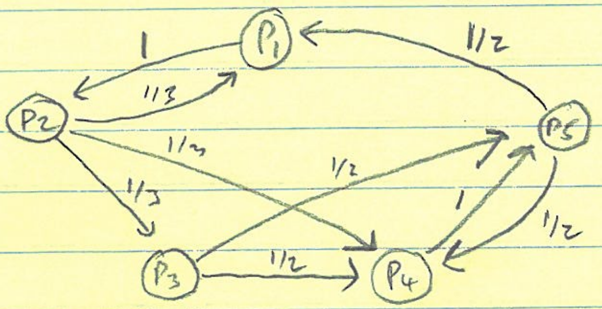
The internet page ranking problem

Suppose you type "XYZABC" into google.
Google identifies relevant pages $P_1, P_2, \ldots, P_n$. (n large).
How does it rank them? It uses the hyperlinks between them.



	ranking
$P_1 \leftarrow P_2, P_5$	2
$P_2 \leftarrow P_1$	4
$P_3 \leftarrow P_2$	4
$P_4 \leftarrow P_2, P_3, P_5$	1
$P_5 \leftarrow P_3, P_4$	2

We say P_4 is an "authority" on "XYZABC" because many pages link to it. But the only link on P_4 is to $\rightarrow P_5$ which "transfers" P_4's authority to P_5. Google: Imagine you surf $P_1, P_2, \ldots, P_n$ taking hyperlinks with <u>equal probability</u>. Let q_i be the probability you are on page P_i after $N =$ steps ($N = 10^9$). This will converge very slowly to

$$\begin{array}{cc} & \text{ranking} \\ q = \begin{bmatrix} 6/29 \\ 6/29 \\ 2/29 \\ 7/29 \\ 8/29 \end{bmatrix} & \begin{matrix} 3 \\ 3 \\ 5 \\ 2 \\ 1 \end{matrix} \end{array}$$

web surfer on P_1

$$R_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow R_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad R_2 = \begin{bmatrix} 1/3 \\ 0 \\ 1/3 \\ 1/3 \\ 0 \end{bmatrix} \quad R_4 = \begin{bmatrix} 0 \\ \frac{1}{3} \\ 0 \\ \frac{1}{6} \\ \frac{1}{6}+\frac{1}{3} \end{bmatrix} \quad \frac{1}{6}+\frac{1}{3} = \frac{1}{2}$$

R_0 is called the initial state or starting point.
The vectors $R_1, R_2, \ldots, R_k$ tell us the probability the web surfer is on page P_i after k steps.
The sequence $R_0, R_1, \ldots, R_{20}, \ldots$ converges rapidly.

Let Q be the $n \times n$ matrix where $Q_{ij} = \text{Prob}(P_j \to P_i)$

$$
Q = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{array}
\begin{array}{c}
\begin{array}{ccccc} P_1 & P_2 & P_3 & P_4 & P_5 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{array} \\
\begin{bmatrix} 0 & \frac{1}{3} & 0 & 0 & \frac{1}{2} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 1 & 0 \end{bmatrix}
\end{array}
\begin{array}{c} R_2 \\ \begin{bmatrix} \frac{1}{3} \\ 0 \\ \frac{1}{3} \\ \frac{1}{3} \\ 0 \end{bmatrix} \end{array}
=
\begin{array}{c} R_3! \\ \begin{bmatrix} 0 \\ \frac{1}{3} \\ 0 \\ \frac{1}{2}\cdot\frac{1}{3} = \frac{1}{6} \\ \frac{1}{2}\cdot\frac{1}{3} + \frac{1}{3} = \frac{1}{2} \end{bmatrix} \end{array}
\qquad \boxed{R_k = Q R_{k-1}}
$$

$R_1 = Q \cdot R_0, \quad R_2 = Q R_1 = Q(QR_0) = Q^2 R_0.$

$R_k = Q R_{k-1} = Q(Q^{k-1} R_{k-1}) = Q^k R_0.$

Section 4.1

Notice that the columns of Q are all probability vectors. Such a matrix is called a Markov matrix or stochastic matrix.

Theorem. If Q is a regular Markov matrix and p is any probability vector then

$$\lim_{k \to \infty} Q^k p = q \quad \text{and} \quad Qq = 1 \cdot q \quad \text{and } q \text{ is a prob. vector.}$$

eigenvector of Q

$\lambda = 1.$

[iterate until this converges] ①

[Solve for q] ②

Method ① [Google]. Starting with $R_0 = [\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}]$ compute $R_1 = Q \cdot R_0$, $R_2 = Q \cdot R_1, \dots, R_{100} = Q \cdot R_{99}, \dots$ until the values have converged to say 5 decimal places.

Example. See Maple handout

Method ② Solve $Qq = q$ for q.

q is in Null Space of $Q - I$.

$$Qq = q \Rightarrow Qq = Iq \Rightarrow Qq - Iq = 0 \Rightarrow (Q-I)q = 0$$

Compute a basis for $\mathrm{NUL}(Q - I)$ and pick q s.t. $\sum q_i = 1$.

Solving $(Q-I)x = 0$ using Maple I get

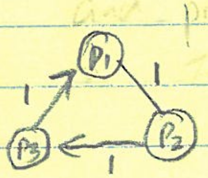
$$x = t \begin{bmatrix} 3/4 \\ 3/4 \\ 1/4 \\ 7/8 \\ 1 \end{bmatrix} \times \frac{8}{29} = \begin{bmatrix} 6/29 \\ 6/29 \\ 2/29 \\ 7/29 \\ 8/29 \end{bmatrix}$$

Σ 29/8

1st.

$\boxed{\text{Theory of Markov matrices 4.9}}$

Consider

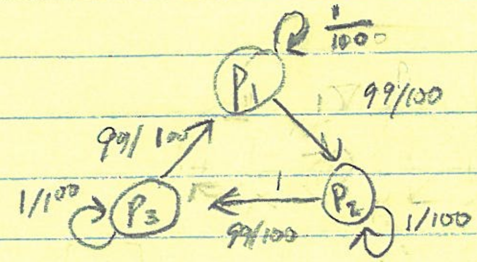


$$\begin{array}{cccc} R_0 & R_1 & R_2 & R_3 \\ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \cdots \end{array}$$

$$Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$ is not regular.

$\lim_{k \to \infty} Q^k R_0$ does not exist.

A fix to make Q regular



$$\begin{array}{ccc} R_0 & R_1 & R_2 \\ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} .01 \\ .99 \\ 0 \end{bmatrix} & \begin{bmatrix} .0001 \\ .0099 + .0099 \\ .9801 \end{bmatrix} \end{array}$$

$$\lim Q^k R_0 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$