PNAS RESEARCH ARTICLE | SOCIAL SCIENCES

A network-based normalized impact measure reveals successful periods of scientific discovery across discipline

Qing Ke[a,1], Alexander J. Gates[b], and Albert-László Barabási[c,d,e,1]

Edited by Kenneth Wachter, University of California, Berkeley, CA; received June 4, 2023; accepted October 19, 2023

The impact of a scientific publication is often measured by the number of citations it receives from the scientific community. However, citation count is susceptible to well-documented variations in citation practices across time and discipline, limiting our ability to compare different scientific achievements. Previous efforts to account for citation variations often rely on a priori discipline labels of papers, assuming that all papers in a discipline are identical in their subject matter. Here, we propose a network-based methodology to quantify the impact of an article by comparing it with locally comparable research, thereby eliminating the discipline label requirement. We show that the developed measure is not susceptible to discipline bias and follows a universal distribution for all articles published in different years, offering an unbiased indicator for impact across time and discipline. We then use the indicator to identify science-wide high impact research in the past half century and quantify its temporal production dynamics across disciplines, helping us identifying breakthroughs from diverse, smaller disciplines, such as geosciences, radiology, and optics, as opposed to citation-rich biomedical sciences. Our work provides insights into the evolution of science and paves a way for fair comparisons of the impact of diverse contributions across many fields.

bibliometrics | scientific impact | citation analysis | science of science

Significance

Distinct citation practices across time and discipline limit our ability to compare different scientific achievements. For example, raw citation counts suggest that advancements in biomedical research have consistently overshadowed the accomplishments from all other disciplines. Here, we introduce a network-based methodology for normalizing citation counts that mitigates the effects of temporal and disciplinary variations in citations. The method allows us to highlight successful periods of scientific discovery across the disciplines and provides insights into the evolution of science.

Today's scientific enterprise is characterized by fierce competitions for limited resources such as funding and academic positions. Consequently, various stakeholders in science, from faculty hiring committees to grant panelists, are tasked to evaluate the scientific accomplishment of individuals, projects, and institutions and project potential future impact. Increasingly, these tasks are assisted by numerical measures that rely on citation data to calibrate the abstract notion of scientific impact, including the h-index and number of papers in high Impact Factor journals (1). The usages of these indicators in evaluation scenarios whose outcome affects researchers' career make it necessary to perform citation analysis in an unbiased manner.

For a scientific publication, the most widely used impact measure is the raw number of citations, C, capturing the volume of subsequent works that build upon it. Citation count is used both as an input for more advanced citation-based indicators for impact, and as a criterion for identifying breakthroughs (2). However, C suffers from two well-recognized biases (3–5): i) temporal bias, reflected by the higher rate of citations accumulated by later papers, an inflation process that makes it difficult to compare the impact of papers published decades apart; and ii) field bias, manifested by systematic differences of C across disciplines, creating the impression that papers in highly cited fields, like cell biology, have inherently bigger impact than, for example, mathematics papers, where citations are fewer.

Many techniques have been proposed to mitigate the influence of temporal and field biases on the assessment of scientific impact (see ref. 6 for a review). These methods typically suppress year- and/or field-level variations in citations by normalizing C^i of a paper i of interest with the average $\langle C^n \rangle_{n\in\mathcal{N}^i}$ of its similar papers $\mathcal{N}^i$. A widely adopted definition of $\mathcal{N}^i$ is the set of papers within the same research field as i, with field either operationalized based on publication venue (7) or constructed algorithmically (8). However, the choice of field classification systems (e.g., Web of Science's Subject Category, Scopus's Subject Area, etc.) can affect the conclusions drawn from normalized indicators based on journal-based classifications of fields (6). Similarly, algorithmically constructed fields may lack the transparency and reproducibility often emphasized in policy-making settings. More critically, partitioning papers into disjoint fields is a priori and assumes that all papers in a field are identical in their subject matter, ignoring significant within-field heterogeneities of subdisciplines and the increasing intermixing between disciplines (9). To avoid the complications of defining crisp disciplines for

Author affiliations: [a]School of Data Science, City University of Hong Kong, Hong Kong, China; [b]School of Data Science, University of Virginia, Charlottesville, VA 22904; [c]Network Science Institute, Northeastern University, Boston, MA 02115; [d]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115; and [e]Department of Network and Data Science, Central European University, Budapest 1051, Hungary

Author contributions: Q.K., A.J.G., and A.-L.B. designed research; Q.K. and A.J.G. performed research; Q.K., A.J.G., and A.-L.B. analyzed data; and Q.K., A.J.G., and A.-L.B. wrote the paper.

Competing interest statement: A.-L.B. is co-scientific founder of and is supported by Scipher Medicine, Inc., which applies network medicine strategies to biomarker development and personalized drug selection, and founder of Naring, Inc. which applies data science to health. The remaining authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: q.ke@cityu.edu.hk or a.barabasi@northeastern.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2309378120/-/DCSupplemental.

Published November 20, 2023.

impact normalization, another set of approaches define $\mathcal{N}^i$ as the set of papers that, together with additional criteria, share bibliographic references with i (coreferenced papers) (10). However, these methods only have limited power in identifying similar papers, as they depend on i's references—rather than the focal paper i itself.

Here, we propose a paper-level, citation-based indicator $\hat{C}$, representing our first contribution. Different from existing methods that require predefined discipline labels of papers for identifying $\mathcal{N}^i$, we consider $\mathcal{N}^i$ as the set of papers with which i is cocited. Our hypothesis is that the frequent coappearance of two papers in the same reference lists captures the scientific community's assessment of the topical relatedness of the two papers (9, 11–13). We show that $\hat{C}$ offers a better ability in identifying papers that the scientific community considers important and corrects for both temporal and field biases. Our proposal is not the first one that relies on cocited papers for citation normalization. Relative citation ratio (RCR), another paper-level citation indicator, is a forerunner in this regard (11). However, there are several key differences between our method and RCR. First, the normalizer used by RCR is the average citation rate of the journals where cocited papers were published, whereas we directly normalize by the average yearly citations of papers. Second, RCR further benchmarks normalized citation rate using papers funded by NIH R01 grants, remaining unclear how to generalize the benchmark to the entire scientific literature. Third, while RCR performs normalization only once, our normalization is performed on a yearly basis. As such, RCR is influenced by papers with different ages, and in theory, it could drop when extending the citation window, whereas $\hat{C}$ accumulates over time and is nondecreasing. Furthermore, our systematic, quantitative comparisons between $\hat{C}$ and RCR indicate that $\hat{C}$ can better correct the field bias than RCR does.

As a second contribution, we use $\hat{C}$ to generate insights into the evolution of science by revealing research fields that continue to produce high impact research over an extended period of time. We achieve this by quantifying the representation of a field in the set of top papers identified by $\hat{C}$, given the time- and field invariance of $\hat{C}$. Our results unveil a diverse set of fields that have been important sources for scientific breakthroughs, allowing us to look beyond the highly cited disciplines.

Results

Defining $\hat{C}$. To define our normalized indicator $\hat{C}$, we look at the total volume of citations generated by all citing papers published in a single year t. These citations are to be distributed among papers already published. For a particular paper i, it receives c_t^i citations (i.e., yearly raw citations). Typically, a large c_t^i is equated with high impact, but to determine the scale for which c_t^i is considered large, we need to compare it with c_t obtained by other papers similar to i in year t, denoted as $\mathcal{N}_t^i$. Here, we leverage cocitations to identify similar papers (9, 11–13) and define $\mathcal{N}_t^i$ as the set of papers that are cocited with i by papers published in year t (Fig. 1). We then define our yearly normalized citations, $\hat{c}_t^i$, as c_t^i normalized by the average yearly citations of the papers in $\mathcal{N}_t^i$, formally,

$$\hat{c}_t^i = \frac{c_t^i}{\langle c_t^n \rangle_{n \in \mathcal{N}_t^i}}. \quad \textbf{[1]}$$

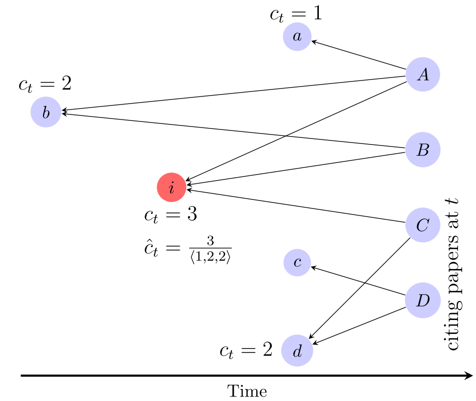


Fig. 1. Defining yearly normalized citations, $\hat{c}_t^i$. For a paper i, its $\hat{c}_t^i$ is defined as yearly raw citations, c_t^i, normalized by the average yearly citations of the papers that are cocited with i by citing papers published in year t. In this figure, paper i at t has $c_t^i = 3$ and is cocited with papers a, b, and d, which have yearly citations of 1, 2, 2, respectively. Therefore, $\hat{c}_t^i = \frac{3}{\langle 1,2,2 \rangle} = 1.8$. In another year, new citing papers are published, which leads to the change of c^i, the list of i's cocited papers, and their yearly citations, resulting in a different $\hat{c}^i$.

By definition, $\hat{c}_t^i = 0$ if $c_t^i = 0$. The total normalized citations in T years after publication is the sum of the yearly contributions: $\hat{C}_T^i = \sum_{t=0}^{T} \hat{c}_t^i$.

Our measure, $\hat{C}$, differs from the many existing measures in two important ways. First, most prevailing normalization methods are designed with a prospective viewpoint that tracks a paper's citations accumulated over T years and normalized to similar papers, which complicates the assessment of citation dynamics. In contrast, our method is designed from a retrospective view, focusing on how citations generated by citing papers published in a single year are distributed to papers already published before and constructed $\hat{C}$ from yearly normalized data. Our rationale is that already published papers compete for citations from citing papers and the comparisons of the collected citations need to be made yearly, since citation volumes are increasing over time, driven by the exponential growth of science publishing and the gradually increasing number of references in a paper (3). Second, most previous normalization procedures involve global partitioning of papers, be it journal based or algorithmically derived, and all the papers within one partition share the same $\mathcal{N}^i$. This choice assumes implicitly that those papers are similar to each other, which is hardly the case: Condensed matter physics, for example, is an umbrella for superconductivity, topological materials, quantum magnetism, subdisciplines with different community sizes, and distinct citation practices. By contrast, our method identifies papers that are locally similar to paper i, and each paper has its own "personalized" $\mathcal{N}^i$ that may change over time.

To further illustrate the advantage of $\hat{C}_T^i$, we compare it with two other popular indicators: total raw citations $C_T^i = \sum_{t=0}^{T} c_t^i$ and $\tilde{C}_T^i$, which normalizes C_T^i with the average C_T of papers in the same year and field, as determined by journals (7). We calculate the three indicators for two exemplar papers published in *Nature* in 1985 (Fig. 2). The first paper, p_1, is in cell biology reporting measured calcium levels in muscle cells (14). The second, p_2, is a geoscience paper that analyzed density contrasts in the Earth's lower mantle (15). Based on C_{10} [T is set to 10 y to tradeoff between including more papers in our analysis while keeping a relatively long citation window, following previous

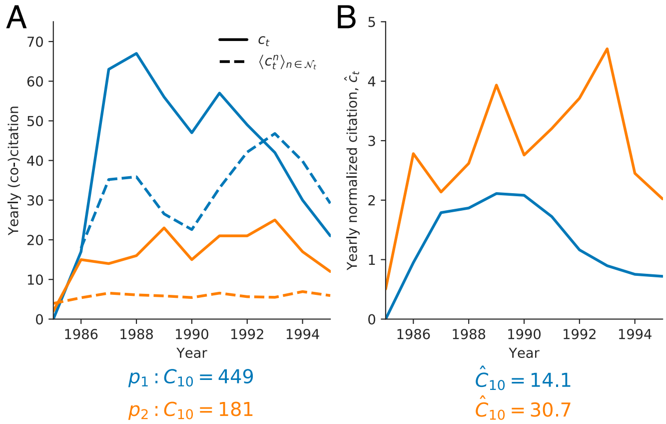


Fig. 2. Comparing $\hat{C}_{10}$ with C_{10} and $\tilde{C}_{10}$, using two papers. The first paper (p_1) refers to ref. 14 (blue), and the second (p_2) refers to ref. 15 (orange). Both were published in 1985 in *Nature*. (*A*) Solid lines represent yearly raw citations, c_t, indicating that p_1 constantly received more citations than p_2 did and thus had a higher impact than p_2, based on C_{10}. Given that both papers were in the same year and journal group, $\tilde{C}_{10}$ also suggests p_1 had a higher impact than p_2. Dashed lines represent average yearly citations of the papers that are cocited with p_1 (p_2), indicating that p_1's cocited papers had more citations than p_2's, i.e., a higher $\langle c_t^n \rangle_{n\in\mathcal{N}_t}$. (*B*) Our yearly normalized measure, $\hat{c}_t$, which compares c_t with $\langle c_t^n \rangle$, suggests that p_1 has a lower impact than p_2.

studies that have also selected similar values (2, 16, 17)], p_1 has a much higher impact than p_2 (449 vs. 181; Fig. 2*A*). But C_{10} does not consider field size; most of p_1's citations came from cell biology and other prolific biomedical fields, while p_2 was mostly cited by papers in geosciences—much smaller fields (*SI Appendix*, Fig. S1). The $\tilde{C}_{10}$ indicator accounts for this field size effect using journal-based categorization of fields. Therefore, it continues to rank p_1 higher (20.5 vs. 8.3) because p_1 and p_2 were published in the same journal and hence share identical field-specific normalizer. By contrast, the proposed $\hat{C}$ automatically identifies the research area of a paper through the list of cocited papers and compares its impact with those papers. Indeed, although p_1 acquired many citations, so did many of the other papers with which it was cocited, i.e., a large $\langle c_t^n \rangle_{n\in\mathcal{N}_t}$ (Fig. 2*A*). This is in contrast to p_2, which received many more citations than its cocited papers (Fig. 2*A*). Therefore, the yearly normalized citations, $\hat{c}_t$, indicate that p_1 has a consistently lower yearly disciplinary impact than p_2 (Fig. 2*B*) and thus a lower total impact (14.1 vs. 30.7). In other words, $\hat{c}_t$ measures the relative impact of each paper within the "discipline" it is embedded in, which may be a single discipline, or a mixture of multiple traditionally defined disciplines.

Validating $\hat{C}$. We validate $\hat{C}_{10}$ using external evaluations from domain experts on the importance of papers. For example, Physical Review Letters (PRL) released a list of 87 "Milestone Letters" that made significant contributions to the development of physics and can be considered as breakthroughs within their subdisciplines (https://journals.aps.org/prl/50years/milestones). First, we confirm that the selection of these papers was not driven by raw citation count, as they are not the most cited papers in the journal and their ranks based on C_{10} among all PRL papers span several orders of magnitude (Fig. 3*C*). Still, milestone papers have higher impact than the average PRL papers, as measured by both C_{10} and $\hat{C}_{10}$ (Fig. 3 *A* and *B*). More importantly, when we compare the ranks of milestone papers based on C_{10} and $\hat{C}_{10}$, we find that these papers are consistently ranked higher if we use $\hat{C}_{10}$ than if we use C_{10} (Fig. 3*C*), indicating a better capability of $\hat{C}_{10}$ to recover important papers. Methods like $\tilde{C}_{10}$ that use journals as the proxy for fields lose such capability, given that all the milestone papers were published in the same journal. The same effect is observed for milestone papers identified by three other journals (*PNAS*, *PRE*, and *Human Relations*), confirming the ability of $\hat{C}_{10}$ to select papers that the scientific community considers important, independent of their citation counts and publication venue (*SI Appendix*, Fig. S2).

$\hat{C}$ Corrects Temporal and Field Biases. One widely known drawback of raw citation count is its dependence on publication time. This can be readily seen from Fig. 4*A*, where we plot, for each year from 1945 to 2007, the distribution of C_{10} for papers published in that single year. It is apparent that these distributions shift systematically rightward over time, indicating an inflation process of C_{10}. For example, the median C_{10} for papers in 1945 is 1, which increased to 9 in 2007; a paper in 1945 with 17 citations was able to become a hit, i.e., a top 5% most cited paper, but achieving the same status required 65 citations for papers in 2007. Such a dependence on time, however, is not presented in $\hat{C}_{10}$. The distributions of $\hat{C}_{10}$ for all the years considered appear to collapse onto a single shape (Fig. 4*B*; see also *SI Appendix*, Fig. S3), demonstrating a universality and lending a strong support for the temporal stability of $\hat{C}_{10}$. We further fit a log-normal distribution for each individual year, since it is one of the most popular functional forms used in previous citation distribution analyses. We find that the distributions of $\hat{C}_{10}$ of papers in individual years are compatible with the log-normal form with nearly the same shape parameter $\sigma = 1$ to 1.2 (*SI Appendix*, Fig. S4). This result is in line with numerous prior studies that identified a similar range of σ for papers in journals (18–20), institutions (20), fields (21–23), as well as Mendeley readerships (24) and patent citations (25).

The inflation of C_{10} generates temporal bias favoring more recent papers. To demonstrate this bias more quantitatively, we rank all papers published in 1945 to 2007 according to C_{10} and calculate the percentage of papers published in each year that appear in the top 5% of the global ranking. If C_{10}-based ranking is fair, it means that each year would contribute 5% of its papers to the top, which is far from the case: There is a systematic increase in the percentage of papers ranked into the top over time (Fig. 4*C*), suggesting that ranking by C_{10} creates bias favoring more recent papers. However, if we rank papers by $\hat{C}_{10}$, yearly contributions fluctuate around the baseline, without exhibiting the recency bias as C_{10} does.

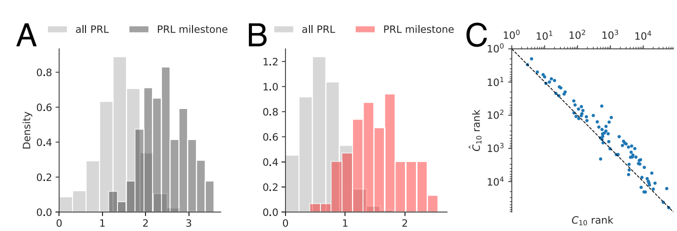


Fig. 3. Validating $\hat{C}_{10}$ using "Milestone Letters" published in Physical Review Letters (PRL). (*A* and *B*) Histograms of C_{10} and $\hat{C}_{10}$ for milestone papers and all PRL papers. (*C*) Comparison between C_{10}- and $\hat{C}_{10}$-based ranks of milestone papers among all PRL papers. Rank 1 corresponds to the paper with the largest C_{10} ($\hat{C}_{10}$). The diagonal dashed line represents equal ranks, and points above the line indicate better ranks based on $\hat{C}_{10}$.

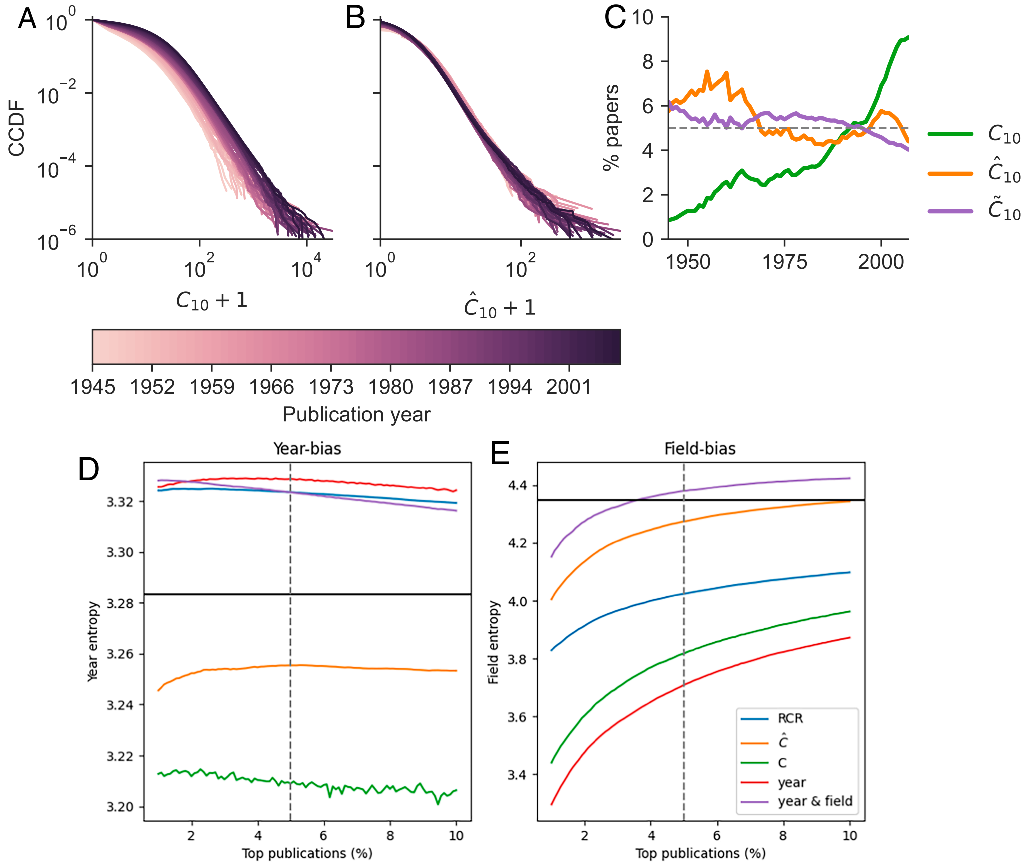


Fig. 4. $\hat{C}_{10}$ corrects the temporal and field bias characterizing C_{10}. (*A* and *B*) The inflation of C_{10} over time, in contrast to the temporal stability of $\hat{C}_{10}$. We plot, for each year from 1945 to 2007, the survival distribution of (*A*) C_{10} and (*B*) $\hat{C}_{10}$ for papers published in that year. The color of a curve encodes publication year. While the distributions of C_{10} shift rightward, the distributions of $\hat{C}_{10}$ are well described by a single shape. (*C*) Percentages of papers published in each year that appear in the top 5% ranked by C_{10}, $\hat{C}_{10}$, or $\tilde{C}_{10}$ among all papers in 1945 to 2007. For clarity, the curve for C_{10} normalized by year is not shown, as it is essentially identical to the $\tilde{C}_{10}$ curve. The RCR case is not applicable due to the unavailability of RCR metrics for the entire WoS corpus. (*D* and *E*) The entropy of (*D*) years and (*E*) fields for the top publications ranked among all papers from 1980 to 2007 using C_{10} (green), C_{10} normalized by each year (red), C_{10} normalized by each year and field (purple), RCR (blue), and $\hat{C}_{10}$ (orange). In both cases, the corpus of papers refers to PubMed where RCR is available (26), the horizontal black line indicates the entropy for the entire corpus, and the vertical dashed line marks the top 5%. See also *SI Appendix*, Fig. S12 where we focus on the absolute entropy difference from the baseline.

Another way to capture the recency bias is to explore the diversity of years present in a ranking of all publications. Here, we take the top *p*-percent of publications as ranked by different citation metrics and then measure the diversity of years using the entropy of the normalized year count distribution. If no bias was present in the ranking and all publications had an equal chance of being at the top, then all years would be present proportional to the total number of publications from that year and we would find an entropy of 3.29 bits, while a smaller entropy reflects less diversity and a bias toward specific years, and a larger entropy occurs for a greater diversity reflecting overcompensation. As shown in Fig. 4*D*, C_{10} is the most biased indicator, selecting more top publications from the same years. On the other hand, C_{10} normalized by the average in each year has a higher entropy than the corpus baseline, indicating an overcompensation with more publications from early years in the overall top ranking than would be expected based on their frequency. Notably, both RCR and C_{10} normalized by each year and field also display this overcompensation and promote more publications from early years. Finally, $\hat{C}_{10}$ shows a compromise between the tendency to over or under compensate and has an entropy of 3.25 bits for the top 5% of publications.

Another potential bias when recognizing top publications using C_{10} comes from the drastically different sizes and citation norms in different disciplines. This can be observed from *SI Appendix*, Fig. S5*A*, where we show the distributions of C_{10} for several fields. We observe that, from cell biology to analytical chemistry to mathematics, there is a systematic, one order of magnitude decease of median C_{10}. These differences nearly disappear if we use $\hat{C}_{10}$ (*SI Appendix*, Fig. S5*B*).

Once again, we can capture the field bias by exploring the diversity of fields present in a ranking of all publications. Here, we take the top *p*-percent of publications as ranked by different citation metrics and then measure the diversity of fields using the entropy of the normalized field count distribution. If no bias was present in the ranking and all publications had an equal chance of being at the top, then all fields would be present proportional to the total number of publications from that field, resulting in an entropy of 4.35 bits. A smaller entropy reflects less diversity and a bias toward specific fields, and a larger entropy occurs for a greater diversity reflecting overcompensation. As shown in Fig. 4*E*, C_{10} is an extremely biased indicator, selecting more top publications from the same few fields. Surprisingly, C_{10} normalized by the average in each year is even more biased toward specific fields—its application increases the tendency to favor publications from specific fields. Notably, both $\hat{C}_{10}$ and C_{10} normalized by each year and field show significantly reduced bias toward a few fields and are consistently the closest to the random baseline. More specifically, C_{10} normalized by each year and field is slightly less bias than $\hat{C}_{10}$ for the top 6% of publications but tends to overcompensate for smaller fields at the 5% mark, while $\hat{C}_{10}$ consistently has a smaller entropy and is closer to the random baseline for top publication sets of 6% to 10%. Finally, RCR shows a reduced bias toward specific fields compared to C_{10}, but is still more biased than $\hat{C}_{10}$.

In summary, based on these diverse perspectives, we conclude that $\hat{C}_{10}$ best corrects both the temporal and field biases characterizing C_{10}.

Fields Producing High-Impact Works. Finally, we leverage the field- and time invariance of $\hat{C}$ to assess science-wide advancement across discipline and time. It has long been posited that science advances through intermittent revolutionary breakthroughs that have long-lasting impact by triggering new directions of research and giving birth to new disciplines (27). While much quantitative attention has focused on characterizing the emergence of individual fields (28, 29), their growth dynamics (30–35), and the interactions between disciplines (36), little is known about the dynamics of individual breakthroughs within and across fields. This paucity of knowledge prompts us to ask the following: What fields produce high impact research and how does a field's ability to stay at the forefront of the research enterprise change over time?

Answers to these questions rely on the accurate identification of breakthroughs. A straightforward way adopted in existing practice considers the top x% most cited papers grouped by year and field as breakthroughs (2). This, on one hand, mitigates the effects of temporal and disciplinary variations in citations but assumes that every field in every year generates breakthroughs at the same rate. Such a strong assumption about the pace of science advancement is highly unlikely to hold, considering, for example, the theoretical argument from Kuhn that scientific revolutions happen only sporadically (27). Since $\hat{C}_{10}$ is not susceptible to temporal and disciplinary biases, we can directly compare the impact of papers across discipline and time. Therefore, we select from all the papers in our sample (including all years and disciplines) the top 5% with the highest $\hat{C}_{10}$ and denote the obtained list as $\hat{H}$. For comparison, we also identify top papers based on C_{10} and $\tilde{C}_{10}$ and denote them respectively as H and $\tilde{H}$. Our first observation is that $\hat{H}$ and H only share 49% of the papers and $\tilde{H}$ and H 59% (*SI Appendix*, Fig. S6). This suggests that papers with low C_{10} can still rank high based on $\hat{C}_{10}$.

Turning to the question of which fields produce high impact papers, we compute the fraction of papers in $\hat{H}$ that belong to a given field, finding that top contributors to $\hat{H}$ include fields in physics, chemistry, and biomedicine (Fig. 5). Particularly interesting are the three of the top four categories, which

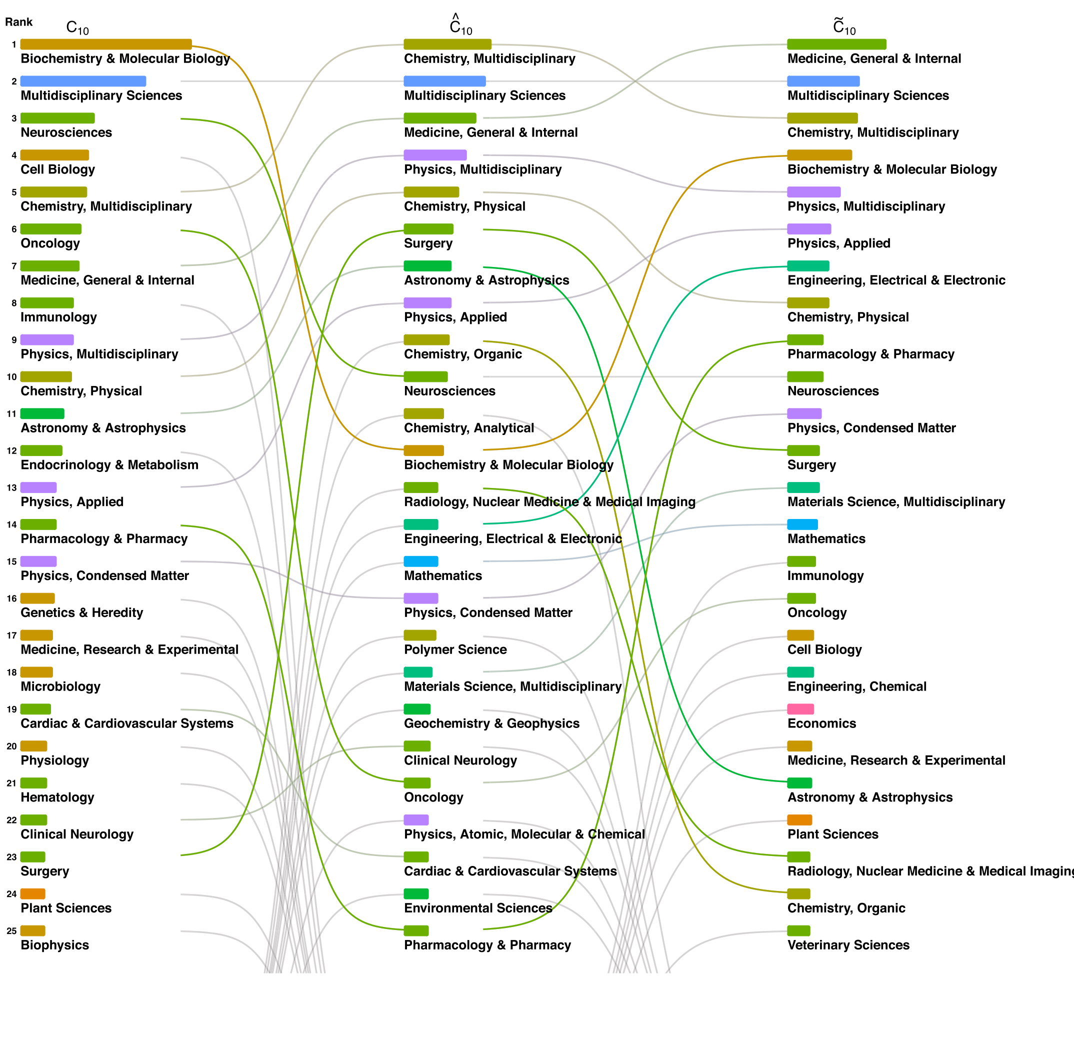


Fig. 5. Rank comparison of fields by their shares in top 5% papers. We identify from all the papers in our sample the top 5% based on C_{10}, $\hat{C}_{10}$, and $\tilde{C}_{10}$ and compute the fraction of these papers that belong to each field. The horizontal bars represent the shares, and the color represents the broad discipline. We only display the top 25 fields and report the fields with the most rank changes in *SI Appendix*, Table S1. Multidisciplinary Sciences is a hybrid category that includes multidisciplinary journals such as *Nature*, *Science*, *PNAS*, *Scientific Reports*, etc.

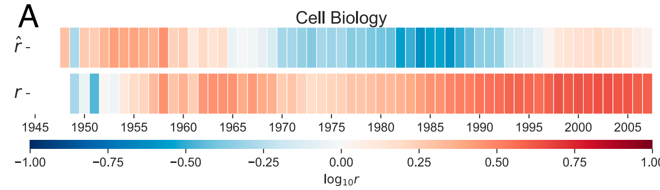
are multidisciplinary chemistry, multidisciplinary sciences, and multidisciplinary physics. These correspond to multidisciplinary journals like *Nature*, *Science*, and *PNAS* that are perceived commonly by the academic community to publish selectively, as well as physics and chemistry journals like *PRL* and *Journal of the American Chemical Society* that publish quality contributions from any topics within the disciplines. *SI Appendix*, Fig. S7 confirms that these journals indeed published the most papers in $\hat{H}$, corroborating the ability of $\hat{C}_{10}$ to pick out high-impact papers. Several fields that account for only a small portion in H become sizable contributors to $\hat{H}$, including mathematics, electrical and electronic engineering, geochemistry and geophysics, organic chemistry, polymer science, etc. Among them, mathematics is noticeable for its steep ascending from a low representation in H (0.19%; ranked 93rd) to a leading position in $\hat{H}$ (1.6%; ranked 15th). On the other hand, many biomedical fields like cell biology, genetics and heredity, and microbiology that are ranked high when using C_{10} move to lower positions when using $\hat{C}_{10}$.

Dynamics of Production of High-Impact Works. We characterize the temporal evolution of the production of high impact papers across fields, by taking field size into consideration, as larger fields would have more top papers by chance. In doing so, we introduce $\hat{r}_{f,t}$ that measures the representation in $\hat{H}$ by papers in field f and year t. Specifically, it is the fraction of papers in $\hat{H}$ that belong to field f and year t, normalized by the fraction of papers of the same group in all the papers. Thus, $\hat{r}_{f,t} > 1$ indicates that the group is overrepresented in $\hat{H}$ and $\hat{r}_{f,t} < 1$ underrepresented, thereby providing a quantification of the field's ability to produce breakthrough articles relative to other fields at t. Similarly, we calculate $r_{f,t}$.

Let us first use Cell Biology as an example (Fig. 6). This field in 2007 accounted for 0.043% and 0.04% in $\hat{H}$ and in all the papers, respectively, indicating its overrepresentation in $\hat{H}$ ($\hat{r} = 1.07$). The heatmaps presented in Fig. 6*A* that encode $\hat{r}_t$ and r_t by color reveal the dynamics of the production of high impact research from Cell Biology over six decades. Based on r_t, it has been consistently a source for revolutionary scientific breakthroughs, which partly reflects high average citations of papers in this field. A much richer dynamics is revealed by $\hat{r}_t$: Relative to other fields, Cell Biology lost its ability to produce breakthroughs between 1965 and 1980s and has regained its leading role in the early 1990s. We hypothesize that the rebound may be related to the emergence of genomics, as exemplified by the Human Genome Project initiated in the 1990s (38). To corroborate this, we identify overrepresented title words of Cell Biology papers published in each decade, compared to papers before 1980, finding that the 1990s were characterized by the rise of studies centered around gene expression (Fig. 6*B*).

We expand the analysis to other fields and group them based on their r_t and $\hat{r}_t$ (Fig. 7). The first group corresponds to fields that continue to produce high impact works disproportionately during the studied period ($\hat{r}_t > 1$, $r_t > 1$), as identified by both measures, including Neuroscience, Astronomy and Astrophysics (Fig. 7*A*). The second group is featured by $\hat{r}_t < 1$ and $r_t > 1$. Those include many Biomedical Research and Clinical Medicine fields that are ranked high only by C_{10} (Fig. 7*B*). The third group includes fields whose importance is dismissed by r_t ($r_t < 1$) but picked up by $\hat{r}_t$ ($\hat{r}_t > 1$). Those fields span diverse disciplines (Fig. 7*C*), including i) within the Earth and Space category, Geochemistry and Geophysics, Meteorology and Atmospheric Sciences, Limnology, and Oceanography; ii) several physics fields, including Fluids and Plasmas Physics, Applied Physics, and Optics; iii) Electrochemistry and Physical Chemistry; iv) Ecology; and v) Surgery, Clinical Neurology, and Radiology from Clinical Medicine. The contrast of the likelihood to produce high impact works for fields in the third category also leads us to ask the following: Are there no periods of time when there are "excitements" going on in those fields, as suggested by r_t? We argue that this is not the case. For example, considering the development of medical imaging, there has been tremendous progresses in this area since the 1970s, and advancements like computer-assisted tomography (CT) and MRI have been quickly applied for medical diagnostics and won the 1979 and 2003 Nobel Prize in Physiology or Medicine (39). The fields of the key breakthroughs behind these development, like radiology ("Radiology, Nuclear Medicine & Medical Imaging"), are dismissed by r_t but captured by $\hat{r}_t$.

A



B

1980s	1990s	2000s
ca-2	expression	expression
monoclonal-antibodies	gene	signaling
growth-factor	protein	apoptosis
monoclonal-antibody	receptor	cancer
invitro	human	kinase
cyto-toxic	novel	protein
cell-line	domain	cell
fibronectin	kinase	regulates
saccharomyces-cerevisiae	ca2	alpha
human-fibroblasts	factor	receptor
flow-cytometry	cloning	stem
gene	transcription	pathway
cyto-toxicity	promoter	beta
invivo	tyrosine	gene
heat-shock	gene-expression	activation
t-cell	activation	novel
messenger-rna	regulation	through
genes	protein-kinase-c	factor
cell-lines	expressed	via
natural-killer	not	arabidopsis

Fig. 6. A case study of Cell Biology. (*A*) Dynamics of the production of top cited papers from Cell Biology. (*B*) Overrepresented title words of Cell Biology papers in each decade, relative to papers before 1980 (37).

Finally, going beyond case studies, we examine which features of a field explain its r_t ($\hat{r}_t$). We hypothesize that field size and number of references by field may correlate with r_t and $\hat{r}_t$, as previous studies have pointed out that the two factors contribute to the temporal and field biases in raw citations (40). We find that field size is less correlated with $\hat{r}_t$ than with r_t (*SI Appendix*, Fig. S8). More importantly, the average number of references per paper is significantly less correlated with $\hat{r}_t$ than with r_t (the median coefficient of determination $R^2 = 0.16$ vs. 0.49; *SI Appendix*, Fig. S9). These results further support the ability of $\hat{C}_{10}$ to correct the systematic biases of C_{10} and $\hat{r}_t$ rather than r_t as a viable indicator for a field's tendency to produce high impact research.

Discussion

Citation-based impact metrics have been increasingly adopted for academic performance evaluations of diverse types of actors—authors (41), institutions (42), and even nations (43), playing an important role in hiring, funding, and promotion (44). It is

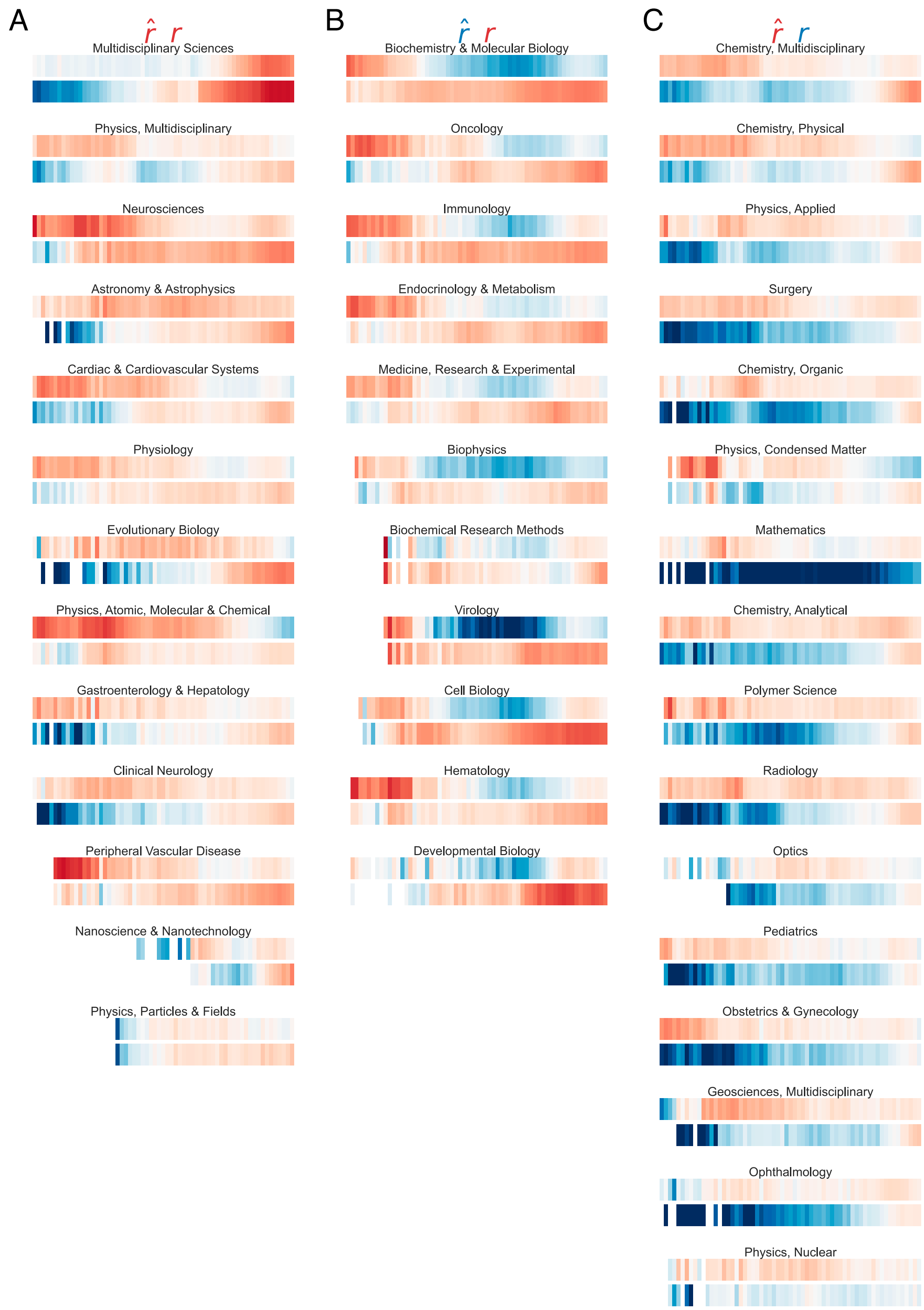


Fig. 7. Production dynamics of top papers in selected fields. We group fields based on their $\hat{r}_t$ and r_t values: (*A*) fields that are overrepresented based on both $\hat{r}_t$ and r_t; (*B*) fields that are underrepresented based on $\hat{r}_t$ and overrepresented based on r_t; and (*C*) fields that are overrepresented based on $\hat{r}_t$ and underrepresented based on r_t.

therefore essential to carry out citation analysis in an unbiased way. Yet, raw citations are known to be biased by variations in citation patterns across discipline and time, prompting us to propose a properly normalized measure that corrects those biases.

The fact that both RCR and $\hat{C}$ rely on cocited papers for citation normalization raises the question of the differences between the two indicators. Our systematic comparisons of them indicate that our $\hat{C}$ can better correct the field bias than RCR does, yet for year bias, all metrics underperform, meaning that they all identify top papers that were more likely to be published recently. Such a undercompensation might indicate that, at least in biomedicine, disruptive science may not be declining as suggested in a previous study (45). Furthermore, $\hat{C}$ identifies a very different set of highly influential papers from the set by RCR, sharing only 54% of the top 5% papers. The rankings of fields based on their shapes in top papers also reveal that, while fields like General and Internal Medicine, Neuroscience, and Surgery are found to be important by both methods, the importance of multidisciplinary physics and chemistry as well as radiology are not recognized by RCR (*SI Appendix*, Fig. S11).

Throughout this work, we used citation data to quantify the scientific impact of a scientific article, aiming to unveil how and when raw citation signals impact. We acknowledge that scientific impact is a complex, multidimensional notion, easier to intuit than to quantify, and getting cited may capture only one aspect of impact as codified in the discourse of science. Supplementing citations with contextual information extracted from text analysis may help capture more nuanced notion of impact (46). Similarly, looking at citations from other domains, like patents and policy documents, may also enrich the multifaceted feature (47, 48). Still, due to the accessibility and quantity of article citation data, citations provide valuable insights into the evolution of scientific discovery and citation-based analyses form the cornerstone of the science of science. In implementing our methodology, given the content and structure of current bibliographic databases, it is easier to calculate our metric for individual papers than existing normalization metrics (*SI Appendix, Text*). Yet, when the number of citations is small, the cocitation neighborhood might also be small and therefore be less reliable for normalization. Mirroring the raw citation count, the cocitation network may also be influenced by self-citations, large authorship teams, or other factors related to the social processes of science that may affect how citations are generated and consequently affect the network (49). For example, publicity, such as comments and promotion in social media, and attention to high Impact Factor journals may induce additional citations. Similarly, social and epistemological considerations may also generate influence citations that are interpreted as expressions of "discursive relation" or "professional relation" to scientific communities. To address these factors, future work can utilize datasets which capture diverse aspects of the social processes, and derive citation networks that better reflect impact. For example, very high profile results are often not cited explicitly, but only mentioned in the text, acquiring hidden citations (50). Our methodology is then ready to be adapted to address the role of these modified graphs. Finally, we focused only on articles, which may not be the main publication medium for some fields in social sciences, humanities, and computer science.

Despite these limitations, our proposal of a citation-based measure that is time invariant and free from discipline bias allows us to compare the impact of papers across years and disciplines. A key contribution of our method is the elimination of the need to assign a publication to a discipline when measuring its impact. Previous research has demonstrated that science is becoming increasingly interdisciplinary (9), complicating the traditional picture of science structured into well-defined research departments and funding programmes. Thus, notions of locally comparable research, such as we introduced here, provide an important step toward studying the interactions between scientific disciplines and the emergence of new research areas. To this end, we demonstrate that contributions to revolutionary breakthroughs in the past half century came from diverse disciplines, such as radiology, applied physics, ecology, and geosciences, as opposed to be dominated by biomedical sciences.

Materials and Methods

We based our analysis on the Web of Science (WoS) database. We only considered "article," "letter," and "note" documents indexed there and limited our attention to 26,792,332 papers published between 1945 and 2007 to ensure a 10-y citation window. Citing documents were constrained within the three selected types; therefore, citations from other types of papers such as editorial and review were not included. Research fields of papers were taken as WoS Subject Category (SC), and papers can have more than one SC.

The $\tilde{C}$ measure uses papers published in the same SC and year as reference set and normalizes the number of citations by the average citations of papers in the set (51, 52). Therefore, this indicator relies on external category labels of papers. Applying the procedure in practice also requires one to make the choice on how to deal with papers with multiple categories, as 36.2% of papers have more than one category (*SI Appendix*, Table S2). There are several ways to handle those papers. A straightforward one is to count them multiple times. This, however, would artificially increase the number of papers drastically, which would lead to the increase in the number of top papers. Here, we first calculate the $\tilde{C}_{10}$ value for each category assigned to a paper and then pick the maximum one. We ignore the 56,991 papers without category labels.

We obtain RCR metrics from the NIH Open Citation Collection (Version 40) (26, 53), which focuses on PubMed papers. We retain only research articles published in 1980 to 2007, as a large fraction of pre-1980 papers do not have RCR values (*SI Appendix*, Fig. S10). We then match the retained papers to WoS using PubMed ID or DOI and drop unmatched papers from our analysis. The final corpus contains 7,710,057 PubMed papers.

A Python implementation of $\hat{C}$ is provided in the pySciSci package (54).

Data, Materials, and Software Availability. Replication data have been deposited in Github (https://github.com/qke/network-normed-c) (55). Some study data available the raw Web of Science data used in this work cannot be shared due to its proprietary nature but is available upon purchase from Clarivate Analytics at https://clarivate.com/contact-us/sales-enquiries/. Other relevant data supporting the replication of this work have been deposited at https://github.com/qke/network-normed-c (55). A Python implementation of the proposed measure is available in the pySciSci package (https://github.com/SciSciCollective/pyscisci).

ACKNOWLEDGMENTS. We thank Onur Varol and Istvan Kovacs for useful discussions and Tongyu Ding for preparing Fig. 5. Q.K. is partially supported by the National Natural Science Foundation of China (72204206), City University of Hong Kong (Project No. 9610552), and Hong Kong Institute for Data Science. A.-L.B. is supported by the Templeton Foundation under contract #61066, the Air Force Office of Scientific Research under award number FA9550-19-1-0354, an European Research Council (ERC) Synergy grant (DYNASNET-810115), the Eric and Wendy Schmidt Fund for Strategic Innovation (G-22-63228), and the National Science Foundation (SES-2219575).

1. C. R. Sugimoto, V. Larivière, *Measuring Research: What Everyone Needs to Know* (Oxford University Press, 2018).
2. B. Uzzi, S. Mukherjee, M. Stringer, B. Jones, Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).
3. L. Egghe, R. Rousseau, *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science* (Elsevier, 1990).
4. O. Persson, W. Glänzel, R. Danell, Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics* **60**, 421–432 (2004).
5. L. Bornmann, H. Daniel, What do citation counts measure? A review of studies on citing behavior. *J. Document.* **64**, 45–80 (2008).
6. L. Waltman, A review of the literature on citation impact indicators. *J. Inf.* **10**, 365–391 (2016).
7. 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).
8. J. Ruiz-Castillo, L. Waltman, Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *J. Inf.* **9**, 102–117 (2015).
9. A. J. Gates, Q. Ke, O. Varol, A. L. Barabási, Nature's reach: Narrow work has broad impact. *Nature* **575**, 32–34 (2019).
10. C. Colliander, A novel approach to citation normalization: A similarity-based method for creating reference sets. *J. Assoc. Inf. Sci. Technol.* **66**, 489–500 (2015).
11. B. I. Hutchins, X. Yuan, J. M. Anderson, G. M. Santangelo, Relative citation ratio (RCR): A new metric that uses citation rates to measure influence at the article level. *PLoS Biol.* **14**, e1002541 (2016).
12. H. W. Shen, A. L. Barabási, Collective credit allocation in science. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12325–12330 (2014).
13. H. Small, Co-citation in the scientific literature: A new measure of the relationship between two documents. *J. Am. Soc. Inf. Sci.* **24**, 265–269 (1973).
14. D. A. Williams, K. E. Fogarty, R. Y. Tsien, F. S. Fay, Calcium gradients in single smooth muscle cells revealed by the digital imaging microscope using Fura-2. *Nature* **318**, 558–561 (1985).

15. B. H. Hager, R. W. Clayton, M. A. Richards, R. P. Comer, A. M. Dziewonski, Lower mantle heterogeneity, dynamic topography and the geoid. *Nature* **313**, 541–545 (1985).
16. R. Sinatra, D. Wang, P. Deville, C. Song, A. L. Barabási, Quantifying the evolution of individual scientific impact. *Science* **354**, aaf5239 (2016).
17. A. Zeng *et al*., Increasing trend of scientists to switch between topics. *Nat. Commun.* **10**, 3439 (2019).
18. M. Thelwall, Citation count distributions for large monodisciplinary journals. *J. Inf.* **10**, 863–874 (2016).
19. M. J. Stringer, M. Sales-Pardo, L. A. Nunes Amaral, Effectiveness of journal ranking schemes as a tool for locating information. *PloS One* **3**, e1683 (2008).
20. A. Chatterjee, A. Ghosh, B. K. Chakrabarti, Universality of citation distributions for academic institutions and journals. *PloS One* **11**, e0146762 (2016).
21. F. Radicchi, C. Castellano, A reverse engineering approach to the suppression of citation biases reveals universal properties of citation distributions. *PLoS One* **7**, e33833 (2012).
22. T. Evans, B. Kaube, N. Hopkins, Universality of performance indicators based on citation and reference counts. *Scientometrics* **93**, 473–495 (2012).
23. M. Golosovsky, Universality of citation distributions: A new understanding. *Quant. Sci. Stud.* **2**, 527–543 (2021).
24. C. A. D'Angelo, S. Di Russo, Testing for universality of Mendeley readership distributions. *J. Inf.* **13**, 726–737 (2019).
25. J. R. Clough, J. Gollings, T. V. Loach, T. S. Evans, Transitive reduction of citation networks. *J. Complex Netw.* **3**, 189–203 (2015).
26. B. I. Hutchins *et al*., The NIH open citation collection: A public access, broad coverage resource. *PLoS Biol.* **17**, e3000385 (2019).
27. T. Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press, 1962).
28. M. Rosvall, C. T. Bergstrom, Mapping change in large networks. *PloS One* **5**, e8694 (2010).
29. N. Shibata, Y. Kajikawa, Y. Takeda, K. Matsushima, Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation* **28**, 758–775 (2008).
30. W. Goffman, V. A. Newill, Generalization of epidemic theory: An application to the transmission of ideas. *Nature* **204**, 225–228 (1964).
31. W. Goffman, Mathematical approach to the spread of scientific ideas-the history of mast cell research. *Nature* **212**, 449–452 (1966).
32. S. M. Kot, The stochastic model of evolution of scientific disciplines. *Scientometrics* **12**, 197–205 (1987).
33. L. M. A. Bettencourt, D. I. Kaiser, J. Kaur, C. Castillo-Chávez, D. E. Wojick, Population modeling of the emergence and development of scientific fields. *Scientometrics* **75**, 495 (2008).
34. M. Szydłowski, A. Krawiec, Growth cycles of knowledge. *Scientometrics* **78**, 99–111 (2009).
35. L. M. A. Bettencourt, J. Kaur, Evolution and structure of sustainability science. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19540–19545 (2011).
36. M. Rosvall, C. T. Bergstrom, Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1118–1123 (2008).
37. B. L. Monroe, M. P. Colaresi, K. M. Quinn, Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Polit. Anal.* **16**, 372–403 (2017).
38. A. J. Gates, D. M. Gysi, M. Kellis, A. L. Barabási, A wealth of discovery built on the Human Genome Project - by the numbers. *Nature* **590**, 212–215 (2021).
39. G. N. Hounsfield, Computerized transverse axial scanning (tomography): Part 1. Description of system. *Br. J. Radiol.* **46**, 1016–1022 (1973).
40. B. M. Althouse, J. D. West, C. T. Bergstrom, T. Bergstrom, Differences in impact factor across fields and over time. *J. Am. Soc. Inf. Sci. Technol.* **60**, 27–34 (2009).
41. J. E. Hirsch, An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16569–16572 (2005).
42. A. L. Kinney, National scientific facilities and their science impact on nonbiomedical research. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17943–17947 (2007).
43. D. A. King, The scientific impact of nations. *Nature* **430**, 311–316 (2004).
44. B. Owens, Research assessments: Judgement day. *Nature* **502**, 288–290 (2013).
45. M. Park, E. Leahey, R. J. Funk, Papers and patents are becoming less disruptive over time. *Nature* **613**, 138–144 (2023).
46. C. Catalini, N. Lacetera, A. Oettl, The incidence and role of negative citations in science. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13823–13826 (2015).
47. Q. Ke, Comparing scientific and technological impact of biomedical research. *J. Inf.* **12**, 706–717 (2018).
48. Q. Ke, Technological impact of biomedical research: The role of basicness and novelty. *Res. Policy* **49**, 104071 (2020).
49. E. Sarigöl, R. Pfitzner, I. Scholtes, A. Garas, F. Schweitzer, Predicting scientific success based on coauthorship networks. *EPJ Data Sci.* **3**, 9 (2014).
50. X. Meng, O. Varol, A. L. Barabási, Hidden citations obscure true impact in science. arXiv [Preprint] (2023). https://arxiv.org/arXiv:2310.16181. Accessed 24 October 2023.
51. S. Fortunato *et al*., Science of science. *Science* **359**, eaao0185 (2018).
52. D. Wang, A. L. Barabási, *The Science of Science* (Cambridge University Press, 2021).
53. iCite, B. I. Hutchins, G. Santangelo, iCite Database Snapshots (NIH Open Citation Collection). *NIH Figs. Archive* (2019). 10.35092/yhjc.c.4586573.
54. A. J. Gates, A. L. Barabási, Reproducible science of science at scale: PySciSci. *Quant. Sci. Stud.*, 1–11 (2023).
55. Q. Ke, A network-based normalized impact measure reveals successful periods of scientific discovery across discipline. GitHub. https://github.com/qke/network-normed-c. Deposited 16 October 2023.