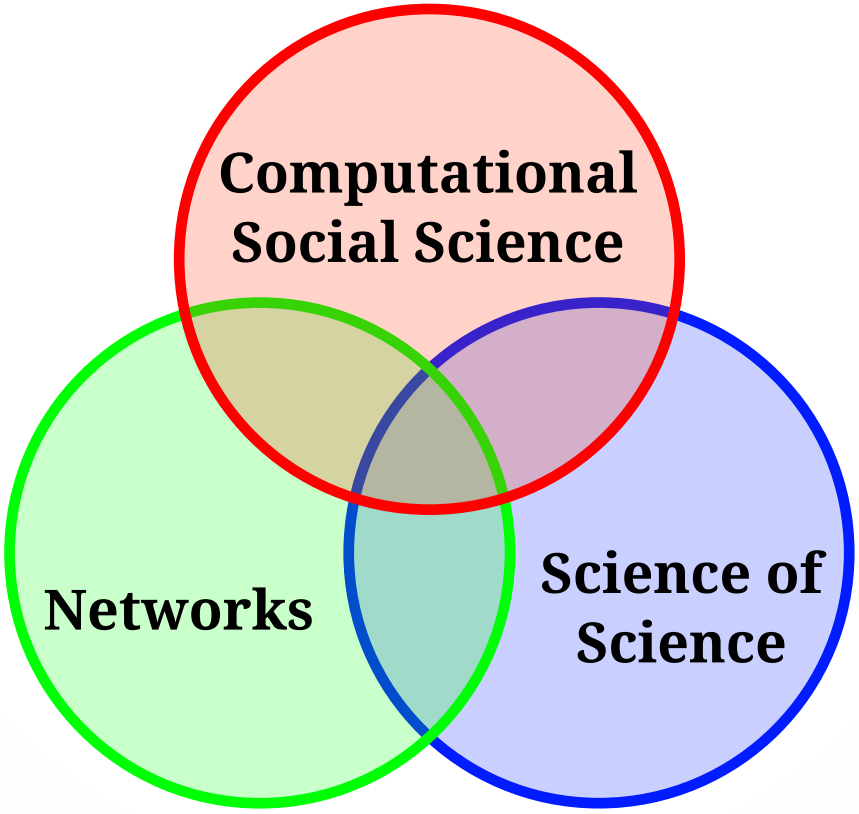
Science of science: universality, reputation, attention, inflation and novelty

Santo Fortunato

INDIANA UNIVERSITY

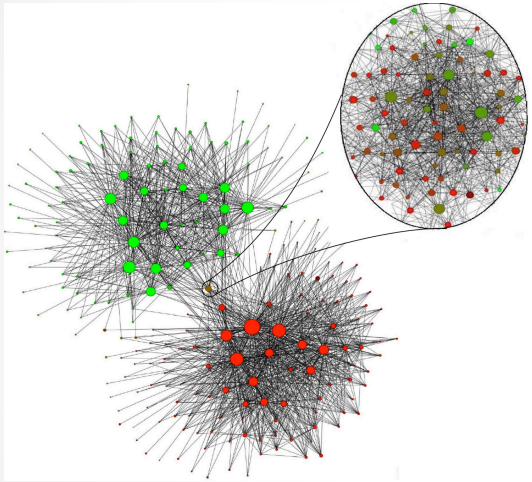
Me in a nutshell…



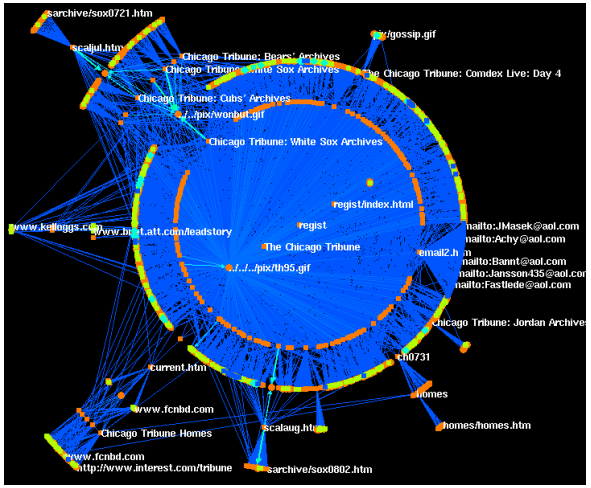
Network science

- **Analysis and modeling**

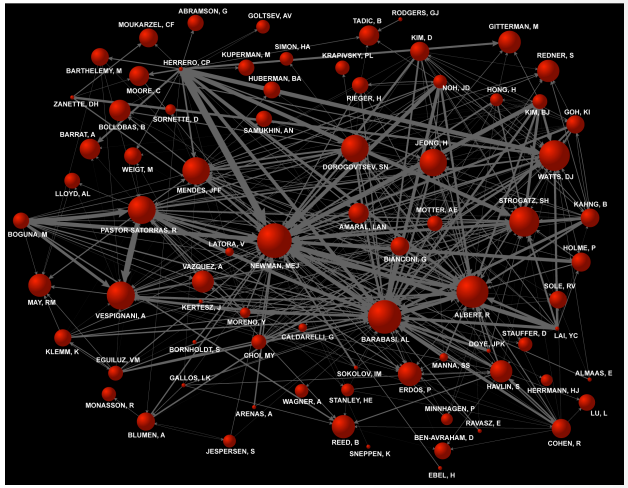
Social



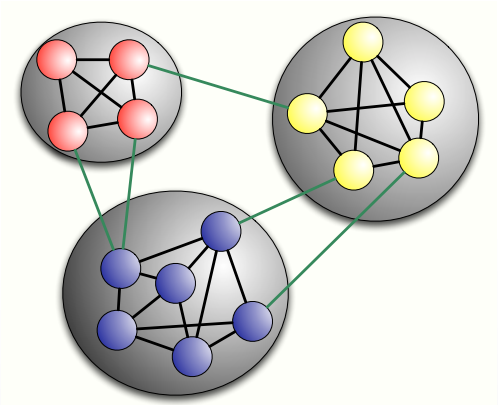
Information: WWW



Information: Citation



- **Community structure**



Network clustering

Physics Reports 486 (2010) 75–174

Contents lists available at ScienceDirect

Physics Reports

journal homepage: www.elsevier.com/locate/physrep

Community detection in graphs

Santo Fortunato *

Complex Networks and Systems Lagrange Laboratory, ISI Foundation, Viale S. Severo 65, 10133, Torino, I, Italy

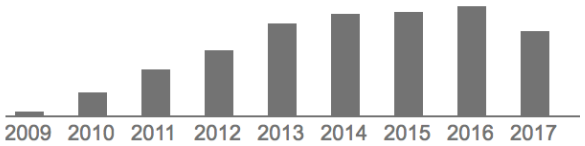
ARTICLE INFO

Article history:
Accepted 5 November 2009
Available online 4 December 2009
editor: I. Procaccia

ABSTRACT

The modern science of networks has brought significant advances to our understanding of complex systems. One of the most relevant features of graphs representing real systems is community structure, or clustering, i.e. the organization of vertices in clusters, with many edges joining vertices of the same cluster and comparatively few edges joining

Total citations Cited by 5914



Scholar articles Community detection in graphs
S Fortunato - Physics reports, 2010
Cited by 5914 Related articles All 44 versions

* Tel.: +39 011 6603090; fax: +39 011 6600049.
E-mail address: fortunato@isi.it.

0370-1573/$ – see front matter
doi: 10.1016/j.physrep.2009.11.002

Physics Reports 659 (2016) 1–44

Contents lists available at ScienceDirect

Physics Reports

journal homepage: www.elsevier.com/locate/physrep

Community detection in networks: A user guide

Santo Fortunato [a,b,*], Darko Hric [b]

[a] *Center for Complex Networks and Systems Research, School of Informatics and Computing, and Indiana University Network Science Institute (IUNI), Indiana University, Bloomington, USA*
[b] *Department of Computer Science, Aalto University School of Science, P.O. Box 15400, FI-00076, Finland*

ARTICLE INFO

Article history:
Accepted 26 September 2016
Available online 6 October 2016
editor: I. Procaccia

Keywords:
Networks
Communities
Clustering

ABSTRACT

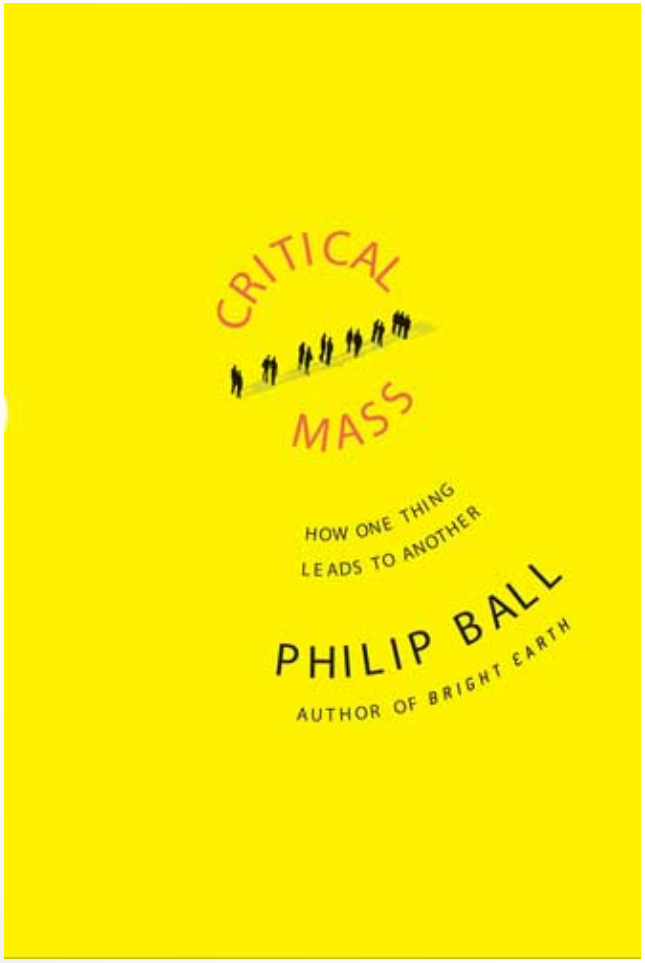
Community detection in networks is one of the most popular topics of modern network science. Communities, or clusters, are usually groups of vertices having higher probability of being connected to each other than to members of other groups, though other patterns are possible. Identifying communities is an ill-defined problem. There are no universal protocols on the fundamental ingredients, like the definition of community itself, nor on other crucial issues, like the validation of algorithms and the comparison of their performances. This has generated a number of confusions and misconceptions, which undermine the progress in the field. We offer a guided tour through the main aspects of the problem. We also point out strengths and weaknesses of popular methods, and give directions to their use.

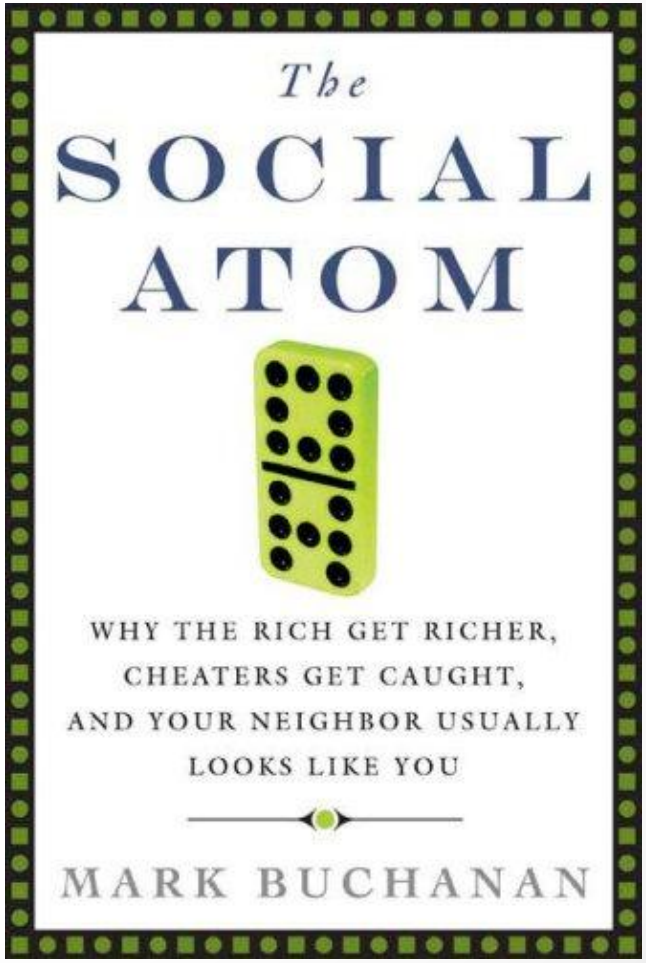
Contents

* Corresponding author at: Center for Complex Networks and Systems Research, School of Informatics and Computing, and Indiana University Network Science Institute (IUNI), Indiana University, Bloomington, USA.
E-mail address: santo@indiana.edu (S. Fortunato).

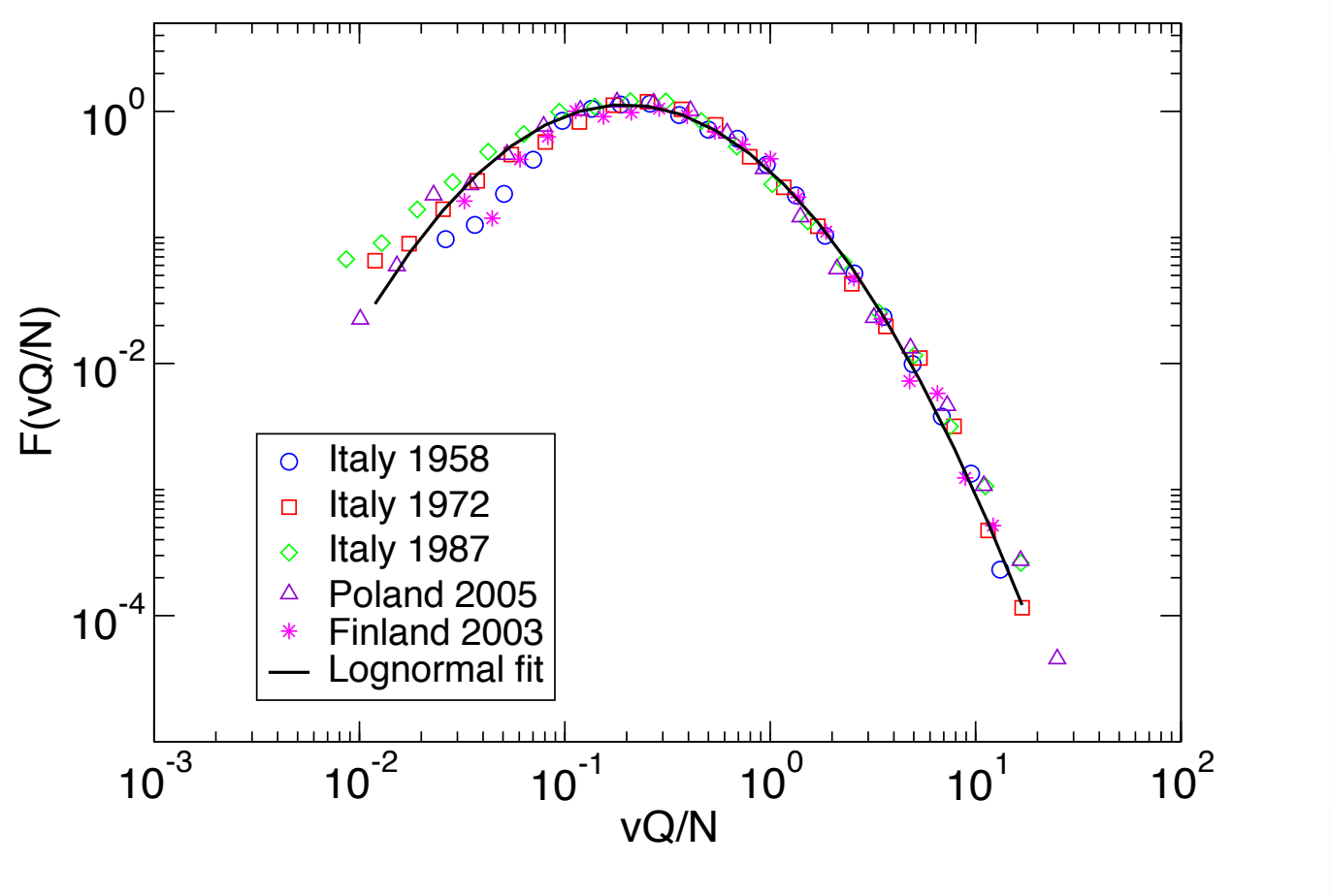
http://dx.doi.org/10.1016/j.physrep.2016.09.002
0370-1573/

Computational social science





Universality in elections



$$F(x) = \frac{1}{\sqrt{2\pi}\,\sigma x} e^{-(\ln(x)-\mu)^2 / 2\sigma^2}$$

$$\mu = -0.45$$

$$\sigma^2 = 0.91$$

Computational social science

REVIEWS OF MODERN PHYSICS, VOLUME 81, APRIL–JUNE 2009

Statistical physics of social dynamics

Claudio Castellano*

SMC, INFM-CNR and Dipartimento di Fisica, "Sapienza" Università di Roma, Piazzale A. Moro 2, 00185 Roma, Italy

Santo Fortunato†

Complex Networks Lagrange Laboratory, ISI Foundation, Viale S. Severo 65, 10133 Torino, Italy

Vittorio Loreto‡

Dipartimento di Fisica, "Sapienza" Università di Roma and SMC, INFM-CNR, Piazzale A. Moro 2, 00185 Roma, Italy
and Complex Networks Lagrange Laboratory, ISI Foundation, Viale S. Severo 65, 10133 Torino, Italy

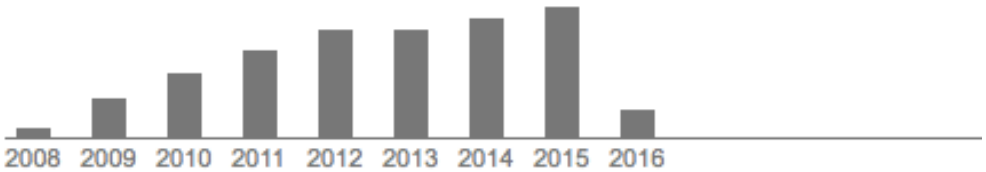
(Published 11 May 2009)

Statistical physics has proven to be a fruitful framework to describe phenomena outside the realm of traditional physics. Recent years have witnessed an attempt by physicists to study collective phenomena emerging from the interactions of individuals as elementary units in social structures. A wide list of topics are reviewed ranging from opinion and cultural and language dynamics to crowd behavior, hierarchy formation, human dynamics, and social spreading. The connections between these problems and other, more traditional, topics of statistical physics are highlighted. Comparison of model results with empirical data from social systems are also emphasized.

DOI: 10.1103/RevModPhys.81.591 PACS number(s): 05.10.−a, 89.20.−a, 89.75.−k

CONTENTS

Total citations Cited by 1912



Scholar articles Statistical physics of social dynamics
C Castellano, S Fortunato, V Loreto - Reviews of modern physics, 2009
Cited by 1912 - Related articles - All 37 versions

SCIENCE WATCH

HOME ABOUT THOMSON REUTERS PRESS RO

ScienceWatch.com › Data & Rankings › New Hot Papers

NEW HOT PAPERS

Outline

- Citation distributions: universality
- Modeling citation dynamics
- The bad consequences of publish or perish:

  a) attention decay

  b) inflation of citation's value
- Novelty: a network-based approach
- The long wait for the Nobel Prize

Citation distributions

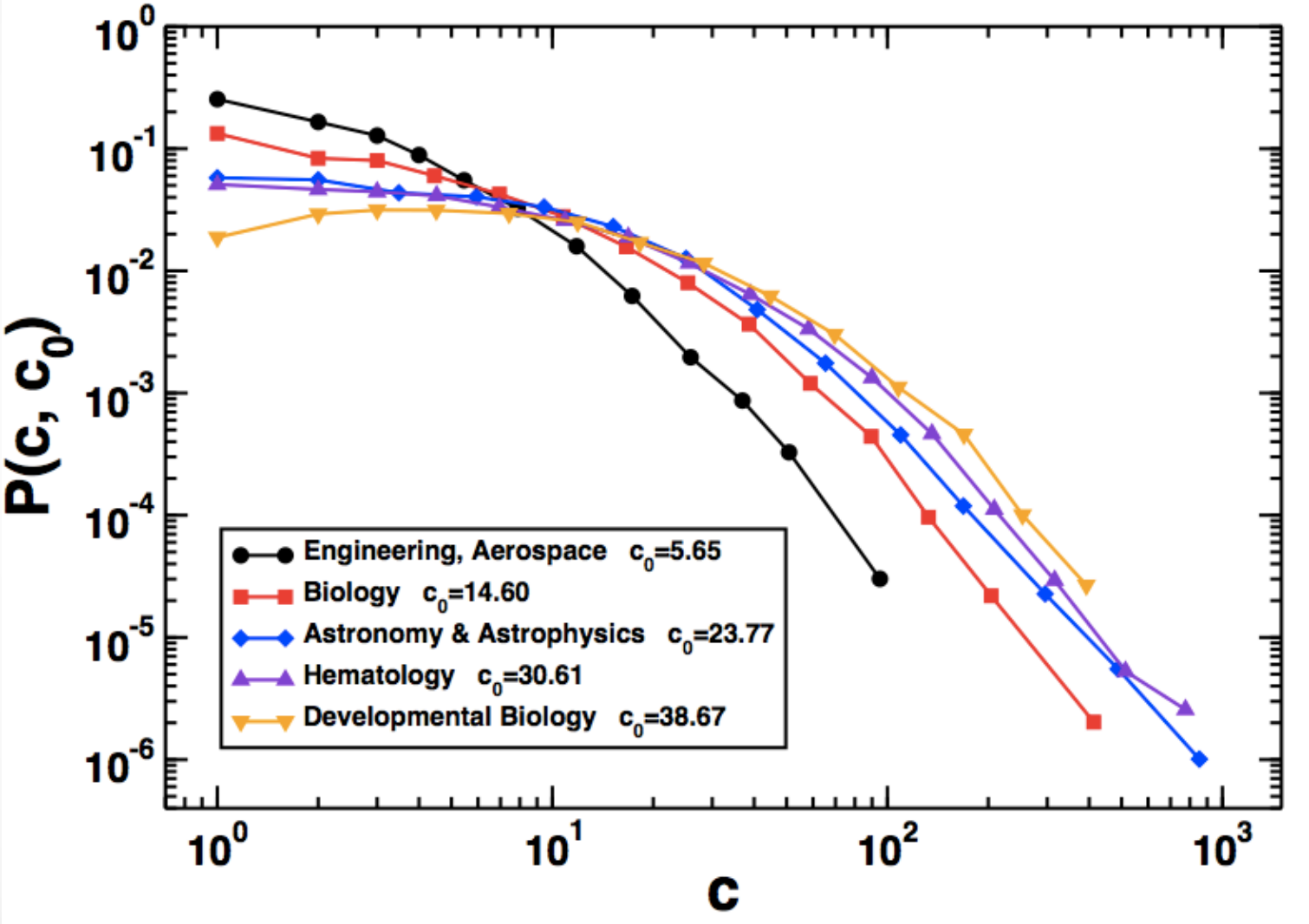
WEB OF KNOWLEDGE℠

THOMSON REUTERS

Index	Subject category	Year	N_p	c_0	c_{max}	σ^2	χ^2/df
1	Agricultural economics and policy	1999	266	6.88	42	1.0 (1)	0.007
2	Allergy	1999	1,530	17.39	271	1.4 (2)	0.012
3	Anesthesiology	1999	3,472	13.25	282	1.8 (2)	0.009
4	Astronomy and astrophysics	1999	7,399	23.77	1,028	1.1 (1)	0.003
5	Biology	1999	3,400	14.6	413	1.3 (1)	0.004
6	Computer science, cybernetics	1999	704	8.49	100	1.3 (1)	0.004
7	Developmental biology	1999	2,982	38.67	520	1.3 (3)	0.002
8	Engineering, aerospace	1999	1,070	5.65	95	1.4 (1)	0.003
9	Hematology	1990	4,423	41.05	1,424	1.5 (1)	0.002
10	Hematology	1999	6,920	30.61	966	1.3 (1)	0.004
11	Hematology	2004	8,695	15.66	1,014	1.3 (1)	0.003
12	Mathematics	1999	8,440	5.97	191	1.3 (4)	0.001
13	Microbiology	1999	9,761	21.54	803	1.0 (1)	0.005
14	Neuroimaging	1990	444	25.26	518	1.1 (1)	0.004
15	Neuroimaging	1999	1,073	23.16	463	1.4 (1)	0.003
16	Neuroimaging	2004	1,395	12.68	132	1.1 (1)	0.005
17	Physics, nuclear	1990	3,670	13.75	387	1.4 (1)	0.001
18	Physics, nuclear	1999	3,965	10.92	434	1.4 (4)	0.001
19	Physics, nuclear	2004	4,164	6.94	218	1.4 (1)	0.001
20	Tropical medicine	1999	1,038	12.35	126	1.1 (1)	0.017

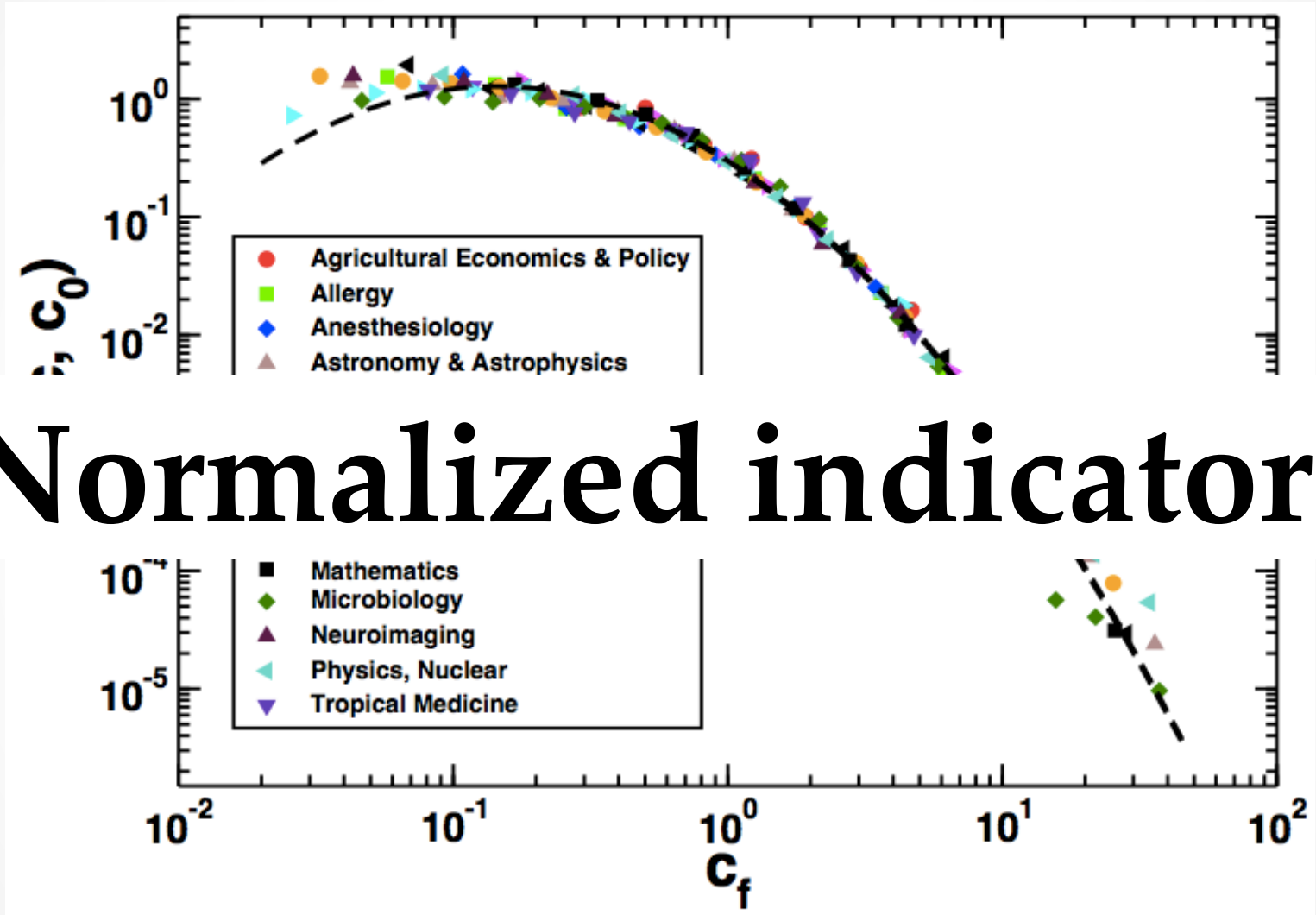
Papers classified in 172 scientific disciplines (from Acoustics to Zoology)

Distribution of cites



Field dependence (Thomson-Reuters category)!

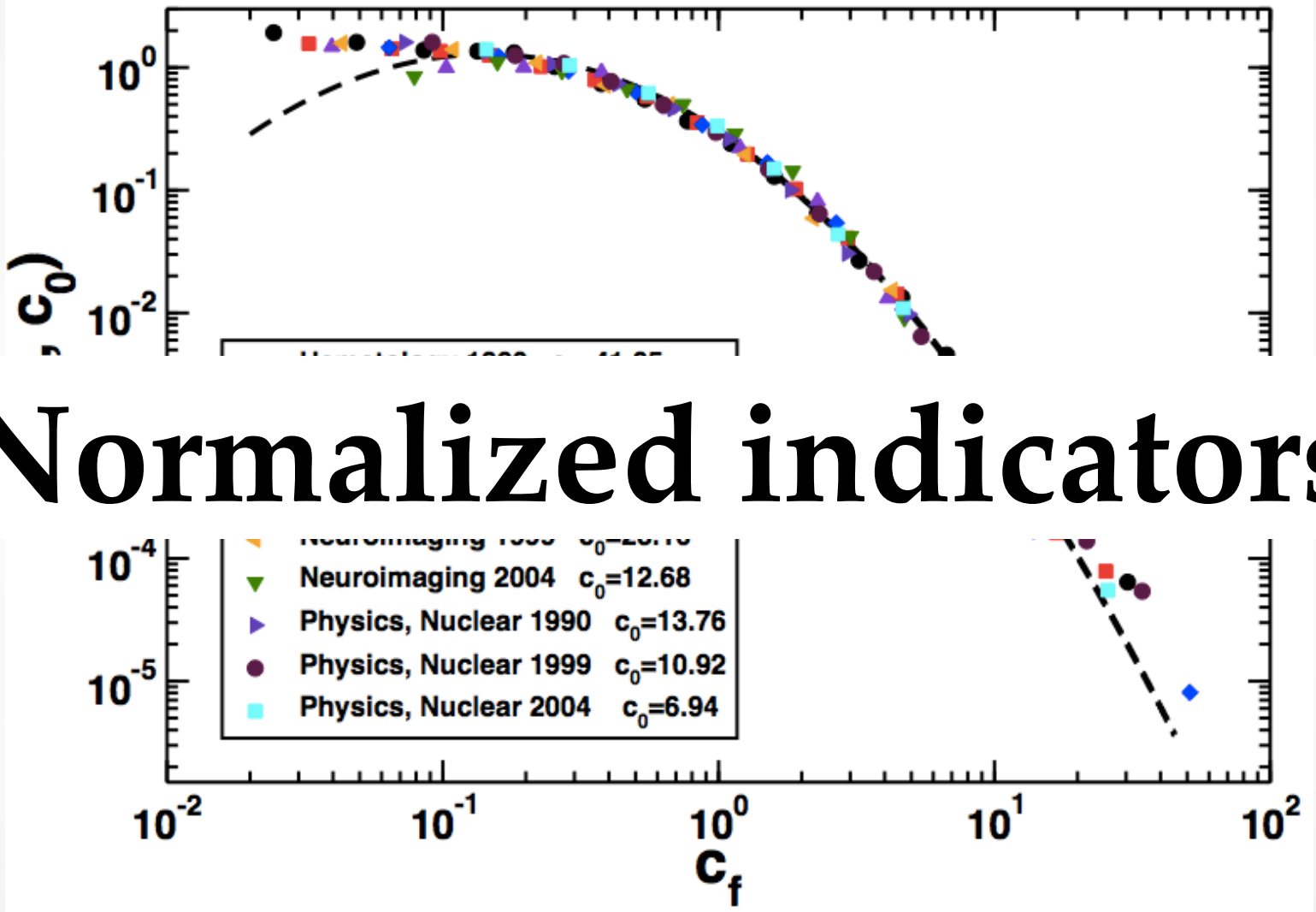
Could c_0 be the reason of the discrepancy?



Normalized indicators?

F. Radicchi, S.F. and C. Castellano,
Proc. Natl. Acad. Sci. USA 105, 17268-17272 (2008)

The universal distribution is **stable** in time!



Normalized indicators?

nature International weekly journal of science

comments on this story

Stories by subject
- Lab life
- Policy
- Mathematics

Stories by keywords
- Citation
- Bibliometrics
- Statistics
- h-index

This article elsewhere
Blogs linking to this article
Add to Digg

Published online 20 October 2008 | Nature | doi:10.1038/news.2008.1169

News

Is physics better than biology?

Citation statistics now comparable across disciplines.

Philip Ball

Is the physics department at your university performing better than the biology department? Answering such questions objectively has been hard, because citation statistics and other bibliometric indicators can't be directly compared across disciplines. But now a team in Italy has found a way to do just that.

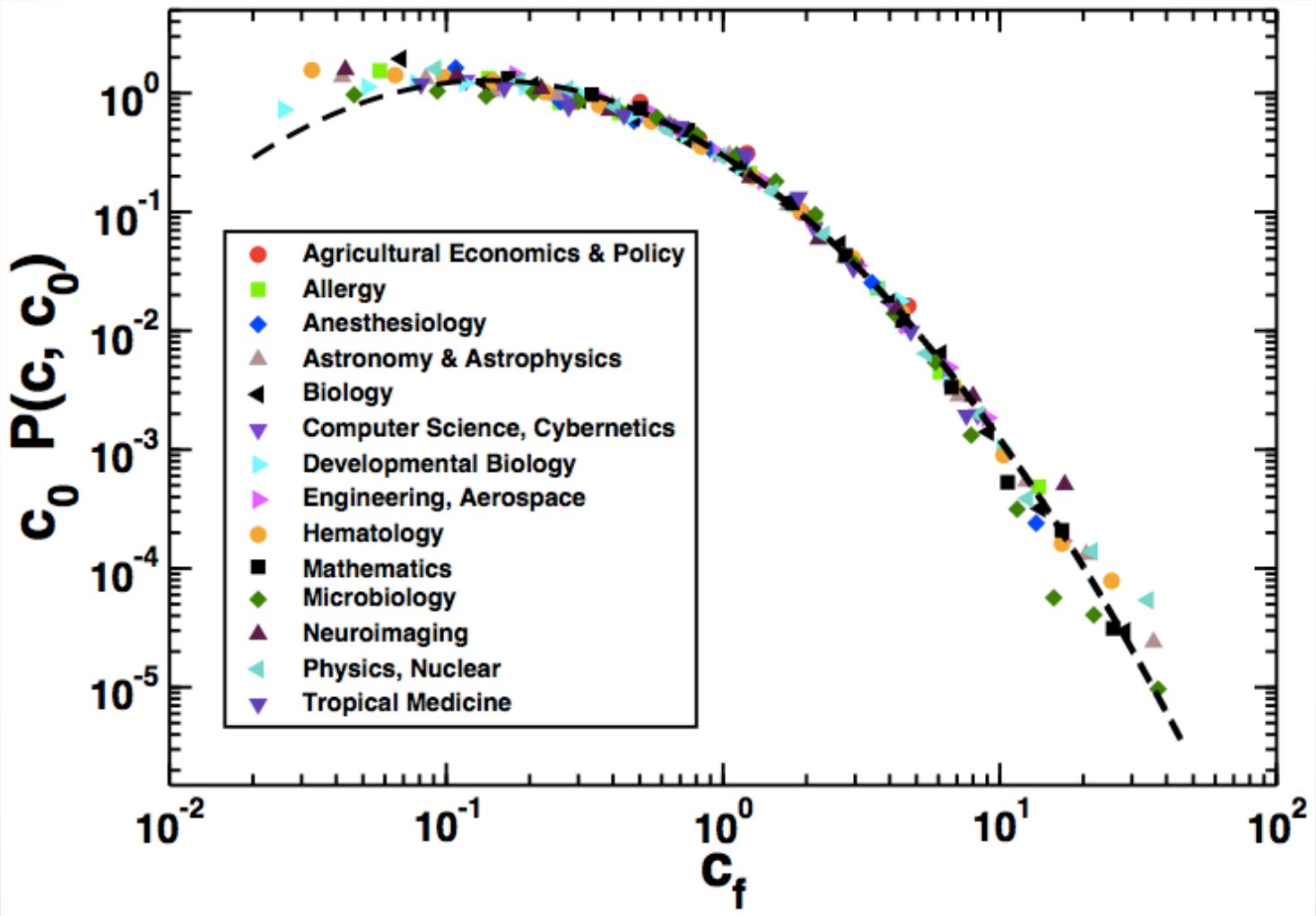
Claudio Castellano at the Sapienza University of

Relate
- A lo 17 S
- Who hori 17 J
- Free mar 02 J
- Cop 22 M

Nature
Tenu / Ass Univ Yale U
Scien Univer
Mo
Pos

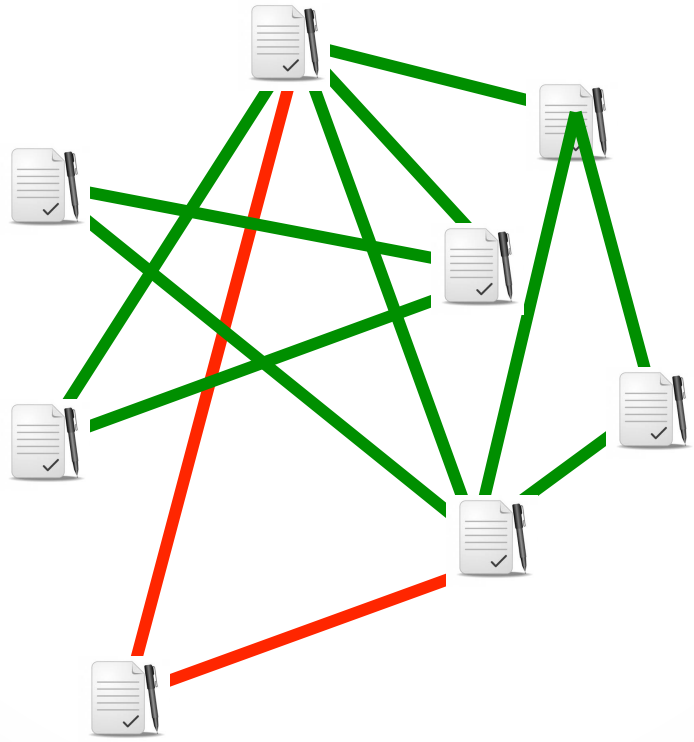
Fitting the universal distribution

Fat-tailed distribution (power law, lognormal)



Modeling citation dynamics

Question: What are the microscopic mechanisms that determine how papers cite each other?



Ingredients: preferential attachment

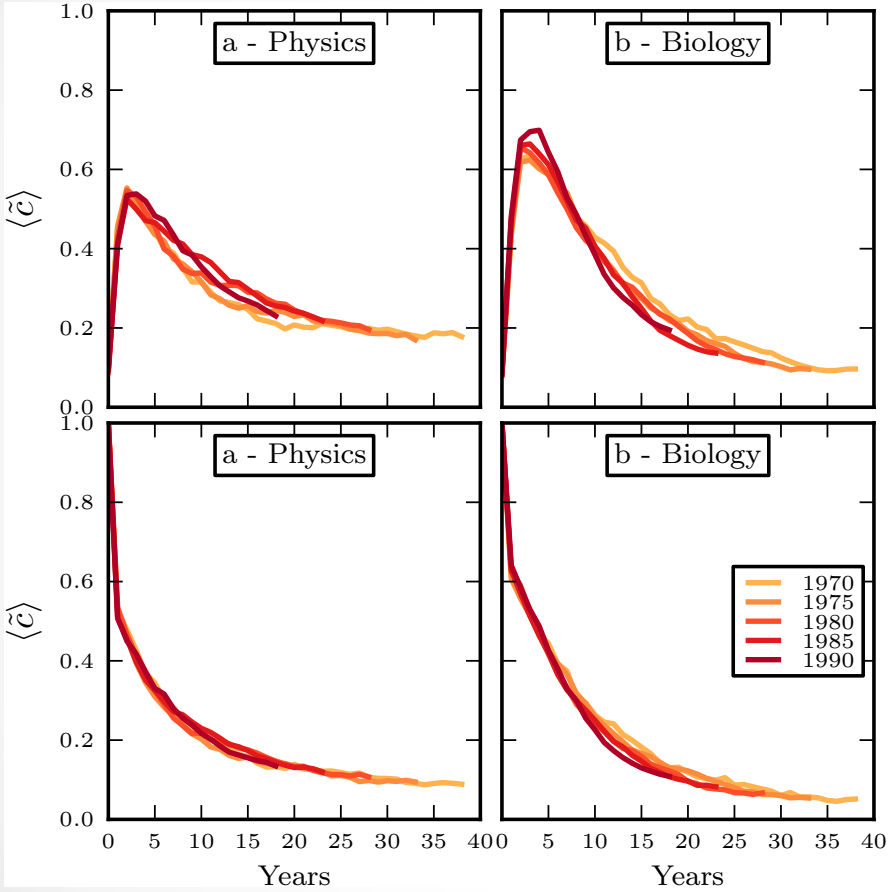
De Solla Price (1976), Barabási & Albert (1999)

Principle: probability for a new paper A to cite an older paper B is proportional to (a power of) the number of citations of B ("rich gets richer")

$$\Pi(A \to B) \propto [c_B]^{\pi}$$

Problem: the rate of citation accumulation grows with time!

Ingredients: life cycle



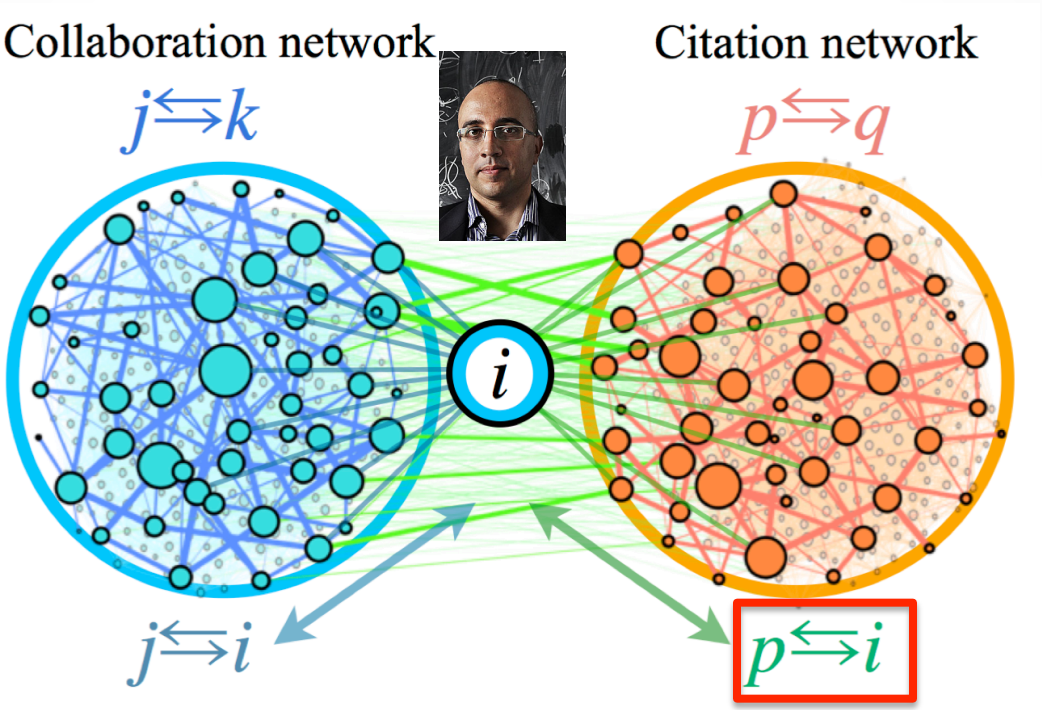
The peak of attention is reached about 5 years after publication, the decay after the peak is exponential

The citation probability should decrease with the difference between the publication times of the citing and the cited item

Ingredients: others

- Fitness (``intrinsic value'') of publications
- Reputation of authors
- Citations of citations
- Collaboration patterns
- Etc.

The role of reputation …



Question: what is the role of author's reputation on citation dynamics?

A. M. Petersen, S. F., R. K. Pan, K. Kaski, O. Penner, A. Rungi, M. Riccaboni, H. E. Stanley, F. Pammolli, Proc. Natl. Acad. Sci. USA 111, 15316-15321 (2014)

Reputation variable

Measure of reputation of author i at year t: total number of citations $C_i(t)$, from the beginning of i's career until year t

Issues:

1. It does not account for quality: high values of $C_i(t)$ can be achieved with many low-impact papers
2. Effect of reputation of coauthors?
3. It can only grow in time, while reputation may also decrease (misconduct, fraud)

A. M. Petersen, S. F., R. K. Pan, K. Kaski, O. Penner, A. Rungi, M. Riccaboni, H. E. Stanley, F. Pammolli, Proc. Natl. Acad. Sci. USA 111, 15316-15321 (2014)

Other variables

$t =$ Career age, number of years since first publication

$\tau =$ Paper age, number of years since first citation

$c_{i,p}(t) =$ Citations of paper p and author i until time t

$\Delta c_{i,p}(t) =$ Citations received by paper p of author i in year t

A. M. Petersen, S. F., R. K. Pan, K. Kaski, O. Penner, A. Rungi, M. Riccaboni, H. E. Stanley, F. Pammolli, Proc. Natl. Acad. Sci. USA 111, 15316-15321 (2014)

Our data

A. Prolific authors of *Physical Review Letters* (PRL) (1958-2008)
B. Highly productive physicists
C. Young assistant professors from US Top 50 physics and astronomy departments
D. Prolific authors of *Cell*
E. Prolific authors of *Annals of Mathematics*

Summary

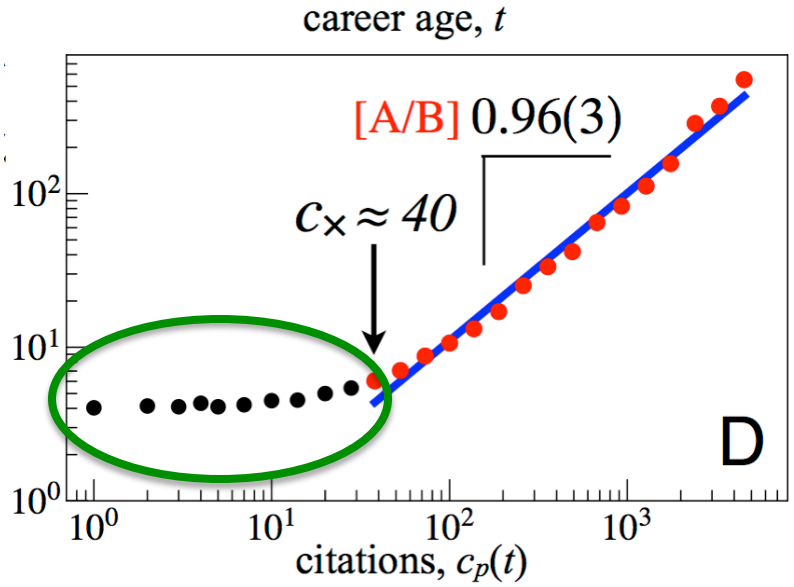
1. 300 physicists
2. 100 biologists
3. 50 mathematicians

A. M. Petersen, S. F., R. K. Pan, K. Kaski, O. Penner, A. Rungi, M. Riccaboni, H. E. Stanley, F. Pammolli, Proc. Natl. Acad. Sci. USA 111, 15316-15321 (2014)

The reputation model

$$\Delta c_{i,p}(t+1) \text{ vs } c_{i,p}(t) \text{ ?}$$

Preferential attachment?



For $c_p(t)$ lower than a threshold c_x the number of cites exceeds that predicted by preferential attachment: why?

The reputation model

Regression model

$$\Delta c_{i,p}(t+1) \equiv \eta \times \Pi_p(t) \times A_\rho(\tau) \times R_i(t) =$$
$$= \eta \, [c_p(t)]^\pi \, \exp(-\tau_p/\bar{\tau}) \, [C_i(t)]^\rho$$

noise

preferential attachment

life cycle

reputation

Main result: reputation term important before threshold c_x and irrelevant above c_x

The reputation model

Citations are initially boosted by author reputation up to a *tipping point* c_x above which the citation rate is sustained mostly by publication reputation

For two scientists such that $C_1(t)=10\ C_2(t)$:

$$\Delta c_1(t)/\Delta c_2(t) \approx 1.66 \qquad (c < c_x)$$

66 % increase for each 10-fold increase in $C_i(t)$!

NATURE | RESEARCH HIGHLIGHTS: SOCIAL SELECTION

Being a big name in science brings benefits

A study that links scientists' reputations with their citations triggers online talk.

Chris Woolston

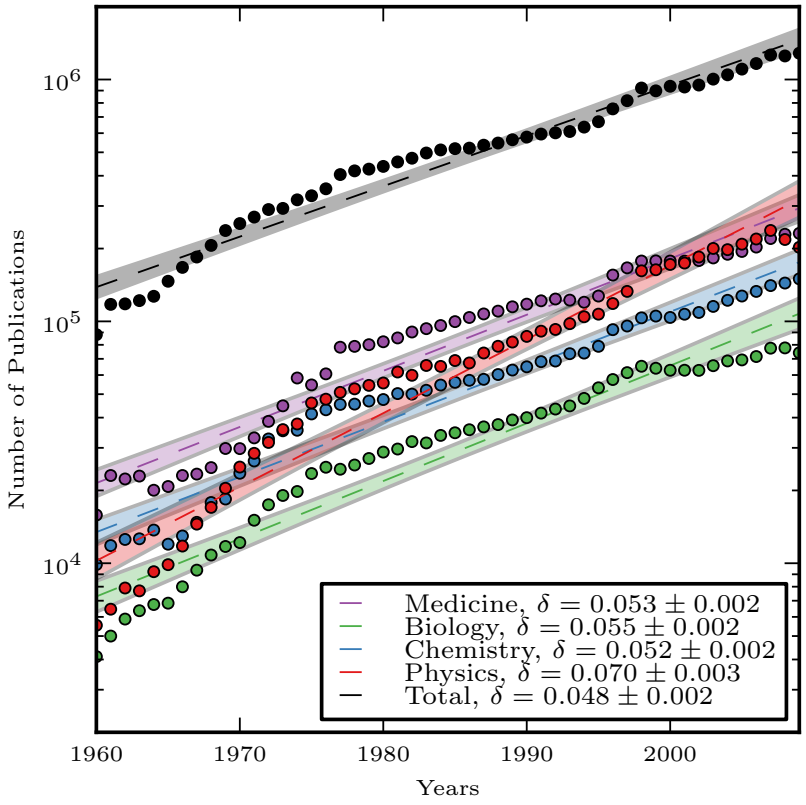
12 November 2014

Scientists develop reputations that often work to their advantage. A study suggests that the presence of a well-known scientist on a list of authors can drive citations of the paper, regardless of the merits of the research — especially soon after its publication. The report rapidly started an online discussion. "How scientists too

Based on data from Altmetric.com. Altmetric is

Publish or perish?

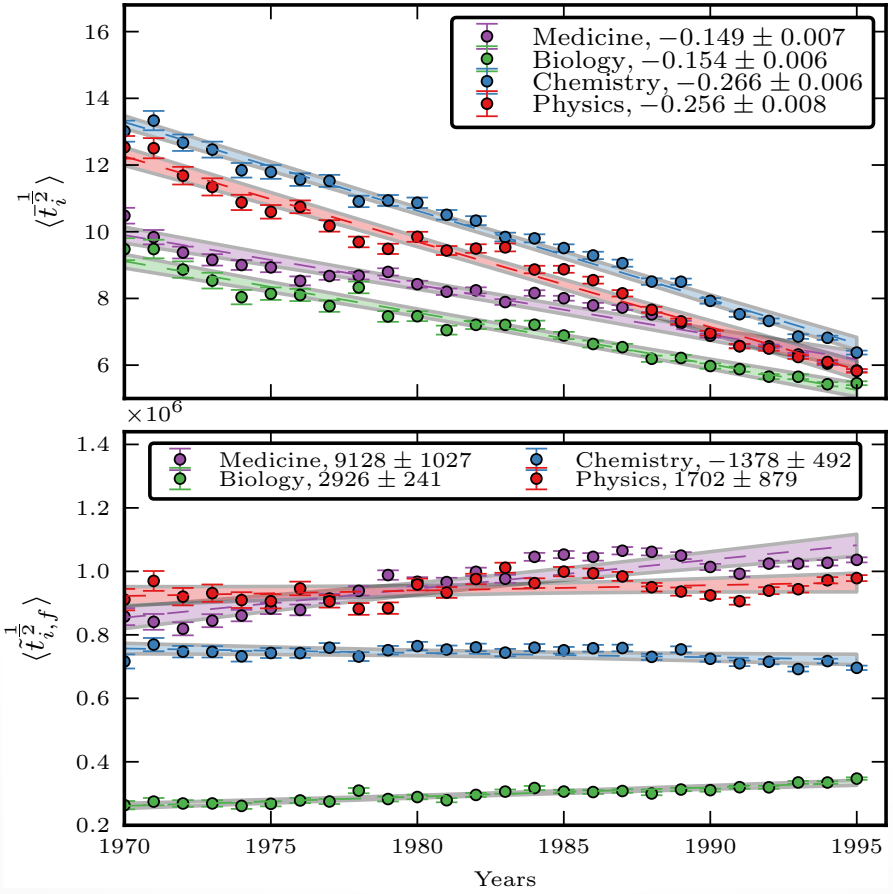
Fact: the number of papers increases exponentially in time



Question: which consequences does this have?

Attention decay

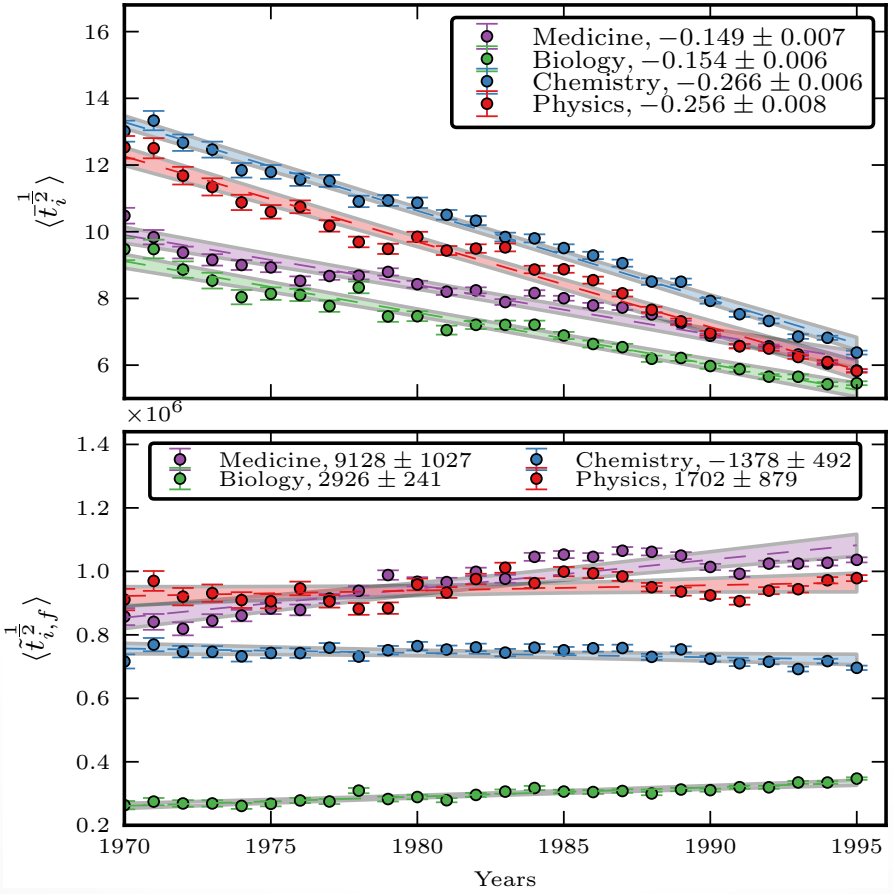
Half-life of a paper: number of years after publication when it collects half of the total number of cites it has now



Result: papers are forgotten more and more easily!

Attention decay

Rescaling: half-life expressed in terms of the number of papers published until half of the citations are collected



Result: half-life approximately stable in time!

Attention decay: summary

- Papers are getting forgotten faster and faster
- The natural unit of time for scholars is the number of published items
- We have a finite memory and keep track of approximately the same number of publications

P. D. B. Parolo, R. K. Pan, R. Ghosh, B. A. Huberman, K. Kaski and S.F., Journal of Informetrics 9, 734-745 (2015)

Mar 29, 2016 | Updated: 09:37 AM EST

The Science Times

New Study Finds That There Are Too Many Scientific Studies

Muneeb Kazi Mar 17, 2015 03:32 PM EDT

Print

According to a new study there are too many scientific studies. These increasing quantities of studies make it difficult for scientists and researchers to keep track of everything. The professors from universities in Finland and California have published a new paper entitled Attention decay in science.

Inflation in science

Example

Within the first 5 years from publication

- Einstein's 1905 paper on special relativity has received **15** citations
- 2004 paper by Geim, Novoselov et al. on graphene has received **2,108** citations!

Question: does contemporary science have more impact than ``classic'' science?

Answer: not really!

Inflation in science

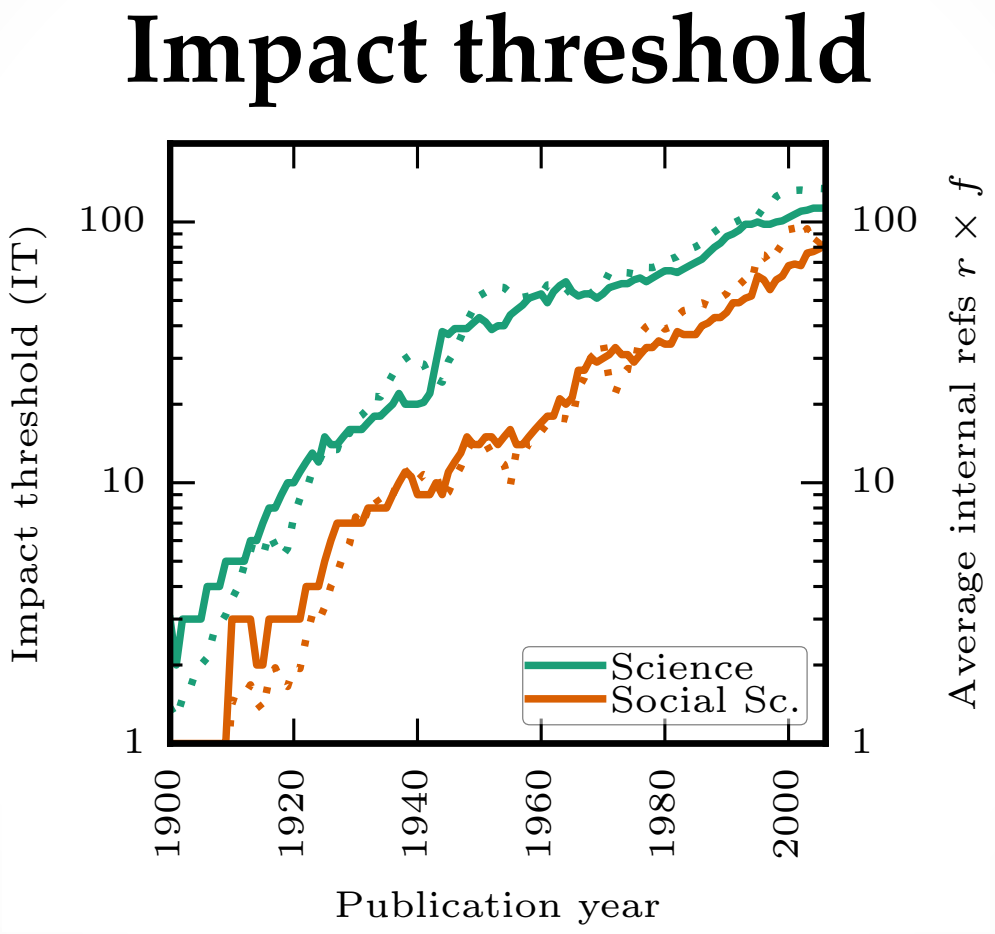
Citations represent the fundamental ***currency*** of scientific credit: what happens when more and more citations are put into the system?

Analogy with money in economics? Inflation?

Assessment: monitoring the ``price'' of a product in time, like in economics

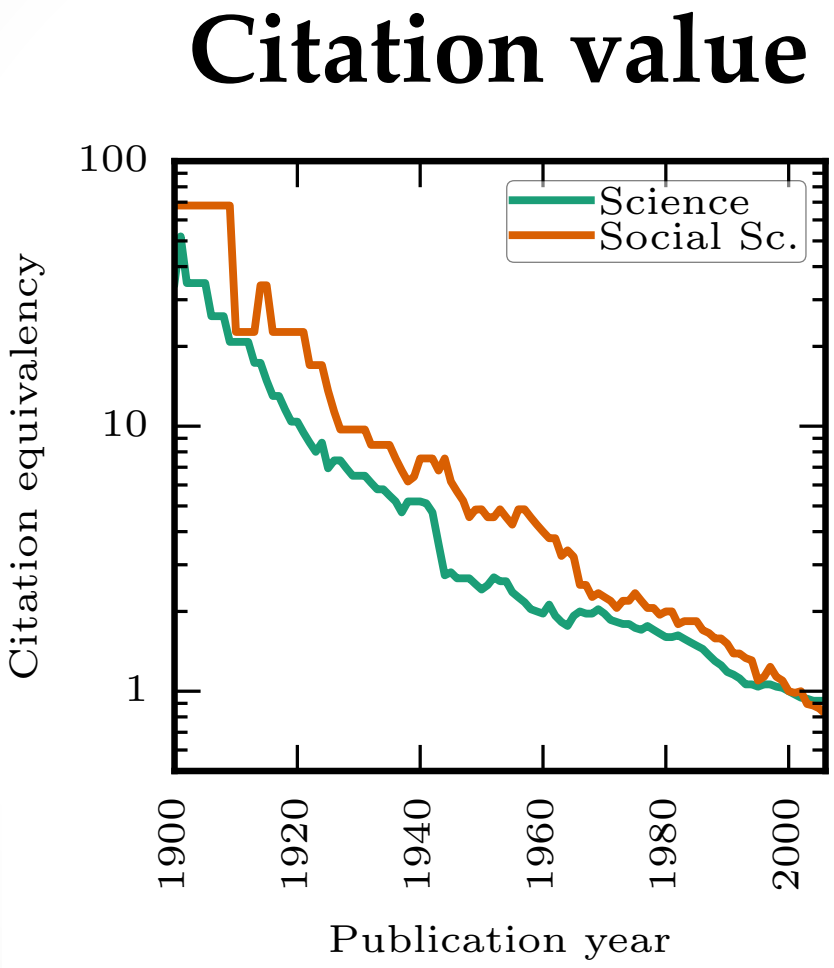
Impact threshold: (minimum) number of citations that a paper has to collect in its first 5 years to be ranked in the top 1% of papers published in the same year

Inflation in science



Result: the IT changed by two orders of magnitude over the past century!

Inflation in science



Result: in 2000 the value of a citation has gone down to about 0.02 % compared to that in 1900!

Inflation in science

Example

Within the first 5 years from publication

- Einstein's 1905 paper on special relativity has received **15** citations
- 1965 paper by Penzias and Wilson on cosmic radiowave background (CRB) has received **193** citations!

After correcting for inflation **Einstein leads by 520 to 372**

Driving factor: growth of the number of references per paper, not growth of the number of publications

Novelty

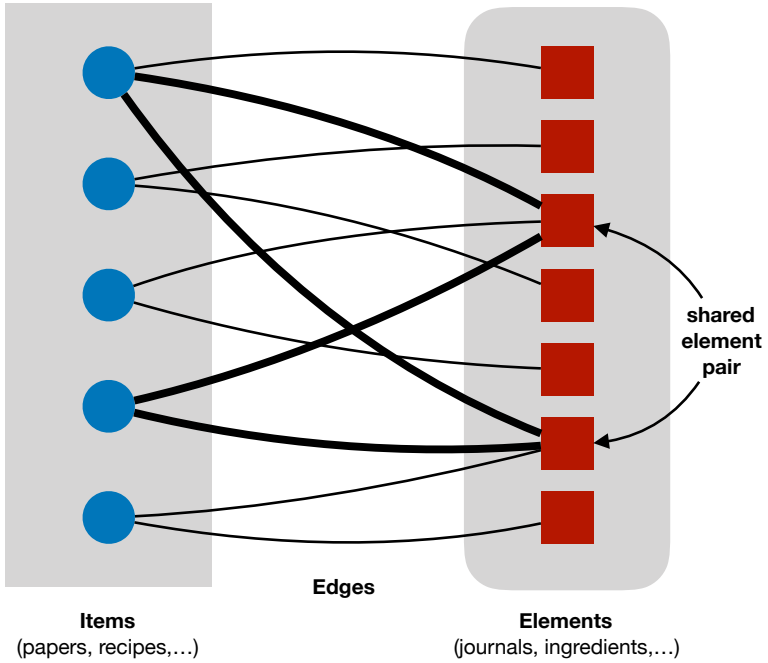
Question: how can we quantify novelty?

Answer: novelty usually derives from *original combinations* of existing elements

Example: for scientific papers, one could check whether there are anomalous pairings of journals in the list of references (Uzzi et al., 2013)

B. Uzzi, S. Mukherjee, M. Stringer, B. F. Jones, *Atypical combinations and scientific impact*. Science **342**, 468-472 (2013).

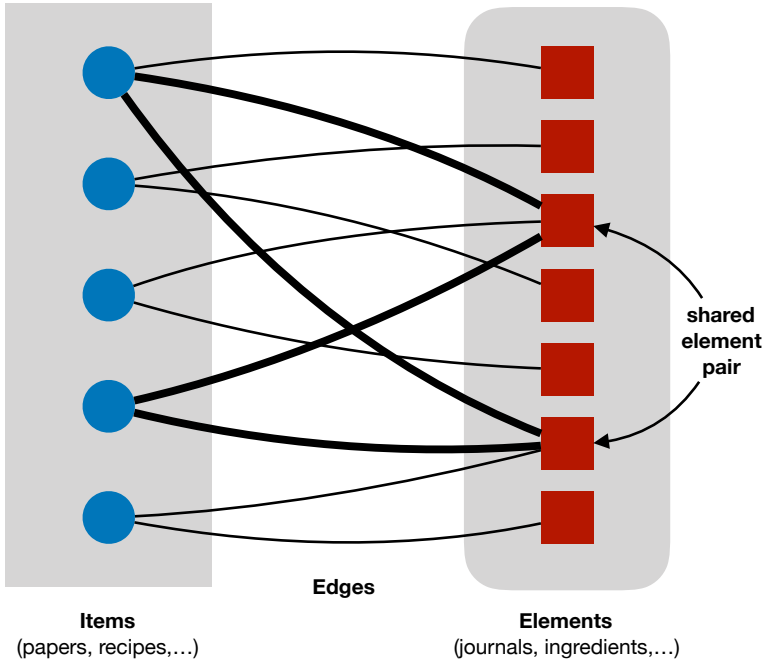
Network-based novelty



Principle:

1) an item is (un)conventional if it features (un)conventional pairings of elements
2) A pairing of elements is (un)conventional if it is featured by a(n) (un)conventional item

Network-based novelty



W_{ij} = 1 if item *i* features element *j*, 0 otherwise

r_i = number of elements of item *i*

n_p = number of items

n_j = number of elements

Network-based novelty

The equations

$$\lambda x_i = \frac{1}{r_i^2} \sum_{j<k}^{n_j} \frac{W_{ij} W_{ik} B_{jk}}{(\sum_{i=1}^{n_p} W_{ij})(\sum_{i=1}^{n_p} W_{ik})}$$

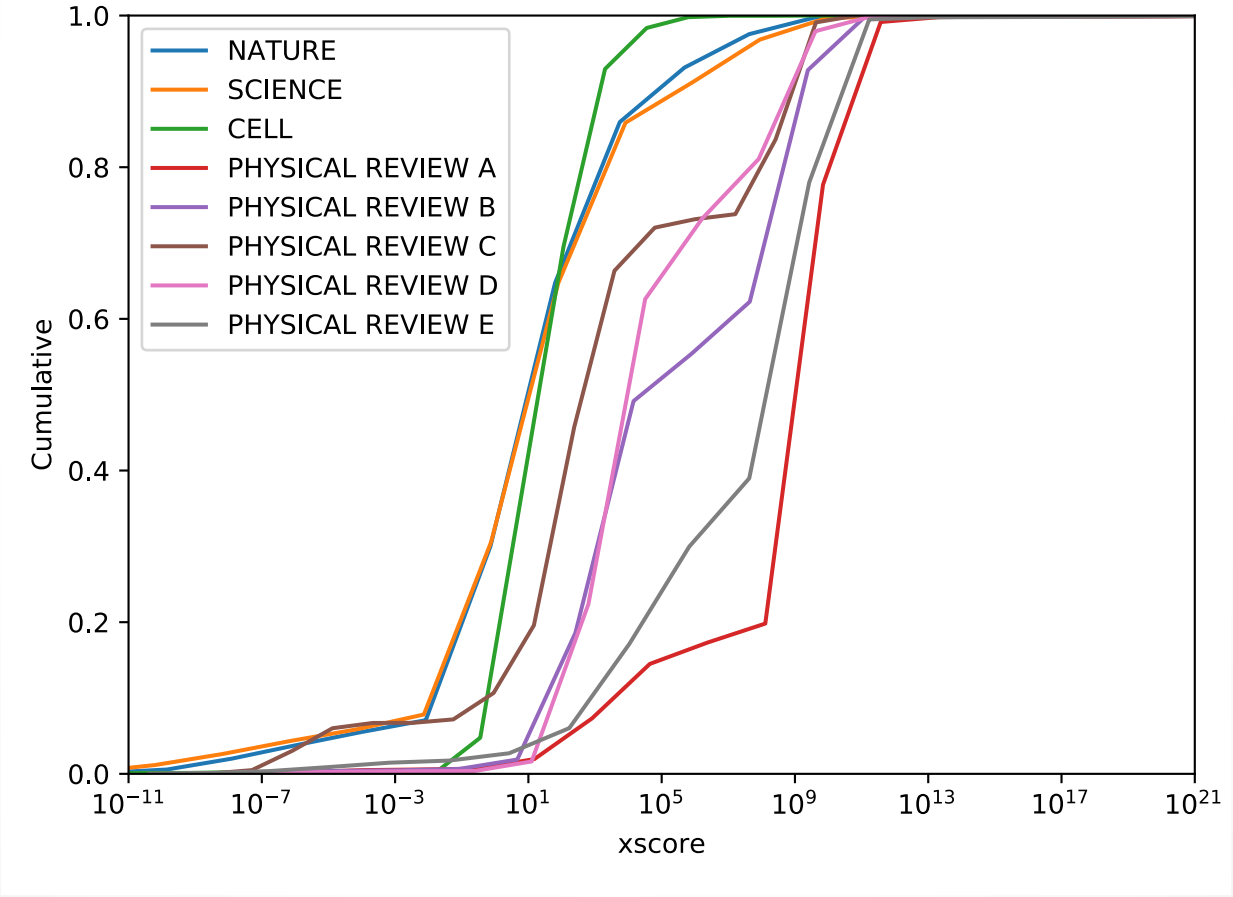
$$\mu B_{jk} = \frac{1}{(\sum_{i=1}^{n_p} W_{ij})(\sum_{i=1}^{n_p} W_{ik})} \sum_{i=1}^{n_p} \frac{W_{ij} W_{ik} x_i}{r_i^2}$$

x_i = conventionality of item i

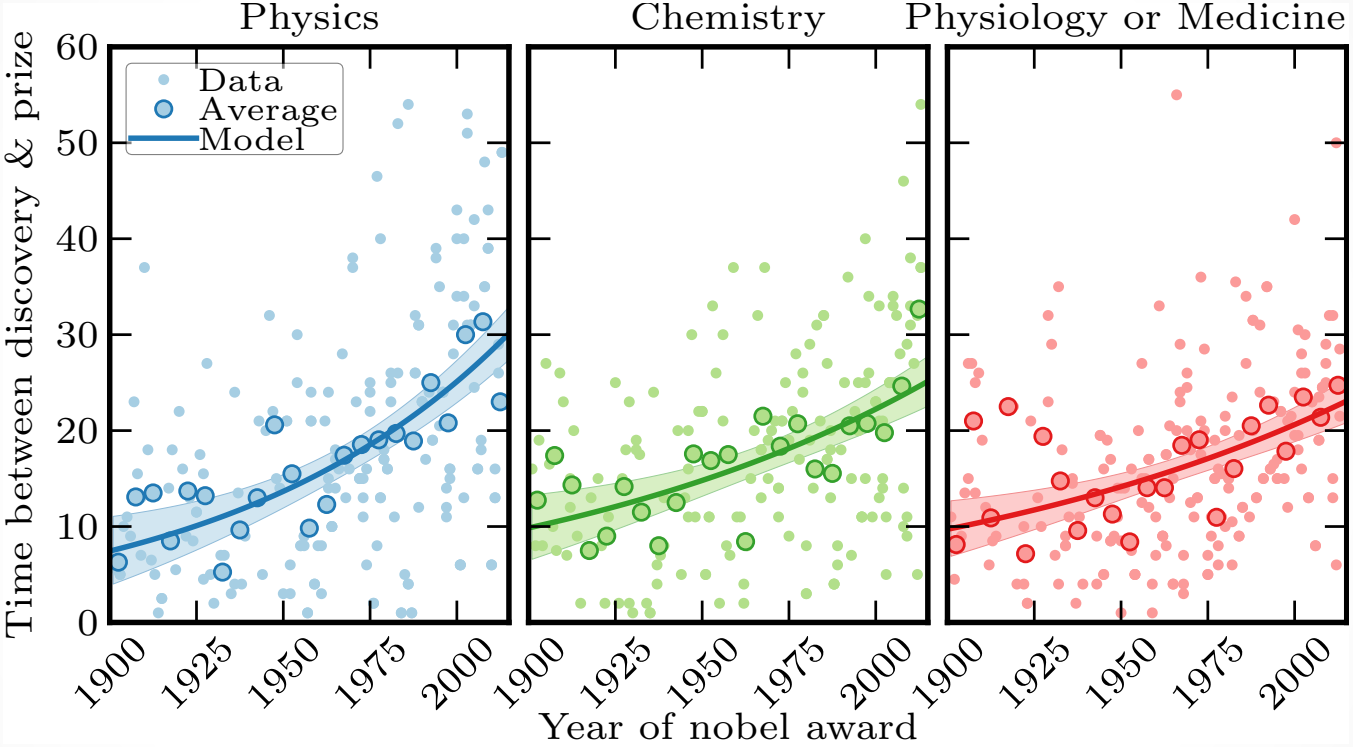
B_{jk} = conventionality of pairing of elements j and k

Preliminary results

Distribution of conventionality of papers

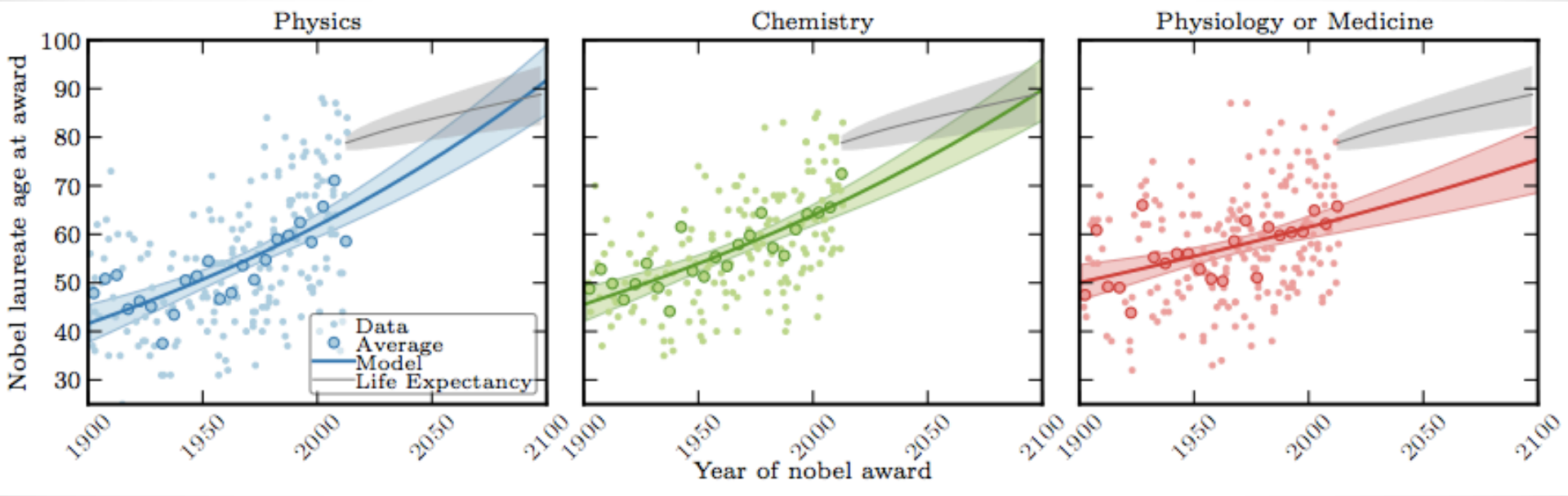


The Nobel Prize Delay



F. Becattini, A. Chatterjee, S. F, M. Mitrović, R. K. Pan, P. Della Briotta Parolo, Nature 508, 186 (2014)
http://scitation.aip.org/content/aip/magazine/physicstoday/news/10.1063/PT.5.2012

The Nobel Prize Delay



F. Becattini, A. Chatterjee, S. F, M. Mitrović, R. K. Pan, P. Della Briotta Parolo, Nature 508, 186 (2014) http://scitation.aip.org/content/aip/magazine/physicstoday/news/10.1063/PT.5.2012

ARTICLE PREVIEW

view full access options ▸

NATURE | CORRESPONDENCE

Prizes: Growing time lag threatens Nobels

Santo Fortunato

Nature **508**, 186 (10 April 2014) | doi:10.1038/508186a

Published online 09 April 2014

The time lag between reporting a scientific discovery worthy of a Nobel prize and the awarding of the medal has increased, with waits of more than 20 years becoming common. If this trend continues, some candidates might not live long enough to attend their Nobel ceremonies.

Before 1940, Nobels were awarded…

Subject terms: Careers · Events · Lab life

Associated links

Correspondence

Prizes: Growing time lag threatens Nobels

Editors' pick

Scientists sound alarm over long wait for Nobel prize

AFP By Peter Harmsen | AFP – Sun, Oct 5, 2014

Outlook

- Citation distributions: universality
- The reputation of an author may boost the impact of a paper in its early stage, afterwards impact is sustained by the paper's own value
- Consequences of exponential growth in the number of publications

  a) papers are forgotten faster and faster

  b) the value of a citation is decreasing
- Novelty can be computed as a recombination process on a network
- Exponentially growing delay between publication of Nobel discovery and reception of the prize: why?

Acknowledgements

Francesco Becattini Arnab Chatterjee Claudio Castellano Lucas Jeub Kimmo Kaski Marija Mitrović

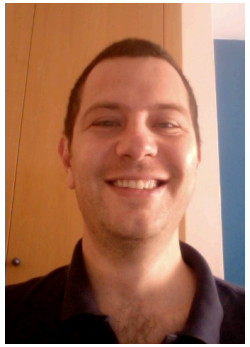
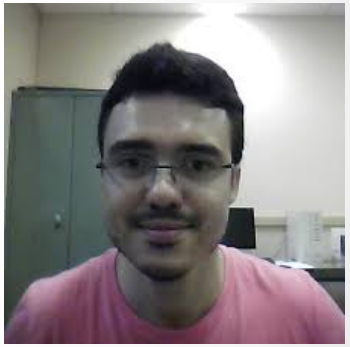





Fabio Pammolli Raj K. Pan Pietro Parolo Orion Penner Alex Petersen Filippo Radicchi Sandro Reia





Massimo Riccaboni Gene Stanley Dashun Wang Yang Wang