

# Milyen pontosan mérhető a tudományos cikkek teljesítménye a hivatkozás számaik és a kulcsszavaik segítségével?

Farkas Illés

MTA-ELTE Statisztikus és Biológiai Fizika Kutatócsoport



This work was partially supported by the  
European Union and the European Social Fund  
through project FuturICT.hu  
(grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013).

National Development Agency  
[www.ujszechenyiterv.gov.hu](http://www.ujszechenyiterv.gov.hu)  
06 40 638 638



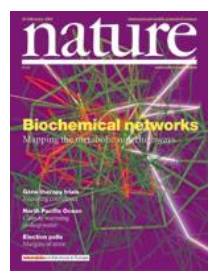
The project is supported by the European Union  
and co-financed by the European Social Fund.



Vicsek Tamás



Collective motion:  
Flocking, Panic,  
Hierarchies,  
Collective robotics



Networks:  
Social Webs, Clusters,  
Biological Networks,  
Hierarchies

European Research Council



National Development Agency  
www.ujszechenyiterv.gov.hu  
06 40 638 638



The project is supported by the European Union and co-financed by the European Social Fund.



This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013).



"WILSON, WHAT EXACTLY IS A KNOWLEDGE WORKER AND DO WE HAVE ANY ON THE STAFF?"

# Knowledge worker = ?

Time spent on producing (modifying) physical objects is decreasing

We need to quantify the „production” of knowledge (benefit to ... society?),

... because we build our decisions on these measurements (already in use)

(1) Roadmaps for newcomers to a field: formal & informal education

(2) Decisions about resources, including: **jobs**



# Knowledge worker = ?



= Data scientist ?

Oct 2012 **Data Scientist: The Sexiest Job of the 21st Century**  
by Thomas H. Davenport and D.J. Patil

Nov 2013 **Still the Sexiest Profession Alive**  
by DJ Patil | 1:03 PM November 21, 2013

= Everyone!

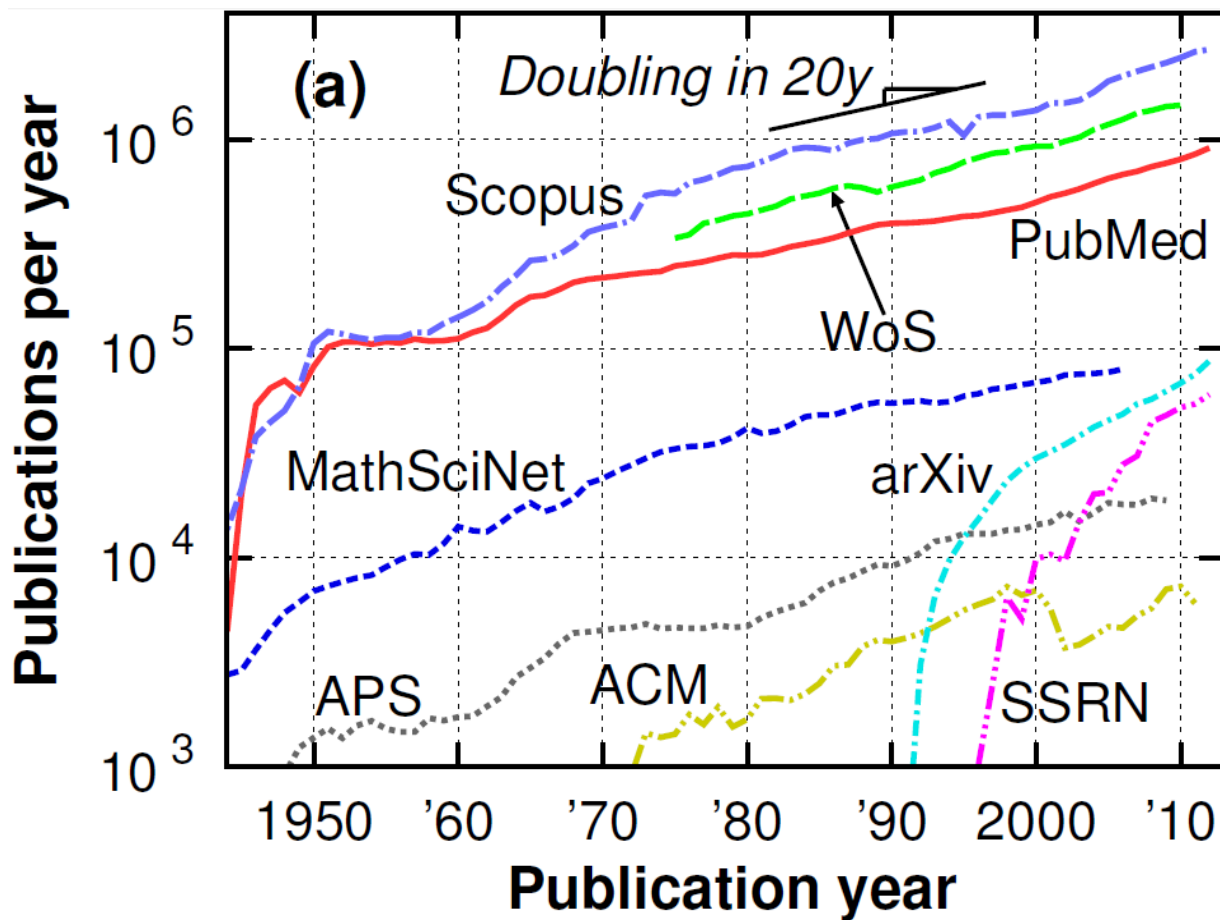
We all think

Example for making collective thinking and improvement the norm:



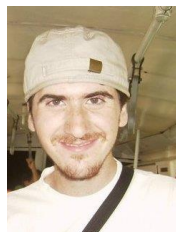
In science: Citation number

→ **Normalize**



arXiv / 1403.2140

to appear in  
*National Science Review*



Szántó-Várnagy  
Ádám



Pollner Péter



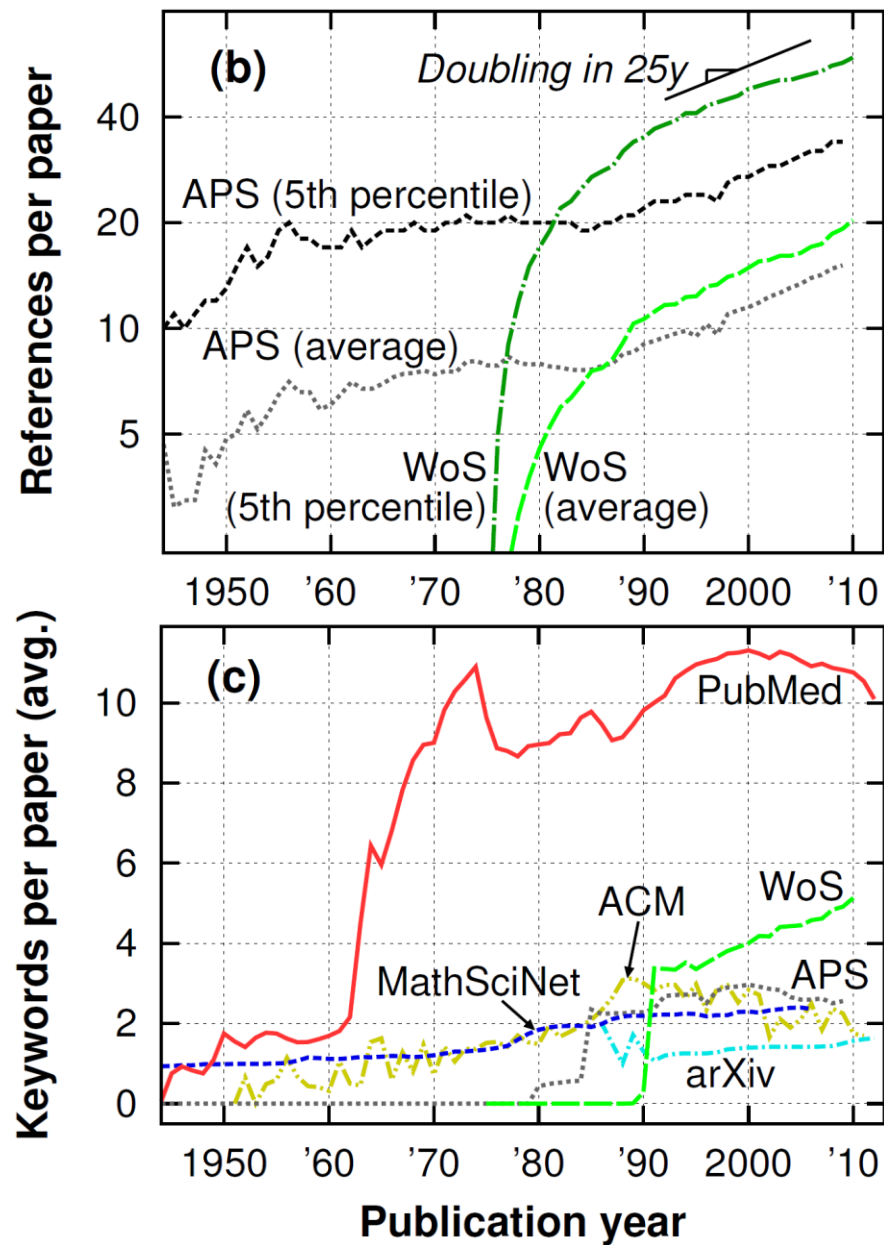
Vicsek Tamás



Farkas Illés

## Citation number

- . Number of citations received until now
- . Growing: Papers per year, References per paper
- . Favors papers that appeared close (in time and topic) to the origins of large and still active research areas
- . Need to compare citation numbers within topics (fields) and years close
- . Large differences: Keyword number per paper  
For both user- and database assigned keywords



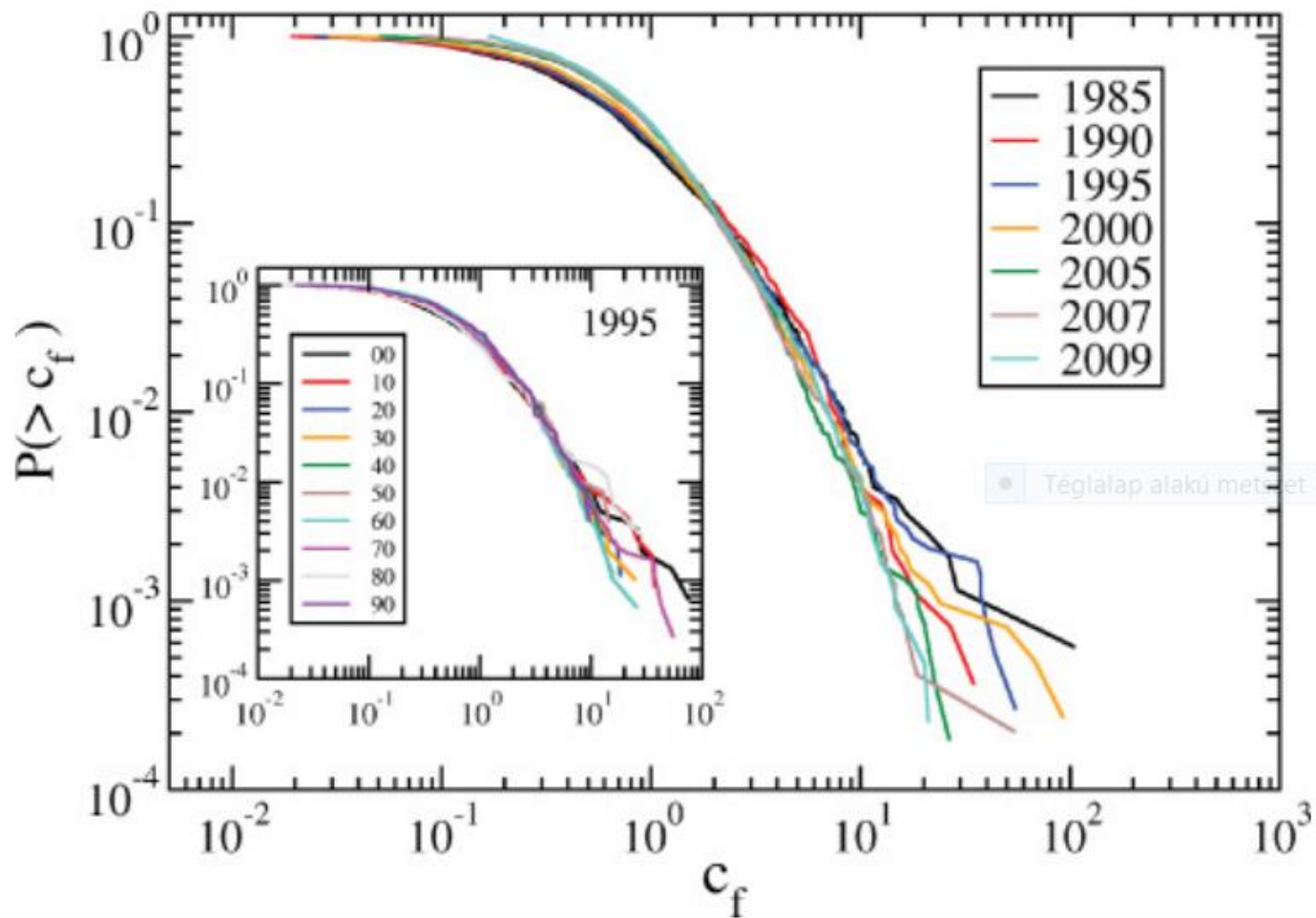
## Scientific metrics

. are built mostly on ALMs (Article-Level Metrics)

. are often mention-counting metrics

? normalization

Radicchi & Castellano  
PRE 2011





## PageRank

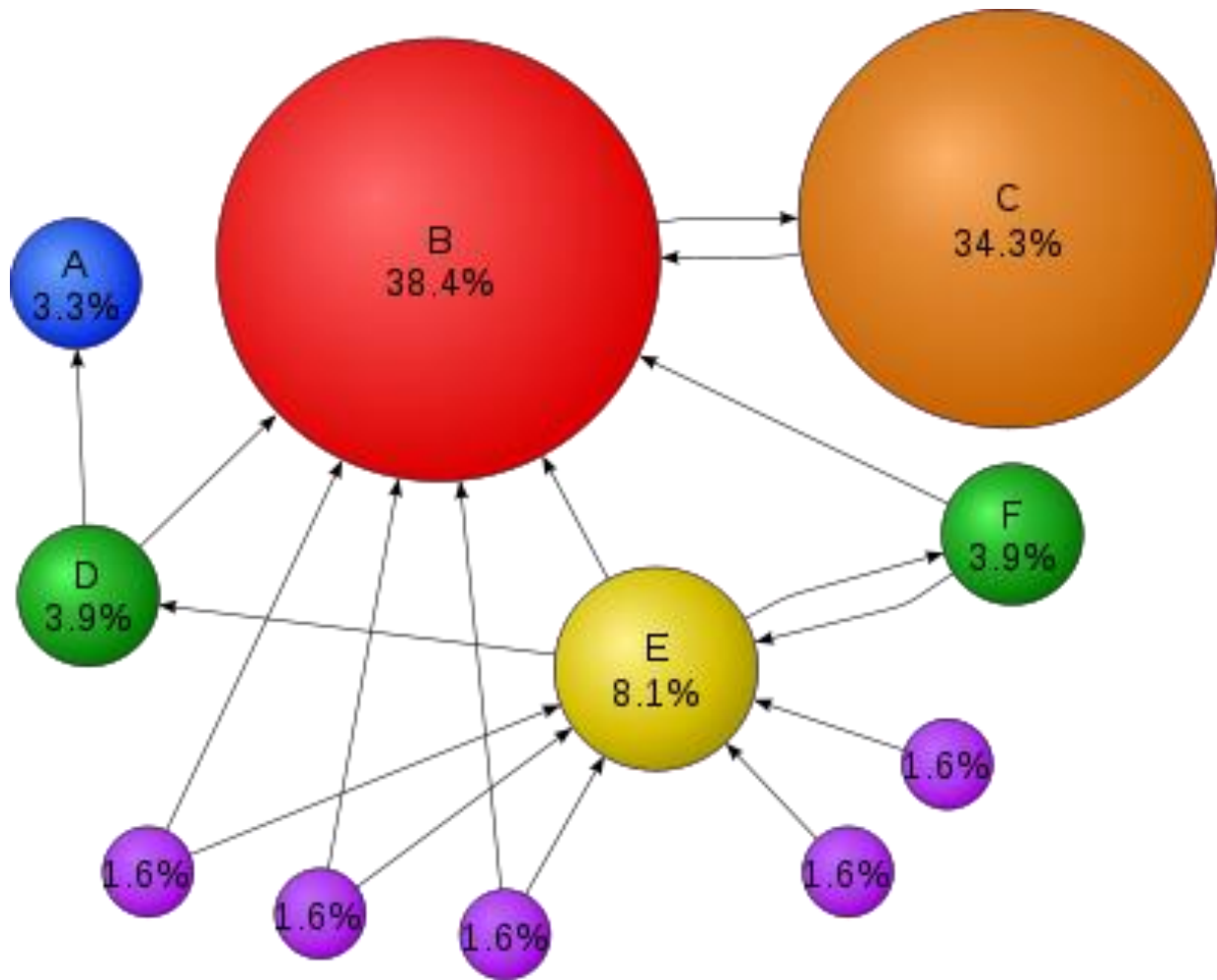
Page et al (1999)

Size of a node is proportional to the sum of the sizes of nodes pointing to it.

Two limits:

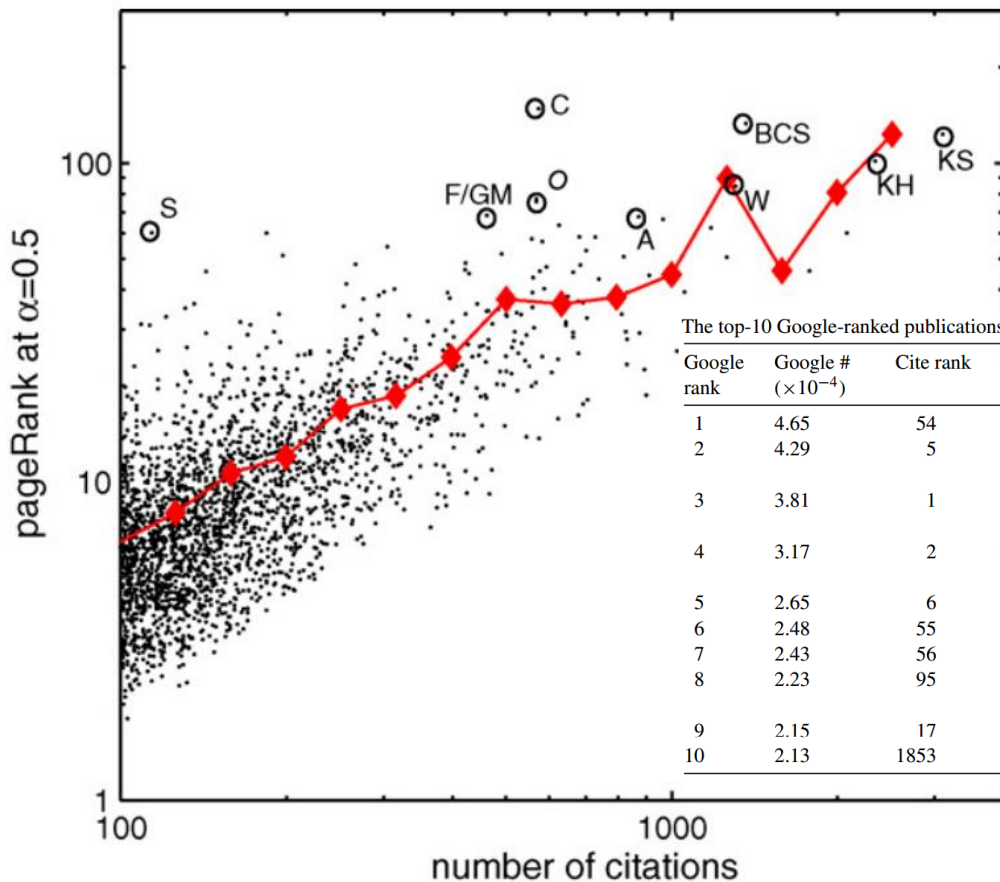
Slope  $\rightarrow 0$

Slope  $\rightarrow \text{inf}$



# CiteRank

Walker et al, J Stat Mech (2007), Chen et al, J Informetrics (2007) (S. Maslov & S. Redner groups)  
 Phys Rev internal citation network: 350k papers, 3.1M citation links



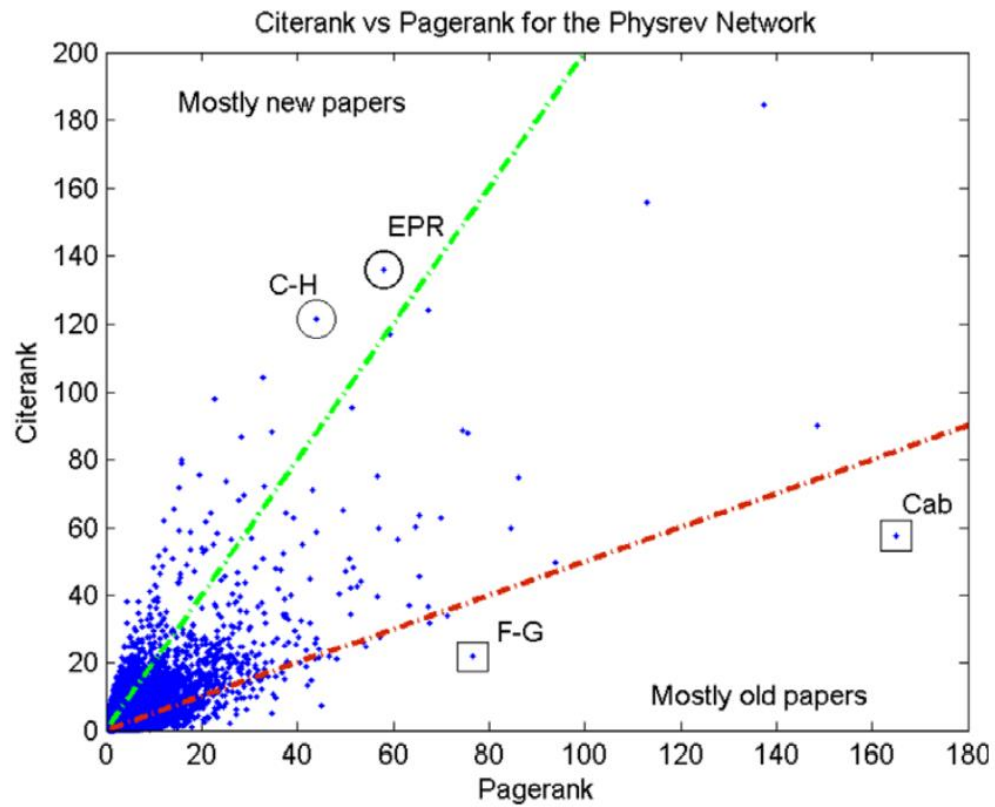
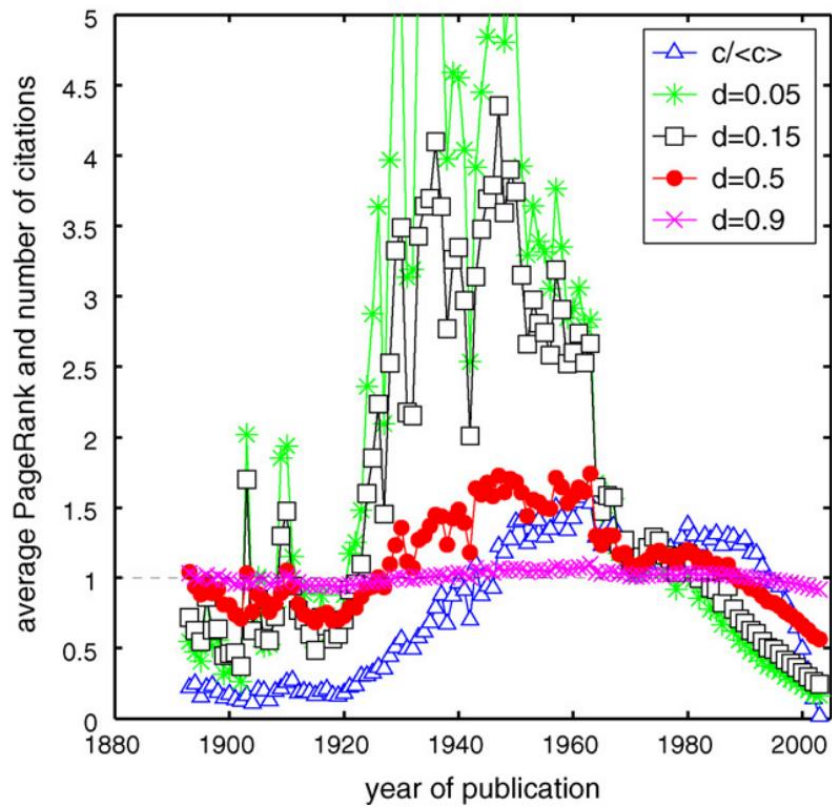
The top-10 Google-ranked publications when  $d = 0.5$

Google rank	Google # ( $\times 10^{-4}$ )	Cite rank	# cites	Publication	Title	Author(s)
1	4.65	54	574	PRL 10 531 1963	Unitary symmetry and leptonic...	N. Cabibbo
2	4.29	5	1364	PR 108 1175 1957	Theory of superconductivity	J. Bardeen, L. Cooper, and J. Schrieffer
3	3.81	1	3227	PR 140 A1133 1965	Self-consistent equations...	W. Kohn and L.J. Sham
4	3.17	2	2460	PR 136 B864 1964	Inhomogeneous electron gas	P. Hohenberg and W. Kohn
5	2.65	6	1306	PRL 19 1264 1967	A model of leptons	S. Weinberg
6	2.48	55	568	PR 65 117 1944	Crystal statistics I	L. Onsager
7	2.43	56	568	RMP 15 1 1943	Stochastic problems in...	S. Chandrasekhar
8	2.23	95	462	PR 109 193 1958	Theory of the Fermi interaction	R.P. Feynman and M. Gell-Mann
9	2.15	17	871	PR 109 1492 1958	Absence of diffusion in...	P.W. Anderson
10	2.13	1853	114	PR 34 1293 1929	The theory of complex spectra	J.C. Slater

# CiteRank

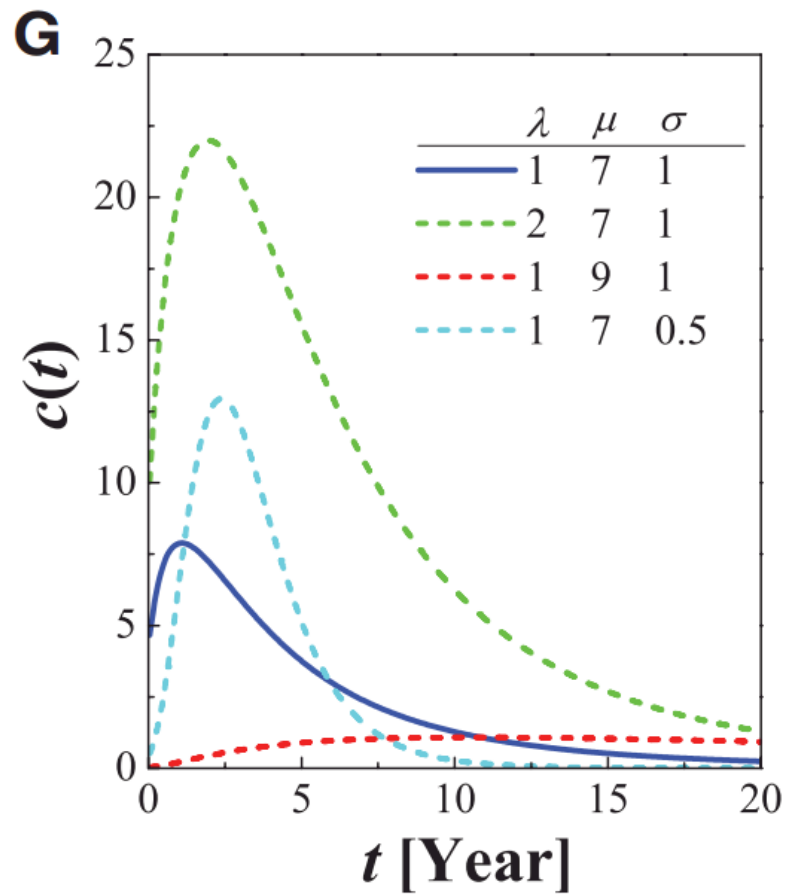
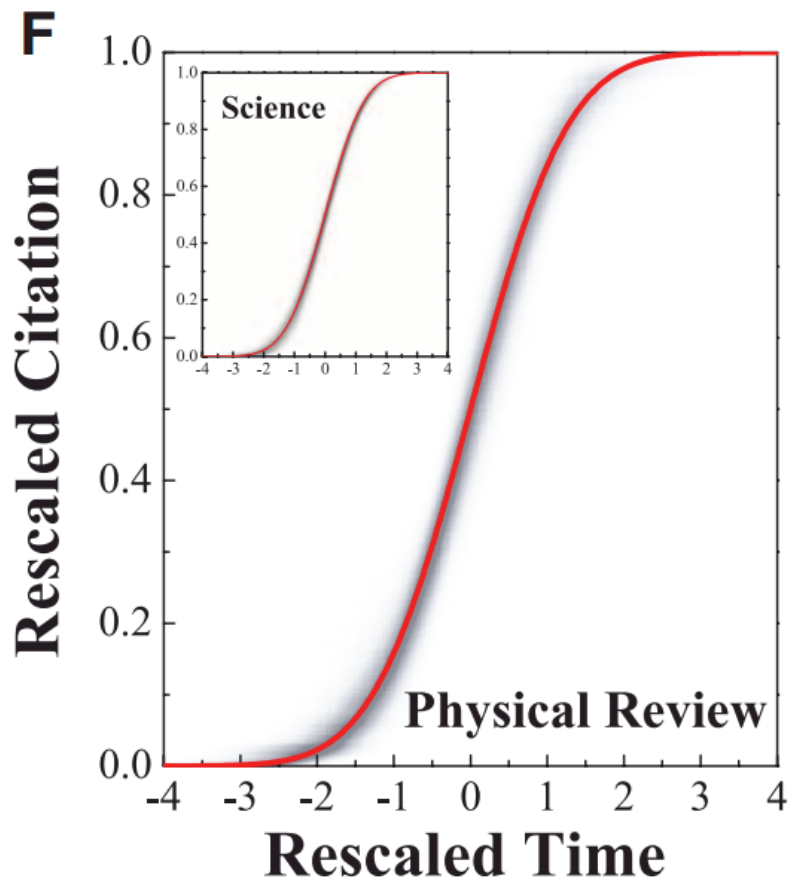
**PageRank:** Start walk from any node

**CiteRank:** from recent nodes



# MiC (Minimal Citation model)

Wang, Song, Barabasi (2013) *Science*



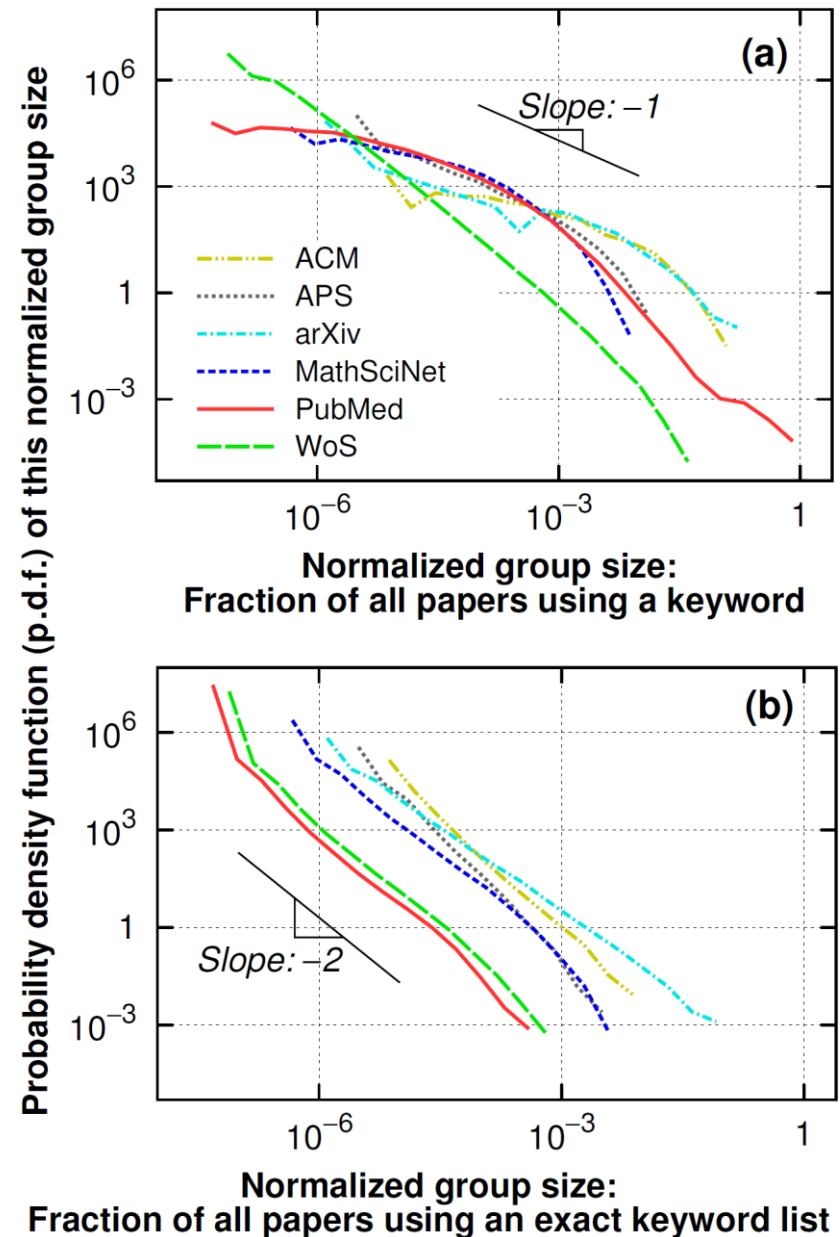
## Identification of topics based on article keywords

Two papers (articles) can be compared, if:

- (1) Their publication years are close
- (2) AND (a) They share at least one keyword
- (b) OR Their keyword lists are identical

**Thematically homogeneous groups of articles are dominantly small**

**→ Statistical errors are high**



## Tag usage distributions on other articles, posts, etc.



Goal: Identify developing communities of articles and  
the initiators of these communities

# Milyen pontosan mérhető a tudományos cikkek teljesítménye a hivatkozás számaik és a kulcsszavaik segítségével?

Farkas Illés

MTA-ELTE Statisztikus és Biológiai Fizika Kutatócsoport



This work was partially supported by the  
European Union and the European Social Fund  
through project FuturICT.hu  
(grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013).

National Development Agency  
[www.ujszachenyiterv.gov.hu](http://www.ujszachenyiterv.gov.hu)  
06 40 638 638



The project is supported by the European Union  
and co-financed by the European Social Fund.