

Matematica del web: il problema di Google

Dario A. Bini

Dipartimento di Matematica, Università di Pisa
www.dm.unipi.it/~bini

Febbraio 2006

- 1 Matematica e applicazioni
- 2 Il problema del Page Rank
- 3 Sistemi Lineari
- 4 Il navigatore virtuale

Luoghi comuni sulla matematica

Esistono molti luoghi comuni (e tante barzellette) sulla matematica

- disciplina fine a se stessa
- astratta e praticamente inutile
- costituita da rigidi meccanismi
- mortifica la fantasia

Il matematico è visto come un personaggio strano che vive nel suo mondo fantastico e si occupa di cose strane e di importanza trascurabile per la vita di ogni giorno.

Luoghi comuni sulla matematica

Esistono molti luoghi comuni (e tante barzellette) sulla matematica

- disciplina fine a se stessa
- astratta e praticamente inutile
- costituita da rigidi meccanismi
- mortifica la fantasia

Il matematico è visto come un personaggio strano che vive nel suo mondo fantastico e si occupa di cose strane e di importanza trascurabile per la vita di ogni giorno.

Luoghi comuni sulla matematica

Esistono molti luoghi comuni (e tante barzellette) sulla matematica

- disciplina fine a se stessa
- astratta e praticamente inutile
- costituita da rigidi meccanismi
- mortifica la fantasia

Il matematico è visto come un personaggio strano che vive nel suo mondo fantastico e si occupa di cose strane e di importanza trascurabile per la vita di ogni giorno.

Luoghi comuni sulla matematica

Esistono molti luoghi comuni (e tante barzellette) sulla matematica

- disciplina fine a se stessa
- astratta e praticamente inutile
- costituita da rigidi meccanismi
- mortifica la fantasia

Il matematico è visto come un personaggio strano che vive nel suo mondo fantastico e si occupa di cose strane e di importanza trascurabile per la vita di ogni giorno.

Luoghi comuni sulla matematica

Esistono molti luoghi comuni (e tante barzellette) sulla matematica

- disciplina fine a se stessa
- astratta e praticamente inutile
- costituita da rigidi meccanismi
- mortifica la fantasia

Il matematico è visto come un personaggio strano che vive nel suo mondo fantastico e si occupa di cose strane e di importanza trascurabile per la vita di ogni giorno.

Pervasività della matematica

Probabilmente alcune di queste cose sono vere. Ma è anche vero che senza la ricerca matematica molte delle funzionalità che utilizziamo, o che sappiamo essere utilizzate ogni giorno, non sarebbero disponibili

La tecnologia avanzata è essenzialmente tecnologia matematica

Ogni giorno ciascuno di noi fa uso inconsapevole di matematica

Pervasività della matematica

Probabilmente alcune di queste cose sono vere. Ma è anche vero che senza la ricerca matematica molte delle funzionalità che utilizziamo, o che sappiamo essere utilizzate ogni giorno, non sarebbero disponibili

La tecnologia avanzata è essenzialmente tecnologia matematica

Ogni giorno ciascuno di noi fa uso inconsapevole di matematica

Pervasività della matematica

Probabilmente alcune di queste cose sono vere. Ma è anche vero che senza la ricerca matematica molte delle funzionalità che utilizziamo, o che sappiamo essere utilizzate ogni giorno, non sarebbero disponibili

La tecnologia avanzata è essenzialmente tecnologia matematica

Ogni giorno ciascuno di noi fa uso inconsapevole di matematica

Alcune applicazioni piacevoli

- **fotografia digitale: jpeg, sharpening, deblurring,...**
- musica digitale: CD, mp3
- film e tv digitali: dvd, dvix, mpeg
- telefonia mobile
- gps, cartografia nautica, computer graphics
- volo automatico
- crittografia
- sport (vela, motociclismo)
- **internet**

Alcune applicazioni piacevoli

- fotografia digitale: jpeg, sharpening, deblurring,...
- musica digitale: CD, mp3
- film e tv digitali: dvd, dvix, mpeg
- telefonia mobile
- gps, cartografia nautica, computer graphics
- volo automatico
- crittografia
- sport (vela, motociclismo)
- **internet**

Alcune applicazioni piacevoli

- fotografia digitale: jpeg, sharpening, deblurring,...
- musica digitale: CD, mp3
- film e tv digitali: dvd, dvix, mpeg
- telefonia mobile
- gps, cartografia nautica, computer graphics
- volo automatico
- crittografia
- sport (vela, motociclismo)
- **internet**

Alcune applicazioni piacevoli

- fotografia digitale: jpeg, sharpening, deblurring,...
- musica digitale: CD, mp3
- film e tv digitali: dvd, dvix, mpeg
- telefonia mobile
- gps, cartografia nautica, computer graphics
- volo automatico
- crittografia
- sport (vela, motociclismo)
- **internet**

Alcune applicazioni piacevoli

- fotografia digitale: jpeg, sharpening, deblurring,...
- musica digitale: CD, mp3
- film e tv digitali: dvd, dvix, mpeg
- telefonia mobile
- gps, cartografia nautica, computer graphics
- volo automatico
- crittografia
- sport (vela, motociclismo)
- **internet**

Alcune applicazioni piacevoli

- fotografia digitale: jpeg, sharpening, deblurring,...
- musica digitale: CD, mp3
- film e tv digitali: dvd, dvix, mpeg
- telefonia mobile
- gps, cartografia nautica, computer graphics
- volo automatico
- crittografia
- sport (vela, motociclismo)
- **internet**

Alcune applicazioni piacevoli

- fotografia digitale: jpeg, sharpening, deblurring,...
- musica digitale: CD, mp3
- film e tv digitali: dvd, dvix, mpeg
- telefonia mobile
- gps, cartografia nautica, computer graphics
- volo automatico
- crittografia
- sport (vela, motociclismo)
- **internet**

Alcune applicazioni piacevoli

- fotografia digitale: jpeg, sharpening, deblurring,...
- musica digitale: CD, mp3
- film e tv digitali: dvd, dvix, mpeg
- telefonia mobile
- gps, cartografia nautica, computer graphics
- volo automatico
- crittografia
- sport (vela, motociclismo)
- **internet**

Alcune applicazioni piacevoli

- fotografia digitale: jpeg, sharpening, deblurring,...
- musica digitale: CD, mp3
- film e tv digitali: dvd, dvix, mpeg
- telefonia mobile
- gps, cartografia nautica, computer graphics
- volo automatico
- crittografia
- sport (vela, motociclismo)
- **internet**

Alcune applicazioni utili

- TAC, la RNM
- modelli cardiaci (potenziale elettrico)
- modelli della circolazione sanguigna (fluidodinamica)
- modelli epidemiologici
- previsioni del tempo
- modelli di code (analisi di rischio)
- modelli economici

Alcune applicazioni utili

- TAC, la RNM
- modelli cardiaci (potenziale elettrico)
- modelli della circolazione sanguigna (fluidodinamica)
- modelli epidemiologici
- previsioni del tempo
- modelli di code (analisi di rischio)
- modelli economici

Alcune applicazioni utili

- TAC, la RNM
- modelli cardiaci (potenziale elettrico)
- modelli della circolazione sanguigna (fluidodinamica)
- modelli epidemiologici
- previsioni del tempo
- modelli di code (analisi di rischio)
- modelli economici

Alcune applicazioni utili

- TAC, la RNM
- modelli cardiaci (potenziale elettrico)
- modelli della circolazione sanguigna (fluidodinamica)
- modelli epidemiologici
- previsioni del tempo
- modelli di code (analisi di rischio)
- modelli economici

Alcune applicazioni utili

- TAC, la RNM
- modelli cardiaci (potenziale elettrico)
- modelli della circolazione sanguigna (fluidodinamica)
- modelli epidemiologici
- previsioni del tempo
- modelli di code (analisi di rischio)
- modelli economici

Alcune applicazioni utili

- TAC, la RNM
- modelli cardiaci (potenziale elettrico)
- modelli della circolazione sanguigna (fluidodinamica)
- modelli epidemiologici
- previsioni del tempo
- modelli di code (analisi di rischio)
- modelli economici

Alcune applicazioni utili

- TAC, la RNM
- modelli cardiaci (potenziale elettrico)
- modelli della circolazione sanguigna (fluidodinamica)
- modelli epidemiologici
- previsioni del tempo
- modelli di code (analisi di rischio)
- modelli economici

Alcune applicazioni militari

- armi “intelligenti”
- aerei killer

Matematica del web

Internet costituisce una sorgente di problemi matematici di particolare interesse teorico e applicativo

- Page ranking (Google)
- Information retrieval
- Gestione del flusso delle informazioni sulla rete

Motori di ricerca e Page Rank

Due studenti dell'università di Stanford, Sergey Brin e Larry Page hanno fatto la loro fortuna inventando "Google"

Problema: Ordinare le pagine presenti sul web in base alla loro importanza

Come si può definire l'importanza (**page rank**) di una pagina?

Motori di ricerca e Page Rank

Due studenti dell'università di Stanford, Sergey Brin e Larry Page hanno fatto la loro fortuna inventando "Google"

Problema: Ordinare le pagine presenti sul web in base alla loro importanza

Come si può definire l'importanza (**page rank**) di una pagina?

Motori di ricerca e Page Rank

Due studenti dell'università di Stanford, Sergey Brin e Larry Page hanno fatto la loro fortuna inventando "Google"

Problema: Ordinare le pagine presenti sul web in base alla loro importanza

Come si può definire l'importanza (**page rank**) di una pagina?

Importanza di una pagina

Varie proposte

- in base al numero di volte che la parola cercata compare
- in base al numero dei link che da essa partono
- in base al numero dei link che ad essa arrivano
- **in base al numero delle pagine importanti che puntano alla pagina**

Importanza di una pagina

Varie proposte

- in base al numero di volte che la parola cercata compare
- in base al numero dei link che da essa partono
- in base al numero dei link che ad essa arrivano
- **in base al numero delle pagine importanti che puntano alla pagina**

Importanza di una pagina

Varie proposte

- in base al numero di volte che la parola cercata compare
- in base al numero dei link che da essa partono
- in base al numero dei link che ad essa arrivano
- **in base al numero delle pagine importanti che puntano alla pagina**

Importanza di una pagina

Varie proposte

- in base al numero di volte che la parola cercata compare
- in base al numero dei link che da essa partono
- in base al numero dei link che ad essa arrivano
- **in base al numero delle pagine importanti che puntano alla pagina**

L'idea di Page e Brin

Ogni pagina ha una sua propria importanza che deriva dalle connessioni (non direttamente dai contenuti)

L'importanza di una pagina viene trasferita in parti uguali alle pagine che essa punta

L'importanza di una pagina è data dalla somma delle frazioni di importanza che gli derivano dalle pagine che ad essa puntano

(Una persona importante dà importanza alle persone che frequenta
Una persona è importante se frequenta molte persone importanti)

Modello matematico

Numeriamo le pagine del web da 1 a n

Definiamo la matrice di connettività nel seguente modo:

$$H = \begin{bmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,n} \\ h_{2,1} & h_{2,2} & \dots & h_{2,n} \\ \vdots & \vdots & & \vdots \\ h_{n,1} & h_{n,2} & \dots & h_{n,n} \end{bmatrix}$$

$h_{i,j} = 1$ se c'è un link dalla pagina i alla pagina j

$h_{i,j} = 0$ altrimenti.

Esempio con $n = 4$

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

①

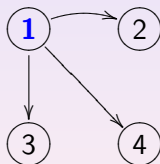
②

③

④

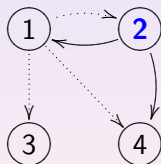
Esempio con $n = 4$

$$H = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$



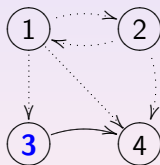
Esempio con $n = 4$

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$



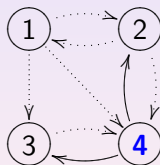
Esempio con $n = 4$

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ 0 & 1 & 1 & 0 \end{bmatrix}$$



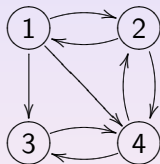
Esempio con $n = 4$

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} \end{bmatrix}$$



Esempio con $n = 4$

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

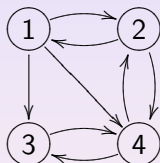


Sommando i valori sulla riga i si trova il numero di link che partono dalla pagina i . Denotiamo con r_i questo numero

Sommando i valori sulla colonna j si trova il numero di pagine che puntano alla pagina j . Denotiamo con c_j questo numero

Esempio con $n = 4$

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

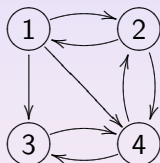


Sommando i valori sulla riga i si trova il numero di link che partono dalla pagina i . Denotiamo con r_i questo numero

Sommando i valori sulla colonna j si trova il numero di pagine che puntano alla pagina j . Denotiamo con c_j questo numero

Esempio con $n = 4$

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$



Sommando i valori sulla riga i si trova il numero di link che partono dalla pagina i . Denotiamo con r_i questo numero

Sommando i valori sulla colonna j si trova il numero di pagine che puntano alla pagina j . Denotiamo con c_j questo numero

Indichiamo con \mathbf{x}_j l'importanza della pagina j

Allora risulta

$$\mathbf{x}_j = h_{1,j} \frac{\mathbf{x}_1}{r_1} + h_{2,j} \frac{\mathbf{x}_2}{r_2} + \cdots + h_{n,j} \frac{\mathbf{x}_n}{r_n}$$

per $j = 1, 2, \dots, n$.

Questo è un **sistema lineare** di n equazioni in n incognite.

Le soluzioni $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ danno il livello di importanza delle singole pagine cioè il **page rank**

Indichiamo con \mathbf{x}_j l'importanza della pagina j

Allora risulta

$$\mathbf{x}_j = h_{1,j} \frac{\mathbf{x}_1}{r_1} + h_{2,j} \frac{\mathbf{x}_2}{r_2} + \cdots + h_{n,j} \frac{\mathbf{x}_n}{r_n}$$

per $j = 1, 2, \dots, n$.

Questo è un **sistema lineare** di n equazioni in n incognite.

Le soluzioni $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ danno il livello di importanza delle singole pagine cioè il **page rank**

Indichiamo con \mathbf{x}_j l'importanza della pagina j

Allora risulta

$$\mathbf{x}_j = h_{1,j} \frac{\mathbf{x}_1}{r_1} + h_{2,j} \frac{\mathbf{x}_2}{r_2} + \cdots + h_{n,j} \frac{\mathbf{x}_n}{r_n}$$

per $j = 1, 2, \dots, n$.

Questo è un **sistema lineare** di n equazioni in n incognite.

Le soluzioni $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ danno il livello di importanza delle singole pagine cioè il **page rank**

L'equazione usata da Google è leggermente diversa

$$\mathbf{x}_j = d\left(h_{1,j}\frac{\mathbf{x}_1}{r_1} + h_{2,j}\frac{\mathbf{x}_2}{r_2} + \cdots + h_{n,j}\frac{\mathbf{x}_n}{r_n}\right) + \frac{1}{n}(1 - d),$$

dove d è un parametro fra 0 e 1, di solito viene posto $d = 0.85$

I valori di \mathbf{x}_j sono compresi fra 0 e 1.

Per calcolare il page rank occorre risolvere un sistema di n equazioni ed n incognite

Alla data di oggi ci sono circa $n = 8.5 \times 10^9$ pagine attive

Come si risolve un sistema lineare?

Se $n = 2$ si ha

$$\begin{cases} ax + by = c \\ dx + ey = f \end{cases}$$

Il metodo di sostituzione detto anche di Eliminazione Gaussiana si applica in generale a sistemi $n \times n$

Esso richiede circa

$$\frac{2}{3}n^3 + \text{spiccioli}$$

operazioni aritmetiche

Come si risolve un sistema lineare?

Se $n = 8.5$ miliardi il metodo di eliminazione richiede circa

$$\frac{2}{3}(8.5 \times 10^9)^3 \approx 4.1 \times 10^{29}$$

(410 miliardi di miliardi di miliardi) operazioni aritmetiche

sono tante?

Complessità del Page Rank

Il calcolatore più veloce esistente al mondo è attualmente il **Blue Gene** dell'IBM

Ha una velocità massima di 360 teraflops cioè 3.6×10^{14} operazioni al secondo (1 tera = 1000 giga)

Per eseguire 4.1×10^{29} operazioni richiederebbe più di 36 milioni di anni

Un tempo “geologico” eppure Larry Page e Sergey Brin **calcolano il page rank ogni mese**

come fanno?

Tecnologia hardware vs tecnologia matematica

Anche se la tecnologia fosse in grado di costruire un computer 1000 volte o un milione di volte più veloce non sarebbe possibile risolvere il problema di Google in tempo reale

Solo sviluppando nuovi metodi matematici è possibile risolvere il sistema con tempi di calcolo brevi

L'algoritmo Page Rank

Equazione

$$\mathbf{x}_j = d\left(h_{1j}\frac{\mathbf{x}_1}{r_1} + h_{2j}\frac{\mathbf{x}_2}{r_2} + \cdots + h_{nj}\frac{\mathbf{x}_n}{r_n}\right) + \frac{1}{n}(1-d),$$

Algoritmo

- 1 Assegna agli \mathbf{x}_j dei valori qualsiasi
- 2 sostituiscili nella parte destra della formula

$$\mathbf{y}_j = d\left(h_{1j}\frac{\mathbf{x}_1}{r_1} + h_{2j}\frac{\mathbf{x}_2}{r_2} + \cdots + h_{nj}\frac{\mathbf{x}_n}{r_n}\right) + \frac{1}{n}(1-d),$$

- 3 e ricava i valori di \mathbf{y}_j , per $j = 1, 2, \dots, n$
- 4 poni $\mathbf{x}_j = \mathbf{y}_j$, per $j = 1, 2, \dots, n$ e proseguì dal punto 2

L' algoritmo Page Rank

Equazione

$$\mathbf{x}_j = d \left(h_{1,j} \frac{\mathbf{x}_1}{r_1} + h_{2,j} \frac{\mathbf{x}_2}{r_2} + \cdots + h_{n,j} \frac{\mathbf{x}_n}{r_n} \right) + \frac{1}{n} (1 - d),$$

Algoritmo

- ① Assegna agli \mathbf{x}_i dei valori qualsiasi
- ② sostituiscili nella parte destra della formula

$$\mathbf{y}_j = d \left(h_{1,j} \frac{\mathbf{x}_1}{r_1} + h_{2,j} \frac{\mathbf{x}_2}{r_2} + \cdots + h_{n,j} \frac{\mathbf{x}_n}{r_n} \right) + \frac{1}{n} (1 - d),$$

- ③ e ricava i valori di \mathbf{y}_j , per $j = 1, 2, \dots, n$
- ④ poni $\mathbf{x}_j = \mathbf{y}_j$ per $j = 1, 2, \dots, n$ e prosegui dal punto 2

L' algoritmo Page Rank

Equazione

$$\mathbf{x}_j = d\left(h_{1,j}\frac{\mathbf{x}_1}{r_1} + h_{2,j}\frac{\mathbf{x}_2}{r_2} + \cdots + h_{n,j}\frac{\mathbf{x}_n}{r_n}\right) + \frac{1}{n}(1 - d),$$

Algoritmo

- ① Assegna agli \mathbf{x}_j dei valori qualsiasi
- ② sostituiscili nella parte destra della formula

$$\mathbf{y}_j = d\left(h_{1,j}\frac{\mathbf{x}_1}{r_1} + h_{2,j}\frac{\mathbf{x}_2}{r_2} + \cdots + h_{n,j}\frac{\mathbf{x}_n}{r_n}\right) + \frac{1}{n}(1 - d),$$

- ③ e ricava i valori di \mathbf{y}_j , per $j = 1, 2, \dots, n$
- ④ poni $\mathbf{x}_j = \mathbf{y}_j$ per $j = 1, 2, \dots, n$ e prosegui dal punto 2

L'algoritmo Page Rank

Equazione

$$\mathbf{x}_j = d\left(h_{1,j}\frac{\mathbf{x}_1}{r_1} + h_{2,j}\frac{\mathbf{x}_2}{r_2} + \cdots + h_{n,j}\frac{\mathbf{x}_n}{r_n}\right) + \frac{1}{n}(1 - d),$$

Algoritmo

- ① Assegna agli \mathbf{x}_j dei valori qualsiasi
- ② sostituiscili nella parte destra della formula

$$\mathbf{y}_j = d\left(h_{1,j}\frac{\mathbf{x}_1}{r_1} + h_{2,j}\frac{\mathbf{x}_2}{r_2} + \cdots + h_{n,j}\frac{\mathbf{x}_n}{r_n}\right) + \frac{1}{n}(1 - d),$$

- ③ e ricava i valori di \mathbf{y}_j , per $j = 1, 2, \dots, n$
- ④ poni $\mathbf{x}_j = \mathbf{y}_j$ per $j = 1, 2, \dots, n$ e prosegui dal punto 2

Viene generata una successione di approssimazioni che converge alla soluzione del sistema **qualunque** siano le approssimazioni iniziali.

$$\mathbf{x}_j^{(1)}, \mathbf{x}_j^{(2)}, \mathbf{x}_j^{(3)}, \dots \rightarrow \mathbf{x}_j$$

Inoltre

$$\begin{aligned} \mathbf{x}_j^{(k)} &\geq 0 \\ \sum_{j=1}^n \mathbf{x}_j^{(k)} &= 1 \end{aligned}$$

Quanto veloce è la convergenza?

L'errore di approssimazione $e^{(k)} = \max_i |x_i^{(k)} - x_i|$ è tale che

$$e^{(k)} \leq \lambda^k \quad \text{con} \quad 0 < \lambda < 1$$

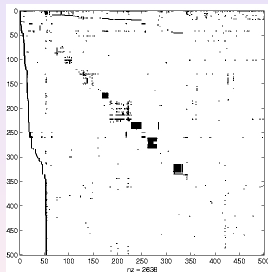
Purtroppo per valori di d vicini a 1 il valore di λ è molto vicino ad 1 e quindi la convergenza è lenta.

λ coincide col modulo del secondo **autovalore** più grande in modulo di una opportuna matrice.

Complessità

Per fare un passo dell'algoritmo del Page Rank bisogna eseguire tante moltiplicazioni quanti sono gli elementi non nulli di H e all'incirca altrettante addizioni

Su ogni riga della matrice H ci sono "pochi" elementi diversi da zero.



Se mediamente ci fossero 50 elementi non nulli su ogni riga, un passo del metodo iterativo eseguito col Blue Gene impiegherebbe 2 millesimi di secondo.

Se anche fossero necessari 1000 passi iterativi basterebbero 2 secondi per approssimare la soluzione di Google.

Interpretazione probabilistica

La soluzione x_j del problema del Page Rank coincide con la probabilità che la pagina j ha di essere visitata da un navigatore virtuale che opera nel seguente modo:

Ad ogni istante (minuto) il navigatore cambia pagina per spostarsi da una pagina all'altra il navigatore

- con probabilità d decide di cliccare su uno dei link presenti sulla pagina corrente scegliendolo a caso con uguale probabilità
- con probabilità $1 - d$ decide di saltare ad un'altra pagina a caso presente sul web scegliendola con uguale probabilità

La probabilità che il navigatore passi dalla pagina i alla pagina j è

$$p_{i,j} = d \frac{h_{i,j}}{r_i} + \frac{1}{n}(1 - d)$$

Denotiamo con

\mathbf{x}_i la probabilità che il navigatore stia visitando la pagina i ad un certo istante

\mathbf{y}_j la probabilità che il navigatore stia visitando la pagina j all'istante successivo

Allora vale

$$\mathbf{y}_j = p_{1,j}\mathbf{x}_1 + p_{2,j}\mathbf{x}_2 + \cdots + p_{n,j}\mathbf{x}_n$$

È il sistema di Google!

$$\mathbf{y}_j = d(h_{1j} \frac{\mathbf{x}_1}{r_1} + h_{2j} \frac{\mathbf{x}_2}{r_2} + \cdots h_{nj} \frac{\mathbf{x}_n}{r_n}) + \frac{1}{n}(1 - d)$$

La soluzione $\mathbf{x}_j^{(k)}$ calcolata dopo k passi dell'algoritmo iterativo di Google fornisce la probabilità che il navigatore virtuale ha di trovarsi nella pagina j dopo k istanti

La soluzione \mathbf{x}_j del sistema fornisce la probabilità che il navigatore virtuale ha di trovarsi nella pagina j dopo “infiniti” istanti di navigazione

La teoria delle catene di Markov e la teoria delle matrici non negative forniscono condizioni di esistenza e unicità della soluzione e di convergenza della successione

Problemi allo studio

- trovare modi diversi di costruire successioni che convergono alla soluzione in modo più rapido
 - Successioni di Krylov
 - metodi del gradiente
 - metodi di aggregazione
 - metodi di estrapolazione
- esprimere la soluzione in funzione di d (come serie di potenze) e vedere come si comporta quando d tende a uno
- studiare modelli matematici diversi, ad esempio: il navigatore salta ad una pagina qualsiasi solo se sulla pagina corrente non ci sono link
- studiare il valore di λ

Bibliografia

- A. ARASU, J. NOVAK, A. TOMKINS, J. TOMLIN, *Page rank computation and the structure of the Web*. (2002)
<http://www2002.org/CDROM/poster/173.pdf>
- A. N. LANGVILLE, C. D. MEYER, *A survey of eigenvector methods for Web information retrieval*, SIAM Rev. **47** (2005), pp. 135–161.
- C. MOLER, *Numerical computing with Matlab*.
<http://www.mathworks.com/moler> sezione 2.11 (2004)
- L. PAGE, S. BRIN, R. MOTWANI, T. WINOGRAD, *The pagerank citation ranking: Bringing order to the Web*,
<http://dbpubs.stanford.edu:8090/pub/1999-66>.

Alcuni tools

Qui trovate alcuni tools per giocare col page rank

Gli indirizzi li ho ottenuti cercando sul web "page rank" tools

- <http://www.rustybrick.com/pagerank-prediction.php>
- <http://www.googlerankings.com>
- <http://www.prchecker.info>
- <http://topologies.net/page-rank-lookups.htm>
- <http://www.seochat.com/seo-tools/future-pagerank>
- http://www.seoutilility.com/it/Page_Rank_Check.htm
- <http://www.markhorrell.com/tools/pagerank.shtml>