Google PageRank

Francesco Ricci
Faculty of Computer Science
Free University of Bozen-Bolzano
fricci@unibz.it

Content

- Linear Algebra
- Matrices
- Eigenvalues and eigenvectors
- Markov chains
- Google PageRank

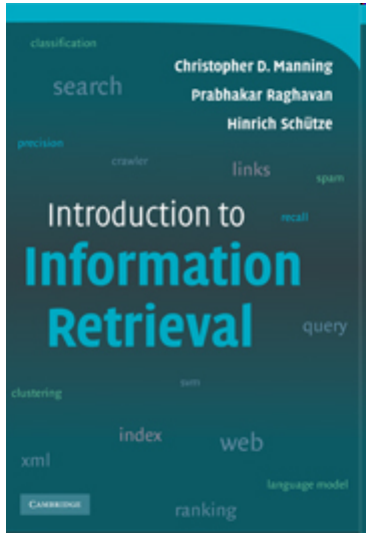
Corsi di ALGEBRA + ANALISI

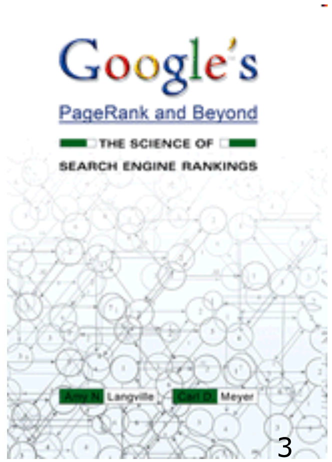
NEW!

OUR TARGET

Literature

- C. D. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008. Chapter 21
- Markov chains description on wikipedia
- Amy N. Langville & Carl D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings,* Princeton University Press, 2006.





Google

- **Google** is the leading search and online advertising company - founded by Larry Page and Sergey Brin (Ph.D. students at Stanford University)
- "googol" or 10^{100} is the mathematical term Google was named after
- Google's success in search is largely based on its **PageRank™** algorithm
- Gartner reckons that Google now make use of more than 1 million servers, spitting out search results, images, videos, emails and ads
- Google reports that it spends some 200 to 250 million US dollars a year on IT equipment.

Matrices

- A **Matrix** is a rectangular array of numbers

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

- a_{ij} is the **element** of matrix **A** in row i and column j
- **A** is said to be a n x m matrix if it has n rows and m columns
- A **square** matrix is a n x n matrix
- The **transpose** $\mathbf{A^T}$ of a matrix **A** is the matrix obtained by exchanging the rows and the columns

$$A^T = \begin{pmatrix} a_{11}^T & a_{12}^T \\ a_{21}^T & a_{22}^T \\ a_{31}^T & a_{32}^T \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

Exercise

- What is the size of these matrices

$$\begin{bmatrix} 1 & 9 & -13 \\ 20 & 5 & -6 \end{bmatrix}. \qquad \begin{bmatrix} 4 \\ 1 \\ 8 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 0 & -1 & -2 & -3 \\ 1 & 0 & -1 & -2 \\ 2 & 1 & 0 & -1 \end{bmatrix}$$

- Compute their transpose

Exercise

- What is the size of these matrices

$$\begin{bmatrix} 1 & 9 & -13 \\ 20 & 5 & -6 \end{bmatrix}.$$

2x3

$$\begin{bmatrix} 4 \\ 1 \\ 8 \end{bmatrix}$$

3x1

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

3X4

- Compute their transpose

$$\begin{bmatrix} 1 & 20 \\ 9 & 5 \\ -13 & -6 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 1 & 8 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \\ -3 & -2 & -1 \end{bmatrix}$$

Matrices

- A square matrix is **diagonal** iff has $a_{ij} = 0$ for all $i \neq j$

$$A = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}$$

- The **Identity** matrix **1** is the diagonal matrix with 1´s along the diagonal

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

- A symmetric matrix **A** satisfy the condition $\mathbf{A}=\mathbf{A}^T$

Exercise

- Is a diagonal matrix symmetric?

- Make an example of a symmetric matrix

- Make an example of a 2x3 symmetric matrix

Exercise

- Is a diagonal matrix symmetric?
  - YES because if it is diagonal then $a_{ij} = 0$ for all $i \neq j$, hence $a_{ij} = a_{ji}$ for all $i \neq j$

- Make an example of a symmetric matrix

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

- Make an example of a 2x3 symmetric matrix
  - Impossible, a symmetric matrix is a square matrix

Vectors

- A vector $\mathbf{v}$ is a one-dimensional array of numbers (is an n x 1 matrix – column vector)
- Example:

$$\mathbf{v} = \begin{pmatrix} 3 \\ 5 \\ 7 \end{pmatrix}$$

- The standard form of a vector is a **column** vector
- The transpose of a column vector $\mathbf{v}^T = (3\ 5\ 7)$ is a row vector.

Operation on matrices

- **Addition: A**=(a_{ij}), **B**=(b_{ij}), **C**=(c_{ij}) = **A**+**B**
  - $c_{ij} = a_{ij} + b_{ij}$
- **Scalar multiplication:** λ is a number, λ **A** = (λa_{ij})
- **Multiplication:** if **A** and **B** are compatible, i.e., the number of *columns* of **A** is equal to the number of *rows* of **B**, then
  - **C**=(c_{ij})= **AB**
  - $c_{ij} = \Sigma_k\, a_{ik}\, b_{kj}$

Examples

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} = \begin{pmatrix} 1*1+2*2+3*3 & 1*4+2*5+3*6 \\ 4*1+5*2+6*3 & 4*4+5*5+6*6 \end{pmatrix} = \begin{pmatrix} 14 & 32 \\ 32 & 77 \end{pmatrix}$$

It is symmetric.
Is it a general fact?
Is AA^T always symmetric?

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

- If **AB=1**, then **B** is said to be the **inverse** of **A** and is denoted with $\mathbf{A^{-1}}$
- If a matrix has an inverse is called **invertible** or **non singular**

Exercise

- Compute the following operations

$$\begin{bmatrix} 1 & 3 & 1 \\ 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 5 \\ 7 & 5 & 0 \end{bmatrix} =$$

$$2 \cdot \begin{bmatrix} 1 & 8 & -3 \\ 4 & -2 & 5 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -6 & 7 \end{bmatrix}^{\mathrm{T}} =$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} =$$

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} =$$

Exercise

- Compute the following operations

$$\begin{bmatrix}1 & 3 & 1\\ 1 & 0 & 0\end{bmatrix} + \begin{bmatrix}0 & 0 & 5\\ 7 & 5 & 0\end{bmatrix} = \begin{bmatrix}1+0 & 3+0 & 1+5\\ 1+7 & 0+5 & 0+0\end{bmatrix} = \begin{bmatrix}1 & 3 & 6\\ 8 & 5 & 0\end{bmatrix}$$

$$2 \cdot \begin{bmatrix}1 & 8 & -3\\ 4 & -2 & 5\end{bmatrix} = \begin{bmatrix}2\cdot 1 & 2\cdot 8 & 2\cdot -3\\ 2\cdot 4 & 2\cdot -2 & 2\cdot 5\end{bmatrix} = \begin{bmatrix}2 & 16 & -6\\ 8 & -4 & 10\end{bmatrix}$$

$$\begin{bmatrix}1 & 2 & 3\\ 0 & -6 & 7\end{bmatrix}^{\mathrm{T}} = \begin{bmatrix}1 & 0\\ 2 & -6\\ 3 & 7\end{bmatrix}$$

$$\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & 3\end{bmatrix},$$

$$\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix} = \begin{bmatrix}3 & 4\\ 0 & 0\end{bmatrix}.$$

Rank of a Matrix

- The row (column) **rank** of a matrix is the maximum number of rows (columns) that are linearly independent
- The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are **linearly independent** iff there is no linear combination $a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n$ (with coefficients a_i not all 0) of the vectors that is equal to 0
- *Example 1:* (1 2 3), (1 4 6), and (0 2 3) are linearly dependent: *show it*
- *Example 2:* (1 2 3) and (1 4 6) are not linearly dependent: *show it*
- The **kernel** of a matrix **A** is the subspace of vectors **v** such that $\mathbf{Av}=0$

Exercise solution

- $1*(1\ 2\ 3)^T - 1*(1\ 4\ 6)^T + 1*(0\ 2\ 3)^T = (0\ 0\ 0)^T$

$$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 4 & 2 \\ 3 & 6 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

- $(1\ {-1}\ 1)^T$ is in the kernel of the matrix:

$$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 4 & 2 \\ 3 & 6 & 3 \end{pmatrix}$$

- $a*(1\ 2\ 3) + b*(1\ 4\ 6) = (0\ 0\ 0)$
  - Then $a=-b$ and also $a = -2b$, absurd.

Rank and Determinant

- **Theorem.** A n x n square matrix is nonsingular iff has full rank (i.e. n).
- **Theorem.** A matrix has full column rank iff it does not have a null vector
- **Theorem.** A n x n matrix **A** is singular iff the det(**A**)=0

$$\det(A) = \begin{cases} a_{11} & \text{if n} = 1 \\ \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{[1j]}) & \text{if } n > 1 \end{cases}$$

- $A_{[ij]}$ is the ij minor, i.e., the matrix obtained by deleting the i-th row and the j-th column from A.

Exercise

- Compute the determinant of the following matrices

$$\begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 4 & 2 \\ 3 & 6 & 3 \end{pmatrix}$$

Exercise

- Compute the determinant of the following matrices

$$\begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix} = 1*4-1*2 = 2$$

$$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 4 & 2 \\ 3 & 6 & 3 \end{pmatrix} = 1*(4*3-2*6)-(2*3-2*3)=0$$

http://www.bluebit.gr/matrix-calculator/

Eigenvectors and Eigenvalues

- **Definition**. If **M** is a square matrix, **v** is a nonzero vector and λ is a number such that
  - **M v** = λ **v**
- then **v** is said to be an (right) *eigenvector* of **A** with *eigenvalue* λ
- If **v** is an eigenvector of **M** with eigenvalue λ, then so is any nonzero multiple of **v**
- **Only the direction matters.**

Example

- The matrix

$$M = \begin{pmatrix} 2 & -3 \\ 1 & -2 \end{pmatrix}$$

Is it singular?

- Has two (right) eigenvectors:
  - $\mathbf{v_1} = (1\ 1)^t$ and $\mathbf{v_2} = (3\ 1)^t$

Prove that

Example

- The matrix

$$M = \begin{pmatrix} 2 & -3 \\ 1 & -2 \end{pmatrix}$$

Is it singular?

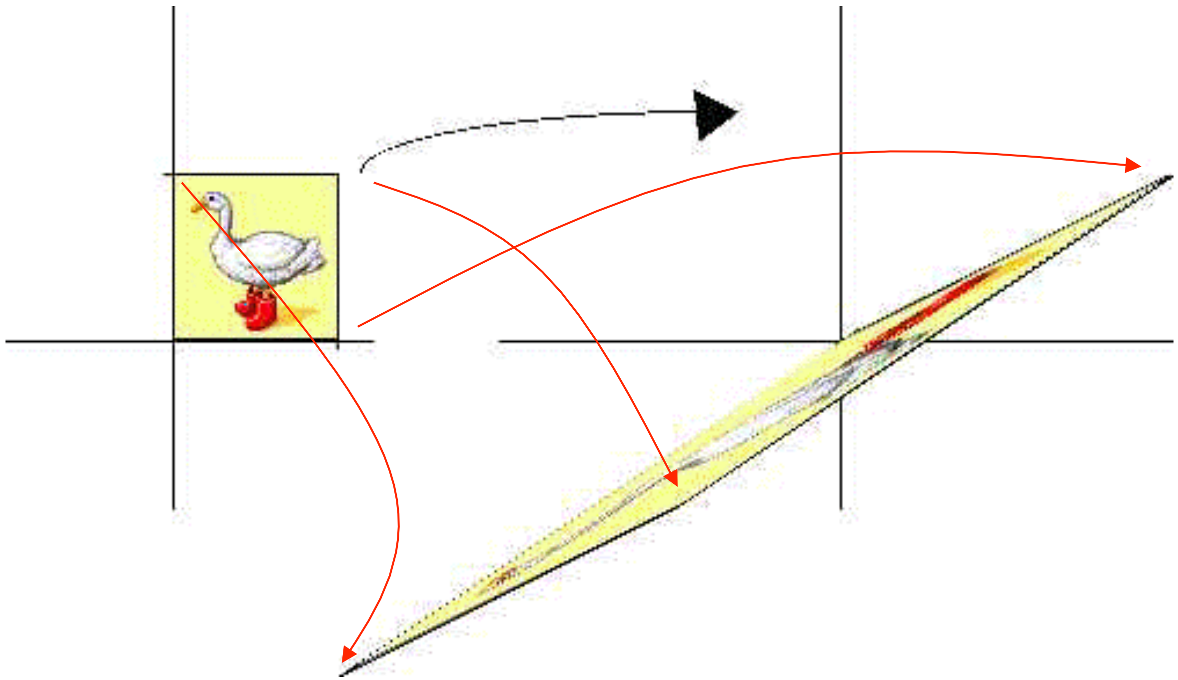
- Has two eigenvectors:
  - $\mathbf{v_1} = (1\ 1)^t$ and $\mathbf{v_2} = (3\ 1)^t$

Prove that

- $\mathbf{Mv_1} = (-1\ -1)^t = -1\ \mathbf{v_1}$
  - The eigenvalue is -1
- $\mathbf{Mv_2} = (3\ 1)^t = 1\ \mathbf{v_2}$
  - The eigenvalue is 1

Transformation

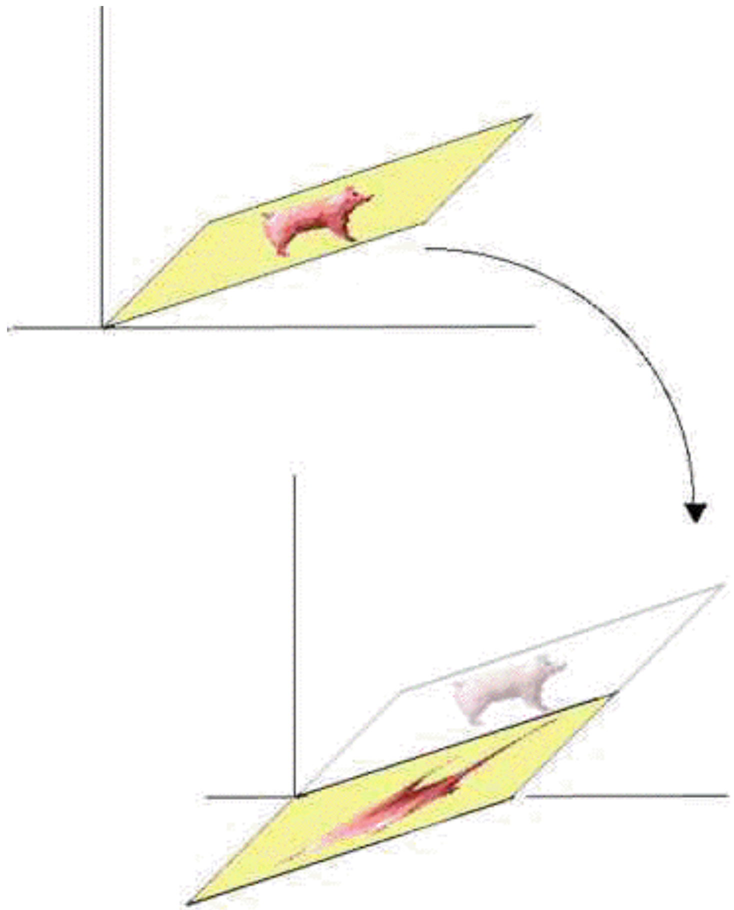
$$M = \begin{pmatrix} 2 & -3 \\ 1 & -2 \end{pmatrix}$$



- There is a lot of distortion in these directions $(1\ 0)^t$, $(1\ 1)^t$, $(0\ 1)^t$

Transformation along eigenvectors

- There are two independent directions which are not twisted at all by the matrix **M**: (1 1) and (3 1)
- one of them is flipped (1 1)
- We see less distortion if our box is oriented in the two special directions.



Results

- **Theorem:** every square matrix has at least one eigenvector
- The usual situation is that an n x n matrix has n linearly independent eigenvectors
- If there are n of them, they are a useful basis for R^n.
- Unfortunately, it can happen that there are fewer than n of them.

Finding Eigenvectors NO !

- $\mathbf{M}\,\mathbf{v} = \lambda\,\mathbf{v}$
  - $\mathbf{v}$ is an eigenvector and is λ an eigenvalue
- If $\lambda = 0$, then finding eigenvectors is the same as finding nonzero vectors in the null space – iff $\det(\mathbf{M}) = 0$, i.e., the matrix is singular
- If $\lambda \mathrel{!}= 0$, then finding the eigenvectors is equivalent to finding the null space for the matrix $\mathbf{M} - \lambda\mathbf{1}$ ($\mathbf{1}$ is the identity matrix)
- The matrix $\mathbf{M} - \lambda\mathbf{1}$ has a non zero vector in the null space iff $\det(\mathbf{M} - \lambda\mathbf{1}) = 0$
- $\det(\mathbf{M} - \lambda\mathbf{1}) = 0$ is called the **characteristic equation**.

Exercise

$$M = \begin{pmatrix} 2 & -3 \\ 1 & -2 \end{pmatrix}$$

Find the eigenvalues and the eigenvectors of this matrix

1) Find the solutions λ of the characteristic equation (eigenvalues)

2) Find the eigenvectors corresponding to the found eigenvalues.

Exercise Solution

$$M = \begin{pmatrix} 2 & -3 \\ 1 & -2 \end{pmatrix}$$

Find the eigenvalues and the eigenvectors of this matrix

- det(**M** – λ**1**) = 0
  - (2 - λ)(-2 - λ) + 3 = $\lambda^2 - 1$
- The solutions are +1 and -1

Exercise Solution

$$M = \begin{pmatrix} 2 & -3 \\ 1 & -2 \end{pmatrix}$$

Find the eigenvalues and the eigenvectors of this matrix

- det(**M** – λ**1**) = 0
  - $(2 - \lambda)(-2 - \lambda) + 3 = \lambda^2 - 1$
- The solutions are +1 and -1
- Now we have to solve the set of linear equations
  - **Mv**=**v** (for the first eigenvalue)

Exercise Solution

$$M = \begin{pmatrix} 2 & -3 \\ 1 & -2 \end{pmatrix}$$

Find the eigenvalues and the eigenvectors of this matrix

- det(**M** – λ**1**) = 0
  - (2 - λ)(-2 - λ) + 3 = $\lambda^2 - 1$
- The solutions are +1 and -1
- Now we have to solve the set of linear equations
  - **Mv**=**v** (for the first eigenvalue)

$$2x - 3y = x$$

$$x - 2y = y$$

  - Has solution x=3y, $(3\ 1)^t$ – and all vectors obtained multiplying this with a scalar.

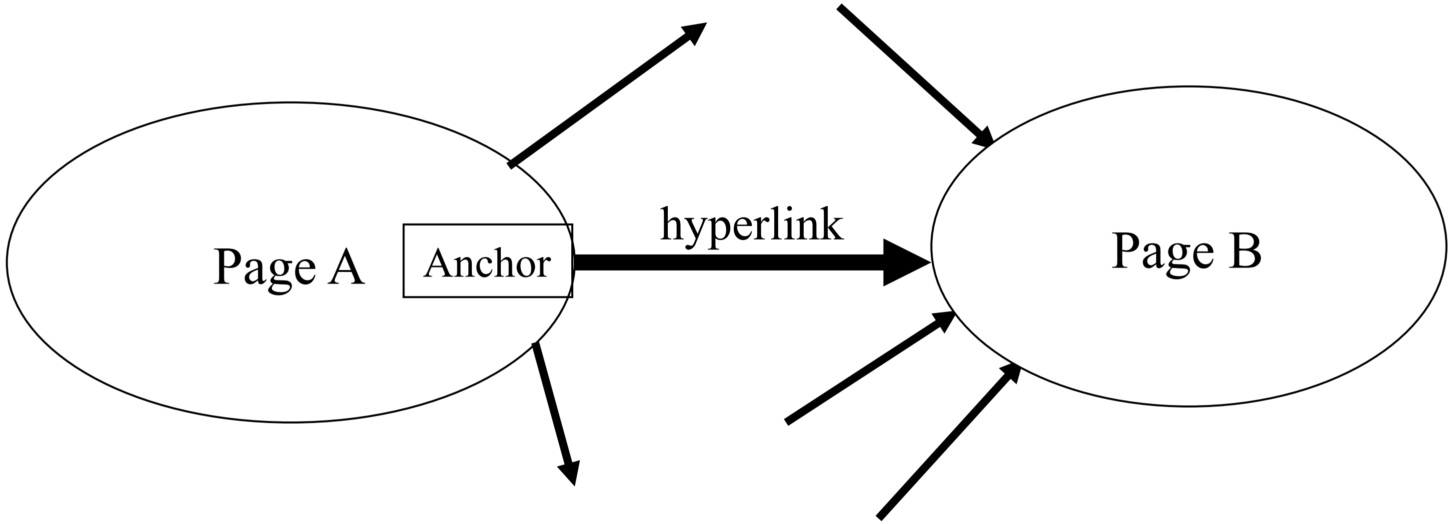
Algorithm

- *To find the eigenvalues and eigenvectors of* **M**:
  - *First find the eigenvalues by solving the characteristic equation*.
    Call the solutions $\lambda_1, \ldots, \lambda_n$. (There is always at least one eigenvalue, and there are at most n of them.)
  - For all λ_k, the existence of a nonzero vector in this null space is guaranteed. Any such vector is an eigenvector.

Graphs

- A **directed graphs G** is a pair (V,E), where V is a finite set and E is a binary relations on V
  - V is the **Vertex set** of G: contains the **vertices**
  - E is the **Edge set** of G: contains the **edges**
- In an undirected graphs G=(V,E) the edges consists of unordered pairs of vertices
- The **in-degree** of a vertex v (directed graph) is the number of edges entering in v
- The **out-degree** of a vertex v (directed graph) is the number of edges leaving v.

The Web as a Directed Graph



Assumption 1: A hyperlink between pages denotes author perceived relevance (quality signal)

Assumption 2: The anchor of the hyperlink describes the target page (textual context)

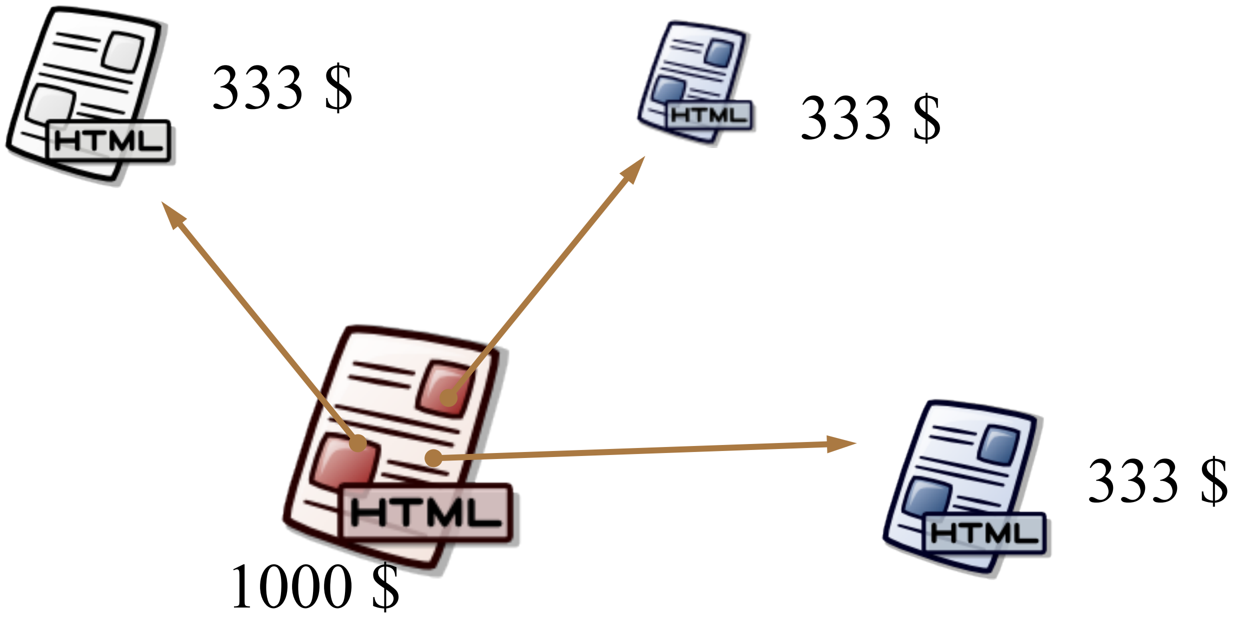
Ranking web pages

- To count *inlinks*: enter in google search form link:*www.mydomain.com*
- Web pages are not equally "important"
  - www.unibz.it vs. www.stanford.edu
  - Inlinks as votes
    - www.stanford.edu has 3200 inlinks
    - www.unibz.it has 352 inlink (Feb 2013)
- Are all *inlinks* equal?
  - Recursive question!

Simple recursive formulation

- Each link's vote is **proportional** to the the importance of its **source** page
- If page **P** with importance **x** has **n** outlinks, each link gets **x/n** votes

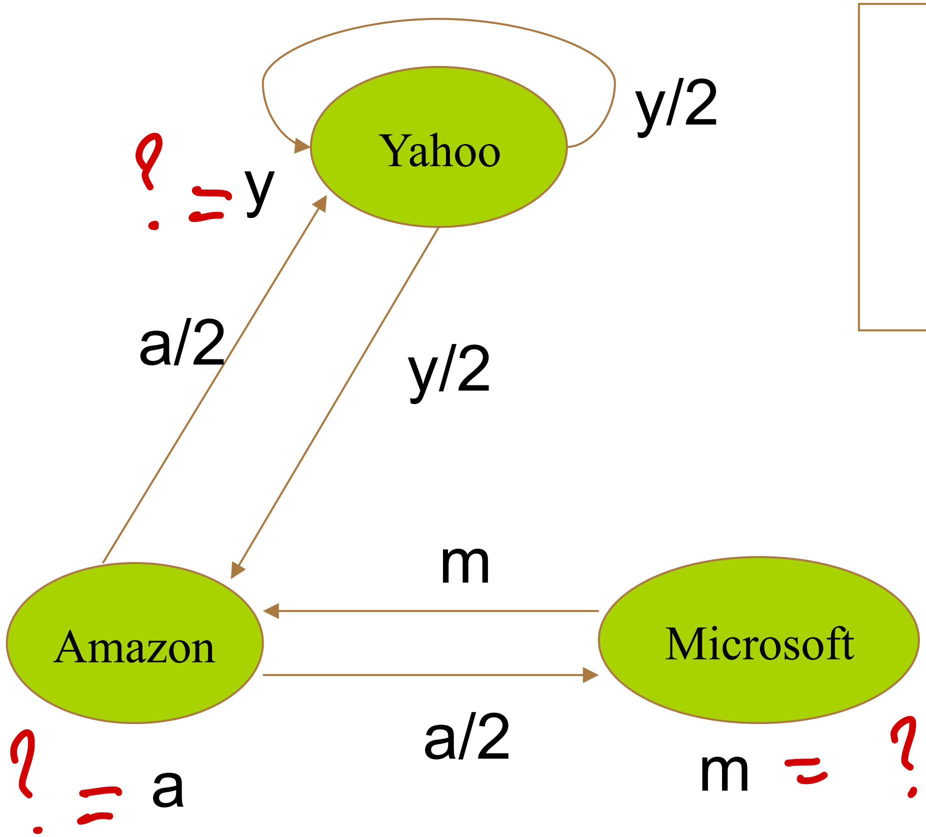
Roles



Simple "flow" model

The web in 1839

incognite = $\begin{pmatrix} y \\ a \\ m \end{pmatrix}$



$$
\begin{aligned}
y &= y/2 + a/2 \\
a &= y/2 + m \\
m &= a/2
\end{aligned}
$$

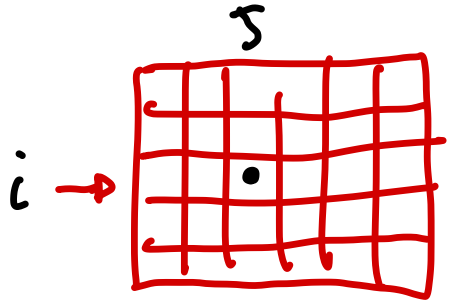
WRITE AS A LINEAR APPLICATIONS MATRIX

a, **m**, and **y** are the importance of these pages

Solving the flow equations

- 3 equations, 3 unknowns, no constants
  - No unique solution
  - If you multiply a solution by a constant (λ) you obtain another solution - try with (2 2 1)
- Additional constraint forces uniqueness
  - y+a+m = 1 (*normalization*) ⇒ The total flow is 1
  - y = 2/5, a = 2/5, m = 1/5
  - *These are the scores of the pages under the assumption of the flow model*
- Gaussian elimination method works for small examples, but we need a better method for large graphs.

Matrix formulation



$M_{i,j} = 1/d_i$

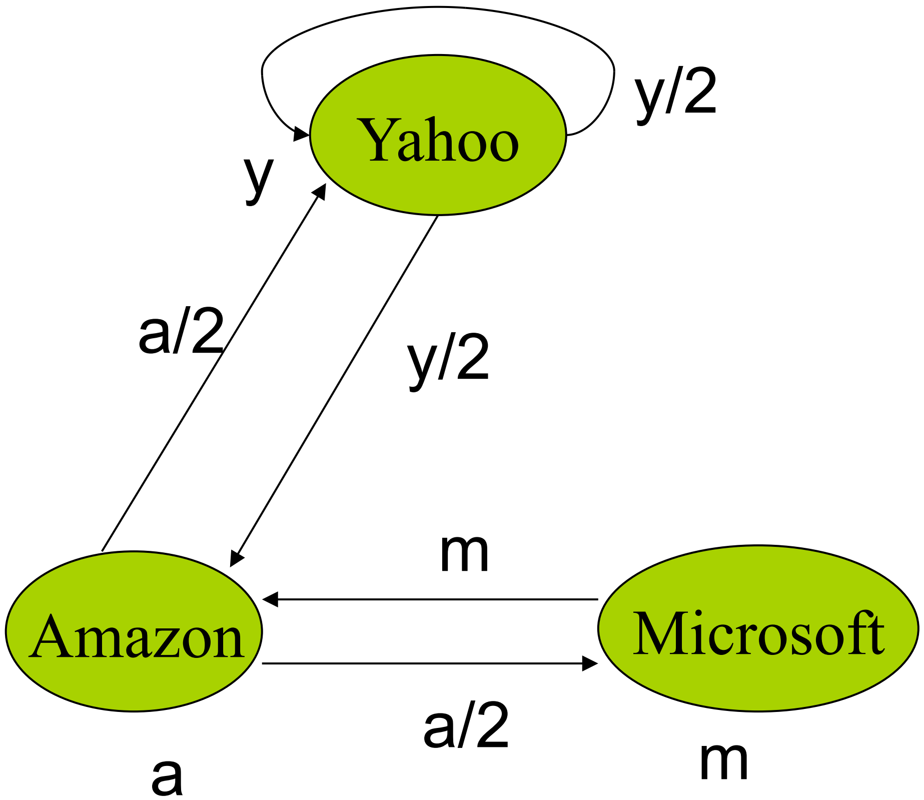
188

$i \to j$

- Matrix **M** has one row and one column for each web page (square matrix)
- Suppose page i has n outlinks
  - If *i links to j*, then $M_{ij}=1/n$
  - Else $M_{ij}=0$
- **M** is a **row stochastic matrix**
  - Rows sum to 1
- Suppose **r** is a vector with one entry per web page
  - r_i is the **importance score** of page i
  - Call it the **rank vector**

NOTA: sia $\bar{r} \in [0,1]^n$ allora $\bar{r}$ riga
$\bar{r} M = \bar{r}$ oppure $M^T \cdot \bar{r}^T = \bar{r}^T$

Example

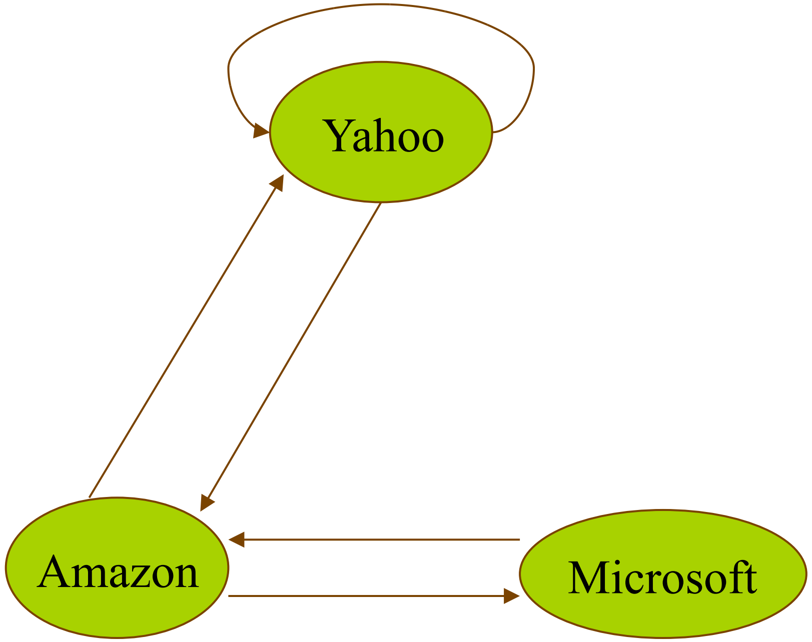


$$
\begin{array}{c|c}
 & \begin{matrix} y & a & m \end{matrix} \\
\begin{matrix} y \\ a \\ m \end{matrix} & \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}
\end{array} = M
$$

$$(y\ a\ m) = (y\ a\ m)\mathbf{M}$$

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

Power Iteration Solution



$$\begin{array}{c c} & \begin{array}{ccc} y & a & m \end{array} \\ \begin{array}{c} y \\ a \\ m \end{array} & \begin{pmatrix} ½ & ½ & 0 \\ ½ & 0 & ½ \\ 0 & 1 & 0 \end{pmatrix} \end{array} = \mathbf{M}$$

(y a m) = (y a m)**M**

(1/3 1/3 1/3) $t=0$

(1/3 1/3 1/3)**M** = (1/3 1/2 1/6) $t=1: r^{(1)} = r^{(0)} \cdot M$

(1/3 1/2 1/6)**M** = (5/12 1/3 1/4) $t=2: r^{(2)} = r^{(1)} \cdot M = r^{(0)} \cdot M^2$

(5/12 1/3 1/4)**M** = (3/8 11/24 1/6)

…

(2/5 2/5 1/5)

Example

Following 6 hours

First 6 hours

	dry	rain	
dry	61	11	72 cases
rain	11	17	28 cases
	72 cases	28 cases	

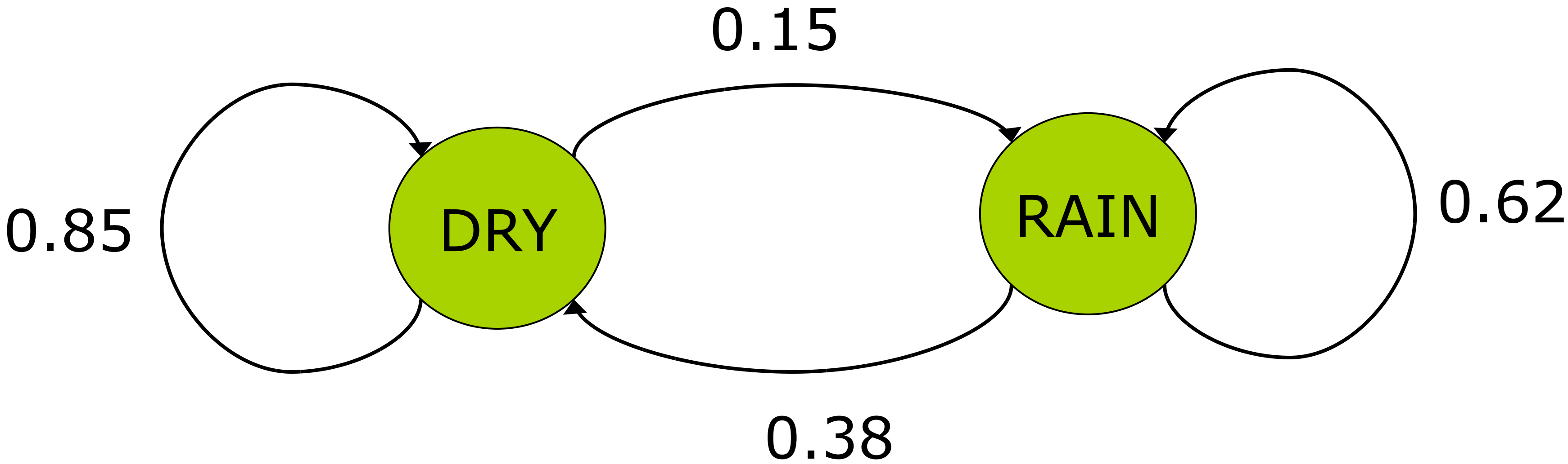
Following 6 hours

First 6 hours

	dry	rain	
dry	.85	.15	72 cases
rain	.38	.62	28 cases
	72 cases	28 cases	

States and probabilities

MARKOV
CHAIN



Composing transitions

$$
\begin{array}{c|cc}
 & \otimes & \mathbf{R} \\
\hline
\otimes & .85 & .15 \\
\mathbf{R} & .38 & .62
\end{array}
$$

What can we say about the 12 h transition? From 00-06 to 12-18?

$\text{Rain} >>> \underset{.62}{} \ \text{Rain} >>> \ \text{Rain} \qquad = .38$

Rain > > > Rain > > > Rain
.62 .62
= .38

Rain > > > dry > > > Rain
.38 .15
= .06

= .44

0.44 = 0.38*0.15+0.62*0.62
What kind of operation is on the matrix?

Composing transitions

$$\begin{array}{ccc} & \text{dry} & \text{rain} \\ \text{dry} & & \\ \text{rain} & & \end{array} \begin{pmatrix} .85 & .15 \\ .38 & .62 \end{pmatrix}\begin{pmatrix} .85 & .15 \\ .38 & .62 \end{pmatrix} = \begin{pmatrix} .78 & .22 \\ .56 & .44 \end{pmatrix} = A^2$$

- The probabilities of the 12hours transitions are given by squaring the matrix representing the probabilities of the 6hours transitions
  - P(rain-in-12hours|rain-now)= P(rain-in-12hours|rain-in-6hours)*P(rain-in-6hours|rain-now)+P(rain-in-12hours|dry-in-6hours)*P(dry-in-6hours|rain-now)=.62*.62+.15*.38=.44
  - P(dry-in-12hours|rain-now)= P(dry-in-12hours|rain-in-6hours)*P(rain-in-6hours|rain-now)+P(dry-in-12hours|dry-in-6hours)*P(dry-in-6hours|rain-now)=38*.62+.85*.38=.56

Behavior in the limit

$$A = \begin{pmatrix} .85 & .15 \\ .38 & .62 \end{pmatrix} \quad A^2 = \begin{pmatrix} .78 & .22 \\ .56 & .44 \end{pmatrix} \quad A^3 = \begin{pmatrix} .75 & .25 \\ .64 & .36 \end{pmatrix}$$

$$A^4 = \begin{pmatrix} .73 & .27 \\ .68 & .32 \end{pmatrix} \quad A^5 = \begin{pmatrix} .72 & .28 \\ .70 & .30 \end{pmatrix} \quad A^6 = \begin{pmatrix} .72 & .28 \\ .71 & .29 \end{pmatrix}$$

$$A^7 = \begin{pmatrix} .72 & .28 \\ .71 & .29 \end{pmatrix} \quad A^8 = \begin{pmatrix} .72 & .28 \\ .72 & .28 \end{pmatrix} \quad A^9 = \begin{pmatrix} .72 & .28 \\ .72 & .28 \end{pmatrix}$$

$$A^\infty = \begin{pmatrix} .72 & .28 \\ .72 & .28 \end{pmatrix}$$

Behavior in the limit

- If a,b <=1, and a+b=1, i.e., (a b) is a generic state with a certain probability a to be dry and b=1-a to be rain, then

$$\begin{pmatrix} a & b \end{pmatrix} A^{\infty} = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} .72 & .28 \\ .72 & .28 \end{pmatrix} = \begin{pmatrix} .72 & .28 \end{pmatrix}$$

- In particular (.72 .28)A=(.72 .28), i.e., it is a (left) **eigenvector** with **eigenvalue** 1
- The eigenvector (.72 .28) represents the limit situation starting from a generic situation (a b): it is called the **stationary distribution**.

Exercise

- Find one (left) eigenvector of the matrix below:
  - Solve first the characteristic equation (to find the eigenvalues)
  - and then find the left eigenvector corresponding to the largest eigenvalue

$$\begin{pmatrix} .85 & .15 \\ .38 & .62 \end{pmatrix}$$

Exercise Solution

- Characteristic equation

$$\det\begin{pmatrix} .85-\lambda & .15 \\ .38 & .62-\lambda \end{pmatrix} = (0.85-\lambda)(0.62-\lambda) - 0.15*0.38 = \lambda^2 - 1.47\lambda + 0.47$$

$$\lambda = \frac{1.47 \pm \sqrt{1.47^2 - 4*0.47}}{2}$$

Solutions $\lambda = 1$ and $\lambda = 0.47$

$$\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} .85 & .15 \\ .38 & .62 \end{pmatrix} = \begin{pmatrix} x & y \end{pmatrix}$$

0.85x + 0.38y=x

x + y = 1

0.85x +0.38(1-x)=x

-0.53x +0.38=0

x = 0.38/0.53=**0.72** y = 1 – 0.72= **0.28**

Markov Chain

n = time slots

- A Markov chain is a sequence X_1, X_2, X_3, ... of random variables ($\Sigma_{v \text{ all possible values of } X}$ P(X=v) = 1) with the property:
- **Markov property**: the conditional probability distribution of the next future state X_{n+1} given the present and past states is a function of the present state X_n alone

$$\Pr(X_{n+1} = x | X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \Pr(X_{n+1} = x | X_n = x_n).$$

- If the state space is finite then the transition probabilities can be described with a matrix $P_{ij}=P(X_{n+1}= j \mid X_n = i)$, i,j =1, …m

j   i → riga

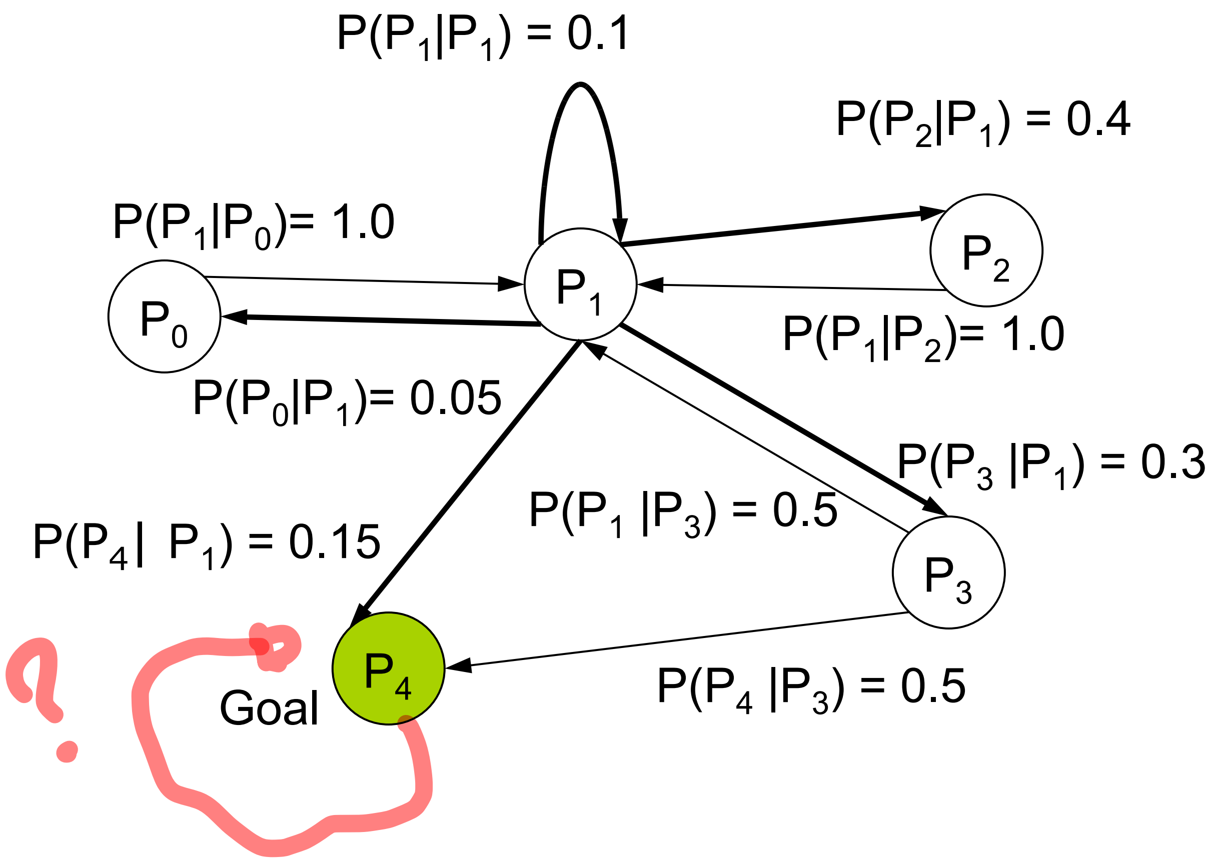
$$\begin{pmatrix} .85 & .15 \\ .38 & .62 \end{pmatrix} = \begin{pmatrix} P(X_{n+1} = 1 \mid X_n = 1) & P(X_{n+1} = 2 \mid X_n = 1) \\ P(X_{n+1} = 1 \mid X_n = 2) & P(X_{n+1} = 2 \mid X_n = 2) \end{pmatrix}$$

Example: Web

- X_t is the page visited by a user (random surfer) at time t;
- At every time t the user can be in one among m pages (states)
- We assume that when a user is on page i at time t, then the probability to be on page j at time $t+1$ depends only on the fact that the user is on page i, and **not on the pages previously visited.**

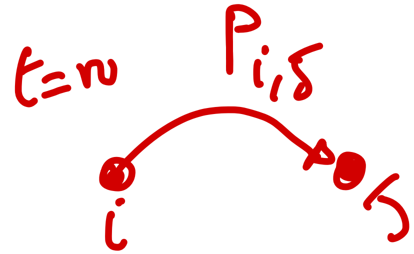
↳ MARKOV CHAIN

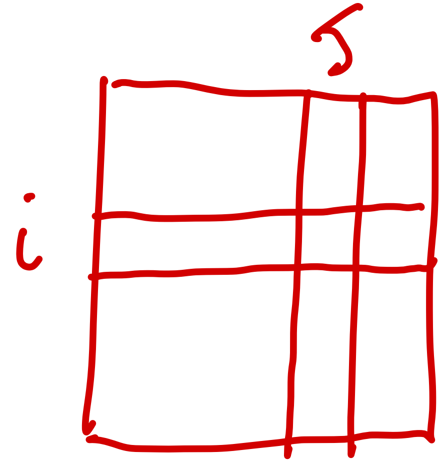
Probabilities



In this example there are 5 states and the probability to jump from a page/state to another is not constant (it is not *1/(#of outlinks of the node)*) … as we have assumed before in the simple web graph

This is not a Markov chain! (why?)





Examples

- $P_{ij}=P(X_{n+1}=j \mid X_n=i)$, i,j =1, …m
- (1, 0, 0, …, 0) P = (P_{11}, P_{12}, P_{13}, …, P_{1n})
  - if at time n it is in **state 1**, then at time n+1 it is in state j with probability P_{1j}, i.e., the first **row** of P_{ij} gives the probabilities to be in the other states
- (0.5, 0.5, 0, …, 0) P = ($P_{11}\cdot 0.5 + P_{21}\cdot 0.5$, …, $P_{1n}\cdot 0.5 + P_{2n}\cdot 0.5$)
  - this is the linear combination of the first two rows.

Stationary distribution

$(\pi_1, \pi_2, \ldots, \pi_m) \cdot \left(\cdot P \right)$

- A **stationary distribution** is a m-dimensional (sum 1) vector which satisfies the equation:

$$\pi^T \mathbf{P} = \pi^T, \qquad \lambda = 1$$

- Where π is a (column) vector and π^T (row vector) is the transpose of π
- A stationary distribution always exists, but is not guaranteed to be unique (*can you make an example of a Markov chain with more than one stationary distribution?*)
- If there is only one stationary distribution then

$$\left[\pi^T = \lim_n x^T P^n \right] \rightarrow \text{POWER METHOD!}$$

- Where x is a generic distribution over the m states (i.e., it is an m-dimensional vector whose entries are <=1 and the sum is 1)

Random Walk Interpretation

- 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 i^{th} component is the probability that the surfer is at page i at time t
  - **p**(t) is a probability distribution on pages

The stationary distribution

$X(t) \equiv$ page of SURFER at time t

- Where is the surfer at time t+1?
  - Follows a link uniformly at random $P = M(i,j) = \Pr[i \to j]$
  - **p**(t+1) = **p**(t)**M** → why? *
- Suppose the random walk reaches a state such that **p**(t+1) = **p**(t)**M** = **p**(t)
  - Then **p**(t) is a **stationary distribution** for the random walk
- Our rank vector **r= p**(t) satisfies **r = rM.**

$P(i,j)$

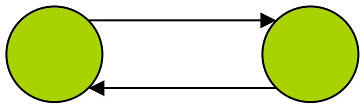
$$* \;\forall j = 1..m: \Pr[X(t+1) = j] = \sum_{i=1}^{m} \Pr[X(t) = i] \cdot \Pr[X(t+1) = j \mid X(t) = i]$$

In vect. form is: $\bar{p}(t+1) = \bar{p}(t) \cdot M$

Ergodic Markov chains

- A Markov chain is **ergodic** if:
  - Informally: *there is a path from any state to any other; and the states are not partitioned into sets such that all state transitions occur cyclically from one set to another.*
  - Formally: for any start state, after a finite transient time T_0, the probability of being in any state **at any fixed time $T>T_0$** is nonzero.

BIPARTITE



Not ergodic (even/ odd).

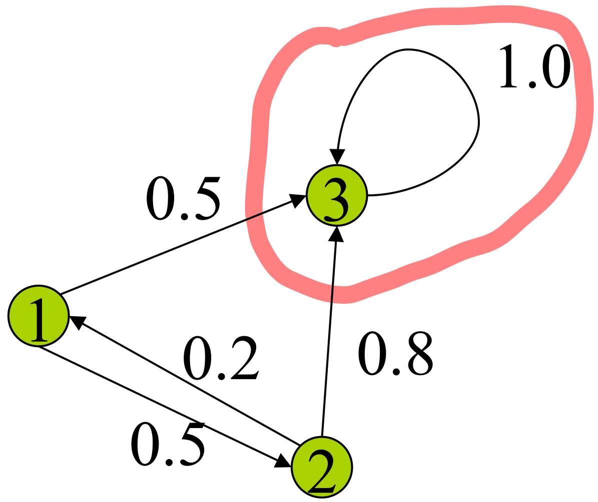
Not ergodic: the probability to be in a state, at a fixed time, e.g., after 500 transitions, is always either 0 or 1 according to the initial state.

Ergodic Markov chains

- For any ergodic Markov chain, there is a **unique long-term visit rate** for each state
  - *Steady-state probability distribution*
- Over a long time-period, we visit each state in proportion to this rate
- It doesn't matter where we start.
- *Note: non ergodic Markov chains may still have a steady state.*

Non Ergodic Example

$$P = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.2 & 0 & 0.8 \\ 0 & 0 & 1 \end{pmatrix}$$



- It is easy to show that the steady state (left eigenvector) is $\pi^T = (0\ 0\ 1)$, $\pi^T P = \pi^T$, i.e., is the state 3
- The user will always reach the state 3 and will stay there (*spider trap*)
- This is a non-ergodic Markov Chain (with a steady-state).

Random teleports

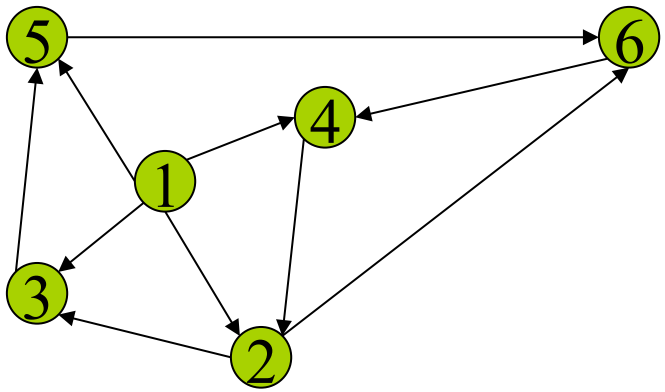
- The Google solution for spider traps (*not for dead ends*)
- At each time step, the random surfer has two options:
  - With probability β, follow a link at random
  - With probability $1-\beta$, jump to some page uniformly at random
  - Common values for β are in the range 0.8 to 0.9
- Surfer will teleport out of spider trap within a few time steps

Matrix formulation

- Suppose there are N pages
  - Consider a page i, with set of outlinks O(i)
  - We have
    - $M_{ij} = 1/|O(i)|$ when *i links j*
    - and $M_{ij} = 0$ otherwise
  - The random teleport is equivalent to
    - adding a **teleport link** from i to every other page with probability $(1-\beta)/N$
    - reducing the probability of following each outlink from $1/|O(i)|$ to $\beta/|O(i)|$
    - Equivalent: tax each page a fraction $(1-\beta)$ of its score and redistribute evenly.

Example

- Simple example with 6 pages



- $P(5|1)=P(4|1)=P(3|1)=P(2|1)= \beta/4 +(1-\beta)/6$
- $P(1|1)=P(6|1)= (1-\beta)/6$
- $P(*|1) = 4[\beta/4 +(1-\beta)/6] + 2(1-\beta)/6 = 1$

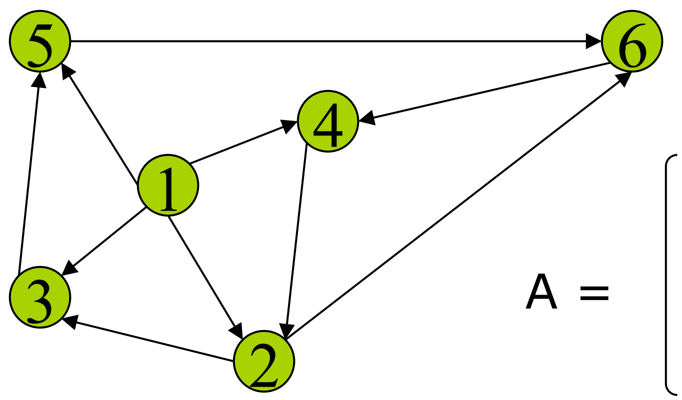
Google Page Rank

- Construct the NxN matrix **A** as follows
  - $A_{ij} = \beta M_{ij} + (1-\beta)/N$
- Verify that **A** is a stochastic matrix
- The **page rank vector r** is the principal eigenvector of this matrix
  - satisfying **r** = **rA**
  - The score of each page r_i satisfies the following:

$$r_i = \beta \left(\sum_{k \in I(i)} \frac{r_k}{|O(k)|} \right) + \frac{(1-\beta)}{N}$$

- I(i) is the set of nodes that have a link **to** page i
- O(k) is the set of links exiting from k
- **r** is the stationary distribution of the random walk with teleports.

Example



$$A = \begin{pmatrix} 0,03 & 0,24 & 0,24 & 0,24 & 0,24 & 0,03 \\ 0,03 & 0,03 & 0,45 & 0,03 & 0,03 & 0,45 \\ 0,03 & 0,03 & 0,03 & 0,03 & 0,88 & 0,03 \\ 0,03 & 0,88 & 0,03 & 0,03 & 0,03 & 0,03 \\ 0,03 & 0,03 & 0,03 & 0,03 & 0,03 & 0,88 \\ 0,03 & 0,03 & 0,03 & 0,88 & 0,03 & 0,03 \end{pmatrix}$$

$P(4|1)=0.24=0.85/4 + 0.15/6$
$P(6|1)=0.03=0.15/6$
$P(4|6)=0.88=0.85/1 + 0.15/6$

$\beta=0.85$

$$A^{30} = \begin{pmatrix} 0,03 & 0,23 & 0,13 & 0,24 & 0,14 & 0,24 \\ 0,03 & 0,23 & 0,13 & 0,24 & 0,14 & 0,24 \\ 0,03 & 0,23 & 0,13 & 0,24 & 0,14 & 0,24 \\ 0,03 & 0,23 & 0,13 & 0,24 & 0,14 & 0,24 \\ 0,03 & 0,23 & 0,13 & 0,24 & 0,14 & 0,24 \\ 0,03 & 0,23 & 0,13 & 0,24 & 0,14 & 0,24 \end{pmatrix}$$

Stationary distribution = (0.03 0.23 0.13 0.24 0.14 0.24)

Dead ends

- Pages with no outlinks are "**dead ends**" for the random surfer (*dangling nodes*)
  - Nowhere to go on next step
- When there are dead ends the matrix is no longer stochastic (the sum of the row elements is not 1)
- This is true even if we add the teleport
  - because the probability to follow a teleport link is only $(1-\beta)/N$ and there are just N of these teleports- hence any of them is $(1-\beta)$

Dealing with dead-ends

- **1) Teleport**
  - Follow random teleport links with probability 1.0 from dead-ends (i.e., for that pages set $\beta = 0$)
  - Adjust matrix accordingly
- **2) Prune and propagate**
  - Preprocess the graph to eliminate dead-ends
  - Might require multiple passes (why?)
  - Compute page rank on reduced graph
  - Approximate values for dead ends by propagating values from reduced graph

Computing page rank

- Key step is matrix-vector multiply
  - $\mathbf{r}^{new} = \mathbf{r}^{old}\mathbf{A}$
- Easy if we have enough main memory to hold $\mathbf{A}$, $\mathbf{r}^{old}$, $\mathbf{r}^{new}$
- Say N = 1 billion pages
  - We need 4 bytes (32 bits) for each entry (say)
  - 2 billion entries for vectors $\mathbf{r}^{new}$ and $\mathbf{r}^{old}$, approx 8GB
  - Matrix A has N^2 entries, i.e., 10^{18}
    - it is a large number!

Sparse matrix formulation

- Although **A** is a dense matrix, it is obtained from a sparse matrix **M**
  - 10 links per node, approx 10N entries
- We can restate the page rank equation
  - $\mathbf{r} = \beta\mathbf{r}\mathbf{M} + [(1-\beta)/N]_N$ *(see slide 63)*
  - $[(1-\beta)/N]_N$ is an N-vector with all entries $(1-\beta)/N$
- So in each iteration, we need to:
  - Compute $\mathbf{r}^{new} = \beta\mathbf{r}^{old}\mathbf{M}$
  - Add a constant value $(1-\beta)/N$ to each entry in $\mathbf{r}^{new}$

Sparse matrix encoding

- Encode sparse matrix using only nonzero entries
  - Space proportional roughly to number of links
  - say 10N, or 4*10*1 billion = 40GB
  - still won't fit in memory, but will fit on disk

source node	degree	destination nodes
0	3	1, 5, 7
1	5	17, 64, 113, 117, 245
2	2	13, 23

Basic Algorithm

- Assume we have enough RAM to fit $\mathbf{r}^{new}$, plus some working memory
  - Store $\mathbf{r}^{old}$ and matrix $\mathbf{M}$ on disk

Basic Algorithm:

- Initialize: $\mathbf{r}^{old} = [1/N]_N$
- Iterate:
  - **Update:** Perform a sequential scan of $\mathbf{M}$ and $\mathbf{r}^{old}$ and update $\mathbf{r}^{new}$
  - Write out $\mathbf{r}^{new}$ to disk as $\mathbf{r}^{old}$ for next iteration
  - Every few iterations, compute $|\mathbf{r}^{new}-\mathbf{r}^{old}|$ and stop if it is below threshold
    - Need to read in both vectors into memory

Update step

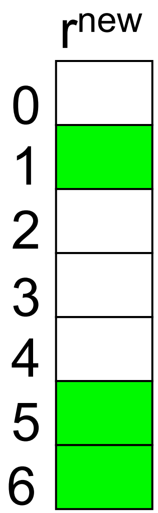
Initialize all entries of $\mathbf{r}^{new}$ to $(1-\beta)/N$

For each page p (out-degree n):

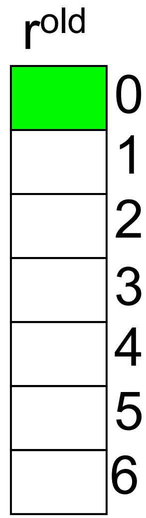
  Read into memory: p, n, $dest_1, \ldots, dest_n$, $r^{old}(p)$

  for j = 1…N:

    $r^{new}(dest_j)$ += $\beta * r^{old}(p)/n$



src	degree	destination
0	3	1, 5, 6
1	4	17, 64, 113, 117
2	2	13, 23



The old value in 0 contributes to updating only the new values in 1,5, and 6.