symmetry

MDPI

Article

A Note on a Minimal Irreducible Adjustment Pagerank

Yuehua Feng [1], Yongxin Dong [2,*] and Jianxin You [2]

[1] School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China
[2] School of Economics and Management, Tongji University, Shanghai 200092, China
* Correspondence: dy893291318@163.com or dongyx@tongji.edu.cn

Abstract: The stochastic modification and irreducible modification in PageRank produce large web link changes correspondingly. To get a minimal irreducible web link adjustment, a PageRank model of minimal irreducible adjustment and its lumping method are discussed by Li, Chen, and Song. In this paper, we provide alternative proofs for the minimal irreducible PageRank by a new type of similarity transformation matrices. To further provide theorems and fast algorithms on a reduced matrix, an 4×4 block matrix partition case of the minimal irreducible PageRank model is utilized and analyzed. For some real applications of our results, a lumping algorithm used for speeding up PageRank vector computations is also presented. Numerical results are also reported to show the efficiency of the proposed algorithm.

Keywords: PageRank; minimal irreducible adjustment; dangling node; lumping; similarity transformation matrix

MSC: 15A06; 15A18; 15A21; 65F10; 65F15; 65B99

Citation: Feng, Y.; Dong, Y.; You, J. A Note on a Minimal Irreducible Adjustment PageRank. *Symmetry* **2022**, *14*, 1640. https://doi.org/10.3390/sym14081640

Academic Editors: Qing-Wen Wang, Zhuo-Heng He, Xuefeng Duan, Xiao-Hui Fu and Guang-Jing Song

Received: 17 July 2022
Accepted: 6 August 2022
Published: 9 August 2022

1. Introduction

In modern Internet search engine and web information retrieval domains, a good ranking method is extremely important due to the fact that it reveals the relative importance of the corresponding web page. As one of the Markov chain applications in web information retrieval, PageRank is such a famous web ranking method. It assumes that the importance of a web page mainly depends on the relative importance of the pages linking to it [1]. In matrix expression, the iterative procedure of the raw PageRank computation is defined by

$$\pi^{(k+1)T} = \pi^{(k)T} H, \text{ where } H \in \mathbb{R}^{n\times n}, \quad \pi^{(k)} \geq 0, \quad \pi^{(k)T} e = 1, \quad k = 0, 1, 2, \cdots, \tag{1}$$

where the matrix $H \in \mathbb{R}^{n\times n}$ models the web link structure matrix, $\pi^{(k)}$, $k = 0, 1, 2, \cdots$ is an iterative vector whose entries represent the importance or weight of each page [2,3]. The web link structure matrix H is given by

$$h_{ij} = \begin{cases} \dfrac{1}{n_i}, & \text{if page } i \text{ links to page } j, \\ 0, & \text{otherwise,} \end{cases}$$

where n_i, $i = 1, 2, \cdots, n$ stands for the number of outlinks of page i. As (1) can not insure the uniqueness of the PageRank vector, Page and Brin design an $n \times n$ matrix by two rank-1 modification procedures [1]

$$G = \alpha S + (1 - \alpha) e v^T, \text{ where } S = H + d w^T, \text{ the damping factor } \alpha \in (0, 1), \tag{2}$$

where the nonnegative vector of unit length $w \in \mathbb{R}^n$ is the dangling node vector, $e \in \mathbb{R}^n$ is a vector of all ones, the nonnegative vector v ($v \in \mathbb{R}^n$, $\|v\|_1 = 1$) is a personalization vector, and the entries of the dangling node indicator vector d are given by

$$d_i = \begin{cases} 1, & \text{if page } i \text{ does not link to other pages,} \\ 0, & \text{otherwise.} \end{cases}$$

As we know, some web pages may not have links to other web pages. If web pages have no outlinks, they are called dangling nodes; otherwise, they are called nondangling nodes, such as videos, jpg files, or pdf files are common dangling nodes in web pages. Thus, the above stochastic modification is necessary. The heart of the unique eigenvector is the irreducibility issue; therefore, the matrix ev^T (where e and v are defined in (2)) is further used to handle this uniqueness problem. The classical PageRank vector is defined by the principle eigenvector of the Google matrix G with unit 1-norm. It is also to find a stationary probability distribution vector of a Markov chain. In other words, the PageRank computation problem is defined by

$$\pi^T G = \pi^T,\ G \in \mathbb{R}^{n\times n}, \pi \geq 0,\ \pi^T e = 1.$$

However, it may take much time merely to compute a large PageRank vector, since hyperlink matrices involved stand for even over a billion pages' link relationships [1,4]. Hence, faster methods are increasingly heated topics. The Arnoldi method can estimate the first few eigenvalues for nonsymmetric eigenvalue problems, while it may cost expensively in each iteration. The inverse iteration is not suitable for PageRank since it may need $O(n^3)$ operations by finding an inversion of a large web link matrix. Instead, methods based on the power method are increasingly attractive recently. The damping factor represents the coupling degree of the corresponding Markov chain. A large value of the damping factor stands for a nearly coupled Markov chain. However, as the damping factor increases, the power method converges slowly [3,5,6]. The lumping or reordering method takes full advantage of the zero entries in the Google matrix [2]. Some recent methods for computing PageRank effectively involve Chebyshev polynomial techniques, matrix splitting methods [7,8], the lumping method [9,10], coupled iteration algorithm [11], Hessenberg-type method [12], a simpler GMRES accelerated by Chebyshev polynomials [13], and hybrid methods [14]. Other accelerated methods, theoretical and numerical results are available, see [7,15,16]. Adding dangling nodes' links to other web sites will change links between web pages maximally. If the number of web pages expands, there will be more dangling node web pages. In addition, every node can be directly connected to other nodes by adding the matrix ev^T; hence, in this case, irreducibility is enforced trivially. In addition, the web link change is increasingly large, so the adjustment (2) is called maximal irreducible adjustment. For more details of the PageRank model and computation, see [1–3,17].

Our recent work provides relatively easy alternative proofs for classical PageRank theoretical analyses by a class of new similarity matrices [18,19]. One may wonder whether the minimal irreducible PageRank model still has similar theoretical results. To further reduce the computational overhead, we discuss a level 4 partition of the minimal irreducible PageRank model matrix. In the level 4 partition of the minimal irreducible Google matrix, we provide a smaller lumping matrix and take full advantage of zero entries in the minimal irreducible Google matrix.

The remainder of this paper is as follows: Section 2 discusses the main theorems of the minimal irreducible PageRank model in [20]. Section 3 presents a class of new similarity matrices which can be used in theoretical analysis. Section 4 provides alternative proofs for the main results due to Li, Chen, and Song [20]. Section 5 discusses the 4×4 level case, and one can derive a solution vector expression by further partitioning the minimal irreducible Google matrix. Hence, the matrix computation dimension involved is reduced and the corresponding computation process can speed up. Finally, Section 6 gives some conclusions of this paper. Throughout the paper, we first revisit the theoretical results of

Li, Chen, and Song on a minimal irreducible PageRank model [20]. In addition, Matlab notation is used if necessary.

2. Some Theoretical Results on a Minimal Irreducible PageRank Model

In this section, we focus on underlying theoretical and computational contributions on a kind of minimal irreducible PageRank by the lumping method. Hence, a minimal irreducible PageRank model is discussed below.

2.1. A Minimal Irreducible PageRank Model

If all dangling nodes are lumped into a single node, then the Google matrix is said to be lumpable [9,21]. In general, the matrix M is lumpable with respect to the partition if

$$M_{ij}e = \mu e, \quad \text{where the scalar } \mu \neq 0, \ e = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T,$$

where $i \neq j$, $1 \leq i, j \leq k+1$, and the matrix

$$PMP^T = \begin{bmatrix} M_{11} & \cdots & M_{1,k+1} \\ \vdots & \ddots & \vdots \\ M_{k+1,1} & \cdots & M_{k+1,k+1} \end{bmatrix},$$

and P is a proper permutation matrix.

Suppose that $H \in \mathbb{R}^{n\times n}$ has k nondangling nodes and $n-k$ dangling nodes, thus there exists a permutation matrix $\Pi \in \mathbb{R}^{n\times n}$, such that

$$\Pi H \Pi^T = \begin{bmatrix} H_{11} & H_{12} \\ 0 & 0 \end{bmatrix}.$$

Hence,

$$\begin{aligned} A = \Pi G \Pi^T =& \alpha \bar{S} + (1-\alpha) e v^T \\ =& \alpha \begin{bmatrix} H_{11} & H_{12} \\ e w_1^T & e w_2^T \end{bmatrix} + (1-\alpha) \begin{bmatrix} e v_1^T & e v_2^T \\ e v_1^T & e v_2^T \end{bmatrix} \\ =& \begin{bmatrix} \alpha H_{11} + (1-\alpha) e v_1^T & \alpha H_{12} + (1-\alpha) e v_2^T \\ e(\alpha w_1^T + (1-\alpha) v_1^T) & e(\alpha w_2^T + (1-\alpha) v_2^T) \end{bmatrix} \\ =& \begin{bmatrix} A_{11} & A_{12} \\ e u_1^T & e u_2^T \end{bmatrix}, \text{ where } \bar{S} = \Pi S \Pi^T, \end{aligned} \tag{3}$$

where the dangling node vector w and personalization vector v are partitioned into $w_1 \in \mathbb{R}^k$, $w_2 \in \mathbb{R}^{n-k}$ and $v_1 \in \mathbb{R}^k$, $v_2 \in \mathbb{R}^{n-k}$. Moreover, if

$$\bar{\pi}^T = \bar{\pi}^T A = \bar{\pi}^T \Pi G \Pi^T, \text{ with } \bar{\pi} \geq 0, \ \|\bar{\pi}\|_1 = 1, \tag{4}$$

where $A = \Pi G \Pi^T$, $G \in \mathbb{R}^{n\times n}$ is the classical Google matrix, and $\bar{\pi}$ is the stationary distribution of A, then the PageRank vector corresponding to (4) is defined by

$$\pi^T = \bar{\pi}^T \Pi.$$

In [20], Li, Chen, and Song discussed a minimal irreducible Google matrix by changing $\bar{S}$ to a bordered matrix

$$\widetilde{S} = \begin{bmatrix} \frac{n}{n+1}\bar{S} & \frac{1}{n+1}e \\ \frac{1}{n+1}e^T & \frac{1}{n+1} \end{bmatrix} = \begin{bmatrix} \frac{n}{n+1}H_{11} & \frac{n}{n+1}H_{12} & \frac{1}{n+1}e \\ \frac{n}{n+1}e w_1^T & \frac{n}{n+1}e w_2^T & \frac{1}{n+1}e \\ \frac{1}{n+1}e^T & \frac{1}{n+1}e^T & \frac{1}{n+1} \end{bmatrix}. \tag{5}$$

The new PageRank vector of order $n+1$ is the solution vector defined by the principle eigenvector of the new Google matrix $\widetilde{S}$,

$$\widetilde{\pi}^T = \widetilde{\pi}^T \widetilde{S}, \text{ or } \widetilde{\pi} = \widetilde{S}^T \widetilde{\pi}, \text{ with } \widetilde{\pi} \geq 0, \ \|\widetilde{\pi}\|_1 = 1.$$

We note that this modification only increases one additional row and column to ensure that the link matrix is strongly connected, in order to ensure that the new Google matrix $\widetilde{S}$ is irreducible. In this way, the Perron–Frobenius theorem can guarantee the uniqueness of a positive principle eigenvector $\widetilde{\pi}$ with unit length ($\|\widetilde{\pi}\|_1 = 1$). Note that the decomposition presented in (5) has been studied before in various forms. See, e.g., [22,23]. In Section 4.3, we mainly studied the first two layers of such decompositions. In addition, another way of creating a minimal irreducible Google matrix is defined by [24,25]

$$\widetilde{S}' = \begin{bmatrix} \alpha\bar{S} & (1-\alpha)e \\ v^T & 0 \end{bmatrix} = \begin{bmatrix} \alpha H_{11} & \alpha H_{12} & (1-\alpha)e \\ \alpha e w_1^T & \alpha e w_2^T & (1-\alpha)e \\ v^T & v^T & 0 \end{bmatrix}. \tag{6}$$

In this paper, we mainly discuss the minimal irreducible Google matrix (5). In addition, we will briefly discuss our results on (6) and leave it as a further question.

In the following, in order to lump the stochastic matrix and show our theoretical results conveniently, we first permutate the rows and columns of $\widetilde{S}$ simultaneously. Thus, we first permutate $\widetilde{S}$, such that

$$\widehat{S} = \widetilde{\Pi}\widetilde{S}\widetilde{\Pi}^T = \begin{bmatrix} \frac{n}{n+1}H_{11} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12} \\ \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{1}{n+1}e^T \\ \frac{n}{n+1}ew_1^T & \frac{1}{n+1}e & \frac{n}{n+1}ew_2^T \end{bmatrix},$$

where the corresponding permutation matrix

$$\widetilde{\Pi} = \begin{bmatrix} I_k & 0 & 0 \\ 0 & 0 & 1 \\ 0 & I_{n-k} & 0 \end{bmatrix}.$$

2.2. Some Related Theorems on the Minimal Irreducible PageRank Model

The similarity transformation matrix

$$L = I_{n-k} - \frac{1}{n-k}\widehat{e}e^T, \quad \text{where} \quad \widehat{e} = e - e_1 = \left[0, 1, \cdots, 1\right]^T \in \mathbb{R}^{n-k}$$

is employed by Ipsen and Selee to analyze the theoretical results, and the identity matrix of order n is denoted by $I_n = \begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix}$, where e_i, $i = 1, 2, \cdots, n$ is the ith column vector of I_n. If large scale computation problems can be reduced to a relatively small scale, the computations can be simplified and accelerated. Fortunately, lumping is one of these methods. In [20], Li, Chen, and Song obtained the following theorems for the new Google matrix. They showed that the minimal irreducible Google matrix can also be lumped. In addition, the spectral distribution of the new Google matrix, the relationship between the eigenvector part associated with the dangling nodes, and the eigenvector part associated with the nondangling nodes is as follows.

Theorem 1 ([20,26]). *Given the stochastic matrix $\bar{S}$ with spectrum $\{1, \lambda_2, \lambda_3, \cdots, \lambda_n\}$. Then, the spectral of the matrix*

$$\widetilde{S} = \begin{bmatrix} \frac{n}{n+1}\bar{S} & \frac{1}{n+1}e \\ \frac{1}{n+1}e^T & \frac{1}{n+1} \end{bmatrix}$$

is $\left\{1, \frac{n}{n+1}\lambda_2, \frac{n}{n+1}\lambda_3, \cdots, \frac{n}{n+1}\lambda_n, 0\right\}$.

Theorem 2 ([20]). *With the above notation, let*

$$X = \begin{bmatrix} I_k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & L \end{bmatrix}, \quad \text{where} \quad L = I_{n-k} - \frac{1}{n-k}\widehat{e}e^T \quad \text{and} \quad \widehat{e} = e - e_1 = \begin{bmatrix} 0 & 1 & \cdots & 1 \end{bmatrix}^T.$$

Then,

$$X\widehat{S}X^{-1} = \begin{bmatrix} \widehat{S}^{(1)} & \widehat{S}^{(2)} \\ 0 & 0 \end{bmatrix}, \quad \text{where} \quad \widehat{S}^{(1)} = \begin{bmatrix} \frac{n}{n+1}H_{11} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}e \\ \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{n-k}{n+1} \\ \frac{n}{n+1}w_1^T & \frac{1}{n+1} & \frac{n}{n+1}w_2^T e \end{bmatrix},$$

so that $\widehat{S}$ *has at least* $n-k-1$ *zero eigenvalues.*

Theorem 3 ([20]). *With the above notation, let*

$$\sigma^T \begin{bmatrix} \frac{n}{n+1}H_{11} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}e \\ \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{n-k}{n+1} \\ \frac{n}{n+1}w_1^T & \frac{1}{n+1} & \frac{n}{n+1}w_2^T e \end{bmatrix} = \sigma^T, \ \sigma \in \mathbb{R}^{k+2} \geq 0, \ \|\sigma\|_1 = 1,$$

and partition $\sigma^T = \begin{bmatrix} \sigma_{1:k}^T & \sigma_{k+1} & \sigma_{k+2} \end{bmatrix}$, *where* σ_{k+1} *and* σ_{k+2} *are scalars. Then, the PageRank vector of* $\widetilde{S}$ *equals*

$$\widetilde{\pi}^T = \begin{bmatrix} \sigma_{1:k}^T & \sigma^T \begin{bmatrix} \frac{n}{n+1}H_{12} \\ \frac{1}{n+1}e^T \\ \frac{n}{n+1}w_2^T \end{bmatrix} & \sigma_{k+1} \end{bmatrix}.$$

3. New Similarity Transformation Matrix

In order to further study the lumping method and make contribution to the PageRank computation, we derive alternative proofs for related theorems by using new similarity transformation matrices [18,19] and try to establish efficient algorithms to compute PageRank. To discuss this topic, we begin to introduce some related work.

Suppose that λ_i is the eigenvalues of $A \in \mathbb{R}^{n\times n}$ ordered as $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n| \geq 0$. In [27], Ding and Zhou proposed the following rank-1 updated spectral theorem.

Theorem 4 ([27]). *Let* u *and* v *be two* $n-$*dimensional column vectors such that* u *is an eigenvector of* A *associated with the eigenvalue* λ_1. *Then, the eigenvalues of* $A + uv^T$ *are* $\{\lambda_1 + v^T u, \lambda_2, \cdots, \lambda_n\}$, *counting algebraic multiplicity.*

Langville and Meyer in their book [3] analyze the spectral distribution of the classical Google matrix by employing a similarity transformation method. To find the relationship between PageRank of the nondangling nodes π_1 and that of dangling nodes π_2, Ipsen and Selee in [9] employed the similarity transformation matrix

$$X = \begin{bmatrix} I_k & 0 \\ 0 & L \end{bmatrix}, \ L = I_{n-k} - \frac{1}{n-k}\widehat{e}e^T, \ \widehat{e} = e - e_1 = \begin{bmatrix} 0 & 1 & \cdots & 1 \end{bmatrix}^T. \tag{7}$$

As it is pointed in [18], the invertible matrix L exists such that the related theorem holds. However, the matrix L is not unique. In the general case of the Google matrix, if the matrix L satisfies the condition

$$Le = e_1, \text{ where } e \text{ is a vector of all ones,} \tag{8}$$

then the proof process can be more simpler and shortened. Motivated by the above conclusion, we further study the case in the new PageRank model. In Theorem 2, if we exploit the matrix [18]

$$\widetilde{X} = \begin{bmatrix} I_k & 0 \\ 0 & \widetilde{L} \end{bmatrix}, \text{ where } \widetilde{L} = I_{n-k} - \widehat{e}e_1^T, \ \widehat{e} = e - e_1 = \begin{bmatrix} 0 & 1 & \cdots & 1 \end{bmatrix}^T, \tag{9}$$

and we separate the first row and column of $\widetilde{L}$

$$\widetilde{L} = \begin{bmatrix} 1 & 0 \\ -e & I_{n-k-1} \end{bmatrix}, \tag{10}$$

and partition L in (7) conformally with $\widetilde{L}$,

$$L = \begin{bmatrix} 1 & 0 \\ -\frac{1}{n-k}e & I_{n-k-1} - \frac{1}{n-k}ee^T \end{bmatrix}, \tag{11}$$

where the 2-by-2 block of L in (11) is

$$I_{n-k-1} - \frac{1}{n-k}ee^T = \begin{bmatrix} \frac{n-k-1}{n-k} & \frac{-1}{n-k} & \cdots & \frac{-1}{n-k} \\ \frac{-1}{n-k} & \frac{n-k-1}{n-k} & \cdots & \frac{-1}{n-k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{-1}{n-k} & \frac{-1}{n-k} & \cdots & \frac{n-k-1}{n-k} \end{bmatrix}.$$

Unfortunately, it is a typical dense matrix; therefore, it is complex when we analyze it, and it is expensive when we compute or store it. Furthermore, the matrix (10) is actually [18]

$$\widetilde{L} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -1 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

because $\widetilde{L}$ is a triangular matrix while L is a dense matrix. Inspired and motivated by the easier similarity transformation matrix $\widetilde{L}$, we consider replacing L with $\widetilde{L}$ in X so that an alternative proof process can be presented, and its process can be simplified and shortened. Moreover, if

$$\widetilde{L} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

in (9), then the proof process of the above theorems also holds. In other words, this similarity transformation matrix is not unique at all. Instead of the matrix L used in [9], we propose a class of invertible matrices $\widehat{L}$, which satisfy the condition (8). Particularly, if $\widehat{L} = L$, it becomes the similarity transformation matrix proposed by Ipsen and Selee in [9]. In the following, as we mainly discuss the minimal irreducible PageRank matrix (5), one may wonder "Could you apply your methods on the results in [24,25]?" Thus, we provide a remark below.

Remark 1. *If we permutate $\widetilde{S}'$ by the permutation matrix $\widetilde{\Pi}$, such that $\widetilde{S}'$ is reordered as*

$$\begin{bmatrix} \alpha H_{11} & (1-\alpha)e & \alpha H_{12} \\ v_1^T & 0 & v_2^T \\ \alpha e w_1^T & (1-\alpha)e & \alpha e w_2^T \end{bmatrix}.$$

Performing a similarity transformation on it, we yield

$$\begin{aligned}
&\begin{bmatrix} I & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & L \end{bmatrix} \begin{bmatrix} \alpha H_{11} & (1-\alpha)e & \alpha H_{12} \\ v_1^T & 0 & v_2^T \\ \alpha e w_1^T & (1-\alpha)e & \alpha e w_2^T \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & L^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \alpha H_{11} & (1-\alpha)e & \alpha H_{12}L^{-1} \\ v_1^T & 0 & v_2^t L^{-1} \\ \alpha L e w_1^T & (1-\alpha)Le & \alpha L e w_2^T L^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \alpha H_{11} & (1-\alpha)e & \alpha H_{12}L^{-1}e_1 & \alpha H_{12}L^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ v_1^T & 0 & v_2^T L^{-1}e_1 & v_2^T L^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ \alpha w_1^T & (1-\alpha) & \alpha w_2^T L^{-1}e_1 & \alpha w_2^T L^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ 0 & 0 & 0 & 0 \end{bmatrix},
\end{aligned}$$

where the invertible matrix L satisfies $Le = e_1$, $e_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$, *and e is a column vector of all ones. Therefore, we preliminarily judge that the new similarity matrix results and theoretical results are also suitable for this form of minimal irreducibility Google matrix. More detailed work is under consideration.*

4. Alternative Proofs

In this section, we will analyze theoretical results of the minimal irreducible PageRank model based on the similarity transformation matrix (9). Here, we mainly provide alternative proofs for the minimal irreducible PageRank model.

4.1. An Alternative Proof of Theorem 1

In [20], Li, Chen, and Song completed the proof of Theorem 1 by using the determinant property: That is,

$$det\left(\begin{bmatrix} \frac{n}{n+1}\bar{S} & \frac{1}{n+1}e \\ \frac{1}{n+1}e^T & \frac{1}{n+1} \end{bmatrix} - \lambda I_{n+1} \right) = \lambda(\lambda-1)\left(\frac{n}{n+1}\lambda_2 - \lambda\right)\cdots\left(\frac{n}{n+1}\lambda_n - \lambda\right),$$

Thus, one can conclude that the eigenvalues of $\widetilde{S}$ is $\{1, \frac{n}{n+1}\lambda_2, \cdots, \frac{n}{n+1}\lambda_n, 0\}$. Here, we show an alternative proof for Theorem 1 by using the similarity transformation matrix.

Proof. A similarity transformation is done as follows:

$$\begin{aligned}
\begin{bmatrix} I & \frac{1}{n}e \\ 0 & 1 \end{bmatrix} \widetilde{S} \begin{bmatrix} I & -\frac{1}{n}e \\ 0 & 1 \end{bmatrix} &= \begin{bmatrix} I & \frac{1}{n}e \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{n}{n+1}\bar{S} & \frac{1}{n+1}e \\ \frac{1}{n+1}e^T & \frac{1}{n+1} \end{bmatrix} \begin{bmatrix} I & -\frac{1}{n}e \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{n}{n+1}\bar{S} + \frac{1}{n(n+1)}ee^T & 0 \\ \frac{1}{n+1}e^T & 0 \end{bmatrix}.
\end{aligned}$$

The eigenvalues of the stochastic matrix $\bar{S}$ are denoted by $\{1, \lambda_2, \lambda_3, \cdots, \lambda_n\}$. By Theorem 4, we obtain that the eigenvalues of $\widetilde{S}$ are

$$\frac{1}{n+1}\left(n + \frac{1}{n}\cdot n\right) = 1, \frac{n}{n+1}\lambda_2, \frac{n}{n+1}\lambda_3, \cdots, \frac{n}{n+1}\lambda_n,$$

and the remaining eigenvalue is 0. Thus, we complete the proof. □

4.2. An Alternative Proof of Theorem 2

Li, Chen, and Song [20] present the proof of Theorem 2 by using the similarity transformation matrix (7). Now, we validate Theorem 2 by a general invertible matrix $\widehat{L}$ satisfying $\widehat{L}e = e_1$ below.

Proof. Since

$$\widetilde{X}^{-1} = \begin{bmatrix} I_k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \widetilde{L}^{-1} \end{bmatrix}, \quad \text{where} \quad \widetilde{L} = I_{n-k} - \widehat{e}e_1^T \quad \text{and} \quad \widetilde{L}^{-1} = I_{n-k} + \widehat{e}e_1^T,$$

then

$$\begin{aligned}
\widetilde{X}\widehat{S}\widetilde{X}^{-1} &= \begin{bmatrix} \frac{n}{n+1}H_{11} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}\widetilde{L}^{-1} \\ \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{1}{n+1}e^T\widetilde{L}^{-1} \\ \frac{n}{n+1}\widetilde{L}ew_1^T & \frac{1}{n+1}\widetilde{L}e & \frac{n}{n+1}\widetilde{L}ew_2^T\widetilde{L}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \frac{n}{n+1}H_{11} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}\widetilde{L}^{-1}e_1 & \frac{n}{n+1}H_{12}\widetilde{L}^{-1}\begin{bmatrix} e_2 & e_3 & \cdots & e_{n-k} \end{bmatrix} \\ \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{1}{n+1}e^T\widetilde{L}^{-1}e_1 & \frac{1}{n+1}e^T\widetilde{L}^{-1}\begin{bmatrix} e_2 & e_3 & \cdots & e_{n-k} \end{bmatrix} \\ \frac{n}{n+1}e_1^T\widetilde{L}ew_1^T & \frac{1}{n+1}e_1^T\widetilde{L}e & \frac{n}{n+1}e_1^T\widetilde{L}ew_2^T\widetilde{L}^{-1}e_1 & \frac{n}{n+1}e_1^T\widetilde{L}ew_2^T\widetilde{L}^{-1}\begin{bmatrix} e_2 & e_3 & \cdots & e_{n-k} \end{bmatrix} \\ 0 & 0 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \frac{n}{n+1}H_{11} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}e & \frac{n}{n+1}H_{12}\widetilde{L}^{-1}\begin{bmatrix} e_2 & e_3 & \cdots & e_{n-k} \end{bmatrix} \\ \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{n-k}{n+1} & \frac{1}{n+1}e^T\widetilde{L}^{-1}\begin{bmatrix} e_2 & e_3 & \cdots & e_{n-k} \end{bmatrix} \\ \frac{n}{n+1}w_1^T & \frac{1}{n+1} & \frac{n}{n+1}w_2^Te & \frac{n}{n+1}w_2^T\widetilde{L}^{-1}\begin{bmatrix} e_2 & e_3 & \cdots & e_{n-k} \end{bmatrix} \\ 0 & 0 & 0 & 0 \end{bmatrix},
\end{aligned}$$

where $e = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T \in \mathbb{R}^{n-k}$, $e^Te = n-k$, $\widetilde{L}^{-1}e_1 = e$, and the last quality follows form $\widetilde{L}e = e_1$ (thus, $e_1^T\widetilde{L}e = 1$). □

Remark 2. *During the above proof of Theorem 2, if $\widetilde{X}$ is generalized to*

$$\widehat{X} = \begin{bmatrix} I_k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \widehat{L} \end{bmatrix}, \text{ where an invertible matrix } \widehat{L} \text{ satisfies } \widehat{L}e = e,$$

the proof also holds.

4.3. An Alternative Proof of Theorem 3

In this alternative proof, the similarity matrix is chosen as $\widehat{X}$. A general invertible matrix $\widehat{L}$ satisfying $\widehat{L}e = e_1$ is applied to validate Theorem 3.

Proof. If

$$\widehat{X}\widehat{S}\widehat{X}^{-1} = \begin{bmatrix} \widehat{S}^{(1)} & \widehat{S}^{(2)} \\ 0 & 0 \end{bmatrix}, \quad \text{where} \quad \widehat{S}^{(1)} = \begin{bmatrix} \frac{n}{n+1}H_{11} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}e \\ \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{n-k}{n+1} \\ \frac{n}{n+1}w_1^T & \frac{1}{n+1} & \frac{n}{n+1}w_2^Te \end{bmatrix},$$

then the stochastic matrix $\widehat{S}^{(1)}$ of order $k+2$ has the same nonzero eigenvalues as $\widehat{S}$. We obtain that

$$\begin{bmatrix} \sigma^T & \sigma^T\widehat{S}^{(2)} \end{bmatrix}$$

is an eigenvector for $\widehat{X}\widehat{S}\widehat{X}^{-1}$ associated with the eigenvalue $\lambda = 1$. This follows from

$$\begin{bmatrix} \sigma^T & \sigma^T\widehat{S}^{(2)} \end{bmatrix}\begin{bmatrix} \widehat{S}^{(1)} & \widehat{S}^{(2)} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sigma^T\widehat{S}^{(1)} & \sigma^T\widehat{S}^{(2)} \end{bmatrix} = \begin{bmatrix} \sigma^T & \sigma^T\widehat{S}^{(2)} \end{bmatrix}.$$

Therefore,

$$\widehat{\pi}^T = \begin{bmatrix} \sigma^T & \sigma^T\widehat{S}^{(2)} \end{bmatrix}\widehat{X}$$

is an eigenvector of $\widehat{S}$ associated with $\lambda = 1$. Since $\widehat{S}$ and $\widehat{S}^{(1)}$ have the same nonzero eigenvalues, and the principle eigenvalue of S or $\widehat{S}$ is distinct, the stationary probability distribution σ of $\widehat{S}^{(1)}$ is unique. We repartition

$$\widehat{\pi}^T = \begin{bmatrix} \sigma_{1:k}^T & \sigma_{k+1} & \begin{bmatrix} \sigma_{k+2} & \sigma^T \widehat{S}^{(2)} \end{bmatrix} \end{bmatrix} \begin{bmatrix} I_k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \widehat{L} \end{bmatrix}.$$

Multiplying out

$$\widehat{\pi}^T = \begin{bmatrix} \sigma_{1:k}^T & \sigma_{k+1} & \begin{bmatrix} \sigma_{k+2} & \sigma^T \widehat{S}^{(2)} \end{bmatrix} \widehat{L} \end{bmatrix}.$$

Since

$$\begin{aligned} \begin{bmatrix} \sigma_{k+2} & \sigma^T \widehat{S}^{(2)} \end{bmatrix} \widehat{L} &= \begin{bmatrix} \sigma^T \widehat{s}^{(1)} e & \sigma^T \widehat{s}^{(1)} \widehat{L}^{-1} \begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \end{bmatrix} \widehat{L} \\ &= \begin{bmatrix} \sigma^T \widehat{s}^{(1)} \widehat{L}^{-1} e_1 & \sigma^T \widehat{s}^{(1)} \widehat{L}^{-1} \begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \end{bmatrix} \widehat{L} \\ &= \sigma^T \widehat{s}^{(1)}, \end{aligned}$$

the first equation uses the fact that

$$\widehat{s}^{(1)} = \begin{bmatrix} \frac{n}{n+1} H_{12}^T & \frac{1}{n+1} e & \frac{n}{n+1} w_2 \end{bmatrix}^T, \quad \sigma_{k+2} = \sigma^T \widehat{s}^{(1)} e \quad \text{and} \quad \widehat{S}^{(2)} = \widehat{s}^{(1)} \widehat{L}^{-1} \begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix},$$

and the second equation uses $e = \widehat{L}^{-1} e_1$. Hence,

$$\widehat{\pi}^T = \begin{bmatrix} \sigma_{1:k}^T & \sigma_{k+1} & \sigma^T \widehat{s}^{(1)} \end{bmatrix}.$$

Furthermore,

$$\widetilde{\pi}^T = \widehat{\pi}^T \widetilde{\Pi} = \begin{bmatrix} \sigma_{1:k}^T & \sigma^T \widehat{s}^{(1)} & \sigma_{k+1} \end{bmatrix}, \quad \text{with} \quad \widehat{s}^{(1)} = \begin{bmatrix} \frac{n}{n+1} H_{12}^T & \frac{1}{n+1} e & \frac{n}{n+1} w_2 \end{bmatrix}^T.$$

As the claim $\widetilde{\pi}$ is unique, we conclude that $\widetilde{\pi}$ is the stationary probability of the PageRank model if $e^T \widetilde{\pi} = 1$. Therefore, we complete the proof. □

5. Results for 4-Level Matrix Partition

We will show that the nondangling nodes can also be further classified into two classes. The nodes which are nondangling but point to only the dangling nodes are further called "weakly nondangling" nodes. The other nondangling nodes are referred to as "strongly nondangling" nodes [10]. Therefore, the reduced matrix obtained by lumping dangling nodes can be further reduced by lumping weakly nondangling nodes, to another single node. In addition, the further reduced matrix is also stochastic with the same nonzero eigenvalues as the Google matrix.

5.1. Derivation

To lump dangling nodes and weakly nondangling nodes [10] to a level 4 matrix partition, we permutate the matrix S in (3) such that

$$\bar{S} = \Pi S \Pi^T = \begin{bmatrix} H_{11}^{(11)} & H_{11}^{(12)} & H_{12}^{(1)} \\ 0 & 0 & H_{12}^{(2)} \\ e w_{k_1}^T & e w_{k_2}^T & e w_2^T \end{bmatrix},$$

where $H_{11}^{(11)} \in \mathbb{R}^{k_1 \times k_1}$, $H_{11}^{(12)} \in \mathbb{R}^{k_1 \times k_2}$, $H_{12}^{(1)} \in \mathbb{R}^{k_1 \times (n-k)}$, $H_{12}^{(2)} \in \mathbb{R}^{k_2 \times (n-k)}$, $w_2 \in \mathbb{R}^{n-k}$, k_1, and k_2 are the number of strongly nondangling nodes and weakly nondangling nodes,

respectively. Moveover, $k = k_1 + k_2$ and $\bar{S}e = e$. After the adjustment by Li, Chen, and Song [20], the new Google matrix in 4×4 block form is

$$G = \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{11}^{(12)} & \frac{n}{n+1}H_{12}^{(1)} & \frac{1}{n+1}e \\ 0 & 0 & \frac{n}{n+1}H_{12}^{(2)} & \frac{1}{n+1}e \\ \frac{n}{n+1}ew_{k_1}^T & \frac{n}{n+1}ew_{k_2}^T & \frac{n}{n+1}ew_2^T & \frac{1}{n+1}e \\ \frac{1}{n+1}e^T & \frac{1}{n+1}e^T & \frac{1}{n+1}e^T & \frac{1}{n+1} \end{bmatrix}. \tag{12}$$

We define the permutation matrix

$$\Pi_1 = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & I & 0 \end{bmatrix}$$

to permutate the third and fourth row and column of G simultaneously. Performing $\Pi_1 G \Pi_1^T$ yields

$$\Pi_1 G \Pi_1^T = \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{11}^{(12)} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(1)} \\ 0 & 0 & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(2)} \\ \frac{1}{n+1}e^T & \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{1}{n+1}e^T \\ \frac{n}{n+1}ew_{k_1}^T & \frac{n}{n+1}ew_{k_2}^T & \frac{1}{n+1}e & \frac{n}{n+1}ew_2^T \end{bmatrix}.$$

Let

$$X = \begin{bmatrix} I_{k_1} & 0 & 0 & 0 \\ 0 & I_{k_2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \widehat{L} \end{bmatrix}, \quad \text{where } \widehat{L} \in \mathbb{R}^{(n-k)\times(n-k)} \text{ is an invertible matrix with } \widehat{L}e = e_1,$$

then

$$\begin{aligned} X\Pi_1 G \Pi_1^T X^{-1} &= \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{11}^{(12)} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(1)}\widehat{L}^{-1} \\ 0 & 0 & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(2)}\widehat{L}^{-1} \\ \frac{1}{n+1}e^T & \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{1}{n+1}e^T\widehat{L}^{-1} \\ \frac{n}{n+1}\widehat{L}ew_{k_1}^T & \frac{n}{n+1}\widehat{L}ew_{k_2}^T & \frac{1}{n+1}\widehat{L}e & \frac{n}{n+1}\widehat{L}ew_2^T\widehat{L}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{11}^{(12)} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(1)}\widehat{L}^{-1} \\ 0 & 0 & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(2)}\widehat{L}^{-1} \\ \frac{1}{n+1}e^T & \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{1}{n+1}e^T\widehat{L}^{-1} \\ \frac{n}{n+1}e_1w_{k_1}^T & \frac{n}{n+1}e_1w_{k_2}^T & \frac{1}{n+1}e_1 & \frac{n}{n+1}e_1w_2^T\widehat{L}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{11}^{(12)} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(1)}e & \frac{n}{n+1}H_{12}^{(1)}\widehat{L}^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ 0 & 0 & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(2)}e & \frac{n}{n+1}H_{12}^{(2)}\widehat{L}^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ \frac{1}{n+1}e^T & \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{n-k}{n+1} & \frac{1}{n+1}e^T\widehat{L}^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ \frac{n}{n+1}w_{k_1}^T & \frac{n}{n+1}w_{k_2}^T & \frac{1}{n+1} & \frac{n}{n+1}w_2^Te & \frac{n}{n+1}w_2^T\widehat{L}^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Denote by

$$G_1 = \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{11}^{(12)} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(1)}e \\ 0 & 0 & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(2)}e \\ \frac{1}{n+1}e^T & \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{n-k}{n+1} \\ \frac{n}{n+1}w_{k_1}^T & \frac{n}{n+1}w_{k_2}^T & \frac{1}{n+1} & \frac{n}{n+1}w_2^Te \end{bmatrix},$$

and then we permutate the second and fourth block row and column of G_1 simultaneously by a permutation matrix Π_2,

$$\Pi_2 G_1 \Pi_2^T = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & I & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{11}^{(12)} & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(1)}e \\ 0 & 0 & \frac{1}{n+1}e & \frac{n}{n+1}H_{12}^{(2)}e \\ \frac{1}{n+1}e^T & \frac{1}{n+1}e^T & \frac{1}{n+1} & \frac{n-k}{n+1} \\ \frac{n}{n+1}w_{k_1}^T & \frac{n}{n+1}w_{k_2}^T & \frac{1}{n+1} & \frac{n}{n+1}w_2^T e \end{bmatrix} \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & I & 0 & 0 \end{bmatrix}^T$$

$$= \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{12}^{(1)}e & \frac{1}{n+1}e & \frac{n}{n+1}H_{11}^{(12)} \\ \frac{n}{n+1}w_{k_1}^T & \frac{n}{n+1}w_2^T e & \frac{1}{n+1} & \frac{n}{n+1}w_{k_2}^T \\ \frac{1}{n+1}e^T & \frac{n-k}{n+1} & \frac{1}{n+1} & \frac{1}{n+1}e^T \\ 0 & \frac{n}{n+1}H_{12}^{(2)}e & \frac{1}{n+1}e & 0 \end{bmatrix}.$$

Denote by

$$X_1 = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \widehat{L} \end{bmatrix}, \quad \text{where } \widehat{L} \in \mathbb{R}^{k_2 \times k_2} \text{ is an invertible matrix with } \widehat{L}e = e_1.$$

Then,

$$X_1 \Pi_2 G_1 \Pi_2^T X_1^{-1} = \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{12}^{(1)}e & \frac{1}{n+1}e & \frac{n}{n+1}H_{11}^{(12)}\widehat{L}^{-1} \\ \frac{n}{n+1}w_{k_1}^T & \frac{n}{n+1}w_2^T e & \frac{1}{n+1} & \frac{n}{n+1}w_{k_2}^T \widehat{L}^{-1} \\ \frac{1}{n+1}e^T & \frac{n-k}{n+1} & \frac{1}{n+1} & \frac{1}{n+1}e^T \widehat{L}^{-1} \\ 0 & \frac{n}{n+1}\widehat{L}H_{12}^{(2)}e & \frac{1}{n+1}\widehat{L}e & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{12}^{(1)}e & \frac{1}{n+1}e & \frac{n}{n+1}H_{11}^{(12)}\widehat{L}^{-1}e_1 & \frac{n}{n+1}H_{11}^{(12)}\widehat{L}^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ \frac{n}{n+1}w_{k_1}^T & \frac{n}{n+1}w_2^T e & \frac{1}{n+1} & \frac{n}{n+1}w_{k_2}^T \widehat{L}^{-1}e_1 & \frac{n}{n+1}w_{k_2}^T \widehat{L}^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ \frac{1}{n+1}e^T & \frac{n-k}{n+1} & \frac{1}{n+1} & \frac{1}{n+1}e^T \widehat{L}^{-1}e_1 & \frac{1}{n+1}e^T \widehat{L}^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ 0 & \frac{n}{n+1} & \frac{1}{n+1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{12}^{(1)}e & \frac{1}{n+1}e & \frac{n}{n+1}H_{11}^{(12)}e & \frac{n}{n+1}H_{11}^{(12)}\widehat{L}^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ \frac{n}{n+1}w_{k_1}^T & \frac{n}{n+1}w_2^T e & \frac{1}{n+1} & \frac{n}{n+1}w_{k_2}^T e & \frac{n}{n+1}w_{k_2}^T \widehat{L}^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ \frac{1}{n+1}e^T & \frac{n-k}{n+1} & \frac{1}{n+1} & \frac{k_2}{n+1} & \frac{1}{n+1}e^T \widehat{L}^{-1}\begin{bmatrix} e_2 & \cdots & e_{n-k} \end{bmatrix} \\ 0 & \frac{n}{n+1} & \frac{1}{n+1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

due to the fact that $H_{12}^{(2)}e = e$, and $e = \widehat{L}^{-1}e_1$.

Let

$$G_2 = \begin{bmatrix} \frac{n}{n+1}H_{11}^{(11)} & \frac{n}{n+1}H_{12}^{(1)}e & \frac{1}{n+1}e & \frac{n}{n+1}H_{11}^{(12)}e \\ \frac{n}{n+1}w_{k_1}^T & \frac{n}{n+1}w_2^T e & \frac{1}{n+1} & \frac{n}{n+1}w_{k_2}^T e \\ \frac{1}{n+1}e^T & \frac{n-k}{n+1} & \frac{1}{n+1} & \frac{k_2}{n+1} \\ 0 & \frac{n}{n+1} & \frac{1}{n+1} & 0 \end{bmatrix}. \tag{13}$$

Then, we obtain a PageRank vector expression for this level 4 partition by using Theorem 3.

Theorem 5. *Using the above notations and defining G_2 by (13), then G_2 of order $k_1 + 3$ with the same nonzero eigenvalues as the full new Google matrix G is stochastic. Let*

$$\widehat{\sigma}^T = \widehat{\sigma}^T G_2, \ \widehat{\sigma} \geq 0, \ \widehat{\sigma}^T e = 1, \ \text{where } G_2 \in \mathbb{R}^{(k_1+3)\times(k_1+3)}.$$

Partition $\widehat{\sigma}^T = \begin{bmatrix} \widehat{\sigma}_{1:k_1+1}^T & \widehat{\sigma}_{k_1+2} & \widehat{\sigma}_{k_1+3} \end{bmatrix}$, *where* $\widehat{\sigma}_{k_1+2}$ *and* $\widehat{\sigma}_{k_1+3}$ *are scalars. Then, we define the vector*

$$\sigma^T = \begin{bmatrix} \widehat{\sigma}_{1:k_1+1}^T & \widehat{\sigma}^T \begin{bmatrix} \frac{n}{n+1} H_{11}^{(12)} \\ \frac{n}{n+1} w_{k_2}^T \\ \frac{1}{n+1} e^T \\ 0 \end{bmatrix} & \widehat{\sigma}_{k_1+2} \end{bmatrix} \in \mathbb{R}^{k_1+k_2+2},$$

and the vector σ *satisfies* $\sigma^T G_1 = \sigma^T$. *Thus, the PageRank vector* π_{n+1} *for* $G \in \mathbb{R}^{(n+1)\times(n+1)}$ *is given by*

$$\pi_{n+1}^T = \begin{bmatrix} \sigma_{1:k_1+k_2}^T & \sigma^T \begin{bmatrix} \frac{n}{n+1} H_{12}^{(1)} \\ \frac{n}{n+1} H_{12}^{(2)} \\ \frac{1}{n+1} e^T \\ \frac{n}{n+1} w_2^T \end{bmatrix} & \sigma_{k_1+k_2+1} \end{bmatrix}.$$

Proof. The proof is technical modifications of the proof in Section 4.3; therefore, it is omitted here. □

Thus, we claim that the new similarity matrix $\begin{bmatrix} I & 0 \\ 0 & L \end{bmatrix}$, where L satisfies (8), can lead to a dimension reduction of an involved minimal irreducible web link matrix in its 4-level matrix partition form. After obtaining a smaller vector $\widehat{\sigma}^T$ of G_2, the PageRank vector can recover according to Theorem 5. Now, we present the output of the corresponding algorithm in Algorithm 1.

Algorithm 1 Lumping algorithm for a minimal irreducible PageRank model (5)

Input: An initial nonnegative vector $\widehat{\sigma}^T = \begin{bmatrix} \widehat{\sigma}_{1:k_1}^T & \widehat{\sigma}_{k_1+1} & \widehat{\sigma}_{k_1+2} & \widehat{\sigma}_{k_1+3} \end{bmatrix}$, $\widehat{\sigma}^T \geq 0$, $\|\widehat{\sigma}^T\| = 1$, a prescribed tolerance ϵ.

Output: The PageRank vector π_n^T.

Choose an initial vector $\widehat{\sigma}^T = \begin{bmatrix} \widehat{\sigma}_{1:k_1}^T & \widehat{\sigma}_{k_1+1} & \widehat{\sigma}_{k_1+2} & \widehat{\sigma}_{k_1+3} \end{bmatrix} \in \mathbb{R}^{k_1+3}$, $\widehat{\sigma}^T \geq 0$, $\|\widehat{\sigma}^T\| = 1$.

Set $\tau = 1$.

while $\tau \geq \epsilon$ **do**

$\widetilde{\sigma}_{1:k_1}^T = \frac{n}{n+1} \widehat{\sigma}_{1:k_1}^T H_{11}^{(11)} + \frac{n}{n+1} \widehat{\sigma}_{k_1+1} w_{k_1}^T + \frac{1}{n+1} \cdot \frac{1}{n+1} e^T$;

$\widetilde{\sigma}_{k_1+1} = \frac{n}{n+1} \widehat{\sigma}_{1:k_1}^T H_{12}^{(1)} e + \frac{n}{n+1} w_2^T e \widehat{\sigma}_{k_1+1} + \frac{n-k}{n+1} \cdot \frac{1}{n+1} + \frac{n}{n+1} \widehat{\sigma}_{k_1+3}$;

% $\widehat{\sigma}_{k_1+2} = \frac{1}{n+1}$ (by direct computation);

$\widetilde{\sigma}_{k_1+3} = 1 - \widehat{\sigma}_{1:k_1}^T e - \widehat{\sigma}_{k_1+1} - \frac{1}{n+1}$;

$\widehat{\sigma} = \widetilde{\sigma}$;

$\tau = \|\widetilde{\sigma} - \widehat{\sigma}\|_1$;

end while

Compute $x = \widehat{\sigma}^T \begin{bmatrix} \frac{n}{n+1} H_{11}^{(12)} & \frac{n}{n+1} w_{k_2}^T & \frac{1}{n+1} e^T & 0 \end{bmatrix}^T$;

Recover σ^T by $\sigma^T = \begin{bmatrix} \widehat{\sigma}_{1:k_1+1}^T & x & \widehat{\sigma}_{k_1+2} \end{bmatrix}$;

Compute $y = \sigma^T \begin{bmatrix} \frac{n}{n+1} H_{12}^{(1)} & \frac{n}{n+1} H_{12}^{(2)} & \frac{1}{n+1} e^T & \frac{n}{n+1} w_2^T \end{bmatrix}^T$;

Obtain the PageRank vector $\pi_{n+1} = \begin{bmatrix} \sigma_{1:k_1+k_2}^T & y & \sigma_{k_1+k_2+1} \end{bmatrix}$.

The power method applied to G_2 involves a sparse matrix vector multiplied by a small matrix of order $k_1 + 3$ as well as several vector operations. The final step in Algorithm 1 recovers minimal irreducible PageRank vector according to Theorem 5. As the dangling nodes are further excluded from each iteration of the power method, Algorithm 1 has the

same convergence rate as that of the power method applied to G but is much faster due to a reduced matrix computation.

5.2. Numerical Experiments

In this subsection, the web link matrix "wiki-Vote" which has 8297 Web pages and 103689 hyperlinks is used in our numerical example. There are 2187 dangling nodes and 905 "weakly nondangling" nodes in this web link matrix. We compare the power method, LDNA [20], and Algorithm 1 for computing (5) as follows:

1. The original method [17], i.e., the power method applied to a full matrix G in (12).
2. LDNA algorithm in [20]. Lumping all dangling nodes into a single node, the power method is applied to the $(k+2) \times (k+2)$ matrix $\widehat{S}^{(1)}$ in Theorem 2.
3. Algorithm 1 (called as Algorithm 1). Lumping two classes of nodes, the power method applied to the $(k_1+3) \times (k_1+3)$ matrix G_2 in (13).

The parameters v and w are chosen as $v = w = \frac{1}{n}e$ in the above algorithms, where e is a vector of all ones. The overall termination criterion is triggered once the residual norm is below $\epsilon = 10 - 10$. Suppose that three algorithms all need N iterations until they reach the convergence criterion. Tables 1 and 2 give the residues and operation costs of three algorithms, respectively. Here, nnz(H) denotes the number of nonzero entries of matrix H. A sparse matrix-vector product requires $\mathcal{O}$(nnz(H)) operations. From Table 1, we can observe that Algorithm 1 produces the least residual norms compared with the other two algorithms. In addition, we also obtain the fact that the above three algorithms have almost the same convergence rate. We can find that the LDNA algorithm saves about $\mathcal{O}((n-k)(N-1))$ operations, and Algorithm 1 saves about $\mathcal{O}((n-k_1)(N-1))$ operations from Table 2. Because Algorithm 1 computes the PageRank vector by lumping the dangling and "weakly nondangling" nodes and applies the power method to the smallest matrix, it needs the least operations.

Table 1. Comparison of residues for three algorithms.

Iteration	Original	LDNA	Algorithm 1
10	$3.0192e-04$	$1.9117e-04$	$1.2928e-04$
20	$1.2539e-06$	$7.0788e-07$	$4.5348e-07$
30	$6.0013e-09$	$3.3144e-09$	$1.9865e-09$
36	$2.4952e-10$	$1.3619e-10$	$7.8191e-11$

Table 2. Comparison of operation costs for three algorithms.

Algorithms	Bounds for Operations
Original	$\mathcal{O}((nnz(H)+n)N)$
LDNA	$\mathcal{O}((nnz(H_{11})+k)N) + O(nnz(H_{12})+k)$
Algorithm 1	$\mathcal{O}((nnz([H_{11}^{(11)}\ H_{12}^{(1)}])+k_1)N) + \mathcal{O}(nnz([H_{11}^{(12)}\ H_{12}])+k_1)$

6. Conclusions

In this paper, we have presented some new proofs and provided a novel efficient algorithm for computing a kind of minimal irreducible PageRank. Firstly, the spectral distribution of the new Google matrix is also validated by a similarity transformation method except for the determinant method. The simple and generalized similarity transformation matrix method is easier for validating Theorems 2 and 3 in the new PageRank model. Secondly, the new Google matrix of an 4×4 block partition is further lumped to reduce the order of the involved computation matrix. Some theoretical results are discussed. Thirdly, as one real application of our results, the power method can operate on a smaller matrix than before. Therefore, the computational cost is reduced. For further work, we can consider how to choose an extrapolation method operating on a reduced matrix to

accelerate the computation of web information retrieval models. The extrapolation method combined with a lumping method is also worth studying.

Author Contributions: Conceptualization, Y.D.; methodology, Y.F. and Y.D.; software, Y.F.; validation, Y.D. and J.Y.; formal analysis, Y.F. and Y.D.; investigation, Y.D.; resources, Y.F.; data curation, Y.F. and Y.D.; writing—original draft preparation, Y.D. and J.Y.; writing—review and editing, Y.F.; supervision, J.Y.; project administration, Y.F. and J.Y.; funding acquisition, Y.F. and J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant Nos. 12001363, 71671125).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the editor and referees who kindly reviewed the earlier version of this manuscript and provided valuable suggestions and comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Page, L.; Brin, S.; Motwani, R.; Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*; Stanford Digital Libraries: 1999. Available online: http://dbpubs.stanford.edu:8090/pub/1999--66 (accessed on 16 July 2022).
2. Feng, Y.H.; You, J.X.; Dong, Y.X. An extrapolation iteration and its lumped type iteration for computing PageRank. *Bulletion Iran. Math. Soc.* **2021**. [CrossRef]
3. Langville, A.N.; Meyer, C.D. *Google's PageRank and Beyond: The Science of Search Engine Rankings*; Princeton University Press: Princeton, NJ, USA, 2006.
4. Brezinski, C.; Redivo-Zaglia, M. The PageRank Vector: Properties, Computation, Approximation, and Acceleration. *SIAM J. Matrix Anal. Appl.* **2006**, *28*, 551–575. [CrossRef]
5. Bai, Z.Z.; Wu, W.T.; Muratova, G.V. The power method and beyond. *Appl. Numer. Math.* **2021**, *164*, 29–42. [CrossRef]
6. Berkhin, P. A survey on PageRank computing. *Internet Math.* **2005**, *2*, 73–120. [CrossRef]
7. Miao, C.Q.; Tan, X.Y. Accelerating the Arnoldi method via Chebyshev polynomials for computing PageRank. *J. Comput. Appl. Math.* **2020**, *377*, 112891. [CrossRef]
8. Tian, Z.L.; Zhang, Y.; Wang, J.X.; Gu, C.Q. Several relaxed iteration methods for computing PageRank. *J. Comput. Appl. Math.* **2021**, *388*, 113295. [CrossRef]
9. Ipsen, I.C.F.; Selee, T.M. PageRank computation with special attention to dangling nodes. *SIAM J. Matrix Anal. Appl.* **2007**, *29*, 1281–1296. [CrossRef]
10. Lin, Y.Q.; Shi, X.H.; Wei, Y.M. On computing PageRank via lumping the Google matrix. *J. Comput. Appl. Math.* **2009**, *224*, 702–708. [CrossRef]
11. Tian, Z.L.; Liu, Z.Y.; Dong, Y.H. The coupled iteration algorithms for computing PageRank. *Numer. Algorithms* **2021**, *89*, 1603–1637. [CrossRef]
12. Gu, X.M.; Lei, S.L.; Zhang, K.; Shen, Z.L.; Wen, C.; Carpentieri, B. A Hessenberg-type algorithm for computing PageRank Problems. *Numer. Algorithms* **2022**, *89*, 1845–1863. [CrossRef]
13. Jin, Y.; Wen, C.; Shen, Z.L.; Gu, X.M. A simpler GMRES algorithm accelerated by Chebyshev polynomials for computing PageRank. *J. Comput. Appl. Math.* **2022**, *413*, 114395. [CrossRef]
14. Hu, Q.Y.; Wen, C.; Huang, T.Z.; Shen, Z.L.; Gu, X.M. A variant of the Power-Arnoldi algorithm for computing PageRank. *J. Comput. Appl. Math.* **2021**, *381*, 113034. [CrossRef]
15. Yu, Q.; Miao, Z.K.; Wu, G.; Wei, Y.M. Lumping Algorithms for Computing Google's PageRank and Its Derivative, with Attention to Unreferenced Nodes. *Inf. Retr.* **2012**, *15*, 503–526. [CrossRef]
16. Dong, Y.X.; Gu, C.Q.; Chen, Z.B. An Arnoldi-Inout method accelerated with a two-stage matrix splitting iteration for computing PageRank. *Calcolo* **2017**, *54*, 857–879. [CrossRef]
17. Gleich, D.F. PageRank beyond the web. *SIAM Rev.* **2015**, *57*, 321–363. [CrossRef]
18. Dong, Y.X.; You, J.X.; Feng, Y.H. Remarks on lumping PageRank results of Ipsen and Selee. *arXiv* **2021**, arXiv:2107.11080.
19. Dong, Y.X.; Feng, Y.H.; You, J.X. On computing HITS ExpertRank via lumping the hub matrix. *arXiv* **2021**, arXiv:2112.13173.
20. Li, L.L.; Chen, X.; Song, Y.Z. The PageRank model of minimal irreducible adjustment and its lumping method. *J. Appl. Math. Comput.* **2013**, *42*, 297–308. [CrossRef]
21. Lee, C.P.C.; Golub, G.H.; Zenios, S.A. *A Fast Two-Stage Algorithm for Computing PageRank and Its Extensions*; Technical Report; Stanford University: Palo Alto, CA, USA, 2003.

22. Avrachenkov, K.; Litvak, N.; Pham, K.S. A singular perturbation approach for choosing the PageRank damping factor. *Internet Math.* **2008**, *5*, 47–69. [CrossRef]
23. Engström, C.; Silvestrov, S. Using graph partitioning to calculate PageRank in a changing network. *Data Anal. Appl. 2 Util. Results Eur. Other Top.* **2019**, *3*, 179–191.
24. Wu, G. Eigenvalues of Certain Augmented Complex Stochastic Matrices with Application in PageRank. Recent Advances in Scientific Computing and Matrix Analysis. 2011. pp. 85–92. Available online: https://www.researchgate.net/profile/Gang-Wu-18/publication/266862168_Eigenvalues_of_certain_augmented_complex_stochastic_matrices_with_applications_to_PageRank/links/571587b708ae1a840264fe1a/Eigenvalues-of-certain-augmented-complex-stochastic-matrices-with-applications-to-PageRank.pdf (accessed on 16 July 2022).
25. Wu, G. On the convergence of the minimally irreducible Markov chain method with applications to PageRank. *Calcolo* **2017**, *54*, 267–279. [CrossRef]
26. Brauer, A. Limits for the characteristic roots of a matrix. IV: Applications to stochastic matrices. *Duke Math. J.* **1952**, *19*, 75–91. [CrossRef]
27. Ding, J.; Zhou, A.H. Eigenvalues of rank-one updated matrices with some applications. *Appl. Math. Lett.* **2007**, *20*, 1223–1226. [CrossRef]