x}_{r+s} = \frac{1}{(\mu - \lambda^*)^m}\mathbf{x}_{r+s} - \ldots + (-1)^{s-1}\frac{q_{s-1}(m)}{(\mu - \lambda^*)^{m+s-1}}\mathbf{x}_{r+1},$$

$$\begin{aligned} &(A - \lambda^* I)^{-m}[\mathbf{x}_{r+1}\ \mathbf{x}_{r+2}\ \mathbf{x}_{r+3}\cdots] \\ &\quad = [\mathbf{x}_{r+1}\ \mathbf{x}_{r+2}\ \mathbf{x}_{r+3}\cdots] \begin{bmatrix} \frac{1}{(\mu-\lambda^*)^m} & -\frac{m}{(\mu-\lambda^*)^{m+1}} & \frac{\frac{1}{2}(m^2+m)}{(\mu-\lambda^*)^{m+2}} & \cdots \\ 0 & \frac{1}{(\mu-\lambda^*)^m} & -\frac{m}{(\mu-\lambda^*)^{m+1}} & \\ 0 & 0 & \frac{1}{(\mu-\lambda^*)^m} & \\ \vdots & & & \ddots \end{bmatrix}. \end{aligned}$$

Set also $\lambda^* := \lambda_i^*$ (λ_i^* is the given approximation of $\lambda = \lambda_i$). Choose $\mathbf{v}_0$, and call α_j the numbers for which $\mathbf{v}_0 = \sum_{j:\,\lambda_j=\lambda} \alpha_j \mathbf{x}_j + \{\ \ldots\ + \sum_{j=r+1}^{r+s} \alpha_j \mathbf{x}_j +\ \ldots\ \}$. Note that $\sum_{j:\,\lambda_j=\lambda} \alpha_j \mathbf{x}_j = \sum_{j=1}^{t} \alpha_j \mathbf{x}_j$. Then

$$\begin{aligned}
&(A-\lambda^* I)^{-m}\mathbf{v}_0\\
&= \textstyle\sum_{j:\,\lambda_j=\lambda} \alpha_j (A-\lambda^* I)^{-m}\mathbf{x}_j + \{\ \cdots\ + \sum_{j=r+1}^{r+s} \alpha_j (A-\lambda^* I)^{-m}\mathbf{x}_j\ \ + \cdots\ \}\\
&= \textstyle\sum_{j:\,\lambda_j=\lambda} \alpha_j \frac{1}{(\lambda-\lambda^*)^m}\mathbf{x}_j + \{\ \cdots\ + [\ \alpha_{r+1}(\frac{1}{(\mu-\lambda^*)^m}\mathbf{x}_{r+1})\\
&+\alpha_{r+2}(\tfrac{1}{(\mu-\lambda^*)^m}\mathbf{x}_{r+2} - \tfrac{m}{(\mu-\lambda^*)^{m+1}}\mathbf{x}_{r+1})\\
&+\alpha_{r+3}(\tfrac{1}{(\mu-\lambda^*)^m}\mathbf{x}_{r+3} - \tfrac{m}{(\mu-\lambda^*)^{m+1}}\mathbf{x}_{r+2} + \tfrac{\frac{1}{2}(m^2+m)}{(\mu-\lambda^*)^{m+2}}\mathbf{x}_{r+1})\\
&+\alpha_{r+4}(\tfrac{1}{(\mu-\lambda^*)^m}\mathbf{x}_{r+4} - \tfrac{m}{(\mu-\lambda^*)^{m+1}}\mathbf{x}_{r+3}\\
&\qquad\qquad + \tfrac{\frac{1}{2}(m^2+m)}{(\mu-\lambda^*)^{m+2}}\mathbf{x}_{r+2} - \tfrac{\frac{1}{6}m^3+\frac{1}{2}m^2+\frac{1}{3}m}{(\mu-\lambda^*)^{m+3}}\mathbf{x}_{r+1})\\
&+\ldots+\alpha_{r+s}(\tfrac{1}{(\mu-\lambda^*)^m}\mathbf{x}_{r+s} -\ \ldots\ + (-1)^{s-1}\tfrac{q_{s-1}(m)}{(\mu-\lambda^*)^{m+s-1}}\mathbf{x}_{r+1})\ ] +\ \cdots\ \}.
\end{aligned}$$

But this implies

$$\begin{aligned}
&(\lambda-\lambda^*)^m (A-\lambda^* I)^{-m}\mathbf{v}_0\\
&= \textstyle\sum_{j:\,\lambda_j=\lambda} \alpha_j \mathbf{x}_j + \{\ \ldots\ + \left(\frac{\lambda-\lambda^*}{\mu-\lambda^*}\right)^m\\
&\Big[\big(\alpha_{r+1} - \alpha_{r+2}\tfrac{m}{\mu-\lambda^*} + \alpha_{r+3}\tfrac{\frac{1}{2}(m^2+m)}{(\mu-\lambda^*)^2} \ldots + (-1)^{s-1}\alpha_{r+s}\tfrac{q_{s-1}(m)}{(\mu-\lambda^*)^{s-1}}\big)\mathbf{x}_{r+1}\\
&+\big(\alpha_{r+2} - \alpha_{r+3}\tfrac{m}{\mu-\lambda^*} + \alpha_{r+4}\tfrac{\frac{1}{2}(m^2+m)}{(\mu-\lambda^*)^2} \ldots + (-1)^{s-2}\alpha_{r+s}\tfrac{q_{s-2}(m)}{(\mu-\lambda^*)^{s-2}}\big)\mathbf{x}_{r+2}\\
&+\ldots+\big(\alpha_{r+s}\big)\mathbf{x}_{r+s}\Big]\ \ + \cdots\ \},
\end{aligned}$$

from which it is clear that the sequence $(\lambda-\lambda^*)^m(A-\lambda^* I)^{-m}\mathbf{v}_0$ converges to an eigenvector of A associated with λ. The assertions about the rate of convergence follow by setting

$$\begin{aligned}
p_{s-1}(\lambda) \quad = \quad & \alpha_{r+1} - \alpha_{r+2}\tfrac{m}{\mu-\lambda^*} + \alpha_{r+3}\tfrac{\frac{1}{2}(m^2+m)}{(\mu-\lambda^*)^2}\\
& -\alpha_{r+4}\tfrac{\frac{1}{6}m^3+\frac{1}{2}m^2+\frac{1}{3}m}{(\mu-\lambda^*)^3} \ldots + (-1)^{s-1}\alpha_{r+s}\tfrac{q_{s-1}(m)}{(\mu-\lambda^*)^{s-1}}.
\end{aligned}$$

Proof (Theorem(power)):
Assume that A is diagonalizable, so $A\mathbf{x}_j = \lambda_j \mathbf{x}_j$, $\{\mathbf{x}_j\}_{j=1}^n$ linearly independent. Choose $\mathbf{v}_0$, and call α_j so that $\mathbf{v}_0 = \sum_{j:\,\lambda_j=\lambda_1} \alpha_j \mathbf{x}_j + \sum_{j:\,\lambda_j\neq\lambda_1} \alpha_j \mathbf{x}_j$. Note that the first sum is non null by assumption. Then we have the equalities:

$$\begin{aligned}
&A^m\mathbf{v}_0 = \lambda_1^m \textstyle\sum_{j:\,\lambda_j=\lambda_1} \alpha_j \mathbf{x}_j + \sum_{j:\,\lambda_j\neq\lambda_1} \alpha_j \lambda_j^m \mathbf{x}_j,\\
&\tfrac{1}{\lambda_1^m} A^m\mathbf{v}_0 = \textstyle\sum_{j:\,\lambda_j=\lambda_1} \alpha_j \mathbf{x}_j + \sum_{j:\,\lambda_j\neq\lambda_1} \alpha_j \big(\frac{\lambda_j}{\lambda_1}\big)^m \mathbf{x}_j.
\end{aligned}$$

The thesis follows letting m go to infinite.

Assume that A is not diagonalizable. Set $\mu := \lambda_j$, where $\lambda_j \neq \lambda_1$, and restrict, as above, the Jordan matrix equation $X^{-1}AX = J$ to a diagonal Jordan block of order s associated with μ. Then

$$\begin{aligned}
&A\mathbf{x}_{r+1} = \mu\mathbf{x}_{r+1},\ \ A^m\mathbf{x}_{r+1} = \mu^m\mathbf{x}_{r+1},\\
&A\mathbf{x}_{r+2} = \mu\mathbf{x}_{r+2} + \mathbf{x}_{r+1},\ \ A^m\mathbf{x}_{r+2} = \mu^m\mathbf{x}_{r+2} + m\mu^{m-1}\mathbf{x}_{r+1},\\[1ex]
&A\mathbf{x}_{r+3} = \mu\mathbf{x}_{r+3} + \mathbf{x}_{r+2},\\
&A^m\mathbf{x}_{r+3} = \mu^m\mathbf{x}_{r+3} + m\mu^{m-1}\mathbf{x}_{r+2} + \tfrac{1}{2}(m^2-m)\mu^{m-2}\mathbf{x}_{r+1},
\end{aligned}$$

$$
\begin{aligned}
&A\mathbf{x}_{r+4} = \mu\mathbf{x}_{r+4} + \mathbf{x}_{r+3},\\
&A^m\mathbf{x}_{r+4} = \mu^m\mathbf{x}_{r+4} + m\mu^{m-1}\mathbf{x}_{r+3} + \tfrac{1}{2}(m^2-m)\mu^{m-2}\mathbf{x}_{r+2}\\
&\qquad +(\tfrac{1}{6}m^3 - \tfrac{1}{2}m^2 + \tfrac{1}{3}m)\mu^{m-3}\mathbf{x}_{r+1},
\end{aligned}
$$

and so on. Set $\lambda := \lambda_1$ and $t = m_a(\lambda_1) = m_g(\lambda_1)$, so we can assume that the upper-left $t \times t$ submatrix of J is diagonal with diagonal entries all equal to λ. Then

$$
\begin{aligned}
A^m\mathbf{v}_0 &= A^m\big(\textstyle\sum_{j=1}^t \alpha_j\mathbf{x}_j + \{\ldots + \sum_{j=r+1}^{r+s}\alpha_j\mathbf{x}_j + \ldots\}\big)\\
&= \textstyle\sum_{j=1}^t \alpha_j\lambda^m\mathbf{x}_j + \{\ldots + \sum_{j=r+1}^{r+s}\alpha_j A^m\mathbf{x}_j + \ldots\}\\
&= \lambda^m \textstyle\sum_{j=1}^t \alpha_j\mathbf{x}_j + \{\ldots + [\ \alpha_{r+1}A^m\mathbf{x}_{r+1} + \alpha_{r+2}A^m\mathbf{x}_{r+2}\\
&\quad +\alpha_{r+3}A^m\mathbf{x}_{r+3} + \alpha_{r+4}A^m\mathbf{x}_{r+4} + \ldots + \alpha_{r+s}A^m\mathbf{x}_{r+s}\ ] + \ldots\}\\
&= \lambda^m \textstyle\sum_{j=1}^t \alpha_j\mathbf{x}_j\\
&\quad +\{\ldots + [\ \alpha_{r+1}(\mu^m\mathbf{x}_{r+1}) + \alpha_{r+2}(\mu^m\mathbf{x}_{r+2} + m\mu^{m-1}\mathbf{x}_{r+1})\\
&\quad +\alpha_{r+3}(\mu^m\mathbf{x}_{r+3} + m\mu^{m-1}\mathbf{x}_{r+2} + \tfrac{1}{2}(m^2-m)\mu^{m-2}\mathbf{x}_{r+1})\\
&\quad +\alpha_{r+4}(\mu^m\mathbf{x}_{r+4} + m\mu^{m-1}\mathbf{x}_{r+3} + \tfrac{1}{2}(m^2-m)\mu^{m-2}\mathbf{x}_{r+2}\\
&\quad +(\tfrac{1}{6}m^3 - \tfrac{1}{2}m^2 + \tfrac{1}{3}m)\mu^{m-3}\mathbf{x}_{r+1})\\
&\quad +\ldots + \alpha_{r+s}(\mu^m\mathbf{x}_{r+s} + \ldots + q_{s-1}(m)\mu^{m-s+1}\mathbf{x}_{r+1})\ ] + \ldots\},
\end{aligned}
$$

$$
\begin{aligned}
A^m\mathbf{v}_0 &= \lambda^m \textstyle\sum_{j=1}^t \alpha_j\mathbf{x}_j\\
&\quad +\{\ldots + [\ (\alpha_{r+1}\mu^m + \alpha_{r+2}m\mu^{m-1} + \alpha_{r+3}\tfrac{1}{2}(m^2-m)\mu^{m-2}\\
&\quad +\alpha_{r+4}(\tfrac{1}{6}m^3 - \tfrac{1}{2}m^2 + \tfrac{1}{3}m)\mu^{m-3}\\
&\quad +\ldots + \alpha_{r+s}q_{s-1}(m)\mu^{m-s+1})\mathbf{x}_{r+1}\\
&\quad +(\alpha_{r+2}\mu^m + \alpha_{r+3}m\mu^{m-1} + \alpha_{r+4}\tfrac{1}{2}(m^2-m)\mu^{m-2} + \ldots\\
&\quad +\alpha_{r+s}q_{s-2}(m)\mu^{m-s+2})\mathbf{x}_{r+2} + \ldots + (\alpha_{r+s}\mu^m)\mathbf{x}_{r+s}\ ] + \ldots\},
\end{aligned}
$$

$$
\begin{aligned}
\tfrac{1}{\lambda^m}A^m\mathbf{v}_0 &= \textstyle\sum_{j=1}^t \alpha_j\mathbf{x}_j + \{\ldots + \left(\tfrac{\mu}{\lambda}\right)^m [\ (\alpha_{r+1} + \alpha_{r+2}\tfrac{m}{\mu} + \alpha_{r+3}\tfrac{\frac{1}{2}(m^2-m)}{\mu^2}\\
&\quad +\alpha_{r+4}\tfrac{\frac{1}{6}m^3 - \frac{1}{2}m^2 + \frac{1}{3}m}{\mu^3} + \ldots + \alpha_{r+s}\tfrac{q_{s-1}(m)}{\mu^{s-1}})\mathbf{x}_{r+1}\\
&\quad +(\alpha_{r+2} + \alpha_{r+3}\tfrac{m}{\mu} + \alpha_{r+4}\tfrac{\frac{1}{2}(m^2-m)}{\mu^2} + \ldots + \alpha_{r+s}\tfrac{q_{s-2}(m)}{\mu^{s-2}})\mathbf{x}_{r+2}\\
&\quad +\ldots + (\alpha_{r+s})\mathbf{x}_{r+s}\ ] + \ldots\}.
\end{aligned}
$$

It is clear that, as m goes to infinite, the sequence $\frac{1}{\lambda^m}A^m\mathbf{v}_0$ converges to an eigenvector of A associated with λ, the dominant eigenvalue of A. Finally, the assertions on the rate of convergence follow by setting

$$
\begin{aligned}
p_{s-1}(m) &= \alpha_{r+1} + \alpha_{r+2}\tfrac{m}{\mu} + \alpha_{r+3}\tfrac{\frac{1}{2}(m^2-m)}{\mu^2}\\
&\quad +\alpha_{r+4}\tfrac{\frac{1}{6}m^3 - \frac{1}{2}m^2 + \frac{1}{3}m}{\mu^3} + \ldots + \alpha_{r+s}\tfrac{q_{s-1}(m)}{\mu^{s-1}}.
\end{aligned}
$$