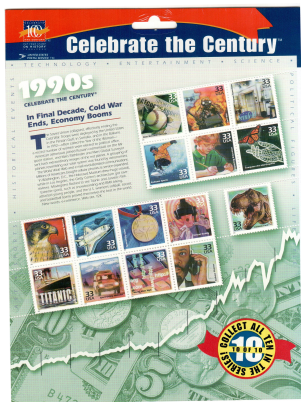


La matematica di

Google

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

The screenshot shows the Excite search engine interface. On the left is a vertical sidebar with the Excite logo and promotional text. The main content area features a search bar with a 'search' button, a 'Where:' dropdown menu set to 'World Wide Web', and links for '[Help]' and '[Advanced Search]'. Below the search bar is a banner for 'INTEGRATED BROWSING, EMAIL, NEWSGROUPS AND PAGE CREATION.' with a Netscape News 3.0 logo. At the bottom, there is a section for 'Excite Reviews' with a grid of category links.

excite

HOLIDAY TOURS

[Celebrate the Season](#)

[New to the Net?](#)

[Excite Talk!](#)

[search](#) [reviews](#) [city.net](#) [live!](#) [tours](#)

[people finder](#) [email lookup](#) [yellow pages](#) [news](#)

Excite Search: twice the power of the competition.

What: [excite search](#)

Where: [\[Help\]](#) [\[Advanced Search\]](#)

INTEGRATED BROWSING, EMAIL, NEWSGROUPS AND PAGE CREATION.

Excite Reviews: site reviews by the web's [best editorial team](#).

- [Arts](#)
- [Business](#)
- [Computing](#)
- [Education](#)
- [Entertainment](#)
- [Health](#)
- [Hobbies](#)
- [Investing](#)
- [Law](#)
- [Life & Style](#)
- [Magazines](#)
- [Movies](#)
- [Music](#)
- [News](#)
- [People Pages](#)
- [Politics](#)
- [Regional](#)
- [Science](#)
- [Shopping](#)
- [Sports](#)

Altavista, come appariva il 19 dicembre 1996

The screenshot shows the Altavista homepage with the following elements:

- Logo:** "ALTA VISTA Technology, Inc." in purple and green.
- Navigation:** "View Multimedia From Our Vantage Point" in red.
- Advertisement:** A red banner for "AUTOMOTIVE" with the text "Car Buying & Car Insurance Pain Relief" and "Buy and insure new cars & trucks online". It includes a "LOW-COST" speech bubble and small car icons.
- Search:** A search bar with "the Web" and "Standard Form" dropdown menus, and a "Submit" button.
- Text:** "Click here for advertising information - reach millions every month!" and "Search with Digital's Alta Vista [Advanced Search]".
- Tip:** A yellow lightbulb icon with the text "Add your URL to the most popular Search Engines".
- Buttons:** Three buttons: "Free Software Download Now...", "Contests Make Me Laugh...", and "Creative Web Create a Site...".
- Footer:** "FREE WEB SITES!" on both sides of the link "Create Your Personal Web Page For Free With Howdy!".

ALTA VISTA
Technology, Inc.

HOTBOT, come appariva il 12 dicembre 1997

The screenshot shows the HotBot search engine interface. At the top, a navigation bar includes links for [HELP](#), [WIRED NEWS](#), [HOTWIRED](#), [WIRED MAG](#), [LIVEWIRED](#), and [SUCK.COM](#). The main header features the HotBot logo and the text "The WIRED Search Center".

The search area includes a text input field with the placeholder "look for all the words" and a red "SEARCH" button. Below this, there are options for "Date" (set to "in the last week") and "Country" (set to "North America (.com)"). A section for "Include media type:" contains checkboxes for "Image", "Audio", "Video", and "Shockwave". The "Return Results:" section shows a count of "10" and a dropdown menu set to "full descriptions". A "CLEAR FORM" button is located at the bottom of the search area.

On the left side, a vertical menu lists various categories: [Search : The Web](#), [Usenet](#), [Top News Sites](#), [Business](#), [People](#), [Email Addresses](#), [Classifieds](#), [Domain Names](#), [Stocks](#), [Discussion Groups](#), and [ShoreWare](#).

On the right side, there are promotional banners. The top one says "Go to the MALL in your Underwear!" with the IMALL logo. Below it, a banner for "Shop WIRED for exclusive deals" includes a red button that says "Click here to order" and a phone number "1-800-FLOWERS®". A link for "Find more deals" is also present. At the bottom right, there is a list of partner logos: Gap, Barnes and Noble.com, Cyberian Outpost, and Microsoft® Expedia™ Travel.

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.



WorldWideWebSize.com
DAILY ESTIMATED SIZE OF THE WORLD WIDE WEB

The size of the World Wide Web (The Internet)

The Indexed Web contains **at least 8.31 billion pages** (Tuesday, 14 February, 2012).

The Dutch Indexed Web contains **at least 823.24 million pages** (Tuesday, 14 February, 2012).

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 ?

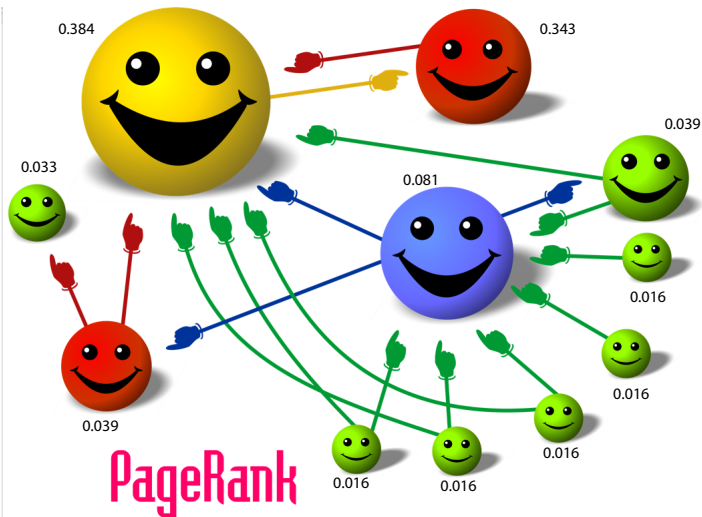
l'algoritmo PageRank,

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!



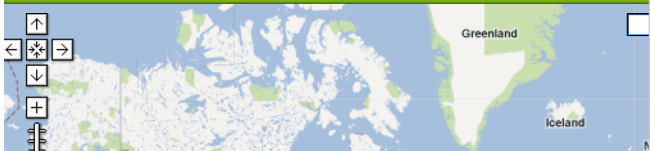
Da cosa dipende la rilevanza di una pagina?

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

Visitor map



P_1, \dots, P_m pagine

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

$x_i =$ "grado di rilevanza" della pagina P_i

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

x_i grande \Leftrightarrow 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 .

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

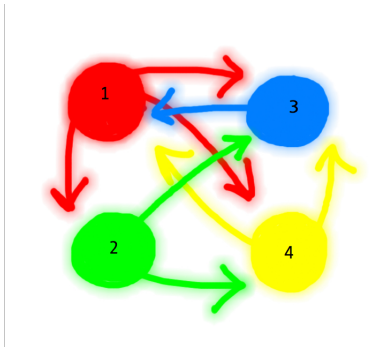
I numeri

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

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

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

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

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

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

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

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

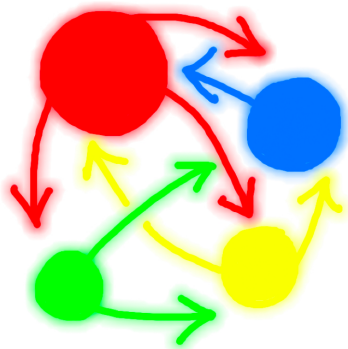
KURT BRYAN[†] AND TANYA LEISE[‡]

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.

