When the hunter becomes the hunted: the science of scientific careers

The Organisation, Economics and Policy of Scientific Research

Collegio Carlo Alberto, Via Real Collegio 30, Moncalieri, Torino.

May 2-3, 2013

Alexander M. Petersen, Massimo Riccaboni, Fabio Pammolli
IMT Institute for Advanced Studies, Lucca Italy

Institutions Markets Technologies
IMT
INSTITUTE FOR ADVANCED STUDIES LUCCA

Outline

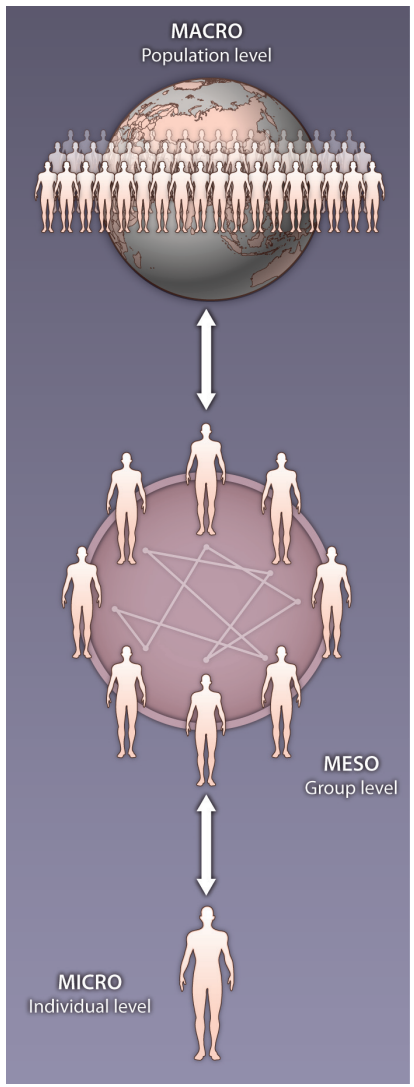
- **Scientific Careers:**
  - quantifying multiple dimensions of career growth
  - Science as a competitive arena: insights from the distribution of career longevity, career achievement, and empirical evidence for cumulative advantage
  - Emergence of "big" team science and measures for team (in)efficiency
  - Closing notes: behavioral / institutional trends in science
    - emergence of competitive strategies
    - cognizant enhancing drugs (CED)
    - is academia becoming more like a professional sport? "Gaming the system", such as strategic "h-index doping", google profile manipulation
- **Institutional trends in Science and their impact on careers**

Science: a multi-scale system with emergent complexity

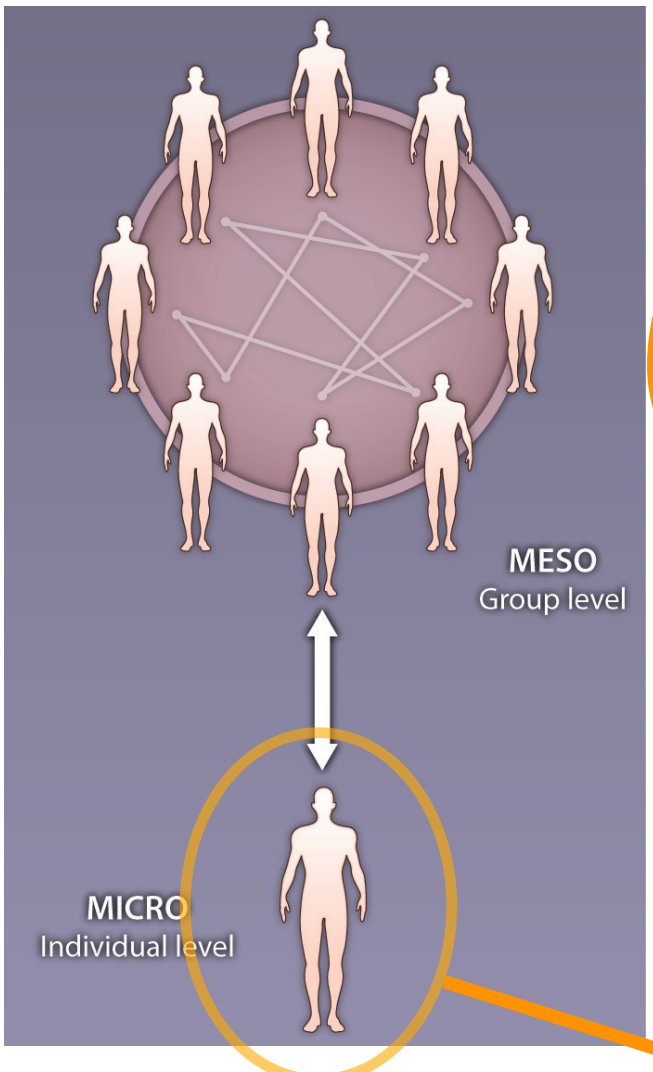
Practical Question: how to measure scientific output and impact at various scales while accounting for systemic heterogeneity

- Country
- Institution
- Lab / Team
- Individual
- Paper

K. Börner, et al. A multi-level systems perspective for the science of team science. Sci. Transl. Med. 2, 49cm24 (2010).



An "atomic" view of Science



K. Börner, et al. A multi-level systems perspective for the science of team science. Sci. Transl. Med. 2, 49cm24 (2010).

Interactions mediated by social "forces":

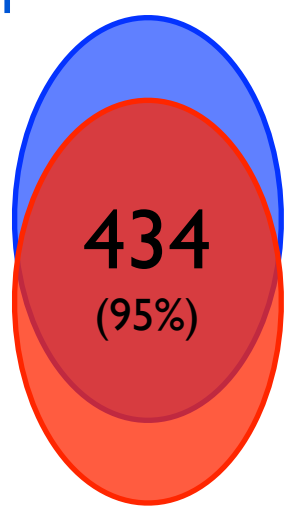
- Collaboration (attractive)
- Competition (repulsive)
- Knowledge (an "exchange particle")

Watson-Crick strategy:

451 publications

* **Michael Stuart Brown**
* **Joseph L. Goldstein**

Recipients of the 1985 Nobel Prize in Physiology or Medicine for describing the regulation of cholesterol metabolism.



458 publications

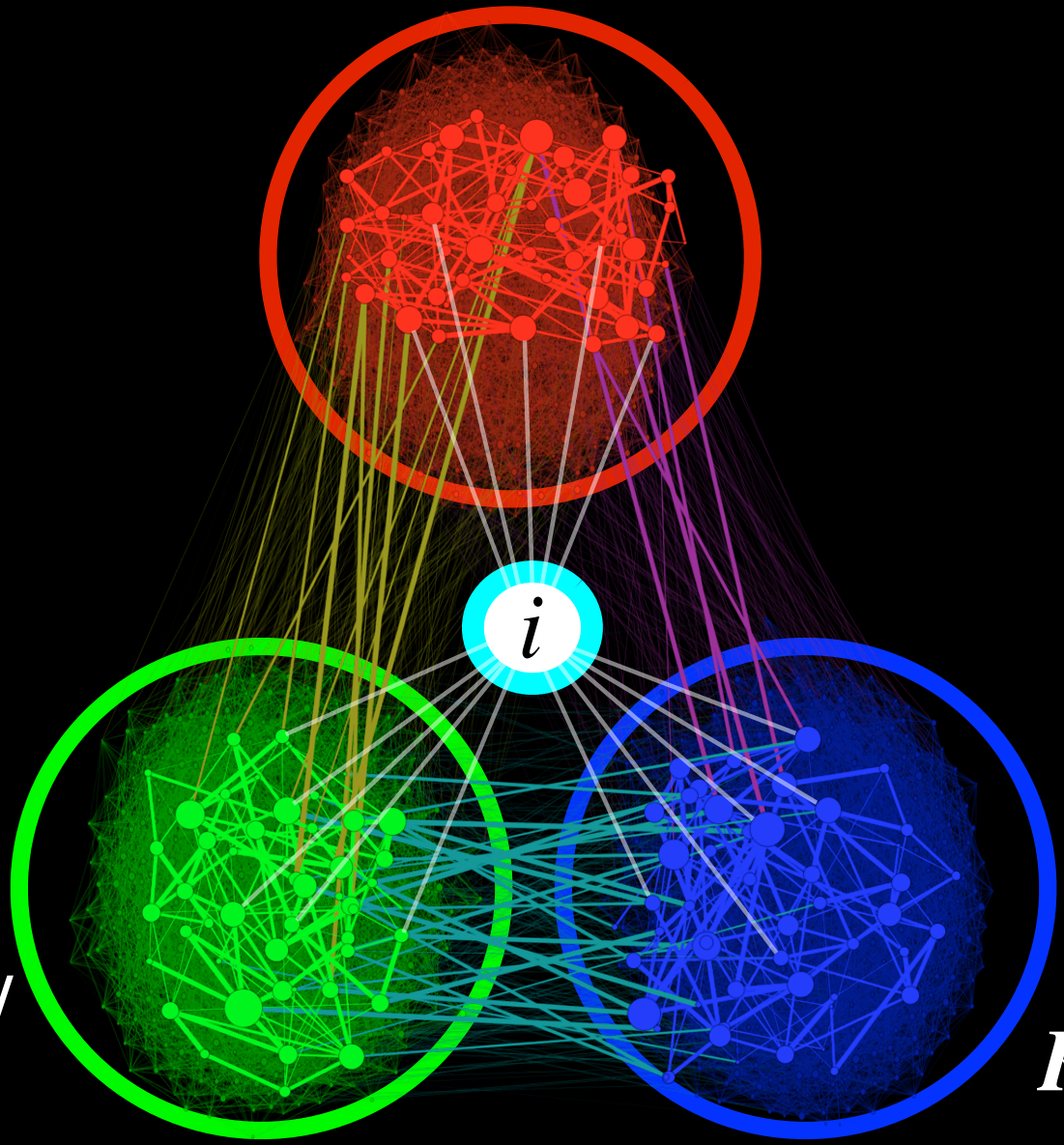
Solo-artist strategy:

* **Marilyn Kozak** (also cell biologist)

$N = 70, N_{solo} = 59$

Institutions Markets Technologies
IMT
INSTITUTE FOR ADVANCED STUDIES LUCCA

career growth as a co-evolving "multiplex"



a data-centric approach aimed at better understanding *"publish or perish"* career growth

Longitudinal career data for 450 top scientists:

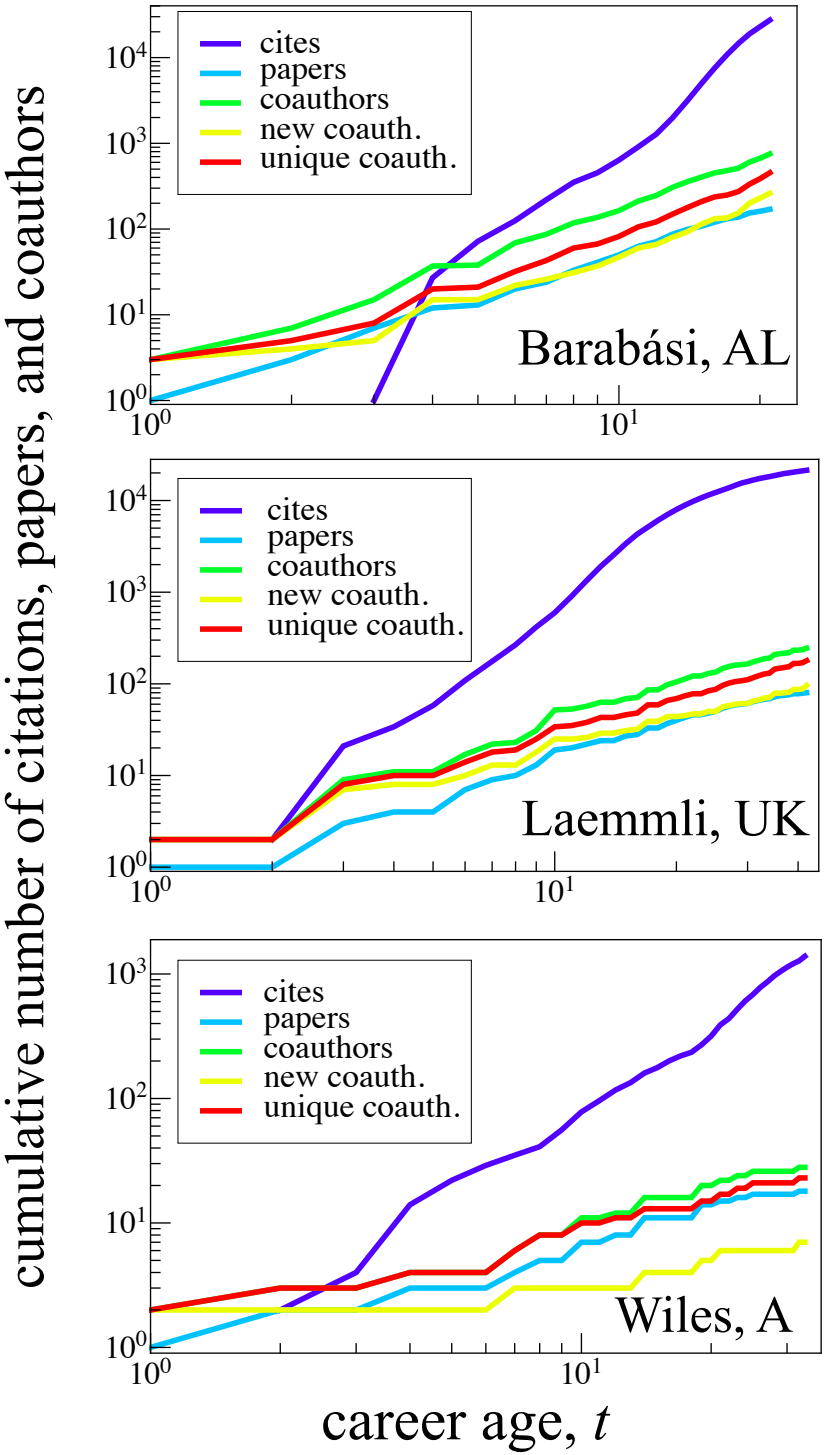
Set A: 100 most-cited physicists, average h-index $\langle h \rangle$ = 61 ± 21

Set B: 100 additional highly-prolific physicists, $\langle h \rangle$ = 44 ± 15

Set C: 100 current assistant professors from 50 US physics depts., $\langle h \rangle$ = 15 ± 7

Set D: 100 most-cited cell biologists, $\langle h \rangle$ = 98 ± 35

Set E: 50 highly-cited pure mathematicians, $\langle h \rangle$ = 20 ± 10



Using comprehensive ISI Web of Science publication data we track the following quantities for each scientific career i in year t:

publication measures

(a) the scientific production is measured by the number $n_i(t)$ of papers published by author i in year t,

$$N_i(t) \equiv \sum_{t'=1}^{t} n_i(t')$$

(b) the impact of paper p is measured by the cumulative number $c_{i,p}(t)$ of citations received up to year t.

$$C_i(t) \equiv \sum_{p=1}^{N_i(t)} c_{i,p}(t)$$

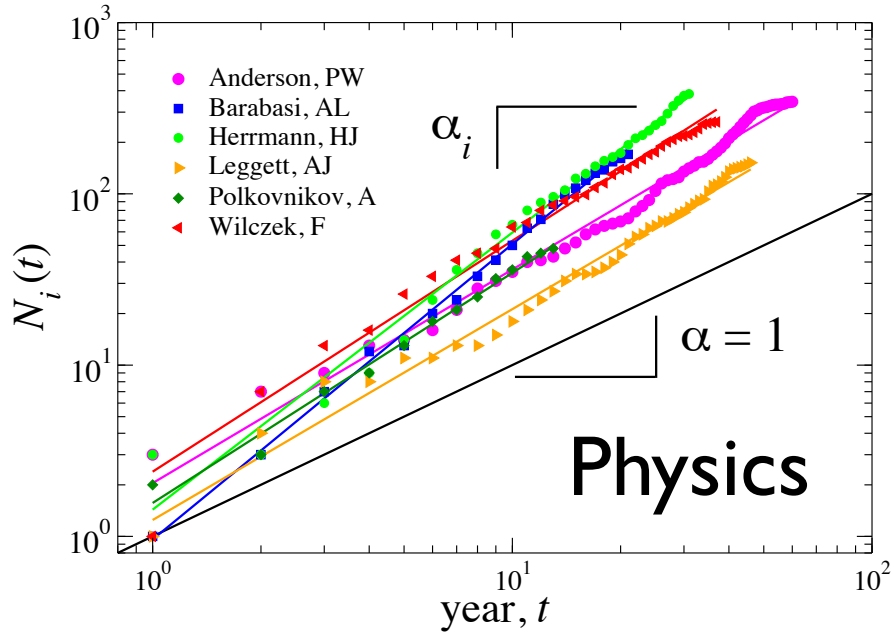
collaboration measures

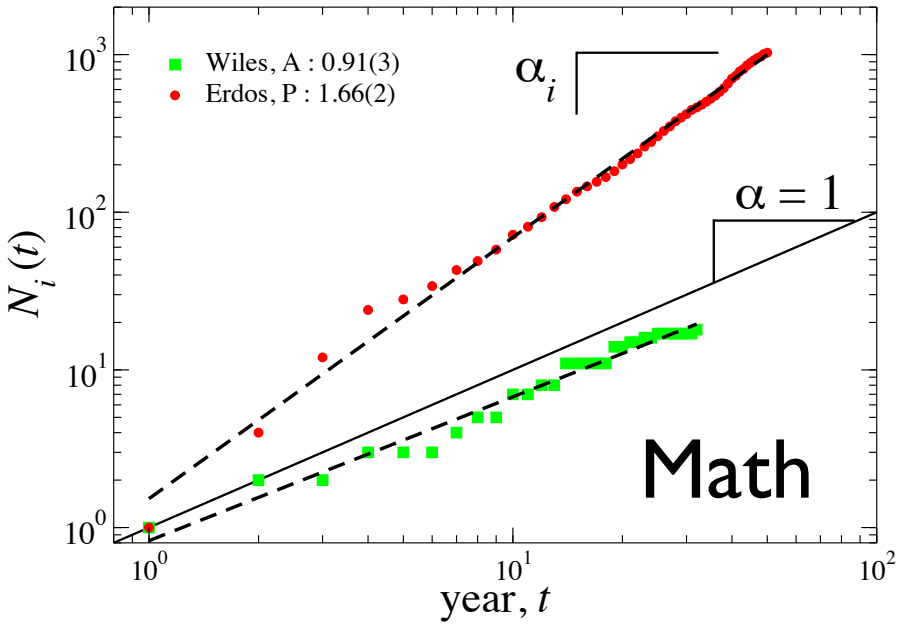
(c) total number of authors on all papers, $k_i^{\mathsf{T}}(t)$

(d) number of distinct coauthors, $k_i(t)$

(e) number of new distinct coauthors $k_i^{New}(t)$

The career trajectory in science: a tale of knowledge, collaboration, and reputation spillovers





Annual production of individual i

$n_i(t)$ number of publications in year t

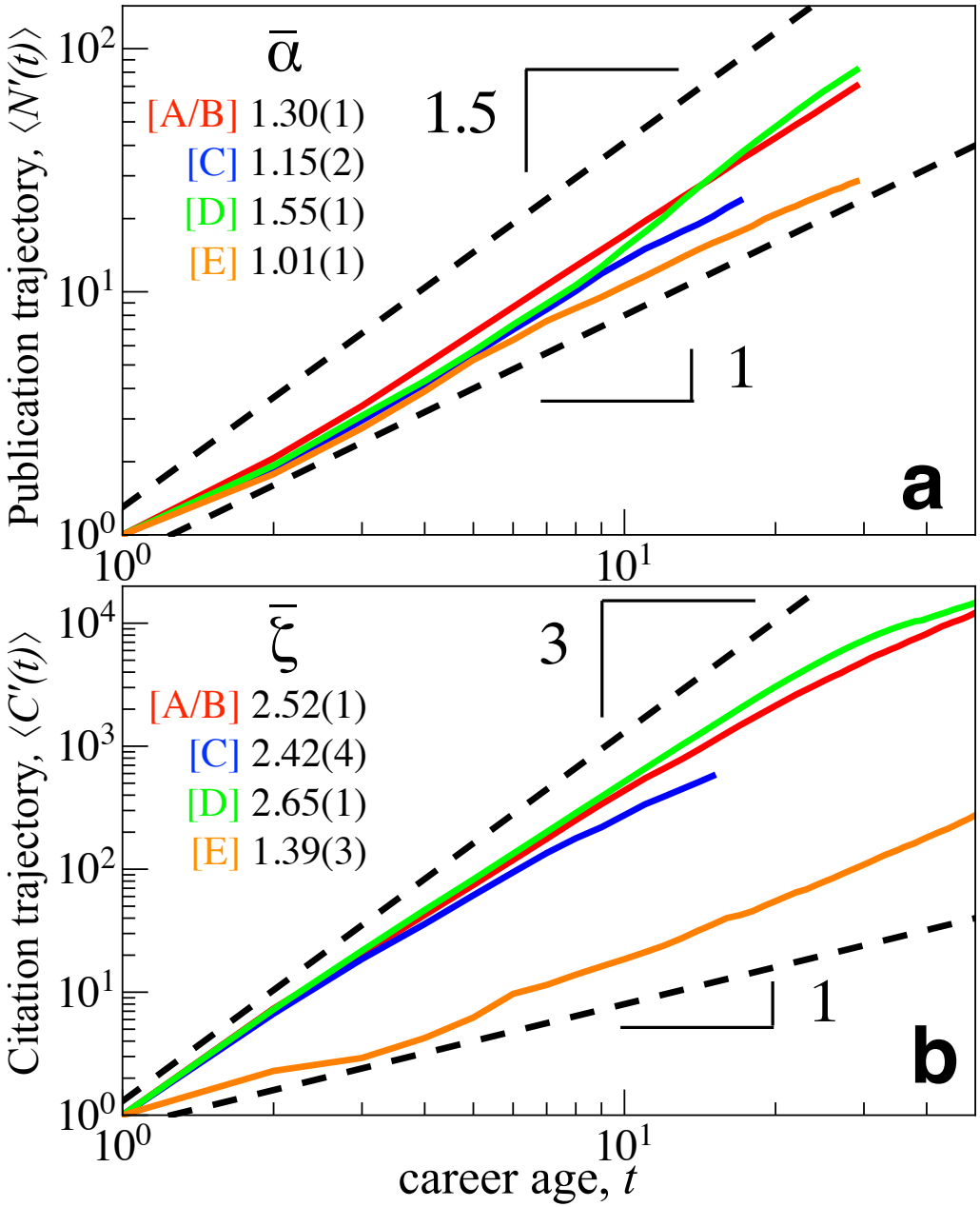
Cumulative production, a proxy for career reputation

$$N_i(t) \equiv \sum_{t'=1}^{t} n_i(t')$$

$$\approx A_i t^{\alpha_i}$$ ← for many prolific careers!

knowledge, reputation, and collaboration spillovers contribute to the increasing returns across the academic career

Are there characteristic career growth patterns?



(normalized) production trajectory $\langle N'(t) \rangle \sim t^{\bar{\alpha}}$

(normalized) cumulative citation trajectory $\langle C'(t) \rangle \sim t^{\bar{\zeta}}$

$\zeta > \alpha > 1 \Rightarrow$ *increasing returns*

Cumulative advantage ~
careers become "attractors" of new opportunities instead of "pursuers"

Reputation and Impact in Academic Careers, (submitted)
A. M. Petersen, S. Fortunato, R. K. Pan, K. Kaski, O. Penner, M. Riccaboni, H. E. Stanley, F. Pammolli

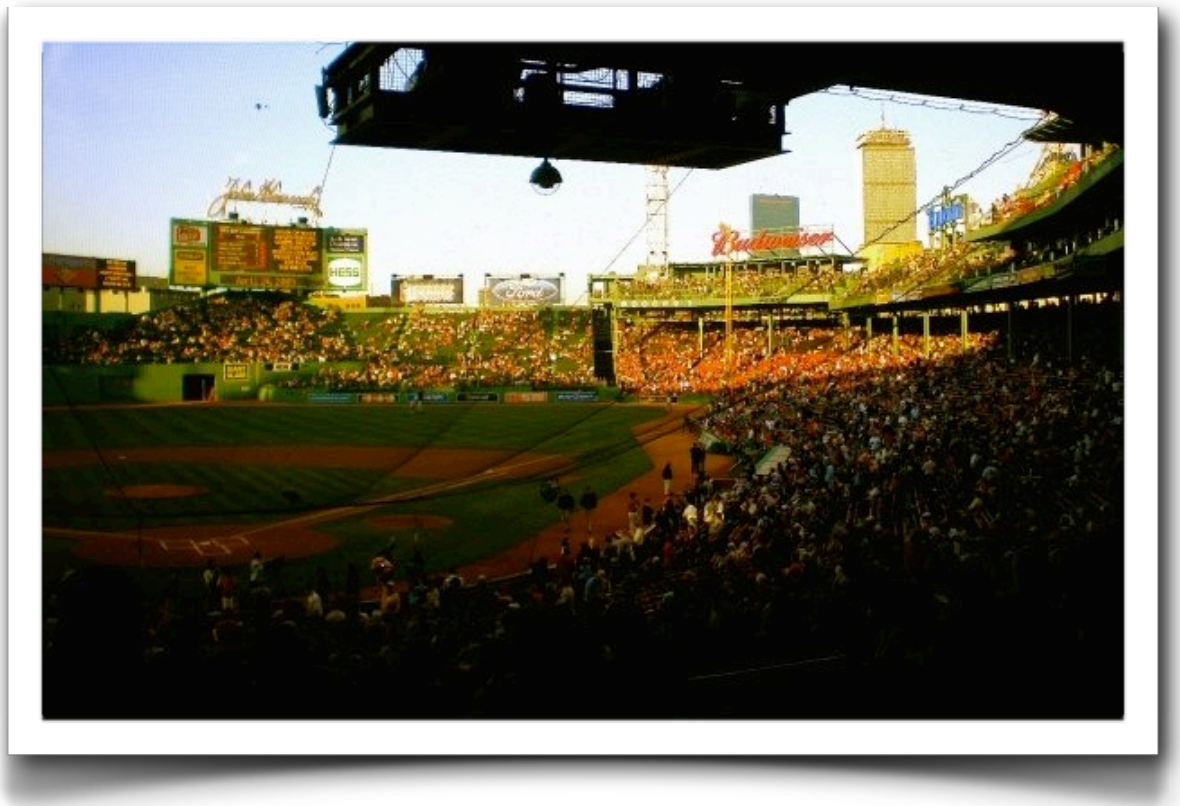
Institutions Markets Technologies
IMT
INSTITUTE FOR ADVANCED STUDIES LUCCA

Competitive arenas in science

Physical Review Letters
moving physics forward

||

Science
AAAS



nature

Cell

The NEW ENGLAND JOURNAL of MEDICINE

PNAS
Proceedings of the National Academy of Sciences of the United States of America

Competitive arenas in science

Physical Review Letters — moving physics forward

||

Science — AAAS

nature

Cell

Journal	Years	Articles	Authors, N^j
CELL	1974-2012	12,349	19,491 (1,753)
Nat./PNAS/Sci.	1958-2012	219,656	112,777 (14,478)
NEJM	1958-2012	18,347	33,149 (2,897)
PRL	1958-2012	98,739	55,827 (10,206)

TABLE I: Summary of journal datsets. N^j is the number of unique surnames we were able to identify in each journal j over the denoted period. The N^j value in parentheses denotes the number of careers with $L_i \geq 5$.

The NEW ENGLAND JOURNAL of MEDICINE

PNAS — Proceedings of the National Academy of Sciences of the United States of America

Disambiguation strategy:

use author profiles with last names that occur with only one first-middle initial

~~Nathan, A~~

~~Nathan, B~~

Nodulman, L → coauthor on 388 PRL articles!!!

(Fermilab scientist, with average # coauthors = 670)

....

~~Smith, A~~

~~Smith, B~~

~~Smith, AB~~

Smithduque, CE → coauthor on 2 Nat./PNAS/Sci.

....

Journal	Years	Articles	Authors, N^j
CELL	1974-2012	12,349	19,491 (1,753)
Nat./PNAS/Sci.	1958-2012	219,656	112,777 (14,478)
NEJM	1958-2012	18,347	33,149 (2,897)
PRL	1958-2012	98,739	55,827 (10,206)

TABLE I: Summary of journal datsets. N^j is the number of unique surnames we were able to identify in each journal j over the denoted period. The N^j value in parentheses denotes the number of careers with $L_i \geq 5$.

Disambiguation strategy:

use author profiles with last names that occur with only one first-middle initial

~~Nathan, A~~

~~Nathan, B~~

Nodulman, L → coauthor on 388 PRL articles!!!

(Fermilab scientist, with average # coauthors = 670)

....

~~Smith, A~~

~~Smith, B~~

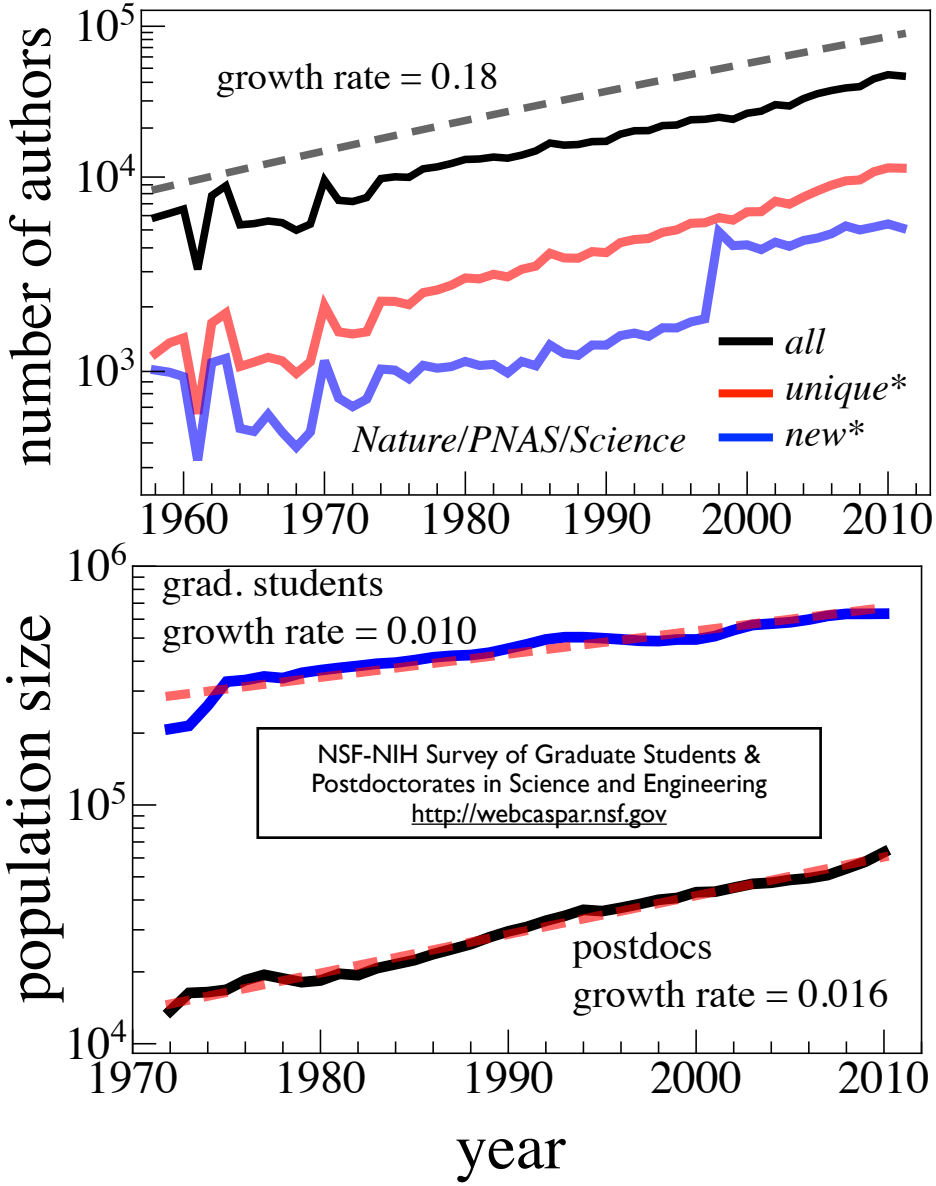
~~Smith, AB~~

Smithduque, CE → coauthor on 2 Nat./PNAS/Sci.

....

Journal	Years	Articles	Authors, N^j
CELL	1974-2012	12,349	19,491 (1,753)
Nat./PNAS/Sci.	1958-2012	219,656	112,777 (14,478)
NEJM	1958-2012	18,347	33,149 (2,897)
PRL	1958-2012	98,739	55,827 (10,206)

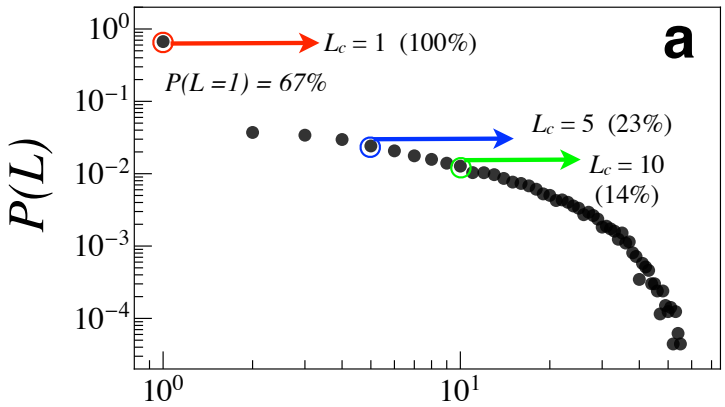
TABLE I: Summary of journal datsets. N^j is the number of unique surnames we were able to identify in each journal j over the denoted period. The N^j value in parentheses denotes the number of careers with $L_i \geq 5$.

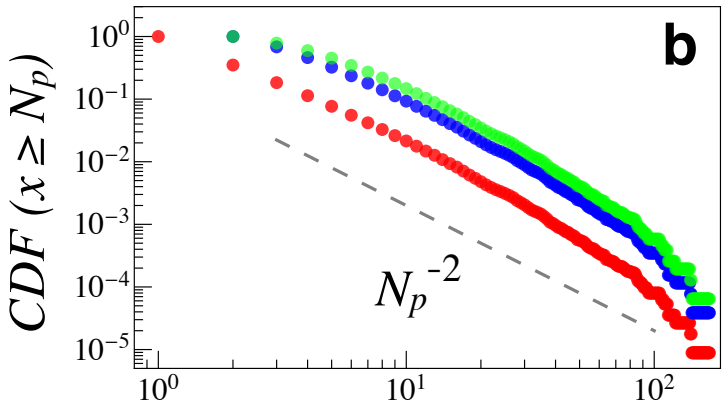


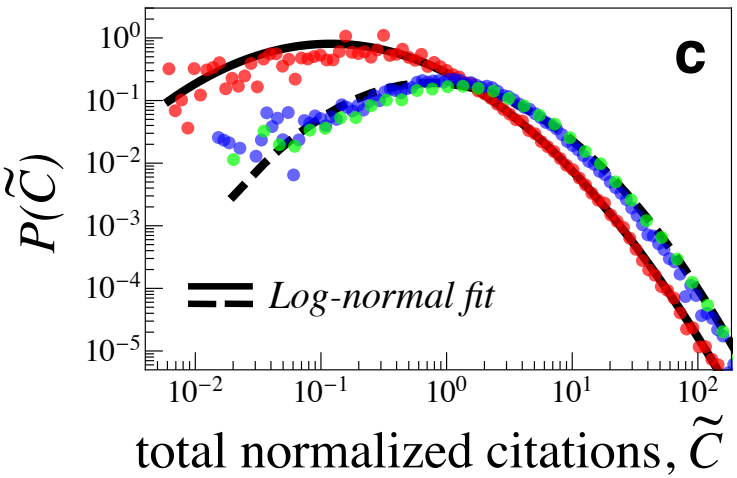
PhD bubble (scientific labor crisis):

decreasing number of long-term contracts in an already competitive and growing system

Nature/PNAS/Science







Peering inside the high-impact arena...

Longevity $L_i^j \equiv t_{i,f}^j - t_{i,0}^j + 1$ *in a given journal set* is extremely right-skewed, in agreement with the quantitative predictions of a rich-get-richer career progress model

> *Quantitative and empirical demonstration of the Matthew effect in a study of career longevity,*
> *A. M. Petersen, W.-S. Jung, J.-S. Yang, H. E. Stanley.*
> *Proc. Natl. Acad. Sci. USA 108, 18-23 (2011).*

Likewise, since production is highly correlated with longevity, the distribution of cumulative publications is also extremely right-skewed

However, the net impact of an author's scientific output is less correlated with an author's longevity and production.

"deflated / detrended" impact measure $\tilde{c} = c_p^j(t) / \langle c^j(t) \rangle$

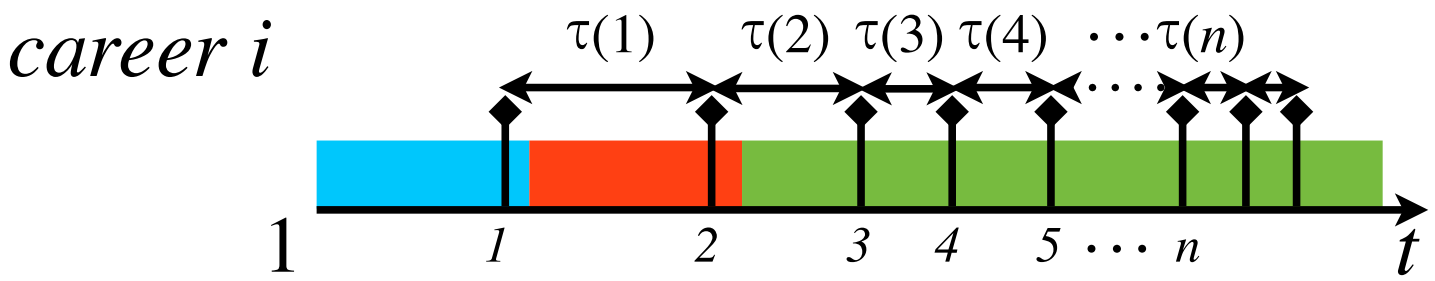
$\Rightarrow$

cumulative impact measure approximately controls for time and discipline

$$\tilde{C}_i = \sum_{p=1}^{N_p} \tilde{c}_{i,p}^j(T_i)$$

tracking the trajectory of "repeat winners"

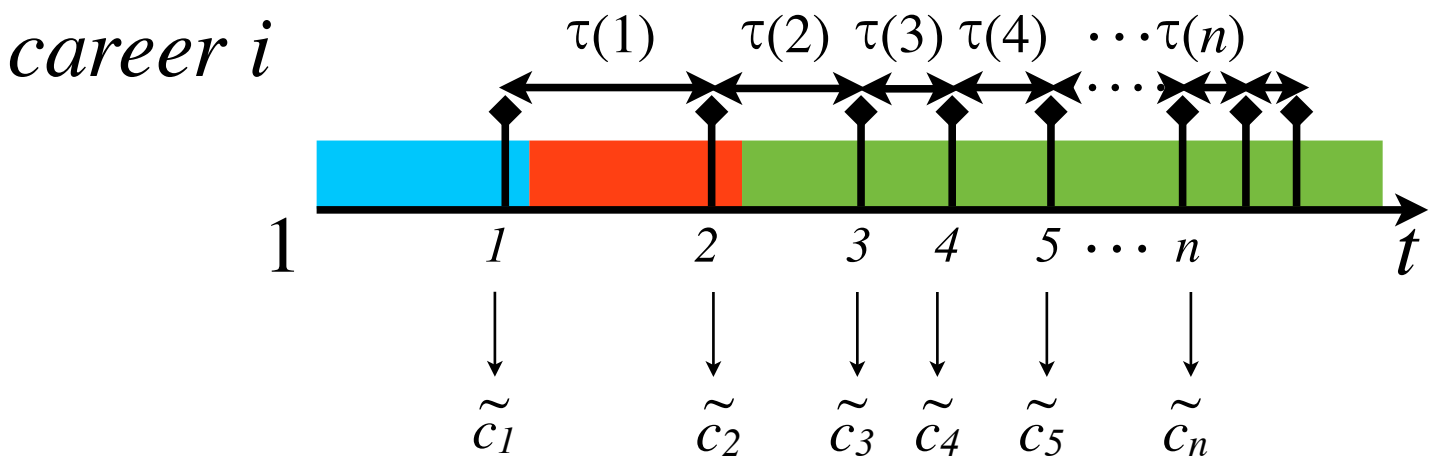
For each career *i* we track his/her longitudinal progress in a given journal



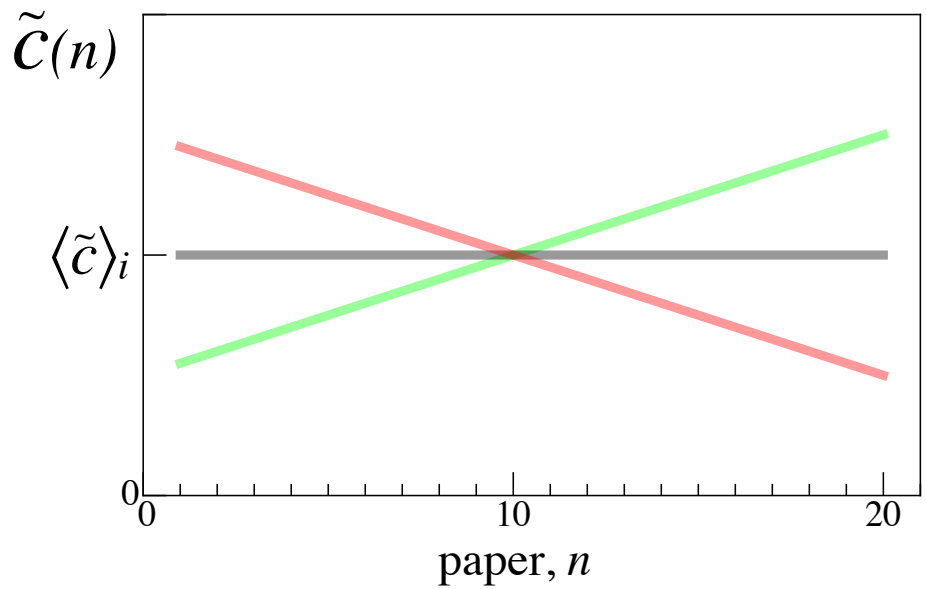
(Q1) Is there a trend in the waiting time $\tau_i(n)$ between an author's n-th paper and (n+1)-th paper?

tracking the trajectory of "repeat winners"

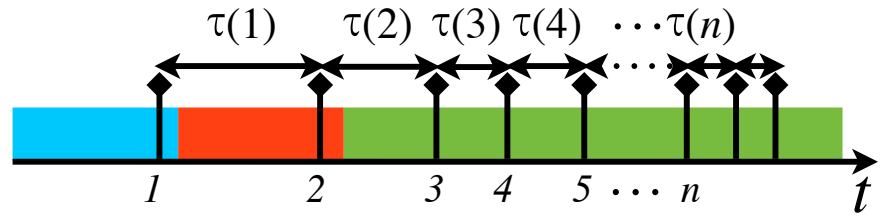
For each career *i* we track his/her longitudinal progress in a given journal

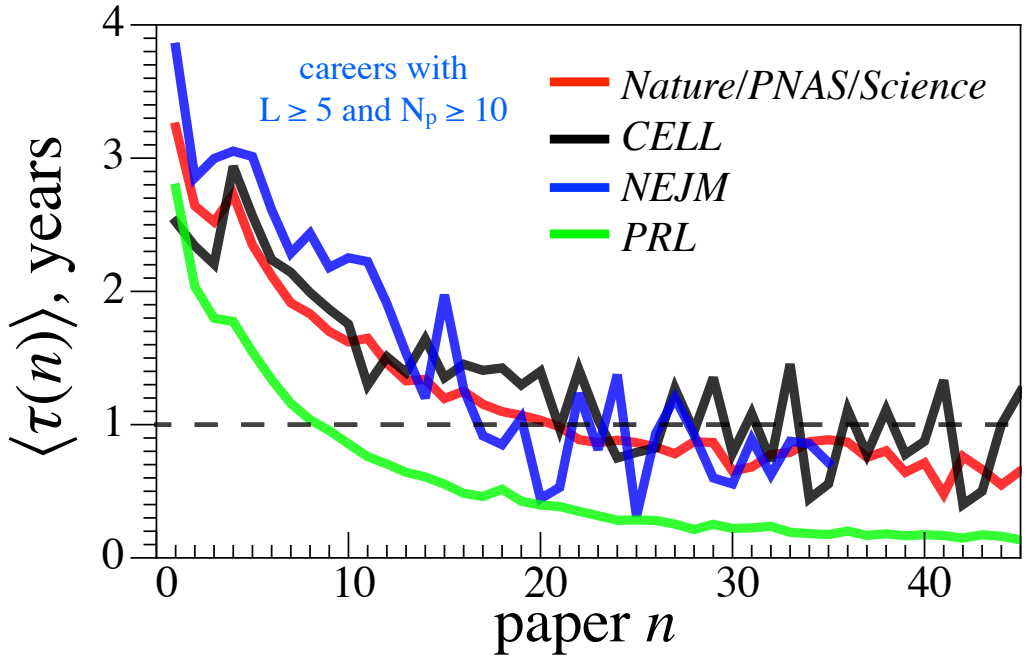


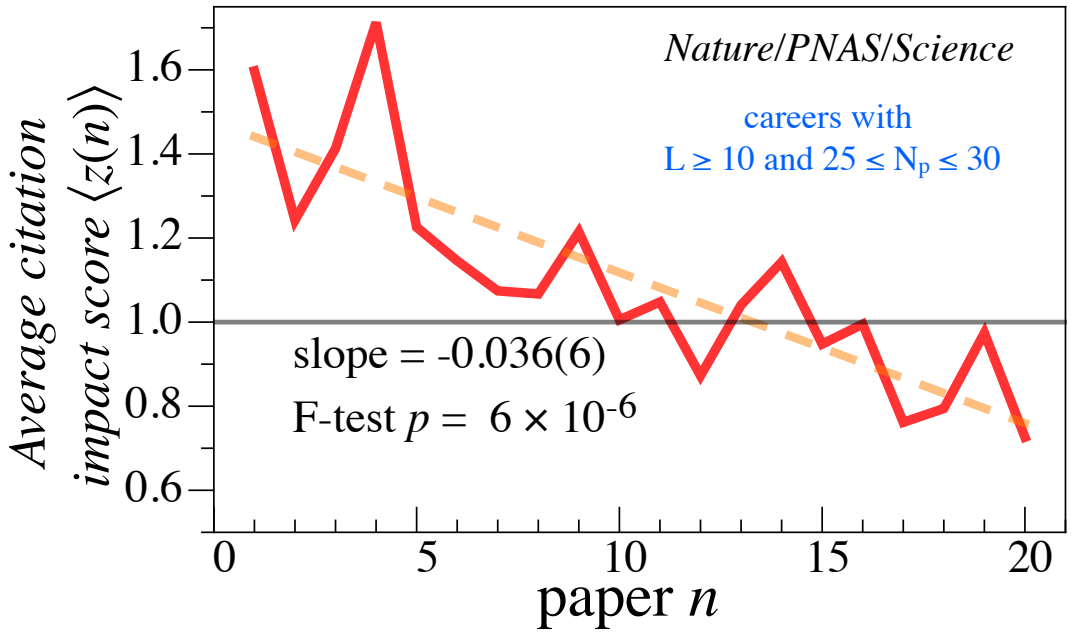
deflated impact measure $\tilde{c} = c_p^j(t)/\langle c^j(t)\rangle$



(Q2) How does the relative impact change over time?







"Cumulative advantage" in high-impact journals

(1) What is the expected waiting time $\tau_i(n)$ between an author's n-th paper and (n+1)-th paper?

"rich clubs" in science

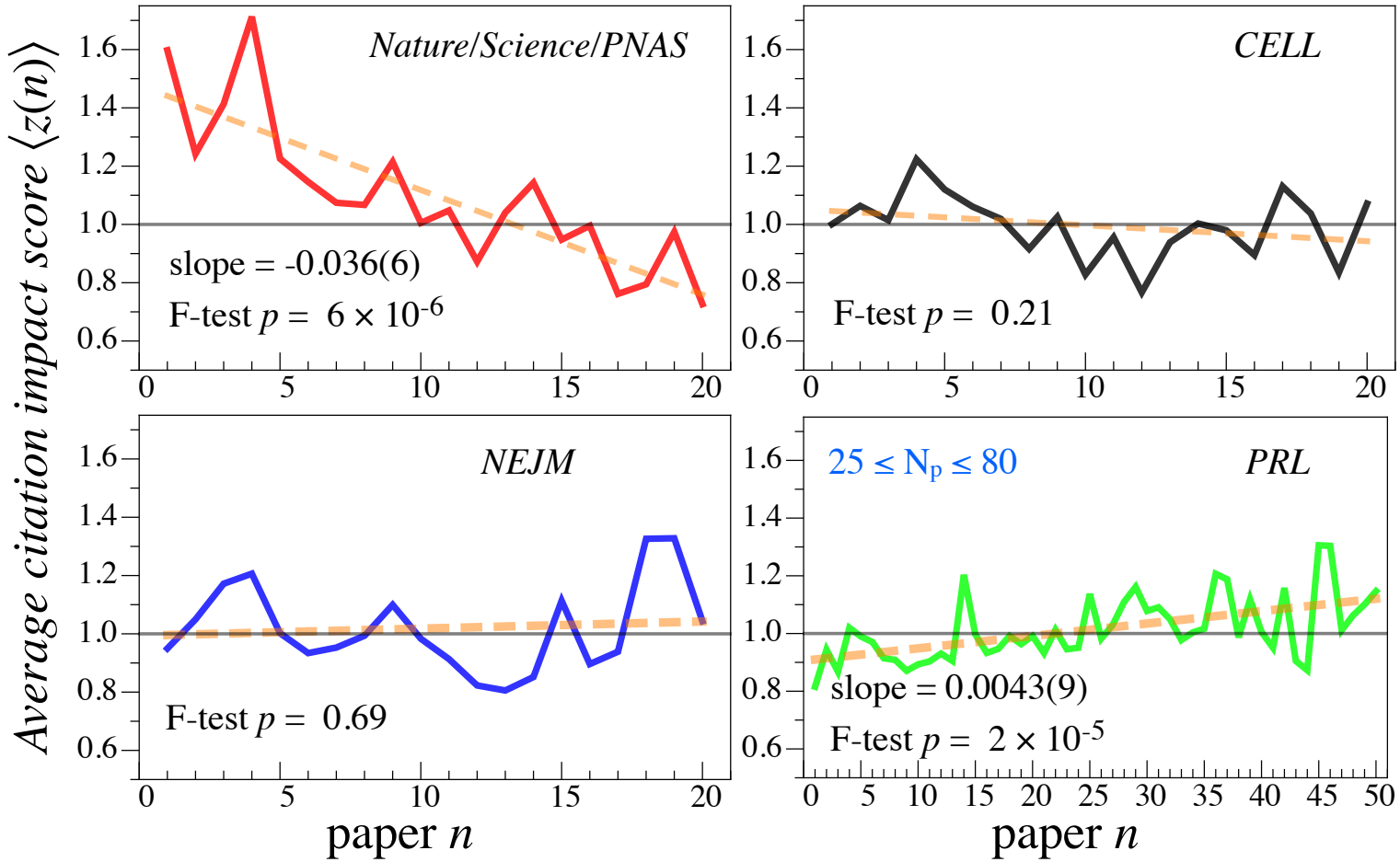
(2) What is the relative impact of an author's n-th paper as compared to their average paper?

$$z_i(n) \equiv \tilde{c}_i(n) / \langle \tilde{c}_i \rangle$$

$$\langle \tilde{c}_i \rangle \equiv N_p^{-1} \sum_{p=1}^{N_p} \tilde{c}_{i,p}$$

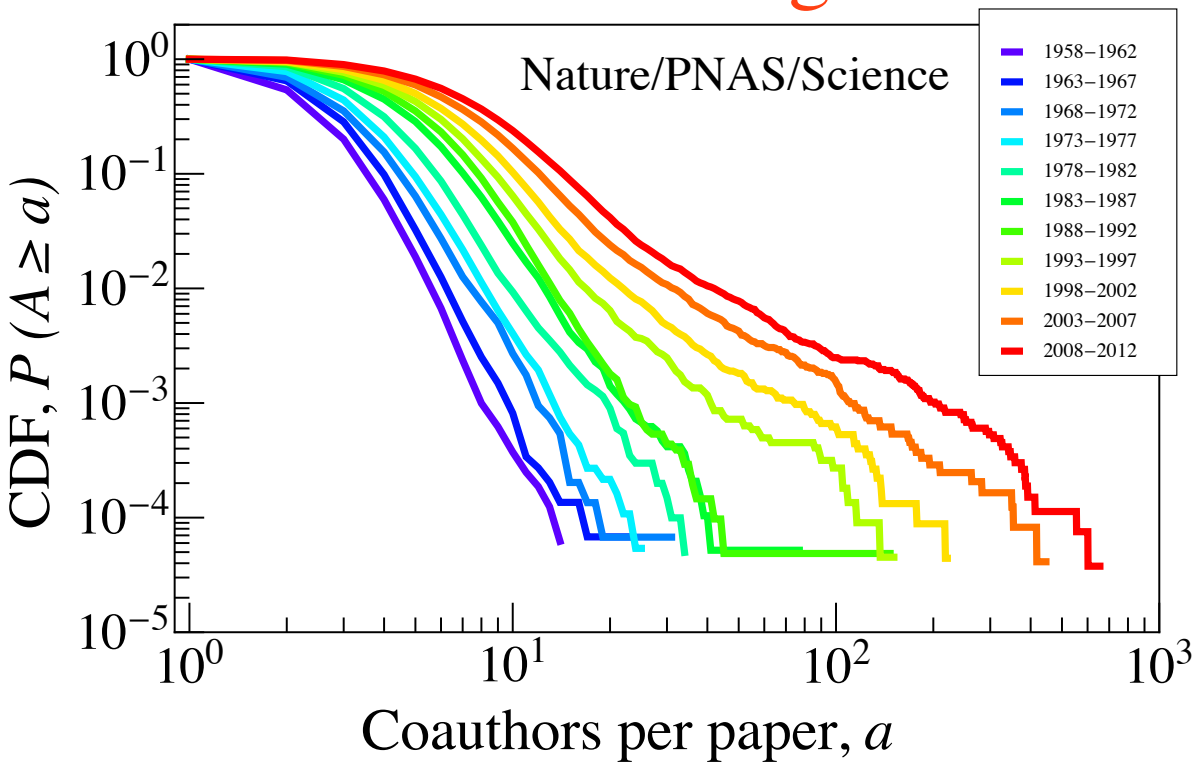
Journal specific variation

calculated using careers with $L \geq 10$ and $25 \leq N_p \leq 30$

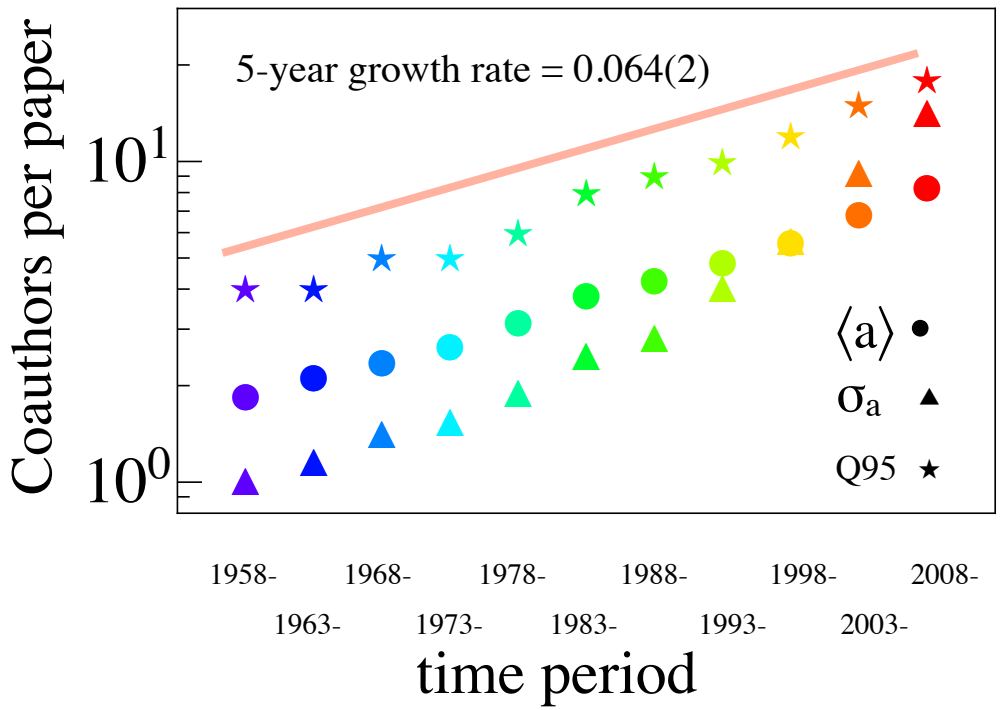


suggests that publication success is related to the role of "arm's length ties" in the editorial process?

Emergence of "big science"



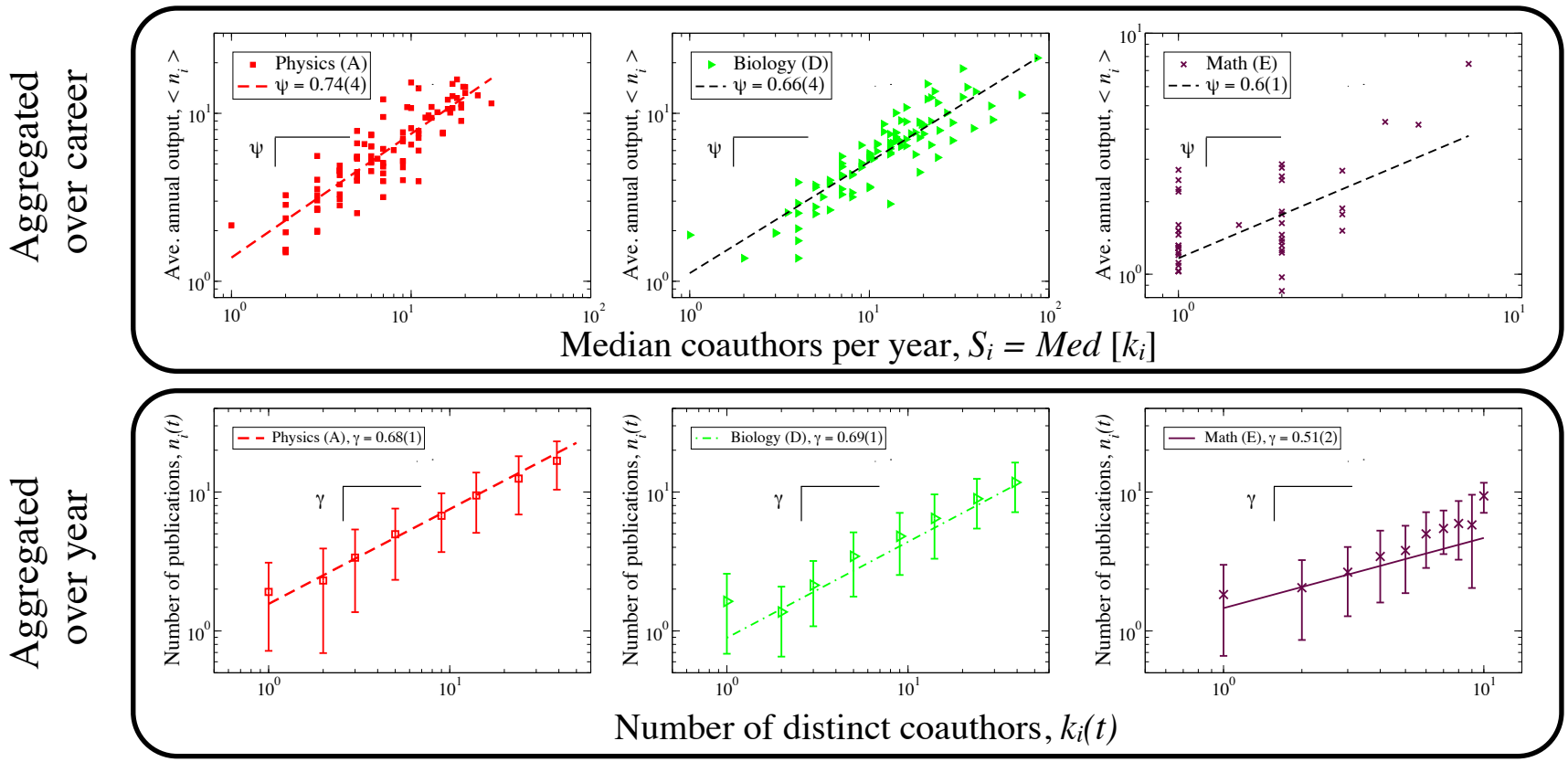
Q: how to "fairly" distribute credit in a system dominated by teams?



0.013 annual growth rate of the mean collaboration size ⟨a⟩ is consistent with the growth in the grad/ postdoc populations

Team (in)efficiency

Q: how does annual productivity depend on the number of “labor inputs” ?



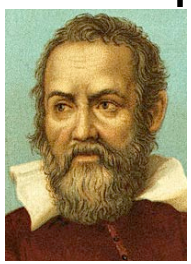
We measure the input-output relation using two aggregation methods, which both yield sub-linear scaling relations with efficiency parameters $\psi \approx \gamma$ and $\psi, \gamma < 1$

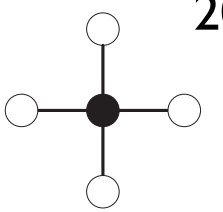
Interestingly, for scientists not in the top cohort we observe smaller ψ and γ values, indicating that team management skills are an important factor related to success

$$\gamma_A = 0.68(1) > \gamma_B = 0.52(1), \gamma_C = 0.51(2)$$

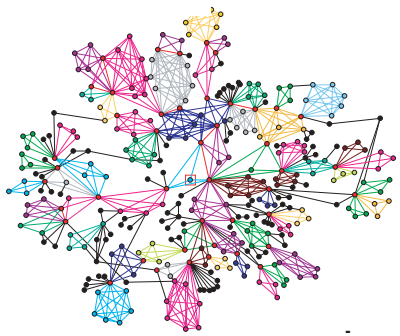
Institutional trends in Science

- emergence of small-world collaboration networks with the increasing role of team-work in science





200+ years



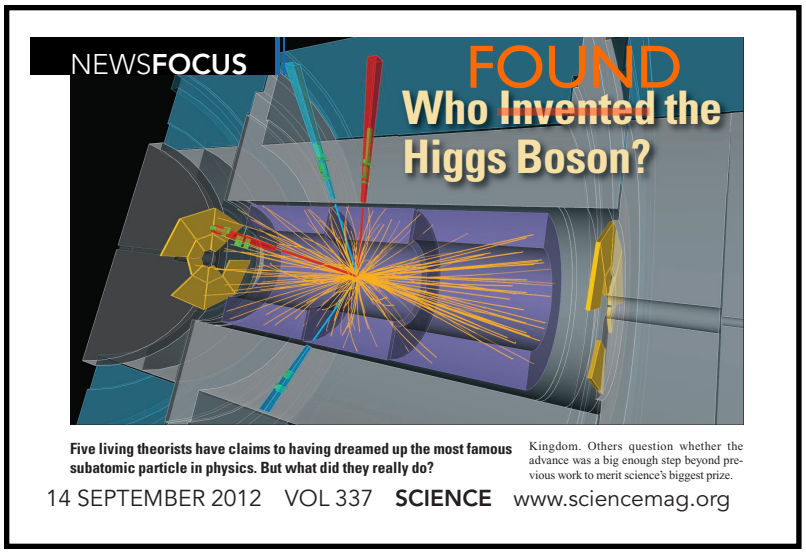
Paul A. David. *The Historical Origins of 'Open Science': An essay on patronage, reputation, and common agency contracting in the scientific revolution.* Capitalism and Society 3(2): Article 5 (2008).

G. Palla, A.-L. Barabasi, T. Vicsek. Quantifying social group evolution. Nature 446, 664-667 (2007)

S. Wuchty, B. F. Jones, B. Uzzi. The increasing dominance of teams in production of knowledge. Science 316, 1036-9 (2007)

- organizational shifts in the business structure of research universities
- shifts away from tenure towards shorter-term contracts + bottle neck in the number of tenure-track positions available
- redefining the role of teaching -vs- research faculty
- shifts in the competitive aspects of science, universities, and scientists: reputation tournaments in omnipresent competition arenas

Increasing team size & changing incentive system



"50-way tie for the Nobel Prize"

www.sciencemag.org **SCIENCE** VOL 336 6 APRIL 2012

Published by AAAS

CITATION IMPACT 9 DECEMBER 2011 VOL 334 **SCIENCE**

Saudi Universities Offer Cash In Exchange for Academic Prestige

Two Saudi institutions are aggressively acquiring the affiliations of overseas scientists with an eye to gaining visibility in research journals

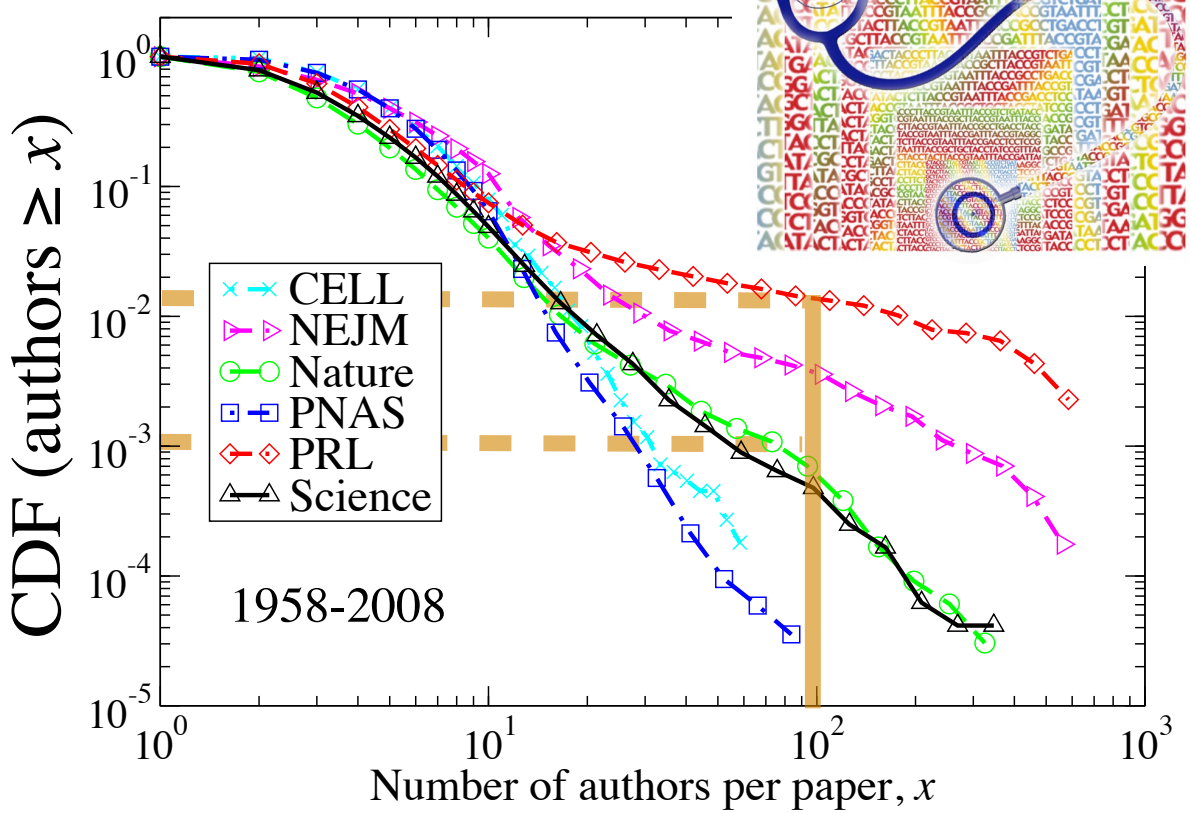
SCIENCE POLICY 5 AUGUST 2011 VOL 333 **SCIENCE**

Changing Incentives to Publish

Chiara Franzoni,[1] Giuseppe Scellato,[2,3] Paula Stephan[4,5,6]*

genome.gov
National Human Genome Research Institute
National Institutes of Health

ENCODE



Ethics in the appraisal of Scientific Careers

- **Competition ("fairness"):**
  - strategizing / extreme behavior, e.g. scientific fraud
  - CED (cognitive enhancing drugs)
  - free-riding + "tragedy of the commons"
- **Funding:**
  - financial incentives & who should subsidize early career risk
  - how to attribute / appraise / reward achievement, especially in the case of extremely large team projects
- **Careers:** predicting future career achievement using incomplete information and poorly understood/ designed achievement measures

Institutions Markets Technologies
IMT
INSTITUTE FOR ADVANCED STUDIES LUCCA

General take-home messages

- Science as an evolving institution: An institutional setting that neglects specific features of academic career trajectories (increasing returns from knowledge spillovers and cumulative advantage, collaboration factors, career uncertainty) is likely to be inefficient and unfair. But what is "fair"?

- Complex career dynamics: Knowledge, reputation, and collaboration spillovers are major factors leading to increasing returns along the scientific career trajectory. A data-centric ("big data") understanding of the production function of individual scientists can improve academic policies aimed at increasing career sustainability and decreasing career risk.

- Competition and Reward: There are many analogies between the superstars in science and the superstars in professional sports, possibly arising from the generic aspects of competition. Currently, the contract length, compensation, and appraisal timescale in these two professions are VERY different. However, is science becoming more like professional sports?

I) "Quantitative and empirical demonstration of the Matthew effect in a study of career longevity,"
A. M. Petersen, W.-S. Jung, J.-S. Yang, H. E. Stanley. Proc. Natl. Acad. Sci. USA 108, 18-23 (2011).

ii) "Statistical regularities in the rank-citation profile of scientists,"
A. M. Petersen, H. E. Stanley, S. Succi. Scientific Reports 1, 181 (2011).

iII) "Persistence and Uncertainty in the Academic Career,"
A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli. Proc. Natl. Acad. Sci. USA 109, 5213-5218 (2012).

Iv) "The case for caution in predicting scientists' future impact"
O. Penner, R. K. Pan, A. M. Petersen, S. Fortunato. Physics Today 66, 8-9 (2013).

v) "Reputation and impact in academic careers" (submitted)
A. M. Petersen, S. Fortunato, R. K. Pan, K. Kaski, O. Penner, M. Riccaboni, H. E. Stanley, F. Pammolli,

vi) "The hunter becomes the hunted: the science of scientific careers" (in preparation).
A. M. Petersen, M. Riccaboni, F. Pammolli. (2013)

Thank You!

A special thanks to my collaborators: Santo Fortunato, Woo-Sung Jung, Fabio Pammolli, Raj Pan, Orion Penner, Massimo Riccaboni, Gene Stanley, Sauro Succi, Fengzhong Wang, and Jae-Sook Yang

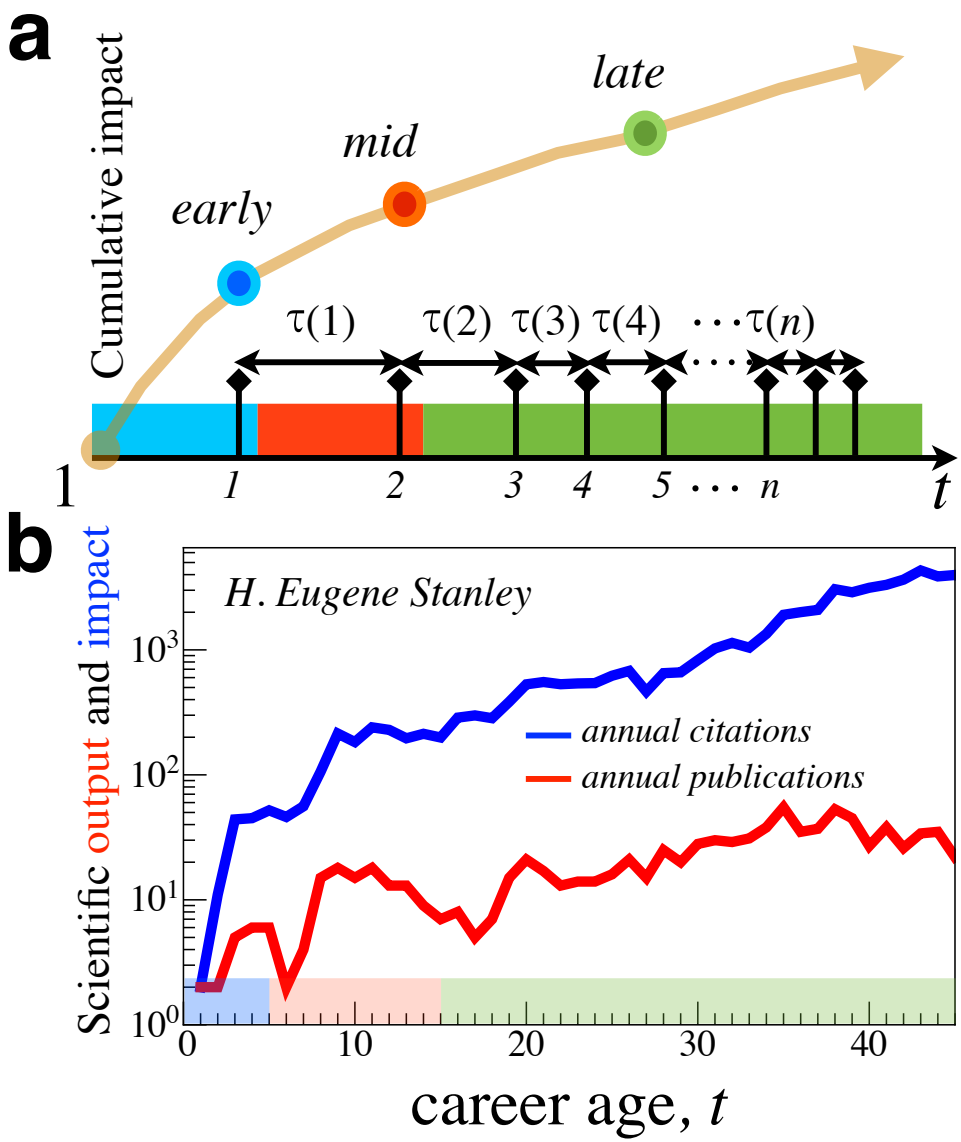
http://physics.bu.edu/~amp17/

When the hunter becomes the hunted: The science of scientific careers

Abstract:

Globalization of the scientific enterprise, the emergence of quantitative publication and impact measures, and shifts in the economics of science have altered the academic career ladder, making scientific careers a topic of increasing interest. Here we analyze two large datasets comprising (i) 450 leading scientists from biology, mathematics, and physics, and (ii) comprehensive publication data for 6 high-impact journals over the 55-year period 1958-2012. We show that top scientists are characterized by increasing returns to scale in their cumulative publication growth, reflecting the amplifying role of underlying social processes. However, for all three disciplines analyzed and for collaboration sizes ranging from 1 up to 100 coauthors per year, we observe a diminishing returns in annual publication rates when controlling for collaboration size, a feature that reflects team management, coordination, and training inefficiencies. These factors will be important considerations in the era of ``big science.'' Using the dynamics of consecutive publications in top journals by distinct authors, we show evidence for cumulative advantage mechanisms, which surprisingly, leads to a negative impact bias in the multidisciplinary journal dataset for Nature/PNAS/Science. This bias has the intriguing implication that the ``rich-get-richer'' effect allows prolific publishers to continue to publish at a discount as their career advances.

Quantitative career appraisal

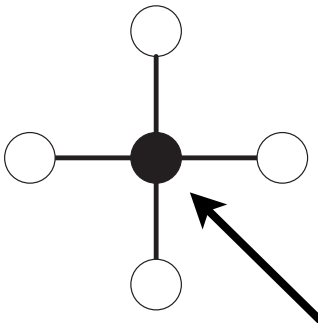


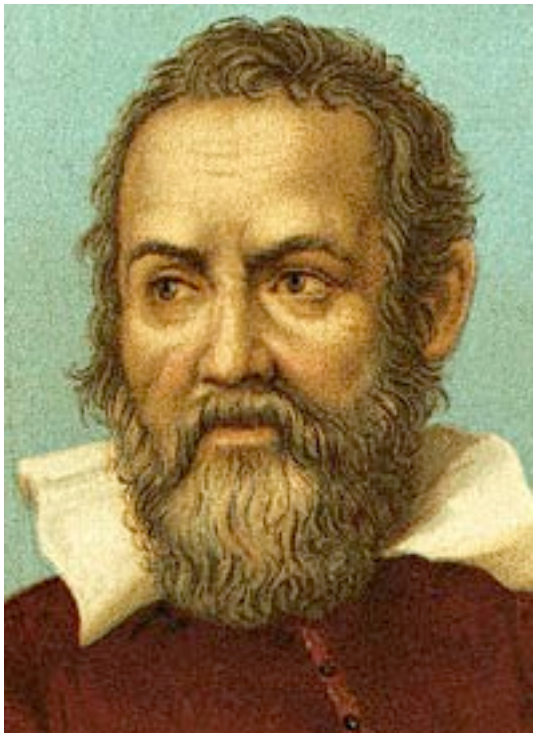
- how to measure career achievements?
- cumulative measures discount short-term uncertainty
- quantitative career evaluation using publication measures typically employs cumulative measures, e.g. the h-index
- What is the appropriate "appraisal time-scale" for academic careers?
  - too-long: reinforces rich-get-richer mechanisms
  - too-short: can induce instability and uncertainty in career growth in publish-or-perish systems
- Measures for "career predictability" must use non-cumulative indicators in order to eliminate spurious correlations

The case for caution in predicting scientists' future impact, Physics Today 66, 2013; Vetting career predictability models, submitted. O. Penner, A. M. Petersen, R. K. Pan, S. Fortunato.

Evolution of Science: "In the beginning..."

Social networks in science:
serve as the backbone for reputation signaling used to overcome the asymmetric information problem
⇒ emerging online reputation tournaments





Galileo Galilei

Noble patron (king, wealthy aristocrat, Pope)

Paul A. David. *The Historical Origins of 'Open Science': An essay on patronage, reputation, and common agency contracting in the scientific revolution.* Capitalism and Society 3(2): Article 5 (2008).