A global bibliometric perspective on the social structure of science

Keywords: bibliometric data; research productivity; scientific collaboration; scientific mobility; scholarly impact and citations

Extended Abstract

Science is a social enterprise with hierarchy systems among its agents (1). The durability of hierarchies depends, among other things, on foundational myths about their inherent necessity. Social scientists have shown how these myths work for the consolidation of national communities (2) and the moral justification of socioeconomic inequality (3). This study presents a quantitative and multivariate assessment of hierarchies in scientific communities across six macro fields of science using large-scale bibliometric data based on 8+ million authors in Scopus from 1996 to 2021. We provide a bibliometric analysis of the scientific hierarchies that, in our view, are underpinned by four myths.

Common beliefs about increased scholars' collaborations (4,5), spatial mobility (6,7), productivity, and subsequently enhanced scholarly works' visibility (8) help invisibilize the depth and entrenched nature of academic hierarchies. Analyzed separately, time trends in scholars' collaboration, mobility, productivity and scholarly work visibility suggest that academia is successfully growing. 1) Number of scholarly publications and scientists' productivity have increased a lot. 2) There is a high and increased share of collaboration and internationalization in the co-authorship of scholarly publications (4,5). 3) There is a vast and growing number of scholars who experience mobility and it positively affects their performance, collaborations and impact (6). 4) Digitization of publication information has made searching and citing works easier and allows more potential visibility and impact for scientific publications.

However, as partially shown by some scholars, these figures about widespread increases in bibliometric measures miss important facts. 1) Many authors have only one article or review publication and the increased share of publications involves those scientists that have more than one publication. They shape the observed trends in research performance (9). 2) Many publications have only one author or group of authors from institutions in one national context. 3) Many authors have been affiliated with a single country throughout their career i.e., have not experienced scholarly mobility, and those mobile, have a very skewed distribution (11). 4) Citations have traditionally had a skewed distribution which is sustained and only a small share of publications, journals and authors receive disproportionately high citations (8).

Hence, increases in academic-success indicators are highly concentrated on a few authors and institutions (8). Additionally, the interrelation between these increasing trends further inequalities among scholars beyond additive effects and becomes a constituent part of scientific hierarchies. Whereas the unequal distribution of scholars' mobility, publications, citations, and collaborations has been documented, there is little knowledge of the interrelatedness of these trends. Are scholarly collaborations, mobility patterns, productivity, and scholarly works' visibility interrelated? If so, how do these interrelations translate into hierarchies in scientific communities?

Answering these questions requires analyzing author-level data and career-based measures embedded in scientific communities. Bibliometric data allows such an approach.

We rely on 12 well-established measures of collaborations (i.e., Number of coauthored publications; N. of nationally coauthored publications; N. of internationally coauthored publications; The average number of coauthors), mobility (i.e., N. of affiliated organizations throughout publication career; N. of national moves based on changes in academic affiliation addresses; N. of international moves), productivity (i.e., N. of publications; N. of publications as the first author; Fractional count of publications which considers the number of contributing authors), and visibility (i.e., an aggregated total count of citations; Average citations throughout publication career) for all authors with at least one publication in the Scopus database since 1996.

Authors' disambiguation allows us to identify groups of authors who publish together and construct a global network of co-authorship. Using these data, we identify the author's location within scientific hierarchies based on their academic achievements. We rely on a global factorial analysis to represent each author as a point in a multidimensional space based on the 12 bibliometric measures (see Figure 1). We use cluster algorithms to group authors with similar academic achievements. The cluster analysis is stratified by groups of academic age, measured as the number of years since the first publication plus one (i.e., One, 2 to 5, 6 to 9, 10 to 14, 15 to 20, and 21 to 25 years). This sample stratification warrants the comparability of authors' academic achievements. Next, we examine the distribution of authors' location within scientific communities in terms of co-authorship and collaboration. For this analysis, we pooled all academic-age groups. The pooled sample allows us to use academic age as a benchmark variable. We compared the distribution of authors within each scientific community according to their academic age and position within the bibliometric-based hierarchy. A side-by-side comparison of the bibliometric-based and academic-age distributions within scientific communities, along with entropy measures for these two distributions allow us to assess the nature and strength of stratification across scientific communities (see Figure 2).

Our results show that the claims of literature on increased productivity, collaboration, and mobility are principally driven by a small fraction of highly prolific, collaborative, mobile, and impactful scientists. These top 10% are driving the observed trends in the bibliometric literature. We find a hierarchically clustered structure with a small top class, and large middle and bottom classes. Investigating the composition of communities of collaboration networks in terms of these top-to-bottom classes and the academic age distribution shows that those at the top succeed by collaborating with a varying group of authors from other classes and age groups. Nevertheless, they are benefiting disproportionately to a much higher degree from this collaboration and its outcome in form of impact and citations.

References

1. Kozlowski D, Larivière V, Sugimoto CR, Monroe-White T. Intersectional inequalities in science. PNAS [Internet]. 2022 Jan 11 [cited 2022 Jan 5];119(2). Available from: https://www.pnas.org/content/119/2/e2113067119
2. Anderson B. Imagined Communities: Reflections on the Origin and Spread of Nationalism. Verso; 2006. 266 p.
3. Pikkety T. Capital et idéologie. 1st ed. Paris: Seuil; 2019. 1232 p.
4. Wuchty S, Jones BF, Uzzi B. The Increasing Dominance of Teams in Production of Knowledge. Science. 2007 May 18;316(5827):1036–9.
5. Jones BF, Wuchty S, Uzzi B. Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. Science. 2008 Nov 21;322(5905):1259–62.
6. Sugimoto CR, Robinson-Garcia N, Murray DS, Yegros-Yegros A, Costas R, Larivière V. Scientists have most impact when they're free to move. Nature. 2017 Oct;550(7674):29–31.
7. Scellato G, Franzoni C, Stephan P. A mobility boost for research. Science. 2017 May 19;356(6339):694–694.
8. Nielsen MW, Andersen JP. Global citation inequality is on the rise. PNAS [Internet]. 2021 Feb 16 [cited 2021 Feb 9];118(7). Available from: https://www.pnas.org/content/118/7/e2012208118
9. Fanelli D, Larivière V. Researchers' Individual Publication Rate Has Not Increased in a Century. PLOS ONE. 2016 Mar 9;11(3):e0149504.
11. Sanliturk E, Zagheni E, Dańko MJ, Theile T, Akbaritabar A. Global patterns of migration of scholars with economic development. Proceedings of the National Academy of Sciences. 2023 Jan 24;120(4):e2217937120.

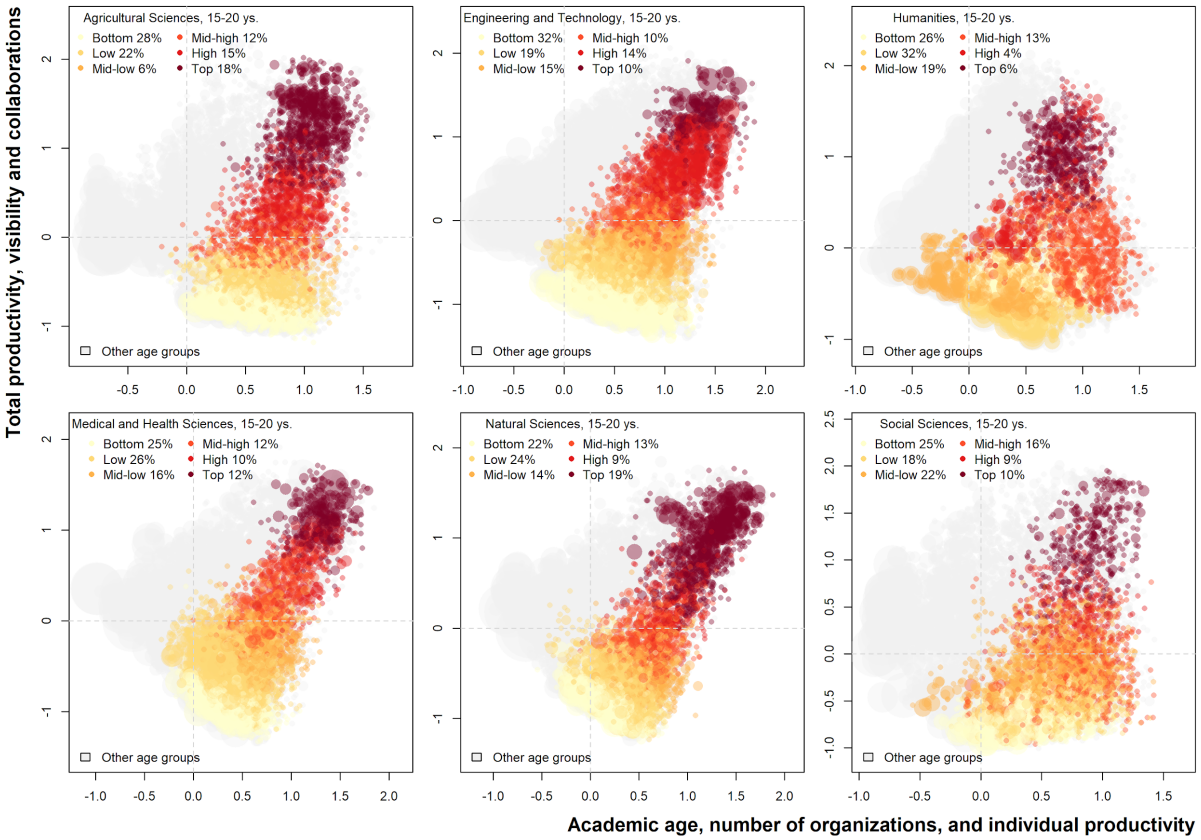

Figure 1. The social structure of science and hierarchy of six identified clusters from top to bottom in six macro fields of science. Multiple Correspondence Analysis (MCA) results using the 12 most widely used bibliometric variables allowed identifying six classes of scientists from Bottom, Low, Middle low, Middle up, High, to Top. In all six fields of science and five-year career groups from a minimum of 1 to a maximum of 25 years of publication career indexed in Scopus, we see the same hierarchical structure appearing. A minority of the top class is identified which consists of less or about 10% of the most successful scientists indicated with dark red colors in the figure.

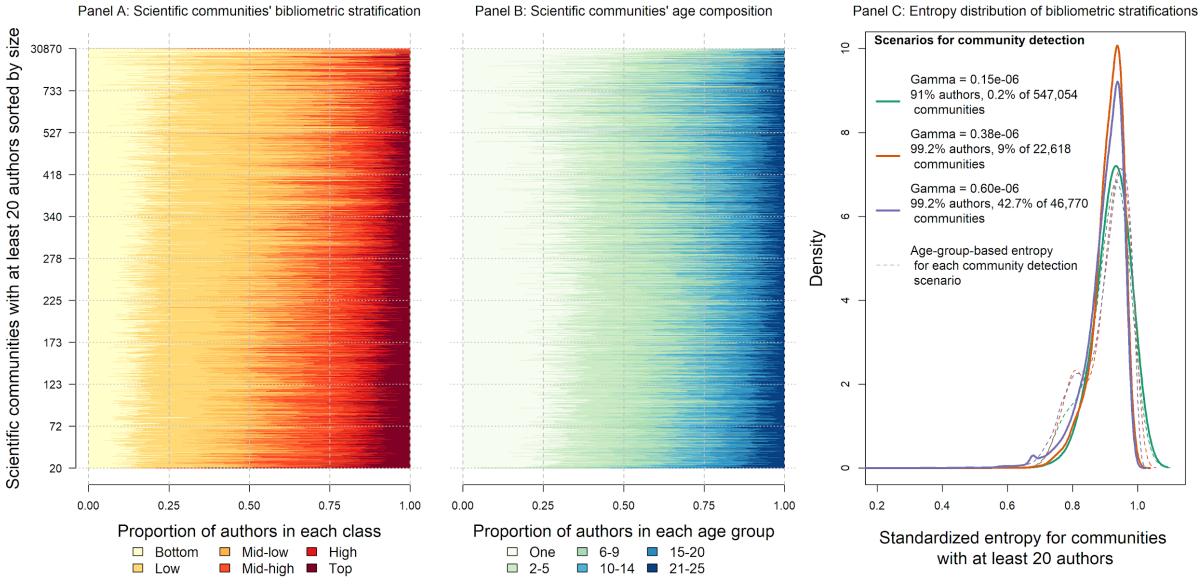

Figure 2. Composition of communities of collaboration in terms of top to bottom classes (left) and age groups (middle) and entropy of stratifications (right). To investigate the trends shown in Figure 1 further and control the collaboration structure among the classes, we turned to co-authorship networks of the studied 28 million publications. Networks of collaboration in terms of co-authoring scientific publications among 8.2 million authors worldwide allowed us to identify communities of collaboration. We used the Constant Potts Model (CPM) and its extension for bipartite networks with a varying range of 18 thresholds for the resolution parameter to detect communities. In all these detected communities (only 3 shown in panel C in the figure to preserve clarity), we investigated the class and age composition of members. Independent from the threshold used, all these communities have a heterogenous composition of classes and age groups and analysis of entropies of this stratification indicates an inter-class and inter-age collaboration structure among the most and least prolific, collaborative/internationalized, and mobile scientists.