PageRank: Splitting Homogeneous Singular Linear Systems of Index One

Douglas V. de Jager and Jeremy T. Bradley

Department of Computing, Imperial College London,
180 Queen's Gate, London SW7 2BZ, United Kingdom

Abstract. The PageRank algorithm is used today within web information retrieval to provide a content-neutral ranking metric over web pages. It employs power method iterations to solve for the steady-state vector of a DTMC. The defining one-step probability transition matrix of this DTMC is derived from the hyperlink structure of the web and a model of web surfing behaviour which accounts for user bookmarks and memorised URLs.
In this paper we look to provide a more accessible, more broadly applicable explanation than has been given in the literature of how to make PageRank calculation more tractable through removal of the dangling-page matrix. This allows web pages without outgoing links to be removed before we employ power method iterations. It also allows decomposition of the problem according to irreducible subcomponents of the original transition matrix. Our explanation also covers a PageRank extension to accommodate TrustRank. In setting out our alternative explanation, we introduce and apply a general linear algebraic theorem which allows us to map homogeneous singular linear systems of index one to inhomogeneous non-singular linear systems with a shared solution vector. As an aside, we show in this paper that irreducibility is not required for PageRank to be well-defined.

1 Introduction

The PageRank metric is a widely-used hyperlink-based estimate of the relative importance of web pages [1, 2]. The standard algorithm for determining PageRank uses power method iterations to solve for the steady-state vector of a DTMC. The one-step probability transition matrix which defines the DTMC is derived from a web graph that reflects the hyperlink structure of the web and a user-centred model of web-surfing behaviour. The model provides a mathematical account of how users make use of web bookmarks and also of what they do when at web pages without any outgoing links.

In July of 2008, it was announced by Google that the company's crawlers had found more than one trillion current unique web pages, and that the web was growing at several billion new pages every day [3]. This is in contrast with

Google's index of 26 million web pages in 1998 [3]. Despite its size, Google's index is only a fraction of the total number of indexable web pages in existence. This is because many sites are currently difficult to index—because of technologies like JavaScript and Flash, and also because certain sites require appropriate handling of forms, drop-downs, and so on. Also, by design, certain sites prevent access by Google's crawlers to many of their constituent web pages through robots.txt and nofollow links.

In [4] an investigation was presented into this rapid growth of the web. It was argued that the recent acceleration of growth has been driven in particular by a growing percentage of web pages without outgoing links—upward of fifty percent—and an analysis was provided into the different sorts of web pages which are classed by search engines as having no outgoing links.

In this paper we show how to reduce the complexity of PageRank calculation by partitioning the treatment of web pages with and without outgoing links, such that only pages with outgoing links are required during the power method iterations. As an added benefit, this approach also permits decomposition of the PageRank problem according to connected subcomponents of the original transition matrix [5]. We show this by presenting PageRank as a special case of a broader class of problem. Our proposal is an alternative formulation of linear algebraic proposals made in [6, 7]—which are extensions of a lumping proposal in [8]. In this paper we also consider a PageRank extension considered via lumpability theory in [9, 10], which allows for TrustRank [11]. We show that this extension is also a special case of the same broader class of problem, and that it can be handled similarly, in linear algebraic fashion. We suggest treating these proposals as companions to work in two other key research areas, research which focuses specifically on the size of the PageRank problem. The two research areas are asynchronous solution methods (where problem size requires the use of hetergeneous computing clusters) [12] and partitioning techniques for the PageRank problem across multiple processors [13, 14, 5].

The kernel of our proposal is a linear algebraic theorem which allows homogeneous singular linear systems of index one to be mapped to inhomogeneous nonsingular linear systems with a shared solution vector. This theorem is a general one. Its formulation was inspired by the PageRank equation. However, the theorem is not restricted in applicability to PageRank; nor indeed to DTMCs. The theorem may be used to apply novel solution methods to eigenvector problems (for example, asynchronous methods which require the spectral radius of the modulus equivalent of the coefficient matrix to be strictly less than one [15]), or it may be used to improve sparsity patterns or conditioning when using traditional solution methods to such problems.

As an aside to the core proposal, we show that, contrary to the standard presentation of PageRank in the literature, irreducibility is not required for PageRank to be well-defined. We show, in particular, that the personalisation vector needs not to be completely dense.

The paper is organised as follows. In Section 2 we review the conceptual model for PageRank. We set out what is now regarded as the standard definition of PageRank, and we present its sparse formulation. In Section 3 we extend the standard PageRank definition to allow for non-dense personalisation vectors. In Section 4 we consider web pages without outgoing links. We introduce a theorem which allows us to map homogeneous singular linear systems of index one to inhomogeneous non-singular linear systems with a shared solution vector. Using this theorem we show how to employ the original transition matrix as a coefficient matrix when solving for standard PageRank—without an adjustment to deal with pages without outgoing links. We then extend our approach to deal with a generalisation of the PageRank definition which accounts for TrustRank.

2 Standard PageRank Definition

PageRank computation for the ranking of hypertext-linked web pages was originally outlined by Page and Brin [1, 2]. Their approach was subsequently amended by Kamvar *et al.* [16]. This alternative formulation of PageRank and its computation is now generally regarded as providing the standard PageRank definition [17, 18].

The standard conceptual model of PageRank is called the *random surfer* model. Consider a surfer who starts at a web page and picks one of the links on that page at random. On loading the next page, this process is repeated. If a *dangling page* (that is, a page without outgoing links—also referred to as a *cul de sac page*) is encountered, then the surfer chooses to visit a random page (as though going to a memorised link, or a bookmarked link). During normal browsing, the user may also decide, with a fixed probability, not to choose a link from the current page, but instead to jump at random to another page. In the latter case, to support both unbiased and personalised surfing behaviour, the model allows for the specification of a probability distribution of target pages.

The PageRank of a page is considered to be the limiting (steady-state) probability that the surfer is visiting a particular page after a large enough number of click-throughs. Calculating this probability vector corresponds to finding a dominant eigenvector of the modified web-graph transition matrix.

2.1 Random Surfer Model

In the random surfer model, the web is represented by a graph $G = (V, E)$, with web pages as the vertices, V, and the links between web pages as the edges, E. If a link exists from page u to page v then $(u \to v) \in E$.

To represent the following of hyperlinks, we construct a transition matrix P from the web graph, setting:

$$P_{ij} = \begin{cases} \frac{1}{\deg(u_i)} & : \text{ if } (u_i \to u_j) \in E \\ 0 & : \text{ otherwise} \end{cases} \tag{1}$$

where $\deg(u)$ is the out-degree of vertex u, i.e. the number of outbound links from page u. From this definition, we see that if a page has no out-links, then this corresponds to a zero row in the matrix P. To represent the surfer's jumping from dangling pages, we construct a second matrix $D = \boldsymbol{d}\boldsymbol{p}^T$, which we refer to as the *dangling-page* matrix, where $\boldsymbol{d}$ and $\boldsymbol{p}$ are both column vectors, and

$$d_i = \begin{cases} 1: & \text{if } \deg(u_i) = 0 \\ 0: & \text{otherwise} \end{cases} \tag{2}$$

and $\boldsymbol{p}$ is the personalisation vector representing the probability distribution of destination pages when a random jump is made. Typically this distribution is taken to be uniform, i.e. $p_i = \frac{1}{n}$ for an n-page graph $(1 \leq i \leq n)$. However, it need not be, as many distinct personalisation vectors may be used to represent different classes of user with different web browsing patterns. This flexibility comes at a cost, though, as each distinct personalisation vector requires an additional PageRank calculation.

Putting together the surfer's following of hyperlinks and their random jumping from dangling pages yields the stochastic matrix $P' = P + D$, where P' is a one-step probability transition matrix of a DTMC.

To represent the surfer's decision not to follow any of the current page links, but to jump instead to a random web page, we construct a *teleportation* matrix E, where $E_{ij} = p_j$ for all i, i.e. this random jump is also dictated by the personalisation vector.

Incorporating this matrix into the model gives:

$$A = cP' + (1 - c)E \tag{3}$$

where $0 < c < 1$, and c represents the probability that the user chooses to follow one of the links on the current pages—i.e. there is a probability of $(1-c)$ that the surfer randomly jumps to another page instead of following links on the current page.

This definition of A avoids two potential problems. The first is that P', although a valid DTMC transition matrix, is not necessarily irreducible and aperiodic. Taken together, these are a sufficient condition for the existence of a unique steady-state distribution [16, 18]. Now, provided $p_i > 0$ for all $1 \leq i \leq n$, irreducibility and aperiodicity are trivially guaranteed.

The second problem relates to the rate of convergence of power method iterations used to compute the steady-state distribution. This rate depends on the reciprocal of the modulus of the subdominant eigenvalue (λ_2). For a general P', $|\lambda_2|$ may be very close to 1, resulting in a very poor rate of convergence. However, it has been shown that in the case of matrix A, $|\lambda_2| \leq c$, thus guaranteeing a good rate of convergence for the widely taken value of $c = 0.85$ [19].

Given the matrix A, we can now define the unique PageRank vector, $\boldsymbol{\pi}$, to be the steady-state vector or the dominant eigenvector that satisfies:

$$\boldsymbol{\pi} A = \boldsymbol{\pi} \tag{4}$$

2.2 Sparse PageRank Definition

Having constructed matrix A we might naïvely attempt to find the PageRank vector of Equation (4) by directly using a power method approach:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} A \tag{5}$$

where $\boldsymbol{x}^{(k)}$ is the k^{th} iterate towards the PageRank vector, $\boldsymbol{\pi}$. However, the web was known to have more than a trillion unique pages in 2008, with several billion new pages being added to this total every day, so it is clear that this is not a practical approach for realistic web graphs [3]. The reason for this is that A is a completely dense matrix, on account of the completely dense teleportation matrix E.

Given the teleportation-matrix density concern, a sparse reduction of the standard equation Equation (4) is typically employed in calculations [18]. The reduction is as follows:

$$\begin{aligned} \boldsymbol{\pi} &= \boldsymbol{\pi} A \\ &= \boldsymbol{\pi}(cP' + (1-c)E) \\ &= c\boldsymbol{\pi}P' + (1-c)\boldsymbol{\pi}E \\ &= c\boldsymbol{\pi}P' + (1-c)\sum_i \boldsymbol{\pi}_i \boldsymbol{p} \\ &= c\boldsymbol{\pi}P' + (1-c)\boldsymbol{p} \end{aligned} \tag{6}$$

where P' is more sparse than the original matrix A.

3 Irreducibility is not required

As given above, it is regularly written in the literature that irreducibility is required for Equation (4) to be well-defined, with a unique steady-state vector. To ensure irreducibility, it is written that a completely dense personalisation vector is required.

Kamvar *et al.* [16] write, “If [the Google matrix] is aperiodic and irreducible, then the Ergodic Theorem guarantees that the stationary distribution of the random walk is unique. In the context of computing PageRank, the standard way of ensuring that [the Google matrix] is irreducible is to add a new set of complete outgoing transitions, with small transition probabilities, to all nodes, creating a complete (and thus an aperiodic and strongly connected) transition graph.”

Langville *et al.* [6] write, "[To guarantee] the existence and uniqueness of the PageRank vector, Brin and Page added another rank-one update, this time an irreducibility adjustment in the form of a dense perturbation matrix [E] that creates direct connections between each page."

In the context of an unbiased representation of how we surf the web (without a back button), a completely dense personalisation vector makes sense. But this density requirement poses a difficulty when we start personalising PageRank. If we were to attempt to categorise users—for example, as avid consumers of sports news—or we were to attempt to model how a particular person, or group of people, surfs the web (as per the intuitive justification for PageRank), then it seems clear that we should allow for zero entries in the personalisation vector, zero entries which correspond to those pages to which the particular person will not teleport.

One might argue that this is not the case, that for any person there is a chance, albeit very small, that this particular person teleports to any web page. But it is difficult to justify intuitively why any personalised categorisation of users should necessarily have non-zero probability of teleporting to every web page. Equally, it is difficult to understand why there cannot be a personalised model of a person for which there is at least one zero personalisation vector entry. The argument for a completely dense personalisation vector seems to be based more on a need for theoretical well-definedness than on any force of intuition.

In this subsection, we recall a theorem from [20] from which it follows that complete density is not required for PageRank to be well-defined.

Let us remove the requirement that $\boldsymbol{p}$ be completely dense, and let us have instead that $\boldsymbol{p}$ is just a probability vector. Let R define the set of indices of non-zero entries, $\{p_t > 0 : 1 \leq t \leq n\}$. Now, if we consider Equation (6), then it is clear that independent of the structure of P, every page is connected to each and every page in set R. Indeed, by definition, these are the pages to which the surfer could teleport whilst on any other page. Accordingly, this set R forms part of a single irreducible component of recurrent pages.

Let us now consider those pages which are not part of R. Any such page is either transient, in which case the page has an outgoing path to some page in R but there is no reciprocal path back from pages in R, or such a page has both an outgoing path to a page in R and has a reciprocal path back from a page in R. In the latter case, the page forms part of an irreducible component of pages of which R is a subset.

Let us now recall the following from [20]:

Proposition 1 *Suppose that a DTMC with one-step probability transition matrix T has just one strongly connected subcomponent of recurrent states. This is equivalent to supposing that there is some state which is reachable from all other states—in matrix form, $\exists j \forall i \exists p (T^p_{ij} > 0)$. Then,*

1. *The transient states all have steady-state entries equal to zero.*
2. *The restriction of matrix T to the recurrent states (removing all rows and columns corresponding to transient states) is an irreducible probability transition matrix.*
3. *There is a corresponding unique steady-state distribution.*
4. *The recurrent states all have positive steady-state entries.*

From this proposition, it follows that we require only that $\boldsymbol{p}$ be a probability vector for PageRank to be well-defined.

4 Computing PageRank without Dangling Pages

4.1 Motivating PageRank Computation without Dangling Pages

Eiron *et al.* write that the number of dangling pages is higher than the number of non-dangling pages [4]. Langville and Meyer write that the number of dangling pages relative to non-dangling is growing, and that some sets of crawled pages show percentages of dangling pages reaching 80% [6].

The reason for this high percentage of dangling pages is two-fold: an increasing number of *pseudo*-dangling pages and an increasing number of *real* dangling pages.

Pseudo-dangling pages are pages which are treated as dangling pages because their outlinks have not (yet) been crawled. There are several reasons for this growing number of pages with uncrawled outlinks.

Firstly, over recent years there has been an ever increasing amount of dynamic content on the web, and also links to such content, and the rapid increase has left crawlers incapable of keeping up. Unlike static pages, which are hand-edited HTML, dynamic pages are database-driven. These dynamic pages are limited in number only by what is available in the database, and, potentially, not even then. For example, the number of pages given by a web calendar might be expected to be (nearly) infinite. Even in more mundane examples the size of the database may not provide an upper bound on the number of potential dynamic pages. For example, session IDs, timestamps, and so on, may further expand the potential number of dynamic pages, as ostensibly the same page is treated differently, because it has a different URL or a different embedded timestamp.

Two further reasons for the growing number of uncrawled links are robots.txt/ nofollow and JavaScript. Robots.txt and nofollow are conventions whereby website owners can demarcate certain parts of their sites as not to be crawled. The outlinked pages in a prohibited part of a site are typically still indexed according to anchor text, but as they are not actually crawled, they are treated as dangling pages [4].

JavaScript links are becoming more prevalent, particularly as part of dynamic AJAX websites. Such links are not evaluated by current search engines [21].

Real dangling pages are those from which there really are no outgoing links. These may be HTML pages without links. However, the main reason for the exploding number of *real* dangling pages is the recent push by the research community to move more and more material online: PDF, postscript and PPT files of papers, presentations, theses, and so on.

4.2 Dangling-Page Matrix Yields Scaling

In the literature, there are two linear algebraic treatments of PageRank whereby the dangling page matrix is removed from the PageRank definition. In [6], a somewhat *ad hoc* proof is presented to show that the dangling-page matrix serves only to scale the solution vector. The proof starts with the identity Equation (19). It then proceeds to show that by reformulating this identity we can get the linear algebraic form of Equation (15). It is *ad hoc* because it proceeds by assuming the given identity is the correct one. It gives no clues as to how one might discover this identity in the first instance. In the later [7], the same fact is proved by way of the Sherman-Morrison formula [22].

In this subsection we look to provide a more accessible explanation for the removal of the dangling-page matrix. Our proposal presents PageRank as a special case of a broader class of problem. The kernel of the explanation is the following theorem which allows us to map homogeneous singular linear systems of index 1 to inhomogeneous non-singular linear systems with the same solution vector.

Theorem 2 *Let us define matrix,* $V^{(K)} \in C^{n\times n}$ $(K \subseteq \{1, 2, \ldots, n\})$*:*

$$V_{ij}^{(K)} = \begin{cases} v_i & \text{if } j \in K \\ 0 & \text{if } j \notin K \end{cases}, \tag{7}$$

where $\boldsymbol{v} \in C^n$.

Now suppose that 1 is not an eigenvalue of $(M - V^{(K)})$*, for some* $M \in C^{n\times n}$*,* $\boldsymbol{v}$ *and* K.[1]

Then, we have the following: If 1 is an eigenvalue of M*, then it is a simple eigenvalue of* M *and there is a corresponding right eigenvector* $\boldsymbol{x}$ *of* M *which is the unique fixed-point of*

$$\boldsymbol{x} = (M - V^{(K)})\boldsymbol{x} + \boldsymbol{v}. \tag{8}$$

Proof: Suppose $\boldsymbol{y} = M\boldsymbol{y}$. Then,

$$\boldsymbol{y} = (M - V^{(K)})\boldsymbol{y} + (\sum_{k\in K} y_k)\boldsymbol{v}. \tag{9}$$

[1] Given an irreducible complex matrix, M, with unit spectral radius, an example of suitably chosen $V^{(K)}$ has $\boldsymbol{v}$ equalling the first column of M and $K = \{1\}$

The last equality holds because

$$V^{(K)}\boldsymbol{y} = \begin{pmatrix} v_1 \sum_{k\in K} y_k \\ \vdots \\ v_n \sum_{k\in K} y_k \end{pmatrix} = (\sum_{k\in K} y_k)\boldsymbol{v}. \tag{10}$$

Now, by supposition, 1 is not an eigenvalue of $(M - V^{(K)})$, so:

$$\sum_{k\in K} y_k \neq 0. \tag{11}$$

Also, for all $\alpha \neq 0$, we have the following:

$$\boldsymbol{z} := (M - V^{(K)})\boldsymbol{z} + \alpha\boldsymbol{v} = \alpha(I - (M - V^{(K)}))^{-1}\boldsymbol{v} \tag{12}$$

This is true because, as 1 is not an eigenvalue of $(M - V^{(K)})$:

$$(I - (M - V^{(K)})) \text{ is invertible} \tag{13}$$

Accordingly, the fixed-point $\boldsymbol{x}$ of

$$\boldsymbol{x} = (M - V^{(K)})\boldsymbol{x} + \boldsymbol{v} \tag{14}$$

is a scalar multiple of $\boldsymbol{y}$, and thus $\boldsymbol{x}$ is a right eigenvector of M which corresponds to eigenvalue 1. Further, given non-singularity, $\boldsymbol{x}$ is the unique fixed-point, and so 1 is a simple eigenvalue of M.

Let us now recall the standard PageRank definition:

$$\boldsymbol{\pi} = \boldsymbol{\pi}(c(P + D) + (1 - c)E) \tag{15}$$

We know from the Perron–Frobenius theorem together with continuity of spectral radius with respect to matrix entries that if we remove cD from the PageRank matrix to yield $(c(P) + (1 - c)E)$, then the resultant matrix does not have 1 as an eigenvalue. So, by Theorem 2:

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{x}(c(P + D) + (1 - c)E - cD) + \boldsymbol{p} \\ &= \boldsymbol{x}(c(P) + (1 - c)E) + \boldsymbol{p} \\ &= c\boldsymbol{x}P + (1 - c)\boldsymbol{x}E + \boldsymbol{p}, \end{aligned} \tag{16}$$

where $\frac{\boldsymbol{x}}{\sum_i x_i} = \boldsymbol{\pi}$.

Now, we know that

$$(1 - c)\boldsymbol{x}E = (1 - c)\sum_i x_i\boldsymbol{p} \tag{17}$$

where $\sum_i x_i > 0$.

So, from Equation (17), we have that

$$\boldsymbol{x} = c\boldsymbol{x}P + \alpha\boldsymbol{p} \tag{18}$$

where α is some positive real scalar.

If we rewrite this equation as an inhomogeneous non-singular linear system, it is clear that the α coefficient serves only to scale. This allows us to solve instead:

$$\boldsymbol{y} = c\boldsymbol{y}P + \boldsymbol{p} \tag{19}$$

where non-negative vector $\boldsymbol{y}$ is such that $\frac{\boldsymbol{y}}{\sum_i y_i} = \boldsymbol{\pi}$.

So, we have defined a scalar multiple of the PageRank vector $\boldsymbol{\pi}$ which makes no appeal to a dangling-page matrix.

Given the removal of the dangling-page matrix, we find in [6] an iterative procedure for removing dangling pages before we employ power iterations to solve for PageRank.

4.3 Different Dangling and Personalisation Vectors

In [9, 10], considerations are presented into generalisations of PageRank which allow the dangling-page vector to differ from the personalisation vector—to account, in particular, for TrustRank [11]. The generalised PageRank definition differs from the standard PageRank in that $D = \boldsymbol{d}\boldsymbol{g}^T$, where dangling-page vector, $\boldsymbol{g}$, is any probability vector. In these two papers, lumpability theory is used to show that, even in this generalised form of PageRank, the dangling-page matrix can be removed.

In this subsection we use Theorem 2 to provide an alternative, linear algebraic reformulation of this generalisation, to remove the dangling-page matrix. We also introduce another theorem, which allows another means of reformulating.

By Theorem 2, reasoning as before, though with the generalised equation,

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{x}(c(P + D) + (1 - c)E - cD) + \boldsymbol{w} \\ &= \boldsymbol{x}(c(P) + (1 - c)E) + \boldsymbol{w} \\ &= c\boldsymbol{x}P + (1 - c)\boldsymbol{x}E + \boldsymbol{w}, \\ &= c\boldsymbol{x}P + \alpha\boldsymbol{p} + \boldsymbol{w}, \end{aligned} \tag{20}$$

where α is a non-zero scalar.

Now, if we solve the two inhomogeneous nonsingular equations:

$$\boldsymbol{y} = c\boldsymbol{y}P + \boldsymbol{p}; \tag{21}$$

$$\boldsymbol{z} = c\boldsymbol{z}P + \boldsymbol{w}. \tag{22}$$

Then, α is easy to determine by substituting $\boldsymbol{x} = \alpha\boldsymbol{y} + \boldsymbol{z}$ into Equation 20.

In the above, we have shown that the dangling page matrix is not required when solving for generalised PageRank, and we did so by appealing to Theorem 2. An alternative approach would be to appeal to the following theorem. It allows us to split an inhomogeneous nonsingular linear system into two constituent systems. The proof is a simplification of the earlier proof.

Theorem 3 *Let matrix $V^{(K)}$ be defined as before in terms of a vector, $\boldsymbol{v}$, and a set of indices, K. Let $M\boldsymbol{x} = \boldsymbol{w}$ where M is non-singular and $\boldsymbol{w}$ is a vector. Then, if $(M - V^{(K)})$ is non-singular, then*

$$(M - V^{(K)})\boldsymbol{x} = \alpha\boldsymbol{v} + \boldsymbol{w}, \tag{23}$$

for scalar $\alpha = \sum_{i \in K} x_i$.

5 Conclusion and Future Work

In this paper we have revisited the assumption that the personalisation vector needs to be completely dense. We used Proposition 1 to show that this is not the case. In so doing we have presented a generalisation of PageRank which better accords with the intuitive justification given in the literature.

We then introduced Theorem 2. This theorem is broad-ranging in terms of its potential applicability. The theorem may be used to apply novel solution methods to eigenvector problems. For example, given some irreducible complex matrix, M, with unit spectral radius of its modulus equivalent, if we choose $\boldsymbol{v}$ to be the first column of M and if we choose $K = \{1\}$, then the theorem allows us to apply asynchronous solution [15] to solve for the dominant eigenvector of M. Equally, the theorem may be used to improve sparsity patterns or conditioning when using traditional solution methods to such problems. We intend to explore both applications in future work. In this paper, however, the application of—and, indeed, inspiration for—Theorem 2 was the PageRank equation. Applied to this special case, we used the theorem to remove the dangling-page matrix from the PageRank definition.

Finally we considered an extension of the PageRank definition which allows the dangling-page vector to be different from the personalisation vector. We showed how the same linear algebraic framework enables the dangling-page matrix to be removed here. We also suggested an alternative approach via Theorem 3.

References

1. L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank citation ranking: Bringing order to the web, Tech. rep., Stanford Digital Library Technologies Project (1998).
2. S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, in: Seventh International World-Wide Web Conference (WWW 1998), 1998.

3. Official Google blog.
URL http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html
4. N. Eiron, K. S. McCurley, J. A. Tomlin, Ranking the web frontier, in: WWW '04: Proceedings of the 13th international conference on World Wide Web, ACM, New York, NY, USA, 2004, pp. 309–318.
5. K. Avrachenkov, N. Litvak, Decomposition of the Google PageRank and Optimal Linking Strategy, Tech. Rep. RR-5101, INRIA (01 2004).
6. A. N. Langville, C. D. Meyer, A reordering for the PageRank problem, SIAM J. Sci. Comput 27 (2004) 2112–2120.
7. G. M. Del Corso, A. Gullí, F. Romani, Fast PageRank computation via a sparse linear system, Internet Mathematics 2 (3).
8. C. P.-C. Lee, G. H. Golub, S. A. Zenios, A fast two-stage algorithm for computing PageRank and its extensions, Technical report, Stanford InfoLab (2003).
9. I. C. F. Ipsen, T. M. Selee, PageRank computation, with special attention to dangling nodes, SIAM J. Matrix Anal. Appl. 29 (4) (2007) 1281–1296.
10. Y. Lin, X. Shi, Y. Wei, On computing PageRank via lumping the Google matrix, J. Comput. Appl. Math. 224 (2) (2009) 702–708.
11. Z. Gyöngyi, H. Garcia-Molina, J. Pedersen, Combating web spam with trustrank, in: VLDB '04: Proceedings of the Thirtieth international conference on Very large data bases, VLDB Endowment, 2004, pp. 576–587.
12. G. Kollias, E. Gallopoulos, D. B. Szyld, Asynchronous iterative computations with web information retrieval structures: The pagerank case, CoRR abs/cs/0606047.
13. A. Cevahir, C. Aykanat, A. Turk, B. B. Cambazoglu, A web-site-based partitioning technique for reducing preprocessing overhead of parallel pagerank computation, in: PARA, 2006, pp. 908–918.
14. J. T. Bradley, D. de Jager, W. J. Knottenbelt, A. Trifunovic, Hypergraph Partitioning for Faster Parallel PageRank Computation, in: Proceedings of 2nd European Performance Evaluation Workshop, Vol. 3670 of LNCS, 2005, pp. 155–171.
15. D. Chazan, W. L. Miranker, Chaotic relaxation, Linear Algebra and Its Applications 2 (1969) 199–222.
16. S. D. Kamvar, T. H. Haveliwala, C. D. Manning, G. H. Golub, Extrapolation methods for accelerating PageRank computations, in: Proceedings of the 12th Int. World Wide Web Conference, 2003.
17. P. Berkhin, A survey on PageRank computing, Internet Mathematics 2 (2005) 73–120.
18. A. N. Langville, C. D. Meyer, Deeper inside PageRank, Internet Mathematics 1 (3) (2004) 335–380.
19. T. Haveliwala, S. Kamvar, The second eigenvalue of the Google matrix, Technical Report 2003–20, Stanford InfoLab (2003).
20. A. Berman, R. J. Plemmons, Nonnegative matrices in the mathematical sciences, New York: Academic Press, New York, 1979.
21. S. Thurow, D. Sullivan, Search Engine Visibility, Pearson Education, 2002.
22. G. H. Golub, C. F. van Loan, Matrix Computations, 3rd Edition, Johns Hopkins University Press, 1996.