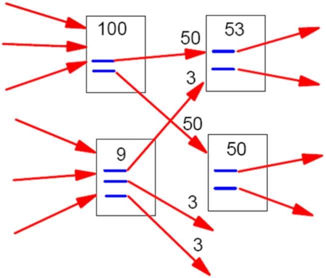
Page Rank Algorithm

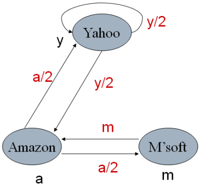
Google PageRank

- PageRank is a link analysis algorithm which assigns a numerical weighting to each Web page, with the purpose of "measuring" relative importance.
- Based on the hyperlinks map
- An excellent way to prioritize the results of web keyword searches

Simple recursive formulation

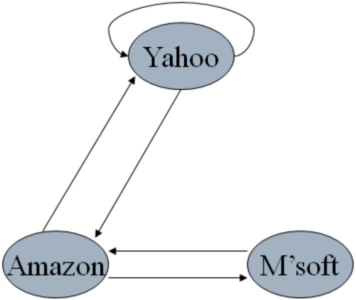


- Each link's vote is proportional to the importance of its source page
- If page P with importance x has n outlinks, each link gets $\frac{x}{n}$ votes
- Page P's own importance is the sum of the votes on its inlinks
- 3 equations, 3 unknowns, no constants – No unique solution
- Additional constraint $(y + a + m = 1)$ forces uniqueness – $y = \frac{2}{5}, a = \frac{2}{5}, m = \frac{1}{5}$
- Gaussian elimination method works for small examples, but we need a better method for large graphs



$$y = y/2 + a/2$$
$$a = y/2 + m$$
$$m = a/2$$

Matrix formulation



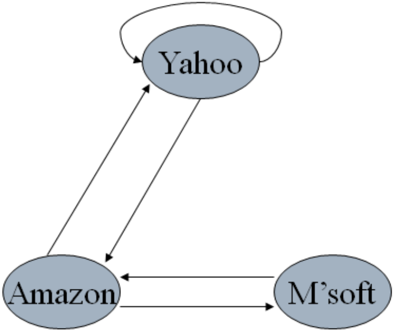
$$y = y/2 + a/2$$
$$a = y/2 + m$$
$$m = a/2$$

$$\begin{array}{c|ccc|} & y & a & m \\ \hline y & 1/2 & 1/2 & 0 \\ a & 1/2 & 0 & 1 \\ m & 0 & 1/2 & 0 \\ \hline \end{array}$$

r = Mr

$$\begin{bmatrix} y \\ a \\ m \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1 \\ 0 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} y \\ a \\ m \end{bmatrix}$$

Power Iteration Example



$$
\begin{array}{c|ccc|}
 & y & a & m \\
\hline
y & 1/2 & 1/2 & 0 \\
a & 1/2 & 0 & 1 \\
m & 0 & 1/2 & 0 \\
\hline
\end{array}
$$

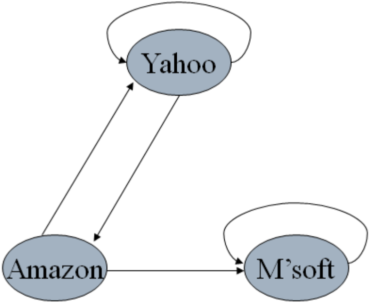
$$
\begin{array}{ccccccccc}
y & & 1/3 & 1/3 & 5/12 & 3/8 & & 2/5 \\
a & = & 1/3 & 1/2 & 1/3 & 11/24 & \ldots & 2/5 \\
m & & 1/3 & 1/6 & 1/4 & 1/6 & & 1/5
\end{array}
$$

Power Iteration Example

- Imagine a random web surfer
  - At any time t, surfer is on some page P
  - At time $t+1$, the surfer follows an outlink from P uniformly at random
  - Ends up on some page Q linked from P
  - Process repeats indefinitely
- Let $p(t)$ be a vector whose ith component is the probability that the surfer is at page i at time t
  - $p(t)$ is a probability distribution on pages
- Where is the surfer at time $t+1$?
  $p(t+1) = M \times p(t)$
- Suppose the random walk reaches a state such that
  $p(t+1) = M \times p(t) = p(t)$ – a stationary distribution for the random walk
- For graphs that satisfy certain conditions, the stationary distribution is unique and eventually will be reached no matter what the initial probability distribution at time $t = 0$.

Spider trap

- A group of pages is a spider trap if there are no links from within the group to outside the group
- Spider traps violate the conditions needed for the random walk theorem

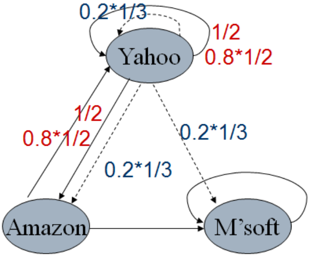


$$
\begin{array}{c|ccc|}
 & y & a & m \\
\hline
y & 1/2 & 1/2 & 0 \\
a & 1/2 & 0 & 0 \\
m & 0 & 1/2 & 1 \\
\hline
\end{array}
$$

$$
\begin{array}{ccccccccc}
y & & 1 & 1 & 3/4 & 5/8 & & 0 \\
a & = & 1 & 1/2 & 1/2 & 3/8 & \dots & 0 \\
m & & 1 & 3/2 & 7/4 & 2 & & 3
\end{array}
$$

Random teleports

- At each time step, the random surfer has two options:
  - With probability β, follow a link at random
  - With probability 1-β, jump to some page uniformly at random
  - Common values for β are in the range 0.8 to 0.9

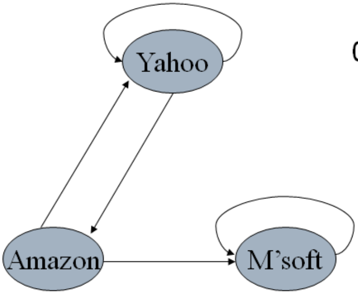


$$\begin{array}{cc} & y \\ \begin{array}{c} y \\ a \\ m \end{array} & \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix} \end{array} \quad 0.8* \begin{array}{c} y \\ \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix} \end{array} + 0.2* \begin{array}{c} y \\ \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \end{array}$$

$$0.8 \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 \\ 0 & 1/2 & 1 \end{bmatrix} + 0.2 \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$\begin{array}{cc} \begin{array}{c} y \\ a \\ m \end{array} & \begin{bmatrix} 7/15 & 7/15 & 1/15 \\ 7/15 & 1/15 & 1/15 \\ 1/15 & 7/15 & 13/15 \end{bmatrix} \end{array}$$

Random teleports ($\beta = 0.8$)



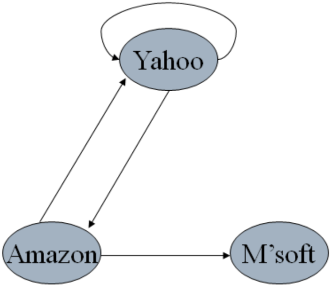
$$0.8 \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 \\ 0 & 1/2 & 1 \end{bmatrix} + 0.2 \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$\begin{array}{c} y \\ a \\ m \end{array} \begin{bmatrix} 7/15 & 7/15 & 1/15 \\ 7/15 & 1/15 & 1/15 \\ 1/15 & 7/15 & 13/15 \end{bmatrix}$$

$$\begin{array}{ccccccccc} y & & 1 & 1.00 & 0.84 & 0.776 & & 7/11 \\ a & = & 1 & 0.60 & 0.60 & 0.536 & \dots & 5/11 \\ m & & 1 & 1.40 & 1.56 & 1.688 & & 21/11 \end{array}$$

Dead ends

- Pages with no outlinks are "dead ends" for the random surfer
- Solutions:
  - Follow random teleport links with probability 1.0 from dead-ends
  - Preprocess the graph to eliminate dead-ends



$$0.8 \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \end{bmatrix} + 0.2 \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$\begin{array}{c} y \\ a \\ m \end{array} \begin{bmatrix} 7/15 & 7/15 & 1/15 \\ 7/15 & 1/15 & 1/15 \\ 1/15 & 7/15 & 1/15 \end{bmatrix}$$

Non-stochastic!

$$\begin{array}{c} y \\ a \\ m \end{array} = \begin{array}{c} 1 \\ 1 \\ 1 \end{array} \quad \begin{array}{c} 1 \\ 0.6 \\ 0.6 \end{array} \quad \begin{array}{c} 0.787 \\ 0.547 \\ 0.387 \end{array} \quad \begin{array}{c} 0.648 \\ 0.430 \\ 0.333 \end{array} \quad \ldots \quad \begin{array}{c} 0 \\ 0 \\ 0 \end{array}$$