Toward the Construction of Geographical Scientometrics

Wataru Souma (CST, Nihon University)
souma.wataru@nihon-u.ac.jp

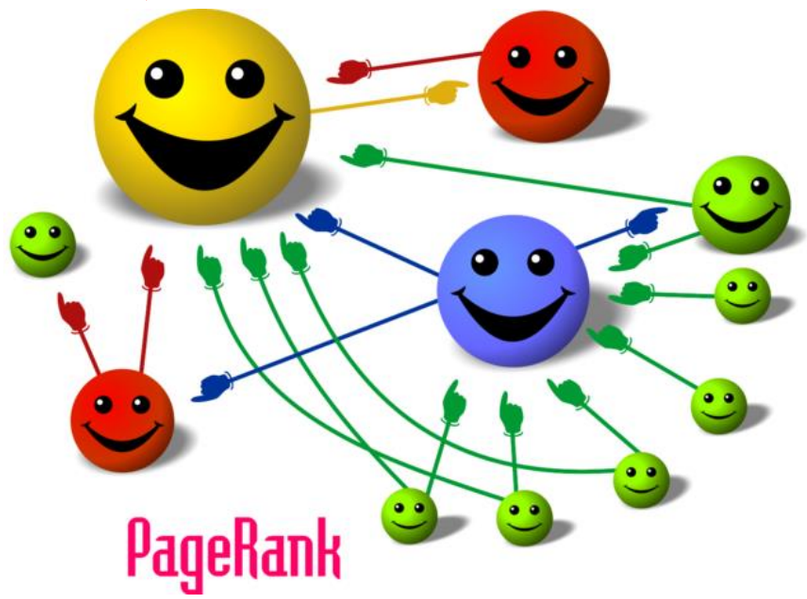
11/8/2019@CIGS

Motive & Data

- The number of citations is the most frequently used measure to quantify the significance of papers.
- Which paper is the most important if there are some papers with the same number of citation?

  -> We need to introduce new measure

  -> PageRank: extract useful pages (papers)
- Whose paper is #1?
- Where is an innovative place?

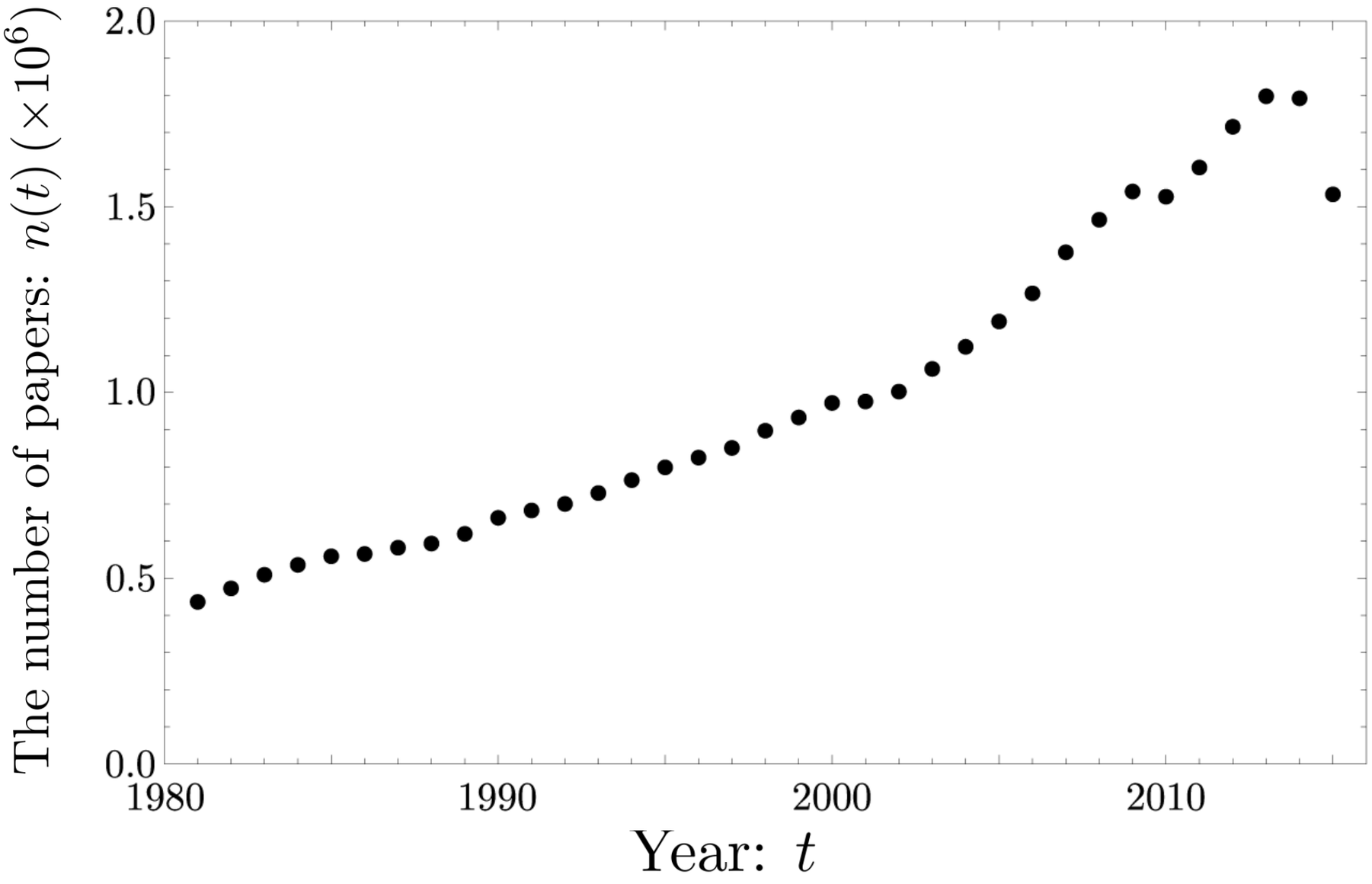


From Wikipedia

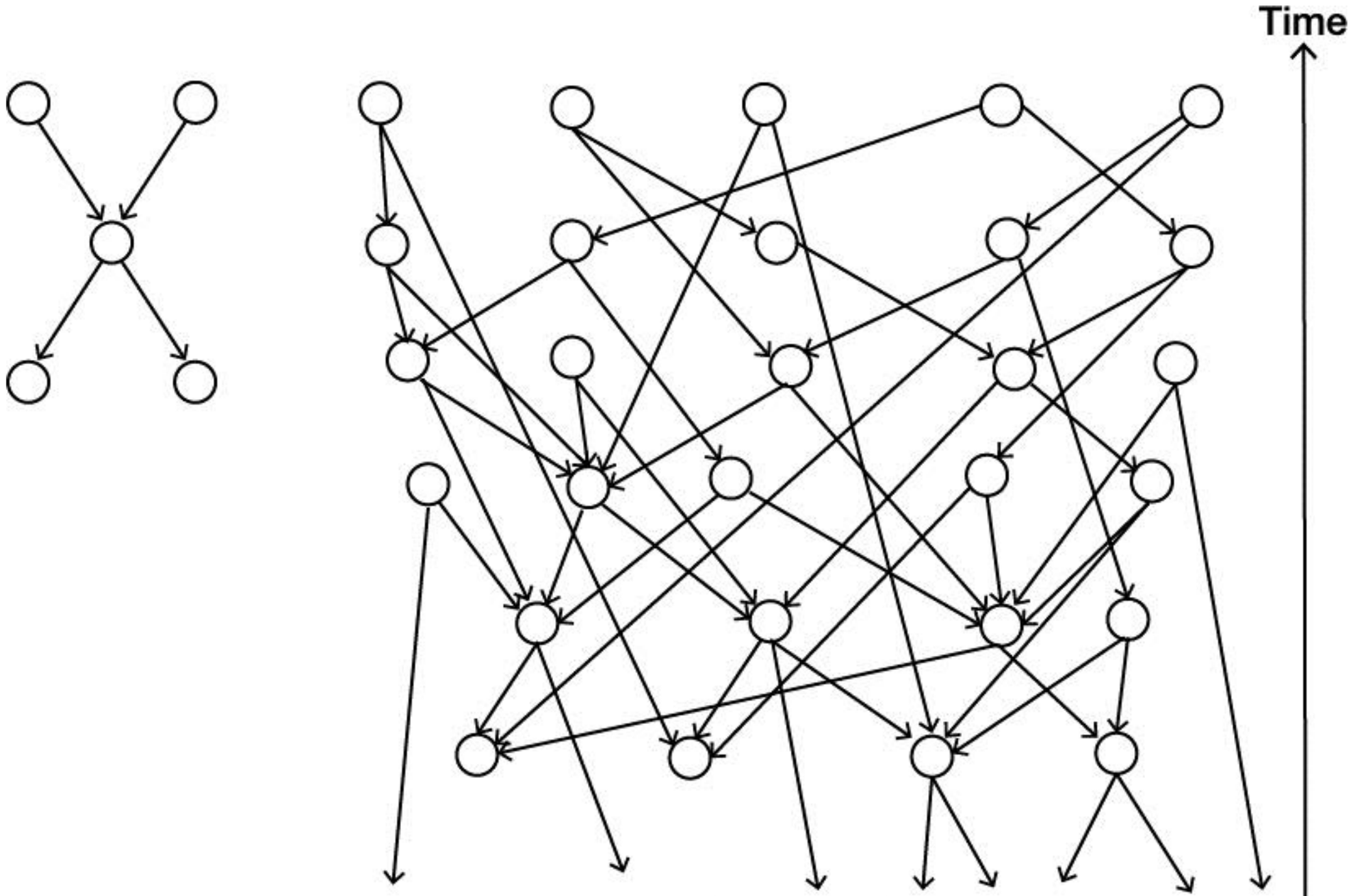
Data

Science Citation Index Expanded (1981 ~ 2015)

- 34,666,719 ≃ 35 million papers
- 591,321,826 ≃ 600 million citations



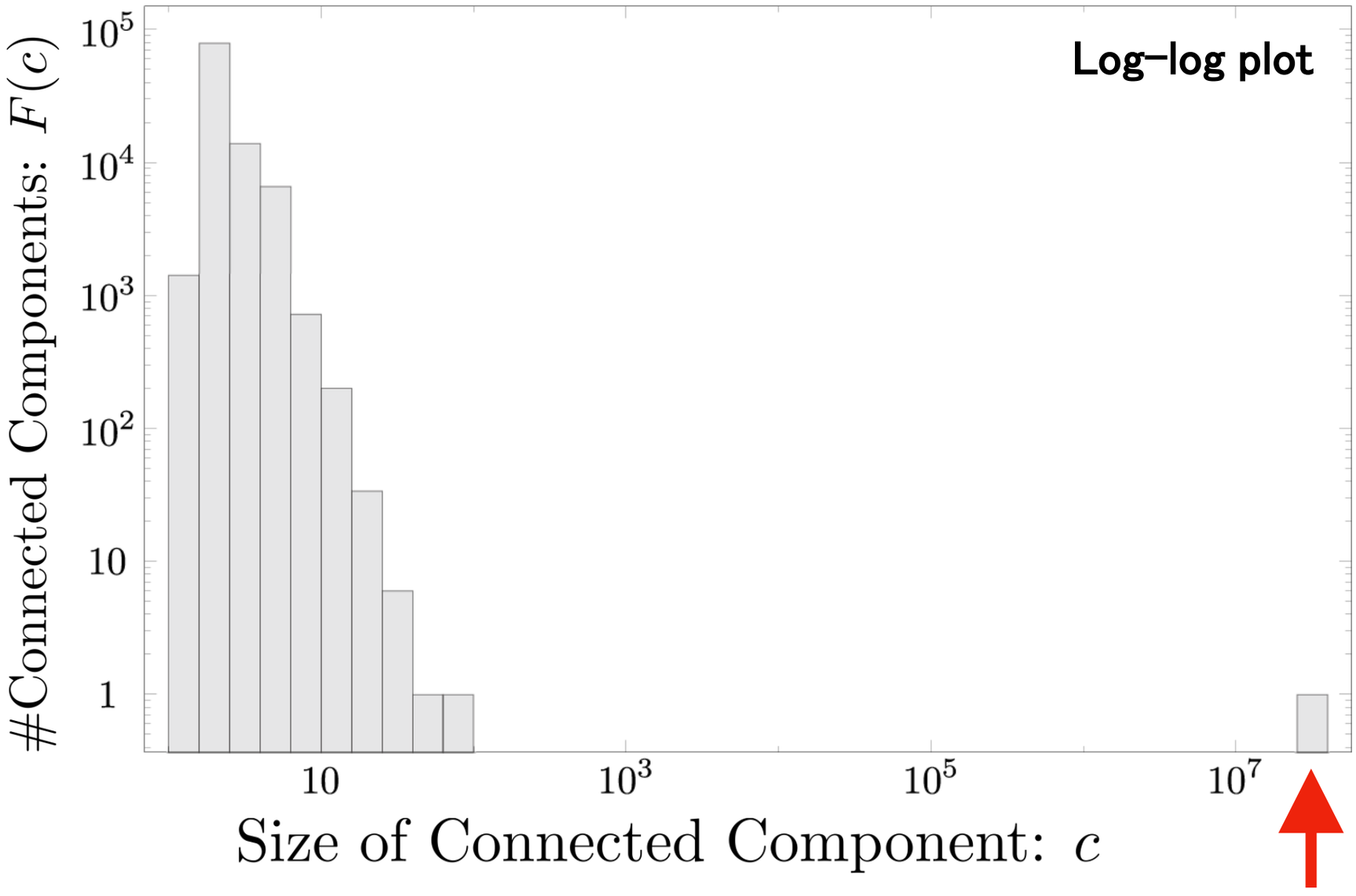
Citation Network



- 34,666,719 ≃ 35 million nodes (papers)
- 591,321,826 ≃ 600 million directed links (citations)

Connected Components

We extracted connected components by ignoring the direction of links



Largest connected component

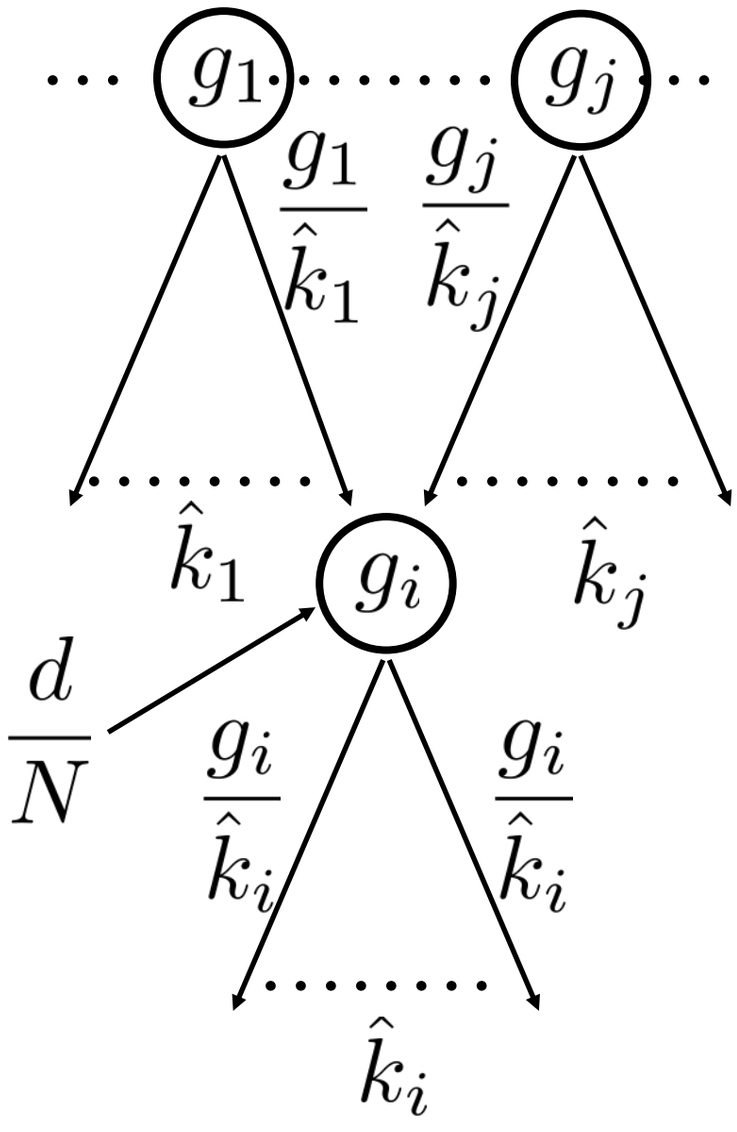
N=34,428,322 papers (99.3%)

591,177,607 citations (99.98%)

Google number

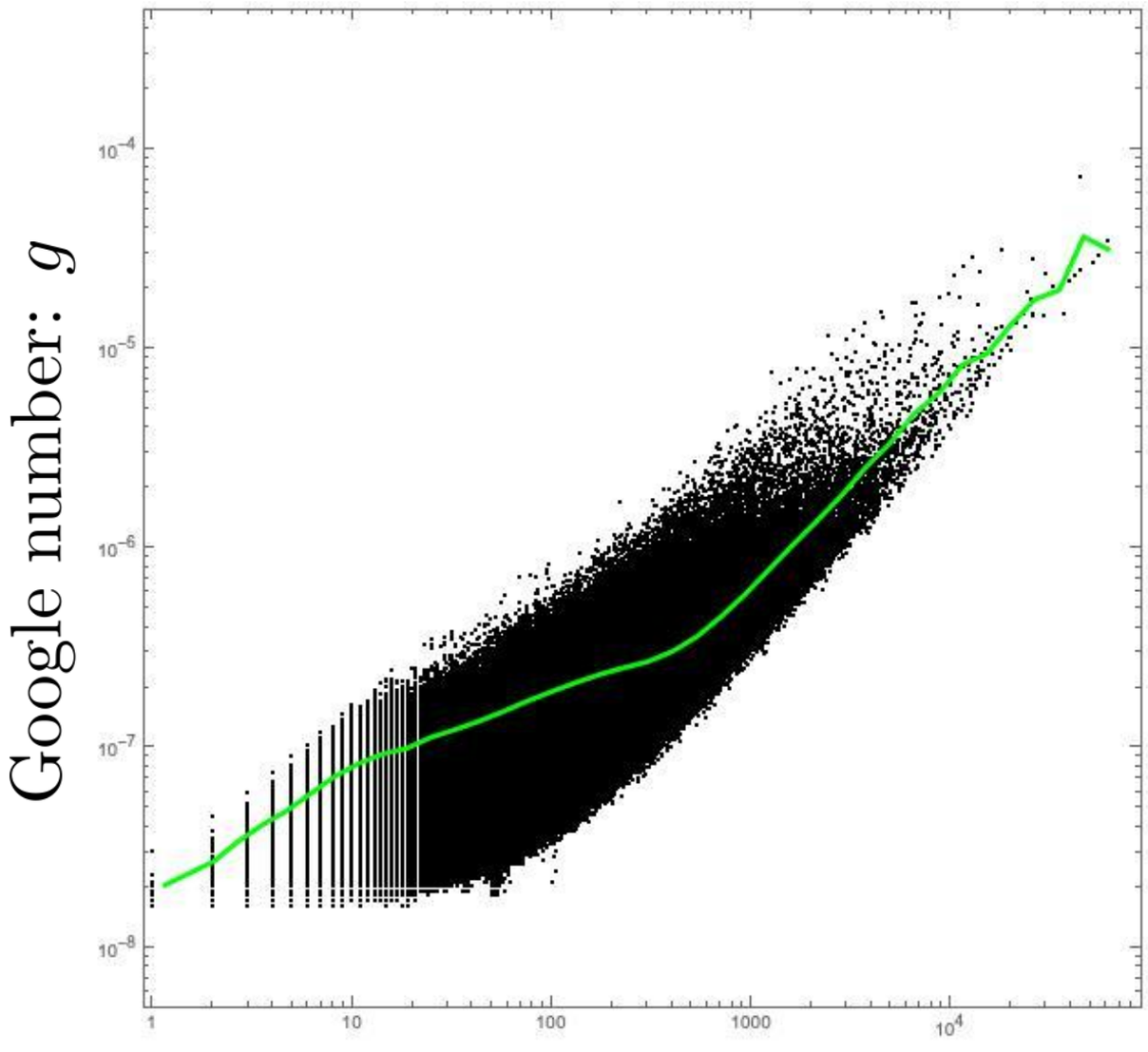
$$g_i = (1-d)\sum_{j\ \mathrm{nn}\ i}\frac{g_j}{\hat{k}_j} + \frac{d}{N}$$

$\hat{k}_j$: The number of citing

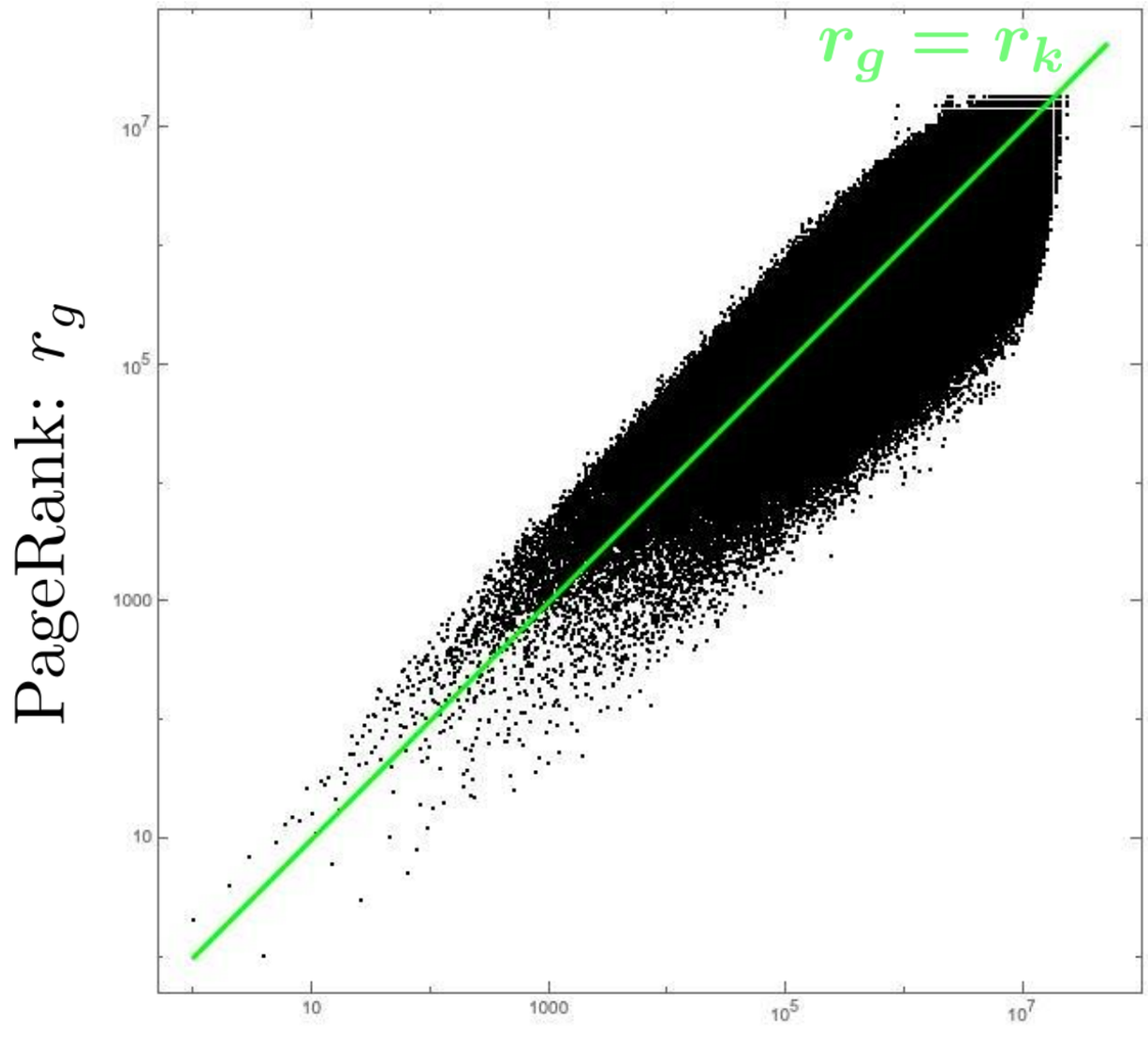


Scatter Plot

Log–log plot　34.4 million dots (papers)



The number of citations: k

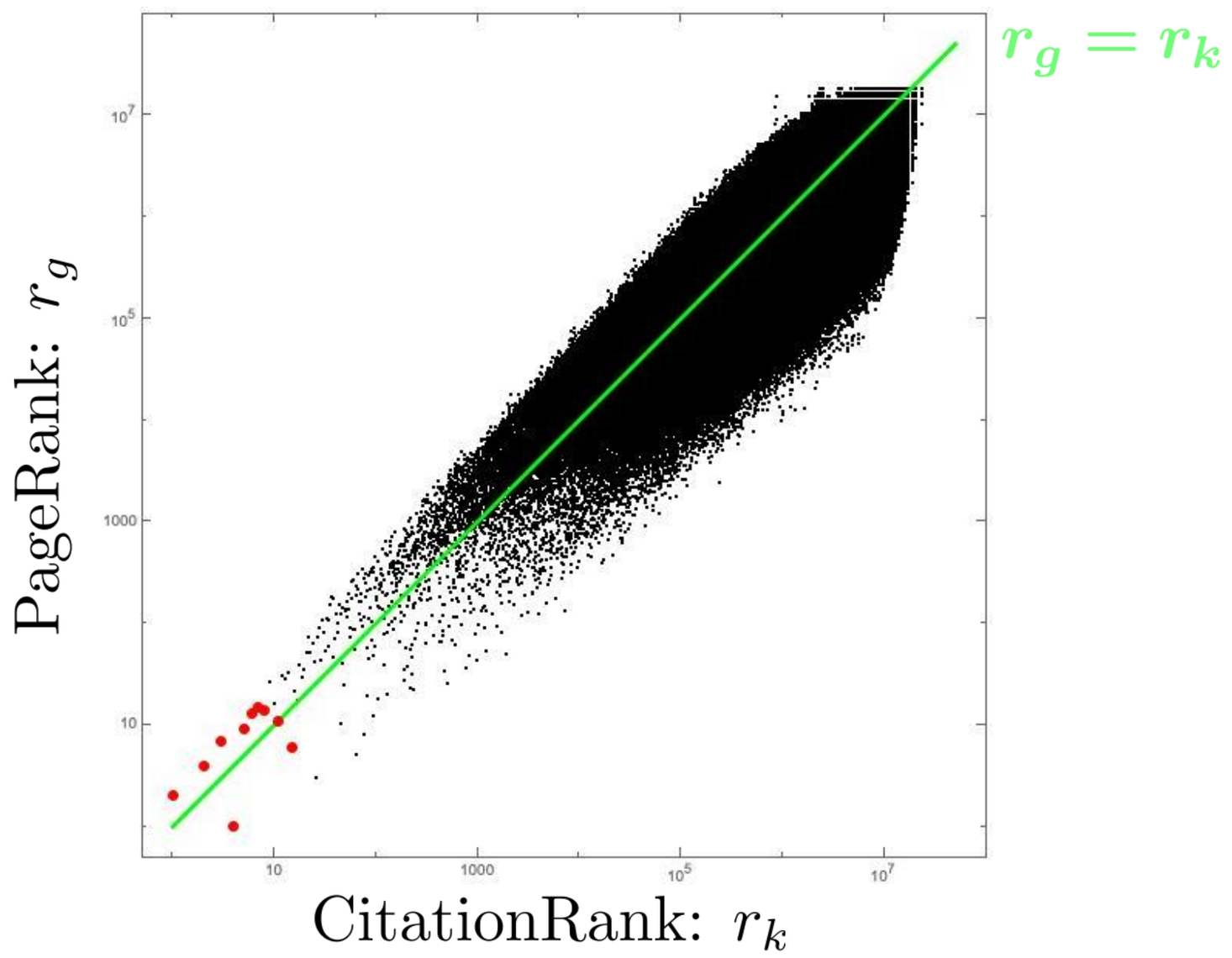


CitationRank: r_k

Popularity Index

$$r_i = \frac{1}{2}\left(r_{k,i} + r_{g,i}\right)$$

CitationRank　　PageRank

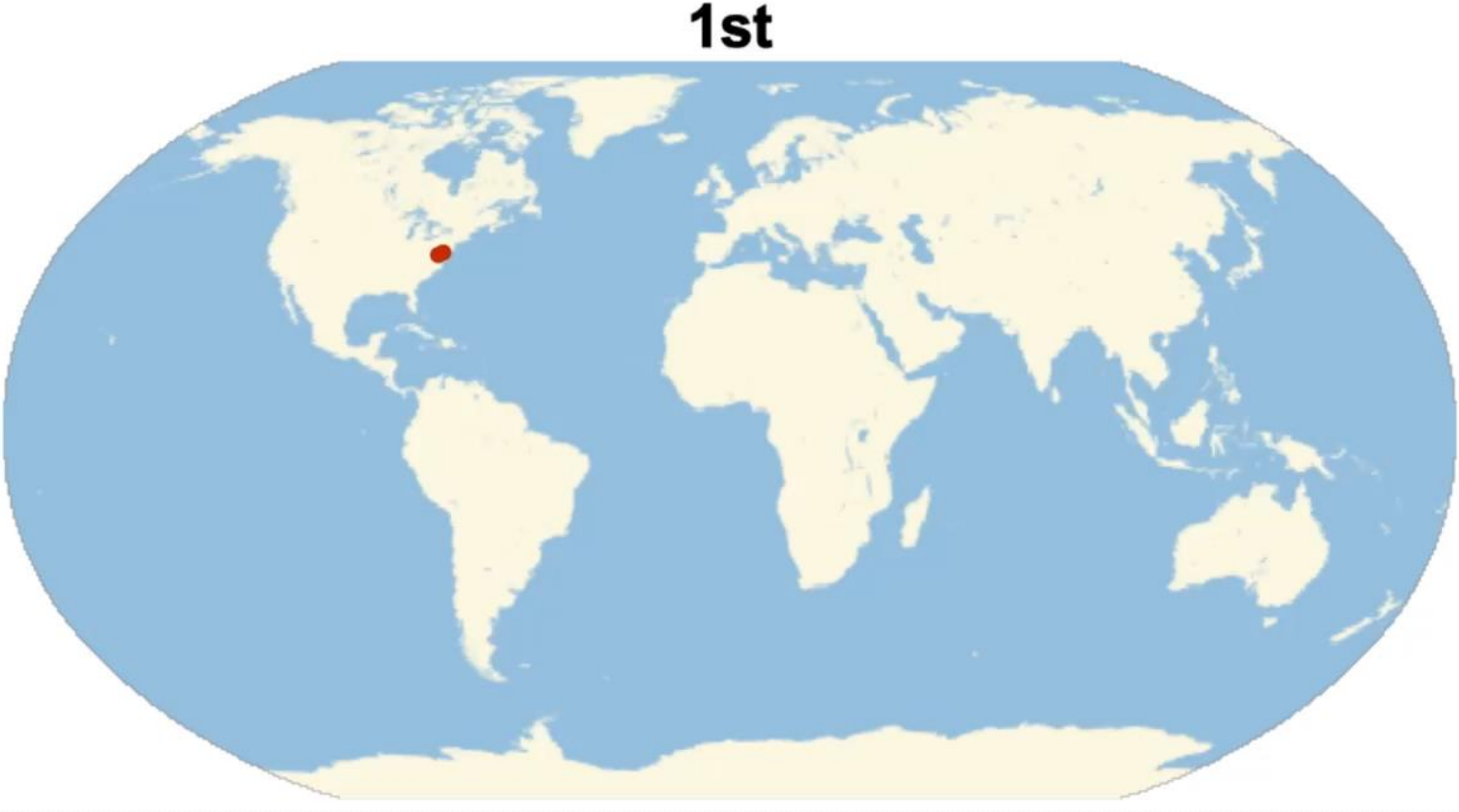


Popular Papers

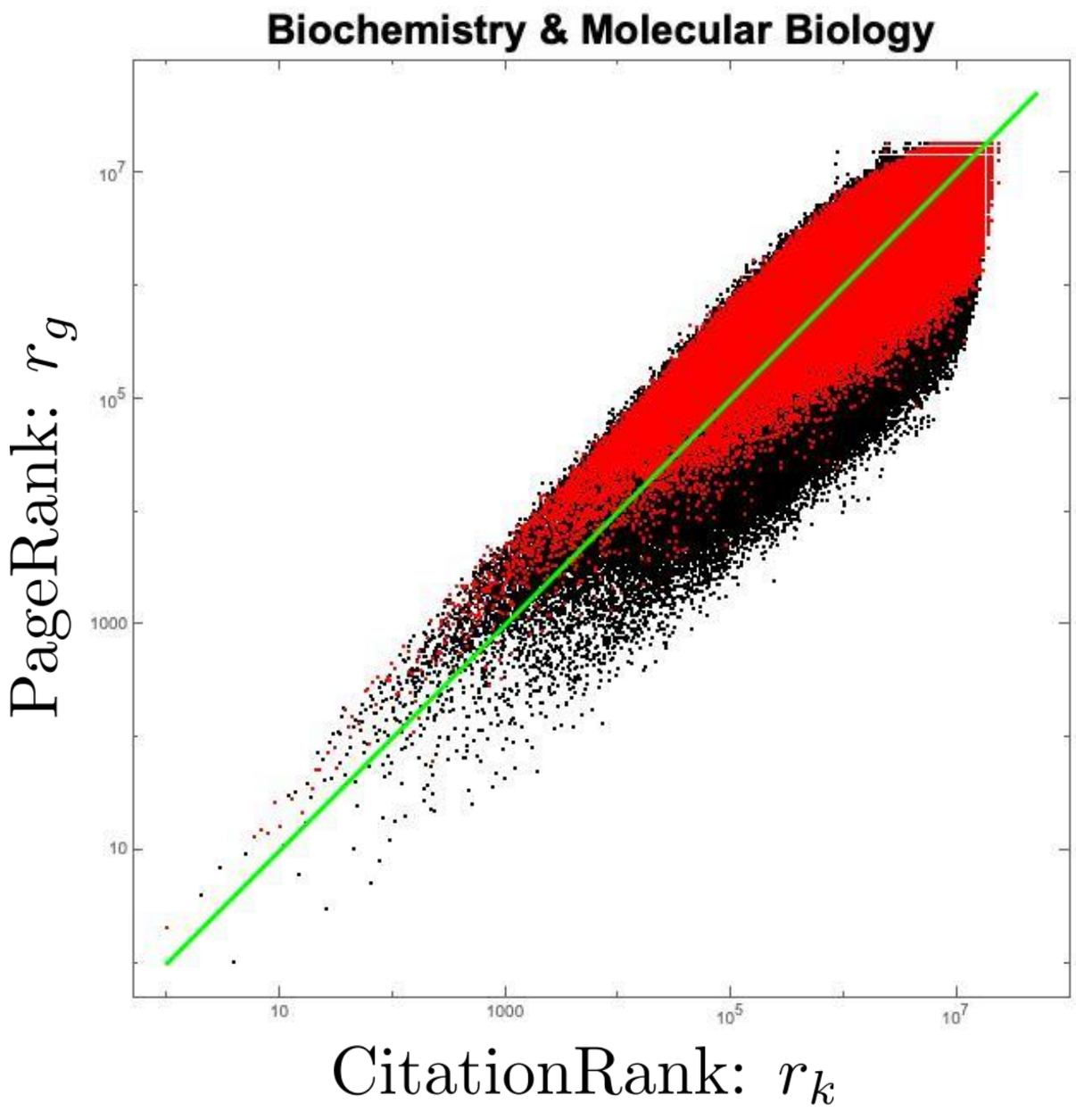
Table 1. Top 10 popular papers ordered by r_i

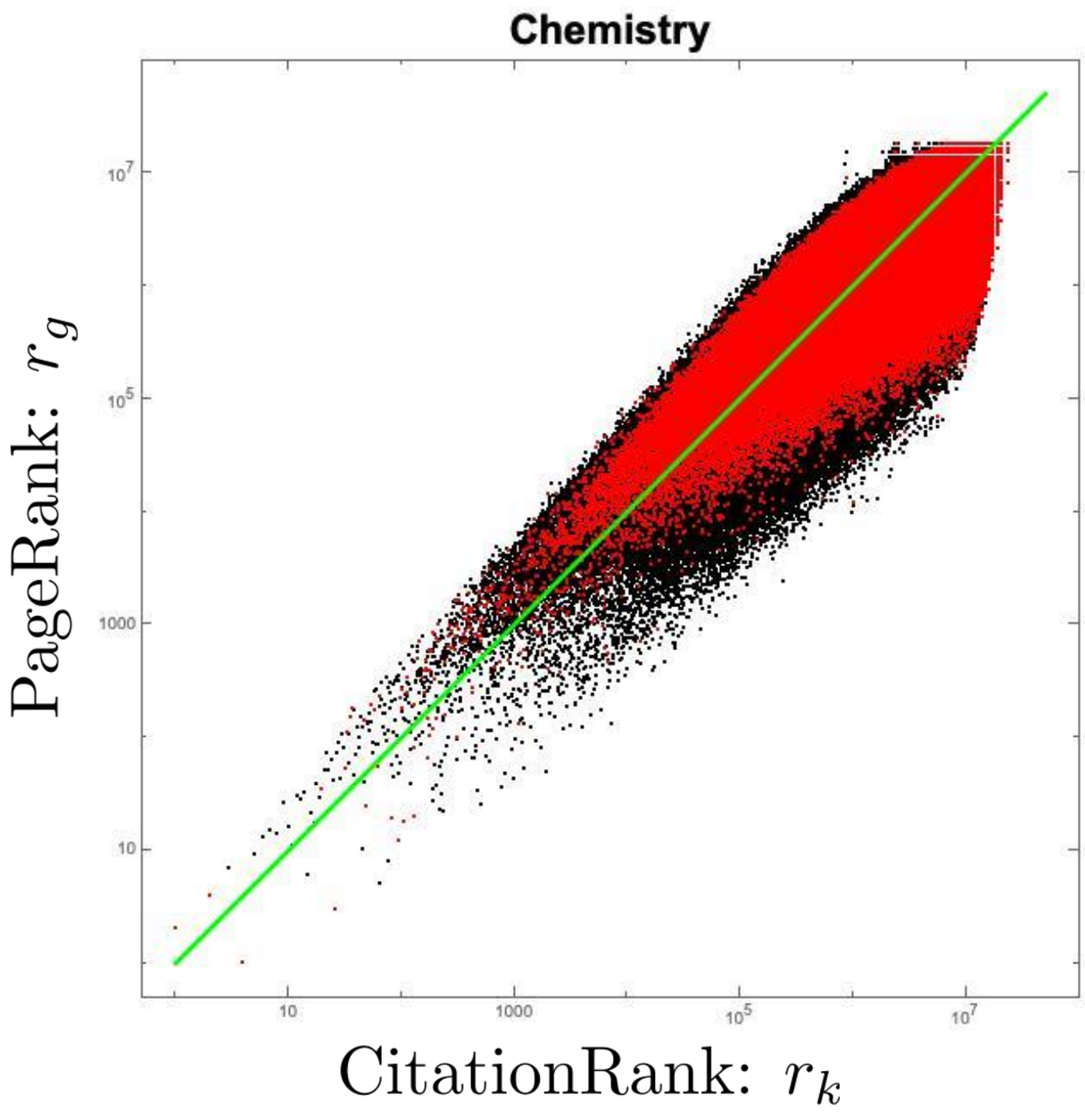
Ranking	r_i	$r_{k,i}$	$r_{g,i}$	*1st author (year)*	*Main Subject*
1	1.5	1	2	Chomczynski (1987)	Biochemistry & Molecular Biology
2	2.5	4	1	Sheldrick (2008)	Chemistry
3	3	2	4	Becke (1993)	Chemistry
4	5	3	7	Lee (1988)	Materials Science
5	7	5	9	Perdew (1996)	Physics
6	9.5	6	13	Thompson (1994)	Biochemistry & Molecular Biology
7	10.5	15	6	Bland (1986)	General & Internal Medicine
8	11	7	15	Altschul (1997)	Biochemistry & Molecular Biology
8	11	8	14	Altschul (1990)	Biochemistry & Molecular Biology
8	11	11	11	Otwinowski (1997)	Biochemistry & Molecular Biology

Geography of Popular Papers



Subject Dependence



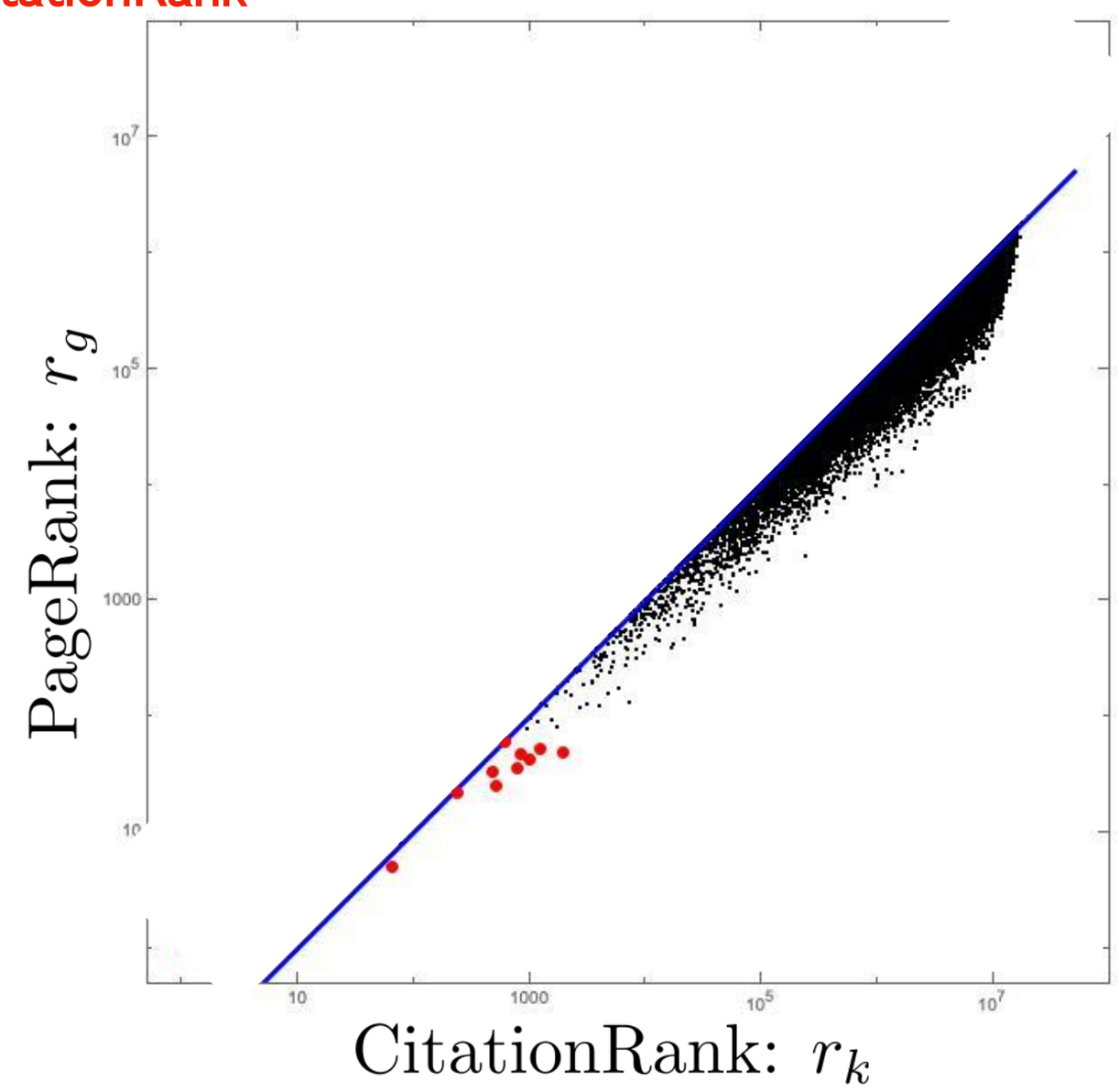


Prestige index

PageRank

$$\text{Ratio: } r_i = \frac{r_{g,i}}{r_{k,i}} \longrightarrow \text{Conditional PageRank: } r_{g,i}(r_i \geq x)$$

CitationRank



$r_g = r_k$

$$r = \frac{r_g}{r_k} = 0.1$$

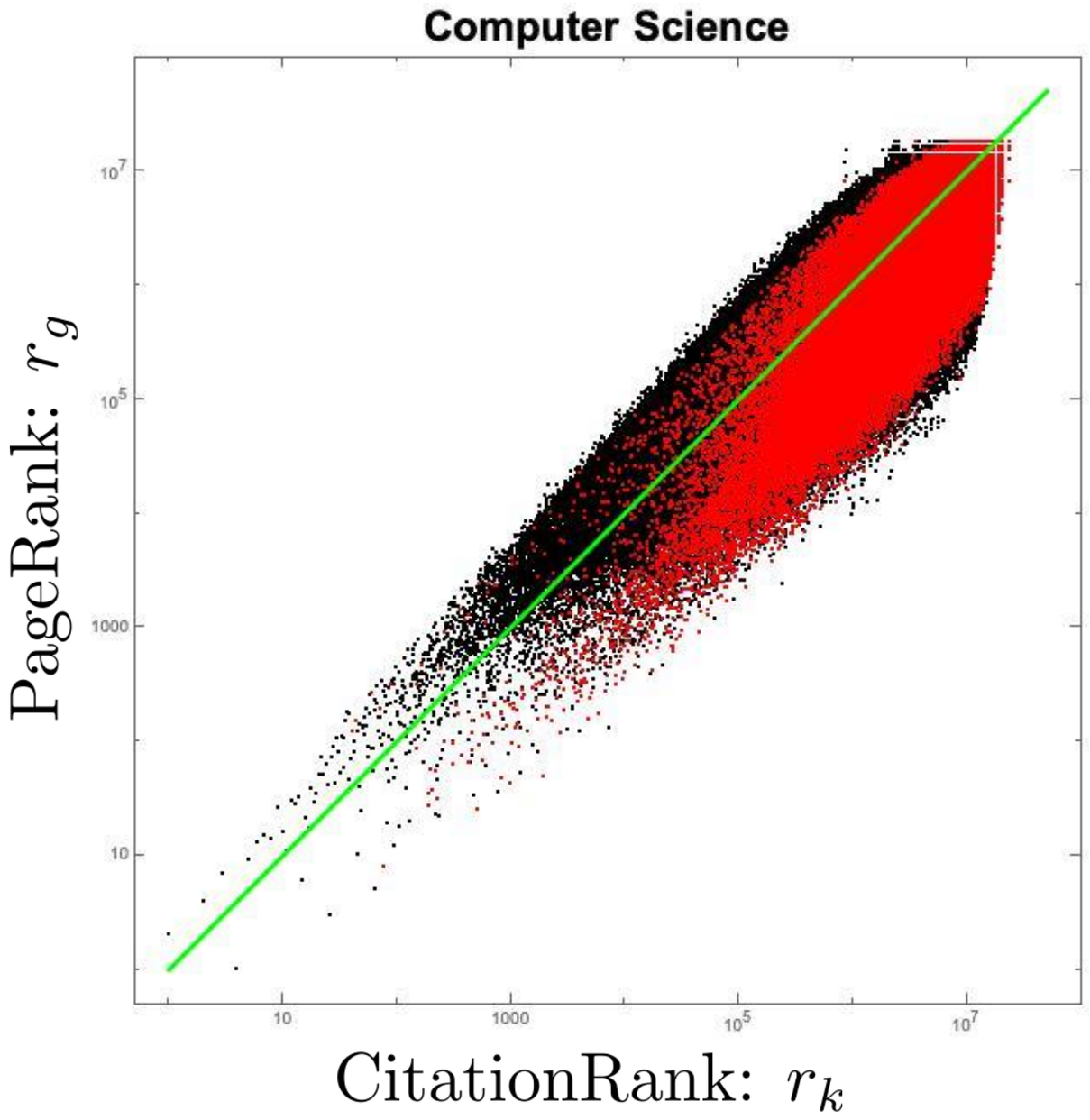
Prestige Papers

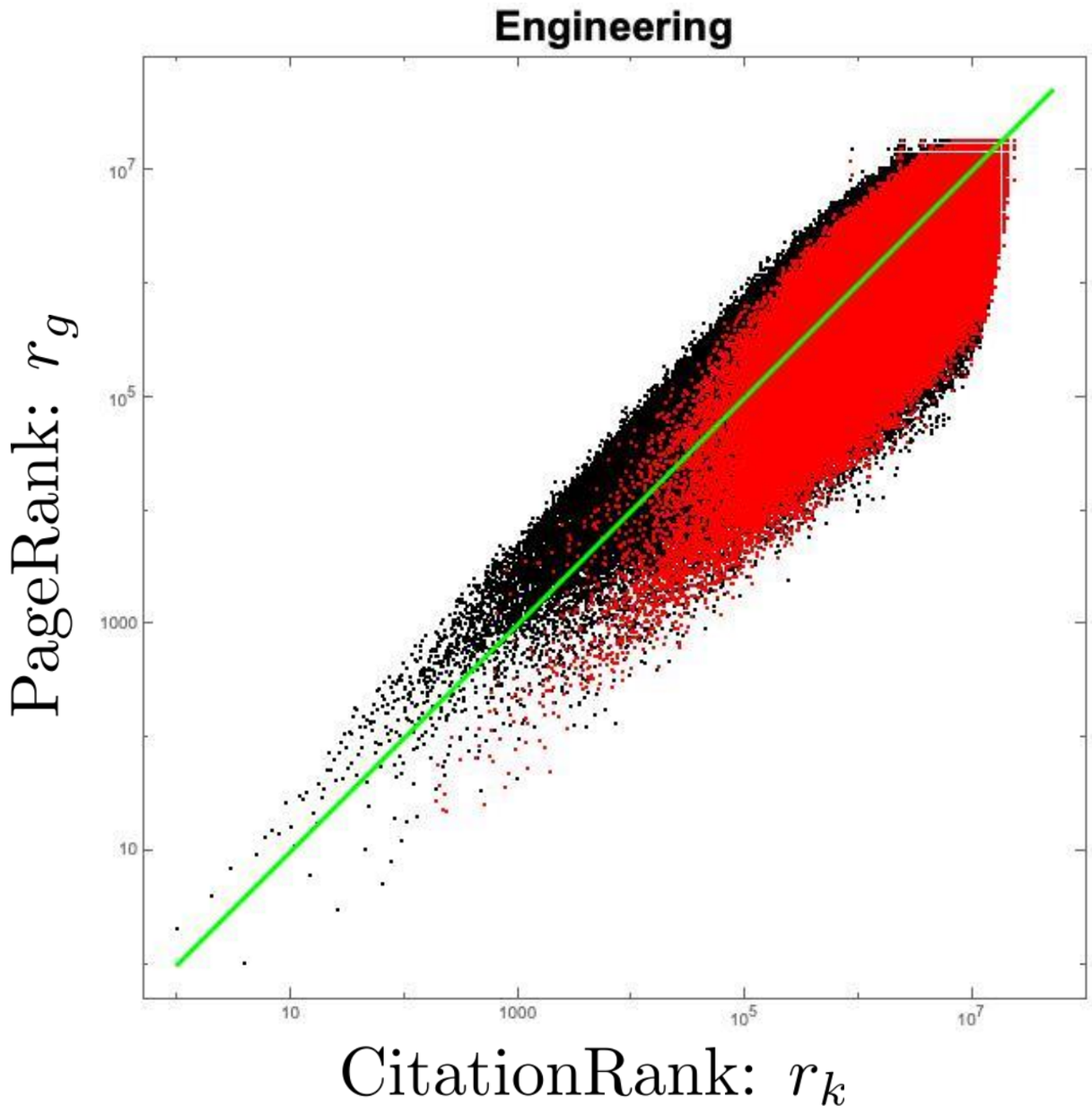
Table 2. Top 10 extremely prestige papers ordered by $r_{g,i}$ under the constraint $x_i = 0.1$

Ranking	$r_{g,i}$	$r_{k,i}$	x_i	*1st author (year)*	*1st Subject*
1	5	64	0.08	Kennedy (1995)	Computer Science
2	22	240	0.09	Alamouti (1998)	Engineering
3	25	516	0.05	Akyildiz (2002)	Computer Science
4	33	481	0.07	Pawlak (1982)	Information Science
5	36	784	0.05	Akyildiz (2002)	Engineering
6	43	998	0.04	Gruber (1994)	Computer Science
7	48	828	0.06	Gupta (2000)	Computer Science
8	49	1916	0.03	Floyd (1993)	Computer Science
9	53	1247	0.04	Bianchi (2000)	Engineering
10	60	609	0.10	Haykin (2005)	Engineering

Geography of Prestige Papers

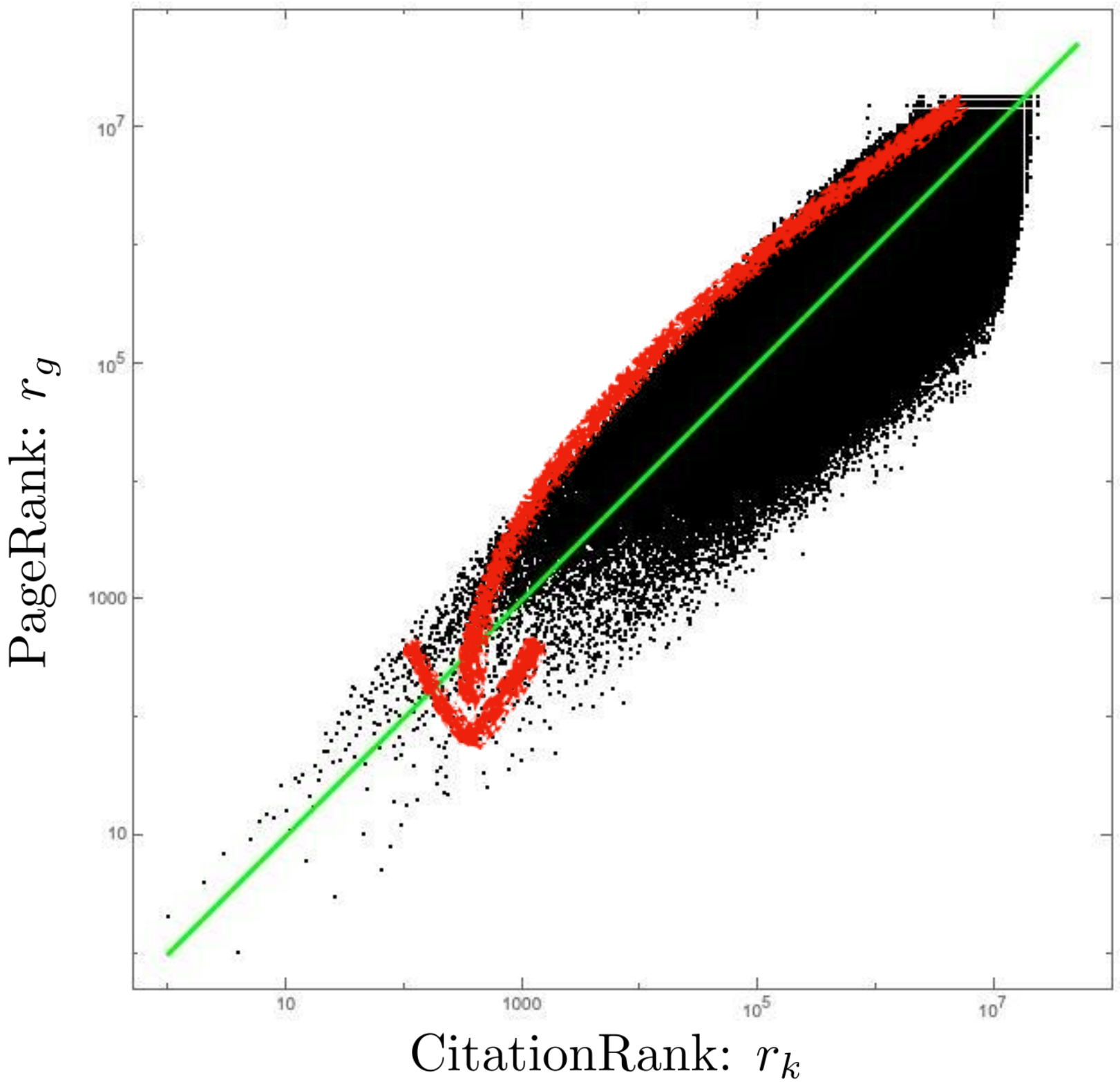
Subject Dependence





Time Dependence

Moving of prestige papers



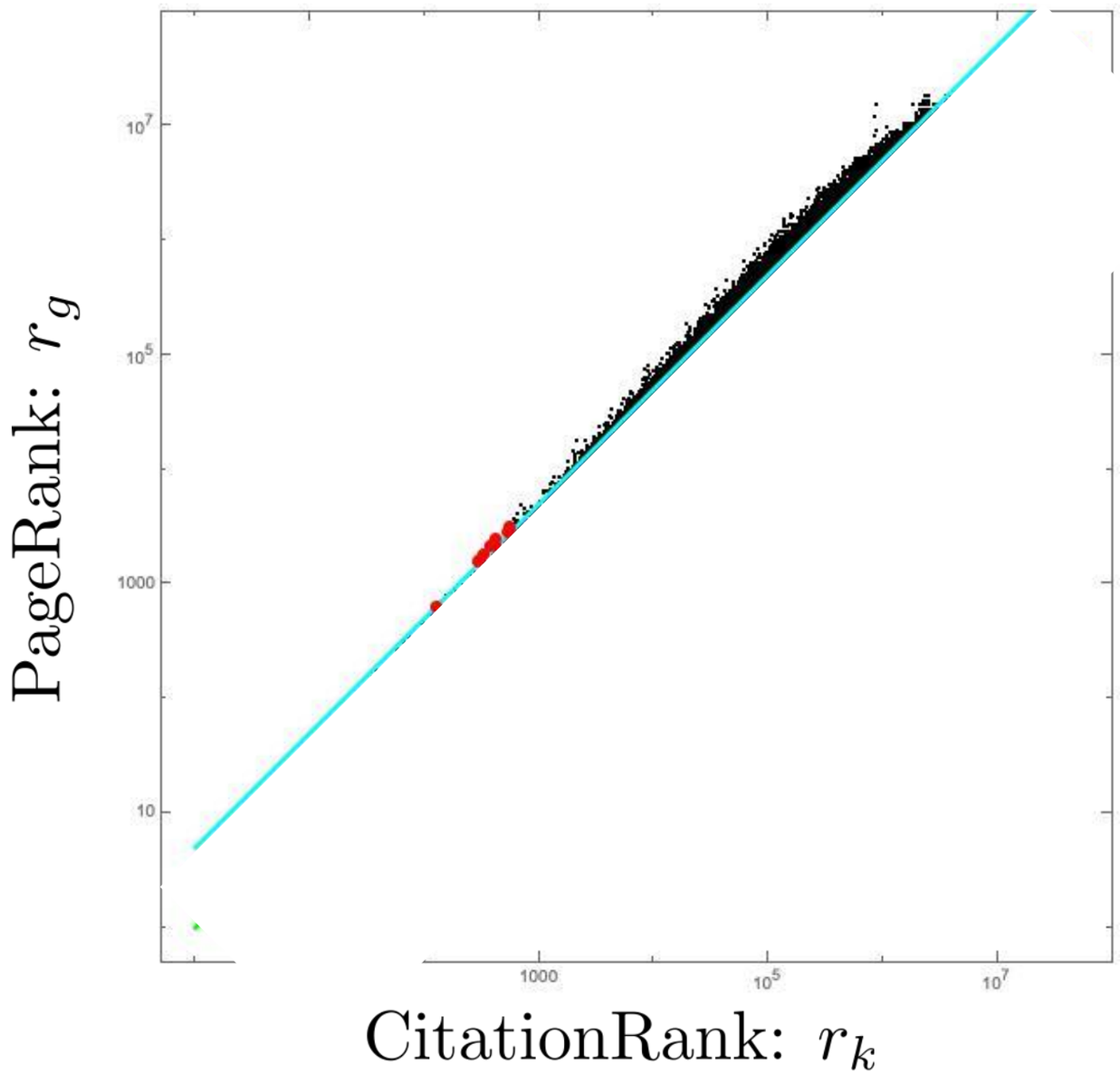
Emerging index

PageRank

Ratio: $r_i = \dfrac{r_{g,i}}{r_{k,i}}$ → Conditional CitationRank: $r_{k,i}(r_i \leq x)$

CitationRank

$$r = \frac{r_g}{r_k} = 5$$

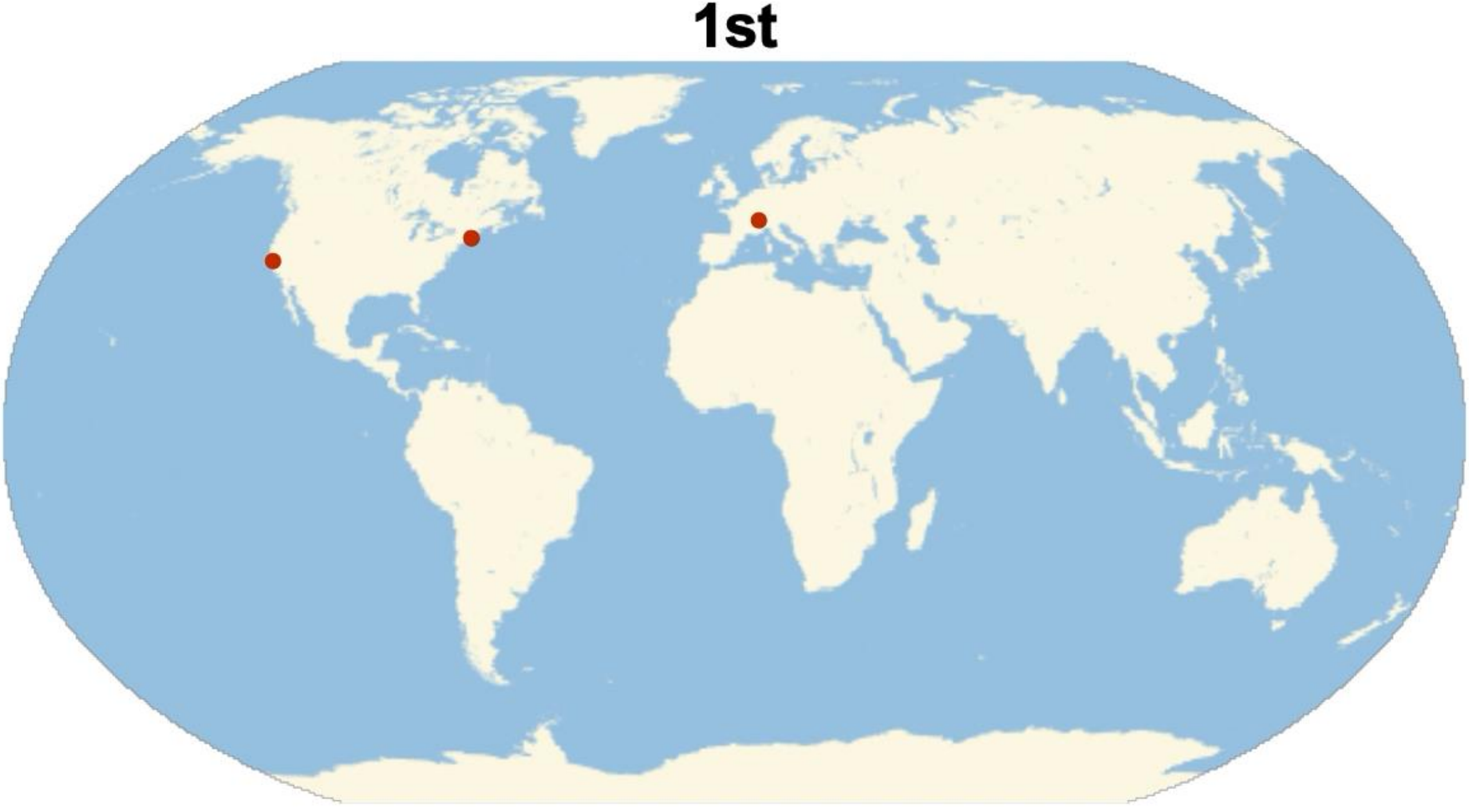


Emerging Papers

Table 3. Top 10 emerging papers ordered by $r_{k,i}$ under the constraint $x_i = 5$

Ranking	$r_{k,i}$	$r_{g,i}$	x_i	*1st author (year)*	*Journal*
1	125	627	5.0	Hanahan (2011)	CELL
2	297	1580	5.3	Huang (2009)	NATURE PROTOCOLS
3	304	1608	5.3	Zhao (2008)	THEORETICAL CHEMISTRY ACCOUNTS
4	327	1810	5.5	Bartel (2009)	CELL
5	375	2128	5.7	Lewis (2005)	CELL
6	414	2506	6.1	Jenuwein (2001)	SCIENCE
7	419	2123	5.1	Li (1997)	CELL
8	535	2802	5.2	Xia (1995)	SCIENCE
9	543	3120	5,7	Lee (2000)	CELL
10	547	2865	5.2	Hall (1998)	SCIENCE

Geography of Emerging Papers



Summary

- We calculated CitationRank and PageRank for papers, which are contained in SCIE (1881~2015)
  - #papers: 34,666,719, #citations: 591,321,826
- We proposed the method to extract popular papers
  - Subjects of popular paper
    - Biology & Molecular Biology, Chemistry
- We proposed the method to extract prestige papers
  - Subjects of prestige paper
    - Computer Science, Engineering
- We proposed the method to extract emerging papers
  - Many papers published in journal “CELL”
  - Central place is Boston (MIT & Harvard)

Appendix

Wataru Souma

College of Science and Technology, Nihon University
Verified email at nihon-u.ac.jp

econophysics network science renormalization group

FOLLOW

TITLE	CITED BY	YEAR
Non-trivial ultraviolet fixed point in quantum gravity W Souma Progress of Theoretical Physics 102 (1), 181-195	291	1999
Do Pareto–Zipf and Gibrat laws hold true? An analysis with European firms Y Fujiwara, C Di Guilmi, H Aoyama, M Gallegati, W Souma Physica A: Statistical Mechanics and its Applications 335 (1-2), 197-216	223	2004
Universal structure of the personal income distribution W Souma Fractals 9 (04), 463-470	176	2001
Growth and fluctuations of personal income Y Fujiwara, W Souma, H Aoyama, T Kaizoji, M Aoki Physica A: Statistical Mechanics and its Applications 321 (3-4), 598-604	165	2003
Pareto's law for income of individuals and debt of bankrupt companies H Aoyama, W Souma, Y Nagahara, MP Okazaki, H Takayasu, ... Fractals 8 (03), 293-300	158	2000
Complex networks and economics W Souma, Y Fujiwara, H Aoyama Physica A: Statistical Mechanics and its Applications 324 (1-2), 396-401	136	2003
A two factor model of income distribution dynamics M Nirei, W Souma Review of Income and Wealth 53 (3), 440-459	129	2007
Growth and Fluctuations of Personal (and Company's) Income II H Aoyama, Y Fujiwara, W Souma The Application of Econophysics, 268-273	101	2004
Rapidly converging truncation scheme of the exact renormalization group KI Aoki, K Morikawa, W Souma, JI Sumi, H Terao Progress of theoretical physics 99 (3), 451-466	95	1998
Econophysics and companies: statistical life and death in complex business networks H Aoyama, Y Fujiwara, Y Ikeda, H Iyetomi, W Souma Cambridge University Press	86	2010

Journal of Computational Social Science
January 2019, Volume 2, Issue 1, pp 33–46 | Cite as

Enhanced news sentiment analysis using deep learning methods

Authors | Authors and affiliations

Wataru Souma, Irena Vodenska, Hideaki Aoyama

Open Access | Research Article
First Online: 08 February 2019

Recent Interest

- Application of machine learning and quantum computer to Finance
- Scientometrics and Informetrics
- Multiplex firm networks
- Application of complex Hilbert principal component analysis to economic data