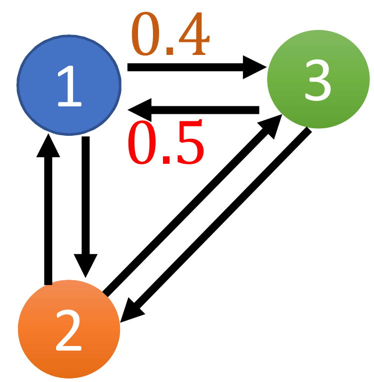
Google PageRank and Eigenvector

(除了標註※之簡報外，其餘採用李宏毅教授之投影片 教材)

Stochastic Matrix



- Consider 3 YouBike stations 1, 2, 3 on campus. Let $p_{i,j}$ be the probability that a bike rented at location $\boldsymbol{j}$ will be returned to location $\boldsymbol{i}$.

$$A = \begin{bmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{bmatrix}$$

Sum = 1

$p_{1,3}$: with prob. 0.5, bike at loc. 3 will be returned to Loc. 1

Day 0 , bike distribution $v_0 = \begin{bmatrix} 10 \\ 5 \\ 8 \end{bmatrix}$

Time 1	Time 2	Time 3	Time 4	Time 5	...
Av_0	A^2v_0	A^3v_0	A^4v_0	A^5v_0	...

Steady state

$Av = v$?

Stochastic Matrix

$$A = \begin{bmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{bmatrix}$$

- A square matrix $\mathbf{A}$ is stochastic of all of its entries are nonnegative, and the sum of each column is 1.
- $\mathbf{A}$ is positive of all of its entries are positive.
- FACT: If A is positive, 1 is an eigenvalue.

$$\underbrace{\begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.4 & 0.4 & 0.2 \\ 0.5 & 0.3 & 0.2 \end{bmatrix}}_{A^T} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

A^T and A have the same eigenvalues

Stochastic Matrix

$$A = \begin{bmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{bmatrix}$$

- FACT: If λ is an eigenvalue, $|\lambda| \leq 1$.

$$\begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.4 & 0.4 & 0.2 \\ 0.5 & 0.3 & 0.2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Suppose $|x_2|$ is the max. among $|x_1|$, $|x_2|$, $|x_3|$

$|\lambda||x_2| = |\lambda x_2| = |\, 0.4\, x_1 + 0.4\, x_2 + 0.2\, x_3| \leq$
$0.4\, |x_1| + 0.4\, |x_2| + 0.2\, |x_3| \leq$
$0.4\, |x_2| + 0.4\, |x_2| + 0.2\, |x_2| \leq |x_2|$

Hence, $|\lambda| \leq 1$

Diagonalizable Stochastic Matrix

$$A = \begin{bmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{bmatrix}$$

Eigenvectors $w_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $w_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

Eigenvalues 1 and $1/2$

$$A(a_1 w_1 + a_2 w_2) = a_1 w_1 + 1/2\, a_2 w_2$$

$$A^2(a_1 w_1 + a_2 w_2) = a_1 w_1 + 1/4\, a_2 w_2$$

$$A^3(a_1 w_1 + a_2 w_2) = a_1 w_1 + 1/8\, a_2 w_2$$

$$\ldots \ldots$$

$$A^n(a_1 w_1 + a_2 w_2) = a_1 w_1 + 1/2^n\, a_2 w_2$$

When n is large, $A^n x$ approaches $a_1 w_1$, which is an eigenvector with eigenvalue 1.

※

Perron–Frobenius Theorem

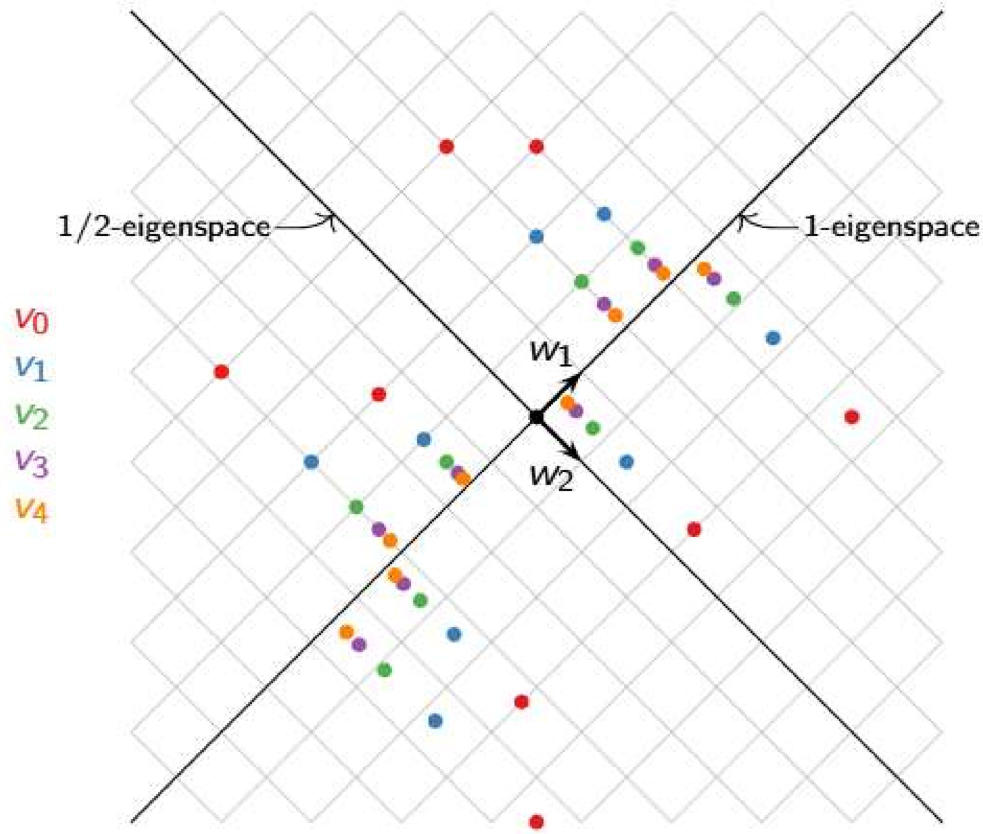
Let $\mathbf{A}$ be a positive stochastic matrix. Then $\mathbf{A}$ admits a unique normalized steady state vector w, which spans the 1-eigenspace.

Moreover, for any vector v^0 with entries summing to some number c, the iterates

$$v^1 = Av^0 \quad v^2 = Av^1 \qquad \ldots\ldots \qquad v^k = Av^{k-1} \quad \ldots\ldots$$

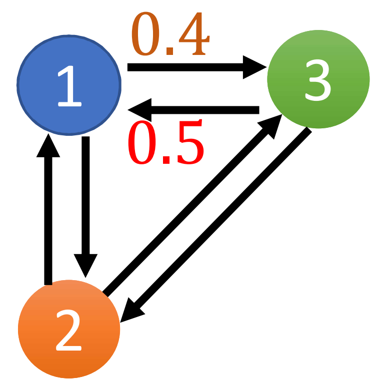
approach cw as k gets large.

Whether the matrix is diagonalizable or not, all vectors are "sucked into the 1-eigenspace," which is a line.



https://services.math.duke.edu/~jdr/1819f-1553/materials/11-12-slides-blank.pdf

Stochastic Matrix



$$A = \begin{bmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{bmatrix}$$

Steady state

Eigenvector $w = \left(\frac{1}{18}\right)\begin{bmatrix} 7 \\ 6 \\ 5 \end{bmatrix}$

Approx. $\begin{bmatrix} 0.39 \\ 0.33 \\ 0.28 \end{bmatrix}$

It says that eventually 39% bikes will be in location 1, 33% will be in location 2, and 28% will be in location 3.

If we start with 100 bikes, eventually we would have (39, 33, 28) bikes in the three locations.

The Most Valuable Eigenvector

Google

THE \$25,000,000,000* EIGENVECTOR
THE LINEAR ALGEBRA BEHIND GOOGLE

KURT BRYAN† AND TANYA LEISE‡

http://userpages.umbc.edu/~kogan/teaching/m430/GooglePageRank.pdf

PageRank

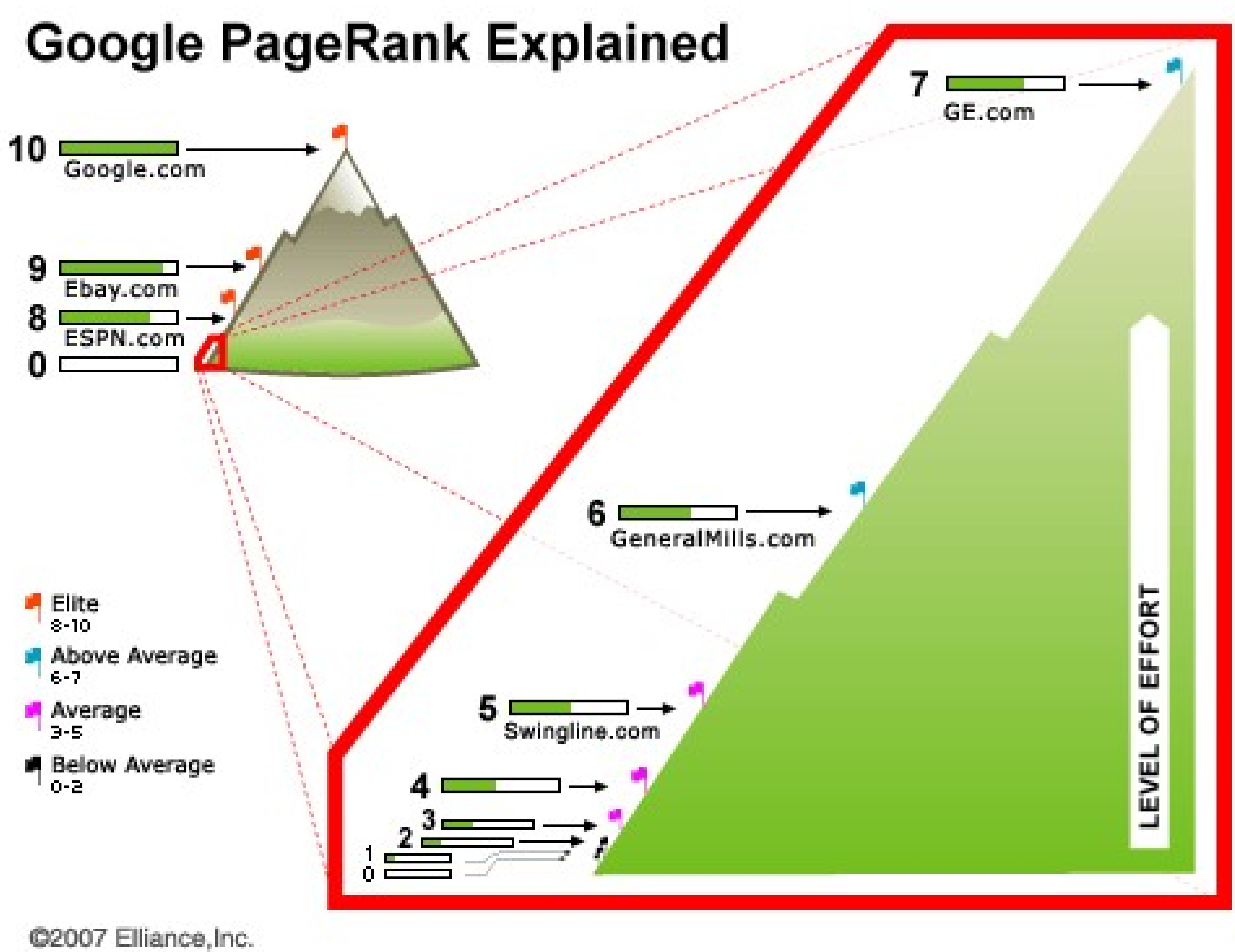
The Anatomy of a Large-Scale Hypertextual Web Search Engine

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

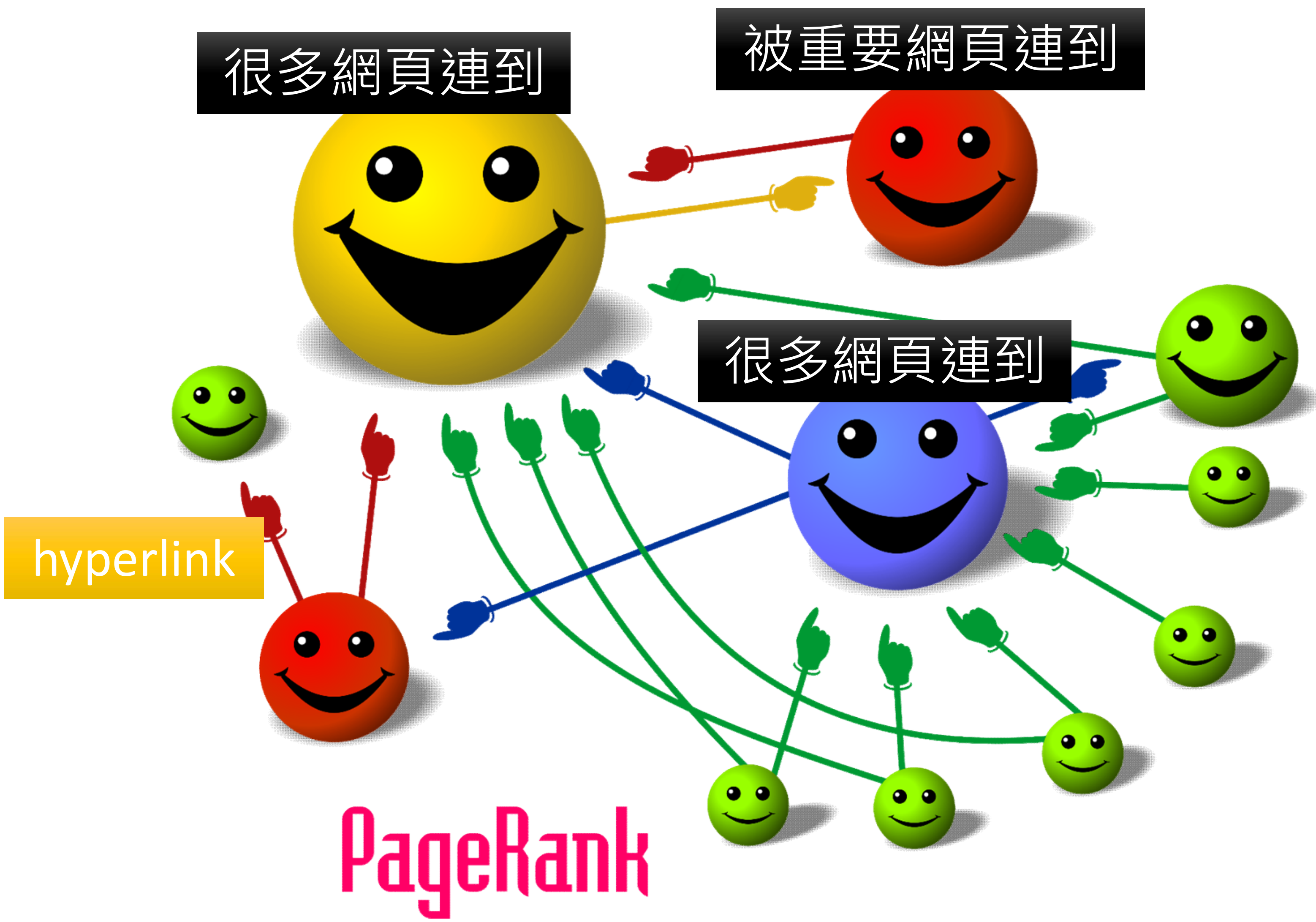


http://www.hobo-web.co.uk/google-pr-update/

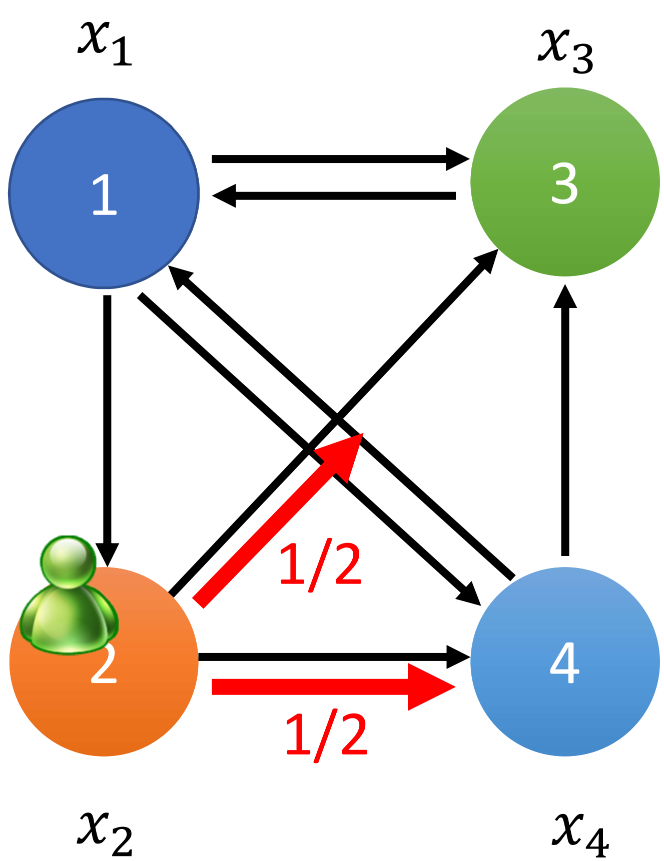
PageRank

- PageRank is a numeric value that represents how important a page is on the web.
- Webpages with a higher PageRank are more likely to appear at the top of Google search results.
- Google interprets a link from page A to page B as a vote, by page A, for page B.
- If page A is more important itself, then the vote of A to B should carry more weight.

Importance



Importance - Formulas



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

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

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

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

Consider a random surfer

Importance - Formulas

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

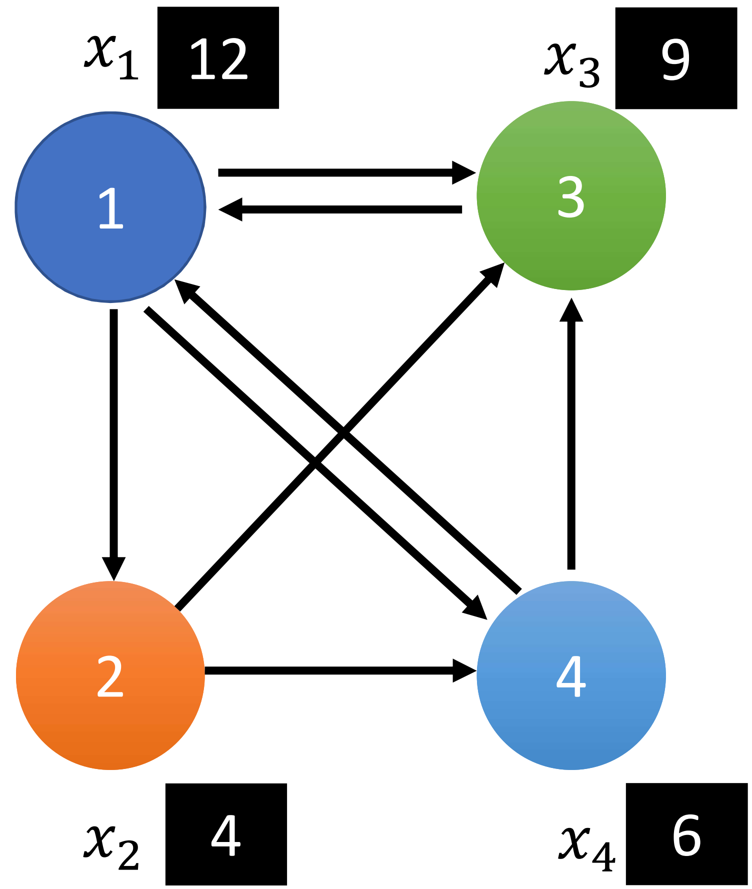
$$Ax = x \longleftarrow$$

$$\boxed{\begin{aligned} x_1 &= x_3 + \frac{1}{2}x_4 \\ x_2 &= \frac{1}{3}x_1 \\ x_3 &= \frac{1}{3}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_4 \\ x_4 &= \frac{1}{3}x_1 + \frac{1}{2}x_2 \end{aligned}}$$

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

Importance - Formulas

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



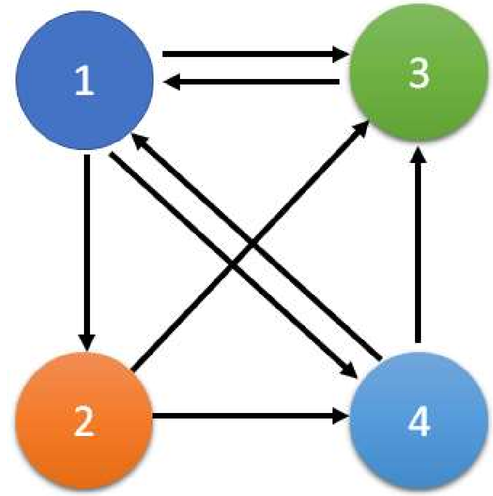
$$Ax = x$$

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

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

Eigenvalue = 1

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



$$Ax = x$$

Column-stochastic Matrix

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

How about the Dangling nodes (只入不出)?

Unique Ranking?

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

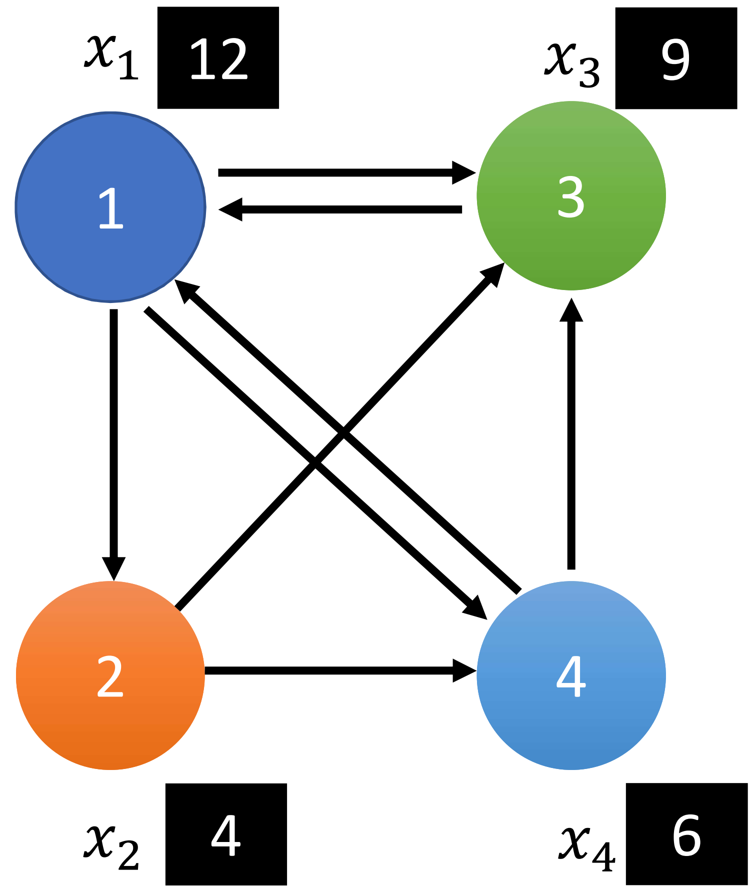
Having eigenvalue $\lambda = 1$

If the dimension of the subspace is 1

Unique Ranking

constraint

Unique Score



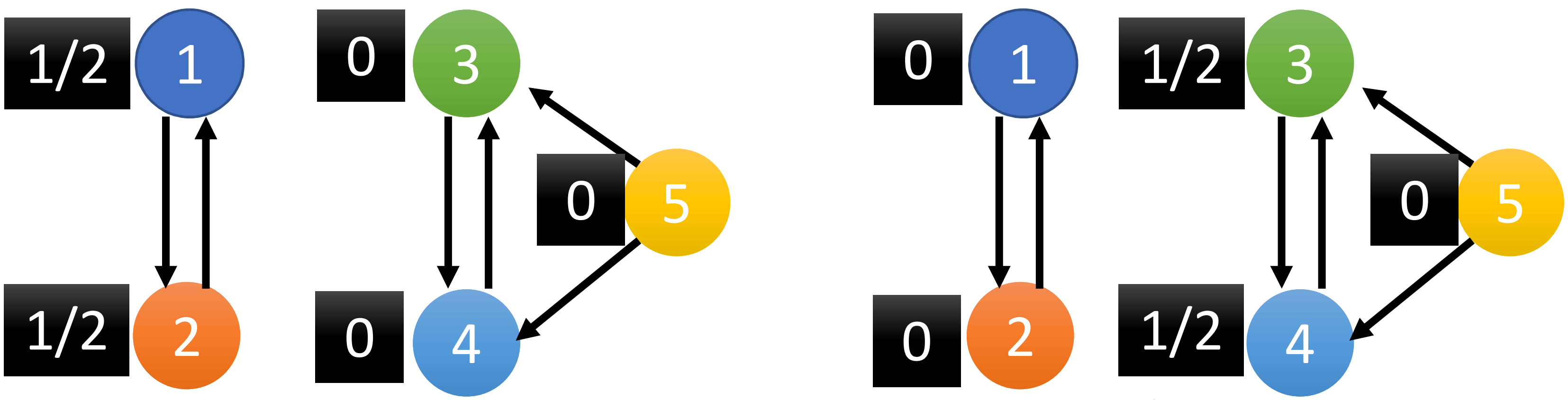
Unique Ranking?

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

How about dimension > 1

Dim for $\lambda = 1$ is 2

Basis:



Any linear combination is in the eigenspace

Not Unique Ranking

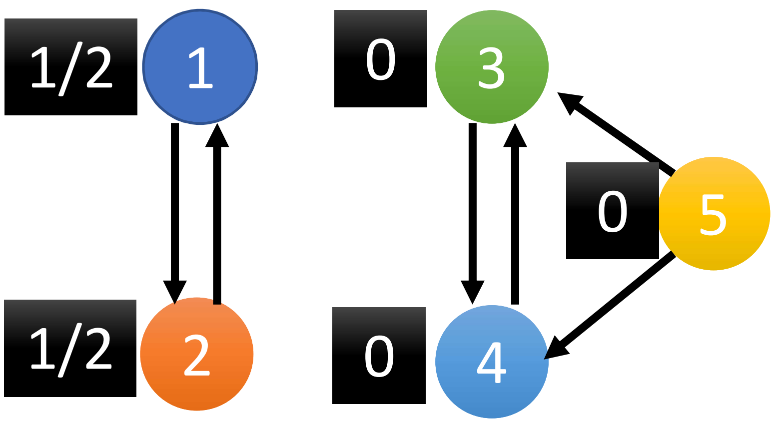
Unique Ranking?

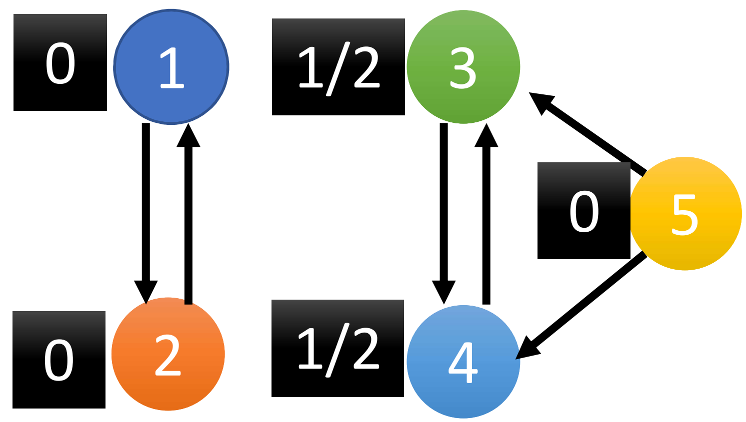
$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

How about dimension > 1

Dim for $\lambda = 1$ is 2

Basis:





Column-stochastic matrix and "positive"

$\Rightarrow$ the dim of the eigenvalue $\lambda = 1$ is 1

Unique Ranking?

Can it be non-uniform?

$$\begin{bmatrix} 1/n & \ldots & 1/n \\ \ldots & \ldots & \ldots \\ 1/n & \ldots & 1/n \end{bmatrix}$$

All entries are 1/n

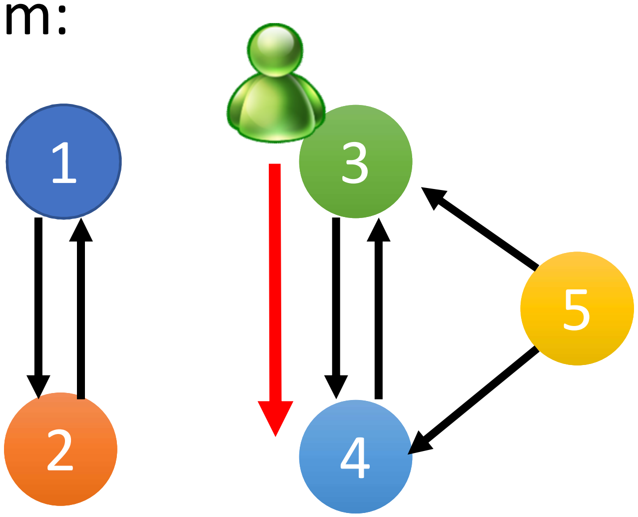
$$\mathbf{M} = (1 - \overset{0.15}{m})\mathbf{A} + m\mathbf{S}$$

Follow the link

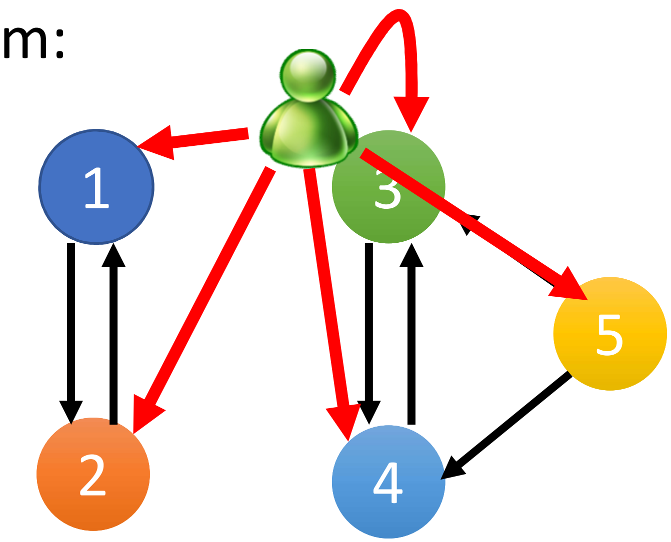
random

There are two ways to surf the web

Prob 1 – m:



Prob m:



Challenges ……

- Google‘s PageRank is an eigenvector of a matrix of order $n > 10^{12}$ (in 2008)

- The matrix A is <u>sparse</u> (tons of zeros)

- One way to compute the eigenvector *x* would be to start with a good approximate solution, such as the PageRanks from the previous month, and simply repeat the assignment .

Power method

Find x^*, such that $x^* = Mx^*$

M is very large

$$x^1 = Mx^0$$

$$x^2 = Mx^1$$

$$\vdots$$

$$x^k = Mx^{k-1}$$

Start from x^0

$$x^0 = \begin{bmatrix} 1/n \\ \vdots \\ 1/n \end{bmatrix}$$

all positive, sum to 1

If $k \to \infty$

$$x_k = x^*$$

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.*“