Supplemental Information

Nature's reach: narrow work has broad impact

Alexander J. Gates[1], Qing Ke[1], Onur Varol[1], and Albert-László Barabási[1,2,3,*]
[1]Center for Complex Network Research, Northeastern University, Boston, Massachusetts 02115, USA
[2]Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA
[3]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA
* To whom correspondence should be addressed: a.barabasi@northeastern.edu

Contents

Acknowledgements

A.J.G., Q.K., O.V. and A.-L.B. were supported in part by the Templeton Foundation, contract #61066, and the Air Force Office of Scientific Research under award number FA9550-19-1-0354.

S1 An Overview of Disciplinary Research

By steering authors and ideas into expert communities that share the same methods, techniques, and employ a common conceptual framework, the emergence of scientific disciplines has deepened scientific discourse and offered intellectual safe harbors for new concepts to develop their true potential. Yet, today many pressing scientific challenges straddle the traditional disciplinary boundaries, requiring tools and expertise that are rarely present within a single discipline[1,2]. Furthermore, scientists often look at other fields for inspiration and tools even as they try to address the challenges of their own discipline[3]. This has fueled interest from the National Science Foundation and other institutions to identify and nurture research that extends beyond a single discipline[2,4,5].

There are multiple ways in which disciplines can be combined and interact, prompting the useful distinction between *crossdisciplinarity*, *multidisciplinarity*, *interdisciplinarity*[6]:

1. **Crossdisciplinarity** occurs when knowledge from one discipline is applied to problems defined by another discipline;

2. **Multidisciplinarity** occurs when knowledge from different disciplines coexists in the same venue, but the disciplinary perspectives themselves do not change[6];

3. **Interdisciplinarity** occurs when knowledge from multiple disciplines is synthesized into a coherent whole[6–8].

Given the importance of *crossdisciplinarity*, *multidisciplinarity*, *interdisciplinarity*, there have been several methods introduced for identifying and quantifying research that extends beyond individual disciplines. The most common techniques focus on the article citation graph[9], and measure the combination of disciplines in the references or citations for an article, journal, author, or institution. To start, the disciplines associated with a scientific publication can be defined according to three distinct dimensions:

a) **Inspiration**, representing the set of concepts, methods and techniques that the work relies on, as captured by the disciplines defined by the *references* chosen by the authors.

b) **Venue**, captured by the choice of the *journal* by the authors, which signals the scientific community the work is intended to speak to and defines the conceptual framework used to communicate its message.

c) **Impact**: The scientific communities that build on the published research, captured by the disciplines that cite the publication.

The most common conceptualization of interdisciplinarity is the integration of knowledge, through looking at the diversity of disciplines, allowing tools originally developed to measure biological diversity to be applied in the bibliometric setting[10]. Measures under this category are one-dimensional, including the number of disciplines[11], the fraction of disciplines[12,13], the Gini impurity index, Simpson's index[14], the Gini coefficient, and discipline entropy[15,16], among others[17–19]. For a more comprehensive review, see ref.[10].

More recent proposals have emphasized that interdisciplinarity is not only about the diversity but also the relatedness of disciplines[20–22]. This has prompted the introduction of multi-dimensional indicators. The most commonly discussed dimensions in the literature are: variety, balance, and disparity[21,22]. Variety counts the number of disciplines, balance quantifies the diversity of disciplines, and disparity measures the relatedness between disciplines.

The most frequently used measure to characterize interdisciplinarity by integrating over these dimensions is the Rao-Stirling index[10,23–26], see Section S5.2 below for more details. A new proposal, DIV*, attempts to separate the influence of balance, variety, and disparity into independent contributing factors[27,28]. However, the debate continues in the literature about the appropriate construction of measures for interdisciplinarity[29,30].

S2 Data Description

The primary source of publication data for this project is the Clarivate Analytics' Web of Science Core Collection (WoS) database, covering the Science Citation Index Expanded and the Social Sciences Citation Index. We considered all articles, reviews, and letters published between 1900 to 2017, and excluded all other types of documents (e.g. editorials, letters to the editor, and book reviews), that are not always peer-reviewed. In total we consider 53,788,499 publications and we extracted all references, resulting in 694,439,758 citation relationships. Finally, we focus our analysis on the 19,252,639 publications that had at least one reference and one citation, and were published before 2010, covering 38,279,893 publications with their disciplinary labels that appeared as a reference or citation to a focus paper. For the journal specific subsets, we focus on the 88,637 articles published in *Nature*, 52,367 articles published in *Science*, 82,456 articles published in *PNAS*.

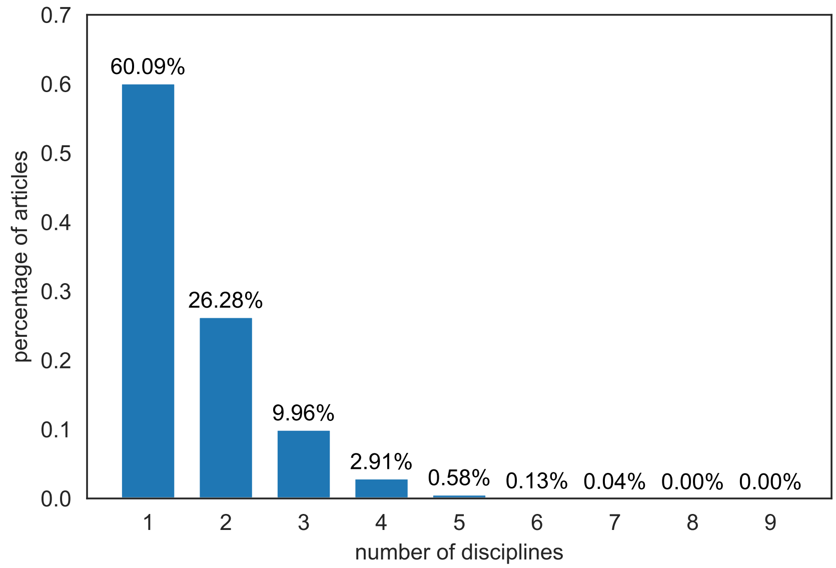

Figure S1. The number of multiple discipline assignments. The percentage of all articles in the WoS assigned to a specific number of disciplines (x-axis). All articles have at least one assignment and more than 99% of articles have 4 or fewer discipline assignments.

The WoS dataset assigns each article to at least one scientific discipline in a three-layer hierarchy over 251 disciplines. For example, a paper is assigned to "Science & Technology" (top layer), "Life Sciences & Biomedicine" (middle layer) and "Biophysics" (leaf layer). The assignment is primarily based on each publication's journal information, but a few multidisciplinary journals provide article-specific categories. For our purposes, we use the 251 disciplines in the bottom layer of the hierarchy, offering the finest differentiation of disciplines.

While many other disciplinary classification schemes exist at both the journal and publication level, the WoS hierarchical classification is regarded as the most consistent with respect to its references[23,31].

S3 Article Vectors

We represent each article by a vector over the 251 disciplines, a, where the entry in dimension a_i represents the article's weighting in discipline i. Let us first consider the construction for article references.

Given the reference list with n references, $\{r_1, \ldots, r_n\}$, we map each reference, r_α, to a set of k_α disciplines, $d_{r_\alpha} = \{d_1, \ldots, d_{k_\alpha}\}$ based on the journal in which the reference was published (Fig. S2). The ith

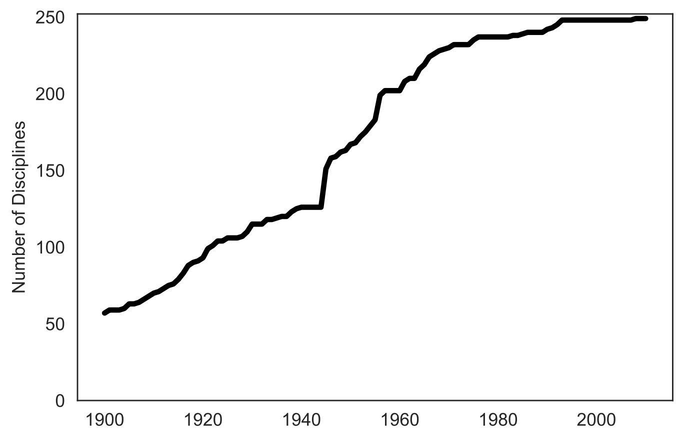

Figure S2. The cumulative number of disciplines with at least one article.

entry in the article reference vector is then given by

$$a_i = \sum_{\alpha=1}^{n} \sum_{\beta=1}^{k_\alpha} \frac{\delta(d_\beta, i)}{k_\alpha}, \tag{S1}$$

where δ is the delta function equaling 1 when discipline d_β is the same as discipline i. Note that the total summation over the vector is equal to the number of references, $\sum_i a_i = n$. The same process is used to construct a representation for citations to an article.

S4 Knowledge Flows

The construction of the knowledge flows starts by identifying the 88,637 *Nature* articles with at least one reference and citation. For each Nature paper we find the matrix formed as the outer product of the reference based article vector and the citation based article vector. Finally, we sum over all of the outer product matrices (one for each paper). The resulting matrix, with rows and columns corresponding to the disciplines, counts the number of papers with a fraction of its references from a specific discipline, which eventually are cited by another discipline.

S5 Co-citation Network

The co-citation network was formed by identifying the 132,169 *Nature* articles with at least one citation in the WoS. We then took all co-citation relationships between these articles from the rest of the web of science. Finally, we filtered out links with low co-citation and identified the largest connected component of the graph. The discipline assignment is based on the most frequently occurring discipline within the citation list for each article (in the case of ties, we randomly select one discipline).

S6 Journal Multidisciplinarity

To characterize the multidisciplinarity of a journal, we first need to assign each article to a discipline. Here, we refer to an article's references, and create the vector representation as in section S3. We then assign each article to the most frequently occurring discipline within the reference list (in the case of ties, we randomly select one discipline).

Each journal is represented by a vector over the 251 disciplines, where each entry counts the number of articles it has published that were assigned to that discipline. We normalize the vector to give the probability that a randomly selected article published by the journal was inspired by references from each discipline. The entropy of this distribution describes the overall spread of the disciplines published by the journal; an entropy of $\log_2(1) = 0$ reflects journals that publish articles solely from a single discipline, while an entropy of $\log_2(251) \approx 8$ reflects journals that publish an equal number of articles from each discipline. To facilitate consistent interpretation, we normalize the entropy by this maximum value, such that the normalized entropy exists between 0 (deeply disciplinary) to 1 (equally across all 251 disciplines). Specifically, for a journal with a article vector p, the normalized multidisciplinary entropy is

$$H(p) = -\frac{1}{\log_2(251)} \sum_{i=1}^{251} p_i \log_2(p_i). \tag{S2}$$

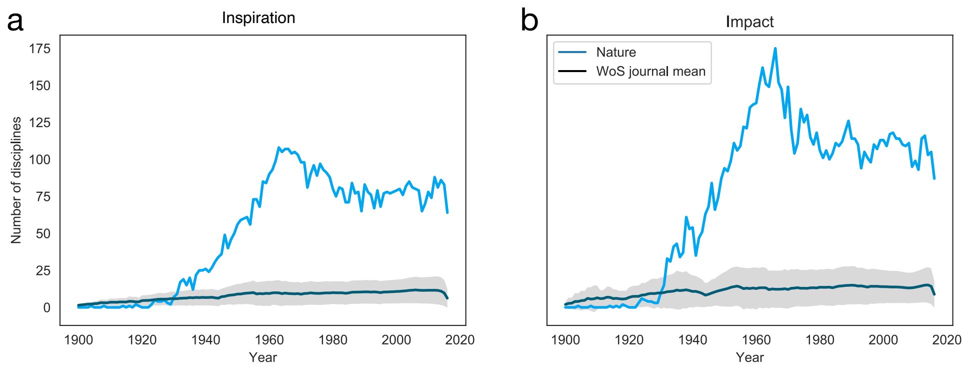

Figure S3. The disciplinary coverage of the average journal in the WoS. For each journal in each year, we find total number of disciplines published by at least one article in that journal when we assign disciplines based on the articles' **a**, inspiration or **b**, impact. The average for all journals in the WoS (black), with the shaded grey region denoting one standard deviation, is considerably smaller than the exact number of disciplines appearing in *Nature* (blue). Note that disciplines appear with varying frequency in the journals, a feature not captured by this figure, but reflected in the entropy measures (Figure S5).

S7 Interdisciplinarity

S7.1 Disciplinary Distances

While any assessment of interdisciplinary diversity will depend on the specific definition, size, and relative scale of the disciplines under consideration, we can mitigate the consequences of these decisions by considering the background distances between the disciplines. Such a distance measure captures the fact that some disciplines are naturally more related to each other, and thus, their combination is less indicative of disciplinary diversity. For example, the knowledge captured by the disciplines Molecular Biology and Organic Chemistry is much more closely related (therefore less diverse) than the knowledge captured by the disciplines Molecular Biology and Nuclear Physics. Here, we capture this overall knowledge level within a discipline by considering the cumulative reference or citation vectors v_i over the set of articles within the discipline, $\mathscr{D}_i$, where $v_i = \sum_{a \in \mathscr{D}_i} a$. The background distance, d_{ij}, between discipline i and discipline j is then

$$d_{ij} = 1 - \frac{v_i \cdot v_j}{|v_i| \cdot |v_j|}. \tag{S3}$$

In other words, disciplines whose articles have very similar reference or citations patterns will have a small distance $d_{ij} \approx 0$, while disciplines whose articles have very different reference or citations patterns will have a large distance $d_{ij} \approx 1$.

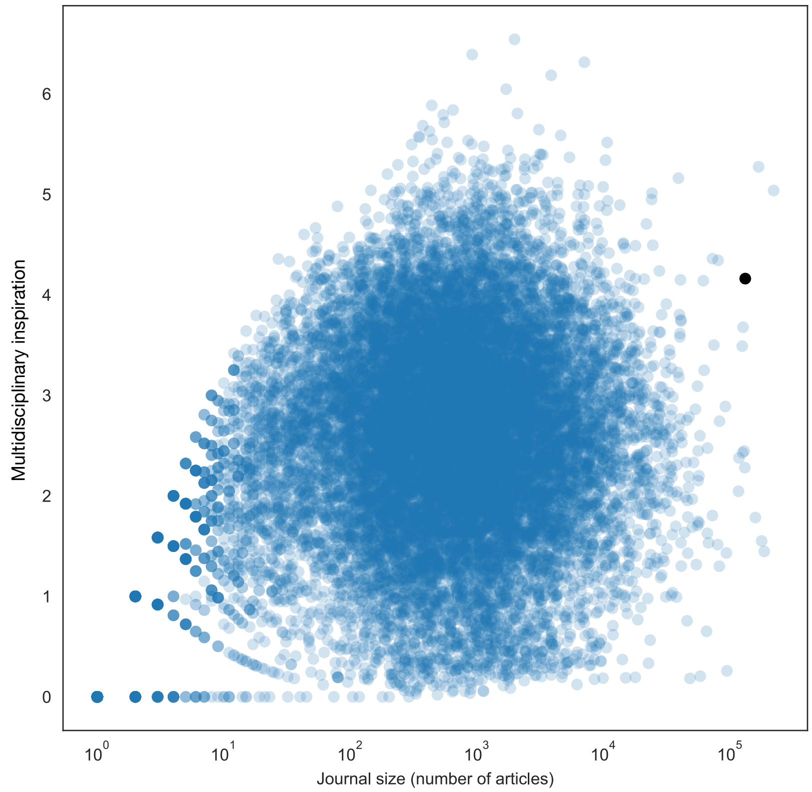

Figure S4. The multidisciplinary inspiration of journals vs their volume. The multidisciplinary inspiration of a journal is independent of the number of articles it has published (Pearson $r = 0.02$). *Nature* appears the as the black dot in the upper righthand side of the figure.

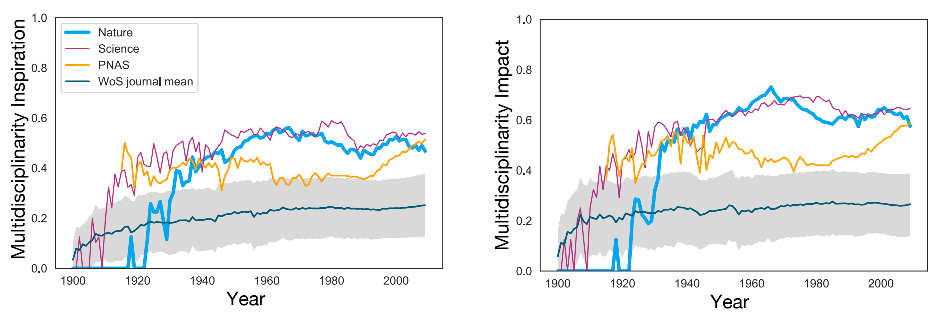

Figure S5. The temporal multidisciplinary inspiration and impact. The **a**, multidisciplinarity inspiration and **b**, multidisciplinarity impact has been growing over the last century.

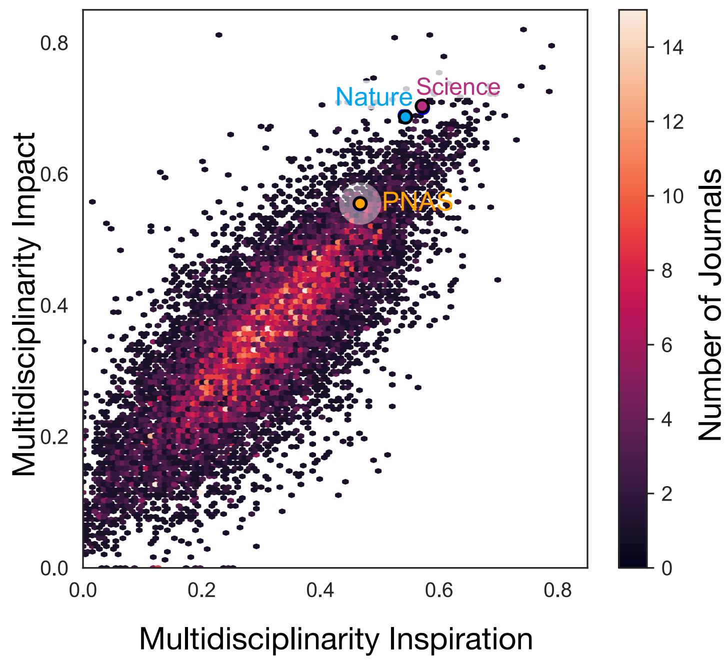

Figure S6. The multidisciplinary inspiration and impact for WoS journals. The multidisciplinarity inspiration vs impact of scientific journals reveals the high degree of variability in how journals combine disciplines. The general science journals *Nature*, *Science*, and *PNAS* all score highly on both measures.

S7.2 Rao-Stirling Diversity

For each article, we now consider the relative proportion of the weight in each dimension by normalizing the vector representation $p = \frac{a}{|a|}$. The Rao-Stirling Diversity heuristic[10] is then given by

$$RS(a) = 2 \cdot \sum_{i \neq j} d_{ij} p_i p_j, \tag{S4}$$

where p_i (p_j) is the fraction of references (or citations) in discipline i (j), and d_{ij} is the aforementioned distance between discipline i and discipline j. A RS score of 0 reflects a set with no diversity (i.e. all references or citations are from the same discipline), while a RS score of 1 denotes maximal diversity. Note that when the distances are all maximal (1) then the RS diversity heuristic is equivalent to the Gini index, a popular measure of inequality.

In addition to using the RS diversity, we also replicated the analysis using DIV* with a fixed number of disciplines. We found that RS and DIV* are fairly correlated at the level of individual articles. Furthermore,

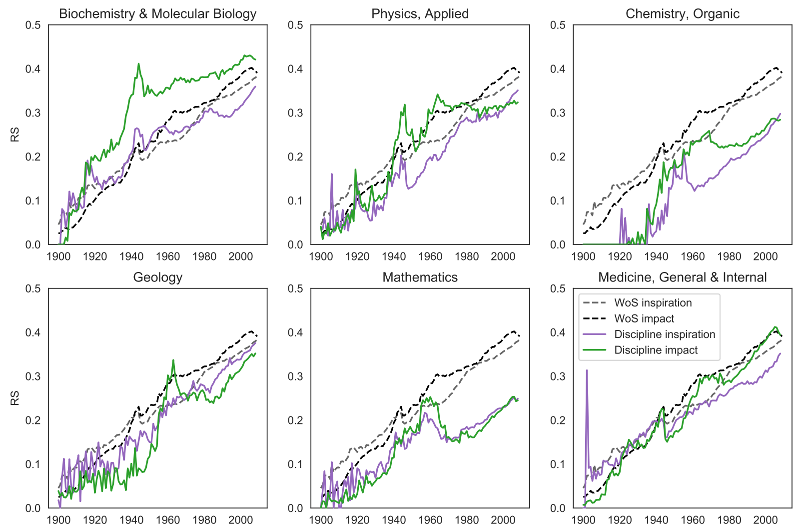

Figure S7. The interdisciplinarity of scientific disciplines. Average Rao-Stirling diversity of scientific publications in 6 scientific disciplines based on (purple) references and (green) citations compared to the average RS across all articles for (black dashed) references and (grey dashed) citations.

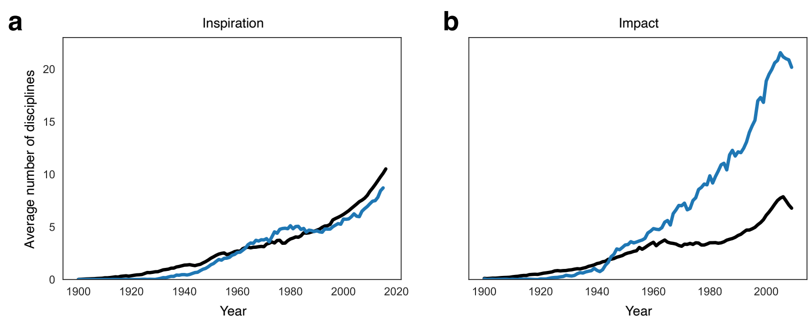

Figure S8. The average number of disciplines inspiring or impacting articles. Average number of disciplines that appear in **a**, references and **b**, citations for articles in the WoS (black) and *Nature* (blue).

the average trends of these replicated calculations qualitatively agree with the original RS results.

S7.3 Finite sampling and the Rao-Stirling Diversity

In general, the RS diversity is dependent on the number of references, and citations for each article. This arises from the fact that more samples allow for a greater spread over the dimensions of the article vectors. For example, consider an article with only 2 references, the maximum RS is then $2 * \frac{1}{2} * \frac{1}{2} = 0.5$. On the other hand, an article with 5 references would have a maximum RS value of $2 * \binom{5}{2} \frac{1}{5} * \frac{1}{5} = 0.8$

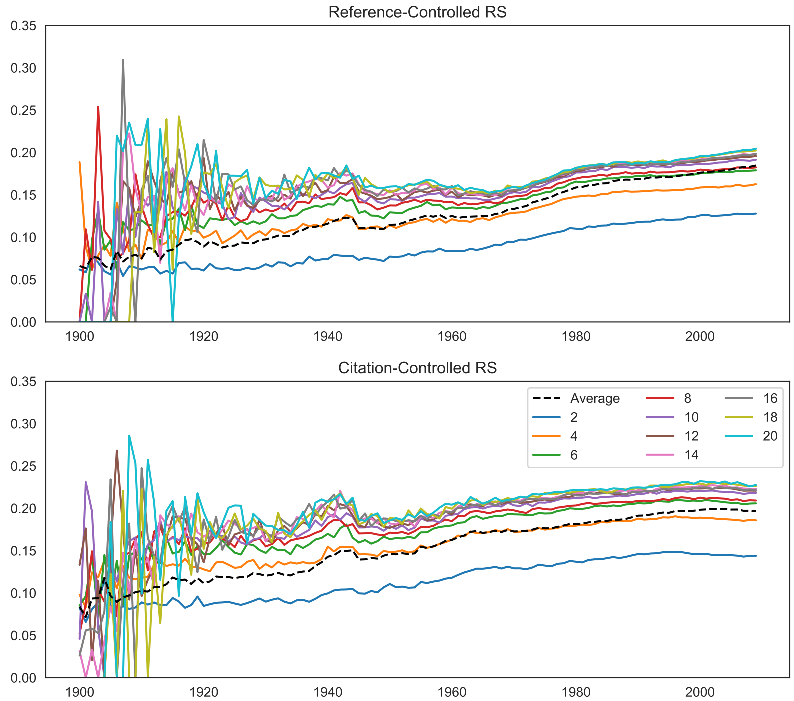

Figure S9. The RS heuristic conditioned on the number of references or citations. Average Rao-Stirling diversity of scientific publications for (top) references and (bottom) citations. The average is considered over (dash-black) all articles in the WoS, or (blue, pink, green, etc.) all articles with the same number of references/citations (2,4,6, etc.).

Here, as shown in Fig. S9**a,b**, we found that the average trends over all articles matches the qualitative trends found conditioned on the number of references or citations. Therefore, to simplify our presentation, we only report the average statistic over all articles regardless of the reference or citation count.

S8 Crossdisciplinarity

To capture the crossdisciplinarity of a publication, we find the distance between the disciplinary inspirations to the article (calculated from the article reference set), and the disciplinary impact of the article (calculated from the article citation set) using the Jensen-Shannon divergence (JSD). The JSD is a symmetric distance measure between two probability distributions. Specifically, given an article reference vector r and an article's citation vector c, as defined in section S3, we find their average probability vector $m = \frac{r+c}{2}$ and calculate the

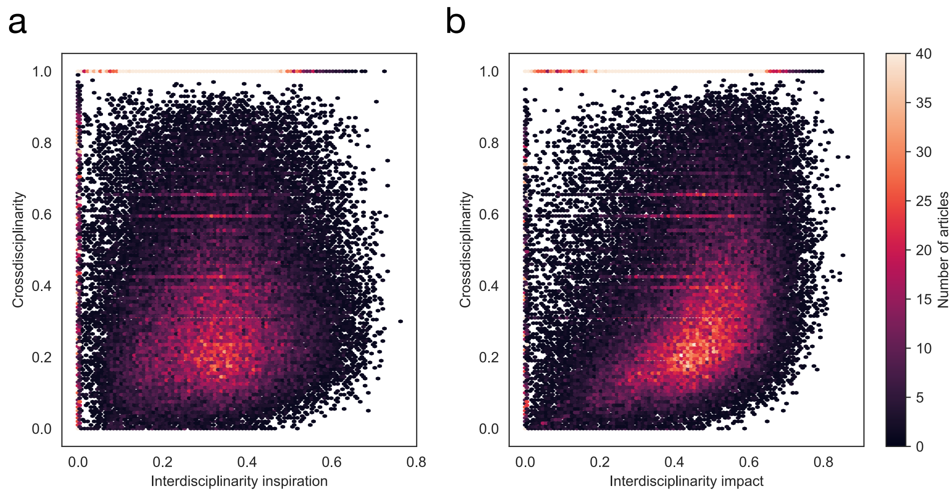

Figure S10. Interdisciplinarity and Crossdisciplinarity of *Nature* articles are independent. The crossdisciplinarity vs **a**, interdisciplinarity inspiration (Pearson r $= -0.17$) and **b**, interdisciplinarity impact (Pearson r $= 0.0$).

JSD as

$$JSD(r,c) = \sum_i \frac{1}{2}\left(r_i \log_2(r_i) + c_i \log_2(c_i)\right) - m_i \log_2(m_i). \tag{S5}$$

A JSD $= 0.0$ implies that the distribution of disciplines that inspired the article was exactly the same as the distribution of disciplines that the article impacted, while a JSD $= 1.0$ denotes maximal crossdisciplinarity in that the disciplines that inspired the article were completely different from the disciplines the article impacted.

References

1. Ledford, H. How to solve the world's biggest problems. *Nature News* **525**, 308 (2015).
2. Lowe, P. & Phillipson, J. Reflexive interdisciplinary research: The making of a research programme on the rural economy and land use. *Journal of Agricultural Economics* **57**, 165–184 (2006).
3. Stember, M. Advancing the social sciences through the interdisciplinary enterprise. *The Social Science Journal* **28**, 1–14 (1991).

4. National Academy of Sciences and National Academy of Engineering and Institute of Medicine. *Facilitating Interdisciplinary Research* (The National Academies Press, Washington, DC, 2005).

5. Cech, T. R. & Rubin, G. M. Nurturing interdisciplinary research. *Nature Structural & Molecular Biology* **11**, 1166–1169 (2004).

6. Klein, J. T. *Interdisciplinarity: History, theory, and practice* (Wayne state university press, 1990).

7. Choi, B. C. K. & Pak, A. W. P. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. definitions, objectives, and evidence of effectiveness. *Clin Invest Med* **29**, 14 (2006).

8. Wagner, C. S. *et al.* Approaches to understanding and measuring interdisciplinary scientific research: A review of the literature. *Journal of informetrics* **5**, 14–26 (2011).

9. Fortunato, S. *et al.* Science of science. *Science* **359** (2018).

10. Stirling, A. A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface* **4**, 707–719 (2007).

11. MacArthur, R. H. Patterns of species diversity. *Biological reviews* **40**, 510–533 (1965).

12. Larivière, V. & Gingras, Y. On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology* **61**, 126–131 (2010).

13. Van Noorden, R. Interdisciplinary research by the numbers. *Nature* **525**, 306–307 (2015).

14. Simpson, E. H. Measurement of diversity. *Nature* **163**, 688 (1949).

15. Carayol, N. & Thi, T. U. N. Why do academic scientists engage in interdisciplinary research? *Research Evaluation* **14**, 70–79 (2005).

16. Silva, F. N., Rodrigues, F. A., Oliveira, O. N. & da F. Costa, L. Quantifying the interdisciplinarity of scientific journals and fields. *Journal of Informetrics* **7**, 469–477 (2013).

17. Steele, T. W. & Stier, J. C. The impact of interdisciplinary research in the environmental sciences: a forestry case study. *Journal of the American Society for Information Science* **51**, 476–484 (2000).

18. Rinia, E. J., van Leeuwen, T. N., van Vuren, H. G. & van Raan, A. F. J. Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research. *Research Policy* **30**, 357–361 (2001).

19. Huutoniemi, K., Klein, J. T., Bruun, H. & Hukkinen, J. Analyzing interdisciplinarity: Typology and indicators. *Research Policy* **39**, 79–88 (2010).

20. Larivière, V., Haustein, S. & Börner, K. Long-distance interdisciplinarity leads to higher scientific impact. *PLOS ONE* **10**, e0122565 (2015).

21. Wang, J., Thijs, B. & Glänzel, W. Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. *PLOS ONE* **10**, e0127298 (2015).

22. Yegros-Yegros, A., Rafols, I. & D'Este, P. Does interdisciplinary research lead to higher citation impact? the different effect of proximal and distal interdisciplinarity. *PLOS ONE* **10**, e0135095 (2015).

23. Porter, A. L., Roessner, D. J. & Heberger, A. E. How interdisciplinary is a given body of research? *Research Evaluation* **17**, 273–282 (2008).

24. Porter, A. & Rafols, I. Is science becoming more interdisciplinary? measuring and mapping six research fields over time. *Scientometrics* **81**, 719–745 (2009).

25. Zhang, L., Rousseau, R. & Glänzel, W. Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the Association for Information Science and Technology* **67**, 1257–1265 (2016).

26. Cassi, L., Champeimont, R., Mescheba, W. & de Turckheim, E. Analysing institutions interdisciplinarity by extensive use of rao-stirling diversity index. *PLOS ONE* **12**, e0170296 (2017).

27. Leydesdorff, L., Wagner, C. S. & Bornmann, L. Betweenness and diversity in journal citation networks as measures of interdisciplinarity—a tribute to eugene garfield. *Scientometrics* **114**, 567–592 (2018).

28. Leydesdorff, L., Wagner, C. S. & Bornmann, L. Interdisciplinarity as diversity in citation patterns among journals: Rao-stirling diversity, relative variety, and the gini coefficient. *Journal of Informetrics* **13**, 255–269 (2019).

29. Wang, Q. & Schneider, J. W. Consistency of interdisciplinarity measures. *arXiv preprint arXiv:1810.00577* (2018).

30. Rousseau, R. On the leydesdorff-wagner-bornmann proposal for diversity measurement. *Journal of Informetrics* **13**, 906–907 (2019). DOI 10.1016/j.joi.2019.03.015.

31. Wang, Q. & Waltman, L. Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics* **10**, 347–364 (2016).

Number	Title	Inspiration RS	Impact RS	JSD
1	*Single-channel current recorded from membrane of denervated frog muscle fibres*	0.233072	0.461258	0.252984
2	*Sexually Mature Individuals of Xenopus laevis from the Transplantation of Single Somatic Nuclei*	0.322432	0.468268	0.311278
3	*Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid*	0.368546	0.53371	0.557003
4	*Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism*	0.494494	0.501228	0.402304
5	*A Jupiter-mass companion to a solar-type star*	0.27947	0.34229	0.609987
6	*Large losses of total ozone in Antarctica reveal seasonal ClOx/NOx interaction*	0.536028	0.803054	0.689392

Table S1. Six focus articles honoring *Nature*'s 150th anniversary. These six articles were selected by *Nature*'s editors to capture the diversity of scientific publications in the journal. Here we use them to illustrate our disciplinary measures.