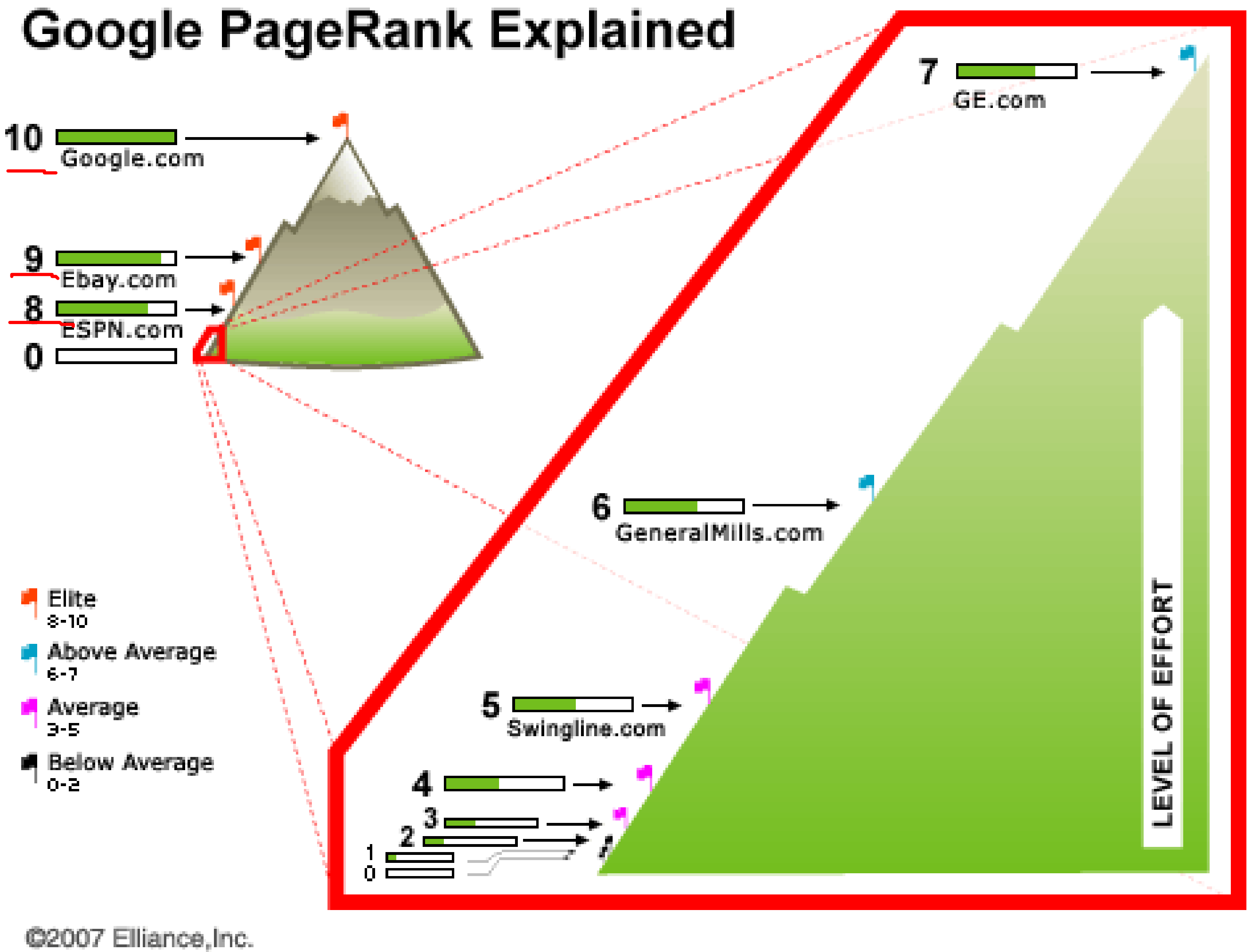
The Most Valuable Eigenvector

Hung-yi Lee

Search

http://incomebully.com/does-pr-pagerank-still-matter/



http://www.hobo-web.co.uk/google-pr-update/

PageRank

- Information of 2008

Rank：8
Google台灣首頁：www.google.com.tw
Youtube台灣首頁：http://tw.youtube.com
台灣大學網站首頁：www.ntu.edu.tw

Rank：7
痞客邦首頁：www.pixnet.net
104人力銀行：www.104.com.tw
無名小站首頁：www.wretch.cc

Rank：6
博客來網站：www.books.com.tw
聯合新聞網首頁：http://udn.com
天下雜誌網站首頁：www.cw.com.tw

Rank：5
推推王網站：http://funp.com
愛情公寓網站：www.i-part.com.tw
KKman網站首頁：
www.kkman.com.tw

Rank：4
工頭堅部落格：http://worker.bluecircus.net
白木怡言部落格：www.yubou.tw
RO仙境傳說網站：http://ro.gameflier.com

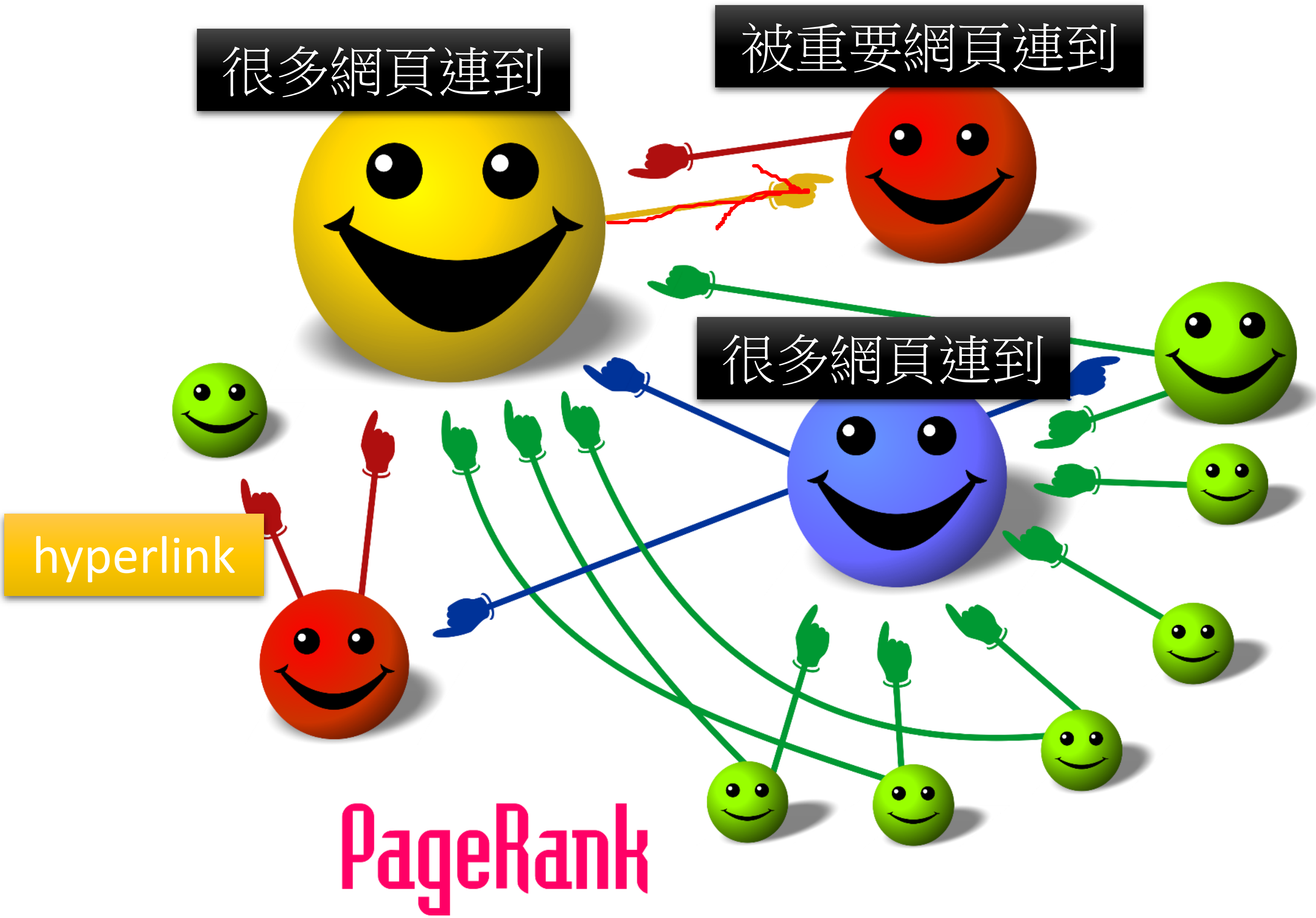
Source: http://terrylogin.blogspot.com/2008/09/pagerank.html

PageRank

PageRank

- *Webpages with a higher PageRank are more likely to appear at the top of Google search results.*
- *PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value.*
- ***Google interprets a link from page A to page B as a vote, by page A, for page B.***

Importance



PageRank

The Anatomy of a Large-Scale Hypertextual Web Search Engine

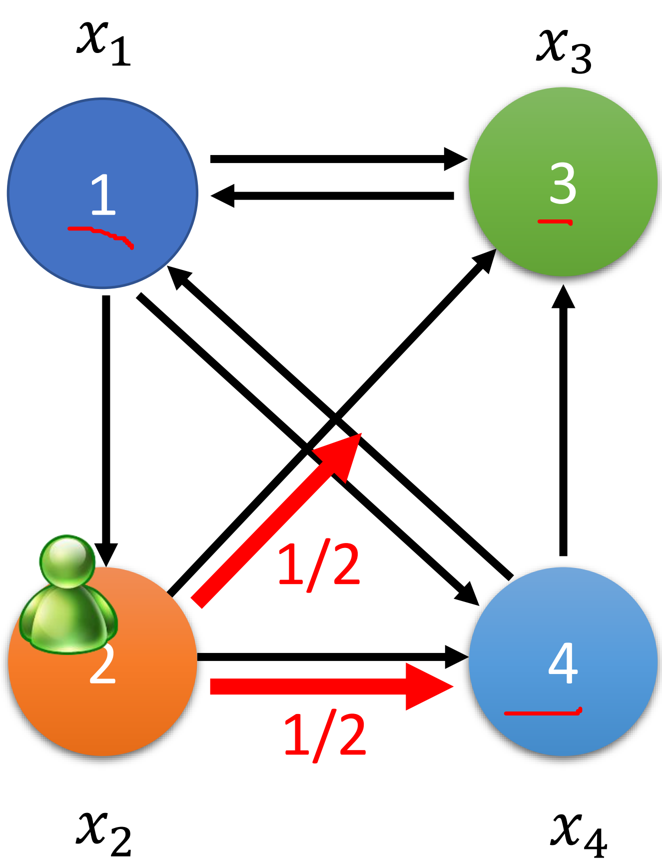
Sergey Brin and Lawrence Page

Computer Science Department,
Stanford University, Stanford, CA 94305, USA
sergey@cs.stanford.edu and page@cs.stanford.edu

Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at http://google.stanford.edu/ To engineer a search engine is a challenging task. Search engines index tens to hundreds of millions of web pages involving a comparable number of distinct terms. They answer tens of

Importance - Formulas



$$x_1 = x_3 + \boxed{\frac{1}{2}} x_4$$

$$x_2 = \frac{1}{3} x_1$$

$$x_3 = \frac{1}{3} x_1 + \frac{1}{2} x_2 + \frac{1}{2} x_4$$

$$x_4 = \frac{1}{3} x_1 + \frac{1}{2} x_2$$

Consider a random surfer

Importance - Formulas

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$Ax = x$$

The solution x is in the eigenspace of eigenvalue $\lambda = 1$

$$x_1 = x_3 + \frac{1}{2} x_4$$

$$x_2 = \frac{1}{3} x_1$$

$$x_3 = \frac{1}{3} x_1 + \frac{1}{2} x_2 + \frac{1}{2} x_4$$

$$x_4 = \frac{1}{3} x_1 + \frac{1}{2} x_2$$

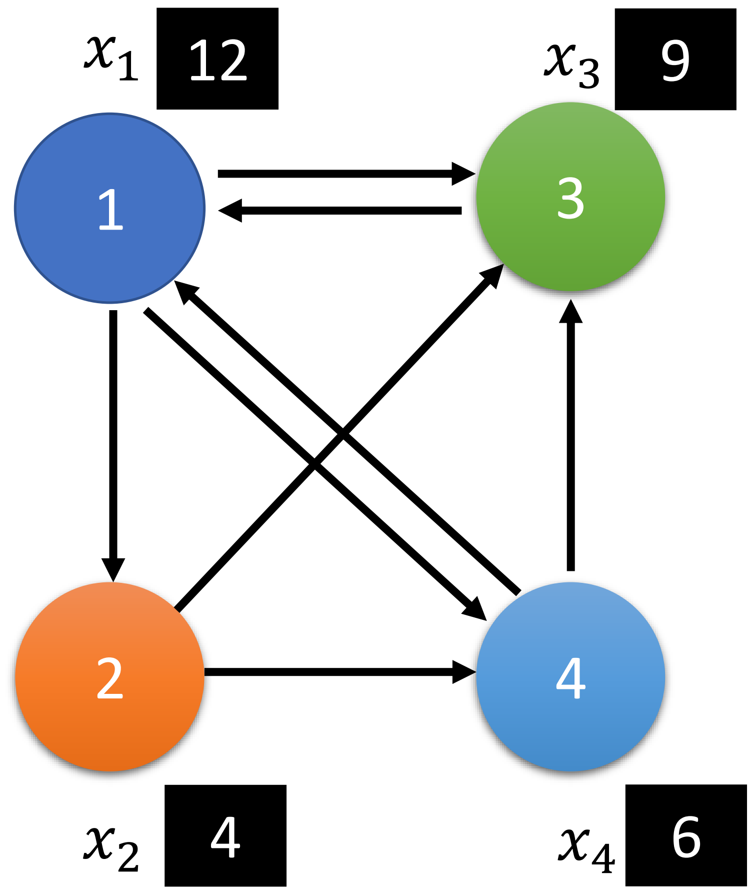
Importance - Formulas

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

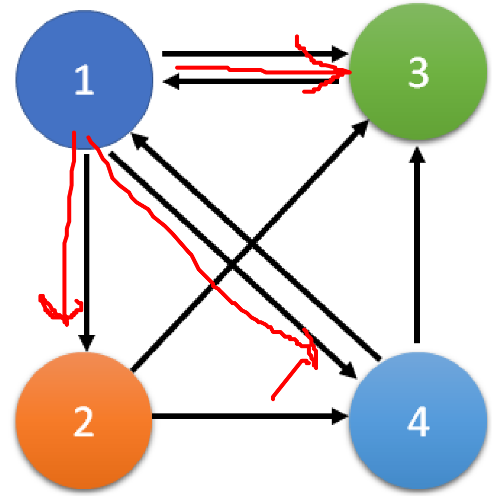
$$Ax = x$$

The solution x is in the eigenspace of eigenvalue $\lambda = 1$

$$Span\{[12 \quad 4 \quad 9 \quad 6]^T\}$$



Eigenvalue = 1



$$Ax = x$$

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \boxed{\frac{1}{3}} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

Column-stochastic Matrix

只要每个column的和都是1

Column-stochastic matrix always have eigenvalue $\lambda = 1$ **Proof**

How about the Dangling nodes (只入不出)?

Unique Ranking?

kb_1

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}$$

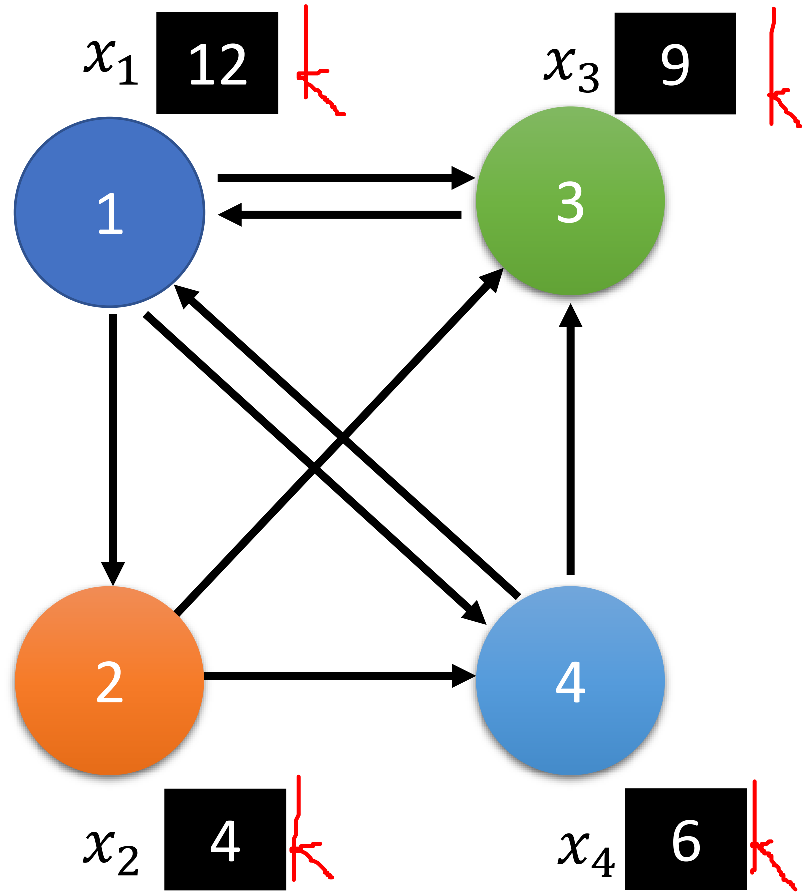
Having eigenvalue $\lambda = 1$

The dimension of the subspace is 1

Unique Ranking

constraint

Unique Score



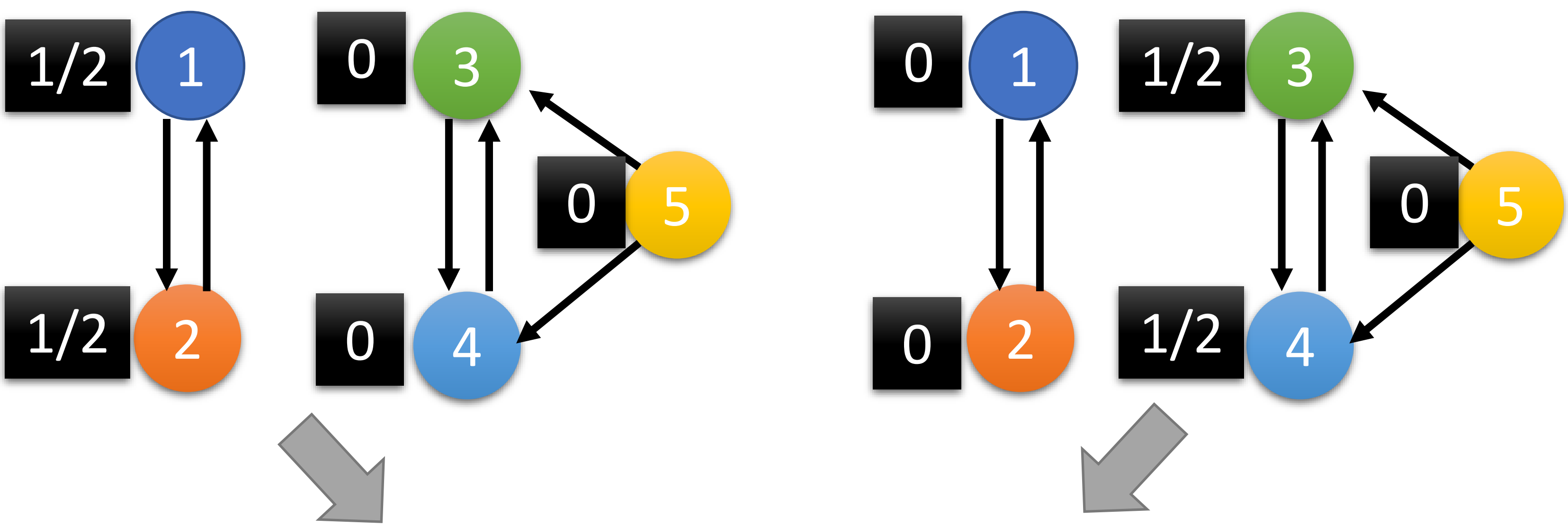
Unique Ranking?

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

How about dimension > 1

Dim for $\lambda = 1$ is 2

Basis:



Any linear combination is in the eigenspace

Not Unique Ranking

Unique Ranking?

Can it be non-uniform?

All entries are 1/n

$$\mathbf{M} = (1 - \overset{0.15}{m})\mathbf{A} + m\mathbf{S}$$

$Mx = x$

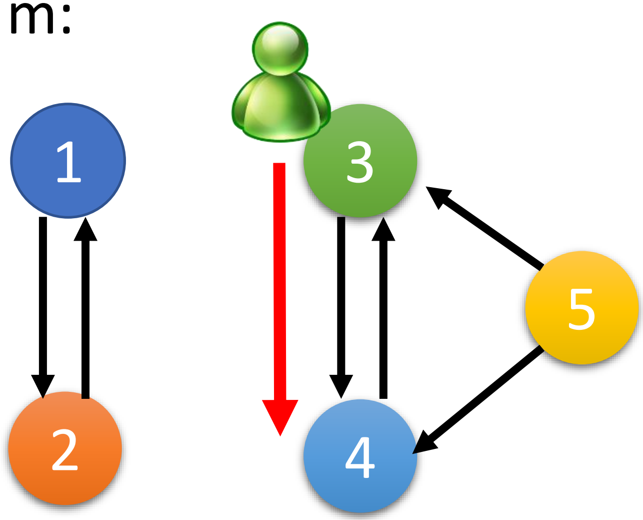
Follow the link

random

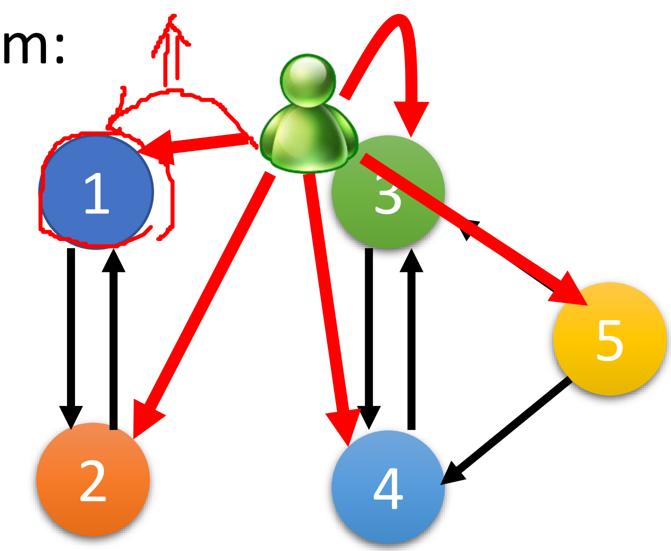
$\frac{1}{3}$ $\frac{1}{3}$ $\frac{1}{3}$

There are two ways to surf the web

Prob 1 – m:



Prob m:



Unique Ranking?

$$\mathbf{M} = (1 - m)\mathbf{A} + m\mathbf{S}$$

- Unique ranking
- For M, the dim of the eigenvalue $\lambda = 1$ is 1

M is Column-stochastic matrix and "positive"

Proof → Dim = 1

Hint: For M, the eigenvectors for eigenvalue $\lambda = 1$ are all "positive" or "negative"

Power method

$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix}$ $x_1 + x_2 + \cdots$

M is very large

Find x^*, such that $x^* = Mx^*$, $\|x^*\|_1 = 1$

Start from x_0, $\|x_0\|_1 = 1$

$$x_1 = Mx_0$$

$$x_2 = Mx_1$$

$$\vdots$$

$$x_k = Mx_{k-1}$$

If $k \to \infty$

$$x_k = x^*$$

Proof

Actually ……

- **The Last Toolbar Pagerank Update was December 2013**
- Google declared thereafter: “P*ageRank is something that we haven’t updated for over a year now, and we’re probably not going to be updating it again going forward, at least the Toolbar version.*“

Reference

- THE $25,000,000,000 EIGENVECTOR: THE LINEAR ALGEBRA BEHIND GOOGLE
  - http://userpages.umbc.edu/~kogan/teaching/m430/GooglePageRank.pdf
- A SURVEY OF EIGENVECTOR METHODS FOR WEB INFORMATION RETRIEVAL
  - http://doradca.oeiizk.waw.pl/survey.pdf