

REVIEW SUMMARY

SCIENCE COMMUNITY

Science of science

Santo Fortunato,* Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, Alessandro Vespignani, Ludo Waltman, Dashun Wang, Albert-László Barabási*

BACKGROUND: The increasing availability of digital data on scholarly inputs and outputs—from research funding, productivity, and collaboration to paper citations and scientist mobility—offers unprecedented opportunities to explore the structure and evolution of science. The science of science (SciSci) offers a quantitative understanding of the interactions among scientific agents across diverse geographic and temporal scales: It provides insights into the conditions underlying creativity and the genesis of scientific discovery, with the ultimate goal of developing tools and policies that have the potential to accelerate science. In the past decade, SciSci has benefited from an influx of natural, computational, and social scientists who together have developed big data–based capabilities for empirical analysis and generative modeling that capture the unfolding of science, its institutions, and its workforce. The value proposition of SciSci is that with a deeper understanding of the factors that drive successful science, we can more effectively address environmental, societal, and technological problems.

ADVANCES: Science can be described as a complex, self-organizing, and evolving network of scholars, projects, papers, and ideas. This representation has unveiled patterns characterizing the emergence of new scientific fields through the study of collaboration networks and the path of impactful discoveries through the study of citation networks. Microscopic models have traced the dynamics of citation accumulation, allowing us to predict the future impact of individual papers. SciSci has revealed choices and trade-offs that scientists face as they advance both their own careers and the scientific horizon. For example, measurements indicate that scholars are risk-averse, preferring to study topics related to their current expertise, which constrains the potential of future discoveries. Those willing to break this pattern engage in riskier careers but become more likely to make major breakthroughs. Overall, the highest-impact science is grounded in conventional combinations of prior work but features unusual combinations. Last, as the locus of research is shifting into teams, SciSci is increasingly focused on

the impact of team research, finding that small teams tend to disrupt science and technology with new ideas drawing on older and less prevalent ones. In contrast, large teams tend to develop recent, popular ideas, obtaining high, but often short-lived, impact.

OUTLOOK: SciSci offers a deep quantitative understanding of the relational structure between scientists, institutions, and ideas because it facilitates the identification of fundamental mechanisms responsible for scientific discovery. These interdisciplinary data-driven efforts complement contributions from related fields such as scientometrics and the economics and sociology of

ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aao0185>

science. Although SciSci seeks long-standing universal laws and mechanisms that apply across various fields of science, a fundamental challenge going forward is accounting for undeniable differences in culture, habits, and preferences between different fields and countries. This variation makes some cross-domain insights difficult to appreciate and associated science policies difficult to implement. The differences among the questions, data, and skills specific to each discipline suggest that further insights can be gained from domain-specific SciSci studies, which model and identify opportunities adapted to the needs of individual research fields. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: santo@indiana.edu (S.F.); barabasi@gmail.com (A.-L.B.)
Cite this article as S. Fortunato *et al.*, *Science* 359, eaao0185 (2018). DOI: 10.1126/science.aao0185

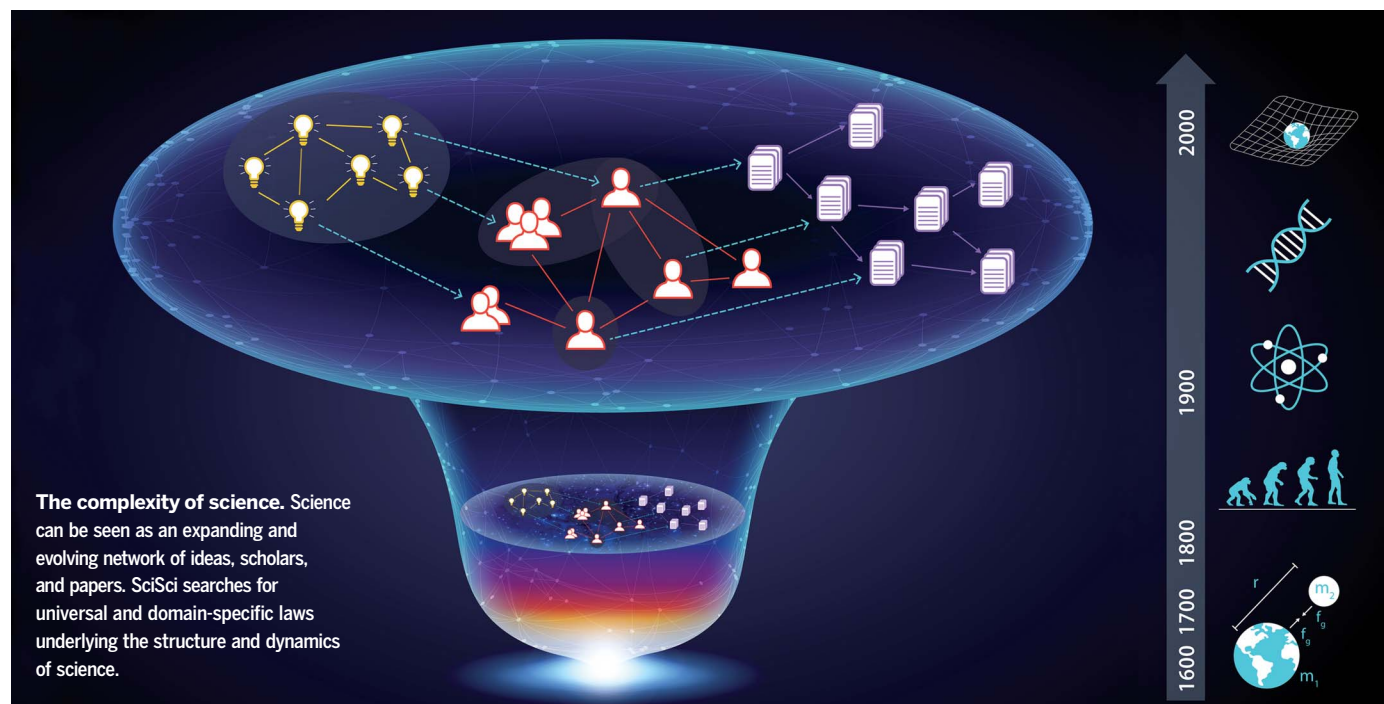


ILLUSTRATION: NICOLE SAMAY

REVIEW

SCIENCE COMMUNITY

Science of science

Santo Fortunato,^{1,2*} Carl T. Bergstrom,³ Katy Börner,^{2,4} James A. Evans,⁵ Dirk Helbing,⁶ Staša Milojević,¹ Alexander M. Petersen,⁷ Filippo Radicchi,¹ Roberta Sinatra,^{8,9,10} Brian Uzzi,^{11,12} Alessandro Vespignani,^{10,13,14} Ludo Waltman,¹⁵ Dashun Wang,^{11,12} Albert-László Barabási^{8,10,16*}

Identifying fundamental drivers of science and developing predictive models to capture its evolution are instrumental for the design of policies that can improve the scientific enterprise—for example, through enhanced career paths for scientists, better performance evaluation for organizations hosting research, discovery of novel effective funding vehicles, and even identification of promising regions along the scientific frontier. The science of science uses large-scale data on the production of science to search for universal and domain-specific patterns. Here, we review recent developments in this transdisciplinary field.

The deluge of digital data on scholarly output offers unprecedented opportunities to explore patterns characterizing the structure and evolution of science. The science of science (SciSci) places the practice of science itself under the microscope, leading to a quantitative understanding of the genesis of scientific discovery, creativity, and practice and developing tools and policies aimed at accelerating scientific progress.

The emergence of SciSci has been driven by two key factors. The first is data availability. In addition to the proprietary Web of Science (WoS), the historic first citation index (*I*), multiple data sources are available today (Scopus, PubMed, Google Scholar, Microsoft Academic, the U.S. Patent and Trademark Office, and others). Some of these sources are freely accessible, covering

millions of data points pertaining to scientists and their output and capturing research from all over the world and all branches of science. Second, SciSci has benefited from an influx of and collaborations among natural, computational, and social scientists who have developed big data-based capabilities and enabled critical tests of generative models that aim to capture the unfolding of science, its institutions, and its workforce.

One distinctive characteristic of this emerging field is how it breaks down disciplinary boundaries. SciSci integrates findings and theories from multiple disciplines and uses a wide range of data and methods. From scientometrics, it takes the idea of measuring science from large-scale data sources; from the sociology of science, it adopts theoretical concepts and social processes;

and from innovation studies, it explores and identifies pathways through which science contributes to invention and economic change. SciSci relies on a broad collection of quantitative methods, from descriptive statistics and data visualization to advanced econometric methods, network science approaches, machine-learning algorithms, mathematical analysis, and computer simulation, including agent-based modeling. The value proposition of SciSci hinges on the hypothesis that with a deeper understanding of the factors behind successful science, we can enhance the prospects of science as a whole to more effectively address societal problems.

Networks of scientists, institutions, and ideas

Contemporary science is a dynamical system of undertakings driven by complex interactions among social structures, knowledge representations, and the natural world. Scientific knowledge is constituted by concepts and relations embodied in research papers, books, patents, software, and other scholarly artifacts, organized into scientific disciplines and broader fields. These social, conceptual, and material elements are connected through formal and informal flows of information, ideas, research practices, tools, and samples. Science can thus be described as a complex, self-organizing, and constantly evolving multiscale network.

Early studies discovered an exponential growth in the volume of scientific literature (2), a trend that continues with an average doubling period of 15 years (Fig. 1). Yet, it would be naïve to equate the growth of the scientific literature with the growth of scientific ideas. Changes in the publishing world, both technological and economic, have led to increasing efficiency in the production of publications. Moreover, new publications in science tend to cluster in discrete areas of knowledge (3). Large-scale text analysis,

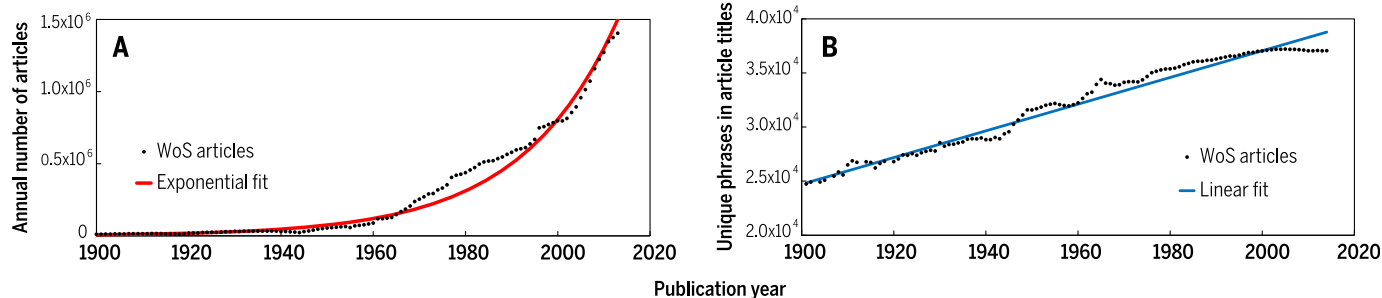


Fig. 1. Growth of science. (A) Annual production of scientific articles indexed in the WoS database. (B) Growth of ideas covered by articles indexed in the WoS. This was determined by counting unique title phrases (concepts) in a fixed number of articles (4).

¹Center for Complex Networks and Systems Research, School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408, USA. ²Indiana University Network Science Institute, Indiana University, Bloomington, IN 47408, USA. ³Department of Biology, University of Washington, Seattle, WA 98195-1800, USA. ⁴Cyberinfrastructure for Network Science Center, School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408, USA. ⁵Department of Sociology, University of Chicago, Chicago, IL 60637, USA. ⁶Computational Social Science, ETH Zurich, Zurich, Switzerland. ⁷Ernest and Julio Gallo Management Program, School of Engineering, University of California, Merced, CA 95343, USA. ⁸Center for Network Science, Central European University, Budapest 1052, Hungary. ⁹Department of Mathematics, Central European University, Budapest 1051, Hungary. ¹⁰Institute for Network Science, Northeastern University, Boston, MA 02115, USA. ¹¹Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA. ¹²Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208, USA. ¹³Laboratory for the Modeling of Biological and Sociotechnical Systems, Northeastern University, Boston, MA 02115, USA. ¹⁴ISI Foundation, Turin 10133, Italy. ¹⁵Centre for Science and Technology Studies, Leiden University, Leiden, Netherlands. ¹⁶Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA.

*Corresponding author. Email: santo@indiana.edu (S.F.); barabasi@gmail.com (A.-L.B.)

using phrases extracted from titles and abstracts to measure the cognitive extent of the scientific literature, have found that the conceptual territory of science expands linearly with time. In other words, whereas the number of publications grows exponentially, the space of ideas expands only linearly (Fig. 1) (4).

Frequently occurring words and phrases in article titles and abstracts propagate via citation networks, punctuated by bursts corresponding to the emergence of new paradigms (5). By applying network science methods to citation networks, researchers are able to identify communities as defined by subsets of publications that frequently cite one another (6). These communities often correspond to groups of authors holding a common position regarding specific issues (7) or working on the same specialized subtopics (8). Recent work focusing on biomedical science has illustrated how the growth of the literature reinforces these communities (9). As new papers are published, associations (hyperedges) between scientists, chemicals, diseases, and methods (“things,” which are the nodes of the network) are added. Most new links fall between things only one or two steps away from each other, implying that when scientists choose new topics, they prefer things directly related to their current expertise or that of their collaborators. This densification suggests that the existing structure of science may constrain what will be studied in the future.

Densification at the boundaries of science is also a signal of transdisciplinary exploration, fusion, and innovation. A life-cycle analysis of eight fields (10) shows that successful fields undergo a process of knowledge and social unification that leads to a giant connected component in the collaboration network (104), corresponding to a sizeable group of regular coauthors. A model in which scientists choose their collaborators through random walks on the coauthorship network successfully reproduces author productivity, the number of authors per discipline, and the interdisciplinarity of papers and authors (11).

Problem selection

How do scientists decide which research problems to work on? Sociologists of science have long hypothesized that these choices are shaped by an ongoing tension between productive tradition and risky innovation (12, 13). Scientists who adhere to a research tradition in their domain often appear productive by publishing a steady stream of contributions that advance a focused research agenda. But a focused agenda may limit a researcher’s ability to sense and seize opportunities for staking out new ideas that are required to grow the field’s knowledge. For example, a case study focusing on biomedical scientists choosing novel chemicals and chemical relationships shows that as fields mature, researchers tend to focus increasingly on established knowledge (3). Although an innovative publication tends to result in higher impact than a conservative one, high-risk innovation strategies are rare, because the additional reward does not compensate for

the risk of failure to publish at all. Scientific awards and accolades appear to function as primary incentives to resist conservative tendencies and encourage betting on exploration and surprise (3). Despite the many factors shaping what scientists work on next, macroscopic patterns that govern changes in research interests along scientific careers are highly reproducible, documenting a high degree of regularity underlying scientific research and individual careers (14).

Scientists’ choice of research problems affects primarily their individual careers and the careers of those reliant on them. Scientists’ collective choices, however, determine the direction of scientific discovery more broadly (Fig. 2). Conservative strategies (15) serve individual careers well but are less effective for science as a whole. Such strategies are amplified by the file drawer problem (16): Negative results, at odds with established hypotheses, are rarely published, leading to a systemic bias in published research and the canonization of weak and sometimes false facts (17). More risky hypotheses may have been tested by generations of scientists, but only those successful enough to result in publications are known to us. One way to alleviate this conservative trap is to urge funding agencies to proactively sponsor risky projects that test truly unexplored hypotheses and take on special interest groups advocating for particular diseases.

Measurements show that the allocation of biomedical resources in the United States is more strongly correlated to previous allocations and research than to the actual burden of diseases (18), highlighting a systemic misalignment between biomedical needs and resources. This misalignment casts doubts on the degree to which funding agencies, often run by scientists embedded in established paradigms, are likely to influence the evolution of science without introducing additional oversight, incentives, and feedback.

Novelty

Analyses of publications and patents consistently reveal that rare combinations in scientific discoveries and inventions tend to garner higher citation rates (3). Interdisciplinary research is an emblematic recombinant process (19); hence, the successful combination of previously disconnected ideas and resources that is fundamental to interdisciplinary research often violates expectations and leads to novel ideas with high impact (20). Nevertheless, evidence from grant applications shows that, when faced with new ideas, expert evaluators systematically give lower scores to truly novel (21–23) or interdisciplinary (24) research proposals.

The highest-impact science is primarily grounded in conventional combinations of prior work, yet it simultaneously features unusual combinations

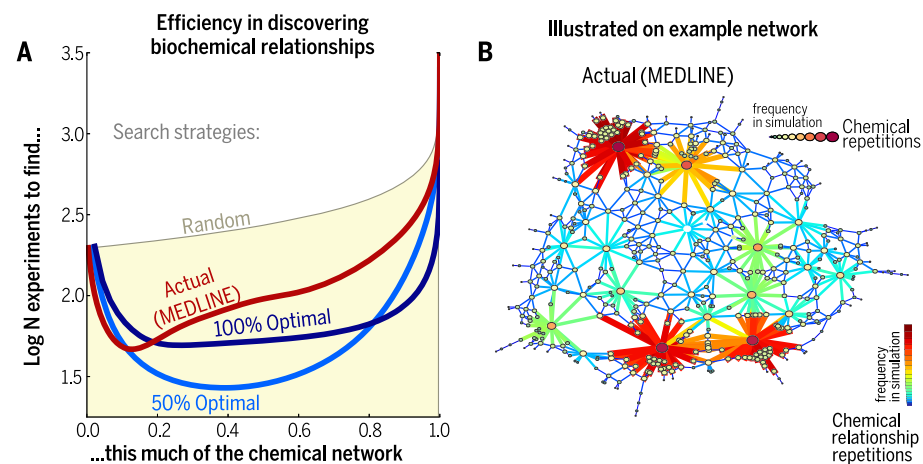


Fig. 2. Choosing experiments to accelerate collective discovery. (A) The average efficiency rate for global strategies to discover new, publishable chemical relationships, estimated from all MEDLINE-indexed articles published in 2010. This model does not take into account differences in the difficulty or expense of particular experiments. The efficiency of a global scientific strategy is expressed by the average number of experiments performed (vertical axis) relative to the number of new, published biochemical relationships (horizontal axis), which correspond to new connections in the published network of biochemicals co-occurring in MEDLINE-indexed articles. Compared strategies include randomly choosing pairs of biochemicals, the global (“actual”) strategy inferred from all scientists publishing MEDLINE articles, and optimal strategies for discovering 50 and 100% of the network. Lower values on the vertical axis indicate more efficient strategies, showing that the actual strategy of science is suboptimal for discovering what has been published. The actual strategy is best for uncovering 13% of the chemical network, and the 50% optimal strategy is most efficient for discovering 50% of it, but neither are as good as the 100% optimal strategy for revealing the whole network. (B) The actual, estimated search process illustrated on a hypothetical network of chemical relationships, averaged from 500 simulated runs of that strategy. The strategy swarms around a few “important,” highly connected chemicals, whereas optimal strategies are much more even and less likely to “follow the crowd” in their search across the space of scientific possibilities. [Adapted from (15)]

(25–27). Papers of this type are twice as likely to receive high citations (26). In other words, a balanced mixture of new and established elements is the safest path toward successful reception of scientific advances.

Career dynamics

Individual academic careers unfold in the context of a vast market for knowledge production and consumption (28). Consequently, scientific careers have been examined not only in terms of individual incentives and marginal productivity (i.e., relative gain versus effort) (29), but also institutional incentives (30, 31) and competition (32). This requires combining large repositories of high-resolution individual, geographic, and temporal metadata (33) to construct representations of career trajectories that can be analyzed from different perspectives. For example, one study finds that funding schemes that are tolerant of early failure, which reward long-term success, are more likely to generate high-impact publications than grants subject to short review cycles (31). Interacting systems with competing time scales are a classic problem in complex systems science. The multifaceted nature of science is motivation for generative models that highlight unintended consequences of policies. For example, models of career growth show that non-tenure (short-term) contracts are responsible for productivity fluctuations, which often result in a sudden career death (29).

Gender inequality in science remains prevalent and problematic (34). Women have fewer publications (35–37) and collaborators (38) and less funding (39), and they are penalized in hiring decisions when compared with equally qualified men (40). The causes of these gaps are still unclear. Intrinsic differences in productivity rates and career length can explain the differences in collaboration patterns (38) and hiring rates (35) between male and female scientists. On the other hand, experimental evidence shows that biases against women occur at very early career stages. When gender was randomly assigned among the curricula vitae of a pool of applicants, the hiring committee systematically penalized female candidates (40). Most studies so far have focused on relatively small samples. Improvements in compiling large-scale data sets on scientific careers, which leverage information from different sources (e.g., publication records, grant applications, and awards), will help us gain deeper insight into the causes of inequality and motivate models that can inform policy solutions.

Scientists' mobility is another important factor offering diverse career opportunities. Most mobility studies have focused on quantifying the brain drain and gain of a country or a region (41, 42), especially after policy changes. Research on individual mobility and its career effect remains scant, however, primarily owing to the difficulty of obtaining longitudinal information about the movements of many scientists and accounts of the reasons underlying mobility decisions. Scientists who left their country of origin outperformed scientists who did not relocate,

according to their citation scores, which may be rooted in a selection bias that offers better career opportunities to better scientists (43, 44). Moreover, scientists tend to move between institutions of similar prestige (45). Nevertheless, when examining changes in impact associated with each move as quantified by citations, no systematic increase or decrease was found, not even when scientists moved to an institution of considerably higher or lower rank (46). In other words, it is not the institution that creates the impact; it is the individual researchers that make an institution.

Another potentially important career factor is reputation—and the dilemma that it poses for manuscript review, proposal evaluation, and promotion decisions. The reputation of paper authors, measured by the total citations of their

previous output, markedly boosts the number of citations collected by that paper in the first years after publication (47). After this initial phase, however, impact depends on the reception of the work by the scientific community. This finding, along with the work reported in (46), suggests that, for productive scientific careers, reputation is less of a critical driver for success than talent, hard work, and relevance.

A policy-relevant question is whether creativity and innovation depend on age or career stage. Decades of research on outstanding researchers and innovators concluded that major breakthroughs take place relatively early in a career, with a median age of 35 (48). In contrast, recent work shows that this well-documented propensity of early-career discoveries is fully explained by productivity, which is high in the early stages of a scientist's career and drops later (49). In other words, there are no age patterns in innovation: A scholar's most cited paper can be any of his or her papers, independently of the age or career stage when it is published (Fig. 3). A stochastic model of impact evolution also indicates that breakthroughs result from a combination of the ability of a scientist and the luck of picking a problem with high potential (49).

Team science

During past decades, reliance on teamwork has increased, representing a fundamental shift in the way that science is done. A study of the authorship of 19.9 million research articles and 2.1 million patents reveals a nearly universal shift toward teams in all branches of science (50) (Fig. 4). For example, in 1955, science and engineering teams authored about the same number of papers as single authors. Yet by 2013, the fraction of team-authored papers increased to 90% (51).

Nowadays, a team-authored paper in science and engineering is 6.3 times more likely to receive 1000 citations or more than a solo-authored paper, a difference that cannot be explained by self-citations (50, 52). One possible reason is a team's ability to come up with more novel combinations of ideas (26) or to produce resources that are later used by others (e.g., genomics). Measurements show that teams are 38% more likely than solo authors to insert novel combinations into familiar knowledge domains, supporting the premise that teams can bring together different specialties, effectively combining knowledge to prompt scientific breakthroughs. Having more collaborations means greater visibility through a larger number of coauthors, who will likely introduce the work to their networks, an enhanced impact that may partially compensate for the fact that credit within a team must be shared with many colleagues (29).

Work from large teams garners, on average, more citations across a wide variety of domains. Research suggests that small teams tend to disrupt science and technology with new ideas and opportunities, whereas large teams develop existing ones (53). Thus, it may be important to

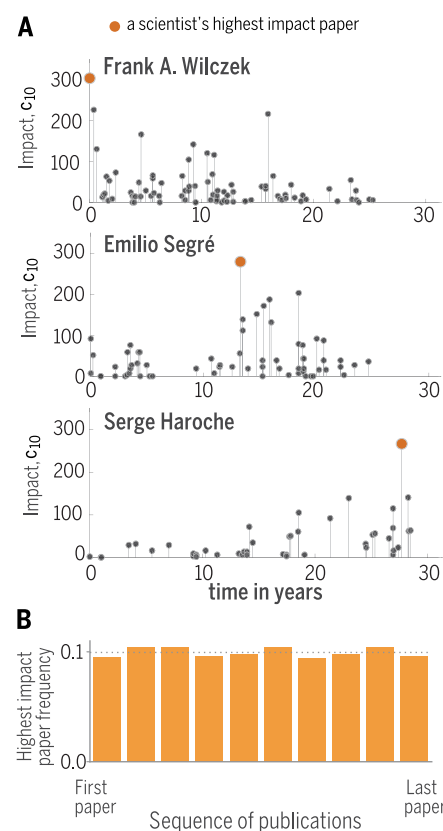


Fig. 3. Impact in scientific careers. (A) Publication record of three Nobel laureates in physics.

The horizontal axis indicates the number of years after a laureate's first publication, each circle corresponds to a research paper, and the height of the circle represents the paper's impact, quantified by C_{10} , the number of citations after 10 years. The highest-impact paper of a laureate is denoted with an orange circle.

(B) Histogram of the occurrence of the highest-impact paper in a scientist's sequence of publications, calculated for 10,000 scientists. The flatness of the histogram indicates that the highest-impact work can be, with the same probability, anywhere in the sequence of papers published by a scientist (49).

fund and foster teams of all sizes to temper the bureaucratization of science (28).

Teams are growing in size, increasing by an average of 17% per decade (50, 54, 105), a trend underlying a fundamental change in team compositions. Scientific teams include both small, stable “core” teams and large, dynamically changing extended teams (55). The increasing team size in most fields is driven by faster expansion of extended teams, which begin as small core teams but subsequently attract new members through a process of cumulative advantage anchored by productivity. Size is a crucial determinant of team survival strategies: Small teams survive longer if they maintain a stable core, but larger teams persist longer if they manifest a mechanism for membership turnover (56).

As science has accelerated and grown increasingly complex, the instruments required to expand the frontier of knowledge have increased in scale and precision. The tools of the trade become unaffordable to most individual investigators, but also to most institutions. Collaboration has been a critical solution, pooling resources to scientific advantage. The Large Hadron Collider at CERN, the world’s largest and most powerful particle collider, would have been unthinkable without collaboration, requiring more than 10,000 scientists and engineers from more than 100 countries. There is, however, a trade-off with increasing size that affects the value and risk associated with “big science” (2). Although it may be possible to solve larger problems, the burden of reproducibility may require duplicating initial efforts, which may not be practically or economically feasible.

Collaborators can have a large effect on scientific careers. According to recent studies (57, 58), scientists who lose their star collaborators experience a substantial drop in their productivity, especially if the lost collaborator was a regular coauthor. Publications involving extremely strong collaborators gain 17% more citations on average, pointing to the value of career partnership (59).

Given the increasing number of authors on the average research paper, who should and does gain the most credit? The canonical theory of credit (mis)allocation in science is the Matthew effect (60), in which scientists of higher statuses involved in joint work receive outsized credit for their contributions. Properly allocating individual credit for a collaborative work is difficult because we cannot easily distinguish individual contributions (61). It is possible, however, to inspect the citation patterns of the coauthors’ publications to determine the fraction of credit that the community assigns to each coauthor in a publication (62).

Citation dynamics

Scholarly citation remains the dominant measurable unit of credit in science. Given the reliance of most impact metrics on citations (63–66), the dynamics of citation accumulation have been scrutinized by generations of scholars. From foundational work by Price (67), we know that the distribution of citations for scientific papers is highly skewed: Many papers are never cited, but

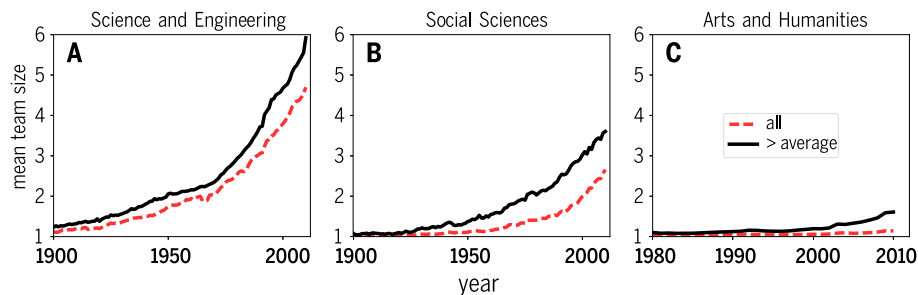


Fig. 4. Size and impact of teams. Mean team size has been steadily growing over the past century. The red dashed curves represent the mean number of coauthors over all papers; the black curves consider just those papers receiving more citations than the average for the field. Black curves are systematically above the dashed red ones, meaning that high-impact work is more likely to be produced by large teams than by small ones. Each panel corresponds to one of the three main disciplinary groups of papers indexed in the WoS: (A) science and engineering, (B) social sciences, and (C) arts and humanities.

seminal papers can accumulate 10,000 or more citations. This uneven citation distribution is a robust, emergent property of the dynamics of science, and it holds when papers are grouped by institution (68). If the number of citations of a paper is divided by the average number of citations collected by papers in the same discipline and year, the distribution of the resulting score is essentially indistinguishable for all disciplines (69, 70) (Fig. 5A). This means that we can compare the impact of papers published in different disciplines by looking at their relative citation values. For example, a paper in mathematics collecting 100 citations represents a higher disciplinary impact than a paper in microbiology with 300 citations.

The tail of the citation distribution, capturing the number of high-impact papers, sheds light on the mechanisms that drive the accumulation of citations. Recent analyses show that it follows a power law (71–73). Power-law tails can be generated through a cumulative advantage process (74), known as preferential attachment in network science (75), suggesting that the probability of citing a paper grows with the number of citations that it has already collected. Such a model can be augmented with other characteristic

features of citation dynamics, such as the obsolescence of knowledge, decreasing the citation probability with the age of the paper (76–79, 106), and a fitness parameter, unique to each paper, capturing the appeal of the work to the scientific community (77, 78). Only a tiny fraction of papers deviate from the pattern described by such a model—some of which are called “sleeping beauties,” because they receive very little notice for decades after publication and then suddenly receive a burst of attention and citations (80, 81).

The generative mechanisms described above can be used to predict the citation dynamics of individual papers. One predictive model (77) assumes that the citation probability of a paper depends on the number of previous citations, an obsolescence factor, and a fitness parameter (Fig. 5, B and C). For a given paper, one can estimate the three model parameters by fitting the model to the initial portion of the citation history of the paper. The long-term impact of the work can be extrapolated (77). Other studies have identified predictors of the citation impact of individual papers (82), such as journal impact factor (72). It has been suggested that the future *h*-index (83) of a scientist can be accurately predicted (84), although the predictive power is reduced when

Box 1. Lessons from SciSci.

- 1. Innovation and tradition:** Left bare, truly innovative and highly interdisciplinary ideas may not reach maximum scientific impact. To enhance their impact, novel ideas should be placed in the context of established knowledge (26).
- 2. Persistence:** A scientist is never too old to make a major discovery, as long as he or she stays productive (49).
- 3. Collaboration:** Research is shifting to teams, so engaging in collaboration is beneficial. Works by small teams tend to be more disruptive, whereas those by big teams tend to have more impact (4, 50, 53).
- 4. Credit:** Most credit will go to the coauthors with the most consistent track record in the domain of the publication (62).
- 5. Funding:** Although review panels acknowledge innovation, they ultimately tend to discount it. Funding agencies should ask reviewers to assess innovation, not only expected success (24).

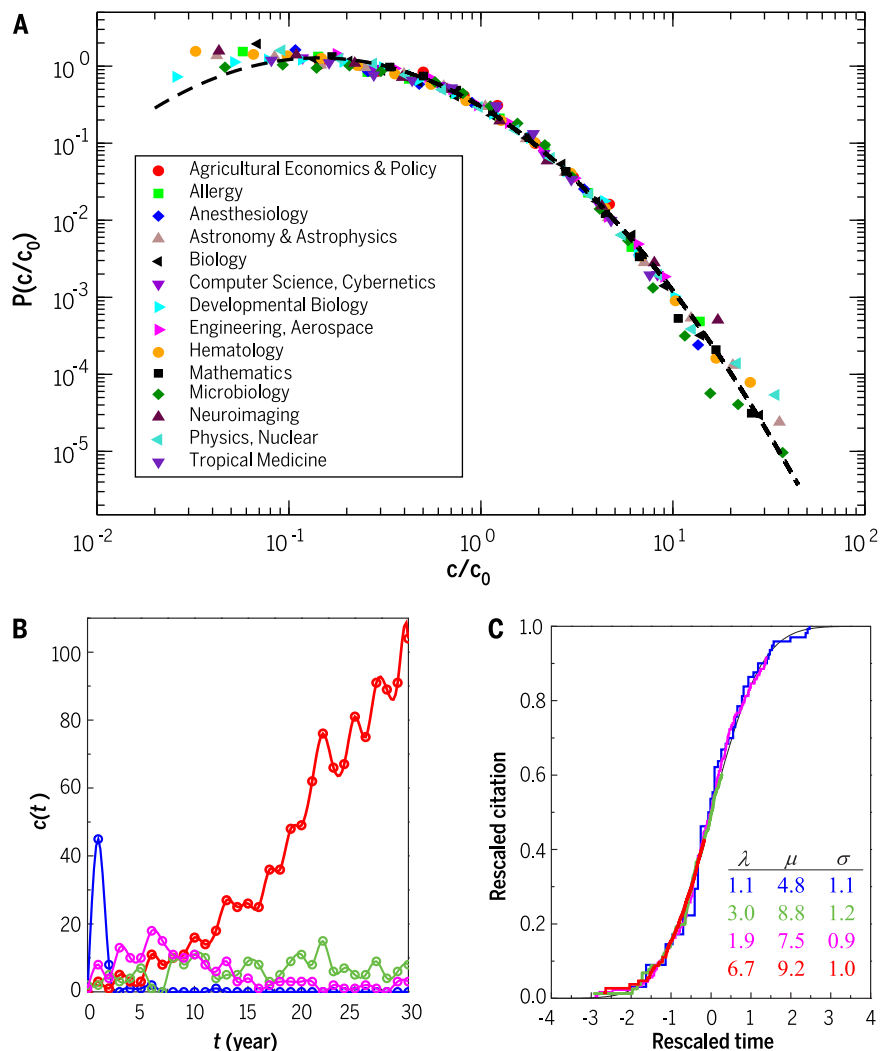


Fig. 5. Universality in citation dynamics. (A) The citation distributions of papers published in the same discipline and year lie on the same curve for most disciplines, if the raw number of citations c of each paper is divided by the average number of citations c_0 over all papers in that discipline and year. The dashed line is a lognormal fit. [Adapted from (69)] (B) Citation history of four papers published in *Physical Review* in 1964, selected for their distinct dynamics, displaying a “jump-decay” pattern (blue), experiencing a delayed peak (magenta), attracting a constant number of citations over time (green), or acquiring an increasing number of citations each year (red). (C) Citations of an individual paper are determined by three parameters: fitness λ , immediacy μ , and longevity σ . By rescaling the citation history of each paper in (B) by the appropriate (λ, μ, σ) parameters, the four papers collapse onto a single universal function, which is the same for all disciplines. [Adapted from (77)]

accounting for the scientist’s career stage and the cumulative, nondecreasing nature of the h -index (85). Eliminating inconsistencies in the use of quantitative evaluation metrics in science is crucial and highlights the importance of understanding the generating mechanisms behind commonly used statistics.

Outlook

Despite the discovery of universals across science, substantial disciplinary differences in culture, habits, and preferences make some cross-domain insights difficult to appreciate within particular fields and associated policies challenging to im-

plement. The differences among the questions, data, and skills required by each discipline suggest that we may gain further insights from domain-specific SciSci studies that model and predict opportunities adapted to the needs of each field. For young scientists, the results of SciSci offer actionable insights about past patterns, helping guide future inquiry within their disciplines (Box 1).

The contribution of SciSci is a detailed understanding of the relational structure between scientists, institutions, and ideas, a crucial starting point that facilitates the identification of fundamental generating processes. Together, these data-driven efforts complement contributions from

related research domains such as the economics (30) and sociology of science (60, 86). Causal estimation is a prime example, in which econometric matching techniques demand and leverage comprehensive data sources in the effort to simulate counterfactual scenarios (31, 42). Assessing causality is one of the most needed future developments in SciSci: Many descriptive studies reveal strong associations between structure and outcomes, but the extent to which a specific structure “causes” an outcome remains unexplored. Engaging in tighter partnerships with experimentalists, SciSci will be able to better identify associations discovered from models and large-scale data that have causal force to enrich their policy relevance. But experimenting on science may be the biggest challenge SciSci has yet to face. Running randomized, controlled trials that can alter outcomes for individuals or institutions of science, which are mostly supported by tax dollars, is bound to elicit criticisms and pushback (87). Hence, we expect quasi-experimental approaches to prevail in SciSci investigations in the near future.

Most SciSci research focuses on publications as primary data sources, implying that insights and findings are limited to ideas successful enough to merit publication in the first place. Yet most scientific attempts fail, sometimes spectacularly. Given that scientists fail more often than they succeed, knowing when, why, and how an idea fails is essential in our attempts to understand and improve science. Such studies could provide meaningful guidance regarding the reproducibility crisis and help us account for the file drawer problem. They could also substantially further our understanding of human imagination by revealing the total pipeline of creative activity.

Science often behaves like an economic system with a one-dimensional “currency” of citation counts. This creates a hierarchical system, in which the “rich-get-richer” dynamics suppress the spread of new ideas, particularly those from junior scientists and those who do not fit within the paradigms supported by specific fields. Science can be improved by broadening the number and range of performance indicators. The development of alternative metrics covering web (88) and social media (89) activity and societal impact (90) is critical in this regard. Other measurable dimensions include the information (e.g., data) that scientists share with competitors (91), the help that they offer to their peers (92), and their reliability as reviewers of their peers’ works (93). But with a profusion of metrics, more work is needed to understand what each of them does and does not capture to ensure meaningful interpretation and avoid misuse. SciSci can make an essential contribution by providing models that offer a deeper understanding of the mechanisms that govern performance indicators in science. For instance, models of the empirical patterns observed when alternative indicators (e.g., distributions of paper downloads) are used will enable us to explore their relationship with citation-based metrics (94) and to recognize manipulations.

- Rev. Econ. Stud. **76**, 283–317 (2009). doi: [10.1111/j.1467-937X.2008.00531.x](https://doi.org/10.1111/j.1467-937X.2008.00531.x)
55. S. Milojević, Principles of scientific research team formation and evolution. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3984–3989 (2014). doi: [10.1073/pnas.1309723111](https://doi.org/10.1073/pnas.1309723111); pmid: 24591626
 56. G. Palla, A.-L. Barabási, T. Vicsek, Quantifying social group evolution. *Nature* **446**, 664–667 (2007). doi: [10.1038/nature05670](https://doi.org/10.1038/nature05670); pmid: 17410175
 57. G. J. Borjas, K. B. Doran, Which peers matter? The relative impacts of collaborators, colleagues, and competitors. *Rev. Econ. Stat.* **97**, 1104–1117 (2015). doi: [10.1162/REST_a_00472](https://doi.org/10.1162/REST_a_00472)
 58. P. Azoulay, J. G. Zivin, J. Wang, Superstar extinction. *Q. J. Econ.* **125**, 549–589 (2010). doi: [10.1162/qjec.2010.125.2.549](https://doi.org/10.1162/qjec.2010.125.2.549)
 59. A. M. Petersen, Quantifying the impact of weak, strong, and super ties in scientific careers. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4671–E4680 (2015). doi: [10.1073/pnas.1501444112](https://doi.org/10.1073/pnas.1501444112); pmid: 26261301
 60. R. K. Merton, The Matthew effect in science. *Science* **159**, 56–63 (1968). doi: [10.1126/science.159.3810.56](https://doi.org/10.1126/science.159.3810.56)
 61. L. Allen, J. Scott, A. Brand, M. Hlava, M. Altman, Publishing: Credit where credit is due. *Nature* **508**, 312–313 (2014). doi: [10.1038/508312a](https://doi.org/10.1038/508312a); pmid: 24745070
 62. H.-W. Shen, A.-L. Barabási, Collective credit allocation in science. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12325–12330 (2014). doi: [10.1073/pnas.1401992111](https://doi.org/10.1073/pnas.1401992111); pmid: 25114238
 63. L. Waltman, A review of the literature on citation impact indicators. *J. Informetr.* **10**, 365–391 (2016). doi: [10.1016/j.joi.2016.02.007](https://doi.org/10.1016/j.joi.2016.02.007)
 64. J. E. Hirsch, An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16569–16572 (2005). doi: [10.1073/pnas.0507655102](https://doi.org/10.1073/pnas.0507655102); pmid: 16275915
 65. H. F. Moed, *Citation Analysis in Research Evaluation* (Springer, 2010).
 66. E. Garfield, Citation analysis as a tool in journal evaluation. *Science* **178**, 471–479 (1972). doi: [10.1126/science.178.4060.471](https://doi.org/10.1126/science.178.4060.471); pmid: 5079701
 67. D. J. de Solla Price, Networks of scientific papers. *Science* **149**, 510–515 (1965). doi: [10.1126/science.149.3683.510](https://doi.org/10.1126/science.149.3683.510); pmid: 14325149
 68. Q. Zhang, N. Perra, B. Gonçalves, F. Ciulla, A. Vespignani, Characterizing scientific production and consumption in physics. *Sci. Rep.* **3**, 1640 (2013). doi: [10.1038/srep01640](https://doi.org/10.1038/srep01640); pmid: 23571320
 69. F. Radicchi, S. Fortunato, C. Castellano, Universality of citation distributions: Toward an objective measure of scientific impact. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17268–17272 (2008). doi: [10.1073/pnas.0806977105](https://doi.org/10.1073/pnas.0806977105); pmid: 18978030
 70. L. Waltman, N. J. van Eck, A. F. J. van Raan, Universality of citation distributions revisited. *J. Assoc. Inf. Sci. Technol.* **63**, 72–77 (2012). doi: [10.1002/asi.21671](https://doi.org/10.1002/asi.21671)
 71. M. Golosovsky, S. Solomon, Runaway events dominate the heavy tail of citation distributions. *Eur. Phys. J. Spec. Top.* **205**, 303–311 (2012). doi: [10.1140/epjst/e2012-01576-4](https://doi.org/10.1140/epjst/e2012-01576-4)
 72. C. Stegheuis, N. Litvak, L. Waltman, Predicting the long-term citation impact of recent publications. *J. Informetr.* **9**, 642–657 (2015). doi: [10.1016/j.joi.2015.06.005](https://doi.org/10.1016/j.joi.2015.06.005)
 73. M. Thelwall, The discretised lognormal and hooked power law distributions for complete citation data: Best options for modelling and regression. *J. Informetr.* **10**, 336–346 (2016). doi: [10.1016/j.joi.2015.12.007](https://doi.org/10.1016/j.joi.2015.12.007)
 74. D. de Solla Price, A general theory of bibliometric and other cumulative advantage processes. *J. Am. Soc. Inf. Sci.* **27**, 292–306 (1976). doi: [10.1002/asi.4630270505](https://doi.org/10.1002/asi.4630270505)
 75. A.-L. Barabási, R. Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999). doi: [10.1126/science.286.5439.509](https://doi.org/10.1126/science.286.5439.509); pmid: 10521342
 76. P. D. B. Parolo *et al.*, Attention decay in science. *J. Informetr.* **9**, 734–745 (2015). doi: [10.1016/j.joi.2015.07.006](https://doi.org/10.1016/j.joi.2015.07.006)
 77. D. Wang, C. Song, A.-L. Barabási, Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013). doi: [10.1126/science.1237825](https://doi.org/10.1126/science.1237825); pmid: 24092745
 78. Y.-H. Eom, S. Fortunato, Characterizing and modeling citation dynamics. *PLOS ONE* **6**, e24926 (2011). doi: [10.1371/journal.pone.0024926](https://doi.org/10.1371/journal.pone.0024926); pmid: 21966387
 79. M. Golosovsky, S. Solomon, Stochastic dynamical model of a growing citation network based on a self-exciting point process. *Phys. Rev. Lett.* **109**, 098701 (2012). doi: [10.1103/PhysRevLett.109.098701](https://doi.org/10.1103/PhysRevLett.109.098701); pmid: 23002894
 80. A. F. J. van Raan, Sleeping Beauties in science. *Scientometrics* **59**, 467–472 (2004). doi: [10.1023/B:SCIE.0000018543.82441.f1](https://doi.org/10.1023/B:SCIE.0000018543.82441.f1)
 81. Q. Ke, E. Ferrara, F. Radicchi, A. Flammini, Defining and identifying Sleeping Beauties in science. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7426–7431 (2015). doi: [10.1073/pnas.1424329112](https://doi.org/10.1073/pnas.1424329112); pmid: 26015563
 82. I. Tahamtan, A. Safiropour Afshar, K. Ahamdzadeh, Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics* **107**, 1195–1225 (2016). doi: [10.1007/s1192-016-1889-2](https://doi.org/10.1007/s1192-016-1889-2)
 83. J. E. Hirsch, Does the *h* index have predictive power? *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19193–19198 (2007). doi: [10.1073/pnas.0707962104](https://doi.org/10.1073/pnas.0707962104); pmid: 18040045
 84. D. E. Acuna, S. Allesina, K. P. Kording, Future impact: Predicting scientific success. *Nature* **489**, 201–202 (2012). doi: [10.1038/489201a](https://doi.org/10.1038/489201a); pmid: 22972278
 85. O. Penner, R. K. Pan, A. M. Petersen, K. Kaski, S. Fortunato, On the predictability of future impact in science. *Sci. Rep.* **3**, 3052 (2013). doi: [10.1038/srep03052](https://doi.org/10.1038/srep03052); pmid: 24165898
 86. J. R. Cole, H. Zuckerman, in *The Idea of Social Structure: Papers in Honor of Robert K. Merton*, L. A. Coser, Ed. (Harcourt Brace Jovanovich, 1975), pp. 139–174.
 87. P. Azoulay, Research efficiency: Turn the scientific method on ourselves. *Nature* **484**, 31–32 (2012). doi: [10.1038/484031a](https://doi.org/10.1038/484031a); pmid: 22481340
 88. M. Thelwall, K. Koussa, Web indicators for research evaluation. Part 1: Citations and links to academic articles from the Web. *Prof. Inf.* **24**, 587–606 (2015). doi: [10.3145/epi.2015.sep.08](https://doi.org/10.3145/epi.2015.sep.08)
 89. M. Thelwall, K. Koussa, Web indicators for research evaluation. Part 2: Social media metrics. *Prof. Inf.* **24**, 607–620 (2015). doi: [10.3145/epi.2015.sep.09](https://doi.org/10.3145/epi.2015.sep.09)
 90. L. Bornmann, What is societal impact of research and how can it be assessed? A literature survey. *Adv. Inf. Sci.* **64**, 217–233 (2013).
 91. C. Haeussler, L. Jiang, J. Thursby, M. Thursby, Specific and general information sharing among competing academic researchers. *Res. Policy* **43**, 465–475 (2014). doi: [10.1016/j.respol.2013.08.017](https://doi.org/10.1016/j.respol.2013.08.017)
 92. A. Oettl, Sociology: Honour the helpful. *Nature* **489**, 496–497 (2012). doi: [10.1038/489496a](https://doi.org/10.1038/489496a); pmid: 23018949
 93. S. Ravindran, “Getting credit for peer review.” *Science*, 8 February 2016; www.sciencemag.org/careers/2016/02/getting-credit-peer-review.
 94. R. Costas, Z. Zahedi, P. Wouters, Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *J. Assoc. Inf. Sci. Technol.* **66**, 2003–2019 (2015). doi: [10.1002/asi.23309](https://doi.org/10.1002/asi.23309)
 95. A. Clauset, D. B. Larremore, R. Sinatra, Data-driven predictions in the science of science. *Science* **355**, 477–480 (2017). doi: [10.1126/science.aal4217](https://doi.org/10.1126/science.aal4217)
 96. S. Wessely, Peer review of grant applications: What do we know? *Lancet* **352**, 301–305 (1998). doi: [10.1016/S0140-6736\(97\)11129-1](https://doi.org/10.1016/S0140-6736(97)11129-1); pmid: 9690424
 97. N. Geard, J. Noble, paper presented at the 3rd World Congress on Social Simulation, Kassel, Germany, 6 to 9 September 2010.
 98. Calm in a crisis. *Nature* **468**, 1002 (2010). doi: [10.1038/4681002a](https://doi.org/10.1038/4681002a); pmid: 21170024
 99. R. Roy, Funding science: The real defects of peer review and an alternative to it. *Sci. Technol. Human Values* **10**, 73–81 (1985). doi: [10.1177/016224398501000309](https://doi.org/10.1177/016224398501000309)
 100. J. Bollen, D. Crandall, D. Junk, Y. Ding, K. Börner, An efficient system to fund science: From proposal review to peer-to-peer distributions. *Scientometrics* **110**, 521–528 (2017). doi: [10.1007/s1192-016-2110-3](https://doi.org/10.1007/s1192-016-2110-3)
 101. M. S. Kohn *et al.*, IBM's health analytics and clinical decision support. *Yearb. Med. Inform.* **9**, 154–162 (2014). doi: [10.15266/iy-2014-0002](https://doi.org/10.15266/iy-2014-0002); pmid: 25123736
 102. J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullainathan, “Human decisions and machine predictions” (National Bureau of Economic Research, 2017).
 103. B. Liu, R. Govindan, B. Uzzi, Do emotions expressed online correlate with actual changes in decision-making?: The case of stock day traders. *PLOS ONE* **11**, e0144945 (2016). doi: [10.1371/journal.pone.0144945](https://doi.org/10.1371/journal.pone.0144945); pmid: 26765539
 104. M. E. J. Newman, The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 404–409 (2001). doi: [10.1073/pnas.021544898](https://doi.org/10.1073/pnas.021544898); pmid: 11149952
 105. R. Guimerà, B. Uzzi, J. Spiro, L. A. N. Amaral, Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**, 697–702 (2005). doi: [10.1126/science.1106340](https://doi.org/10.1126/science.1106340); pmid: 15860629
 106. M. Stringer, M. Sales-Pardo, L. A. Amaral, Effectiveness of journal ranking schemes as a tool for locating information. *PLOS ONE* **3**, e1683 (2008). doi: [10.1371/journal.pone.0001683](https://doi.org/10.1371/journal.pone.0001683); pmid: 18301760

ACKNOWLEDGMENTS

This work was supported by Air Force Office of Scientific Research grants FA9550-15-1-0077 (A.-L.B., R.S., and A.V.), FA9550-15-1-0364 (A.-L.B. and R.S.), FA9550-15-1-0162 (J.A.E. and D.W.), and FA9550-17-1-0089 (D.W.); National Science Foundation grants NCSE 1538763, EAGER 1566393, and NCN CP supplement 1553044 (K.B.) and SBE1158803 (J.A.E.); National Institutes of Health awards P01 AG039347 and U01CA198934 (K.B.) and IIS-0910664 (B.U.); Army Research Office grant W911NF-15-1-0577 and Northwestern University Institute on Complex Systems (B.U.); DARPA (Defense Advanced Research Projects Agency) Big Mechanism program grant 14145043 and the John Templeton Foundation's grant to the Metaknowledge Network (J.A.E.); Intellectual Themes Initiative “Just Data” project (R.S.); and European Commission H2020 FETPROACT-GSS CIMPACT grant 641191 (R.S. and A.-L.B.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of our funders.

10.1126/science.aa0185

Science

Science of science

Santo Fortunato, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Stasa Milojevic, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, Alessandro Vespignani, Ludo Waltman, Dashun Wang and Albert-László Barabási

Science **359** (6379), eaao0185.
DOI: 10.1126/science.aao0185

The whys and wherefores of SciSci

The science of science (SciSci) is based on a transdisciplinary approach that uses large data sets to study the mechanisms underlying the doing of science—from the choice of a research problem to career trajectories and progress within a field. In a Review, Fortunato *et al.* explain that the underlying rationale is that with a deeper understanding of the precursors of impactful science, it will be possible to develop systems and policies that improve each scientist's ability to succeed and enhance the prospects of science as a whole.

Science, this issue p. eaao0185

ARTICLE TOOLS

<http://science.sciencemag.org/content/359/6379/eaao0185>

REFERENCES

This article cites 93 articles, 26 of which you can access for free
<http://science.sciencemag.org/content/359/6379/eaao0185#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2018 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works