GOOGLE PAGERANKING PROBLEM: THE MODEL AND THE ANALYSIS

ANTONIO CICONE* AND STEFANO SERRA-CAPIZZANO†

Abstract. The spectral and Jordan structures of the Web hyperlink matrix $G(c) = cG + (1-c)ev^T$ have been analyzed when G is the basic (stochastic) Google matrix, c is a real parameter such that $0 < c < 1$, v is a nonnegative probability vector, and e is the all-ones vector. Typical studies have relied heavily on special properties of nonnegative, positive, and stochastic matrices. There is a unique nonnegative vector $y(c)$ such that $y(c)^T G(c) = y(c)^T$ and $y(c)^T e = 1$. This *PageRank* vector $y(c)$ can be computed effectively by the power method.

We consider a square complex matrix A and nonzero complex vectors x and v such that $Ax = \lambda x$ and $v^* x = 1$. We use standard matrix analytic tools to determine the eigenvalues, the Jordan blocks, and a distinguished left λ-eigenvector of $A(c) = cA + (1-c)\lambda x v^*$ as a function of a complex variable c. If λ is a semisimple eigenvalue of A, there is a uniquely determined projection N such that $\lim_{c\to 1} y(c) = Nv$ for all v; this limit may fail to exist for some v if λ is not semisimple. As a special case of our results, we obtain a complex analog of PageRank for the Web hyperlink matrix $G(c)$ with a complex parameter c. We study regularity, limits, expansions, and conditioning of $y(c)$ and we propose algorithms (e.g., complex extrapolation, power method on a modified matrix etc.) that may provide an efficient way to compute PageRank also with c close or equal to 1. An interpretation of the limit vector Nv and a related critical discussion on the model, on its adherence to reality, and possible ways for its improvement, represent the contribution of the paper on modeling issues.

Keywords. Google matrix, PageRanking, surfing model, rank-one perturbation, Brauer's Theorem, Jordan Canonical Form, principle of biorthogonality, extrapolation formulae.

AMS subject classifications. 65F10, 65F15, 65Y20, 15A18, 15A21, 15A51.

This material was partly presented at the Seminar "Web Information Retrieval and Linear Algebra Algorithms", Schloss Dagstuhl-Wadern (Germany), february 2007, see [38], and at the Workshop "Numerical Linear Algebra in Internet Algorithms", Monopoli - Bari (Italy), September 2007. The resulting paper could be viewed into two parts: the part regarding the model is new and the part regarding the analysis is based for large part on a joint work with Roger Horn [24] and on [37].

1. Introduction. As customary in the literature (see e.g. [31]), the Web can be regarded as a huge directed graph whose n nodes are all the Web pages and whose edges are constituted by all the direct links between pages. If $\deg(i)$ indicates the cardinality of the pages different from i which are reached by a direct link from page i, the simplest Google matrix G is defined as $G_{i,j} = 1/\deg(i)$ if $\deg(i) > 0$ and there exists a link in the Web from page i to a certain page $j \neq i$. In the case where $\deg(i) = 0$ (the so-called dangling nodes), we set $G_{i,j} = 1/n$ where n is the size of the matrix, i.e., the cardinality of all the Web pages. This definition is a model for the behavior of a generic Web user: if the user is visiting page i with $\deg(i) > 0$, then with probability $1/\deg(i)$ he/she will move to one of the pages $j \neq i$ linked by i; if i is a dangling node, i.e., it has no links, then the user will make just a random choice with uniform distribution $1/n$. The basic PageRank is an n sized vector which gives a measure of the importance of every page in the Web and this notion of importance of a given page is measured according to the limit probability that a generic user

*Dipartimento di Matematica, Università degli Studi di L'Aquila, Via Vetoio, Loc. Coppito, 67100 L'Aquila, Italia (antonio.cicone@univaq.it)

†Dipartimento di Fisica e Matematica, Università dell'Insubria - Sede di Como, Via Valleggio 11, 22100 Como, Italia (stefano.serrac@uninsubria.it, serra@mail.dm.unipi.it).

reaches that page asymptotically, i.e., after infinitely many clicks: this is the surfing model. On the other hand, we would like to have a more intrinsic and intuitive notion of importance or ranking of the Web pages. Indeed, taking inspiration from social sciences, the following ideas are quite natural:

- a page j is more important if there exists a page i referring to it,
- if i is a "very important page" and is referring j, then the importance of j is increased,
- if i is referring to many pages including $j \neq i$, i.e. $\deg(i)$ is large, then this adds little importance to j.

It is worth mentioning that the idea contained in the above itemized sentences is exactly a quantification of the notion of VIP (very important people) appearing in social sciences or, quite equivalently, according to a famous sentence of the PopArt master Andy Warhol "Don't pay attention to what they write about you. Just measure it in inches" (with several distinguished precursors "I don't care what you say about me, as long as you say something about me, and as long as you spell my name right" (George Cohan), "The only thing worse than being talked about is not being talked about" (Oscar Wilde), etc.). By the way, this basic observation shows the large potential of these researches in terms of the broad range of possible applications; see e.g. [4] for a recent study in the context of bibliometry.

Now we translate in formulae these concepts. More in detail, after a reasonable normalization, for every $j = 1, \ldots, n$, the importance $y[j]$ of page j is defined as follows

$$y[j] = \sum_{i \to j} \frac{y[i]}{\deg(i)}, \quad y[j] \geq 0, \qquad \sum_{i=1}^{n} y[i] = 1.$$

The definition is nice in principle and can be interpreted in matrix-vector terms as $y^T \hat{G} = y^T$, $y[j] \geq 0$, for all j, $\sum_{i=1}^{n} y[i] = 1$, where $\hat{G}_{i,j} = G_{i,j}$ if there exists in the Web a link from i to j and $\hat{G}_{i,j} = 0$ otherwise: G and $\hat{G}$ are the same with the exception of the management of dangling nodes. However, even by interpreting the above relations as an eigenvector problem with respect to the eigenvalue 1, either 1 may belong or may fail to belong to the spectrum of the resulting matrix. Explicit and very simple examples can be constructed: take e.g. the matrix

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

associated to a toy Web with only two nodes i, j with $i < j$ and a unique link from i to j; it is clear that the problem defined by $y^T \hat{G} = y^T$, $y[j] \geq 0$, for all j, $\sum_{i=1}^{n} y[i] = 1$ has no solution, since 1 is not in the spectrum of $\hat{G}$. The reason is again the presence of dangling nodes that in turn implies the existence of identically zero rows. Hence, for giving a solution to the above mathematical incongruence, we define

$$\begin{aligned} &\deg^*(i) = \deg(i), \quad \text{if} \quad \deg(i) > 0, \\ &\deg^*(i) = n, \quad \text{if} \quad \deg(i) = 0, \end{aligned} \tag{1.1}$$

and we correct accordingly the relations concerning $y[j]$ in the following way:

$$y[j] = \sum_{i \to j} \frac{y[i]}{\deg^*(i)}, \quad y[j] \geq 0, \qquad \sum_{i=1}^{n} y[i] = 1.$$

Putting the above relations in matrix terms, and introducing the l^1 norm of a real or complex vector w as $\|w\|_1 = \sum_{j=1}^{n} |w[j]|$, we have

$$y^T G = y^T, \quad y \geq 0, \quad \|y\|_1 = 1. \tag{1.2}$$

Interestingly enough, it should be observed that any vector y solution to (1.2) represents also a solution in the sense of the surfing model and vice versa. Therefore, in other words, with the above choice, there is an identification, which can be criticized between the surfing model and the definition of importance: in fact the definition of G referred to the dangling nodes is perfectly coherent in the surfing model, while is not justified at all when defining a notion of importance (see Section 2).

Now looking at (1.2), a basic PageRank is a nonnegative left eigenvector y of G associated to the dominating eigenvalue 1 normalized so that $\|y\|_1 = y^T e = 1$, e being the vector of all ones (see e.g. [35, 27]). Since the matrix G is nonnegative and has row sum equal to 1 it is clear that a (canonical) right eigenvector related to 1 is e and that all the other eigenvalues are in modulus at most equal to 1.

As a consequence, the good news is that a solution always exists; the bad news is that there might be multiple independent nonnegative solutions. And even if there is a unique solution, computing it by standard methods such as the power method [21] may fail, because G has one or more eigenvalues different from 1 that have modulus 1, see [22].

In fact, the structure of G is such that we have no guarantee for its aperiodicity and for its irreducibility: therefore the gap between 1 and the modulus of the second largest eigenvalue can be zero, see [22]. This means that the computation of the PageRank by the application of the standard power method (see e.g. [21]) to the matrix G^T (or one of its variations for our specific problem) is not convergent in general. A solution is found by considering a change in the model: given a value $c \in [0, 1)$, from the basic Google matrix G we define the parametric Google matrix $G(c)$ as $cG + (1 - c)ev^T$ with $v[i] \geq 0$, $\|v\|_1 = 1$. This change corresponds to the following user behavior: if the user is visiting page i, then the next move will be with probability c according to the rule described by the basic Google matrix G and with probability $1 - c$ according to the rule described by v. We notice that this change is again meaningful in terms of the surfing model, but there is no clear interpretation in terms of notion of importance. Generally a value of c as 0.85 is considered in the literature (see e.g. [27]). For $c \ll 1$, the good news is that the $y(c)$, i.e., the left dominating nonnegative eigenvector solution of (1.2) with $G = G(c)$, is unique and can be computed in a fast way since $G(c)$ has second eigenvalue whose modulus is dominated by c (see [31] and references therein): therefore the convergence to $y(c)$ is such that the error at step k decays in the generic case as c^k. Of course the computation becomes slow if c is chosen close to 1 and there is no guarantee of convergence if $c = 1$.

In this paper we have four main expository and research goals.

The first concerns a discussion on the model and on its adherence to the reality: a basic example presented in Section 2 is used to point out pathologies and limitations of the actual model and to propose some possible improvements to the model.

Second we would like to understand the characteristics of the matrix $G(c)$ as a function of the parameter c (by completing the analysis in [37, 8]): we are interested in the eigenvalues and in the eigenvector structure, so that the analysis of canonical forms (Jordan, Schur etc.) is of prominent interest.

Third we would like to understand the behavior (regularity, expansions, limits, conditioning etc.) of the PageRank vector $y(c)$ as a function of c also for c close or equal to 1, and fourth we are interested in using the analytical characterization of $y(c)$ for computational purposes. In particular, it is known that for $c = 1$ the problem is ill-posed since there exist infinitely many solutions, forming a convex set, satisfying relations (1.2). On the other hand, for $c \in [0, 1)$, the solution exists and is unique, but the known algorithms become very slow when c is close to 1. Our interest is to compute $y(c)$ in these difficult cases, especially in the limit as c tends to 1.

The philosophical message that can be extracted from the latter three points is as follows: it has been said that the "PageRank problem is closely related to Markov chains" [13, p. 553]; however, framing the PageRank problem in the general setting of standard matrix-analytic properties of complex matrices can liberate one's imagination and stimulate novel approaches that might not be considered in the context of Markov chains.

More in concrete terms, our results are the following:

1. the eigenvalues of $G(c)$, $\forall c \in \mathbb{C}$;
2. the canonical forms of $G(c)$, $\forall c \in \mathbb{C}$ such that $|c| < 1$ (in fact the condition $|c| < 1$ can be replaced by the less restrictive (*): $\forall c \in \mathbb{C}$ such that $c\lambda_j \neq 1$, $j = 2, \ldots, n$, $\lambda_1 = 1, \lambda_2, \ldots, \lambda_n$ being the eigenvalues of $G = G(1)$);
3. a rational expansion for $y(c)$, $\forall c \in \mathbb{C} : |c| < 1$ (in fact only (*) is required);
4. for $c = 1$ the problem (1.2) is ill-posed, but
$$\lim_{\substack{c \to 1 \\ (*) \text{ holds}}} y(c) = \lim_{\substack{c \to 1^- \\ c \in \mathbb{R}}} y(c) = \tilde{y}$$
and $\tilde{y}$ is a solution of (1.2);
5. for this special solution $\tilde{y}$ we show that it coincides with Nv where v is the personalization vector and N^T is a nonnegative projector coinciding with the Cesaro mean $\lim_{r\to\infty} \frac{1}{r+1} \sum_{j=0}^{r} G^j(1)$: the result is due to Lasserre, who calls N the ergodic projector, in a context of probability theory [30], and is known in the field of Web searching engines thanks to [9];
6. if we set $y(1) = \tilde{y}$, then $y(c)$ is analytic in a proper neighborhood of 1 and its sensitivity $\kappa(y(c), \delta) = \frac{\|y(\tilde{c}) - y(c)\|}{\|(y(c)\|}$, $\tilde{c} = c(1 + \delta)$, δ complex parameter of small modulus, is defined by the quantity κ_c where $\kappa(y(c), \delta) = \kappa_c |c\delta|(1 + O(\delta))/\|y(c)\|$ with respect to a generic induced norm $\|\cdot\|$; moreover, in a proper neighborhood of $c = \mu^{-1}$, μ belonging to the spectrum of $G = G(1)$ and up to a function independent of c, the factor κ_c grows generically as
$$\max_{\mu \neq 1, \mu \in \Sigma(G(1)), n(\mu) \in S(\mu)} F(\mu, c), \tag{1.3}$$
$$F(\mu, c) = \left| z_1 (1 - c\mu)^{-n(\mu)} + z_2 (1 - c\mu)^{-n(\mu)-1} (1 - c) \right| |c|^{n(\mu)-2}, \tag{1.4}$$
with $z_1 = (n(\mu) - 1)(1 - c) - c$, $z_2 = c\mu n(\mu)$, $\Sigma(W)$ denoting the spectrum of a given square matrix W and $S(\mu)$ denoting the set of all possible sizes of the Jordan blocks related to the eigenvalue μ.
7. numerical procedures of extrapolation type, based on the third item, for the computation of $y(c)$, when c is close or equal to 1 (i.e. the limit Cesaro vector $\tilde{y}$ of the fourth item and fifth item).

The results in 1) follows from Brauer's Theorem. We discuss the proof and we provide a short historical account both in Section 4 and 9. Findings 2)-7) are obtained in the more general setting of a special type of rank one perturbations. More specifically, instead of $G(c)$ we consider $A(c) := cA + (1-c)\lambda x v^*$ where $Ax = \lambda x$, $v \in \mathbb{C}^n$, $c \in \mathbb{C}$, and $v^*x = 1$. It is clear that our setting is a special instance of the latter: however it is important to stress that this generality allows to clarify and even to simplify the mathematical reasoning and the proofs of the results. For instance, the existence of the limit as c tends to 1 via any complex path requires that the eigenvalue λ is semisimple and this is the case in the Google setting; it is worthwhile observing that such an assumption is essential as the following two examples show.

Example 1. Consider

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix},\ \lambda = 1,\ x = y = e_1,\ v^* = \begin{bmatrix} 1 & \alpha & \beta \end{bmatrix}$$

and the vector $y(c)$ defined by

$$y(c)^* = \begin{bmatrix} 1 & \alpha & \frac{(c-1)\beta}{2c-1} \end{bmatrix}.$$

$y(c)^*$ is the normalized left eigenvector of

$$cA + (1-c)\lambda x v^* = \begin{bmatrix} \lambda + c(1-\lambda) & \lambda(1-c)\alpha & \lambda(1-c)\beta \\ 0 & c & 0 \\ 0 & 0 & 2c \end{bmatrix}$$

associated with the eigenvalue $\lambda = 1$. Moreover,

$$\lim_{c\to 1} y(c)^* = \begin{bmatrix} 1 & \alpha & 0 \end{bmatrix}.$$

Although $\lambda = 1$ is not a simple eigenvalue of A, it is semisimple.

Example 2. Consider

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix},\ \lambda = 1,\ x = y = e_1,\ v^* = \begin{bmatrix} 1 & \alpha & \beta \end{bmatrix}$$

and the vector $y(c)$ defined by

$$y(c)^* = \begin{bmatrix} 1 & \alpha & \beta + \frac{c\alpha}{1-c} \end{bmatrix}.$$

$y(c)^*$ is the normalized left eigenvector of

$$cA + (1-c)\lambda x v^* = \begin{bmatrix} \lambda + c(1-\lambda) & \lambda(1-c)\alpha & \lambda(1-c)\beta \\ 0 & c & c \\ 0 & 0 & c \end{bmatrix}$$

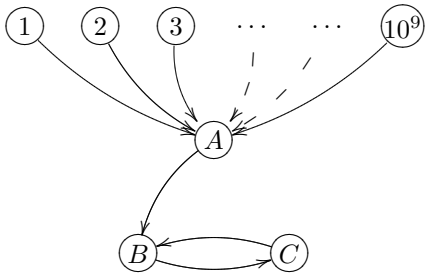
associated with the eigenvalue $\lambda = 1$. However, $\lim_{c\to 1} y(c)^*$ does not exist unless $\alpha = 0$. In this case, $\lambda = 1$ is not semisimple.

Here we report some further comments.

- The results in items 3)-4) were obtained in [37] with the constraint that $c \in [0, 1)$. Moreover, item 4) shows that the parameter c acts like a regularization parameter that stabilizes problem (for a general treatment of regularization techniques see [19]): the nice thing is that, as c tends to 1, we obtain a limit vector $\tilde{y}$, one of the solutions of the original problem.
- The above analysis can be attacked also by transforming the PageRank problem into an equivalent linear system formulation (see e.g. [31]). In that case we write $y(c) = (1 - c)(I - cG^T)^{-1}v$ and indeed this is a compact version of formula (7.6), used in Section 7 for a detailed analysis of $y(c)$ and of its conditioning as a function of the parameter c.
- The algorithms that we propose are new and are partly based on specialized extrapolation procedures discussed in [15, 13]; moreover, we should benefit from the choice of a complex parameter thanks to items 2)-4) with respect to [14], especially in terms of stability.
- Another important issue is how to interpret the computed vector $y(c)$ when c is equal or close to 1 and how to distinguish it from the infinitely many solutions existing when $c = 1$.

The paper is organized as follows. Section 2 is devoted to a critical view of the basic Google model and to propose improvements that prevent pathologic behaviors. In Section 3, we set notation and terminology for the basic matrix-theoretic concepts that we employ to analyze a generalization of $G(c)$: for a square complex matrix A, nonzero complex vectors x and v such that $Ax = \lambda x$ and $v^*x = 1$, and a complex variable c, we study $A(c) = cA + (1 - c)\lambda xv^*$. In Section 4 we explain how Alfred Brauer used the *classical principle of biorthogonality* in 1952 to prove a theorem that reveals the eigenvalues of $A(c)$. In Section 5 we introduce the *complete principle of biorthogonality* and use it to obtain the Jordan blocks of $A(c)$ under the assumption that there is a nonzero vector y such that $y^*A = \lambda y^*$ and $y^*x = 1$. In particular, such a vector y exists if λ is a simple or semisimple eigenvalue of A. In Section 6 we derive a representation for a distinguished left λ-eigenvector $y(c)$ of $A(c)$; this representation is an explicit rational vector-valued function of the complex variable c. In Subsection 6.1 we study $\lim_{c\to 1} y(c)$, and in Subsection 6.2 we focus on the special case in which A is the basic Google matrix G, $\lambda = 1$, $x = e$, and v is a nonnegative probability vector; in this respect Section 7 is devoted to a comparison with the explicit formulae of $y(c)$ and of the Jordan form in [37], to a detailed analysis of the conditioning of $y(c)$ also for $c = 1$, and of the eigenvector structure of $G(c)$. In Section 8, we propose few algorithmic ideas for computing PageRank. The first exploits properties of $G(c)$ as a function of the complex variable c, especially in the unit open disk and in a proper disk centered at $c = 1$, while the second is based on a proper shift of the matrix G. Furthermore some remarks on the interpretation of the limit vector $\tilde{y}$ are given. Section 9 mentions some prior work and Section 10 is devoted to final conclusions and future work.

2. Comments and proposals on the model. Let us start our discussion on the model, by considering in detail the following (extreme) example:



According to the classical algorithm in the ideal case (i.e. for $c = 1$, $y(1) = \tilde{y}$) the page A has zero PageRank (as the 10^9 pages in the first row) and the importance is concentrated in B and C. In some sense the obtained ranking is against common sense since, given the topology of the graph, it is clear that page A should have a significant PageRank measure. This and other related pathologies need further investigations and this is the goal of the rest of this section in which we will suggest a revision of the PageRank model.

2.1. A monotonicity principle and the transient behavior. As already mentioned, according to the classical algorithm in the "ideal case", page A has zero ranking (as the 10^9 pages in the first row) and the importance is concentrated in B and C. This ranking is highly counter-intuitive and indeed wrong: if you are a leader of 10^9 people, you are really powerful no matter if any of your followers has low ranking, i.e., he/she is not important ....

Now suppose that C is deleted, i.e., the considered Web page is deleted: this can be also interpreted as merging B and C in a unique node that we can still call B. Then the rankings of any of 10^9 pages in the first row will move slightly from 0 to $1/(3(10^9 + 1))$; on the other hand the ranking of B goes down dramatically from $1/2$ to $1/3 + 1/(3(10^9 + 1))$ and A becomes really a leader moving from 0 to $1/3$. Again this sharp modification of the ranking is highly counter-intuitive and indeed wrong. At least, one would expect that the cumulative ranking of B and C equal to 1, before deletion of C, and the ranking of B alone after deletion of C should remain roughly speaking equal: a sort of monotonicity which is substantially violated by the actual model, which, on the other hand, induces an unmotivated discontinuous behavior in the solutions. In this respect, the original reason of such a pathology relies on the opposite extremal behavior of functions $\deg(\cdot)$ and $\deg^*(\cdot)$ in which a zero row is replaced by e^T/n. We notice that in the literature a wider idea has been considered by replacing e^T/n by any stochastic vector w^T: however the discontinuity in the model still remains and in the following different solutions to the problem are considered.

A strong and macroscopic evidence of the problems in the actual model is that for most of the nodes in real Web examples (what is called "core" in the literature) the ranking is zero. Indeed, only the use of values of the teleportation parameter c far away from 1 (e.g. 0.85) partly alleviates the problem (in Latin "ex falso quod libet" ..., is an expression capturing the fact that from something wrong anything can derive and, by coincidence, also good things...). In actuality, a basic error is the confusion between the notion of "importance" (PageRanking) and the stochastic model for surfing on the web. We can identify two critical points.

We have a somehow unnatural (wrong) treatment of dangling nodes: with the actual model, there is no monotonicity as the example of deletion of node C in the

above graph shows. In a new model, the management of dangling nodes should be changed for insuring a sort of monotonicity.

The other substantial problem is that the transient effects are not taken into account. A user is on the Web for a finite number of clicks, at every visit. This implies that looking at the stationary vector, as the number of clicks tends to infinity, is just theoretical and far from reality. A new model has to incorporate the transient behavior (see also the functional approach in [1]), which would give the right importance to a node as A in our example.

2.2. New proposals. Following Del Corso, Gullí and Romani [17], one objective of the section is a precise policy for providing to any dangling node a link to itself or to itself and its parents, with a given distribution (in order to impose monotonicity, at least in a weak sense). A link to itself models a reload action and hence it could be also used for the other non-dangling nodes.

A second and more relevant objective is to incorporate the transient behavior for differentiating our ranking from the limit solution to the surfing model. This will be done:

- by eliminating the teleportation parameter that induces a confusion between the notion of ranking and the surfing model,
- by introducing a weighting by experience for reinforcing the role of the transient phenomena,
- by using the Cesaro mean for avoiding oscillatory phenomena at the limit,
- by defining a nonlinear model, in the spirit of a dynamical system, for using the computed PageRank at time t in order to update the PageRank at time $t + \Delta t$.

2.2.1. A new policy for dangling nodes. Now we describe a way for implementing weighted self-loops and weighted links to parents.

Let $\mathrm{in}(A)$ be the set of ingoing edges to node A including possibly the node A itself, let $\deg^-(A)$ be its cardinality (this number could be theoretically zero), $\mathrm{out}(A)$ be the set of outgoing edges from node A including possibly the node A itself, and let $\deg(A)$ be its cardinality (this number could be theoretically zero and in this case, as already observed, node A is a dangling node). In the following $v[\mathrm{in}(A)]$ is the sum of $v[B]$ for B such that the edge from B to A exists, $v[\mathrm{out}(A)]$ is the sum of $v[B]$ for B such that the edge from A to B exists, v being as usual the personalization vector, and g is a positive damping parameter. A reasonable choice is g such that $g/(1+g) = 1/10$, i.e., $g = 1/9$ (ref. Fig. 2.1).

The resulting policy will be the following (any form $0/0$ is set to zero).

Case 1 If $\deg^-(A) = \deg(A) = 0$ then there will be a unique edge from A to A (a loop) with weight of the node equal to 1; in case $v[A] = 0$ the node is simply deleted with its edges.

Case 2 If $\deg(A) = 0$ and $\deg^-(A) > 0$ then there will an edge from A to A with weight

$$v[A]/(v[\mathrm{in}(A)]/\deg^-(A) + v[A])$$

and $\deg^-(A)$ edges from A to B, with B belonging to the set of ingoing nodes of A and with weight

$$v[B]/(v[\mathrm{in}(A)] + \deg^-(A)v[A]);$$

in case $v[\mathrm{in}(A)] + v[A] = 0$ the node is simply deleted with its edges.

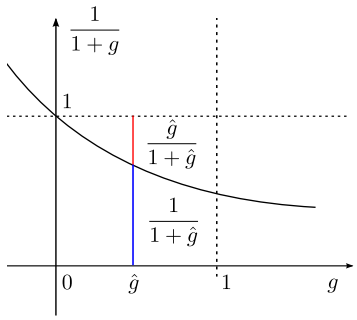


FIG. 2.1.

Case 3 If $\deg(A) > 0$ and $\deg^-(A) = 0$, then there will an edge from A to A with weight

$$gv[A]/(v[\text{out}(A)]/\deg(A) + gv[A])$$

and $\deg(A)$ edges from A to B, with B belonging to the set of outgoing nodes of A and with weight

$$v[B]/(v[\text{out}(A)] + g\deg(A)v[A]);$$

in case $v[\text{out}(A)] + v[A] = 0$ the node is simply deleted with its edges.

Case 4 Otherwise, there will be an edge from A to A with weight

$$gv[A]\deg^-(A)/\{(1+g)(v[\text{in}(A)] + \deg^-(A)v[A])\},$$

$\deg(A)$ edges from A to B, with B belonging to the set of outgoing nodes of A and with weight

$$v[B]/\{(1+g)v[\text{out}(A)]\},$$

and $\deg^-(A)$ edges from A to B, with B belonging to the set of ingoing nodes of A and with weight

$$gv[B]/\{(1+g)(v[\text{in}(A)] + \deg^-(A)v[A])\}.$$

in case $v[\text{in}(A)] + v[\text{out}(A)] + v[A] = 0$ the node is simply deleted with its edges.

To have an idea in a concrete but exemplified case, by setting $g = 1/9$, i.e., $g/(1+g) = 1/10$ and by supposing the personalization vector v uniform, i.e., with all entries equal to $1/n$, the described policy amounts to the following scheme: If $\deg^-(A) = \deg(A) = 0$ then there will be a unique edge from A to A (a loop) with weight of the node equal to 1; If $\deg(A) = 0$ and $\deg^-(A) > 0$ then there will an edge from A to A with weight $1/2$ and $\deg^-(A)$ edges from A to B, with B belonging to the set of ingoing nodes of A and with weight $1/(2\deg^-(A))$; If $\deg(A) > 0$ and $\deg^-(A) = 0$, then there will an edge from A to A with weight $1/10$ and $\deg(A)$ edges from A to B, with B belonging to the set of outgoing nodes of A and with weight $9/(10\deg(A))$; Otherwise, there will be an edge from A to A with weight $1/20$,

$\deg(A)$ edges from A to B, with B belonging to the set of outgoing nodes of A and with weight $9/(10\deg(A))$, and $\deg^-(A)$ edges from A to B, with B belonging to the set of ingoing nodes of A and with weight $1/(20\deg^-(A))$.

We observe that is it possible to apply this new policy in the construction of the matrix G of the PageRank model. In this way we get a matrix G that can be reduced into the direct sum of irreducible and primitive blocks and the Perron Frobenius theory ensures us that there exists always a unique ranking, function of the personalization vector v. It is possible to obtain this solution applying the power method to the matrix $G(1) = G$. At this point the problem related to the presence of dangling nodes is solved, but transient phenomena are not yet taken into account. A further modification of the PageRank model must be considered.

2.2.2. Introducing transient phenomena. An improved ranking may be computed starting from the uniform vector u with components all equal to $w[0]/n$ (with n being the size of the Web) and adding all the vectors $w[j]P^j e/n$, where $P = G^T$, $G = G(1)$, is the transpose of the Google matrix with parameter $c = 1$, j ranges from 0 to a reasonable number of clicks, $w[j] > 0$ is the jth term of a sequence forming a convergent series. Here the Google matrix is that of the old model with $c = 1$ and classical treatment of dangling nodes: moreover, the present proposal is not limited to choosing a hyperlink at random within a page with uniform distribution; if statistical data are known about actual usage patterns, then that information can be included since any arbitrary distribution u describing the choice of the hyperlink can be considered. Here the speed of the decay of $w[j]$ to zero, as j tends to infinity, can be used for deciding to give more or less importance to the stationary limit distribution (solution to the surfing model) or to the transient behavior. Indeed, if one should choose a page where to put the advertisement of a new product, the user would prefer a page with high transient ranking (transient, i.e., for j moderate e.g. at most 10, 15) because many people will have a chance of looking at it, instead of a page with low transient ranking and high final ranking (final, i.e., as j tends to infinity). In fact no user will wait so much or, if he/she waits on the Web, then he/she will be probably terribly tired and unable to appreciate any commercial suggestion. This can motivate a first concrete proposal of $w[0] = w[1] = \ldots = w[k] = (p-1)/(pk)$, for a reasonably moderate k (e.g. k integer with k in the interval $[7, 20]$), p belonging to $[2, 10]$, and the remaining $w[j]$, $j > k$, such that $w[k+1] + w[k+2] + \ldots = 1/p$. In practice, for j larger of any reasonable number of clicks, dictated e.g. by the "physical resistance" of a generic user, we could set $w[j] = 0$. Furthermore, since the Cesaro sum of the $P^j u$ tends to a stationary distribution (as in the Google model) and this stationary distribution is the limit as the teleportation parameter c tends to 1 of $y(c)$, $y(c)$ being the PageRank, instead of the general condition $w[k+1] + w[k+2] + \ldots = 1/p$ we can safely choose $w[k+1] = 1/p$, $w[m] = 0$ for every m larger than $k+1$ and the classical $y(1)$ instead of $P^{k+1}u$. The choice of $y(1)$ is recommended for stabilizing the computation: indeed the sequence $P^j u$ may fail to converge, while its Cesaro mean converges to the ergodic projector.

A natural problem at this point is: how to manage SPAM pages? An interesting idea used in the previous model is based on a careful choice of the personalization vector v (see below): hence as before, in the previous sum, the uniform vector u is replaced by the personalization vector v.

A second natural problem is the computation of $y(1)$ intended, by definition, as the limit as the teleportation parameter c tends to 1 of $y(c)$ with generic personalization vector v.

In fact from the analysis in [37, 24] we know that $y(c)$ is an analytic function of c on the complex plane, except for a finite number of points different from 1 outside the open unit disk (see Sections 6.2 and 7). Therefore $y(1)$ can be approximated, just by continuity, by $y(c)$ with c close to 1 $(0.9, 0.99)$: there is a lot of work by Golub and Greif (using Arnoldi, see [20]), Del Corso, Gullí, Romani (using the linear system representation and preconditioned GMRES, see [17]), Breziski et al. (vector extrapolation based on explicit rational formulae of $y(c)$, see [13, 14, 15]) etc. for making such computations fast. Otherwise the straightforward but effective algorithm in Subsection 8.2 can be conveniently employed.

An appropriate choice of the involved parameters, based on the experience, is also possible with special reference to k, p and to the weights $w[j]$. Here is a first embodiment: a visit to the page A will make A more important if it is longer: following this principle the value $w[j]$ could be decided as a monotone function of the average time of a generic user between the click j and the click $j+1$ (see below). While the previous model is trying to rank the importance at the limit (the asymptotic stationary distribution, i.e., the solution to the surfing model), the present approach can be seen as a global ranking, i.e., as a weighted integral over the discrete time (decided by clicks on the Web) of the ranking. Of course, as already informally observed, the weights $w[j]$, like in any weighted quadrature formula, decide where to put the attention for giving the final decision on the global ranking.

Another healthy effect of the integral approach is the stabilization of the involved quantities which prevent from spurious oscillations and this stabilization is typical of any Cesaro like process. Indeed, by considering again the example above, with the old model the ranking of page B and C are oscillating. Depending on starting distribution vector, they exchange the first and the second top positions at every j and the difference between their ranking is not negligible. Of course, the use of teleportation just alleviates the phenomenon, which is eliminated at the limit, but in practice it remains well visible. The averaging implied by the integral approach substantially reduces this fact as any Cesaro like process does: however, it should be noticed that a plain Cesaro approach would again give emphasis only on the limit behavior, since its representing matrix would converge to the spectral projector (see again [30, 38]).

Furthermore, let us give more details on a more accurate proposal for the determination of the sequence $w[j]$, based on experience. Consider for a moment to have the following information on all the visits on the Web for a certain window of observation (one week for instance). Let surfing[j] with $j \geq 0$ be a nonnegative integer that represents the number of visits to the Web that last at least j clicks. If you are on the Web and you change Web page not clicking, but by writing explicitly the address, then this is counted as a restart, i.e., in the number surfing[0]. Moreover, there exists only a finite number of indices j with nonzero surfing[j], due to the finiteness of the time interval and due to the physical resistance of the generic user. Now we make a statistic on the lengths t_{j+1}, with $j \geq 0$, of the time intervals between the click number j and the click $j+1$, if the click j is not the last click, or the time intervals between the click number j and the exit, if the click j is the last. Let us denote by $T[j+1]$ the average value of these t_{j+1} based on our observations over all the visits. Then calling $\gamma[j] = \text{surfing[j]} \cdot \text{T[j + 1]}$, $j \geq 0$, and $s[h]$ the sum of all $\gamma[j]$ with $j \geq h$, our integral will be

$$y = F(P, v, w) = w[0]v + w[1]Pv + \cdots + w[k]P^k v + w[k+1]y(1) \qquad (2.1)$$

with $w[j] = \gamma[j]/s[0]$, $j = 0, \ldots, k$, and $w[k+1] = s[k+1]/s[0]$.

In this way more influence is given to $P^s v$ if the "area" $w[s]$ is maximal: $w[j]$ may be viewed as the area of a rectangle where the length of the basis is the average time between click j and click $j+1$ and the length of the height is equal to the value surfing[j] i.e. the number of visits that last at least j clicks. It is not excluded that the behavior of such a sequence $w[j]$ can be roughly approximated by a Poisson distribution with a given mean.

Along the same line the personalization vector v can be described. It should be nonnegative and with unit l^1 norm (just a matter of scaling). Moreover $v[j]$ should be put at zero if j is recognized as a SPAM page and for the other pages the value $v[j]$ has to be proportional to the sum over the visits to j at the first click of the visit-time.

Of course these parameters have to be estimated, but the leaders of Web-Searching Market (as Google, Microsoft, Yahoo etc.) for sure have access to such information.

We can now apply the new policy for nodes, described in the former subsection, in the construction of the matrix $P = G^T$ that appears in the formula (2.1). Thanks to the structure of this new P there are no more spurious oscillations in the terms $P^j v$ for j increasing. So instead of $y(1)$ in (2.1) we can consider safely the term $P^{k+1}v$.

$$y = F(P, v, w) = w[0]v + w[1]Pv + \cdots + w[k+1]P^{k+1}v \tag{2.2}$$

with $w[j]$ evaluated as previously proposed.

The ranking coming out of the joining of these two techniques seems to be exempt from the pathologies of the classical PageRank.

Finally, the latter statement suggests to look at the problem in a time dependent and nonlinear way, since the Web evolves in time and the expected values of the various time intervals, i.e. $T[j]$ for $j = 1, 2, \dots$, also depend on the ranking that we attribute to Web pages. A concrete proposal is the following: if $\hat{y}(t)$ denotes this new definition of the PageRank according to the formula (2.2), then we define the new ranking at $t + \Delta t$ as

$$\hat{y}(t+\Delta t) = F(P(t+\Delta t), z, w(t+\Delta t)), \qquad z = m\hat{y}(t) + (1-m)v(t+\Delta t), \;\; 0 \le m \le 1,$$

where $P(t+\Delta t)$ is the Web matrix at the time $t+\Delta t$, $w(t+\Delta t)$ is the vector of the weights at the time $t+\Delta t$, and where z is defined as a convex combination of $v(t+\Delta t)$ (the personalization vector defined as before at time $t+\Delta t$) and $\hat{y}(t)$ which carries the information on ranking at the older time t. The nonnegative parameters m and $1-m$ of the convex combination can be interpreted as weights that measure the level of fidelity, which is based on the "past importance".

2.2.3. Further possibilities. In summary two goals are achieved by the new model. The actual efficiency (fast computation) is preserved, since the new computation will involve at most two vectors, which already were computed in the preceding model, and it seems that the old pathologies are removed without introducing new ones. The new ranking method according to the proposal may be called the VisibilityRank or the CommercialRank, since a query-independent measure is given of the "fair value" of any Web page for deciding e.g. the cost of putting an advertisement in that page, as in the determination of the cost of renting a space for advertisement in a given place of a given street, square in a given town etc.

As a final remark on this model part, it is worth mentioning that this model could be of interest not only in Web ranking, but also in political/social sciences e.g.

for ranking who/what is influential and who/what is not (as an example one could be interested in answering to the following questions: Bill Clinton's opinion is really influential and at which level? How to rank immaterial forces such as a religious authority vs material forces such as economic/military powers?), in many aspects of marketing, for ranking human resources, for ranking the importance of a paper and/or of a researcher looking in scientific databases, see [4]. Let us think to MathSciNet for Mathematicians where a generic node is any paper in the database and a link from A to B is just a bibliographic reference to paper B in paper A. For evaluating the impact (i.e. the ranking) of a paper the very same model and the same procedure as described before could be applied to the related graph. For evaluating or ranking a researcher (a very hot topic nowadays in several countries) it would be enough to modify the graph where every single node is a researcher and a link from A to B means that the researcher A has written at least one paper referring to at least one paper of the researcher B: the links have to be weighted and the related weights will be proportional to the number of such papers and will be properly normalized according to the number of authors in the referring papers of A and in the referred papers of B. The algorithm will be again the same and again the same idea would work for ranking researcher groups or Institutions such as Departments, Faculties, Universities (see e.g. the hierarchical approach in European Patent 1, 653, 380 A1). In addition it is worth stressing that the described procedures for defining the graph and for computing the ranking are unchanged in any Scientific homogeneous community.

Of course, for modeling in a convincing way such complex phenomena, it would be recommended to enrich the structure of the graph by adding to nodes and/or to edges more information (meta-graph? ...). However, the essential basic idea for defining and computing the ranking has to remain virtually the same.

3. Terminology and notation. All the matrices and vectors that we consider have real or complex entries. We denote the conjugate transpose of an m-by-n matrix $X = [x_{ij}]$ by $X^* = [\bar{x}_{ji}]$. For $p \in [1, \infty)$ for a vector $w \in \mathbb{C}^n$, its l^p norm is given by $\|w\|_p = \left[\sum_{j=1}^n |w[j]|^p\right]^{1/p}$ while its l^∞ norm is $\|w\|_\infty = \max_{j=1,\ldots,n} |w[j]|$; for a square matrix A and for $p \in [1, \infty]$, $\|A\|_p$ is the associated induced norm. If A is a square matrix, its *characteristic polynomial* is $p_A(t) := \det(tI - A)$; the (complex) zeroes of $p_A(t)$ are the *eigenvalues* of A. A complex number λ is an eigenvalue of A if and only if there are nonzero vectors x and y such that $Ax = \lambda x$ and $y^*A = \lambda y^*$; x is said to be an *eigenvector* (more specifically, a *right eigenvector*) of A associated with λ and y is said to be a *left eigenvector* of A associated with λ. If λ is an eigenvalue of A, its *algebraic multiplicity* is its multiplicity as a zero of $p_A(t)$; its *geometric multiplicity* is the maximum number of linearly independent eigenvectors associated with it. The geometric multiplicity of an eigenvalue is never greater than its algebraic multiplicity. An eigenvalue whose algebraic multiplicity is one is said to be *simple*. An eigenvalue whose algebraic and geometric multiplicities are equal is said to be *semisimple*; an eigenvalue λ of A is semisimple if and only if $\text{rank}(A - \lambda I) = \text{rank}(A - \lambda I)^2$.

We let e_1 indicate the first column of the identity matrix I: $e_1 = [1\ 0\ \ldots\ 0]^T$. We let $e = [1\ 1\ \ldots\ 1]^T$ denote the all-ones vector. Whenever it is useful to indicate that an identity or zero matrix has a specific size, e.g., r-by-r, we write I_r or 0_r.

Two vectors x and y of the same size are *orthogonal* if $x^*y = 0$. The *orthogonal complement* of a given set of vectors is the set (actually, a vector space) of all vectors that are orthogonal to every vector in the given set.

An n-by-r matrix X has *orthonormal* columns if $X^*X = I_r$. A square matrix U

is *unitary* if it has orthonormal columns, that is, if U^* is the inverse of U.

A square matrix A is a *projection* if $A^2 = A$.

A square matrix A is *row-stochastic* if it has real nonnegative entries and $Ae = e$, which means that the sum of the entries in each row is 1; A is *column-stochastic* if A^T is row-stochastic. We say that A is *stochastic* if it is either row-stochastic or column-stochastic.

The *direct sum* of k given square matrices $X_1, \ldots, X_k$ is the block diagonal matrix

$$\begin{bmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_k \end{bmatrix} = X_1 \oplus \cdots \oplus X_k.$$

The k-by-k *Jordan block* with eigenvalue λ is

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}, \quad J_1(\lambda) = [\lambda].$$

Each square complex matrix A is similar to a direct sum of Jordan blocks, which is unique up to permutation of the blocks; this direct sum is the *Jordan canonical form* of A. The algebraic multiplicity of λ as an eigenvalue of $J_k(\lambda)$ is k; its geometric multiplicity is 1. If λ is a semisimple eigenvalue of A with multiplicity m, then the Jordan canonical form of A is $\lambda I_m \oplus J$, in which J is a direct sum of Jordan blocks with eigenvalues different from λ; if λ is a simple eigenvalue, then $m = 1$ and the Jordan canonical form of A is $[\lambda] \oplus J$.

Suppose that a square matrix A is similar to the direct sum of a zero matrix and a nonsingular matrix, that is,

$$A = S \begin{bmatrix} 0_m & 0 \\ 0 & B \end{bmatrix} S^{-1}, \ B \text{ is nonsingular.} \tag{3.1}$$

The matrix

$$A^D = S \begin{bmatrix} 0_m & 0 \\ 0 & B^{-1} \end{bmatrix} S^{-1}$$

is called the *Drazin inverse* of A; it does not depend on the choice of S or B in the representation (3.1). [16, Chapter 7] Moreover, both $AA^D = A^D A$ and $I - AA^D$ are projections. If X and Y have m columns, $S = [X \ S_2]$, and $(S^{-1})^* = [Y \ Z_2]$, then $A^D = S_2 B^{-1} Z_2^*$ and $I - AA^D = XY^*$.

In a block matrix, the symbol $\bigstar$ denotes a block whose entries are not required to take particular values. Finally we consider $A^0 = I$.For a systematic discussion of a broad range of matrix analysis issues, see [23].

4. Basic principle of biorthogonality and eigenvalues. The following observation about left and right eigenvectors associated with different eigenvalues is the *basic principle of biorthogonality* [23, Theorem 1.4.7].

LEMMA 4.1. *Let A be a square complex matrix and let x and y be nonzero complex vectors such that $Ax = \lambda x$ and $y^*A = \mu y^*$. If $\lambda \neq \mu$, then $y^*x = 0$ (that is, x and y are orthogonal).*

Proof. Compute y^*Ax in two ways: (i) as $y^*(Ax) = y^*(\lambda x) = \lambda(y^*x)$, and (ii) as $(y^*A)x = (\mu y^*)x = \mu(y^*x)$. Since $\lambda(y^*x) = \mu(y^*x)$ and $\lambda \neq \mu$, it follows that $y^*x = 0$. ∎

For a given vector v and a matrix A with eigenvalue λ and associated eigenvector x, how are the eigenvalues of $A + xv^*$ related to those of A? This question was asked and answered by Alfred Brauer in 1952 [11, Theorem 26]:

THEOREM 4.2 (Brauer). *Let A be an n-by-n complex matrix and let x be a nonzero complex vector such that $Ax = \lambda x$. Let*

$$\lambda, \lambda_2, \ldots, \lambda_n$$

be the eigenvalues of A. Then for any complex n-vector v the eigenvalues of $A + xv^*$ are

$$\lambda + v^*x, \lambda_2, \ldots, \lambda_n.$$

Brauer's proof involved three steps: (i) Compute

$$(A + xv^*)x = Ax + xv^*x = \lambda x + (v^*x)x = (\lambda + v^*x)x,$$

which shows that $\lambda + v^*x$ is an eigenvalue of $A + xv^*$. (ii) If μ is an eigenvalue of A that is different from λ, and if y is a left eigenvector of A associated with μ, then Lemma 4.1 ensures that

$$y^*(A + xv^*) = y^*A + y^*xv^* = \mu y^* + (y^*x)v = \mu y^* + 0 \cdot v = \mu y^*.$$

Thus, the distinct eigenvalues of A that are different from λ are all eigenvalues of $A+xv^*$, but perhaps not with the same multiplicities. (iii) Brauer completed his proof with a continuity argument to show that the multiplicities of the common eigenvalues of A and $A + xv^*$ (setting aside the respective eigenvalues λ and $\lambda + v^*x$) are the same.

Brauer's theorem tells us something interesting about the eigenvalues of $A(c)$.

COROLLARY 4.3. *Let A be an n-by-n complex matrix. Let λ be an eigenvalue of A, let x and v be nonzero complex vectors such that $Ax = \lambda x$ and $v^*x = 1$, and let $A(c) = cA + (1 - c)\lambda xv^*$. Let*

$$\lambda, \lambda_2, \ldots, \lambda_n$$

be the eigenvalues of A. Then for any complex number c, the eigenvalues of $A(c)$ are

$$\lambda, c\lambda_2, \ldots, c\lambda_n.$$

Proof. In the statement of Brauer's Theorem, replace A and v by cA and $(1-\bar{c})\bar{\lambda}v$, respectively. The eigenvalues of cA are $c\lambda, c\lambda_2, \ldots, c\lambda_n$, x is an eigenvector of cA associated with the eigenvalue $c\lambda$, and Brauer's Theorem tells us that the eigenvalues of $cA + x((1 - \bar{c})\bar{\lambda}v)^* = cA + (1 - c)\lambda xv^*$ are $c\lambda + (1 - c)\lambda v^*x, c\lambda_2, \ldots, c\lambda_n$, which are $\lambda, c\lambda_2, \ldots, c\lambda_n$ since $v^*x = 1$. ∎

Robert Reams [36, p. 368] revisited Brauer's theorem in 1996. He observed that the Schur triangularization theorem [23, Theorem 2.3.1] can be used to prove Brauer's

Theorem without a continuity argument: Let $S = [x\ S_1]$ be any nonsingular matrix that upper triangularizes A as

$$S^{-1}AS = \begin{bmatrix} \lambda & \bigstar & \cdots & \bigstar \\ & \lambda_2 & \ddots & \vdots \\ & & \ddots & \bigstar \\ 0 & & & \lambda_n \end{bmatrix}$$

and whose first column is an eigenvector x associated with the eigenvalue λ. Since $I = S^{-1}S = [S^{-1}x\ \ \bigstar]$, we see that $S^{-1}x = e_1$. Compute

$$S^{-1}(xv^*)S = \left(S^{-1}x\right)(v^*S) = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} v^*x & \bigstar & \cdots & \bigstar \end{bmatrix}$$
$$= \begin{bmatrix} v^*x & \bigstar & \cdots & \bigstar \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Therefore, the similarity

$$S^{-1}(A + xv^*)S = \begin{bmatrix} \lambda + v^*x & \bigstar & \cdots & \bigstar \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \bigstar \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix},$$

reveals both the eigenvalues of $A + xv^*$ and their multiplicities.

A new alternative proof that the eigenvalues of $A(c)$ are $\lambda, c\lambda_2, \ldots, c\lambda_n$, only based on polynomial identities, is proposed below.

For any n-by-k complex matrices Z and W with $n \geq k$, the n eigenvalues of ZW^* are the k eigenvalues of W^*Z together with $n-k$ zero eigenvalues [23, Theorem 1.3.20]. In particular, for any vectors $z, w \in \mathbb{C}^n$ the n eigenvalues of zw^* are $w^*z, 0, \ldots, 0$, so the n eigenvalues of $I+zw^*$ are $1+w^*z, 1, \ldots, 1$. It follows that $\det(I+zw^*) = 1+w^*z$.

Since $(tI - cA)x = (t - c\lambda)x$, we have $(tI - cA)^{-1}x = (t - c\lambda)^{-1}x$ for any $t \neq c\lambda$. For any $z \in \mathbb{C}^n$ and for $t \neq c\lambda$, compute

$$\begin{aligned} p_{cA+xz^*}(t) &= \det(tI - (cA + xz^*)) \\ &= \det((tI - cA) - xz^*) \\ &= \det(tI - cA)\det(I - (tI - cA)^{-1}xz^*) \\ &= p_{cA}(t)\det(I - (t - c\lambda)^{-1}xz^*) \\ &= p_{cA}(t)\left(1 - (t - c\lambda)^{-1}z^*x\right) \\ &= \frac{p_{cA}(t)\,(t - c\lambda - z^*x)}{t - c\lambda}. \end{aligned}$$

Thus, for any $z \in \mathbb{C}^n$ we have the polynomial identity

$$(t - c\lambda)p_{cA+xz^*}(t) = (t - (c\lambda + z^*x))p_{cA}(t), \tag{4.1}$$

where it is again legal to have $t = c\lambda$ by continuity arguments. The zeroes of the left-hand side are $c\lambda$ together with the eigenvalues of $cA+xz^*$; the zeroes of the right-hand side are $c\lambda + z^*x, c\lambda, c\lambda_2, \ldots, c\lambda_n$. It follows that the eigenvalues of $cA + xz^*$ are $c\lambda + z^*x, c\lambda_2, \ldots, c\lambda_n$.

Now set $z = (1 - \overline{c})\overline{\lambda}v$, use the condition $v^*x = 1$, and conclude that the eigenvalues of $A(c)$ are $\lambda, c\lambda_2, \ldots, c\lambda_n$ for any $c \in \mathbb{C}$.

Finally, it is worth mentioning a two-lines proof of Brauer's theorem due to Iannazzo [25] which could be considered a special case of a proof trick used in the functional formulation of the shift [6][Section 3.2], in a structured Markov chains context. Based on the matrix-polynomial identity and $Axv^* = \lambda xv^*$

$$(A + xv^* - \mu I)(\mu - \lambda)I = (A - \mu I)((\mu - \lambda)I - xv^*),$$

by taking the determinant of both members and using the formula for the characteristic polynomial of a dyad, it holds that

$$\begin{aligned} p_{A+xv^*}(\mu)(\mu - \lambda)^n &= (-1)^n p_A(\mu) p_{x^*v}(\mu - \lambda) \\ &= (-1)^n p_A(\mu)(\mu - \lambda)^{n-1}(\mu - \lambda - v^*x). \end{aligned}$$

The unique factorization theorem for polynomials achieves the proof.

It is worthwhile to remark that the interest of Iannazzo for Brauer's theorem does not come from the Google matrix, but from fast Markov chains computations, Riccati matrix equations etc. See [5] and references reported therein.

5. Complete principle of biorthogonality and Jordan blocks. Brauer used the basic principle of biorthogonality to analyze the *eigenvalues* of $A + xv^*$. We now want to analyze the *Jordan blocks* of $A + xv^*$.

The basic principle of biorthogonality is silent about what happens when $\lambda = \mu$. In that event, there are three possibilities: (i) $y^*x = 0$ (we can normalize so that $x^*x = y^*y = 1$); (ii) $y^*x \neq 0$ (we can normalize so that $y^*x = 1$); or (iii) $x = \alpha y$ (we can normalize so that $x = y$ and $x^*x = 1$). The following *complete principle of biorthogonality* addresses all the possibilities and describes reduced forms for A that can be achieved in each case.

THEOREM 5.1. *Let A be an n-by-n complex matrix and let x and y be nonzero complex vectors such that $Ax = \lambda x$ and $y^*A = \mu y^*$.*

(a) Suppose that $\lambda \neq \mu$ and $x^*x = y^*y = 1$. Then $y^*x = 0$. Let $U = [x\ y\ U_1]$, in which the columns of U_1 are any given orthonormal basis for the orthogonal complement of x and y. Then U is unitary and

$$U^*AU = \begin{bmatrix} \lambda & \bigstar & \bigstar \\ 0 & \mu & 0 \\ 0 & \bigstar & B \end{bmatrix}, \quad B = U_1^*AU_1 \text{ is } (n-2)\text{-by-}(n-2).$$

(b) Suppose that $\lambda = \mu$, $y^*x = 0$, and $x^*x = y^*y = 1$. Let $U = [x\ y\ U_1]$, in which the columns of U_1 are any given orthonormal basis for the orthogonal complement of x and y. Then U is unitary, the algebraic multiplicity of λ is at least two, and

$$U^*AU = \begin{bmatrix} \lambda & \bigstar & \bigstar \\ 0 & \lambda & 0 \\ 0 & \bigstar & B \end{bmatrix}, \quad B = U_1^*AU_1 \text{ is } (n-2)\text{-by-}(n-2).$$

(c) Suppose that $\lambda = \mu$ and $y^*x = 1$. Let $S = [x\ S_1]$, in which the columns of S_1 are any given basis for the orthogonal complement of y. Then S is nonsingular, $(S^{-1})^* = [y\ Z_1]$, the columns of Z_1 are a basis for the orthogonal complement of x, and

$$S^{-1}AS = \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix}, \quad B = Z_1^*AS_1 \text{ is } (n-1)\text{-by-}(n-1). \tag{5.1}$$

(d) Suppose that $\lambda = \mu$, $x = y$, and $x^*x = 1$. Let $U = [x\ U_1]$, in which the columns of U_1 are any given orthonormal basis for the orthogonal complement of x. Then U is unitary and

$$U^*AU = \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix}, \quad B = U_1^*AU_1 \text{ is } (n-1)\text{-by-}(n-1). \tag{5.2}$$

Proof. (a) Lemma 4.1 ensures that x and y are orthogonal. Let $U = [x\ y\ U_1]$, in which the columns of U_1 are a given orthonormal basis for the orthogonal complement of x and y. The n columns of U are an orthonormal set, so U is unitary. Compute the unitary similarity

$$\begin{aligned} U^*AU &= \begin{bmatrix} x^* \\ y^* \\ U_1^* \end{bmatrix} A[x\ y\ U_1] = \begin{bmatrix} x^*Ax & x^*Ay & x^*AU_1 \\ y^*Ax & y^*Ay & y^*AU_1 \\ U_1^*Ax & U_1^*Ay & U_1^*AU_1 \end{bmatrix} \\ &= \begin{bmatrix} \lambda x^*x & x^*Ay & x^*AU_1 \\ \lambda y^*x & \mu y^*y & \mu y^*U_1 \\ \lambda U_1^*x & U_1^*Ay & U_1^*AU_1 \end{bmatrix} = \begin{bmatrix} \lambda & \bigstar & \bigstar \\ 0 & \mu & 0 \\ 0 & \bigstar & U_1^*AU_1 \end{bmatrix}. \end{aligned}$$

(b) As in (a), construct a unitary matrix $U = [x\ y\ U_1]$, in which the columns of U_1 are a given orthonormal basis for the orthogonal complement of x and y. The reduced form of A under unitary similarity via U is the same as in (a), but with $\lambda = \mu$. The characteristic polynomial of A is

$$p_A(t) = \det(tI - A) = \det \begin{bmatrix} t-\lambda & \bigstar & \bigstar \\ 0 & t-\lambda & 0 \\ 0 & \bigstar & tI - B \end{bmatrix}$$

A Laplace expansion by minors down the first column gives

$$p_A(t) = (t-\lambda) \det \begin{bmatrix} t-\lambda & 0 \\ \bigstar & tI - B \end{bmatrix}.$$

Finally, a Laplace expansion by minors across the first row gives

$$p_A(t) = (t-\lambda)^2 \det(tI - B) = (t-\lambda)^2 p_B(t),$$

so λ is a zero of $p_A(t)$ with multiplicity at least two.

(c) Let the columns of S_1 be a given basis for the orthogonal complement of y and let $S = [x\ S_1]$. The columns of S_1 are linearly independent, so S is singular only if x is a linear combination of the columns of S_1, that is, only if $x = S_1\xi$ for some vector ξ.

But then $1 = y^*x = y^*S_1\xi = 0\xi = 0$. This contradiction shows that S is nonsingular. Partition $(S^{-1})^* = [\eta\ Z_1]$ and compute

$$I = S^{-1}S = \begin{bmatrix} \eta^* \\ Z_1^* \end{bmatrix} \begin{bmatrix} x & S_1 \end{bmatrix} = \begin{bmatrix} \eta^*x & \eta^*S_1 \\ Z_1^*x & Z_1^*S_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & I_{n-1} \end{bmatrix}. \tag{5.3}$$

Thus, the $n-1$ columns of Z_1, necessarily linearly independent, are orthogonal to x, so they are a basis for the orthogonal complement of x. Also, $\eta^*S_1 = 0$ means that η is orthogonal to the orthogonal complement of y, so $\eta = \alpha y$. But $1 = \eta^*x = (\alpha y)^*x = \bar{\alpha}y^*x = \bar{\alpha}$, so $\alpha = 1$ and $\eta = y$. Finally, compute the similarity

$$\begin{aligned} S^{-1}AS &= \begin{bmatrix} y^* \\ Z_1^* \end{bmatrix} A \begin{bmatrix} x & S_1 \end{bmatrix} = \begin{bmatrix} y^*Ax & y^*AS_1 \\ Z_1^*Ax & Z_1^*AS_1 \end{bmatrix} \\ &= \begin{bmatrix} \lambda y^*x & \lambda y^*S_1 \\ \lambda Z_1^*x & Z_1^*AS_1 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & Z_1^*AS_1 \end{bmatrix}. \end{aligned}$$

(d) Let the columns of U_1 be a given orthonormal basis for the orthogonal complement of x. Then the n columns of $U = [x\ U_1]$ are an orthonormal set, so U is unitary. Compute the unitary similarity

$$\begin{aligned} U^*AU &= \begin{bmatrix} x^* \\ U_1^* \end{bmatrix} A \begin{bmatrix} x & U_1 \end{bmatrix} = \begin{bmatrix} x^*Ax & x^*AU_1 \\ U_1^*Ax & U_1^*AU_1 \end{bmatrix} \\ &= \begin{bmatrix} \lambda x^*x & \lambda x^*U_1 \\ \lambda U_1^*x & U_1^*AU_1 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & U_1^*AU_1 \end{bmatrix}. \end{aligned}$$

□

We now use the complete principle of biorthogonality to establish an analog of Brauer's Theorem 4.2 for Jordan blocks.

THEOREM 5.2. *Let A be an n-by-n complex matrix. Let $\lambda, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of A, and let x and y be nonzero complex vectors such that $Ax = \lambda x$ and $y^*A = \lambda y^*$. Assume that $y^*x = 1$. Then the Jordan canonical form of A is*

$$[\lambda] \oplus J_{n_1}(\nu_1) \oplus \cdots \oplus J_{n_k}(\nu_k)$$

for some positive integers $k, n_1, \ldots, n_k$ and some set of eigenvalues $\{\nu_1, \ldots, \nu_k\} \subseteq \{\lambda_2, \ldots, \lambda_n\}$. For any complex vector v such that $\lambda + v^*x \neq \lambda_j$ for each $j = 2, \ldots, n$, the Jordan canonical form of $A + xv^*$ is

$$[\lambda + v^*x] \oplus J_{n_1}(\nu_1) \oplus \cdots \oplus J_{n_k}(\nu_k). \tag{5.4}$$

Proof. The hypotheses and Theorem 5.1(c) ensure that

$$S^{-1}AS = \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix} \tag{5.5}$$

for some nonsingular S of the form $S = [x\ S_1]$, so that $S^{-1}x = e_1$. The eigenvalues of B are $\lambda_2, \ldots, \lambda_n$; let

$$J_{n_1}(\nu_1) \oplus \cdots \oplus J_{n_k}(\nu_k)$$

be the Jordan canonical form of B. Just as in Reams' proof of Brauer's Theorem, we have

$$S^{-1}(xv^*)S = \left(S^{-1}x\right)(v^*S) = e_1 \begin{bmatrix} v^*x & v^*S_1 \end{bmatrix} = \begin{bmatrix} v^*x & w^* \\ 0 & 0 \end{bmatrix}, \tag{5.6}$$

in which we set $w^* := v^* S_1$. Combining the similarities (5.5) and (5.6), we see that

$$S^{-1}(A + xv^*)S = \begin{bmatrix} \lambda + v^*x & w^* \\ 0 & B \end{bmatrix}.$$

Now let ξ be any given $(n-1)$-vector, verify that

$$\begin{bmatrix} 1 & \xi^* \\ 0 & I \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -\xi^* \\ 0 & I \end{bmatrix},$$

and compute the similarity

$$\begin{bmatrix} 1 & -\xi^* \\ 0 & I \end{bmatrix} \begin{bmatrix} \lambda + v^*x & w^* \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & \xi^* \\ 0 & I \end{bmatrix} = \begin{bmatrix} \lambda + v^*x & w^* + \xi^*((\lambda + v^*x)I - B) \\ 0 & B \end{bmatrix}.$$

We have assumed that $\lambda + v^*x$ is not an eigenvalue of B, so we may take

$$\xi^* := -w^*((\lambda + v^*x)I - B)^{-1},$$

in which case $w^* + \xi^*((\lambda + v^*x)I - B) = 0$ and $A + xv^*$ is revealed to be similar to

$$\begin{bmatrix} \lambda + v^*x & 0 \\ 0 & B \end{bmatrix}.$$

Thus, the Jordan canonical form of $A + xv^*$ is (5.4): the direct sum of $[\lambda + v^*x]$ and the Jordan canonical form of B. □

The following result strengthens the conclusion of Corollary 4.3 to describe not only the eigenvalues of $A(c)$, but also its Jordan blocks.

COROLLARY 5.3. *Let A be an n-by-n complex matrix. Let $\lambda, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of A; let x, y, and v be nonzero complex vectors such that $Ax = \lambda x$, $y^*A = \lambda y^*$, and $v^*x = 1$; and let $A(c) = cA + (1-c)\lambda xv^*$. Assume that $y^*x = 1$ and integers $k, n_1, \ldots, n_k$ and the set $\{\nu_1, \ldots, \nu_k\}$ are defined as in the previous Theorem. Let the Jordan canonical form of A be*

$$[\lambda] \oplus J_{n_1}(\nu_1) \oplus \cdots \oplus J_{n_k}(\nu_k).$$

Then for any nonzero complex number c such that

$$c\lambda_j \neq \lambda \text{ for each } j = 2, \ldots, n, \tag{5.7}$$

the Jordan canonical form of $A(c)$ is

$$[\lambda] \oplus J_{n_1}(c\nu_1) \oplus \cdots \oplus J_{n_k}(c\nu_k).$$

Proof. We proceed as in the proof of Corollary 4.3. In the statement of Theorem 5.2, replace A and v, respectively, by cA and $(1-\bar{c})\bar{\lambda}v$, respectively. For any c, cA is similar to

$$[c\lambda] \oplus cJ_{n_1}(\nu_1) \oplus \cdots \oplus cJ_{n_k}(\nu_k),$$

but if $c \neq 0$, we can say more: this direct sum is similar to

$$[c\lambda] \oplus J_{n_1}(c\nu_1) \oplus \cdots \oplus J_{n_k}(c\nu_k).$$

Moreover, x is an eigenvector of cA associated with the eigenvalue $c\lambda$, the remaining eigenvalues of cA are $c\lambda_2, \ldots, c\lambda_n$, and

$$c\lambda + ((1-\bar{c})\bar{\lambda}v)^* x = c\lambda + (1-c)\lambda v^* x = c\lambda + (1-c)\lambda = \lambda.$$

Thus, our assumption (5.7) and Theorem 5.2 ensure that the Jordan canonical form of

$$cA + x((1-\bar{c})\bar{\lambda}v)^* = cA + (1-c)\lambda x v^* = A(c)$$

is

$$[\lambda] \oplus J_{n_1}(c\nu_1) \oplus \cdots \oplus J_{n_k}(c\nu_k).$$

□

In the above analysis, often the matrix B is determined only up to similarity. If convenient, we can take B to be a Jordan canonical form, upper triangular, a real Jordan form (if A is real), a Schur canonical form, etc. Perhaps this flexibility can be exploited to achieve a computational advantage.

Finally we stress a pleasant contrast between Corollary 4.3 and Corollary 5.3. In Corollary 4.3 the hypothesis is weaker than that of Corollary 5.3, and of course a weaker conclusion is obtained. However, Corollary 4.3 is of independent interest, since it gives a broader context for the famous eigenvalue properties of the Google matrix perturbation: for instance, similar problems appear and Corollary 4.3 is useful in the context of iterative solvers for algebraic Riccati equation, for accelerating the convergence of cyclic reduction based algorithms (see [7, 5] and references therein and [28] for further applications of mathematical physics).

6. The normalized left λ-eigenvector of $A(c)$. If $\lambda \neq 0$, Corollary 4.3 ensures that it is a simple eigenvalue of $A(c)$ for all but finitely many values of c. We would like to have an explicit expression for its associated left eigenvector $y(c)$, normalized so that $y(c)^* x = 1$.

THEOREM 6.1. *Let A be an n-by-n complex matrix. Let $\lambda, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of A; let $\mu_1, \ldots, \mu_d$ be the nonzero eigenvalues of A that are different from λ; let x and v be nonzero complex vectors such that $Ax = \lambda x$ and $v^* x = 1$; and let $A(c) = cA + (1-c)\lambda x v^*$. Assume that $\lambda \neq 0$.*
(i) Suppose that there is a complex vector y such that $y^* A = \lambda y^*$ and $y^* x = 1$. Assume that $c\lambda_j \neq \lambda$ for each $j = 2, \ldots, n$. Let S_1, Z_1, and B be defined as in Theorem 5.1(c). Then λ is not an eigenvalue of cB. Define the vector $y(c)$ by

$$y(c)^* = y^* + (1-c)\lambda v^* S_1 (\lambda I_{n-1} - cB)^{-1} Z_1^*. \tag{6.1}$$

Then $y(c)$ is the only vector that satisfies the conditions

$$y(c)^* A(c) = \lambda y(c)^* \text{ and } y(c)^* x = 1. \tag{6.2}$$

If λ is a simple eigenvalue of A, then it is not an eigenvalue of B.
(ii) Suppose that λ is a semisimple eigenvalue of A with multiplicity $m \geq 2$ and suppose that

$$c\mu_j \neq \lambda \text{ for each } j = 1, \ldots, d. \tag{6.3}$$

Let $S = [X\ S_2]$ be any nonsingular matrix such that X has m columns and

$$S^{-1}AS = \begin{bmatrix} \lambda I_m & 0 \\ 0 & E \end{bmatrix}, \ E \text{ is } (n-m)\text{-by-}(n-m). \tag{6.4}$$

Then λ is not an eigenvalue of cE or E. Partition $(S^{-1})^* = [Y\ Z_2]$, in which Y has m columns. Then $AX = \lambda X$, $Y^*A = \lambda Y^*$, and $Y^*X = I_m$. Moreover, the columns of X may be chosen to be any m linearly independent right λ-eigenvectors of A, and

$$XY^* = I - (\lambda I - A)(\lambda I - A)^D \tag{6.5}$$

is a projection that is determined uniquely by A and λ, regardless of the choice of columns of X. Define the vector $y(c)$ by

$$y(c)^* = v^*XY^* + (1-c)\lambda v^* S_2(\lambda I_{n-m} - cE)^{-1} Z_2^*. \tag{6.6}$$

Then $y(c)$ satisfies the conditions (6.2); if, in addition, $c \neq 1$, then $y(c)$ is the only vector that satisfies these conditions. If both A and λ are real, then XY^* is a real projection.

(iii) Suppose that λ is a semisimple eigenvalue of A with multiplicity m. Let K be a given compact complex set that does not contain any of the points $\lambda\mu_1^{-1}, \dots, \lambda\mu_d^{-1}$. Let $\tilde{c}$ and c be distinct points in K. If $m \geq 2$, let $y(\cdot)^*$ be defined by (6.6). Then

$$\frac{y(\tilde{c})^* - y(c)^*}{\tilde{c} - c} = \lambda v^* S_2(\tilde{c}E - \lambda I)^{-1}(E - \lambda I)(cE - \lambda I)^{-1} Z_2^*; \tag{6.7}$$

the derivative of $y(c)$ is

$$y'(c)^* = \lambda v^* S_2(cE - \lambda I)^{-2}(E - \lambda I) Z_2^*; \tag{6.8}$$

the derivative of $y(c)^*$ at $c = 0$ is

$$y'(0)^* = \lambda^{-1} v^* S_2(E - \lambda I) Z_2^* = \lambda^{-1} v^*(A - \lambda I); \tag{6.9}$$

and the derivative of $y(c)^*$ at $c = 1$ is

$$y'(1)^* = \lambda v^* S_2(E - \lambda I)^{-1} Z_2^* = \lambda v^*(A - \lambda I)^D. \tag{6.10}$$

If $m = 1$ and $y(\cdot)$ is defined by (6.1), then the four preceding identities are correct if we replace E with B, S_2 with S_1, and Z_2 with Z_1. Finally, independently of $m \geq 1$, for each given vector norm $\|\cdot\|$ there is a positive constant M (depending on A, λ, v, and K) such that

$$\|y(\tilde{c}) - y(c)\| \leq M|\tilde{c} - c| \text{ for all } \tilde{c}, c \in K. \tag{6.11}$$

Proof. (i) The similarity (5.1) shows that the eigenvalues of B are $\lambda_2, \dots, \lambda_n$, so our assumption that $\lambda \neq c\lambda_j$ for all $j = 2, \dots, n$ ensures that λ is not an eigenvalue of cB. If λ is an eigenvalue of B it must have multiplicity at least two as an eigenvalue of A, so if it is a simple eigenvalue of A it is not an eigenvalue of B. The vector $y(c)$ defined by (6.1) satisfies the condition $y(c)^*x = 1$ because $y^*x = 1$ and $Z_1^*x = 0$. To show that it is a left λ-eigenvector of $A(c)$, we begin by combining (5.5) and (5.6):

$$S^{-1}(cA + (1-c)\lambda xv^*)S = \begin{bmatrix} \lambda & (1-c)\lambda v^* S_1 \\ 0 & cB \end{bmatrix}. \tag{6.12}$$

A calculation verifies that the vector $\eta(c)$ defined by

$$\eta(c)^* = [1 \;\; (1-c)\lambda v^* S_1(\lambda I_{n-1} - cB)^{-1}]$$

is a left λ-eigenvector of the matrix in (6.12) and $\eta(c)^* e_1 = 1$; if $c \neq 1$, it is the only such vector. Therefore, the vector $y(c)$ defined by

$$\begin{aligned} y(c)^* = \eta(c)^* S^{-1} &= [1 \;\; (1-c)\lambda v^* S_1(\lambda I_{n-1} - cB)^{-1}] \begin{bmatrix} y^* \\ Z_1^* \end{bmatrix} \\ &= y^* + (1-c)\lambda v^* S_1(\lambda I_{n-1} - cB)^{-1} Z_1^* \end{aligned}$$

is a normalized left λ-eigenvector of $A(c)$, and it is the only vector that satisfies the conditions (6.2).

(ii) Let D denote the block diagonal matrix in (6.4), and let S be any nonsingular matrix such that $S^{-1}AS = D$. Partition $S = [X \; S_2]$ and $(S^{-1})^* = [Y \; Z_2]$, in which X and Y have m columns. Then

$$[AX \;\; AS_2] = AS = SD = [\lambda X \;\; S_2 D],$$

and

$$\begin{bmatrix} Y^* A \\ Z_2^* A \end{bmatrix} = S^{-1}A = DS^{-1} = \begin{bmatrix} \lambda Y^* \\ E Z_2^* \end{bmatrix},$$

which tells us that the columns of X are a linearly independent set of right λ-eigenvectors of A and the columns of Y are a linearly independent set of left λ-eigenvectors of A. The identity $S^{-1}S = I$ tells us that $Y^*X = I_m$ and hence that $X^*Y = (Y^*X)^* = I_m^* = I_m$.

Now let R be any given nonsingular m-by-m matrix, let $\hat{S} = [XR \;\; S_2] := [\hat{X} \;\; S_2]$, partition $(\hat{S}^{-1})^* = [\hat{Y} \;\; \hat{Z}_2]$, compute $(\hat{S}^{-1})^* = [Y(R^{-1})^* \;\; Z_2]$, and notice that $\hat{Y}\hat{X}^* = YX^*$. We draw two conclusions from these observations: (1) We are free to let the columns of X be *any* linearly independent set of right λ-eigenvectors of A; and (2) Regardless of the choice of columns of X, the product YX^* remains the same. Moreover, $(YX^*)^2 = Y(X^*Y)X^* = YI_mX^* = YX^*$, so YX^* (and hence also XY^*) is a projection.

This second conclusion also follows from a useful representation for XY^*. We have

$$\lambda I - A = S \begin{bmatrix} 0 & 0 \\ 0 & \lambda I - E \end{bmatrix} S^{-1} \text{ and } (\lambda I - A)^D = S \begin{bmatrix} 0 & 0 \\ 0 & (\lambda I - E)^{-1} \end{bmatrix} S^{-1},$$

and hence

$$I - (\lambda I - A)(\lambda I - A)^D = I - S \begin{bmatrix} 0 & 0 \\ 0 & I_{n-m} \end{bmatrix} S^{-1} = [X \; S_2] \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Y^* \\ Z_2^* \end{bmatrix} = XY^*.$$

Let the first column of X be the given λ-eigenvector x such that $v^*x = 1$, and write $X = [x \; \tilde{X}]$. Then x is the first column of S, so $S^{-1}x = e_1$ and

$$v^*S = [v^*X \;\; v^*S_2] = [v^*x \;\; v^*\tilde{X} \;\; v^*S_2] = [1 \;\; v^*\tilde{X} \;\; v^*S_2].$$

Thus,

$$S^{-1}(xv^*)S = (S^{-1}x)(v^*S) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & v^*\tilde{X} & v^*S_2 \end{bmatrix} = \begin{bmatrix} 1 & v^*\tilde{X} & v^*S_2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{6.13}$$

and so

$$S^{-1}\left(cA+(1-c)\lambda xv^*\right)S=\begin{bmatrix}\lambda & (1-c)\lambda v^*\tilde{X} & (1-c)\lambda v^*S_2\\ 0 & c\lambda I_{m-1} & 0\\ 0 & 0 & cE\end{bmatrix}. \tag{6.14}$$

The assumption (6.3) (which is satisfied for $c=1$) ensures that λ is not an eigenvalue of cE, and a calculation verifies that $\eta(c)$ defined by

$$\begin{aligned}\eta(c)^* &= [1 \;\; v^*\tilde{X} \;\; (1-c)\lambda v^*S_2(\lambda I-cE)^{-1}]\\ &= [v^*X \;\; (1-c)\lambda v^*S_2(\lambda I-cE)^{-1}]\end{aligned}$$

is a left λ-eigenvector of the matrix in (6.14) and $\eta(c)^*e_1=1$; if $c\neq 1$ it is the unique such vector. Therefore, $y(c)$ defined by

$$\begin{aligned}y(c)^*=\eta(c)^*S^{-1} &= \begin{bmatrix} v^*X & (1-c)\lambda v^*S_2\left(\lambda I-cE\right)^{-1}\end{bmatrix}\begin{bmatrix}Y^*\\ Z_2^*\end{bmatrix}\\ &= v^*XY^*+(1-c)\lambda v^*S_2(\lambda I-cE)^{-1}Z_2^*\end{aligned}$$

satisfies the conditions (6.2); if $c\neq 1$ it is the only vector that satisfies these conditions.

If A and λ are real, the matrix $S=[X\;S_2]$ that gives the reduced form (6.4) may be taken to be real (one may reduce to the real Jordan form, for example [23, Theorem 3.4.5]). Then $(S^{-1})^*=[Y\;Z_2]$ is real, so the uniquely determined product XY^* must always be real, regardless of the choice of X.

(iii) Using the identity $\alpha R^{-1}-\beta T^{-1}=R^{-1}(\alpha T-\beta R)T^{-1}$, we compute

$$\begin{aligned}y(\tilde{c})^*-y(c)^* &= \lambda v^*S_2((1-\tilde{c})(\lambda I-\tilde{c}E)^{-1}-(1-c)(\lambda I-cE)^{-1})Z_2^*\\ &= (\tilde{c}-c)\lambda v^*S_2(\tilde{c}E-\lambda I)^{-1}(E-\lambda I)(cE-\lambda I)^{-1}Z_2^*.\end{aligned}$$

This identity verifies (6.7). One obtains (6.8) by letting $\tilde{c}\to c$; (6.9) and (6.10) follow by setting $c=1$ and $c=0$, respectively. The bound (6.11) follows from taking the norm of both sides of (6.7) and observing that the right-hand side is a continuous function on a compact set, so it is bounded. ∎

The vector function $y(c)$ defined by (6.6) is a complex analytic function at all but finitely many points in the complex plane, provided that λ is a nonzero semisimple eigenvalue of A. The points $c=0$ and $c=1$ are of special interest.

- The condition (6.3) is satisfied for all c such that $|c|<\min\{|\lambda\mu_j^{-1}|:j=1,\ldots,d\}$. Thus, $y(c)$ is analytic in an open neighborhood of $c=0$ and can be represented there by a Maclaurin series obtained from (6.6) by expanding $(\lambda I_{n-m}-cE)^{-1}$ as a power series in c:

$$\begin{aligned}y(c)^* &= v^*\left(I+\sum_{k=1}^{\infty}\lambda^{-k}\left(S_2(E-\lambda I)E^{k-1}Z_2^*\right)c^k\right)\\ &= v^*\left(I+\sum_{k=1}^{\infty}\lambda^{-k}\left((A-\lambda I)A^{k-1}\right)c^k\right).\end{aligned} \tag{6.15}$$

  This representation reveals all of the derivatives of $y(c)$ at $c=0$.

- The condition (6.3) is also satisfied for all c such that $|c-1| < \min\{|\lambda\mu_j^{-1}-1| : j = 1, \ldots, d\}$. Thus, $y(c)$ is analytic in an open neighborhood of $c = 1$. If we let $\gamma = c - 1$, use (6.6), and expand

$$(\lambda I - cE)^{-1} = (\lambda I - E)^{-1}\left(I - \gamma E(\lambda I - E)^{-1}\right)^{-1}$$

as a power series in γ, we obtain

$$y(\gamma+1)^* = v^*\left(XY^* - \lambda\sum_{k=1}^{\infty}\left(S_2(\lambda I - E)^{-k}E^{k-1}Z_2^*\right)\gamma^k\right). \tag{6.16}$$

This series reveals all the derivatives of $y(c)$ at $c = 1$. We can use the Drazin inverse to write this series as

$$y(\gamma+1)^* = v^*\left(XY^* - \lambda\sum_{k=1}^{\infty}\left(((\lambda I - A)^D)^k A^{k-1}\right)\gamma^k\right). \tag{6.17}$$

In particular, the first derivative is

$$y'(1)^* = \lambda v^* S_2(E - \lambda I)^{-1}Z_2^* = \lambda v^*(A - \lambda I)^D. \tag{6.18}$$

6.1. The behavior of $y(c)$ as $c \to 1$. We are interested in the behavior of the left eigenvector $y(c)$ defined by (6.1) as $c \to 1$ in the complex plane, and to understand it better we considered two examples in the Introduction. These two examples are not exceptional: when $\lambda \neq 0$, the essential hypothesis required to ensure that $\lim_{c\to 1} y(c)$ exists for all choices of v is that λ is semisimple. The following theorem verifies this assertion and gives an explicit formula for the limit.

THEOREM 6.2. *Let A be an n-by-n complex matrix with eigenvalues $\lambda, \lambda_2, \ldots, \lambda_n$. Suppose that λ is a nonzero semisimple eigenvalue of A with multiplicity $m \geq 1$; let x and v be given nonzero complex vectors such that $Ax = \lambda x$ and $v^*x = 1$; and let $A(c) = cA + (1-c)\lambda xv^*$. If $m = 1$, let y be the unique vector such that $y^*A = \lambda y$ and $y^*x = 1$. If $m > 1$, let $XY^* = I - (\lambda I - A)(\lambda I - A)^D$ be the projection defined in Theorem 6.1(ii). Then*

(i) For some $\epsilon > 0$ and all complex c such that $0 < |c-1| < \epsilon$, as well as for all complex c such that $\lambda \neq c\lambda_j$ for all $j = 2, \ldots, n$, the vector $y(c)$ defined by (6.1) when λ is simple, or by (6.6) when it is not, is the unique vector that satisfies $y(c)^*A(c) = \lambda y(c)^*$ and $y(c)^*x = 1$.

(ii) If λ is a simple eigenvalue of A, then $\lim_{c\to 1} y(c) = yx^*v = y$.

(iii) If $m > 1$, then

$$\lim_{c\to 1} y(c) = YX^*v = (XY^*)^*v. \tag{6.19}$$

Proof. (i) If λ and 0 are the only eigenvalues of A, then any positive value of ϵ will do. If the nonzero eigenvalues of A that are different from λ are $\mu_1, \ldots, \mu_d$, let

$$\epsilon = \min\{|1 - \lambda\mu_1^{-1}|, \ldots, |1 - \lambda\mu_d^{-1}|\}.$$

Then the hypothesis (5.7) is satisfied and the assertion follows from Theorem 6.1. Since $y(c)$ is defined in a punctured open complex neighborhood of the point $c = 1$,

it is reasonable to ask about the limit of $y(c)$ (as a function of the complex variable c) as $c \to 1$.

(ii) The assertion follows from (6.1) since λ is not an eigenvalue of B:

$$\begin{aligned}\lim_{c\to 1} y^*(c) &= y^* + \lim_{c\to 1}\left((1-c)\lambda v^* S_1(\lambda I - cB)^{-1}Z_1^*\right)\\ &= y^* + \lim_{c\to 1}(1-c)\cdot \lambda v^* S_1 \cdot \lim_{c\to 1}(\lambda I - cB)^{-1}Z_1^*\\ &= y^* + \left(0\cdot \lambda v^* S_1(\lambda I - B)^{-1}Z_1^*\right) = y^* = v^* x y^*.\end{aligned}$$

(iii) This assertion follows in the same way from (6.6):

$$\begin{aligned}\lim_{c\to 1} y(c)^* &= v^* X Y^* + \lim_{c\to 1}\left((1-c)\lambda v^* S_2(\lambda I_{n-m} - cE)^{-1}Z_2^*\right)\\ &= v^* X Y^* + \lim_{c\to 1}(1-c)\cdot \lambda v^* S_2 \cdot \lim_{c\to 1}(\lambda I_{n-m} - cE)^{-1}Z_2^*\\ &= v^* X Y^* + 0\cdot \lambda v^* S_2 \cdot (\lambda I_{n-m} - E)^{-1}Z_2^* = v^* X Y^*.\end{aligned}$$

□

6.2. A special case: The general parametric Google matrix. We begin with a summary of the properties of a row-stochastic matrix that are relevant to our analysis of the general parametric Google matrix.

LEMMA 6.3. [2, 24] *Let A be a row-stochastic matrix. Then*

(i) $\lambda = 1$ is an eigenvalue of A associated with the right eigenvector $x = e$.

(ii) Every entry of A is in the real interval $[0, 1]$.

(iii) For each $k = 1, 2, \ldots$, A^k is row-stochastic, so its entries remain bounded as $k \to \infty$.

(iv) Every eigenvalue of A has modulus at most 1.

(v) Every eigenvalue of A that has modulus 1 is semisimple.

(vi) If the eigenvalue 1 has multiplicity m, then the Jordan canonical form of A is

$$I_m \oplus J_{n_1}(\nu_1) \oplus \cdots \oplus J_{n_k}(\nu_k),$$

in which each $\nu_j \neq 1$, each $|\nu_j| \leq 1$, and $n_j = 1$ if $|\nu_j| = 1$.

(vii) If 1 is a simple eigenvalue of A, then there is a unique vector y with nonnegative entries such that $y^T A = y^T$ and $y^T e = 1$.

Since the basic Google matrix G has all the properties stated in the preceding lemma, and since these properties are special cases of the key hypotheses in our analyses in the preceding sections, specialization of our general results permits us to identify several pleasant and useful properties of the *general parametric Google matrix* $G(c) = cG + (1-c)xv^*$ with complex c and v. In fact the following theorem is an interpretation of Theorems 6.1 and 6.2 when A is the Google matrix and hence $\lambda = 1$.

THEOREM 6.4. *Let G be an n-by-n row stochastic matrix, and let its eigenvalue $\lambda = 1$ (necessarily semisimple) have multiplicity $m \geq 1$. If $m = 1$, let y be the unique vector with nonnegative entries such that $y^T G = y^T$ and $y^T e = 1$. If $m > 1$, let the m columns of X be any linearly independent set of right 1-eigenvectors of G, and let Y be the matrix defined in Theorem 6.1(ii); its columns are an independent set of left 1-eigenvectors of G. Let v be a given complex vector such that $v^* e = 1$, let c be a complex number, and let $G(c) = cG + (1-c)ev^*$. Let $1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of G, let $\mu_1, \ldots, \mu_d$ be the nonzero eigenvalues of G that are different from 1, let*

$$\epsilon = \min\{|1 - \mu_1^{-1}|, \ldots, |1 - \mu_d^{-1}|\},$$

and let

$$I_m \oplus J_{n_1}(\nu_1) \oplus \cdots \oplus J_{n_k}(\nu_k), \ \ \textit{each } \nu_j \neq 1 \tag{6.20}$$

be the Jordan canonical form of G, with $\{\mu_1, \ldots, \mu_d\} \subseteq \{\nu_1, \ldots, \nu_k\} \subseteq \{\lambda_2, \ldots, \lambda_n\}$. Then

(i) The eigenvalues of $G(c)$ are $1, c\lambda_2, \ldots, c\lambda_n$, and $|c\lambda_j| \leq |c|$ for each $j = 2, \ldots, n$.

(ii) In the Jordan canonical form (6.20), $n_j = 1$ for each j such that $|\nu_j| = 1$.

(iii) If $0 < |c| < 1$ (or, more generally, if $c \neq 0$ and $1 \neq c\mu_j$ for each $j = 1, \ldots, d$), then the Jordan canonical form of $G(c)$ is

$$[1] \oplus cI_{m-1} \oplus J_{n_1}(c\nu_1) \oplus \cdots \oplus J_{n_k}(c\nu_k)$$

if $m > 1$; it is

$$[1] \oplus J_{n_1}(c\nu_1) \oplus \cdots \oplus J_{n_k}(c\nu_k)$$

if $m = 1$.

(iv) Suppose either that $|c| < 1$ or that $0 < |1 - c| < \epsilon$. Then 1 is a simple eigenvalue of $G(c)$.

(v) Suppose either that $|c| < 1$ or that $0 < |1 - c| < \epsilon$. If $m > 1$, the unique left 1-eigenvector $y(c)$ of $G(c)$ such that $y(c)^* e = 1$ is defined by

$$y(c)^* = v^* XY^* + (1 - c)v^* S_2 (I_{n-m} - cE)^{-1} Z_2^*, \tag{6.21}$$

and

$$\lim_{c \to 1} y(c) = YX^* v. \tag{6.22}$$

The matrices S_2, E, and Z_2 are defined in Theorem 6.1(ii); 1 is not an eigenvalue of E. The matrix

$$YX^* = I - (I - G^T)(I - G^T)^D \tag{6.23}$$

is a real projection with nonnegative entries.

(vi) Suppose either that $|c| < 1$ or that $0 < |1 - c| < \epsilon$. If $m = 1$, the unique left 1-eigenvector $y(c)$ of $G(c)$ such that $y(c)^* e = 1$ is defined by

$$y(c)^* = y^* + (1 - c)v^* S_1 (I_{n-1} - cB)^{-1} Z_1^*, \tag{6.24}$$

and

$$\lim_{c \to 1} y(c) = y. \tag{6.25}$$

The matrices S_1, Z_1, and B are defined in Theorem 6.1(i); 1 is not an eigenvalue of B.

(vii) The vector function $y(c)$ defined by (6.21) if $m > 1$, and by (6.24) if $m = 1$, is analytic in the unit disk $\{c : |c| < 1\}$ and is represented there by the Maclaurin series

$$y(c)^* = v^* \left(I + \sum_{k=1}^{\infty} \left((G - I)G^{k-1} \right) c^k \right). \tag{6.26}$$

(viii) Let $\gamma = c - 1$. The vector function $y(c)$ defined by (6.21) if $m > 1$, and by (6.24) if $m = 1$, is analytic in the disk $\{c : |1 - c| < \epsilon\}$ and is represented there by the power series

$$y(c)^* = y(\gamma + 1)^* = v^* \left(XY^* - \sum_{k=1}^{\infty} \left(((I - G)^D)^k G^{k-1} \right) \gamma^k \right). \tag{6.27}$$

In particular, the first derivative at $c = 1$ is

$$y'(1)^* = v^*(G - I)^D. \tag{6.28}$$

(ix) Let K be a given compact complex set that does not contain any of the points $\mu_1^{-1}, \ldots, \mu_d^{-1}$. Define $y(c)$ on K by (6.21) if $m > 1$ and by (6.24) if $m = 1$. Then $\|y(c)\|_1 \geq 1$ for all $c \in K$ and there is a positive constant M such that

$$\frac{\|y(\tilde{c}) - y(c)\|_1}{\|y(c)\|_1} \leq \|y(\tilde{c}) - y(c)\|_1 \leq M|\tilde{c} - c| \text{ for all } \tilde{c}, c \in K.$$

The assertions in (vii) and (viii) of the preceding Theorem follow from (6.15), (6.17), and (6.18). The assertion (ix) follows from Theorem 6.1(iii) and the observation that $1 = |y(c)^* e| \leq \|y(c)^*\|_1$.

We emphasize that the representations (6.21) and (6.24) for the unique normalized left 1-eigenvector of $G(c)$ are valid not only for all real $c \in (0, 1)$, but also for all complex c in the open unit disk, as well as for all c in a punctured open neighborhood of the point 1 in the complex plane. The limits (6.22) and (6.25) are to be understood as limits of functions of a complex variable; the existence of these limits ensures that they may be computed via any sequence of values of c that tends to 1.

The preceding comments have an important consequence. Suppose the vector v has positive real entries and satisfies $v^T e = 1$. Then for all real c such that $0 < c < 1$, $G(c)$ has positive entries. The Perron-Frobenius Theorem ensures that each such $G(c)$ has a unique left 1-eigenvector $y(c)$ that has positive entries and satisfies $y(c)^T e = 1$. Theorem 6.4 ensures that $\lim_{c \to 1} y(c) = \tilde{y}$ exists, so if we take this limit with $c \in (0, 1)$ we know that $\tilde{y}$ has real nonnegative entries. However, we can also take this limit with c tending to 1 along some non-real path in the complex plane. Regardless of the path taken, and even though $y(c)$ can be non-real on that path, nevertheless the limit obtained is always the nonnegative vector $\tilde{y}$ (this further degree of freedom is exploited in the algorithm presented in Section 8).

We can draw one more conclusion from the preceding discussion, which is the last statement in Theorem 6.4(v). For each given nonnegative vector v, we have argued that the vector

$$\tilde{y} = \lim_{c \to 1} y(c) = YX^* v$$

has nonnegative entries. But a matrix N has the property that the entries of Nv are nonnegative whenever the entries of v are nonnegative if and only if all the entries of N are nonnegative. Thus, the projection

$$YX^* = [\eta_1 \ \ldots \ \eta_n] = I - (I - G^T)(I - G^T)^D$$

is both real and nonnegative. Its columns $\eta_1, \ldots, \eta_n$ are a uniquely determined set of nonnegative left 1-eigenvectors of G such that, for any given nonnegative probability vector v, $\lim_{c \to 1} y(c) = v_1\eta_1 + \cdots + v_n\eta_n$ is a convex combination of them.

Of course the nonnegativity of N could be obtained elementarily also by using Markov chains arguments.

7. Comparison with an explicit prior expression of Google Jordan form. We now consider a result from [37] and we ask ourselves how one can obtain, extend, and interpret them, by employing our findings in the previous sections and by allowing the parameter c in the complex field.

Indeed, we study the Jordan form general case, in which G is not necessarily diagonalizable, where the decomposition $G = SJS^{-1}$,

$$J = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_2 & \blacklozenge & & \\ \vdots & & \ddots & \ddots & \\ \vdots & & & \lambda_{n-1} & \blacklozenge \\ 0 & \cdots & \cdots & 0 & \lambda_n \end{bmatrix} \tag{7.1}$$

is the Jordan Canonical Form of G and $\blacklozenge$ denotes a value that can be either 0 or 1.

THEOREM 7.1. *Let G be a row stochastic matrix of size n, let $c \in (0,1)$, and suppose that v is a nonnegative n-vector whose entries add to 1. Consider the matrix $G(c) = cG + (1-c)ev^T$ and let $G = SJ(1)S^{-1}$, $S = [e\ x_2\ \ldots\ x_n]$, $[S^{-1}]^T = [y\ \ y_2\ \ldots\ y_n]$, and*

$$J(c) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & c\lambda_2 & c\cdot\blacklozenge & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & c\lambda_{n-1} & c\cdot\blacklozenge \\ 0 & \cdots & \cdots & 0 & c\lambda_n \end{bmatrix},$$

$$J(c) = D^{-1} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & c\lambda_2 & \blacklozenge & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & c\lambda_{n-1} & \blacklozenge \\ 0 & \cdots & \cdots & 0 & c\lambda_n \end{bmatrix} D \tag{7.2}$$

in which $D = \mathrm{diag}(1, c, \ldots, c^{n-1})$ and $\blacklozenge$ denotes a value that can be 0 or 1. Then

$$G(c) = ZJ(c)Z^{-1},$$

in which

$$Z = SR^{-1},$$

$$R = I_n + e_1 w^T, \quad w^T = [0\ w[2]\ \ldots\ w[n]],$$

$$w[2] = (1-c)v^T x_2/(1-c\lambda_2), \tag{7.3}$$

$$w[j] = [(1-c)v^T x_j + [J(c)]_{j-1,j} w[j-1]]/(1-c\lambda_j),\ j = 3, \ldots, n. \tag{7.4}$$

In particular

$$y(c) = y + \sum_{j=2}^{n} w[j]y_j \tag{7.5}$$

where $y = y(1)$ *if the eigenvalue* 1 *of* $G = G(1)$ *is simple and where the quantities* $w[j]$ *are expressed as in (7.3)-(7.4). Conversely,* y *is one of the basic PageRank vectors when the eigenvalue* 1 *of* $G(1)$ *is semisimple but not simple.*

Notice that in the original paper [37], there is a typo since D and D^{-1} are exchanged in (7.2): we thank Gang Wu and Yimin Wei for pointing this out to our attention, see [40].

7.1. Matching old and new representations. Here we make a critical analysis of the above results in the light of the conclusions in Subsection 6.2. From Lemma 6.3(vi) and Theorem 6.4 we know that the eigenvalue 1 in the matrix $G = G(1)$ is semisimple with multiplicity m. Therefore $[J(c)]_{j-1,j} = 0$ and $1 - c\lambda_j = 1 - c$, $j = 2, \ldots, m$. Hence, as already acknowledged in [37][Section 3], the coefficient $w[j]$, $j = 2, \ldots, m$, is equal to $v^T x_j = x_j^T v$ and then

$$y(c) = y + \sum_{j=2}^{m} y_j (x_j^T v) + \sum_{j=m+1}^{n} w[j]y_j.$$

Therefore the Cesaro averaging projector N already discussed in the previous sections has the form $N = YX^* = [y\ y_2\ \cdots\ y_m][e\ x_2\ \cdots\ x_m]^T$ and hence $y(c) = Nv + \sum_{j=m+1}^{n} w[j]y_j$. Moreover the eigenvalue λ_j, $j \geq m+1$, is different from 1, is in modulus bounded by 1, and if unimodular then it is semisimple. As a consequence, by (7.3)–(7.4), we obtain $\lim_{c\to 1} w[j] = 0$, $j = m+1, \ldots, n$, so that

$$\lim_{c\to 1} y(c) = Nv$$

which agrees with (6.19), (6.22), and (6.23): moreover, by the general reasoning at the end of Subsection 6.2, we deduce that N is entry-wise nonnegative.

Now, by taking into account the notations in (6.20) considered in Theorem 6.4, and by looking carefully at the expression of coefficients $w[j]$, $j = m+1, \ldots, n$, in (7.3)–(7.4), we can rewrite the vector $y(c)$ as

$$y(c) = Nv + \sum_{j=m+1}^{n} w[j]y_j = Nv + (1-c)\sum_{j=1}^{k}\sum_{s=1}^{n_j}\sum_{t=1}^{n_j+1-s} c^{t-1}(1-c\nu_j)^{-t}(x_{j,s}^T v)\ y_{j,t}, \tag{7.6}$$

where the vectors x_j, y_j, $j = m+1, \ldots, n$, in the former representation, have been reorganized according to the Jordan structure as $x_{j,s}, y_{j,s}$, $j = 1, \ldots, k$, $s = 1, \ldots, n_j$ (ref. Theorem 6.4). If we compare the latter equation with the Toeplitz matrices (Toeplitz, i.e., constant along diagonals, see e.g. [10]) of size n_j

$$J_{n_j}(\nu_j) = \begin{bmatrix} \nu_j & 1 & 0 & \cdots & 0 \\ 0 & \nu_j & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \nu_j & 1 \\ 0 & \cdots & \cdots & 0 & \nu_j \end{bmatrix},$$

$$T_{n_j}(c) = \begin{bmatrix} 1-c\nu_j & -c & 0 & \cdots & 0 \\ 0 & 1-c\nu_j & -c & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & 1-c\nu_j & -c \\ 0 & \cdots & \cdots & 0 & 1-c\nu_j \end{bmatrix}^{-1} =$$

$$= \frac{1}{1-c\nu_j} \begin{bmatrix} 1 & \frac{c}{1-c\nu_j} & \frac{c^2}{(1-c\nu_j)^2} & \cdots & \frac{c^{n_j-1}}{(1-c\nu_j)^{n_j-1}} \\ 0 & 1 & \frac{c}{1-c\nu_j} & \cdots & \frac{c^{n_j-2}}{(1-c\nu_j)^{n_j-2}} \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & 1 & \frac{c}{1-c\nu_j} \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix},$$

we observe $T_{n_j}(c) = (I_{n_j} - cJ_{n_j}(\nu_j))^{-1}$ and therefore

$$\sum_{s=1}^{n_j} \sum_{t=1}^{n_j+1-s} c^{t-1}(1-c\nu_j)^{-t}(x_{j,s}^T v)\ y_{j,t} = [y_{j,1}\ \cdots\ y_{j,n_j}]T_{n_j}^T(c)[x_{j,1}\ \cdots\ x_{j,n_j}]^T v.$$

Hence, taking into account (7.6), we can write

$$\begin{aligned} y(c) &= Nv + (1-c)\sum_{j=1}^{k}[y_{j,1}\ \cdots\ y_{j,n_j}]\cdot \\ &\quad \cdot(I_{n_j} - cJ_{n_j}^T(\nu_j))^{-1}[x_{j,1}\ \cdots\ x_{j,n_j}]^T v \\ &= (1-c)(I-cG^T)^{-1}v \end{aligned} \tag{7.7}$$

which coincides with the general representation (6.6), where $X = [e\ x_2\ \cdots\ x_m]$, $Y = [y\ y_2\ \cdots\ y_m]$, $N = YX^T$, $E = J_{n_1}(\nu_1)\oplus\cdots\oplus J_{n_k}(\nu_k)$, as in the expression (6.20), and $S_2 = [X_1\ \cdots\ X_k]$, $Z_2 = [Y_1\ \cdots\ Y_k]$, $X_j = [x_{j,1}\ \cdots\ x_{j,n_j}]$, $Y_j = [y_{j,1}\ \cdots\ y_{j,n_j}]$, $j = 1, \ldots, k$.

7.2. Eigenvector structure of $G(c)$, discontinuity points in its Jordan form. When writing the Jordan form in Theorem 7.1, the matrix D is chosen as

$$\mathrm{diag}(1, c, \ldots, c^{n-1}).$$

However, that matrix is not unique: for instance the matrix

$$\hat{D} = I_m \oplus \mathrm{diag}(1, c, \ldots, c^{n-m-1})$$

is also a feasible choice, since the Jordan structure of $G(c)$ is equally obtained as $DJ(c)D^{-1} = \hat{D}J(c)\hat{D}^{-1}$. Indeed, following the Jordan blocks structure in (6.6), we can define a new optimal diagonal matrix $\tilde{D}$ of minimal conditioning with the constraint that $DJ(c)D^{-1} = \tilde{D}J(c)\tilde{D}^{-1}$. This optimal matrix takes the form

$$\tilde{D} = I_m \oplus \mathrm{diag}(1, c, \ldots, c^{n_1-1}) \oplus \mathrm{diag}(1, c, \ldots, c^{n_2-1}) \oplus \cdots \oplus \mathrm{diag}(1, c, \ldots, c^{n_k-1}).$$

Therefore, switching from $G = G(1)$ to $G(c)$, while the eigenvalues change in a smooth way since $1 \to 1$ with the same multiplicity m, $\nu_j \to c\nu_j$, $j = 1, \ldots, k$, the left and

right vectors change as follows

$$\begin{array}{lll} x_{j,t} \to c^{1-t}\left[x_{j,t} - (1-c)\frac{c^t}{1-c\nu_j}e\right], & y_{j,t} \to c^{t-1}y_{j,t}, & t = 1,\dots,n_j, \\ x_t \to x_t - e, & y_t \to y_t, & t = 2,\dots,m, \\ x_1 \equiv e \to e, & y_1 \equiv y \to y(1) = Nv. & \end{array}$$

Therefore, in the given representation and under the assumption of nondiagonalizable G, the Jordan canonical form has a discontinuity at $c = 0$, while it behaves smoothly at $c = 1$. In fact, $\lim_{c\to 0} G(c) = ev^T$ is not normal in general but it is diagonalizable, while $G(c)$ with $c \neq 0$ is not diagonalizable in general: in fact $G(c)$ has the same Jordan pattern as $G(1)$ for $c \neq 0$ while it is diagonalizable for $c = 0$. Hence, as emphasized in the previous displayed equations, it is clear that the discontinuity/degeneracy is located in the left and right vectors associated to nontrivial Jordan blocks. As a consequence the matrix $G(c)$ is continuous at $c = 0$, but it is not so for its Jordan representation. On the other hand the other discontinuities at $c = \nu_j^{-1}$, for every $j = 1,\dots,k$, are essential not only in the representations, but also in the matrix $G(c)$, and at the point $c = 1$ every involved quantity is analytic.

Finally it should be noted the following "surprising" fact: not only nothing bad happens at $c = 1$, but indeed nothing bad happens for $c > 1$ (at least, a little bit bigger than 1) and this is not seen by the power series representations of $y(c)$ described in the literature, which diverge for $c > 1$ (see [8]).

7.3. Condition number of $y(c)$: general derivation. Given its relevance for numerical stability, we consider in some detail the conditioning of $y(c)$ in several norms and especially in the more natural l^1 norm. More precisely, we are interested in estimating

$$\kappa(y(c),\delta) = \frac{\|y(\tilde{c}) - y(c)\|}{\|y(c)\|},$$

with $\tilde{c} = c(1+\delta)$, δ complex parameter of small modulus, K compact set in the complex field nonintersecting $\{\mu_j^{-1} : \ j = 1,\dots,d\}$, and $c, \tilde{c} \in K$. Since $y(c)$ is analytic in its domain, it is clear that

$$\kappa(y(c),\delta) = \kappa_c \frac{|c\delta|}{\|y(c)\|}\,(1 + O(\delta))$$

with $\kappa_c = \|y'(c)\|$. Our next task is the differentiation of $y(c)$ in the light of (7.7), and especially its norm evaluation. We have

$$\begin{aligned} y'(c) = &-\sum_{j=1}^{k} [y_{j,1} \ \cdots \ y_{j,n_j}](I_{n_j} - cJ_{n_j}^T(\nu_j))^{-1}[x_{j,1} \ \cdots \ x_{j,n_j}]^T v + \\ &+(1-c)\sum_{j=1}^{k} [y_{j,1} \ \cdots \ y_{j,n_j}](I_{n_j} - cJ_{n_j}^T(\nu_j))^{-2} J_{n_j}^T(\nu_j)[x_{j,1} \ \cdots \ x_{j,n_j}]^T v, \end{aligned} \tag{7.8}$$

which of course agrees with the differentiation of (7.6), after observing that

$$(I_{n_j} - cJ_{n_j}(\nu_j))^{-2} J_{n_j}(\nu_j) = \frac{1}{1-c\nu_j} \begin{bmatrix} t_0 & t_1 & t_2 & \cdots & t_{n_j-1} \\ 0 & t_0 & t_1 & \cdots & t_{n_j-2} \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & t_0 & t_1 \\ 0 & \cdots & \cdots & 0 & t_0 \end{bmatrix}$$

with $t_s = \frac{sc^{s-1}}{(1-c\nu_j)^s} + \frac{(s+1)\nu_j c^s}{(1-c\nu_j)^{s+1}}$, $s = 0, \dots, n_j - 1$. In fact, upper triangular Toeplitz matrices form a commutative algebra and the generic coefficient on the diagonal in the result is a simple convolution of the coefficients of the factors. Therefore, putting the two terms of (7.8) together, we find

$$\begin{aligned} y'(c) &= \sum_{j=1}^{k} [y_{j,1} \ \cdots \ y_{j,n_j}] \tilde{T}_{n_j}^T(c) [x_{j,1} \ \cdots \ x_{j,n_j}]^T v \\ &= \sum_{j=1}^{k} Y_j \tilde{T}_{n_j}^T(c) X_j^T v \\ &= Z_2 \left[\oplus_{j=1}^{k} \tilde{T}_{n_j}^T(c) \right] S_2^T v, \end{aligned} \tag{7.9}$$

with

$$\tilde{T}_{n_j}(c) = \frac{1}{1-c\nu_j} \begin{bmatrix} \tilde{t}_0 & \tilde{t}_1 & \tilde{t}_2 & \cdots & \tilde{t}_{n_j-1} \\ 0 & \tilde{t}_0 & \tilde{t}_1 & \cdots & \tilde{t}_{n_j-2} \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \tilde{t}_0 & \tilde{t}_1 \\ 0 & \cdots & \cdots & 0 & \tilde{t}_0 \end{bmatrix},$$

$\tilde{t}_s = -\frac{c^s}{(1-c\nu_j)^s} + (1-c)\left[\frac{sc^{s-1}}{(1-c\nu_j)^s} + \frac{(s+1)\nu_j c^s}{(1-c\nu_j)^{s+1}}\right]$, $s = 0, \dots, n_j - 1$, and with $S_2 = [X_1 \ \cdots \ X_k]$, $Z_2 = [Y_1 \ \cdots \ Y_k]$.

Therefore looking at the dependence with respect to the parameter c we find that κ_c grows generically, in a neighborhood of 1, μ_j^{-1}, $j = 1, \dots, d$, as

$$\max_{j=1,\dots,k} \left| \frac{\tilde{t}_{n_j-1}}{1-c\nu_j} \right|, \quad \frac{\tilde{t}_{n_j-1}}{1-c\nu_j} = \left[z_1(1-c\nu_j)^{-n_j} + z_2(1-c\nu_j)^{-n_j-1}(1-c) \right] c^{n_j-2}, \tag{7.10}$$

$z_1 = (n_j - 1)(1-c) - c$, $z_2 = c\nu_j n_j$, which agrees with the estimate in the introduction (see (1.3)-(1.4)). More precisely, for almost every v nonnegative and with unit l^1 norm, there exists a positive constant $\theta = \theta(S, v)$, independent of c, such that

$$\kappa_c \geq \theta \max_{j=1,\dots,k} \left| z_1(1-c\nu_j)^{-n_j} + z_2(1-c\nu_j)^{-n_j-1}(1-c) \right| |c|^{n_j-2}. \tag{7.11}$$

In fact by elementary measure theory argument, the set of all possible v such that $x_{j,s}^T v = 0$ for at least one index $j = 1, \dots, k$ and one index $s = 1, \dots, n_j$ has zero Lebesgue measure. On the other hand, taking into account (7.9), a direct majorization of the quantity κ_c leads to

$$\kappa_c \leq \sum_{j=1}^{k} \|Y_j\| \|\tilde{T}_{n_j}^T(c)\| \|X_j^T v\|$$

and to the more appealing

$$\kappa_c \leq \|Z_2\| \cdot \| \oplus_{j=1}^{k} \tilde{T}_{n_j}^T(c)\| \cdot \|S_2^T v\|. \tag{7.12}$$

If we take reasonable norms as the l^p norms with $p \in [1, \infty]$, then by recalling $G = G(1) = SJS^{-1}$ and since S_2, Z_2 are submatrices of S, S^{-1} respectively, the bound in

(7.12) directly implies the following

$$\kappa_c \leq \kappa(S) \max_{j=1,\ldots,k} \left\| \tilde{T}_{n_j}^T(c) \right\| \|v\|. \tag{7.13}$$

In other words, the first part which does not depend on c, tells us that the conditioning of $y(c)$ can be associated with the (lack of) orthogonality of the left and right vectors in the Jordan form of G not associated to the eigenvalue 1, while the second part carries the information on the parameter c. Notice that the generic, lower, and upper bounds in (7.10)–(7.13) are all well defined also at $c = 1$, and indeed the latter improves the estimates known in the literature, where, for $0 \leq c < 1$, the amplification factor upper-bound grows as $(1-c)^{-1}$ and blows up at $c = 1$: see [29] and references there reported; however, it has to be pointed out that these estimates in [29] are more general since they are based on an arbitrary perturbation $\tilde{G}$ of $G = G(1)$ subject to the only constraint that $\tilde{G}$ is irreducible and stochastic. Furthermore, for c in the unit disk and far away from $c = 1$, the obtained amplification factor is simpler and more useful, i.e., $(1-|c|)^{-1} + |1-c|(1-|c|)^{-2}$ which reduces to $2(1-c)^{-1}$ for $0 \leq c < 1$: therefore our more detailed analysis is of interest essentially in the vicinity of critical points $c = \mu_j^{-1}$, $j = 1, \ldots, d$, $c = 1$, all outside or on the frontier of the unity disk.

7.4. Condition number of $y(c)$: norm analysis of $\tilde{T}_{n_j}^T(c)$. A critical analysis of (7.13) shows that the quantities $\kappa(S)$ and $\|v\|$ are fixed data of the problem ($G = G(1)$ and v); in particular, since $\|v\|_1 = 1$ and $\|\cdot\|_p \leq \|\cdot\|_1$, $p \in [1, \infty]$, we uniformly have $\|v\|_p \leq 1$. Hence we should focus our attention on $\left\| \tilde{T}_{n_j}^T(c) \right\|$, $j = 1, \ldots, k$.

For instance, by considering the l^1 and the l^∞ norms, we have

$$\left\| \tilde{T}_{n_j}^T(c) \right\|_1 = \left\| \tilde{T}_{n_j}^T(c) \right\|_\infty = \sum_{s=0}^{n_j-1} \left| \frac{\tilde{t}_s}{1-c\nu_j} \right|,$$

which grows as $\frac{\tilde{t}_{n_j-1}}{1-c\nu_j}$, for c in a neighborhood of μ_j^{-1}. However, $|\mu_j^{-1}| \geq 1$ while, especially for computational purposes, we are more interested in the behavior of the conditioning for c of modulus at most 1.

In such a case, independently of the chosen norm among l^1, l^2, l^∞, we observe the following: for c such that $|1-c\nu_j| < |c|$, the conditioning of $\tilde{T}_{n_j}^T(c)$ grows exponentially with the size n_j of the Jordan blocks; of course, also for Jordan blocks of moderate size, the conditioning can become very high. For $|1-c\nu_j| = |c|$, it is clear that the conditioning grows as n_j^2 which can become large only for quite high-dimensional Jordan blocks. For $|1-c\nu_j| > |c|$ the situation is very interesting because, irrespectively of the size n_j the conditioning is bounded. Indeed, by looking at the induced l^2 (the spectral norm), classical results on Toeplitz operators (see the Szegö distribution result in the classical Böttcher, Silbermann book [10]) tell us that there exists a proper function $g_{j,c}(t)$ defined on $[0, 2\pi)$

$$\left\| \tilde{T}_{n_j}^T(c) \right\|_2 \leq \|g_{j,c}(t)\|_\infty, \qquad \lim_{n_j \to \infty} \left\| \tilde{T}_{n_j}^T(c) \right\|_2 = \|g_{j,c}(t)\|_\infty.$$

That function $g_{j,c}(t)$ called symbol is obtained through the coefficients of $\tilde{T}_{n_j}(c)$ in the sense that these coefficients are Fourier coefficients of $g_{j,c}(t)$. In our specific setting a straightforward computation shows that the symbol $g_{j,c}(t)$ is

$$\frac{\partial}{\partial c}(1-c)(1-c[\nu_j+\exp(-\mathbf{i}t)])^{-1} = (\nu_j-1+\exp(-\mathbf{i}t))(1-c[\nu_j+\exp(-\mathbf{i}t)])^{-2}, \quad \mathbf{i}^2 = -1.$$

Therefore the quantity

$$\max_{t\in[0,2\pi)} \left|(\nu_j - 1 + \exp(-\mathbf{i}t))(1 - c[\nu_j + \exp(-\mathbf{i}t)])^{-2}\right|$$

represents a tight measure, irrespectively of n_j, of the contribution of $\tilde{T}^T_{n_j}(c)$ to the conditioning of $y(c)$ in l^2 norm. In this context a tight measure of the l^1 norm would have been more desirable, since the l^1 norm represents the most natural choice for the problem at hand.

7.5. Condition number of $y(c)$: extremal examples. Here we are interested in showing two extremal examples taken from very structured Web graphs. The Web graph is the one produced by a unique huge loop: page i links only to page $i+1$, $i = 1, \ldots, n-1$, page n links only to page 1; since the set of dangling nodes is empty the matrix $G = G(1)$ is a special cyclic permutation matrix which generates the algebra of circulants. Circulant matrices are normal and diagonalized by the discrete Fourier transform so that in the Jordan form we have $x_j = f_j$, $y_j = \bar{f}_j$ with

$$f_j = \frac{1}{\sqrt{n}} \left(\exp\left(-\frac{\mathbf{i}2\pi jk}{n}\right)\right)_{k=0}^{n-1}, \quad j = 0, \ldots, n-1.$$

The eigenvalues of $G = G(1)$, accordingly to the same ordering of the Fourier eigenvectors, are $\omega_j = \exp\left(-\frac{\mathbf{i}2\pi j}{n}\right)$, $j = 0, \ldots, n-1$ (the n roots of unity). We notice that e, the used vector of all ones, coincides with $\sqrt{n}f_0$. Therefore if the rank-one correction is chosen with $v = e/n = f_0/\sqrt{n}$, then ev^T is also a circulant matrix. In this specific example the computed vector $y(c)$ coincides with v independently of c and therefore $k_c = 0$. Therefore for this given graph, the chosen vector v lies in the zero measure set excluded when deriving (7.11). In fact, for this graph and for this vector v we have that the vectors x_j, $j = m+1, \ldots, n$, $m = 1$, are all orthogonal to v and then the whole expression in (7.8) trivially vanishes.

More delicate is to try to satisfy (7.13) with equality. For important examples the estimate is not tight, but it is not too bad at least in a neighborhood of $c = 1$. Take the above graph, consider $v = e_1$ with $c = 1$. In such a case the estimate (7.13) of κ_1 gives

$$|1 - \omega_1|^{-1} = \left[|1 - \cos(2\pi(n))|^2 + \sin^2(2\pi/n)\right]^{-1/2} \sim n/2\pi.$$

A direct computation of $y'(c)$ at $c = 1$ gives the expression

$$y'(1) = -\sum_{j=1}^{n-1} \bar{f}_j (1 - \omega_j)^{-1} (f_j^T e_1)$$

Since $\bar{f}_j$, $j = 1, \ldots, n-1$, are orthonormal and since $|f_j^T e_1| = 1/\sqrt{n}$ it easily follows that

$$\|y'(1)\|_2 = \sqrt{\sum_{j=1}^{n-1} (\sqrt{n}|1 - \omega_j|)^{-2}} \sim \sqrt{n}\sqrt{\sum_{j=1}^{n-1} (2\pi j)^{-2}}$$

so that the real l^2 norm of $y'(c)$ differs, asymptotically, from the bound (7.13) by a factor $\sqrt{n}$.

8. Computational suggestions. The spectral structure of $G(c)$ was first comprehended in the context of sophisticated results about Markov chains, which required that $c \in [0,1)$ and $v \geq 0$. We now know that the spectral (indeed, the Jordan, Schur etc.) structure of $G(c)$ follows from basic matrix analytic facts that permit both c and v to be complex. This new freedom in the Google perturbation is exploited to compute the PageRank more efficiently, especially when c is close to 1 or even equal to 1.

The algorithms that we propose have to be regarded as a preliminary step that, in our opinion, merits further research.

We choose p small integer number (let us say $p = 10$) and we compute $y(c_j)$, $j = 0, \ldots, p-1$, at equally-spaced points c_j, $j = 0, \ldots, p-1$, on the complex circle of radius (let us say) 0.5 or 0.25. The computations are extremely fast since the standard power method at the kth iteration converges with a relative reduction error of at least $|c|^k$ (see [21, Chapter 7, p. 330]), which is independent of the huge size of the problem; indeed, the nature of our data permits us to use a vector-valued DFT procedure, whose numerical stability is excellent. We employ these p vectors as a starting point for a specific extrapolation algorithm at $c = 0.85$ or $c = 0.99$, whose details are given in [15, 13]. The idea is to use the expansion of $y(c)$ around $c = 1$ as in (6.27) with $\gamma = c + 1$ or as in (6.21) or as in (7.3)-(7.5), and to employ linear combinations in order to cancel out certain terms in the remainder $y(c) - \tilde{y}$, $\tilde{y} = y(1) = Nv$, and to increase the accuracy; see [12, Chapter 4] for details. The vector $\hat{y}(c)$, computed by extrapolation, will be corrupted by errors of approximation and due to roundoff: therefore, since we know in advance that $y(c)$ has to be nonnegative and normalized, we set to zero the real part whenever negative and the imaginary part, and we normalize the resulting nonnegative vector, by dividing by its l^1 norm (in this case the sum of all the coefficients). Finally we can use a standard iterative procedure (the power method or iterative techniques for an equivalent linear system [27, 32, 17, 26]) as an iterative refinement to increase the precision. We remind that computing the PageRank with $c = 0.99$ or 1 is very difficult by straightforward techniques, due to slow convergence or even to lack of convergence for $c = 1$; see [17] and references therein, and [31, Section 6.1] for a specific discussion on the case $c = 0.99$.

All this comprises a new scheme to compute the PageRank, with c equal to 1 or very close to it, is:

- Step 1: Compute $y(c_j)$, $c_j = 0.25 \cdot \exp(\mathbf{i} 2j\pi/p)$, $\mathbf{i}^2 = -1$, $j = 0, \ldots, p-1$ (*Evaluation via vector DFT*).
- Step 2: *Vector Extrapolation* at the desired (difficult) $c \approx 1$ (e.g. $c = 0.85$, $c = 0.99$, $c = 1$) to obtain $\hat{y}(c)$.
- Step 3: *Project* $\hat{y}(c)$ into the nonnegative cone and do l^1 *normalization.*
- Step 4: Apply *Iterative Refinement* by classical procedures. Since $c \approx 1$, it is advisable to use preconditioning and Krylov techniques, see [17].

We finally remark that the complex Google setting implicit in Sections 4-6.2 is useful not only for matrix theoretic purposes, but also for computation; all the needed formulae (also those in Theorem 7.1, see also [37]) are well defined for c in the open unit disk and in a proper disk around $c = 1$. In fact it will be interesting to see whether an algorithm that exploits complex parameters will work well in practice and will enhance the numerical stability as expected. The results of numerical experiments for n of moderate size have been promising. See also [14] for a successful numerical experimentation with real parameters.

A second simpler and maybe more promising possibility comes from looking at

the power series in (6.27). The idea is the following: we can zero out the first-order term by forming $y(+\gamma) + y(-\gamma)$. We can zero out the first and second order terms by forming this sum with γ replaced by $\pm i\gamma$, and so on. This looks appealing, but the practical problem is that it requires solution of some large linear systems in a parameter range where the power method diverges. The equations certainly have solutions and can be computed by Krylov techniques (see [17]), but they cannot be obtained by the power method.

8.1. Comments on the "ideal" PageRank vector $\tilde{y}$. First we look at the PageRank problem as an ill-posed problem and we draw some analogies with another famous case of ill-posedness, i.e., the image restoration problem [3, Chapter 1]. Then we provide an interpretation on the vector $\tilde{y}$, the limit as c tends to 1 of our regularized solutions $y(c)$.

When one considers the pure Google matrix with $c = 1$, i.e., problem (1.2), finding the PageRank (that is, a nonnegative left 1-eigenvector whose entries sum to one) is an ill-posed problem (according to Hadamard [19, Section 2, p. 31]): infinitely many solutions exist and they can all be described as convex combinations of basic nonzero, nonnegative vectors $Z[i]$, $i = 1, \ldots, m$ [37, Section 4], where m is the multiplicity of the eigenvalue 1 of G, i.e., the number of irreducible components of the Markov chain represented by G (see e.g. [22]). These basic vectors are somehow *local* or sparse in the sense that they have a huge number of zero entries: in fact, the reason of such a locality relies on the fact that any $Z[i]$, $i = 1, \ldots, m$, is associated to a single irreducible component of G. On the other hand, when we consider instead $G(c)$ with a parameter $c \in [0, 1)$ (or c in the complex open unit disk), we make a sort of regularization that forces stability of the associated numerical problem and uniqueness of the solution. Furthermore, just as in the image restoration problem, our ill-posed problem requires nonnegativity of the solution: in this direction, we may ask if classical procedures used to solve the image restoration problem can be adapted to the PageRank computational problem. Indeed, concerning the algorithm sketched in the Section 8, we already exploited this similarity in the regularization Step 1 and in the limit process in Step 2, while we borrowed Step 3 again from standard image restoration techniques. Pushing further this reasoning, we may ask in addition if the SPAM pages [31, Section 9.2] can be considered as a noise disturbance, whose effect has to be diminished or eliminated.

Finally let us briefly mention some features of the vector $\tilde{y}$. Indeed, in the limit as c tends to 1, we obtain a special convex combination of nonnegative solutions, but it is much *less* local: it has a larger support (i.e. the set of indices related to nonzero entries), which clearly depends on the personalization vector v, since $\tilde{y} = Nv$ with N being the Cesaro averaging projector. For the modeler, this is a good thing, since all of the Web is taken into account, not just a smaller irreducible subset as in the local vectors $Z[i]$, $i = 1, \ldots, m$. The nature of the dependence of the support on v is not yet completely understood and deserves further investigation. However, even the vector $\tilde{y}$ in the real Web shows still a huge number of components with zero ranking; not only this, but many of these pages with zero PageRank are quite important according to common sense, see [8] and the discussion and the new proposals of Section 2.

8.2. A plain alternative for computing $\tilde{y} = \lim_{c \to 1} y(c)$. Here we make a plain algebraic modification of the matrix G in such a way that the set of solutions identified by (1.2) remains the same, but the power method converges unconditionally.

The main idea is to modify the row stochastic Google matrix G via a convex sum

with the matrix I and more precisely for $\delta \in (0,1)$ we set

$$G_\delta = \delta G + (1-\delta)I.$$

We apply the power method to this new matrix G_δ, that has $\lambda_1(\delta) = 1$ as spectral radius and eigenvalue of (geometric and algebraic) multiplicity m and $|\lambda_j(\delta)| < 1$ for every $m+1 \le j \le n$. Therefore we will observe convergence with an asymptotical rate given by

$$\max_{j \in \{m+1,\ldots,n\}} |\lambda_j(\delta)| < 1. \tag{8.1}$$

Of course the strictly dominating eigenvalue 1 will have an algebraic multiplicity and geometrical multiplicity $m \ge 1$ as in (1.2). So the power method will give back an eigenvector that is function of the initial choice x_0, but, may be surprisingly, not depending on the parameter δ.

An important question arises: how to choose x_0 for which the solution of the power method applied to G_δ coincides with $\tilde{y} = \lim_{c \to 1} y(c)$?

The interesting fact is that $G_\delta = N^T \oplus R_\delta$ where any eigenvalue of R_δ is of the form $1 - \delta + \delta\lambda_j$, $j = m+1, \ldots, n$, N is the nonnegative projector given in (7.6) and previously described, and the λ_j's are the eigenvalues of the pure Google matrix G. We know that $|\lambda_j| \le 1$ and $\lambda_j \ne 1$ for $j = m+1, \ldots, n$. Hence for any $\delta \in (0,1)$ we have $|1 - \delta + \delta\lambda_j| < 1$ for $j = m+1, \ldots, n$. As a consequence the unique solution of the power method applied to G_δ^T with starting vector x_0 is exactly Nx_0. We notice that if x_0 is strictly positive then every iterate is also strictly positive but many of the entries of the limit vector could be zero. Therefore for computing numerically $\tilde{y} = \lim_{c \to 1} y(c) = Nv$ it is sufficient to set $\delta \in (0,1)$ and to apply the power method to G_δ^T with initial guess v. As already observed the convergence is unconditional, but the speed of convergence depends on δ. In conclusion a proper choice of δ for maximizing the convergence rate of the power method is an interesting issue that we discuss in the next subsection.

8.3. Rate of convergence of the power method. As we will see the matrix G_δ with eigenvalues $\lambda_j(\delta)$ is such that the power method shows a rate of convergence given by (8.1): now we allow the value $\delta = 1$, i.e., we consider $\delta \in (0,1]$. The question is which eigenvalue $\lambda_j(\delta)$ is of maximal modulus for $j \in \{m+1, \ldots, n\}$, with n size of G_δ, and how to choose δ in order to maximize the rate of convergence. We know that for $\delta = 1$ every eigenvalue $\lambda_j(1) = \lambda_j$ that lies on the unit circle in the complex plane has a maximal modulus. In this case there is no convergence since $|\lambda_j| = |e^{i\varphi}| = 1$ for some $j \ge m+1$, $\varphi \in \mathbb{R}$.

In general for $\lambda \in \{\lambda_{m+1}, \ldots, \lambda_n\}$ we set $\lambda = re^{\mathbf{i}\varphi}$ and then, since $\lambda_j \ne 1$ for $j \in \{m+1, \ldots, n\}$ (i.e. we cannot have simultaneously $r = 1$ and $\cos(\varphi) = 1$), we find

$$\lambda_j(\delta) = \delta\lambda_j + 1 - \delta = \delta r\cos(\varphi) + 1 - \delta + \mathbf{i}\delta r\sin(\varphi).$$

Hence

$$\begin{aligned}
|\lambda_j(\delta)|^2 &= \delta^2 r^2\cos^2(\varphi) + \delta^2 r^2\sin^2(\varphi) + (1-\delta)^2 + 2\delta r(1-\delta)\cos(\varphi) \\
&= \delta^2 r^2 + (1-\delta)^2 + 2\delta r(1-\delta)\cos(\varphi) \\
&\le_{\text{setting } \cos(\varphi)=1} \delta^2 r^2 + (1-\delta)^2 + 2\delta r(1-\delta) \\
&= (\delta r + 1 - \delta)^2 \\
&\le_{\text{setting } r=1} (\delta + 1 - \delta)^2 = 1
\end{aligned}$$

and where equality to 1 is impossible since we cannot have at the same time $r = 1$ and $\cos(\varphi) = 1$. This proves the unconditioned convergence of the power method for $\delta \in (0, 1)$.

This result can be seen also graphically (ref. Fig. 8.1) since for $\delta \in (0, 1)$ all the eigenvalues $\lambda_j(\delta)$ with $j \in \{m+1, \ldots, n\}$ lie in the disc with boundary given by $\mathcal{C}_\delta/\{1\}$.

Now the question is how to maximize the rate of convergence i.e. how to choose $\delta \in (0, 1]$ for minimizing $s(\delta) = \max_{j\in\{m+1,\ldots,n\}} |\lambda_j(\delta)|$. This translates into a typical min-max problem:

$$\hat{g} \equiv \min_{\delta\in(0,1]} \max_{j\in\{m+1,\ldots,n\}} \delta^2 r_j^2 + (1-\delta)^2 + 2\delta r_j(1-\delta)\cos(\varphi_j) \tag{8.2}$$

with $\lambda_j(1) = \lambda_j = r_j e^{\mathbf{i}\varphi_j}$. Indeed for $\delta = 0$, $G_\delta = I$ and therefore $s(0) = 1$ so that the minimum exists in the set $(0, 1]$ and is located in the open set $(0, 1)$ if, as it usually happens for large Web matrices, at least one r_j equals 1 for $j \geq m + 1$.

Looking at the function $\delta^2 r_j^2 + (1-\delta)^2 + 2\delta r_j(1-\delta)\cos(\varphi_j)$ as a function of the radius r_j we notice that it is increasing for $\delta \in (0, 1)$ and $\cos(\varphi_j)$; moreover, setting $x_j = r_j \cos(\varphi_j)$ the real part of λ_j, the same function $\delta^2 r_j^2 + (1-\delta)^2 + 2\delta(1-\delta)x_j$ as a function of x_j is increasing again for $\delta \in (0, 1)$. Therefore, if $\bar{x}$ is the maximal real part of the eigenvalues λ_j, $j = m + 1, \ldots, n$, then it is evident that $|\bar{x}| < 1$ and

$$f_{\bar{x}}(\delta) \equiv \delta^2 + (1-\delta)^2 + 2\delta(1-\delta)\bar{x} \geq \max_{j\in\{m+1,\ldots,n\}} \delta^2 r_j^2 + (1-\delta)^2 + 2\delta r_j(1-\delta)\cos(\varphi_j)$$

so that, by minimizing $f_{\bar{x}}(\delta)$ with respect to δ, we find $\delta_{\text{opt}} = 1/2$ and the upper-bound

$$\hat{g} \leq f_{\bar{x}}(1/2) = \frac{1}{2}(1 + \bar{x}) \in [0, 1). \tag{8.3}$$

Finally, if the eigenvalue of G_δ coming from that of G with maximal real part is the one maximizing $\delta^2 r_j^2 + (1-\delta)^2 + 2\delta(1-\delta)x_j$ over $j = m + 1, \ldots, n$, then we can give interesting lower bounds that is

$$\hat{g} \geq \frac{1}{4}(1 + 2\bar{x} + \bar{r}^2) \geq \frac{1}{4}(1 + \bar{x})^2 \tag{8.4}$$

and

$$\hat{g} \geq \frac{\bar{r}^2 - \bar{x}^2}{1 + \bar{r}^2 - 2\bar{x}}, \tag{8.5}$$

where $\bar{r} \in [\bar{x}, 1]$ is the modulus of the eigenvalue with real part equal to $\bar{x}$ and $\bar{\varphi}$ its angle.

This last relation it is obtained evaluating

$$\frac{\partial}{\partial\delta}\left(\delta^2\bar{r}^2 + (1-\delta)^2 + 2\delta\bar{x}(1-\delta)\right) = 0 \tag{8.6}$$

and substituting the result

$$\delta = \frac{1-\bar{x}}{1 + \bar{r}^2 - 2\bar{x}} \quad \text{for} \quad \cos(\bar{\varphi})^2 \neq 1 \tag{8.7}$$

in the square modulus of the eigenvalue with maximal real part, that becomes exactly $(\bar{r}^2 - \bar{x}^2)/(1 + \bar{r}^2 - 2\bar{x})$.

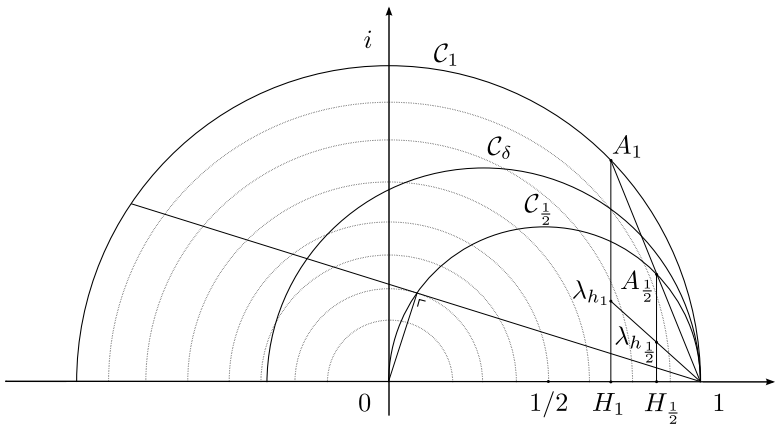


FIG. 8.1. *Geršgorin circles of G_δ for $\delta \in [0,1]$*

These results can also be represented graphically. We start plotting in the complex plane $\mathcal{C}_1$ i.e. the upper half boundary of the Geršgorin region of matrix G (ref. Fig. 8.1), since G is real we have specularity with respect to the real axis. If we consider the matrix G_δ for $\delta \in (0,1)$ we get the circle $\mathcal{C}_\delta$. Finally, when $\delta = 0$ the circle collapses in the point 1 on the real axis.

We can observe that for $\delta \in (0,1)$ the trajectory of the generic eigenvalue of G_δ, $\lambda_j(\delta)$ with $j \in \{m+1,\ldots,n\}$, is given by the convex combination of the two vectors $[\,r_j\,\cos(\varphi_j), r_j\,\sin(\varphi_j)\,]^T$ and $[\,1,0\,]^T$. It is plain that the minimal modulus of any of these eigenvalues is achieved for the unique δ such that the single trajectory intersects the circumference of radius $1/2$ centered in $(1/2,0)$. This coincides with $\mathcal{C}_{1/2}$, i.e. boundary of the Geršgorin region associated to G_δ for $\delta = 1/2$, that is given by the points satisfying the relation $r = \cos(\varphi)$.

As δ varies in the interval $[0,1]$, it is straightforward that the relative position of the eigenvalues does not change. This means that if we suppose to know, for a particular value of $\delta \in (0,1]$, the eigenvalue with maximal real part, $\lambda_h(\delta)$ with $h \in \{m+1,\ldots,n\}$, this will remain always the one with maximal real part for every $\delta \in [0,1]$.

Interesting enough we observe that given $\delta \in [0,1/2]$ the problem stated in (8.2) (i.e. minimize, with respect to δ, the maximal modulus of the eigenvalues $\lambda_j(\delta)$ of G_δ for $j \in \{m+1,\ldots,n\}$) becomes simply $\hat{g} \equiv \max_{j\in\{m+1,\ldots,n\}} 1/4[r_j^2 + 1 + 2r_j\cos(\varphi_j)]$ since for δ decreasing from $1/2$ to 0 every eigenvalue increases its modulus. Hence we can restate (8.2) as

$$\hat{g} \equiv \min_{\delta\in\left[\frac{1}{2},1\right]} \max_{j\in\{m+1,\ldots,n\}} \delta^2 r_j^2 + (1-\delta)^2 + 2\delta r_j(1-\delta)\cos(\varphi_j) \tag{8.8}$$

Furthermore, for every $\delta \in (0,1)$, eigenvalues different from 1 with maximal distance from the origin are the ones on the circumference $\mathcal{C}_\delta$ and among these the greatest are the ones with maximal real part.

Now, assuming that we know the eigenvalue $\lambda_h(1)$, except 1 of course, with maximal real part $\bar{x} = \bar{r}\cos(\bar{\varphi})$. We draw in the complex plane the vertical line that passes

through this eigenvalue. This line intersects the circumference $\mathcal{C}_1$ in A_1 and the real axis in H_1. For a generic δ we do the same and we get the points A_δ and H_δ with real part equal to $\delta\bar{x} + 1 - \delta$. In particular $A_{1/2}$ has a minimal distance from the origin among all the possible A_δ. The previous considerations allow us to state that the quantity $\hat{g}$ will be bounded by

$$\left|\overline{OH}_{1/2}\right|^2 \leq \hat{g} \leq \left|\overline{OA}_{1/2}\right|^2 \tag{8.9}$$

i.e. relations (8.3) and (8.4).

If we suppose that the trajectory, in function of δ, of the eigenvalue with maximal real part λ_h intersects the circumference $\mathcal{C}_{1/2}$ in the point B, we can rewrite relation (8.5) as

$$\hat{g} \geq \left|\overline{OB}\right|^2 \tag{8.10}$$

Hence if λ_h lies inside the circle $\mathcal{C}_{1/2}$ the above relation becomes simply $\hat{g} \geq \bar{r}^2$. In fact if we impose in (8.7) that $\delta \in [0, 1]$ we get $\bar{r} \geq \cos(\bar{\varphi})$ i.e. the relation (8.5) is valid when λ_h lies outside circle $\mathcal{C}_{1/2}$.

We observe that in the light of the model proposed in Section 2, all these reasonings hold. We add only that in this last case we do not have roots of unity among the λ_j for $j \in \{m+1, \ldots, n\}$ since the graph associated to the matrix G can be reduced into the direct sum of irreducible and primitive blocks. This implies that relation (8.3) becomes a strict inequality and that the power method will converge even in the case of $\delta = 1$, but of course not necessarily with a maximal rate of convergence.

Furthermore, we observe that $G_\delta = \delta G + (1-\delta)I$, $\delta \in (0, 1]$, is a linear polynomial of G with the condition that the eigenvalue 1 is a fixed point of the transformation and the coefficients are nonnegative. If instead of G_δ we consider any polynomial $p(G)$ of G with nonnegative coefficients and such that $p(1) = 1$, then we could have a larger degree of freedom for maximizing the convergence rate but, of course, the already difficult min-max problem (8.2) would become analytically very intricate. This and other issues such as a more careful study of the min-max problem (8.2) will be the subject of future researches.

9. Some comments about prior work. The eigenvalues of the standard real parametric Google matrix $G(c)$ were analyzed by Haveliwala and Kamvar [22] (only the second eigenvalue), Eldén [18], and Langville and Meyer [31] (their proof is the same as that of Reams). A different approach via the characteristic polynomial is suggested in [34, Problem 7.1.17, p. 502]. These authors were apparently unaware of the prior work of Brauer [11] and Reams [36].

Relying on sophisticated results about Markov chains, [37] gives an analysis of the Jordan canonical form of the standard real $G(c)$; it also gives a rational representation for $y(c)$ and computes its limit as $c \to 1$, again in the standard real case only. The Maclaurin series for $y(c)$ was studied in [8], where the partial sums of (6.26) for nonnegative real v and $0 < c < 1$ were identified as the iterates obtained in solving $y^T G(c) = y^T$ with the power method starting at v. Finally, comparing our findings with the results in [8, 29], one important message of the present paper is that the point $c = 1$ is not a singularity point for $y(c)$, and hence limits and conditioning of $y(c)$ can be derived and safely handled, both in theory and in practical computations.

10. Concluding remarks and future work. As a final remark, we stress that the analysis of our matrix-theoretic oriented approach is also valid for the modified

enhanced models proposed e.g. in [1, 39] etc. or discussed here in Section 2. Indeed the interest in the general matrix-theoretic analysis relies on its level of adaptability. In fact the results and the conclusions of Sections 4–7 are virtually unchanged if one considers a different way of handling nodes or if one allows self-links giving raise to a different definition of $G(1)$. Moreover, in this context we must no forget that there exist completely different applications [6, 28] including dynamical agents theory [33, 41], where the idea and the computational suggestions in Section 8 have a lot of potential to be further developed and studied.

REFERENCES

[1] R. Baeza-Yates, P. Boldi, C. Castillo *Generalizing PageRank: Damping Functions for Link-Based Ranking Algorithms*, Proc. ACM SIGIR 2006, Seattle, Washington, USA, August 2006, ACM Press, pp. 308-315.

[2] A. Berman, R. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. Classics in Appl. Math., SIAM Publ., Philadelphia, PA, 1994.

[3] M. Bertero, P. Boccacci, *Introduction to Inverse Problems in Imaging*. Inst. of Physics Publ. Bristol and Philadelphia, London, 1998.

[4] D. Bini, G. Del Corso, F. Romani *A Google-like algorithm for ranking scientific publications* presentation at the Workshop *Numerical Linear Algebra in Internet Algorithms* (Monopoli - Bari (Italy), September 11th–15th 2007).

[5] D. Bini, B. Iannazzo, B. Meini, G. Latouche, *On the solution of Riccati equations arising in fluid queues*, Linear Algebra Appl., 413-2/3 (2006) pp. 474–494.

[6] D Bini, G. Latouche, B. Meini, *Numerical Methods for Structured Markov Chains*. Oxford University Press, Oxford, 2005.

[7] D. Bini, B. Meini, I. M. Spitkovsky, *Shift techniques and canonical factorizations in the solution of M/G/1-type Markov chains*, Stochastic Models, 21-2/3 (2005) pp. 279–302.

[8] P. Boldi, M. Santini, S. Vigna, *PageRank as a function of the damping factor*, Proc. 14th Intenational World Wide Web Conference (WWW 2005), Chiba, Japan 2005, ACM, pp. 557–566.

[9] P. Boldi, Violetta Lonati, M. Santini, S. Vigna, *Graph fibrations, Graph isomorhism, and PageRank*, RAIRO Inform. Thor., 40 (2006) pp. 227–253.

[10] A. Böttcher, B. Silbermann, *Introduction to Large Truncated Toeplitz Matrices*. Springer-Verlag, New York, NY, 1999.

[11] A. Brauer, *Limits for the characteristic roots of a matrix. IV: Applications to stochastic matrices*, Duke Math. J., 19 (1952) pp. 75–91.

[12] C. Brezinski, M. Redivo Zaglia, *Extrapolation Methods. Theory and Practice*. North-Holland, Amsterdam, 1991.

[13] C. Brezinski, M. Redivo Zaglia, *The PageRank vector: properties, computation, approximation, and acceleration*, SIAM J. Matrix Anal. Appl., 28 (2006) pp. 551–575.

[14] C. Brezinski, M. Redivo Zaglia, *Rational extrapolation for the PageRank vector*, Math. Comput., 77 (2008) pp. 1585–1598.

[15] C. Brezinski, M. Redivo Zaglia, S. Serra-Capizzano, *Extrapolation methods for PageRank computations*, Comptes Rendus Acad. Sci. Paris - Ser. I, 340 (2005) pp. 393–397.

[16] S. Campbell, C. Meyer, *Generalized Inverses of Linear Transformations*. Dover Publications, New York, 1979.

[17] G. Del Corso, A. Gullí, F. Romani, *Fast PageRank computation via a sparse linear system*, Internet Math., 3-2 (2005), pp. 259–281.

[18] L. Eldén, *The eigenvalues of the Google matrix*, Report LiTH-MAT-R-04-01, Department of Mathematics, Linkping University, Sweden, 2003.

[19] H. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht, 1996.

[20] G. Golub, C. Greif, *An Arnoldi-type algorithm for computing PageRank*, BIT, 46 (2006), pp. 759–771.

[21] G. Golub, C. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 1983.

[22] H. Haveliwala, S.D. Kamvar, *The second eigenvalue of the Google matrix*, Stanford University (2003).

[23] R. Horn, C. Johnson, *Matrix Analysis*. Cambridge University Press, New York, 1985.

[24] R. Horn, S. Serra-Capizzano, *A general setting for the parametric Google matrix*, Internet Math., 3-4 (2008) pp. 385–411.

[25] B. Iannazzo, private communication (2007).

[26] I. Ipsen, S. Kirkland, *Convergence analysis of a PageRanking updating algorithm by Langville and Meyer*, SIAM J. Matrix Anal. Appl., 27-4 (2006) pp. 952–967.

[27] S. Kamvar, H. Haveliwala, C. Manning, G. Golub, *Extrapolation methods for accelerating PageRank computations*, Proc. 12th International World Wide Web Conference (WWW 2003), Budapest, Hungary 2003, ACM, pp. 680–689.

[28] R. Kerner, *Models of Agglomeration and Glass Transition*. Imperial College Press, 2006.

[29] S. Kirkland, *Conditioning of the entries in the stationary vector of a Google-type matrix*, Linear Algebra Appl., 418 (2006) pp. 665–681.

[30] J. Lasserre, *A formula for singular perturbations of Markov Chains*, J. Appl. Prob., 31-2 (1994)

pp. 829–833.
[31] A. Langville, C. Meyer, *Deeper inside PageRank*, Internet Math., 1-3 (2005) pp. 335–380.
[32] A. Langville, C. Meyer, *A survey of eigenvector methods for WEB information retrieval*, SIAM Review, 47-1 (2005) pp. 135–161.
[33] N. Lynch, *Distributed Algorithms*. Morgan Kaufmann, San Mateo, CA, 1997.
[34] C. Meyer, *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, 2000.
[35] L. Page, S. Brin, R. Motwani, T. Winograd, *The Page-Rank citation ranking: bringing order to the web*, Stanford Digital Libraries Working Paper, 1998.
[36] R. Reams, *An inequality for nonnegative matrices and the inverse eigenvalue problem*, Linear Multilinear Algebra, 41 (1996) pp. 367–375.
[37] S. Serra-Capizzano, *Jordan canonical form of the Google matrix: a potential contribution to the PageRank computation*, SIAM J. Matrix Anal. Appl., 27 (2005) pp. 305–312.
[38] S. Serra-Capizzano, *Google PageRanking Problem: The Model and the Analysis* presented at the Dagstuhl-Seminar 07071 *Web Information Retrieval and Linear Algebra Algorithms* (Schloss Dagstuhl-Wadern (Germany), February 11th–16th 2007) and at the Workshop *Numerical Linear Algebra in Internet Algorithms* (Monopoli - Bari (Italy), September 11th–15th 2007): http://drops.dagstuhl.de/opus/volltexte/2007/1069/pdf/07071.SerraCapizzanoStefano.Paper.1069.pdf
[39] S. Vigna, *TruRank: taking PageRank to the limit*, Proc. 14th International World Wide Web Conference (WWW 2005), Special Interest Tracks & Posters, Chiba, Japan 2005, ACM, pp. 976–977.
[40] Y. Wei, G. Wu, *Comments on "Jordan Canonical Form of the Google Matrix"*, SIAM J. Matrix Anal. Appl., 30-1 (2008) pp. 364–374.
[41] L. Xiao, S. Boyd, *Fast linear iterations for distributed averaging*, Syst. Control Letters, 53-1 (2004) pp. 65–78.