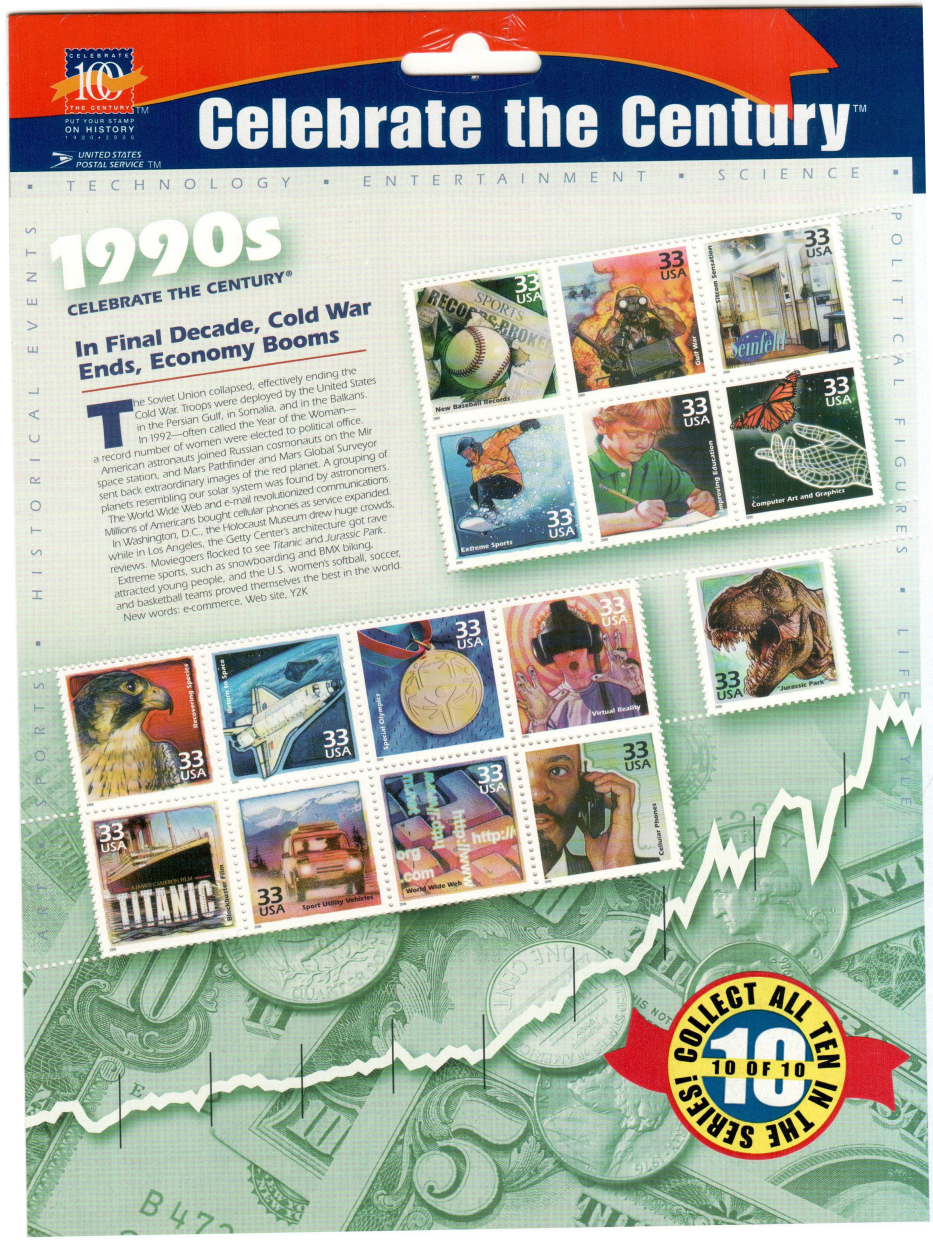
La matematica di

Laura Geatti
Università di Roma "Tor Vergata"
ScienzaOrienta 2018

Roma, 22 febbraio 2018


Celebrate the Century™
PUT YOUR STAMP ON HISTORY 1900•2000
UNITED STATES POSTAL SERVICE TM
TECHNOLOGY • ENTERTAINMENT • SCIENCE
HISTORICAL EVENTS
POLITICAL FIGURES
SPORTS
LIFESTYLES
ARTS
1990s
CELEBRATE THE CENTURY®
In Final Decade, Cold War Ends, Economy Booms
The Soviet Union collapsed, effectively ending the Cold War. Troops were deployed by the United States in the Persian Gulf, in Somalia, and in the Balkans. In 1992—often called the Year of the Woman—a record number of women were elected to political office. American astronauts joined Russian cosmonauts on the Mir space station, and Mars Pathfinder and Mars Global Surveyor sent back extraordinary images of the red planet. A grouping of planets resembling our solar system was found by astronomers. The World Wide Web and e-mail revolutionized communications. Millions of Americans bought cellular phones as service expanded. In Washington, D.C., the Holocaust Museum drew huge crowds, while in Los Angeles, the Getty Center's architecture got rave reviews. Moviegoers flocked to see Titanic and Jurassic Park. Extreme sports, such as snowboarding and BMX biking, attracted young people, and the U.S. women's softball, soccer, and basketball teams proved themselves the best in the world. New words: e-commerce, Web site, Y2K
New Baseball Records
Gulf War
Seinfeld
Extreme Sports
Improving Education
Computer Art and Graphics
Reopening Space
Special Olympics
Virtual Reality
Jurassic Park
TITANIC
Blockbuster Film
Sport Utility Vehicles
World Wide Web
Cellular Phones
33 USA
COLLECT ALL TEN IN THE SERIES!
10 OF 10

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."

World Wide Web

The WorldWideWeb (W3) is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an executive summary of the project, Mailing lists , Policy , November's W3 news , Frequently Asked Questions .

What's out there?
: Pointers to the world's online information, subjects , W3 servers, etc.

Help
: on the browser you are using

Software Products
: A list of W3 project components and their current state. (e.g. Line Mode ,X11 Viola , NeXTStep , Servers , Tools , Mail robot , Library)

Technical
: Details of protocols, formats, program internals etc

Bibliography
: Paper documentation on W3 and references.

People
: A list of some people involved in the project.

History
: A summary of the history of the project.

How can I help ?
: If you would like to support the web..

Getting code
: Getting the code by anonymous FTP , etc.

Marzo-aprile 1994: il **World Wide Web Worm**

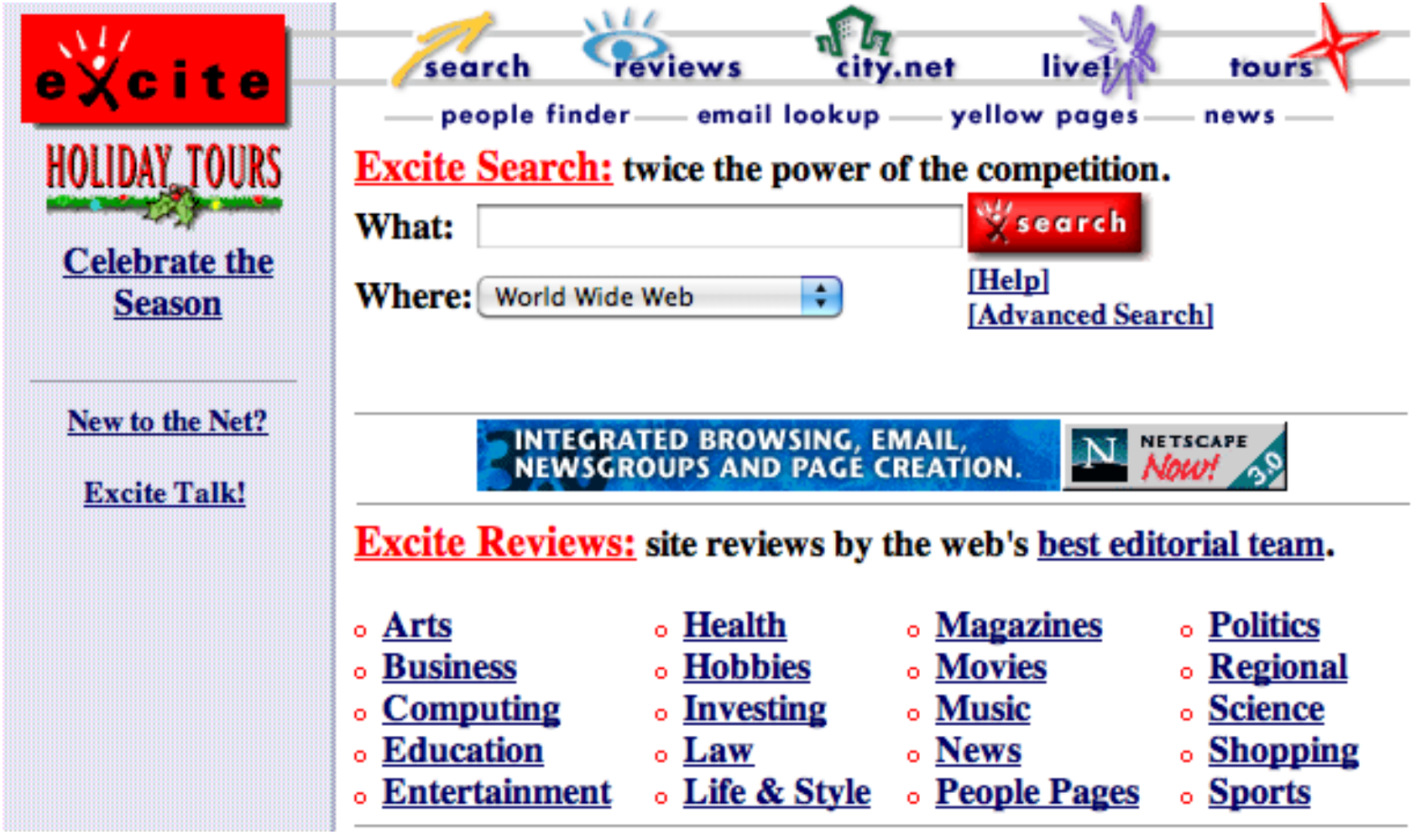
un indice di 110.000/150.000 pagine

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

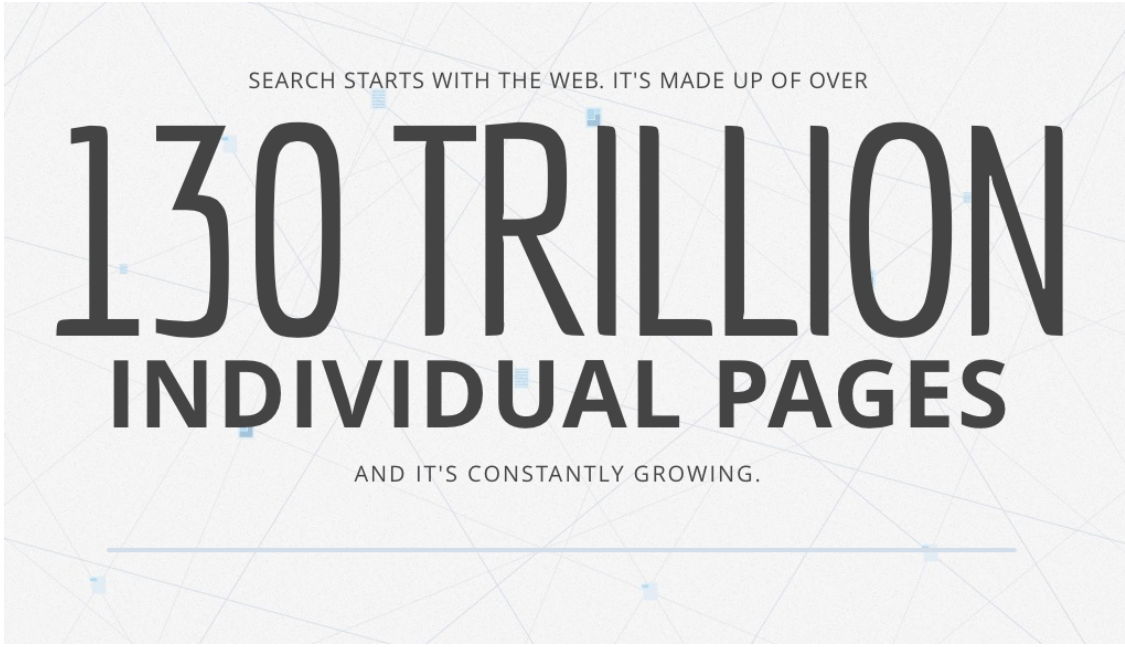

excite
search
reviews
city.net
live!
tours
people finder
email lookup
yellow pages
news
HOLIDAY TOURS
Celebrate the Season
New to the Net?
Excite Talk!
Excite Search: twice the power of the competition.
What:
search
Where: World Wide Web
[Help]
[Advanced Search]
INTEGRATED BROWSING, EMAIL, NEWSGROUPS AND PAGE CREATION.
NETSCAPE Now! 3.0
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
Technology, Inc.
View Multimedia From Our Vantage Point
AUTOBYTEL
USA CANADA
Buy and insure new cars & trucks online
Car Buying & Car Insurance
Pain Relief
LOW-COST
Click here for advertising information - reach millions every month!
Search the Web and Display the Results in Standard Form
Submit
Search with Digital's Alta Vista [Advanced Search]
Add your URL to the most popular Search Engines
Free Software
Download Now...
Contests
Make Me Laugh...
Creative Web
Create a Site...
FREE WEB SITES !
Create Your Personal Web Page For Free With Howdy!
FREE WEB SITES !


HELP
WIRED NEWS HOTWIRED WIRED MAG LIVEWIRED SUCK.COM
HOTBOT
The WIRED Search Center
Search : The Web
Usenet
Top News Sites
Businesses
People
Email Addresses
Classifieds
Domain Names
Stocks
Discussion Groups
ShareWare
look for all the words
SEARCH
for more options use SuperSearch
Date in the last week
Continent North America (.com)
Include media type:
Image Audio Video Shockwave
Return Results:
10 full descriptions
CLEAR FORM
CYBRARIAN
Wired editors' hip-pocket guide to the Web's best sources.
REFERENCE
Statistics
People finder
Dictionaries
Style guides
Atlases
Maps
TECHNOLOGY
News
Computing
Downloads
Web tools
Programming
Organizations
CURRENT AFFAIRS
News
Commentary
Cyber rights
US government
World
BUSINESS
News
Advertising
Marketing
Yellow pages
Organizations
Careers
INVESTING/ FINANCE
Stocks
Indices
SEC filings
IPOs
Research
HEALTH/ SCIENCE
Wellness
Resources
Organizations
Research
Go to the MALL in your Underwear! CLICK HERE
iMALL
Shop WIRED for exclusive deals
Click here to order
1-800-FLOWERS®
Find more deals
Gap
Barnesand Noble.com
Cyberian Outpost
Microsoft® Expedia™ Travel
ONSALE Auctions
1-800-FLOWERS
GetSmart Mortgage Finder
Kodak
Music Boulevard
Microsoft® CarPoint™
Video Serve.com

Oggi:

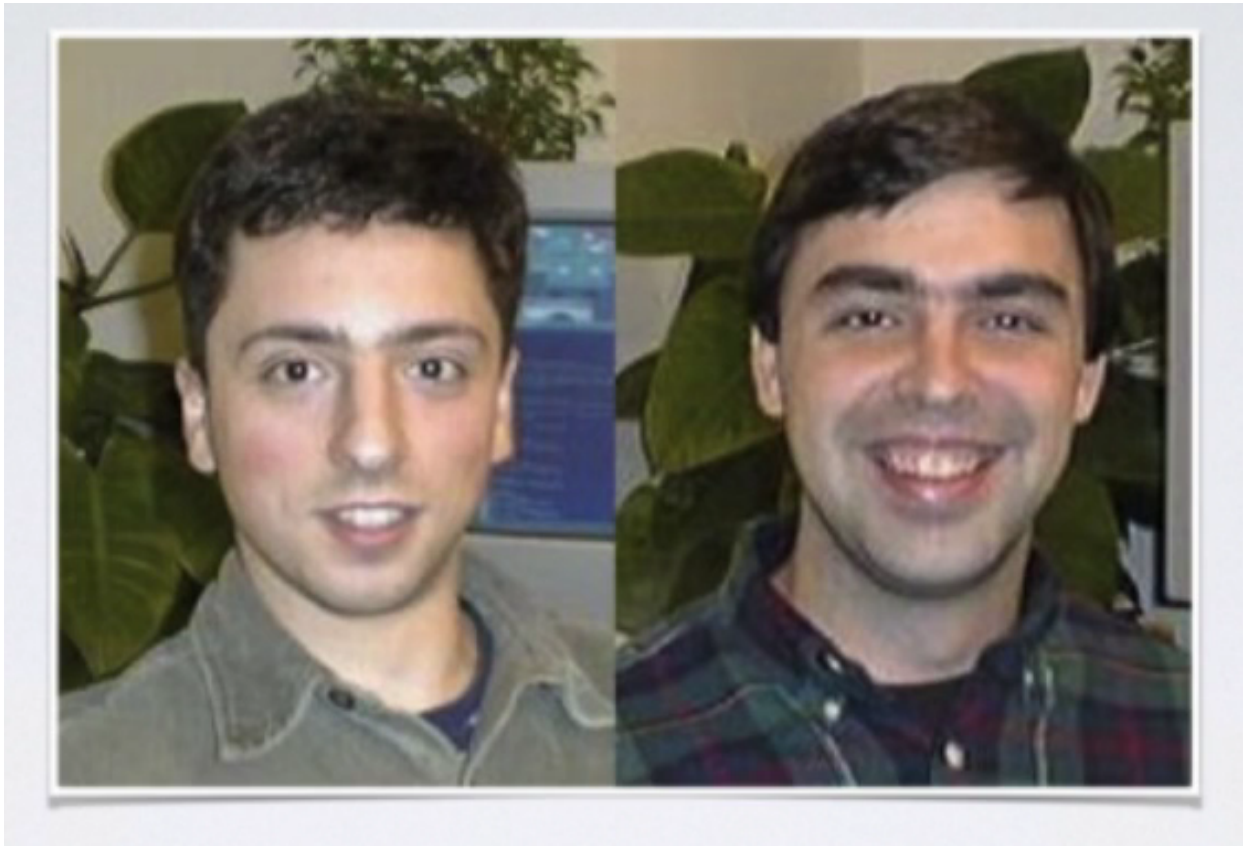
il World Wide Web contiene milioni di milioni di documenti.



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

Gli inventori: Larry Page & Sergey Brin

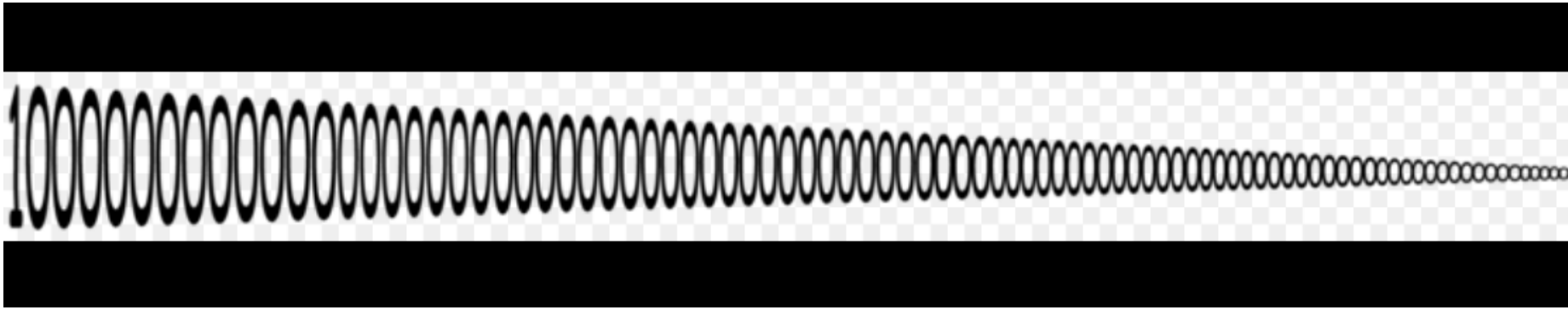


Stanford, 1998

1 googol =

$$1.0 \times 10^{100}$$

. . . visto da Hydro Yot . . .



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 la rilevanza di ogni pagina del web.

Due miti da sfatare

Le pagine importanti sono le più visitate?
FALSO!

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

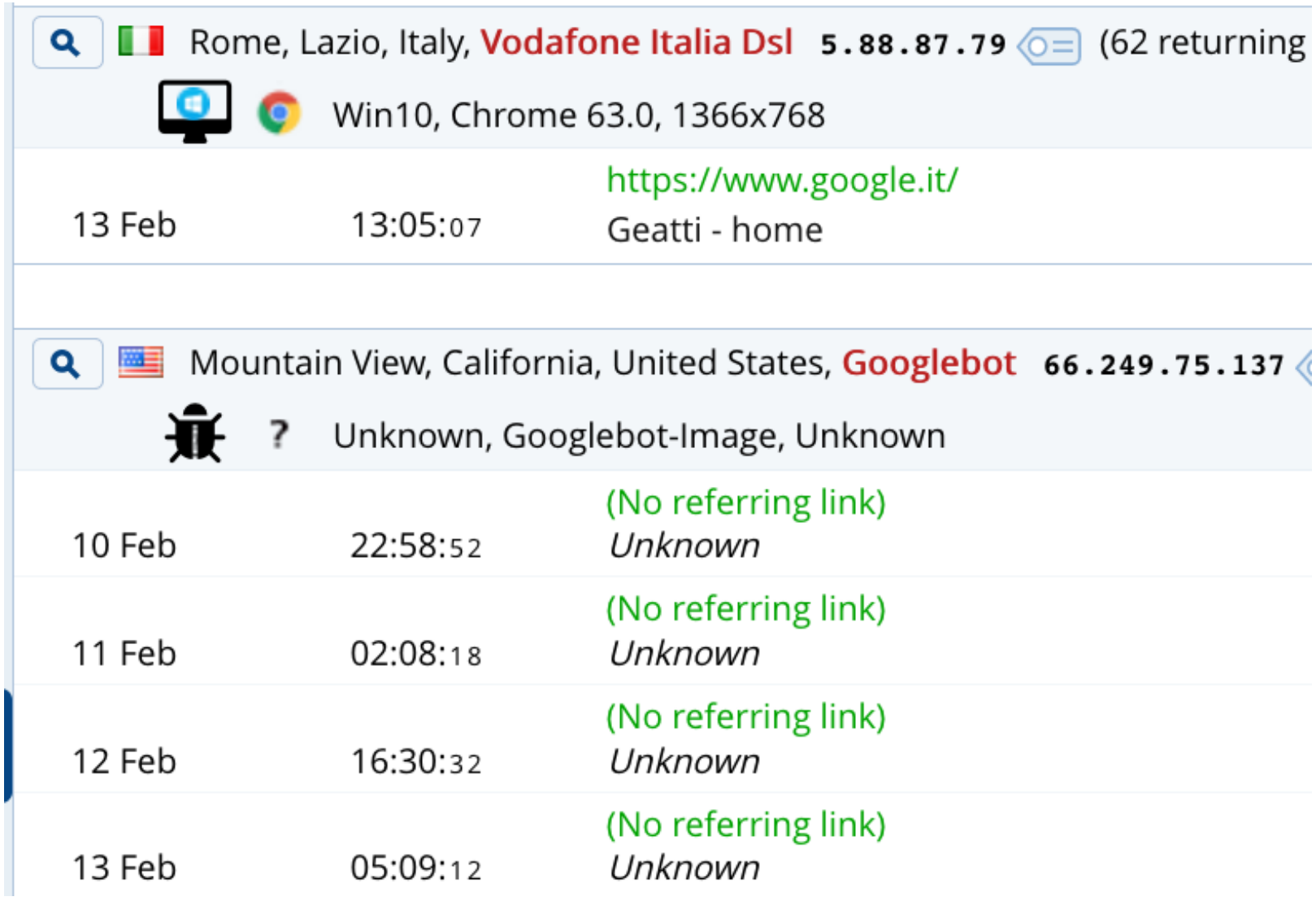
Da cosa dipende la rilevanza di una pagina?

Dipende dalla rilevanza

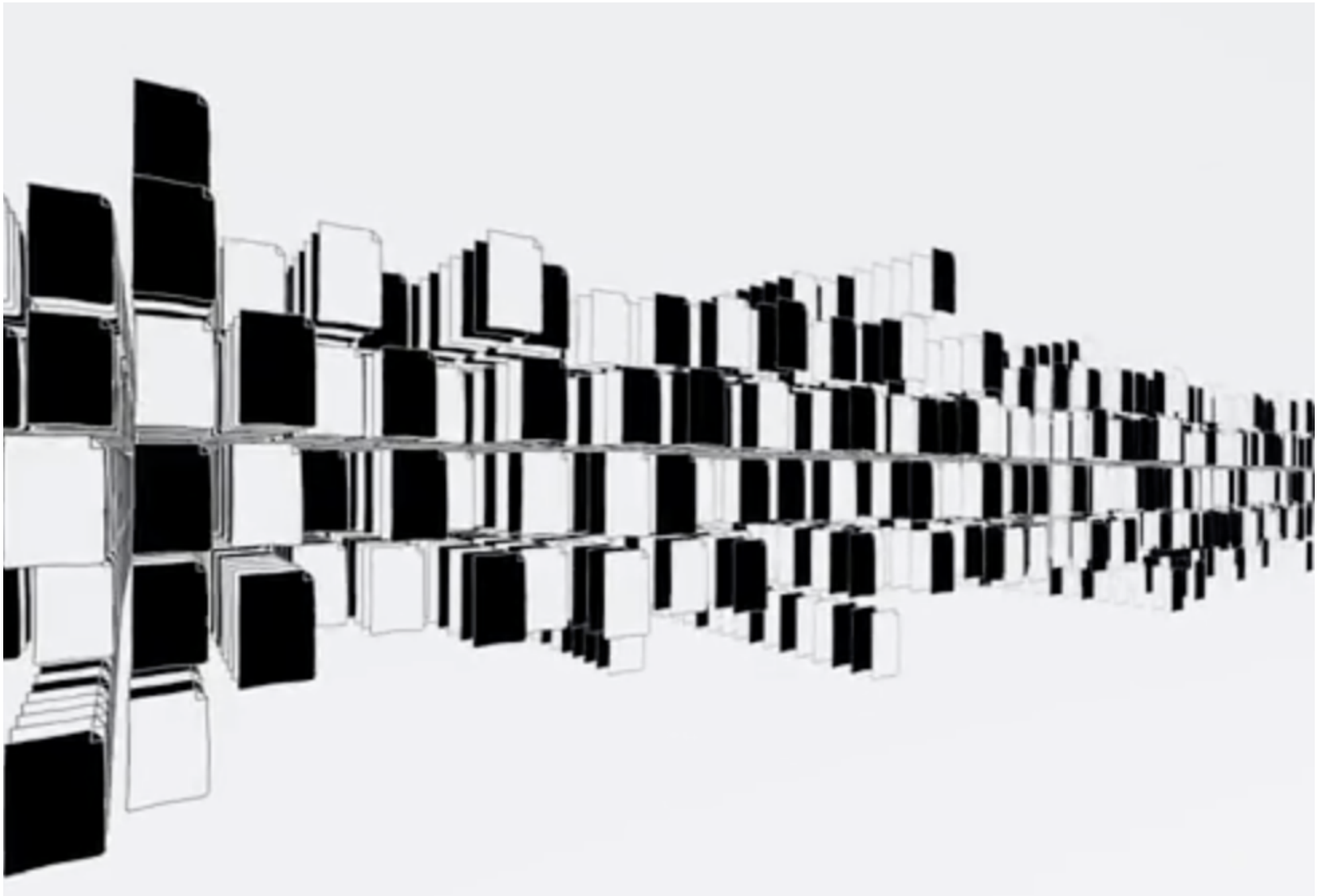
delle pagine che hanno un link verso di essa.

Google scansiona e indicizza le pagine del Web

Il Google crawler sulla mia pagina...



Google crea un indice: analizza e ordina i risultati in base al contenuto.



L'indice di Google contiene

centinaia di miliardi di pagine web

occupa piu' di 100.000.000 di gigabytes

distribuiti in varie localita' del mondo

Google determina la rilevanza di ogni pagina mediante l'algoritmo PageRank

Google conta i link IN & OUT di ogni pagina

$P_1, \ldots, P_m$ le pagine del web

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

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

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

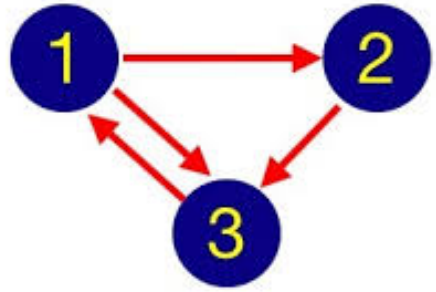
x_i grande $\Leftrightarrow$ pagina P_i importante.

La rilevanza di una pagina è data dalla *somma pesata delle rilevanze delle pagine che hanno un link verso di essa*.....
N.B.: i link a se stessi non contano...!!!

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

dove j indica le pagine P_j che puntano verso P_i,

ed n_j è il numero di link da P_j verso altre pagine.



$$x_1 = 0 \cdot x_1 + 0 \cdot x_2 + 1 \cdot x_3$$
$$x_2 = \frac{1}{2} \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3$$
$$x_3 = \frac{1}{2} \cdot x_1 + 1 \cdot x_2 + 0 \cdot x_3$$
$$x_1 + x_2 + x_3 = 1$$

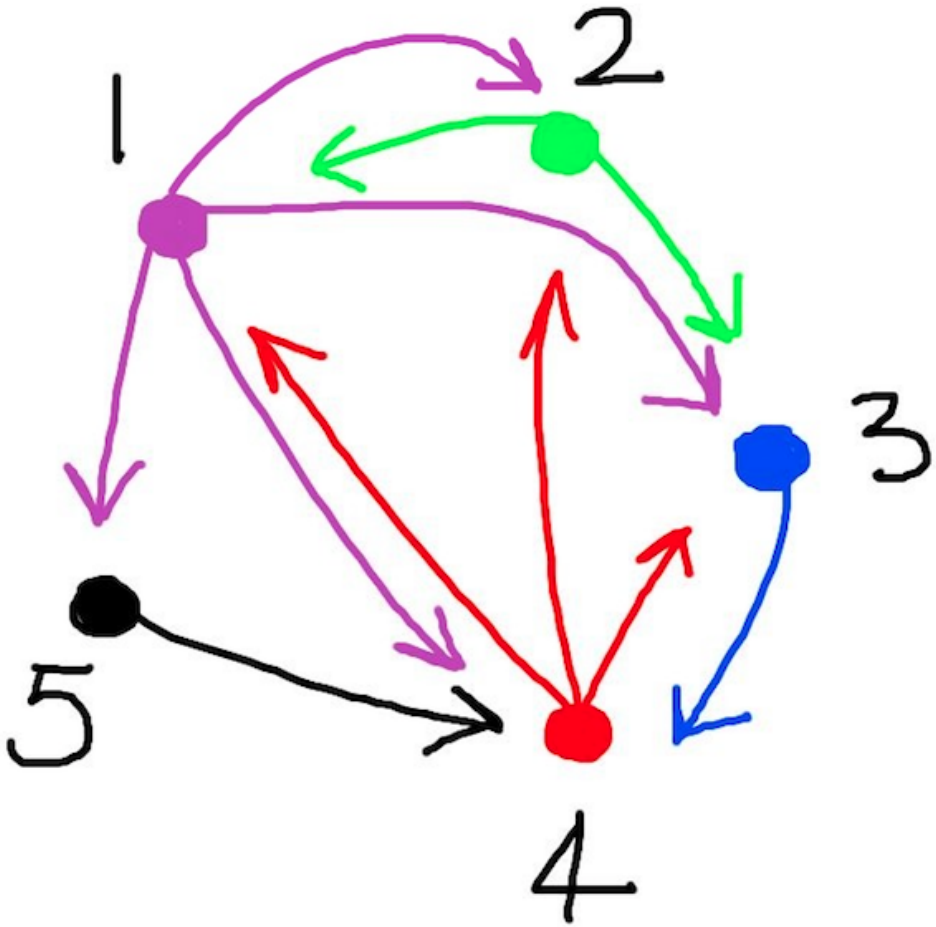
Soluzione $x_1 = x_3 = 0.4, \quad x_2 = 0.2.$

I numeri $x_1, x_2, \ldots, x_m$ sono soluzioni di un sistema lineare:

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

LA TEORIA: *per i sistemi di questo tipo* $x_1, x_2, \ldots, x_m$ sono univocamente determinati.

Proviamo PageRank su un altro mini-esempio:



Ci sono 5 pagine: $P_1, \quad P_2, \quad P_3, \quad P_4, \quad P_5$

$$n_1 = 4, \quad n_2 = 2, \quad n_3 = 1, \quad n_4 = 3, \quad n_5 = 1.$$

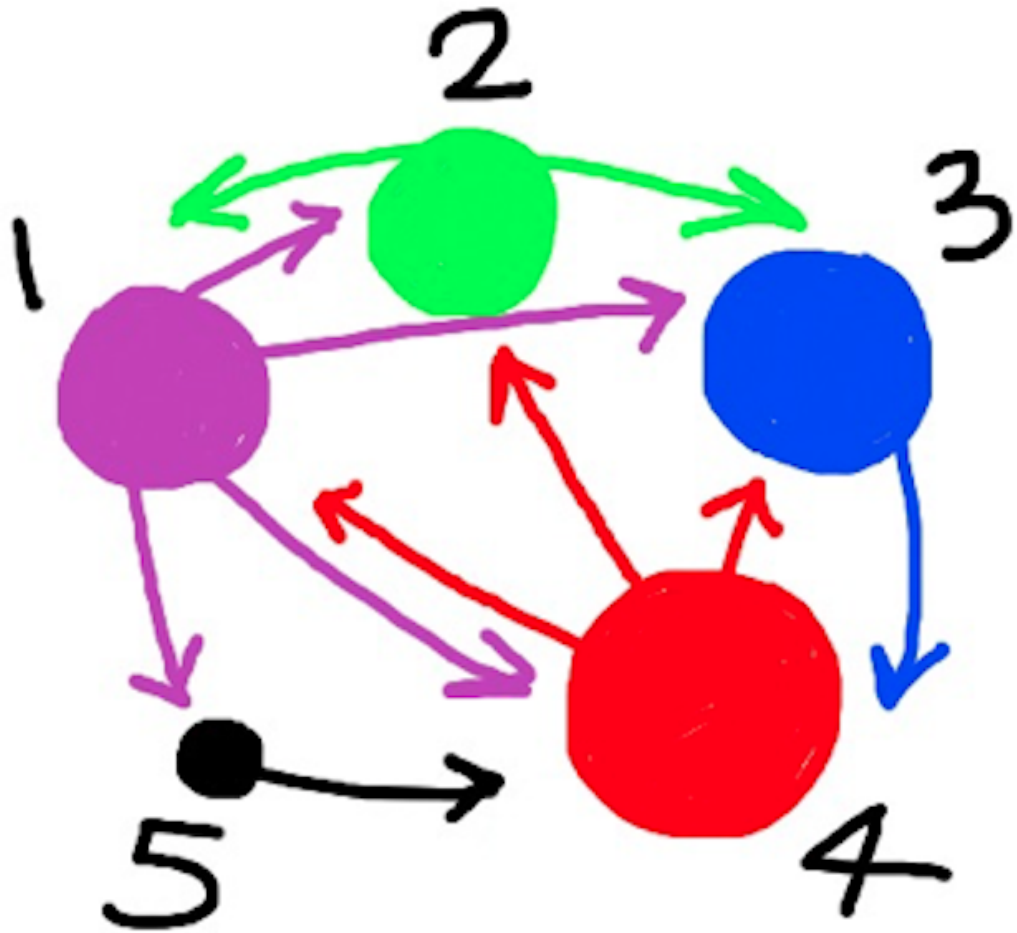
$$\begin{aligned}
x_1 &= \tfrac{1}{2}x_2 + \tfrac{1}{3}x_4 \\
x_2 &= \tfrac{1}{4}x_1 + \tfrac{1}{3}x_4 \\
x_3 &= \tfrac{1}{4}x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{3}x_4 \\
x_4 &= \tfrac{1}{4}x_1 + x_3 + x_5 \\
x_5 &= \tfrac{1}{4}x_1
\end{aligned}$$

$$x_1 + x_2 + x_3 + x_4 + x_5 = 1$$

Risolvendo il sistema troviamo:

$x_4 = 0.33,\ x_3 = 0.26,$

$x_1 = 0.19,\quad x_2 = 0.16,\quad x_5 = 0.04.$



Nel linguaggio dell'ALGEBRA LINEARE

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$$

è un AUTOVETTORE della matrice

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

La matrice di Google è immensa....

$$N \times N, \qquad N > 60 \cdot 10^{12}$$

Calcolare l'autovettore di Google richiede giorni.

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

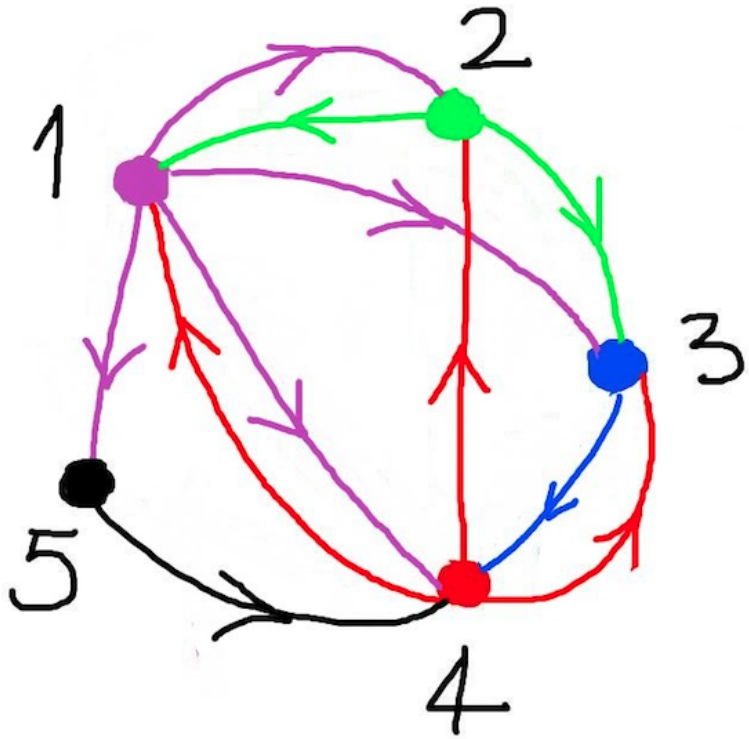
KURT BRYAN† AND TANYA LEISE‡

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.

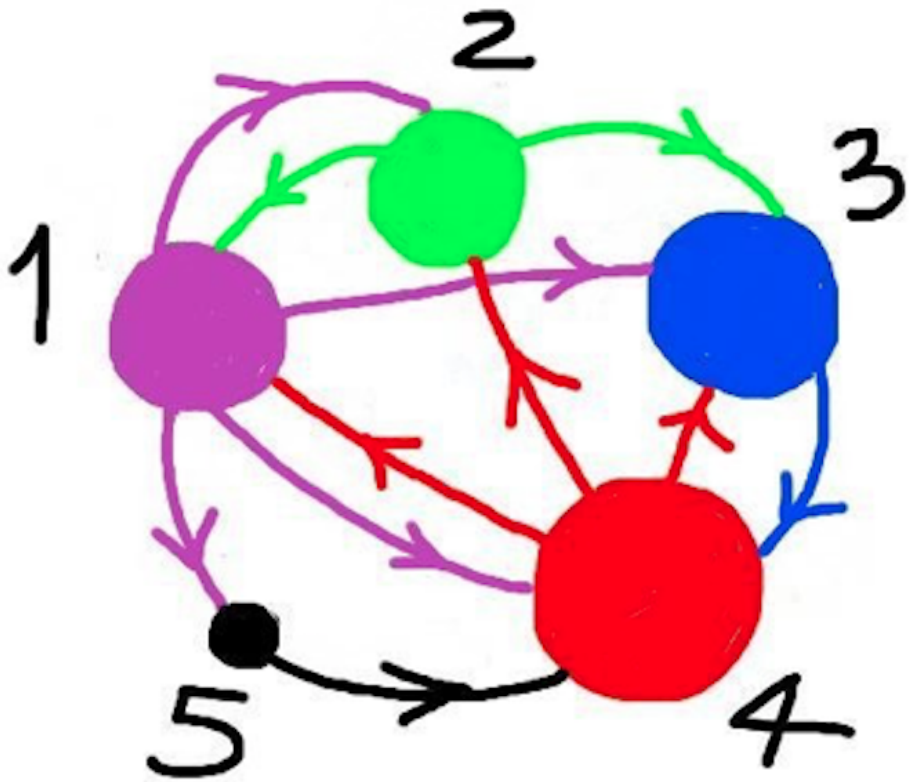
Key words. linear algebra, PageRank, eigenvector, stochastic matrix

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

Interpretazione probabilistica del PageRank di un sito: la rete come grafo orientato.

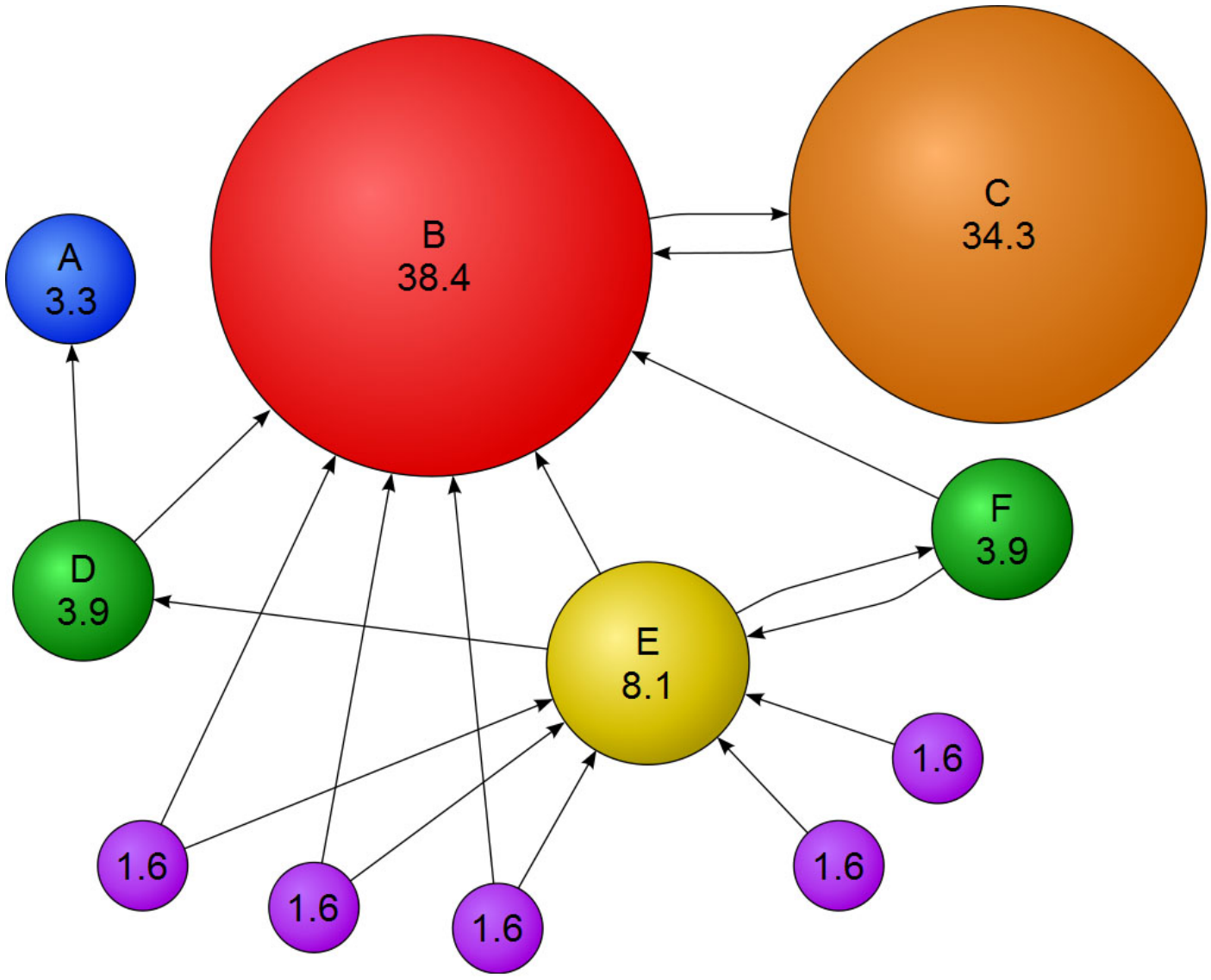


Se facciamo una passeggiata casuale di 100 passi sul grafo... probabilmente passeremo

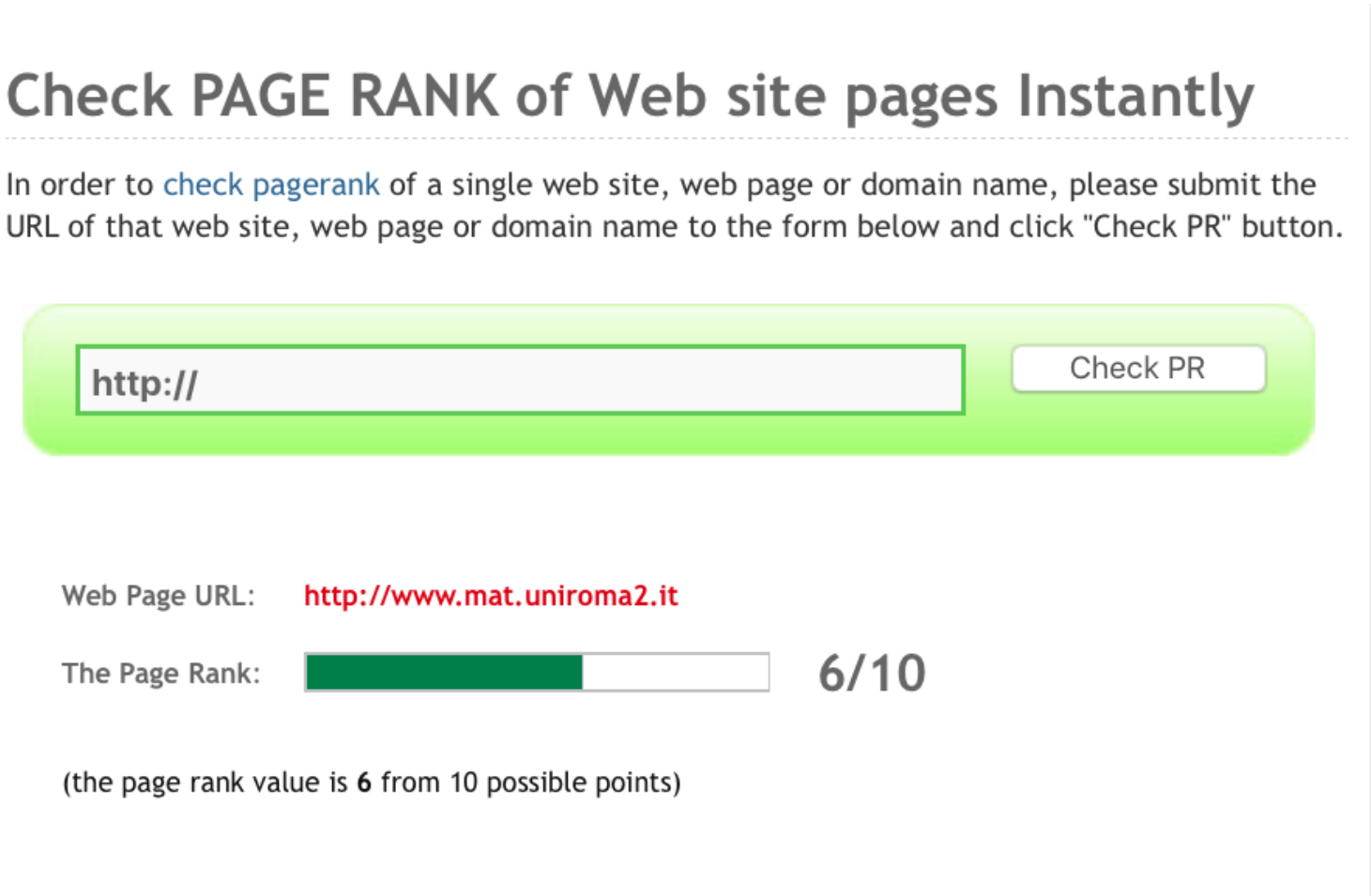


33 volte dal nodo 4, 26 volte dal nodo 3, 19 volte dal nodo 1, 16 volte dal nodo 2, 4 volte dal nodo 5.

Il secondo classificato ha solo un link in arrivo...ma molto importante...



Sul web si trovano diversi programmi che dicono di saper calcolare il PageRank di qualunque sito:



Check PAGE RANK of Web site pages Instantly

In order to check pagerank of a single web site, web page or domain name, please submit the URL of that web site, web page or domain name to the form below and click "Check PR" button.

Web Page URL: http://www.quirinale.it

The Page Rank: 7/10

(the page rank value is **7** from 10 possible points)

Check PAGE RANK of Web site pages Instantly

In order to check pagerank of a single web site, web page or domain name, please submit the URL of that web site, web page or domain name to the form below and click "Check PR" button.

http://

Check PR

Web Page URL: http://web.mit.edu

The Page Rank: 9/10

(the page rank value is **9** from 10 possible points)

27 settembre 2017: Google ha compiuto 19 anni.

Alcuni siti che abbiamo visitato insieme. . .

- La nascita del World Wide Web

`https://home.cern/topics/birth-web`

- La prima pagina web

`http://info.cern.ch/hypertext/WWW/TheProject.html`

- La schermata di un computer prima di internet

`http://line-mode.cern.ch/www/hypertext/WWW/TheProject.h`

- Notizie dal WWW: novembre 1992

`http://info.cern.ch/hypertext/WWW/News/9211.html`

- Come funziona la ricerca di GOOGLE

`https://www.google.com/search/howsearchworks/`

- I centri dati di GOOGLE

`https://www.google.com/about/datacenters`

Bibliografia

• S. Brin, L. Page, *The anatomy of a large scale Hypertextual web search engine*, `www-db.stanford.edu/~sergey`

• K. Bryan, T. Leise, *The $ 25.000.000.000 eigenvector. The linear algebra behind Google*, Siam Review, **48**(3) (2006), 569-581.

• P. Fernández Gallardo, *Google's secret and Linear Algebra*, Newsletter European Math. Soc. N.63, March 2007.