Radical Row-Stochastic Rankings

Yuchen Xin, Aidan Shah, Darby Brillon

1 Introduction

How do you rank a set of objects that are interconnected with each other?

When you look something up in Google, an algorithm decides which webpages to display on top of others. You might have realized that most of the time, you click on the first five webpages — rarely do you ever scroll down at all. This is all thanks to a statistical algorithm called PageRank that Google uses to rank webpages.

PageRank ranks webpages by scoring each webpage based on how likely a random web surfer is to arrive there, either by clicking on hyperlinks between webpages or by the occasional direct URL search. For a webpage to have a higher PageRank score, either there are a lot of hyperlinks from other webpages directing to it, or the webpages directing to it are highly ranked themselves.

Mathematically speaking, a matrix G is used to encode all the information about the hyperlinks between webpages, as well as how likely a random web surfer will type in URLs to certain webpages. Then, we represent the state of a random web surfer as a discrete probability distribution vector $\mathbf{v}$, whose i^{th} is the probability that the random web surfer ends up on webpage i.

Multiplying $\mathbf{v}$ with G will return the new discrete probability distribution after the web surfer clicks on hyperlinks or manually types in URLs. Finally, after multiplying by G an arbitrary amount of times, we hope to find a steady state probability vector that represents an "average" long-term behavior of the random web-surfer. If such a steady state vector exists, the steady state probability for being on each webpage corresponds with that webpage's PageRank score.

In this paper, we first introduce some basic theory about the matrix G, which turns out to be a special case of the more general class of stochastic, Markov-chain transition matrices. We then describe the PageRank algorithm in more mathematical detail, explaining how to encode arbitrary networks of webpages and random web surfer behavior in the matrix G. Next, we will give an example ranking for a hypothetical network of webpages. Finally, we will also explain how to extend this method to other types of rankings, namely sports rankings.

2 Background and Theory

First, we state some useful definitions:

Definition. *A stochastic or probability vector* $\mathbf{v} \in \mathbb{R}^n$ *is a real vector with all nonnegative elements that sum up to 1.*

Definition. *A row stochastic matrix is a square matrix where each row is a probability vector.*

Now, we state and prove a fundamental theorem:

Theorem 2.1. *If A is an $m \times n$ matrix and B an $n \times p$ matrix such that every row of A and B is a probability vector, then every row of the product AB is also a probability vector.*

Proof. Let $C = AB$ be the $m \times p$ product. Since each row of A and B is a probability vector, all elements of A and B are nonnegative. Then, for all $1 \leq i \leq m$ and $1 \leq j \leq p$,

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \geq 0$$

Since the product and sum of nonnegative numbers is still nonnegative.

It remains to check that each row of C sums up to 1. For each row $1 \leq i \leq m$,

$$\sum_{j=1}^{p} c_{ij} = \sum_{j=1}^{p} \left(\sum_{k=1}^{n} a_{ik} b_{kj} \right) = \sum_{k=1}^{n} \left(\sum_{j=1}^{p} a_{ik} b_{kj} \right) = \sum_{k=1}^{n} a_{ik} \left(\sum_{j=1}^{p} b_{kj} \right) = \sum_{k=1}^{n} a_{ik}(1) = 1$$

□

Corollary 2.1.1. *The product of two same dimension, row stochastic matrices is row stochastic. By induction, the product of any number of same dimension, row stochastic matrices is row stochastic.*

Later, we will use row stochastic matrices to describe transitions between states. In particular, given a row stochastic matrix G, each element g_{ij} represents the probability of transitioning to state j, *given* we start on state i. In probability notation, we have the conditional probability

$$g_{ij} = P(j' \mid i) = \frac{P(j' \cap i)}{P(i)}$$

where j' denotes being at state j after one transition.

Then, for some initial probability row vector $\mathbf{v}$ where each element $v_i = P(i)$ is the probability of starting at state i, performing the **left multiplication**[1] $\mathbf{v}G$ gives a new probability vector (Theorem 2.1), whose j^{th} element is

$$\sum_i v_i g_{ij} = \sum_i P(i) P(j' \mid i) = \sum_i P(i) \frac{P(j' \cap i)}{P(i)} = \sum_i P(j' \cap i) = P(j')$$

The summation gives the total probability of starting at any state with probability described by $\mathbf{v}$ and then transitioning to state j, which equals the probability of being at state j after one transition.

[1] Left multiplication of row stochastic matrices by probability vectors is used by convention in this field. In our computations later, we will switch back to the right multiplication we are familiar with.

If we multiply again by G, we get another probability vector representing the probabilities of being at each state after two transitions. In general, the j^{th} element of $\mathbf{v}G^k$ is the probability of being at state j after k transitions, starting at probability v_j. Here, G^k can be interpreted as the "k-step transition matrix". In the following section, we will discuss properties of the matrix G^k and the vector $\mathbf{v}G^k$ as $k \to \infty$.

Steady State Behavior

As discussed in the introduction, given some stochastic matrix G representing the hyperlink and URL probabilities, we want to find a steady state vector, defined below:

Definition. *A steady state vector for an $n \times n$ row stochastic matrix G is a probability row vector $\mathbf{w} \in \mathbb{R}^n$ such that for any initial probability row vector $\mathbf{v} \in \mathbb{R}^n$,*

$$\lim_{k\to\infty} \mathbf{v}G^k = \mathbf{w}$$

We will now prove the existence of the limit and the properties of the steady state vector $\mathbf{w}$.

Theorem 2.2. *If G is an $n \times n$ row stochastic matrix with no zero elements, then the limit*

$$\lim_{k\to\infty} G^k = W$$

exists, where W is a matrix with all rows equal to each other, and all elements positive.

Proof. Let G be an $n \times n$ row stochastic matrix with all elements greater than 0. If $n = 1$, then G is a 1×1 matrix, and since G is stochastic, it must be that $G = 1$. Then

$$\lim_{k\to\infty} G^k = \lim_{k\to\infty} 1^k = 1 = W.$$

Otherwise, suppose $n \geq 2$. Let $k \geq 1$ be an arbitrary integer.

Let $\mathbf{y} \in \mathbb{R}^n$ be an arbitrary column vector with each element between 0 and 1, inclusive. Additionally, let m_k be the minimum element and M_k the maximum element of $G^k\mathbf{y}$. We will prove that as $k \to \infty$, $G^k\mathbf{y} \to$ a column vector with equal elements to eventually show that $G^k \to W$ where W has all rows equal to each other (see footnote[2]).

We first want to show that the sequences m_k and M_k converge. Since G is row stochastic, G^k is also row stochastic by Corollary 2.1.1, so all elements of G are between 0 and 1, inclusive. Also, since each element of $\mathbf{y}$ is also between 0 and 1, inclusive, we know by definition of matrix multiplication that $G^k\mathbf{y}$ is a column vector with each element between 0 and 1, inclusive. So,

$$0 \leq m_k, M_k \leq 1 \quad \text{for all} \quad k \geq 1$$

We now show that m_k is monotonically increasing and M_k is monotonically decreasing. Let

$$G^{k-1}\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \text{so that} \qquad G^k\mathbf{y} = G(G^{k-1}\mathbf{y}) = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

[2] We use right multiplication only in the body of this proof. At the end, we show how right multiplication by $\mathbf{y}$ can be used to prove facts about left multiplication.

Suppose the minimum element of $G^{k-1}\mathbf{y}$ is $y_i = m_{k-1}$, and suppose row $\mathbf{f}$ of G multiplies with $G^{k-1}\mathbf{y}$ to produce the new minimum element m_k of $G^k\mathbf{y}$. In other words,

$$m_k = \mathbf{f} \cdot G^{k-1}\mathbf{y} = f_1y_1 + f_2y_2 + \cdots + f_iy_i + \cdots + f_ny_n$$

Since G is row stochastic we can write

$$f_i = 1 - f_1 - \cdots - f_{i-1} - f_{i+1} - \cdots f_n$$

Substituting this back in, we get

$$\begin{aligned} m_k &= f_1y_1 + \cdots + \Big(1 - f_1 - \cdots - f_{i-1} - f_{i+1} - \cdots f_n\Big)\, y_i + \cdots + f_ny_n \\ &= f_1y_1 + \cdots + \Big(y_i - f_1y_i - \cdots - f_{i-1}y_i - f_{i+1}y_i - \cdots f_ny_i\Big) + \cdots + f_ny_n \\ &= y_i + f_1(y_1 - y_i) + \cdots + f_{i-1}(y_{i-1} - y_i) + f_{i+1}(y_{i+1} - y_i) + \cdots + f_n(y_n - y_i) \end{aligned}$$

Now, for any $1 \leq r \leq n$ with $r \neq i$, we know $f_r(y_r - y_i) \geq 0$ because $f_r > 0$ as G was given to have no zero elements and $y_r \geq y_i$ as y_i is the minimum element. Therefore

$$m_k = y_i + f_1(y_1 - y_i) + \cdots + f_n(y_n - y_i) \quad \geq \quad y_i = m_{k-1}$$

and so the sequence m_k is monotonically increasing.

By the same argument, if we assumed $y_j = M_{k-1}$ and $\mathbf{h}$ to be the row that multiplies with $G^{k-1}\mathbf{y}$ to produce the new M_k, then we would have gotten

$$M_k = y_j + h_1(y_1 - y_j) + \cdots + h_{j-1}(y_{j-1} - y_j) + h_{j+1}(y_{j+1} - y_j) + \cdots + h_n(y_n - y_j)$$

Similarly, for any $1 \leq r \leq n$ with $r \neq j$, all $h_r(y_r - y_j) \leq 0$ because $h_r > 0$ by assumption and $y_r \leq y_j$ as y_j is the maximum element. So

$$M_k = y_j + h_1(y_1 - y_j) + \cdots + h_n(y_n - y_j) \quad \leq \quad y_j = M_{k-1}$$

which proves the sequence M_k is monotonically decreasing.

Since both sequences m_k and M_k are monotonic and bounded, they both converge to a limit.

Given that the limits exist, let $m_k \to m$ and $M_k \to M$, so that the limit $\lim_{k\to\infty}(M_k - m_k)$ can be split up to equal $M - m$. We can now prove that $m = M$ by showing $M - m = \lim\limits_{k\to\infty}(M_k - m_k) = 0$. We do so by bounding

$$0 \leq M_k - m_k \leq M_{k,\max} - m_{k,\min}$$

where $M_{k,\max}$ is an upper bound for M_k and $m_{k,\min}$ is a lower bound for m_k.

Recall that $y_i = m_{k-1}$ and $y_j = M_{k-1}$ were the respective minimum and maximum elements of $G^{k-1}\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$. Also recall that $\mathbf{f}$ is the row of G such that

$$m_k = \mathbf{f} \cdot G^{k-1}\mathbf{y} = f_1y_1 + \cdots + f_ny_n.$$

Since all elements of $\mathbf{f}$ are strictly positive, m_k attains a lower bound when all elements of $\mathbf{y}$ are m_{k-1} except $y_i = M_{k-1}$ (as defined earlier):

$$\begin{aligned} m_{k,\min} = f_1m_{k-1} + \cdots + f_jM_{k-1} + \cdots + f_nm_{k-1} \quad &= \quad f_jM_{k-1} + m_{k-1}\sum_{i=1,i\neq j}^{n} f_i \\ = f_jM_{k-1} + (1 - f_j)m_{k-1} \quad &= \quad m_{k-1} + (M_{k-1} - m_{k-1})f_j \end{aligned}$$

where we use the fact that the elements of $\mathbf{f}$ sum up to 1. Now, let $d > 0$ be the minimum element of G. If f_j is minimized (that is, $f_j = d$), then the lower bound becomes

$$m_{k,\min} = m_{k-1} + (M_{k-1} - m_{k-1})d$$

Similarly, we can show that M_k is bounded above by

$$M_{k,\max} = M_{k-1} + (m_{k-1} - M_{k-1})d$$

Then,[3]

$$\begin{aligned} M_k - m_k \le M_{k,\max} - m_{k,\min} &= (M_{k-1} - m_{k-1}) + 2(m_{k-1} - M_{k-1})d \\ &= (1-2d)(M_{k-1} - m_{k-1}) \end{aligned}$$

Inductively, we see that

$$0 \le M_k - m_k \le (1-2d)^k (M_0 - m_0)$$

where $0 \le M_k - m_k$ by definition of maximum and minimum. Finally, since the $n \times n$ matrix G is stochastic with $n \ge 2$, the minimum element $d > 0$ is bounded by $0 < d \le \frac{1}{2}$. Multiplying by -2 and then adding 1 on all sides,

$$1 > 1 - 2d \ge 0$$

so as $k \to \infty$, $(1-2d)^k \to 0$. Thus,

$$0 \le \lim_{k\to\infty} (M_k - m_k) \le \lim_{k\to\infty} (1-2d)^k (M_0 - m_0) = 0$$

Implying

$$0 = \lim_{k\to\infty} (M_k - m_k) = M - m$$

which proves $M = m$. But every element of $G^k\mathbf{y}$ is bounded between m_k and M_k, so as k gets arbitrarily large, all elements of $G^k\mathbf{y}$ approach the same number $u = M = m$. In other words, for all probability vectors $\mathbf{y} \in \mathbb{R}^n$,

$$\lim_{k\to\infty} G^k \mathbf{y} = \mathbf{u}$$

with $\mathbf{u}$ an n-column vector with each element being u.

Now, for each $1 \le i \le n$, pick $\mathbf{y}$ with the i^{th} component equal to 1 and all other components equal to 0. By definition of matrix multiplication, we can see that $\mathbf{u} = \lim\limits_{k\to\infty} G^k\mathbf{y}$ which equals the i^{th} column of $\lim\limits_{k\to\infty} G^k$. So all elements of the i^{th} column of $\lim\limits_{k\to\infty} G^k$ are equal to u.

In fact, since m_1 is the minimum value of $G\mathbf{y}$ which equals the i^{th} column of G, we must have $m_1 > 0$ be definition of G. Then, since we previously showed that m_k was monotonically increasing, we know $0 < m_1 \le m_2 \le \cdots \le \lim\limits_{k\to\infty} m_k = u$, so $u > 0$. So all elements of the i^{th} column of $\lim\limits_{k\to\infty} G^k$ are positive.

Therefore, all columns of $\lim_{k\to\infty} G^k$ exist, and have equal elements which are all strictly positive. We can write this as

$$\lim_{k\to\infty} G^k = W$$

with W a matrix with all rows equal to each other, and all elements positive. □

[3]See the remark after this proof regarding the inequality.

Remark. *The above proof for Theorem 2.2 showed* $\lim_{k\to\infty} G^k = W$ *by proving the inequality*

$$M_k - m_k \le (1-2d)(M_{k-1} - m_{k-1})$$

where we showed $0 \le 1-2d < 1$*. But this inequality looks very similar to a contraction mapping!*

If we rewrote our proof and based it on right multiplication by G*, then by deriving a similar inequality, we could have invoked the Banach Fixed-Point Theorem to show that* G^k *converges to a fixed point* W *such that* $WG = W$*. This is actually the topic of the next Theorem 2.3, although we do not use the Banach Fixed-Point Theorem.*

With Theorem 2.2, we can now show that the steady state vector is a fixed point of G.

Theorem 2.3. *Let* G *be a row stochastic matrix with no zero elements. Then* $\lim_{k\to\infty} G^k = W$*, where all rows of* W *are equal to* $\mathbf{w}$*. Additionally,* $\mathbf{w}$ *is the* ***unique*** *probability vector satisfying*

$$\mathbf{w}G = \mathbf{w}.$$

In other words, if we find a probability left eigenvector $\mathbf{w}$ *of* G *corresponding with eigenvalue 1, then* $\lim_{k\to\infty} G^k = W$ *where* W *is a square matrix whose rows are all equal to* $\mathbf{w}$*.*

Proof. Let G be a $n \times n$ row stochastic matrix with no zero elements. Then by Theorem 2.2, we know $\lim_{k\to\infty} G^k = W$, with W a square matrix with all rows equal to $\mathbf{w} \in \mathbb{R}^n$.

We can say that

$$\lim_{k\to\infty} G^{k+1} = \left(\lim_{k\to\infty} G^k\right) G = WG$$

but we also know that

$$\lim_{k\to\infty} G^{k+1} = \lim_{k\to\infty} G^k = W$$

so we can conclude that $W = WG$. But all rows of W are equal to $\mathbf{w}$, so by definition of matrix multiplication, we see that $\mathbf{w} = \mathbf{w}G$.

To prove the uniqueness of $\mathbf{w}$, let $\mathbf{v} \in \mathbb{R}^n$ be any probability row vector satisfying $\mathbf{v}G = \mathbf{v}$. Multiplying by G on both sides, we get $\mathbf{v}G^2 = \mathbf{v}G = \mathbf{v}$, and by induction, we can see that $\mathbf{v}G^k = \mathbf{v}$ for any integer $k \ge 1$. Taking limits, we have

$$\mathbf{v} = \lim_{k\to\infty} \mathbf{v} = \lim_{k\to\infty} \mathbf{v}G^k = \mathbf{v} \lim_{k\to\infty} G^k = \mathbf{v}W$$

Now, we note that by Corollary 2.1.1, W is a row stochastic matrix. Let a_{ij} be the elements of W, and let w_i be the elements of $\mathbf{w}$. Then, for any $1 \le i \le n$, the i^{th} component of $\mathbf{v}$ is

$$\begin{aligned} v_i &= \sum_{j=1}^{n} v_j a_{ji} && \text{Since } \mathbf{v} = \mathbf{v}W \\ &= \sum_{j=1}^{n} v_j w_i && \text{Since all rows of } W \text{ are the same} \\ &= w_i \sum_{j=1}^{n} v_j = w_i(1) = w_i && \text{Since } \mathbf{v} \text{ is a probability vector} \end{aligned}$$

So $\mathbf{v} = \mathbf{w}$. ∎

We are now ready for the final theorem that equates the steady state vector with the probability left eigenvector of G corresponding to eigenvalue 1.

Theorem 2.4. *Let G be a row stochastic matrix with no zero elements. If $\mathbf{w}$ is a probability vector satisfying $\mathbf{w}G = \mathbf{w}$, then for any initial probability vector $\mathbf{v}$,*

$$\lim_{k\to\infty} \mathbf{v}G^k = \mathbf{w}$$

Proof. Let G be an $n \times n$ row stochastic matrix, and $\mathbf{v} \in \mathbb{R}^n$ any probability row vector. Let $\mathbf{w} \in \mathbb{R}^n$ be a probability row vector satisfying $\mathbf{w}G = \mathbf{w}$. By Theorem 2.3, $\mathbf{w}$ is unique, and

$$\lim_{k\to\infty} G^k = W$$

with W a matrix with all rows equal to $\mathbf{w}$. Multiplying on the left by $\mathbf{v}$, we have

$$\mathbf{v} \lim_{k\to\infty} G^k = \lim_{k\to\infty} \mathbf{v}G^k = \mathbf{v}W$$

Now, let w_i be the elements of $\mathbf{w}$, and a_i be the elements of $\mathbf{v}W$. Then for each $1 \leq i \leq n$,

$$\begin{aligned} a_i &= \sum_{j=1}^{n} v_j w_i && \text{Since all rows of } W \text{ equal } \mathbf{w} \\ &= w_i \sum_{j=1}^{n} v_j = w_i(1) = w_i && \text{Since } \mathbf{v} \text{ is a probability vector} \end{aligned}$$

Thus, $\mathbf{v}W = \mathbf{w}$, and we have

$$\lim_{k\to\infty} \mathbf{v}G^k = \mathbf{w}$$

□

We have completed all the necessary proofs and are now able to treat any problem regarding finding the steady state vector of a stochastic matrix as an eigenvector problem. Indeed, Theorem 2.3 implies that for any row stochastic matrix G with no zero elements, there exists a unique probability row vector $\mathbf{w}$ such that $\mathbf{w}G = \mathbf{w}$. In other words, there is guaranteed to exist a unique left probability eigenvector corresponding with eigenvalue 1. Then by Theorem 2.4, we know $\mathbf{w}$ is exactly the steady state probability vector of G.

But how do we compute the probability row $\mathbf{w}$? Since $\mathbf{w}G = \mathbf{w}$, we can take the transpose of both sides to get

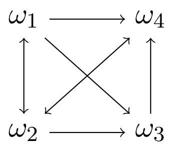
$$(\mathbf{w}G)^T = \mathbf{w}^T = G^T\mathbf{w}^T$$

Clearly, we can find $\mathbf{w}^T$ by finding the column eigenvector of G^T corresponding with eigenvalue 1, and normalizing the resulting vector to be a probability vector. Taking the transpose will give us the desired steady state vector $\mathbf{w}$.

In the following section, we will explain how Google's PageRank algorithm uses this exact problem solving strategy to rank webpages.

3 Google's PageRank Algorithm

We now describe how to implement PageRank with an example. Assume there are only four webpages $\omega_1, \omega_2, \omega_3, \omega_4$, with hyperlinks between them represented by the following graph:



This graph can be represented by a *hyperlink matrix* $H = (h_{ij})$, where

$$h_{ij} = \begin{cases} 1 & \text{there is a hyperlink from } \omega_i \text{ to } \omega_j \\ 0 & \text{otherwise} \end{cases}$$

In our example, we get the matrix

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Now, we seek to construct a transition matrix representing where a random web surfer will go for any webpage that they start on. We first only consider hyperlink transitions between webpages, given by H.

If a web surfer begins on webpage ω_i, we assume they are equally likely to click on any outgoing hyperlink from ω_i. To do so, we will convert the i^{th} row of H into a *probability vector*, representing where the web surfer is likely to go next upon arriving on webpage ω_i.

Since the web surfer will click on some hyperlink, the probabilities of each row must add up to 1. Thus, we will normalize each row by dividing each row by the sum of the elements in each row. For example, row 1 has elements $\begin{bmatrix} 0 & 1 & 1 & 1 \end{bmatrix}$, so we divide the row by $1 + 1 + 1 = 3$ to get $\begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$. The new row can be interpreted as "a random web surfer at ω_1 about to click on an outgoing hyperlink has a $1/3$ probability of going to either ω_2, ω_3, or ω_4".

Repeating for all rows, we get the *row stochastic matrix* S of probability vectors:

$$S = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

S models how a random web surfer would travel between webpages by only clicking on hyperlinks, with an equal likelihood of clicking on each hyperlink. However, to account for the ability of a random web surfer to manually type in URLs to navigate to different webpages, we introduce the personalization vector $\mathbf{p}$.

Definition. *The personalization vector* $\mathbf{p}$ *is a probability vector representing how likely each webpage is to be visited by typing in URLs manually, instead of via direct hyperlinks.* $\mathbf{p}$ *is constant and independent of which webpage the random web surfer is currently on.*

The personalization vector that Google currently uses is unknown, so for simplicity, we will use $\mathbf{p} = \begin{bmatrix} \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix}$ where n is the number of webpages. With this definition of $\mathbf{p}$, a random web surfer is equally likely to navigate to any webpage by typing in URLs manually.

If we just consider only the personalization vector, we get another row stochastic matrix

$$P = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} = \mathbf{zp}$$

where $\mathbf{z}$ is a column vector of ones. We note that each row is identical because $\mathbf{p}$ was defined to be a probability vector independent of which webpage the random web surfer is currently on.

Finally, we can combine S and P as a weighted sum to get the *Google matrix* G:

$$G = \alpha S + (1 - \alpha)P = \alpha S + (1 - \alpha)\mathbf{zp}$$

where α is the dampening factor, defined below.

Definition. *The dampening factor* $0 \leq \alpha \leq 1$ *weights the likelihood of navigating webpages via hyperlinks against navigating webpages via typing in URLs manually. A higher value of* α *indicates a greater reliance on hyperlinks to navigate webpages.*

Since S and P are both row stochastic matrices, and $0 \leq \alpha \leq 1$, it is clear that the final Google matrix G is also row stochastic. To be clear, each element g_{ij} completely describes the probability of a random web surfer going from webpage ω_i to webpage ω_j by any means.

In most studies, the dampening factor was set between 0.85 and 0.99. In our example, we will use $\alpha = 0.9$. We then have

$$G = \frac{9}{10} \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} + \frac{1}{10} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} = \frac{1}{40} \begin{bmatrix} 1 & 13 & 13 & 13 \\ 13 & 1 & 13 & 13 \\ 1 & 1 & 1 & 37 \\ 1 & 37 & 1 & 1 \end{bmatrix}$$

Our goal is to rank each page by how likely a random web surfer will arrive at them. This is described by the steady state vector $\mathbf{w}$ of the matrix, which we proved in Theorem 2.4[4] to be the long term probability distribution of a web surfer arriving at particular webpages, starting with any initial probability distribution. The i^{th} element of $\mathbf{w}$, the likelihood a random web surfer ends up at webpage ω_i in the long term, corresponds to the PageRank score for ω_i.

As discussed at the end of previous section, we can find $\mathbf{w}^T$ as the eigenvector corresponding with eigenvalue 1 for the matrix G^T. Since we know $\mathbf{w}$ exists, we can skip finding eigenvalues and directly work with the matrix $G^T - I$:

$$G^T - I = \frac{1}{40} \begin{bmatrix} 1 & 13 & 1 & 1 \\ 13 & 1 & 1 & 37 \\ 13 & 13 & 1 & 1 \\ 13 & 13 & 37 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \frac{1}{40} \begin{bmatrix} -39 & 13 & 1 & 1 \\ 13 & -39 & 1 & 37 \\ 13 & 13 & -39 & 1 \\ 13 & 13 & 37 & -39 \end{bmatrix}$$

[4]We can apply Theorem 2.4 because every element of G is strictly positive, which is guaranteed by the $\mathbf{zp}$ term and our definition of $\mathbf{p}$ as having all positive elements.

Now we perform row reduction to find the kernel of $G^T - I$:

$$\frac{1}{40}\begin{bmatrix} -39 & 13 & 1 & 1 \\ 13 & -39 & 1 & 37 \\ 13 & 13 & -39 & 1 \\ 13 & 13 & 37 & -39 \end{bmatrix} \longrightarrow \begin{bmatrix} -39 & 13 & 1 & 1 \\ 13 & -39 & 1 & 37 \\ 13 & 13 & -39 & 1 \\ 13 & 13 & 37 & -39 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & -104 & 4 & 112 \\ 13 & -39 & 1 & 37 \\ 0 & 52 & -40 & -36 \\ 0 & 0 & 76 & -40 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 13 & -39 & 1 & 37 \\ 0 & -104 & 4 & 112 \\ 0 & 0 & -38 & 20 \\ 0 & 0 & 76 & -40 \end{bmatrix} \longrightarrow \begin{bmatrix} 13 & -39 & 1 & 37 \\ 0 & -104 & 4 & 112 \\ 0 & 0 & -38 & 20 \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 13 & -39 & 1 & 37 \\ 0 & -104 & 4 & 112 \\ 0 & 0 & 1 & -\frac{10}{19} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 13 & -39 & 0 & \frac{713}{19} \\ 0 & -104 & 0 & \frac{2168}{19} \\ 0 & 0 & 1 & -\frac{10}{19} \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 13 & 0 & 0 & -\frac{800}{152} \\ 0 & -104 & 0 & \frac{2168}{19} \\ 0 & 0 & 1 & -\frac{10}{19} \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & -\frac{800}{1976} \\ 0 & 1 & 0 & -\frac{2168}{1976} \\ 0 & 0 & 1 & -\frac{10}{19} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

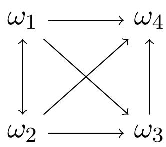
From this we get the system of equations

$$x_1 - \frac{800}{1976}x_4 = 0, \quad x_2 - \frac{2168}{1976}x_4 = 0, \quad x_3 - \frac{10}{19}x_4 = 0$$

So the kernel is generated by the vector $\begin{bmatrix}\frac{800}{1976} & \frac{2168}{1976} & \frac{10}{19} & 1\end{bmatrix}^T \approx \begin{bmatrix}0.405 & 1.097 & 0.526 & 1\end{bmatrix}^T$. We can find $\mathbf{w}$ by taking the transpose and normalizing this vector to a probability vector, but since that does not change the relative PageRank scores, we can skip finding $\mathbf{w}$ and directly order the elements from greatest to least to get the final webpage ranking: $\omega_2, \omega_4, \omega_3, \omega_1$ (from highest to lowest ranked).

Dangling Nodes

A dangling node is a node that does not map to another node. An example of is given in the following webpage diagram:



Here, ω_4 is a dangling node as it does not map to any other node in the diagram. This would create a row of all zeroes in our hyperlink matrix, which would make it not row stochastic and thus cause the Google matrix to not be row stochastic. If the Google matrix is not stochastic, there is no guaranteed steady state probability vector.

To resolve this, we convert every row of all zeroes in H (which are the dangling node rows) to the personalization row vector $\mathbf{p}$ in S, and then evaluate G normally. Since $G = \alpha S + (1-\alpha)\mathbf{zp}$, dangling nodes are maintained as $\mathbf{p}$ rows in G. This correctly models a random web surfer relying only on typing in URLs to leave a webpage that has no outgoing hyperlinks. In this example, we have

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad S = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

Then, using $\alpha = 0.9$ and $\mathbf{p} = \begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}$ again, we get

$$G = \frac{9}{10}\begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} + \frac{1}{10}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} = \frac{1}{40}\begin{bmatrix} 1 & 13 & 13 & 13 \\ 13 & 1 & 13 & 13 \\ 1 & 1 & 1 & 37 \\ 10 & 10 & 10 & 10 \end{bmatrix}$$

We can now proceed to find the steady state vector. As in the previous example, we will find the steady state vector by finding the kernel of $G^T - I$:

$$G^T - I = \frac{1}{40}\begin{bmatrix} 1 & 13 & 13 & 13 \\ 13 & 1 & 13 & 13 \\ 1 & 1 & 1 & 37 \\ 10 & 10 & 10 & 10 \end{bmatrix}^T - I = \frac{1}{40}\begin{bmatrix} 1 & 13 & 1 & 10 \\ 13 & 1 & 1 & 10 \\ 13 & 13 & 1 & 10 \\ 13 & 13 & 37 & 10 \end{bmatrix} - I = \frac{1}{40}\begin{bmatrix} -39 & 13 & 1 & 10 \\ 13 & -39 & 1 & 10 \\ 13 & 13 & -39 & 10 \\ 13 & 13 & 37 & -30 \end{bmatrix}$$

Doing matrix reduction on $G^T - I$ (by first eliminating the scalar $\frac{1}{40}$), we get

$$\begin{bmatrix} -39 & 13 & 1 & 10 \\ 13 & -39 & 1 & 10 \\ 13 & 13 & -39 & 10 \\ 13 & 13 & 37 & -30 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & -104 & 4 & 40 \\ 13 & -39 & 1 & 10 \\ 0 & 52 & -40 & 0 \\ 0 & 0 & 76 & -40 \end{bmatrix} \longrightarrow \begin{bmatrix} 13 & -39 & 1 & 10 \\ 0 & -104 & 4 & 40 \\ 0 & 52 & -40 & 0 \\ 0 & 0 & 76 & -40 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 13 & -39 & 1 & 10 \\ 0 & -104 & 4 & 40 \\ 0 & 0 & -38 & 20 \\ 0 & 0 & 76 & -40 \end{bmatrix} \longrightarrow \begin{bmatrix} 13 & -39 & 1 & 10 \\ 0 & -104 & 4 & 40 \\ 0 & 0 & -38 & 20 \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 13 & -39 & 1 & 10 \\ 0 & -104 & 4 & 40 \\ 0 & 0 & 1 & -\frac{10}{19} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 13 & -39 & 0 & \frac{200}{19} \\ 0 & -104 & 0 & \frac{800}{19} \\ 0 & 0 & 1 & -\frac{10}{19} \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 13 & 0 & 0 & -\frac{800}{152} \\ 0 & -104 & 0 & \frac{800}{19} \\ 0 & 0 & 1 & -\frac{10}{19} \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & -\frac{800}{1976} \\ 0 & 1 & 0 & -\frac{800}{1976} \\ 0 & 0 & 1 & -\frac{10}{19} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

From this we get the system of equations

$$x_1 - \frac{800}{1976}x_4 = 0, \quad x_2 - \frac{800}{1976}x_4 = 0, \quad x_3 - \frac{10}{19}x_4 = 0$$

So the kernel is generated by the vector $\begin{bmatrix} \frac{800}{1976} & \frac{800}{1976} & \frac{10}{19} & 1 \end{bmatrix}^T \approx \begin{bmatrix} 0.405 & 0.405 & 0.526 & 1 \end{bmatrix}^T$. Ordering the elements from greatest to least, we get the following ranking: ω_4 in first, ω_3 in second, and ω_1 and ω_2 tied for third.

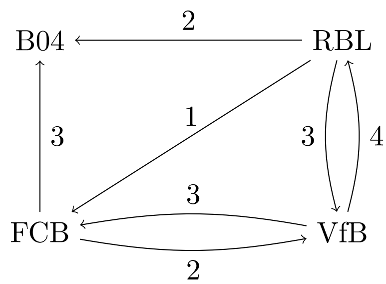
4 Extensions

The ranking algorithm described previously can be extended to analyzing other networks. In fact, we can generalize PageRank's method by representing an arbitrary network as a matrix of weighted connections.

An example of a use of this method is within sports. We can model each node as a team and each connection as the difference in scores for one game played between two teams. These connections will point from the losing team to the winning team. In contrast to the PageRank algorithm described earlier, each connection is weighted by the score difference, which can each be greater than 1. If two teams played multiple games against each other and one side won all of them, then the score differences are added together. If the wins were split, then there would be connections going both ways, each weighted with the sum of the score differences for all the wins of one team against the other. This is called the GEM method (as introduced by Zack et al.).

As an example, we will model the games played between the top 4 soccer teams in the German Bundesliga. We have four nodes: B04 (Bayern Leverkusen), FCB (Bayern Munich), VfB (VfB Stuttgart), and RBL (RB Leipzig).

In the 2023-2024 season, there were two games played between FCB - VfB with the scores 3 - 0 and 1 - 3. FCB won one game with a score difference of 3, while VfB won the other with a score difference of 2. Repeating for all other games in the season, we get the following diagram:



This would correspond with the hyperlink matrix

$$H = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 3 & 0 & 2 & 0 \\ 0 & 3 & 0 & 4 \\ 2 & 2 & 3 & 0 \end{bmatrix} \quad \begin{array}{l} \textit{(outgoing connections from B04)} \\ \textit{(outgoing connections from FCB)} \\ \textit{(outgoing connections from VfB)} \\ \textit{(outgoing connections from RBL)} \end{array}$$

Since B04 is a dangling node, we convert it to the personalization vector $\mathbf{p} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$ in our stochastic matrix S:

$$S = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{3}{5} & 0 & \frac{2}{5} & 0 \\ 0 & \frac{3}{7} & 0 & \frac{4}{7} \\ \frac{2}{7} & \frac{2}{7} & \frac{3}{7} & 0 \end{bmatrix}$$

Again, using $\alpha = 0.9$, we get

$$G = \frac{9}{10} \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{3}{5} & 0 & \frac{2}{5} & 0 \\ 0 & \frac{3}{7} & 0 & \frac{4}{7} \\ \frac{2}{7} & \frac{2}{7} & \frac{3}{7} & 0 \end{bmatrix} + \frac{1}{10} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{113}{200} & \frac{1}{40} & \frac{77}{200} & \frac{1}{40} \\ \frac{1}{40} & \frac{115}{280} & \frac{1}{40} & \frac{151}{280} \\ \frac{79}{280} & \frac{79}{280} & \frac{115}{280} & \frac{1}{40} \end{bmatrix}$$

We can now find the steady state vector by finding of $G^T - I$, as we have done before:

$$G^T - I = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{113}{200} & \frac{1}{40} & \frac{77}{200} & \frac{1}{40} \\ \frac{1}{40} & \frac{115}{280} & \frac{1}{40} & \frac{151}{280} \\ \frac{79}{280} & \frac{79}{280} & \frac{115}{280} & \frac{1}{40} \end{bmatrix}^T - I = \begin{bmatrix} \frac{1}{4} & \frac{113}{200} & \frac{1}{40} & \frac{79}{280} \\ \frac{1}{4} & \frac{1}{40} & \frac{115}{280} & \frac{79}{280} \\ \frac{1}{4} & \frac{77}{200} & \frac{1}{40} & \frac{115}{280} \\ \frac{1}{4} & \frac{1}{40} & \frac{151}{280} & \frac{1}{40} \end{bmatrix} - I$$

$$= \frac{1}{1400} \begin{bmatrix} -1050 & 791 & 35 & 395 \\ 350 & -1365 & 575 & 395 \\ 350 & 539 & -1365 & 575 \\ 350 & 35 & 755 & -1365 \end{bmatrix}$$

Eliminating the $\frac{1}{1400}$ scalar and doing row reduction, we get

$$\begin{bmatrix} -1050 & 791 & 35 & 395 \\ 350 & -1365 & 575 & 395 \\ 350 & 539 & -1365 & 575 \\ 350 & 35 & 755 & -1365 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & -3304 & 1760 & 1580 \\ 350 & -1365 & 575 & 395 \\ 0 & 1904 & -1940 & 180 \\ 0 & 1400 & 180 & -1760 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 350 & -1365 & 575 & 395 \\ 0 & -3304 & 1760 & 1580 \\ 0 & 1904 & -1940 & 180 \\ 0 & 1400 & 180 & -1760 \end{bmatrix} \longrightarrow \begin{bmatrix} 350 & -1365 & 575 & 395 \\ 0 & -3304 & 1760 & 1580 \\ 0 & 0 & -\frac{3058720}{3304} & \frac{3603040}{3304} \\ 0 & 0 & \frac{3058720}{3304} & -\frac{3603040}{3304} \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 350 & -1365 & 575 & 395 \\ 0 & -3304 & 1760 & 1580 \\ 0 & 0 & 1 & -\frac{3603040}{3058720} \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 350 & -1365 & 0 & \frac{2928520}{2731} \\ 0 & -3304 & 0 & \frac{9976900}{2731} \\ 0 & 0 & 1 & -\frac{3603040}{3058720} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 350 & 0 & 0 & -\frac{2386585}{5462} \\ 0 & -3304 & 0 & \frac{9976900}{2731} \\ 0 & 0 & 1 & -\frac{3603040}{3058720} \\ 0 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & -\frac{2386585}{1911700} \\ 0 & 1 & 0 & -\frac{9976900}{9023224} \\ 0 & 0 & 1 & -\frac{3603040}{3058720} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

From this we get the system of equations

$$x_1 - \frac{2386585}{1911700} x_4 = 0, \quad x_2 - \frac{9976900}{9023224} x_4 = 0, \quad x_3 - \frac{3603040}{3058720} x_4 = 0$$

So the kernel is generated by the vector $\begin{bmatrix} \frac{2386585}{1911700} & \frac{9976900}{9023224} & \frac{3603040}{3058720} & 1 \end{bmatrix}^T$ which is approximately equal to $\begin{bmatrix} 1.248 & 1.106 & 1.178 & 1 \end{bmatrix}^T$. This gives us the final ranking: Bayern Leverkusen in first, VfB Stuttgart in second, Bayern Munich in third, and then RB Leipzig in fourth. If we alternatively ranked these four teams using the soccer scoring system of 0 points for a loss, 1 point for a draw, and 3 points for a win, we would get the same ranking. So the GEM method clearly produces accurate ranking results in sports.

5 Conclusion

PageRank is a powerful algorithm that has in part shaped our modern world due to the role it plays in Google's search engine. Underpinning it is simple linear algebra that ends up doing very powerful things. By just finding an eigenvector of an easily computable matrix, we can ensure that the most useful webpages are shown to readers first. We have seen that this method can be used to rank webpages for every search on Google.com, but also for any arbitrary network, such as sports tournaments. This is just one of the few examples where linear algebra is used to model complex systems and describe its properties.

References

[1] Bundesliga: Matchday 34: Season 2023-2024. Bundesliga. https://www.bundesliga.com/en/bundesliga/matchday/2023-2024.

[2] Grinstead, C., & Snell, L. (2023, August 4). Introductory Probability. Statistics LibreTexts. https://stats.libretexts.org/Bookshelves/Probability_Theory/Book%3A_Introductory_Probability_(Grinstead_and_Snell).

[3] Wikipedia contributors. (2024, May 15). PageRank. Wikipedia. https://en.wikipedia.org/wiki/PageRank.

[4] Zack, L., Lamb, R., & Ball, S. (2013). An application of Google's PageRank to NFL rankings. Involve, 5(4), 463–471.