BIBLIOMETRIC ANALYSIS OF SENIOR US MATHEMATICS FACULTY

JOSHUA PAIK AND IGOR RIVIN

Abstract. We introduce a methodology to analyze citation metrics across fields of Mathematics. We use this methodology to collect and analyze the MathSciNet (http://www.ams.org/mathscinet) profiles of Full Professors of Mathematics at all 131 R1, research oriented US universities. The data recorded was citations, field, and time since first publication. We perform basic analysis and provide a ranking of US math departments, based on age corrected and field adjusted citations.

Introduction

For as long as the authors can remember, there has been discussion of comparable quality of various researchers (in all fields of research, but the authors are most familiar with mathematics, so this paper concerns itself with mathematics exclusively). While such a comparison is not strictly speaking possible (mathematics is not like competitive swimming, where a single number determines the better swimmer), those of us who have been on hiring committees have needed to compare researchers in diverse fields, and those of us who have had students (or job offers) have had to have some sort of estimate of the quality of people independently of age and field (and building on that, to have some reasonably gauge of the quality of departments). The (admittedly ambitious) purpose of this note is to propose an objective metric, based entirely on citations data (as such, it can be gamed, as can any metric be).

Briefly, we normalize the number of MathSciNet citations by dividing it by the number of years since the author’s first paper raised to the magical power 1.3. We further segment mathematics into a number of “major fields”, assign mathematicians to fields (this is very difficult for some people, including, ironically, the authors of this paper), and compute the z-score of the normalized citation number. For each department we then compute the mean z-score of faculty to compute the department’s ranking.

1991 Mathematics Subject Classification. 00A99.
Key words and phrases. citations, ranking.
1. Data

Data was collected between January 16 - January 24, 2020. After accessing the list of R1 schools, we found the faculty lists from the relevant departmental web page, determined their level of seniority, and searched their profiles on MathSciNet. In total we collected citation records for 2807 math professors at 131 different institutions. We then collected the total citations, the year of earliest indexed publication, and the field of the most cited publication for each mathematician. A second pass through the data set occurred between January 30 - February 2, 2020, and discovered errors were corrected. This data set is as complete as possible.

There were initially 65 fields that were most cited as classified by MathSciNet, which we reduced to 20 fields, using the mapping in the appendix. This mapping was constructed using general expertise on the way each field worked, and is recorded in the appendix.

2. Exploratory Data Analysis

2.1. Distribution. Citations and Citations/Year\(^{1.3}\) appear exponentially distributed after transformation by a square root.

![Distribution of citations and an exponential qq-plot of square root citations.](image1)

Figure 1: Distribution of citations and an exponential qq-plot of square root citations.

![Distribution of citations per year\(^{1.3}\) and an exponential qq-plot of square root citations per year\(^{1.3}\).](image2)

Figure 2: Distribution of citations per year\(^{1.3}\) and an exponential qq-plot of square root citations per year\(^{1.3}\).
2.2. Citations/Year\textsuperscript{1.3} vs. Age. As noted in earlier work \cite{2}, citations and year are positively correlated, but citations per year\textsuperscript{1.3} and year are not correlated\footnote{The exponent 1.3 was determined via a statistical analysis}. We repeat this analysis to check for robustness and find that when linearly regressed, citations/year\textsuperscript{1.3} and year have a slope of 0.0338 with a 95\% confidence interval of $[-0.00899, 0.0766]$. The $p$-value is 0.122, so we fail to reject the null hypothesis that the slope is zero, and $R^2 = 0.03$. We conclude that citations/year\textsuperscript{1.3} and year are not correlated.

![Scatter plot of citations/year\textsuperscript{1.3} and Years elapsed since first publication. The red line indicates the regressed line with equation $0.0338\text{Year} + 8.40 = \text{Citations/Year}^{1.3}$. The $p$-value for slope is 0.122 and the $R^2 = 0.03$. We fail to reject the null hypothesis that the slope is zero and conclude they are not correlated.]

2.3. Factored Linear Models. Before proceeding with the analysis, we should assess the importance of the explanatory variables when looking at differences in citations between mathematicians. We do this by constructing nested linear models with all three combinations of \textit{Year} and \textit{Field}, and determine the best model with Akaike Information Criterion, AIC [1]. Per standard interpretation the lower the AIC, the better the model. Let $C =$Citations, $a =$Age, and $f =$Field.

$$
\begin{array}{|c|c|}
\hline
\text{Model} & \text{AIC score} \\
\hline
\log(C) = \beta_0a + \beta_1f + \epsilon & 9122.124 \\
\log(C) = \beta_0f + \epsilon & 9347.318 \\
\log(C) = \beta_0a + \epsilon & 9514.644 \\
\hline
\end{array}
$$

![Table of AIC scores for tested linear models. As the AIC score is lowest, the model consisting of both Age and Field is best, where Age impacts citations positively. For more detailed information on the model, refer to the ANOVA tables]

Figure 3: Scatter plot of citations/year\textsuperscript{1.3} and Years elapsed since first publication. The red line indicates the regressed line with equation $0.0338\text{Year} + 8.40 = \text{Citations/Year}^{1.3}$. The $p$-value for slope is 0.122 and the $R^2 = 0.03$. We fail to reject the null hypothesis that the slope is zero and conclude they are not correlated.
in the RMarkdown on Github. It is clear that certain fields contribute negatively to overall citations and other fields contribute positively.

While it is not appropriate to pick a model for the sole reason that it minimizes AIC, it makes sense to consider both age and field.

2.4. Fields. Different fields in mathematics have different citation practices. Some fields like Partial Differential Equations have more mathematicians, whereas some fields like Number Theory have fewer mathematicians. Some results from fields like topology are widely applicable across disciplines, whereas more obscure results are not. We quantify the bibliographic differences between fields. Note that the major fields below are larger categories containing potentially multiple MathSciNet tags, and the mappings are recorded in the appendix.

| Major Field               | Mean Citations | S.D. Citations | Mean Cit/Year$^{1.3}$ | Count |
|---------------------------|----------------|----------------|-----------------------|-------|
| PDE                       | 1472.07        | 2182.45        | 14.58                 | 372   |
| Computer Science          | 1260.44        | 2223.06        | 14.08                 | 225   |
| Probability               | 1165.92        | 1401.07        | 12.06                 | 137   |
| Harmonic Analysis         | 1120.12        | 1336.62        | 10.51                 | 200   |
| Combinatorics             | 1023.24        | 1673.53        | 10.08                 | 116   |
| Algebra                   | 934.42         | 1310.59        | 9.12                  | 220   |
| Algebraic Geometry        | 846.62         | 1308.80        | 9.51                  | 169   |
| Geometry                  | 890.68         | 1486.72        | 8.87                  | 311   |
| Number Theory             | 742.66         | 920.31         | 7.38                  | 159   |
| Dynamics                  | 560.44         | 555.17         | 7.33                  | 68    |
| Mathematical Physics      | 643.01         | 716.41         | 7.25                  | 96    |
| Analysis                  | 977.18         | 1951.95        | 7.15                  | 45    |
| Applied Mathematics       | 646.60         | 976.98         | 6.87                  | 299   |
| Group Theory              | 686.38         | 1151.64        | 6.74                  | 81    |
| Logic                     | 634.00         | 690.14         | 6.32                  | 55    |
| Complex Analysis          | 612.86         | 725.15         | 6.17                  | 115   |
| Lie Groups                | 512.02         | 590.59         | 4.78                  | 43    |
| Statistics                | 220.73         | 331.15         | 3.10                  | 83    |
| History                   | 74.0           | 104.65         | 0.677                 | 2     |
| Other                     | 5              | 6.61           | 0.074                 | 11    |

Figure 5: Mean citations and citations per year$^{1.3}$ including counts, split by field, from top to bottom ranked by mean citations per year$^{1.3}$ to account for age.

We ran a permutation test between each field to verify the observed partial order. We report the inconclusive differences ($p$-value greater than 0.05) between fields when comparing citations per year$^{1.3}$ in the appendix.
2.5. Ranking of Departments. The above figures shows that comparing mathematicians in two different fields is akin to comparing apples and oranges. The cleanest way to standardize this is to compute an interfield z-score of the citations per year\(^1\), and hence associating a “rank” to each mathematician. Then we computed the mean of the interfield z-scores of each full professor at the respective institution. We report here the top 20 schools and record the remaining schools in the appendix\(^2\).

(1) Princeton University
(2) Harvard University
(3) Stanford University
(4) University of Chicago
(5) Columbia University in the City of New York
(6) Massachusetts Institute of Technology
(7) University of California, Los Angeles
(8) University of Miami
(9) Yale University
(10) Brown University
(11) University of California, Berkeley
(12) New York University
(13) University of Oregon
(14) California Institute of Technology
(15) Duke University
(16) Stony Brook University
(17) Rutgers University-New Brunswick
(18) University of Virginia
(19) Texas A&M University
(20) Northwestern University

3. Conclusion

The rankings based on our normalized z-score (call it the PR score) correspond reasonably well with the “folk” rankings of mathematicians. While we do not want to flatter or insult individuals by giving their scores here, we do give a ranking of departments, and we see that it, again, corresponds well with the “folk” rankings. If they do not, we encourage the reader to look at the faculty pages of the departments in question. It seems, therefore, that there is, indeed, a fully quantitative way to produce meaningful rankings which work at least in a statistical sense - they fail for polymaths, and they also are less successful for mathematicians the bulk of whose

\(^2\)Where by "Schools" we mean "Mathematics departments" - for Universities with separate Pure and Applied math departments, the ranking will be different if the departments were to be combined.
work is not indexed by MathSciNet - in particular those who do interdisciplinary work.

REFERENCES

[1] Christopher M Bishop. *Machine learning and pattern recognition*. 2006.

[2] Joshua Paik and Igor Rivin. Data analysis of the responses to professor abigail thompson’s statement on mandatory diversity statements, 2020.

4. APPENDIX

4.1. Code and Data. Available at https://github.com/joshp112358/Differences.

4.2. Classifications.
| Major Field     | Sub Fields                                                                 |
|-----------------|-----------------------------------------------------------------------------|
| Algebra         | Algebraic Topology; Associative Rings and Algebra; Category theory, Homological algebra; Commutative rings and algebras; Field theory; General algebraic systems; K-theory; Linear and Multilinear Algebra, matrix theory; Associative rings and algebras; Order, lattices, ordered algebraic structures. |
| Algebraic Geometry | Algebraic Geometry                                                          |
| Analysis        | Difference and functional equations; Integral equations; Integral transforms, operational calculus; Ordinary differential equations; Real functions; Special functions. |
| Applied Mathematics | Approximations and expansions; Biology other natural sciences; Calculus of variations and optimal control, optimization; Fluid mechanics, Game theory, economics, social and behavioral sciences; Geophysics, Mechanics of deformable sciences; Mechanics of solids, Operations research, mathematical programming; Systems theory, control. |
| Combinatorics   | Combinatorics                                                              |
| Complex Analysis | Functions of a complex variable; Potential theory; Several complex variables and analytic spaces |
| Computer Science | Computer Science; Numerical Analysis; Information and communication, circuits. |
| Dynamics        | Dynamical Systems and Ergodic Theory                                        |
| Geometry        | Convex and discrete geometry; Differential Geometry; General topology; Geometry; Manifolds and cell complexes; |
| Group theory    | Group theory and generalizations.                                           |
| Harmonic analysis | Abstract harmonic analysis; Fourier analysis; Functional analysis; Global analysis, analysis on manifolds; Measure and integration, Operator theory. |
| History         | History and biography.                                                     |
| Lie Groups      | Topological Groups, Lie Groups.                                            |
| Logic           | Logic and foundations; Mathematical logic and foundations; Set theory       |
| Major Field          | Sub Fields                                                                 |
|---------------------|-----------------------------------------------------------------------------|
| Mathematical Physics| Classical thermodynamics, heat transfer; Mechanics of particles and systems; Optics, electromagnetic theory; Quantum theory; Relativity and gravitational theory; Statistical mechanics, structure of matter. |
| Number Theory       | Number Theory                                                              |
| Other               | Other                                                                       |
| PDEs                | Partial Differential Equations; Global Analysis, Analysis on manifolds     |
| Probability         | Probability theory and stochastic processes                                |
| Statistics          | Statistics                                                                  |

4.3. **Ranking of Departments.**

1. Princeton University
2. Harvard University
3. Stanford University
4. University of Chicago
5. Columbia University in the City of New York
6. Massachusetts Institute of Technology
7. University of California, Los Angeles
8. University of Miami
9. Yale University
10. Brown University
11. University of California, Berkeley
12. New York University
13. University of Oregon
14. California Institute of Technology
15. Duke University
16. Stony Brook University
17. Rutgers University-New Brunswick
18. University of Virginia
19. Texas A&M University
20. Northwestern University
21. University of Michigan
22. Rice University
23. The University of Texas at Austin
24. Carnegie Mellon University
25. University of Illinois at Chicago
26. University of California, Irvine
(27) University of Pittsburgh
(28) Georgia Institute of Technology
(29) University of Minnesota
(30) Vanderbilt University
(31) Indiana University Bloomington
(32) SUNY at Albany
(33) University of California, San Diego
(34) University of North Texas
(35) University of Washington
(36) University of Connecticut
(37) Arizona State University
(38) Pennsylvania State University
(39) University of Southern California
(40) Purdue University
(41) University of Illinois at Urbana-Champaign
(42) Cornell University
(43) University of Maryland - College Park
(44) University of Utah
(45) North Carolina State University
(46) Johns Hopkins University
(47) University of California, Riverside
(48) University of California, Santa Cruz
(49) Washington University in St. Louis
(50) Wayne State University
(51) University of Pennsylvania
(52) Brandeis University
(53) Colorado State University
(54) University of Notre Dame
(55) University of California, Santa Barbara
(56) University of North Carolina at Chapel Hill
(57) University of Houston
(58) University of Iowa
(59) The Ohio State University
(60) University of South Florida
(61) Michigan State University
(62) University of California, Davis
(63) Virginia Polytechnic Institute and State University
(64) University of Missouri
(65) University of Wisconsin - Madison
(66) University of Massachusetts Amherst
(67) University of South Carolina
(68) Emory University
(69) University of Central Florida
(70) University of Kentucky
(71) University of Florida
(72) University of Delaware
(73) Louisiana State University
(74) Syracuse University
(75) Georgia State University
(76) University of Colorado Denver
(77) Boston University
(78) Tulane University of Louisiana
(79) Clemson University
(80) University of Kansas
(81) University of Southern Mississippi
(82) Boston College
(83) Mississippi State University
(84) University of Rochester
(85) CUNY Graduate School and University Center
(86) The University of Tennessee
(87) George Washington University
(88) Georgetown University
(89) Florida State University
(90) Iowa State University
(91) University at Buffalo
(92) Northeastern University
(93) Tufts University
(94) University of Nebraska-Lincoln
(95) University of Georgia
(96) University of New Hampshire
(97) Virginia Commonwealth
(98) University of Cincinnati
(99) Dartmouth College
(100) Rensselaer Polytechnic Institute
(101) University of Nevada, Reno
(102) West Virginia University
(103) Auburn University
(104) The University of Texas at Arlington
(105) Texas Tech University
(106) University of Arizona
4.4. Inconclusive Permutation Tests between Fields. We report the results of a one sided permutation test, when comparing cit/year\(^{1.3}\) which failed to be significant at the 0.05 level. We record the hypothesis on the left and the \(p\)-value to the right.

- \(\text{PDE} \geq \text{Computer Science} \quad p\)-value: 0.397
- \(\text{PDE} \geq \text{Probability} \quad p\)-value: 0.0768
- \(\text{Computer Science} \geq \text{Probability} \quad p\)-value: 0.156
- \(\text{Probability} \geq \text{Harmonic Analysis} \quad p\)-value: 0.113
- \(\text{Probability} \geq \text{Combinatorics} \quad p\)-value: 0.1049
- \(\text{Harmonic Analysis} \geq \text{Combinatorics} \quad p\)-value: 0.3824
- \(\text{Harmonic Analysis} \geq \text{Algebra} \quad p\)-value: 0.0961
- \(\text{Harmonic Analysis} \geq \text{Algebraic Geometry} \quad p\)-value: 0.1807
- \(\text{Combinatorics} \geq \text{Algebra} \quad p\)-value: 0.2181
- \(\text{Combinatorics} \geq \text{Algebraic Geometry} \quad p\)-value: 0.3265
Applied Mathematics $\geq$ Complex Analysis — $p$-value: 0.2362
Applied Mathematics $\geq$ Lie Groups — $p$-value: 0.0616
Applied Mathematics $\geq$ History — $p$-value: 0.0534
Group Theory $\geq$ Logic — $p$-value: 0.3728
Group Theory $\geq$ Complex Analysis — $p$-value: 0.2834
Group Theory $\geq$ Lie Groups — $p$-value: 0.0513
Logic $\geq$ Complex Analysis — $p$-value: 0.4276
Logic $\geq$ Lie Groups — $p$-value: 0.0613
Complex Analysis $\geq$ Lie Groups — $p$-value: 0.0871
Lie Groups $\geq$ History — $p$-value: 0.0531
Statistics $\geq$ History — $p$-value: 0.2517
History $\geq$ Other — $p$-value: 0.1533

School of Mathematics and Statistics, University of St Andrews
Current address: Department of Mathematics, The Pennsylvania State University
E-mail address: joshdpaik@gmail.com

Mathematics Department, Temple University
Current address: Research Department, Edgestream Partners, LLP
E-mail address: rivin@temple.edu