Google Matrix and PageRank: Foundations, Computations and Convergence

Ashwary Sharma, Hardik Prabhu, Kshitish Krit N, Nachiket Dravid

April 2020

Abstract

The PageRank algorithm for ordering web-pages is introduced as a fitting solution to the information retrieval problem. The structural foundation of Markov Chains is introduced, followed by the power method for computating the pagerank values for the Google Matrix. Lastly, the impact of second eigenvalue is explored on the convergence. All the while, major focus is on how linear algebra and its numerical techniques form the core of the stochastic manipulations as well as computational part of the algorithm. Finally, a sample web graph and matrix is created and is shown to resemble that of the actual web and its matrix.

1 Introduction: The Information Question

The advent of computers and sub-sequently the internet has profoundly affected the way humans interact with 'information' - especially the textual and graphic kind, and increasingly the rest. This transformation has posed its own challenges, and PageRank seeks to answer one of them.
The World Wide Web provides access to billions of webpages via servers to any machine running the client - a browser. Accessing desired information through textual queries provides two major challenges: indexing and ordering. Indexing is concerned with associating webpages with relevant keywords to enable searching such large numbers of webpages via text based search queries. Search engines like Google use very interesting programs called 'web-crawlers' to this effect. Our focus, although, will be on the ordering challenge.

1.1 The Importance and Ordering

The next milestone is about being able to 'sort' webpages by 'importance' with respect to each other. It is a seemingly naive question, but is devilishly complex. Relative importance or relevance is a very 'reference dependent' notion. Meaning, it depends on who you ask! So how can one do this, with the information that's available?

Multiple ways were tried, but none were as successful as PageRank. The initial version viewed the web as a directed graph, owing to the existence of in-links/hyperlinks between pages, and assigned a score to each webpage based on its 'centrality'. This score was called its PageRank. This method has its roots in the field of 'citation analysis', where 'centrality' of research papers is an important concern.
Although Google's information retrieval systems have evolved to more sophisticated ways, this principle remains as the core.

2 The PageRank Algorithm

2.1 Intuitive Idea

Let us imagine a random surfer starting at any webpage. The surfer traverses the links from that web-page with equal probability, and repeats the process at each web-page encountered. This gives an idea of the random traversal of the directed graph.
But, one may reach a page with no outbound links. A small modification simplifies things - and also makes the mathematical theory much simpler. At any page, let the surfer not only follow the links from the page randomly...but also randomly 'teleport' to any webpage in the set. This is done with the help of the teleportation constant $\alpha \epsilon (0, 1)$. For example, if a webpage has k outbound links, then it chooses one of them with probability $(1 - \alpha)/k$. It chooses any of the remaining webpages with probability α/M, where M denotes the total number of webpages.
Now, the idea is that the webpages that are visited more and more times as number of iterations increase, are more 'important'. This relative proportion of visits is the desired 'pagerank'.

2.2 The Algorithm

Consider the set of indexed webpages. Let the webpages correspond to the vertices of the directed graph $G = (V, E)$, and let there be a directed edge for every inlink: from the pointing webpage(vertex) to the pointed webpage(vertex). Let A be the matrix representation of this directed graph.
The algorithm takes in the directed graph A, the teleportation coefficient α. It outputs a probability vector p. The corresponding entry for each vertex in p os called its pagerank - and denotes the relative centrality of the webpage.
The algorithm to calculate Google Matrix G is as follows,

1. Divide each in A be number of 1's in the row.
2. Multiply resulting matrix by 1-α.
3. Add α/M to every entry of the resulting matrix to obtain probability/transition/Google matrix G.

Pagerank vector π is calculated by calculating the left-eigenvector of G, that is, π such that

$$\pi G = \pi \tag{1}$$

One of the ways to do it is to keep iteratively multiplying G with itself. Convergence is deductively ensured within the framework of markov chains, to a matrix P such that each row of P is the left-eigenvector π.

3 Statistical Foundations and Matrix Representation

The random surfer with teleportation operation can mathematically be modelled through markov chains. Theorems are merely stated and not proved in this section, for brevity and to preserve the linear algebraic exposition of the document.

Definition 1. *A sequence of random variables $(X_0, X_1...)$ is called a discrete time markov chain with finite state space if it follows the 'markov' property, ie*

$$P(X_{n+1} = j|X_n = i_n, ..., X_0 = i_0) = P(X_{n+1} = j|X_n = i) \tag{2}$$

$\forall n \ \epsilon \ \mathbb{N}$, and X_i are discrete random variables taking values in state space $\{1, ..., M\}$ but are not necessarily identically distributed.

In the theory of markov chains, we are usually interested in calculating marginal distributions f_{X_i} or the long run behaviour; ie, the proportion of states visited as $n \to \infty$. Working with distributions is massively simplified by a neat trick called the 'transition matrix', and things thought to be impossible to compute are suddenly within reach!

Definition 2. *A transition matrix for a markov chain is a matrix P such that,*

$$P_{ij} = P(X_{n+1} = j|X_n = i) \tag{3}$$

It has dimension $M \times M$, for state space $1, ..., M$.

Theorem 1. *(Conditional distribution of X_n):*

$$P(X_n = j|X_0 = i) = [P^n]_{ij} \tag{4}$$

Theorem 2. *(Marginal distribution of X_n):*

$$P(X_n = i) = \sum_{j=1}^{n} [P^n]_{ij} P(X_0 = j) \tag{5}$$

And so we see how beautifully simplified things have become!

Definition 3. *A probability vector(vector with sum of values as 1) is called a stationary distribution for a markov chain if*

$$\pi P = \pi \tag{6}$$

where π *is a column vector of dimensions* $1 \times M$*.*

Stationary distributions don't always exist, but are ensured if the chain or the matrix P follows certain properties.

Definition 4. *A markov chain is called irreducible if for any two states i and j, it is always possible to go from i to j in finite number of steps. In other words,* $\exists n \,\epsilon\, \mathbb{N}$ *such that* $[P^n]_{ij}$ *is positive, where* P *is the transition matrix for the markov chain.*

Theorem 3. *(Existence and uniqueness): A stationary distribution exists and is unique for an irreducible markov chain.*

Definition 5. *Period of a state of markov chain is the gcd of all number of steps where it is 'possible' to return to i, when starting at i. In other words, period of a state i is gcd of set* $n : [P^n]_{ii} > 0$*.*

A chain is called aperiodic if all the states are aperiodic, meaning all states have gcd 1.

Theorem 4. *A chain is irreducible and aperiodic iff* $\exists n \,\epsilon\, \mathbb{N}$ *such that* $[P^n]_{ij}$ *is positive* $\forall i, j$*.*

Theorem 5. *(Convergence to stationary distribution): For an irreducible and periodic markov chain with stationary distribution s and transition matrix* P*,* $P(X_n = i)$ $n \to \infty$ *is* s_i*. For the matrix,* P^n *converges s.t each row converges to* s*.*

And now, the Google Matrix. Let A be the matrix representation of the web graph with M vertices, and let $\alpha \,\epsilon\, (0, 1)$ be the teleportation coefficient as previously discussed, then the Google Matrix G is

$$G = (1 - \alpha)P + \alpha Q/M \tag{7}$$

where Q is an $M \times M$ matrix with all entries 1 and P is the transition matrix obtained by dividing each row of A by number of non-zero entries in that row. Matrix G is indeed a transition matrix, with each row having sum as 1 and all entries are positive values. Hence, it is also a transition matrix for a markov chain that is irreducible and aperiodic. The stationary vector exists, and is a representative of the proportion of visits a random surfer would make to every webpage over a large number of iterations.
Point to be highlighted is how the 'domain' within which the question is solved changes while it all ties back. Specifically, we started with an information retrieval question. Tried a stochastic model and solution. And finally, the final

implementational details are linear algebraic. All of these 'transformations' help solve the question first intended to be solved.
Now, the question is about finding the left-eigenvector of a sparse web-matrix. The power method is suitable for this purpose, and we proceed to why as well as how.

4 Linear Algebra Preliminaries

Definition: A square matrix A is called nilpotent if $A^p = 0$ for some positive integer p.

Lemma: Upper triangular matrix with zero diagonal entries is nilpotent.
Proof:
Let $U \in C^{n\times n}$ be an nxn matrix such that $U_{ii} = 0, \forall\ 1 \leq i \leq n$.
The first column of U is 0.
The j^{th} column of U, for $j \neq 1$,

$$U_{(j)} = Ue_j \in Span\{e_1, e_2, e_3 ... e_{j-1}\}$$

where, e_i is a unit vector with only nonzero i^{th} element $= 1$.
Therefore, the j^{th} column of U^2,

$$U^2_{(j)} = UU_{(j)} = U(Ue_j) = U(\sum_{i=1}^{j-1} a_i e_i)$$

where, for all i, $a_i \in C$ is a constant.

$$U(\sum_{i=1}^{j-1} a_i e_i) = \sum_{i=1}^{j-1} a_i U_{(i)}$$

$$\Rightarrow U^2_{(j)} \in Span\{e_1, e_2 ... e_{j-2}\}$$

Using induction, we can show that $\forall\ 1 \leq k < j$, $U^k_{(j)} \in Span\{e_1, e_2 ... e_{j-k}\}$
Therefore, $U^{j-1}_{(j)} = \alpha e_1$

$$\Rightarrow U^j_{(j)} = U^j e_j = U(U^{j-1} e_j) = U(\alpha e_1) = \alpha U_{(1)} = 0$$

$\Rightarrow U^m_{(j)} = 0$ for $m \geq j$, $\forall\ 1 \leq j \leq n$

$\Rightarrow U^n = 0$

Hence, the proof of the lemma is complete. Moreover, we have shown that for any upper triangular matrix with zero diagonal entries, when raised to its size, is a zero matrix.

Definition: A jordan block of size m corresponding to λ is defined as a $m \times m$ square matrix of the form,

$$J_\lambda = \begin{bmatrix} \lambda & 1 & 0 & ... & 0 \\ 0 & \lambda & 1 & ... & 0 \\ 0 & 0 & \lambda & ... & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & ... & \lambda \end{bmatrix}$$

Theorem: For a jordan block of size m corresponding to eigenvalue λ, $lim_{r\to\infty} J^r$ exists if and only if $|\lambda| < 1$ or $\lambda = 1$, and if $\lambda = 1$ then $m = 1$.
Proof:
Let J be $m \times m$ Jordan block corresponding to eigenvalue λ. Let $N = J - \lambda I_m$. I_m is $m \times m$ identity matrix.

$$\Rightarrow N = \begin{bmatrix} \\ 0 & e_1 & e_2 & ... & e_{m-1} \\ \\ \end{bmatrix}$$

It is clear that N is upper triangular with zero diagonal entries. From the previous lemma, $N^m = 0$. Let $D = \lambda I_m$.

$$\Rightarrow ND = N(\lambda I_m) = \lambda N = (\lambda I_m)N = DN$$

$ND = DN$, N and D commute.
Therefore, for any natural number r,

$$J^r = (N + D)^r$$

Applying binomial theorem,

$$\Rightarrow J^r = \sum_{k=0}^{r} \binom{r}{k} N^k D^{r-k}$$

For further simplification, let us examine the matrix N^k.

Consider the matrix N^2, the j^{th} column,

$$N^2_{(j)} = N(N_{(j)})$$

$$\Rightarrow N^2_{(1)} = N(0) = 0$$

$$\Rightarrow N^2_{(2)} = N(e_1) = N_{(1)} = 0$$

for $j > 2$,

$$N^2_{(j)} = N(e_{j-1}) = N_{(j-1)} = e_{j-2}$$

Therefore, $N^2 = \begin{bmatrix} \\ 0 & 0 & e_1 & e_2 & ... & e_{m-2} \\ \\ \end{bmatrix}$

Using induction we can show that, $\forall\ 1 < k < m,\ k \in \mathbb{N}$

$N^k = \begin{bmatrix} \\ 0 & ... & 0 & e_1 & ... & e_{m-k} \\ \\ \end{bmatrix}$

First k columns are zero vectors. Which can be also as stated as,

$$N_{ij}^k = \begin{cases} 1 & j = i + k \\ 0 & otherwise \end{cases}$$

Continuing from where we left,

$$J^r = \sum_{k=0}^{r} \binom{r}{k} N^k D^{r-k}$$

If r is very large compared to m. ($r >> m$)

$$\Rightarrow J^r = \sum_{k=0}^{m-1} \lambda^{r-k} \binom{r}{k} N^k$$

Since $N^m = 0$ and $D = \lambda I_m$

$$\Rightarrow J_{ij}^r = \sum_{k=0}^{m-1} \lambda^{r-k} \binom{r}{k} N_{ij}^k$$

$$\Rightarrow J_{ij}^r = \sum_{k=1}^{m-1} \lambda^{r-k} \binom{r}{k} N_{ij}^k + (\lambda^r I_m)_{ij}$$

$$\Rightarrow J_{ij}^r = \begin{cases} \lambda^r & j = i \\ \binom{r}{j-i} \lambda^{r-(j-i)} & j > i \\ 0 & j < i \end{cases}$$

Hence,

$$J^r = \begin{bmatrix} \lambda^r & \binom{r}{1}\lambda^{r-1} & \binom{r}{2}\lambda^{r-2} & ... & \binom{r}{m-1}\lambda^{r-m+1} \\ 0 & \lambda^r & \binom{r}{1}\lambda^{r-1} & ... & \binom{r}{m-2}\lambda^{r-m+2} \\ 0 & 0 & \lambda^r & ... & \binom{r}{m-3}\lambda^{r-m+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & ... & \lambda^r \end{bmatrix}$$

Lemma: $S = \{x \in C : |x| < 1 \text{ or } x = 1\}$ be a set of complex number. Then for all x $\lim_{n\to\infty} x^n$ exists if and only if $x \in S$.

Consider $\lim_{r\to\infty} J^r$,

$$lim_{r\to\infty} J^r = \begin{bmatrix} \lim_{r\to\infty}\lambda^r & \lim_{r\to\infty}\binom{r}{1}\lambda^{r-1} & \lim_{r\to\infty}\binom{r}{2}\lambda^{r-2} & ... & \lim_{r\to\infty}\binom{r}{m-1}\lambda^{r-m+1} \\ 0 & \lim_{r\to\infty}\lambda^r & \lim_{r\to\infty}\binom{r}{1}\lambda^{r-1} & ... & \lim_{r\to\infty}\binom{r}{m-2}\lambda^{r-m+2} \\ 0 & 0 & \lim_{r\to\infty}\lambda^r & ... & \lim_{r\to\infty}\binom{r}{m-3}\lambda^{r-m+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & ... & \lim_{r\to\infty}\lambda^r \end{bmatrix}$$

By using the above stated lemma,
$lim_{r\to\infty} J^r$ exists if and only if $\lambda \in S$ and if $\lambda = 1$, then $m = 1$. Hence, we have proved the theorem.

Now we show the necessary and sufficient conditions for existence of limit of powers of any complex square matrix.

Lemma: For any square matrix, if its characteristic polynomial splits then it has a Jordan canonical form which is similar to it.

Let $A \in C^{n\times n}$ be a complex square matrix. From the fundamental theorem of algebra, we know that, any polynomial in complex field splits. Therefore, the characteristics polynomial of A splits. By using the above stated lemma, $\exists$ $V \in C^{n\times n}$ and $\exists$ $J \in C^{n\times n}$ such that,

$$A = V^{-1}JV$$

Where J is the jordan cononical form.

$$J = \begin{bmatrix} J_1 & & & & \\ & J_2 & & & \\ & & J_3 & & \\ & & & \ddots & \\ & & & & J_p \end{bmatrix}$$

Where each jordan block J_i is a jordan block corresponding to an eigenvalue λ_i of size m_i. Therefore,

$$A^r = V^{-1} J^r V$$

and

$$J^r = \begin{bmatrix} J_1^r & & & & \\ & J_2^r & & & \\ & & J_3^r & & \\ & & & \ddots & \\ & & & & J_p^r \end{bmatrix}$$

which implies that $lim_{r \to \infty} A^r$ exists if and only if $lim_{r \to \infty} J_i^r$ exists, $\forall\ 1 \leq i \leq p$. By using previously stated theorem, $lim_{r \to \infty} A^r$ exists if and only if all the eigenvalues $\lambda_i \in S$ and if $\lambda_i = 1$, then $m_i = 1$. (Result 1)

Definition: Let $\lambda_1,\ \lambda_2,\ \lambda_3,...\lambda_p$ be eigenvalues of matrix $A \in C^{n \times n}$. Then its spectral radius $\rho(A)$ is defined as,

$$\rho(A) = max\{|\lambda_1|, |\lambda_2|, |\lambda_3|, ...|\lambda_p|\}$$

Lemma: If $A \in C^{n \times n}$ then $\rho(A) \leq \|A\|$, for all induced matrix norms.
Proof:
Let λ be an eigenvalue of A and x be the eigenvector corresponding to it.

$$Ax = \lambda x$$

$$\Rightarrow |\lambda| \|x\| = \|Ax\| \leq \|A\| \|x\|$$

$$\Rightarrow |\lambda| \leq \|A\|$$

$$\Rightarrow \rho(A) \leq \|A\|$$

Definition: A vector with all non-negative real components which sum up to 1 is defined as probability vector.
Definition: A transition matrix (column stochastic) is defined as a square matrix whose columns are probability vectors.

Lemmas:
M is matrix of size n with non-negative entries. M is a transition if and only if $M^{\top}e$ = e, where $e = [1, 1, 1, ...1]^{\top}$.
Let p be a vector with all components non zero. p is probability vector if and only if $e^{\top}p = 1$.
Proofs:
If M is a transition matrix,

$$M^{\top}e = \begin{bmatrix} M_{(1)}e \\ M_{(2)}e \\ M_{(3)}e \\ \vdots \\ M_{(n)}e \end{bmatrix}$$

Where $M_{(j)}e$ is the sum of all the terms in jth column. Since M is a transition matrix, the sum is 1.

$$\Rightarrow M^{\top}e = e$$

Conversely, if $M^{\top}e = e$, then each column of M is a probability vector. Hence, we proved the first lemma.

If p is a probability vector, then $e^{\top}p$ is the sum of all components of p, which is equal to 1.
Conversely, if $e^{\top}p = 1$, then sum of all components of p=1. Therefore, p is a probability vector.

By using above stated lemmas, we can show that product of two transition matrix is a transition matrix and product of a transition matrix and probability vector is a probability vector.

Theorem: Every transition matrix has an eigenvalue 1 and its spectral radius is 1.
Proof:
Let $A \in C^{n\times n}$ be a transition matrix. By using previously stated lemma $A^{\top}e = e$. Therefore, 1 is eigenvalue of $A^{\top}$. Since A has real entries, $A^{\top} = A^{*}$. We know that A and A^{*} has same set of eigenvalues. Therefore, 1 is an eigenvalue of A.

By using previously stated lemma, $\rho(A) \leq \|A\|_1$.
For a transition matrix, column sum for all columns is 1. Therefore, $\|A\|_1 = 1$.
Hence, $\rho(A) = 1$.

We will look at the theorems associated with Transition matrix with all entries positive. These theorems are the reason why the Google transition matrix is build such way that all entries are positive.

Theorem: $A \in C^{n\times n}$ be a matrix which has each entries positive and let λ be an eigenvalue of A such that $|\lambda| = \|A\|_\infty$. Then $\lambda = \|A\|_\infty$ and the eigenspace corresponding to it, $E_\lambda = Span\{e\}$.
Proof:
Let v be an eigenvector of A corresponding to $|\lambda| = \|A\|_\infty$.

$$v = [v_1, v_2, v_3, ..v_n]^\top \text{ and } Av = \lambda v$$

Let $|v_k| = max\{|v_1|, |v_2|, |v_3|, ...|v_n|\}$.
Observe that $|v_k| \neq 0$, because otherwise v=0.
The kth row of Av is λv_k.

$$\Rightarrow |\lambda v_k| = |\sum_{j=1}^{n} A_{kj} vj|$$

$$\Rightarrow |\lambda v_k| \leq \sum_{j=1}^{n} |A_{kj} v_j| \qquad (1)$$

And by inspection we also have these inequalities.

$$\sum_{j=1}^{n} |A_{kj} vj| \leq \sum_{j=1}^{n} |A_{kj}||v_k| \qquad (2)$$

$$(\sum_{j=1}^{n} |A_{kj}|)|v_k| \leq \|A\|_\infty |v_k| \qquad (3)$$

By putting (1),(2) and (3) together,

$$|\lambda||v_k| \leq \|A\|_\infty |v_k|$$

Since $|\lambda| = \|A\|_\infty$, all the three inequalities are equalities. Therefore, rewriting them as equalities, we have

$$|\lambda v_k| = |\sum_{j=1}^{n} A_{kj} v_j| = \sum_{j=1}^{n} |A_{kj} v_j| \qquad (1)$$

$$\sum_{j=1}^{n} |A_{kj} vj| = \sum_{j=1}^{n} |A_{kj}||v_k| \qquad (2)$$

$$(\sum_{j=1}^{n} |A_{kj}|)|v_k| = \|A\|_\infty |v_k| \qquad (3)$$

Before going further, we will look at the lemma regarding equality in triangle inequality for complex numbers.

Lemma: If $z_1, z_2, z_3, ...z_n$ are non-zero complex numbers such that $\sum_{i=1}^{n} |z_i| = |\sum_{i=1}^{n} z_i|$. Then $\exists\ z \in C$ such that,

$$z_i = t_i z \qquad \forall 1 \leq i \leq n$$

Where t_i's are positive constants.
Proof:
We will prove this lemma for n= 2 and then generalize the result.
Let z, w be non-zero complex numbers such that, $|z + w| = |z| + |w|$. Consider $(|z + w|)^2$

$$|z + w|^2 = (z + w)(\overline{z + w})$$

$$\Rightarrow |z + w|^2 = |z|^2 + 2\mathfrak{Re}(z\overline{w}) + |w|^2$$

$$\Rightarrow |z + w|^2 = |z|^2 + 2\mathfrak{Re}(z\overline{w}) + |w|^2$$

$$\Rightarrow |z + w|^2 = (|z| + |w|)^2 + 2(\mathfrak{Re}(z\overline{w}) - |z||w|)$$

Since $(|z| + |w|)^2 = |z + w|^2$, $\mathfrak{Re}(z\overline{w}) = |z||w|$. Consider the expression $|z - tw|^2$ where t be a positive real number.

$$\Rightarrow |z - tw|^2 = |z|^2 - 2\mathfrak{Re}(z\overline{w})t + |w|^2 t^2$$

$$\Rightarrow |z - tw|^2 = |z|^2 - 2|z||w|t + |w|^2 t^2 = (|z| - t|w|)^2$$

If $t = \dfrac{|z|}{|w|}$, then $|z - tw| = 0$.

$$\Rightarrow z = tw$$

Hence, proving the lemma for n=2.

Continuing from where we left,
Since $|v_k| \neq 0$, from (3) we get

$$\sum_{j=1}^{n} |A_{kj}| = \|A\|_\infty$$

From (2) we get

$$\sum_{j=1}^{n} |A_{kj}||vj| = \sum_{j=1}^{n} |A_{kj}||v_k| \text{ and } |v_j| \leq |v_k|$$

$$\Rightarrow |v_j| = |v_k| \qquad \forall\, 1 \le j \le n$$

From (1),

$$|\sum_{j=1}^{n} A_{kj} v_j| = \sum_{j=1}^{n} |A_{kj} v_j|$$

Since A has all entries positive and $|v_j| \neq 0$, $A_{kj} v_j \neq 0$. Therefore, by using the above stated lemma, $\exists z \in C$ such that $A_{kj} v_j = c_j z$, $c_j > 0$, $\forall\, 1 \le j \le n$. Without loss of generality, we assume $|z| = 1$.

$$\Rightarrow |v_k| = |v_j| = \frac{|C_j z|}{|A_{kj}|} = \frac{C_j}{A_{kj}}$$

$$\Rightarrow v_j = |v_k| z, \qquad \forall\, 1 \le j \le n$$

Let $|v_k| z = b$. Therefore, $v = [b, b, b ... b]^\top = be$. This implies $v \in Span\{e\}$ which means

$$E_\lambda = Span\{e\}$$

Finally, observe that all the entries of Ae are positive because the same is true for the entires of A and e. But $Ae = \lambda e$, and hence $\lambda > 0$. Therefore $\lambda = |\lambda| = \|A\|_\infty$.

Theorem: $A \in C^{n \times n}$ be a matrix which has each entries positive and let λ be an eigenvalue of A such that $|\lambda| = \|A\|_1$. Then $\lambda = \|A\|_1$ and the eigenspace corresponding to it, has dimension 1.
Proof:
Observe that $A^\top$ and A have same set of eigenvalues, and $\|A\|_1 = \|A^\top\|_\infty$. We know that for any square matrix M, $det(M) = det(M^\top)$ and $rank(M) = rank(M^\top)$. Therefore, $det(A - \lambda I_n) = det(A^\top - \lambda I_n)$ and $rank(A - \lambda I_n) = rank(A^\top - \lambda I_n)$. By applying rank-nullity theorem, dimensions for eigenspaces of a common eigenvalue are the same for A and $A^\top$. Therefore, by applying the previous theorem for matrix $A^\top$ we get the theorem for $|\lambda| = \|A\|_1$.

We have already shown that for a transition matrix A, $\|A\|_1 = 1$ and 1 is also an eigenvalue. Combining this fact with the above stated theorem, we get an important result which is written in form of a theorem below.

Theorem: $A \in C^{n \times n}$ be a transition matrix which has each entries positive and then $\rho(A) = 1$ and the eigenspace corresponding to eigenvalue 1, has dimension 1. Moreover, if for eigenvalue λ, $|\lambda| = 1$, then $\lambda = 1$.

If we recall (Result 1), the above stated theorem is not enough to guarantee limit of powers of our transition matrix. The eigenspace of eigenvalue 1 has dimension 1, but the multiplicity may be greater. We have to also ensure that for $\lambda = 1$, the size of jordan blocks is 1.

Theorem: $A \in C^{n\times n}$ be a transition matrix. Then for the Jordan blocks corresponding to eigenvalue 1 is 1.
Proof:
Defination: For $A \in C^{n\times n}$, we define a matrix norm $\|A\| = max\{|A_{ij}| : 1 \leq i, j \leq n\}$
Let J be the canonical form of A.

$$A^r = V^{-1}J^rV$$

and

$$J^r = \begin{bmatrix} J_1^r & & & & \\ & J_2^r & & & \\ & & J_3^r & & \\ & & & \ddots & \\ & & & & J_p^r \end{bmatrix}$$

$$J^r = VA^rV^{-1}$$

lemma: $\|A^m\| \leq 1$
Proof:
A is a transition matrix. Therefore, A^m is also a transition matrix. All entries of A^m are less than equal to 1.

$$\Rightarrow \|A^m\| = max\{|A^m_{ij}| : 1 \leq i, j \leq n\} \leq 1$$

lemma:$A \in C^{n\times n}$,$B \in C^{n\times n}$. Then $\|AB\| \leq n\|A\|\|B\|$
Proof:

$$\|AB\| = max\{|AB_{ij}| : 1 \leq i, j \leq n\}$$

$$\Rightarrow [\|AB\| = max\{|\sum_{k=1}^{n} A_{ik}B_{kj}| : 1 \leq i, j \leq n\}$$

Consider $A_{ik}B_{kj}$,

$$A_{ik}B_{kj} \leq \|A\|\|B\|$$

$$\Rightarrow \sum_{k=1}^{n} A_{ik}B_{kj} \leq n\|A\|\|B\|$$

$$\Rightarrow max\{|\sum_{k=1}^{n} A_{ik}B_{kj}| : 1 \leq i, j \leq n\} \leq n\|A\|\|B\|$$

Therefore, $\|AB\| \leq n\|A\|\|B\|$.

By using the lemmas,

$$J^r = VA^rV^{-1}$$

$$\Rightarrow \|J^r\| \leq n^2\|V\|\|A^r\|\|V^{-1}\|$$

$$\Rightarrow \|J^r\| \leq n^2 \|V\| \|V^{-1}\|$$

$\exists\, c \in \mathbb{R}$ such that $\forall\, r \in \mathbb{N}$,

$$\|J^r\| \leq c$$

Let K be a Jordan block corresponding to eigenvalue 1. If size of K is greater than 1, then as previously derived, $K_{12}^r = r\lambda^{r-1} = r$. r is an entry in J^r.
For $r > c$,

$$r \leq \|J^r\| \leq c$$

Which is a contradiction. Therefore, size of $K = 1$.

Therefore, from (Result 1) we can say that for any transition matrix $A \in C^{n\times n}$ which has all entries positive $\lim_{m\to\infty} A^m$ exists. The eigenspace of 1 has dimension 1 and block size cannot be greater than 1 which implies multiplicity of eigenvalue 1 is 1.

In the random surfer model, a web surfer bounces along randomly following the hyperlink structure of the Web. That is, when the surfer arrives at a page with several outlinks by choosing one at random, hyperlinks to this new page or by entering a new destination in the browsers URL line (teleports) and continues this random decision process indefinitely. In the long run, the proportion of time the random surfer spends on a given page is a measure of the relative importance of that page. Following theorem suggests that given an arbitrary starting probability for the n webpages, we arrive at a unique probability vector which determines the pagerank. Observe that with inital probabilty p and transition matrix A, $A^m p$ denotes the probabilty of our random surfer reaching each of the webpages after m iteration in time.

Theorem: For any probability vector P and $A \in C^{n\times n}$ be a transition matrix with positive entries.
Then

$$\lim_{m\to\infty} A^m p = v$$

Where v is the unique probability eigenvector corresponding to 1.
Proof:
From previous theorem, $\lim_{m\to\infty} A^m$ exists.

$$\lim_{m\to\infty} A^m = L$$

Lemma: L is also a transition matrix.
Proof:

$$e^\top L = e^\top \lim_{m\to\infty} A^m = \lim_{m\to\infty} e^\top A^m = \lim_{m\to\infty} e^\top = e^\top$$

$$\Rightarrow L^{\top}e = e$$

Lemma: Columns of L are eigenvectors of A corresponding to eigenvalue 1.
Proof:

$$AL = A \lim_{m\to\infty} A^m = lim_{m\to\infty} A^{m+1} = L$$

Consider the jth column of AL.

$$AL_{(j)} = L_{(j)}$$

Observe that the columns of L have to be multiples of each other because dimension of eigenspace of 1 is 1. Therefore, each column of L is an unique probability vector.

Let $y = \lim_{m\to\infty} A^m p$

$$\Rightarrow y = Lp$$

Therefore, y is a probability vector. And from the above stated lemmas,

$$Ay = ALP = Lp = y$$

y is also an eigenvector corresponding to eigenvalue 1.
Hence, the proof of the theorem is complete.

For computing the eigenvector, we start with an arbitrary initial vector $x^{(0)}$ and we form the vector sequence $\{x^{(k)}\}_{k=0}^{\infty}$ by defining $x^{(k)} = A^k x^{(0)}$. This method is called the power method. We will elaborate on it as well as the convergence rate of the algorithm in the next section.

5 Power method

In this section we consider the simplest method to compute the eigenvector corresponding to the largest eigenvalue, called vector iteration or power method. Let $A \in C^{n \times n}$. We form the vector sequence $\{x^{(k)}\}_{k=0}^{\infty}$ by defining $x^{(k)} = Ax^{(k-1)}$.
Starting with an arbitrary initial vector $x^{(0)}$. It is clear that $x^{(k)} = A^k x^{(0)}$.

5.1 The algorithm:

1: Choose a starting vector $x^{(0)}$.
2: $k = 0$
3: **repeat**
4: $k := k + 1$
5: $x^{(k)} = Ax^{(k-1)}$
6: **stop** when convergence criteria is satisfied.

Let $A = XJY$ be the Jordan normal form of A with $Y = X^{-1}$. Then,

$$Yx^k = JYx^{(k-1)} \text{ and } Yx^k = J^k Y x^{(0)}$$

If the sequence $\{x^{(k)}\}_{k=0}^{\infty}$ converges to x_* then the sequence $\{y^{(k)}\}_{k=0}^{\infty}$ with $y^{(k)} = Yx^{(k)}$ converges to $y_* = Yx_*$. Therefore, for the convergence analysis, we may assume without loss of generality that A is in Jordan canonical form matrix.

Definition: The angle θ between two nonzero vectors x and y is given by,

$$\theta = \arccos\left(\frac{|x^* y|}{\|x\|\|y\|}\right) = \arcsin\left(\| \left(I - \frac{xx^*}{\|x\|^2}\right) \frac{y}{\|y\|} \|\right)$$

Observe that, with the above stated definition we are only concerned with the acute angle between two vectors. When investigating the convergence behaviour of eigensolvers we usually show that the angle between the approximating and the desired vector tends to zero as the number of iterations increases. In fact it is more convenient to work with the sine of the angle.

5.2 Convergence analysis:

Let us assume that A has Jordan block form,

$$A = \begin{bmatrix} \lambda_1 & & & & \\ & J_2 & & & \\ & & J_3 & & \\ & & & \ddots & \\ & & & & J_p \end{bmatrix} = \begin{bmatrix} \lambda_1 & \\ & \mathbf{J_2} \end{bmatrix}$$

where λ_1 has multiplicity 1 and $|\lambda_1| = \rho(A)$. Let J_2 correspond to jordan block of 2nd largest eigenvalue. Then, the eigenvector of A corresponding to its largest eigenvalue $lambda_1$ is e_1. We will now show that the iterates $x^{(k)}$ converge to e_1. More precisely, we will show that the sin of angle between $x^{(k)}$ and e_1 goes to zero with k tends to ∞. Let

$$x^{(k)} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \\ \vdots \\ x_n^{(k)} \end{bmatrix} = \begin{bmatrix} x_1^{(k)} \\ \mathbf{x_2^{(k)}} \end{bmatrix}$$

The *sine* of angle between $x^{(k)}$ and e_1 is given as,

$$\sin\theta^{(k)} = \|(I - e_1 e_1^*)\frac{x^{(k)}}{\|x^{(k)}\|}\| = \frac{\|\mathbf{x_2^{(k)}}\|}{\|x^{(k)}\|}$$

We know that $x^{(k)} = Ax^{(k-1)}$,

$$\Rightarrow x^{(k)} = \begin{bmatrix} x_1^{(k)} \\ \mathbf{x_2^{(k)}} \end{bmatrix} = \begin{bmatrix} \lambda_1 & \\ & \mathbf{J_2} \end{bmatrix} \begin{bmatrix} x_1^{(k-1)} \\ \mathbf{x_2^{(k-1)}} \end{bmatrix} = \begin{bmatrix} \lambda_1 & \\ & \mathbf{J_2} \end{bmatrix}^k \begin{bmatrix} x_1^{(0)} \\ \mathbf{x_2^{(0)}} \end{bmatrix}$$

Defining

$$y^{(k)} = \frac{1}{\lambda_1^k} x^{(k)}$$

we have

$$y^{(k)} = \begin{bmatrix} 1 & \\ & \frac{1}{\lambda_1}\mathbf{J_2} \end{bmatrix} y^{(k-1)}$$

Let us assume that $y_1^{(0)} = 1$. Then $y_1^{(k)} = 1$ for all k and

$$\mathbf{y_2^{(k)}} = \frac{1}{\lambda_1}\mathbf{J_2}\mathbf{y_2}^{(k-1)}, \ \frac{1}{\lambda_1}\mathbf{J_2} = \begin{bmatrix} \frac{1}{\lambda_1}J_2 & & & & \\ & \frac{1}{\lambda_1}J_3 & & & \\ & & \frac{1}{\lambda_1}J_4 & & \\ & & & \ddots & \\ & & & & \frac{1}{\lambda_1}J_p \end{bmatrix} = \begin{bmatrix} \mu_2 & \\ & J' \end{bmatrix}$$

Where μ_2 is $\frac{\lambda_2}{\lambda_1}$, λ_2 is the 2nd largest eigenvector.

Theorem: Let $\|.\|$ be any matrix norm. Then for any square matrix M, $lim_{k\to\infty}\|M^k\|^{1/k} = \rho(M)$.
This theorem is also known as Gelfand's formula. The proof for the theorem is outside the scope of our disscusion.

Therefore, by applying the Gelfand's formula, for any $\epsilon > 0$ there exist $K_\epsilon \in \mathbb{N}$ such that,

$$\left\| \left(\frac{1}{\lambda_1} \mathbf{J_2} \right)^k \right\|^{1/k} \leq |\mu_2| + \epsilon, \qquad \forall\, k > K_\epsilon$$

we can choose ϵ such that

$$|\mu_2| + \epsilon < 1$$

Then the angle between $y^{(k)}$ and e_1,

$$\sin\theta^{(k)} = \|(I - e_1 e_1^*) \frac{x^{(k)}}{\|y^{(k)}\|}\| = \frac{\|\mathbf{y_2^{(k)}}\|}{\|y^{(k)}\|}$$

$$= \frac{\|\mathbf{y_2}^{(k)}\|}{\sqrt{1 + \|\mathbf{y_2}^{(k)}\|^2}} \leq \|\mathbf{y_2}^{(k)}\| \leq \|\frac{1}{\lambda_1} \mathbf{J_2} \mathbf{y_2}^{(k-1)}\| \leq \|\frac{1}{\lambda_1^k} \mathbf{J_2}^{(k)} \mathbf{y_2}^{(0)}\| \leq (|\mu_2| + \epsilon)^k \|\mathbf{y_2}^{(0)}\|$$

Thus, the angle between $y^{(k)}$ and e_1 goes to zero with a rate $\mu_2 + \epsilon$ for any positive ϵ. Since $x^{(k)}$ is a scalar multiple of $y^{(k)}$, the same holds for the angle between $x^{(k)}$ and e_1. Since we can choose ϵ arbitrarily small, we have proved that the *sine* of angle between $x^{(k)}$ and e_1,

$$\sin\theta^{(k)} \leq c \left| \frac{\lambda_2}{\lambda_1} \right|^k$$

provided $x_1^{(0)} \neq 0$.
Returning to a general matrix $A \in C^{n \times n}$ with Jordan canonical form $A = XJY$. The sequence $y^{(k)} = Yx^{(k)}$ converges to $y_* = \alpha e_1$ with $\alpha \neq 0$. Therefore, $x^{(k)}$ converges to a multiple of Xe_1, which is an eigenvector associated with the largest eigenvalue λ_1.

5.3 Convergence Analysis for PageRank

It is proved in the previous section that for any probability vector P and $A \in C^{n \times n}$ be a transition matrix with positive entries.
Then

$$\lim_{m \to \infty} A^m p = v$$

Where v is the unique probability eigenvector corresponding to 1.
For computing the eigenvector, we start with an arbitrary initial vector $x^{(0)}$ such that $x_1^{(0)} \neq 0$ and apply the power method to get the eigenvector.

It can be seen that the power method converges fast if $\dfrac{\lambda_2}{\lambda_1}$ is small, and slowly if $\dfrac{\lambda_2}{\lambda_1}$ is close to 1.

The convergence rate is geometric in the sense that the angle between $x^{(k)}$ and the eignevector is bounded by $\left|\frac{\lambda_2}{\lambda_1}\right|^k$ at the kth iteration.

We know that for our transition matrix, $\lambda_1 = 1$. Therefore the rate of convergence is given by simply $|\lambda_2|$. This means we have no control over the convergence. We aren't computing λ_2 value because it is of no use determining the ranking of the pages. Thankfully the structure of our transition matrix comes to our aid.

Our tranistion matrix A which is used to rank n web pages is created in the following way. P is an $n \times n$ row-stochastic matrix. E is the $n \times n$ rank-one row-stochastic matrix $E = ev^\top$, where e to be the n-vector whose elements are all $e_i = 1$. A is the $n \times n$ column-stochastic matrix:

$$\mathbf{A} = [\mathbf{cP} + (\mathbf{1} - \mathbf{c})\mathbf{E}]^\top$$

We will look at each of the components and then combine it all together.

P: P is the hyperlink matrix of our collection of webpages. P is row stochastic and P_{ij} denotes the probability that our random surfer, who is on website i, by following any of the outlinks reaches website j. Observe that if there is no hyperlink connecting the two sites $p_{ij} = 0$.

E: E is the rank-one row-stochastic matrix defined as, $E = ev^\top$. v is the probability vector that denotes the probability of jumping to a new destination by entering the name in the browsers URL line (teleport). Observe that all the rows of E are the same. It means, the probability distribution remains the same irrespective of the current webpage of our random surfer.

This two are combined together to form A, the tranistion matrix(column stochastic) with all positive entries. The constant c influences the behaviour of our random surfer. If c=1, then our random surfer never teleport and likewise, if c=0, then our random surfer never clicks on any hyperlinks and relies solely on browser's URL.
Theorem: The second eigenvalue of A is bounded by c.

$$|\lambda_2| \leq c.$$

Proof:
CASE 1: c = 0
If c = 0, then, from equation , $A = E^\top$. Since E is a rank-one matrix, $\lambda_2 = 0$. Thus, Theorem 1 is proved for c=0.

CASE 2: c =1

$|\lambda_2| < 1$ for A.

CASE 3: $0 < c < 1$

Lemma: If A is transition matrix with positive entries. Then e and v_2 are orthagonal. v_2 is a eigenvector corresponding to λ_2.
Proof:
We have already shown that $|\lambda_2| < 1$. e is an eigenvector of $A^\top$ because $A^T e = e$.

$$\Rightarrow e^\top v_2 = (A^\top e)^\top v_2$$

$$\Rightarrow e^\top v_2 = e^\top (A v_2) = \lambda_2 e^\top v_2$$

$$\Rightarrow (1 - \lambda_2) e^\top v_2 = 0$$

$$\Rightarrow e^\top v_2 = 0$$

Lemma: $E^\top v_2 = 0$
Proof:
By definition, $E = ev^\top$, and $E^\top = ve^\top$. Thus, $E^\top v_2 = ve^\top v_2$. From previous Lemma , $e^\top v_2 = 0$. Therefore, $E^\top v_2 = 0$.
Lemma: The second eigenvector v_2 of A must be an eigenvector y_i of $P^\top$, and the corresponding eigenvalue is $\gamma_i = \lambda_2 / c$.
Proof:
From the equation

$$\mathbf{A} = [\mathbf{cP} + (\mathbf{1} - \mathbf{c})\mathbf{E}]^\top$$

We get

$$cP^\top v_2 + (1 - c) E^\top v_2 = \lambda_2 v_2$$

$$\Rightarrow cP^\top v_2 = \lambda_2 v_2$$

We can divide through by c to get:

$$P^\top v_2 = \frac{\lambda_2}{c} v_2$$

If we let $y_i = v_2$ and $\gamma_i = \lambda_2 / c$, we can rewrite equation,

$$P^\top y_i = \gamma_i y_i$$

Therefore, v_2 is also an eigenvector of $P^\top$, and the relationship between the eigenvalues of A and $P^\top$ that correspond to v_2 is given by:

$$\lambda_2 = c\gamma_i$$

Observe that $P^\top$ is column stochastic matrix. Thus,

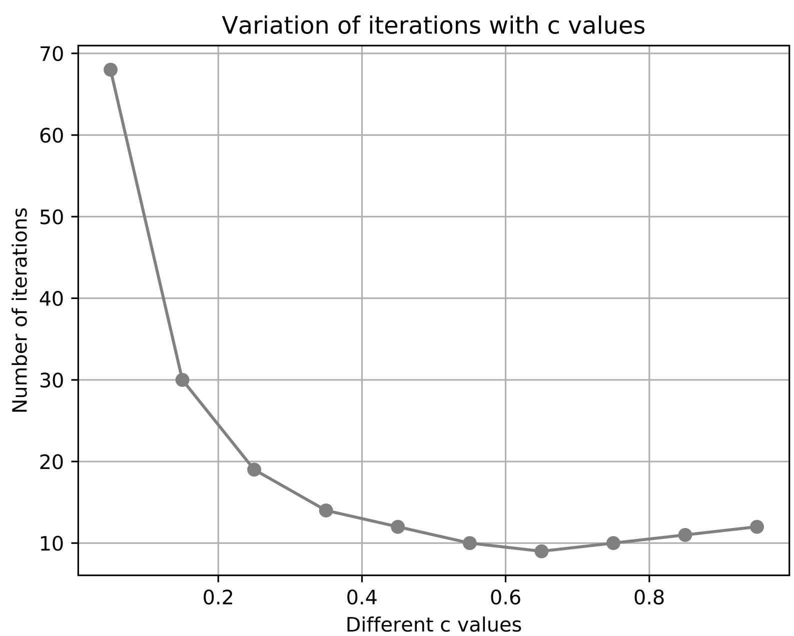
$$|\lambda_2| = |c\lambda_2| \leq c$$

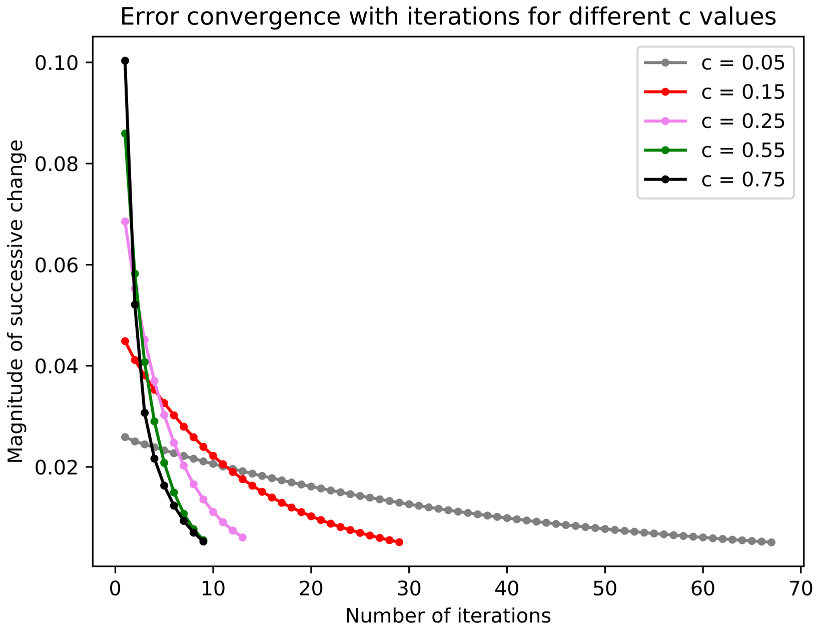
5.4 Convergence Summary For PageRank

The PageRank algorithm uses the power method to compute the principal eigenvector of A. The rate of convergence of the power method is given by $|\lambda_2|$. For PageRank, the typical value of c has been given as 0.85. For this value of c, the convergence rate of the power method for the transition matrix A is determined by 0.85 and its powers. Therefore, the convergence rate of PageRank will be fast, even as the web scales.

6 Sample Web Matrix, Results and Convergence

In the spirit of truly testing and implementing the algorithm, a sample web graph was created with the root at 'https://www.cmi.ac.in' and a maximum depth of 3 by crawling the webpages in a breadth-first manner. The number of webpages or vertices registered were 5243. The said adjacency matrix was converted into irreducible and aperiodic row stochatic matrices for different values of c, c ϵ $\{0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95\}$. Subsequently, power method was used with an equiprobable column vector with each of the stochastic matrices.
Fun Trivia - The websites ordered by their PageRank are[c = 0.85]:
'https://www.cmi.ac.in/admissions/', 'https://www.cmi.ac.in//',
'https://www.cmi.ac.in/about/', 'https://www.cmi.ac.in/people/',
'https://www.cmi.ac.in/teaching/' etc.





A few key points to note,
beginenumerate

Algorithm converged for all values of c it was implemented for.

The marginal correction between successive iterations, shown in Fig 1, rapidly converges to 0 for all values of c it was implemented for.
The similar behaviour of sample web graph and matrix to the actual web matrix and its computation provide a fitting conclusion to the analysis of Google Matrix and PageRank algorithm, since both are formulated on and hence follow the same mathematical principles.