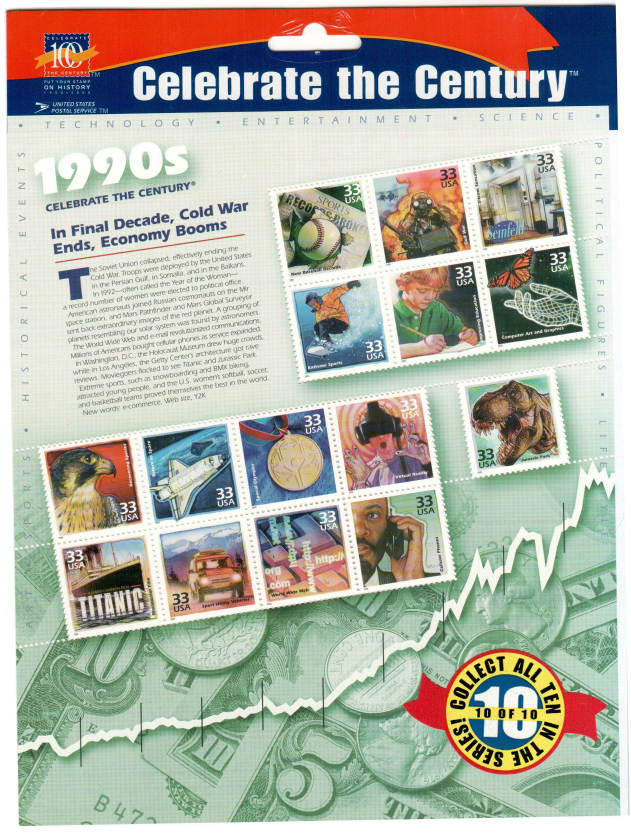
La matematica di

Laura Geatti
Università di Roma "Tor Vergata"

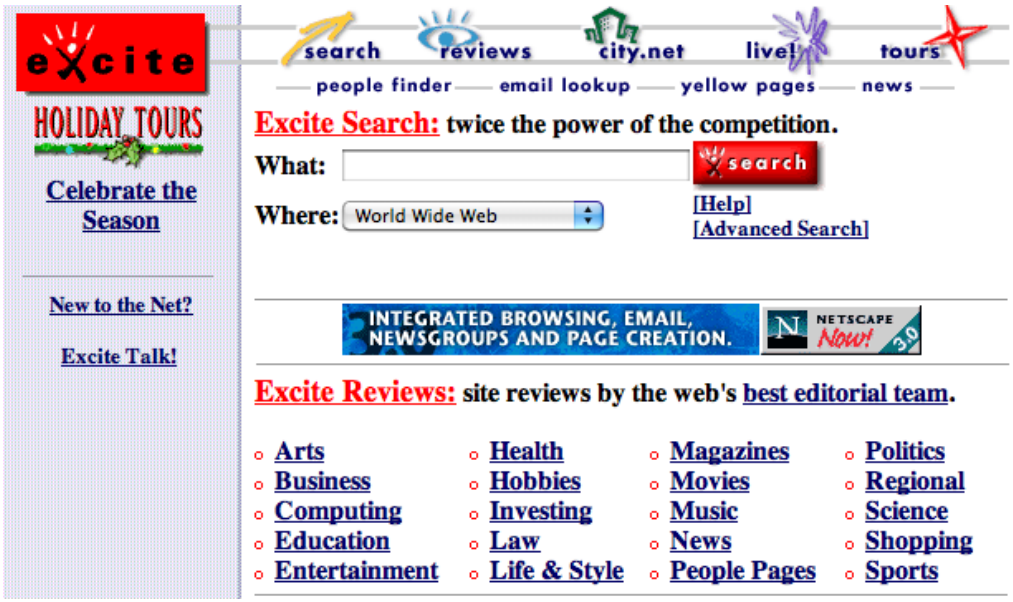
All'inizio degli anni '90....



"Il **World Wide Web** ravviva **Internet**, fino allora *solo-testo*, con immagini, suoni e filmati. Milioni di persone accedono a Internet per lavoro, studio e divertimento."

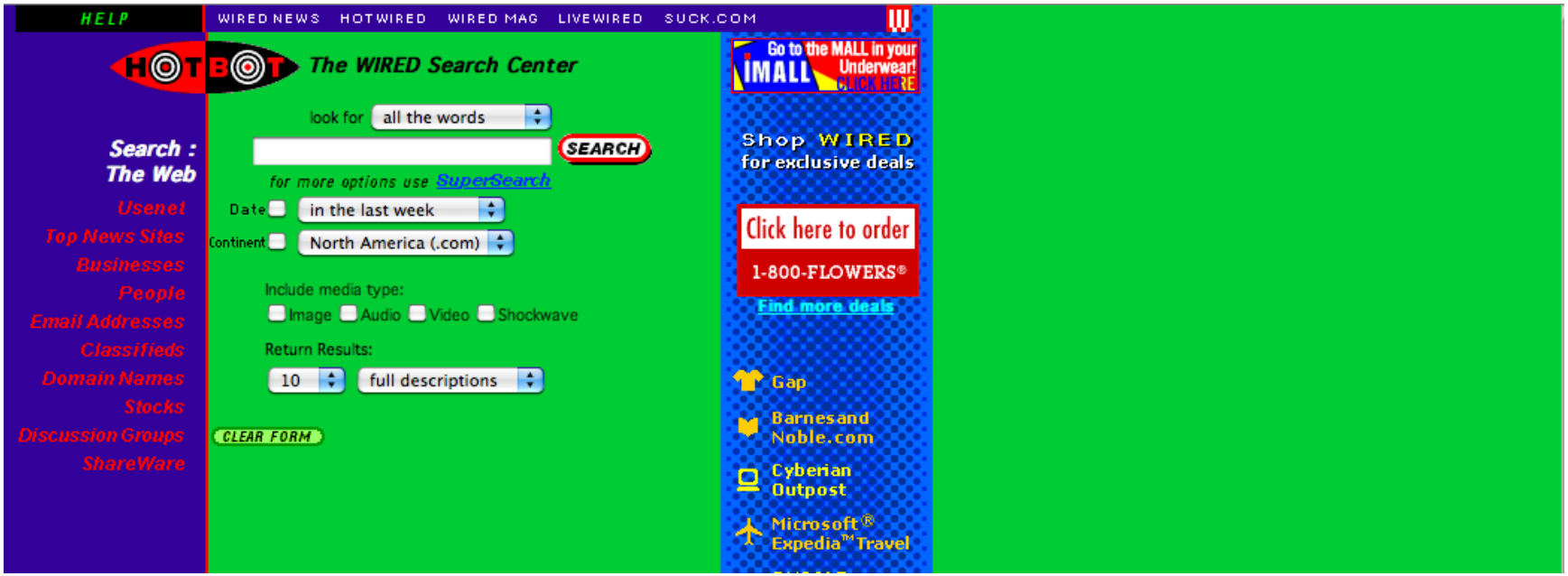
Alcuni motori di ricerca di quel periodo

eXcite, come appariva il 19 dicembre 1996



Altavista, come appariva il 19 dicembre 1996

HOTBOT, come appariva il 12 dicembre 1997



Marzo-aprile 1994: il motore di ricerca **World Wide Web Worm**

un indice di 150.000 pagine

circa 1500 ricerche al giorno

Novembre 1997: i migliori motori di ricerca

un indice fra 2.000.000 e 100.000.000 pagine

circa 20.000.000 ricerche al giorno

Il World Wide Web contiene ALMENO 8 bilioni di pagine.

Senza un buon criterio per ordinare i risultati di una ricerca

l'informazione cercata è sommersa

in un mare di risultati irrilevanti.

Il 27 settembre 1998, appare Google

Google, la pagina dell'11 novembre 1998

Gli inventori: Larry Page & Sergey Brin

Presto Google diventa il motore di ricerca più usato.

Google presenta i risultati più rilevanti in testa.

Il segreto di Google ?

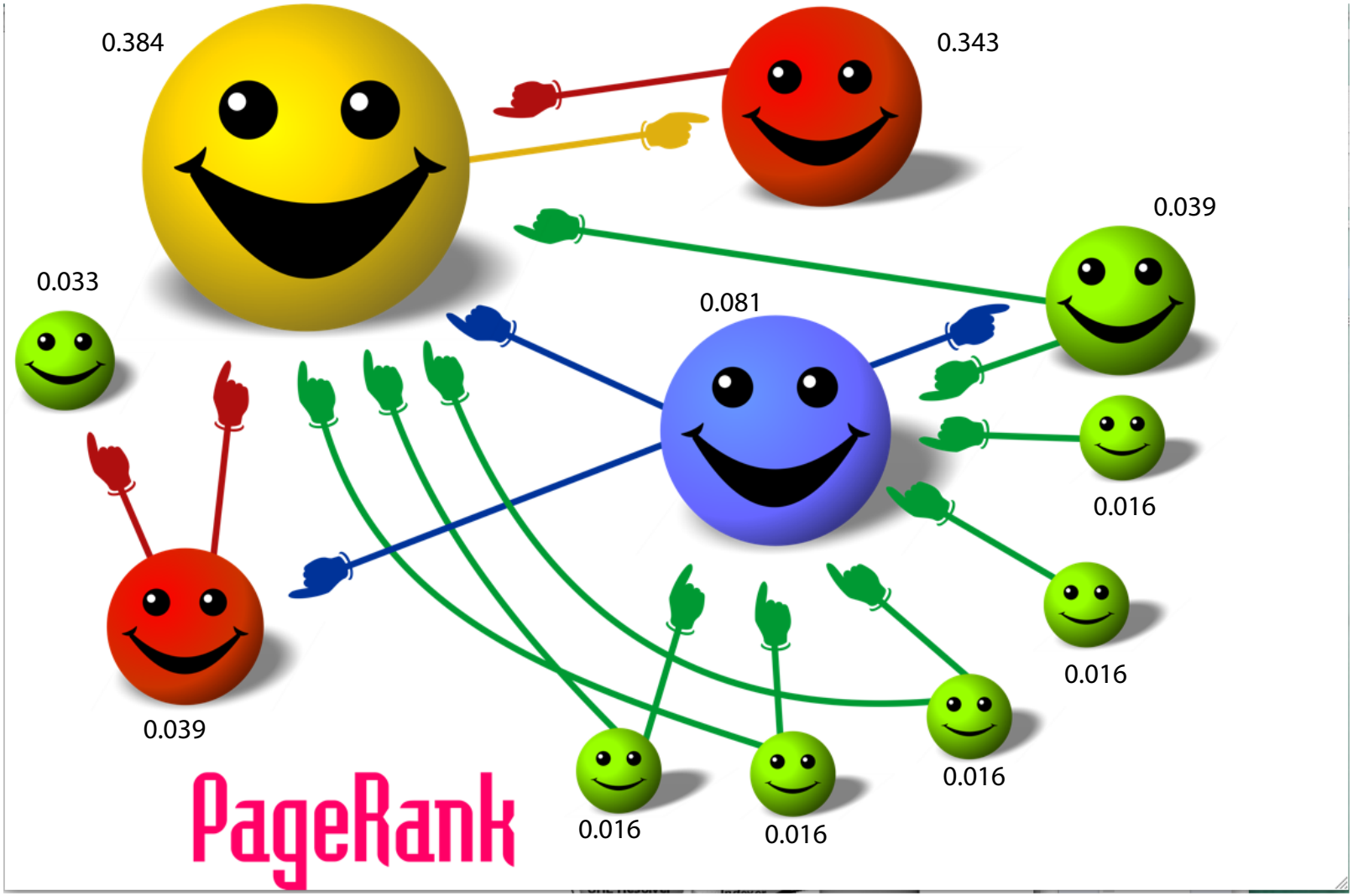
l'algoritmo PageRank,

che calcola il grado di rilevanza di ogni pagina del web.

Due miti da sfatare

Le pagine importanti sono le più visitate?
FALSO!

Sono quelle che hanno più links da altre pagine?
FALSO!


0.384
0.343
0.039
0.033
0.081
0.016
0.016
0.039
0.016
0.016
0.016
PageRank

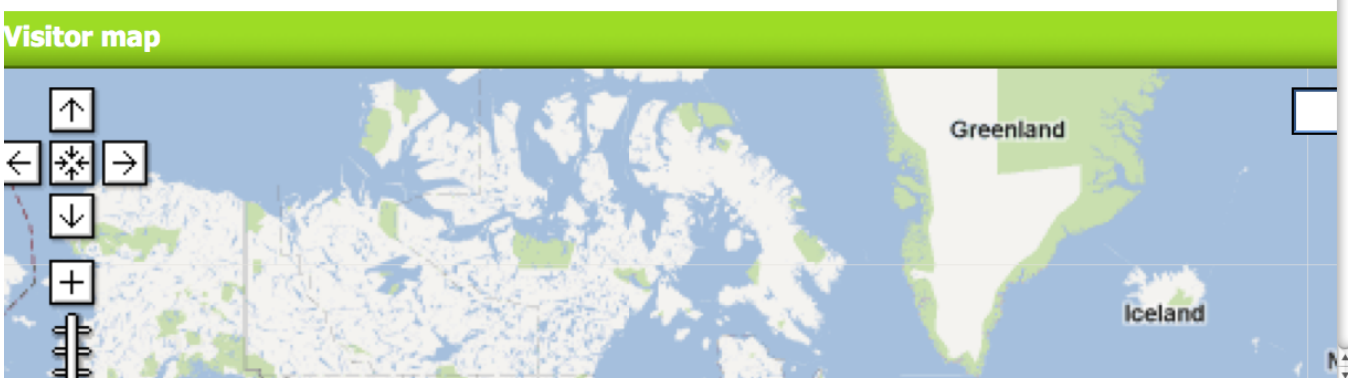
Da cosa dipende la rilevanza di una pagina?

Dipende dalla rilevanza

delle pagine che hanno un link verso di essa.

Google indicizza le pagine del Web

4. 29 January	11:34		Fastweb, Milan, Lombardia, Italy
5. 30 January	13:15		H3G Italy, Manoppello, Abruzzi, Italy
6. 31 January	08:02		Fastweb, Rome, Lazio, Italy
7. 31 January	20:14		Telecom Italia, Parma, Emilia-Romagna, Italy
8. 7 February	15:23		Universita' di Ferrara, Ferrara, Emilia-Romagna, Italy
9. 10 February	17:10		Infostrada IUnet, Rome, Lazio, Italy
10. 12 February	14:48		Google, Mountain View, California, United States



$P_1, \quad \ldots, P_m$ pagine

L'algoritmo PageRank assegna ad ogni pagina un numero $x_i \geq 0$

$$x_i = \text{“grado di rilevanza” della pagina } P_i$$

$$x_1 + x_2 + \ldots + x_m = 1$$

x_i grande $\quad \Leftrightarrow \quad$ pagina P_i importante.

L'algoritmo PageRank

CONTA

n_i = il numero di links da P_i verso altre pagine

ASSUME

che la rilevanza x_i della pagina P_i soddisfi la relazione lineare

$$x_i = \sum_{j \neq i} \frac{1}{n_j} x_j,$$

dove j indicizza le pagine P_j che hanno un link verso P_i.

$$
\begin{aligned}
x_1 &= \frac{1}{n_2}x_2 + \frac{1}{n_3}x_3 + \dots \\
x_2 &= \frac{1}{n_1}x_1 + \frac{1}{n_3}x_3 + \dots \\
\vdots &= \quad \vdots \quad \vdots \\
x_m &= \frac{1}{n_1}x_1 + \frac{1}{n_3}x_2 + \dots
\end{aligned}
$$

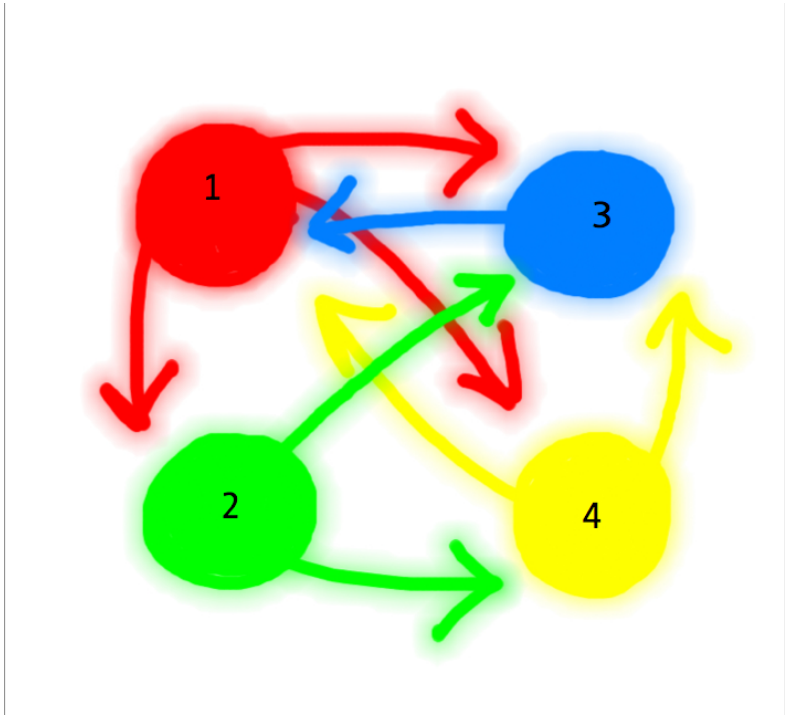
I numeri

$$x_1, x_2, \dots, x_m$$

sono soluzioni di un sistema lineare.

LA TEORIA: sotto certe condizioni $x_1, x_2, \dots, x_m$ sono univocamente determinati.

Proviamo l'algoritmo PageRank su un modello più piccolo:



Ci sono 4 pagine P_1, P_2, P_3, P_4

$n_1 = 3$, $n_2 = 2$, $n_3 = 1$, $n_4 = 2$.

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

Nel linguaggio dell'ALGEBRA LINEARE

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

è un AUTOVETTORE della matrice

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

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

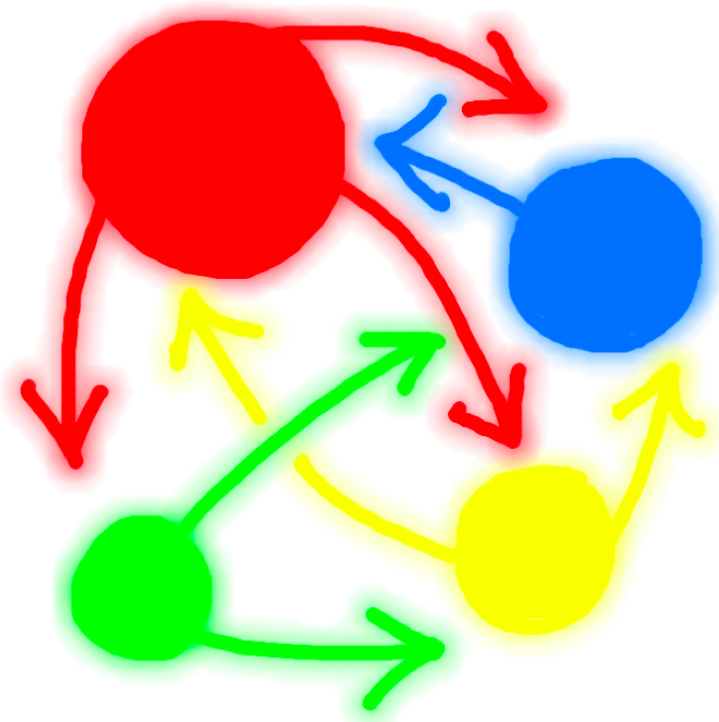
KURT BRYAN† AND TANYA LEISE‡

Abstract. Google's success derives in large part from its PageRank algorithm, which ranks the importance of webpages according to an eigenvector of a weighted link matrix. Analysis of the PageRank formula provides a wonderful applied topic for a linear algebra course. Instructors may assign this article as a project to more advanced students, or spend one or two lectures presenting the material with assigned homework from the exercises. This material also complements the discussion of Markov chains in matrix algebra. Maple and Mathematica files supporting this material can be found at www.rose-hulman.edu/∼bryan.

Key words. linear algebra, PageRank, eigenvector, stochastic matrix

AMS subject classifications. 15-01, 15A18, 15A51

Risolvendo il sistema



$$x_1 = 12/31 \sim 0.387$$
$$x_2 = 4/31 \sim 0.129$$
$$x_3 = 9/31 \sim 0.290$$
$$x_4 = 6/31 \sim 0.194$$

27 settembre 2011: GOOGLE compie 13 anni.