The Field-Dependent Nature of PageRank Values in Citation Networks

This manuscript (<u>permalink</u>) was automatically generated from <u>greenelab/indices manuscript@b56c696</u> on January 5, 2023.

Authors

- · Benjamin J. Heil
 - © 0000-0002-2811-1031 · ♠ ben-heil · ❤ autobencoder

 Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania · Funded by The Gordon and Betty Moore Foundation (GBMF4552)
- Casey S. Greene [™]
 - © 0000-0001-8713-9213 · ♀ cgreene · У greenescientist

Department of Pharmacology, University of Colorado School of Medicine; Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine · Funded by The Gordon and Betty Moore Foundation (GBMF4552); The National Human Genome Research Institute (R01 HG10067)

Abstract

The value of scientific research can be easier to assess at the collective level than at the level of individual contributions. Several journal-level and article-level metrics aim to measure the importance of journals or individual manuscripts. However, many are citation-based and citation practices vary between fields. To account for these differences, scientists have devised normalization schemes to make metrics more comparable across fields. We use PageRank as an example metric and examine the extent to which field-specific citation norms drive estimated importance differences. In doing so, we recapitulate differences in journal and article PageRanks between fields. We also find that manuscripts shared between fields have different PageRanks depending on which field's citation network the metric is calculated in. We implement a degree-preserving graph shuffling algorithm to generate a null distribution of similar networks and find differences more likely attributed to field-specific preferences than citation norms. Our results suggest that while differences exist between fields' metric distributions, applying metrics in a field-aware manner rather than using normalized global metrics avoids losing important information about article preferences. They also imply that assigning a single importance value to a manuscript may not be a useful construct, as the importance of each manuscript varies by the reader's field.

Introduction

There are more academic papers than any human can read in a lifetime. Attention has been given to ranking papers, journals, or researchers by their "importance," assessed via various metrics. Citation count assumes the number of citations determines a paper's importance. The h-index and Journal Impact Factor focus on secondary factors like author or journal track records. Graph-based methods like PageRank or disruption index use the context of the citing papers to evaluate an article's relevance [1,2,3,4]. Each of these methods has limitations, and permutations exist that attempt to shore up specific weaknesses [5,6,7,8].

One objection to such practices is that "importance" is subjective. The San Francisco Declaration on Research Assessment (DORA) argues against using Journal Impact Factor, or any journal-based metric, to assess individual manuscripts or scientists [9]. DORA further argues in favor of evaluating the scientific content of articles and notes that any metrics used should be article-level (https://sfdora.org/read/). However, even article-level metrics often ignore that the importance of a specific scientific output will fundamentally differ across fields. Even Nobel prize-winning work may be unimportant to a cancer biologist if the prize-winning article is about astrophysics.

Because there are differences between fields' citation practices [10], scientists have developed strategies including normalizing the number of citations based on nearby papers in a citation network, rescaling fields' citation data to give more consistent PageRank results, and so on [5,11,12,13]. Such approaches normalize away field-specific effects, which might help to compare one researcher with another in a very different field. However, they do not address the difference in the relevance of a topic between fields. This phenomenon of field-specific importance has been observed at the level of journal metrics. Mason and Singh recently noted that depending on the field, the journal *Christian Higher Education* is either ranked as a Q1 (top quartile) journal or a Q4 (bottom quartile) journal [14].

It is possible that, while global journal-level metrics fail to capture field-specific importance, article-level metrics are sufficiently granular that the importance of a manuscript remains constant across fields. We investigate the extent to which article-level metrics generalize between fields. We examine this using MeSH terms to define fields and use field-specific citation graphs to assess their importance within the field. While it is trivially apparent that journals or articles that do not have cross-field citations will have variable importance, we ignore these cases. We include only those including

citations in both fields, where we expect possible consistency. We first replicate previous findings that journal-level metrics can differ substantially among fields. We also find field-specific variability in importance at the article level. We make our results explorable through a web app that shows metrics for overlapping papers between pairs of fields.

Our results show that even article-level metrics can differ substantially among fields. We recommend that metrics used for assessing research outputs include field-specific, in addition to global, ones. While qualitative assessment of the content of manuscripts remains time-consuming, our results suggest that within-field and across-field assessment remains key to assessing the importance of research outputs.

Results

Journal rankings differ between fields

In an attempt to quantify the relative importance of journals, scientists have created rankings using metrics the Journal Impact Factor, which essentially uses citations per article, and those that rely on more complex representations like Eigenfactor [15]. Previous reports note that journal rankings differ substantially between fields using metrics based on citation numbers [14]. We calculated a field-specific PageRank-based score for each journal as the median PageRank of manuscripts published in that journal for that field (Fig. 1 A). We first sought to understand the extent to which PageRank replicated journal ranking differences across fields.

To begin, we compared the differences in ranking between the top fifty journals in nanotechnology and their corresponding ranks in microscopy. While the ranks were correlated (r=.75) there was a great deal of variance, especially for journals outside the top 20 in nanotechnology (Fig. 1 B). We then examined the top-ranked journal in each of our 45 fields to determine whether the top-ranking journal was consistent across fields (Fig. 1 C). We found that the most commonly top-ranked journal was *Science*. This was unsurprising, given that it tends to rank highly among global journal-level metrics such as eigenfactor. However, while *Science* was the top-ranked journal in a plurality of fields, approximately 80% of fields had a different journal in that spot.

We also investigated the presence of single-topic journals in our dataset, as MeSH headings reflect a different type of aggregation than journals do [16]. Of the 5,178 journals with at least 50 articles in our dataset, the median number of fields publishing in a given journal is 15 (Fig. 1 D). In the context of MeSH, specialty journals are rare. Most journals publish manuscripts with in one-third or more of the MeSH headings in our dataset.

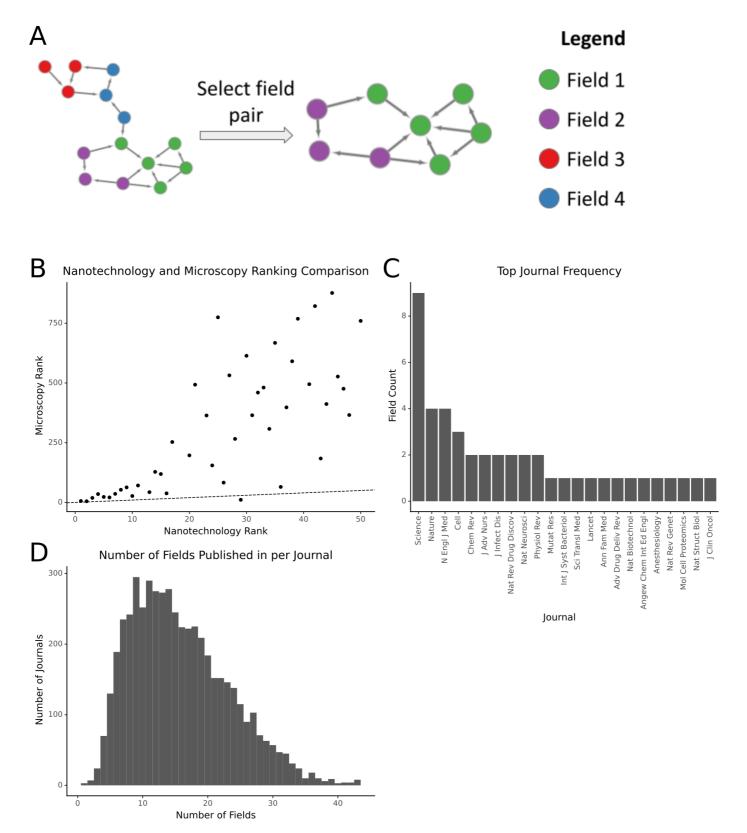


Figure 1: Journals' PageRank-derived rankings differ between fields. A) A schematic showing how paired networks are derived from the full citation network. B) A comparison of the ranks of the top 50 journals by PageRank in nanotechnology and their rank in microscopy. Top-50 nanotechnology journals with no papers in microscopy have been omitted. C) The frequency with which journals in the dataset are the top journal for a field. D) The distribution of fields published per journal. The X-axis corresponds to the number of fields for which a journal has at least one paper within the field. All plots restrict the set of journals to those with at least 50 papers in the dataset.

Manuscript PageRanks differ between fields

We split the citation network into its component fields and calculated the PageRank for each article (Fig. 2 A). We examined the distribution of PageRanks across fields and found that they differed

greatly (Fig. $\underline{2}$ B). We first examined whether the citation practices of fields contributed to importance differences. Investigating manuscripts that appeared in pairs of fields, we found that the distribution of importances matched the network more than that of the alternative topic area of the manuscript (Fig. $\underline{2}$ B, C, D).

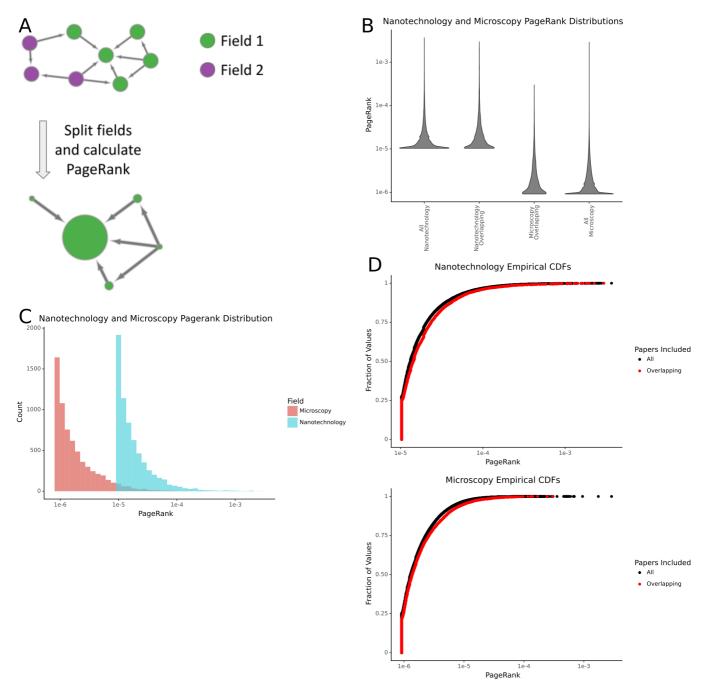


Figure 2: Differences in the distribution of PageRanks between fields. A) A schematic showing how field pairs are split and their PageRanks are calculated. B) The distribution of article PageRanks for nanotechnology and microscopy. The distributions marked with 'All' contain all the papers for the given field in the dataset, while those marked 'overlapping' contain only articles present in both fields. C) The empirical cumulative density functions of nanotechnology and microscopy. D) The differences in distribution of the PageRanks of articles shared by nanotechnology and microscopy. E) A density plot showing the joint distribution of PageRanks for papers overlapping in nanotechnology and microscopy.

Fields' differences are not solely driven by differences in citation practices

We devised a strategy to generate an empirical null for a field pair under the assumption that the field pair represented a single, homogenous field (Fig. 3 A). For each field-pair intersection, we performed a degree-distribution preserving permutation. We created 100 permuted networks for each field pair.

We then split the networks into their constituent fields and calculated a percentile using the number of permuted networks with a lower PageRank for a manuscript than the true PageRank. A manuscript with a PageRank higher than all networks has a percentile of 100, and one lower than all permuted networks has a percentile of zero. We used the difference in the percentile in each field as the field-specific affinity for a given paper. This percentile score allowed us to control for the differing degree distributions between fields by comparing papers based on their expected PageRank in a random network with the same node degrees.

We selected field pairs with varying degrees of correlation between their PageRanks (Fig. 3 B). By examining the fields' PageRank percentiles, we found that many articles had large differences in their perception between fields (Fig. 3 C). In nanotechnology and microscopy, papers with high nanotechnology percentiles and low microscopy percentiles tended towards applications of nanotechnology, while their counterparts with high microscopy percentiles and low nanotechnology percentiles were often papers about technological developments in microscopy (Fig. 3 A, Table 1). Immunochemistry-favored papers are largely applications of immunochemical methods, while anatomy-favored articles tend to focus experiments on a single anatomical region (Fig. 3 B, Table 2). Proteomics and metabolomics tend to use similar methods, so the fields on either end are largely (though not entirely) field-specific applications of those methods (Fig. 3 C, Table 3). Manuscripts favored in computational biology were similarly applications-focused. However, those with more importance in human genetics tended towards policy papers due to its MeSH heading (H01.158.273.343.385) excluding fields like genomics, population genetics, and microbial genetics (Fig. 3 D, Table 4). In addition to papers with large differences between fields, each field pair has papers with high PageRanks and similar percentiles. While some papers may be influential in multiple fields, others have more field-specific import.

It is impossible to describe all the field pairs and relevant differences between fields within the space of a journal article. Instead, we have developed a web server that displays the percentiles for all pairs of fields in our dataset with at least 1000 shared articles (Fig. 3 D), which can be accessed at https://www.indices.greenelab.com. We hope that the availability of the web server and the reproducibility of our code will assist other scientists in uncovering new insights from this dataset.

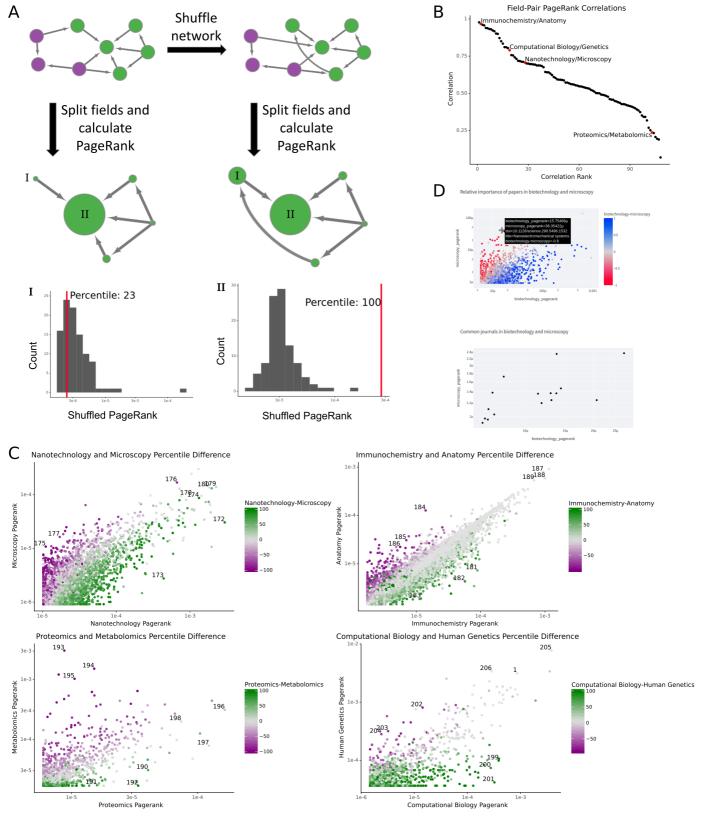


Figure 3: Field-specific preferences in papers. A) A schematic showing how networks are shuffled and how articles' percentile scores are calculated. The histograms at the bottom of the figure correspond to the distribution of PageRanks for the shuffled networks, while the red lines correspond to an article's PageRank in the true citation network. B) The Pearson correlation of PageRanks between fields. The red points are the field pairs expanded in panel C. C) The percentile scores and PageRanks for overlapping articles in various fields. Points are colored based on the difference in percentile scores in the fields e.g. "Nanotechnology-Microscopy" corresponds to the difference between the nanotechnology and microscopy percentile scores. The numbers next to points are the reference number for the article in the bibliography. D) A screenshot of the webserver showing the percentile score difference and journal median PageRank plot functionality.

Nanotechnology Percentile	Microscopy Percentile	Title	Reference
100	4	A robust DNA mechanical device controlled by hybridization topology	[17]
100	5	Bioadhesive poly(methyl methacrylate) microdevices for controlled drug delivery	[18]
99	2	DNA-templated self- assembly of protein arrays and highly conductive nanowrires	[19]
0	100	Photostable luminescent nanoparticles as biological label for cell recognition of system lupus erythematosus patients	[20]
5	90	WSXM: a software for scanning probe microscopy and a tool for nanotechnology	[21]
0	77	Measuring Distances in Supported Bilayers by Fluorescence Interference- Contrast Microscopy: Polymer Supports and SNARE Proteins	[22]
100	99	Toward fluorescence nanoscopy	[23]
100	86	In vivo imaging of quantum dots encapsulated in phospholipid micelles	[24]
100	99	Water-Soluble Quantum Dots for Multiphoton Fluorescence Imaging in Vivo	[25]

 Table 1: Nanotechnology/microscopy papers of interest

Immunochemistry Percentile	Anatomy Percentile	Title	Reference
100	45	Immunoelectron microscopic exploration of the Golgi complex	[26]
100	14	Immunocytochemical and electrophoretic analyses of changes in myosin gene expression in cat posterior temporalis muscle during postnatal development	[27]

Immunochemistry Percentile	Anatomy Percentile	Title	Reference
98	5	Electron microscopic demonstration of calcitonin in human medullary carcinoma of thyroid by the immuno gold staining method	[28]
12	100	Grafting genetically modified cells into the rat brain: characteristics of E. coli β-galactosidase as a reporter gene	[29]
12	100	Vitamin-D-dependent calcium-binding-protein and parvalbumin occur in bones and teeth	[30]
3	100	Mapping of brain areas containing RNA homologous to cDNAs encoding the alpha and beta subunits of the rat GABAA gamma-aminobutyrate receptor	[31]
100	100	Studies of the HER-2/neu Proto-Oncogene in Human Breast and Ovarian Cancer	[32]
100	100	Expression of c-fos Protein in Brain: Metabolic Mapping at the Cellular Level	[33]
100	100	Proliferating cell nuclear antigen (PCNA) immunolocalization in paraffin sections: An index of cell proliferation with evidence of deregulated expression in some neoplasms	[34]

 Table 2: Immunochemistry/anatomy papers of interest

Proteomics Percentile	Metabolomics Percentile	Title	Reference
67	2	Proteomics Standards Initiative: Fifteen Years of Progress and Future Work	[35]
99	0	Limited Environmental Serine and Glycine Confer Brain Metastasis Sensitivity to PHGDH Inhibition	[36]
100	0	A high-throughput processing service for retention time alignment of complex proteomics and metabolomics LC-MS data	[37]

Proteomics Percentile	Metabolomics Percentile	Title	Reference
0	100	MeltDB: a software platform for the analysis and integration of metabolomics experiment data	[38]
0	98	In silico fragmentation for computer assisted identification of metabolite mass spectra	[39]
0	100	The Metabonomic Signature of Celiac Disease	[40]
91	70	Visualization of omics data for systems biology	[41]
0	16	FunRich: An open access standalone functional enrichment and interaction network analysis tool	[42]
0	5	Proteomic and Metabolomic Characterization of COVID- 19 Patient Sera	[43]

 Table 3: Proteomics/metabolomics papers of interest

Computational Biology Percentile	Human Genetics Percentile	Title	Reference
99	0	Development of Human Protein Reference Database as an Initial Platform for Approaching Systems Biology in Humans	[44]
100	1	A database for post- genome analysis	[<u>45</u>]
100	1	Use of mass spectrometry- derived data to annotate nucleotide and protein sequence databases	[46]
12	100	Genetic Discrimination: Perspectives of Consumers	[47]
0	81	Committee Opinion No. 690: Carrier Screening in the Age of Genomic Medicine	[48]
23	100	Public health genomics: The end of the beginning	[49]
100	99	Initial sequencing and analysis of the human genome	[50]
100	100	An STS-Based Map of the Human Genome	[51]

Computational Biology Percentile	Human Genetics Percentile	Title	Reference
100	100	A New Five-Year Plan for the U.S. Human Genome Project	[<u>52</u>]

Table 4: Computational biology/human genetics papers of interest

Methods

COCI

We used the March 2022 version of the COCI citation index [53] as the source of our citation data. This dataset contains around 1.3 billion citations from ~73 million bibliographic resources.

Selecting fields

To differentiate between scientific fields, we needed a way to map papers to fields. Fortunately, all the papers in Pubmed Central (https://www.ncbi.nlm.nih.gov/pmc/) have corresponding Medical Subject Headings (MeSH) terms. While MeSH terms are varied and numerous, the subheadings of the Natural Science Disciplines (H01) category fit our needs. However, MeSH terms are hierarchical and vary greatly in their size and specificity. To extract a balanced set of terms, we recursively traversed the tree and selected headings having at least 10000 DOIs without having multiple children that also meet the cutoff. Our resulting headings were comprised of 45 terms, from "Acoustics" to "Water Microbiology."

Building single heading citation networks

The COCI dataset consists of pairs of Digital Object Identifiers (DOIs). To change these pairs into a form we could run calculations on, we needed to convert them into networks. To do so, we created 45 empty networks, one for each previously selected MeSH term. We then iterated over each pair of DOIs in COCI and added them to a network if the DOIs corresponded to two journal articles written in English, both of which were tagged with the corresponding MeSH heading.

Because we were interested in the differences between fields, we also needed to build networks from pairs of MeSH headings. These networks were built via the same process, except that instead of keeping articles corresponding to a single DOI we added a citation to the network if both articles were in the pair of fields, even if the citation occurred across fields. Running this network-building process yielded 990 two-heading networks.

Sampling a graph from the degree distribution while preserving the distribution of degrees in the network was challenging. Because citation graphs are directed, it is impossible to simply swap pairs of edges and end up with a graph uniformly sampled from the space. Instead, a more sophisticated three-edge swap method must be used [54]. Because this algorithm had not been implemented yet in NetworkX [55], we implemented the code to perform shuffles and submitted our change to the library (https://github.com/networkx/networkx/pull/5663). With the shuffling code implemented, we created 100 shuffled versions of each of our combined networks to act as a background distribution against which we could compare metrics.

Once we had a collection of shuffled networks, we needed to split them into their constituent fields. To do so, we reduced the network to solely the nodes that were present in the single heading citation

network and kept only citations between these nodes.

Metrics

We used the NetworkX implementation of PageRank with default parameters to evaluate paper importance within fields. To determine the degree to which the papers' PageRank values were higher or lower than expected, we compared the PageRank values calculated for the true citation networks to the values in the shuffled networks for each paper. We then recorded the percent of shuffled networks where the paper had a lower PageRank than the true network to derive a single number that described these values. For example, if a paper had a higher PageRank in the true network than in all the shuffled networks it received a percentile of 100. Likewise, if it had a lower PageRank in the true network than in all the shuffled networks it received a percentile of 0.

A convenient feature of the percentiles was that they were directly comparable between fields. For manuscripts represented in two fields, the difference in scores was used to estimate its variability in importance. For example, if a paper had a score of 100 in field A (indicating a higher PageRank in the field than expected given its number of citations and the network structure) and a score of 0 in field B (indicating a lower than expected PageRank), then the large difference in scores indicated the paper was more highly valued in field A than field B. If the paper had similar scores in both fields, it indicated that the paper was similarly valued in the two fields.

Hardware/runtime

We ran the full analysis pipeline on the RMACC Summit cluster at the University of Colorado. The pipeline took about a week to run, from downloading the data to analyzing it to visualizing it. Performance in other contexts will depend heavily on details such as the number of CPU nodes available and the network speed.

Server details

Our webserver is built by visualizing our data in Plotly (https://plotly.com/python/plotly-express/) on the Streamlit platform (https://streamlit.io/). The field pairs made available by the frontend are those with at least 1000 shared papers after filtering out papers with more than a 5% missingness level of their PageRanks after shuffling. The journals available for visualization are those with at least 25 papers for the given field pair.

Discussion/Conclusion

We analyze hundreds of field-pair citation networks to examine the extent to which article-level importance metrics vary between fields. As previously reported, we find systematic differences in PageRanks between fields [7,56] that would warrant some form of normalization when making crossfield comparisons with global statistics. However, we also find that field-specific differences are not driven solely by differences in citation practices. Instead, the importance of individual papers appears to differ meaningfully between fields. Global rankings or efforts to normalize out field-specific effects obscure meaningful differences in manuscript importance between communities.

As with any study, this research has certain limitations. One example is our selection of MeSH terms to represent fields. We used MeSH because it is a widely-annotated set of subjects in biomedicine and thresholded MeSH term sizes to balance having enough observations to calculate appropriate statistics with having sufficient granularity to capture fields. This selection process resulted in fields at the granularity of "biophysics" and "ecology." We also have to select a number of swaps to generate a

background distribution of PageRanks for each field pair. We selected three times as many swaps as edges, where each swap modifies three edges, but certain network structures may require a different number.

We also note that there are inherent issues with the premise of ranking manuscripts' importance. We sought to understand the extent to which such rankings were stable between fields after correcting for field-specific citation practices. We found limited stability between fields, mostly between closely-related fields, suggesting that the concept of a universal ranking of importances is difficult to justify. In the way that reducing a distribution to a Journal Impact Factor distorts assessment, attempting to use a single universal score to represent importance across fields poses similar challenges at the level of individual manuscripts. Furthermore, this work's natural progression would extend to estimating the importance of individual manuscripts to individual researchers. Thus, a holistic measure of importance would need to include a distribution of scores not only across fields but across researchers. It may ultimately be impossible to calculate a meaningful importance score. The lack of ground truth for importance is an inherent feature, not a bug, of science's step-wide progression.

Shifting from the perspective of evaluation to discovery can reveal more appropriate uses for these types of statistics. Field-pair calculations for such metrics may help with self-directed learning of new fields. An expert in one field, e.g., computational biology, who aims to learn more about genetics may find manuscripts with high importance in genetics and low importance in computational biology to be important reads. These represent manuscripts not currently widely cited in one's field but highly influential in a target field. Our application can reveal these manuscripts for MeSH field pairs, and our source code allows others to perform our analysis with different granularity.

Code and Data Availability

The code to reproduce this work can be found at https://github.com/greenelab/indices. The data used for this project is publicly available and can be downloaded with the code provided above. Our work meets the bronze standard of reproducibility [57] and fulfills aspects of the silver and gold standards including deterministic operation.

Acknowledgements

We would like to thank Jake Crawford for reviewing code that went into this project and Faisal Alquaddoomi for figuring out the web server hosting. We would also like to thank the past and present members of GreeneLab who gave feedback on this project during lab meetings. This work utilized resources from the University of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236).

Funding

This work was supported by grants from the National Institutes of Health's National Human Genome Research Institute (NHGRI) under award R01 HG010067 and the Gordon and Betty Moore Foundation (GBMF 4552) to CSG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. An index to quantify an individual's scientific research output

JE Hirsch

Proceedings of the National Academy of Sciences (2005-11-07) https://doi.org/cbq6dz

DOI: <u>10.1073/pnas.0507655102</u> · PMID: <u>16275915</u> · PMCID: <u>PMC1283832</u>

2. New Tools for Improving and Evaluating The Effectiveness of Research

Irving H Sher, Eugene Garfield Research Program Effectiveness (1965-06-27)

3. The PageRank Citation Ranking: Bringing Order to the Web.

Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd *Stanford InfoLab* (1999)

4. Large teams develop and small teams disrupt science and technology

Lingfei Wu, Dashun Wang, James A Evans

Nature (2019-02) https://doi.org/gfvnb9

DOI: <u>10.1038/s41586-019-0941-9</u> · PMID: <u>30760923</u>

5. Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level

Blan Hutchins, Xin Yuan, James M Anderson, George M Santangelo *PLOS Biology* (2016-09-06) https://doi.org/f88zk2

DOI: 10.1371/journal.pbio.1002541 · PMID: 27599104 · PMCID: PMC5012559

6. Measuring contextual citation impact of scientific journals

Henk F Moed

Journal of Informetrics (2010-07) https://doi.org/dpbgj9

DOI: <u>10.1016/j.joi.2010.01.002</u>

7. Collective topical PageRank: a model to evaluate the topic-dependent academic impact of scientific papers

Yongjun Zhang, Jialin Ma, Zijian Wang, Bolun Chen, Yongtao Yu *Scientometrics* (2017-12-23) https://doi.org/gc4b2s

DOI: <u>10.1007/s11192-017-2626-1</u>

8. Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers

Lutz Bornmann, Sitaram Devarakonda, Alexander Tekles, George Chacko *Quantitative Science Studies* (2020-08) https://doi.org/gg2ts5

DOI: 10.1162/qss a 00068

9. **Impact Factor Distortions**

Bruce Alberts

Science (2013-05-17) https://doi.org/mjm

DOI: 10.1126/science.1240319 · PMID: 23687012

10. Citation patterns in economics and beyond

Matthias Aistleitner, Jakob Kapeller, Stefan Steinerberger

Science in Context (2019-12) https://doi.org/gg62s8

DOI: <u>10.1017/s0269889720000022</u> · PMID: <u>32202238</u>

11. Disruptive papers published in Scientometrics: meaningful results by using an improved variant of the disruption index originally proposed by Wu, Wang, and Evans (2019)

Lutz Bornmann, Sitaram Devarakonda, Alexander Tekles, George Chacko *Scientometrics* (2020-03-14) https://doi.org/ggzzxd

DOI: 10.1007/s11192-020-03406-8

12. Quantifying and suppressing ranking bias in a large citation network

Giacomo Vaccario, Matúš Medo, Nicolas Wider, Manuel Sebastian Mariani *Journal of Informetrics* (2017-08) https://doi.org/gbzjdh

DOI: <u>10.1016/j.joi.2017.05.014</u>

13. Quantitative evaluation of alternative field normalization procedures

Yunrong Li, Filippo Radicchi, Claudio Castellano, Javier Ruiz-Castillo *Journal of Informetrics* (2013-07) https://doi.org/f48tvt

DOI: 10.1016/j.joi.2013.06.001

14. When a journal is both at the 'top' and the 'bottom': the illogicality of conflating citation-based metrics with quality

Shannon Mason, Lenandlar Singh

Scientometrics (2022-05-25) https://doi.org/gq2468

DOI: 10.1007/s11192-022-04402-w

15. **Eigenfactor: Measuring the value and prestige of scholarly journals**

Carl Bergstrom

College & Research Libraries News (2007-05-01) https://doi.org/gf24tg

DOI: <u>10.5860/crln.68.5.7804</u>

16. Cited references and Medical Subject Headings (MeSH) as two different knowledge representations: clustering and mappings at the paper level

Loet Leydesdorff, Jordan A Comins, Aaron A Sorensen, Lutz Bornmann, Iina Hellsten *Scientometrics* (2016-10-08) https://doi.org/gc8zk4

DOI: <u>10.1007/s11192-016-2119-7</u> · PMID: <u>27942085</u> · PMCID: <u>PMC5124055</u>

17. A robust DNA mechanical device controlled by hybridization topology

Hao Yan, Xiaoping Zhang, Zhiyong Shen, Nadrian C Seeman

Nature (2002-01) https://doi.org/czh8hg

DOI: <u>10.1038/415062a</u> · PMID: <u>11780115</u>

18. Bioadhesive poly(methyl methacrylate) microdevices for controlled drug delivery

Sarah L Tao, Michael W Lubeley, Tejal A Desai

Journal of Controlled Release (2003-03) https://doi.org/c7fpg4

DOI: 10.1016/s0168-3659(03)00005-1 · PMID: 12628329

19. DNA-Templated Self-Assembly of Protein Arrays and Highly Conductive Nanowires

Hao Yan, Sung Ha Park, Gleb Finkelstein, John H Reif, Thomas H LaBean *Science* (2003-09-26) https://doi.org/bfgvgf

DOI: 10.1126/science.1089389 · PMID: 14512621

20. Photostable Luminescent Nanoparticles as Biological Label for Cell Recognition of System Lupus Erythematosus Patients

Xiaoxiao He, Kemin Wang, Weihong Tan, Jun Li, Xiaohai Yang, Shasheng Huang, Dan Xiao *Journal of Nanoscience and Nanotechnology* (2002-07-01) https://doi.org/dcj5cg

DOI: <u>10.1166/jnn.2002.105</u> · PMID: <u>12908257</u>

21. **WSXM:** A software for scanning probe microscopy and a tool for nanotechnology I Horcas, R Fernández, JM Gómez-Rodríguez, J Colchero, J Gómez-Herrero, AM Baro

Review of Scientific Instruments (2007-01)

DOI: <u>10.1063/1.2432410</u>

22. Measuring Distances in Supported Bilayers by Fluorescence Interference-Contrast Microscopy: Polymer Supports and SNARE Proteins

Volker Kiessling, Lukas K Tamm

Biophysical Journal (2003-01) https://doi.org/dqsg2c

DOI: <u>10.1016/s0006-3495(03)74861-9</u> · PMID: <u>12524294</u> · PMCID: <u>PMC1302622</u>

23. Toward fluorescence nanoscopy

Stefan W Hell

Nature Biotechnology (2003-10-31) https://doi.org/dnzt3b

DOI: 10.1038/nbt895 · PMID: 14595362

24. In Vivo Imaging of Quantum Dots Encapsulated in Phospholipid Micelles

Benoit Dubertret, Paris Skourides, David J Norris, Vincent Noireaux, Ali H Brivanlou, Albert Libchaber

Science (2002-11-29) https://doi.org/dd6sqp

DOI: 10.1126/science.1077194 · PMID: 12459582

25. Water-Soluble Quantum Dots for Multiphoton Fluorescence Imaging in Vivo

Daniel R Larson, Warren R Zipfel, Rebecca M Williams, Stephen W Clark, Marcel P Bruchez, Frank W Wise, Watt W Webb

Science (2003-05-30) https://doi.org/cn9j76

DOI: 10.1126/science.1083780 · PMID: 12775841

26. Immunoelectron microscopic exploration of the Golgi complex.

JW Slot, HJ Geuze

Journal of Histochemistry & Cytochemistry (1983-08) https://doi.org/dxxzxg

DOI: <u>10.1177/31.8.6863900</u> · PMID: <u>6863900</u>

27. Immunocytochemical and electrophoretic analyses of changes in myosin gene expression in cat posterior temporalis muscle during postnatal development

JFY Hoh, S Hughes, C Chow, PT Hale, RB Fitzsimons

Journal of Muscle Research and Cell Motility (1988-02) https://doi.org/d72278

DOI: 10.1007/bf01682147 · PMID: 3392187

28. Electron microscopic demonstration of calcitonin in human medullary carcinoma of thyroid by the immuno gold staining method

J Dämmrich, W Ormanns, R Schäffer

Histochemistry (1984) https://doi.org/ct253c

DOI: 10.1007/bf00514331 · PMID: 6511490

29. Grafting genetically modified cells into the rat brain: characteristics of E. coli β-galactosidase as a reporter gene

S Shimohama, MB Rosenberg, AM Fagan, JA Wolff, MP Short, XO Breakefield, T Friedmann, FH Gage

Molecular Brain Research (1989-06) https://doi.org/dptnbm

DOI: 10.1016/0169-328x(89)90061-2 · PMID: 2501628

30. Vitamin-D-dependent calcium-binding-protein and parvalbumin occur in bones and teeth

MR Celio, AW Norman, CW Heizmann

Calcified Tissue International (1984-12) https://doi.org/fdnfdg

DOI: 10.1007/bf02405306 · PMID: 6423230

31. Mapping of brain areas containing RNA homologous to cDNAs encoding the alpha and beta subunits of the rat GABAA gamma-aminobutyrate receptor.

JM Séquier, JG Richards, P Malherbe, GW Price, S Mathews, H Möhler *Proceedings of the National Academy of Sciences* (1988-10) https://doi.org/fv2p49
DOI: 10.1073/pnas.85.20.7815 · PMID: 2845424 · PMCID: PMC282284

32. Studies of the HER-2/neu Proto-Oncogene in Human Breast and Ovarian Cancer

Dennis J Slamon, William Godolphin, Lovell A Jones, John A Holt, Steven G Wong, Duane E Keith, Wendy J Levin, Susan G Stuart, Judy Udove, Axel Ullrich, Michael F Press *Science* (1989-05-12) https://doi.org/cngtqx

DOI: <u>10.1126/science.2470152</u> · PMID: <u>2470152</u>

33. Expression of c-fos Protein in Brain: Metabolic Mapping at the Cellular Level

SM Sagar, FR Sharp, T Curran

Science (1988-06-03) https://doi.org/b39h2t
DOI: 10.1126/science.3131879 • PMID: 3131879

34. Proliferating cell nuclear antigen (PCNA) immunolocalization in paraffin sections: An index of cell proliferation with evidence of deregulated expression in some, neoplasms

PA Hall, DA Levison, AL Woods, CC-W Yu, DB Kellock, JA Watkins, DM Barnes, CE Gillett, R Camplejohn, R Dover, ... DP Lane

The Journal of Pathology (1990-12) https://doi.org/cntmbr

DOI: 10.1002/path.1711620403 · PMID: 1981239

35. Proteomics Standards Initiative: Fifteen Years of Progress and Future Work

Eric W Deutsch, Sandra Orchard, Pierre-Alain Binz, Wout Bittremieux, Martin Eisenacher, Henning Hermjakob, Shin Kawano, Henry Lam, Gerhard Mayer, Gerben Menschaert, ... Andrew R Jones

Journal of Proteome Research (2017-09-15) https://doi.org/gbw99d

DOI: <u>10.1021/acs.jproteome.7b00370</u> · PMID: <u>28849660</u> · PMCID: <u>PMC5715286</u>

36. Limited Environmental Serine and Glycine Confer Brain Metastasis Sensitivity to PHGDH Inhibition

Bryan Ngo, Eugenie Kim, Victoria Osorio-Vasquez, Sophia Doll, Sophia Bustraan, Roger J Liang, Alba Luengo, Shawn M Davidson, Ahmed Ali, Gino B Ferraro, ... Michael E Pacold *Cancer Discovery* (2020-09-01) https://doi.org/ghf85j

DOI: <u>10.1158/2159-8290.cd-19-1228</u> · PMID: <u>32571778</u> · PMCID: <u>PMC7483776</u>

37. A high-throughput processing service for retention time alignment of complex proteomics and metabolomics LC-MS data

Isthiaq Ahmad, Frank Suits, Berend Hoekman, Morris A Swertz, Heorhiy Byelas, Martijn Dijkstra, Rob Hooft, Dmitry Katsubo, Bas van Breukelen, Rainer Bischoff, Peter Horvatovich *Bioinformatics* (2011-02-23) https://doi.org/cxsszv

DOI: <u>10.1093/bioinformatics/btr094</u> · PMID: <u>21349866</u>

38. MeltDB: a software platform for the analysis and integration of metabolomics experiment data

Heiko Neuweger, Stefan P Albaum, Michael Dondrup, Marcus Persicke, Tony Watt, Karsten Niehaus, Jens Stoye, Alexander Goesmann

Bioinformatics (2008-09-02) https://doi.org/fds6vt

DOI: <u>10.1093/bioinformatics/btn452</u> · PMID: <u>18765459</u>

39. In silico fragmentation for computer assisted identification of metabolite mass spectra

Sebastian Wolf, Stephan Schmidt, Matthias Müller-Hannemann, Steffen Neumann *BMC Bioinformatics* (2010-03-22) https://doi.org/d7gpf5

DOI: <u>10.1186/1471-2105-11-148</u> · PMID: <u>20307295</u> · PMCID: <u>PMC2853470</u>

40. The Metabonomic Signature of Celiac Disease

Ivano Bertini, Antonio Calabrò, Valeria De Carli, Claudio Luchinat, Stefano Nepi, Berardino Porfirio, Daniela Renzi, Edoardo Saccenti, Leonardo Tenori

Journal of Proteome Research (2008-12-11) https://doi.org/c6sdnp

DOI: <u>10.1021/pr800548z</u> · PMID: <u>19072164</u>

41. Visualization of omics data for systems biology

Nils Gehlenborg, Seán I O'Donoghue, Nitin S Baliga, Alexander Goesmann, Matthew A Hibbs, Hiroaki Kitano, Oliver Kohlbacher, Heiko Neuweger, Reinhard Schneider, Dan Tenenbaum, Anne-Claude Gavin

Nature Methods (2010-03) https://doi.org/cp9zgi

DOI: 10.1038/nmeth.1436 · PMID: 20195258

42. FunRich: An open access standalone functional enrichment and interaction network analysis tool

Mohashin Pathan, Shivakumar Keerthikumar, Ching-Seng Ang, Lahiru Gangoda, Camelia YJ Quek, Nicholas A Williamson, Dmitri Mouradov, Oliver M Sieber, Richard J Simpson, Agus Salim, ... Suresh Mathivanan

PROTEOMICS (2015-06-17) https://doi.org/f278rp

DOI: 10.1002/pmic.201400515 · PMID: 25921073

43. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera

Bo Shen, Xiao Yi, Yaoting Sun, Xiaojie Bi, Juping Du, Chao Zhang, Sheng Quan, Fangfei Zhang, Rui Sun, Liujia Qian, ... Tiannan Guo

Cell (2020-07) https://doi.org/gg2cck

DOI: 10.1016/j.cell.2020.05.032 · PMID: 32492406 · PMCID: PMC7254001

44. Development of Human Protein Reference Database as an Initial Platform for Approaching Systems Biology in Humans

Suraj Peri, JDaniel Navarro, Ramars Amanchy, Troels Z Kristiansen, Chandra Kiran Jonnalagadda, Vineeth Surendranath, Vidya Niranjan, Babylakshmi Muthusamy, TKB Gandhi, Mads Gronborg, ... Akhilesh Pandey

Genome Research (2003-10) https://doi.org/bc8cnv

DOI: 10.1101/gr.1680803 · PMID: 14525934 · PMCID: PMC403728

45. A database for post-genome analysis

Minoru Kanehisa

Trends in Genetics (1997-09) https://doi.org/cfgb98

DOI: <u>10.1016/s0168-9525(97)01223-7</u> · PMID: <u>9287494</u>

46. Use of mass spectrometry-derived data to annotate nucleotide and protein sequence databases

Matthias Mann, Akhilesh Pandey

Trends in Biochemical Sciences (2001-01) https://doi.org/ch565r

DOI: 10.1016/s0968-0004(00)01726-6 · PMID: 11165518

47. **Genetic Discrimination: Perspectives of Consumers**

EVirginia Lapham, Chahira Kozma, Joan O Weiss

Science (1996-10-25) https://doi.org/df7k88

DOI: 10.1126/science.274.5287.621 · PMID: 8849455

48. Committee Opinion No. 690: Carrier Screening in the Age of Genomic Medicine

Obstetrics & Gynecology

Ovid Technologies (Wolters Kluwer Health) (2017-03) https://doi.org/f92g56

DOI: 10.1097/aog.000000000001951 · PMID: 28225425

49. Public health genomics: The end of the beginning

Muin J Khoury

Genetics in Medicine (2011-03) https://doi.org/bjsxzk
DOI: 10.1097/gim.0b013e31821024ca · PMID: 21311338

50. Initial sequencing and analysis of the human genome

Eric S Lander, Lauren M Linton, Bruce Birren, Chad Nusbaum, Michael C Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William FitzHugh, ... Michael J Morgan *Nature* (2001-02-15) https://doi.org/bfpgjh

DOI: 10.1038/35057062 · PMID: 11237011

51. An STS-Based Map of the Human Genome

Thomas J Hudson, Lincoln D Stein, Sebastian S Gerety, Junli Ma, Andrew B Castle, James Silva, Donna K Slonim, Rafael Baptista, Leonid Kruglyak, Shu-Hua Xu, ... Eric S Lander *Science* (1995-12-22) https://doi.org/ftf

DOI: 10.1126/science.270.5244.1945 · PMID: 8533086

52. A New Five-Year Plan for the U.S. Human Genome Project

Francis Collins, David Galas

Science (1993-10) https://doi.org/fwkrnb

DOI: <u>10.1126/science.8211127</u> · PMID: <u>8211127</u>

53. Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations

Ivan Heibi, Silvio Peroni, David Shotton

Scientometrics (2019-09-14) https://doi.org/ggzz8b

DOI: <u>10.1007/s11192-019-03217-6</u>

54. A simple Havel-Hakimi type algorithm to realize graphical degree sequences of directed graphs

Péter L Erdős, István Miklós, Zoltán Toroczkai *arXiv* (2010-01-21) https://arxiv.org/abs/0905.4913

55. Exploring Network Structure, Dynamics, and Function using NetworkX

Aric A Hagberg, Daniel A Schult, Pieter J Swart Proceedings of the 7th Python in Science conference (2008)

56. Topic-based Pagerank: toward a topic-level scientific evaluation

Erjia Yan

Scientometrics (2014-05-06) https://doi.org/f6br99

DOI: <u>10.1007/s11192-014-1308-5</u>

57. Reproducibility standards for machine learning in the life sciences

Benjamin J Heil, Michael M Hoffman, Florian Markowetz, Su-In Lee, Casey S Greene, Stephanie C Hicks

Nature Methods (2021-08-30) https://doi.org/gmnnqh

DOI: 10.1038/s41592-021-01256-7 · PMID: 34462593 · PMCID: PMC9131851