Nanotechnology Citation Analysis: Dynamics Field Evolution

Juan D. Rogers
School of Public Policy
Georgia Institute of Technology

Presentation Outline

- Citations in evaluation of research
- Nanotechnology citations data
- Cumulative citation analysis
  - Density estimation using recent statistical techniques
- Results for two cohorts

Citations in Evaluation

- Acknowledged measure of impact
- Difficult and confusing use of citations in actual evaluation
- Comparison of evaluation levels must take into account population characteristics
  - What does an above average citation rate mean?
- Background of Normal distribution confuses measures
- Many measures emphasize early citation
  - Impact factor looks at a window of two years

Citation Networks are Complex Systems

- Nodes: peer-reviewed papers
  - They don't change once in place
- Edges: citations
  - In-degree: edges "aimed" at previous papers
- Networks derive their properties from interactions of its components
- Exhibit power law distributions
  - $F(x) = cx^{-\alpha}$ for $x \geq x_{min}$
  - α is a "scaling" factor: $F(ax) = ca^{-\alpha}x^{-\alpha}=d\ x^{-\alpha}$: scale free
  - $F(x)$ is linear on a log-log scale: $\log(F) = \log(c) - \alpha.\log(x)$
- Estimated with iterative ML with K-S technique

Clauset, A., Shalizi, C., Newman, M. 2007. "Power-Law Distributions in Empirical Data", Physica A 386, pp. 414-438.

Consequences for citation indicators

- Some power laws have “fat tails”:
  - Higher than expected probability of extreme events
- For ranges of the scaling factor have infinite variance and mean
  - $1 < \alpha \leq 2$, has infinite variance and mean
  - $2 < \alpha < 3$, has infinite variance and finite mean
  - $\alpha \geq 3$, has finite variance and mean
- Central limit theorem doesn’t work with infinite variance
  - Estimation of moments from samples is inconsistent
  - Measures of central tendency are uninformative
- Reflects the highly interdependent nature of items in the network
  - Comparison of individual items to each other is not germane to the dynamics of the network

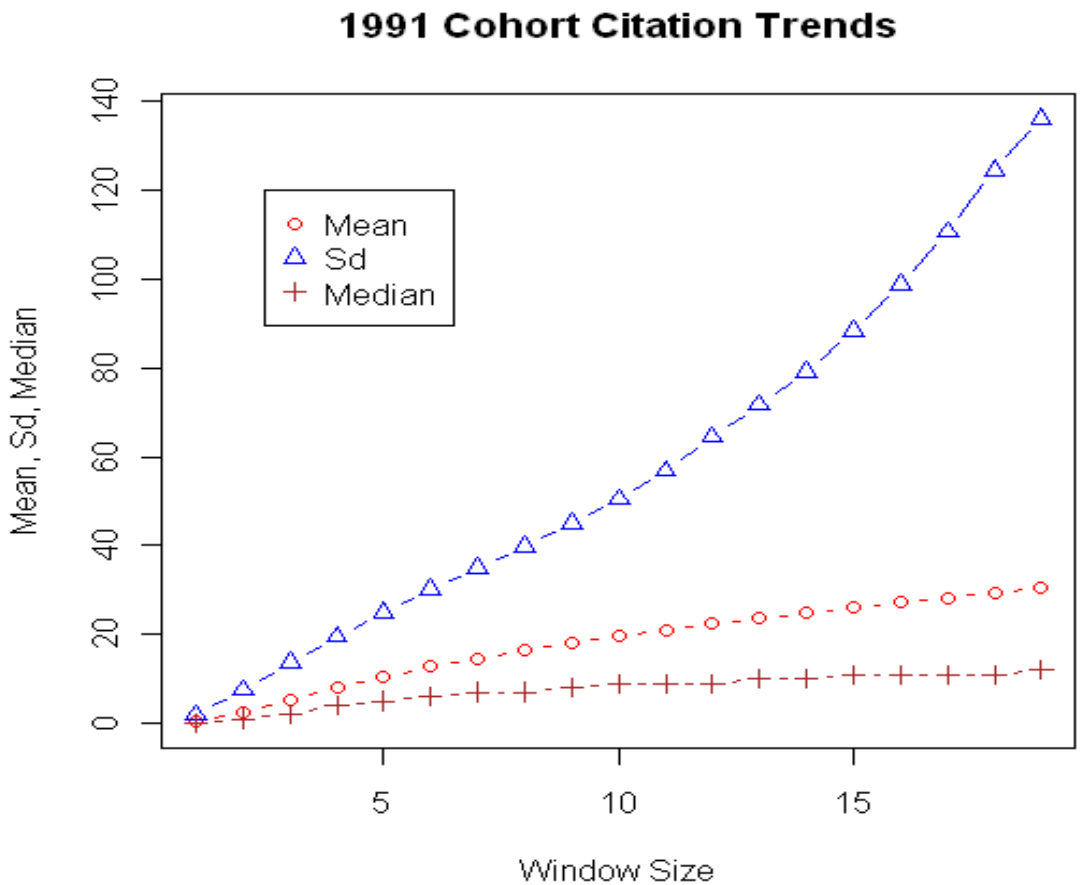
Nanotechnology Publications Data

- Field defined by search strategy documented in:

Porter, A., Youtie, J., Shapira, P., Schoeneck, D. 2008. "Refining Search Terms for Nanotechnology," Journal of Nanoparticle Research, 10(5):715-728.

- Well established definition of field boundaries
- Two early cohorts of emerging field
  - 1991: 7,533 papers
  - 1992: 9,664 papers
- Yearly citations for each individual paper

Cumulative citation descriptive statistics (1991)



Average:

10.4; 19.65; 26.03; 30.46

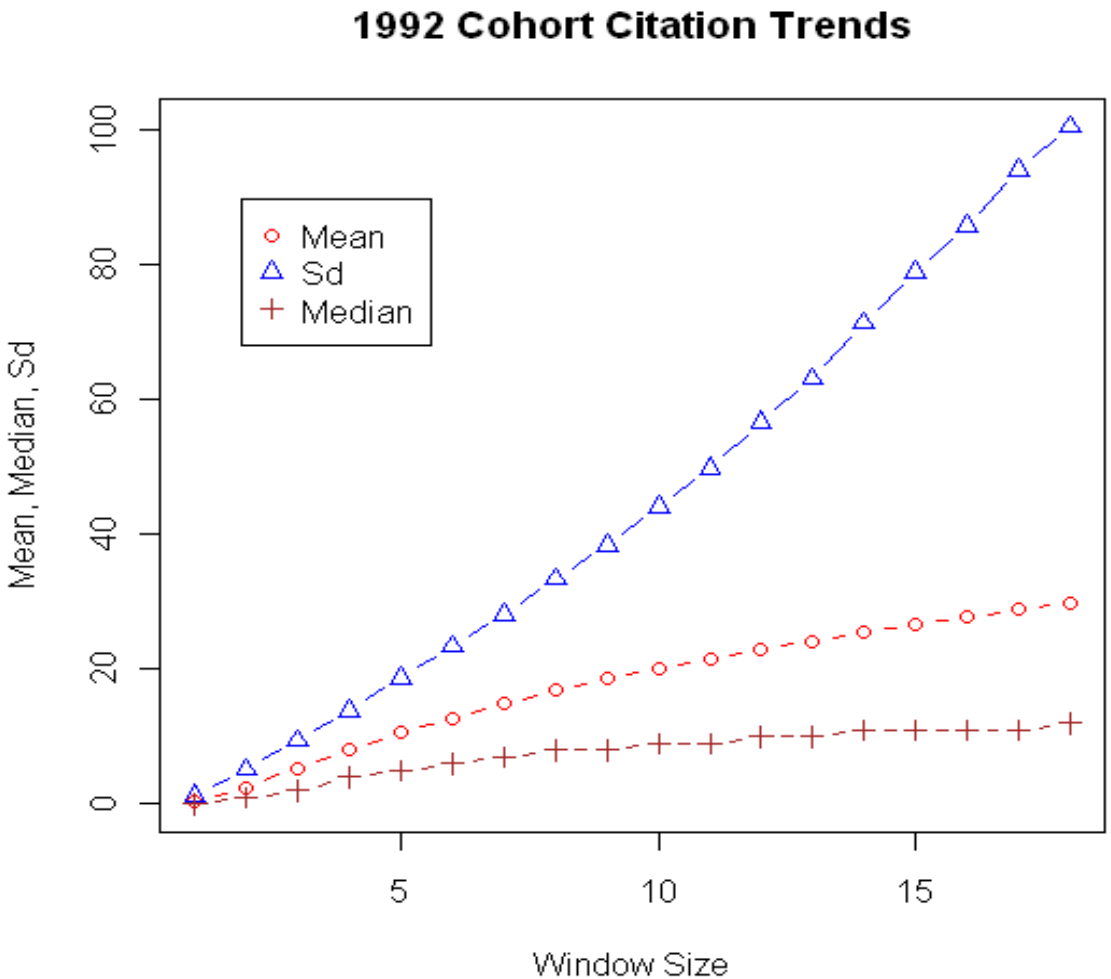
Median:

5; 9; 11; 12

St. Dev:

24.88; 50.5; 88.4; 136.22

Cumulative citation descriptive statistics (1992)



Average:

10.51; 20.16; 26.72; 29.83

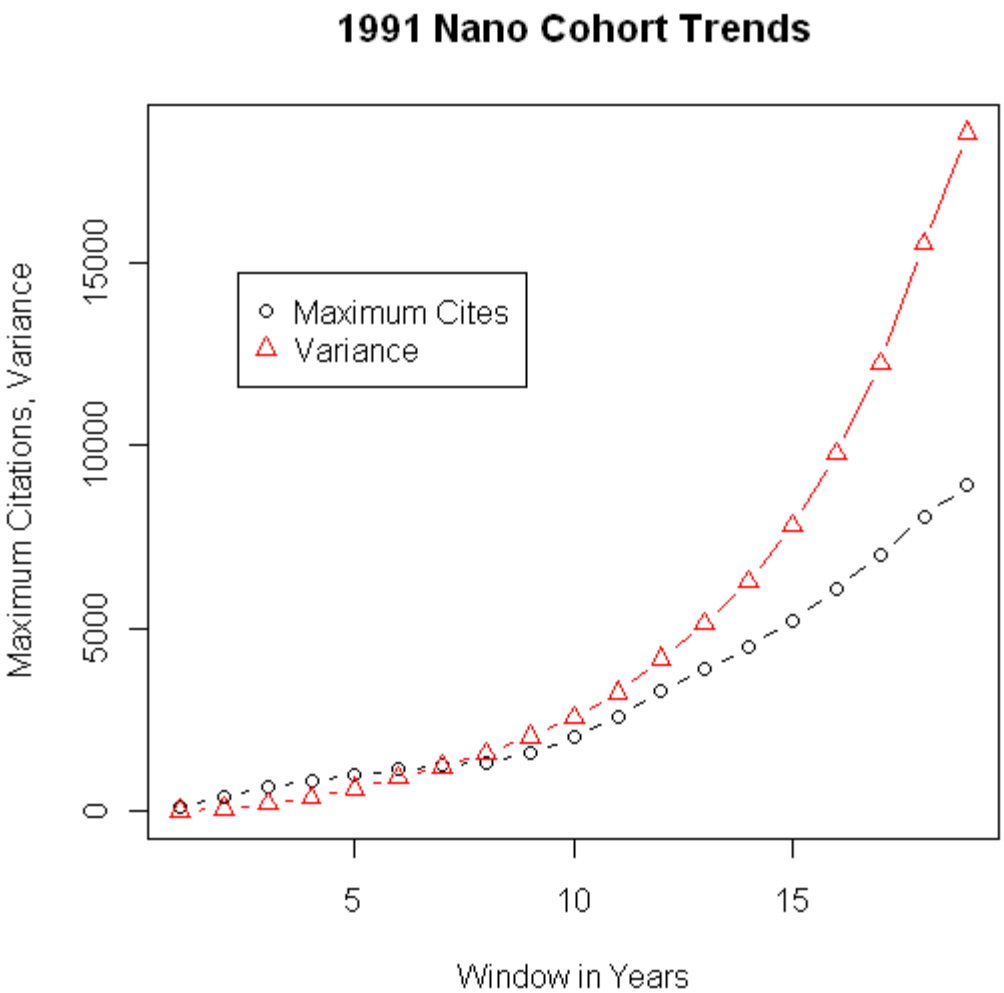
Median:

5; 9; 11; 12

St. Dev:

18.75; 44.13; 78.97; 100.60

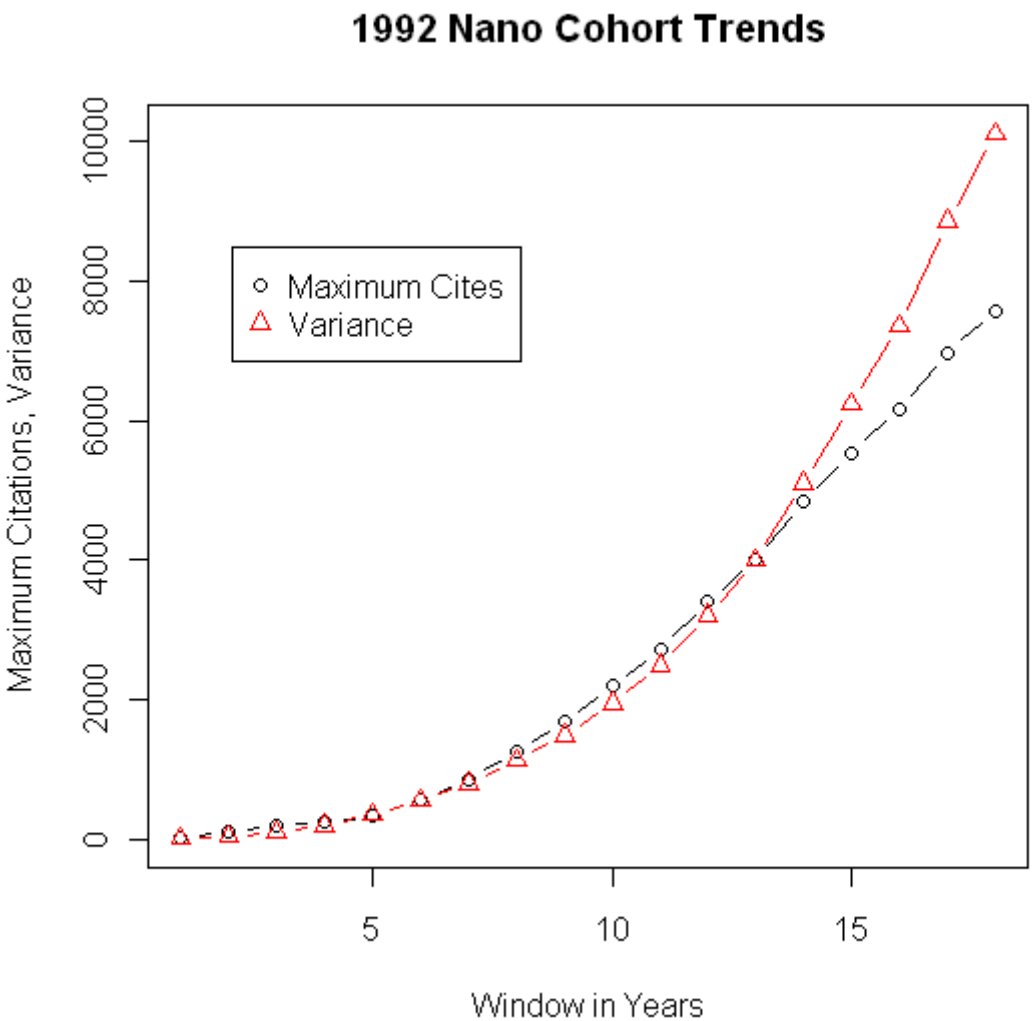
Trends for maximum cumulative citations and variance (1991)



Maximum cumulative Citations:
990; 2009; 5196; 8893

Variance:
618.82; 2550.38; 7815.37; 18555.47

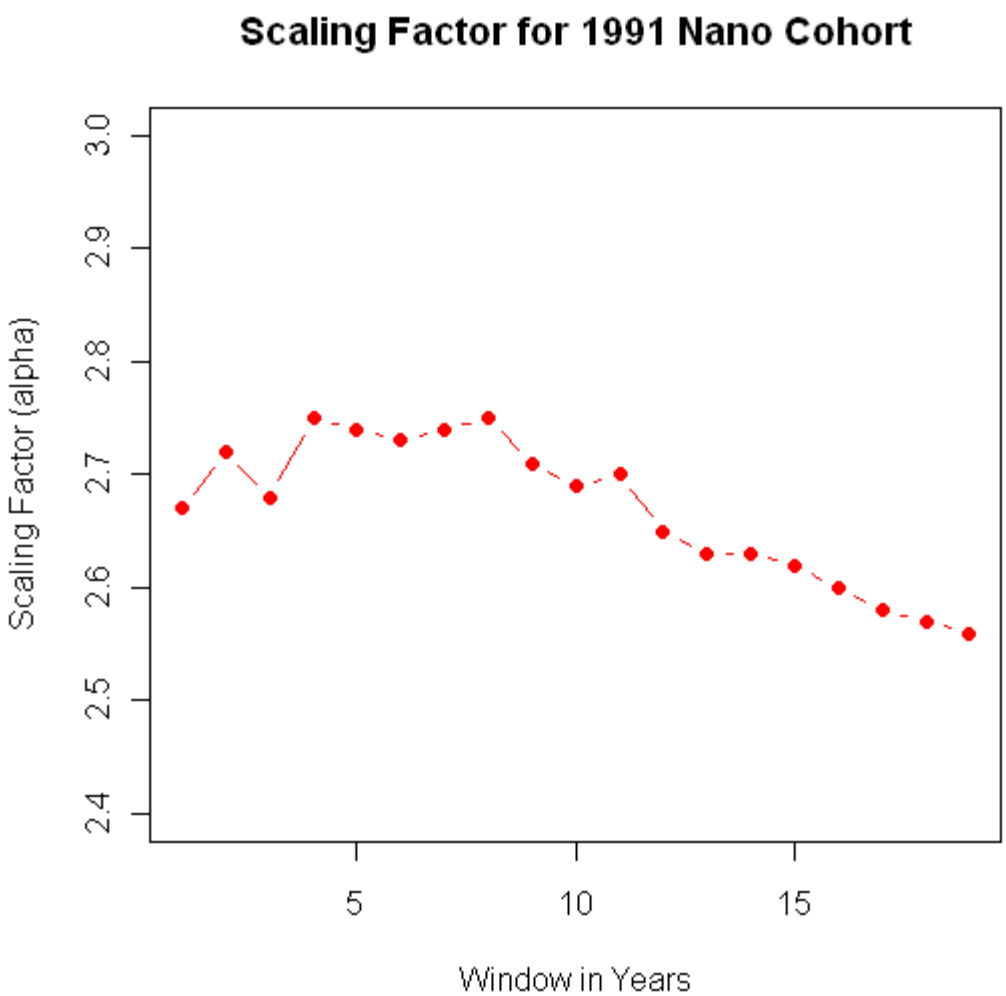
Trends for maximum cumulative citations and variance (1992)



Maximum cumulative Citations:
320; 2188; 5541; 7576

Variance:
351.38; 1947.57; 6245.67; 10120.18

Cumulative Citations Density Estimation (1991)

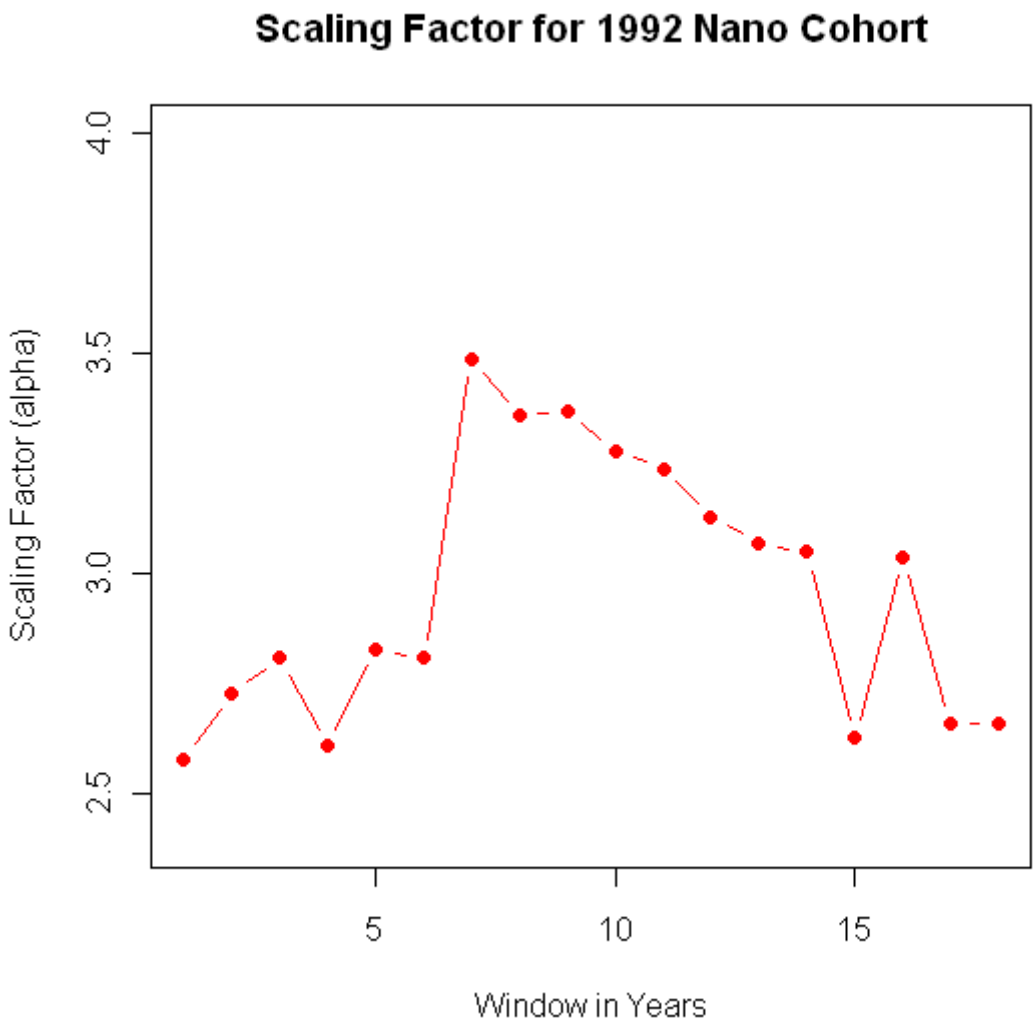


Power Law Scaling Factor (Alpha):
2.74; 2.69; 2.62; 2.56

Lower bound of power law (Xmin):
28; 49; 61; 60

Number of papers in tail:
572; 635; 673; 836

Cumulative Citations Density Estimation (1992)

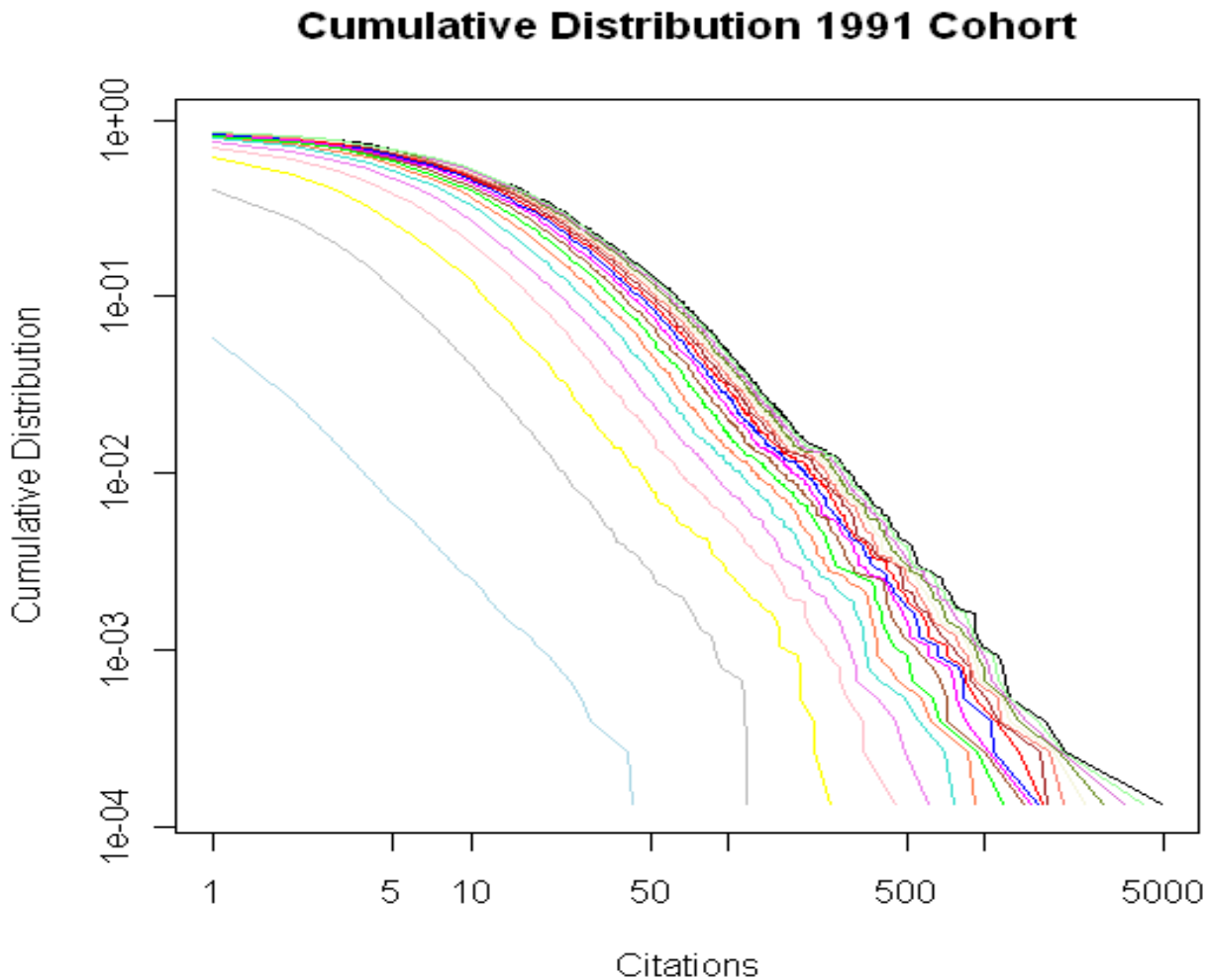


Power Law Scaling Factor (Alpha):
2.83; 3.28; 2.63; 2.66

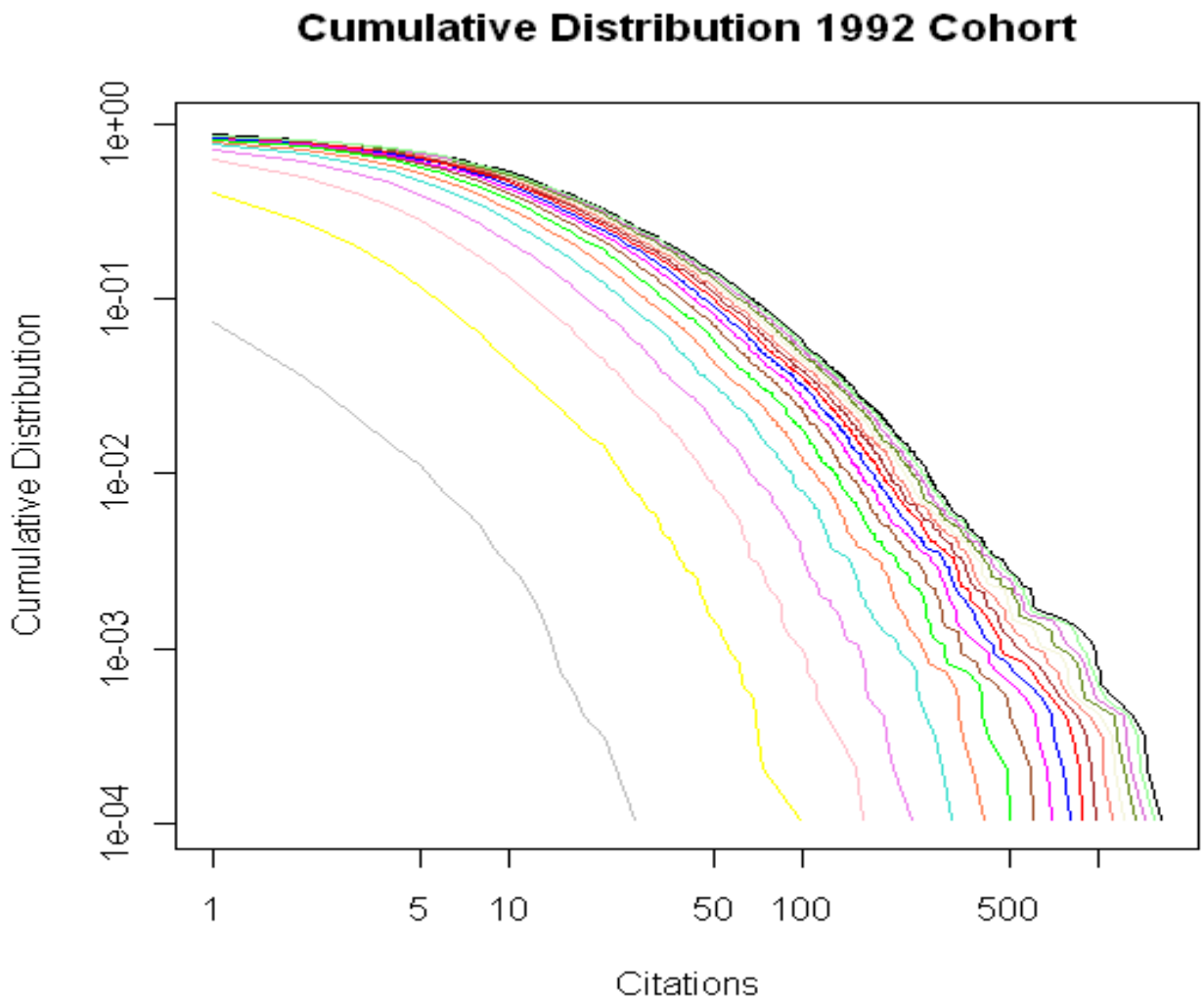
Lower bound of power law (Xmin):
31; 127; 64; 82

Number of papers in tail:
929; 196, 925; 753

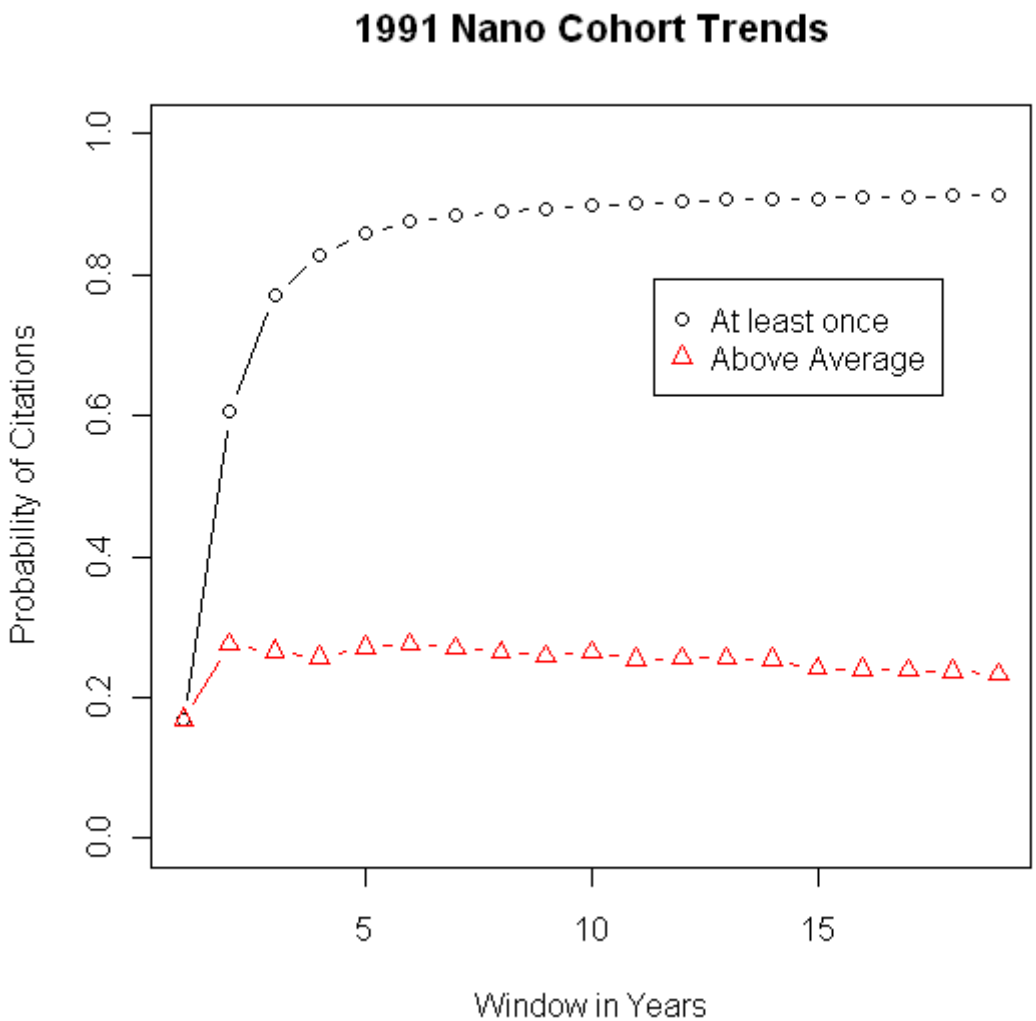
Cumulative Distribution by Window Size (1991)



Cumulative Distribution by Window Size (1992)



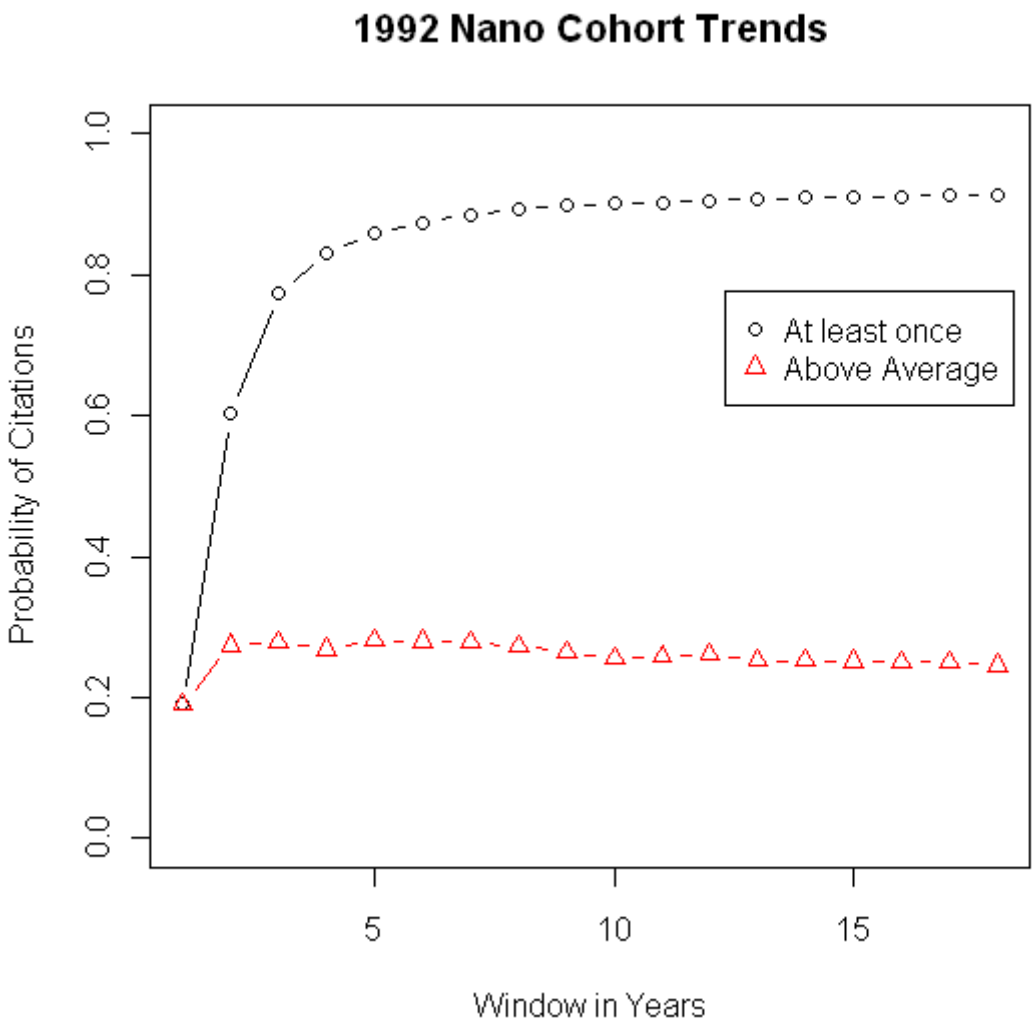
Probability of being cited (1991)



Probability of being cited:
.828; .898; .909; .913

Probability of average citation rate:
.271; .264; .242; .234

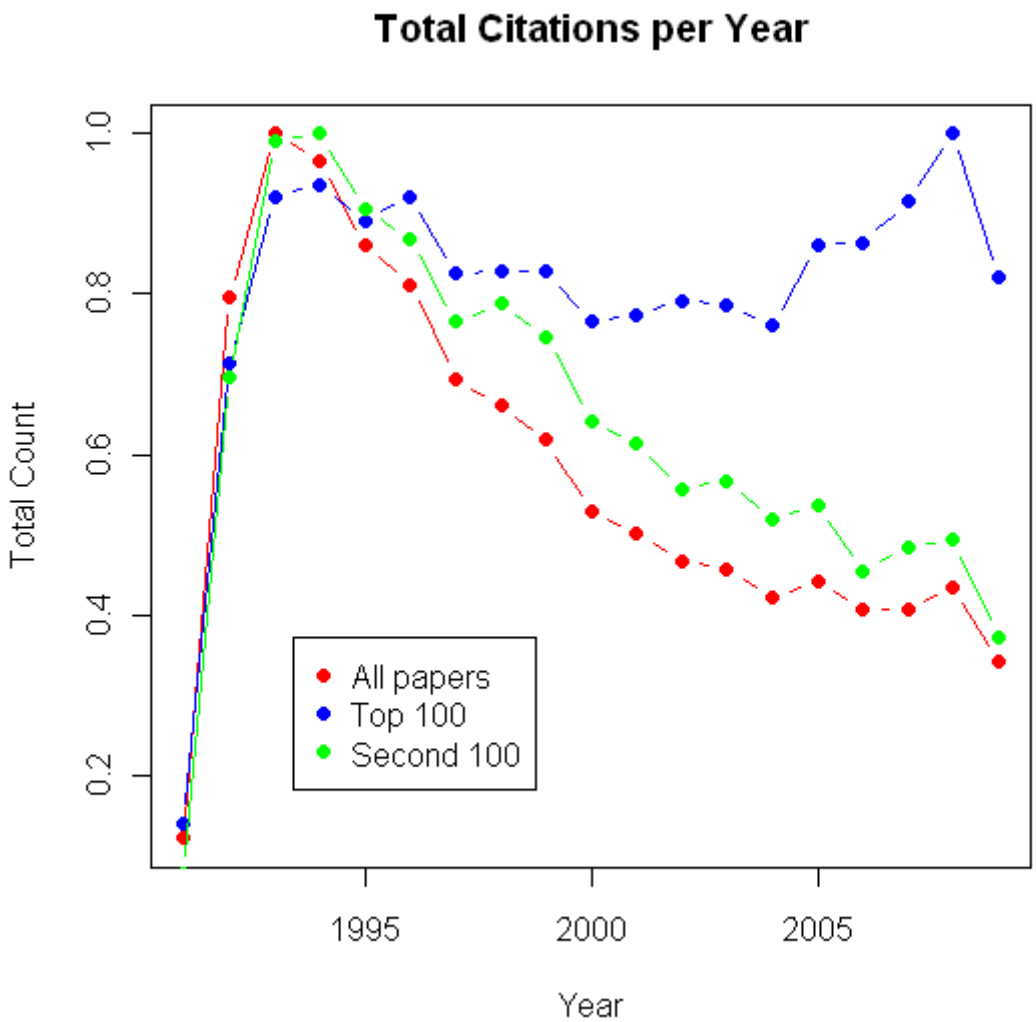
Probability of being cited (1992)



Probability of being cited:
.859; .900; .910; .913

Probability of average citation rate:
.281; .256; .252; .246

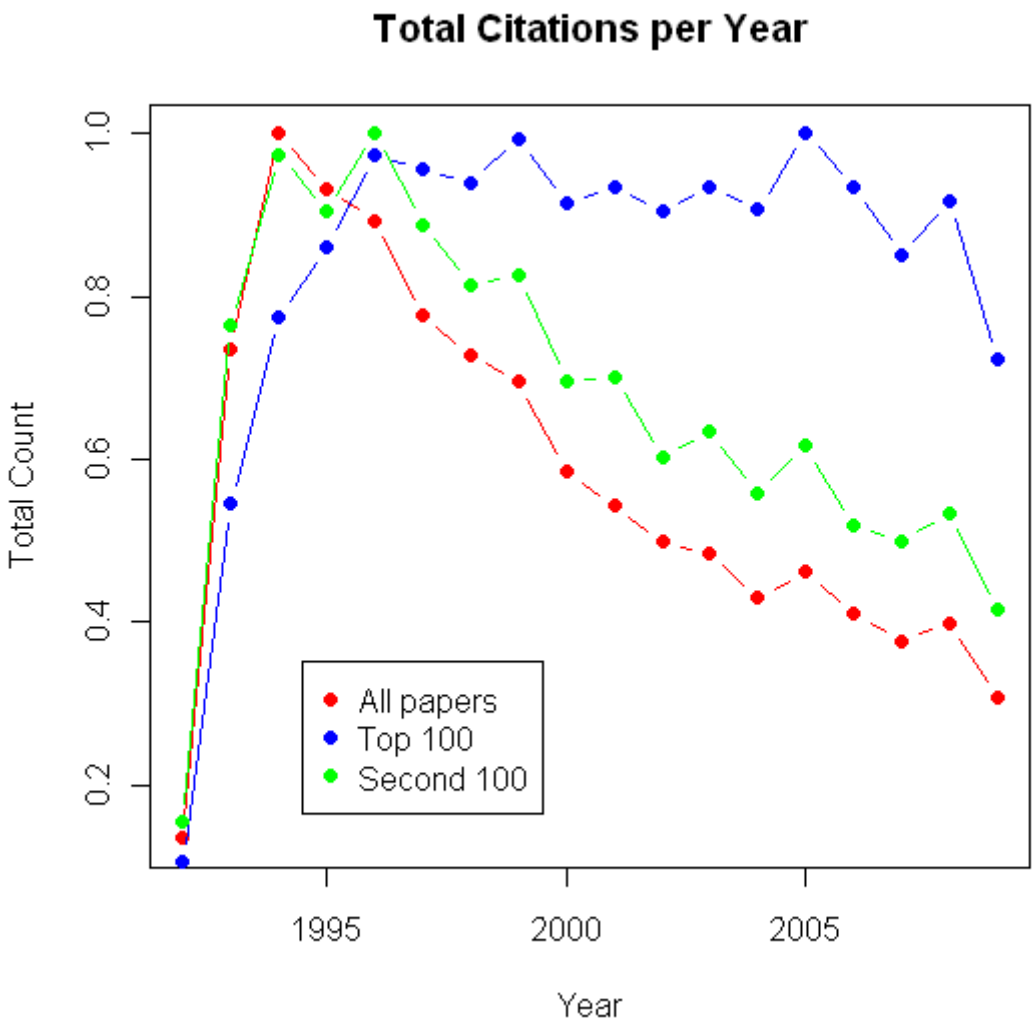
Evolution of Yearly Citations (1991)



Maximum per group:

Top 100:	4016
Second 100:	1462
All:	20974

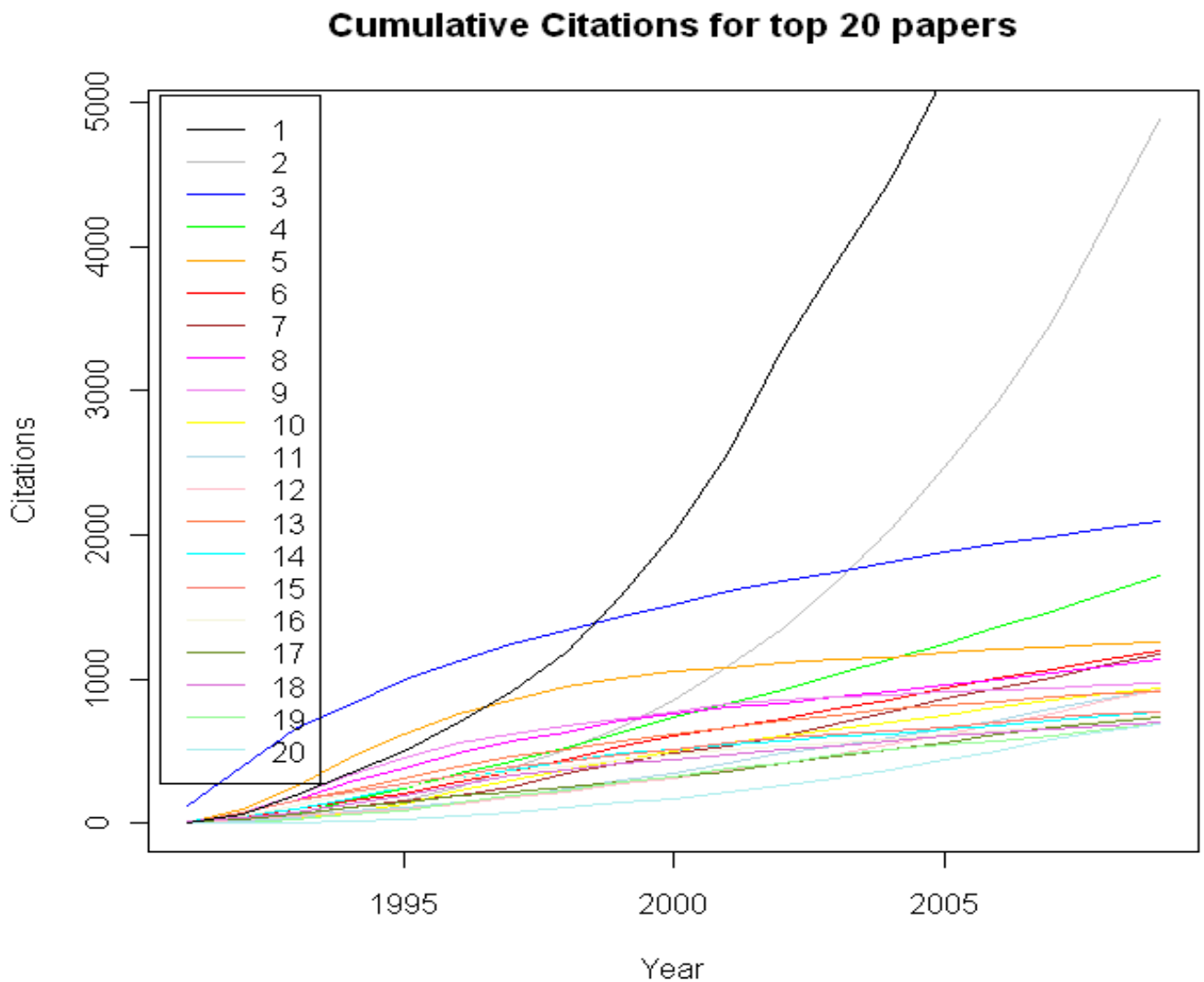
Evolution of Yearly Citations (1992)



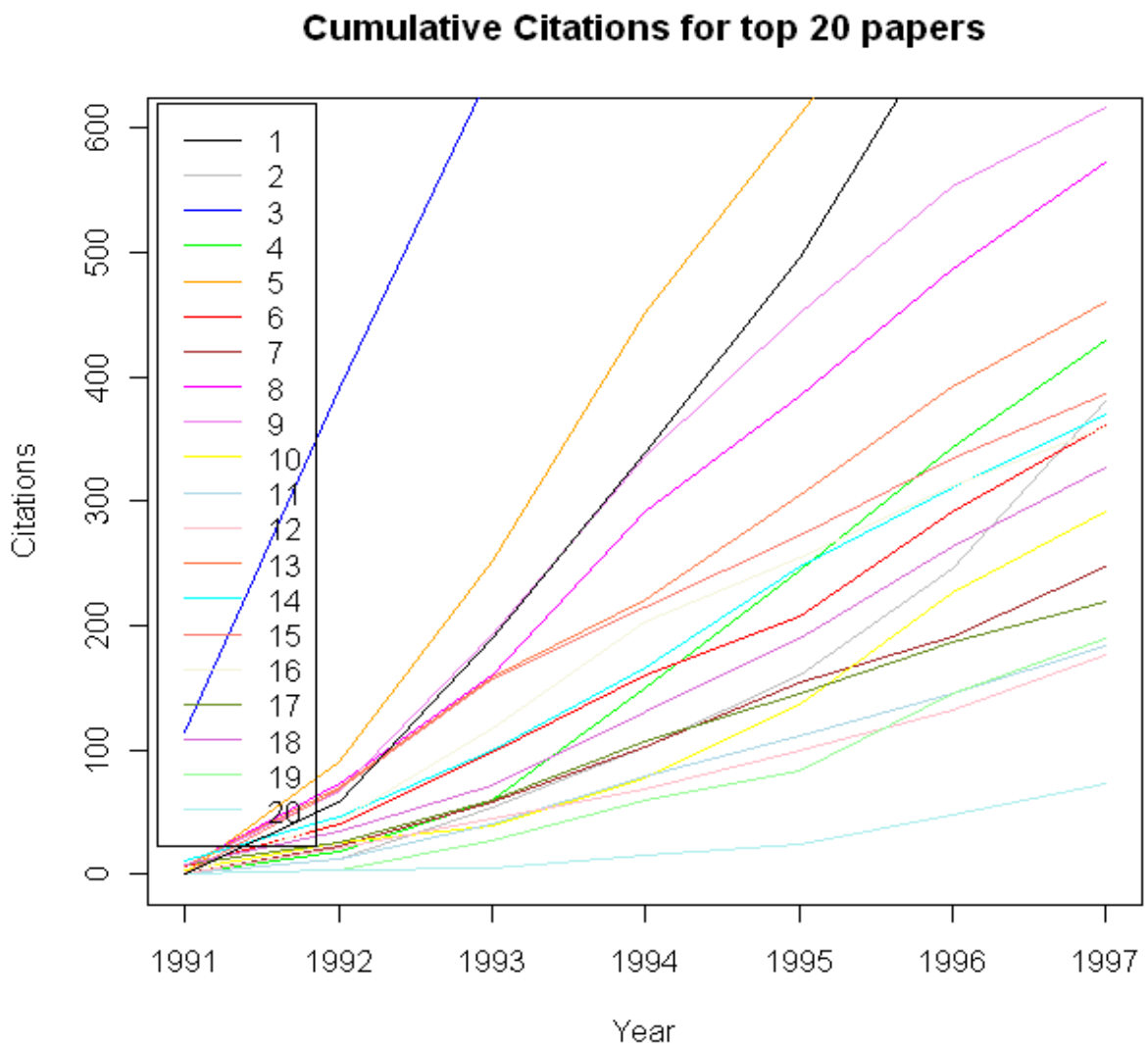
Maximum per group:

Top 100:	3607
Second 100:	1822
All:	27753

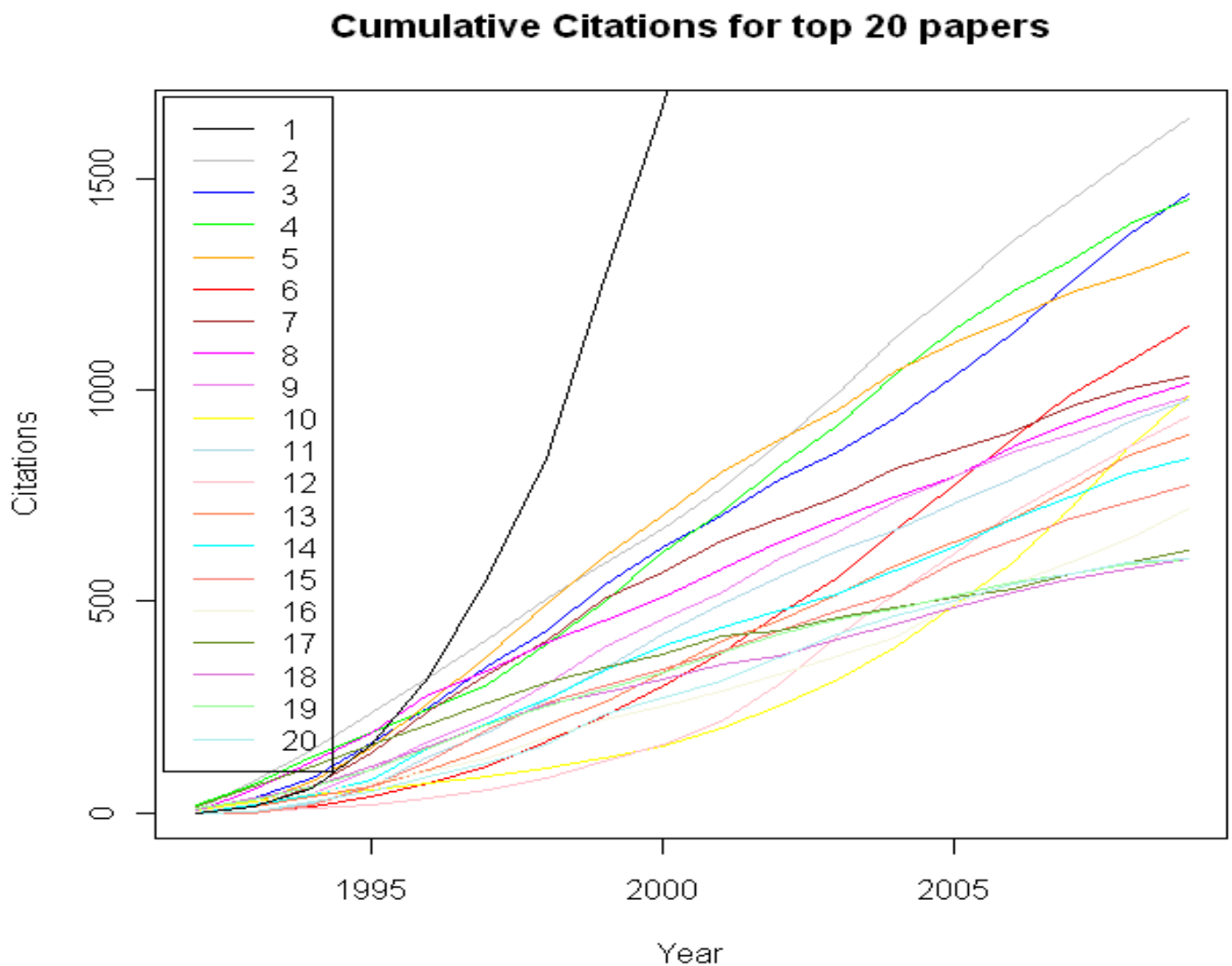
Evolution of Cumulative Citations Top Papers (1991)



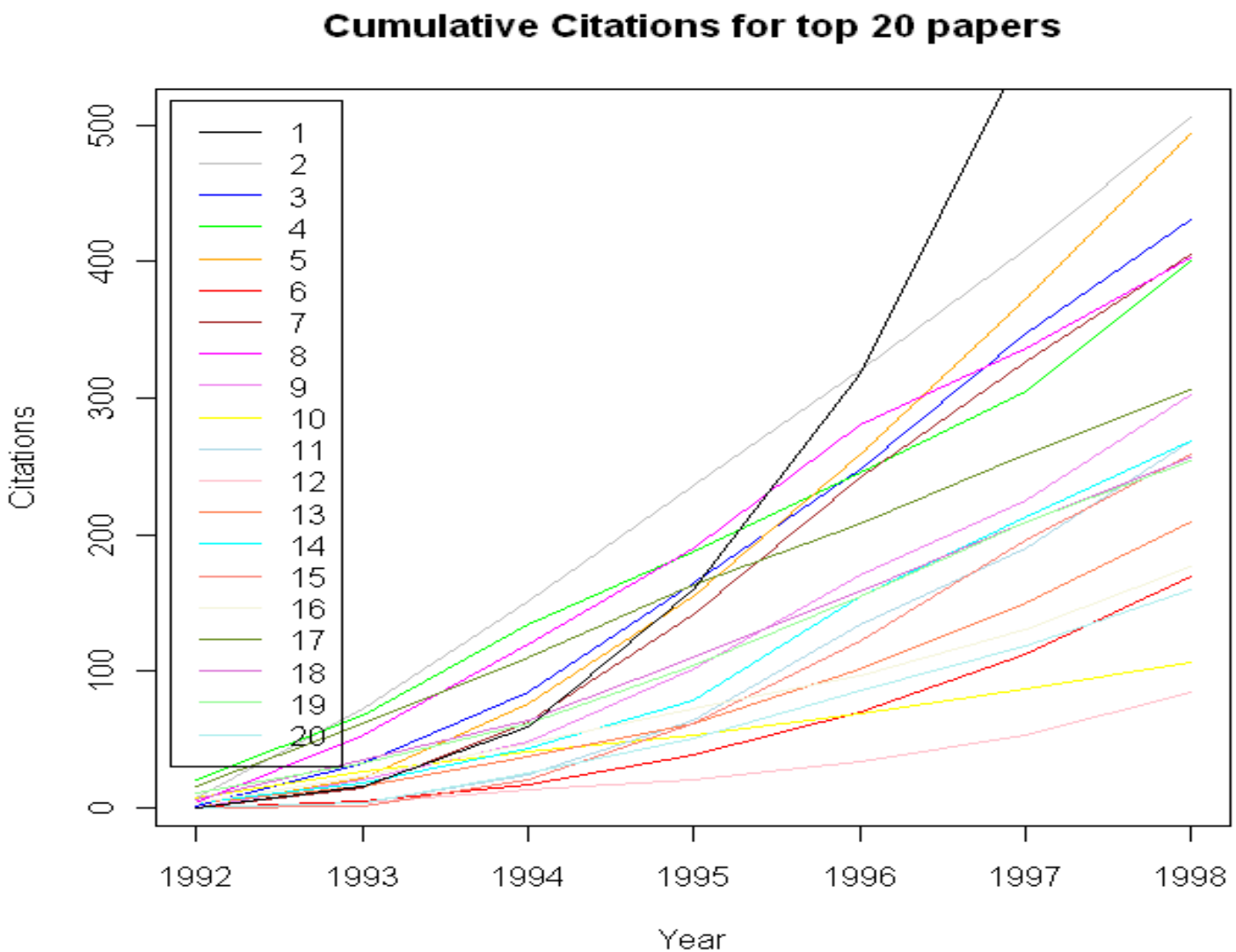
Detail of first few years (1991)



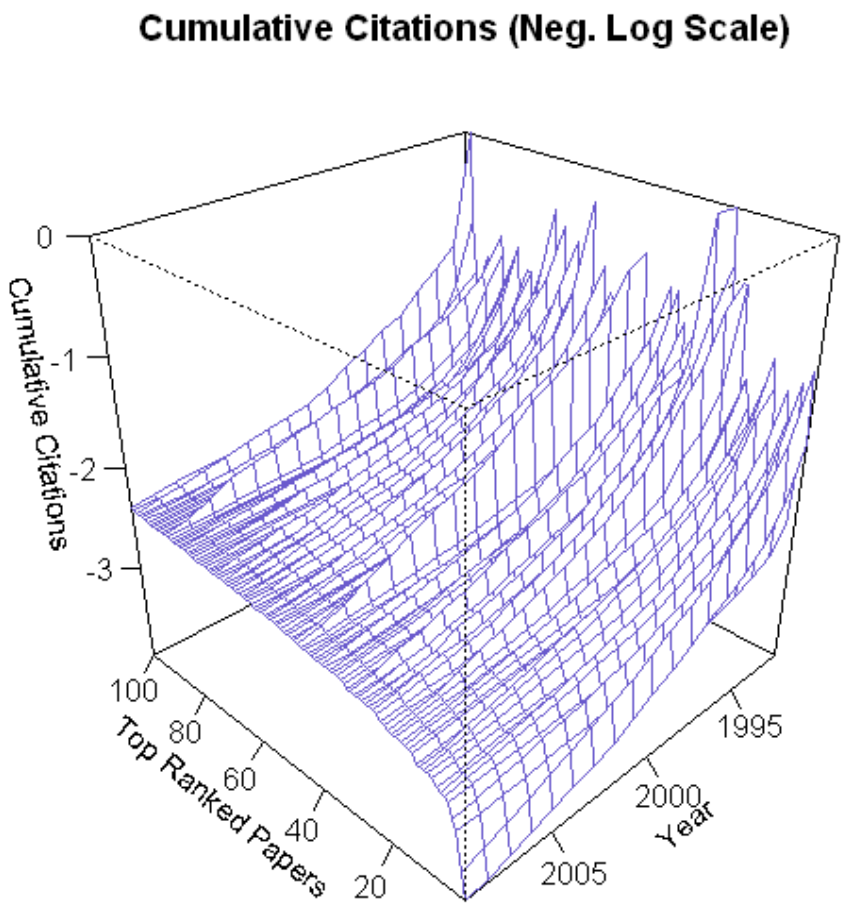
Evolution of Cumulative Citations Top Papers (1992)



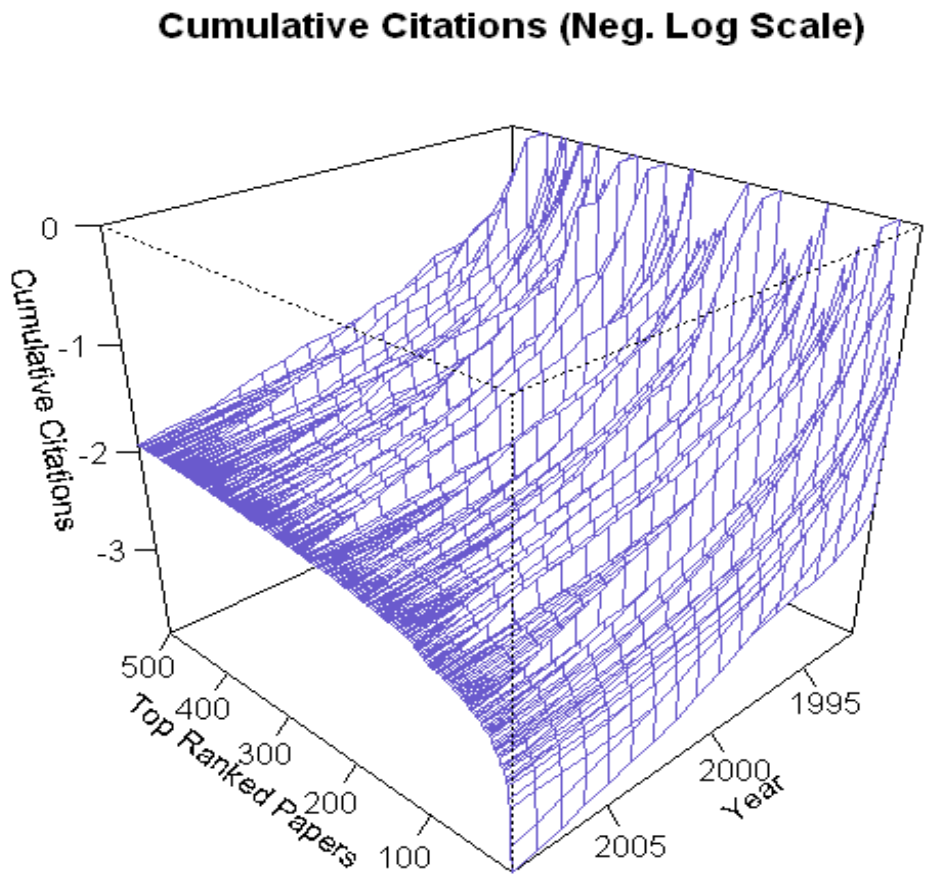
Details of first few years (1992)



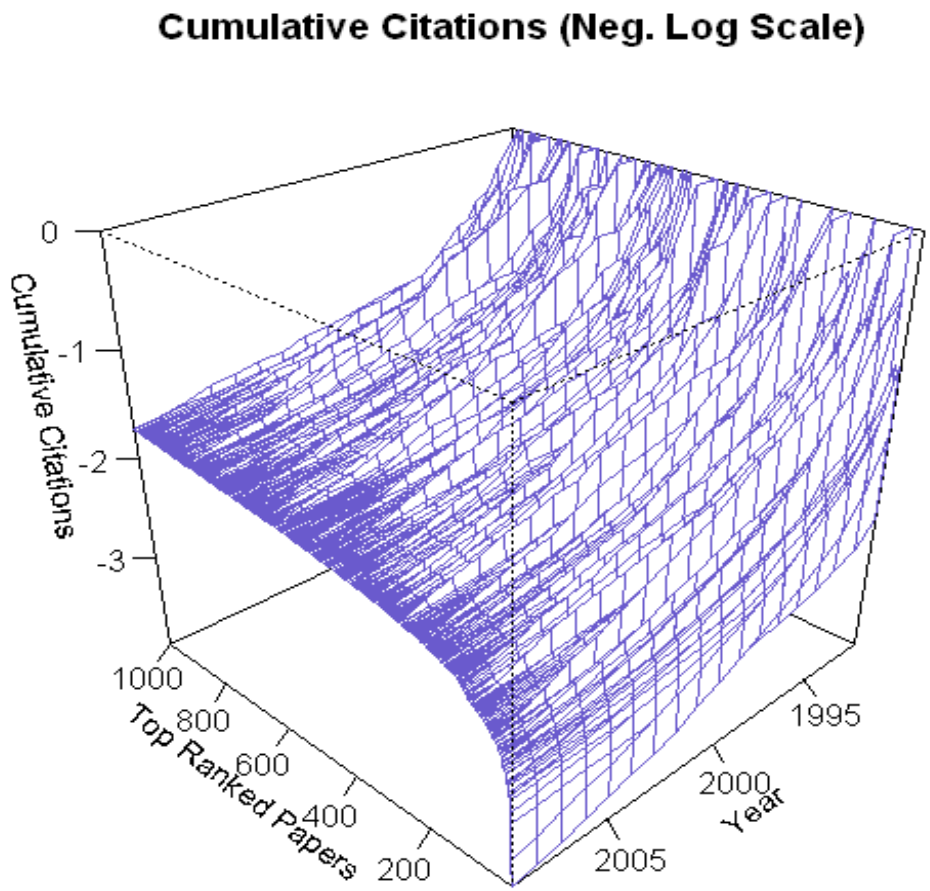
3D Representation of Cumulative Citations (Top 100 1991)



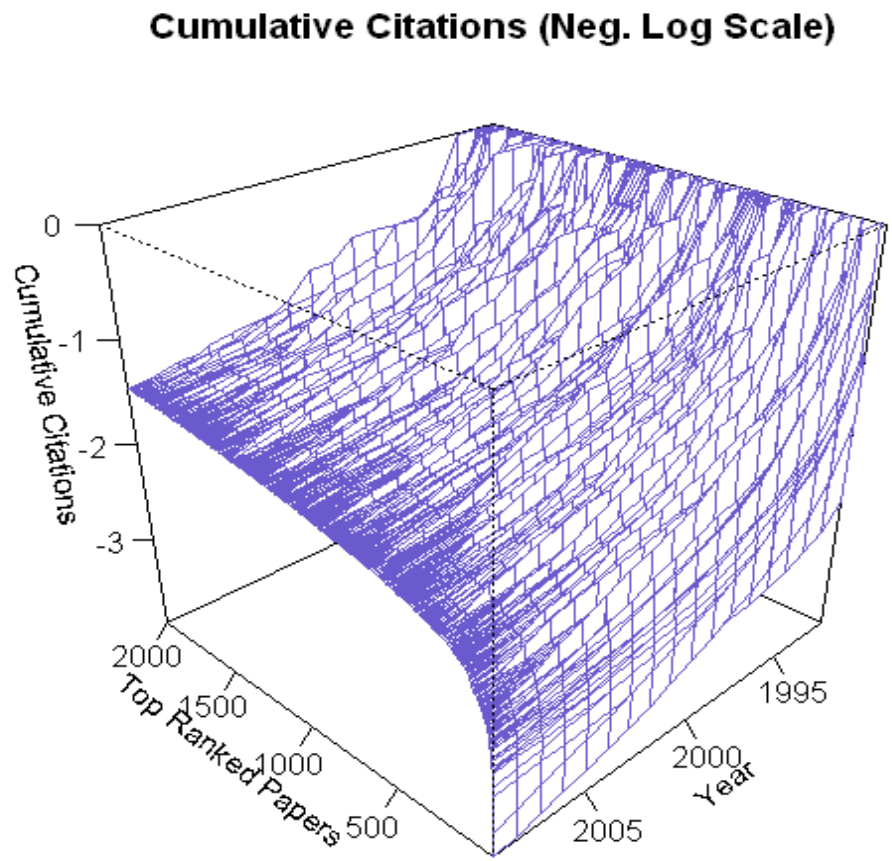
3D Representation of Cumulative Citations (Top 500 1991)



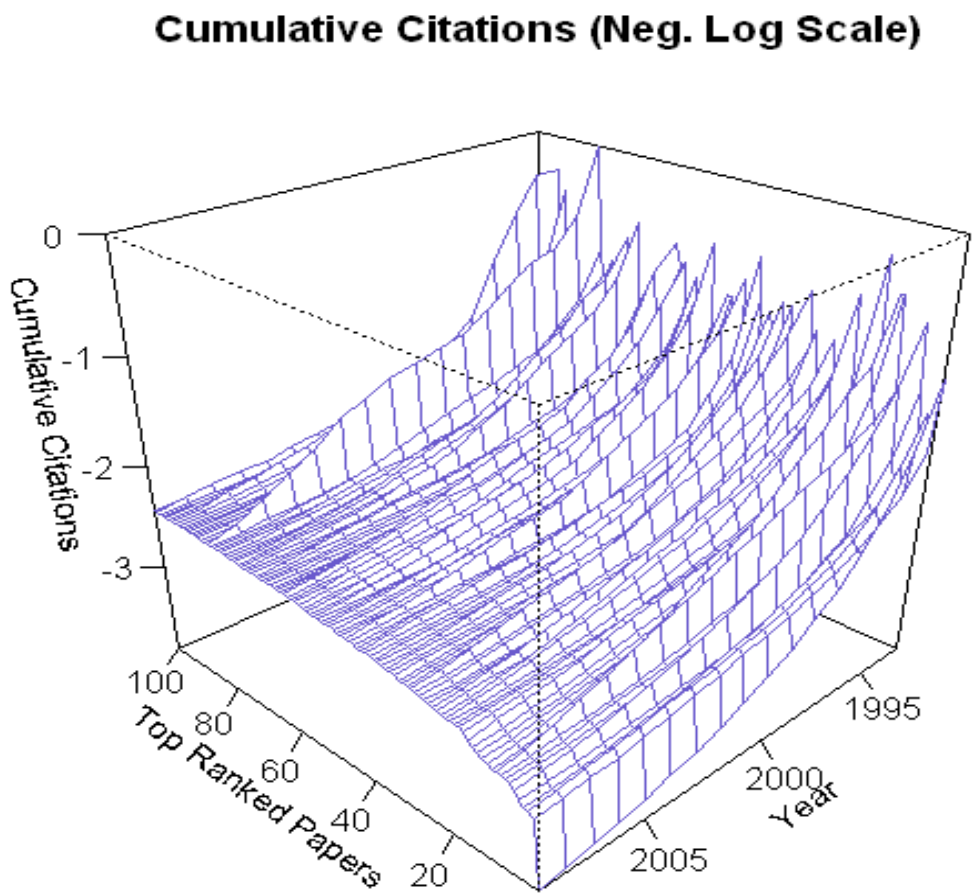
3D Representation of Cumulative Citations (Top 1000 1991)



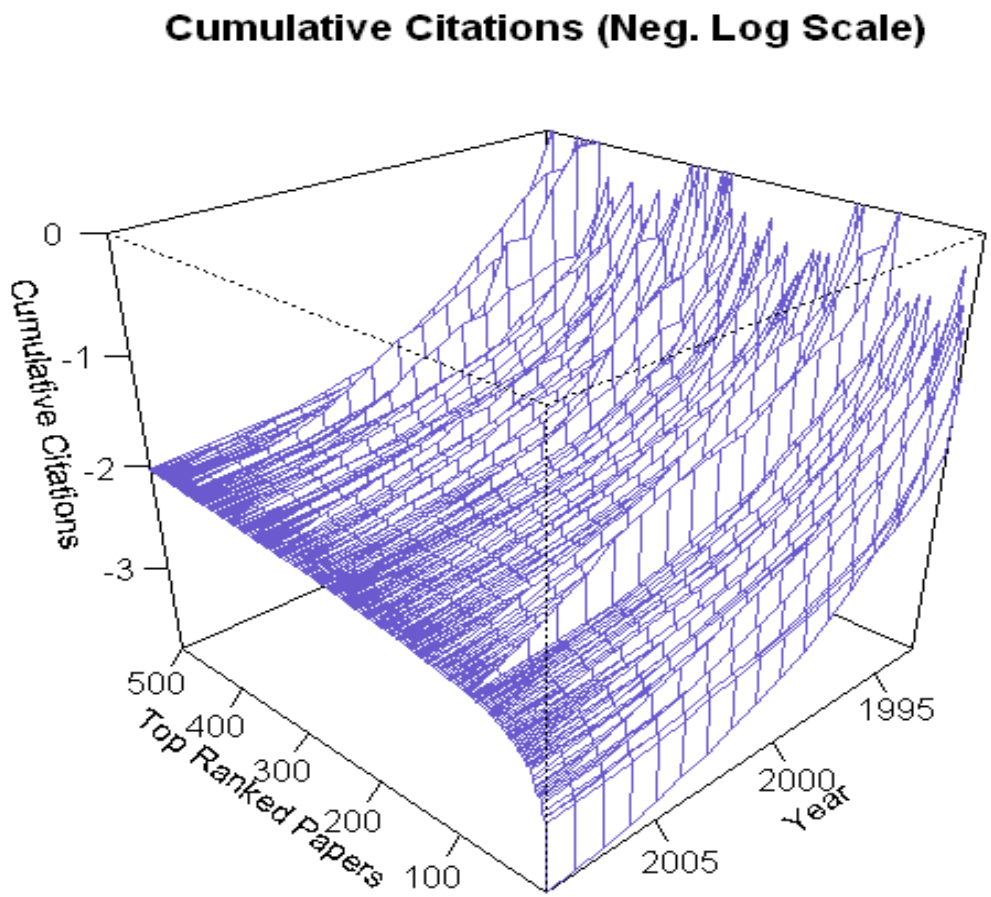
3D Representation of Cumulative Citations (Top 2000 1991)



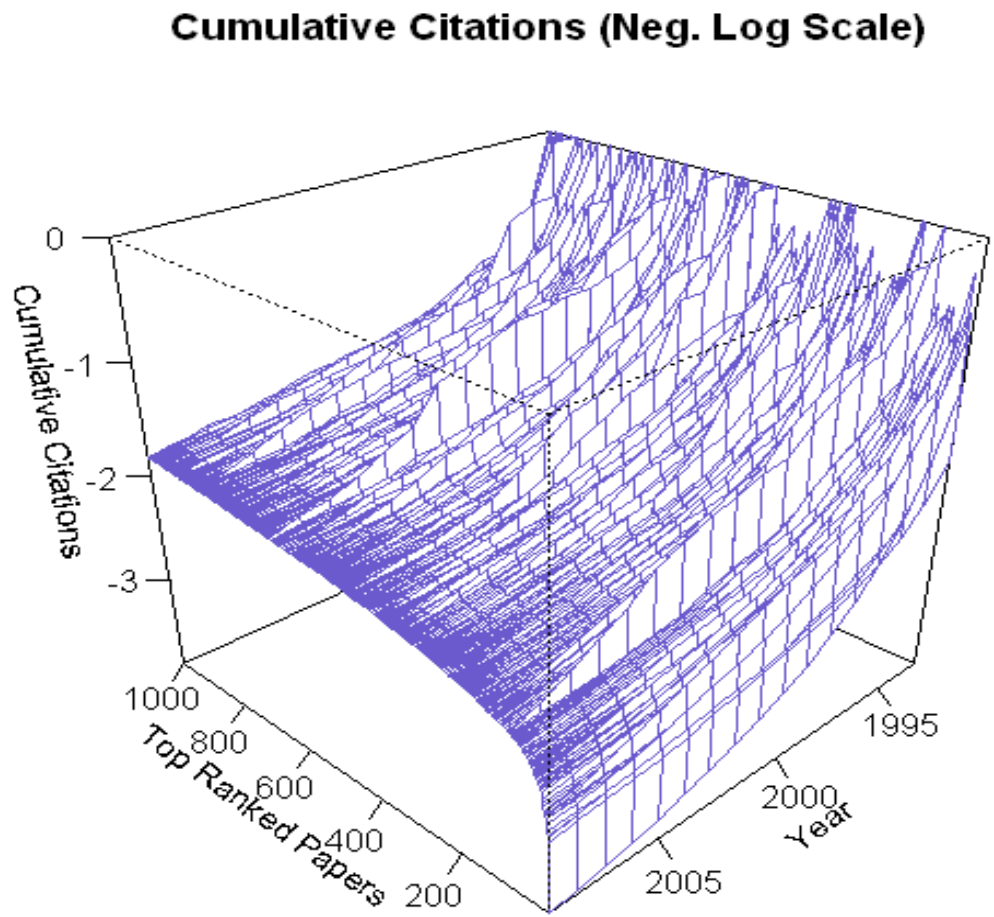
3D Representation of Cumulative Citations (Top 100 1992)



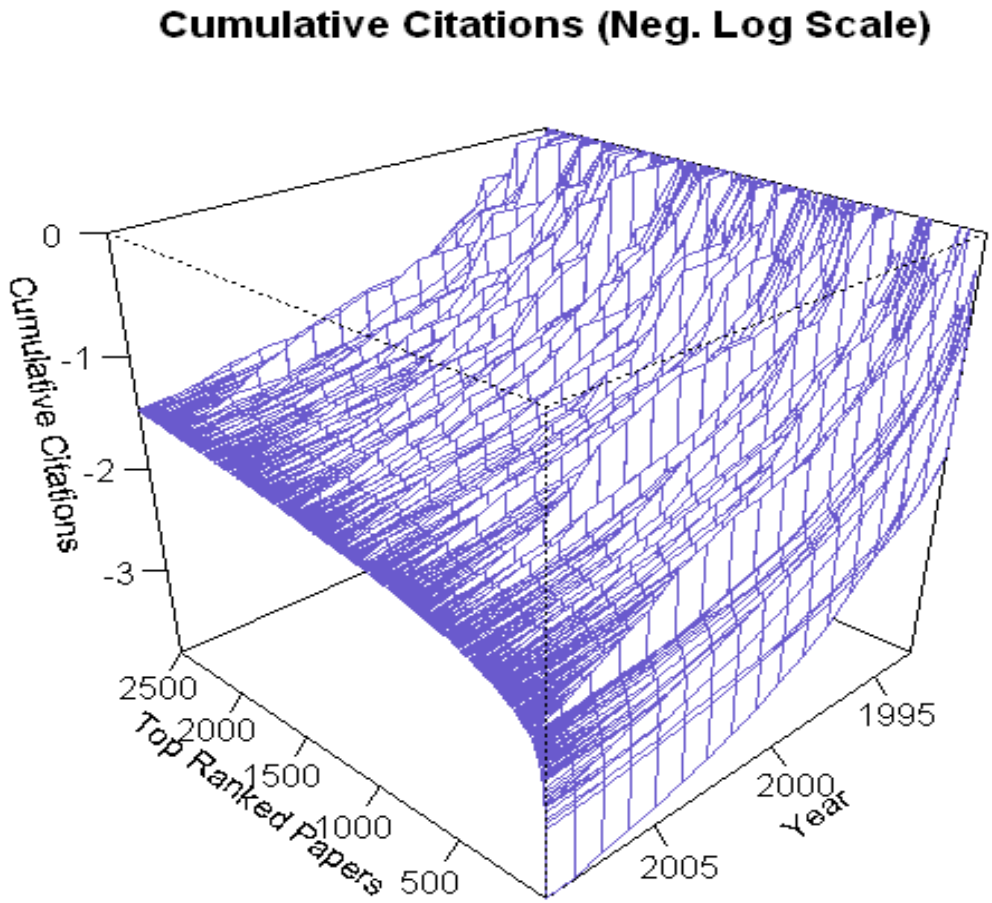
3D Representation of Cumulative Citations (Top 500 1992)



3D Representation of Cumulative Citations (Top 1000 1992)



3D Representation of Cumulative Citations (Top 2500 1992)



Conclusions

- Citations are a network phenomenon showing impact at the network level
- Effects of papers continue for more than a decade
- Bursts of citations for papers across the rank many years later
- The highly influential papers are many
- High probability of being cited at all