# Reputation and impact in academic careers

Alexander Michael Petersen[a,1], Santo Fortunato[b,1], Raj K. Pan[b], Kimmo Kaski[b], Orion Penner[c], Armando Rungi[a], Massimo Riccaboni[c,d], H. Eugene Stanley[e,1], and Fabio Pammolli[a,e]

[a]Laboratory for the Analysis of Complex Economic Systems, Institutions, Markets, Technologies (IMT) Lucca Institute for Advanced Studies, 55100 Lucca, Italy; [b]Department of Biomedical Engineering and Computational Science, Aalto University School of Science, FI-00076 Aalto, Finland; [c]Laboratory of Innovation Management and Economics, IMT Lucca Institute for Advanced Studies, 55100 Lucca, Italy; [d]Department of Managerial Economics, Strategy and Innovation, Katholieke Universiteit Leuven, 3000 Leuven, Belgium; and [e]Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215

Reputation 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. Here, we develop an original framework for measuring how a publication's citation rate $\Delta c$ depends on the reputation of its central author $i$, in addition to its net citation count $c$. To estimate the strength of the reputation effect, we perform a longitudinal analysis on the careers of 450 highly cited scientists, using the total citations $C_i$ of each scientist as his/her reputation measure. We find a citation crossover $c_\times$, which distinguishes the strength of the reputation effect. For publications with $c < c_\times$, the author's reputation is found to dominate the annual citation rate. Hence, a new publication may gain a significant early advantage corresponding to roughly a 66% increase in the citation rate for each tenfold increase in $C_i$. However, the reputation effect becomes negligible for highly cited publications meaning that, for $c \geq c_\times$, the citation rate measures scientific impact more transparently. In addition, we have developed a stochastic reputation model, which is found to reproduce numerous statistical observations for real careers, thus providing insight into the microscopic mechanisms underlying cumulative advantage in science.

computational sociology | science of science | networks of networks | Matthew effect | sociophysics

Citation counts are widely used to judge the impact of both scientists and their publications (1–4). Although it is recognized that many factors outside the pure merit of the research or the authors influence such counts, little effort has been devoted to identifying and quantifying the role of the author-specific factors. Recent investigations have begun to study the impact the individual scientists have through collaboration and reputation spillovers (5, 6), two integrative features of scientific careers that contribute to cumulative advantage (7–9). However, the majority of citation models avoid author-specific effects, mainly due to the difficulty in acquiring comprehensive disambiguated career data (10–13). As the measures are becoming increasingly common in evaluation scenarios throughout science, it is crucial to better understand what the citation measures actually represent in the context of scientists' careers. Moreover, how does reputation affect a scientist's access to key resources, the incentives to publish quality over quantity, and other key decisions along the career path (14–18)? In addition, what role does reputation play in the mentor-matching process within academic institutions, in the effectiveness of single/double blinding in peer review, and in the reward system of science (14, 15, 19)?

It is against this background that we have developed a quantitative framework with the goal of isolating the effect of author reputation upon citation dynamics. Specifically, by controlling for time- and author-specific factors, we quantify the role of author reputation on the citation life cycle of individual publications at the micro level. We use a longitudinal career dataset from Thomson Reuters Web of Science comprising 450 highly cited scientists, 83,693 articles, and 7,577,084 citations tracked over 387,103 publication years. Dataset [A] refers to 100 top-cited physicists, [B] to another set of 100 highly prolific physicists, [C] to 100 assistant professors in physics, [D] to 100 top-cited cell biologists, and [E] to 50 top-cited pure mathematicians (for further data elaboration, see *SI Appendix*). For each central scientist $i$, we analyze the scientific production measured by the number $n_i(t)$ of publications published in year $t$, the cumulative number of citations $c_{i,p}(t)$ received by publication $p$, and our quantitative reputation measure defined here as the net citations aggregated across all publications $C_i(t) = \sum_p c_{i,p}(t)$.

We begin with a description of our reputation model, followed by an empirical analysis of career trajectories, establishing $C_i$ as a good quantitative measure of reputation. We then establish quantitative benchmarks from the citation distribution within individual publication portfolios and also quantify features of the citation life cycle, both of which are crucial components of our reputation effect model. Combining several empirical features of our analysis, we then investigate the role of the reputation effect, showing that author reputation accounts for a significant fraction of the citation rate of young publications, thus providing a testable mechanism underlying cumulative advantage in science (7–9). We conclude by developing a stochastic Monte Carlo (MC) reputation model, one which matches the microscopic and macroscopic citation dynamics.

## Results

**Reputation Signaling.** Academic career growth is a complex process emerging from the institutional, social, and cognitive aspects of science. Conceptually, each career $i$ is embedded in two fundamental

### Significance

Over a scientist's career, a reputation is developed, a standing within a research community, based largely upon the quantity and quality of his/her publications. Here, we develop a framework for quantifying the influence author reputation has on a publication's future impact. We find author reputation plays a key role in driving a paper's citation count early in its citation life cycle, before a tipping point, after which reputation has much less influence relative to the paper's citation count. In science, perceived quality, and decisions made based on those perceptions, is increasingly linked to citation counts. Shedding light on the complex mechanisms driving these quantitative measures facilitates not only better evaluation of scientific outputs but also a more transparent evaluation of the scientists producing them.

www.manaraa.com

networks that are interconnected: the nodes in the first network represent scientists and in the second network represent publications. The links within the first network represent collaborations between scientists, and within the second network they represent citations between publications; the cross-links represent the associations between individuals and their publications.

Because these networks are dynamic, it is difficult to fully understand for any given individual, let alone the entire system, the complex information contained by all associations. As a result, scientific reputation has emerged as a key signaling mechanism to address the dilemma of excessive information that arises, for example, in the task of evaluating, comparing, and ranking publication profiles in academic competitions. Reputation signals can flow between scientists, $j \rightleftarrows k$; between publications, $p \rightleftarrows q$; and between a publication and a scientist, $p \rightleftarrows i$. The focus of our analysis is on this latter dependency, $i \rightarrow p$, whereby author reputation can impact the citation rate of his/her publications, generating subsequent reputation feedback, $p \rightarrow i$.

Reputation plays an important role as a signal of trustworthiness and quality, a role that addresses directly the "agency problem" characterizing the reward system in science (14). Moreover, reputation signaling in scientific networks is used to overcome information asymmetries between scientists and other academic agents; in this role, it will become increasingly important as the rate of science publication grows and scientists have less time to absorb relevant advancements (14, 19–21). With little time to read every paper on a given topic, this trustworthiness signal is anecdotally consistent with the common practice of perusing the author names when preliminarily evaluating the relevance of a newly found publication. In the past, an author's identity and associated reputation was mainly linked to reference lists and personal interactions. Nowadays, an author's reputation is becoming increasingly visible through searchable publication databases, laboratory websites, press, and other media, in addition to citations.

We measure the author reputation by $C_i(t)$, which measures not only the number of times his/her $N_i(t)$ publications have been referenced (an indication of overall scientific impact) but also the number of appearances of his/her name in the literature, thereby providing a name association visibility. What $C_i$ does not account for is intrinsic research quality, e.g., the quality ratio $C_i/N_i$ is broadly distributed across scientists. Because quantitative proxies for quality are limited to citation counts, it is presently difficult to distill the role played by quality in assessing overall scientific impact.

By analyzing the top scientists, we reduce the compound reputation effect occurring when two or more scientists of comparable reputation are coauthors on a publication, a scenario where it may be difficult to estimate the differential impact of these scientists on the citation rate. Due to data limitations requiring author name disambiguation and career data for all coauthors $j$, we assume that a majority of the reputation signal is attributable to the central scientist $i$ by the following approximation: $C(t) \approx \sum_j C_j(t) \approx C_i(t)$. Also, by analyzing top-cited cohorts, we can establish an upper bound to the strength of the reputation effect. We note that $C_i$ possibly discounts the role of mentor reputation effects early in the career (22). Nevertheless, by analyzing top scientists, the signaling advantage received early in their careers by associating with prestigious mentors/coauthors should be negligible over the long run (20).

To measure the role of author reputation vis-à-vis publication impact, we use a regression model that correlates the increase in the number of citations $\Delta c_{i,p}(t+1)$ for a given paper $p$ in year $t+1$ using three explanatory variables: (i) the role played by the net number of citations $c_p(t)$ accrued up to paper age $\tau_p$ quantified by the power-law regression parameter $\pi$, (ii) the role of publication age and the obsolescence of knowledge quantified by the exponential regression parameter $\overline{\tau}$; and (iii) the role of author reputation $C_i(t)$ quantified by the power-law regression parameter $\rho$.

Together, these three features are (i) the publication citation effect $\Pi_p(t) \equiv [c_p(t)]^{\pi}$, (ii) the life cycle effect $A_p(\tau) \equiv \exp[-\tau_p/\overline{\tau}]$, and (iii) the author reputation effect $R_i(t) \equiv [C_i(t)]^{\rho}$. We perform a multivariate regression to estimate the $\pi$, $\overline{\tau}$, and $\rho$ values that parameterize the citation model,

$$\Delta c_{i,p}(t+1) \equiv \eta \times \Pi_p(t) \times A_p(\tau) \times R_i(t), \qquad [1]$$

with the multiplicative log-normal noise term $\eta$. In *SI Appendix*, we perform an additional fixed-effects regression using year as well as author variables to better control for the overall growth in scientific output across time. To fully justify our reputation effect model, in what follows we first account for two key features: measures for cumulative career reputation and obsolescence features of the citation life cycle.

**Patterns of Growth for Longitudinal Reputation Measures.** In this section, we investigate the patterns of cumulative publication and citation growth across the career. A striking statistical patterns observed for top scientists is the faster than linear growth in time, both in cumulative publication number $N_i(t) \equiv \sum_{t'=1}^{t} n_i(t')$ and in cumulative citation count $C_i(t) \equiv \sum_{p=1}^{N_i(t)} c_{i,p}(t)$ for a large part of a scientist's "growth phase," which we find to be $\approx 30$ y after their first publication. Fig. 1 *A* and *B* shows the characteristic growth trajectories $\langle N'(t) \rangle \sim t^{\overline{\alpha}}$ and $\langle C'(t) \rangle \sim t^{\overline{\zeta}}$, calculated by an appropriate average over individual $N_i(t)$ and $C_i(t)$, respectively. To facilitate visual comparison, we use arbitrary normalized ordinate units so that each curve starts from the same point, $\langle N'(1) \rangle = \langle C'(1) \rangle \equiv 1$. The growth trajectories are characterized by superlinear algebraic growth, with $\overline{\alpha} \gtrsim 1$ and $\overline{\zeta} > \overline{\alpha}$ (values shown in Fig. 1). Individual exponents $\alpha_i$ and $\zeta_i$ are also calculated for the $N_i(t)$ and $C_i(t)$ of each author (in addition to multiple other quantitative measures; *SI Appendix*, Tables S1–S9). We averaged both $\alpha_i$ and $\zeta_i$ within each dataset, confirming that $\langle \alpha_i \rangle \cong \overline{\alpha}$ and $\langle \zeta_i \rangle \cong \overline{\zeta}$, confirming that the aggregate patterns hold for the individual scale. In *SI Appendix*, we control for the exponential growth in scientific publication rates, which can
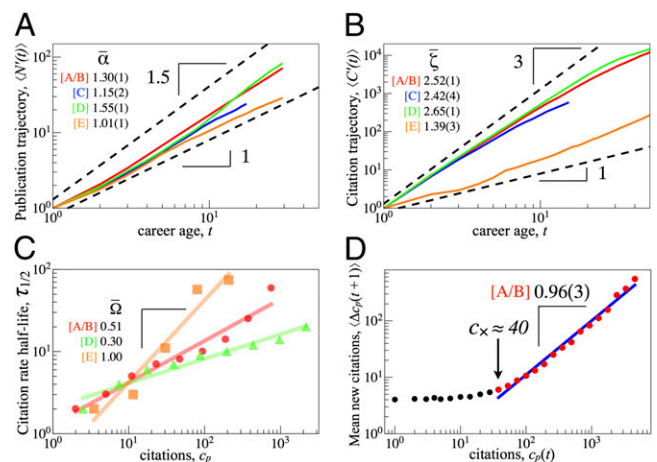


**Fig. 1.** Quantifying cumulative reputation measures and citation dynamics. (*A* and *B*) Growth trajectories of the cumulative publications $N'(t)$ and citations $C'(t)$, appropriately rescaled to start from unity in each ordinate. The characteristic $\overline{\alpha}$ and $\overline{\zeta}$ exponents shown in each legend are calculated over the growth phase of the career. The mathematicians [E] have distinct career trajectories, with $\overline{\alpha} \approx 1$ because collaboration spillovers via division of labor likely play a smaller role in publication rate growth. See *SI Appendix*, Tables S1–S9, for $\alpha_i$ and $\zeta_i$ values calculated for individual careers. (*C*) Relation between $\tau_{1/2}$ and cumulative citations $c_p$. (*D*) PA dynamics with $\pi \approx 1$ break down for $c < c_\times$. The reputation effect provides a citation boost above the baseline PA citation rate attributable to $c_p(t)$ only.

APPLIED PHYSICAL SCIENCES

SOCIAL SCIENCES

www.manaraa.com

contribute to the longitudinal growth in $C_i(t)$. We define "deflated" citation counts $\Delta c_{i,p}^D(t) \equiv \Delta c_{i,p}(t)/D(t)$, which are normalized by the number of publications $D(t)$ within a given discipline (because a new publication can cite an old publication only once). For each discipline, we observe a 5% exponential growth in $D(t)$ over the last half century. After deflating each $C_i(t)$, the net affect is only to reduce the estimated $\zeta_i$ values by roughly 15%, meaning that the growth exponents $\zeta_i \gtrsim 2$ reflect significant growth above the underlying baseline growth trend in science.

Hence, we use $C_i(t)$ as a quantitative measure of reputation owing to the fact that the time dependence is readily quantified by a single parameter $\zeta_i$. We also use the power-law scaling of $C_i(t)$ as a benchmark for the stochastic career model we develop in the final section. Fig. 2 shows two additional empirical benchmarks: (i) the microscopic citation dynamics of individual publications comprising the publication portfolio, and (ii) the rank-citation profile, which is the Zipf distribution of the publications ranked in decreasing order of rank $r$, $c_i(1) \geq c_i(2) \geq \ldots \geq c_i(N_i)$. We confirm that the individual curves $c_i(r)$ belong to the class of the discrete generalized $\beta$ distributions (DGBDs), which in the general form reads $c(r) \propto r^{-\beta}(N+1-r)^\gamma$ (4). We validate the DGBD fits using the $\chi^2$ test (*SI Appendix*).

**Variability in the Citation Life Cycle.** Important scientific discoveries can cause paradigm shifts and significantly boost the reputation of scientists associated with the discovery (23). To measure the reputation effect, one must also account for obsolescence features of scientific knowledge. It is also important to account for the variations in scientific impact, because most publications report results that are not seminal breakthroughs, but, rather, report incremental advances that are likely to have relatively short-term relevance.
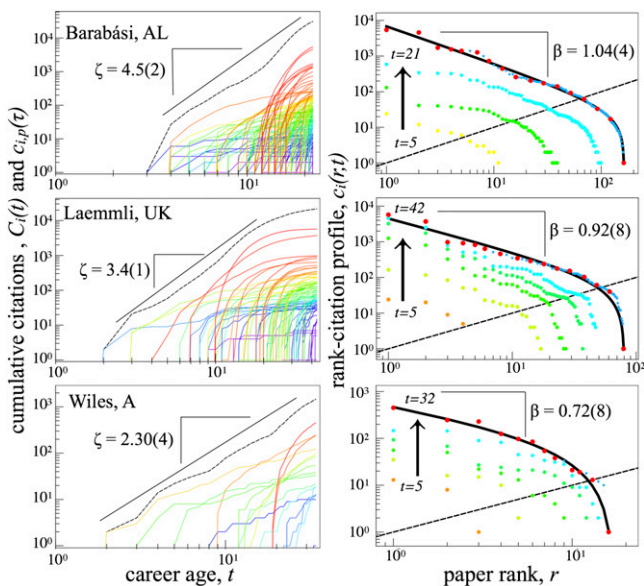


**Fig. 2.** Quantitative patterns in the growth and size-distribution of the publication portfolio for scientists from 3 disciplines. (*Left*) $c_{i,p}(t)$ for each author's most cited papers (colored according to net citations in 2010) along with $C_i(t) \sim t^{\zeta_i}$ (dashed black curve). (*Right*) The evolution of each author's rank-citation profile using snapshots taken at 5-y intervals. The darkest blue data points represent the most recent $c_i(r,t)$, and the subset of red data points indicate the logarithmically spaced data values used to fit the empirical data to our benchmark DGBD rank-citation distribution model (4) (solid black curve; *SI Appendix*). The intersection of $c_i(r,t)$ with the dashed black line corresponds to the author's h-index $h_i(t)$.

In this section, we analyze the dynamics of the citation trajectory $\Delta c_p(\tau)$, the number of new citations received in publication year $\tau$, where $\tau$ is the number of years since the publication was first cited. We analyze $\Delta c_p(\tau)$ at two levels of aggregation: (i) for each discipline, we calculate an averaged $\Delta c_p(\tau)$ by collecting publications with similar total citation counts $c_p$. To achieve a scaled trajectory that is better suited for averaging, we normalize each individual $\Delta c_p(\tau)$ by its peak citation value, $\Delta c_p'(\tau) \equiv \Delta c_p(\tau)/\text{Max}[\Delta c_p(\tau)]$. In Fig. 3, the panels on the *Left* show the characteristic citation trajectory of publications belonging to each of the top 5 quintiles of each disciplinary citation distribution. Each curve represents the average trajectory $\langle \Delta c'(\tau) \rangle \equiv N_q^{-1} \sum_p \Delta c_p'(\tau)$ calculated from the $N_q$ publications in quintile $q$. (ii) For each career $i$, we calculate $\langle \Delta c_i'(\tau) \rangle$ by averaging over groups of ranked citation sets within his/her publication portfolio. The panels on the *Right* of Fig. 3 show that, even within prestigious careers, there is significant variation in the citation life cycle.

At both levels of aggregation, the impact life cycle typically peaks before publication age $\tau \approx 5$ y. Counterexamples likely correspond to publications that receive a delayed secondary attention, e.g., receiving subsequent experimental validation of a previous theoretical prediction, and vice versa. We define the half-life $\tau_{1/2}$ as the time to reach half the peak citation rate, $\Delta c'(\tau_{1/2}) = 1/2$ in the decay phase. Papers in the theoretical domains of mathematics and physics can exhibit $\tau_{1/2} > 40$ y. Remarkably, some mathematics publications even have $\tau_{1/2}$ that span nearly the entire data sample duration of 100 y, reflecting the indisputable and foundational nature of progress by proof. This is in contrast to top-cited cell biology publications, whereby for even the top 20% of most cited works, the value $\tau_{1/2} \approx 10$ y. This relatively short decay timescale likely arises from the large scale of research output in biomedical fields, which leads to a significantly higher discovery rate, and likewise, a relatively faster obsolescence rate.

The relation between the decay timescale $\bar{\tau}$ and $c_p$ provides insight into the knowledge diffusion rate. Fig. 1C shows an approximate scaling relation $\tau_{1/2} \sim c_p^{\overline{\Omega}}$ when grouping publications into logarithmically spaced $c_p$ bins. Physics and biology differ mainly for the highly cited publications with $c_p \gtrsim 40$, whereas mathematics shows larger variation in $\tau_{1/2}$ per citation. The $\overline{\Omega}$ value provides an approximate relation between citations and time. In mathematics, $\tau_{1/2} \propto c_p$, indicating that the impact is distributed roughly uniformly across time. However, for biology publications, the sublinear relation with $\overline{\Omega} \approx 0.30$ indicates that for two publications, one with twice the citation impact as the other, the more cited publication gained twice the number of citations over a time period $\tau_{1/2}$ that was less than twice as large as the $\tau_{1/2}$ of the less-cited publication. The differences in $\overline{\Omega}$ are possibly related to discipline-dependent bursts in technological advancement, funding initiatives (15), and other social aspects of science that are related to nonlinearities in scientific advancement.

**Baseline Citation Model.** To provide an initial test for basic mechanistic differences between the citation dynamics of highly cited publications and less-cited publications, in this subsection we analyze the relation between $\Delta c_p(t+1)$ and $c_p(t)$ representing the standard baseline preferential attachment (PA) model (corresponding to the limit $\bar{\tau} \to \infty$ and $\rho = 0$). Grouping together papers by $c_p(t)$ (using logarithmic bins), we calculate for each group the mean number of new citations in the following year, $\langle \Delta c_p(t+1) \rangle$. Fig. 1D shows the empirical relation for physicists in datasets [A/B], indicating that publications with citations above a gradual but substantial citation crossover value $c_\times$ obey a distinct scaling law that matches approximately linear ($\pi \approx 1$) PA dynamics (see *SI Appendix*, Fig. S8, for other disciplines). However, below $c_\times$, the citation rates are in excess of the citation rate expected from
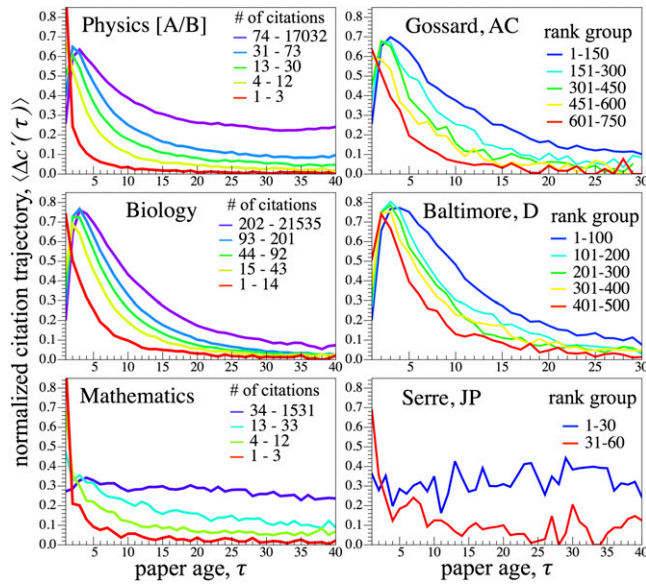
Petersen et al.

www.manaraa.com

**Fig. 3.** The citation life cycle reflects both the intrinsic pace of discovery and the obsolescence rate of new knowledge, two features that are discipline dependent. (*Left*) For each of three disciplines, the averaged citation trajectory $\langle \Delta c'(\tau) \rangle$ is calculated for papers in the $n$th quintile with the corresponding citation range indicated in each legend. For example, for physicists in dataset [A], the top 20% of papers have between 74 and 17,032 citations, and the papers in percentile 21–40 have between 31 and 73 citations. (*Right*) $\langle \Delta c'(\tau) \rangle$ calculated for rank-ordered groups of papers (listed in each legend) for three authors chosen from each discipline.

linear PA alone, reflecting the citation premium that can be achieved via reputation.

**Quantifying the Role of the Reputation Effect.** By analyzing the publications of highly cited scientists, we have shown that the basic citation dynamics above and below the citation crossover value $c_\times$ vary considerably. In this subsection, we investigate the role played by the reputation effect for publications with $c_p(\tau) \geq c_\times$ compared with publications with $c_p(\tau) < c_\times$. Based upon the assessment of the growth dynamics (*SI Appendix*, Figs. S8 and S9), we choose the crossover values $c_\times \equiv 40$ [A/B], $c_\times \equiv 10$ [C], $c_\times \equiv 100$ [D], and $c_\times \equiv 20$ [E]. Our results are not strongly dependent on reasonable variations around our choice of $c_\times$. Table 1 shows the $\pi_i$, $\overline{\tau}_i$, and $\rho_i$ estimates, above and below $c_\times$, for the individual careers highlighted in Figs. 1 and 3.

Our main result is a robust pattern of role switching by author- and publication-specific effects, specifically,

$$\rho(c < c_\times) > \rho(c \geq c_\times) \quad \text{and} \quad \pi(c < c_\times) < \pi(c \geq c_\times). \quad [2]$$

For example, for the aggregate dataset [A/B] representing prolific physicists, we estimate the values $\rho(c < 40) \approx 0.2$, $\rho(c \geq 40) \approx 0$, $\pi(c < 40) \approx 0.4$, and $\pi(c \geq 40) \approx 1$. To emphasize the role of reputation on new publications, consider two scientists separated by a factor of 10 in their cumulative citations, $C_1(t) = 10C_2(t)$. All other things being equal, the citation premium attributable to reputation alone for publications in the reputation regime ($c < c_\times$) is $\Delta c_1(t)/\Delta c_2(t) = 10^\rho \approx 1.66$ (using the value $\rho = 0.22$ for dataset [A]). Hence, there is a 66% increase in the citation rate for each 10-fold increase in $C_i(t)$, which integrated over a career can provide significant positive feedback. A pattern that emerges independent of discipline is $\rho(c \geq c_\times) \approx 0$, meaning that reputation only plays a significant role for $c < c_\times$. In *SI Appendix*, section S6, we test the robustness of this result by implementing a fixed-effects regression, the result of which reaffirms the distinct roles of $\pi$ and $\rho$ above and below $c_\times$. Hence, these two inequalities in Eq. **2** indicate that publications are initially boosted by author reputation to a citation "tipping point" $c_{i,p} \approx c_\times$, above which the citation rate is sustained in large by publication reputation. These findings show how microscopic reputation mechanisms contribute to cumulative "rich-get-richer" processes in science (7, 9).

**Simulating Synthetic MC Careers with the Reputation Model.** Here, we discuss three variants of a MC career growth model that simulates the dynamics of $\Delta c_{i,p}(t+1)$ for each publication $p$ in each time period $t$ of the career of synthetic author $i$. With each variant, we introduce progressively a new feature of publication citation trajectories. (*i*) We begin with a basic linear PA model whereby $\Delta c_{i,p}(t+1) \propto c_{i,p}(t)$. Model *ii* (PA-LC) includes a life-cycle (LC) obsolescence effect, $A_p(\tau)$. Fig. 4 compares models *i* and *ii*, which do not incorporate author specific factors, with the reputation model *iii* given by Eq. **1**. The PA model fails to reproduce the characteristic trajectories of real publications, because there is a clear first-mover advantage (24) for the first publications published in the career, as well as non–power-law growth of $C_i(t)$.

We use quantitative patterns demonstrated for real careers in Figs. 2 and 3 as empirical benchmarks to distinguish models *ii* and *iii*. We confirm that the reputation model *iii* satisfies the empirical benchmark characteristics in all three graphical categories (Fig. 4). We also confirm for the model *iii*, but not for the model *ii*, that there is a distinction between $\langle \Delta c'(\tau_p) \rangle$ for different rank sets. Furthermore, for model *iii*, we quantitatively confirm that $C(t) \sim t^\zeta$ with $2 \lesssim \zeta \lesssim 3$. For sufficiently large $t$, we also confirm that $c(r,t)$ belongs to the class of DGBD distributions, with $\beta$ values within the range of values observed empirically. In *SI Appendix*, we further demonstrate how the model can be used

**Table 1. Best-fit parameters for each effect ($\pm$SEs), both for individuals and the average values ($\pm$SD) within each dataset**

| Name | $c(t-1) < c_\times$ | | | $c(t-1) \geq c_\times$ | | |
|---|---|---|---|---|---|---|
| | $\pi_i$ (paper) | $\overline{\tau}_i$ (life cycle) | $\rho_i$ (reputation) | $\pi_i$ (paper) | $\overline{\tau}_i$ (life cycle) | $\rho_i$ (reputation) |
| Gossard, A. C. | $0.34 \pm 0.027$ | $4.92 \pm 0.261$ | $0.25 \pm 0.008$ | $0.80 \pm 0.048$ | $4.73 \pm 0.184$ | $0.09 \pm 0.024$ |
| Barabási, A. L. | $0.42 \pm 0.036$ | $3.00 \pm 0.155$ | $0.29 \pm 0.010$ | $1.06 \pm 0.016$ | $3.65 \pm 0.111$ | $0.01 \pm 0.011$ |
| Average $\pm$ SD [A] | $0.43 \pm 0.14$ | $5.67 \pm 2.52$ | $0.22 \pm 0.06$ | $0.96 \pm 0.19$ | $8.93 \pm 4.09$ | $-0.07 \pm 0.11$ |
| Baltimore, D. | $0.32 \pm 0.018$ | $4.64 \pm 0.148$ | $0.28 \pm 0.006$ | $0.62 \pm 0.047$ | $5.92 \pm 0.250$ | $0.15 \pm 0.026$ |
| Laemmli, U. K. | $0.54 \pm 0.036$ | $5.09 \pm 0.297$ | $0.21 \pm 0.014$ | $1.09 \pm 0.025$ | $6.40 \pm 0.255$ | $-0.12 \pm 0.019$ |
| Average $\pm$ SD [D] | $0.40 \pm 0.14$ | $6.64 \pm 6.24$ | $0.26 \pm 0.05$ | $0.99 \pm 0.22$ | $9.55 \pm 26.30$ | $-0.06 \pm 0.14$ |
| Serre, J. P. | $0.33 \pm 0.095$ | $15.90 \pm 3.724$ | $0.14 \pm 0.026$ | $0.66 \pm 0.065$ | $20.50 \pm 3.862$ | $-0.03 \pm 0.039$ |
| Wiles, A. | $0.56 \pm 0.208$ | $5.23 \pm 1.187$ | $0.24 \pm 0.052$ | $0.70 \pm 0.059$ | $9.04 \pm 0.633$ | $0.10 \pm 0.042$ |
| Average $\pm$ SD [E] | $0.27 \pm 0.17$ | $30.60 \pm 56.80$ | $0.14 \pm 0.07$ | $0.54 \pm 0.25$ | $21.40 \pm 54.30$ | $0.01 \pm 0.11$ |

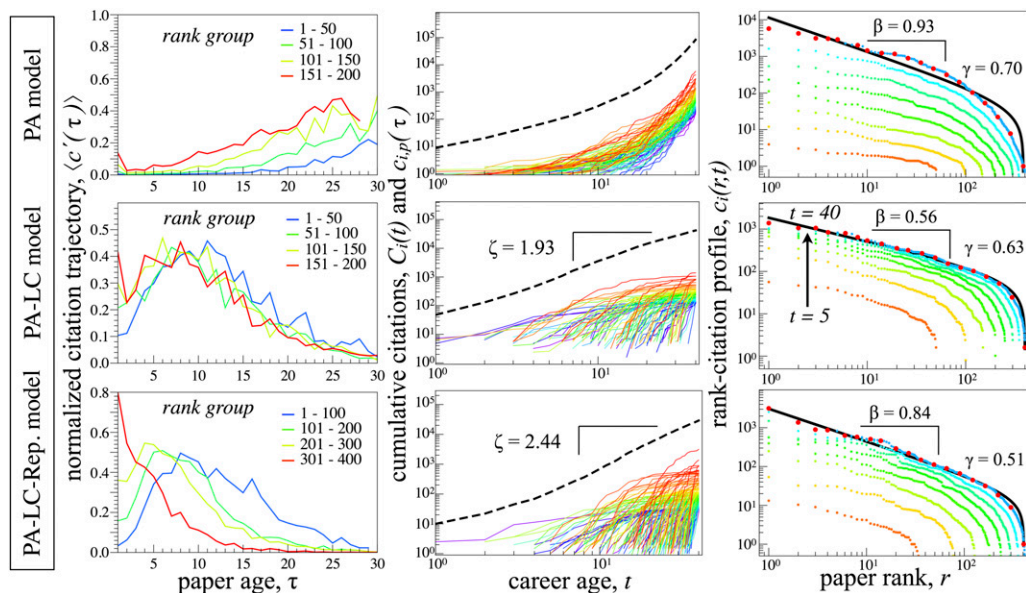For statistical significances, see *SI Appendix*, Tables S10–S22.

**Fig. 4.** Comparison of three MC career models against empirical benchmarks demonstrated in Figs. 1–3 and *SI Appendix*, Figs. S1–S3. For each model, we show $\langle \Delta c'(\tau) \rangle$ for the top four groups of ranked papers, the evolution of $c_{i,p}(\tau)$ and $C_i(t)$ (dashed black curve), and the evolution of the rank-citation profile $c_i(r)$ at 5-period intervals. The best-fit DGBD $\beta$ and $\gamma$ parameters are also useful as quantitative benchmarks. For each model, we evolve the system over $T \equiv 40$ periods, each period representative of a year. See *SI Appendix* for further elaboration of the model parameters used in the MC simulation.

to estimate properties of "average" careers for a given MC parameter set. For example, *SI Appendix*, Fig. S11, shows excellent agreement between the reputation model's prediction and empirical data when estimating the fraction $f_{\geq c_x}$ of publications with $c_p \geq c_{\times}$ for a given career age $t$. Empirically, we observe saturation $f_{\geq c_x} \approx 0.20$–$0.30$ for large $t$.

## Discussion

Social networks in science are characterized by heterogeneous structure (25) that provides opportunities for intellectual and social capital investment (26) and influences scientists' research strategies (21). Identifying patterns of career growth is becoming increasingly important, largely due to the widespread emergence of quantitative evaluation processes and recent efforts to develop quantitative models of career development. However, difficulties in accounting for complex social mechanisms, in addition to nonlinearities and nonstationarities in the career growth process, highlight the case for caution in the development of predictive career models (16, 17). Without a better understanding of the institutional features and scientific norms that affect scientific careers, along the variable path from apprentice to group leader and mentor, there is a possibility to misuse quantitative career metrics in the career evaluation process.

Toward the goal of better understanding career growth, with potential policy implications for the quantitative career evaluation process, we have analyzed the effect of reputation on the microlevel processes underlying the dynamics of a scientist's research impact. We used a regression model for the citation rate $\Delta c_{i,p}$, which accounts for the role of publication impact ($\pi$), the role of knowledge obsolescence ($\bar{\tau}$), and the role of author reputation ($\rho$). Interestingly, we find that the reputation parameter $\rho(c \geq c_{\times}) \approx 0$, meaning that in the long run the reputation effect makes a negligible contribution to the citation rate of papers with large $c_p$. However, we identify caveats concerning the way publications can become highly cited. By analyzing the variation of $\rho$ and $\pi$ for publications above and below a citation threshold $c_{\times}$, we identify the advantageous role that author reputation plays in the citation dynamics of new publications, finding that future publications can gain roughly a 66% increase in $\Delta c$ for each 10-fold increase in reputation $C_i$. We note that it is also likely that both institutional affiliation and journal reputation also play a role in the citation dynamics; however, disentangling

the interaction between the multiple reputation sources will likely be challenging and remains an open avenue for investigation.

In the process of analyzing the effect of reputation on career growth, it was necessary to also quantify two essential features of our model, namely patterns of cumulative productivity and impact across the career, and patterns of obsolescence in the citation life cycle of individual publications. For prolific scientists, we have identified a robust pattern of growth for two cumulative reputation measures, $N_i(t)$ and $C_i(t)$, each of which are quantifiable by a single scaling parameter, $\alpha_i$ and $\zeta_i$, respectively. These regularities suggest that underlying social processes sustain career growth via reinforcing coevolution of scientific collaboration and publication (6, 27–29). We also introduced a citation deflator index to control for the increased supply of citations arising from the exponential 5% growth (per year) in the total publication output. Analyzing the growth of "deflated" citation trajectories, $C_i^D(t)$, we observed $\zeta_i \gtrsim 2$ values which confirms that the observed career growth is significantly above the baseline inflation rate of science. We note that in using nondecreasing cumulative reputation measure $C_i(t)$, we have overlooked the possibility that reputation can significantly decrease, as occurs when a scientist is associated with invalidated and/or fraudulent science. Indeed, recent evidence indicates that the retraction of a publication can have a negative impact on the potential growth of $C_i$ (30). As a robustness check, we also used the annual citation rate $\Delta C_i(t)$ as an additional (noncumulative) reputation measure, one that is more amenable to controlling for secular growth trends. We applied a multivariate fixed effects regression using $\Delta C_i(t)$ as the reputation measure (*SI Appendix*, section S6), which reconfirms the role of reputation in citation dynamics.

Our analysis tracks the evolution of each scientist's publication portfolio across the career, suitably illustrated by the rank-citation profile $c_i(r)$, which highlights the skewed distribution of $c_{i,p}$, even within a career. Arising from the power-law features of $c_i(r)$ (4), we emphasize the disproportionate fraction of a scientist's total citations $C_i$ owed to the $c_i(r = 1)$ citations coming from his/her highest-cited publication. For example, the average and SD of the ratio $c_i(1)/C_i$ is $0.15 \pm 0.13$ for the physicists, $0.09 \pm 0.08$ for the biologists, and $0.16 \pm 0.08$ for the mathematicians we analyzed, which emphasizes the potentially large reputation boost that can follow from just a single high-impact publication. With rapidly increasing numbers of journals accompanied by the opportunity for rapid publication, the reputation effect provides an incentive to aim for

quality over quantity in the publication process, reinforcing a research strategy that is beneficial for science and scientists.

It is also important to consider the role of reputation in light of the increasing orientation of science around team endeavors characterized by multiple levels of hierarchy and division of labor (31). Because it is difficult to evaluate and assign credit to individual contributions in a team setting, there may be an increase in the role and strength of the reputation in overcoming the problem associated with asymmetric and incomplete information. In addition to the collaboration network, reputation also plays a key role in numerous other scientific inputs (money, labor, knowledge, etc.) that inevitably affect the overall quantity and quality of scientific outputs. It will become increasingly important to understand the relation between these inputs and outputs to efficiently allocate scientific resources (6, 15, 18).

In light of individual careers, an institutional setting based on quantitative appraisal that neglects these complex relations may inadvertently go against the goal of sustaining the careers of talented and diligent young academics (6). For example, our finding of a crossover behavior around $c_\times$ shows how young scientists lacking reputation can be negatively affected by social stratification in science. The appealing competitive advantage gained by working with a prestigious mentor may be countered by the possibility that it may not be the ideal mentor–advisee match. Despite having analyzed cohorts of highly cited scientists, our results have broad implications across the scientific population when one considers the numerous careers that interact with top scientists via collaboration or mentorship.

In excess, the reputation effect may also negatively affect science, especially considering how online visibility has become a relatively new reputation platform in an increasingly competitive environment. As such, strategies of self-promotion may emerge as scientists try to "game" with reputation systems. In such scenarios, it may be hard to disentangle fair from foul play.

For example, it may be difficult to distinguish self-citation strategies aimed at boosting $C_i$ from the natural tendency for scientists who are crossing disciplinary borders to self-cite with the intention to send credibility signals (32). Reputation will also become increasingly important in light of preferential treatment in search queries, e.g., Google Scholar, which provide query results ordered according to citation measures. These systemic search and retrieval features may further strengthen association of reputation between publications and authors. In all, our results should motivate future research to inspire institutional and funding body evaluation schemes to appropriately account for the roles that reputation and social context play in science. For example, our results can be used in support of the double-blind review system, which by reducing the role of reputation, is perceived to have advantages due to its objectivity and fairness (33). We conclude with a general note that the data deluge brought forth during the past decade is fueling extensive efforts in the computational social sciences (34) to identify and study the so-called "social atom" (35). Because our methodology is general, we speculate that other social networks characterized by trust and partial/asymmetric information are also based on similar reputation mechanisms. Indeed, it is likely that agent-based reputation mechanisms will play an increasing role in society due to the omnipresence of online recommender systems governed by reputation dynamics operating as a general diffusive contagion phenomena (36).

1. Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: Toward an objective measure of scientific impact. *Proc Natl Acad Sci USA* 105(45): 17268–17272.
2. Radicchi F, Fortunato S, Markines B, Vespignani A (2009) Diffusion of scientific credits and the ranking of scientists. *Phys Rev E Stat Nonlin Soft Matter Phys* 80(5 Pt 2):056103.
3. Petersen AM, Wang F, Stanley HE (2010) Methods for measuring the citations and productivity of scientists across time and discipline. *Phys Rev E Stat Nonlin Soft Matter Phys* 81(3 Pt 2):036114.
4. Petersen AM, Stanley HE, Succi S (2011) Statistical regularities in the rank-citation profile of scientists. *Sci Rep* 1:181.
5. Azoulay P, Zivin JSG, Wang J (2010) Superstar extinction. *Q J Econ* 125(2):549–589.
6. Petersen AM, Riccaboni M, Stanley HE, Pammolli F (2012) Persistence and uncertainty in the academic career. *Proc Natl Acad Sci USA* 109(14):5213–5218.
7. Merton RK (1968) The Matthew effect in science. *Science* 159(3810):56–63.
8. De Solla Price D (1976) A general theory of bibliometric and other cumulative advantage processes. *J Am Soc Inf Sci* 27(8):292–306.
9. Petersen AM, Jung WS, Yang JS, Stanley HE (2011) Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *Proc Natl Acad Sci USA* 108(1):18–23.
10. Peterson GJ, Pressé S, Dill KA (2010) Nonuniversal power law scaling in the probability distribution of scientific citations. *Proc Natl Acad Sci USA* 107(37):16023–16027.
11. Eom Y-H, Fortunato S (2011) Characterizing and modeling citation dynamics. *PLoS One* 6(9):e24926.
12. Medo M, Cimini G, Gualdi S (2011) Temporal effects in the growth of networks. *Phys Rev Lett* 107(23):238701.
13. Golosovsky M, Solomon S (2012) Stochastic dynamical model of a growing citation network based on a self-exciting point process. *Phys Rev Lett* 109(9):098701.
14. Stephan PE (1996) The economics of science. *J Econ Lit* 34(3):1199–1235.
15. Stephan PE (2012) *How Economics Shapes Science* (Harvard Univ Press, Cambridge, MA).
16. Penner O, Petersen AM, Pan RK, Fortunato S (2013) The case for caution in predicting scientists' future impact. *Phys Today* 66(4):8–9.
17. Penner O, Pan RK, Petersen AM, Kaski K, Fortunato S (2013) On the predictability of future impact in science. *Sci Rep* 3:3052.
18. Duch J, et al. (2012) The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PLoS One* 7(12):e51332.
19. David PA (2008) 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.
20. Ductor L, Fafchamps M, Goyal S, van der Leij MJ Social networks and research output. *Rev Econ Stat*, 10.1162/REST_a_00430.
21. Shane S, Cable S (2002) Network ties, reputation, and the financing of new ventures. *Manage Sci* 48(3):364–381.
22. Malmgren RD, Ottino JM, Nunes Amaral LA (2010) The role of mentorship in protégé performance. *Nature* 465(7298):622–626.
23. 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.
24. Newman MEJ (2009) The first-mover advantage in scientific publication. *Europhys Lett* 86(6):68001.
25. Burt RS (1992) *Structural Holes* (Harvard Univ Press, Cambridge, MA).
26. Nahapiet J, Ghoshal S (1998) Social capital, intellectual capital, and the organizational advantage. *Acad Manage Rev* 23(2):242–266.
27. Börner K, Maru JT, Goldstone RL (2004) The simultaneous evolution of author and paper networks. *Proc Natl Acad Sci USA* 101(Suppl 1):5266–5273.
28. Guimerà R, Uzzi B, Spiro J, Amaral LAN (2005) Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308(5722):697–702.
29. Palla G, Barabási AL, Vicsek T (2007) Quantifying social group evolution. *Nature* 446(7136):664–667.
30. Lu SF, Jin GZ, Uzzi B, Jones B (2013) The retraction penalty: Evidence from the Web of Science. *Sci Rep* 3:3146.
31. Wutchy S, Jones BF, Uzzi B (2008) Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* 322(5905):1259–1262.
32. Hellsten I, et al. (2007) Self-citations, co-authorships and keywords: A new approach to scientists' field mobility? *Scientometrics* 72:469–486.
33. Ware M (2008) *Peer Review: Benefits, Perceptions and Alternatives*, PRC Summary Papers 4 (Publishing Research Consortium, London), December, pp 1–20.
34. Lazer D, et al. (2009) Life in the network: The coming age of computational social science. *Science* 323(5915):721–723.
35. Buchanan M (2007) *The Social Atom* (Bloomsbury Publishing, New York).
36. Vespignani A (2012) Modelling dynamical processes in complex socio-technical systems. *Nat Phys* 8(1):32–39.

www.manaraa.com