nature physics

SUPPLEMENTARY INFORMATION
https://doi.org/10.1038/s41567-018-0314-6

In the format provided by the authors and unedited.

A Nobel opportunity for interdisciplinarity

Michael Szell [1,2,3,4], Yifang Ma [5,6] and Roberta Sinatra [1,2,3,7,8*]

[1]Department of Network and Data Science, Central European University, Budapest, Hungary. [2]Center for Complex Network Research, Northeastern University, Boston, MA, USA. [3]Complexity Science Hub, Vienna, Austria. [4]MTA KRTK Agglomeration and Social Networks Lendulet Research Group, Centre for Economic and Regional Studies, Hungarian Academy of Sciences, Budapest, Hungary. [5]Northwestern Institute on Complex Systems (NICO) and Kellogg School of Management, Northwestern University, Evanston, IL, USA. [6]Key Laboratory of Mathematics, Informatics and Behavioral Semantics, Ministry of Education, Beijing, China. [7]Department of Mathematics, Central European University, Budapest, Hungary. [8]ISI Foundation, Torino, Italy.
*e-mail: robertasinatra@gmail.com

Interdisciplinarity: A Nobel Opportunity
Supporting Information

Michael Szell[1,2,3], Yifang Ma[4,5], and Roberta Sinatra[1,6,2,3,7*]

SELECTION OF NOBEL PRIZE WINNING PAPERS

Our initial data set consists of the Nobel Prize winning papers identified in Ref. [1], covering all papers up to 2013 since the Nobel committee started offering a detailed explanation with references for the prize, and additionally the Nobel Prize winning papers of the years 2014, 2015, 2016, and 2017 identified with the information provided on the nobelprize.org page. Of all these papers, we were not able to consider the following three types: 1) Papers of 4 laureates that are not contained in the Web of Science database, for example Tu et al. (1981), Acta pharmaceutica Sinica, 16(5), 366-370, 2) 6 papers that were published after 2002, since our citation data is available until 2012 and citations 10 years after publication cannot be determined here, for example Novoselov et al. (2004), Science 306, 666-669, and 3) 17 papers that were sleeping beauties, i.e. papers that had an insufficient number of citations 10 years after publication from life sciences, physics and chemistry to be placed in the triangle of Fig. 2. with sufficient statistical confidence, for example Kroemer (1996), Physica Scripta T68, 32-45, which had only 5 citations 10 years after publication. These discarded, and the resulting 108 Nobel Prize winning papers, their identifiers and bibliometric data, are provided in the SI Data file. Citations from different fields can be fractional numbers when citing papers belong to multiple disciplines. Paper disciplines are identified by using the journal subject categories of Web of Science. Note that we did not consider Nobel Prizes in economics because of the sparsity of economics papers in the Web of Science data set. The data are available as supplementary data set.

SELECTION OF TOP 10,000 PAPERS

The top 10,000 papers shown in Fig. 2b are the 10,000 papers in the whole Web of Science data set that had the most citations 10 years after publication, published before 2002 for the same reason as above.

VALIDATION OF RESULTS WITH A SECOND DATA SET

Because Nobel Prize winning papers are sometimes inconsistently or not explicitly identified on the nobelprize.org website, we cross-checked our results with another data set in which the key works of Nobel Prize winners were manually curated [2], available from http://www.nber.org/nobel/. Using the year of "key research" from this data set [2], we identified in the Web of Science database for each Nobel laureate (who is part of our data set) all papers in that year. Due to name disambiguation issues we manually removed the papers that were wrongly attributed in the Web of Science search query to a Nobel laureate; for example papers of Diana M. Lee (attributed to laureate David M. Lee) or Roger J. Davis (attributed to laureate Raymond Davis Jr.). Finally, for each Nobel laureate we selected their most cited paper in that year. The resulting 99 papers are visualized in Fig. SI 1. This second data set is not perfectly overlapping with our main data set, nevertheless the conclusions are identical.

[1]Department of Network and Data Science, Central European University, 1051 Budapest, Hungary. [2]Center for Complex Network Research, Northeastern University, Boston, MA 02115, USA. [3] Complexity Science Hub, 1080 Vienna, Austria. [4]Northwestern Institute on Complex Systems (NICO) and Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA. [5]Key Laboratory of Mathematics, Informatics and Behavioral Semantics, Ministry of Education, Beijing 100191, China. [6]Department of Mathematics, Central European University, 1051 Budapest, Hungary. [7]ISI Foundation, 10126 Torino, Italy.
To whom correspondence should be addressed. Email: robertasinatra@gmail.com

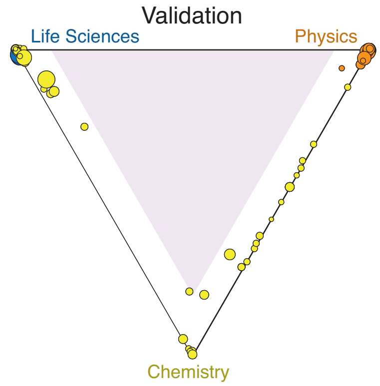

FIG. SI 1. **Validation of results with a second data set.** Using the 99 papers of Nobel laureates that were published in their year of "key research" (data from Ref. [2]) shows the same results as Fig. 2.

INCREASING NUMBER OF PAPERS AND INTERDISCIPLINARITY

A substantial time delay exists between the date of publication of a Nobel Prize paper and the date of its Nobel Prize award. This delay has increased over time, having typically reached on average around 20 years [3]. Given the increase of interdisciplinary impact over time, Fig. 3, we ask: is our analysis not relevant because of the recency of the papers studied? As we show below the answer is no, as we have already reached a few years ago the boundary of typical time delay where the issue has become truly pressing.

The 220 interdisciplinary papers that fall into the shaded area of Fig. 2 follow a comparable distribution of publication years as all the top 10,000 papers. This comparison is shown in Fig. SI 2. Consistent with the measured increase of interdisciplinarity since the mid 1990s, Fig. 3, the distribution of publication years of the interdisciplinary papers, Fig. SI 2b, is slightly more right-skewed than the distribution of all 10,000 papers, Fig. SI 2a. About half of the 220 interdisciplinary papers were published up to 1997, more than 20 years in the past, while the other half was published between 1998 and 2002, now between 15 and 20 years ago. These observations imply that enough time has passed to account for award delays for our dataset of papers to be relevant. The longitudinal analysis demonstrates that, although increased focus on interdisciplinary research is a relatively recent phenomenon, it is now high time to develop an adequate credit system for outstanding scientific accomplishments of any kind.

SIMPLE GINI MEASURE OF INTERDISCIPLINARY IMPACT

We measure the interdisciplinary impact I of papers as $I = 1 - G$, Fig. 3, where G is the Gini coefficient, a standard measure for inequality, applied to the number of citations coming from each discipline within 10 years after publication. The value of I ranges from 0 to 1. If a paper has $I = 1$, then it received an equal amount of citations from each discipline; if $I = 0$, it received citations only from one field. All the different disciplines of journals are given by Web of Science as Life Science, Physics, Chemistry, Computers, Engineering, Mathematics, Agriculture & Environment, Humanity, Others. When the number of fields is restricted to only Life Science, Physics, Chemistry, we find the same result of increased interdisciplinarity since the mid 1990s, only with naturally larger fluctuations, showing robustness of the method.

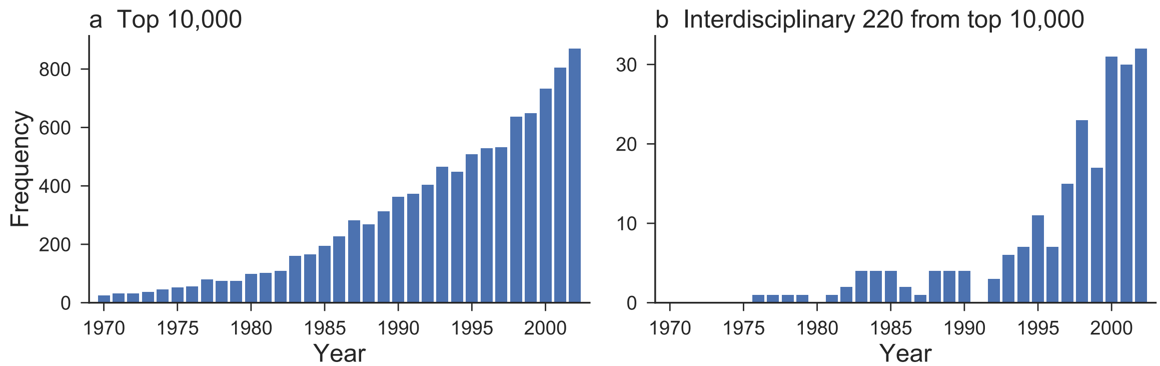

FIG. SI 2. **Distribution of publication dates of the top 10,000 papers.** a) The publication years of the 10,000 highest impact papers (in terms of citations after 10 years) follow a right-skewed distribution, with more papers having been published in more recent years, in line with the well documented growth of scientific output over time. b) The distribution for the subset of 220 interdisciplinary papers that fall into the shaded area of Fig. 2b, is also right-skewed. Here the skewness is much stronger, with half of the papers coming from the last five years 1998-2002 of the data set, consistent with the increase of interdisciplinarity since the mid 1990s, Fig. 3.

Subject category	Citations	%	Papers	%
Agriculture & Environment	36,699,873	6.2	2,513,478	6.6
Chemistry	52,039,544	8.8	4,087,563	10.7
Computers	5,040,600	0.9	600,146	1.6
Engineering	48,455,305	8.2	4,548,102	11.9
Humanity	4,238,596	0.7	404,308	1.1
Life Science	358,507,967	60.6	20,450,178	53.4
Mathematics	7,660,741	1.3	1,079,410	2.8
Others	25,053,535	4.2	1,146,265	3.0
Physics	53,500,509	9.0	3,437,279	9.0

TABLE SI 1. **Absolute and relative number of citations and papers in Web of Science by top-level subject category.** The distribution is skewed towards life sciences, which dominate all other fields, motivating the granularity analysis.

THE EFFECT OF GRANULARITY IN CITING JOURNAL CATEGORIES

To understand whether the analyzed Nobel Prize winning papers feature multidisciplinary nuances on a more fine-grained level, we re-apply our method after increasing the granularity of subject categories of the journals of papers which cite them. One good reason for such an additional analysis is the dominance of Life Science papers (60.6%) and citations (53.4%), Table SI 1. Ignoring the different impacts within this discipline might mask multidisciplinarity within the biological and medical sciences [4]. In particular, we use the 2-letter code, combined subject categories that are assigned by Thomson Reuters to journals. These assignments are not unique; a journal can have multiple such low-level categories.

Doing so reveals that the 108 Nobel papers are cited by papers from 163 different low-level categories. At first glance this variety seems to identify multidisciplinary impacts *within* top-level categories: most physiology/medicine and chemistry Nobel papers are cited by both *Biochemistry & Molecular Biology* and *Cell Biology*, and most physics Nobel papers are cited by both *Physics, Multidisciplinary* and *Physics, Particles & Fields*, Fig. SI 3. For each discipline we identify the top 3 low-level categories of citing journal papers, and draw the corresponding triangle plots. As the plots show, Fig. SI 3 bottom, within a top-level category, impact as measured by these low-level categories seems more spread than impact between the top-level categories, Fig. 2. Unfortunately, it is not at all clear how to interpret this

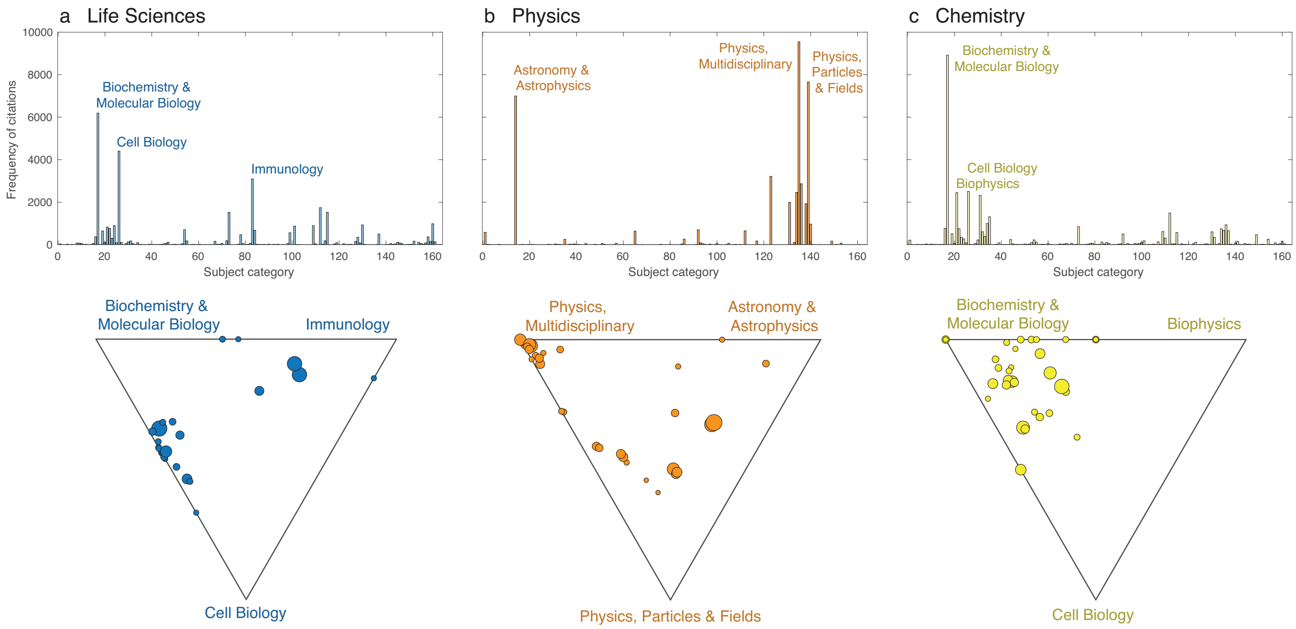

FIG. SI 3. **Distribution of subject categories of papers that cite the Nobel papers and corresponding triangle plots of the top 3 categories from each field.** a) Life Sciences, b) Physics, c) Chemistry.

higher spread due to a number of reasons. First, journals are often assigned multiple low-level categories by Thomson Reuters. This is especially true for the two categories of *Biochemistry & Molecular Biology* and *Cell Biology* – many journals in biology tend to be assigned both. Second, although the low-level category *Physics, Multidisciplinary* sounds as if it signifies multidisciplinarity by itself, it actually refers to a large number of common physics journals (740 out of 2462) such as Physical Review Letters that publish papers in a variety of physics topics and in which the papers themselves are not necessarily multidisciplinary. Thus, such a "multidisciplinary" category signifies the diversity of papers within a journal and should not be mistaken for multidisciplinarity of single papers within a journal. Third, even without the aforementioned issues, it should be expected that the closer sub-fields we consider, the more likely we should find impact in multiple ones. This expectation is confirmed, for example, in studies of the PACS hierarchy [5] in the corpus of papers of the American Physical Society. In other words, multidisciplinary impact is naturally resolution-dependent – it is in fact an open question after which level of zooming into sub-categories we are still justified to use the term "multidisciplinary".

While all of these analyses of low-level categories reveal some subtle intricacies in the way journals are categorized and raise interesting caveats, they do not affect our main point: On the relevant, least granular level of the Nobel Prize categories, only discoveries in clear-cut fields receive an award.

BIBLIOMETRIC INTERDISCIPLINARITY INDICES

Additionally to the network-based metrics that produced Figs. 1 and 2 in the main text, we performed an analysis with interdisciplinarity indices used in scientometrics [6], calculating and discussing a number of metrics for all the Nobel Prize winning papers. Given the different fields $f = 1, \dots, F$, then the occurrence of each discipline is accounted in the occurrence vector $\boldsymbol{o} = \{o_1, \dots, o_F\}$, $o_f \geq 0$, or, similarly, in its associated frequency vector $\boldsymbol{p} = \{p_1, \dots, p_F\}$, where $p_f = o_f / \sum_f o_f$. We calculated for the 108 Nobel Prize winning papers the following metrics [6, 7]:

- Variety: $\sum_{i=1}^{F} 1$

- Balance: $-\sum_{i=1}^{F} p_i \log p_i$
- normalized balance: $-\frac{1}{\log F}\sum_{i=1}^{F} p_i \log p_i$
- Integration: $I = \sum_{i,j=1}^{F} p_i p_j d_{ij}$, where d_{ij} is the distance between fields i and j (using cosine similarity based on the number of co-citations)

The resulting histograms of these metrics, applied to each the references and citations of the papers, are reported in Fig. SI 4. Due to the small number of 108 papers, a statistically reliable interpretation of the index distributions is difficult. The most notable differences in the reference indices (top 4 rows) is maybe the relatively large fraction of zero-index papers in physics, absent in medicine/physiology and chemistry. This spike means that only in physics, Nobel Prizes are sometimes awarded to discoveries that build exclusively on papers from their own field. It is hard to make out any other notable differences in the reference indices. For the citation indices (bottom 4 rows), again the distribution masses are concentrated more to the left for physics than for medicine and chemistry, showing that also the diversity of impacted fields is slightly smaller in physics. The main insights offered in Figs. 1 and 2 of the manuscript are not identifiable from these bibliometric indicators due to their aggregated nature. Indeed, while these indicators do an excellent job at quantifying the heterogeneity of references and citations of papers, they are not designed to distinguish which combination of disciplines produces that heterogeneity.

[1] H.-W. Shen, A.-L. Barabasi, *Proceedings of the National Academy of Sciences* (2014).
[2] B. F. Jones, B. A. Weinberg, *Proceedings of the National Academy of Sciences* **108**, 18910 (2011).
[3] S. Fortunato, *Nature* **508**, 186 (2014).
[4] A. Porter, I. Rafols, *Scientometrics* **81**, 719 (2009).
[5] R. K. Pan, S. Sinha, K. Kaski, J. Saramäki, *Scientific reports* **2**, 551 (2012).
[6] I. Rafols, M. Meyer, *Scientometrics* **82**, 263 (2010).
[7] A. Stirling, *Journal of the Royal Society Interface* **4**, 707 (2007).

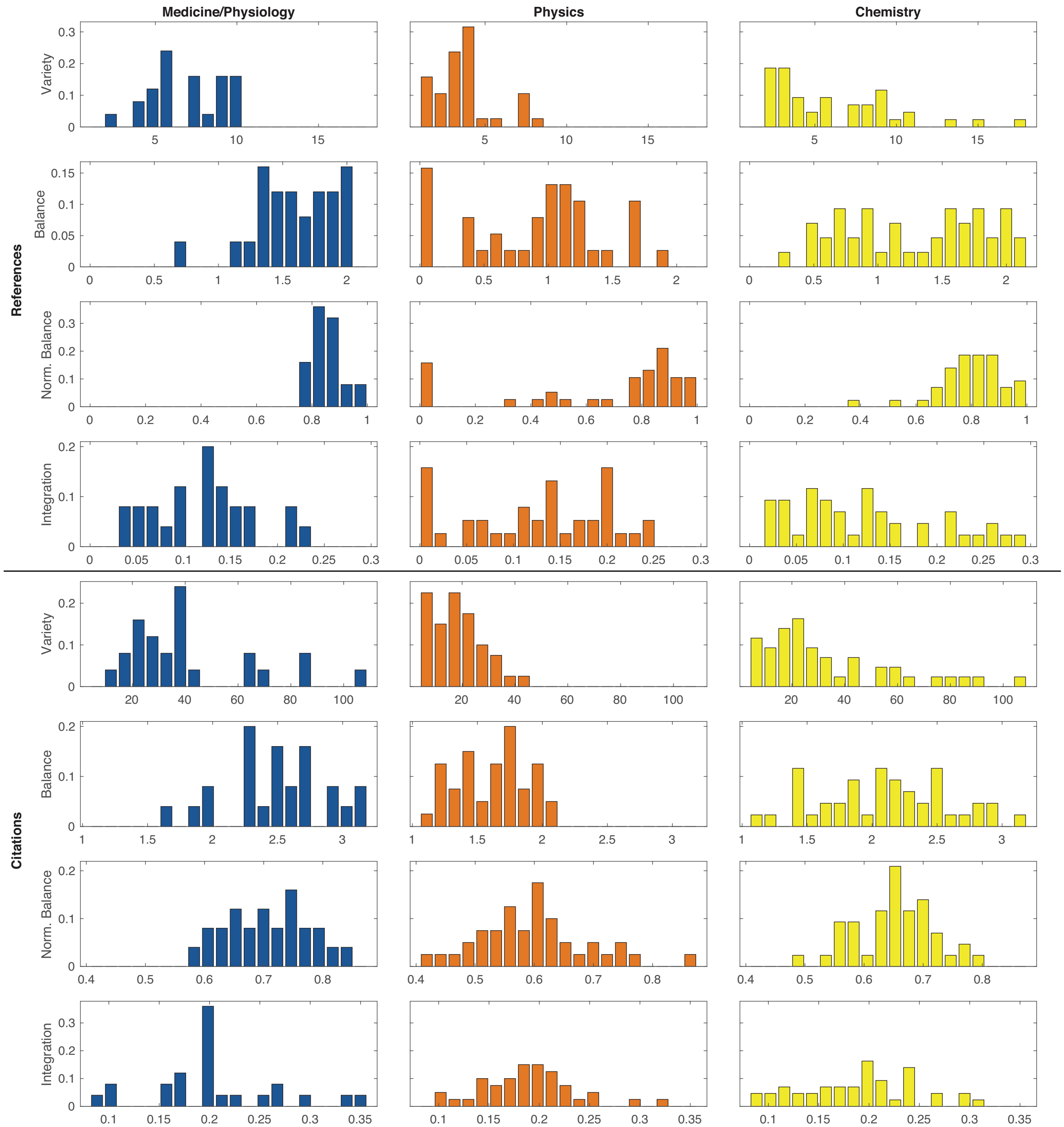

FIG. SI 4. **Bibliometric interdisciplinarity indices for references and citations of the Nobel Prize winning papers.** The distributions for physics Nobel papers are concentrated more to the left than in medicine/physiology and chemistry, indicating that here Nobel Prizes tend to be awarded to discoveries that build or have impact mostly on papers from their own field.