Taboos in PageRank Computation *

Frank van Rest† Flora Spieksma‡

October 22, 2008

Abstract

In [11] Ipsen and Selee study the effect of lumping dangling nodes on the PageRank vector, i.e. the order in which Google displays the web pages. It appears that the PageRank of a non-dangling node is independent of whether dangling nodes are lumped or not. Ipsen and Selee show this result by expressing lumping as a similarity transform. In this paper we will use a probabilitistic method using taboo sets for an alternative proof clarifying this result. We will also present an aggregation/disaggregation algorithm for computing PageRank, which is based on a double taboo decomposition and which has a local character.

AMS subject classifications: 15A51, 60J10, 65A06, 65C40, 65F10, 65F15, 68P20.
Keywords: Markov chains, stochastic matrix, stochastic complement, stationary distribution, taboo probabilities, power algorithm.
Running title: Taboos in PageRank computation.

1 Introduction

The PageRank vector π is a ranking of web pages that reflects the importance of a page. Mathematically it is calculated as the stationary distribution of a discrete time Markov chain $\{X_n\}_n$, that lives on the set $\mathcal{V}$ of all web pages, say $\mathcal{V} = \{1, \ldots, n\}$, where n is the total number of web pages. The total number of web pages has been recently claimed to exceed one trillion [1]. Web experiments suggest that the proportion of dangling nodes is substantial (more than 50% [6]). A huge reduction of computational effort might be achieved, if one can lump dangling nodes without affecting the PageRank of non-dangling nodes. This is precisely the result in Ipsen and Selee [11]. We will first derive the same result in a probabilistic manner that we believe to clarify the result. More precisely, we use a generalisation of the notion that, upto a multiplicative constant, the stationary probability of a state is the expected number of visits to that state between two successive visits to a (fixed) reference state, provided that the Markov chain under consideration is irreducible. This amounts to applying a taboo decomposition, where the reference state plays the role of the taboo set. In the second part of this paper, we apply a double taboo decomposition to derive an aggregation/disaggregation algorithm for computing PageRank.

Let us first introduce our notation. Vectors are always meant to be column vectors, the i-th co-ordinate is denoted by subscript i. By the (column) vector e we mean the vector consisting of ones only, indexed by elements from the space under consideration. Let v be any vector on the space

*Please send all correspondence to the second author
†Mathematics Institute, University of Leiden, P.O.Box 9512, 2300RA Leiden, The Netherlands. Email: frest@math.leidenuniv.nl
‡Mathematics Institute, University of Leiden, P.O.Box 9512, 2300RA Leiden, The Netherlands. Email: spieksma@math.leidenuniv.nl

and A a subset of elements. Then v_A is the projection of v onto the subspace A, in other words v_A has i-th component $v_{A,i} = v_i$, $i \in A$, and $v_{A,i} = 0$ otherwise. Vectors are always assumed to have the right dimension.

The directed graph with vertices $\mathcal{V}$ and edges corresponding to the hyperlinks is called the *web graph*. The adjacency matrix with normalised rows is denoted by H. The zero rows of H correspond to dangling nodes. Denote by $\mathcal{D}$ the set of dangling nodes.

With each dangling node we associate a probability distribution. This probability distribution may be different across dangling nodes, in order to incorporate the possibility of personalisation (cf. Ipsen and Selee [11]). Formally, we assume that $\mathcal{D}$ is partitioned into m classes $\mathcal{D}_1, \ldots \mathcal{D}_m$. With class $\mathcal{D}_l$ we associate a probability distribution w^l on $\mathcal{V}$, $l = 1, \ldots, m$. The vectors w^l are called dangling node vectors. The matrix $H + \sum_{l=1}^m e_{\mathcal{D}_l}^T w^l$ is a stochastic matrix. The interpretation of this matrix is, that on web page $i \in \mathcal{V} \setminus \mathcal{D}$, the next web page to be visited is chosen uniformly from the links on i. On the other hand, if the present page $i \in D_l$, then the next page chosen is drawn from distribution w^l.

The Markov chain associated with the web, is constructed as follows. A each time instant, with probability α the next web page is drawn from the row of $H + \sum_{l=1}^m e_{\mathcal{D}_l}^T w^l$ that corresponds to the presently visited web page, and with probability $(1-\alpha)$ the next page is drawn from a probability distribution v on $\mathcal{V}$.

The transition matrix G associated with $\{X_n\}_n$ is called the *Google matrix*. It has the following form

$$G = \alpha(H + \sum_{l=1}^m e_{\mathcal{D}_l} w^{l,T}) + (1-\alpha) e v^T, \quad 0 < \alpha < 1.$$

For the present computation of PageRank it seems that $m = 1$, $w^1 = v$ and that $\alpha = 0.85$.

The lumped Google matrix $\hat{G}$ is constructed as follows. The state space $\mathcal{V}$ is collapsed into $\hat{\mathcal{V}} = (\mathcal{V} \setminus \mathcal{D}) \cup \hat{\mathcal{D}}$, where $\hat{\mathcal{D}} = \{\mathcal{D}_1, \ldots, \mathcal{D}_m\}$. Then

$$\hat{G} = \alpha(\hat{H} + \sum_l e_{\mathcal{D}_l} \hat{w}^{l,T}) + (1-\alpha) e \hat{v}^T,$$

with $\hat{v}_i = v_i$ for $i \in \mathcal{V} \setminus \mathcal{D}$, and $\hat{v}_l = \sum_{d \in \mathcal{D}_l} v_d$. Further, $\hat{H}$ and $\hat{w}$ are the lumped versions of H and w_l respectively. The PageRank vector associated with the lumped Google matrix is denoted by $\hat{\pi}$. The following statement asserts that the PageRank vector associated with the lumped Google matrix equals the lumped PageRank vector of the full Google matrix. It has been proved in [11] Theorems 3.2 and 3.3.

t:t1

Theorem 1.1 *One has $\hat{\pi}_i = \pi_i$, for $i \in \mathcal{V} \setminus \mathcal{D}$, that is, PageRank and lumped PageRank are equal for non-dangling nodes. Moreover, $\hat{\pi}_{\mathcal{D}_l} = \pi_{\mathcal{D}_l}$, $l = 1, \ldots, m$. For $i \in D_l$,*

(e:PReq)
$$\pi_i = \alpha \sum_{j \in \mathcal{N}} \pi_j H_{ji} + (1-\alpha) v_i + \sum_k \hat{\pi}_{\mathcal{D}_k} w_l(i). \tag{1.1}$$

In the next section we will prove this theorem by a probabilistic argument.

2 Taboo matrix and proof of Theorem 1.1

Consider an irreducible, aperiodic Markov chain $\{X_n\}_n$ on a finite state space $\boldsymbol{S}$, with transition matrix P. Partition

(e:ep)
$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \tag{2.1}$$

where P_{11}, P_{22} are square matrices. The stochastic complement [16] of P_{22} in P is given by

$$S_{22} = P_{22} + P_{21}(\mathbf{I} - P_{11})^{-1}P_{12}.$$

S_{22} is an irreducible, stochastic matrix. It is aperiodic if P_{22} is (cf. Theorem 5.1 [16]). We denote the stationary distribution of $\{X_n\}_n$ by π say, which is partitioned as $\pi = (\pi_1, \pi_2)$ in correspondance with the partition of P. The following lemma holds.

t:ll1

Lemma 2.1

(e:eee)
$$\pi_1^T = \pi_2^T P_{21}(\mathbf{I} - P_{11})^{-1} = \pi_2^T P_{21} \sum_{n\geq 0} P_{11}^n. \tag{2.2}$$

Upto a multiplicative constant, π_2 is the stationary distribution of the Markov chain with transition matrix S_{22}. If P_{22} is a 1×1-matrix, then

(e:eee2)
$$\pi_2 = \frac{1}{1 + P_{21}(\mathbf{I} - P_{11})^{-1}e}. \tag{2.3}$$

The first two statements of the lemma are proved in [16] Corollary 4.1 and [10] (cf. equation (3.2) of that paper). For P_{22} a 1×1-matrix, the first and last statements are well-known renewal theoretical results (cf. Chung [5] §I.15, see also Avrachenkov and Litvak [3]). We will show the first and last statements by decomposition arguments. The second follows immediately from (2.2) and the fact that a primitive stochastic matrix has a unique left eigenvector (upto a multiplicative constant) to eigenvalue 1.

For the first statement, let set A be the states corresponding to the rows of P_{22}. Then $\boldsymbol{S} \setminus A$ are the states corresponding to the rows of P_{11}. For $i \notin A$

$$\begin{aligned}
\pi_i &= \sum_j \pi_j\, p_{ji} \\
&= \sum_{a\in A} \pi_a\, p_{ai} + \sum_{j\notin A} \pi_j\, p_{ji} \\
&= \sum_{a\in A} \pi_a\, p_{ai} + \sum_{j\notin A} \Big(\sum_{a\in A} \pi_a\, p_{aj} + \sum_{k\notin A} \pi_k\, p_{kj} \Big)\, p_{ji} \\
&= \pi_2^T P_{21}(\mathbf{I} + P_{11})e_i + \pi_1^T P_{11}^2 e_i.
\end{aligned}$$

Iterating this, we obtain for each $i \notin A$ and $n \geq 1$

$$\pi_i = \pi_2^T P_{21}(\mathbf{I} + P_{11} + \cdots + P_{11}^n)e_i + \pi_1^T P_{11}^{n+1} e_i.$$

Take the limit $n \to \infty$. Then $P_{11}^n \to 0$, by irreducibility and the fact that P_{11} is a substochastic matrix. This implies (2.2).

Next, note that

$$e_a^T(P_{21}P_{11}^{n-1})e_j = \mathsf{P}\{X_n = j, X_1, \ldots, X_n \notin A \,|\, X_0 = a\},$$

equals the probability that the Markov chain is in state j at time n, without having passed through states of A after time 0, given that it starts in state $a \in A$. In other words, the set A is a *taboo set* (cf. Chung [5]) and the above probabilities are taboo probabilities.

Denote by $T_A = \min\{n \geq 1 \,|\, X_n \in A\}$ the first hitting time of A, and by $N_i = \sum_{n\geq 1} \mathbf{1}_{\{X_n=i, T_A>n\}}$ the total number of visits to state $i \notin A$ before hitting set A. Then

$$\begin{aligned}
\mathsf{E}\{N_i \,|\, X_0 = a\} &= \mathsf{E}\{\sum_{n\geq 1} \mathbf{1}_{\{X_n=i,T_A>n\}} \,|\, X_0 = a\} \\
&= \sum_{n\geq 1} \mathsf{P}\{X_n = i, T_A > n \,|\, X_0 = a\} \\
&= \sum_{n\geq 1} (P_{21}P_{11}^n)_{ai} = e_a^T P_{21}(\mathbf{I} - P_{11})^{-1} e_i, \quad i \notin A, a \in A.
\end{aligned}$$

Hence, an alternative formulation of (2.2) is

(e:eee1)
$$\pi_i = \sum_{a\in A} \pi_a \mathsf{E}\{N_i \,|\, X_0 = a\}, \quad i \notin A. \tag{2.4}$$

For $A = \{a\}$ consisting of one element, this implies that π_i/π_a equals the expected number of visits to state i before hitting set A.

Rewrite $\mathsf{P}\{T_A > n \,|\, X_0 = a\} = e_a^T P_{21} P_{11}^{n-1} e_{A^C}$, for $n \geq 1$. Then

$$\begin{aligned}
\mathsf{E}\{T_A \,|\, X_0 = a\} &= \sum_{n\geq 0} \mathsf{P}\{T_A > n \,|\, X_0 = a\} \\
&= 1 + e_a^T P_{21}(\mathbf{I} - P_{11})^{-1} e_{A^C} \\
&= 1 + \sum_{i\notin A} \mathsf{E}\{N_i \,|\, X_0 = a\}.
\end{aligned}$$

Again, considering the case of $A = \{a\}$ consisting of one taboo state, we have $P_{22} = (\,p_{aa})$, $P_{21} = (\,p_{aj})_{j\neq a}$. The summation over all stationary probabilities yields

$$1 = \sum_i \pi_i = \sum_{i\neq a} \pi_a \mathsf{E}\{N_i \,|\, X_0 = a\} + \pi_a = \pi_a(1 + \mathsf{E}\{T_a \,|\, X_0 = a\}) = \pi_a(1 + P_{21}(\mathbf{I} - P_{11})^{-1} e),$$

whence (2.3) follows

hier nog even latie

$$\pi_a = \frac{1}{1 + \mathsf{E}\{T_a \,|\, X_0 = a\}} = \frac{1}{1 + P_{21}(\mathbf{I} - P_{11})^{-1} e},$$

in other words, π_a equals one over the return time to a.

Observe that S_{22} is the transition matrix of the Markov chain embedded on the instants of visits to states of A. With the above interpretation, the validity of the second statement of Lemma 2.1 is intuitively obvious: given a selected state $a \in A$, for any $i \in A \setminus \{a\}$, π_i/π_a equals the expected number of visits to i before returning to a. But this number is the same for original Markov chain and the embedded one on the instants of visits of A.

For the proof of Theorem 1.1 it is convenient not to work with partitions of P, but to use the so-called taboo matrix (cf. Chung [5]).

For taboo set A, define the taboo matrix ${}_AP$ by ${}_Ap_{ij} = p_{ij}$ for $j \notin A$ and ${}_Ap_{ij} = 0$ for $j \in A$. In terms of the above partition of P, this means that

$${}_AP = \begin{pmatrix} P_{11} & 0 \\ P_{21} & 0 \end{pmatrix}.$$

Put ${}_AP^{(n)} = ({}_AP)^n$, so that ${}_AP^{(0)} = \mathbf{I}$. Rewriting (2.2) yields

$$\pi_i = \pi_A^T (\mathbf{I} - {}_AP)^{-1} e_i, \quad i \notin A.$$

Proof of Theorem 1.1. Choose state $i \notin \mathcal{D}$ and use taboo set $\{i\}$. Note first that

$$ {}_iG_{ij}^{(n)} = {}_i\hat{G}_{ij}^{(n)}, \quad j \notin \mathcal{D}. \tag{2.5} $$

(e:e1)

The easiest way to see this, is by drawing directed transition graphs associated with Google matrix and lumped version. With the Google matrix we associate the standard transition graph in the following way. Associate a vertex i with each web page i. Draw an arrow $i \to j$ if $G_{ij} > 0$ and assign weight G_{ij} to this arrow.

With the lumped Google matrix we associate a slightly different transition graph. It is the lumped Google graph but with preserved arrows. More precisely, it has vertex set $\hat{\mathcal{V}}$. With web pages $i, j \notin \mathcal{D}$ we associate an arrow $i \to j$ with weight G_{ij}. Let $i \in \hat{\mathcal{V}}$, and suppose $j = \mathcal{D}_l$. Let $A_l \subset \mathcal{D}_l$ with $A_l = \{d \in \mathcal{D}_l \mid G_{id} > 0\}$. Put $|A_l|$ parallel arrows $i \to j$, and assign respective weights G_{id}, $d \in A_l$, to the successive parallel arrows.

The weight of a path in these graphs is simply the product of the weights along the path. Hence it reflects the probability of this path. It is now immediate that for $j \notin \mathcal{D}$ with each path $i \rightsquigarrow j$ we can associate one path $i \rightsquigarrow j$ in the lumped graph of equal length and weight, and vice versa. Similarly, with each path $i \rightsquigarrow d \in \mathcal{D}_l$ in the Google graph, we can associate precisely one path $i \rightsquigarrow \mathcal{D}_l$ in the lumped graph of equal length and weight and vice versa. (2.5) follows from the fact that ${}_iG_{ij}^{(n)}$ is the sum over the weights of all paths $i \rightsquigarrow j$ of length n in the Google graph, and similarly for ${}_i\hat{G}_{ij}^{(n)}$. Analogously,

$$ \sum_{d \in \mathcal{D}_l} {}_iG_{id}^{(n)} = {}_i\hat{G}_{i\mathcal{D}_l}^{(n)}. \tag{2.6} $$

(e:e1a)

From (2.5) and (2.6) we have

$$ \sum_{j \in \mathcal{V}} {}_iG_{ij}^{(n)} = \sum_{j \in \hat{\mathcal{V}}} {}_i\hat{G}_{ij}^{(n)}, $$

so that by virtue of Lemma 2.1

$$ \frac{1}{\pi_i} = \sum_{n \geq 0} \sum_j {}_iG_{ij}^{(n)} = \sum_{n \geq 0} \sum_j {}_i\hat{G}_{ij}^{(n)} = \frac{1}{\hat{\pi}_i}, $$

where we have taken taboo set $A = \{i\}$. Hence, $\pi_i = \hat{\pi}_i$ for $i \notin \mathcal{D}$.

Next, use (2.6) and Lemma 2.1 to obtain that

$$ \sum_{d \in \mathcal{D}_l} \pi_d = \pi_i \sum_{d \in \mathcal{D}_l} \sum_n {}_iG_{id}^{(n)} = \pi_i \sum_n \sum_{d \in \mathcal{D}_l} {}_iG_{id}^{(n)} = \hat{\pi}_i \sum_n {}_i\hat{G}_{i\mathcal{D}_l}^{(n)} = \hat{\pi}_{\mathcal{D}_l}, $$

where we have used Fubini's theorem in the last equality. The last statement in Theorem 1.1 follows directly from the linear equation $\pi = \pi\mathcal{G}$. QED

Note that similar derivations using a taboo set can be applied to general non-negative primitive matrices. Consequently, one can lump equal rows of non-negative primitive matrices without affecting the components of the left eigenvector to the largest positive eigenvalue that are associated with non lumped states.

3 Revised power algorithm

By virtue of Lemma 2.1 one can compute π_A, upto a multiplicative constant, as the stationary distribution of the embedded Markov chain on set A, for any subset $A \subset \boldsymbol{S}$, or, in an alternative formulation, as the stationary distribution of the stochastic complement corresponding to the set A.

This is the basis of the iterative disaggregation/aggregation algorithm in [13], [14], which has been refined in [10]. The rate of convergence of this algorithm is equal to the second eigenvalue of the stochastic complement. For the Google matrix [10] shows that the second eigenvalue is strictly smaller than α, which is the rate of convergence of the power algorithm, provided A is chosen suitably. In particular, $\boldsymbol{S}\setminus A$ is a collection of states, one from each closed class in the Markov chain associated with $H + \sum_l e_{D_l} w^{l,T}$, such that there is no transition from each state to itself.

We will propose an alternative disaggregation/aggregation algorithm based on a double taboo decomposition. To explain it, we partition P as (2.1), where the states have been permuted in such a way that the rows of P_{22} correspond to the states of A, and the rows of P_{11} to the states of $B = \boldsymbol{S}\setminus A$.

Applying Lemma 2.1, we can do two successive taboo decompositions to get

$$\pi_1^T = \pi_2^T P_{21}(\mathbf{I} - P_{11})^{-1} = \pi_1^T P_{12}(\mathbf{I} - P_{22})^{-1} P_{21}(\mathbf{I} - P_{11})^{-1}.$$

This means that π_1 can be calculated as the left eigenvector corresponding to eigenvalue 1 of the matrix $M = P_{12}(\mathbf{I} - P_{22})^{-1} P_{21}(\mathbf{I} - P_{11})^{-1}$, where M is a $|B| \times |B|$-matrix, which may be chosen to have a much smaller dimension than P. In general M is not a stochastic matrix as can be seen from the next example.

Example 3.1 (Example 6.3 from [10]) Let

$$P = \begin{pmatrix} 5/6 & 0 & 1/6 \\ 3/4 & 1/6 & 1/12 \\ 2/3 & 1/3 & 0 \end{pmatrix}.$$

The matrix P has two eigenvalues, 0 and 1, with 0 an eigenvalue of algebraic multiplicity 2. The rate of convergence of the iterative disaggregation/aggregation algorithm in [13], [14], [10] is equal to the second largest eigenvalue of the particular stochastic complement considered. The three possibilities for 2×2 stochastic complements yield convergence rates 1/6, 2/15 and 5/36. This is worse than the convergence rate of the power algorithm applied to P, which is 0.

Denote

$$P_i = \begin{pmatrix} P_{i,11} & P_{i,12} \\ P_{i,21} & P_{i,22} \end{pmatrix},$$

where we have permuted row and column i to the *last* position, and where $P_{i,11}$ is a 2×2 matrix. Put $M_i = P_{i,12}(\mathbf{I} - P_{i,22})^{-1} P_{i,21}(\mathbf{I} - P_{i,11})^{-1}$. Calculation yields

$$M_1 = \begin{pmatrix} 9/29 & 45/58 \\ 8/29 & 20/29 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0 & 0 \\ 58/10 & 1 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 29/30 & 2/30 \\ 29/60 & 1/30 \end{pmatrix},$$

all of which have eigenvalues 1 and 0.

This suggests the following aggregation/disaggregation algorithm for computing PageRank. We consider the lumped Google matrix, that we denote by G. For simplicity we assume only one (lumped) dangling node d, and denote the associated probability distribution by w. Recall that an approximation of the PageRanks of the unlumped dangling nodes can be obtained from (1.1).

Use the partition $G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}$, where the rows of G_{22} correspond to a suitably chosen set A, such that $(\mathbf{I} - G_{22})^{-1}$ is easily calculated, and such that $\sum_{n\geq 0} G_{11}^n$ has good convergence properties.

To this end, we introduce an exra dangling node $\mathbf{v}$. The extended state space is $\boldsymbol{\mathcal{V}} = \mathcal{V} \cup \{\mathbf{v}\}$ and we extend the link matrix H accordingly. The transition probabilities are then given by

$$\boldsymbol{G} = \alpha(e_{\mathcal{V}} H + e_d w^T) + (1-\alpha) e_{\mathcal{V}} e^T + e_{\mathbf{v}} v^T.$$

Consider the taboo set $\{i\}$, with $i \neq \mathbf{v}$. Given that we start at web page i, it follows that the expected number of visits of $j \neq \mathbf{v}$ before returning to i, is the same for original and extended process. Denote by $\boldsymbol{\pi}$ the stationary distribution corresponding to $\boldsymbol{G}$. By virtue of (2.2) $\pi_j/\pi_i = \boldsymbol{\pi}_j/\boldsymbol{\pi}_i$. A simple calculation shows that by this extension the stationary probabilities of the original web pages are reduced by a factor $2-\alpha$, i.e. $\boldsymbol{\pi}_i = \pi_i/(2-\alpha)$, $i \neq \mathbf{v}$.

Next, let $A \subset \boldsymbol{\mathcal{V}}$ be an arbitrary subset containing the extra dangling node $\mathbf{v}$. Then

$$\begin{pmatrix} \boldsymbol{G}_{12} \\ \boldsymbol{G}_{22} \end{pmatrix} e_{\mathbf{v}} = \begin{pmatrix} (1-\alpha)e_{\mathcal{V}} \\ 0 \end{pmatrix}, \quad e_{\mathbf{v}}^T (\boldsymbol{G}_{21}, \boldsymbol{G}_{22}) = (v^T, 0).$$

Algorithm 1
% Inputs $\boldsymbol{G}$, α, initial distribution $\sigma^0 = (\sigma_1^0, \sigma_2^0)$, output approximation $\tilde{\pi}$ of π.
% Step 1 Aggregation
% Power method applied to the matrix $M = \boldsymbol{G}_{12}(\mathbf{I} - \boldsymbol{G}_{22})^{-1}\boldsymbol{G}_{21}(\mathbf{I} - \boldsymbol{G}_{11})^{-1}$.
Put $t = 0$.
While not converged

$$\sigma_1^{t+1,T} = \sigma_1^{t,T} M.$$

$t := t + 1$.
end while
% Step 2 Disaggregation
% Recover PageRank for pages in A.
Put $\sigma_{2,i}^t = \sigma_1^{t,T} G_{12}(\mathbf{I} - G_{22})^{-1} e_i$ for $i \in A$.
Put $\tilde{\pi} = \sigma_{\mathcal{V}}^t/(\sigma_{\mathcal{V}}^t, e)$.

t:tt1
Theorem 3.1 *Assume that $v_j > 0$ for all $j \in \mathcal{V}$ and that $\mathbf{v} \in A$, where the rows of $\boldsymbol{G}_{22}$ correspond to the states of A. Then $\boldsymbol{G}$ and M are primitive, non-negative matrices. Algorithm 1 converges at rate α^2. In particular, with $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$ the PageRank vector associated with the extended Googlematix $\boldsymbol{G}$*

$$\|\sigma_1^t - \frac{(\sigma_1^0, \boldsymbol{G}_{12}e)}{(\boldsymbol{\pi}_1, \boldsymbol{G}_{12}e)}\boldsymbol{\pi}_1\|_1 \leq \frac{2(\sigma_1^0, \boldsymbol{G}_{12}e)}{1-\alpha}\alpha^{2(t-1)}.$$

The positivity assumption on v is not strictly necessary, but especially lateron it will make our arguments somewhat simpler.

Proof. Irreducibility is immediate from the assumptions. Moreover, the Google graph associated with $\boldsymbol{G}$ has cycles of length 2: $i \to \mathbf{v} \to i$, if $v_i > 0$. If there is a link from page i to page j, then $i \to j \to \mathbf{v} \to i$ is a cycle of length 3. Hence the greatest common divisor of all cycles is 1. It follows that $\boldsymbol{G}$ is primitive (cf. Chung [5] for the case of a stochastic matrix).

As regards M, note that $M_{ii} \geq (1-\alpha)v_i > 0$. Primitivity follows in the same way as in the above.

Consider now also the matrix $\hat{M} = (\mathbf{I} - \boldsymbol{G}_{22})^{-1}\boldsymbol{G}_{21}(\mathbf{I} - \boldsymbol{G}_{11})^{-1}\boldsymbol{G}_{12}$, which is related to M by

(e:bga)
$$\boldsymbol{G}_{12}\hat{M}^t = M^t\boldsymbol{G}_{12}. \tag{3.1}$$

Both $(\mathbf{I} - \boldsymbol{G}_{22})^{-1}\boldsymbol{G}_{21}$ and $(\mathbf{I} - \boldsymbol{G}_{11})^{-1}\boldsymbol{G}_{12}$ are stochastic matrices ([16] Theorem 2.1). Consequently, $\hat{M}$ is a stochastic matrix and $\sum_{t\geq 0}\hat{M}^t z^t$ converges on the unit disc in the complex plane. Then also

$$\boldsymbol{G}_{12}\sum_{n\geq 0}\hat{M}^t z^t = \sum_{n\geq 0} M^t z^t \boldsymbol{G}_{12}$$

converges on the unit disc. Since 1 is an eigenvalue of M, it necessarily follows that 1 is the largest eigenvalue (in absolute value).

It now follows from Perron Frobenius theory (cf. Karlin [12]) that to eigenvalue 1 there exist unique (upto a multiplicative constant) left and right eigenvectors with all components positive. All other eigenvalues are smaller in absolute value.

The left eigenvector equals $\boldsymbol{\pi}_1$. For the right eigenvector, note that e_A is a right eigenvector of $\hat{M}$ to eigenvalue 1. In the course of the proof we will use the notation e_A and e_B to indicate the vectors on A and B respectively, with all components equal to 1. In the formulation of the theorem we have suppressed this dependance. It follows from (3.1) that $\boldsymbol{G}_{12}e_A$ is a right eigenvector of M and $\boldsymbol{\pi}_1\boldsymbol{G}_{12}$ is a left eigenvector of $\hat{M}$, both to eigenvalue 1. From Perron Frobenius theory we then have that

$$M^n \to \frac{1}{(\boldsymbol{\pi}_1, \boldsymbol{G}_{12}e_A)}\boldsymbol{G}_{12}e_A\boldsymbol{\pi}_1^T, \quad n \to \infty.$$

We will next show that

(e:split)
$$\hat{M} = \alpha^2 P + (1-\alpha)e_A f^T + \alpha(1-\alpha)e_A p^T, \tag{3.2}$$

for some stochastic matrix P and probability distributions $f, p \geq 0$ on A.

As a shorthand notation, we write $Q = H + e_d w^T$, so that $\boldsymbol{G} = \alpha Q + (1-\alpha)e_{\mathcal{V}}e^T + e_{\mathbf{v}}v^T$. Q is a stochastic matrix and we partition it in accordance with the partition of $\boldsymbol{G}$. Note that $\boldsymbol{G}_{11} = \alpha Q_{11}$. For $i \in A, i \neq \mathbf{v}$ we have

$$\hat{M}_{ij} = (1-\alpha)\hat{M}_{\mathbf{v}j} + e_i^T(\mathbf{I} + \alpha Q_{22}(\mathbf{I} - \boldsymbol{G}_{22})^{-1})\boldsymbol{G}_{21}(\mathbf{I} - \boldsymbol{G}_{11})^{-1}\boldsymbol{G}_{12}e_j,$$

where the first term denotes the contribution to $\hat{M}_{ij}$ of the paths passing through $\mathbf{v}$ at the first step, and the second term the contribution of the remaining paths. Obviously one has

$$\hat{M}_{\mathbf{v}j} = (1-\alpha)\hat{M}_{\mathbf{v}j} + \alpha\hat{M}_{\mathbf{v}j},$$

and so we may choose

$$f_j = \hat{M}_{\mathbf{v}j}, \quad j \in A.$$

For $i \in A, \neq \mathbf{v}$

(e:eqqq)
$$\begin{aligned} e_i^T(\mathbf{I} + \alpha Q_{22}(\mathbf{I} - \boldsymbol{G}_{22})^{-1})\boldsymbol{G}_{21}(\mathbf{I} - \boldsymbol{G}_{11})^{-1}\boldsymbol{G}_{12}e_{\mathbf{v}} &\geq e_i^T(\mathbf{I} + \alpha Q_{22}(\mathbf{I} - \boldsymbol{G}_{22})^{-1})\boldsymbol{G}_{21}e_B(1-\alpha) \\ &\geq e_i^T(\alpha Q_{21}e_B + \alpha Q_{22}e_A)(1-\alpha) = \alpha(1-\alpha). \end{aligned} \tag{3.3}$$

Also $\hat{M}_{\mathbf{vv}} \geq (1-\alpha)$ and so we may choose $p = \delta_{\mathbf{v}}$. This clearly implies (3.2), since $\hat{M}$ is a stochastic matrix. A standard coupling argument (cf. Asmussen [2]) now yields that

$$\sum_j |\hat{M}_{ij}^t - \hat{M}_{kj}^t| \leq 2\alpha^{2t}.$$

We may replace fixed initial states i and k by initial probability distributions on A. Choose initial distributions $p = \sigma_1^{0,T}\boldsymbol{G}_{12}/(\sigma_1^0, \boldsymbol{G}_{12}e_A)$ and $q = \boldsymbol{\pi}_1^T\boldsymbol{G}_{12}/(\boldsymbol{\pi}_1, \boldsymbol{G}_{12}e_A)$. We have already deduced that $\pi_1\boldsymbol{G}_{12}$ is a left eigenvector of $\hat{M}$. It follows that

(e:bbbound)
$$\|p^T\hat{M}^t - q^T\hat{M}^t\|_1 \leq \sum_{i,k,j} p_i q_k |\hat{M}_{i,j}^t - \hat{M}_{kj}^t| \leq 2\alpha^{2t}. \tag{3.4}$$

As a consequence

$$\begin{aligned}
&\|\frac{1}{(\sigma_1^0, \boldsymbol{G}_{12}e_A)}\sigma_1^{0,T}M^t - \frac{1}{(\boldsymbol{\pi}_1, \boldsymbol{G}_{12}e_A)}\boldsymbol{\pi}_1\|_1 \\
&\quad = \|\frac{1}{(\sigma_1^0, \boldsymbol{G}_{12}e_A)}\sigma_1^{0,T}M^t - \frac{1}{(\boldsymbol{\pi}_1, \boldsymbol{G}_{12}e_A)}\boldsymbol{\pi}_1 M^t\|_1 \\
&\quad = \|(p^T\hat{M}^{t-1} - q^T\hat{M}^{t-1})(\mathbf{I}-\boldsymbol{G}_{22})^{-1}\boldsymbol{G}_{21}(\mathbf{I}-\boldsymbol{G}_{11})^{-1}\|_1 \\
&\quad \leq \|p^T\hat{M}^{t-1} - q^T\hat{M}^{t-1}\|_1 \cdot \|\mathbf{I}-\boldsymbol{G}_{22})^{-1}\boldsymbol{G}_{21}\|_\infty \cdot \|(\mathbf{I}-\boldsymbol{G}_{11})^{-1}\|_\infty \\
&\quad \leq \frac{2}{1-\alpha}\alpha^{2(t-1)},
\end{aligned}$$

since $\|\boldsymbol{G}_{11}\|_\infty \leq \alpha$. QED

The idea is now to take a taboo set A, such that $(\mathbf{I}-\boldsymbol{G}_{22})^{-1}$ is easily calculated. We follow [10] and take A such that there are no links between nodes of A. Assume that the dangling node $d \in A$. Put $\sum_{i\in A, i\neq \mathbf{v},d} v_i = \bar{v}$ and $\sum_{i\in A, i\neq \mathbf{v},d} w_i = \bar{w}$. For computing $(\mathbf{I}-\boldsymbol{G}_{22})^{-1}$ it is sufficient to first compute the inverse

$$\Gamma = \left(\mathbf{I} - \begin{pmatrix} 0 & 0 & 1-\alpha \\ \alpha\bar{w} & \alpha w_d & 1-\alpha \\ \bar{v} & v_d & 0 \end{pmatrix}\right)^{-1}.$$

Then we have for $j \in A$, $j \neq d, \mathbf{v}$

$$e_i^T(\mathbf{I}-\boldsymbol{G}_{22})^{-1}e_j = \delta_{ij} + \Gamma_{i\mathbf{v}}v_j + \Gamma_{id}\alpha w_j,$$

with δ_{ij} the Kronecker delta. This means that only have to compute the inverse of a 3×3 matrix, regardless of the number of nodes contained in A. If we allow m dangling nodes with different associated probability distributions, apart from the extra dangling node $\mathbf{v}$, then we have to compute the inverse of an $(m+2)\times(m+2)$-matrix.

As said, we can choose A in such a way that $(\mathbf{I}-\boldsymbol{G}_{22})^{-1}$ is easily computed. Unfortunately, this implies that $(\mathbf{I}-\boldsymbol{G}_{11})^{-1}$is not so easily computed. We propose to approximate it by a suitably normalised finite sum $D_N\sum_{n=0}^N \boldsymbol{G}_{11}^n$, with D_N a diagonal matrix with

$$D_{N,ii} = 1/e_i^T\sum_{n=0}^N G_{11}^n\boldsymbol{G}_{12}e_A$$

and N a fixed parameter.

Put $M_N = \boldsymbol{G}_{12}(\mathbf{I}-\boldsymbol{G}_{22})^{-1}\boldsymbol{G}_{21}D_N\sum_{n=0}^N\boldsymbol{G}_{11}^n$ and $\hat{M}_N = (\mathbf{I}-\boldsymbol{G}_{22})^{-1}\boldsymbol{G}_{21}D_N\sum_{n=0}^N\boldsymbol{G}_{11}^n\boldsymbol{G}_{12}$. By our choice of D_N, $\hat{M}_N$ is still a stochastic matrix and the analogon of (3.1) continues to hold:

(e:bgb)
$$\boldsymbol{G}_{12}\hat{M}_N^t = M_N^t\boldsymbol{G}_{12}. \tag{3.5}$$

In Algorithm 1 we replace the term $(\mathbf{I}-\boldsymbol{G}_{11})^{-1}$ by $D_N\sum_{t=0}^N\boldsymbol{G}_{11}$. The resulting Algorithm 2 has a *local character*, since only the contribution of pages within distance N of the pages in the taboo set A, measured along paths not passing through pages in A, is taken into account. That is, the pages within distance N of pages in A in the web sub graph, in which the arrows pointing to pages of A have been deleted.

Algorithm 2

`% Inputs` $\boldsymbol{G}$`,` α`,` N`,` `initial distribution` $\sigma^0 = (\sigma_1^0, \sigma_2^0)$`,`

% output approximation $\tilde{\pi}$ of π.
% **Step 1 Aggregation**
% Power method applied to the matrix $M_N = \boldsymbol{G}_{12}(\mathbf{I} - \boldsymbol{G}_{22})^{-1}\boldsymbol{G}_{21}D_N \sum_{n=0}^{N} \boldsymbol{G}_{11}^n$.
Put $t = 0$.
While not converged

$$\sigma_1^{t+1,T} = \sigma_1^{t,T} M_N.$$

$t := t + 1$.
end while
% **Step 2 Disaggregation**
% Recover PageRank for pages in A.
Put $\sigma_{2,i}^t = \sigma_1^{t,T} G_{12}(\mathbf{I} - G_{22})^{-1} e_i$ for $i \in A$.
Put $\tilde{\pi} = \sigma_{\mathcal{V}}^t/(\sigma_{\mathcal{V}}^t, e)$.

We will derive an estimate on how the outcome of this algorithm compares to PageRank.

t:tt2

Theorem 3.2 *Assume that $v_j > 0$ for all $j \in \mathcal{V}$ and that $\mathbf{v} \in A$, where the rows of $\boldsymbol{G}_{22}$ correspond to the states of A. Then $\boldsymbol{G}$ and M_N are primitive, non-negative matrices. Algorithm 2 converges at rate α^2 to a positive vector $\boldsymbol{\pi}_{N,1}$, with*

(e:boundd)

$$\|\sigma_1^t - \frac{(\sigma_1^0, \boldsymbol{G}_{12}e)}{(\boldsymbol{\pi}_{N,1}, \boldsymbol{G}_{12}e)}\boldsymbol{\pi}_{N,1}\|_1 \leq \frac{2(\sigma_1^0, \boldsymbol{G}_{12}e)}{1-\alpha}\alpha^{2(t-1)}. \tag{3.6}$$

Moreover,

(e:bound)

$$\|\frac{1}{(\boldsymbol{\pi}_{N,1}\boldsymbol{G}_{12}e)}\boldsymbol{\pi}_{N,1} - \frac{1}{(\boldsymbol{\pi}_1, \boldsymbol{G}_{12}e)}\boldsymbol{\pi}_1\|_1 \leq \frac{2\delta}{1-\alpha} \cdot \alpha^{N+1}, \tag{3.7}$$

where

$$\delta = \frac{3 - \alpha^2 - 6\alpha^{N+1} + \alpha^{N+3} + 2\alpha^{2N+2}}{(1-\alpha^2)(1-\alpha^{N+1}) - 2\alpha^{N+1}(2-\alpha^{N+1})}$$

t:cor

Corollary 3.3

$$\|\sigma_1^t - \frac{(\sigma_1^0, \boldsymbol{G}_{12}e)}{(\boldsymbol{\pi}_1, \boldsymbol{G}_{12}e)}\boldsymbol{\pi}_1\|_1 \leq \frac{2(\sigma_1^0, \boldsymbol{G}_{12}e)}{1-\alpha}(\alpha^{2(t-1)} + \delta\alpha^{N+1}).$$

Proof of Theorem 3.2. We may copy the proof of Theorem 3.1. The only point that requires some attention, is the analogon of inequality (3.3). Note that

$$\begin{aligned} e_j D_N \sum_{t=0}^{N} \boldsymbol{G}_{11}^t \boldsymbol{G}_{12} e_{\mathbf{v}} &= (1-\alpha) e_j D_N \sum_{t=0}^{N} \boldsymbol{G}_{11}^t e_B \\ &= (1-\alpha)\frac{e_j \sum_{t=0}^{N} \boldsymbol{G}_{11}^t e_B}{e_j \sum_{t=0}^{N} \boldsymbol{G}_{11}^t \boldsymbol{G}_{12} e_A} \\ &\geq (1-\alpha), \end{aligned}$$

since $e_j \boldsymbol{G}_{12} e_A \leq 1$ for all $j \in B$. This yields convergence rate and bound on the distance between σ^t and $\boldsymbol{\pi}_{N,1}(\sigma^0, \boldsymbol{G}_{12}e_A)/(\boldsymbol{\pi}_{N,1}, \boldsymbol{G}_{12}e_A)$.

We still have to derive the bound in (3.7). To this end, we use a perturbation bound from [7]. Denote by $\hat{\pi}$ and $\hat{\pi}_N$ the stationary distributions of the Markov chains with transition matrices $\hat{M}$ and $\hat{M}_N$ respectively. These live on the space A.

We further define the so-called deviation matrix $\hat{D}$ of the Markov chain associated with $\hat{M}$ by

$$\hat{D} = \sum_{t=0}^{\infty} (\hat{M}^t - e_A \hat{\pi}^T). \tag{3.8}$$

(e:dev)

The matrix $\hat{D}$ represents the total deviation from stationarity in the Markov chain associated with $\hat{M}$. Then (cf. [7] equation (14), [8] equation (4))

$$\hat{\pi}_N - \hat{\pi} = \hat{\pi}^T \sum_{t \geq 1} ((\hat{M}_N - \hat{M})\hat{D})^t, \tag{3.9}$$

(e:dev2)

so that

$$\|\hat{\pi}_N - \hat{\pi}\|_1 \leq \sum_{t \geq 1} \|\hat{M}_N - \hat{M}\|_\infty^t \cdot \|\hat{D}\|_\infty^t. \tag{3.10}$$

(e:a3)

By (3.4)

$$\|\hat{M}^t - e_A \hat{\pi}^T\|_\infty \leq 2\alpha^{2t},$$

and so

$$\|\hat{D}\|_\infty \leq 2 \sum_{t \geq 0} \alpha^{2t} = \frac{2}{1-\alpha^2}. \tag{3.11}$$

(e:a1)

Moreover, since $(\mathbf{I} - \boldsymbol{G}_{22})^{-1}\boldsymbol{G}_{21}$ is a stochastic matrix,

$$\begin{aligned} \|\hat{M}_N - \hat{M}\|_\infty &\leq \|D_N \sum_{t=0}^{N} \boldsymbol{G}_{11}^t \boldsymbol{G}_{12} - \sum_{t \geq 0} \boldsymbol{G}_{11}^t \boldsymbol{G}_{12}\|_\infty \\ &\leq \|D_N - \mathbf{I}\|_\infty \|\sum_{t=0}^{N} \boldsymbol{G}_{11}^t \boldsymbol{G}_{12}\|_\infty + \|\sum_{t \geq N+1} \boldsymbol{G}_{11}^t \boldsymbol{G}_{12}\|_\infty. \end{aligned} \tag{3.12}$$

(e:a2)

We derive some useful bounds. First,

$$\begin{aligned} \|\sum_{t=0}^{N} \boldsymbol{G}_{11}^t \boldsymbol{G}_{12}\|_\infty &= \max_i e_i^T \sum_{t=0}^{N} \boldsymbol{G}_{11}^t \boldsymbol{G}_{12} e_A \\ &= \max_i \|\sum_{t=0}^{N} (\boldsymbol{G}_{11}^t e_B - \boldsymbol{G}_{11}^{t+1} e_B)\|_\infty \\ &= \max_i (1 - e_i^T \boldsymbol{G}_{11}^{N+1} e_B) \geq 1 - \alpha^{N+1}. \end{aligned} \tag{3.13}$$

(e:aa2)

Here we have used that $\boldsymbol{G}_{11}e_B + \boldsymbol{G}_{12}e_A = e_B$ in the second equality. In the last inequality we have used that $\mathbf{v} \in \mathbf{A}$, so that $e_i \boldsymbol{G}_{11}^{N+1} e_B \leq \alpha^{N+1}$. Using that $\sum_{t=0}^{N} \boldsymbol{G}_{11}^t \boldsymbol{G}_{12}$ is a substochastic matrix we find

$$\|\sum_{t=0}^{N} \boldsymbol{G}_{11}^t \boldsymbol{G}_{12}\|_\infty \leq 1. \tag{3.14}$$

(e:aa3)

Next, the same reasoning leading to (3.13) yields

$$\|\sum_{t \geq N+1} \boldsymbol{G}_{11}^t \boldsymbol{G}_{12}\|_\infty = \|\boldsymbol{G}_{11}^{N+1}\|_\infty \leq \alpha^{N+1}. \tag{3.15}$$

(e:aa4)

Finally, $\|D_N - \mathbf{I}\|_\infty$ can be appropriately bounded by using (3.13). The latter implies that $1 \leq e_i^T D_N e_i \leq (1-\alpha^{N+1})^{-1}$, so that

$$\|D_N - \mathbf{I}\|_\infty = \max_i |e_i^T D_N e_i - 1| \leq \frac{\alpha^{N+1}}{1-\alpha^{N+1}}.$$

Combination with (3.10), (3.11), (3.12), (3.14) and (3.15) yields

$$\|\hat{\pi}_N - \hat{\pi}\|_1 \leq \gamma \cdot \alpha^{N+1},$$

for

$$\gamma = \frac{2(2-\alpha^{N+1})}{(1-\alpha^2)(1-\alpha^{N+1}) - 2\alpha^{N+1}(2-\alpha^{N+1})}.$$

We next use that

$$\begin{aligned}
\frac{1}{(\boldsymbol{\pi}_{N,1}, \boldsymbol{G}_{12}e_A)}\boldsymbol{\pi}_{N,1} - \frac{1}{(\boldsymbol{\pi}_1, \boldsymbol{G}_{12}e_A)}\boldsymbol{\pi}_1 &= \hat{\pi}_N P D_N \sum_{t=0}^{N} \boldsymbol{G}_{11}^t - \hat{\pi} P \sum_{t\geq 0} \boldsymbol{G}_{11}^t \\
&= \hat{\pi}_N P (D_N \sum_{t=0}^{N} \boldsymbol{G}_{11}^t - \sum_{t\geq 0} \boldsymbol{G}_{11}^t) + (\hat{\pi}_N - \hat{\pi}) P \sum_{t\geq 0} \boldsymbol{G}_{11}^t \\
&= \hat{\pi}_N P (D_N - \mathbf{I}) \sum_{t=0}^{N} \boldsymbol{G}_{11}^t - \hat{\pi}_N P \sum_{t\geq N+1} \boldsymbol{G}_{11}^t + \\
&\qquad (\hat{\pi}_N - \hat{\pi}) P \sum_{t\geq 0} \boldsymbol{G}_{11}^t,
\end{aligned}$$

where $P = (\mathbf{I} - \boldsymbol{G}_{22})^{-1}\boldsymbol{G}_{21}$ is a stochastic matrix. By similar arguments as above, we derive

$$\begin{aligned}
\|\frac{1}{(\boldsymbol{\pi}_{N,1}, \boldsymbol{G}_{12}e_A)}\boldsymbol{\pi}_{N,1} - \frac{1}{(\boldsymbol{\pi}_1, \boldsymbol{G}_{12}e_A)}\boldsymbol{\pi}_1\|_1 &\leq \|D_N - \mathbf{I}\|_\infty \|\sum_{t=0}^{N} \boldsymbol{G}_{11}^t\|_\infty + \|\sum_{t\geq N+1} \boldsymbol{G}_{11}^t\|_\infty \\
&\qquad + \|\hat{\pi}_N - \hat{\pi}\|_1 \|\sum_{t\geq 0} \boldsymbol{G}_{11}^t\|_\infty \\
&\leq \frac{2+\gamma}{1-\alpha}\alpha^{N+1},
\end{aligned}$$

which is precisely (3.7). QED

Several remarks are due. A rough estimate on the size of t and N needed to have tolerance level 10^{-6} in both (3.6) and (3.7) yields that

$$10^{-6} \approx \frac{2}{1-\alpha}\alpha^{2(t-1)}, \quad 10^{-6} \approx \frac{2(3-\alpha^2)}{(1-\alpha)(1-\alpha^2)}\alpha^{N+1},$$

so that

$$\begin{aligned}
t &\approx 1 + \frac{-6 - {}^{10}\log 2 + {}^{10}\log(1-\alpha)}{2 \cdot {}^{10}\log \alpha} \approx 50 \\
N &\approx \frac{-6 - {}^{10}\log 2 - {}^{10}\log(3-\alpha^2) + {}^{10}\log(1-\alpha) - {}^{10}\log(1-\alpha^2)}{{}^{10}\log \alpha} - 1 \approx 112.
\end{aligned}$$

Note first that the derived bounds are independent of the choice of set A, as long as $\mathbf{v} \in A$.

Further, the algorithm only considers pages at maximum distance N to pages of A, measured along paths not crossing A. The small world effect tells that the average connected distance in the web graph was of the order 16 in 2000 [4]. On the other hand, in [4] it is reported that with probability at least 3/4 randomly chosen starting and final nodes are not connected in the web graph. By judiciously choosing non-interlinked pages in A, the small world effect is expected to be largely annulled. That is, the average connected distance in the web subgraph on the nodes of $\mathcal{V} \setminus A$ is expected to increase with increasing set A. We then expect that the number of pages at maximum distance N from pages of A, measured along paths not passing A, is relatively small. At the same time, with increasing A, the iterates of the resulting matrix $\boldsymbol{G}_{11}$ are expected to have a larger convergence rate than α. A smaller value of N then would suffice to obtain the same tolerance level. Obviously, the above algorithms require testing.

Deviation matrix and group inverse As a last remark, we would like to point out that the deviation matrix $\hat{D}$ (equation (3.8)) is the group inverse of the matrix $\mathbf{I}-\hat{M}$ (cf. [15]). For discussing these notions, consider a Markov chain with transition matrix P, which is aperiodic. Let $\Pi = \lim_{n\to\infty} P^{(n)}$ be the stationary matrix. Then the deviation matrix $D = \sum_{n\geq 0}(P^{(n)} - \Pi)$ is the unique matrix satisfying

$$\Pi D = D\Pi = 0, \quad (\mathbf{I} - P)D = D(\mathbf{I} - P) = \mathbf{I} - \Pi.$$

If B is a set of states containing precisely one state of each closed class, then D has the following taboo decomposition (cf. [9])

$$D = (\mathbf{I} - \Pi)\sum_{n\geq 0} {}_B P^{(n)}(\mathbf{I} - \Pi).$$

The group inverse $Z^{\#}$ of the matrix $Z = \mathbf{I} - P$ is the unique matrix satisfying [15]

$$ZZ^{\#}Z = Z, \quad Z^{\#}ZZ^{\#} = Z^{\#}, \quad ZZ^{\#} = Z^{\#}Z.$$

It is directly checked that $Z^{\#} = D$. Consider a perturbation Q of P (with Q a stochastic matrix). Let Π_Q be the corresponding stationary matrix. Suppose that the number of closed classes in both Markov chains equals 1. Then the stationary matrices Π and Π_Q each have all rows indentically equal to their respective stationary distributions π and π_Q. In [7] (cf. also [8]) it is shown that

(e:dev3)
$$\pi_Q - \pi = \pi^T \sum_{t\geq 1}((Q - P)D)^t = \pi^T(Q - P)D(\mathbf{I} - (Q - P)D)^{-1}. \tag{3.16}$$

In case that Q and P differ only in one row, it is straightforward to check that (3.16) reduces to the perturbation formula in [13] Theorem 3.1. It is interesting to note that Langville and Meyer claim that one should perform sequential row updates to allow more general perturbations. Apparently this is related to the fact that the perturbation formula in [13] is based on the Sherman-Morrison formula. On the other hand, (3.16) (and (3.9)) holds regardless of the size of the perturbation.

References

[1] `http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html`

[2] S. Asmussen (1987), *Applied Probability and Queues*. J.Wiley&Sons, Chichester.

[3] K. Avrachenkov and N. Litvak (2006), The effect of new links on Google PageRank. *Stochastic Models* **22**, 319–331.

[4] A. Broder, R. Kumar, F. Mahgul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener (2000), Graph structure in the Web. *Computer Networks* **33**, 309–320.

[5] K.L. Chung (1960), *Markov Chains with Stationary Transition Probabilities*. Springer-Verlag, Berlin.

[6] N. Eiron, K.S. McCurley and J.A. Tomlin (2004), Ranking the Web frontier. In: *Proc. Thirteenth Int. World Wide Web Conf.*, ACM Press, 309-319.

[7] B. Heidergott, A. Hordijk and M. van Uitert (2007), Series expansions for finite-state Markov chains. *Prob. Engin. Inf. Sciences* **21**, 381–400.

[8] B. Heidergott (2008), Perturbation analysis of Markov chains. In: *Int. Workshop on DES (WODES'08) Göteborg*, Sweden, 99–104.

[9] A. Hordijk and F.M. Spieksma (1994), A new formula for the deviation matrix. In: *Probability, Statistics and Optimisation: a Tribute to Peter Whittle*, 497–507, F. P. Kelly (Edt.), Wiley, New York.

[10] I.C.F. Ipsen and S. Kirkland (2006), Convergence analysis of a PageRank updating algorithm by Langville and Meyer. *SIAM J. Matrix Anal. Appl.* **27**, 952–967.

[11] I.C.F. Ipsen and T.M. Selee (2007), PageRank computation, with special attention to dangling nodes. *SIAM J. Matrix Anal. Appl.* **29**, 1281–1296.

[12] S. Karlin (1966), *A First Course in Stochastic Processes*. Academic Press, New York.

[13] A.M. Langville and C.D. Meyer (2006), Updating Markov chains with an eye on Google's PageRank. *SIAM J. Matrix Anal. Appl.* **27**, 968–987.

[14] A.M. Langville and C.D. Meyer (2006), *Google's PageRank and Beyond*. Princeton University Press, Princeton.

[15] A.M. Langville and C.D. Meyer (2006), Fiddling with PageRank. *Proc. Markov Anniversary Meeting*. Boson Press, Nashville.

[16] C.D. Meyer (1989), Stochastic complementation, uncoupling Markov chains and the theory of nearly irreducible systems. *SIAM Review* **31**, 240–272.