# Stochastic matrices, PageRank, and the Google matrix

Siltanen/Railo/Kaarnioja

Spring 2018
Applications of matrix computations

# Introduction

**Example**

In Crazy town, even the weather is strange.

- If today is sunny, then tomorrow is sunny with probability 0.7.
- If today is rainy, then tomorrow is rainy with probability 0.8.

What is the proportion of sunny days to rainy days?

Let the space of possible states be $S = \{x_1, x_2\}$, where

| | |
|---|---|
| $x_1$ | sunny day |
| $x_2$ | rainy day |

We can collect all different transitions $x_i \to x_j$ between days into an array:

| *tomorrow / today* | $x_1$ | $x_2$ |
|---|---|---|
| $x_1$ | 0.7 | 0.3 |
| $x_2$ | 0.2 | 0.8 |

| tomorrow / today | $x_1$ | $x_2$ |
|---|---|---|
| $x_1$ | 0.7 | 0.3 |
| $x_2$ | 0.2 | 0.8 |

$$\Rightarrow \quad P := \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} := \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}$$

Let $\pi_1^t \in [0, 1]$ be the probability that day $t$ is sunny and let $\pi_2^t = 1 - \pi_1^t$ be the probability that day $t$ is rainy.

Day $t + 1$ is sunny with probability

$$\pi_1^{t+1} = p_{11}\pi_1^t + p_{21}\pi_2^t. \qquad \text{("law of total probability")}$$

Day $t + 1$ is rainy with probability

$$\pi_2^{t+1} = p_{12}\pi_1^t + p_{22}\pi_2^t. \qquad \text{("law of total probability")}$$

In matrix form,

$$\boldsymbol{\pi}_{t+1}^{\mathrm{T}} = \boldsymbol{\pi}_t^{\mathrm{T}} P, \quad \text{where } \boldsymbol{\pi}_t^{\mathrm{T}} = [\pi_1^t, \pi_2^t].$$

| tomorrow / today | $x_1$ | $x_2$ |
|---|---|---|
| $x_1$ | 0.7 | 0.3 |
| $x_2$ | 0.2 | 0.8 |

$$\Rightarrow \quad P := \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} := \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}$$

It is easy to diagonalize $P$ as

$$P = \begin{bmatrix} 1 & -3/2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 2/5 & 3/5 \\ -2/5 & 2/5 \end{bmatrix} \tag{1}$$

so we obtain the *limiting probability distribution*

$$\lim_{t \to \infty} \boldsymbol{\pi}_t^{\mathrm{T}} = \lim_{k \to \infty} \boldsymbol{\pi}_0^{\mathrm{T}} P^k = \left[2/5, 3/5\right]$$

*regardless of the initial state* $\boldsymbol{\pi}_0^{\mathrm{T}} = [\pi_1^0, \pi_2^0]$, $\pi_1^0 + \pi_2^0 = 1$.

As we will see, it is not a coincidence that this is also an *equilibrium state* $[2/5, 3/5] = [2/5, 3/5]P$, i.e., a left-eigenvector, of $P$. Therefore, the fact that this equilibrium appears on the right-hand side modular matrix in equation (1) is no coincidence either! Why? :)

Discrete time Markov chains

# Discrete time Markov chains

Let $S = \{x_1, \ldots, x_n\}$ be a finite set of possible states.[1] A sequence $(X_t)_{t=0}^{\infty}$ of random variables is called a *Markov chain* if the value of any $X_{t+1}$ depends only on the previous state $X_t$:

$$\Pr(X_{t+1} = x | X_t, \ldots, X_0) = \Pr(X_{t+1} = x | X_t).$$

Furthermore, we assume that the Markov process is time homogeneous, i.e., the transition probabilities do not change over time.

[1] The definition naturally extends to countably infinite state spaces, but here we only consider the finite case.

We can tabulate all possible transition probabilities as

$$p_{i,j} = \Pr(X_{t+1} = x_j | X_t = x_i) \in [0, 1]$$

and we require that $p_{i,1} + \cdots + p_{i,n} = 1$ for all $i \in \{1, \ldots, n\}$ to conserve the probability. Because of time homogeneity, these values do not change as time increases.

The $n \times n$ matrix containing $p_{i,j}$ as its $(i, j)$ element,

$$P = \begin{bmatrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \ddots & \vdots \\ p_{n,1} & \cdots & p_{n,n} \end{bmatrix},$$

is called a *(row) stochastic matrix*[2] and it can be used to nicely encapsulate the properties of a Markov chain. Note that each row sums to 1.

[2] Sometimes it is also convenient to consider column stochastic matrices $P^{\mathrm{T}}$, where each column sums to 1.

Let $(X_t)_{t=0}^{\infty}$ be a discrete time Markov chain over a finite state space. Let the PMF of $X_t$ be given by the vector $\boldsymbol{\pi}_t^{\mathrm{T}} = [\pi_1^t, \ldots, \pi_n^t]$. Then the PMF $\boldsymbol{\pi}_{t+1}^{\mathrm{T}} = [\pi_1^{t+1}, \ldots, \pi_n^{t+1}]$ of $X_{t+1}$ is given by the equations

$$\begin{cases} \pi_1^{t+1} = p_{1,1}\pi_1^t + \cdots + p_{n,1}\pi_n^t, \\ \vdots \\ \pi_n^{t+1} = p_{1,n}\pi_1^t + \cdots + p_{n,n}\pi_n^t, \end{cases} \qquad \text{(“law of total probability”)}$$

i.e., $\boldsymbol{\pi}_{t+1}^{\mathrm{T}} = \boldsymbol{\pi}_t^{\mathrm{T}} P$.

In particular, the $k^{\text{th}}$ state can be obtained via power iteration:

$$\begin{aligned} \boldsymbol{\pi}_1^{\mathrm{T}} &= \boldsymbol{\pi}_0^{\mathrm{T}} P \\ \boldsymbol{\pi}_2^{\mathrm{T}} &= \boldsymbol{\pi}_1^{\mathrm{T}} P = \boldsymbol{\pi}_0^{\mathrm{T}} P^2, \\ \boldsymbol{\pi}_3^{\mathrm{T}} &= \boldsymbol{\pi}_2^{\mathrm{T}} P = \boldsymbol{\pi}_0^{\mathrm{T}} P^3, \\ &\vdots \\ \boldsymbol{\pi}_k^{\mathrm{T}} &= \boldsymbol{\pi}_{k-1}^{\mathrm{T}} P = \boldsymbol{\pi}_0^{\mathrm{T}} P^k. \end{aligned}$$

# Some elementary properties of stochastic matrices

- If $P$ is a stochastic matrix, then $P^k$ is a stochastic matrix for all $k \in \mathbb{Z}_+$.

Let $P = (p_{i,j})_{i,j}$ with $\sum_j p_{i,j} = 1$ for all $j$. Let us check that this condition holds for $P^2$:

$$\sum_j (P^2)_{i,j} = \sum_j \sum_k p_{i,k} p_{k,j} = \sum_k p_{i,k} \sum_j p_{k,j} = 1.$$

The case of general $k \in \mathbb{Z}_+$ is proven by induction.

- Let $\Delta^n = \{[x_1, \ldots, x_n] \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 1,\ x_i \geq 0 \text{ for all } 1 \leq i \leq n\}$. If $\mathbf{x}^{\mathrm{T}} \in \Delta^n$, then $\mathbf{x}^{\mathrm{T}} P \in \Delta^n$ for any (row) stochastic $n \times n$ matrix $P$.

Let $P = (p_{i,j})_{i,j}$ with $\sum_j p_{i,j} = 1$ for all $j$. This is proved analogously to the previous case:

$$\sum_i (\mathbf{x}^{\mathrm{T}} P)_i = \sum_i \sum_k x_k p_{k,i} = \sum_k x_k \sum_i p_{k,i} = 1.$$

In the following, let $P$ be an $n \times n$ (row) stochastic matrix corresponding to some Markov chain with $n$ states.

**Definition**

A Markov chain is said to have a *limiting probability distribution* if $\boldsymbol{\pi}_\infty^{\mathrm{T}} = \lim_{k\to\infty} \boldsymbol{\pi}_0^{\mathrm{T}} P^k$ exists for any initial distribution $\boldsymbol{\pi}_0$.

The limiting probability distribution is the ultimate distribution that the random variables $(X_t)_{t=0}^{\infty}$ tend toward – making the Markov chain "forget" its starting state. It can be used to describe the fraction of time that the Markov process stays on each state as $t \to \infty$.

**Definition**

The probability distribution $\boldsymbol{\pi}$ satisfying $\boldsymbol{\pi}^{\mathrm{T}} = \boldsymbol{\pi}^{\mathrm{T}} P$ is called the *equilibrium state*.

That is, the equilibrium state is a probability distribution which remains *unchanged* in the Markov process.

# The equilibrium state

**Proposition**

*An equilibrium state* $\boldsymbol{\pi}^{\mathrm{T}} = \boldsymbol{\pi}^{\mathrm{T}} P$ *always exists, where* $\boldsymbol{\pi}^{\mathrm{T}} = [\pi_1, \ldots, \pi_n]$ *is a nonnegative vector normalized such that* $\pi_1 + \cdots + \pi_n = 1$.

To see this, let us consider the eigenvalue problem
$\lambda\boldsymbol{\pi}^{\mathrm{T}} = \boldsymbol{\pi}^{\mathrm{T}} P \;\Leftrightarrow\; P^{\mathrm{T}}\boldsymbol{\pi} = \lambda\boldsymbol{\pi}$.

- 1 is an eigenvalue of $P^{\mathrm{T}}$.

Let $\mathbf{1} = [1, \ldots, 1]^{\mathrm{T}}$. 1 is obviously an eigenvalue of the matrix $P$ since $P\mathbf{1} = [p_{1,1} + \cdots + p_{1,n}, \ldots, p_{n,1} + \cdots + p_{n,n}]^{\mathrm{T}} = \mathbf{1}$.

Since $\det(P - \lambda I) = \det(P^{\mathrm{T}} - \lambda I)$, the matrices $P$ and $P^{\mathrm{T}}$ have the same eigenvalues (but generally *not* the same eigenvectors!).

- All eigenvalues of $P^{\mathrm{T}}$ satisfy $|\lambda| \leq 1$.

We note that it is sufficient to consider the eigenvalues of $P$. Normalize the eigenvector $\mathbf{x}$ so that index $i$ corresponds to the maximal component $x_i = 1$ and $|x_j| \leq 1$ for all $j \neq i$. Then we obtain from $P\mathbf{x} = \lambda\mathbf{x}$ that

$$\sum_j p_{i,j} x_j = \lambda x_i \Leftrightarrow p_{i,i} x_i + \sum_{j \neq i} p_{i,j} x_j = \lambda x_i \Leftrightarrow \lambda - p_{i,i} = \sum_{j \neq i} p_{i,j} x_j.$$

Taking absolute values on both sides and using the triangle inequality, we obtain

$$|\lambda - p_{i,i}| \leq \sum_{j \neq i} p_{i,j} = 1 - p_{i,i}.$$

Since $|\lambda - p_{i,i}| \geq ||\lambda| - |p_{i,i}|| \geq |\lambda| - p_{i,i}$, we have

$$|\lambda| \leq 1$$

and the claim follows since $\lambda$ was arbitrary.

Thus, we know that there exists some $\boldsymbol{\pi} \in \mathbb{R}^n$ such that $\boldsymbol{\pi}^{\mathrm{T}} = \boldsymbol{\pi}^{\mathrm{T}} P$.

To see why its elements are nonnegative, we must invoke the spectral theory of nonnegative matrices.

### Theorem (Weak Perron–Frobenius theorem[3])

*Let $A$ be an $n \times n$ matrix with nonnegative entries. Then*

- *There exists a positive eigenvalue $\lambda_1 > 0$ such that for all other eigenvalues $\lambda$ of $A$ it holds that $|\lambda| \leq \lambda_1$.*
- *The eigenvector $\mathbf{x}$ corresponding to $\lambda_1$ contains nonnegative entries.*

By the previous discussion, $\lambda_1 = 1$ for column stochastic $P^{\mathrm{T}}$ and thus the eigenvector $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_n]^{\mathrm{T}}$ corresponding to it must have nonnegative entries. Naturally, the eigenvector $\boldsymbol{\pi}$ can be normalized s.t. $\pi_1 + \cdots + \pi_n = 1$ allowing interpretation as a probability distribution.

[3] Notice that the weak formulation does not imply that the largest eigenvalue is simple! Thus, without additional assumptions, it leaves open the possibility of multiple equilibra!

Hence an equilibrium state $\boldsymbol{\pi}^{\mathrm{T}} = \boldsymbol{\pi}^{\mathrm{T}} P$ always exists for a Markov chain.

On the other hand, if the limiting probability distribution $\boldsymbol{\pi}_{\infty}^{\mathrm{T}} = \lim_{k\to\infty} \boldsymbol{\pi}_{0}^{\mathrm{T}} P^{k}$ exists, then it is an equilibrium state: $\boldsymbol{\pi}_{\infty}^{\mathrm{T}} = \boldsymbol{\pi}_{\infty}^{\mathrm{T}} P$.

However:

- A Markov chain may have an equilibrium but no limiting distribution independent of the starting state, e.g., `P`=[0,1;1,0].
- Multiple equilibra $\leftrightarrow$ different starting states either converge to different limits or may not converge at all.

The interesting questions are:

- When is the equilibrium solution $\boldsymbol{\pi}^{\mathrm{T}} = \boldsymbol{\pi}^{\mathrm{T}} P$ unique?
  $\rightarrow$ This means that *IF* the power method converges, then it must converge toward the unique limiting/equilibrium state.
- When does the power iteration converge?
  $\rightarrow$ For a system of sufficient size (as we shall see with the Google PageRank), calling `eig` (or `eigs`) may not be reasonable. Instead, iterative solvers must be used to compute the limiting distribution and, ideally, one wants to be sure that the iterative scheme converges.

In essence, our Markov chain needs to be "sufficiently nice" for both of these points to hold.

The $n \times n$ matrix $A$ is called *primitive* if there exists some $k \in \mathbb{Z}_+$ such that $(A^k)_{i,j} > 0$ for all $i, j \in \{1, \ldots, n\}$. Note that here we do not assume that $A$ is stochastic!

Uniqueness of the equilibrium vector can be ensured, e.g., under the following circumstances.

**Theorem (Perron)**

*A real primitive matrix $P$ has*

- *a unique eigenvector* $\mathbf{r}$ *with all positive entries,*
- *the eigenvalue $\lambda$ corresponding to* $\mathbf{r}$ *is dominant and has multiplicity one.*

*In addition, the Collatz–Wielandt formula holds:*

$$\lambda = \max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \mathbf{x} \geq \mathbf{0}}} \min_{i : x_i > 0} \frac{(A\mathbf{x})_i}{\mathbf{x}_i} = \min_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \mathbf{x} \geq \mathbf{0}}} \max_{i : x_i > 0} \frac{(A\mathbf{x})_i}{\mathbf{x}_i}.$$

# Power iteration (Collatz–Wielandt)

When the matrix $A$ is primitive, the Collatz–Wielandt formula implies a stronger version of the power method!
(See [Golub and Van Loan 4th edition, Exercise P7.3.4] and [Varga].)

**Algorithm**

*Let $A$ be a primitive $n \times n$ matrix.*
*Start with an initial guess $\mathbf{x}^0 \in \mathbb{R}^n_+$.*
**for** $k = 1, 2, \ldots$ **do**
  *Set* $\mathbf{y}^k = \mathbf{x}^{k-1} / \|\mathbf{x}^{k-1}\|$;
  *Compute* $\mathbf{x}^k = A\mathbf{y}^k$;
  *Set* $\overline{\lambda_k} = \max_{1 \le i \le n} \frac{(\mathbf{x}^k)_i}{(\mathbf{y}^k)_i}$ *and* $\underline{\lambda_k} = \min_{1 \le i \le n} \frac{(\mathbf{x}^k)_i}{(\mathbf{y}^k)_i}$;
**end for**
*Then* $\underline{\lambda}_0 \le \underline{\lambda}_1 \le \cdots \le \lambda \le \cdots \le \overline{\lambda}_1 \le \overline{\lambda}_0$. *In addition,* $(\overline{\lambda}_k, \mathbf{x}^k) \to (\lambda, \mathbf{x})$ *and* $(\underline{\lambda}_k, \mathbf{x}^k) \to (\lambda, \mathbf{x})$, *where* $(\lambda, \mathbf{x})$ *is the dominant eigenpair of A.*

### Definition

Let $P$ be a stochastic matrix. $P$ is called *regular* if there exists $k \in \mathbb{Z}_+$ such that $(P^k)_{i,j} > 0$ for all $i, j \in \{1, \dots, n\}$.

In other words, if there is a nonzero probability to reach each state $x_j$ from any state $x_i$ in $k$ steps, then the stochastic matrix $P$ is called regular.
$\rightarrow$ A regular matrix $P$ is primitive.
$\rightarrow$ A regular matrix $P$ satisfies the conditions of Perron's theorem and therefore there exists a unique equilibrium $\boldsymbol{\pi}$ such that

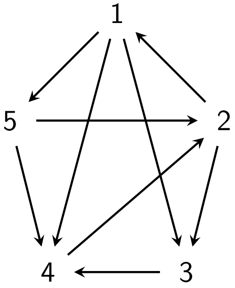
$$\boldsymbol{\pi}^{\mathrm{T}} = \lim_{k \to \infty} \boldsymbol{\pi_0}^{\mathrm{T}} P^k$$

for any $\boldsymbol{\pi}_0^{\mathrm{T}} = [\pi_1^0, \dots, \pi_n^0] \in \mathbb{R}_+^n$ such that $\pi_1^0 + \cdots + \pi_n^0 = 1$.

*Remark*. The assumption of primitivity can be relaxed. If the Markov chain is irreducible and aperiodic, then the [Perron–Frobenius theorem] asserts the uniqueness of the equilibrium solution. However, we omit this discussion.

PageRank

# Five-page internet




Let us consider an internet of five pages, which link to each other according to the directed graph on the left.

First, we create an adjacency matrix $A$ to represent this network. We set $A_{i,j} = 1$ if page $j$ has a link to page $i$ and $A_{i,j} = 0$ otherwise. In this case,

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We wish to associate a ranking $r_i$ to page $i$ in such a way that $r_i > r_j$ indicates that page $i$ is more popular than page $j$.

# Probabilistic interpretation

Let's think of the ranking problem in terms of Markov chains.

In the following, let $n$ denote the number of webpages on the internet.

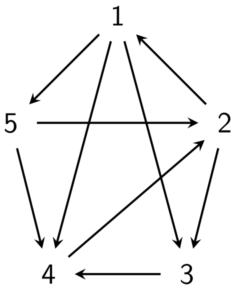
We make the following modeling assumptions:

- Links from page $i$ to itself are ignored.
- Multiple links to page $j$ from page $i$ are ignored.

One click of a hyperlink is modeled as one time step in the discrete time Markov process. We can replace the adjacency matrix with a *column* stochastic matrix $P$ by normalizing column $j$ with the number of outbound links from page $j$:

$$P_{i,j} = A_{i,j} / \sum_i A_{i,j}.$$

# Five-page internet




Adjacency matrix:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Column stochastic matrix:

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1/2 \\ 1/3 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 1 & 0 & 1/2 \\ 1/3 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

# PageRank

If the equilibrium solution

$$P\mathbf{r} = \mathbf{r}$$

is unique, it can be used to model the proportion of visits to each page on the internet.

**Definition (PageRank)**

The $i^{\text{th}}$ component of the equilibrium vector $\mathbf{r}$ is called the PageRank of page $i$.

Recall that the equilibrium solution is the dominant eigenvector of the Markov matrix $P \rightarrow$ In Matlab, `[r,~]=eigs(P,1)`

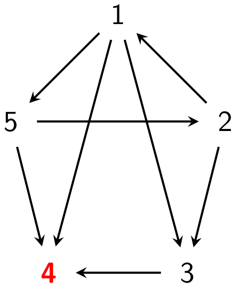
The indexed Web contains more than **4 billion webpages.**

Luckily, the adjacency matrix $A$ and, in consequence, the system $P$ are sparse. Since we are only interested in the dominant eigenpair, the power method provides excellent grounds for determining the equilibrium distribution!

However, there are certain network configurations where the power method fails to converge.

# Dangling node ($P$ matrix fails to be stochastic)

What if site 4 decides it wants to get all of the visitors and removes the link to site 2 from its webpage?




Site 4 is now a dangling node with no outgoing links.

Adjacency and Markov matrices:

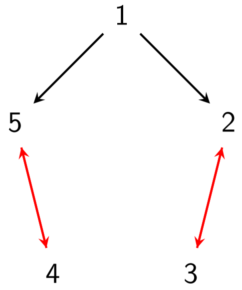
$$A = \begin{pmatrix} 0 & 1 & 0 & \mathbf{0} & 0 \\ 0 & 0 & 0 & \mathbf{0} & 1 \\ 1 & 1 & 0 & \mathbf{0} & 0 \\ 1 & 0 & 1 & \mathbf{0} & 1 \\ 1 & 0 & 0 & \mathbf{0} & 0 \end{pmatrix},$$

$$P = \begin{pmatrix} 0 & 1/2 & 0 & \mathbf{0} & 0 \\ 0 & 0 & 0 & \mathbf{0} & 1/2 \\ 1/3 & 1/2 & 0 & \mathbf{0} & 0 \\ 1/3 & 0 & 1 & \mathbf{0} & 1/2 \\ 1/3 & 0 & 0 & \mathbf{0} & 0 \end{pmatrix}.$$

Power method tends to 0!

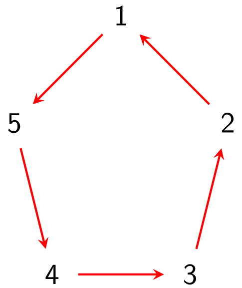
# Other potential issues

**Reducible network**




Different inital states may or may not converge to different distributions. *It is possible that the system has multiple equilibra.*

**Periodic network**




$P^5\mathbf{x} = \mathbf{x}$; we return to the same state infinitely many times. *Power iteration may not converge.*

These examples are overly simplified; it is enough that only one part of the network contains a loop/is separated from rest of the network.

# How does Google do it?

Suppose that we are given an arbitrary adjacency matrix $A$. For a large enough system, $A$ is likely to be reducible, periodic, or nonstochastic.

**The random surfer model**: with some fixed probability $\alpha \in [0, 1]$, a visitor on page $i$ clicks a link directing to page $j$; conversely, with probability $1 - \alpha$ the visitor visits another page at random with transition probability $1/n$.

In addition, if the site is a dangling node, then the visitor continues to an arbitrary webpage with probability $1/n$.

Enter the Google matrix:

$$G_{i,j} = \begin{cases} \alpha \frac{A_{i,j}}{\sum_i A_{i,j}} + (1-\alpha)\frac{1}{n}, & \text{if } \sum_i A_{i,j} \neq 0, \\ \frac{1}{n} & \text{otherwise.} \end{cases}$$

As the damping parameter, Google uses the value $\alpha = 0.85$.

$$G_{i,j} = \begin{cases} \alpha \frac{A_{i,j}}{\sum_i A_{i,j}} + (1-\alpha)\frac{1}{n}, & \text{if } \sum_i A_{i,j} \neq 0, \\ \frac{1}{n} & \text{otherwise,} \end{cases} \qquad \alpha = 0.85.$$

- The matrix $G$ is a column stochastic matrix by construction $\to$ the dominant eigenvector $\mathbf{r}$ is such that $G\mathbf{r} = \mathbf{r}$.
- The matrix $G$ is positive $\to$ by Perron's theorem, $\mathbf{r} \in \mathbb{R}_+^n$, and we can normalize it so that $\|\mathbf{r}\|_1 = 1$ is the unique equilibrium probability distribution of $G\mathbf{r} = \mathbf{r}$.
- 1 is a simple eigenvalue $\to$ for all eigenvalues $\lambda \neq 1$ of $G$, $|\lambda| < 1 \to$ the power method converges (see also the note on the Collatz–Wielandt formula).
- Actually, it can even be shown for the second largest eigenvalue of $G$ that $|\lambda_2| \leq \alpha \to$ the power method converges at rate $\mathcal{O}(\alpha^k)$ (cf., e.g., [Langville and Meyer, Theorem 4.7.1]).

# Computational remarks

$$G_{i,j} = \begin{cases} \alpha \frac{A_{i,j}}{\sum_i A_{i,j}} + (1-\alpha)\frac{1}{n}, & \text{if } \sum_i A_{i,j} \neq 0, \\ \frac{1}{n} & \text{otherwise,} \end{cases} \qquad \alpha = 0.85.$$

The Google matrix can be written as

$$G = \alpha P + \frac{\alpha}{n}\mathbf{1}\mathbf{v}^{\mathrm{T}} + \frac{1-\alpha}{n},$$

where $P$ is an $n \times n$ matrix such that

$$P_{i,j} = \begin{cases} \frac{A_{i,j}}{\sum_i A_{i,j}}, & \text{if } \sum_i A_{i,j} \neq 0 \\ 0 & \text{otherwise,} \end{cases}$$

$\mathbf{1} = [1, \ldots, 1]^{\mathrm{T}}$ is an $n$-vector of ones, and $\mathbf{v} \in \{0, 1\}^n$ is a Boolean vector, whose $j^{\text{th}}$ component is 1 if and only if $\sum_i A_{i,j} = 0$.

Now the action of the matrix

$$G = \alpha P + \frac{\alpha}{n}\mathbf{1}\mathbf{v}^{\mathrm{T}} + \frac{1-\alpha}{n}$$

can be expressed by

$$G\mathbf{x} = \alpha P\mathbf{x} + \frac{\alpha}{n}(\mathbf{v}^{\mathrm{T}}\mathbf{x})\mathbf{1} + \frac{1-\alpha}{n}\sum_{i=1}^{n} x_i, \quad \mathbf{x} = [x_1, \ldots, x_n]^{\mathrm{T}},$$

where $P$ is sparse so computing $P\mathbf{x}$ is possible and $(\mathbf{v}^{\mathrm{T}}\mathbf{x})$ is now simply a dot product of a sparse vector $\mathbf{v}$ against $\mathbf{x}$.

# Google PageRank algorithm

**Algorithm**

*Input: sparse* $P \in \mathbb{R}^{n \times n}$, $\mathbf{v} \in \{0,1\}^n$, *and* $\alpha \in [0,1]$ *as above.*
*Initialize* $\mathbf{r}^0 = \mathbf{1}/n$.
**for** $k = 1, 2, \ldots$ **do**
    *Compute* $\mathbf{r}^k = \alpha P \mathbf{r}^{k-1} + \frac{\alpha}{n}(\mathbf{v}^{\mathrm{T}} \mathbf{r}^{k-1})\mathbf{1} + \frac{1-\alpha}{n}$;
**end for**
*Since* $G$ *is a column stochastic matrix, each iteration is automatically normalized as* $\|G\mathbf{r}\|_1 = 1$ *for any* $\|\mathbf{r}\|_1 = 1$.

# Bibliography


- G. Golub and C. Van Loan. *Matrix Computations*, 4th edition, Johns Hopkins University Press, 2013.
- R. A. Horn and C. R. Johnson. *Matrix Analysis*, 1st paperback edition, Cambridge University Press, 1985.
- A. N. Langville and C. D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, 2006.
- D. Poole. *Linear Algebra: A Modern Introduction*. Thomson Brooks/Cole, 2005.
- R. S. Varga. *Matrix Iterative Analysis*, Springer Berlin Heidelberg, 1999.

Appendix

- Let $P$ be a row stochastic $n \times n$ matrix. Then there exists a vector $\boldsymbol{\pi} \in \mathbb{R}^n$ with nonnegative entries satisfying $\boldsymbol{\pi}^{\mathrm{T}} = \boldsymbol{\pi}^{\mathrm{T}} P$.

Let us denote the standard Euclidean $n$-simplex by

$$\Delta^n := \left\{ [x_1, \ldots, x_n] \in \mathbb{R}^n \mid \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0 \text{ for all } 1 \leq i \leq n \right\}.$$

Define the mapping

$$T(\mathbf{x}^{\mathrm{T}}) = \mathbf{x}^{\mathrm{T}} P.$$

Now we note that $\mathbf{x}^{\mathrm{T}} \in \Delta^n \Rightarrow \mathbf{x}^{\mathrm{T}} P \in \Delta^n$. Hence we have the following:

- $T \colon \Delta^n \to \Delta^n$ is continuous as a (bounded) linear mapping.
- The set $\Delta^n$ is convex and compact.

$\therefore$ [Brouwer's fixed-point theorem] implies that there exists $\mathbf{x}_0^{\mathrm{T}} \in \Delta^n$ such that $\mathbf{x}_0^{\mathrm{T}} = T(\mathbf{x}_0^{\mathrm{T}}) = \mathbf{x}_0^{\mathrm{T}} P$.

This essentially proves the weak Perron–Frobenius theorem for stochastic matrices!