

# Sperimentazione sull'influenza del parametro $\gamma$ sull'ordinamento del PageRank Corso di LSMC, a.a. 2017-2018

Davide Gori  
550282

April 4, 2018

## 1 Obiettivi e descrizione della sperimentazione

Vogliamo valutare come e quanto il parametro  $\gamma$  influenza l'ordinamento PageRank. In particolare confronteremo il caso  $\gamma = 0.85$  (valore consigliato da Google) e il caso  $\gamma = 0.99$  (in cui il vettore di personalizzazione ha una bassa incidenza).

## 2 Sperimentazione

Generiamo una matrice di adiacenza con  $10n$  elementi non nulli; usiamo i seguenti parametri di **PageRank**:  $itmax = 1000$ ,  $v(i) = 1$  per ogni  $0 \leq i \leq n$ . Di seguito il procedimento eseguito:

- Calcoliamo i due vettori **y1**, **y2** di ranking con la funzione **PageRank** usando i parametri appena specificati e rispettivamente  $\gamma_1 = 0.85$ ,  $\gamma_2 = 0.99$ .
- Calcoliamo le due permutazioni **sigma1** e **sigma2** che permutano le pagine in ordine crescente di importanza.
- Calcoliamo il vettore che nella posizione  $i$  ha la posizione nel ranking  $\gamma_1$  dell'elemento che occupa l' $i$ -esimo posto nel ranking  $\gamma_2$ .
- Tracciamo il grafico delle prime  $k = 1000$  componenti e poi facciamo lo stesso per  $k = 2000, 10000$  (basterà decommentare e commentare lo script).

## 3 Lo script

Questo è lo script che realizza la sperimentazione:

```

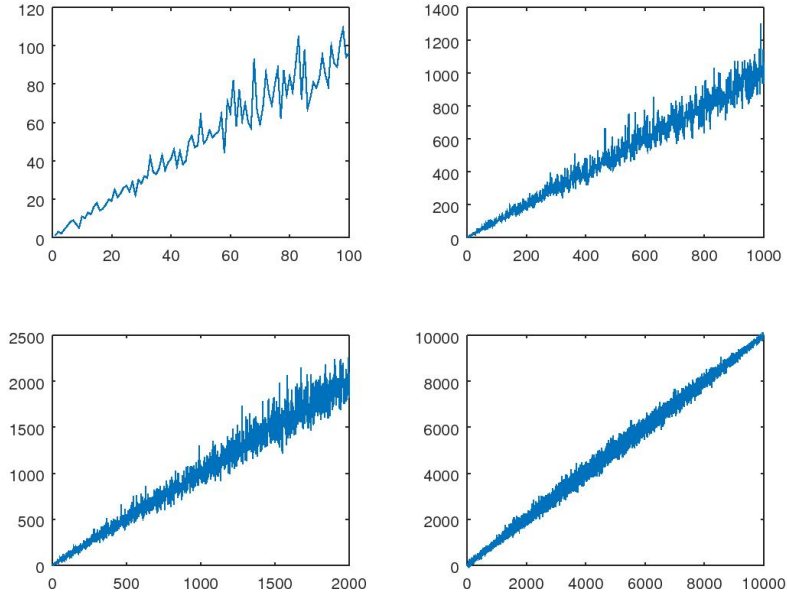
1 n=10000;
2 d=10/n;
3 y=[1:n];
4 H=sprand(n, n, d) ~0;
5 itmax=1000;
6 v=ones(1,n);
7 y1=PageRank(H, v, 0.85, itmax);
8 y2=PageRank(H, v, 0.99, itmax);
9 [val1, sigma1] = sort(y1);
10 [val2, sigma2] = sort(y2);
11 w=zeros(1, n);
12 for i=1:n
13     w(i)=find(sigma1==sigma2(i));
14 end
15 clf
16 for i=1:4
17     subplot(2, 2, i);
18     plot(w(1:[100 1e3 2e3 1e4](i)));
19 end

```

Dove la funzione `PageRank` è la solita funzione usata nella precedente sperimentazione.

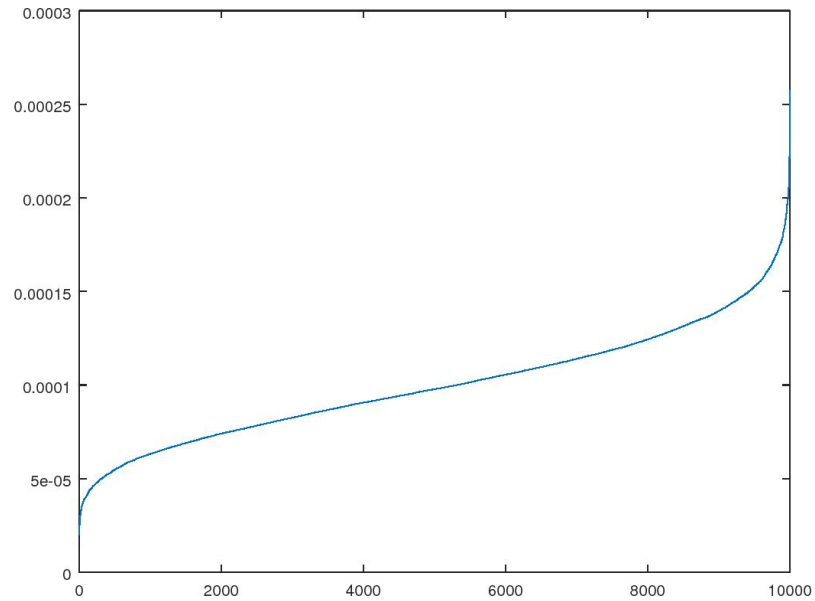
## 4 Risultati

Riportiamo i grafici delle prime 100, 1000, 2000, 10000, componenti di  $w$ :



Possiamo notare che nelle prime e ultime componenti del vettore  $w$  è presente un minor discostamento dalla bisettrice, mentre nelle posizioni centrali questo fenomeno è più evidente. Questo a mio avviso è causato da due fattori:

- Effetto bordo: le prime e le ultime posizioni del vettore di pagerank possono essere "spostate" dalla permutazione  $w$  solo in un senso (nel caso delle prime ho che la posizione non può diminuire e analogamente per le ultime non può aumentare)
- Concentrazione dei valori di PageRank: osservando il grafico che plotta i valori di PageRank delle pagine in ordine (cioè plotta *val1*) si vede che la "derivata" di questa curva è maggiore all'inizio e alla fine, quindi le pagine nel "mezzo" del vettore  $w$  hanno rank molto vicini e cambiare  $\gamma$  fa variare molto di più le posizioni mediane che quelle estremali.



## 5 Sperimentazione variando $n$ e densità della matrice sparsa

Proviamo ora a variare la dimensione della matrice e la densità applicando sempre la stessa sperimentazione.

## 6 Lo script

Questo è lo script che realizza la sperimentazione:

```

1 N=[10000, 10000, 10000, 5000, 5000, 5000];
2 D=[10/10000, 100/10000, 1000/10000, 10/5000, 100/5000, 1000/5000];
3 for j=1:6
4     n=N(j);
5     d=D(j);
6     y=[1:n];
7     H=sprand(n, n, d) ~0;
8     itmax=1000;
9     v=ones(1,n);
10    y1=PageRank(H, v, 0.85, itmax);
11    y2=PageRank(H, v, 0.99, itmax);
12    [val1, sigma1] = sort(y1);
13    [val2, sigma2] = sort(y2);

```

```

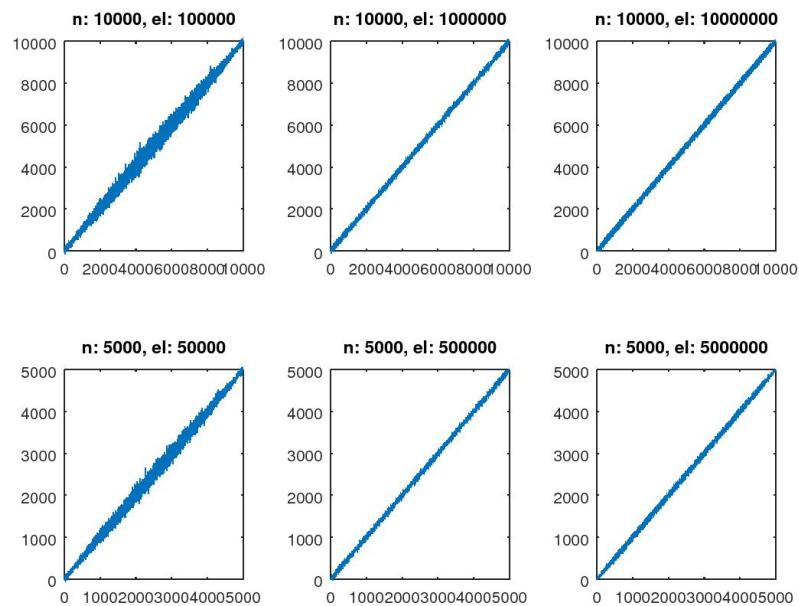
14 w=zeros(1, n);
15 for i=1:n
16     w(i)=find(sigma1==sigma2(i));
17 end
18 subplot(2, 3, j);
19 plot(w);
20 title(sprintf("n: %d, el: %d", n, n*n*d))
21 drawnow;
22 end

```

Dove la funzione PageRank è la solita funzione usata nella precedente sperimentazione.

## 7 Risultati

Riportiamo i grafici di  $w$  per i vari valori di  $n$  e  $n^2 \cdot d$  in sovrapposizione:



Si può notare che aumentando il numero di elementi non nulli si ha un discostamento minore del grafico dalla bisettrice.