Perron Frobenius Theorem and Google Page Rank

Parth Chavan

November 22, 2020

1 Introduction

The importance of a Webpage is an intrinsically subjective matter which depends on the surfers interests. But there is a lot that can be said about the importance of webpages according the perspective of the surfer. This paper describes PageRank, a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them. The Pagerank algorithm gives each page its own rating according to its importance from the perspective of the surfer where a page becomes important if important pages link to it. Other way to think of this is to imagine a random surfer on the web following links from page to page . The pagerank of a page is the probability that the surfer lands on that page as if a page is important , the surfer is more likely to land on that page. The behaviour of the surfer is an example of a markov process as the next webpage the surfer visits depends only uopn the webpage the surfer is currrently on. In this paper we introduce the reader to Google's Pagerank algorithm. We assume that the reader is familiar with working knowledge of Linear algebra and Markov chains .

2 Some Algebraic Graph Theory

Associating a matrix to a graph can be a powerful concept because we can make use of all the machinery of linear algebra and matrix computations. In next section we can see that, if matrix associated to a graph has special properties (primitivity, irreducibility), then more can be said about the corresponding graph.

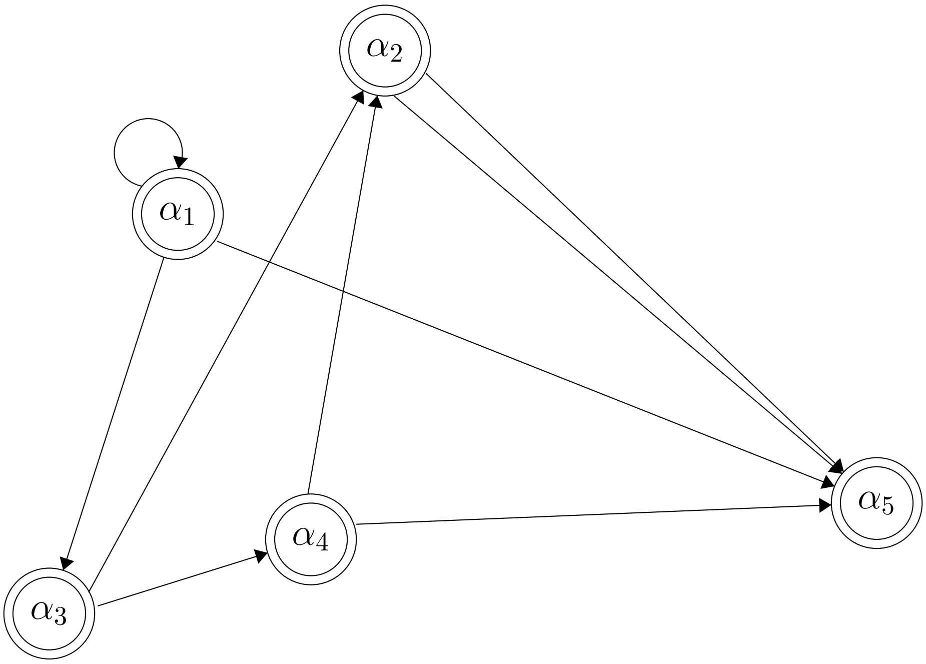

Figure 1

2.1 Terminologies

Definition 1. A directed multigraph G is a pair (V, E) where V is the set of vertices and E is the multiset built from a subset of $V \times V$

Edges of the form (u, v) are represented are represented by arrows going from u to v where, u is called the origin and v is called the destination. A digraph(directed multigraph) is simple if for all $v \in V, (v, v) \notin E$

Definition 2. Let $G = (V, E)$ be a directed multigraph. The adjacency matrix of G is a square matrix $A(G) = (A(G))_{u,v \in V}$ of dimension $\#(V)$ indexed by the vertices of V where, for all u, $v \in V$, $[A(G)]_{u,v}$ is the number of edges from u to v

In a directed multigraph of finite degree, the indegree of the vertex v is the number of incoming edges to v or with destination v. It is denoted by $deg^{-}(v)$ and $\omega^{-}(v)$ denotes the set whose elements are the vertices with v as the destination. The outdegree of the vertex v, denoted by $deg^{+}(v)$, is the number of outgoing edges from v or with origin v and $\omega^{+}(v)$ denotes the set of vertices with v as the origin. The successors (respectively, predecessors) of a vertex v are the vertices w (respectively, u) such that (v, w) (respectively, (u, v)) belongs to $\omega^{+}(v)$ (respectively, $\omega^{-}(v)$). The set of successors (respectively, predecessors) of v is denoted by $\text{succ}(v)$ (respectively, $\text{pred}(v)$). Thus there is a loop on v iff $\text{succ}(v) \cap \text{pred}(v) \neq \emptyset$. We call a vertex source if it's indegree is zero and we call it a sink if it's outdegree is zero.

The Adjacency matrix of the graph in Figure 1 is

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Definition 3. We say that two graphs (E_1, V_1) and (E_2, V_2) are isomorphic if there exists a bijective mapping

$$\psi : V_1 \to V_2$$

such that, $\deg^+ v = \deg^+ \psi(v)$ (respectively $\deg^-$) denoted by $G_1 \cong G_2$
As a consequence, we have $A(G_1) = PA(G_2)P^{-1}$ iff $G_1 \cong G_2$ where $P \in \{0, 1\}^{n \times n}$ is a permutation matrix .
It is important to note that a result stated for directed graphs can also be applied for undirected graphs but the converse is not necessarily true. (In particular we have A(G) is symmetric for undirected graphs). We also have that

$$A(G)e = (deg^+ v_1 \ldots deg^+ v_n)^T \quad e^T A(G) = (deg^- v_1 \cdots deg^- v_n)$$

Where e is the $n \times 1$ column vector with all entries 1 .

Now let's see how adjacency matrix helps us calculate number of walks of length n between any two vertices.

Proposition 1. Let G be a directed multigraph of order k. The number of walks of length n from u to v where, $u, v \in V(G)$ is $[A(G)^n]_{u,v}$.

Proof. Let $A(G) = (a)_{ij}$We have that,

$$[A(G)^n]_{u,v} = \sum a_{uu_1} a_{u_1 u_2} \cdots a_{u_{n-1} v}$$

where, the sum ranges over all ranges of the sequence $(i_1, i_2, \cdots i_{n-1})$ and $1 \leq i_k \leq k$ which is a consequence of matrix multiplication. But a_{ij} is the number of edges from a_i to a_j.Hence summing over all $(i_1, \cdots i_{n-1})$ just gives the total number of walks of length n from v_i to v_j , as desired. ■

3 Perron Frobenius Theorem

3.1 Essential Linear Algebra

In this section we give some necessary tools to prove Perron-Frobenius Theorem.

Definition 4. Let $A \in \mathbb{C}^{n\times n}$. We define the algebraic multiplicity of an eigenvalue λ to be its multiplicity as a root of characteristic polynomial of A and geometric multiplicity to be the dimension of eigenspace with respect to λ. The spectrum of A is the set of all eigenvalues of A and the spectral radius of A is $\max\{|\lambda_1|, \cdots, |\lambda_k|\}$ where $\{\lambda_1, \cdots, \lambda_k\}$ are eigenvalues of A and is denoted by $\rho(A)$.

We let M_n denote the set of $n \times n$ matrices with complex entries , p_A denote the characetristic polynomial of $A \in M_n$ and adj$A$, tr$A$ denote the trace and adjugate of A respectively.

Lemma 1. Let $A \in M_n$, $\lambda \in \mathbb{C}$, and nonzero vectors $x, y \in \mathbb{C}^n$ be given. Suppose that λ has geometric multiplicity 1 as an eigenvalue of A, $Ax = \lambda x$ and $y^*A = \lambda y^*$. There is also a $\gamma \in \mathbb{C}$ such that $adj(\lambda I - A) = \gamma xy^*$

We let A^* denote the conjugate transpose of A.

Proof. We have $rank(\lambda I - A) = n - 1$ and hence rank(adj $(\lambda I - A)) = 1$, i.e. $\text{adj}(\lambda I - A) = \xi\eta^*$ for some $\xi, \eta \in \mathbb{C}^n$. But $(\lambda I - A)(\ \text{adj}(\lambda I - A)) = det(\lambda I - A)I = 0$. So $(\lambda I - A)\xi\eta^* = 0$ and $(\lambda I - A)\xi = 0$ which implies that, $\xi = \alpha x$ for some scalar α. Using the identity $(\lambda I - A)(\ \text{adj}(\lambda I - A))$, in a asimilar fashion we conclude $\eta = \beta y$ for some non-zero scalar β. Thus $\text{adj}(\lambda I - A) = \alpha\beta xy^*$. ∎

Now we give an important result whose proof relies on Rolle's theorem and the fact that $p'_A(\lambda) = \text{tr}(\text{adj}\lambda I - A)$

Theorem 1. Let $A \in M_n$, $\lambda \in \mathbb{C}$ and non zero vectors $x, y \in C^n$. Suppose λ is an eigenvalue of A and x, y^* are right and left eigenvectors respectively. Then,
(a) If λ has algebraic multiplicity 1, then $y^*x \neq 0$.
(b) If λ has geometric multiplicity 1, then it has algebraic multiplicity 1 iff $y^*x \neq 0$.

Proof. In both the cases we have assumed that the geometric multiplicity of λ is 1. Preceeding lemma tells us that $\text{adj}(\lambda I - A) = \gamma xy^*$. Then $p_A(\lambda) = 0$ and $p'_A(\lambda) = \text{tr adj}(\lambda I - A) = \gamma y^*x$. In (a) we assume that the algebraic multiplicity is 1, so $p'_A(\lambda) \neq 0$ and hence $y^*x \neq 0$. In (b) we assume that $y^*x \neq 0$ so $p'_A(\lambda) \neq 0$ and hence λ is not a root of p' giving us that its algebraic multiplicity is 1. ∎

The set of all $m - by - n$ matrices over field F is denoted by $M_{m,n}(F)$.
A function $||| \cdot ||| : M_n \to \mathbb{R}$ is a matrix norm if it satisfies the following conditions

- $|||A||| \geq 0$ and equality occours iff $A = 0$.
- $|||cA||| = |c|\ |||A|||\ \forall c \in \mathbb{C}$

- $|||A+B||| \leq |||A||| + |||B|||$
- $|||AB||| \leq |||A|||\ |||B|||$

We say that a real matrix T is non-negative (or positive) if all the entries of T are non-negative (or positive).We write $T \geq 0$ or $T > 0$ respectively. The Frobenius norm of a matrix A is $\sqrt{\text{tr}(AA^*)}$ and is denoted by $\|A\|_2$

Proposition 2. Let $A, B \in M_n$ then we have
(a) If $|A| \leq |B|$ then, $\|A\|_2 \leq \|B\|_2$
(b) $\|A\|_2 = \|\ |A|\ \|_2$

Proof. For part (a) , concluding that $tr(BB^*) \geq tr(AA^*)$ is trivial using the fact $a\overline{a} = |a|^2\ \forall\ a \in \mathbb{C}$ from which the inequality follows. Similar logic can be used for part B ∎

Now we give some results using these inequalities.

Proposition 3. (Gelfand's formula) Let $|||\cdot|||$ be a matrix norm on M_n. Then $\rho(A) = \lim_{k\to\infty} |||A^k|||^{1/k}$

Proof for this proposition can be found in matrix analysis.

Theorem 2. Let $A, B \in M_n$ and suppose that B is nonnegative. If $|A| \leq B$, then $\rho(A) \leq \rho(|A|) \leq \rho(B)$.

Proof. As $|A| \leq B$ we have $|A|^m \leq B^m$ and hence, $|A^m| \leq |A|^m \leq B^m$. Invoking the above inequalities we have

$$||A^m||_2 \leq ||\ |A|^m\ ||_2 \leq ||B^m||_2 \text{ and } ||A^m||_2^{1/m} \leq ||\ |A|^m\ ||_2^{1/m} \leq ||B^m||_2^{1/m}$$

forall m $\in \mathbb{N}$. Applying Gelfands formula we have $\rho(A) \leq \rho(|A|) \leq \rho(B)$ ∎

We now state two Lemmas concerning norm and spectral radius.

Lemma 2. Let $A = [a_{ij}] \in M_n$ be nonnegative. Then

$$\rho(A) \leq |||A|||_\infty = max_{1\leq i\leq n} \sum_{j=1}^{n} a_{i,j}$$

and

$$\rho(A) \leq |||A|||_1 = max \leq j \leq n \sum_{i=1}^{n} a_{i,j}.$$

If all the row sums of A are equal, then $\rho(A) = |||A|||_\infty$; if all the column sums of A are equal, then $\rho(A) = |||A|||_1$.

Proof. We know that , $|\lambda| \leq \rho(A) \leq |||A|||$ for any eigenvalue λ of A and any matrix norm $||| \cdot |||$. If all the row sums of A are equal, then $e = [1, \cdots, 1]^T$ is an eigenvector of A with eigenvalue $\lambda = |||A|||_\infty$ and so $|||A|||_\infty = \lambda \leq \rho(A) = |||A|||_\infty$. Similarly the statement for columns can be applies to A^T. ∎

Lemma 3. Let $A = [a_{ij}] \in M_n$ be nonnegative. Then

$$\min_{1\leq i\leq n} \sum_{j=1}^{n} a_{i,j} \leq \rho(A) \leq \max_{1\leq i\leq n} \sum_{j=1}^{n} a_{i,j}$$

and

$$\min_{1\leq j\leq n} \sum_{i=1}^{n} a_{i,j} \leq \rho(A) \leq \max_{1\leq j\leq n} \sum_{i=1}^{n} a_{i,j}$$

Proof. Let $\alpha = \min_{1\leq i\leq n} \sum_{j=1}^{n} a_{i,j}$. If $\alpha = 0$, let $B = 0$. If $\alpha > 0$ let $B = [b_{i,j}]$ by letting $b_{i,j} = \alpha a_{i,j} (\sum_{j=k}^{n} a_{i,k})^{-1}$. Then $A \geq B \geq 0$ and $\sum_{j=1}^{n} b_{i,j} = \alpha$. The preceeding lemmas tell us that $\rho(B) = \alpha$ and $\rho(B) \leq \rho(A)$. The upper bound in the first chain of inequalities is the norm bound in previous lemma. The column sum bounds follows by applying row sum bound to A^T. ∎

Corollary 1. Let $A = [a_{ij}] \in M_n$. If A is nonnegative and either $\sum_{j=1}^{n} a_{i,j} > 0$ for all $i = 1, \cdots, n$ or $\sum_{i=1}^{n} aij > 0$ for all $j = 1, \cdots, n$, then $\rho(A) > 0$. In particular, $\rho(A) > 0$ if $n \geq 2$ and A is irreducible and nonnegative.

Now we state an important theorem for non-negative matrices.

Theorem 3. If$A \in M_n$ is nonnegative, then $\rho(A)$ is an eigenvalue of A and there is a nonnegative nonzero vector x such that $Ax = \rho(A)x$.

Proof. For any $\epsilon > 0$ define $A(\epsilon) = A + \epsilon J_n$. Let $x(\epsilon) = [x(\epsilon_i)]$ be the perron vector of $A(\epsilon)$ so that $\epsilon > 0, \sum_{i=1}^{n} e_i = 1$. Since vectors $\{x(\epsilon) : \epsilon > 0\}$ is contained in the set $\{x : x \in \mathbb{C}^n, ||x||_1 \leq 1\}$ there is a monotone decreasing sequence $\epsilon_1 \geq \epsilon_2, \cdots$ with $\lim_{n\to\infty} \epsilon_n = 0$ such that , $\lim_{n\to\infty} x(\epsilon_n) = x$. Since $x(\epsilon_i >)0, ||x(\epsilon_i)||_1 = 1$ forall$i \in \mathbb{N}$ the limit vector $\lim_{n\to\infty} x(\epsilon_n) = x$ must be nonnegative and non-zero.Theorem 4 ensures us that $\rho(A(\epsilon_k)) \geq \rho(A(\epsilon_{k+1})) \geq \cdots \leq \rho(A)$ forall$k \in \mathbb{N}$, so $\rho = \lim_{k\to\infty} \rho(A(\epsilon_k))$ exists and $\rho \geq \rho(A)$. However $x \neq 0$ and

$$\lim_{k\to\infty} A(\epsilon_k)x(\epsilon_k) = \lim_{k\to\infty} \rho(A(\epsilon_k))x(\epsilon_k) = \lim_{k\to\infty} \rho(A(\epsilon_k)) \lim_{k\to\infty} x(\epsilon_k) = \rho x$$

It follows that ρ is an eigenvalue of A and $\rho = \rho(A)$. ∎

Lemma 4. If $\lambda_1, \cdots, \lambda_n$ are eigenvalues of $A \in M_n$ then $\lambda_1 + 1, \cdots, \lambda_n + 1$ are eigenvalues of $A + I$ and $\rho(I + A) \leq \rho(A) + 1$.If A is nonnegative then, $\rho(I + A) = \rho(A) + 1$.

Proof. As $\lambda_1, \cdots, \lambda_n$ are eigenvalues of A We have that

$$det \begin{pmatrix} a_{11} - \lambda_i & \cdots & a_{1,n} \\ \vdots & \ddots & \cdots \\ a_{n,1} & \cdots & a_{nn} - \lambda_i \end{pmatrix} = 0$$

Thus ,

$$det \begin{pmatrix} a_{11} + 1 - (\lambda_i + 1) & \cdots & a_{1,n} \\ \vdots & \ddots & \cdots \\ a_{n,1} & \cdots & a_{nn} - (\lambda_i + 1) \end{pmatrix} = 0$$

and $\rho(I+A) = \max_{1 \leq i \leq n} |\lambda_i + 1| \leq \max_{1 \leq i \leq n} |\lambda_i| + 1 = \rho(A)+1$.. However, previuos lemma ensures that $\rho(A) + 1$ is an eigenvalue of $A + I$ if $A \geq 0$. Thus in this case , $\rho(I + A) = \rho(A) + 1$. ■

3.2 Primitive Graphs and Perron Frobenius Theorem

Definition 5. Let $t \geq 1$ be an integer. A matrix $A \in R^{t \times t}$ with non-negative entries is primitive if there exists an integer N such that A^N is positive. A directed multigraph is primitive if its adjacency matrix is primitive.

Definition 6. *A* matrix $A \in R^{t \times t}_{\geq} 0$ is irreducible if, for all indices $i, j \in 1, ..., t$, there exists an integer N such that $[A^N]_{i,j}$ is positive. We say that a directed multigraph is irreducible if its adjacency matrix is irreducible. A graph is calld irreducible if its adjacency matrix is irreducible.

Note that a directed multigraph G is strongly connected if and only if $A(G)$ is irreducible. The definition readily applies to the undirected connected multigraphs. We say that a real matrix T is non-negative (or positive) if all the entries of T are non-negative (or positive).We write $T \geq 0$ or $T > 0$ respectively. Note that if a matrix T is irreducible , $I + T$ is primitive. For matrices A and B we write $A > B$ if all entries (component wise) af A are greater than that of B.

Theorem 4. (Perron) Let $A \in M_n$ be positive. Then

- $\rho(A) > 0$
- $\rho(A)$ is algebraically a simple eigenvalue of A.

- There is a unique real positive vector $x = [x_i]$ such that $A(x) = \rho(x)$ and $\sum x_i = 1$.
- There is a unique vector $y = [y_i]$ such that $y^T A = \rho(A)y^T$ and $\langle x, y\rangle = 1$.
- $\lambda < \rho(A)$ for every eigenvalue of A other than $\rho(A)$

Proof. For the first statement we have that $A|x| = z > 0$ where, $\rho(A) = \lambda$ and x is an eigenpair.We have $z = A|x| \geq |Ax| = |\lambda x| = |\lambda||x| = |\rho(A)||x|$. So $y = z - \rho(A)x \geq 0$. If $y = 0$, $\rho(A)|x| = A|x| > 0$ and thus, $\rho(A) > 0$. If $y \neq 0$ then $0 < Ay = Az - \rho(A)A|x| = Az - \rho(A)z$ and this implies $A(z) > \rho(A)z$ which is a contradiction
.Now we prove existence of y. There is an eigenpair λ, x of A with $|\lambda| = \rho(A)$. Invoking preceeding lemma tells us that $\rho(A), x$ is an eigenpair and existence of y follows considering A^T.
Suppose $w, z \in \mathbb{C}^n$ are vectors such that , $Aw = \rho(A)w$ and $Az = \rho(A)z$. Then $w = \alpha z$ for some $\alpha \in \mathbb{C}$. There are real numbers θ_1, θ_2 such that $p = [p_j] = e^{-i\theta_1 z} > 0$ and $q = [q_j] = e^{-i\theta_2} > 0$. Let $\beta = \min_{1\leq i\leq n} q_i p_i^{-1}$. Let $r = q - \beta p$. Notice $r \geq 0$ and at east one enry of r is 0. If $r \neq 0$ then $0 < Ar = Aq - A\beta p = \rho(A)q - \beta\rho(A)p = \rho(A)r$. This is a contradiction. We conclude $r = 0$ and $q = \beta p$ and $w = \beta e^{i(\theta_2-\theta_1)}z$. Thus the geometric multiplicity is one and there exist eigenvector $x = [x_i]$ such that $\sum x_i = 1$ and previous result shows us that this is unique. This is called right Perron vector.
A^T is also positve and all results can also be applied to A^T . An eigenvector of A^T corrosponding th $\rho(A)$ normalized so that $\sum x_i y_i = 1$ is left Perron vector. Now $y^* x = y^T x = 1$. Thus Theorem 3 tells us that the algebraic multiplicity of x is 1. ∎

Theorem 5. (Perron - Frobenius) Let T be an irreducible matrix

- (a) T has a positive (real) eigenvalue λ_{max} such that all other eigenvalues of T satisfy
$$|\lambda| \leq \rho(A)$$
- (b) Furthermore $\rho(A)$ has algebraic and geometric multiplicity one, and has an eigenvector x with $x > 0$.
- (c) any non-negative eigenvector is amultiple of x.
- (d) There is a unique real vector $x = [x_i]$ such that, $Ax = \rho(A)x$ and $\sum x_i = 1$; this vector is positive.

- There is a unique real vector $y = [y_i]$ such that $y^T(A) = \rho(A)y^T$ and $\sum x_i y_i = 1$; this vector is positive.

In particular, the spectral radius of an irreducible matrix is one of its eigenvalues.

Proof. Lemma 2 shows that $\rho(A) > 0$ in conditions even weaker than irreducibility. If algebraic multiplicity of $\rho(A)$ is greater than 1 then the algebraic multiplicity of $\rho(A)+1$ in $A+I$ is also greater than 1 which contradicts Perron's theorem. Previuos Lemma ensures us that , there is a nonzero vector x such that , $Ax = \rho(A)x$. Then, $(A+I)^{n-1}x = (\rho(A)+1)^{n-1}x$. Now $x = (\rho(A)+1)^{1-n}(A+I)^{n-1}x$, is positive. If we impose normalization $e^T x = 1$, part(b) ensures x is unique. Part (d) follows on applying part (c) to A^T. ■

Thus Perron - Frobenius Theorem ensures us that the left and right eigenspaces of irreducible matrices are one-dimensional. Now we give an application of Perron-Frobenius Theorem.

Lemma 5. Let P be the transition of an irreducible Markov chain. Then 1 is an eigenvalue of P, and its eigenspace has dimension 1.

Proof. We first show that 1 is an eigenvalue of P. To do this, we simply produce a (right) eigenvector with eigenvalue 1, which is e, the vector all of whose entries are 1. This is an eigenvector with eigenvalue 1 because all the rows of P sum to 1. Now we show that the 1-eigenspace of P is 1-dimensional, as follows. Suppose v is another eigenvector of P with eigenvalue 1, which isn't a scalar multiple of e. By scaling if necessary, we may assume that all of the entries of v are at least -1, and that one of the entries of v is exactly equal to -1. Since v is not a scalar multiple of e, not all the entries of v are equal to -1. Now let $w = e + v$. By the construction of v, all the entries of w are nonnegative, and they aren't all 0, but some entry of w is positive.

Next, note that since the sum of two eigenvectors of P with eigenvalue 1 is again an eigenvector with eigenvalue 1, so w is an eigenvector of P with eigenvalue 1. This means that $Pw = w$. Multiplying by P, we find that $P^2w = Pw = w$, so w is an eigenvector of P^2 with eigenvalue 1, and more generally for each positive integer k, w is an eigenvector P^k with eigenvalue 1.

Let i be some index such that $w_i = 0$, which we know must exist, and let j be some index such that $w_j \neq 0$ (hence $w_j > 0$), which also exists. Now let k be such that $p_{ij}^{(k)} > 0$. Then we have

$$0 = w_i = \sum_{\ell \in \Omega} p_{i\ell}^{(k)} w_\ell \geq p_{ij}^{(k)} w_j > 0.$$

This is a contradiction, and we can therefore conclude that the 1-eigenspace of P is 1-dimensional. ■

Corollary 2. Let P be a primitive nonnegative stochastic matrix . Then there is a unique stochastic (row) vector π such that $\pi P = \pi$. Moreover, all the entries of π are positive.

Proof. Since e is a right eigenvector for P with eigenvalue 1, the Perron–Frobenius Theorem guarantees that 1 is the Perron–Frobenius eigenvalue for P, so there is a unique left probability eigenvector with eigenvalue 1. ■

4 Google's Page Rank

4.1 Defining The Google Matrix

A positive real number called PageRank is associated with every Webpage.We will use the words PageRank and Score equally. Let $V = \{1, ..., n\}$ be the set of Webpages on the World Wide Web. If π_j denotes the page rank of j we set

$$\pi_j = \sum_{i \in pred(j)} \frac{\pi_i}{deg^+ i}$$

Lets assume that the scores are normailzed so that, $\sum_{j=1}^{n} \pi_j = 1$, and we can assume it as a probability distribution. But the problem here is that the values of π_i are not known. To eradicate this problem we assume that every page has the same rank in the beginning. Now the rule given is used to calculate the page rank at every step. This is an iterative procedure. Let $r_{k+1}(\pi_i)$ denote the page rank of page i after $k+1$ iterations. Then we have

$$r_{k+1}(\pi_i) = \sum_{j \in pred(i)} \frac{r_k(\pi_j)}{deg^+ j}$$

This is also called as Power method. So this rrocess is started at $r_0(\pi_i) = 1/n$ hoping that this process will converge at some stable value.

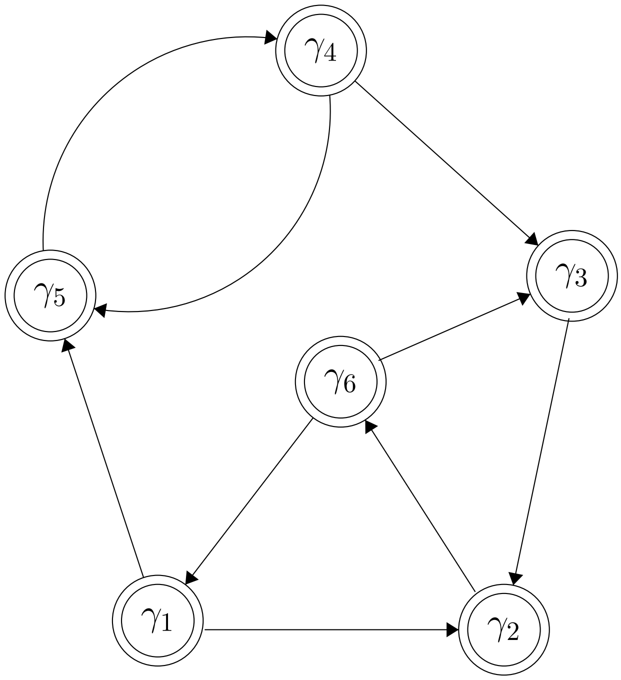

Figure 2

iteration 0	iteration 1	iteration 2
1/6	1/12	1/24
1/6	1/4	3/8
1/6	1/6	1/6
1/6	1/6	1/6
1/6	1/6	1/6
1/6	1/6	1/6

It can also be interpreted as follows : Consider a digraph G with each webpage as a vertex and let P_1, P_2 be any two webpages. We have that, $(P_1, P_2) \in E(G)$ iff there is a link on page P_1 referencing page P_2 amd the Page rank of page P_i is sum of all rankings of all pages pointing towards page P_i.

Definition 7. Let $G = (\{1, ..., n\}, E)$ be a digraph with adjacency matrix $A(G)$. The hyperlink matrix H is defined by

$$H_{i,j} = \begin{cases} A(G)_{i,j}/deg^+ i & if deg^+ i > 0 \\ 0 & if deg^+ i = 0 \end{cases}$$

Compared with H, note that S is stochastic. We have that, $\Pi H = \Pi$ where, $\Pi = (\pi_1, \cdots, \pi_n)$. Thus we are looking for the eigenvector associated with eigenvalue 1.

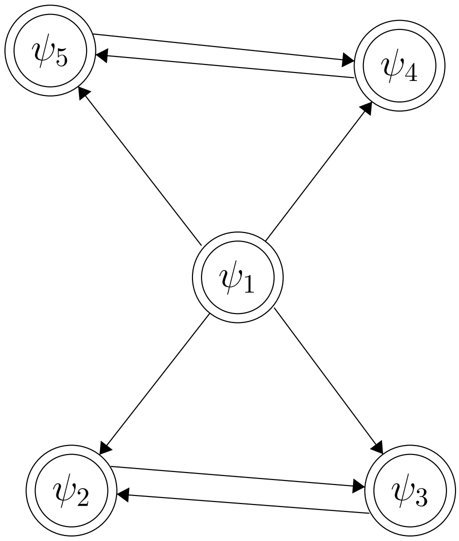

Figure 3

Let $\pi^{(k)T}$ be the pagerank vector at k^{th} iteration. Using this, we can write the the pagerank vector at $(k+1)^{th}$ iteration as

$$\pi^{(k+1)T} = \pi^{(k)T} H.$$

But still there are some problems. For example sinks (dead ends) in the web graph those which accumulate more and more PageRank at each iteration, monopolizing scores and refusing to share. There's also a problem of cycles. For example in Figure, Suppose the iterative process starts with $\pi^{(0)T} = (1\ 0)$. Here, the iterates will not converge no matter how long the process is run. The iterates $\pi^{(k)T}$ alter between (1 0) and (0 1) as k is even or odd. If we modified the hyperlink matrix so that it becomes primitive, then it can be considered as a transition matrix of an ergodic markov chain and using Convergence for ergodic markov chains we know that the page ranks converge to a stable state called stationary distribution.
Consider the following scenario : Imagine a random web surfer who bounces along randomly following the hyperlink structure of the web. In other words he chooses one of the outlinks (A link which carries him onto a different page) uniformly and this process goes on indefinitely. In the long term, the proportion of time spent on a specific page characterizes the relative importance of that page. Unfortunately, the random surfer may get stuck on a sink node (dead end) and never escape. And on the web there are plenty on sinks eg: pdf files , image files , videos, etc. To fix this we do a stochastic adjustment where the 0^T rows are replaced with $1/ne^T$ thereby making H stochastic.

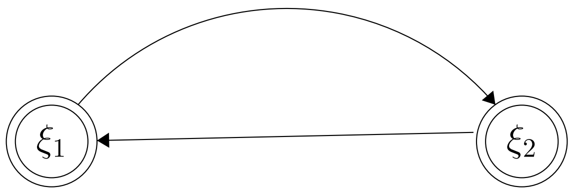

Figure 4

Let us introduce a useful variant of hyperlink matrix,

$$S_{i,j} = \begin{cases} A(G)_{i,j}/\deg^+ i & \text{if} \deg^+ i > 0 \\ 1/n & \text{if} \deg^+ i = 0 \end{cases}$$

It's important to note that the hyperlink matrix here is stochastic because there are no isolated vertices but this isn't generally true. Assume that you are browsing web pages and you reach a dead end i.e. there are no external links on that page. After this an average surfer will either hit the back button or will go to a new search engine. The point here is that after visiting a page which is a dead end, the surfer could virtually be on any page of the Web. Thus, this is reasonable to get rid of sinks proceeding in that way. Having no sinks does not mean that the digraph is strongly connected and hence we cannot apply Perron-Frobenius theorem as it requires a irreducible (Strongly connected directed graph) matrix. There is one last trick to consider

Definition 8. Let$G = (\{1, ..., n\}, E)$ be a digraph with adjacency matrix $A(G)$. The Google matrix G is defined by

$$G = \alpha S + (1-\alpha)1/nee^T$$

where S has been given definition and J is a $n \times n$ matrix whose entries are all equal to 1 and α is a fixed real number in $(0, 1)$.

In this model α is a parameter that controls the time spent on travelling on hyperlinks opposed to getting bored and being present virtually anywhere (teleporting). The teleporting matrix $\mathbf{E} = 1/nee^T$ meaning that the surfer is equally likely when teleporting to be present on any page. As α gets closer to 1 we get rid of sinks (dead ends) and as α gets closer to 0 we start getting a complete graph made up of links. The value of α has to be carefully chosen. Note that any positive $\alpha \neq 1$ does a good job. Also every entry in G is positive and hence G is primitive. Now G is primitive and hence a transition matrix of an ergodic markov chain. We also have that,

$$G = \alpha S + (1-\alpha)1/nee^T = \alpha(H + 1/nae^T) + (1-\alpha)1/nee^T = \alpha H + (\alpha a + (1-\alpha)e)1/ne^T$$

where, a is the vector with $a_1 = 1/n$ if vertex i is a sink and 0 otherwise. In the summary, Google's adjusted Page rank method is

$$\pi^{(k+1)T} = \pi^{(k)T} G$$

The hyperlink matrix H of the graph in Figure 3 is

$$H = \begin{pmatrix} 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The vectors $\Pi^T = (0 \quad 0 \quad 0 \quad 1/2 \quad 1/2)$ and $\Pi'^T = (0 \quad 1/2 \quad 1/2 \quad 0 \quad 0)$ are both pageranks. Thus in even such simple situations we get two different page ranks.

4.2 Computation of the Pagerank vector

The Pagerank vector can be computed in two ways

(a) Solving for the eigenvector problem π^T

$$\pi^T = \pi^T G$$

$$\pi^T e = 1$$

(b) Solve the following linear homogeneous system for π^T

$$\pi^T (I - G) = 0^T$$

$$\pi^T e = 1$$

In the first system we know from Lemma 4 and Corollary 2 that 1 is the Perron-Frobenius Eigenvalue of G so the goal is to find the Perron-Frobenius eigenvector. In the second system, the aim is to find the left hand normalized null vector of $I - G$.
For example consider the webpage graph in figure 5 , taking $\alpha = .9$.

$$G = .9H + (.9(0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0)^T + .1(1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1)^T) + 1/6(1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1)$$

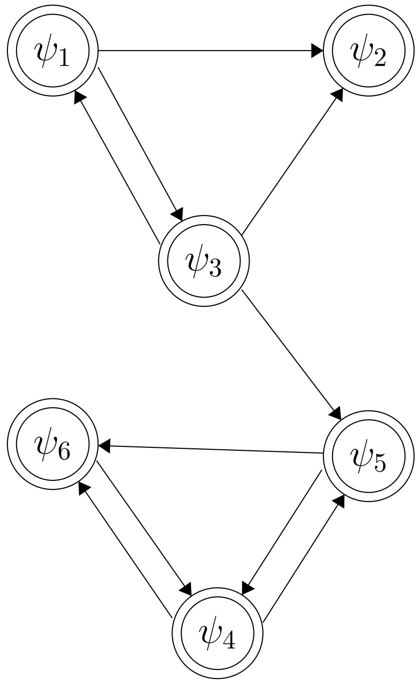

Figure 5

$$= \begin{pmatrix} 1/60 & 7/15 & 7/15 & 1/60 & 1/60 & 1/60 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 19/60 & 19/60 & 1/60 & 1/60 & 1/60 & 1/60 \\ 1/60 & 1/60 & 1/60 & 1/60 & 7/15 & 7/15 \\ 1/60 & 1/60 & 1/60 & 7/15 & 1/60 & 7/15 \\ 1/60 & 1/60 & 1/60 & 11/12 & 1/60 & 1/60 \end{pmatrix}$$

Google's PageRank vector is the stationary vector of G and is given by

$$\pi^T = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \\ (.03721 & .05396 & .04151 & .3751 & .206 & .2862) \end{matrix}$$

Its interpretation is that 3.721% of the time the surfer is on page 1. Thus the page rank here is $(4 \quad 6 \quad 5 \quad 2 \quad 3 \quad 1)$

One of the reasons power method is used is the number of iterations it requires. Brin and Page reported in their 1998 papers, and others have confirmed, that only $50 - 100$ power iterations are needed before the iterates have converged, giving a satisfactory approximation to the exact PageRank vector. Each iteration of the power method requires $O(n)$ effort because H is so sparse. As a result, it's hard to find a method that can beat $50O(n)$ power iterations.The next logical question is: why does the power method applied to G require only about 50 iterations to converge? Is there something about the structure of G that indicates this speedy convergence? The answer comes from the theory of Markov chains. In general, the asymptotic rate of convergence of the power method applied to a matrix depends on the ratio of the two eigenvalues that are largest in magnitude, denoted λ_1 and λ_2. Precisely, the asymptotic convergence rate is the rate at which $|\lambda_2/\lambda_1|^k \to 0$. For stochastic matrices such as G, $\lambda_1 = 1$, so $|\lambda_2|$ governs the convergence. Since G is also primitive, $|\lambda_2| < 1$. Fortunately, for the PageRank problem, it's easy to show that if the respective spectrums are $\sigma(S) = 1, \mu_2, ..., \mu_n$ and $\sigma(G) = 1, \lambda_2, ..., \lambda_n$, then $\lambda_k = \alpha\mu_k$ for $k = 2, 3, \cdots, n$(Theorem 6). Furthermore, the link structure of the Web makes it very likely that $|\mu_2| = 1$ (or at least $|\mu_2| \sim 1$), which means that $|\lambda_2(G)| = \alpha$ (or $|\lambda_2(G)| \sim \alpha$). As a result, the convex combination parameter α explains the reported convergence after just 50 iterations. In their papers, Google founders Brin and Page use $\alpha = .85$, and at last report, this is still the value used by Google. $\alpha^{50} = .85^{50} \sim .000296$, which implies that at the 50^{th} iteration one can expect roughly $2-3$ places of accuracy in the approximate PageRank vector. This degree of accuracy is apparently adequate for Google's ranking needs. Mathematically, ten places of accuracy may be needed to distinguish between elements of the PageRank vector but when PageRank scores are combined with content scores, high accuracy may be less important. Here we conclude with our paper with proof of Theorem 6.

Theorem 6. If the spectrum of S is $\{1, \lambda_2, \cdots, \lambda_n\}$ then the spectrum of the google matrix $G = \alpha S + \big(1-\alpha\big)ev^T$ is $\{1, \alpha\lambda_2, \cdots, \alpha\lambda_n\}$

Since S is stochastic, $(1, e)$ is an eigenpair of S. Let $Q = (eX)$ be a nonsingular matrix that has the eigenvector e as its first column. Let

$$Q^{-1} = \begin{pmatrix} y^T \\ Y^T \end{pmatrix} \implies QQ^{-1} = \begin{pmatrix} y^T e & y^T X \\ Y^T e & Y^T X \end{pmatrix}$$

which gives two useful identities, $y^T e = 1$ and $Y^T e = 0$. As a result, the similarity transformation

$$QSQ^{-1} = \begin{pmatrix} y^T e & y^T SX \\ Y^T e & Y^T SX \end{pmatrix} = \begin{pmatrix} 1 & y^T SX \\ 0 & Y^T SX \end{pmatrix}$$

reveals that $Y^T SX$ contains the remaining eigenvalues of $S, \lambda_2, \cdots, \lambda_n$ Applying the similarity transformation to $G = \alpha S + \big(1-\alpha\big)ev^T$ gives

$$Q^{-1} = \big(\alpha S + (1-\alpha)ev^T\big)Q = \alpha Q^{-1}SQ = (1-\alpha)Q^{-1}ev^TQ$$

$$= \begin{pmatrix} \alpha & \alpha y^T SX \\ 0 & \alpha Y^T SX \end{pmatrix} + (1-\alpha)\begin{pmatrix} y^T e \\ Y^T e \end{pmatrix}\begin{pmatrix} v^T e & v^T X \end{pmatrix} = \begin{pmatrix} \alpha & \alpha y^T SX \\ 0 & \alpha Y^T SX \end{pmatrix}\begin{pmatrix} (1-\alpha) & (1-\alpha)v^T X \\ 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \alpha y^T S X + (1-\alpha) v^T X \\ 0 & \alpha Y^T S X \end{pmatrix}$$

Therefore, the eigenvalues of $G = \alpha S + \big(1-\alpha\big) e v^T$ are $\{1, \alpha\lambda_2, \cdots, \alpha\lambda_n\}$. Thus we conclude with the proof.

References

[HJ12] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[LM11] Amy N Langville and Carl D Meyer. *Google's PageRank and beyond: The science of search engine rankings*. Princeton university press, 2011.