THE $25,000,000,000* EIGENVECTOR(本征向量) THE LINEAR(线的) ALGEBRA(线性代数) BEHIND GOOGLE

KURT BRYAN† AND TANYA LEISE‡

Abstract. Google's success derives in large part(大型模制品;大规格模制品) from its PageRank algorithm(算法), which ranks the importance of webpages according(相符合) to an eigenvector(本征向量) of a weighted link matrix. Analysis of the PageRank formula provides a wonderful applied topic for a linear(线的) algebra(线性代数) course. Instructors(教员) may assign(分配) this article as a project to more advanced students, or spend one or two(一、二个) lectures presenting(send的过去式和过去分词) the material with assigned homework from the exercises. This material also complements(补充) the discussion of Markov chains(马尔可夫链) in matrix algebra(矩阵代数)(矩阵代数). Maple and Mathematica files suppor this material can be found at www.rose-hulman.edu/∼bryan.

Key words(关键字). linear(线的) algebra(线性代数), PageRank, eigenvector(本征向量), stochastic matrix(随机矩阵)

AMS(存取方法服务) subject classifications(主题分类). 15-01, 15A18, 15A51

1. Introduction. When Google went online in the late 1990's, one thing that set it apart from other search engines was that its search res(远程输入服务)ult listings(列表) always seemed deliver the "good stuff" up front(在最前面)(在最前面). With other search engines you often had to wade through screen(费力地完成) after screen of links to irrelevant(不恰当的) web pages that just happened to match the search text. Part of the magic behind Google is its PageRank algorithm(算法), which quantitatively(量) rates the importance of each page on the web, allowing Google to rank the pages and thereby(因此) present to(出现在...) the user the more(越发) important (and typically most relevant and helpful) pages first.

Understanding how to calculate PageRank is essential(要素) for anyone designing a web page that they want people to access frequently(频繁), since getting listed first in a Google search leads to(导致) many people looking at(看) your page. Indeed, due to Google's(由于...) prominence(突起) as a search engine, its ranking system has had a deep influence on the development and structure(结构) of the internet, and on what kinds of information and services get accessed most frequently(频繁) . Our goal in this paper is to explain one of the core(核心) ideas behind how Google calculates web page rankings(等级). This turns out to be a delightful(令人愉快的) application of standard linear algebra.

Search engines such as(例如...) Google have to do three basic things:

1. Crawl(爬行) the web and locate(定居) all web pages with public access.
2. Index(索引) the data from step 1, so that(所以) it can be searched efficiently(生效) for relevant keywords(关键字) or phrases.
3. Rate the importance of each page in the database, so that(所以) when a user does a search and the subset(子集) of pages in the database with the desired information has been(be的过去分词) found, the more(越发) important pages can be presented first.

This paper will focus on step 3. In an interconnected(使互相连接) web of pages, how can one meaningfully define(定义) and quantify(定量) the "importance" of any(在所有的...当中) given(赠予的) page?

The rated importance of web pages is not the only factor(因素) in how links are presented, but it is a significant(重要的) one. There are also successful ranking algorithms(算法) other than(不同于) PageRank. The interested(感兴趣的 reader(读者) will find a wealth of information about ranking algorithms and search engines, and we list just a few(几个) references for getting started (see the extensive(广的) bibliography(参考书目) in [9], for example(例如), for a more complete list). F or a brief overview(概述) of how Google handles the entire process see [6], and for an in-depth(深入的) treatment of PageRank see [3] and a companion article [9]. Another article with good(粘性物) concrete examples is [5]. F or more background on PageRank and explanations of essential(要素) principles of web design to maximize(取...最大值) a website's PageRank, go to(转到) the websites [4, 11, 14]. T o find out more about search engine principles in general(大多数) and other ranking algorithms, see [2] and [8]. Finally , for an account of some newer approaches to searching the web, see [12] and [13].

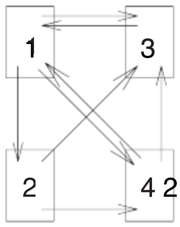
2. Developing(发展中的) a formula(公式) to rank pages.

---

*THE APPROXIMATE(接近于) MARKET V ALUE OF GOOGLE WHEN THE COMPANY WENT PUBLIC(摊牌) IN 2004.

†Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, IN 47803; email: kurt.bryan@rose-hulman.edu; phone: 812) 877-8485; fax: (812)877-8883.

‡Mathematics and Computer Science Department, Amherst College, Amherst, MA 01002; email: tleise@amherst.edu; phone: (413)542-5411; fax: (413)542-2550.



FIG(无花果). 2.1. *An example of a web with only four pages. An arrow from page A to page B indicates a link from page A to page B.*

2.1. The basic idea. In what follows we will use the phrase “importance score” or just “score” for any quantitative(数量的) rating of a web page’s importance. The importance score for any web page will always be a non-negative real number(实数). A core idea in assigning a score to any given web page is that the page’s score is derived from(得自) the links made to that page from other web pages. The links *to* a given(赠予的) page are called the *backlinks* for that page. The web thus becomes a democracy(民主政治) where pages vote for(赞成) the importance of other pages by linking to them.

Suppose the web of interest contains n pages, each page indexed by an integer(整数) k, $1 \leq k \leq n$. A typical example is illustrated(举例说明) in Figure 2.1, in which an arrow from page A to page B indicates a link from page A to page B. Such a web is an example of a *directed graph.* [1] W e’ll use x_k to denote(指示) the importance score(核心) of page k in the web. The x_k is non-negative and $x_j > x_k$ indicates that page j is more important than page k (so $x_j = 0$ indicates page j has the least possible importance score).

A very simple approach is to take x_k as the number of backlinks for page k. In the example in Figure 2.1, we have $x_1 = 2$, $x_2 = 1$, $x_3 = 3$, and $x_4 = 2$, so that(所以) page 3 would be the most important, pages 1 and 4 tie for second, and page 2 is least important. A link to page k becomes a vote for(赞成) page k’s importance.

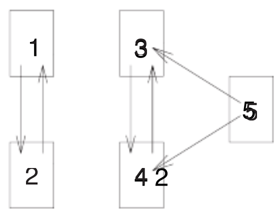
This approach ignores an important feature one would expect a ranking algorithm to have, namely(即) , that a link to page k from an important page should boost(推进) page k’s importance score more than a link from an unimportant(不重要的) page. F or example, a link to your homepage directly(径直地) from Y ahoo ought to(应该)(应该) boost(推进) your page’s score(核心) much more than(比...多) a link from, say , `www.kurtbryan.com` (no relat to the author). In the web of Figure 2.1, pages 1 and 4 both have two backlinks: each links to the other, but page 1’s second backlink is from the seemingly(看来似乎) important page 3, while page 4’s second backlink is from the relatively(相对地) unimportant(不重要的) page 1. As such(照比), perhaps we should rate page 1’s importance higher than that of page 4.

As a first attempt at incorporating(合并的) this idea let’s compute the score of page j as the sum of the scores of all pages(在所有的...中偏偏...) linking to page j. F or example, consider the web of Figure 2.1. The score of page 1 would be determined by the relation $x_1 = x_3 + x_4$. Since x_3 and x_4 will depend on(依靠) x_1 this scheme(方案) seems strangely(奇妙地) self-referential, but it is the approach we will use, with one more modification(修正). Just as in elections(选举), we don’t want a single individual(人) to gain(获利) influence merely by casting(投掷) multiple(倍数) votes. In the same vein(血管), we seek a scheme(方案) in which a web page doesn’t gain(获利) extra influence simply(简单地) by linking to lots of other pages. If page j contains n_j links, one of which links to page k, then we will boost(推进) page k’s score by x_j/n_j rather than by x_j. In this scheme(方案) each web page gets a total of one vote, *weighted by that web page’s score*, that is evenly(平衡地) divided up among all of(实足) its outgoing(外出) links. T o quantify(定量) this for a web of n pages, let $L_k \subset \{1, 2, \ldots, n\}$ denote(指示) the set of pages with a link to page k, that is, L_k is the set of page k’s backlinks. F or each k we require

$$x_k = \sum_{j \in L_k} \frac{x_j}{n_j}, \qquad (2.1)$$

where n_j is the number of outgoing(外出) links from page j (which must be positive since if $j \in L_k$ then

[1] A graph consists of(由...组成) a set of *vertices(顶点)* (in this context(上下文), the web pages) and a set of *edges*. Each edge joins a pair of vertices(顶点). The graph is *undirected* if the edges have no direction. The graph is *directed* if each edge (in(挤进) the web context(上下文), the links) has a direction, that is, a starting(开端) and ending vertex(顶点).



Fig(无花果). 2.2. *A web of five pages, consisting of(由…组成) two disconnected(断开) "subwebs" W_1 (pages 1 and 2) and W_2 (pages 3, 4, 5).*

page j links to at least(至少)(至少) page k!). We will assume that a link from a page to itself will not be counted. In this "democracy of the web" you don't get to(到达) vote for(赞成) yourself !

Let's apply this approach to the four-page web of Figure 2.1. For page 1 we have $x_1 = x_3/1 + x_4/2$, since pages 3 and 4 are backlinks for page 1 and page 3 contains only one link, while page 4 contains two links (splitting its vote in(选出) half). Similarly(相像地), $x_2 = x_1/3$, $x_3 = x_1/3 + x_2/2 + x_4/2$, and $x_4 = x_1/3 + x_2/2$. These linear equations(线性方程) can be written(书面的) $\mathbf{Ax} = \mathbf{x}$, where $\mathbf{x} = [x_1\ x_2\ x_3\ x_4]^T$ and

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}. \tag{2.2}$$

This transforms the web ranking problem into the "standard" problem of finding an eigenvector for a square matrix!(方阵) (Recall(回想) that the eigenvalues(特征值) λ and eigenvectors $\mathbf{x}$ of a matrix $\mathbf{A}$ satisfy the equation(等式) $\mathbf{Ax} = \lambda\mathbf{x}$, $\mathbf{x} \neq \mathbf{0}$ by definition.) We thus seek an eigenvector $\mathbf{x}$ with eigenvalue(特征值) 1 for the matrix $\mathbf{A}$. We will refer to(查阅) $\mathbf{A}$ as the "link matrix" for the given(赠予的) web.

It turns out(打扫)(打扫) that the link matrix $\mathbf{A}$ in equation(等式) (2.2) does indeed have eigenvectors with eigenvalue 1, namely(即) , all multiples(倍数) of the vector $[12\ 4\ 9\ 6]^T$ (recall(回想) that any non-zero multiple(倍数) of an eigenvector is again an eigenvector). Let's agree to scale(按比例缩小) these "importance score eigenvectors" so that the components sum to 1. In this case(既然这样) we obtain $x_1 = \frac{12}{31} \approx 0.387$, $x_2 = \frac{4}{31} \approx 0.129$, $x_3 = \frac{9}{31} \approx 0.290$, and $x_4 = \frac{6}{31} \approx 0.194$. Note that this ranking differs from(不同) that generated(产生) by simply(简单地) counting backlinks. It might seem surprising that page 3, linked to by all other pages, is not the most important. To understand this, note that page 3 links only to page 1 and so casts its entire vote for page(投票选某人) 1. This, with the vote of page 2, results in(导致) page 1 getting the highest(最高) importance score.

More generally , the matrix $\mathbf{A}$ for any web must have 1 as an eigenvalue(特征值) if the web in question(正被讨论) has no *dangling(垂悬的) nodes(节点)* (pages with no outgoing(外出) links). To see this, first note that for a general(橙花醛;β-柠檬 of n pages formula(公式) (2.1) gives rise to(引起) a matrix $\mathbf{A}$ with $A_{ij} = 1/n_j$ if page j links to page i, $A_{ij} = 0$ otherwise. The jth column(专栏) of $\mathbf{A}$ then contains n_j non-zero entries, each equal to $1/n_j$, and the column(专栏) thus sums to(共计) 1. This motivates(给与动机) the following definition(定义), used in the study of Markov chains:(

Definition(定义) 2.1. *A square matrix(方阵) is called a* ***column-stochastic matrix*** *if all of(实足) its entries are nonnegative and the entries in each column(专栏) sum to one.*

The matrix $\mathbf{A}$ for a web with no dangling(垂悬的) nodes(节点) is column-stochastic. We now prove

Proposition(建议) 1. *Every column-stochastic matrix has* 1 *as an eigenvalue.* *Proof(证据).* Let $\mathbf{A}$ be an $n \times n$ column-stochastic matrix and let $\mathbf{e}$ denote(指示) an n dimensional(…维的) column vector(列向量)(列向量) with all entries eq to 1. Recall(回想) that $\mathbf{A}$ and its transpose(互换位置) $\mathbf{A}^T$ have the same eigenvalues. Since $\mathbf{A}$ is column-stochastic it is easy to see that $\mathbf{A}^T\mathbf{e} = \mathbf{e}$, so that 1 is an eigenvalue for $\mathbf{A}^T$ and hence(因此) for $\mathbf{A}$. □

In what follows we use $V_1(\mathbf{A})$ to denote the eigenspace for eigenvalue 1 of a column-stochastic matrix $\mathbf{A}$.

2.2. Shortcomings. Several difficulties arise with using formula(公式) (2.1) to rank websites. In this section we discuss two issues: webs with non-unique rankings(等级) and webs with dangling(垂悬的) nodes(节点).

2.2.1. Non-Unique Rankings(等级). For our rankings it is desirable(令人想望的) that the dimension of $V_1(\mathbf{A})$ equal one, so that there is a unique eigenvector $\mathbf{x}$ with $\sum_i x_i = 1$ that we can use for importance scores. This is true in the web of Figure 2.1 and more generally is always true for the special case of a strongly(强有力地) connected web (that is, you can get from any page to any other page in a finite(有限的) number of steps); see Exercise 10 below.

Unfortunately(恐怕) , it is not always true that the link matrix $\mathbf{A}$ will yield(出产) a unique ranking for all(尽管) webs. Consider the web in Figure 2.2, for which the link matrix is

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

W e find here that $V_1(\mathbf{A})$ is two-dimensional;(两维的) one possible pair of basis vectors(基底向量)(基底向量) is $\mathbf{x} = [1/2, 1/2, 0, 0, 0]^T$ and $\mathbf{y} = [0, 0, 1/2, 1/2, 0]^T$. But note that any linear combination(线性组合) of these two vectors yields(出产) another vector in $V_1(\mathbf{A})$, e.g., $\frac{3}{4}\mathbf{x} + \frac{1}{4}\mathbf{y} = [3/8, 3/8, 1/8, 1/8, 0]^T$. It is not clear which, if any(若有的话) , of these eigenvectors we should use for the rankings!

It is no coincidence that for the web of Figure 2.2 we find that dim((暗淡的)$V_1(\mathbf{A})) > 1$. It is a consequence of the fact that if a web W, considered as an undirected graph (ignoring which direction each arrows points), consists of r disconnected(断开) subwebs $W_1, \ldots, W_r$, then dim((使暗淡)$V_1(\mathbf{A})) \geq r$, and hence(因此) there is no unique importance sc(理学士)ore(矿石) vector $\mathbf{x} \in V_1(\mathbf{A})$ with $\sum_i x_i = 1$. This makes intuitive sense: if a web W consists of r disconnected(断开) subwebs $W_1, \ldots, W_r$ then one would expect difficulty in finding a common reference frame(参考帧)(参考帧) for comparing the scores of pages in one(成为一体) subweb with those in another subweb.

Indeed, it is not hard to see why a web W consisting of r disconnected(断开) subwebs forces dim($V_1(\mathbf{A})) \geq r$. Suppose a web W has n pages and r component subwebs $W_1, \ldots, W_r$. Let n_i denote the number of pages in W_i. Index(索引) the pages in W_1 with indices(索引) 1 through n_1, the pages in W_2 with indices $n_1 + 1$ through $n_1 + n_2$, the pages in W_3 with $n_1 + n_2 + 1$ through $n_1 + n_2 + n_3$, etc(及其他). In general(大多数), let $N_i = \sum_{j=1}^{i} n_j$ for $i \geq 1$, with $N_0 = 0$, so W_i contains pages $N_{i-1} + 1$ through N_i. F or example, in the web of Figure 2 we can take $N_1 = 2$ and $N_2 = 5$, so W_1 contains pages 1 and 2, W_2 contains pages 3, 4, and 5. The web in Figure 2.2 is a particular example of the general case, in which the matrix $\mathbf{A}$ assumes a block diagonal(对角线的) structure(结构)

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_r \end{bmatrix},$$

where $\mathbf{A}_i$ denotes the link matrix for W_i. In fact(事实上) W_i can be considered as a web in its own right. Each $n_i \times n_i$ matrix $\mathbf{A}_i$ is column-stochastic, and hence(因此) possesses some eigenvector $\mathbf{v}^i \in \mathbb{R}^{n_i}$ with eigenvector 1. F or each i between 1 and r construct a vector $\mathbf{w}^i \in \mathbb{R}^n$ which has 0 components for all elements(元件) corresponding to(相当于...) blocks other than(不同于) block i. F or example,

$$\mathbf{w}^1 = \begin{pmatrix} \mathbf{v}^1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{w}^2 = \begin{pmatrix} 0 \\ \mathbf{v}^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots$$

Then it is easy to see that the vectors $\mathbf{w}^i$, $1 \leq i \leq r$, are linearly(线) independent eigenvectors for $\mathbf{A}$

with eigenvalue 1 because

$$\mathbf{A}\mathbf{w}^{i} = \mathbf{A}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \mathbf{v}^{i} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{w}^{i}.$$

Thus $V_1(\mathbf{A})$ has dimension at least(至少) r.

2.2.2. Dangling(悬挂的) Nodes. Another difficulty may arise when using the matrix **A** to generate(产生) rankings. A web with dangling nodes produces a matrix **A** which contains one or more columns of all zeros. In this case(既然这样) **A** is *column-substochastic*, that is, the column sums of **A** are all less than or equal to one. Such a matrix must have all eigenvalues less than or equal to 1 in magnitude(巨大), but 1 need not actually be an eigenvalue for **A**. Nevertheless, the pages in a web with dangling nodes can still be ranked use a similar technique. The corresponding substochastic matrix must have a positive eigenvalue $\lambda \leq 1$ and a corresponding eigenvector $\mathbf{x}$ with non-negative entries (called(入口调用) the *Perron(露天台阶) eigenvector*) that can be used to rank the web pages. See Exercise 4 below. W e will not further consider the problem of dangling nodes here, however.

Exercise 1. *Suppose the people who own page 3 in the web of Figure 1 are infuriated(狂怒的) by the fact that its importance score, computed(计算) using formula (2.1), is lower(低的) than the score of page 1. In an attempt to boost page 3's score, they create a page 5 that links to page 3; page 3 also links to page 5. Does this boost page 3's score above that of page 1?*

Exercise 2. *Construct a web consisting of(由...组成) three or more subwebs and verify(证明$_1$(**A**))that* dim((暗淡的)V *equals (or exceeds)(超过) the number of the components in the web.*

Exercise 3. *Add a link from page 5 to page 1 in the web of Figure 2. The resulting web, considered as an undirected graph, is connected. What is the dimension of $V_1(\mathbf{A})$?*

Exercise 4. *In the web of Figure 2.1, remove the link from page 3 to page 1. In the resulting(结果的) web page 3 is now a dangling node. Set up the corresponding(相当的) substochastic matrix and find its largest positive (Perron) eigenvalue. Find a non-negative Perron(露天台阶) eigenvector for this eigenvalue, and scale(剥落) the vector so that components sum to one. Does the resulting(结果的) ranking seem reasonable?*

Exercise 5. *Prove that in any web the importance score of a page with no backlinks is zero.*

Exercise 6. *Implicit(暗示的) in our analysis up to(从事于(坏事))(从事于(坏事)) this point is the assertion(断言) that the man the pages of a web W are indexed has no effect on the importance score assigned to any given page. Prove this, as follows: Let W contains n pages, each page assigned an index 1 through n, and let **A** be the resulting link matrix. Suppose we then transpose(互换位置) the indices of pages i and j (so page i is now page j and vice-versa). Let $\tilde{\mathbf{A}}$ be the link matrix for the relabelled web.*

- *Argue that $\tilde{\mathbf{A}} = \mathbf{P}\,\mathbf{A}\mathbf{P}$, where $\mathbf{P}$ is the elementary matrix(初等矩阵) obtained by transposing(调换) rows i and j of the $n \times n$ identity matrix(单位矩阵)(单位矩阵). Note that the operation $\mathbf{A} \to \mathbf{P}\,\mathbf{A}$ has the effect of swapping rows i and j of $\mathbf{A}$, while $\mathbf{A} \to \mathbf{A}\mathbf{P}$ swaps columns i and j. Also, $\mathbf{P}^2 = \mathbf{I}$, the identity matrix.*
- *Suppose that $\mathbf{x}$ is an eigenvector for $\mathbf{A}$, so $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ for some λ. Show that $\mathbf{y} = \mathbf{P}\mathbf{x}$ is an eigenvector for $\tilde{\mathbf{A}}$ with eigenvalue λ.*
- *Explain why this shows that transposing the indices of any(在所有的...当中) two pages leaves the importance scores unchanged, and use this result to argue that any permutation(交换) of the page indices(索引) leaves the importance scores unchanged.*

3. A remedy(药物) for dim(V_1(**A**))(暗淡的)An enormous amount of computing(计算) resources are needed to determine an eigenvector for the link matrix corresponding to(相当于...) a web containing billions of pages. It is thus important to know that our algorithm will yield(出产) a unique set of sensible(有感觉的) web rankings. The analysis above shows that our first attempt to rank web pages leads to difficulties(通向...) if the web isn't connected. And the worldwide web, treated as an undirected graph, contains many disjoint((使)脱节) components; see [9] for some interesting statistics concerning the structure(结构) of the web.

Below we present and analyze a modification(修正) of the above method((某人)不懂/不能解决) that is guaranteed to overcome this shortcoming. The analysis that follows is basically(基本上) a special case of the Perron-F robenius theorem(定理), and we only prove what we need for this application. F or a full statement and proof(证据) of the Perron-F robenius theorem(定理), see chapter 8 in [10].

3.1. A modification to the link matrix A. F or an n page web with **no dangling nodes** we can generate(产生) unambiguous importance scores as follows, including cases of web with multiple subwebs.

Let $\mathbf{S}$ denote(指示) an $n \times n$ matrix with all entries $1/n$. The matrix $\mathbf{S}$ is column-stochastic, and it is easy to check that $V_1(\mathbf{S})$ is one-dimensional. W e will replace the matrix $\mathbf{A}$ with the matrix

$$\mathbf{M} = (1 - m)\mathbf{A} + m\mathbf{S}, \tag{3.1}$$

where $0 \leq m \leq 1$. M is a weighted average(加权平均值) of $\mathbf{A}$ and $\mathbf{S}$. The value of m originally(本来) used by Google is reportedly(根据传说) 0.15 [9, 11]. F or any $m \in [0, 1]$ the matrix $\mathbf{M}$ is column-stochastic and we show below that $V_1(\mathbf{M})$ is always one-dimensional if $m \in (0, 1]$. Thus $\mathbf{M}$ can be used to compute(计算) unambiguous importance scores. In the case when $m = 0$ we have the original problem(初始问题), for then $\mathbf{M} = \mathbf{A}$. At the other extreme is $m = 1$, yielding $\mathbf{M} = \mathbf{S}$. This is the ultimately(最后) egalitarian(平等主义的) case: the only normalized eigenvector $\mathbf{x}$ with eigenvalue 1 has $x_i = 1/n$ for all(尽管) i and all(连同其他一切) web pages are rated equally(相等地) important.

Using $\mathbf{M}$ in place of(代替) $\mathbf{A}$ gives a web page with no backlinks (a dangling node) the importance score of m/n (Exercise 9), and the matrix $\mathbf{M}$ is substochastic for any $m < 1$ since the matrix $\mathbf{A}$ is substochastic. Therefore the modified(修正) formula yields nonzero importance scores for dangling links (if $m > 0$) but does not resolve(使分解) the issue of dangling nodes. In the remainder(削价出售(图书)) of this article, we only consider webs with no dangling nodes.

The equation(等式) $\mathbf{x} = \mathbf{M}\mathbf{x}$ can also be cast as

$$\mathbf{x} = (1 - m)\mathbf{A}\mathbf{x} + m\mathbf{s}, \tag{3.2}$$

where $\mathbf{s}$ is a column vector(列向量) with all entries $1/n$. Note that $\mathbf{S}\mathbf{x} = \mathbf{s}$ if $\sum_i x_i = 1$.

W e will prove below that $V_1(\mathbf{M})$ is always one-dimensional, but first let's look at(看) a couple of(一对...) examples.

Example 1: F or the web of four pages in Figure 2.1 with matrix $\mathbf{A}$ given by (2.2), the new formula gives (with $m = 0.15$)

$$\mathbf{M} = \begin{bmatrix} 0.0375 & 0.0375 & 0.8875 & 0.4625 \\ 0.32083\bar{3} & 0.0375 & 0.0375 & 0.0375 \\ 0.32083\bar{3} & 0.4625 & 0.0375 & 0.4625 \\ 0.32083\bar{3} & 0.4625 & 0.0375 & 0.0375 \end{bmatrix},$$

and yields importance scores $x_1 \approx 0.368$, $x_2 \approx 0.142$, $x_3 \approx 0.288$, and $x_4 \approx 0.202$. This yields the same ranking of pages as the earlier computation, but the scores are slightly different.

Example 2 shows more explicitly the advantages of using(有用) $\mathbf{M}$ in place of(代替) $\mathbf{A}$.

Example 2: As a second example, for the web of Figure 2.2 with $m = 0.15$ we obtain the matrix

$$\mathbf{M} = \begin{bmatrix} 0.03 & 0.88 & 0.03 & 0.03 & 0.03 \\ 0.88 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.88 & 0.455 \\ 0.03 & 0.03 & 0.88 & 0.03 & 0.455 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \end{bmatrix}. \tag{3.3}$$

The space $V_1(\mathbf{M})$ is indeed one-dimensional, with normalized(使常态化) eigenvector com(计算机输出缩微胶片;计算机输出胶片缩 0.2, $x_2 = 0.2$, $x_3 = 0.285$, $x_4 = 0.285$, and $x_5 = 0.03$. The modification(修正), using $\mathbf{M}$ instead of(代替) $\mathbf{A}$, allows us to compare pages in different subwebs.

Each entry M_{ij} of $\mathbf{M}$ defined(定义) by equation (3.1) is strictly(严格地) positive, which motivates(给与动机) the following(下列 definition(定义).

Definition 3.1. *A matrix* $\mathbf{M}$ *is* **positive** *if* $M_{ij} > 0$ *for all(尽管)* i *and* j. This is the key property(财产) that guarantees $\dim(V_1(\mathbf{M})) = 1$, which we prove in the next section.

3.2. Analysis of the matrix M. Note that Proposition(建议) 1 shows that $V_1(\mathbf{M})$ is nonempty since $\mathbf{M}$ is stochastic(随机的) . The goal of this section is to show that $V_1(\mathbf{M})$ is in fact(事实上) one-dimensional. This is a consequence of the following(下列各项) two propositions(建议).

Proposition 2. *If* $\mathbf{M}$ *is positive and column-stochastic, then any eigenvector in* $V_1(\mathbf{M})$ *has all positive or all negative(否定) components.* *Proof(证据).* We use proof(证据) by contradiction(矛盾证明). First note that in the standard triangle inequality(不平等) $|\sum_i y_i| \leq \sum_i |y_i|$ (with all y_i real) the inequality(不平等) is strict when the y_i are of mixed sign. Suppose $\mathbf{x} \in V_1(\mathbf{M})$ contains elements(元件) of mixed sign. From(只读存储器) $\mathbf{x} = \mathbf{Mx}$ we have $x_i = \sum_{j=1}^n M_{ij} x_j$ and the summands(被加数) $M_{ij} x_j$ are of mixed sign (since $M_{ij} > 0$). As a result(结果) we have a strict inequality(不平等)

$$|x_i| = \left| \sum_{j=1}^n M_{ij} x_j \right| < \sum_{j=1}^n M_{ij} |x_j|. \tag{3.4}$$

Sum both sides of inequality (3.4) from $i = 1$ to $i = n$, and swap the i and j summations(总和). Then use the fact that $\mathbf{M}$ is column-stochastic ($\sum_i M_{ij} = 1$ for all(尽管) j) to find

$$\sum_{i=1}^n |x_i| < \sum_{i=1}^n \sum_{j=1}^n M_{ij} |x_j| = \sum_{j=1}^n \left(\sum_{i=1}^n M_{ij} \right) |x_j| = \sum_{j=1}^n |x_j|,$$

a contradiction(反驳). Hence $\mathbf{x}$ cannot contain both positive and negative(否定) elements(元件). If $x_i \geq 0$ for all i (and not all x_i are zero) then $x_i > 0$ follows immediately from $x_i = \sum_{j=1}^n M_{ij} x_j$ and $M_{ij} > 0$. Similarly(相像地) $x_i \leq 0$ for all i implies(暗示) that each $x_i < 0$. □

The following proposition will also be useful for analyzing(分析的) $\dim(V_1(\mathbf{M}))$:

Proposition 3. *Let* $\mathbf{v}$ *and* $\mathbf{w}$ *be linearly independent vectors(线性无关向量) in* $\mathbb{R}^m$, $m \geq 2$. *Then, for some values of* s *and* t *that are not both zero, the vector* $\mathbf{x} = s\mathbf{v} + t\mathbf{w}$ *has both positive and negative(否定) components.* *Proof(证据).* Linear independence implies(暗示) neither $\mathbf{v}$ nor $\mathbf{w}$ is zero. Let $d = \sum_i v_i$. If $d = 0$ then $\mathbf{v}$ must contain components of mixed sign, and taking $s = 1$ and $t = 0$ yields the conclusion. If $d \neq 0$ set $s = -\frac{\sum_i w_i}{d}$, $t = 1$, and $\mathbf{x} = s\mathbf{v} + t\mathbf{w}$. Since $\mathbf{v}$ and $\mathbf{w}$ are independent $\mathbf{x} \neq \mathbf{0}$. However, $\sum_i x_i = 0$. We conclude that $\mathbf{x}$ has both positive and negative components. □

We can now prove that using $\mathbf{M}$ in place of(代替) $\mathbf{A}$ yields an unambiguous ranking for any web with no dangling nodes.

Lemma(辅助定理) 3.2. *If* $\mathbf{M}$ *is positive and column-stochastic then* $V_1(\mathbf{M})$ *has dimension* 1. *Proof(证据).* We again use proof by contradiction(矛盾证明). Suppose there are two linearly(线) independent eigenvectors(石山) $\mathbf{v}$ and $\mathbf{w}$ in the subspace(子空间) $V_1(\mathbf{M})$. For any real numbers(实数) s and t that are not both zero, the nonzero vector $\mathbf{x} = s\mathbf{v} + t\mathbf{w}$ must be in $V_1(\mathbf{M})$, and so have components that are all negative or all positive. But by Proposition 3, for some choice of s and t the vector $\mathbf{x}$ must contain components of mixed sign, a contradiction(反驳). We conclude that $V_1(\mathbf{M})$ cannot contain two linearly independent vectors(线性无关向量), and so has dimension one. □

Lemma(辅助定理) 3.2 provides the "punchline" for our analysis of the ranking algorithm using the matrix $\mathbf{M}$ (for $0 < m < 1$). The space $V_1(\mathbf{M})$ is one-dimensional, and moreover(而且), the relevant eigenvectors have entirely(完全) positive or negative components. We are thus guaranteed the existence of a unique eigenvector $\mathbf{x} \in V_1(\mathbf{M})$ with positive components such that $\sum_i x_i = 1$.

Exercise 7. *Prove that if* $\mathbf{A}$ *is an* $n \times n$ *column-stochastic matrix and* $0 \leq m \leq 1$, *then* $\mathbf{M} = (1 - m)\mathbf{A} + m\mathbf{S}$ *is also a column-stochastic matrix.*

Exercise 8. *Show that the product(产品) of two column-stochastic matrices is also column-stochastic.*

Exercise 9. *Show that a page with no backlinks is given importance score* $\frac{m}{n}$ *by formula (3.2).*

Exercise 10. *Suppose that* $\mathbf{A}$ *is the link matrix for a strongly(强有力地) connected web of* n *pages (any page can be reached from any other page by following a finite(有限的) number of links). Show that* $\dim(V_1(\mathbf{A})) = 1$ (暗淡的) *as follows. Let* $(\mathbf{A}^k)_{ij}$ *denote the* (i, j)*-entry of* $\mathbf{A}^k$.

- *Note that page* i *can be reached from page* j *in one(成为一体) step if and only* $A_{ij} > 0$ *(since* $A_{ij} > 0$ *means there's a link from* j *to* i*!) Show that* $(\mathbf{A}^2)_{ij} > 0$ *if and only if(只要…就) page* i *can be reached from page* j *in exactly two steps. Hint:(暗示)* $(\mathbf{A}^2)_{ij} = \sum_k A_{ik} A_{kj}$; *all* A_{ij} *are non-negative, so* $(\mathbf{A}^2)_{ij} > 0$ *implies(暗示) that for some* k *both* A_{ik} *and* A_{kj} *are positive.*

- *Show more generally that* $(\mathbf{A}^p)_{ij} > 0$ *if and only if(只要...就) page i can be reached from page j in EXACTLY p steps.*
- *Argue that* $(\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^p)_{ij} > 0$ *if and only if(只要...就) page i can be reached from page j in p or fewer steps (note $p = 0$ is a legitimate(合法的) choice—any page can be reached from itself in zero steps!)*
- *Explain why* $\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^{n-1}$ *is a positive matrix if the web is strongly(强有力地) connected.*
- *Use the last part (and Exercise 8) so show that* $\mathbf{B} = \frac{1}{n}(\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^{n-1})$ *is positive and column-stochastic (and hence by Lemma(辅助定理) 3.2,* $\dim(\text{(暗淡的)}V_1(\mathbf{B})) = 1$*).*
- *Show that if* $\mathbf{x} \in V_1(\mathbf{A})$ *then* $\mathbf{x} \in V_1(\mathbf{B})$*. Why does this imply that* $\dim(\text{(暗淡的)}V_1(\mathbf{A})) = 1$*?*

Exercise 11. *Consider again the web in Figure 2.1, with the addition of a page 5 that links to page 3, where page 3 also links to page 5. Calculate the new ranking by finding the eigenvector of* $\mathbf{M}$ *(corresponding to(相当于...)* $\lambda = 1$*) that has positive components summing(求和) to one. Use* $m = 0.15$.

Exercise 12. *Add a sixth page that links to every page of the web in the previous exercise, but to which no other page links. Rank the pages using* $\mathbf{A}$*, then using* $\mathbf{M}$ *with* $m = 0.15$*, and compare the results.*

Exercise 13. *Construct a web consisting of two or more subwebs and determine the ranking given by formula (3.1).*

At present(现在) the web contains at least eight billion pages—how does one compute an eigenvector for an eight billion by eight billion matrix? One reasonable approach is an iterative procedure(迭代过程) called(过程调用) the *power method* (along with modifications)(同...一起) that we will now examine for the special case at hand(在手边). It is worth(值得(做)的) noting that(并不是说...) there is much additional(附加的) analysis one can do, and many improved methods for the computation(计算) of PageRank. The reference [7] provides a typical example and additional(附加的) references.

4. Computing(计算) the Importance Score Eigenvector. The rough idea behind the power method[2] for computing an eigenvector of a matrix $\mathbf{M}$ is this: One starts with(从...开始) a "typical" vector $\mathbf{x}_0$, then generates the sequence(序列) $\mathbf{x}_k = \mathbf{M}\mathbf{x}_{k-1}$ (so $\mathbf{x}_k = \mathbf{M}^k\mathbf{x}_0$) and lets k approaches infinity(无限大). The vector $\mathbf{x}_k$ is, to good(粘性物) approximation(接近), an eigenvector for the dominant(占优势的) (largest magnitude)(巨大) eigenvalue of $\mathbf{M}$. However, depending on(依靠) the magnitude(巨大) of this eigenvalue, the vector $\mathbf{x}_k$ may also grow without bound(凝固) or decay(衰退) to the zero vector. One thus typically rescales at each iteration, say by computing $\mathbf{x}_k = \frac{\mathbf{M}\mathbf{x}_{k-1}}{\|\mathbf{M}\mathbf{x}_{k-1}\|}$, where $\|\cdot\|$ can be any vector norm(基准). The method generally requires that the corresponding eigenspace be one-dimensional(...维的), a condition that is satisfied in the case when $\mathbf{M}$ is defined(定义) by equation (3.1).

To use the power method on the matrices $\mathbf{M}$ that arise from the web ranking problem we would generally need to know that any other eigenvalues λ of $\mathbf{M}$ satisfy $|\lambda| < 1$. This assures(保证) that the power method will converge to the eigenvector we want. Actually, the following proposition provides what we need, with no reference to any other eigenvalues of $\mathbf{M}$!

Definition 4.1. *The 1-norm of a vector* $\mathbf{v}$ *is* $\|\mathbf{v}\|_1 = \sum_i |v_i|$.

Proposition 4. *Let* $\mathbf{M}$ *be a positive column-stochastic* $n \times n$ *matrix and let* V *denote the subspace(子空间) of* $\mathbb{R}^n$ *consisting of vectors(由...组成)* $\mathbf{v}$ *such that* $\sum_j v_j = 0$*. Then* $\mathbf{M}\mathbf{v} \in V$ *for any* $\mathbf{v} \in V$*, and*

$$\|\mathbf{M}\mathbf{v}\|_1 \le c\|\mathbf{v}\|_1$$

for any $\mathbf{v} \in V$*, where* $c = \max_{1\le j\le n} |1 - 2\min_{1\le i\le n} M_{ij}| < 1$.

Proof(证据). To see that $\mathbf{M}\mathbf{v} \in V$ is straightforward: Let $\mathbf{w} = \mathbf{M}\mathbf{v}$, so that $w_i = \sum_{j=1}^n M_{ij}v_j$ and

$$\sum_{i=1}^n w_i = \sum_{i=1}^n \sum_{j=1}^n M_{ij}v_j = \sum_{j=1}^n v_j \left(\sum_{i=1}^n M_{ij}\right) = \sum_{j=1}^n v_j = 0.$$

Hence $\mathbf{w} = \mathbf{M}\mathbf{v} \in V$. To prove the bound in the proposition note that

$$\|\mathbf{w}\|_1 = \sum_{i=1}^n e_i w_i = \sum_{i=1}^n e_i \left(\sum_{j=1}^n M_{ij}v_j\right),$$

[2] See [15] for a general introduction to the power method and the use of spectral(光谱的) dec(美国数字电子公司)omposition to find the rate of convergence(集中) of the vectors $\mathbf{x}_k = \mathbf{M}^k\mathbf{x}_0$.

where $e_i = \mathrm{sgn}(w_i)$. Note that the e_i are not all of(实足) one sign, since $\sum_i w_i = 0$ (unless $\mathbf{w} \equiv \mathbf{0}$ in which case the bound clearly holds). Reverse the double sum to obtain

$$\|\mathbf{w}\|_1 = \sum_{j=1}^{n} v_j \left(\sum_{i=1}^{n} e_i M_{ij} \right) = \sum_{j=1}^{n} a_j v_j, \tag{4.1}$$

where $a_j = \sum_{i=1}^{n} e_i M_{ij}$. Since the e_i are of mixed sign and $\sum_i M_{ij} = 1$ with $0 < M_{ij} < 1$, it is easy to see that

$$-1 < -1 + 2 \min_{1 \le i \le n} M_{ij} \le a_j \le 1 - 2 \min_{1 \le i \le n} M_{ij} < 1.$$

We can thus bound(凝固)

$$|a_j| \le |1 - 2 \min_{1 \le i \le n} M_{ij}| < 1.$$

Let $c = \max_{1 \le j \le n} |1 - 2 \min_{1 \le i \le n} M_{ij}|$. Observe that $c < 1$ and $|a_j| \le c$ for all(尽管) j. From(只读存储器) equation (4.1) we have

$$\|\mathbf{w}\|_1 = \sum_{j=1}^{n} a_j v_j = \left| \sum_{j=1}^{n} a_j v_j \right| \le \sum_{j=1}^{n} |a_j||v_j| \le c \sum_{j=1}^{n} |v_j| = c\|\mathbf{v}\|_1,$$

which proves the proposition.

Proposition 4 sets the stage for the following proposition.

Proposition 5. *Every positive column-stochastic matrix* $\mathbf{M}$ *has a unique vector* $\mathbf{q}$ *with positive components such that* $\mathbf{Mq} = \mathbf{q}$ *with* $\|\mathbf{q}\|_1 = 1$. *The vector* $\mathbf{q}$ *can be computed as* $\mathbf{q} = \lim_{k\to\infty} \mathbf{M}^k \mathbf{x}_0$ *for any initial guess* $\mathbf{x}_0$ *with positive components such that* $\|\mathbf{x}_0\|_1 = 1$.

Proof(证据). From(只读存储器) Proposition 1 the matrix $\mathbf{M}$ has 1 as an eigenvalue and by Lemma 3.2 the subspace(子空间) $V_1(\mathbf{M})$ is one-dimensional. Also, all non-zero vectors in $V_1(\mathbf{M})$ have entirely(完全) positive or negative components. It is clear that there is a unique vector $\mathbf{q} \in V_1(\mathbf{M})$ with positive components such that $\sum_i q_i = 1$.

Let $\mathbf{x}_0$ be any vector in $\mathbb{R}^n$ with positive components such that $\|\mathbf{x}_0\|_1 = 1$. We can write $\mathbf{x}_0 = \mathbf{q} + \mathbf{v}$ where $\mathbf{v} \in V$ (V as in Proposition 4). We find that $\mathbf{M}^k\mathbf{x}_0 = \mathbf{M}^k\mathbf{q} + \mathbf{M}^k\mathbf{v} = \mathbf{q} + \mathbf{M}^k\mathbf{v}$. As a result(结果)

$$\mathbf{M}^k\mathbf{x}_0 - \mathbf{q} = \mathbf{M}^k\mathbf{v}. \tag{4.2}$$

A straightforward induction(归纳法) and Proposition 4 shows that $\|\mathbf{M}^k\mathbf{v}\|_1 \le c^k\|\mathbf{v}\|_1$ for $0 \le c < 1$ (c as in Proposition 4) and so $\lim_{k\to\infty} \|\mathbf{M}^k\mathbf{v}\|_1 = 0$. From equation (4.2) we conclude that $\lim_{k\to\infty} \mathbf{M}^k\mathbf{x}_0 = \mathbf{q}$. □

Example: Let $\mathbf{M}$ be the matrix defined by equation (3.3) for the web of Figure 2.2. We take $\mathbf{x}_0 = [0.24, 0.31, 0.08, 0.18, 0.19]^T$ as an initial guess; recall that we had $\mathbf{q} = [0.2, 0.2, 0.285, 0.285, 0.03]^T$. The table below shows the value of $\|\mathbf{M}^k\mathbf{x}_0 - \mathbf{q}\|_1$ for several values of k, as well as the(象...一样的...) ratio(比) $\|\mathbf{M}^k\mathbf{x}_0 - \mathbf{q}\|_1 / \|\mathbf{M}^{k-1}\mathbf{x}_0 - \mathbf{q}\|_1$. Compare this ratio(比) to c from Proposition 4, which in this case(既然这样) is 0.94.

k	$\|\mathbf{M}^k\mathbf{x}_0 - \mathbf{q}\|_1$	$\frac{\|\mathbf{M}^k\mathbf{x}_0 - \mathbf{q}\|_1}{\|\mathbf{M}^{k-1}\mathbf{x}_0 - \mathbf{q}\|_1}$
0	0.62	
1	0.255	0.411
5	0.133	0.85
10	0.0591	0.85
50	8.87×10^{-5}	0.85

It is clear that the bound $\|\mathbf{M}^k\mathbf{x}_0 - \mathbf{q}\|_1 \le c^k\|\mathbf{x}_0 - \mathbf{q}\|_1$ is rather pessimistic (note 0.85 is the value $1 - m$, and 0.85 turns out to be the second largest eigenvalue for $\mathbf{M}$). One can show that in general(大多数) the power method will converge asymptotically(渐近) according(相符合) to $\|\mathbf{M}\mathbf{x}_k - \mathbf{q}\|_1 \approx |\lambda_2|\|\mathbf{x} - \mathbf{q}\|_1$, where λ_2

is the second largest eigenvalue of $\mathbf{M}$. Moreover(而且), for $\mathbf{M}$ of the form $\mathbf{M} = (1-m)\mathbf{A} + m\mathbf{S}$ with $\mathbf{A}$ column-stochastic and all(连同其他一切) S can be shown(show的过去分词) (see, e.g., [1], Theorem(定理) 5.10). As a result(结果), the power method will converge(使集合) much more(更加) rapidly(飞快地) than indicated by c. Nonetheless(然而), the value of c in Proposition 4 provides a very simple bound(凝固) on the convergence(集中) of the power method here. It is easy to see that since all entries of $\mathbf{M}$ are at least m/n, we will always have $c \leq 1 - 2m/n$ in Proposition 4.

As a practical(实际的) matter, note that the $n \times n$ positive matrix $\mathbf{M}$ has no non-zero elements, so the multiplication(乘法) $\mathbf{Mv}$ for $\mathbf{v} \in \mathbb{R}^n$ will typically take $O(n^2)$ multiplications(乘法) and additions, a formidable(巨大的) computation(计算) if $n = 8,000,000,000$. But equation (3.2) shows that if $\mathbf{x}$ is positive with $\|\mathbf{x}\|_1 = 1$ then the multiplication(乘法) $\mathbf{Mx}$ is equivalent to $(1$(等于)$-m)\mathbf{Ax} + m\mathbf{s}$. This is a far more efficient(有效率的) computation(计算), since $\mathbf{A}$ can be expected to contain mostly(大概) zeros (most web pages link to only a few(几个) other pages). W e've now proved our main theorem:

THEOREM 4.2. *The matrix $\mathbf{M}$ defined by (3.1) for a web with no dangling nodes will always be a positive column-stochastic matrix and so have a unique $\mathbf{q}$ with positive components such that $\mathbf{Mq} = \mathbf{q}$ and $\sum_i q_i = 1$. The vector $\mathbf{q}$ may be computed as the limit of iterations(重复) $\mathbf{x}_k = (1-m)\mathbf{Ax}_{k-1} + m\mathbf{s}$, where $\mathbf{x}_0$ is any initial vector with positive components and $\|\mathbf{x}_0\|_1 = 1$.*

The eigenvector $\mathbf{x}$ defined by equation (3.2) also has a probabilistic(概率性的) interpretation(解释). Consider a web-surfer on a web of n pages with no dangling nodes. The surfer begins at some web page (it doesn't matter where) and randomly(偶然) moves from web page to web page according(相符合) to the following(下列各项) procedure: If the surfer is currently(现在) at a page with r outgoing links, he either randomly(偶然) chooses any one of these links with uniform probability(可能性) $\frac{1-m}{r}$ OR he jumps to any randomly(偶然) selected page on the web, each with probability(可能性) $\frac{m}{n}$ (note that $r\frac{1-m}{r} + n\frac{m}{n} = 1$, so this accounts for everything(解释...) he can do). The surfer repeats this page-hopping procedure ad infinitum(无限地)(无限地). The component x_j of the normalized vector $\mathbf{x}$ in equation (3.2) is the fraction(小部分) of time that the surfer spends, in the long run(长远), on page j of the web. More important pages tend to(注意) be linked to by many other pages and so the surfer hits those most often.

EXERCISE 14. *For the web in Exercise 11, compute the values of $\|\mathbf{M}^k\mathbf{x}_0 - \mathbf{q}\|_1$ and $\frac{\|\mathbf{M}^k\mathbf{x}_0-\mathbf{q}\|_1}{\|\mathbf{M}^{k-1}\mathbf{x}_0-\mathbf{q}\|_1}$ for $k = 1, 5, 10, 50$, using an initial guess $\mathbf{x}_0$ not too close to the(太...以致不能...) actual eigenvector $\mathbf{q}$ (so that you can watch the convergence). Determine $c = \max_{1\leq j\leq n} |1 - 2\min_{1\leq i\leq n} M_{ij}|$ and the absolute value(绝对值) of the second largest eigenvalue of $\mathbf{M}$.*

EXERCISE 15. *T o see why the second largest eigenvalue plays a role in(在...起作用) bounding $\frac{\|\mathbf{M}^k\mathbf{x}_0-\mathbf{q}\|_1}{\|\mathbf{M}^{k-1}\mathbf{x}_0-\mathbf{q}\|_1}$, consider an $n \times n$ positive column-stochastic matrix $\mathbf{M}$ that is diagonalizable(对角化的). Let $\mathbf{x}_0$ be any vector with non-negative components that sum to(共计) one. Since $\mathbf{M}$ is diagonalizable(对角化的), we can create a basis of eigenvectors $\{\mathbf{q}, \mathbf{v}_1, \ldots, \mathbf{v}_{n-1}\}$, where $\mathbf{q}$ is the steady state(稳态) vector(状态向量), and then(于是) write $\mathbf{x}_0 = a\mathbf{q} + \sum_{k=1}^{n-1} b_k\mathbf{v}_k$. Determine $\mathbf{M}^k\mathbf{x}_0$, and then(于是) show that $a = 1$ and the sum of the components of each $\mathbf{v}_k$ must equal 0. Next apply Proposition 4 to prove that, except for the(除...外) non-repeated eigenvalue $\lambda = 1$, the other eigenvalues are all strictly(严格地) less than one in absolute value(绝对值). Use this to(用...来(做)...) evaluate $\lim_{k\to\infty} \frac{\|\mathbf{M}^k\mathbf{x}_0-\mathbf{q}\|_1}{\|\mathbf{M}^{k-1}\mathbf{x}_0-\mathbf{q}\|_1}$.*

EXERCISE 16. *Consider the link matrix*

$$\mathbf{A} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \\ 1 & \frac{1}{2} & 0 \end{bmatrix}.$$

Show that $\mathbf{M} = (1-m)\mathbf{A} + m\mathbf{S}$ (all $S_{ij} = 1/3$) is not diagonalizable(对角化的) for $0 \leq m < 1$.

EXERCISE 17. *How should the value of m be chosen(choose的过去分词)? How does this choice affect the rankings and the computation time(计算时间)(计算时间)?*

REFERENCES

[1] A. BERMAN AND R. PLEMMONS, *Nonnegative Matrices in the Mathematical(数学的) Sciences*, Academic Press(学术出版社)(学术出版社), New York, 1979.
[2] M. W. BERRY(浆果) AND M. BROWNE, *Understanding Search Engines: Mathematical(数学的) Modeling(数学模拟) and Text Retrieval(取回), Second Edition*, SIAM(暹罗), Philadelphia(费城), 2005.
[3] M. BIANCHINI, M. GORI, AND F. SCARSELLI, *Inside PageRank*, ACM(计算机协会) Trans. Internet Tech., 5 (2005), pp. 92–128.

[4] S. Brin and L. Page, *The anatomy(剖析) of a large-scale(大规模的) hypertextual web search engine*, http(超文本传输协议) : //www(全球信息网) − db.stanford.edu/ ∼ backrub/google.html (accessed August 1, 2005).
[5] A. F arahat, T. Lofaro, J. C. Miller(厂主), G. Rae, and L.A. W ard(急性呼吸道[疾]病), *Authority Rankings from HITS, PageRank, and SALSA: Existence, Uniqueness,(唯一性) and Effect of Initialization(初始化)*, SIAM(暹罗) J. Sci(串行通信接口). Comput., 27 (2006), pp. 1181-1201.
[6] A. Hill, *Google Inside Out(彻底地)*, Maximum PC, April 2004, pp. 44-48.
[7] S. Kamvar, T. Haveliwala, and G. Golub, *Adaptive methods(自适应方法) for the computation of PageRank*, Linear Algebra Appl., 386 (2004), pp. 51–65.
[8] A. N. Langville and C. D. Meyer, *A survey(调查) of eigenvector methods of web information retrieval(取回)*, SIAM(暹罗) Review, 47 (2005), pp. 135-161.
[9] A. N. Langville and C. D. Meyer, *Deeper inside PageRank*, Internet Math., 1 (2005), pp. 335–380.
[10] C. D. Meyer, *Matrix Analysis(矩阵分析) and Applied(应用的) Linear Algebra*, SIAM, Philadelphia(费城), 2000.
[11] Cleve Moler, *The world's largest matrix computation(矩阵计算)(矩阵计算)*, http(超文本传输协议) : //www(全球信息网).mathworks.com/com(计算机输出 /clevescorner/oct02_cleve.html(超文本标记语言) (accessed August 1, 2005).
[12] Mostafa, J., *Seeking better web searches*, Sci(串行通信接口). Amer., 292 (2005), pp. 66-73.
[13] Sara(萨拉族) Robinson, *The Ongoing(前进) search for(搜查) efficient(有效率的) web search algorithms(搜索算法)*, SIAM News, 37 (Nov 2004).
[14] Ian Rogers(对!), *The Google Pagerank algorithm and how it works*, http(超文本传输协议) : //www(全球信息网).iprcom.com(计算机输出缩微胶片;计算机辅 (accessed August 1, 2005).
[15] W. J. Stew art, *An Introduction to the Numerical(数字的) Solution(解决) of Markov(马尔可夫) Chains*, Princeton(普林斯顿) University Press, Princeton(普林斯顿), 1994.