

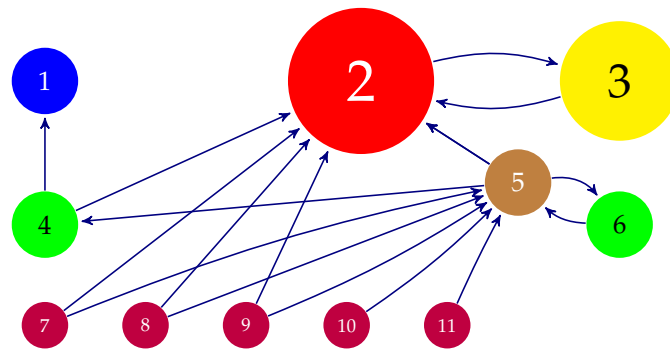
An Application of Eigenvectors: The PageRank

Aim: To rank results of a web search engine (such as Google) according to the “importance” of the web pages.

1998: For this purpose, Larry Page and Sergei Brin developed the PageRank algorithm as the basis of the Google empire.

Assumption: “important” means more links from other (important) web pages.

Idea: Let us think of the web as a directed graph, i.e., web pages are nodes and links from one page to another, i.e, from one node to another, are modeled as directed edges. For example a web structure consisting of 11 web pages could look as follows:



We now randomly place a random surfer according to the initial probability distribution $x^0 = (x_1^0, \dots, x_n^0)$ on this graph. Here, n (in the example above $n = 11$) denotes the number of web pages and x_i^0 denotes the probability that the random surfer starts at web page i . Further let $e := (1, \dots, 1)^T$, then the fact that x^0 is a probability distribution (i.e., probabilities sum up to 1) translates into $e^T x^0 = x_1^0 + \dots + x_n^0 = 1$. Now we make the assumption that the random surfer moves...

- (1) ...with probability $\alpha \in (0, 1)$ according to the link structure (with equal preferences to outgoing links)
 - (2) ...with probability $(1 - \alpha)$ he can teleport to a random page (with equal probability) to prevent stranding in deadlocks
- Pages, where the random surfer is more likely to appear in the long run based on the web's structure are considered more important.

These two movements can be described by multiplying the current probability distribution with the two following matrices:

$$P_1 = \begin{pmatrix} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 1 & 1 & & & 1/2 & & & & & & & \\ 2 & & & 1 & 1/2 & 1/3 & & 1/2 & 1/2 & 1/2 & & \\ 3 & & 1 & & & & & & & & & \\ 4 & & & & 1/3 & & & & & & & \\ 5 & & & & & 1 & 1/2 & 1/2 & 1/2 & 1 & 1 & \\ 6 & & & & 1/3 & & & & & & & \\ 7 & & & & & & & & & & & \\ 8 & & & & & & & & & & & \\ 9 & & & & & & & & & & & \\ 10 & & & & & & & & & & & \\ 11 & & & & & & & & & & & \end{pmatrix}, \quad P_2 := \frac{1}{n} e e^T = \left(\frac{1}{n} \right)_{ij}$$

More precisely,

- (1) **Link structure:** P_1 is the probability matrix (column stochastic) defined by $P_1^{ij} :=$ Probability that random surfer moves from page j to page i defined by the link structure
- (2) **Jumps:** P_2 is the probability matrix (column stochastic) defined by $P_2^{ij} := \frac{1}{n} =$ Probability that random surfer jumps from page j to page i

The movement of the random surfer is then completely defined by the probability matrix

$$P = \alpha P_1 + (1 - \alpha) P_2.$$

This matrix is also known as the **Google Matrix**. For the next time instances we therefore obtain

$$\begin{aligned} x^1 &= \alpha P_1 x^0 + (1 - \alpha) P_2 x^0 = P x^0 \\ x^2 &= \alpha P_1 x^1 + (1 - \alpha) P_2 x^1 = P x^1 \\ x^{k+1} &= \alpha P_1 x^k + (1 - \alpha) P_2 x^k = P x^k = P^{k+1} x^0 \\ x^* &= \lim_{k \rightarrow \infty} x^k =: \text{PageRank} \end{aligned}$$

Observations:

- One can easily show that P_1 , P_2 and thus P are column stochastic (i.e., $e^T P = e^T$)
- Consequently, since x^0 is a probability distribution (i.e., $e^T x^0 = 1$), also $e^T x^k = 1$ for all k and $e^T x^* = 1$

Question:

Does this sequence $\{x^k\}_{k \in \mathbb{N}}$ of vectors converge (to a steady state)? More precisely, is there a $x^* = \lim_{k \rightarrow \infty} x^k = \lim_{k \rightarrow \infty} P^k x^0$, so that

$$P x^* = 1 x^*. \quad (1)$$

→ With other words, is there an **eigenvector** x^* to the **eigenvalue** 1 of the matrix P ?

→ **Eigenvalue algorithms** are developed to solve such problems. One of them is the **Power iteration**, which, applied to the eigenvalue problem above, produces precisely the sequence

$$x^k = P^k x^0.$$

Intuitively, in the limit most of the “mass” would be located at web pages that have many incoming links and would therefore be ranked as being more important. In fact, the i -th component of x^* is called the **PageRank** of the web page i .

Remark: *Perron Theorem*

A positive damping factor $\alpha > 0$ is also technically necessary as it assures that the matrix P has only strictly positive coefficients. The Perron Theorem then states that its largest eigenvalue is strictly larger than all other eigenvalues (in magnitude). Thus the convergence of the Power method is guaranteed. Since the matrix is column stochastic one can further show that the largest eigenvalue is 1.

Solution: