A century of physics

Roberta Sinatra, Pierre Deville, Michael Szell, Dashun Wang and Albert-László Barabási

An analysis of Web of Science data spanning more than 100 years reveals the rapid growth and increasing multidisciplinarity of physics — as well its internal map of subdisciplines.

The conventional narrative of physics is one of paradigm shifts[1]: from the Copernican Revolution to Einstein's *annus mirabilis*[2]. And for many, the stories would seem to involve genius in isolation — the lone physicist divorced from other sciences, unperturbed by societal beliefs. But the reality is quite different: physics has always been in a constant dialogue with other disciplines, be it mathematics, chemistry or theology. This dialogue is largely driven by methodology: what traverses disciplinary boundaries is the idea that complex phenomena can be understood in terms of a small number of universal laws[3].

In this era of interdisciplinary science, of biological physics, network science and econophysics, defining physics as the science of the properties of matter and energy[3] is increasingly outdated and inaccurate. We are therefore prompted to ask anew: what is physics? When two engineers accidentally discover cosmic microwave background radiation, is that physics or engineering? When a physicist uncovers the structure of DNA, is that biology or physics? Is the interdisciplinary role of physics something new — a potential fad — or has it always been an integral part of the field? Is physics dying or thriving, becoming more insular or more interdisciplinary? To answer these questions, we will rely on the very framework physics pioneered: collecting data from which to draw our conclusions.

What is physics?

The late Cambridge physicist Sam Edwards once remarked that "Physics is what physicists do."[4] Following in his footsteps, we define physics not from an epistemological point of view[3], but look instead at what physicists do. We do so by focusing on the research papers through which we communicate our basic discoveries — forcing us to ask: what exactly is a physics paper? A simplistic answer would be: it is a paper published in a physics journal. This narrow, yet obvious definition allows us to construct a core physics dataset of ~2.4 million papers published in 242 physics journals and documented in Web of Science (WoS) between 1900 and 2012.

The problem with the above definition is that many influential physics papers, and an increasing fraction associated with Nobel prizes, are published in interdisciplinary journals, such as *Nature* or *Science*. Furthermore, many papers of interest to physicists are published in journals of other disciplines. Take for example the founding paper of chaos theory, a thriving subfield of statistical physics, which was published in *Journal of the Atmospheric Sciences*[5]. This and many similar examples force us to address an important question: how do we identify papers that are not published in the physics literature, but given their subject matter and their impact on the evolution of physics, could or should have been?

To map out the complete physics literature, we compared the references and citations of all papers in WoS to a null model in which each paper's citations are assigned randomly, regardless of a paper's journal or research area[6]. A paper is a potential physics publication if its references and citations to the core physics literature are significantly higher than expected by chance[6,7]. Our algorithm recursively scans the 40 million papers published between 1900 and 2012 and documented in WoS, identifying ~5.1 million papers of potential interest to the physics community outside the core (Box 1).

This corpus contains two classes of papers: the first class consists of 4.5 million papers whose references are significantly biased towards core physics papers; this is the body of literature within the physics influence sphere. The second class contains 3.8 million papers heavily cited by core physics papers, representing papers of direct interest to the physics community. The intersection of these two classes consists of 3.2 million papers, distinguished by the fact that they reference the physics literature and are also cited by it in a statistically significant fashion. Hence, these papers are indistinguishable from the physics core, apart from their place of publication, prompting us to call them interdisciplinary physics papers.

Taken together, we find that the literature of direct interest to the physics community is more than twice that published by physics journals: on top of the 2.4 million core physics papers (Box 1c, blue), there are 3.2 million interdisciplinary papers published in non-physics journals (Box 1c, red), that, based on their referencing and citation patterns, are indistinguishable from papers published in physics journals. We identified six physics Nobel winning publications[8] in this interdisciplinary set, and many other highly influential physics papers, such as Hubbard's 1963 model of interacting particles[9] and Hopfield's 1982 paper on neural networks[10].

The growth of physics

Throughout the history of physics, major paradigm shifts, such as the development of quantum physics, have spurred significant new research, resulting in a burst of publications and giving birth to new and enduring subfields, from nuclear to condensed-matter physics[11]. The very existence of this growth is supported by the number of physics papers published each year (Fig. 1a), which has been increasing roughly exponentially for the past 110 years, an expansion that was halted temporarily only by the two World Wars. Note, however, that the growth rate of physics is indistinguishable from the growth of science in general[12]. Hence, the field's exponential growth is not driven by paradigm changes, but by societal needs, and capped by access to resources. This growth was particularly remarkable following World War II, when the physics literature doubled every 6.5 years. And yet, after 1970 this growth slowed, settling on its current rate of doubling every 18.7 years. Once again, the recent slowdown is not unique to physics, but characterizes the whole scientific literature contained in WoS. Finally, whereas pre-1910 physics literature was limited to physics journals, since the 1920s the growth of the core and interdisciplinary physics literature have been indistinguishable, indicating that publishing outside the physics core has been integral to the development of physics throughout the last century.

Box 1 | Defining physics

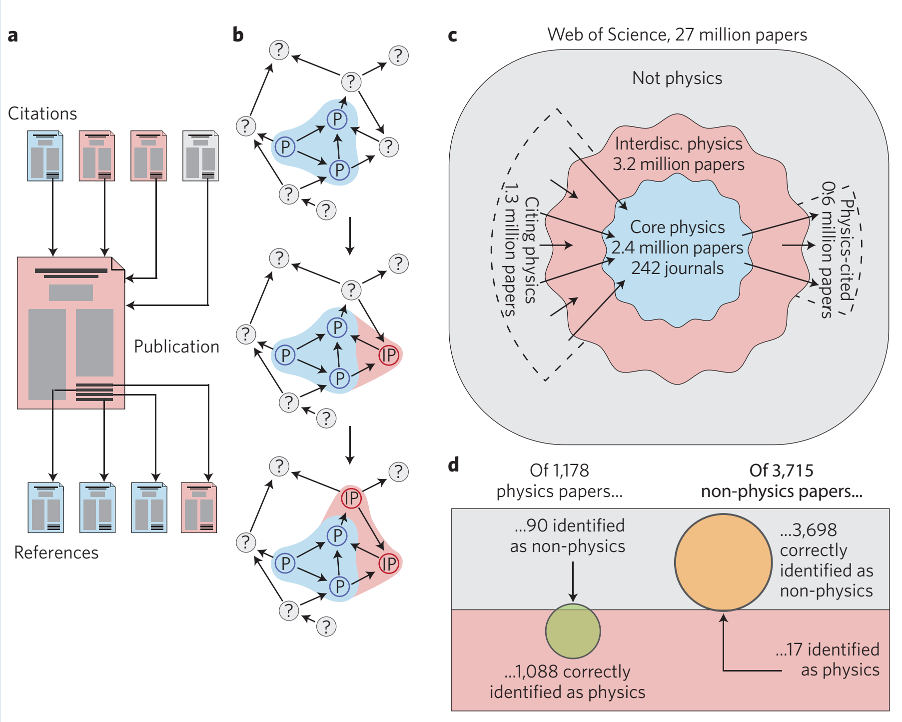
The key unit of communication within science is a research publication, whose references and citations contain considerable contextual information about the topic and the discipline of the paper (panel **a**). We mined this information to understand the nature and evolution of physics.

To identify the whole corpus of core physics papers, we started from 2.4 million papers published in 242 physics journals indexed in Web of Science (WoS), the list of journals being extracted by combining information from Wikipedia, Scopus and Scimago (www.scimagojr.com). These core physics papers are shown as blue nodes and marked with P in panel **b**.

To identify papers that are not in the core, but nevertheless belong to the physics literature, for each paper connected to the core by a reference or a citation, we measured the fraction of its links to the core physics literature. Then we estimated the expected fraction if references and citations were randomly reassigned to any paper in the WoS dataset, regardless of the discipline. If the observed fraction of references and citations to core physics papers was significantly larger than in the null model, we labelled the paper an interdisciplinary physics paper, shown as red nodes and marked with IP in panel **b**. The first three steps of this procedure depicted in panel **b** were repeated until the algorithm converged and no new papers were added.

In this way, our algorithm identified 3.2 million papers published outside the main physics literature that, with respect to their reference and citation patterns, are indistinguishable from the core physics papers (panel **c**).

To validate our algorithm, we measured its ability to detect a set of papers published outside the physics core that are classified as physics in *Science* and *Proceedings of the National Academy of Sciences USA* (true positives). We also tested the algorithm's ability to minimize the inclusion of papers that are known to have no relation to physics (false positives). The algorithm identifies 92.4% of the 1,178 physics papers as true positives, and includes only 0.5% of the 3,715 non-physics papers as false positives (panel **d**). To check that the limited coverage of WoS (for example, due to ceased journals before the database was created in the 1960s) did not influence our results, we ran all analyses on the dataset of American Physical Society papers, which contains a complete set of physics papers and citations from 1893 to 2010. We found no qualitative differences in the results.



The literature published in physics journals went from representing around 4% of the scientific literature in 1945 to about 10% after 1980, and has been approximately steady since then (Fig. 1b). Interdisciplinary physics has followed a similar pattern, growing from 6% in 1945 to a maximum of 18% in 1964, and stabilizing at 12% after 1980. The wider physics literature represents around 22% of all scientific literature since the 1980s, a remarkable fraction that documents the profound role and embeddedness of physics within the larger scientific enterprise.

Is the exponential growth of the physics literature driven by an exponential growth in the number of physicists, or by gradually increasing productivity? To answer this question, we used the disambiguated authorships of papers published by the American Physical Society (ref. 13 and Sinatra *et al.*, manuscript in preparation), finding that the number of authors has increased at the same rate as the number of papers (Fig. 1c). This leads us to conclude that the growth of physics literature is driven solely by the increasing number of authors. We do observe, however, nontrivial shifts in productivity. Indeed, whereas before 2000, a typical physicist co-authored fewer than one publication per year, in the past 15 years, the number of papers co-authored by each physicist jumped above one for the first time (Fig. 1d, black curve). Yet, this remarkable growth in productivity did not boost the field's overall productivity (Fig. 1c), as the total papers-per-author ratio dropped slightly (Fig. 1d, red curve). This indicates that the observed singular growth in individual productivity has its origins in collaborative effects: although on average each physicist continues to write fewer than one paper (Fig. 1a), she/he ends up as co-author of multiple publications. In other words, during the past decade, collaborative work has significantly boosted individual productivity when measured as whole paper counts[14,15], while leaving the field's overall productivity unchanged.

Impact and its variations

Is the steady growth of physics accompanied by a steady growth in impact? To address this question we used the average number of citations each paper acquires within 10 years of its publication, $\langle c_{10}\rangle$ (Fig. 2a) as a proxy for impact[16,17]. We find a puzzling peak in impact for papers published between 1955 and 1965, hinting at the existence of a 'golden age of physics' around 1960.

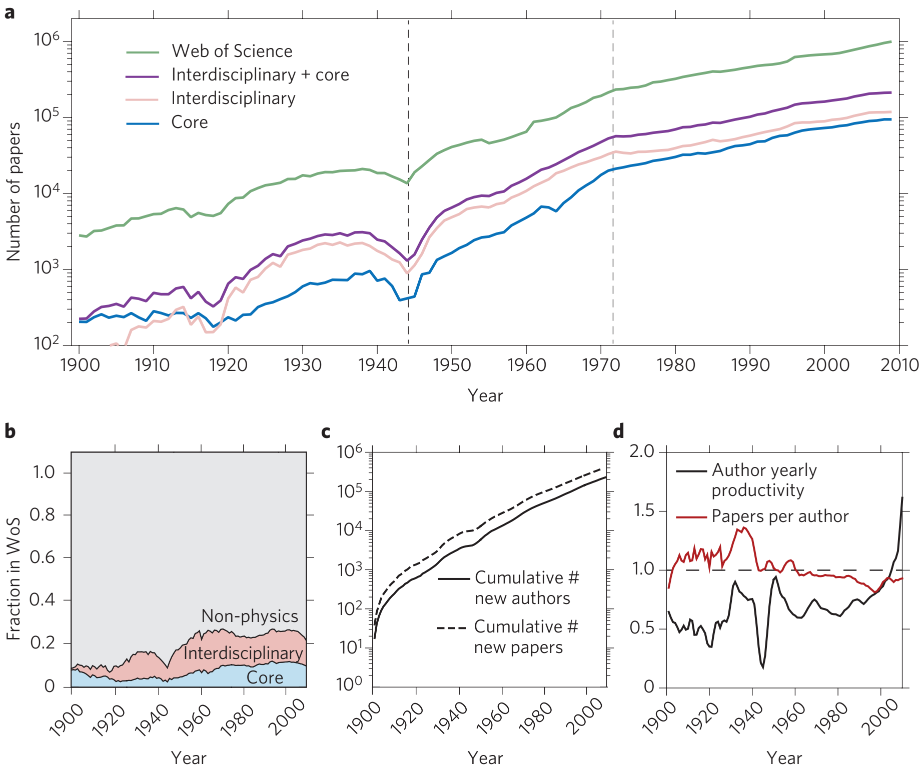


Figure 1 | The evolution of physics. **a**, Changes in the number of papers in physics and the Web of Science (WoS) data over the past century. The plot shows that the literature has been growing exponentially over time. The growth was steepest between 1950 and 1970, doubling every 6.5 years. After 1970, the doubling time increased to 18.7 years. Growth was only disrupted around 1915 and 1945 due to the World Wars. **b**, The fraction of physics core papers grew slowly over time, from 4% in 1920 to about 10% in 2000 (blue). The fraction of interdisciplinary physics papers (red) showed a dip during World War II, but recovered to a stable fraction of about 12% of the whole scientific literature. **c**, During the past century the number of authors increased at the same rate as the number of papers. **d**, Overall productivity, representing the number of papers per author, the ratio of the curves shown in panel **c**, has been fluctuating between 1900 and 1950 in the vicinity of one. In contrast, the number of papers co-authored by each physicist has been less than one during the past 100 years, but increased sharply in the past 15 years. This growth is a direct consequence of collaborative effects: individual productivity is boosted because physicists end up on many more papers as co-authors.

Interdisciplinary physics papers also had a small peak in the 60s, but in the 1990s their impact overcame their 60s peak. These peaks prompt us to ask: did core physics really peak in the 1960s?

A steady growth of citations is typically fuelled by two effects: a growing number of papers, and a growing number of references per paper. None of these curves shows discontinuities around the 1960s; however, the exponential growth in the literature was unabated in this period (Fig. 1a) and the average number of references per paper has been growing steadily from two references in 1900 to approximately 15 in 2000 (Fig. 2b). There was, however, a clear discontinuity in the way we cite papers, occurring during the 1960s. Indeed, in the late 1930s, physics was the most myopic it has ever been: citations were largely to papers published within the previous four years (Fig. 2c,d). After World War II, the average age of references dramatically increased, probably because there were too few papers published during the war period[18,19]. The scientific community simply continued where it left off before the war, citing pre-war papers.

As the literature recovered, the myopic citation style set in again — the age of references reaching a new low in 1960. This was followed, however, by a gradual but sustained change in citation habits that persists even today: physicists are systematically reaching deeper into the literature, citing older and older papers (Fig. 2c,d). Our hypothesis is that this dramatic change is a consequence of peer review. Indeed, before the 1960s, papers were accepted solely on editorial discretion[20]. The increasing specialization of science has created the need for expert opinion, which was difficult to implement until photocopying became available in 1959 (ref. 21).

Peer review became common in the 1960s — *Nature* officially introduced it in 1967. The reviewer's ability to point out relevant work not only boosted references to the older literature, but also induced a behavioural change, prompting the authors to more carefully credit earlier work. This change in referencing habits can explain the 1960s peak in impact. Indeed, whereas papers published in the late 1950s were still highly cited by the papers immediately following them, they also got an additional boost from later papers. And yet, papers published from the late 1960s suddenly had to compete for citations with much earlier papers, and in doing so they lost citations. Indeed, a null model indicates that a change in the referencing habits documented in Fig. 2d does induce a peak in citations for papers published in the 1960s (Fig. 2a). Hence, the 1960s peak in impact may not necessarily represent evidence of a golden age of physics; it is likely an unintended consequence of the emergence of deeper referencing, probably induced by the gradual introduction of peer review.

Figure 2a also documents a systematic gap in impact between core and interdisciplinary papers. To understand its origin, we explored the citation patterns of core and interdisciplinary physics in 1960, when the impact gap was the largest, and in 2000, when it was close to vanishing given the steady rise in the impact of interdisciplinary physics papers. As Fig. 2e indicates, in the decade 1950–1960, core physics was extremely self-referential, sending 11.1 times more citations to papers in physics journals than expected by chance. Interdisciplinary physics was much less self-referential, and with a fourfold increase in references towards physics journals, it was strongly coupled to core physics. In 2000, however, the self-referential nature of core physics significantly decreased, while that of interdisciplinary physics slightly increased. In other words, over the past four decades, core physics has lost some of its tribal nature, becoming more open to the vast non-physics literature. In contrast, interdisciplinary physics started to act as a separate discipline, tightly coupled with physics, but also increasingly tribal.

The inner structure of physics

Physics is not monolithic, but consists of a number of subfields, each with its own intellectual challenges, methodologies and culture. We used the Physics and Astronomy Classification Scheme (PACS) by the American Physical Society to classify the entire physics literature (core and interdisciplinary) into ten major subfields.

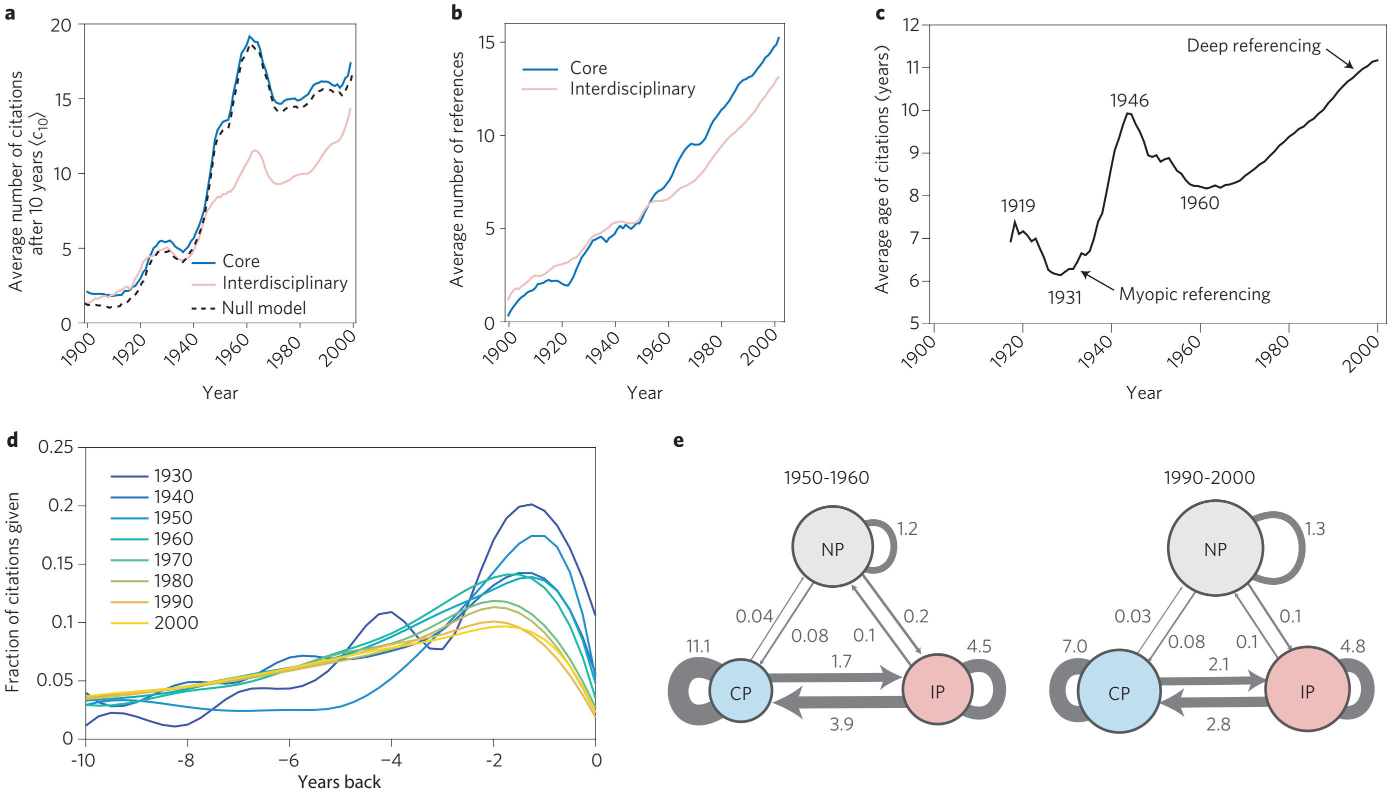


Figure 2 | The evolution of impact. **a**, We used the number of citations collected over 10 years, ⟨c_{10}⟩, as a proxy of a publication's long-term impact. The average impact of physics papers has grown from 1900 to 1950, and papers published between 1950 and 1960 received the most citations. The dotted line corresponds to a random null model, obtained by using the growth in number of papers (Fig. 1a), the increasing number of references (panel **b**) and different citation age distribution (panel **d**). These three time-dependent factors together can reproduce the observed impact peak in the 1960s. **b**, The average number of references per paper has been increasing steadily over time. **c**, The average age of the references of papers published in American Physical Society journals, documenting large variations in the depth of referencing over the past century. **d**, The probability of citing past papers for different years shows again remarkable shifts in citation patterns from myopic referencing in the 1930s–1950s to the current, increasingly deep referencing. **e**, The flow of citations between the core physics (CP), interdisciplinary physics (IP) and the non-physics (NP) literature in the decade before 1960 and before 2000. Node sizes are proportional to the logarithm of the number of papers published in each area. The weight of each link corresponds to the number of citations divided by expected citations, for instance, a weight 2 indicates that citations are twice the number of expected citations in a null model, where citations are randomly redistributed between areas.

We find that these subfields naturally cluster into three domains (Fig. 3a)[22]. The first and largest domain corresponds to condensed matter and interdisciplinary physics and related areas (IPR), where IPR corresponds to the subset of papers pertaining to PACS code 80, distinct from the interdisciplinary physics corpus defined in Box 1. The second domain incorporates topics from electromagnetism to atomic and plasma physics, whereas the third domain contains papers on particle, nuclear and astrophysics. The subfields within each domain cite each other in a statistically significant fashion, and tend not to cite subfields in other domains. These three domains are held together by general physics, which is significantly co-cited by multiple subfields from each domain.

The loops in Fig. 3a indicate the degree to which papers in each subfield cite papers in their own area, uncovering a remarkable pattern: the smaller a subfield, the more self-referential it is. Indeed, we find that a subfield's degree of self-referencing decreases as N^{-1} with the number of papers, N (Fig. 3b). For example, papers in plasma, nuclear and astrophysics, the smallest subfields, are 23 times more likely to reference themselves than expected by chance. In contrast, condensed matter, IPR and general physics, the four largest subfields, cite themselves only about four times more than expected by chance. This self-referential nature results in systematic citation barriers between some subfields (Fig. 3c). Nuclear physics suffers the most from this, having up to seven times fewer citations than expected by chance from the four largest subfields, and displaying a 23-fold self-referencing. General physics is absent in Fig. 3c because it is not undercited by any other subfield, affirming its central role in Fig. 3a.

To understand the long-term impact of papers in each subfield, we selected all papers published in 2000 within each PACS code, and fitted a citation model[17] to their citation trajectories, predicting their average number of citations over the next 20 years (Fig. 3d). These fits provide two key parameters: the ultimate impact, c^{∞}, representing the total number of citations a typical paper will collect during its lifetime, and impact time, T^{*}, representing the characteristic time over which a paper remains cited. As Fig. 3e indicates, papers published in electromagnetism and interdisciplinary physics have the largest ultimate impact, a typical paper in these areas collecting more than 50 citations over its lifetime. They also have a much longer impact time than other subfields (T^{*} = 11.74 and 14 years, respectively), indicating that high-impact papers require a long time to acquire their citations.

On the other extreme of the spectrum are papers published in particle and nuclear physics that burn out after 6–7 years, consequently collecting much fewer citations

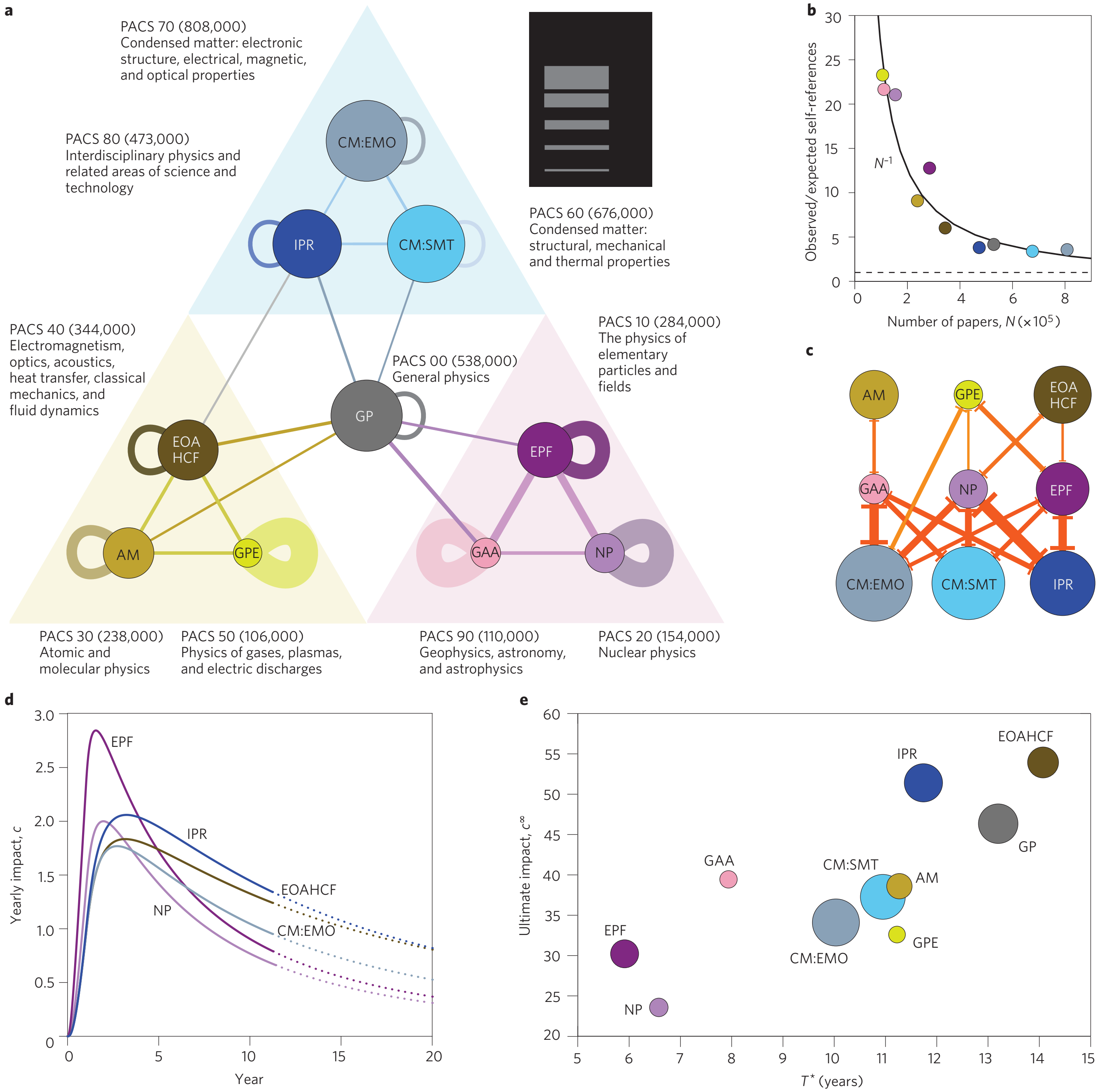


Figure 3 | The anatomy of physics. We classified each physics paper from the core and interdisciplinary literature into one of the ten major Physics and Astronomy Classification Scheme (PACS)-based subfields of physics, available after 1975. Since only 5% of these papers have PACS numbers, 3.7 million of them do not show subfield-specific citation bias, hence could not be assigned a subfield by our algorithm. The remaining 1.9 million papers were successfully assigned a subfield, allowing us to explore the inner structure of the physics literature. Note that 1.1 million papers have multiple PACS numbers, hence they may belong to multiple subfields. **a**, The co-citation patterns between the different physics subfields. Node sizes are proportional to the number of papers published in each subfield. Two subfields are connected if the number of citations between them significantly exceeds the expected citations, the line widths for all links, shown in the key, correspond to how many times the observed reference exceeds the random expectation. We find that core physics naturally clusters into three domains, as indicated by the colours of the clusters. General physics plays a central role, linking the three domains together. The loops on each node show the degree of self-referencing of the corresponding field. **b**, The smaller a subfield, the more its reference list is biased towards papers in the same subfield. **c**, A link between two subfields is evidence of citation barriers, where two areas have a significantly smaller number of citations than expected by chance. The highest citation barriers, up to nine times fewer citations than expected by chance, are between GAA and CM:EMO, and between NP and IPR. **d**, The yearly citations, c, of papers published in 2000 in selected subfields. Dotted curves indicate predictions. **e**, The ultimate impact, c^∞, representing the total number of citations a typical paper receives (that is, the area under the curves shown in panel **d**), versus impact time, T^*, for papers published in 2000 in each subfield, representing the typical time over which a paper collects its citations. The symbol size is proportional to the number of papers published in each subfield.

during their lifetime. Their low impact is partly explained by the insular nature of these subfields: both fields are about sevenfold undercited by the largest domain of physics (Fig. 3c), and show the highest degree of self-referencing. Taken together, Fig. 3 documents the rather heterogeneous nature of physics, consisting of subfields with widely different impact, longevity, internal culture and ability to interact with the rest of physics.

Where to from here?

Many of us involved in hiring committees or thesis defences in physics have been confronted with the question: can a particular body of work be considered physics, or is a particular scientist a physicist? The futility of the debate is often unnerving, and the analysis of the physics literature shows us why: there is not a single standard of what physics is. On one end there is the considerable corpus of work published in physics journals that, by the virtue of their publication venue, tends to automatically get a stamp of approval by most physicists. Yet, as we showed here, there exists a much larger physics literature, that in its subject matter and referencing patterns is indistinguishable from the core physics literature, yet it is published outside of physics journals. It contains many papers without which physics, as we know it, could not flourish.

The analysis of this extended corpus offers a treasure trove of quantitative information, unveiling the anatomy of the discipline. It demonstrates, for example, that our ability to define a field such as physics using sets of journals is long gone, and exposes the tribal nature of the different subdisciplines of physics: the smaller the subfield, the more self-referential it becomes. Nuclear and particle physics stand out in this context: they are not only the most self-referential subfields of physics, but are also separated by significant citation barriers from most other subfields. This isolation brings significant impact penalties: papers in those areas burn out very fast and have much lower ultimate impact than their peers in other subdisciplines. These patterns hide important vulnerabilities: research in ecology and technology adoption indicates that such monocultures, unable to link to their environment, are often destined to extinction[23].

Our understanding of the dialogue that physics shares with other disciplines is still in its infancy. Stumbling blocks include a lack of accurate records for the publications of each scientist, and the paucity of tools capable of mining the text of research papers. But given the rapid rate at which such tools are developed, we are advancing on an era in which hypotheses about the nature and evolution of a discipline or discovery will be testable in minute detail. Such advances offer a chance to go beyond sweeping generalizations such as the theory of paradigms[1], and ask deeper questions about the way physics changes, and how a certain discovery influences subsequent work[18], within and outside of physics. Given that the most fruitful discoveries come from the cross-pollination of different fields — recombining previously disconnected ideas and resources[24] — such quantitative understanding could help us identify as yet unexplored research areas. A better understanding of the optimal paths of innovation may also change the way we fund scientists and institutions to unlock their creative potential and enhance their long-term impact. ❐

Roberta Sinatra, Pierre Deville, Michael Szell and Albert-László Barabási* are at the Center for Complex Network Research and Physics Department, Northeastern University, Boston, Massachusetts 02115, USA. Dashun Wang is at the College of Information Sciences and Technology, Pennsylvania State University, University Park, Pennsylvania 16801, USA. P.D is also in the Department of Applied Mathematics, Université catholique de Louvain, Louvain-la-Neuve 1348, Belgium. A.L.B. is also at the Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA, and in the Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA, and at the Center for Network Science, Central European University, Budapest 1051, Hungary.*
**e-mail: r.sinatra@neu.edu; alb@neu.edu*

References

1. Kuhn, T. *The Structure of Scientific Revolutions* (Univ. Chicago Press, 1996).
2. Barabasi, A-L. *Nature Phys.* **8,** 14–16 (2011).
3. Main, P. & Tracy, C. *Phys. World* **26** (4), 17–18 (2013).
4. Donald, A. *Phys. Biol.* **11,** 053008 (2014).
5. Lorenz, E. N. *J. Atmos. Sci.* **20,** 130–141 (1963).
6. Fortunato, S. *Phys. Rep.* **486,** 75–174 (2010).
7. Zhu, X. *Semi-Supervised Learning Literature Survey* Tech. Rep. 1530 (Univ. Wisconsin-Madison, 2005).
8. Shen, H-W. & Barabasi, A-L. *Proc. Natl Acad. Sci. USA* **111,** 12325–12330 (2014).
9. Hubbard, J. *Proc. R. Soc. Lond. A* **276,** 238–257 (1963).
10. Hopfield, J. J. *Proc. Natl Acad. Sci. USA* **79,** 2554–2558 (1982).
11. Mehra, J. *The Golden Age of Theoretical Physics* (World Scientific, 2001).
12. Bornmann, L. & Mutz, R. *J. Assoc. Inf. Sci. Technol.* (2015).
13. Deville, P. *et al. Sci. Rep.* **4,** 4770 (2014).
14. Wuchty, S., Jones, B. & Uzzi, B. *Science* **316,** 1036–1039 (2007).
15. Pavlidis, I., Petersen, A. M. & Semendeferi, I. *Nature Phys.* **10,** 700–702 (2014).
16. Radicchi, F. & Castellano, C. *Phys. Rev. E* **83,** 046116 (2011).
17. Wang, D., Song, C. & Barabási, A-L. *Science* **342,** 127–132 (2013).
18. Redner, S. *Phys. Today* **58** (6), 49–54 (2005).
19. Nakamoto, H. *Synchronous and Diachronous Citation Distribution* (Elsevier, 1988).
20. Burnham, J. C. *J. Am. Med. Assoc.* **263,** 1323–1329 (1990).
21. Spier, R. *Trends Biotechnol.* **20,** 357–358 (2002).
22. Pan, R. K., Sinha, S., Kaski, K. & Saramaki, J. *Sci. Rep.* **2,** 551 (2012).
23. Arthur, W. B. *The Nature of Technology: What it is and How it Evolves* (Simon and Schuster, 2009).
24. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. *Science* **342,** 468–472 (2013).

Acknowledgements

R.S., M.S. and A.L.B. are supported by the AFOSR grant FA9550-15-1-0077. P.D. is supported by the National Foundation for Scientific Research (FNRS) and by the Research Department of the Communaute Francaise de Belgique (Large Graph Concerted Research Action). D.W. is supported by the AFOSR, grant FA9550-15-1-0162. R.S. is also supported by the James S. McDonnell Foundation. A.L.B. is also supported by the Future and Emerging Technologies Project 317 532 'Multiplex' financed by the European Commission. We wish to thank I. Georgescu for many fruitful discussions and ideas throughout this project and J. De Nicolo for support with handling the WoS data.