University of Nottingham
UK | CHINA | MALAYSIA

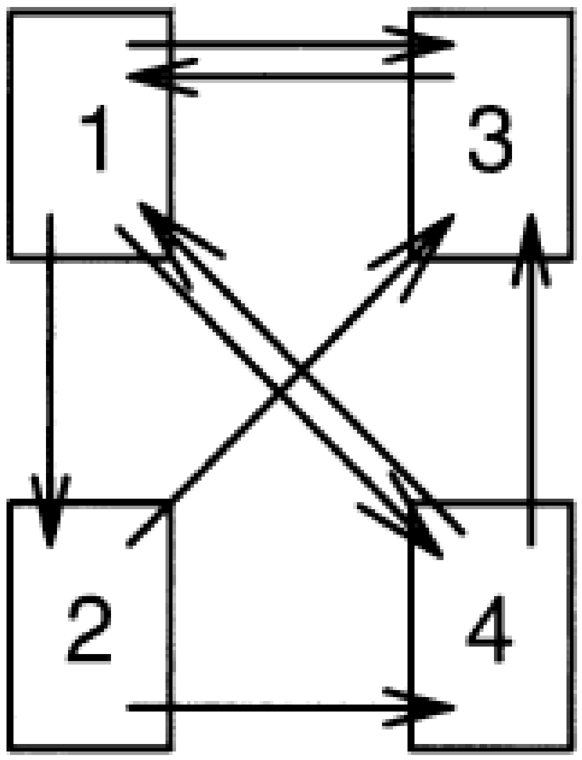
The $25,000,000,000 Eigenvector:
The Linear Algebra behind Google

Notes based on the article
by Kurt Bryan and Tanya Leise

PageRank : The Idea

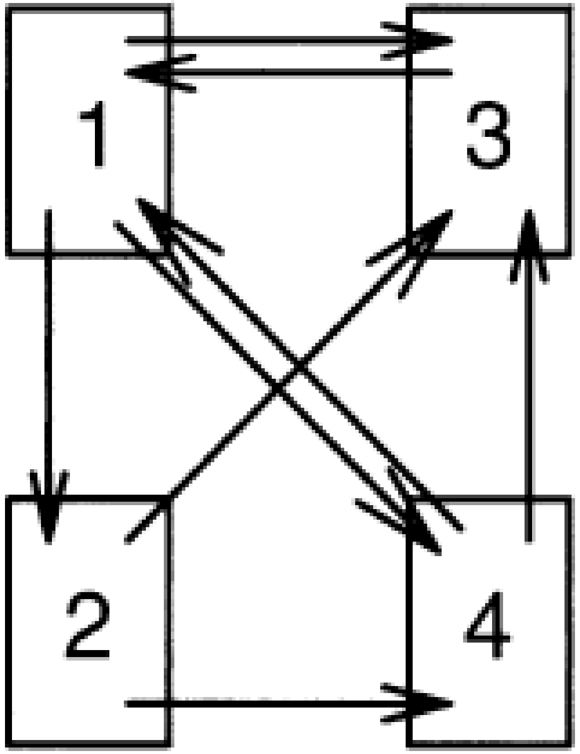
- Google's success derives in large part from its PageRank algorithm.
- PageRank ranks the importance of web pages according to an eigenvector of a weighted link matrix.
- The core idea is to a assign “an importance score” to any given web page based on the links made to that page from other pages.
- The weight of the link from a page is itself based on the importance of that page.

PageRank : The Idea



- Consider this simple example of a web with only 4 pages.
- An arrow from page A to page B indicates a link from A to B.
- In the general case, suppose the web of interest contains n pages, each page indexed by $k \in \{1, \cdots, n\}$.
- Let l_{jk} be an indicator (0/1) that page j links to page k.
- For example, $l_{12} = 1$, $l_{21} = 0$, $l_{23} = 1$, $l_{34} = 0$.

How do we calculate the Importance Scores?

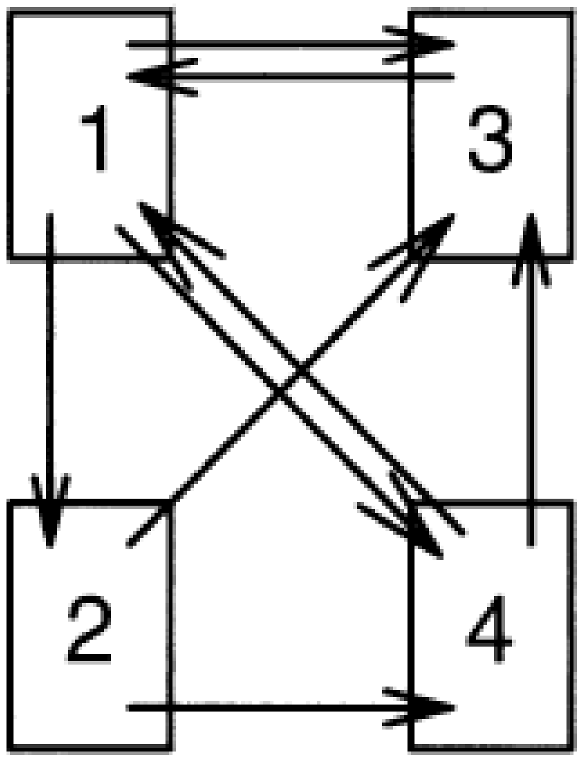


- Let x_k denote the importance score of page k.
- A simple idea is to just set importance as the number of links to a page:

$$x_k = \sum_{j=1}^{n} l_{jk}$$

- So for the example to the left,
  $x_1 = 2, x_2 = 1,\ x_3 = 3,\ x_4 = 2.$
- But a link to page k from an important page should raise k's importance score more than a link from an unimportant page.

How do we calculate the Importance Scores?



- So alternatively take the weight according to importance of linking page j, sharing its importance across all links:

$$x_j/n_j$$

where $n_j = \sum_{k=1}^{n} l_{jk}$

- In general,

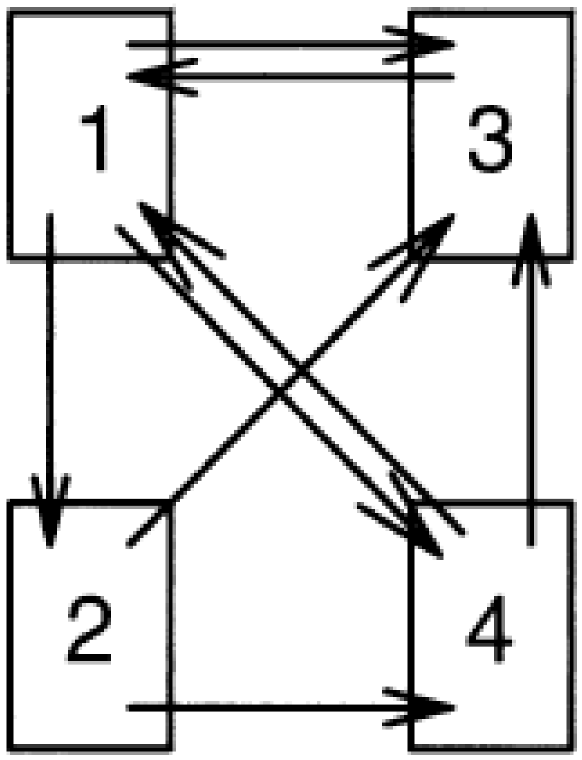
$$x_k = \sum_{j=1}^{n} \frac{l_{jk}}{n_j} x_j$$

- So for the example to the left,

$$x_1 = \tfrac{1}{1}x_3 + \tfrac{1}{2}x_4, \qquad x_2 = \tfrac{1}{3}x_1,$$

$$x_3 = \tfrac{1}{3}x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{2}x_4, \qquad x_4 = \tfrac{1}{3}x_1 + \tfrac{1}{2}x_2.$$

How do we calculate the Importance Scores?

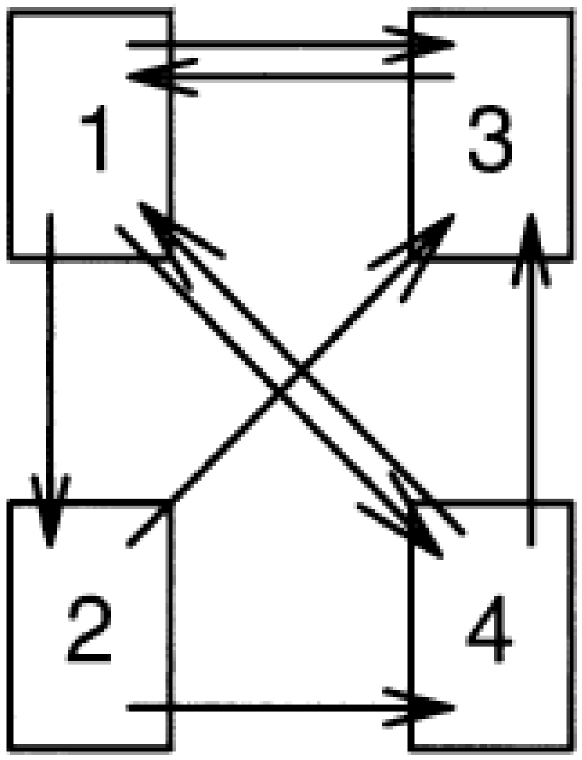


- How to solve for x_1, x_2, x_3, x_4?
- Write out the equations:

$$x_1 = \frac{l_{11}}{n_1} x_1 + \frac{l_{21}}{n_2} x_2 + \cdots + \frac{l_{n1}}{n_n} x_n$$
$$x_2 = \frac{l_{12}}{n_1} x_1 + \frac{l_{22}}{n_2} x_2 + \cdots + \frac{l_{n2}}{n_n} x_n$$
$$\cdots$$
$$x_n = \frac{l_{1n}}{n_1} x_1 + \frac{l_{2n}}{n_2} x_2 + \cdots + \frac{l_{nn}}{n_n} x_n$$

- Or using matrix notation, $\mathbf{x} = \mathbf{A}\mathbf{x}$

Importance Scores are Eigenvectors



$$\mathbf{x} = \mathbf{A}\mathbf{x}$$

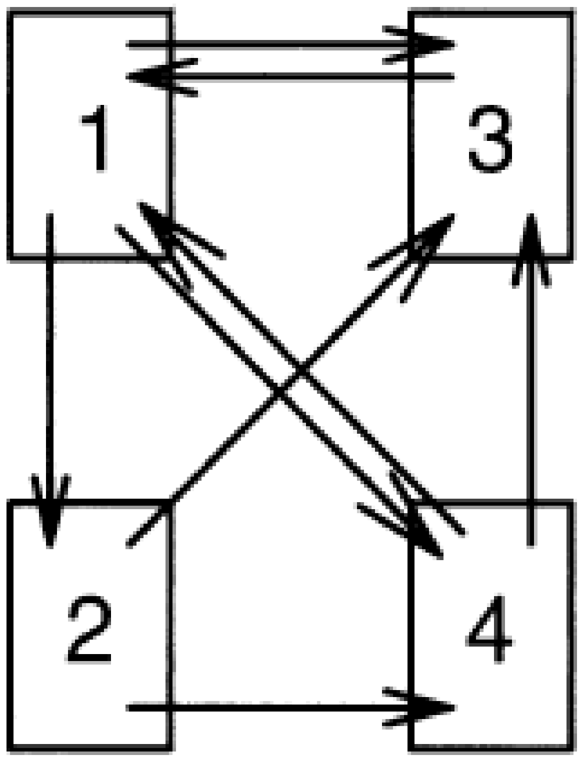
where $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$ and

$$\mathbf{A} = \begin{pmatrix} \frac{l_{11}}{n_1} & \cdots & \frac{l_{n1}}{n_n} \\ \vdots & \ddots & \vdots \\ \frac{l_{1n}}{n_1} & \cdots & \frac{l_{nn}}{n_n} \end{pmatrix}$$

This shows that $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda = 1$.

Hence find eigenvalues of $\mathbf{A}$ to compute Importance Scores.

Importance Scores are Eigenvectors



For our example,

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

Solving for $\lambda = 1$ gives eigenvectors

$$\alpha(12, 4, 9, 6\ )$$

So PageRank of

- page 1 is 12,
- page 2 is 4,
- page 3 is 9,
- page 4 is 6.