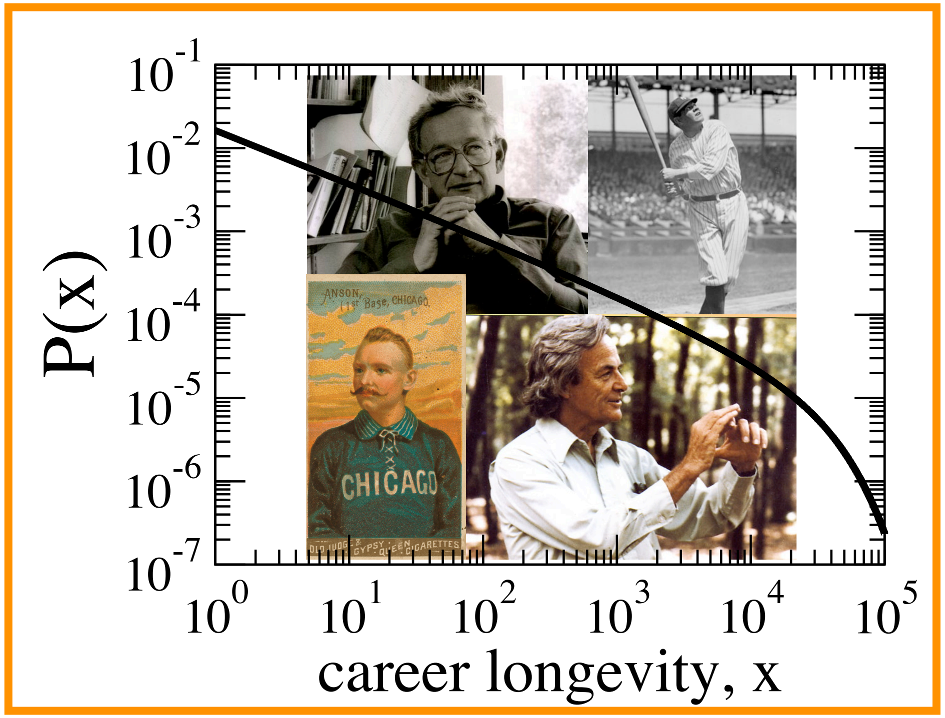
Quantifying Statistical Regularities
in the
Career Achievements of Scientists and Athletes

IES Mini-Workshop
Physics and Complexity in Society
February 2nd-4th, 2012 • POSTECH, Pohang, Korea

Institutions
Markets
Technologies
IMT
INSTITUTE
FOR ADVANCED
STUDIES
LUCCA



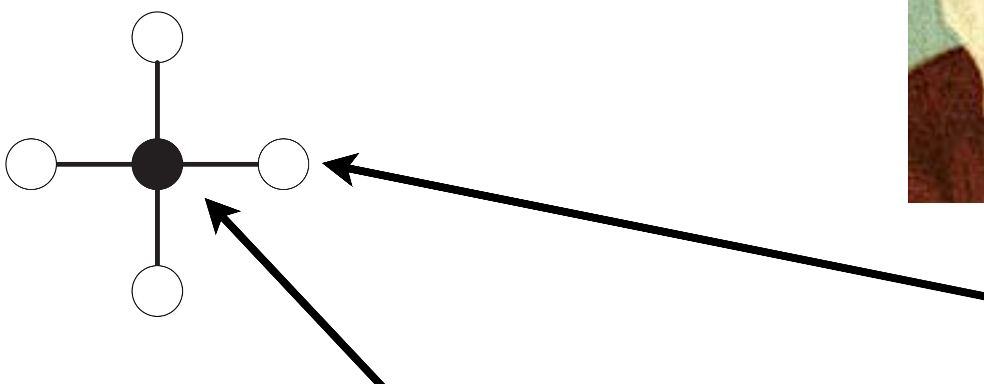
Alexander M. Petersen
IMT Institute for Advanced Studies, Lucca Italy

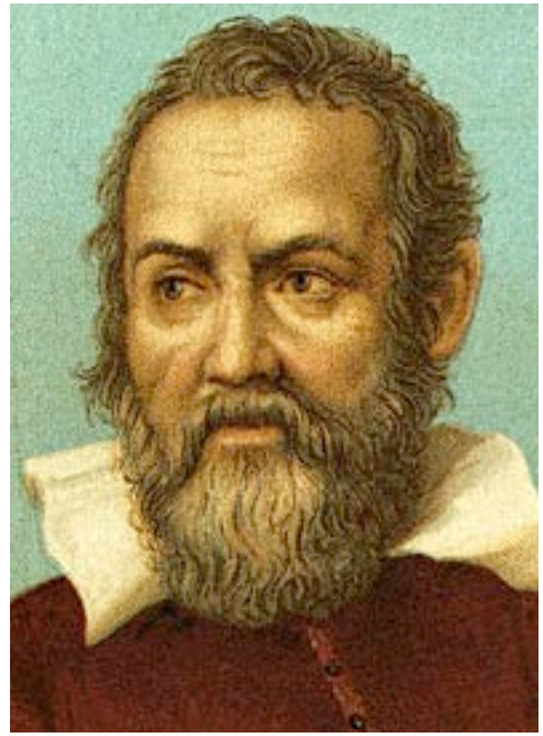
Evolution of Science: "In the beginning..."

"Leaning" Tower of Pisa

- scientific patronage during the renaissance
- the emergence of "open science"

<u>Early scientific labor system</u>



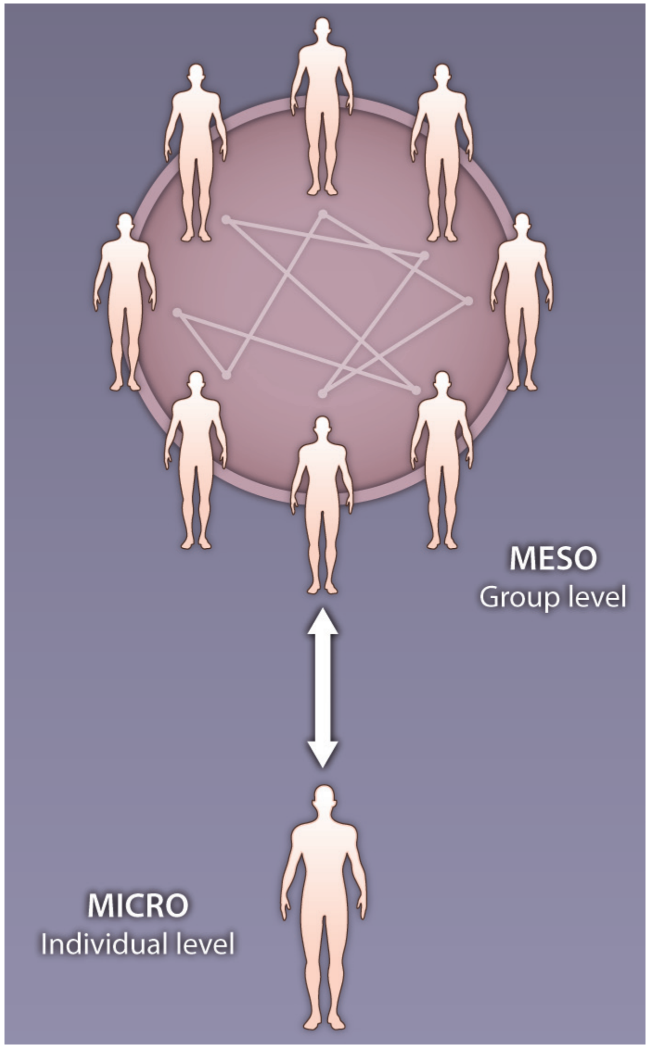


Galileo Galilei

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).

Sociophysics: scientists as interacting "atoms"



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

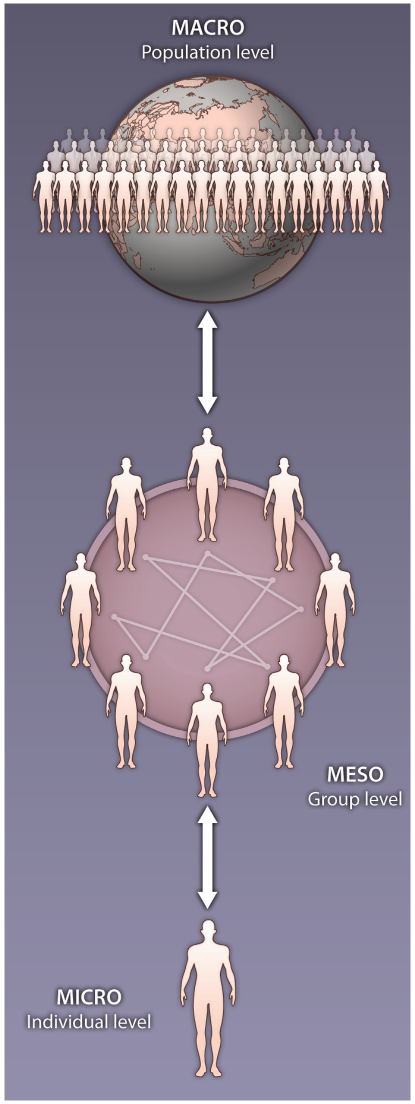
Interactions mediated by social "forces":

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

Institutions Markets Technologies
IMT
INSTITUTE FOR ADVANCED STUDIES LUCCA

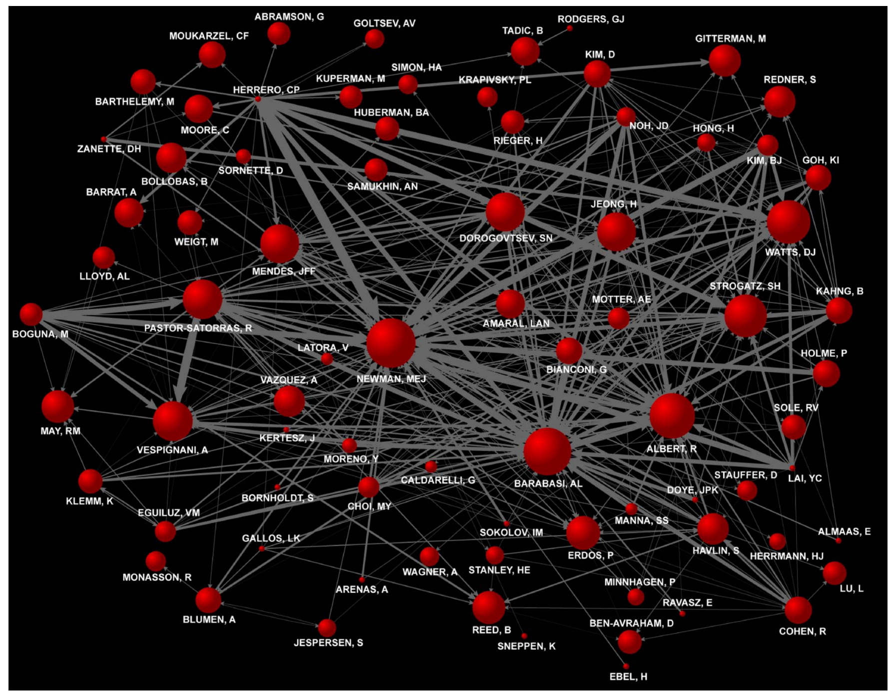
Collaboration & the "Invisible College"

Multi-level system



- The increase in the typical size of scientific collaborations has led to the increasingly difficult task of allocating funding and assigning recognition
- The scientific collaboration radius can change dramatically over the course of a career
- Economics of collaboration:
  - collaboration efficiency
  - knowledge spillovers
  - competition: science is a market for knowledge-based goods
  - size-dependent growth fluctuations (uncertainty)

The network of network scientists



F. Radicchi, et al. Phys. Rev. E **80**, 056103, 2009

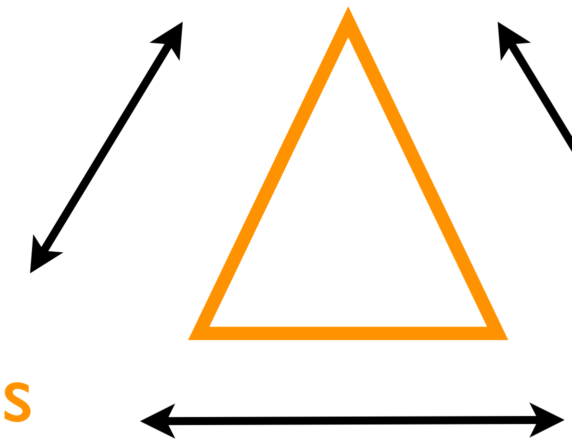
K. Börner, et al. A multi-level systems perspective for the science of team science. Sci. Transl. Med. 2, 49cm24 (2010).

Science of Science

Why study science itself?

- “When the hunter becomes the hunted....”

Academia
(Universities, research institutes)



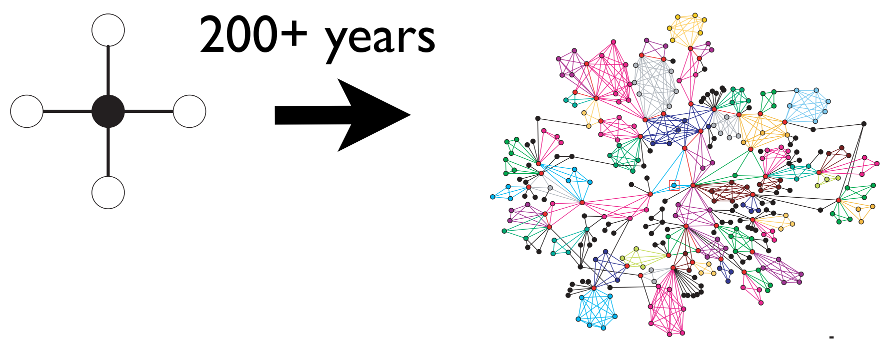
Funding bodies
(Gov’t / non-profit)

Industrial (proprietary)
Science

“the Triple Helix of Science” (Loet Leydesdorff, Univ. of Amsterdam)

Emerging trends in Science

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



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)

Chait RP, ed. The Questions of Tenure. (Harvard University Press, Cambridge USA, 2002).

- 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

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

- shifts in the competitive aspects of science, universities, and scientists: reputation tournaments in omnipresent competition arenas

... Career briefs on the Academic labor market

280 | NATURE | VOL 472 | 21 APRIL 2011

RETHINKING PHDS

Fix it, overhaul it or skip it completely — institutions and individuals are taking innovative approaches to postgraduate science training.

ACADEMIA

The changing face of tenure

Although still highly desirable, tenure is not as prevalent as it was in some places — and that may not be a bad thing.

NOVEMBER 2010 | VOL 468 | NATURE | 123

NATURE | VOL 472 | 21 APRIL 2011

Issues & Perspectives

Academia's Crooked Money Trail

Science Careers From the journal *Science*

10.1126/science.caredit.a1200001

Undergraduates also carry an increasing share of the load, she adds: Their tuition, often paid with student loans, rises as more funds go to research. Their teachers, meanwhile, increasingly are cut-rate adjuncts rather than the famous professors the recruiting brochures boast about.

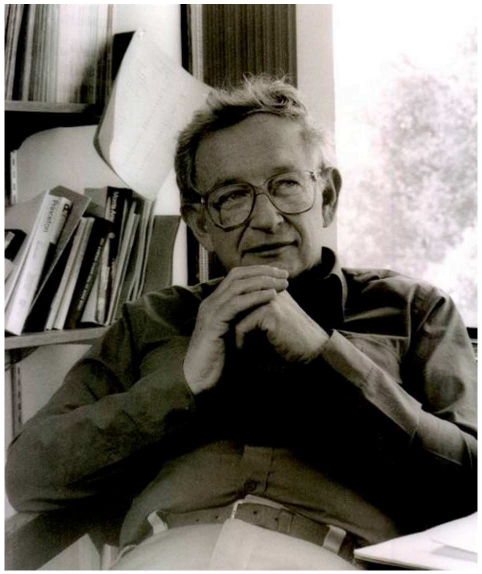
UNITED STATES

Mid-career crunch

Some senior scientists feel neglected by the National Institutes of Health's grant formula.

17 MARCH 2011 | VOL 471 | NATURE | 399

Opening Questions



Using quantitative methods developed in statistical physics to address questions in *sociology*.....

- Are stellar careers an anomaly?
- Are there statistical regularities in *success*?
- Are there universal mechanisms that guide success?

Outline of Quantitative Analysis

1. Empirical evidence: production, impact, longevity
2. Quantitative career longevity model: based on the *Matthew "rich-get-richer" effect*
3. Empirical evidence: Patterns in the growth dynamics of careers
4. Quantitative preferential capture model: featuring the competition for limited opportunities

1. Empirical Facts

Quantifying impact and productivity in science

"Math-letes"

Publication careers of individual scientists within individual journals

Phys. Rev. Lett. 42, 673–676 (1979)

Scaling Theory of Localization: Absence of Quantum Diffusion in Two Dimensions

Abstract | References | Citing Articles (2,099) | Page

E. Abrahams
Serin Physics Laboratory, Rutgers University, Piscataway, New Jersey 08854

P. W. Anderson*, D. C. Licciardello, and T. V. Ramakrishnan†
Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08540

Received 7 December 1978; published in the issue dated 5 March 1979

Arguments are presented that the T=0 conductance G of a disordered electronic system depends on its length scale L in a universal manner. Asymptotic forms are obtained for the scaling function $\beta(G)=d\ln G/d\ln L$, valid for both $G \ll G_c \simeq e^2/\hbar$ and $G \gg G_c$. In three dimensions, G_c is an unstable fixed point. In two dimensions, there is no true metallic behavior; the conductance crosses over smoothly from logarithmic or slower to exponential decrease with L.

For example, P.W. Anderson:
(n = 64 articles
published in *PRL* over this
51-year period)

TABLE I. Summary of data set size for each journal. Total number N of unique (but possibly degenerate) name identifications.

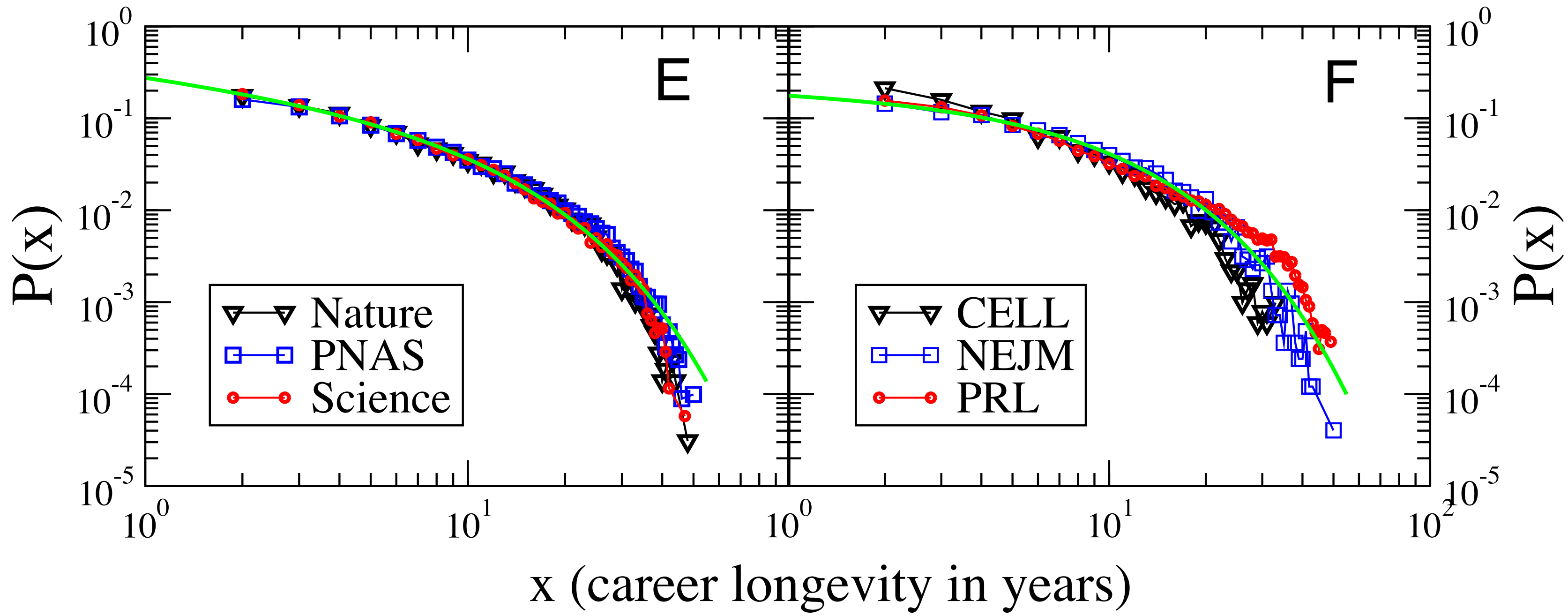
Journal	Years	Articles	Authors, N
CELL	1974–2008	53290	31918
NEJM	1958–2008	17088	66834
Nature	1958–2008	65709	130596
PNAS	1958–2008	84520	182761
PRL	1958–2008	85316	112660
Science	1958–2008	48169	109519

Measures for "career longevity" in academia

- Each author i has n articles in a given journal j. As a proxy for career longevity in academia, we define the journal longevity x as the number of years separating his/her first and last publication in journal j :

$$x_{i,j} = y_{i,j}(f) - y_{i,j}(0) + 1$$

Empirical distributions in high-impact journals



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

- Each author i has n articles in a given journal j. As a proxy for career longevity in academia, we define the journal longevity x as the number of years separating his/her first and last publication in journal j :

$$x_{i,j} = y_{i,j}(f) - y_{i,j}(0) + 1$$

Empirical evidence for the *Matthew Effect*

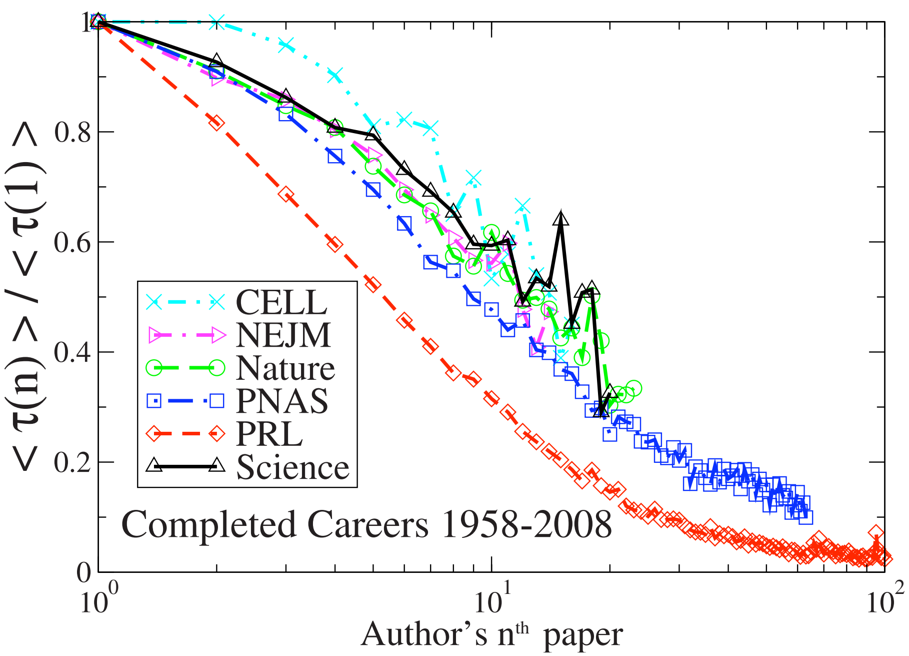
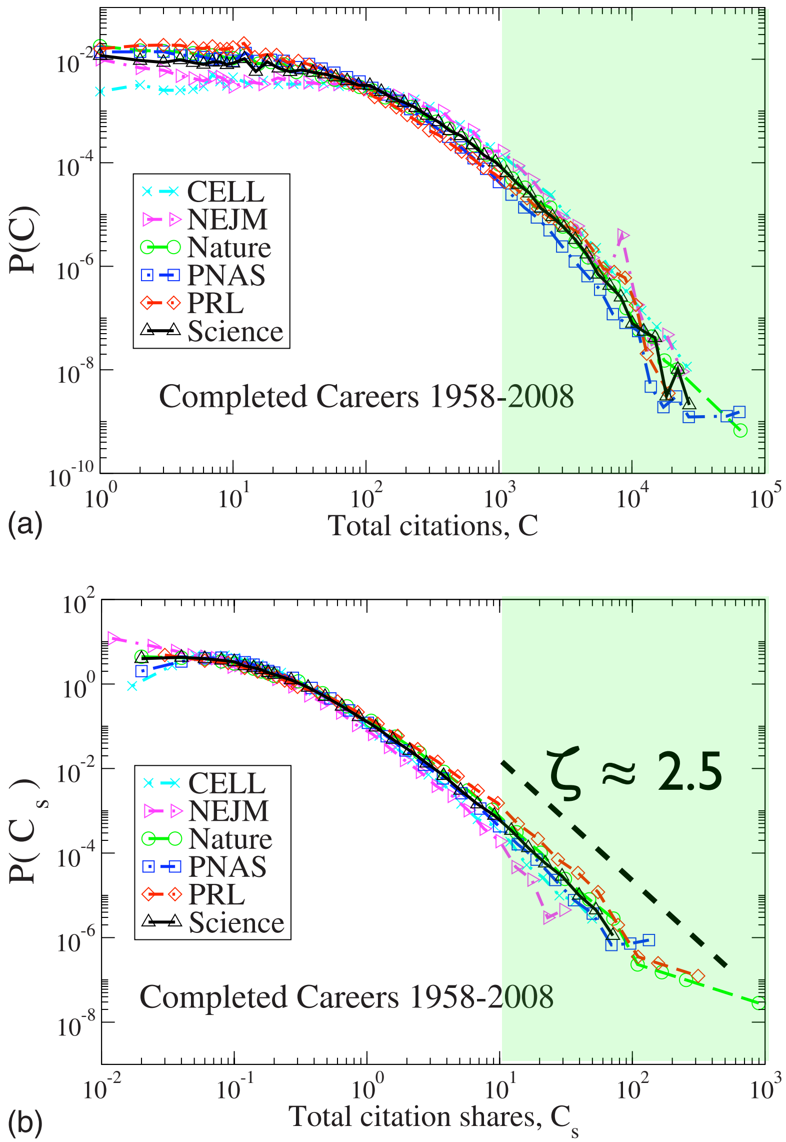


FIG. 7. (Color online) A decreasing waiting time $\tau(n)$ between publications in a given journal suggests that a longer publication career (larger n) facilitates future publications, as predicted by the Matthew effect. We plot $\langle\tau(n)\rangle/\langle\tau(1)\rangle$, the average waiting time $\langle\tau(n)\rangle$ between paper n and paper $n+1$, rescaled by the average waiting time between the first and second publication, $\langle\tau(1)\rangle$. The values of $\langle\tau(1)\rangle$ are 2.2 (*CELL*, *PRL*), 3.0 (*Nature*, *PNAS*, *Science*), and 3.5 (*NEJM*) years.

- For a given journal: the waiting time $\tau(n)$ is the number of years between an author's paper n and paper n+1
- A decreasing $\tau(n)$ indicates that it becomes "easier" to publish in a journal with each successive publication

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

Journals as “arenas for competition”



Each author has n articles in a given journal j.
Each article i, published in year y, can be quantified by the number of citations c_i it has received at the time of data extraction.
(May, 2009)

Two possible ways to measure citations:

(i) Total citations:

$$C = \sum_{i=1}^{n} c_i .$$

(ii) Total citations ``shares”:

$$C_s = \sum_{i=1}^{n} \frac{1}{a_i} \frac{c_i(y)}{\langle c(y) \rangle} .$$

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

Top-20 “champions” of *Physical Review Letters*

Total citations ``shares”:

$$C_s = \sum_{i=1}^{n} \frac{1}{a_i} \frac{c_i(y)}{\langle c(y) \rangle}.$$

Traditional total citations:

$$C = \sum_{i=1}^{n} c_i.$$

PRL		
Name	C_s	n
WEINBERG, S	313.3	49
ANDERSON, PW	137.4	64
WILCZEK, F	120.0	62
TERSOFF, J	105.1	76
HALDANE, FDM	102.3	38
YABLONOVITCH, E	87.5	21
PERDEW, JP	78.3	20
LEE, PA	74.6	76
PENDRY, JB	74.1	29
PARRINELLO, M	72.8	68
FISHER, ME	71.6	67
CIRAC, JI	66.7	97
HALPERIN, BI	66.7	50
RANDALL, L	63.4	14
BURKE, K	63.2	18
JOHN, S	62.8	20
GEORGI, H	61.9	26
CAR, R	59.8	51
GLASHOW, SL	59.6	37
CEPERLEY, DM	58.9	39

2. Longevity Model

The "righ-get-richer" Matthew Effect:

"For to all those who have, more will be given, and they will have an abundance"

Gospel of St. Matthew 25: 29

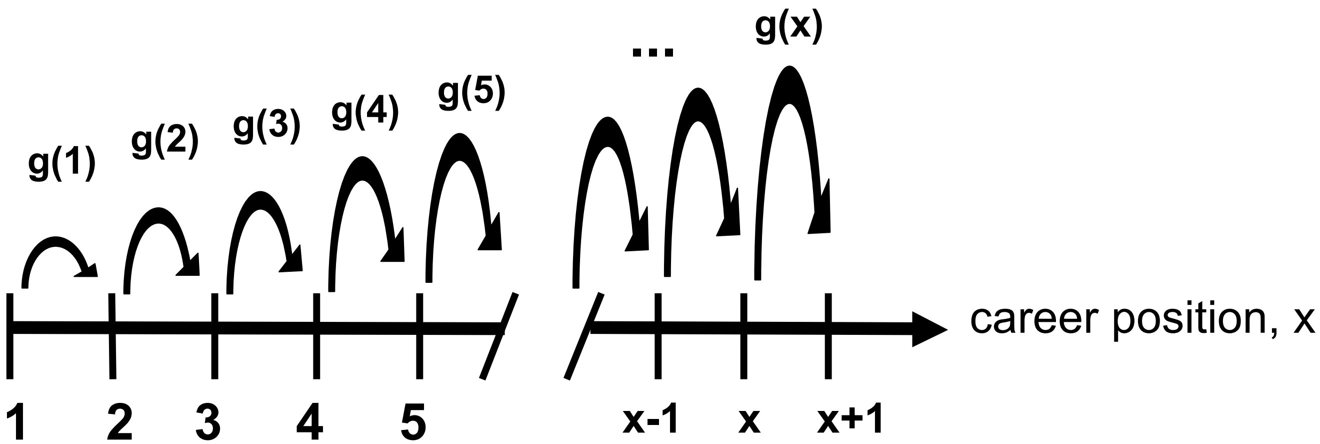
~ 2000+ year old sociological effect!

A stochastic model for career longevity

- Ingredient I: Random forward progress
  Experience and reputation can provide positive feedback in sustaining a career (generic "rich-get-richer" effect)
- Ingredient II: Random termination time
  Career must survive through a horizon of hazards which eventually terminate the career

Ingredient 1: Random forward progress

- Forward progress is made according to the "progress rate" $g(x)$
- Matthew Effect: $g(x)$ increases with career position x



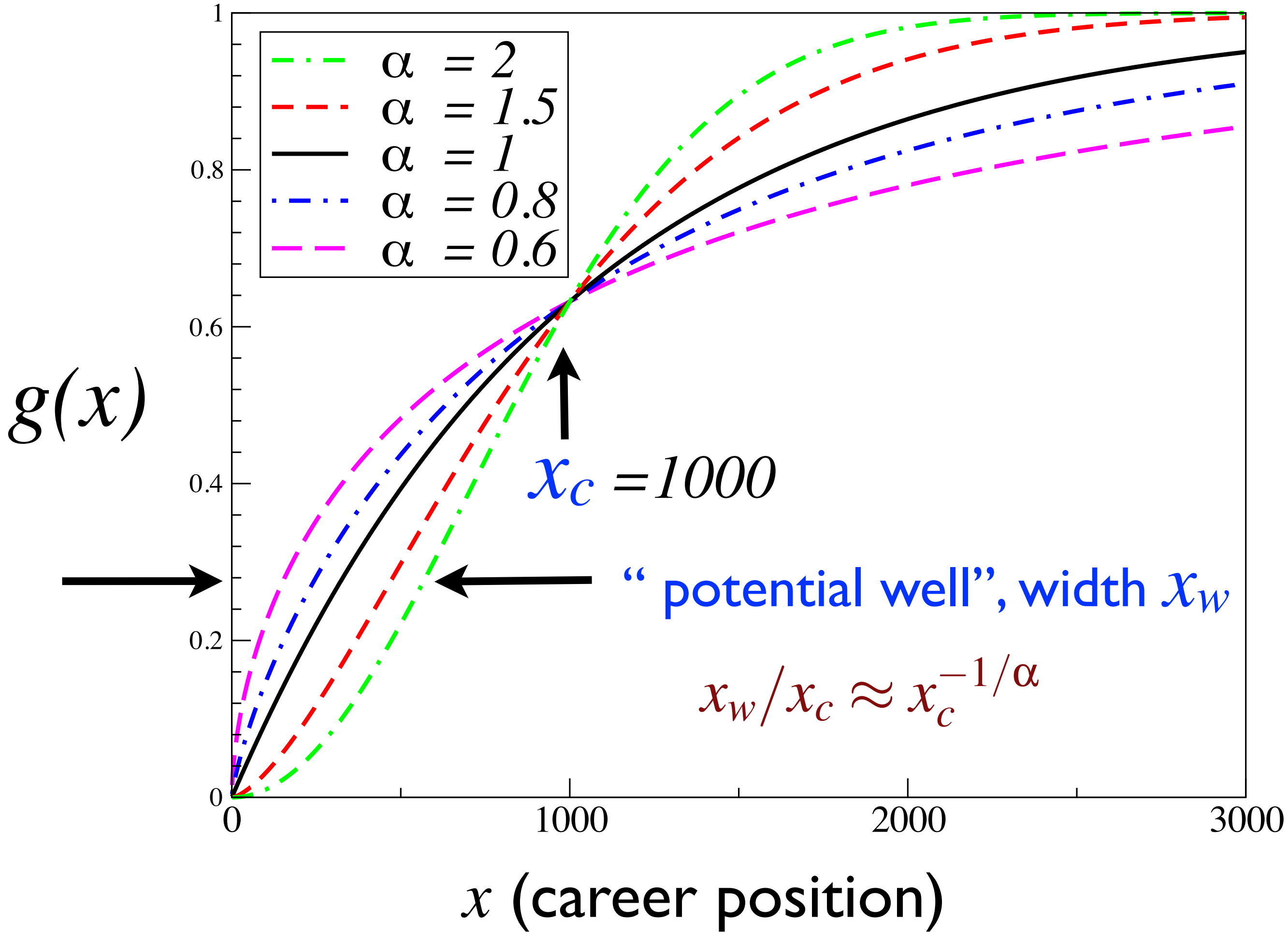
$P(x\,,\,t)$ = probability that career is at position x at time t

Master Equation approach → Poisson Distribution

$$P(x,t) = \frac{e^{-\lambda t}(\lambda t)^{x-1}}{(x-1)!}$$

$$\lambda \equiv g(x)$$

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



- x_c ≡ career position time-scale which separates veterans from newcomers.
- α ≡ quantifies the rate at which an individual climbs the "career ladder" : $g(x) \sim x^{\alpha}$ *for* $x \ll x_c$

Ingredient II: Random Termination Time

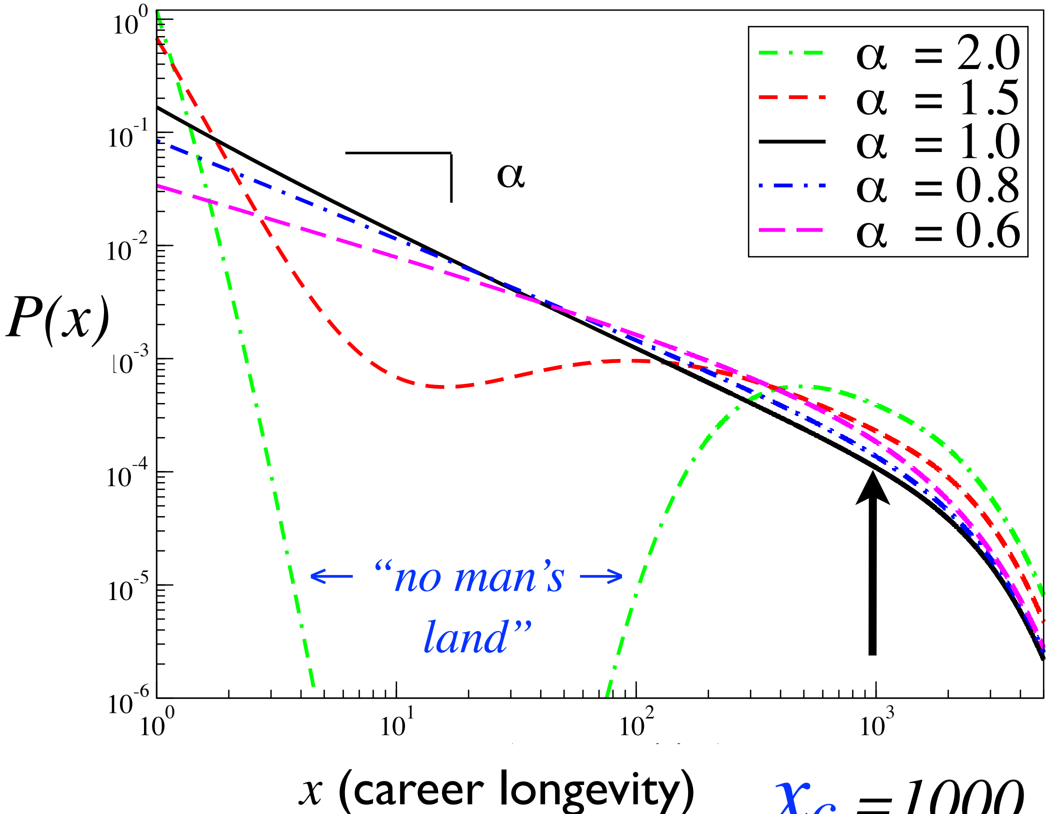
- Termination of career can occur for many reasons:
  - decreased work performance
  - economic downturn
  - health and other biological factors
  - forced retirement / change of profession
- Hence, the career position at termination time T = career longevity. Hazard rate $H(T)$: conditional probability that termination will occur at time $(T + \delta T)$ given that termination has not yet occurred at time T

$$H(T) = \frac{r(T)}{S(T)} = -\frac{\partial}{\partial T} \ln S(T) \qquad S(T) = 1 - \int_0^T r(t)dt$$

- So we choose a constant hazards model, $H(T) = 1/x_c$ which corresponds to a pdf of termination times:

$$r(T) = \exp[-T/x_c]/x_c$$

Progress rate $g(x)$ + Career Longevity *pdf* $P(x)$



x_c =1000

$$P(x) = \frac{g(x)^{x-1}}{x_c[\frac{1}{x_c} + g(x)]^x} \approx \frac{1}{g(x)x_c} e^{-\frac{x}{g(x)x_c}}$$

for convex $\alpha > 1$:

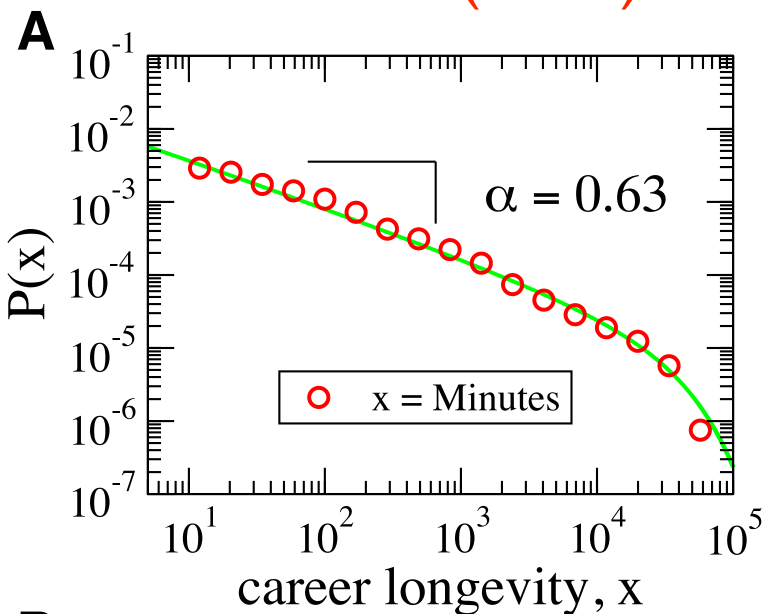
Bimodal

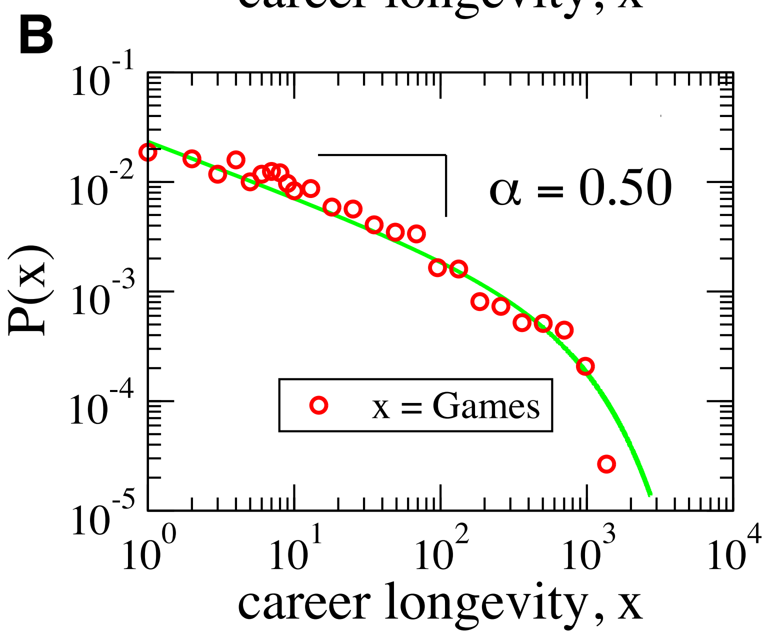
for concave $\alpha < 1$:

$$P(x) \propto \begin{cases} x^{-\alpha} & x < x_c \\ e^{-(x/x_c)} & x > x_c \end{cases}$$

- α ≡ power-law exponent for career longevity, which is intrinsically related to the rate at which individuals establish their reputation and secure future opportunity based on prior success.

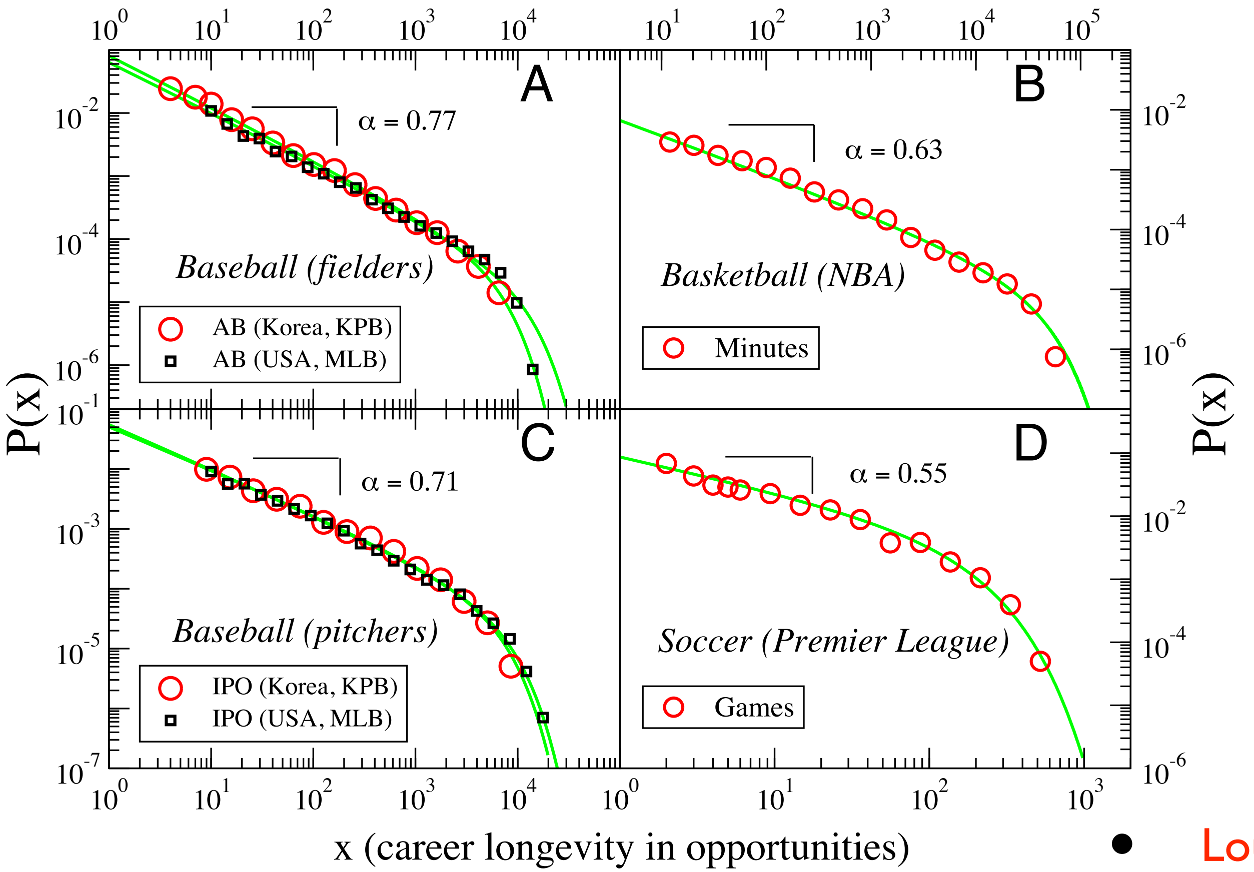
Empirical Results: Career longevity in professional basketball (NBA)





- Analyzed 2700+ completed careers over the 59-yr period 1946-2004
- $x \equiv$ *career longevity* (e.g. min. or games played)
- $P(x)$: probability density function (pdf) of career longevity x
- $P(x)$ is truncated power-law:
  - scaling exponent $\alpha \lesssim 1$
  - Exponential cutoff x_c: Finite-lifetime
  - Scale Free behavior: $P(x_1)/P(x_2) \cong (x_2/x_1)^{\alpha}$ *for* $x < x_c$
- 3% of players played between 1-12 minutes in their entire career! However, the average career length is approx. <x> = 6,500 min., Max(x) = 57,446 min. (Kareem A.-Jabbar)
- 2% of players played in only 1 game in their entire career! <x> = 273 games ~ 3 seasons, Max(x) = 1,611 games (R. Parish)

Career Longevity in 4 sports leagues



opportunities ~ time duration

Major League Baseball

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

``One-hit wonders''

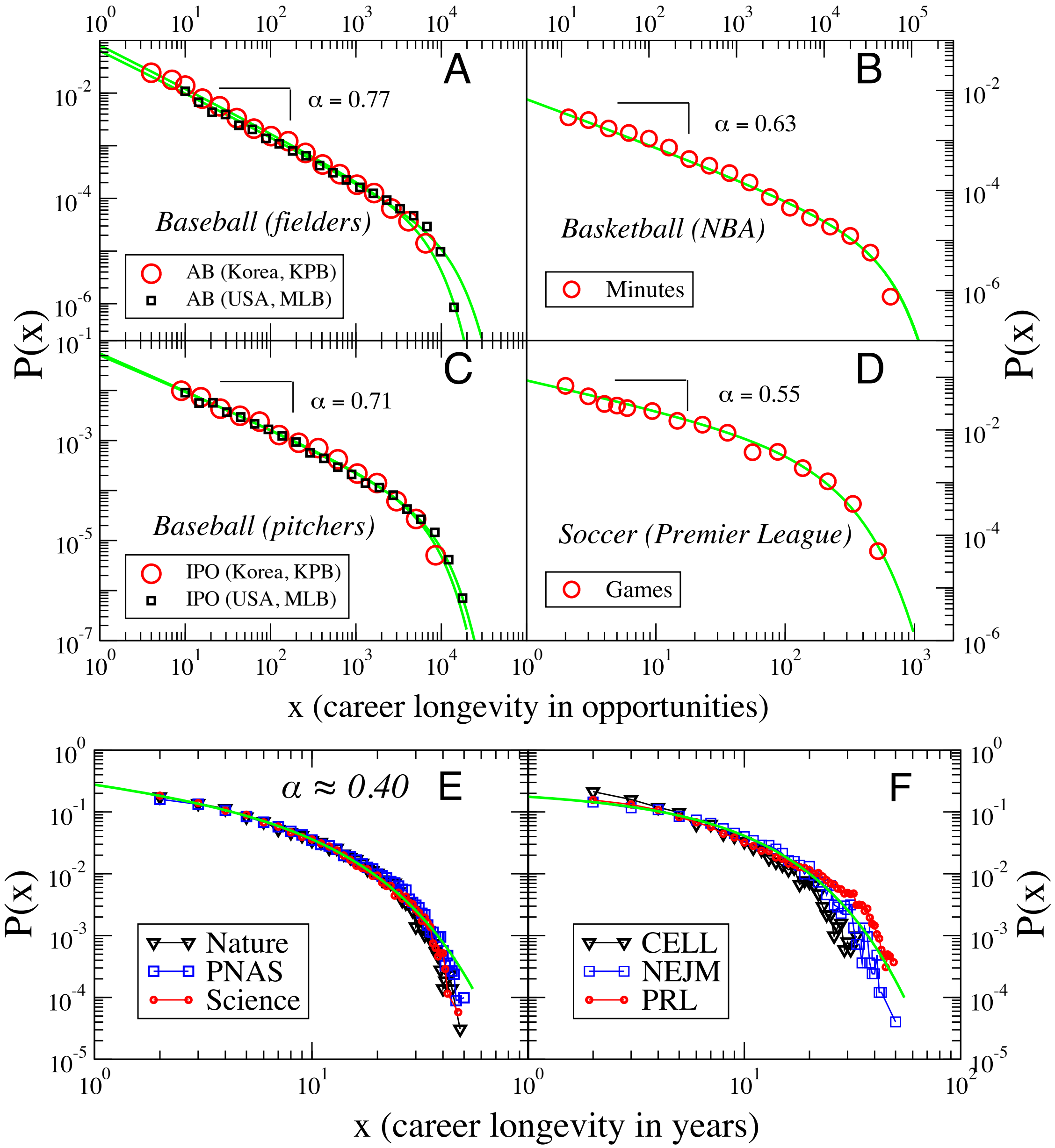
- 3% of all fielders finish their career with ONE at-bat!
- 3% of all pitchers finish their career with less than one inning pitched!

``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

New statistical laws for success!

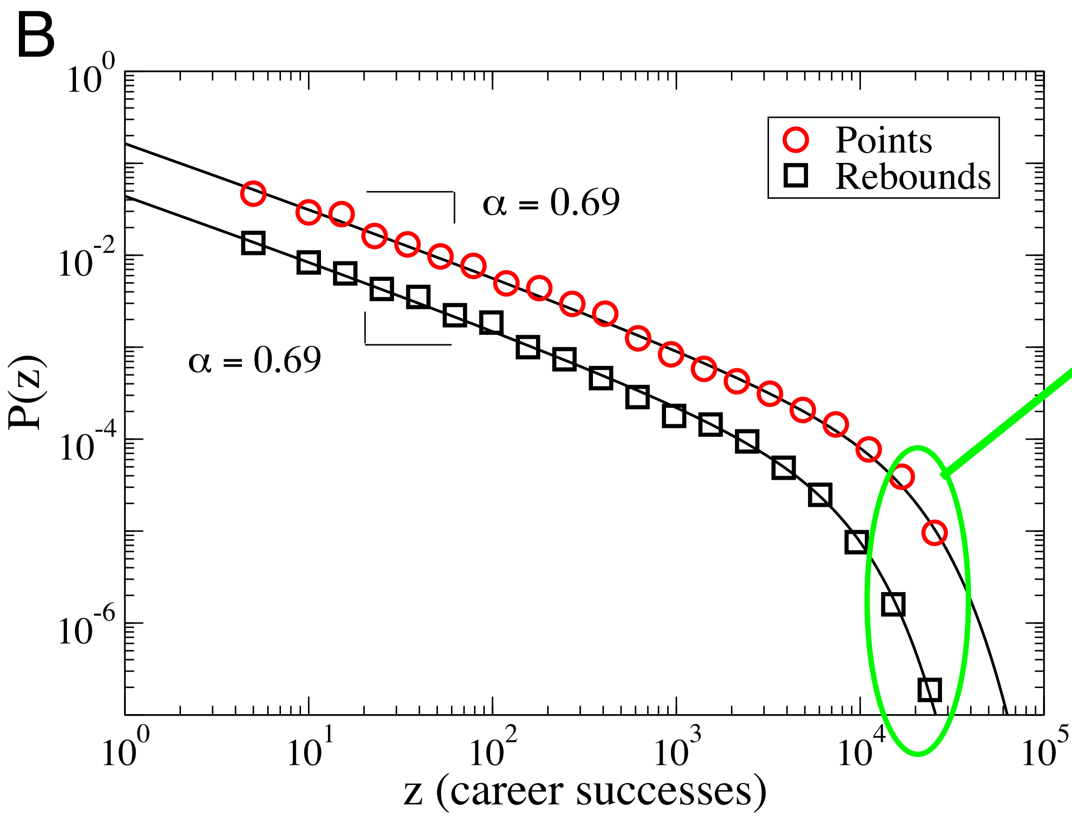

시사코리아저널
Sisa Korea Journal
경제
POSTECH-美보스턴대 공동연구팀, 성공의법칙 물리학으로 풀었다!!



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

Implications of longevity on career success

American Basketball (NBA + ABA): 1946-2004



Wilt
Chamberlain !

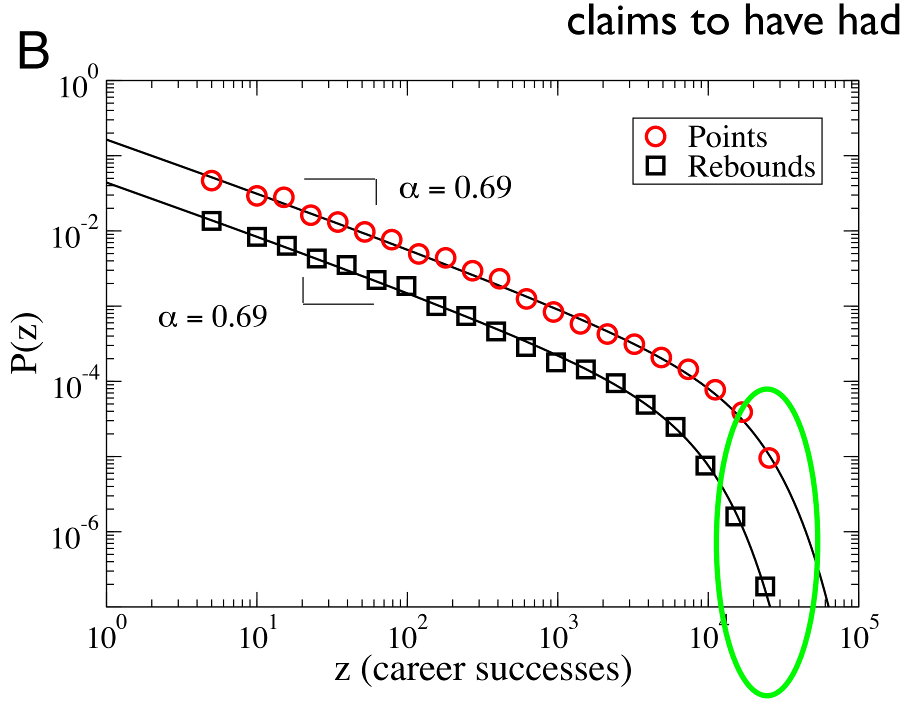
31,419 Points (#4 all-time)
23,924 Rebounds (#1)

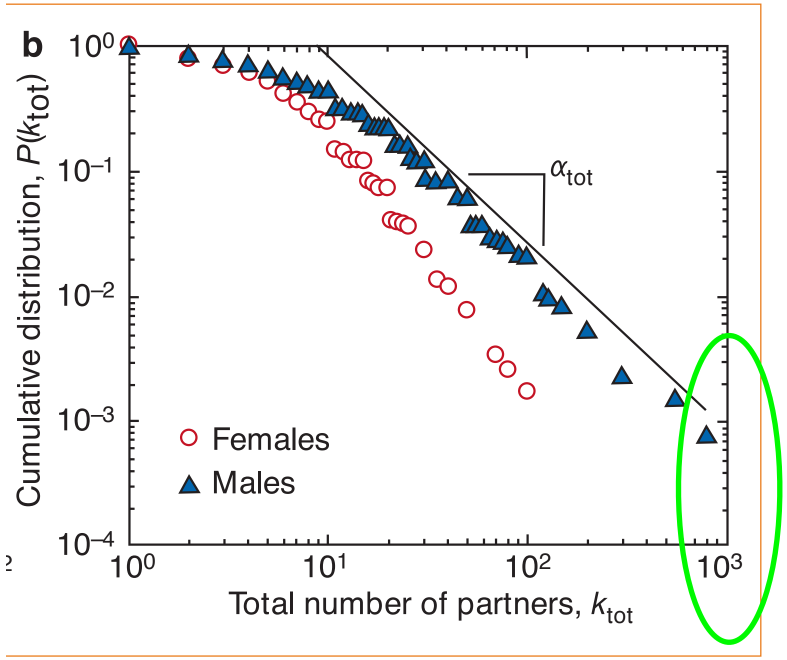
z = career success total $\propto$ career longevity

- Career longevity exponents α carry over naturally into career statistics

Right-skewed phenomena in the social sciences

Wilt
Chamberlain:
claims to have had over 20,000+ sexual partners....





F. Liljeros, et al.,"*The web of human sexual contacts,*" *Nature* **411**, 907 (2001)

Superstars play an important role in society, as reputation hubs, as social trademarks used for economic gain, and as cultural role models

Significant implications for theoretical models of disease propogation, quarantine, and treatment!

"superstars" are not outliers, but are predicted and consistent with empirical heavy tailed distributions

3. Empirical Facts

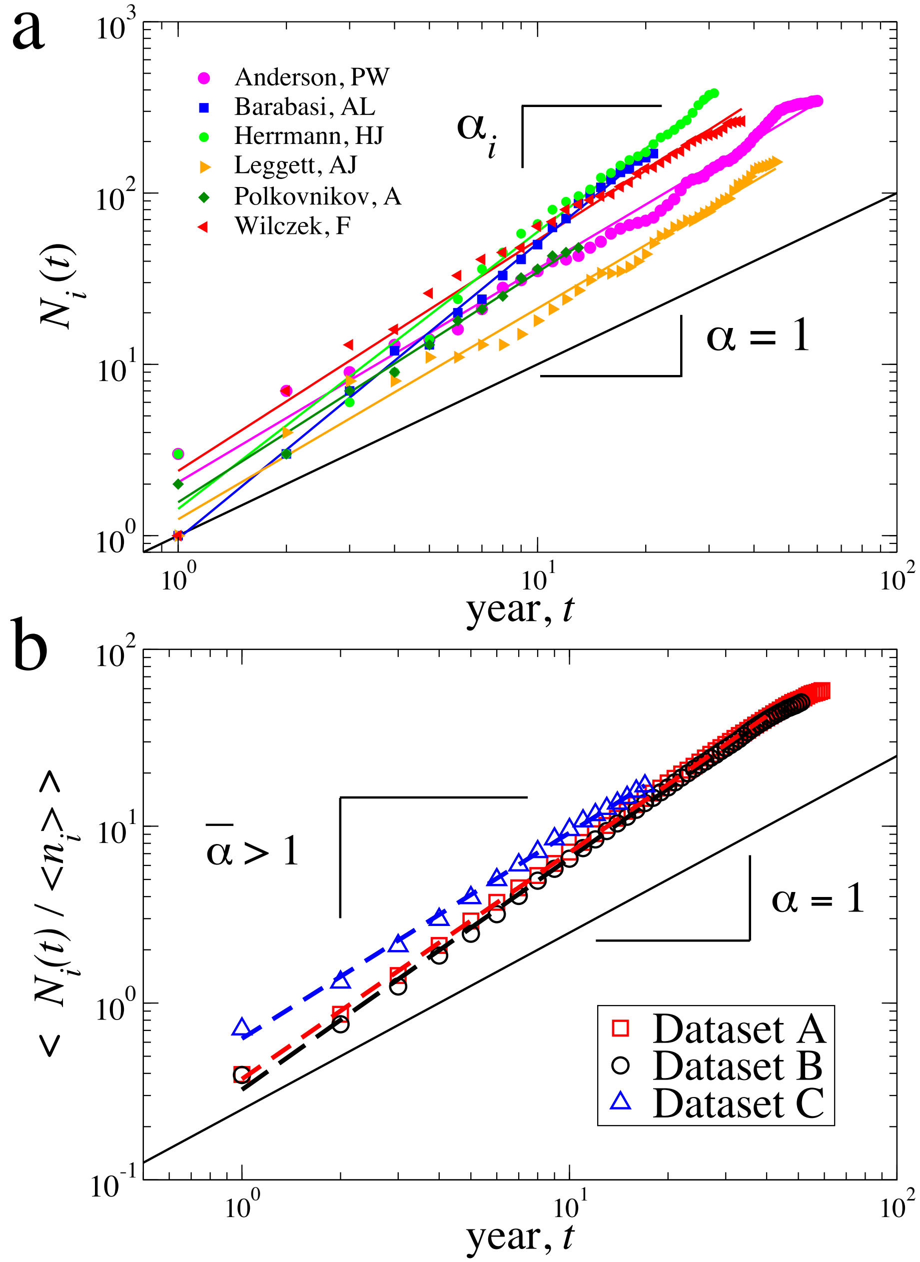
고집 and 위험

Persistency and Uncertainty
in the career trajectory

"publish or perish"

The academic career trajectory: a tale of knowledge and reputation spillovers

다른 일 or 작은 폭포



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

A proxy for the "cumulative" reputation

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

$$\approx A_i t^{\alpha_i}$$

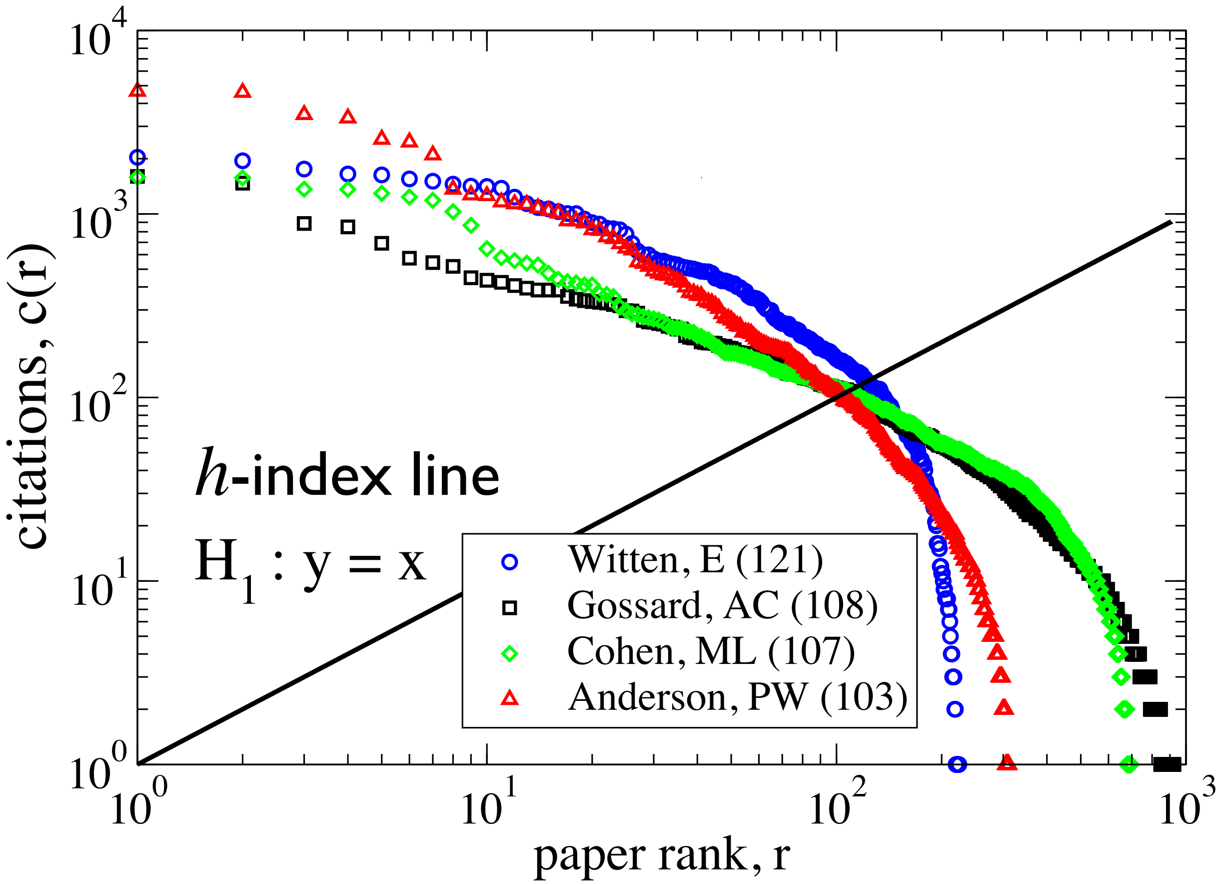
[A] 100 "top" physicists ($\bar{\alpha}$ = 1.28 ± 0.01)

[B] 100 (prolific) control physicists ($\bar{\alpha}$ = 1.31 ± 0.01)

[C] 100 asst. professors (physics) ($\bar{\alpha}$ = 1.15 ± 0.02)

As a result of knowledge spillovers (both in time and (collaboration) space) successful careers become "attractors"

Publication impact: the rank-citation profile $c_i(r)$

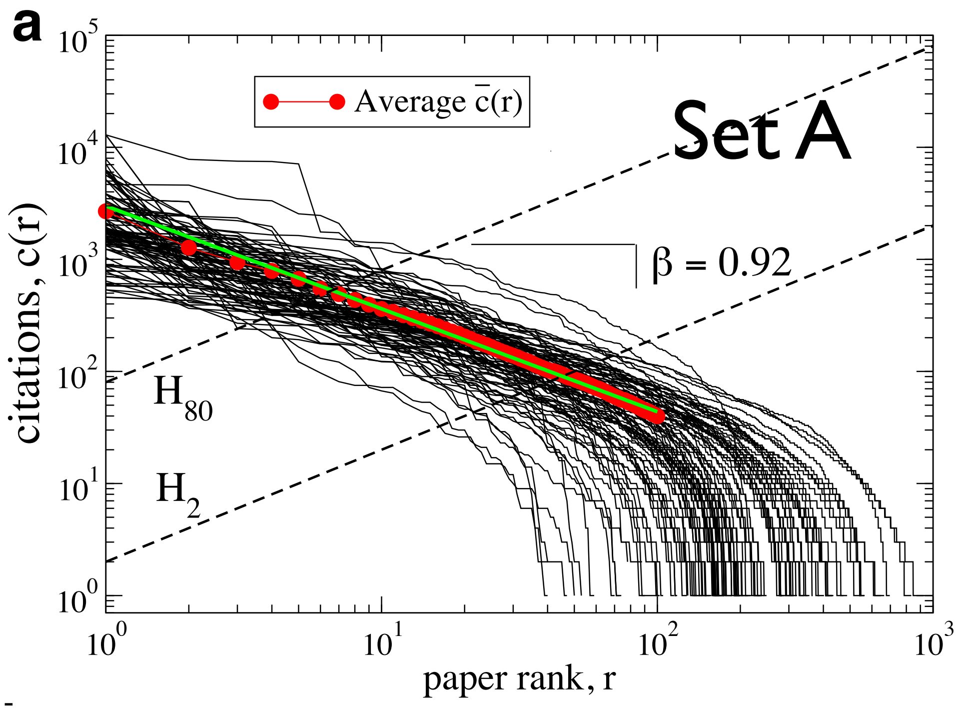


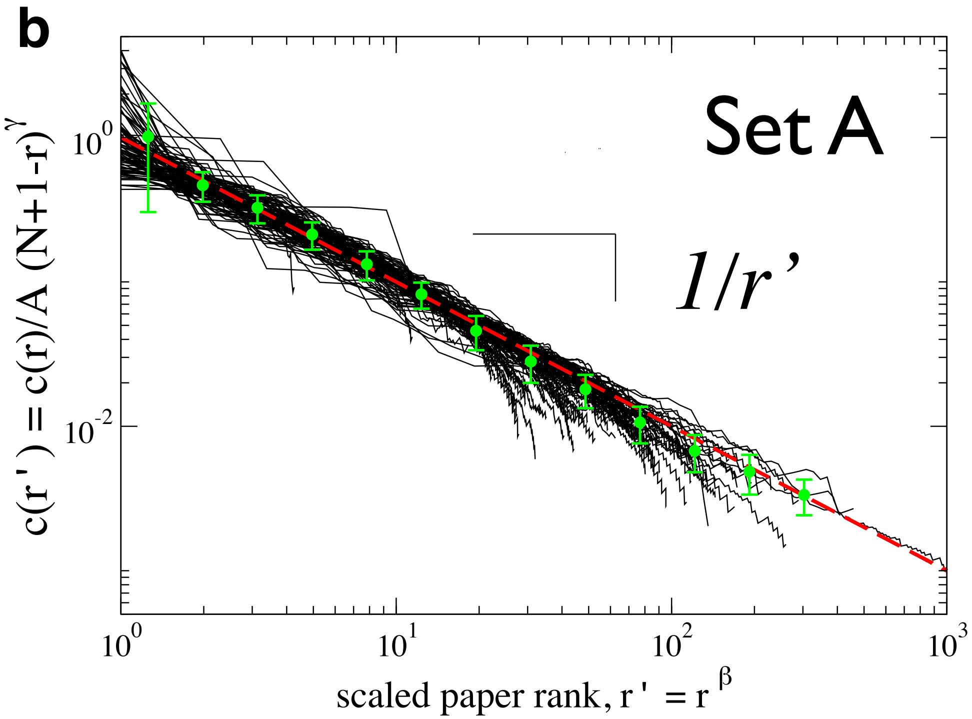
$c_i(r)$ is the rank-ordered (Zipf) citation distribution of the N papers published by individual i in his/her entire career

Reputation spillovers also contribute to increasing returns. Interestingly, even the top scientists have a significant number of papers that go relatively un-cited.

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

A comparison of $c_i(r)$ the top-100 "champions" of PRL (Set A) with average h-index $\langle h \rangle = 61 \pm 21$





Discrete Generalized Beta Distribution(DGBD):

$$c(r) \equiv Ar^{-\beta}(N+1-r)^{\gamma} \; .$$

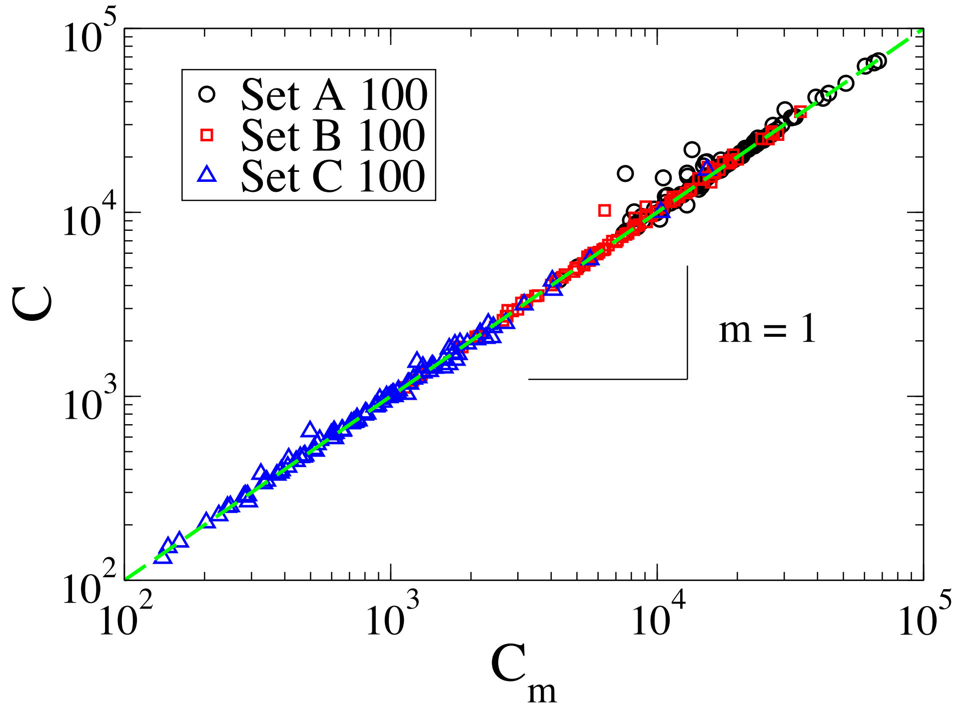
Martinez-Mekler, et al. "Universality of rank-ordering distributions in the arts and sciences." PLoS ONE 4: e4791 (2009).

Average values of the DGBD model parameters:

$$\langle\beta\rangle = 0.83 \pm 0.23 \quad \text{and} \quad \langle\gamma\rangle = 0.67 \pm 0.19$$

Further validation of the DGBD model, comparing the *predicted* and *actual* total number of citations, C_i

$$C_i = \sum_{r=1}^{r=N} c_i(r)$$



$$C_{i,m} = \sum_{r=1}^{r=N} A r^{-\beta} (N+1-r)^{\gamma}$$

Scaling relation between C, h, and β

$$C_{\beta,h} \sim h^{1+\beta}$$

$$\beta \cong 1$$

* $C \approx 4h^2$

* S. Redner, "On the meaning of the h-index." J. Stat. Mech. 2010, L03005 (2010).

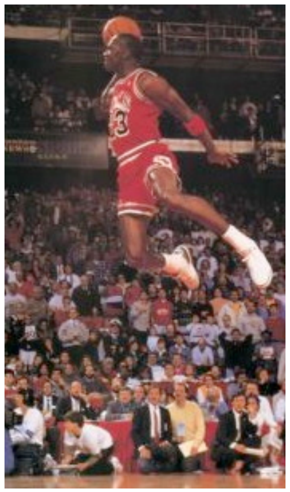
4. Preferential Capture Model

Competition for limited resources

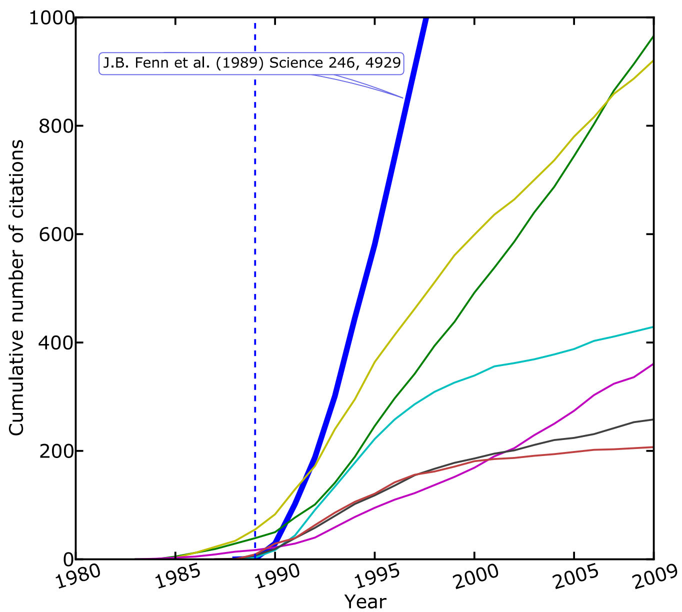
- $$$ (funding, salary share)
- Fame / Recognition
- production opportunities (publication)

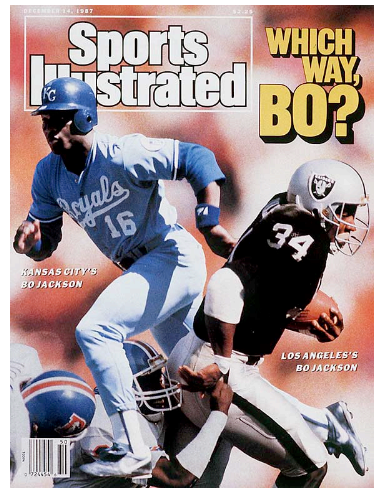
Understanding the role of "career shocks"

Persistence and Uncertainty



scientific "career shocks" following big discovery can trigger future recognition and reward, resembling the cascading earthquake dynamics

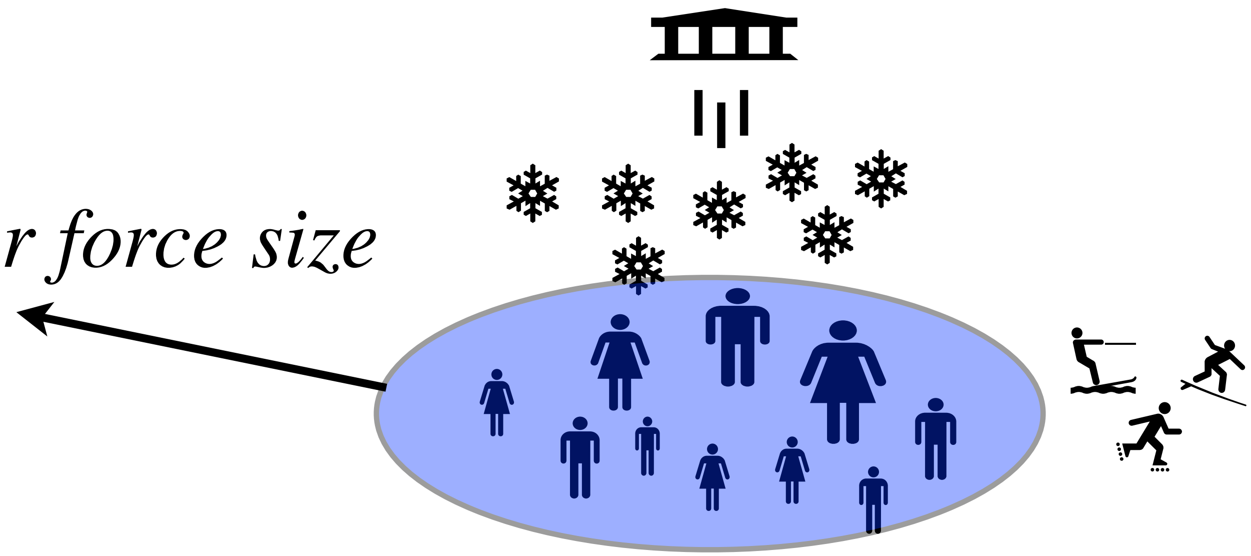




Mazloumian A, Eom Y-H, Helbing D, Lozano S, Fortunato S (2011) How Citation Boosts Promote Scientific Paradigm Shifts and Nobel Prizes. PLoS ONE 6(5): e18975.

Production opportunities falling down from heaven

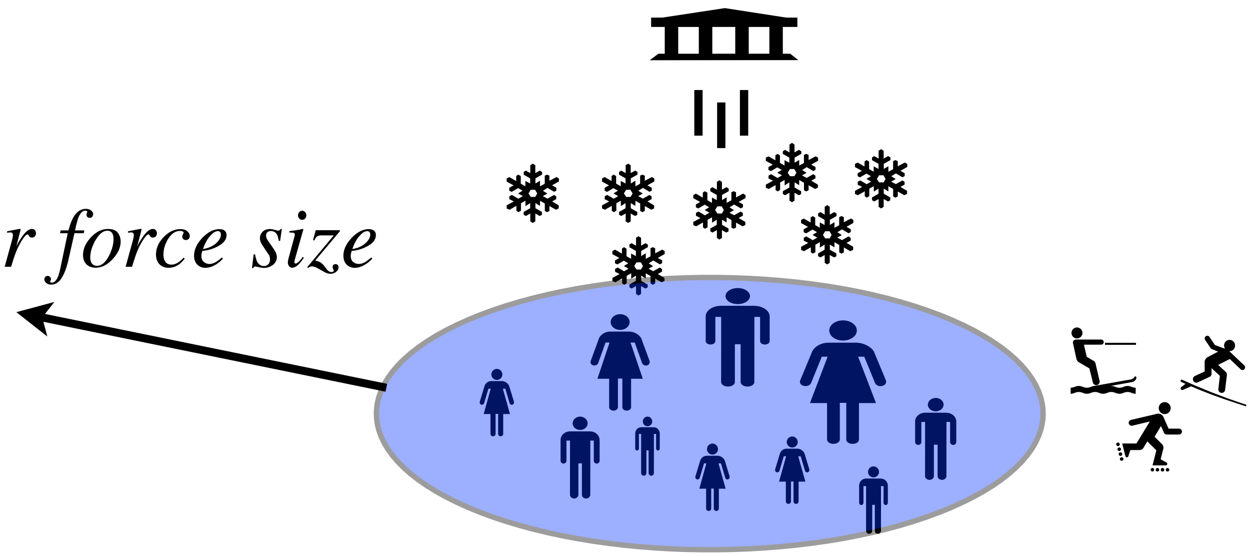
I = *labor force size*



A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli.
"'Persistence and Uncertainty in the Academic Career."
Under review.

Monte Carlo Model

I = labor force size



1) The system consists of $I \equiv 1000$ agents competing for P opportunities that are allocated in a single period. There is no entry, hence the number I is kept constant. Also, P is also kept constant, so there is no growth in the labor supply.

2) We run the Monte Carlo (MC) simulation for $T \equiv 100$ time periods and all agents are by construction from the same age cohort (born at same time).

3) Each time period corresponds to the allocation of $P \equiv \sum_{i=1}^{I} n_{0,i}$ opportunities, sequentially one at a time, to randomly assigned agents i, where $n_{0,i} \equiv 1$ is the potential production capacity of a given individual.

4) The assignment of a given opportunity is proportional to the time-dependent weight (capture rate) $w_i(t)$ of each agent. Hence, the assignment of 1 opportunity to agent i at period t results in the production (achievement) $n_i(t)$ to increase by one unit: $n_i(t) \rightarrow n_i(t) + 1$. In the next time period $t+1$, we update the weight $w_i(t+1)$ to include the performance $n_i(t)$ in the current period.

A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli. "'Persistence and Uncertainty in the Academic Career." Under review.

Preferential Capture Model

1) In each Monte Carlo step we allocate one opportunity to a randomly chosen individual i so that $n_i(t) \to n_i(t)+1$

2) The individual i is chosen with probability $\mathcal{P}_i(t)$ proportional to $[w_i(t)]^\pi$

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

where the value $w_i(t)$ is given by an exponentially weighted sum over the entire achievement history

$$w_i(t) \equiv \sum_{\Delta t=1}^{t-1} n_i(t-\Delta t) e^{-c\Delta t} \ .$$

*(appraisal = 값 매김)

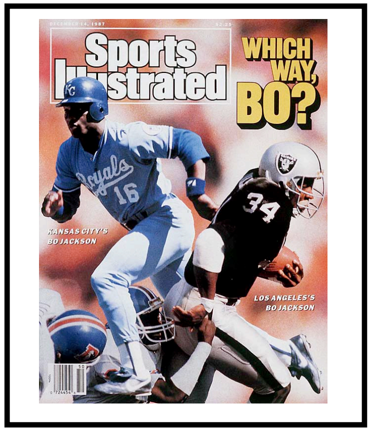
The details of the appraisal determines how much the past record of accomplishment determines the ability to capture new opportunities

Reputation is cumulative → what you do as a grad student matters!

Numerous Nobel Prizes awarded for work primarily done during the PhD

If the appraisal timescale is too short, than a Nobel worthy phd thesis loses it's weight overnight ! In sports this is exemplified.

→ This is why “sudden death” careers occur in sports so frequently!



Preferential Capture Model

1) In each Monte Carlo step we allocate one opportunity to a randomly chosen individual i so that $n_i(t) \rightarrow n_i(t)+1$

2) The individual i is chosen with probability $\mathcal{P}_i(t)$ proportional to $[w_i(t)]^\pi$

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

where the value $w_i(t)$ is given by an exponentially weighted sum over the entire achievement history

$$w_i(t) \equiv \sum_{\Delta t=1}^{t-1} n_i(t-\Delta t) e^{-c\Delta t} \ .$$

The parameter $c \geq 0$ is a memory parameter which determines how the record of accomplishments in the past affect the ability to obtain new opportunities in the current period, and therefore, the future. The limit $c = 0$ rewards long-term accomplishment by equally weighting the entire history of accomplishments. Conversely, when $c \gg 1$ the value of $w_i(t)$ is largely dominated by the performance $n_i(t-1)$ in the previous period, corresponding to increased emphasis on short-term accomplishment in the immediate past. Intermediate values $0 < c < 1$ weight more equally the immediate past and the entire history of accomplishment.

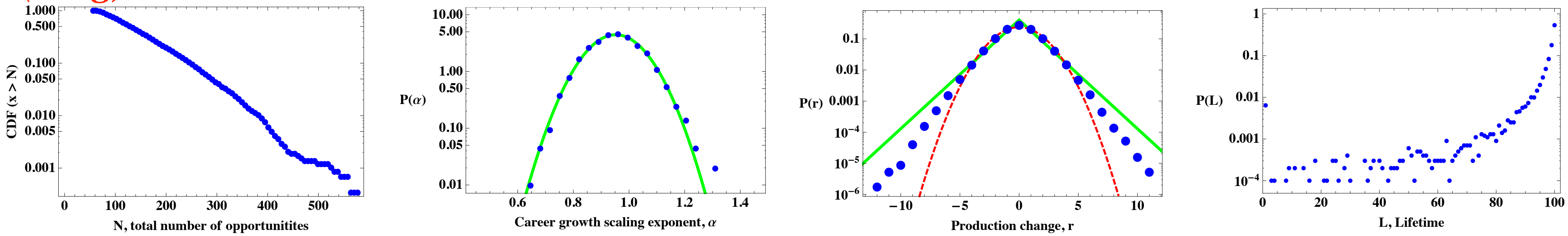
*(apprasal = 값 매김)

$c \rightarrow 0$: appraisal over the lifetime of achievements (~ tenure system)

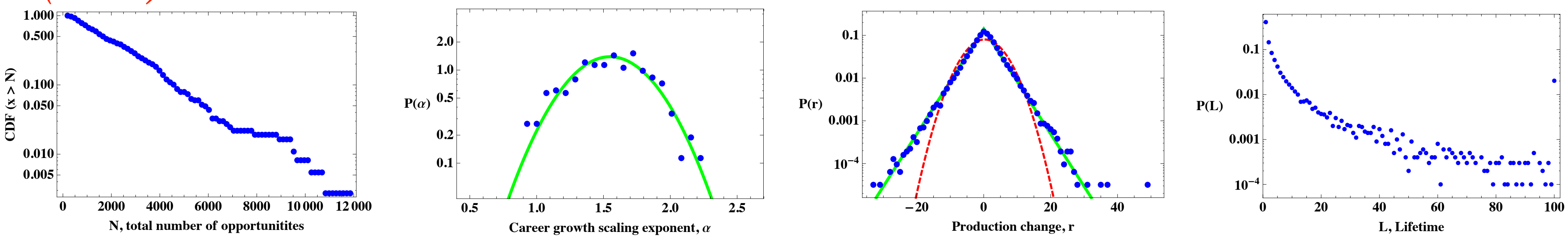
$c > 1$: appraisal over only recent achievements (short-term contract system)

Results for the linear ($\pi = 1$) preferential capture model

c = 0 (long) *N(total)* α *Production change, r* *Longevity, L*



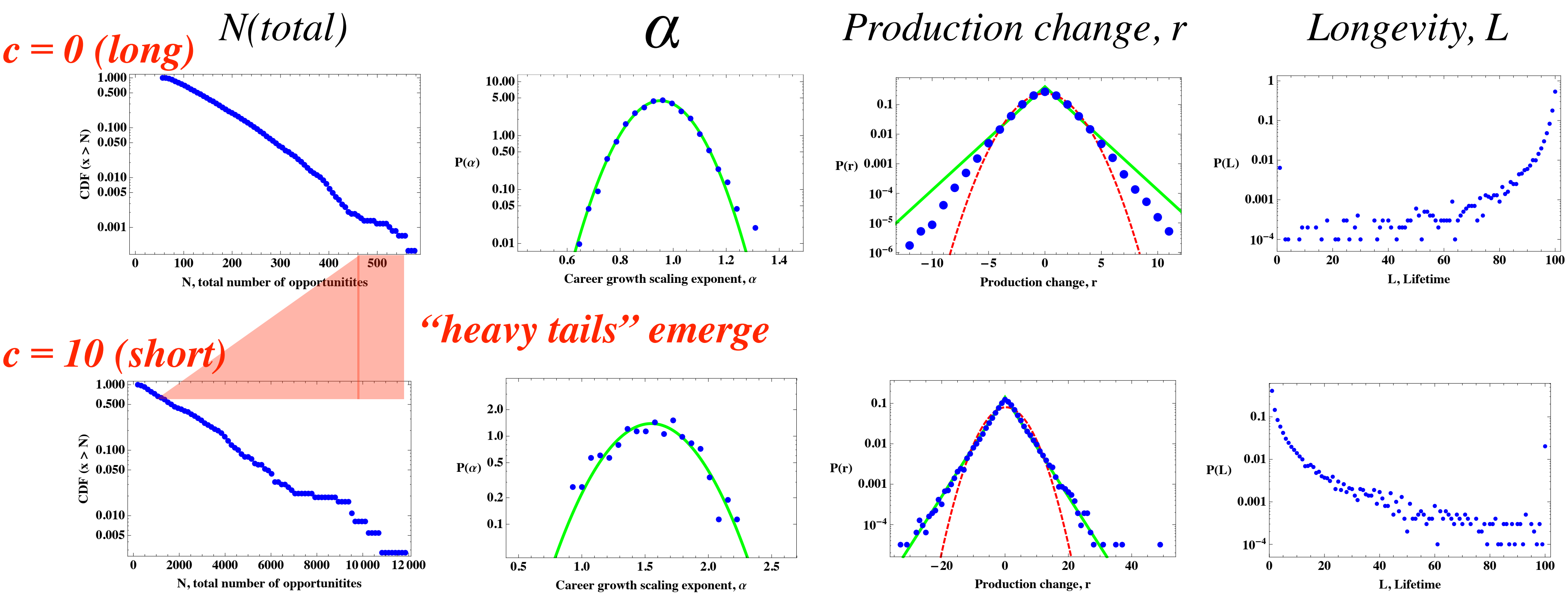
c = 10 (short)



c $\rightarrow$ 0 : appraisal over the lifetime of achievements (~ tenure system): career is better protected from intrinsic negative production shocks (periods of lull) and as a result, most careers sustain production throughout the career, reaching the maximum career lifespan T.

$c > 1$: appraisal over only recent achievements (short-term contract system): most individuals experience "sudden death" termination relatively early in the career. Meanwhile, a small number of "kings" survive the initial selection process, which is governed primarily by random chance, and dominate the system.

“Sudden death” careers and “dragon kings”

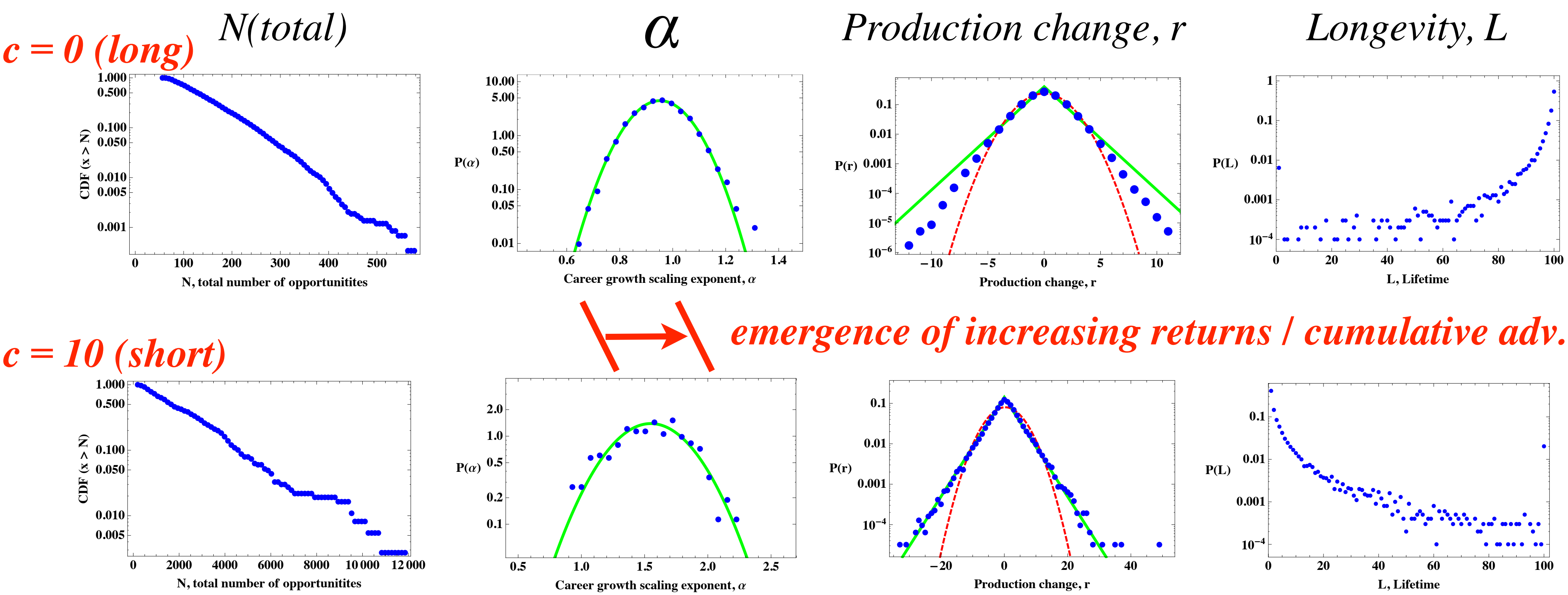


c → 0 : appraisal over the lifetime of achievements (~ tenure system): career is better protected from intrinsic negative production shocks (periods of lull) and as a result, most careers sustain production throughout the career, reaching the maximum career lifespan T.

$c > 1$: appraisal over only recent achievements (short-term contract system): most individuals experience “sudden death” termination relatively early in the career. Meanwhile, a small number of “kings” survive the initial selection process, which is governed primarily by random chance, and dominate the system.

“Sudden death” careers and “dragon kings”

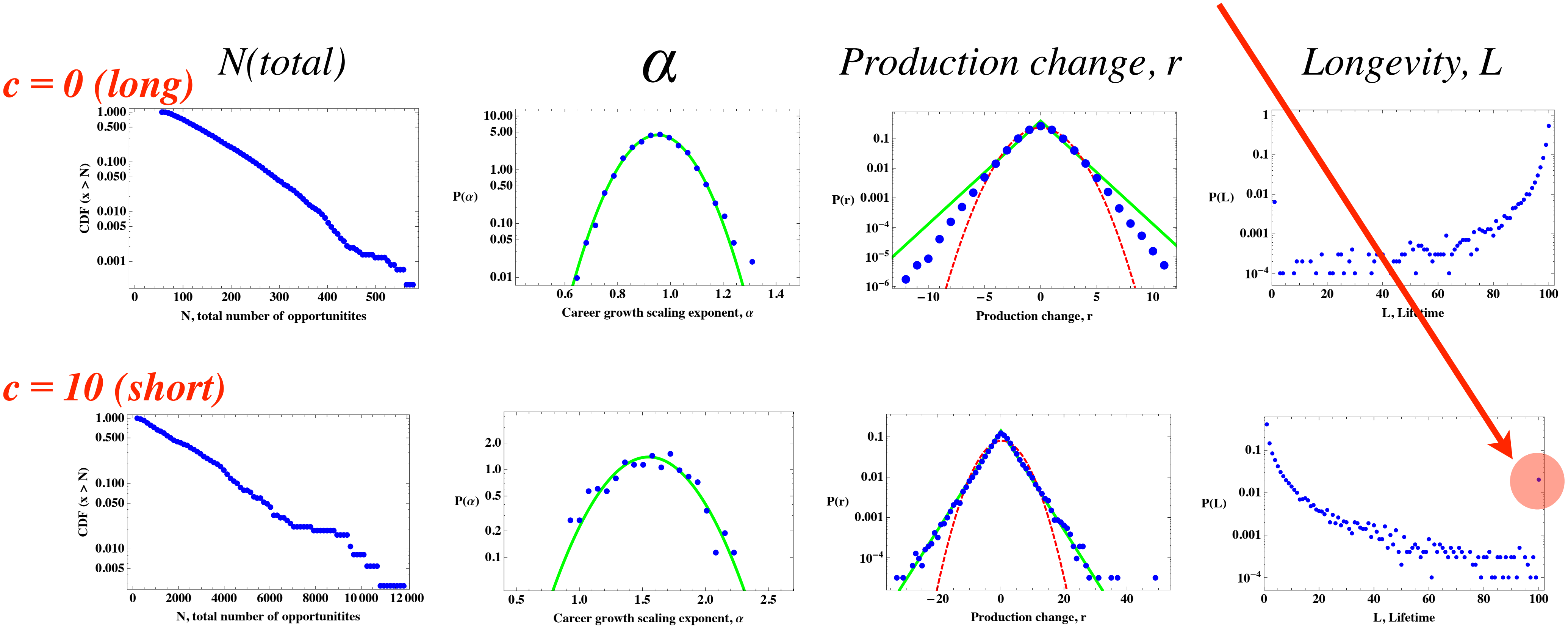
Distributions of 4 career measures:



c → 0 : appraisal over the lifetime of achievements (~ tenure system): career is better protected from intrinsic negative production shocks (periods of lull) and as a result, most careers sustain production throughout the career, reaching the maximum career lifespan T.

$c > 1$: appraisal over only recent achievements (short-term contract system): most individuals experience “sudden death” termination relatively early in the career. Meanwhile, a small number of “kings” survive the initial selection process, which is governed primarily by random chance, and dominate the system.

“Sudden death” careers and “dragon kings”

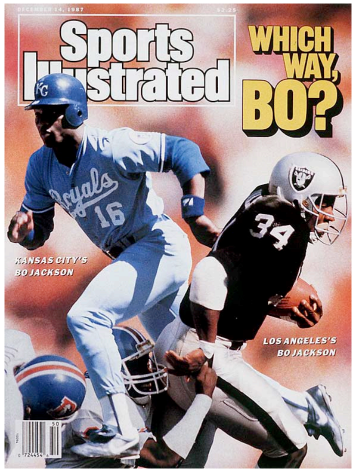


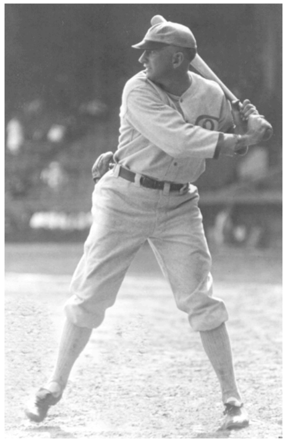
$c \to 0$: appraisal over the lifetime of achievements (~ tenure system): career is better protected from intrinsic negative production shocks (periods of lull) and as a result, most careers sustain production throughout the career, reaching the maximum career lifespan T.

$c > 1$: appraisal over only recent achievements (short-term contract system): most individuals experience “sudden death” termination relatively early in the career. Meanwhile, a small number of “kings” survive the initial selection process, which is governed primarily by random chance, and dominate the system.

“Sudden death” careers: associated with the unforeseeable future and the risks of stagnant production

Our model shows that high competition levels can make careers vulnerable to termination relatively early in the career as a result of negative production fluctuations and not necessarily due to lack of individual persistence. Many professions are marked by competitive features that can stunt the growth of inexperienced individuals and can lead to early career termination.





Sudden career termination in science due to ethical scandals

Jan Hendrik Schön Scandal (2001)

On October 31, 2002, *Science* withdrew eight papers written by Schön

On December 20, 2002, *Physical Review* withdrew six papers

On March 5, 2003, *Nature* withdrew seven papers

Diederik Alexander Stapel Scandal (2011)

Social psychologist (ironically) made up data for at least 30 publications according to preliminary investigation, which is still ongoing.

Take home messages

- There are many analogies between the superstars in science and the superstars in professional sports, possibly arising from the generic aspects of competition.
- There is a beautiful statistical regularity that "bridges the gap" between the relatively short careers and the extremely long "stellar" careers.
  - Stellar careers are not an anomaly! They are predicted by pdf $P(x)$
- The Matthew "rich-get-richer" effect can be used to explain the extremely right-skewed probability distributions that quantify both longevity and success.
  - evidence in the decreasing time duration $\tau(n)$ between publications and a model that predicts two classes of $P(x)$ depending on the choice of $g(x)$ There are many analogies between the superstars in science and the superstars in professional sports, possibly arising from the generic aspects of competition.
- Knowledge and collaboration spillovers result in increasing returns within the scientific career trajectory
- An institutional setting that neglects the specific features of academic career trajectories (increasing returns from knowledge spillovers and cumulative advantage, collaboration factors, career uncertainty) may inadvertently expose temporarily "cold" careers, leaving them out to freeze.
- A quantitative picture of Science is emerging, but there are still many open questions!

Thank You!

A special thanks to my collaborators:

Woo-Sung Jung, Orion Penner, Gene Stanley, Sauro Succi, Fengzhong Wang, Jae-Sook Yang, Massimo Riccaboni and Fabio Pammolli

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

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

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

III) A. M. Petersen, W.-S. Jung, H. E. Stanley, "On the distribution of career longevity and the evolution of home run prowess in professional baseball." Europhysics Letters 83, 50010 (2008).

IV) A. M. Petersen, O. Penner, H. E. Stanley , "'Methods for detrending success metrics to account for inflationary and deflationary factors." Eur. Phys. J. B 79, 67-78 (2011).

V) A. M. Petersen, O. Penner. "A method for the unbiased comparison of MLB and NBA career statistics across era." Submitted to the MIT Sloan Sports Analytics Conference, 2012.

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

VII) A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli "'Persistence and Uncertainty in the Academic Career." Under review, PNAS.

5. Universal Growth Patterns

How similar are "career atoms" to the analogous "atoms" in other socio-economic systems?

General principles of interacting systems?

VIII) A. M. Petersen, J. Tenenbaum, S. Havlin, H. E. Stanley. "'Statistical laws governing the fluctuations in word use from word birth to word death." Submitted.

Growth dynamics in complex systems: growing evidence of common underlying principles

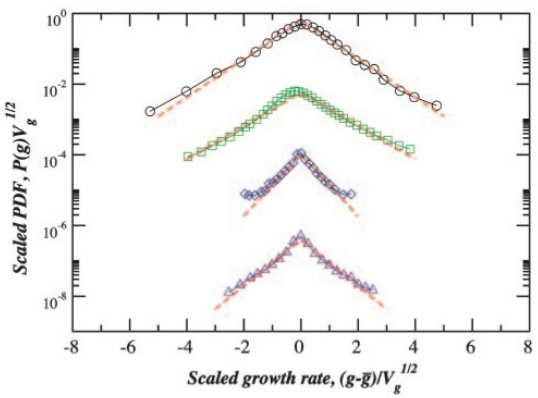
Q: How do complex systems grow ?

Q: How big are the rare events (often neglected by simple models) ?

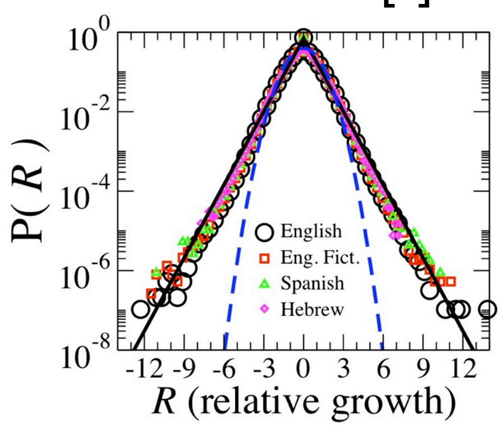
- Size *S(t)* at time *t*
- Growth rate R(t) = $g(t) \equiv \log\left(\frac{S(t+1)}{S(t)}\right) = \log S(t+1) - \log S(t)$
- Laplace PDF $P(R) = \frac{1}{\sqrt{2}\,\sigma(R)} e^{-(\sqrt{2}|R-\langle R\rangle|/\sigma(R))}$

Excess number of large growth (+/-) events as compared to the Gibrat multiplicative growth model which predicts a Gaussian distribution for P(R)

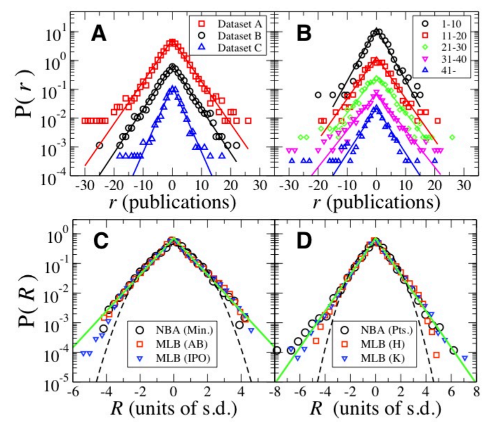
Firm size / Country GDP [1]



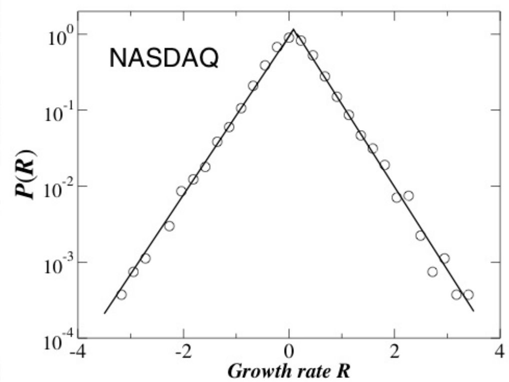
Word Use [2]



Individual Productivity [3]



Stock Price [4]



[1] D. Fu, et al., **The Growth of Business Firms: Theoretical Framework and Empirical Evidence.** PNAS **102**, 18801 (2005).
[2] A M. Petersen, et al., **Statistical Laws Governing Fluctuations in Word Use from Word Birth to Word Death**, submitted.
[3] A M. Petersen, et al., **Persistency and Uncertainty in the Academic Career**, in preparation.
[4] B. Podobnik, et al., **Common scaling behavior in finance and macroeconomics.** Eur. Phys. J. B **76**, 487 (2010).

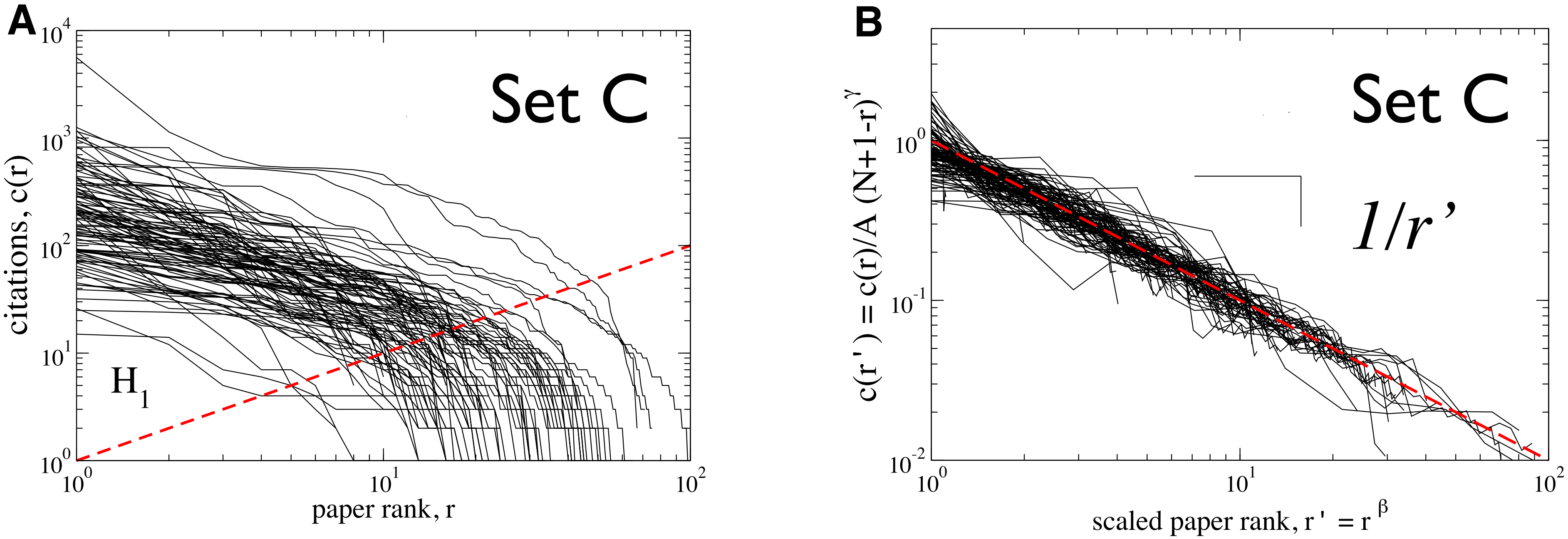
Quantifying Statistical Regularities in the Career Achievements of Scientists and Athletes

Abstract:

For many professions, the quantitative analysis of individual careers is made difficult by the lack of comprehensive data and the difficulty in defining measures for productivity and longevity. However, comprehensive career data is recorded in professional sports and is perfectly tailored for studying human productivity. Similarly, the publication careers of scientists are also quantifiable using similar measures. Since both professions are subject to the common forces of competition, one motivating question in this talk is: “What are the statistical regularities in career achievement across an entire cohort of competitors?”

In this talk I will discuss the statistical regularities that describe the everyday topic of career achievement using comprehensive career data. In the first part of the talk, I will discuss the topic of career longevity, using as example the 60+ year history of the National Basketball Association and 2700+ complete careers over the period 1946-2004. Surprisingly, we find that a common career longevity distribution describes the careers of 20,000+ athletes from 4 sports leagues and 400,000+ scientists from 6 high-impact journals, where each journal serves as a generic arena for competition. In order to account for the regularities we observe across several professions, we develop an exactly solvable model for career longevity based on the Matthew “rich-get-richer” effect. Our model is in excellent agreement with empirical career longevity distributions for each profession analyzed. Our model follows from two general assumptions: (i) that there is random forward progress in the career, whereby it becomes easier to make progress the further along one is in his/her career, and (ii) that career termination follows from random hazards that are present throughout the career. The findings suggests that there is a common underlying mechanism which underlies career development in competitive professions. In the second part of the talk, I will discuss the publication careers of 300 individual scientists (ranging from very the very famous to current Assistant professors) and find remarkable statistical regularity in the functional form of the rank-citation distribution (analogous to the Zipf rank-frequency distribution) for each scientist studied.

Common functional form also describes even Assistant Professors with average h-index $\langle h \rangle = 15 \pm 7$

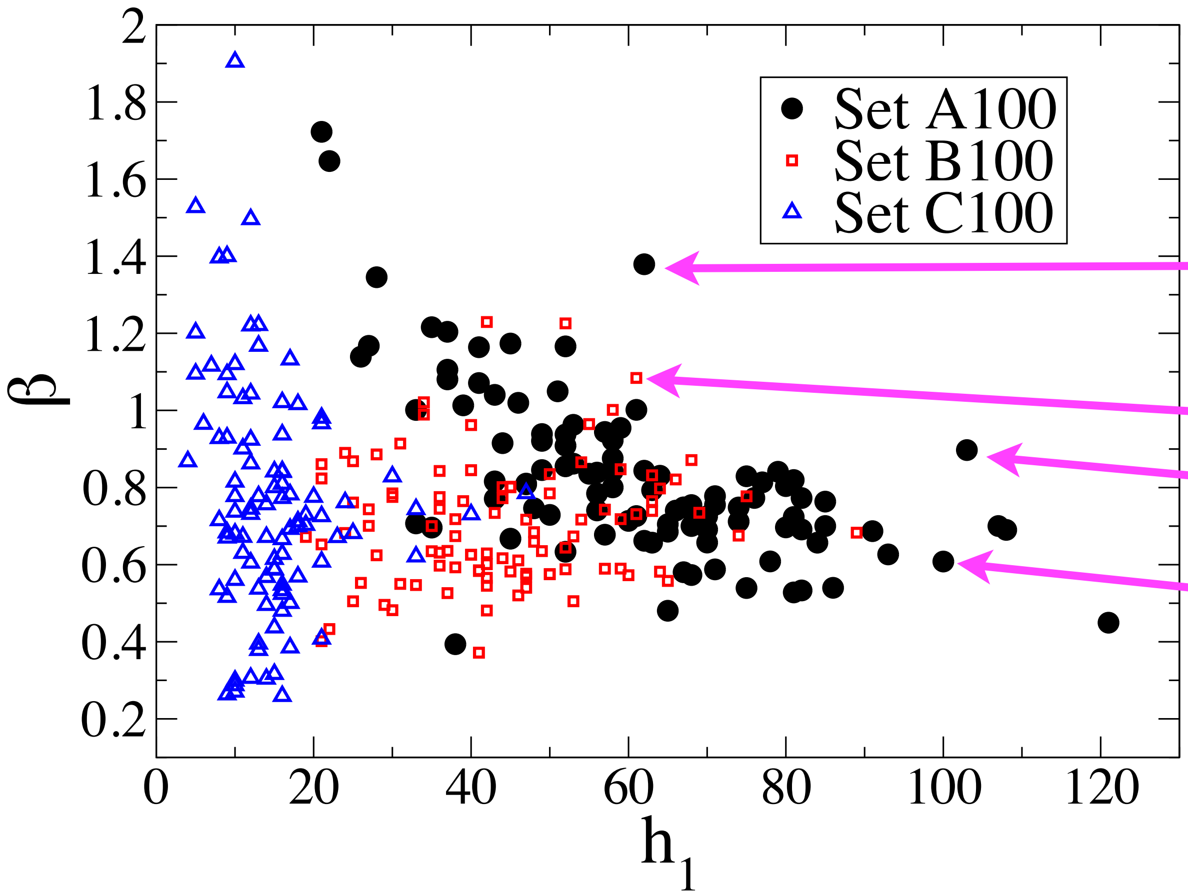


Set C: 100 Asst. Professors, 2 chosen from each of the top-50 U.S. physics departments

Average values of the DGBD model parameters:

$$\langle \beta \rangle = 0.79 \pm 0.38 \quad \text{and} \quad \langle \gamma \rangle = 0.89 \pm 0.36$$

The β -vs- h parameter space



Matthias Ernzerhof (U. Montreal)

c (1) = 16,314

"Generalized gradient approximation made simple"

Perdew, JP; Burke, K; Ernzerhof, M

PRL 1996

A.-L. Barabasi

P. W. Anderson

H. E. Stanley

For a given *h*, a large β value corresponds to a larger total citations,

$C_i \sim h^{1+\beta}$,

which is a proxy for career publication impact