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science and econophysics, defining physics as the science of the properties of matter and energy[3] is increasingly outdated and inaccurate.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

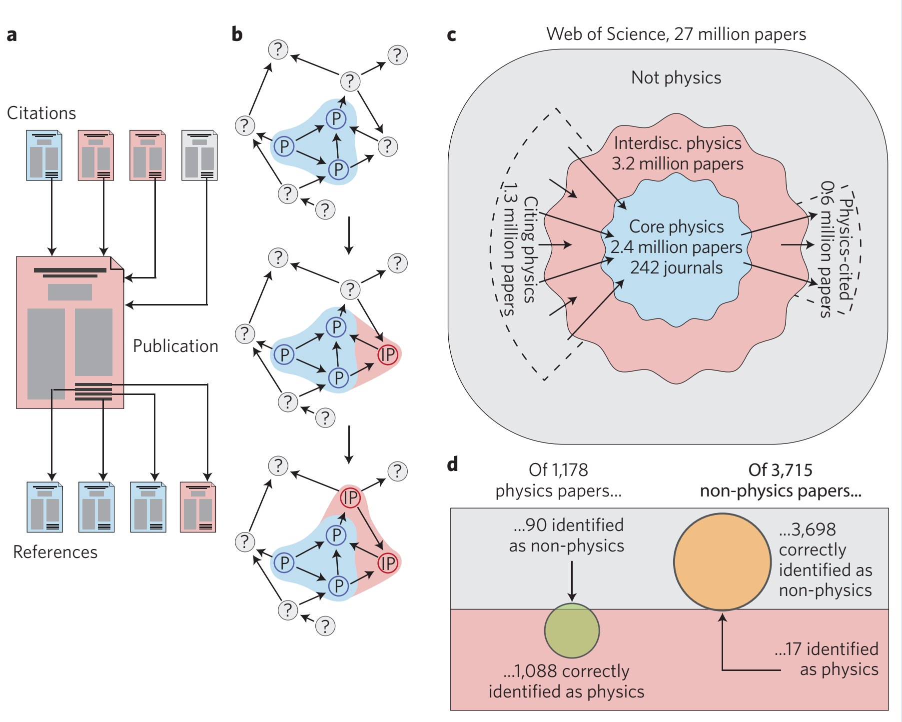
The key unit of communication within science is a research publication, whose references and citations contain considerable contextual information about the topic and the discipline of the paper (panel **a**). We mined this information to understand the nature and evolution of physics.

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.