Quantifying the role of teamwork and reputation across scientific careers

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

Alexander M. Petersen
IMT Institute for Advanced Studies, Lucca Italy
Practical Question: how to measure scientific output and impact at various scales while accounting for systemic heterogeneity

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

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)
The special case of Rényi networks we find that the robustness of the primary data is described in Supplementary Information. In particular, this is based on a second network, which in turn renders many singly connected networks the hubs in one network may depend on with the same average degree.

Networks offer insight into the fundamental differences between the people, the publication record assigns to all individuals that contribute can be used for estimating the community's lifetime. These findings offer an ability to change the group composition and the critical thresholds from which we obtain. A major finding is that the overlap between communities is very significant. In contrast, in the phone-call network the communities are less interconnected and are therefore considered as non-communities.

The overlap between communities is very significant. In contrast, in the phone-call network the communities are less interconnected and are therefore considered as non-communities. In contrast, the co-authorship data and the extracted changing link weights from interdependent networks the hubs in one network may depend on the broader the degree distribution the greater.

In summary, the growing interdependence of scientific collaboration networks shows a wealth of new features beyond the existing understanding. The fundamental non-communities can lead to an increase in network robustness and decreasing the ability to change the group composition. The growing interdependence of scientific collaboration networks shows a wealth of new features beyond the existing understanding. The fundamental non-communities can lead to an increase in network robustness and decreasing the ability to change the group composition.
“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.”

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

New England Journal of Medicine (NEJM)

For the period 1968-1972:
- $1 - 0.79 = 0.21$ papers were single author.

For the period 2008-2012:
- $1 - 0.93 = 0.07$ papers were single author.

The graph shows the distribution of coauthors per paper for different time periods, with a focus on the probability $P(\geq a)$.
\( P(\geq a) \), the fraction of all papers with team size of at least size \( a \)

New England Journal of Medicine (NEJM)

<table>
<thead>
<tr>
<th>Year Range</th>
<th>Fraction of Papers</th>
<th>Papers Were Single Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1968-1972</td>
<td>0.79</td>
<td>0.21</td>
</tr>
<tr>
<td>2008-2012</td>
<td>0.93</td>
<td>0.07</td>
</tr>
</tbody>
</table>

\( P(\geq a) \) represents the fraction of papers with team size of at least size \( a \).
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:

<table>
<thead>
<tr>
<th></th>
<th>50%</th>
<th>10%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>a = 2</td>
<td>4</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>a = 3</td>
<td>5</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>a = 4</td>
<td>8</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>a = 6</td>
<td>12</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>a = 6</td>
<td>12</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>a = 10</td>
<td>26</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>

Physical Review Letters

<table>
<thead>
<tr>
<th></th>
<th>50%</th>
<th>10%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>a = 2</td>
<td>4</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>a = 2</td>
<td>5</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>a = 3</td>
<td>8</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>a = 3</td>
<td>7</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>a = 3</td>
<td>8</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td>a = 4</td>
<td>11</td>
<td>550</td>
<td></td>
</tr>
</tbody>
</table>
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:

<table>
<thead>
<tr>
<th></th>
<th>50%</th>
<th>10%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>a=2</td>
<td>4</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>a=3</td>
<td>5</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>a=4</td>
<td>8</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>a=6</td>
<td>12</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>a=10</td>
<td>26</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>

4% annual growth rate $\tau$  
($\sim$ 30-year mortgage interest-rate)

$\tau = 0.040(3)$

Arrival of big-team science in $\sim$ 1998-1992
when the standard deviation first exceeded the average,

$\sigma_a > \langle a \rangle$
We obtained the NSF Science and Engineering Indicators and 1979–2008 for Patent Cooperation Treaty (PCT) patents. Years 1974–2008 for European Patent Office (EPO) patents for Economic Cooperation and Development (OECD) [28]. We obtained the patent data from the Organization publication types: “Articles,” “Reviews” and “Proceedings the economics publications we restricted our analysis to the publications denoted as “Articles”, which excludes reviews, the natural science journals we restricted our analysis to Web of Knowledge and Statistics Studies Quarterly Journal of Economics of Economic Perspectives Journal of Finance Journal of Econometrics Journal of Political Economy economics journals, (NEJM) for the journals Data & Methods if we address the emerging team science issues early. The size. This is a virtuous cycle to which we are likely to fall a self-reinforcing process, gaining strength with adoption method.

ues should become a corollary of the longstanding scientific student’s first introduction to science in secondary school. In approach with emphasis on humanistic values, starting with a gap, there is need for policies that aim to cultivate moral-gradual erosion of ethical standards across science. To fill values from mentor to mentee, undermining the building of large team endeavors due to the transparency problem. ethical standards and sanctioning misbehavior is difficult in the problems raised here will be challenging since monitoring ethically conscious scientist. However, providing solutions to A body of ethical scientists is indeed an invaluable method.

An insidious problem highlighted is how a large team en-

Diverse disciplines

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

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

**Collaboration radius and team efficiency**

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

Towards a micro-level production function:

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

average number of publications per year  
$S_i$ is median number of coauthors per year

Output change ("growth fluctuation"),

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

std. deviation of publication change

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

team efficiency parameter $\psi$

---

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 $\psi$ and $\gamma$ values, suggesting that team management skills are an important factor related to success

$\gamma_{\text{Top100 Physics}} = 0.68(1) > \gamma_{\text{Prolific Physics}} = 0.52(1), \gamma_{\text{AsstProfessor Physics}} = 0.51(2)$
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?
The growth of team endeavors across multiple size scales requires individual introspection and institutional revision of the norms of team ethics:

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

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 December 2008

Professor’s little helper

The use of cognitive-enhancing drugs by both ill and healthy individuals raises ethical questions that should not be ignored, argue Barbara Sahakian and Sharon Morein-Zamir.

NATURE|Vol 450|20/27 December 2007

NATURE|Vol 452|10 April 2008

Poll results: look who’s doping

In January, Nature launched an informal survey into readers’ use of cognition-enhancing drugs. Brendan Maher has waded through the results and found large-scale use and a mix of attitudes towards the drugs.

“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 heterogeneous members.

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?

\( \approx \) factor of 20 increase in retractions from 2000 - 2010

The retraction penalty: Evidence from the web of science.
What makes science special (complex)?

Interactions mediated by social “forces”:

- Collaboration (attractive)
- Competition for priority (repulsive)
- Knowledge (an “exchange particle”)

Collaboration network

principal investigator

FIG. 1: Longitudinal analysis of publication and citation growth patterns.

(a,b) Growth curves, appropriately rescaled to start from unity, show the characteristic career trajectories of the scientists in each cohort. The characteristic \( \alpha \) and \( \beta \) exponents shown in each legend are calculated over the growth phase of the career, in (a) over the first 30 years and in (b) over the first 20 years. The mathematicians [E] have distinct career trajectories, with \( \beta \) since collaboration spillovers play a smaller role in their production growth.

(c) Schematic illustration of the multiplex scientific network surrounding career \( i \). Links in the upper network represent the dynamic collaborations between scientists (nodes); links within the lower network represent the citation network between papers (nodes); the cross-links between the networks represent the association between individual careers and the corresponding publication portfolio, serving as a platform for reputation signaling [14, 21, 23].

What makes science special (complex)?
Diverse collaboration strategies

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%)
Science: a co-evolving network of networks

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

\[ n_i(t) \] 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!

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

\[ N_i(t) \]

\[ \text{year, } t \]

\[ \alpha = 1 \]

\[ \alpha_i \]

Persistence and Uncertainty in the Academic Career,
A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli.
Common growth patterns observed across discipline

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 $\langle h \rangle = 61 \pm 21$

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

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

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

Set E: 50 highly-cited pure mathematicians, $\langle h \rangle = 20 \pm 10$
Models of science

A) microscopic reputation mechanisms
B) cumulative advantage mechanism
C) competition for limited opportunities
A) Reputation flows in the collaboration-citation network

What is the role of the network?
It constitutes the channels for reputation signaling, a mechanism used to overcome problems associated with incomplete information / reproducibility / and the “agency problem” in Science [P. Stephan, J. Econ. Lit 34. 1996]

⇒ Author-specific factors matter!
⇒ evidence is in the citation rates ( p ⇔ i )
Reputation effect citation model

# 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 \( \Pi_p(t) \equiv [c_p(t)]^\pi \)
2. citation life-cycles \( A_p(\tau) \equiv \exp[-\tau_p/\bar{\tau}] \)
3. author reputation effect \( R_i(t) \equiv [C_i(t)]^\rho \)
Author-specific features: $\pi_i$, $\Phi_i$, $\rho_i$

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

**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

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

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

### 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 \( \pi \), the paper citation effect, \( \overline{\pi} \), the life-cycle effect, and \( \rho \), the reputation effect.

<table>
<thead>
<tr>
<th>Name</th>
<th>( \pi_i )</th>
<th>( \overline{\pi}_i )</th>
<th>( \rho_i )</th>
<th>( \pi_i )</th>
<th>( \overline{\pi}_i )</th>
<th>( \rho_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOSSARD, AC</td>
<td>0.34 ± 0.027</td>
<td>4.92 ± 0.261</td>
<td>0.25 ± 0.008</td>
<td>0.80 ± 0.048</td>
<td>4.73 ± 0.184</td>
<td>0.09 ± 0.024</td>
</tr>
<tr>
<td>BARABÁSI, AL</td>
<td>0.42 ± 0.036</td>
<td>3.00 ± 0.155</td>
<td>0.29 ± 0.010</td>
<td>1.06 ± 0.016</td>
<td>3.65 ± 0.111</td>
<td>0.01 ± 0.011</td>
</tr>
<tr>
<td>Ave. ± Std. Dev. [A]</td>
<td>0.43 ± 0.14</td>
<td>5.67 ± 2.52</td>
<td>0.22 ± 0.06</td>
<td>0.96 ± 0.19</td>
<td>8.93 ± 4.09</td>
<td>-0.07 ± 0.11</td>
</tr>
<tr>
<td>BALTIMORE, D</td>
<td>0.32 ± 0.018</td>
<td>4.64 ± 1.48</td>
<td>0.28 ± 0.006</td>
<td>0.62 ± 0.047</td>
<td>5.92 ± 0.250</td>
<td>0.15 ± 0.026</td>
</tr>
<tr>
<td>LAEMMLI, UK</td>
<td>0.54 ± 0.036</td>
<td>5.09 ± 0.297</td>
<td>0.21 ± 0.014</td>
<td>1.09 ± 0.025</td>
<td>6.40 ± 0.255</td>
<td>-0.12 ± 0.019</td>
</tr>
<tr>
<td>Ave. ± Std. Dev. [D]</td>
<td>0.40 ± 0.14</td>
<td>6.64 ± 6.24</td>
<td>0.26 ± 0.05</td>
<td>0.99 ± 0.22</td>
<td>9.55 ± 26.30</td>
<td>-0.06 ± 0.14</td>
</tr>
<tr>
<td>SERRE, JP</td>
<td>0.33 ± 0.095</td>
<td>15.90 ± 3.724</td>
<td>0.14 ± 0.026</td>
<td>0.66 ± 0.065</td>
<td>20.50 ± 3.862</td>
<td>-0.03 ± 0.039</td>
</tr>
<tr>
<td>WILES, A</td>
<td>0.56 ± 0.208</td>
<td>5.23 ± 1.187</td>
<td>0.24 ± 0.052</td>
<td>0.70 ± 0.059</td>
<td>9.04 ± 0.633</td>
<td>0.10 ± 0.042</td>
</tr>
<tr>
<td>Ave. ± Std. Dev. [E]</td>
<td>0.27 ± 0.17</td>
<td>30.60 ± 56.80</td>
<td>0.14 ± 0.07</td>
<td>0.54 ± 0.25</td>
<td>21.40 ± 54.30</td>
<td>0.01 ± 0.11</td>
</tr>
</tbody>
</table>

Consider 2 scientists (with roughly equivalent paper lifecycle factor \( \overline{\pi} \)), and one with 10x as many total citations as the other, 

\[
C_1(t) = 10 \times C_2(t)
\]

then for 2 papers with the same # of citations \( c < c_x \) (in the strong reputation regime)

\[
\frac{\Delta c_{1,p}(t+1)}{\Delta c_{2,p}(t+1)} = 10^\rho = 1.66
\]
Benchmark patterns of microscopic career growth dynamics

\[ c_p(\tau) = \sum_{t} \Delta c_p(t) \]  
 cumulative # of citations at paper age \( \tau \)

\[ C_i(t) = \sum_{r=1}^{N_i(t)} c_i(r, t) \sim t^{\zeta_i} \]  
 cumulative citations by career age \( t \)

The rank-citation profile illustrates the evolution of the publication-impact portfolio

\[ c(r) \equiv A r^{-\beta} (N + 1 - r)^{\gamma} \]  
 discrete generalized Beta function (DGBD)

\[ C_i \sim h_i^{1+\beta_i} \]  
 simple scaling relation between the \( h \)-index and \( C \)

B) Modeling “Cumulative advantage”

career i

\[ \tau(1) \quad \tau(2) \quad \tau(3) \quad \tau(4) \quad \cdots \tau(n) \]

\[ t \sim \text{career position} \]
Empirical evidence for cumulative advantage

For each career $i$ we track his/her longitudinal publication rate by aggregating over publications in a *specific set* of high-impact journals.

**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 $\sim$ 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.)

$$g(x) = \frac{1}{\langle \tau(x) \rangle}$$

The progress probability $g$ is the inverse of the mean waiting time $\tau$.

[Graph showing the relationship between career position and mean waiting time]


Statistical regularities in the career longevity distribution

Pro Sports

• 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

C) Competition and contract length

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 \( t \)

\[ I = \text{finite labor force size} \]

Persistence and Uncertainty in the Academic Career,
A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli.
Appraising prior achievement

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

The cohort of $I$ agents compete for a fixed number of opportunities in each period over a lifespan of $t = 1 \ldots T$ periods.

In each period, the capture rate of a given individual $i$ is calculated by an appraisal of the achievement 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** $1/c$

$c \to 0$ : appraisal over all lifetime achievements ( ~ tenure system)
$c > 1$ : appraisal over only recent achievements (short-term contract system)

\[\text{exponential discount factor} \quad e^{-c\Delta t}\]
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.
Q: Is there an optimal appraisal (contract) length?

\[ c = 0.1 \text{ (~ long term appraisal)} \]

---

**Linear capture**

\[ \pi = 1.0 \]

---

**Super-linear capture**

\[ \pi = 1.2 \]

---

non-linear preferential capture model

\[ P_i(\cdot) = \frac{i(\cdot)^\pi}{\sum_{i=1}^{I} i(\cdot)^\pi} \]

---

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/


• 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.
Life cycles
Dynamic network characterized by life-cycles

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

least-cited papers

Collaboration network

Citation network

I

FIG. 1: Longitudinal analysis of publication and citation growth patterns.

(a,b) Growth curves, appropriately rescaled to start from unity, show the characteristic career trajectories of the scientists in each cohort. The characteristic \( \alpha \) and \( \beta \) exponents shown in each legend are calculated over the growth phase of the career, in (a) over the first 30 years and in (b) over the first 20 years. The mathematicians [E] have distinct career trajectories, with \( \beta \) since collaboration spillovers play a smaller role in their production growth. (c) Schematic illustration of the multiplex scientific network surrounding career \( i \). Links in the upper network represent the dynamic collaborations between scientists (nodes); links within the lower network represent the citation network between papers (nodes); the cross-links between the networks represent the association between individual careers and the corresponding publication portfolio, serving as a platform for reputation signaling [14, 21, 23].
FIG. 1: Longitudinal analysis of publication and citation growth patterns.

(a,b) Growth curves, appropriately rescaled to start from unity, show the characteristic career trajectories of the scientists in each cohort. The characteristic ** and ** exponents shown in each legend are calculated over the growth phase of the career, in (a) over the first 30 years and in (b) over the first 20 years. The mathematicians ** have distinct career trajectories, with ** since collaboration spillovers play a smaller role in their production growth. (c) Schematic illustration of the multiplex scientific network surrounding career **. Links in the upper network represent the dynamic collaborations between scientists (nodes); links within the lower network represent the citation network between papers (nodes); the cross-links between the networks represent the association between individual careers and the corresponding publication portfolio, serving as a platform for reputation signaling [14, 21, 23].

**Dynamic network characterized by life-cycles**

- Collaboration network
- Citation network

**Weak ties ~ 1-3 years**
**Strong ties ~ lifetime**

Q: Is the “Invisible college” held together by weak ties?