PROCEEDINGS OF SCIENCE

A scientometric analysis of diversity in HEP over the past three decades

Maria Grazia Pia[*†]
INFN Sezione di Genova
E-mail: mariagrazia.pia@ge.infn.it

Tullio Basaglia
CERN
E-mail: tullio.basaglia@cern.ch

Zane W. Bell
ORNL
E-mail: bellzw@ornl.gov

Arnold Burger
Fisk University
E-mail: aburger@fisk.edu

Paul V. Dressendorfer
IEEE
E-mail: p.dressendorfer@ieee.org

This study addresses various aspects of diversity in high energy physics through a scientometric analysis of the literature spanning approximately three decades, from the LEP and Tevatron era to the LHC era. It concerns both fundamental physics and technological research related to high energy physics, and compares the evolution of some diversity parameters in this field with that in other research domains, such as nuclear physics and astrophysics. The data are collected from the Web of Science and are analyzed by means of econometric methods and techniques pertaining to statistical ecology.

The 39th International Conference on High Energy Physics (ICHEP2018)
4-11 July, 2018
Seoul, Korea

[*]Speaker.

[†]The author thanks the University of Genova, Department of Physics, for providing her access to the Web of Science (Clarivate Analytics).

PoS(ICHEP2018)511

1. Introduction

This paper summarizes a scientometric analysis of publications in scholarly journals pertaining to the fields of high energy physics (HEP), astrophysics and nuclear physics. The period covered by the analysis, from 1985 to 2017, encompasses the activity at major colliders (LEP, Tevatron, beauty factories and LHC), as well as the lifecycle of several fixed target and astroparticle experiments.

The analysis examines social and scientific characteristics of the publications, such as their geographical distribution, the distribution of participating research institutions and the spectrum of scholarly journals where they are published. The data are collected from the Web of Science [1] and are analyzed by means of econometric methods and techniques pertaining to statistical ecology.

Due to the page limit in the conference proceedings, this papers reports only a brief overview of the methods and results, which will be more extensively documentated in a forthcoming publication.

2. Analysis method

An original methodology, based on statistical techniques, has been developed to evaluate scientometric data with respect to diversity. It encompasses measures of diversity [2], derived from the domains of information theory and ecology, and measures of inequality [3] pertaining to econometrics, whose evolution is objectively appraised by means of trend tests [4, 5, 6]. The scientometric analysis concerns physics-oriented and technology-oriented publications in high energy physics, astrophysics and nuclear physics. The analysis software has been developed in R [7].

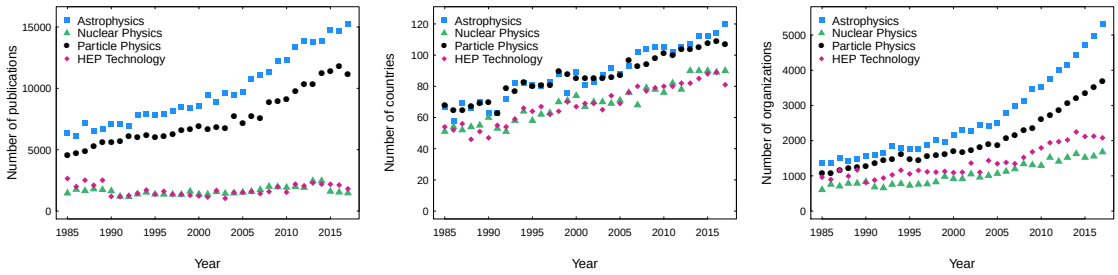
Simpson index and Shannon entropy are commonly used as measures of diversity. Shannon entropy, first created in the context of information theory [8], measures the minimum volume of communication required to code a message and is related to the concept of multiplicity of states. Simpson index [9] encodes the probability that two entities randomly taken from a data set represent the same type. Renyi's entropy [10] is a generalization of Shannon Entropy; a parameter in its formulation controls the relative importance of rare species. Hill numbers[11] are a mathematical family of diversity indices, where a parameter accounts for the effective number of species.

3. Overview of the results

One can observe a general increase of the number of publications, of journals, of participating countries, organizations and authors over the period subject to analysis; a sample of results is shown in Figure 1. Although the yearly data distributions exhibit low and approximately constant median values over the whole period, outliers extend up to very large number of publications.

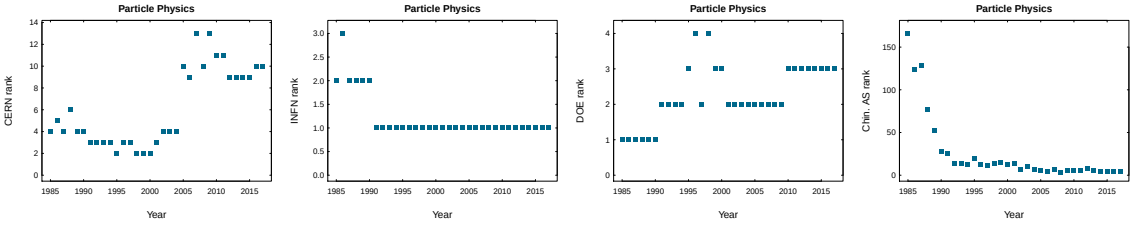
The scientometric data show the evolution of the role of entities traditionally active in HEP research as well the appearance of new players on the scene. As a example, Figure 2, which reports the relative position of a few representaive organizations as a function of time: lower ranks reported in the figures correspond to more prominent position in terms of number of HEP publications.

As an example, Figure 3, shows the Hill number of order 1, corresponding to the exponential of Shannon entropy, as a function of time: one can observe a general evolution towards greater diversity of countries, organizations and authors in all the examined research domains. Sensitivity



(a) Number of papers (b) Number of countries (c) Number of organizations

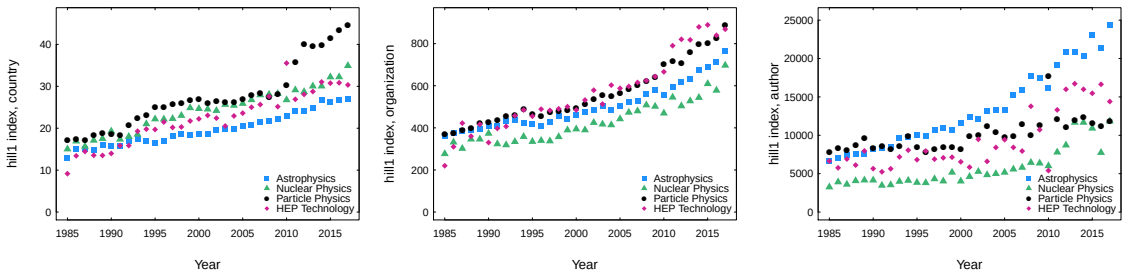
Figure 1: Evolution of publications in HEP journals.



(a) CERN (b) INFN (c) USA DOE (d) Chinese Acad. Science

Figure 2: Rank of representative organizations regarding the number of published papers.

to scarcely represented entities is highlighted by the measurement of Renyii's entropy at various orders; an example is shown in Figure 4.



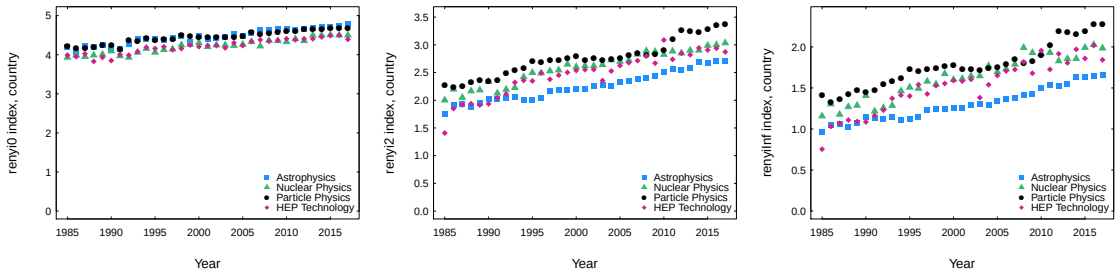
(a) Countries (b) Organizations (c) Authors

Figure 3: Diversity of countries, organizations and authors, measured by Hill number of order 1.

Inequality has decreased across countries, but HEP publications are more and more concentrated within a small number of organizations and authors. Figure 5 shows this evolution; similar trends are observed in Gini [12], Pietra [13] and Atkinson [14] inequality indices. Organization and author inequality in technological publications is lower than in physics publications.

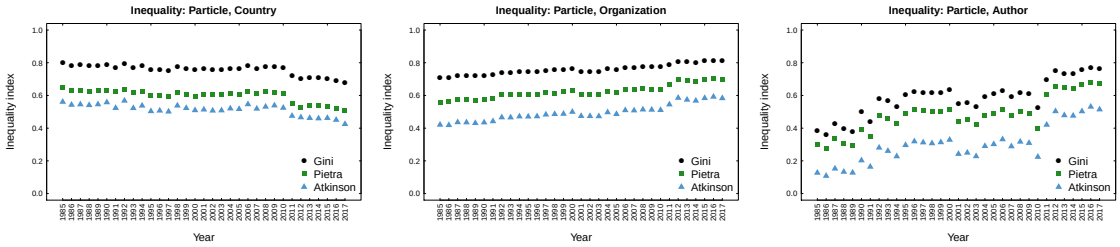
4. Conclusions

In general, one observes evolution towards greater diversity in HEP, but increasing concentration in a small number of organizations. Physics and technological papers exhibit different patterns.



(a) Order: 0 (b) Order: 2 (c) Order: infinite

Figure 4: Diversity in countries measured by Renyi's entropy of order 0, 2 and infinite.



(a) Countries (b) Organizations (c) Authors

Figure 5: Inequality measured by Gini, Pietra and Atkinson indices.

References

[1] Clarivate Analytics, *Web of Science*, http://apps.webofknowledge.com.

[2] A. E. Magurran, *Measuring biological diversity*, Blackwell Publishing, Oxford, UK, 2004.

[3] F. A. Cowell, *Measurement of Inequality*, in *Handbook of Income Distribution*, North Holland, 2015.

[4] H. B. Mann, *Non-parametric tests against trend*, Econometrica, vol. 13, pp. 163-171, 1945.

[5] M. G. Kendall, *Rank Correlation Methods*, 4^{th} edition, Charles Griffin, London, 1975.

[6] D. R. Cox, A. Stuart, *Some quick sign test for trend in location and dispersion*, Biometr. 42:80, 1955.

[7] R Core Team, *R: A language and environment for statistical computing* R Foundation for Statistical Computing, Vienna, Austria, 2018. URL https://www.R-project.org/.

[8] C. Shannon, *A mathematical theory of communication*, Bell System Tech. J. 27: 379, 1948.

[9] E. H. Simpson, *Measurement of diversity*, Nature 163, p. 688, 1949.

[10] A. Renyi, *On measures of entropy and information*, in *Proc. 4th Berkeley Symp. Math. Statist. Probabil.*, Vol. 1, pp. 547-561, Univ. of California Press, 1961.

[11] M. Hill, *Diversity and evenness: a unifying notation and its consequences*, Ecology, 54 (2) 427, 1973.

[12] C. Gini, *Variabilità e Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche*, C. Cuppini Edition, Bologna, 1912.

[13] G. Pietra, *Studi di Statistica Metodologica*, Giuffré, Milano, 1948.

[14] A. B. Atkinson, *On the measurement of inequality*, J. Econ. Theory 2 (3) 244–263, 1970.