Aspetti politici

- I dati relativi alle pubblicazioni devono essere pubblici.
- La validazione deve essere data dalla comunità stessa.
- Gli algoritmi devono essere pubblici.
- Ognuno deve potere verificare il ranking eseguendo in proprio gli algoritmi.

… purtroppo

- I dati sono considerati una proprietà di valore da custodire gelosamente.
- Quelli pubblici sono in uno stato pietoso, tale da renderli praticamente inutilizzabili.
- Chi pubblica sull'argomento spesso si procura i dati per vie traverse e non è disposto a condividerli con la comunità scientifica.

I dati sintetici

- Possono essere generati facilmente in qualunque taglio.
- Possono essere progettati per simulare situazioni particolari.
- Permettono una specie di *prova sul banco* degli algoritmi.

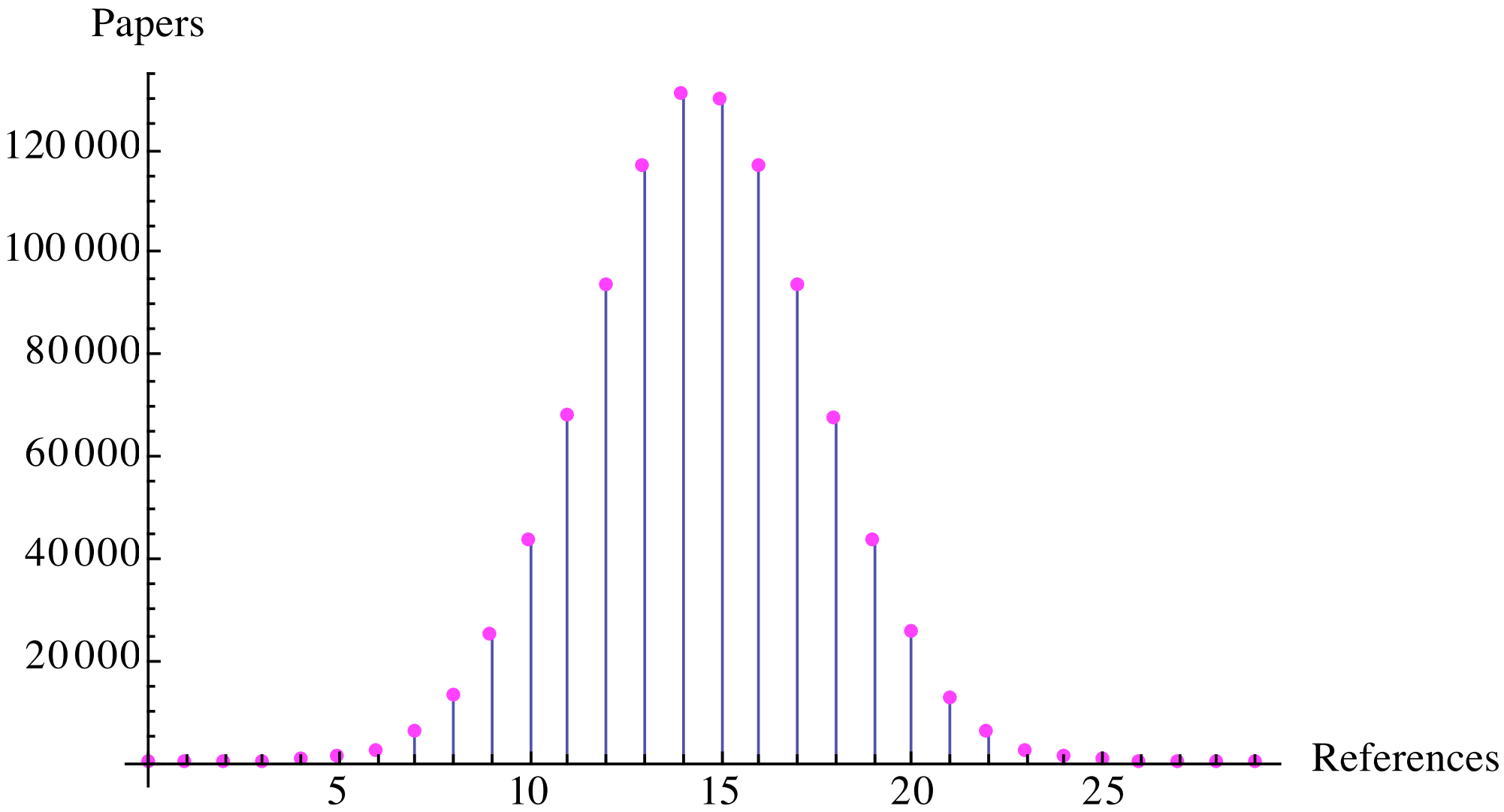
  ovviamente
- Non rispecchiano fedelmente la realtà.

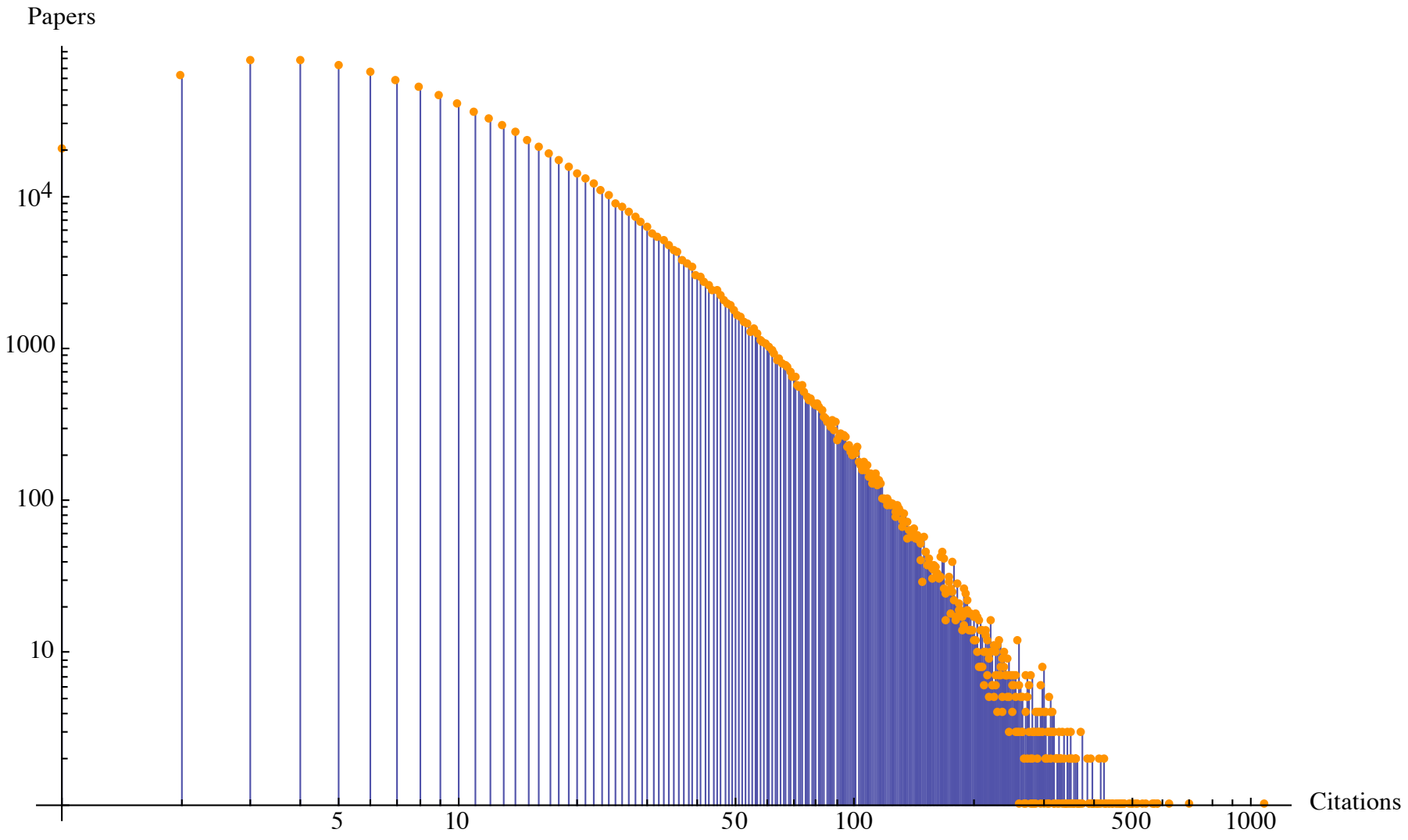
Esempio di generazione di dati sintetici

Citazioni

- Il numero di citazioni in uscita viene estratto casualmente con una distribuzione gaussiana.
- Ad ogni articolo viene assegnato un valore di qualità e il numero di citazioni che riceve viene calcolato sulla base di tale valore.
- Il numero totale delle citazioni in uscita è uguale a quello delle citazioni in entrata.
- L'anno di pubblicazione viene scelto casualmente all'interno di un intervallo definito nel programma.
- Vengono assegnate le citazioni, rispettando le distribuzioni prima generate, garantendo che:
  - un articolo non ne citi uno più giovane;
  - un articolo possa citare lo stesso articolo una sola volta.

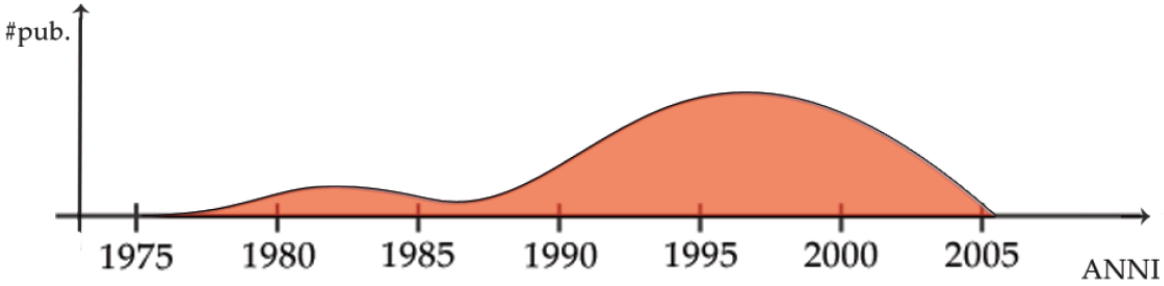
N.B. ora e nel seguito i grafici si riferiscono ad un set di dati sintetici con 1.000.000 articoli, 500.000 autori e 50.000 riviste.

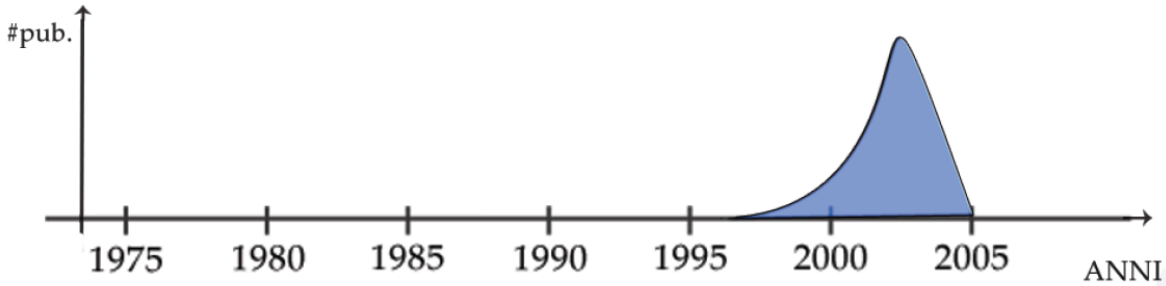

Papers
120 000
100 000
80 000
60 000
40 000
20 000
5
10
15
20
25
References


Papers
10^4
1000
100
10
5
10
50
100
500
1000
Citations

Autori

Ogni autore è caratterizzato da una distribuzione delle sue pubblicazioni negli anni: può avere un periodo di maggior produttività, l'attività di ricerca può essere più o meno concentrata negli anni.





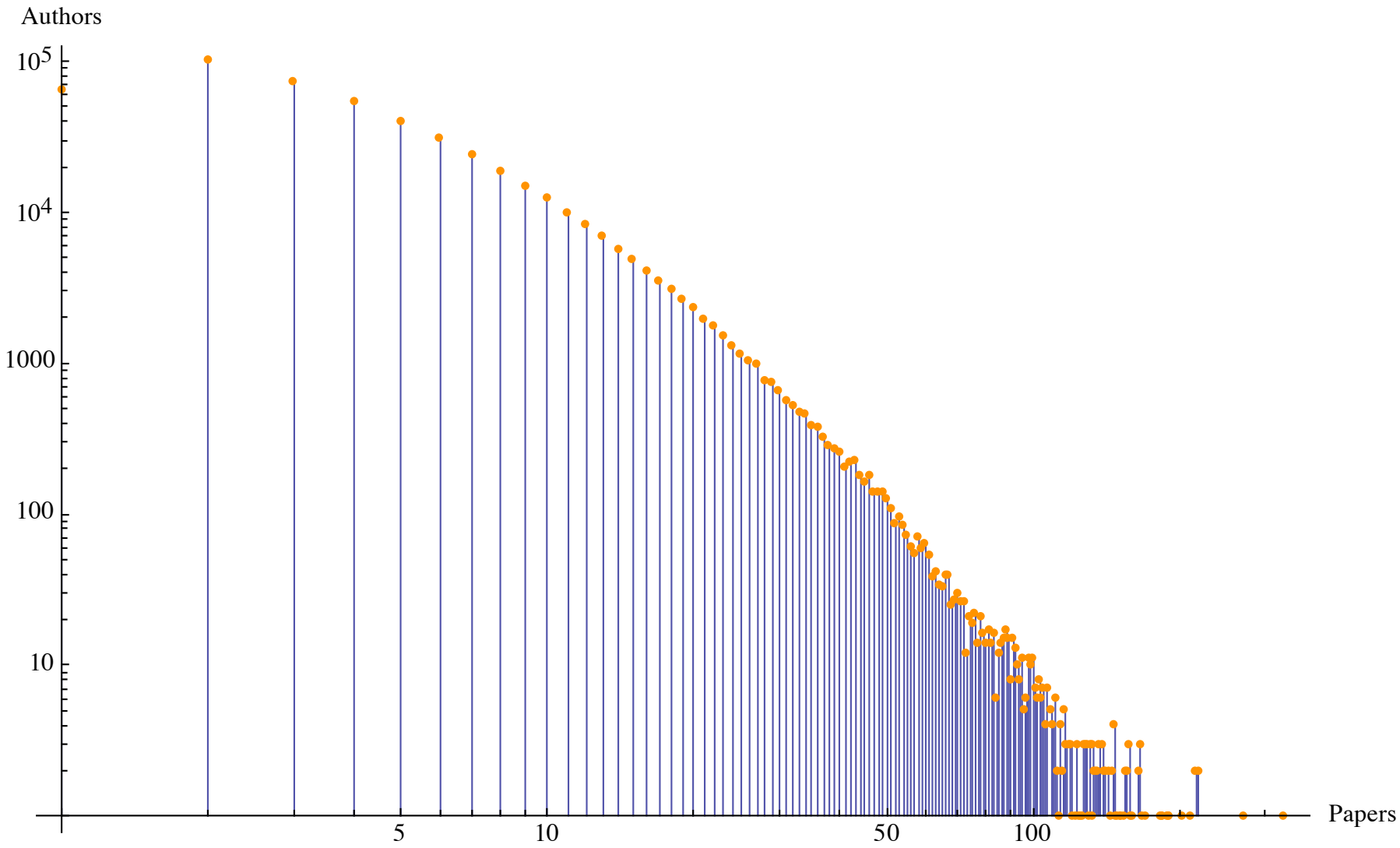
Questo comportamento può essere simulato con una distribuzione delle pubblicazioni negli anni che segue una gaussiana nella quale:

- il valore medio indica l'anno di maggior produttività;
- la deviazione standard rappresenta la dispersione dell'attività di ricerca.
-

La distribuzione delle pubblicazioni per autore (considerando l'insieme degli autori), segue approssimativamente una legge di Zipf.

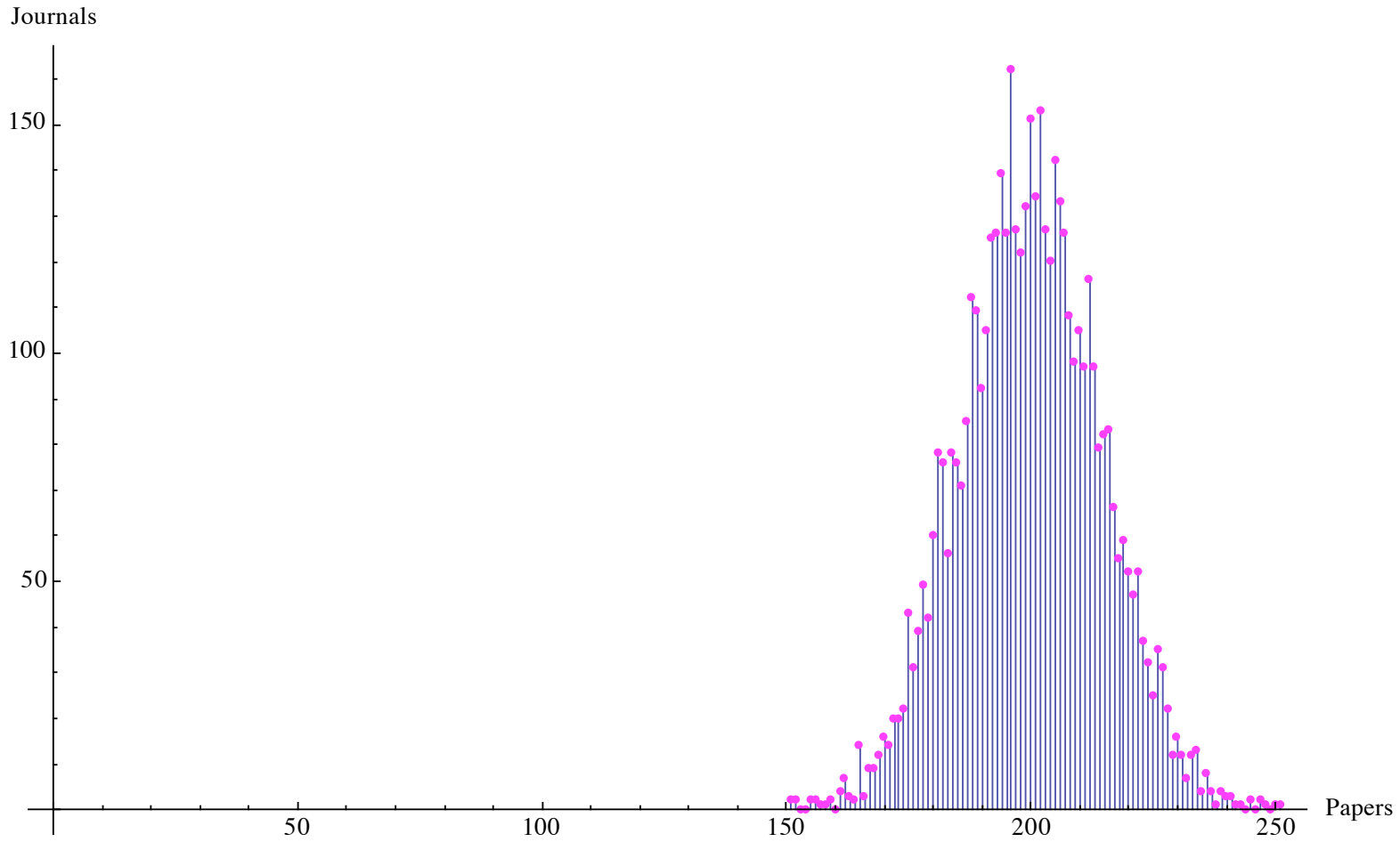
Viene costruita la matrice K garantendo che:

- ogni articolo venga pubblicato
- ogni autore pubblichi almeno un articolo

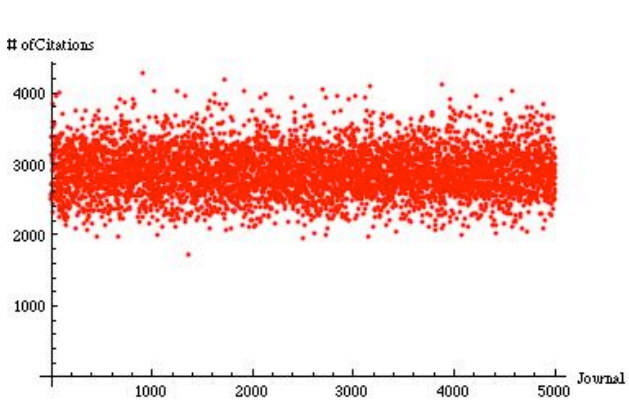

Authors
10^5
10^4
1000
100
10
5
10
50
100
Papers

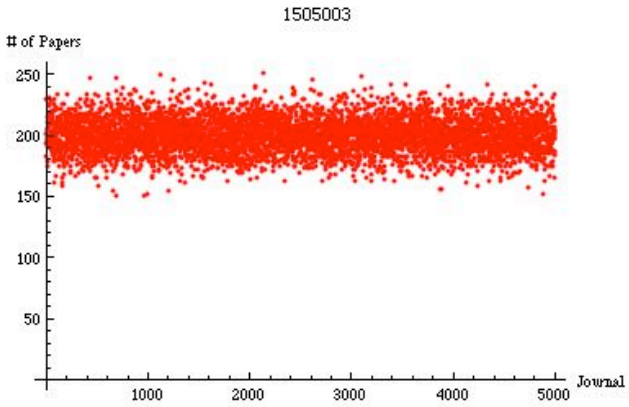
Riviste

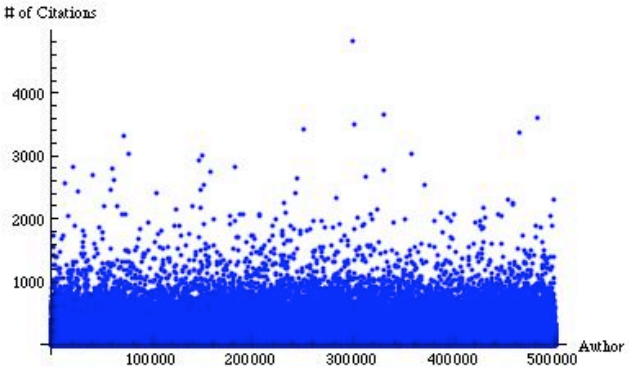
Per la matrice F, una rivista ha una distribuzione delle pubblicazioni per rivista che segue all'incirca una gaussiana.

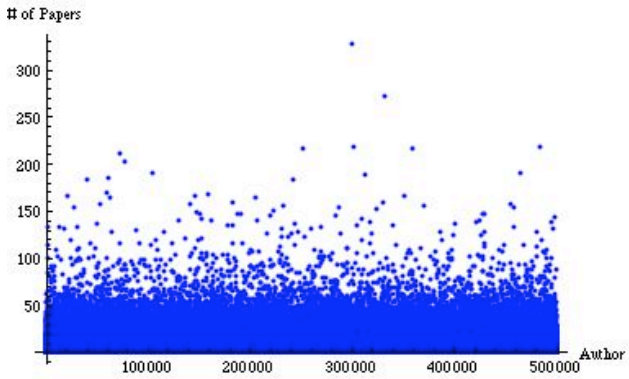


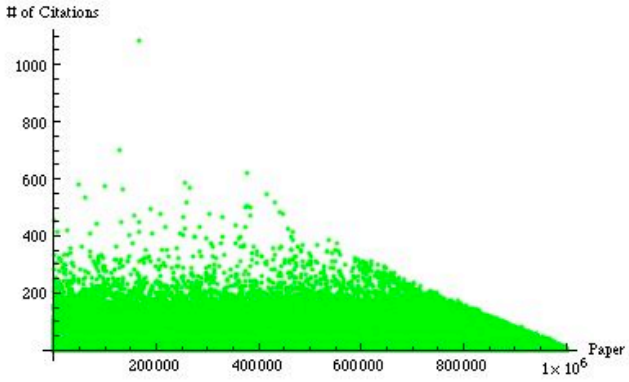
Visione d'insieme

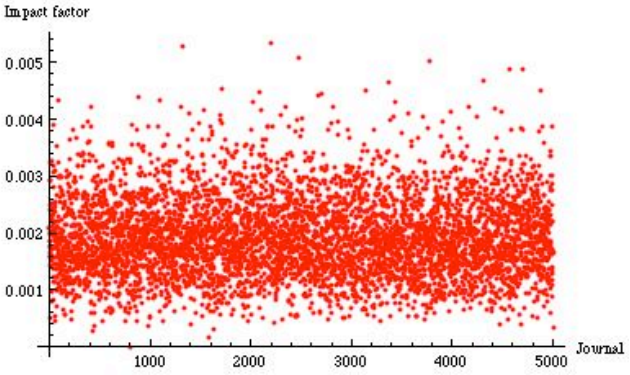




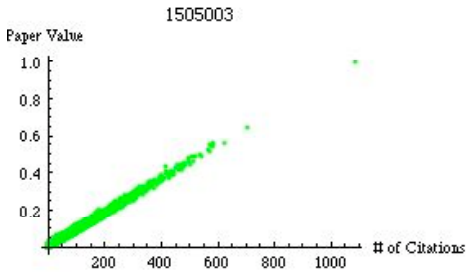


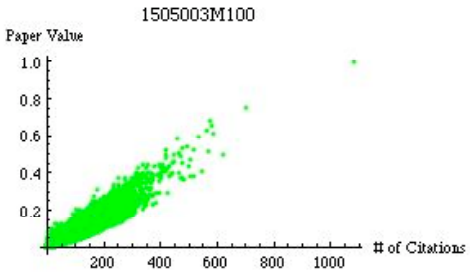


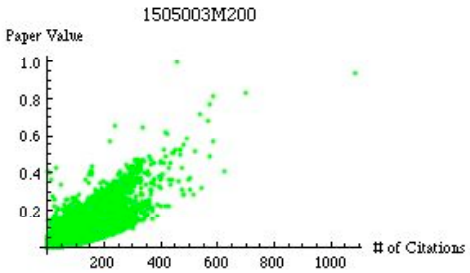


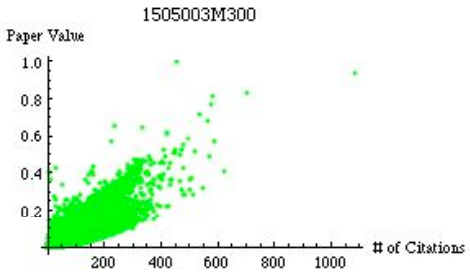


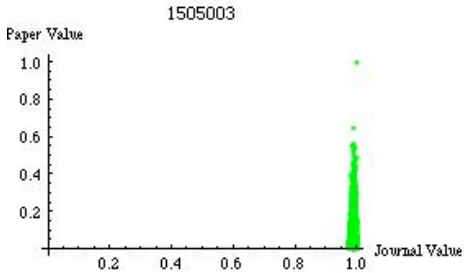
Esempi di risultati del calcolo del rango

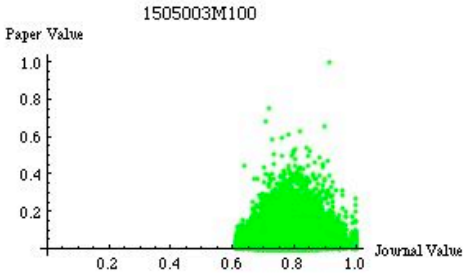


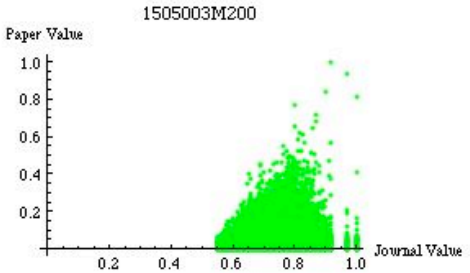


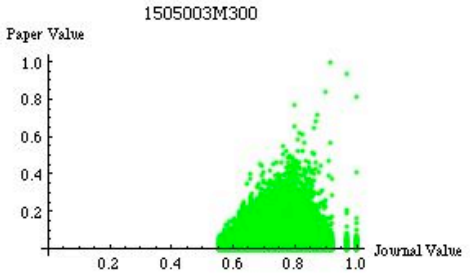


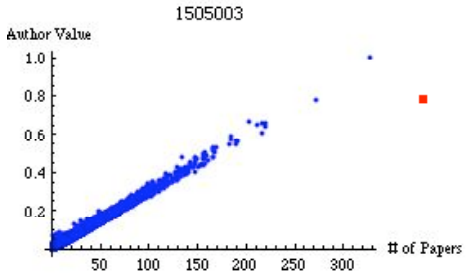


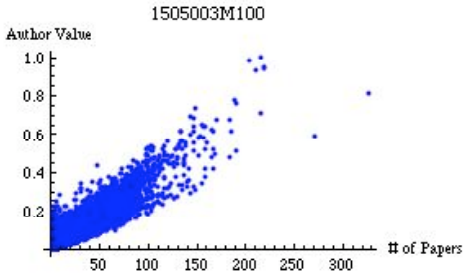


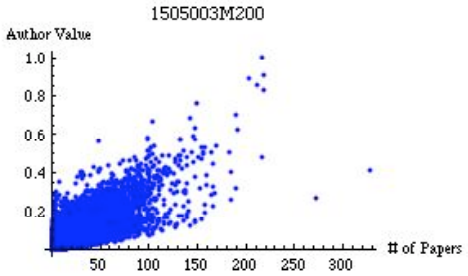


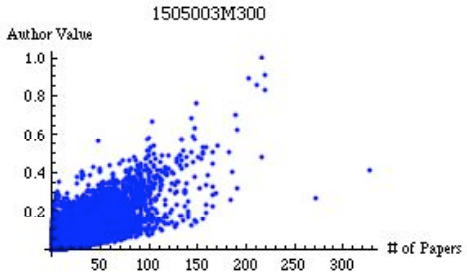


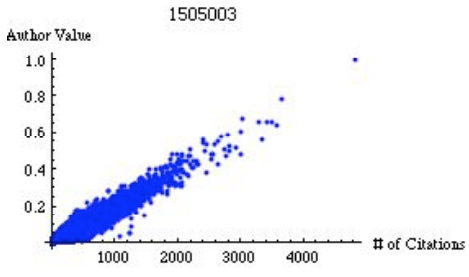


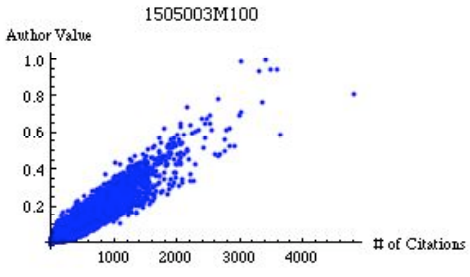

1505003
Author Value
1.0
0.8
0.6
0.4
0.2
50
100
150
200
250
300
# of Papers

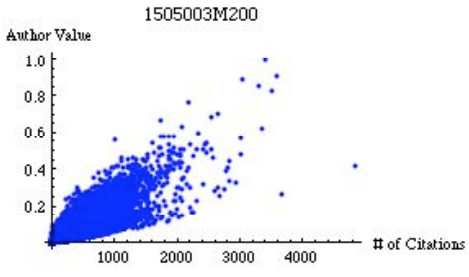

1505003M100
Author Value
1.0
0.8
0.6
0.4
0.2
50
100
150
200
250
300
# of Papers

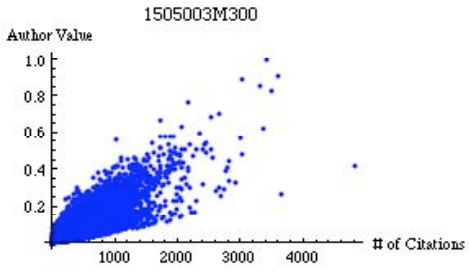

1505003M200
Author Value
1.0
0.8
0.6
0.4
0.2
50
100
150
200
250
300
# of Papers

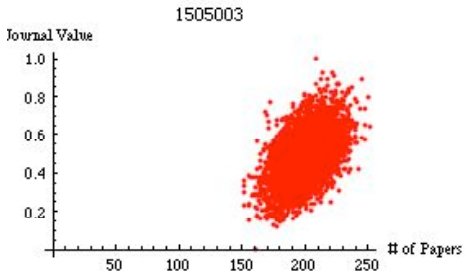

1505003M300
Author Value
1.0
0.8
0.6
0.4
0.2
50
100
150
200
250
300
# of Papers

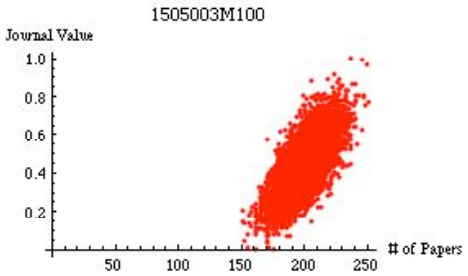

1505003
Author Value
1.0
0.8
0.6
0.4
0.2
1000
2000
3000
4000
# of Citations

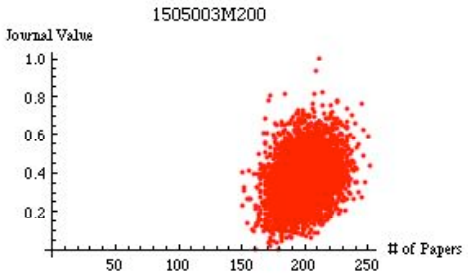

1505003M100
Author Value
1.0
0.8
0.6
0.4
0.2
1000
2000
3000
4000
# of Citations

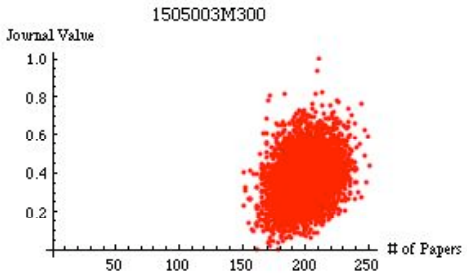

1505003M200
Author Value
1.0
0.8
0.6
0.4
0.2
1000
2000
3000
4000
# of Citations

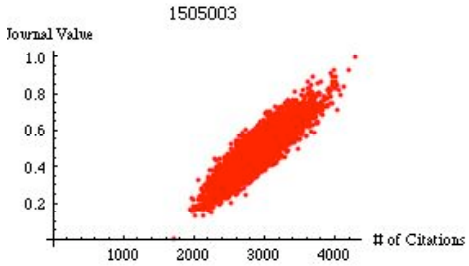

1505003M300
Author Value
1.0
0.8
0.6
0.4
0.2
1000
2000
3000
4000
# of Citations

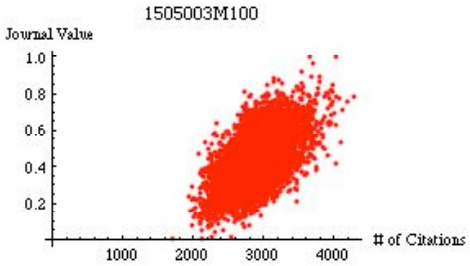

1505003
Journal Value
1.0
0.8
0.6
0.4
0.2
50
100
150
200
250
# of Papers

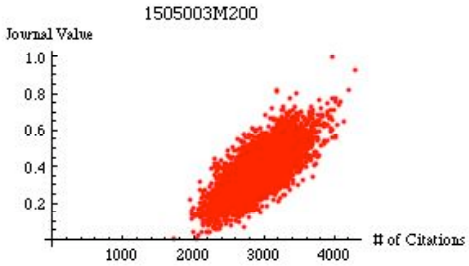

1505003M100
Journal Value
1.0
0.8
0.6
0.4
0.2
50
100
150
200
250
# of Papers

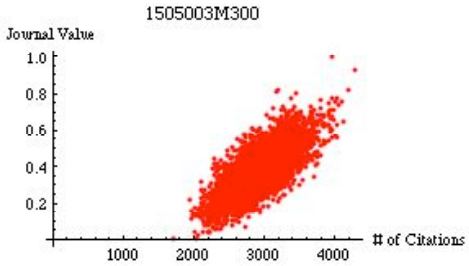

1505003M200
Journal Value
1.0
0.8
0.6
0.4
0.2
50
100
150
200
250
# of Papers

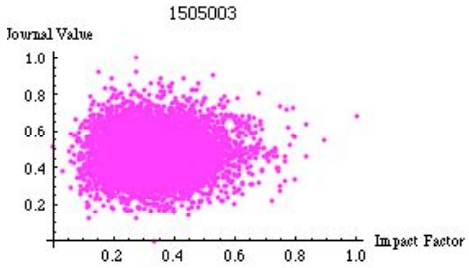

1505003M300
Journal Value
1.0
0.8
0.6
0.4
0.2
50
100
150
200
250
# of Papers

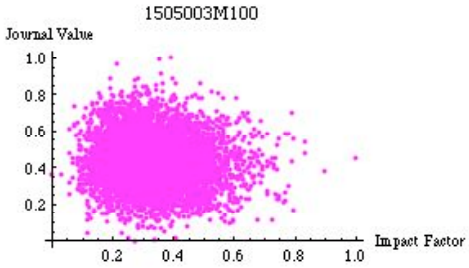

1505003
Journal Value
1.0
0.8
0.6
0.4
0.2
1000
2000
3000
4000
# of Citations

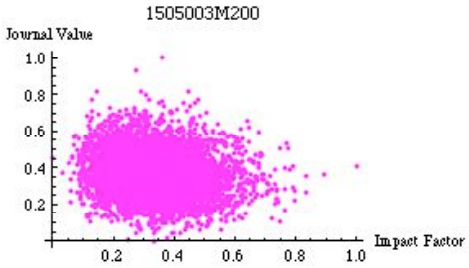

1505003M100
Journal Value
1.0
0.8
0.6
0.4
0.2
1000
2000
3000
4000
# of Citations

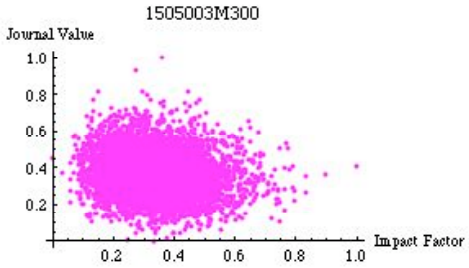

1505003M200
Journal Value
1.0
0.8
0.6
0.4
0.2
1000
2000
3000
4000
# of Citations


1505003M300
Journal Value
1.0
0.8
0.6
0.4
0.2
1000
2000
3000
4000
# of Citations


1505003
Journal Value
1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0
Impact Factor


1505003M100
Journal Value
1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0
Impact Factor


1505003M200
Journal Value
1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0
Impact Factor


1505003M300
Journal Value
1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0
Impact Factor