Quantifying the role of teamwork and reputation across scientific careers

Lorentz Center
International Center for workshops in the Sciences

Simulating the Social Processes of Science
from 7 Apr 2014 through 11 Apr 2014

IMT INSTITUTE FOR ADVANCED STUDIES LUCCA

Alexander M. Petersen
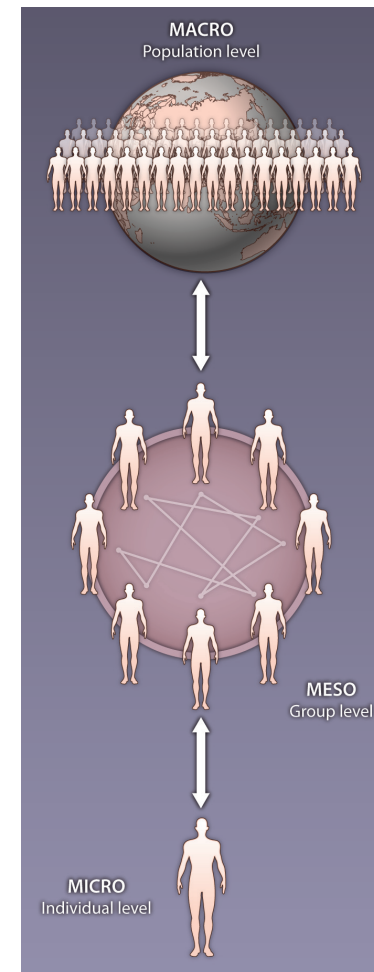
IMT Institute for Advanced Studies, Lucca Italy

Science of Science

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

- Science
- 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).

Models of science:
## don't throw the baby out with the bathwater!

Macro (institutions)

- Exponential growth of Science
- Economics of research universities and national funding programs
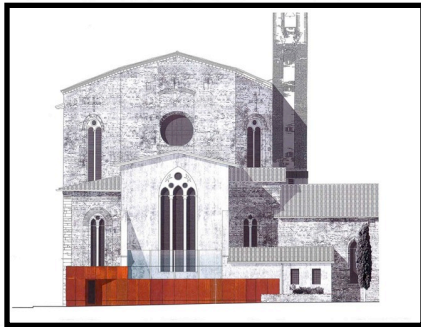- Increasing role of teams (division of labor) in science

Micro (individual careers)

- Growth of careers
- Collaboration patterns within careers
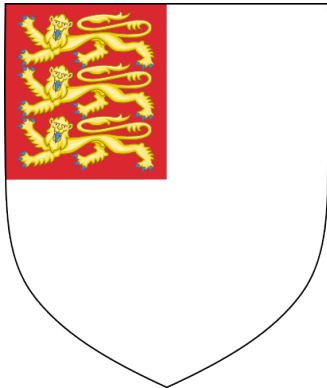- Competition
- Issues of ethics (rules of the game)

Limited complexity in small knowledge networks
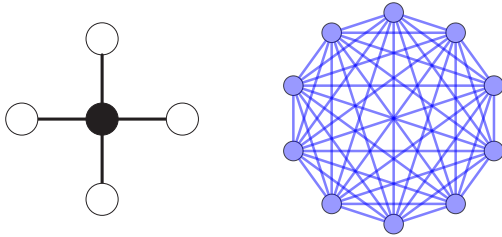
Early scholarly societies, e.g. national societies, scholastic monasteries, noble courts

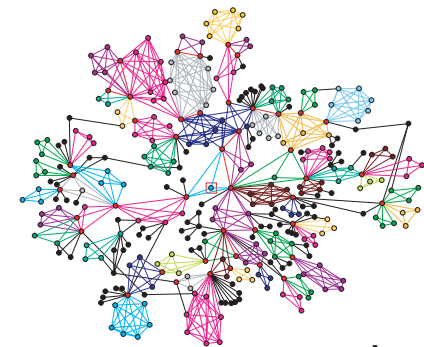**The Royal Society of London for Improving Natural Knowledge, Established 1660**

Paradigm shifts arising from Growth and increasing organizational complexity

⟹

Emergent complexity in large knowledge networks

a Co-authorship

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)

Urban property
210 acres (85 ha) (Main campus)
21 acres (8.5 ha) (Medical campus)
360 acres (150 ha) (Allston campus)
4,500 acres (1,800 ha) (other holdings)

Academic staff
2,100

Admin. staff
2,500 non-medical
11,000 medical

Endowment
US$30 billion (2012) (Large-cap company,
e.g. same market capitalization as Enel and Mitsubishi)
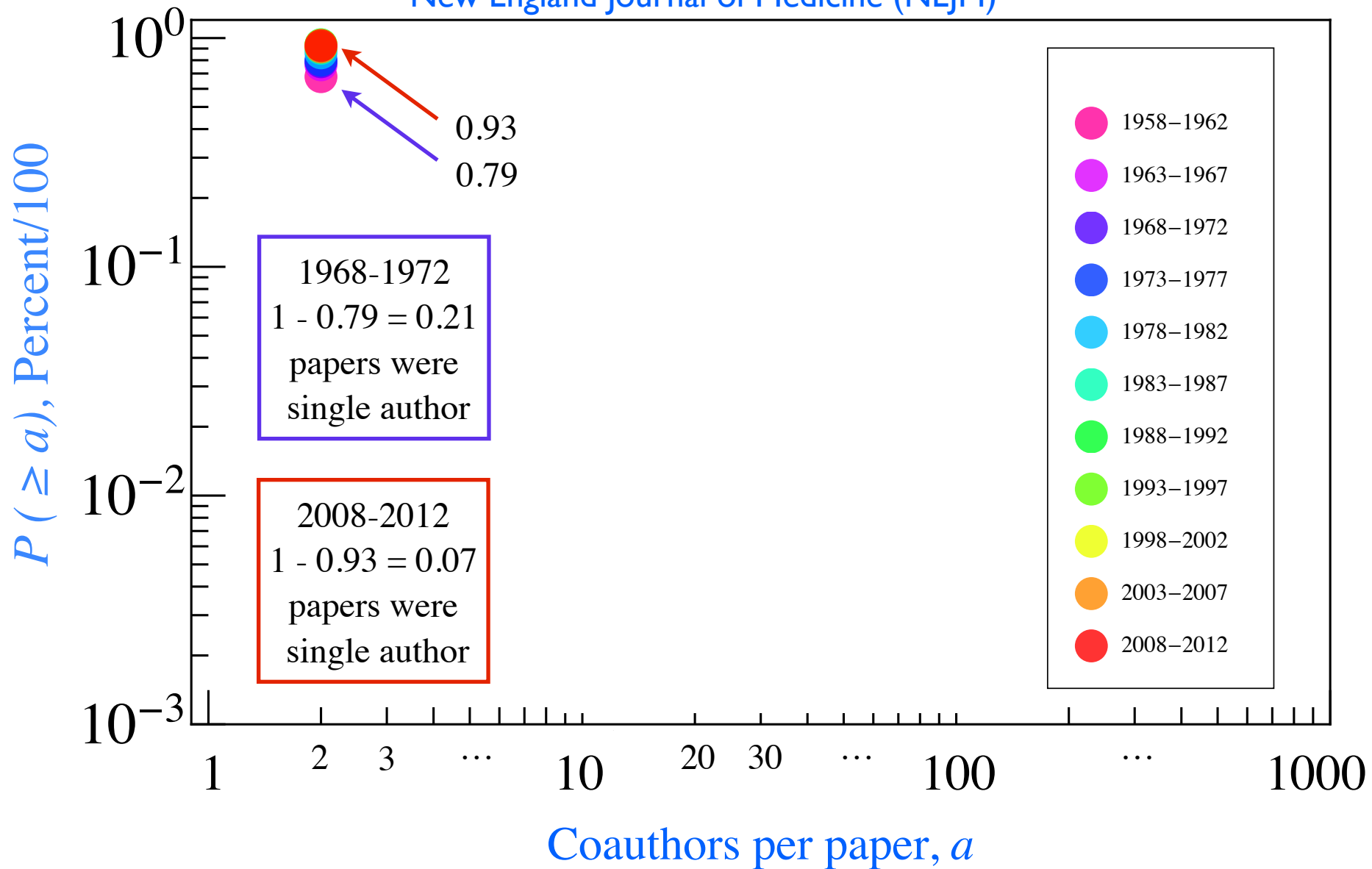
Overcoming grand challenges via teamwork

"Cooperation has created a conundrum for generations of evolutionary scientists. If natural selection among individuals favors the survival of the fittest, why would one individual help another at a cost to itself? .... Cooperation leads to integration, and integration to the complexity we see in modern life... So pervasive is cooperation that Martin Novak of Harvard University ranks it as the third pillar of evolution, alongside of mutation and natural selection."

On the origin of cooperation (2009) E. Pennisi. *Science 325*

$P (\geq a)$, the fraction of all papers with team size of at least size a

New England Journal of Medicine (NEJM)

0.93

0.79

1968-1972
1 - 0.79 = 0.21
papers were
single author

2008-2012
1 - 0.93 = 0.07
papers were
single author

1958–1962
1963–1967
1968–1972
1973–1977
1978–1982
1983–1987
1988–1992
1993–1997
1998–2002
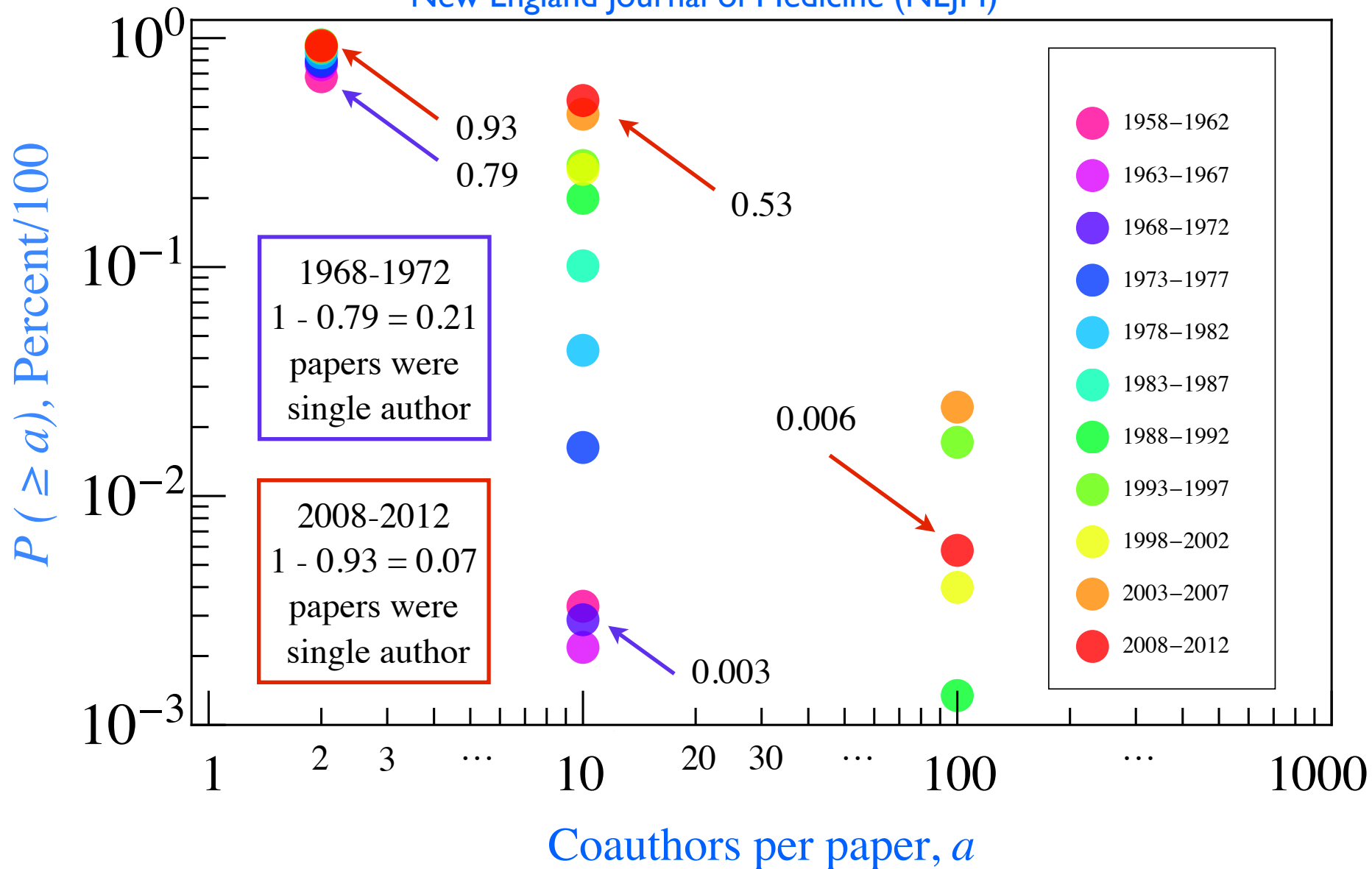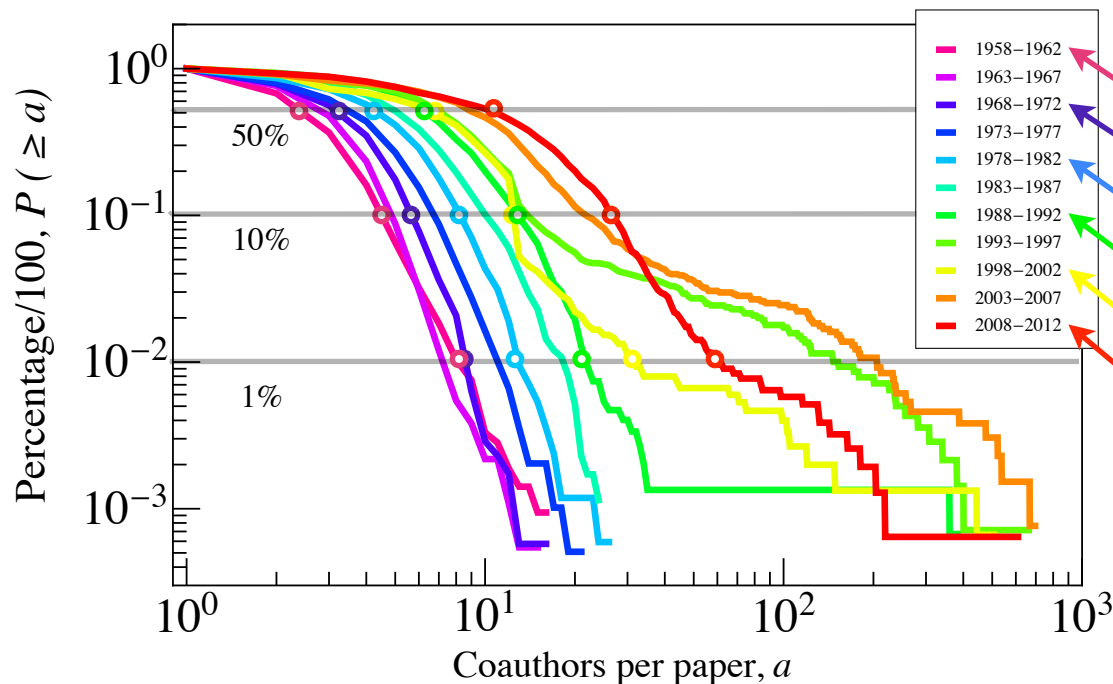2003–2007
2008–2012

$P (\geq a)$, Percent/100

Coauthors per paper, a

$P (\geq a)$, the fraction of all papers with team size of at least size a

New England Journal of Medicine (NEJM)

$P (\geq a)$, Percent/100

Coauthors per paper, a

0.93
0.79
0.53
0.006
0.003

1968-1972
$1 - 0.79 = 0.21$
papers were single author

2008-2012
$1 - 0.93 = 0.07$
papers were single author

1958–1962
1963–1967
1968–1972
1973–1977
1978–1982
1983–1987
1988–1992
1993–1997
1998–2002
2003–2007
2008–2012

Connecting the dots reveals the persistent growth of team size in R&D

New England Journal of Medicine (*NEJM*)

Percentage/100, $P(\geq a)$ vs Coauthors per paper, a

Legend:
- 1958–1962
- 1963–1967
- 1968–1972
- 1973–1977
- 1978–1982
- 1983–1987
- 1988–1992
- 1993–1997
- 1998–2002
- 2003–2007
- 2008–2012

50%, 10%, and 1% of team sizes are greater than:

NEJM

	50%	10%	1%
$a =$	2	4	8
$a =$	3	5	8
$a =$	4	8	12
$a =$	6	12	21
$a =$	6	12	30
$a =$	10	26	60

Physical Review Letters

	50%	10%	1%
$a =$	2	4	10
$a =$	2	5	13
$a =$	3	8	38
$a =$	3	7	69
$a =$	3	8	226
$a =$	4	11	550

Connecting the dots reveals the persistent growth of team size in R&D

New England Journal of Medicine (*NEJM*)

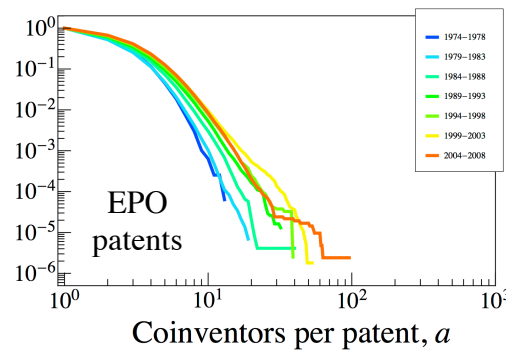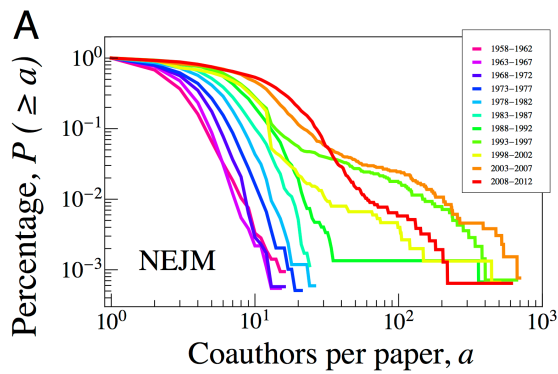50%, 10%, and 1% of team sizes are greater than:

	50%	10%	1%
$a =$	2	4	8
$a =$	3	5	8
$a =$	4	8	12
$a =$	6	12	21
$a =$	6	12	30
$a =$	10	26	60

NEJM

4% annual growth rate τ
(~ 30-year mortgage interest-rate)

$\tau = 0.040(3)$

$\langle a \rangle$ ●
σ_a ▲
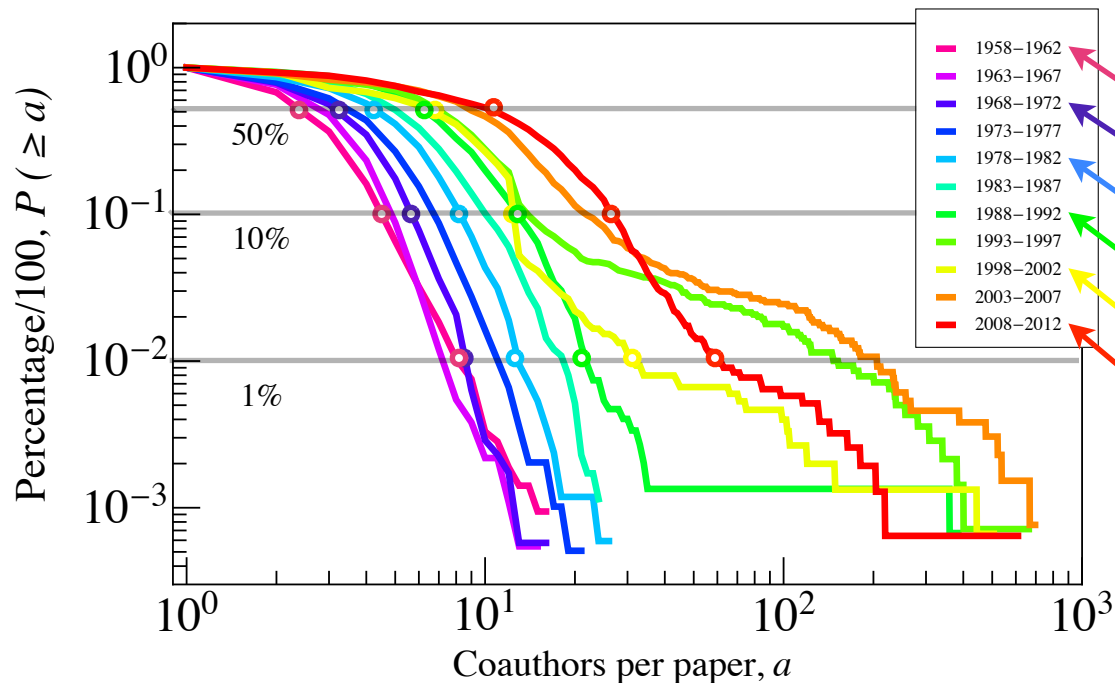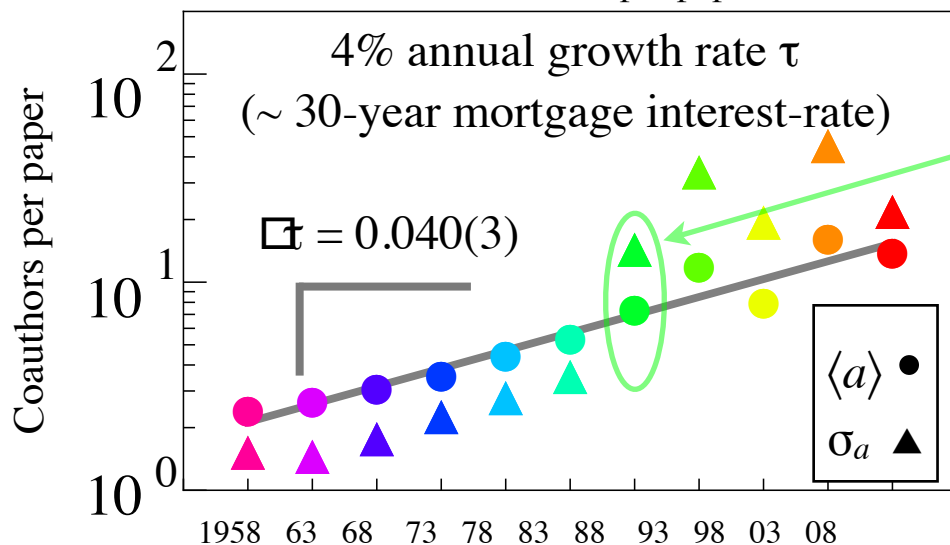
Arrival of big-team science
in ~ 1998-1992
when the standard deviation
first exceeded the average,

$$\sigma_a > \langle a \rangle$$
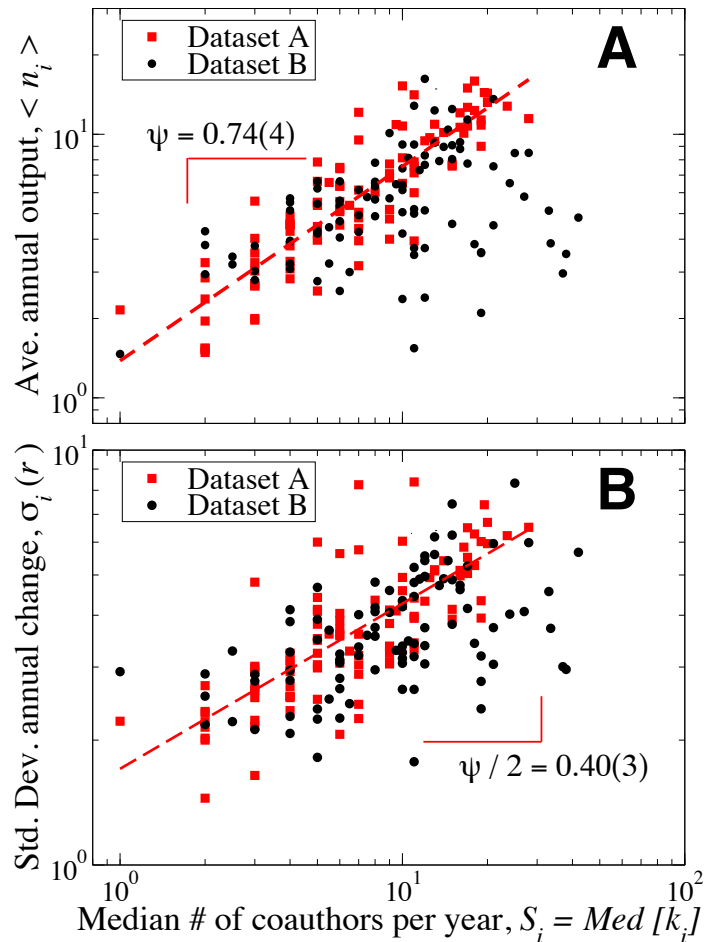
A

Diverse disciplines

Dataset	Years	Articles / Patents	Team size growth rate τ
Cell	1978 – 2012	11,637	0.035(1)
14 Economics journals	1958 – 2012	36,466	0.013(1)
New England J. Medicine	1958 – 2012	18,347	0.040(3)
Physical Review Letters	1958 – 2012	98,739	0.045(4)
European Patent Office	1974 – 2008	2,207,204	0.011(1)
Patent Cooperation Treaty	1979 – 2008	1,695,339	0.018(2)

Regularities allow for Future projections....
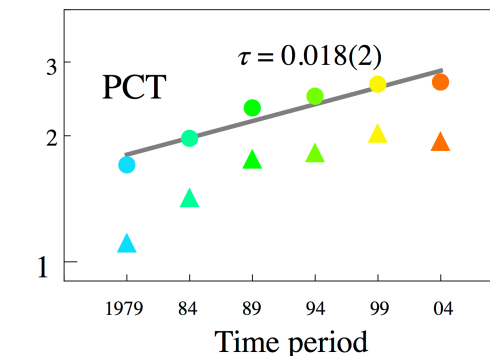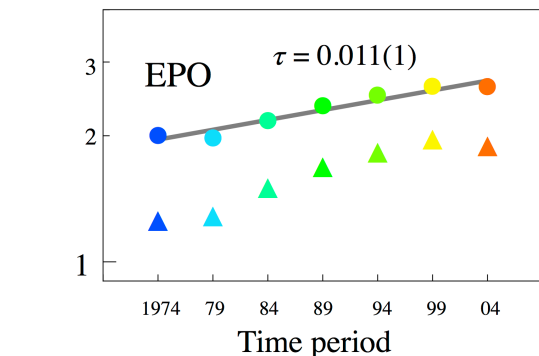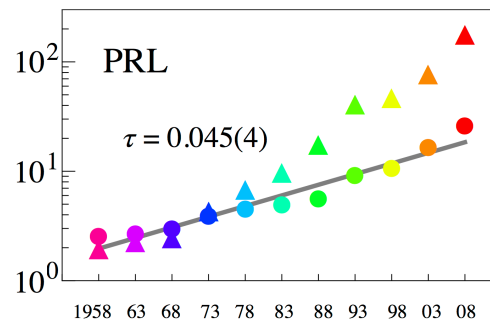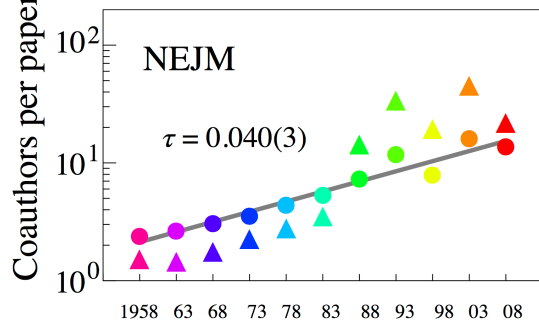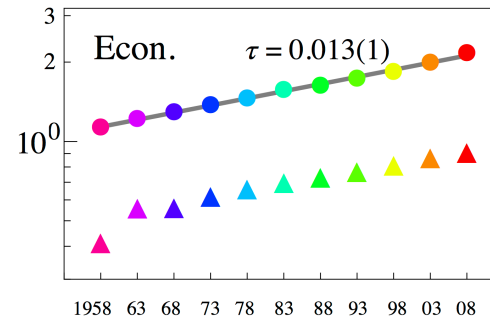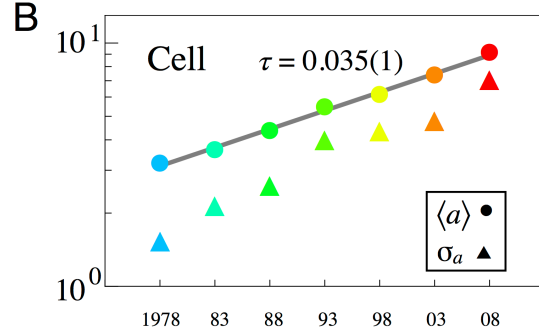
For example, if we extend the growth trend observed for the journal *Cell* over the last 35 years, extrapolating to the year 2050, the average team size is likely to be around 34 coauthors per paper. For PRL and NEJM the predictions are 105 and 74 coauthors per publication, respectively.

For comparison, repeating the same extrapolation for the European Patent Office (EPO) growth trend, suggests that by 2050 the average patent will have roughly 4.2 coinventors, the same average team size for Cell publications in 1988.

Number of coauthors, $k_i(t)$

Ave. annual out

Std. Dev. annual change, $\sigma_i(r)$

Dataset A
Dataset B

B

$\psi / 2 = 0.40(3)$

Median # of coauthors per year, $S_i = Med[k_i]$

Dataset A: Top physicists
Dataset B: random set of prolific physicists

Towards a micro-level production function:

$$\langle n_i \rangle \sim S_i^{\psi}$$

average number of publications per year

S_i is median number of coauthors per year

Output change ("growth fluctuation"),

$$r_i(t) \equiv n_i(t) - n_i(t - \Delta t)$$

There is a decreasing marginal returns (inefficiencies aggregate sub-linearly, $\psi < 1$) with increasing collaboration radius S, likely attributable to team management inefficiencies,

std. deviation of publication change

$$\sigma_i(r) \sim S_i^{\psi/2}.$$

productivity fluctuation scale

team efficiency parameter ψ

Team (in)efficiency

Q: How does annual productivity depend on the number of "labor inputs" ?

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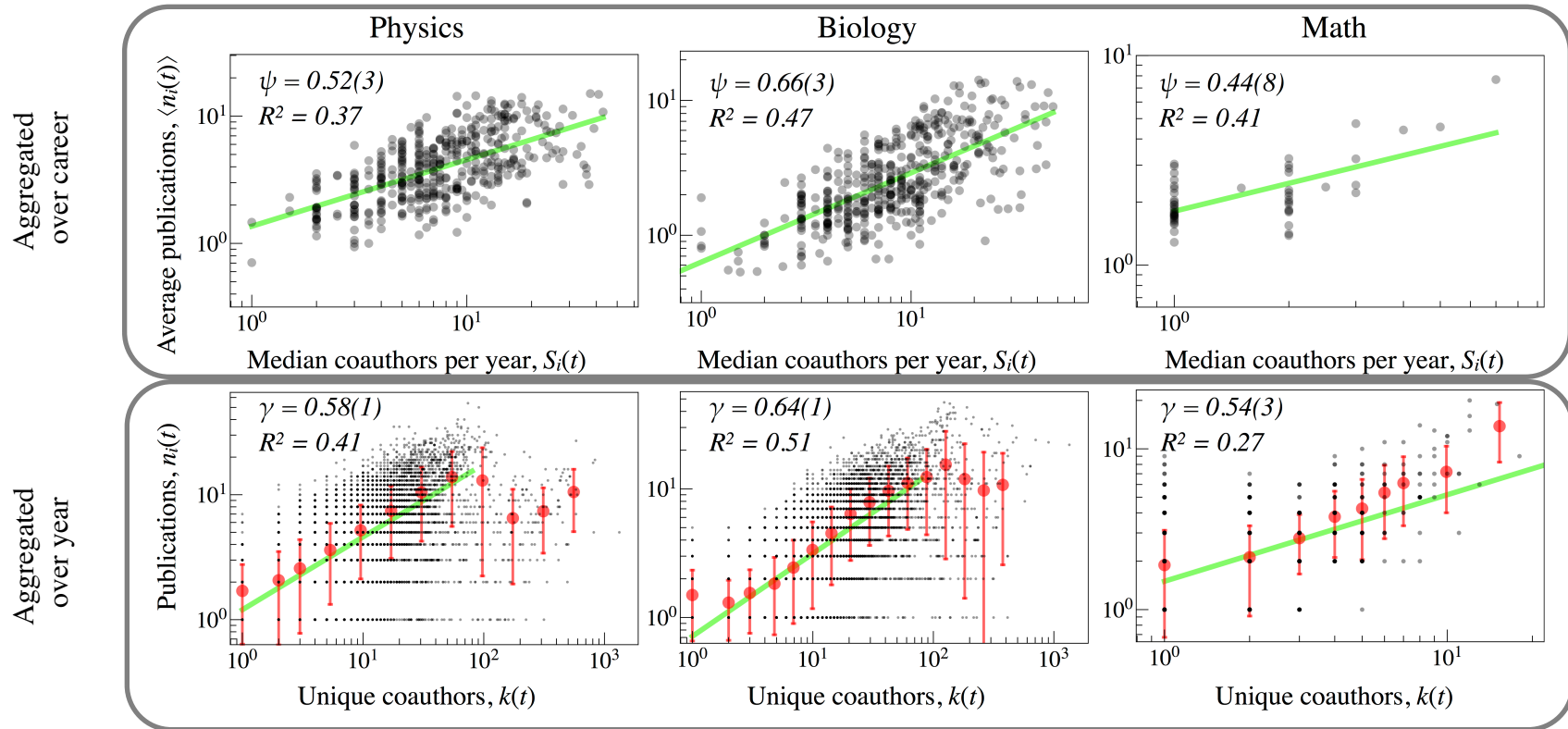
We measure the input-output relation using two aggregation methods, which both yield sub-linear scaling relations with efficiency parameters $\psi \approx \gamma$ and $\psi, \gamma < 1$

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

$\gamma_{\text{Top100 Phyics}} = 0.68(1) > \gamma_{\text{Prolific Physics}} = 0.52(1), \gamma_{\text{AsstProfessor Physics}} = 0.51(2)$

Patterns of collaboration tie-strength
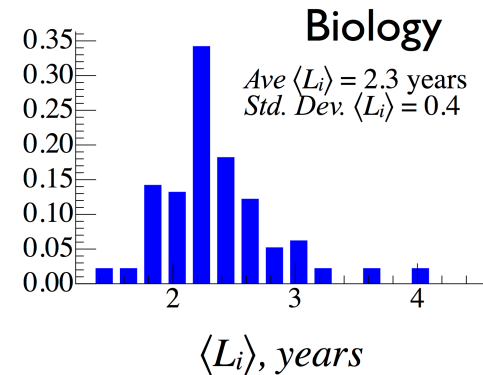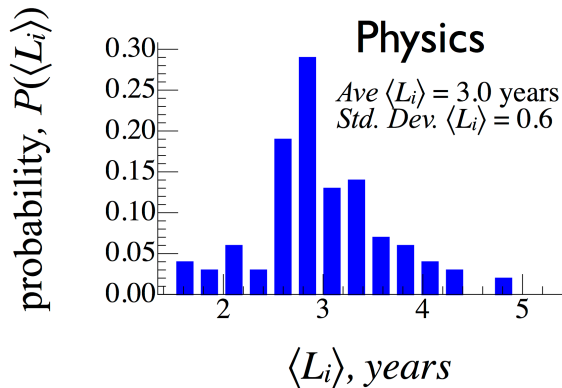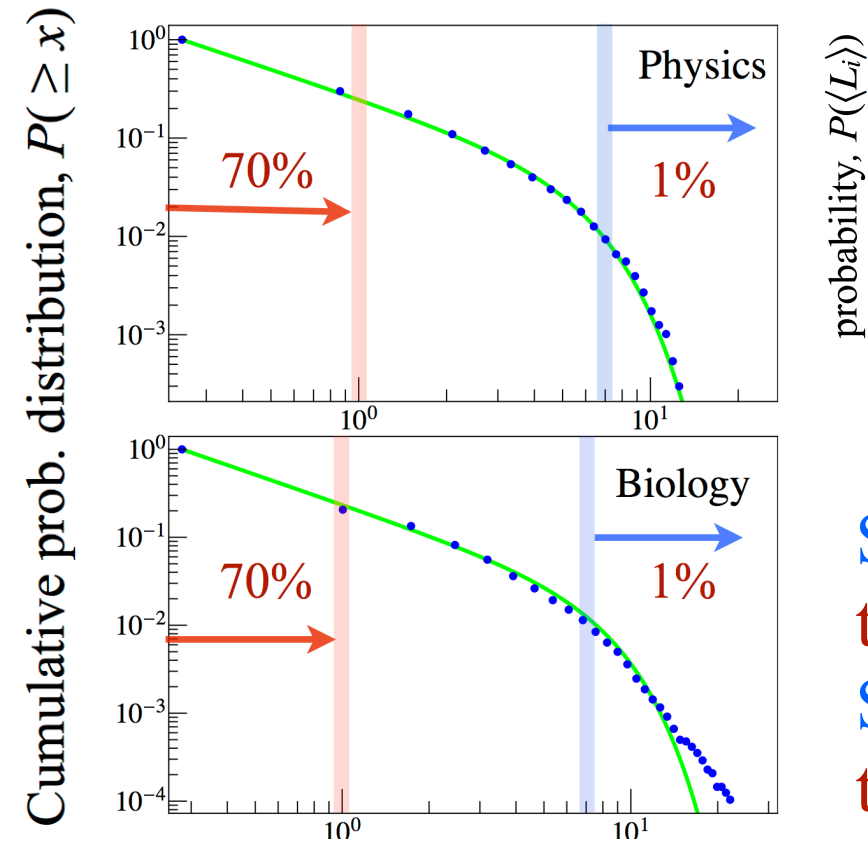
Spurious ties: 70% of collaboration ties last less than $<L>$

Strong ties: only ~1% last longer than $7 <L>$

Is the "invisible college" held together by weak-ties ? (short-term grad/postdoc collaborations) How much does this contribute to team inefficiency?

Paradigm shif... ...ficiency

The growth of team e... ... multiple size scales re... introspection and insti... of the norms of team e...

Scientist incentives are changing

6 ethics issues in team settings:
(i)   Credit/Blame
(ii) Parasitic coauthorship (freeloading), and sanctioning of bad behavior in team setting
(iii) Conflicts of interest
(iv) Breakdown of the mentor-trainee relation and virtue ethics
(v) International variations in ethics codes
(vi) Universality "One-size-fits-all" of team ethics

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]*

Ethical scandals reveal the price of success

"...one survey estimated that almost 7% of students in US universities have used prescription stimulants [Adderall and Ritalin] in this way, and that on some campuses, up to 25% of students had used them in the past year. These students are early adopters of a trend that is likely to grow, and indications suggest that they're not alone."

Towards responsible use of cognitive-enhancing drugs by the healthy

Society must respond to the growing demand for cognitive enhancement. That response must start by rejecting the idea that 'enhancement' is a dirty word, argue **Henry Greely and colleagues**.

NATURE|Vol 456|11 D

Professor'

The use of cognitive-enhanci... ...questions that should not be ignored, argue...

NATURE|Vol 450|20

NATURE|Vol 452|10 April 2008

Poll results: lo

In January, *Nature* launched an informa... **Maher** has waded through the results a...

"One in five respondents said they had used drugs for non-medical reasons to stimulate their focus, concentration or memory. Use did not differ greatly across age-groups..., which will surprise some. "

Team Ethics: Credit distribution in large team science

The reward system in science developed during a period when teams were relatively small. Hence, there is an inherent difficulty in distributing fairly sliced credits in large modular teams comprised of *heterogenous* members
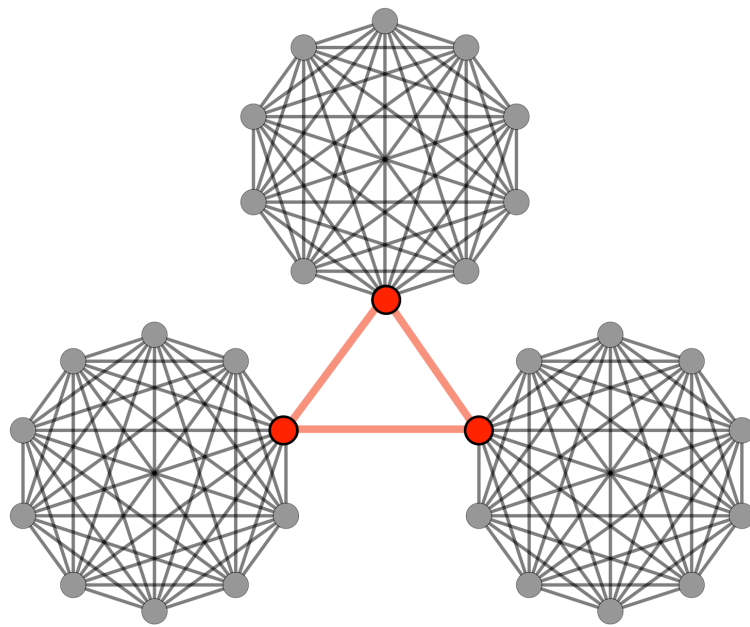
$a = 30, N = 138$

2008-2012
NEJM (Medicine), $P (\geq 30) = 0.065$
PRL (Physics), $P (\geq 30) = 0.040$
Cell (Biology), $P (\geq 30) = 0.017$

Cutting the "credit pie" fairly:
Who gets credit? "Who's on first"?

Citation (impact) credit:
- Is it shared equally amongst a coauthors?

Fraud/Retraction anti-credit:
- can impact all a coauthors
- If credit is shared equally then should blame also?

~ factor of 20 increase in retractions from 2000 - 2010
The retraction penalty: Evidence from the web of science.
Lu SF, Jin GZ, Uzzi B, Jones B. Scientific Reports 3, 3146 (2013).

What makes science special (complex)?
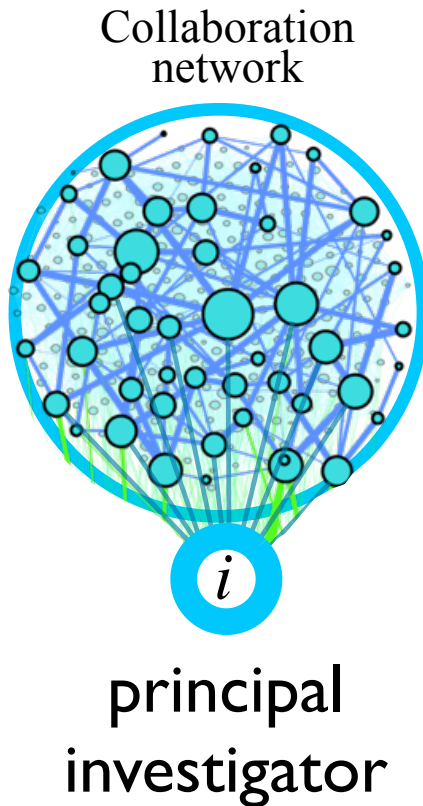
Collaboration network

principal investigator

Interactions mediated by social "forces":

- Collaboration (attractive)

- Competition for priority (repulsive)

- Knowledge (an "exchange particle")

Diverse collaboration strategies

Collaboration
network
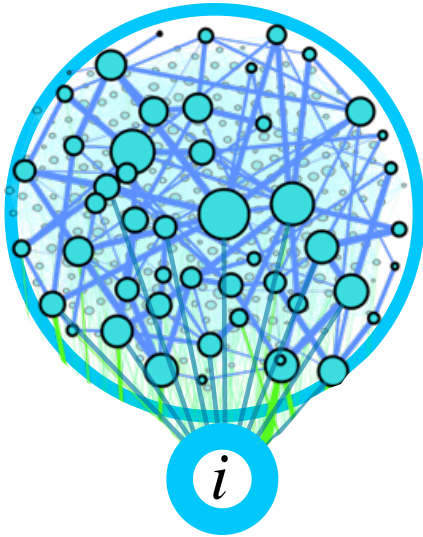
diverse collaboration
strategies even within the
same field!

Interactions mediated by social "forces":

- Collaboration (attractive)

- Competition for priority (repulsive)

- Knowledge (an "exchange particle")

Watson-Crick strategy:

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

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

Solo-artist strategy:

* **Marilyn Kozak**

$N = 70$, $N_{solo} = 59$ (84%)

451
publications

434
(95%)
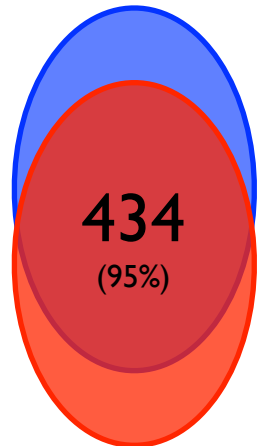
458
publications

Science: a co-evolving network of networks

Collaboration
network

i

Citation network
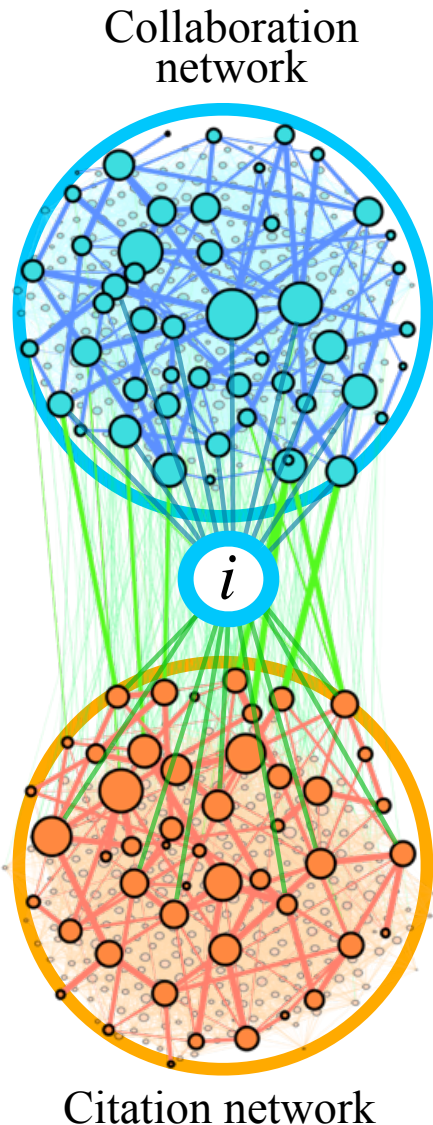
Complexity

- coevolutionary system:
  - knowledge
  - institutions
  - careers
- social processes:
  - behavioral aspects
  - economic incentives
  - cumulative advantage mechanisms
  - collaboration / competition

The context: Stellar (career) growth
a tale of knowledge, collaboration, and reputation spillovers
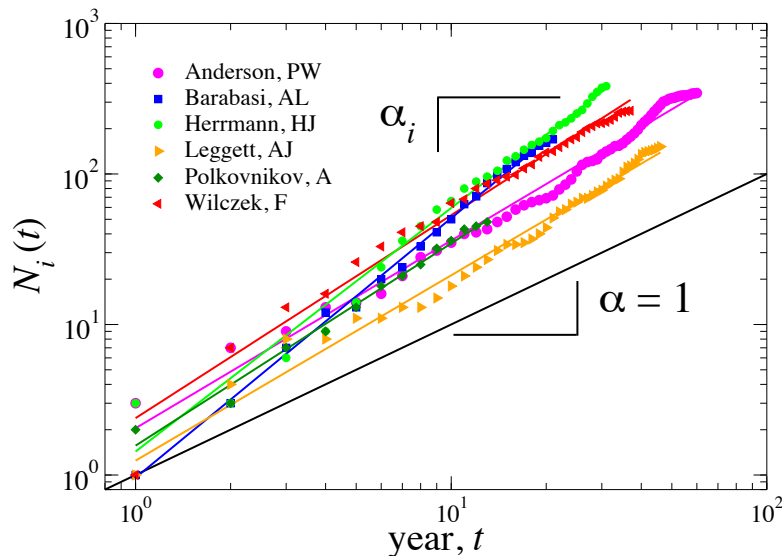
Annual production of individual i

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

Cumulative production, a proxy for career reputation

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

$$\approx A_i t^{\alpha_i} \quad \longleftarrow \quad \text{for many prolific careers!}$$

$\alpha_i > 1$: knowledge, reputation, and collaboration spillovers contribute to sustainable growth across the academic career

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).
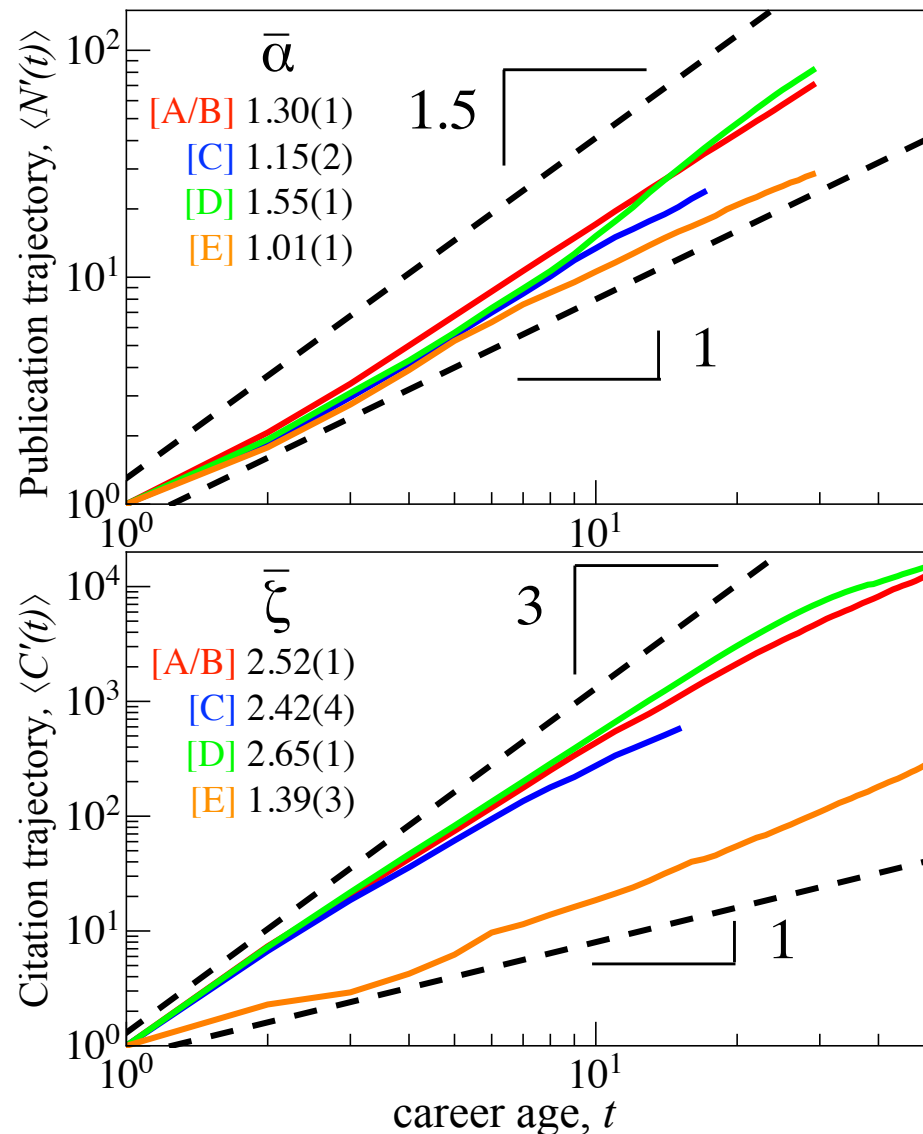
The data:
longitudinal Web of Science publication and citation data for 450 top scientists; 83,693 papers, 7,577,084 citations tracked over 387,103 years

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

Publication trajectory, $\langle N'(t) \rangle$

$\bar{\alpha}$
[A/B] 1.30(1)
[C] 1.15(2)
[D] 1.55(1)
[E] 1.01(1)

1.5

1

Citation trajectory, $\langle C'(t) \rangle$

$\bar{\zeta}$
[A/B] 2.52(1)
[C] 2.42(4)
[D] 2.65(1)
[E] 1.39(3)

3

1

career age, t

Models of science

A) microscopic reputation mechanisms
B) cumulative advantage mechanism
C) competition for limited opportunities
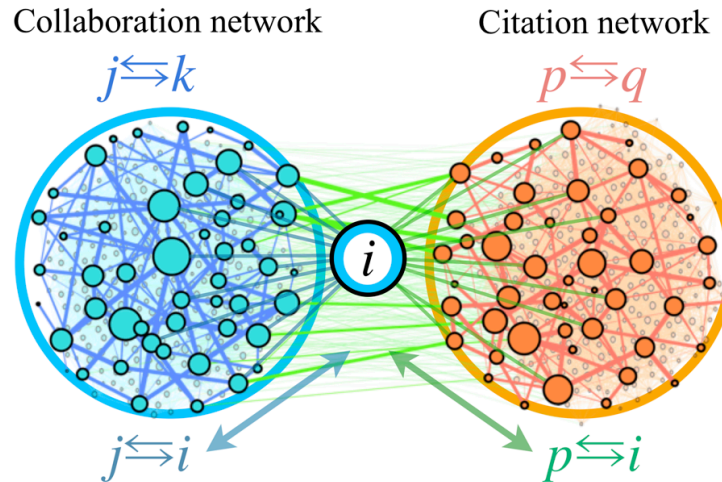
$$j \overset{\longleftarrow}{\longrightarrow} i$$

j

=

$$p \rightleftarrows i$$

C
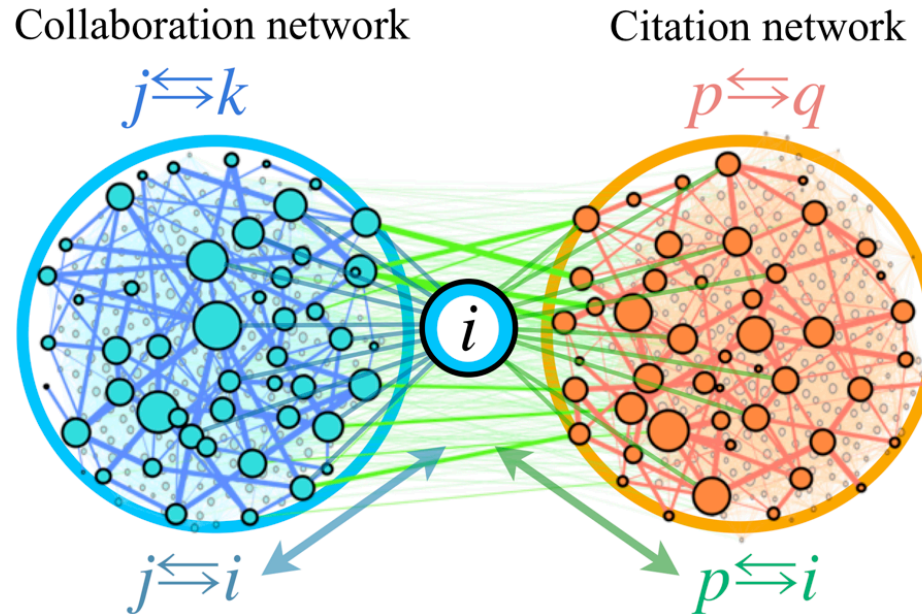
# of new citations in year *t+1* = $\Delta c_{i,p}(t+1) \equiv \eta \times \Pi_p(t) \times A_p(\tau) \times R_i(t)$

1. preferential attachment $\quad \Pi_p(t) \equiv [c_p(t)]^\pi$

2. citation life-cycles $\quad A_p(\tau) \equiv \exp[-\tau_p/\bar{\tau}]$

3. author reputation effect $\quad R_i(t) \equiv [C_i(t)]^\rho$

Author-specific features: $\pi_i, \bar{\tau}_i, \rho_i$

TABLE I: Best-fit parameters for individual careers and the average values within disciplinary datasets. The three features of the citation model are parameterized by π, the paper citation effect, $\bar{\tau}$, the life-cycle effect, and ρ, the reputation effect.

Name	$c(t-1) < c_\times$			$c(t-1) \geq c_\times$			$c_\times$
	π_i	$\bar{\tau}_i$	ρ_i	π_i	$\bar{\tau}_i$	ρ_i	
GOSSARD, AC	0.34 ± 0.027	4.92 ± 0.261	0.25 ± 0.008	0.80 ± 0.048	4.73 ± 0.184	0.09 ± 0.024	
BARABÁSI, AL	0.42 ± 0.036	3.00 ± 0.155	0.29 ± 0.010	1.06 ± 0.016	3.65 ± 0.111	0.01 ± 0.011	40
Ave. ± Std. Dev. [A]	0.43 ± 0.14	5.67 ± 2.52	0.22 ± 0.06	0.96 ± 0.19	8.93 ± 4.09	-0.07 ± 0.11	
BALTIMORE, D	0.32 ± 0.018	4.64 ± 0.148	0.28 ± 0.006	0.62 ± 0.047	5.92 ± 0.250	0.15 ± 0.026	
LAEMMLI, UK	0.54 ± 0.036	5.09 ± 0.297	0.21 ± 0.014	1.09 ± 0.025	6.40 ± 0.255	-0.12 ± 0.019	100
Ave. ± Std. Dev. [D]	0.40 ± 0.14	6.64 ± 6.24	0.26 ± 0.05	0.99 ± 0.22	9.55 ± 26.30	-0.06 ± 0.14	
SERRE, JP	0.33 ± 0.095	15.90 ± 3.724	0.14 ± 0.026	0.66 ± 0.065	20.50 ± 3.862	-0.03 ± 0.039	
WILES, A	0.56 ± 0.208	5.23 ± 1.187	0.24 ± 0.052	0.70 ± 0.059	9.04 ± 0.633	0.10 ± 0.042	20
Ave. ± Std. Dev. [E]	0.27 ± 0.17	30.60 ± 56.80	0.14 ± 0.07	0.54 ± 0.25	21.40 ± 54.30	0.01 ± 0.11	

(left margin, bottom to top:) math | biology | physics

Take home message:

1) The reputation effect is strong for papers not yet highly cited

$$\rho(c < c_\times) > \rho(c \geq c_\times)$$

2) The citation rate of highly-cited papers is largely independent of the author reputation

$$\pi(c < c_\times) < \pi(c \geq c_\times)$$

$$\rho(c \geq c_\times) \approx 0$$

$$\pi(c \geq c_\times) \approx 1 \quad \text{(linear pref. attachment)}$$

Citation boosts attributable to author reputation

TABLE I: Best-fit parameters for individual careers and the average values within disciplinary datasets. The three features of the citation model are parameterized by π, the paper citation effect, $\overline{\tau}$, the life-cycle effect, and ρ, the reputation effect.

Name	$c(t-1) < c_\times$			$c(t-1) \geq c_\times$			$c_\times$
	π_i	$\overline{\tau}_i$	ρ_i	π_i	$\overline{\tau}_i$	ρ_i	
GOSSARD, AC	0.34 ± 0.027	4.92 ± 0.261	0.25 ± 0.008	0.80 ± 0.048	4.73 ± 0.184	0.09 ± 0.024	
BARABÁSI, AL	0.42 ± 0.036	3.00 ± 0.155	0.29 ± 0.010	1.06 ± 0.016	3.65 ± 0.111	0.01 ± 0.011	40
Ave. $\pm$ Std. Dev. [A]	0.43 ± 0.14	5.67 ± 2.52	0.22 ± 0.06	0.96 ± 0.19	8.93 ± 4.09	-0.07 ± 0.11	
BALTIMORE, D	0.32 ± 0.018	4.64 ± 0.148	0.28 ± 0.006	0.62 ± 0.047	5.92 ± 0.250	0.15 ± 0.026	
LAEMMLI, UK	0.54 ± 0.036	5.09 ± 0.297	0.21 ± 0.014	1.09 ± 0.025	6.40 ± 0.255	-0.12 ± 0.019	100
Ave. $\pm$ Std. Dev. [D]	0.40 ± 0.14	6.64 ± 6.24	0.26 ± 0.05	0.99 ± 0.22	9.55 ± 26.30	-0.06 ± 0.14	
SERRE, JP	0.33 ± 0.095	15.90 ± 3.724	0.14 ± 0.026	0.66 ± 0.065	20.50 ± 3.862	-0.03 ± 0.039	
WILES, A	0.56 ± 0.208	5.23 ± 1.187	0.24 ± 0.052	0.70 ± 0.059	9.04 ± 0.633	0.10 ± 0.042	20
Ave. $\pm$ Std. Dev. [E]	0.27 ± 0.17	30.60 ± 56.80	0.14 ± 0.07	0.54 ± 0.25	21.40 ± 54.30	0.01 ± 0.11	

(left margin: math | biology | physics)

The reputation premium: A 66% increase in the citation rate for every 10-fold increase in reputation!

Incentive for Quality > Quantity!
Since ~ 10-15% of an author's C comes from their highest-cited paper alone

consider 2 scientists (with roughly equivalent paper lifecycle factor $\overline{\tau}$), and one with 10× as many total citations as the other,
$$C_1(t) = 10\, C_2(t)\,,$$
then for 2 papers with the same # of citations $c < c_x$ (in the strong reputation regime)

$$\frac{\Delta c_{1,p}(t+1)}{\Delta c_{2,p}(t+1)} = 10^\rho = 1.66$$

c

$$-\chi(z)$$

axis labels: 10^0, 10^{-1}, 10^{-2}

$-\overline{\chi}$

$t\, \zeta_i$

centered rank, $z = r - (N+1)/2$

x-axis: -150, -100, -50, 0, 50, 100, 150

...tic properties of the DGBD.** We graphically illustrate the derivation of the characteristic $c_i(r)$ crossover values that ...imes of $c_i(r)$, in particular, the distinguished "peak" paper regime corresponding to paper ranks $r \leq r^*$ (shaded region). ...n two scaling regimes suggests a complex reinforcement relation between the impact of a scientist's most famous papers ...her other papers. **(a)** The $c_i(r)$ plotted on log-log axes with $N = 278$, $\beta = 0.83$ and $\gamma = 0.67$, corresponding to the ...Dataset [A] scientists. The hatched magenta curve is the $H_1(z)$ line on the log-linear scale with corresponding h-index ...r^* value for $c_i(r)$ is not visibly obvious. **(b)** We plot on log-linear axes the centered citation profile $c_i(z)$ (solid black ...mmetric rank transformation $z = r - z_0$ in Eq. [7]. This representation better highlights the peak paper regime, but fails ...r-law β scaling. **(c)** We plot the corresponding logarithmic derivative $\chi(z)$ of $c(z)$ (solid black curve), which represents ...$c(z)$. The dashed red line corresponds to $-\overline{\chi}$, where $\overline{\chi}$ is the average value of $\chi(z)$ given by Eq. [12]. The values of $\overline{z}_{\pm}$, ...vertical green lines, are defined as the intersection of $\overline{\chi}$ with $\chi(z)$ given by Eq. [13]. The regime $z < \overline{z}_-$ corresponds to ...iven author. The hatched blue line corresponds to $z_{\times}^-$ which marks the crossover between the β and γ scaling regimes.

..., Eom, Y-H., Helbing, D., Lozano, S., Fortu-
...ation boosts promote scientific paradigm shifts
... . *PLoS ONE* **6(5)**, e18975 (2011).
...he Matthew effect in science. *Science* **159**, 56–

...he Matthew effect in science, II: Cumulative

[5] Guimera, R., Uzzi, B., Spiro, J., Amaral, L. A. N. Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**, 697–702 (2005).
[6] Malmgren, R. D., Ottino, J. M., Amaral, L. A. N. The role of mentorship in protégé performance. *Nature* **463**, 622 – 626 (2010).

B) Modeling "Cumulative advantage"

career i
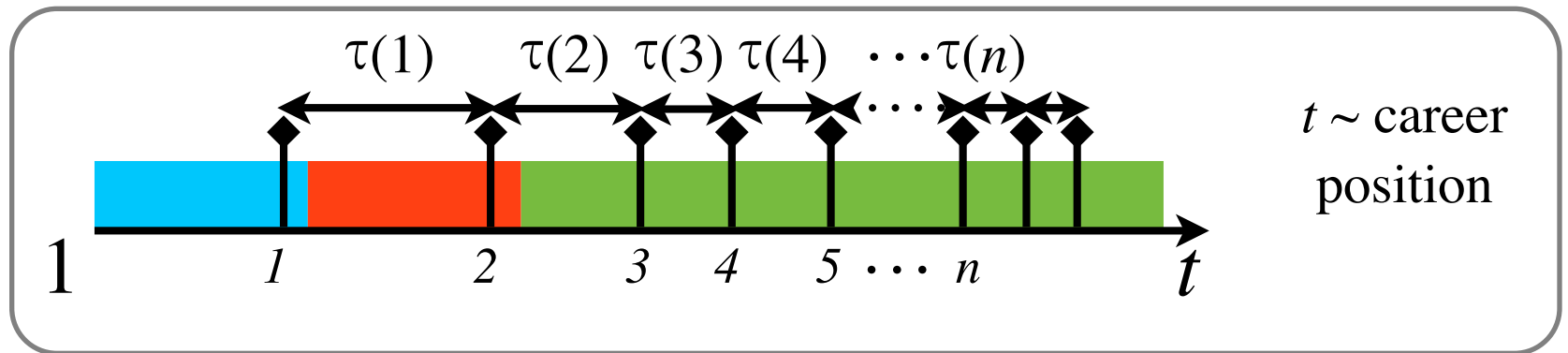
Evidence for cumulative advantage

... his/her longitudinal publication rate by aggregating over ... ns in a *specific set* of high-impact journals

$\tau(1)$ $\tau(2)$ $\tau(3)$ $\tau(4)$ $\cdots \tau(n)$

careers with
$L \geq 5$ and $N_p \geq 10$

— *Nature/PNAS/Science*
— *CELL*
— *NEJM*
— *PRL*
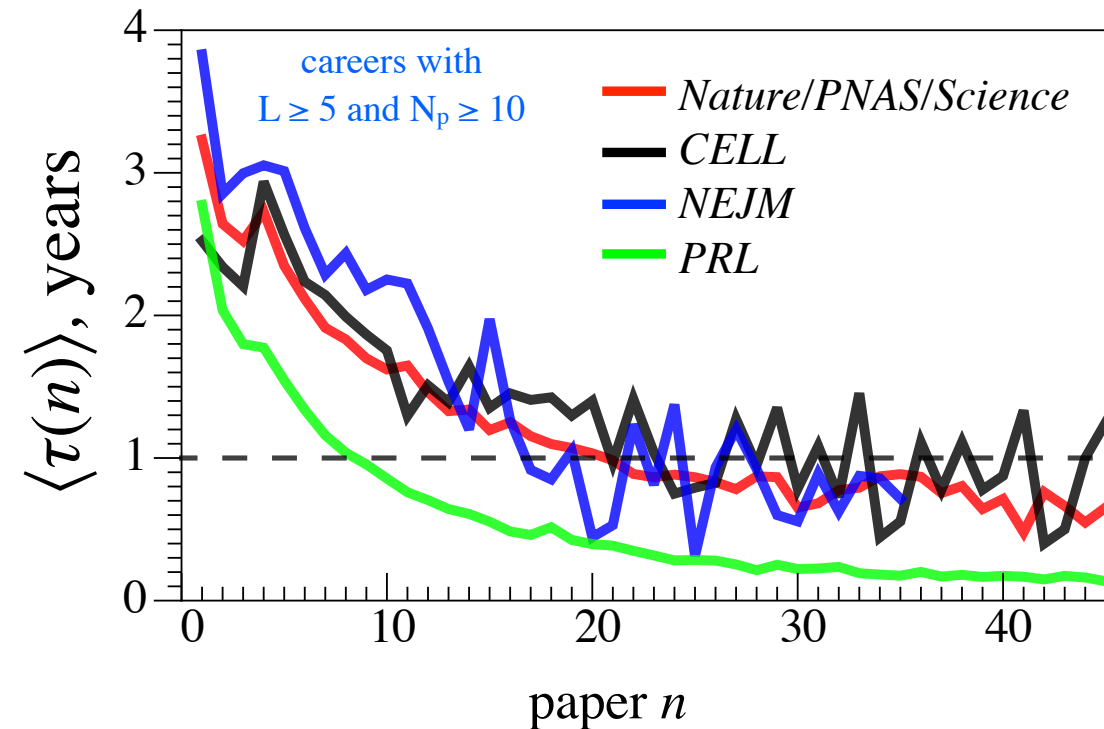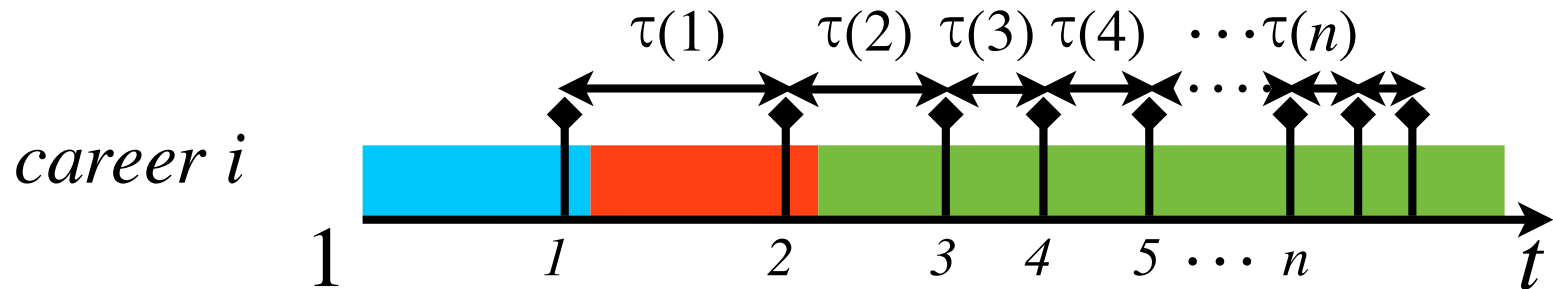
Q: What is the characteristic waiting time $\tau_i(n)$ between an author's n^{th} paper and $(n+1)^{th}$ paper?

By the 10th paper, the waiting time between publications has decreased by ~ factor of 2!

Two main ingredients of the model

1) Forward progress follows a stochastic "progress rate" $g(x)$. Cumulative advantage corresponds to $g(x)$ increasing with career position x

2) Random termination of the career due to hazards (e.g. decreased work performance, economic down, economic downturn, health, retirement, etc.)
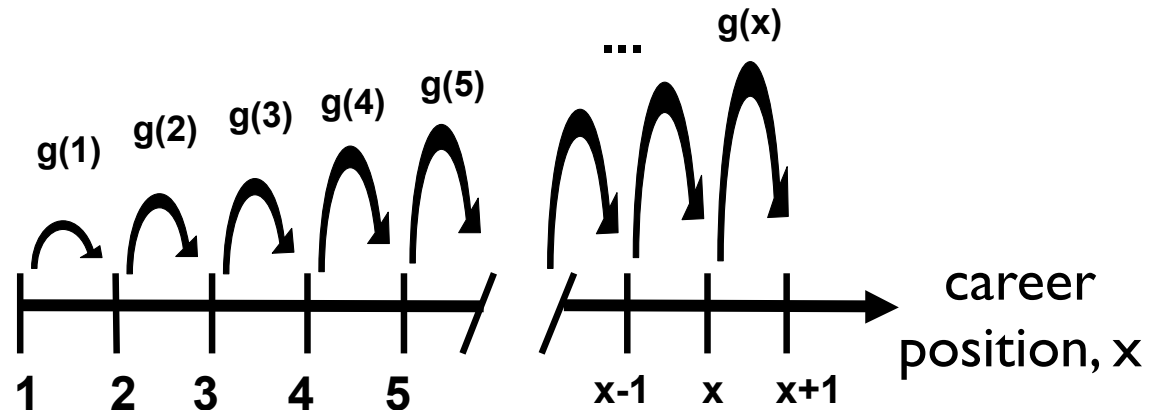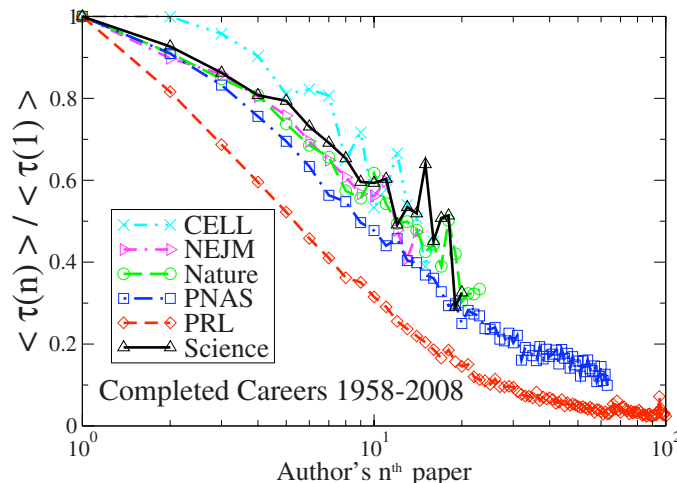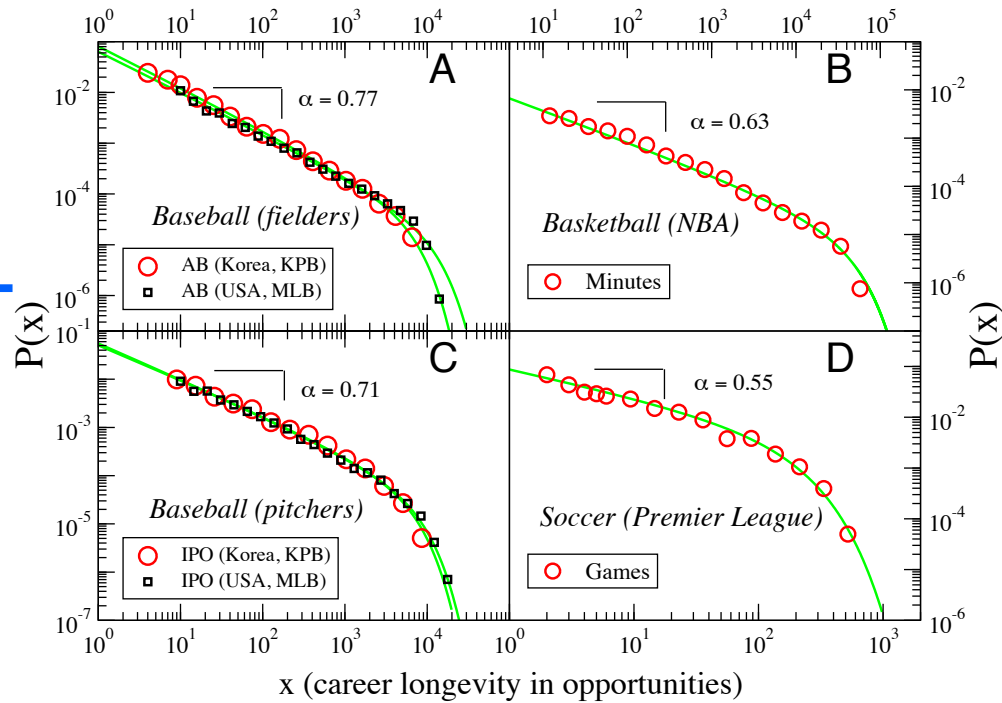

$$g(x) = 1 / \langle \tau(x) \rangle$$

The progress probability g is the inverse of the mean waiting time τ

Methods for measuring the citations and productivity of scientists across time and discipline, A. M. Petersen, F. Wang, H. E. Stanley. Phys. Rev. E 81, 036114 (2010).

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).

$$g(x) \equiv 1 - e^{-(x/x_c)^\alpha}$$

Statistical regularities in the career longevity distribution

Pro Sports

Academia

opportunities ~ time duration

Major League Baseball
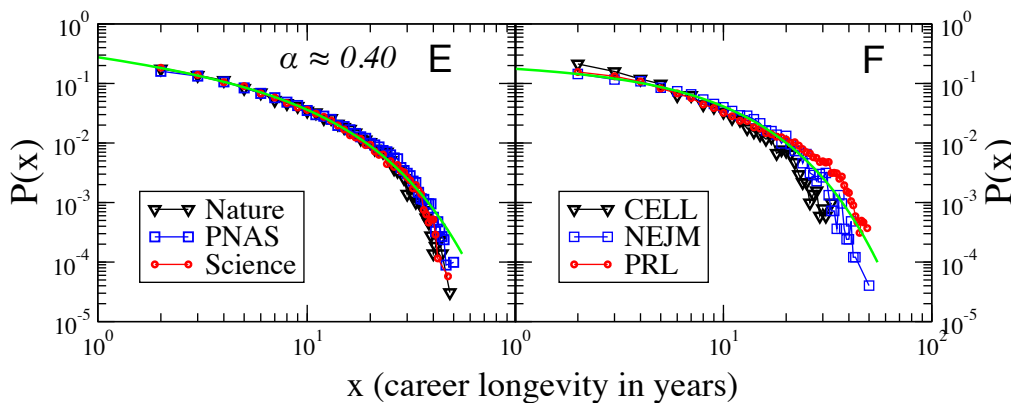
- 130+ years of player statistics, ~ 15,000 careers

``One-hit wonders''

- 3% of all fielders finish their career with ONE at-bat!

- 3% of all pitchers finish their career with less than one inning pitched!

``Iron horses''

- Lou Gehrig (the Iron Horse): NY Yankees (1923-1939)
- Played in 2,130 consecutive games in 15 seasons! 8001 career at-bats!
- Career & life stunted by the fatal neuromuscular disease, amyotrophic lateral sclerosis (ALS), aka Lou Gehrig's Disease
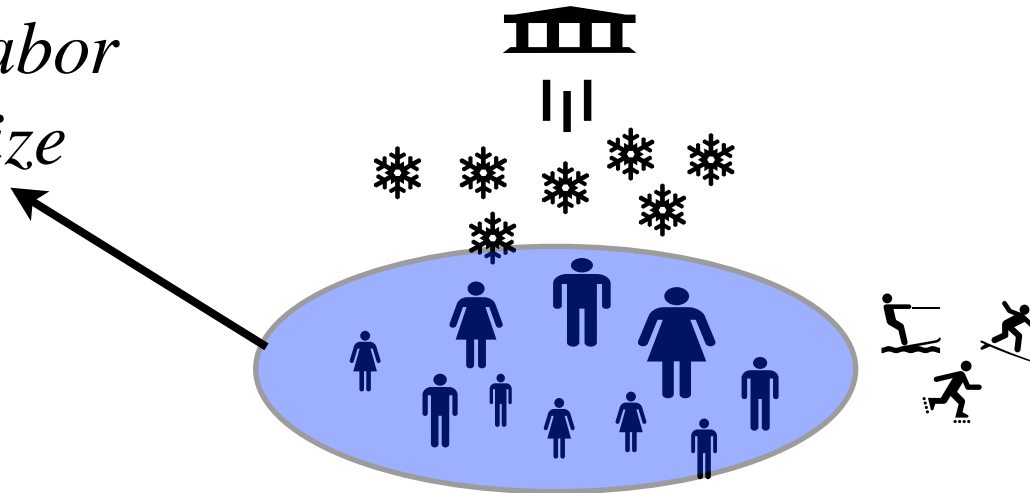
Agent-based competition model with cumulative achievement appraisal (evaluation)

Achievement measured by $n_i(t)$, the number of opportunities (ex. publications) captured in time period t

I = finite labor force size

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).

Appraising prior achievement

Achievement measured by $n_i(t)$, the number of opportunities captured in time period t

gents compete for a fixed number of opportunities in a lifespan of $t = 1...T$ periods.

e capture rate of a given individual i is calculated by an hievement history

$$\text{capture rate} \propto w_i(t) \equiv \sum_{\Delta t=1}^{t-1} n_i(t - \Delta t)e^{-c\Delta t}$$

Appraisal timescale *1/c*

exponential discount factor

er all lifetime achievements (~ tenure system)

only recent achievements (short-term contract system)

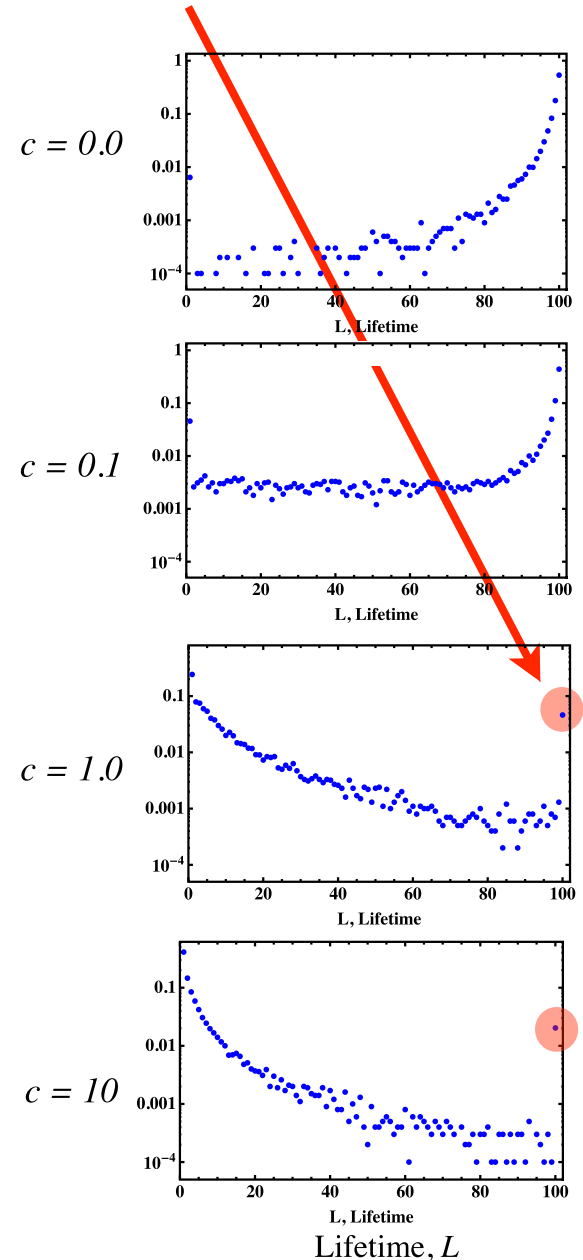# Crowding out by "kingpins"

Our theoretical model suggests that

short-term appraisal systems:

* can amplify the effects of competition and uncertainty making careers more vulnerable to early termination, not necessarily due to lack of individual talent and persistence, but because of random negative production shocks.

* effectively discount the cumulative achievements of the individual.

* may reduce the incentives for a young scientist to invest in human and social capital accumulation.

Longevity probability distributions

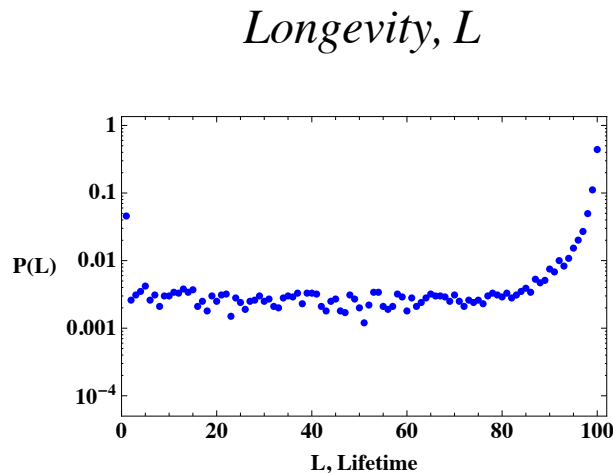Appraisal timescale *1/c*

Long-term

Short-term

$c = 0.0$

$c = 0.1$

$c = 1.0$

$c = 10$

Lifetime, L

Q: Is the...

c = 0.1

linear capture

π = 1.0

π = 1.2

super-linear capture

Plots:
- CDF (x > N) vs N, total number of opportunities
- P(α) vs Career growth scaling exponent, α
- P(r) vs Production change, r
- P(L)

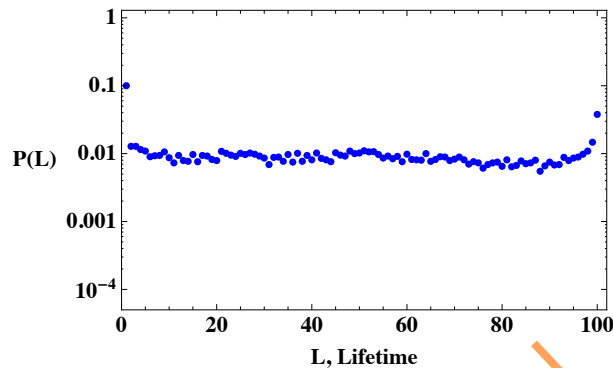**non-linear preferential capture model**

$$\mathcal{P}_i(\ ) = \frac{i(\ )^\pi}{\sum_{i=1}^{I} i(\ )^\pi}$$

Hazard rate $H(L) = -d/dL \, [\ln P(L)]$: conditional probability that failure will occur at time $(L + \delta L)$ given that termination has not yet occurred at time L

$H(L) \approx 0$

hazard rate is not dependent on career position!

Thank you!

A special thanks to my collaborators:

Joel Tenenbaum, Matjaz Perc, Shlomo Havlin, Gene Stanley, Santo Fortunato, Woo-Sung Jung, Kimmo Kaski, Fabio Pammolli, Raj Pan, Orion Penner, Massimo Riccaboni, Sauro Succi, Fengzhong Wang, Jae-Sook Yang, Ionnis Pavlidis, Ionna Semendeferi

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

- *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).*

- *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).*

- *Methods for measuring the citations and productivity of scientists across time and discipline, A. M. Petersen, F. Wang, H. E. Stanley. Phys. Rev. E 81, 036114 (2010).*

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

- *On the Predictability of Future Impact in Science, O. Penner, R. K. Pan, A. M. Petersen, K. Kaski, S. Fortunato. Scientific Reports 3, 3052 (2013).*

- *Reputation and impact in academic careers, A. M. Petersen, S. Fortunato, R. K. Pan, K. Kaski, O. Penner, M. Riccaboni, H. E. Stanley, F. Pammolli, (Under review) ArXiv:1303.7274*

- *As team science evolves, Academia needs new ethics and paradigms, I. Pavlidis, A. M. Petersen, I. Semendeferi, (submitted 2014)*

- *A quantitative perspective on ethics in large team science, A. M. Petersen, I. Pavlidis, I. Semendeferi, ArXiv: 1404.0191*

Title: Quantifying the role of teamwork and reputation across 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. Using comprehensive career data for 450 leading scientists from biology, mathematics, and physics I will discuss patterns of career growth, reflecting on the amplifying role of underlying social processes such as team work and reputation. In the case of teamwork, 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 in light of the increasing prevalence of ``big science''. Indeed, the gradual crowding out of singleton and small team science by large team endeavors is challenging key features of research culture. It is therefore important for the future of scientific practice to reflect upon the scientists' ethical responsibilities within teams. Reputation, on the other hand, is an important social construct in science, which enables informed quality assessments of both publications and careers of scientists in the absence of complete systemic information. However, the relation between reputation and career growth of an individual remains poorly understood, despite recent proliferation of quantitative research evaluation methods. I will discuss an original framework for measuring how a publication's citation rate depends on the reputation of its central author, in addition to its net citation count. I will show how a new publication may gain a significant early advantage corresponding to roughly a 66% increase in the citation rate for each tenfold increase in author reputation.
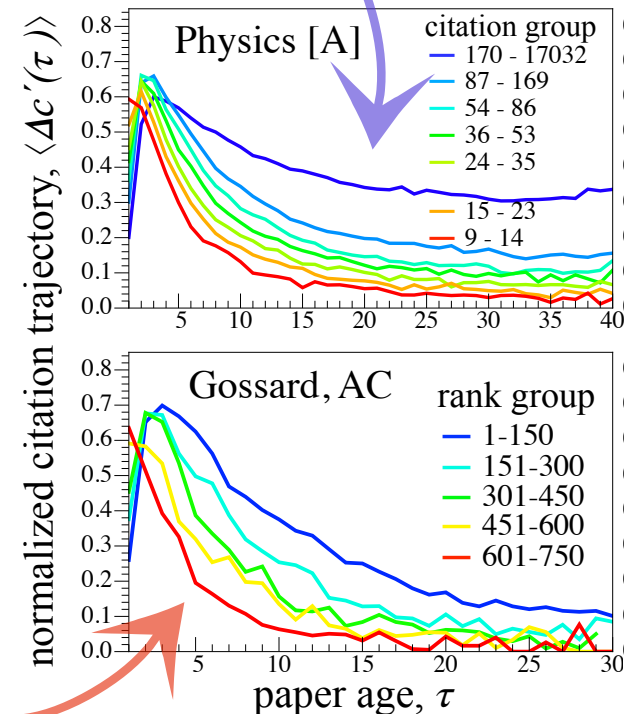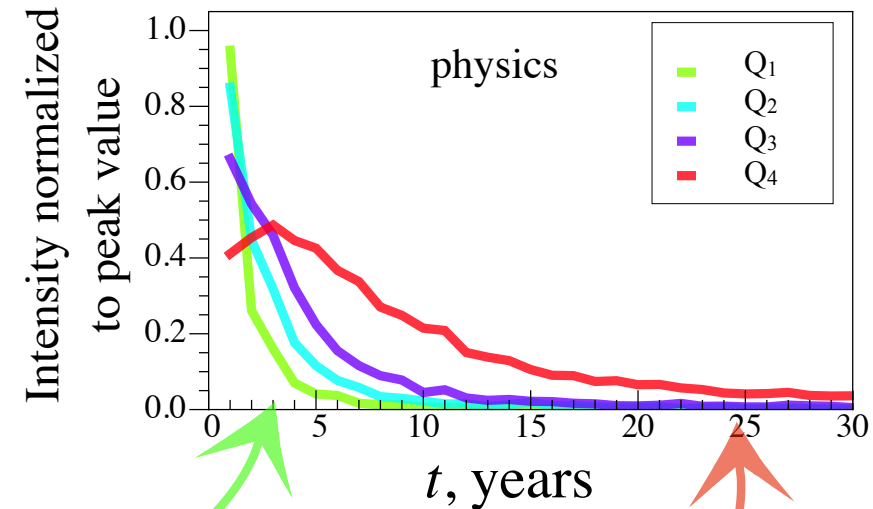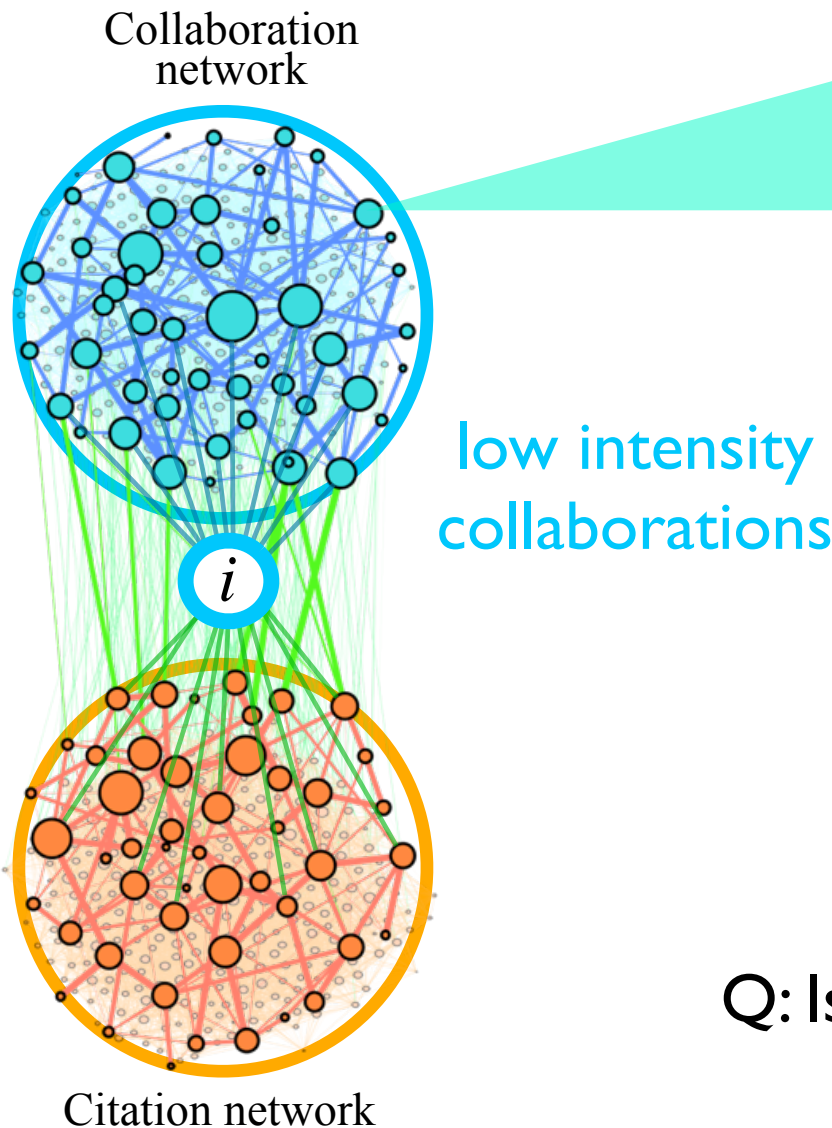
128 Te

Dynamic network characterized by life-cycles

Collaboration network
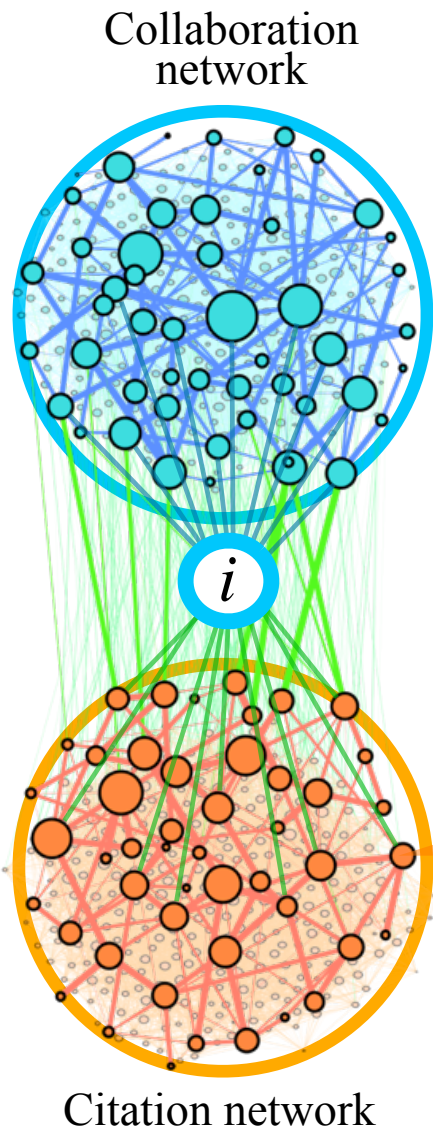
highly-cited papers: extremely long half-life, likely associated w/ axiomatic knowledge or foundational method

i

least-cited papers

Citation network

Physics [A]

citation group
- 170 - 17032
- 87 - 169
- 54 - 86
- 36 - 53
- 24 - 35
- 15 - 23
- 9 - 14

Gossard, AC

rank group
- 1-150
- 151-300
- 301-450
- 451-600
- 601-750

normalized citation trajectory, $\langle \Delta c'(\tau) \rangle$

paper age, τ

Dynamic network characterized by life-cycles

Collaboration network

Citation network

low intensity collaborations

high intensity collaborations

Weak ties ~ 1-3 years
Strong ties ~ lifetime
Q: Is the "Invisible college" held together by weak ties?

physics

Q_1
Q_2
Q_3
Q_4

Intensity normalized to peak value

t, years