

A comparison of three multidisciplinary indices based on the diversity of Scopus subject areas of authors' documents, their bibliography and their citing papers

Ugo Moschini · Elena Fenialdi · Cinzia Daraio · Giancarlo Ruocco · Elisa Molinari

THIS VERSION IS THE PRE-PRINT VERSION OF THE PAPER PUBLISHED IN SCIENTOMETRICS UNDER DOI <https://doi.org/10.1007/s11192-020-03481-x>

Abstract In this paper, we compare the distribution of Elsevier Scopus subject areas of authors documents, their bibliographical references and their citing documents. We compute the complement of the Herfindahl-Hirschman (CHH) index as a measure of multidisciplinary. We analyse a sample of 120 researchers belonging to two groups, one from the Italian Institute of Technology (IIT, whose work is expected to be highly multidisciplinary) and one from the National Institute for Nuclear Physics (INFN, whose work is expected to be much less multidisciplinary). We show that the two groups are distinguishable through the measured index values. By using the subject areas of authors bibliographical references we obtain a better identification of the two groups than relying on the subject areas of the authors documents. We then extend the analysis to 3,317 researchers belonging to seven Italian Scientific-Disciplinary sectors (SSD) providing insights about the degree of multidisciplinary within each SSD. The results seem interesting for assessing the interdisciplinarity of younger researchers with scarce scientific output and few citations.

Keywords Multidisciplinary · Herfindahl-Hirschman Index · Science mapping

Moschini Ugo, Ruocco Giancarlo, Molinari Elisa
Istituto Italiano di Tecnologia, Italy
E-mail: ugo.moschini, giancarlo.ruocco, elisa.molinari@iit.it

Fenialdi Elena
University of Genoa, Italy
E-mail: elena.fenialdi@gmail.com

Daraio Cinzia
Sapienza University of Roma, Italy
E-mail: daraio@dis.uniroma1.it

Article Highlights

1. Analysis of the Herfindahl-Hirschman index as a measure of multidisciplinary for more than 3000 researchers.
2. Index analysed on distributions of Scopus disciplines of documents, bibliographical references and citing documents.
3. The index performs well and using the authors bibliographical references seems a good way forward.

1 Introduction

Nowadays, measuring and understanding *multidisciplinary* research is of prominent interest for both researchers and funders or evaluators (Wagner et al. 2011; Wang et al. 2015), also to choose proper assessment parameters that go along with the characteristics typical of a particular scientific field. Multidisciplinary works and consequent applications are generally considered to have a stronger impact on the society and the scientific development (Rafols and Meyer 2007). In spite of this growing interest, both defining and measuring multidisciplinary is not a trivial task. Several definitions have been given to describe the multiple modalities in which disciplines can interact (Stokols et al. 2003; Choi and Pak 2006; Porter et al. 2007). For a systematic review, see (Rousseau et al. 2019). We use the term multidisciplinary in a broad sense, indicating that elements from different disciplines are present. The term "multidisciplinary" is used in the following and the concept of interdisciplinarity is inherent to it. Since many years, a majority of quantitative indicators of the degree of multidisciplinary of a researcher are based on bibliometric methods (Porter et al. 2007). These methods are commonly classified in bottom-up and top-down approaches (Wagner et al. 2011). Bottom-up approaches are based on grouping and forming sets of articles according to a criterion, like building a co-citation network (Boyack and Klavans 2010). This approach is suitable for finding emerging fields, in which there is no classification available, very few publications and no a priori taxonomy (Leydesdorff 2007; Leydesdorff et al. 2013). Top-down approaches are dependent on some kind of classification available. They are suitable for a large-scale analysis, especially when dealing with big amount of data (Porter and Rafols 2009; Leydesdorff et al. 2013). In fact, published manuscripts are commonly included and indexed in several databases such as Elsevier Scopus and Clarivate Web of Science, the two main databases due to their curated large coverage catalogues. A feature offered by such databases is that main topics or research areas are assigned to articles and journals via (semi-)automatic methods. Bibliometric data and their classification into topics are the raw material used to measure interdisciplinarity. In 2007, Stirling introduced a general framework for analysing diversity in science, that takes into account variety, balance and disparity, that are properties of a *diversity* measure (Stirling 2007). Not only the number of disciplines in articles (or in their reference lists) are taken into account, but also the *distance* between them: he suggested to apply the Rao index developed in biology (Rao 1982) to measure research multidisciplinary, defining the so-called Rao-Stirling index. The Rao-Stirling index is a popular indicator of inter-

disciplinarity, used in many works (Rafols and Meyer 2007; Leydesdorff and Rafols 2011; Leydesdorff et al. 2013; Wagner et al. 2011). However, creating a meaningful distance measure is far from being an easy task (Rafols and Meyer 2007). This requires the selection of a context (a set of papers), the identification of attributes on which a distance measure is based and so on.

In our work, we are going to analyse the behaviour of the Rao-Stirling index when no distance measure is used. This index was demonstrated to be equal to the Herfindahl-Hirschman index, also known as Simpson (Simpson 1949) index, very popular in economics (Rhoades 1993; Rousseau 2018), with different choices of parameters (Porter and Rafols 2009). To measure multidisciplinary, we use the complement of the Herfindahl-Hirschman index (CHH), named Herfindahl's diversity in Porter and Rafols (2009). To define disciplines, we use the top-down classification provided by the Elsevier Scopus. The disciplines are named *subject areas* (SA), as in Scopus. We use the CHH index to analyse the differences in the distribution of subject areas in three cases: the subject areas of an authors documents, of his/her documents bibliographical references and the citing documents of his/her documents. The evaluation is conducted on two sets of researchers and their publications in the years 2010-2018. The first set is made of researchers working at the Italian Institute of Technology (IIT) that is a Foundation established by Law no. 326/2003, and financed by the Italian State to conduct scientific research in the public interest, for the purpose of technological development (see further information at: <https://www.iit.it/about-us/institute>) and according to its strategic plan is engaged in highly multidisciplinary research activities. These researchers, then, are expected to be highly multidisciplinary. The second set contains researchers affiliated with the National Institute for Nuclear Physics (INFN, further information are available at <http://home.infn.it/en/>) in Italy. These researchers are expected to be much less multidisciplinary. Section 5 extends the analysis of Moschini et al. (2019) by investigating the multidisciplinary of 3,317 researchers. This larger set contains all the researchers active in Italy in seven Scientific-Disciplinary Sectors (SSD). Section 6 puts forward possible extensions of the CHH index and summarizes the main results.

2 Distribution of subject areas

For the all the sets of researchers in the paper (IIT, INFN, researchers in SSDs), we considered only the documents published in the period 2010-2018, as indexed in Scopus. Documents were accessed through the Author Scopus identifiers of each researcher profile on Scopus: the subject areas were retrieved via a script based on the Scopus Application Programming Interface (Elsevier 2019). The Scopus *subject areas* (SAs) are classified in 334 categories. The 334 categories are grouped in 27 macro-categories. Scopus assigns a varying number of subject areas to the *source medium* in which a document appears (i.e., journal, conference proceedings, book, ...), rather than to the document itself. Since Scopus assigns subject areas to journals or conferences and not to the article itself, the assumption is that an article inherits the SAs of the source where it is published. In our analysis, we distinguished three differ-

ent ways to consider the subject areas that could potentially describe the disciplines belonging to the scientific production of a researcher:

1. SAs of the documents written by a researcher (**DOC**).
2. SAs of the bibliographical references of every document written by a researcher (**BIB**).
3. SAs of the publications that cite the documents written by a researcher (**CIT**).

We computed the percentage of publications falling in each subject area, for each of the three case. Note that Scopus frequently assigns two or more SAs to a publication: the percentages were then normalized to range from 0 to 100. In the following, the subject area "Multidisciplinary" (MULT) is excluded: it indicates journals that publish articles coming from several disciplines (e.g., "Nature" is classified as MULT in Scopus), rather than inherently multidisciplinary articles and it was removed from the analysis. We are going to refer to a total of 333 SAs and 26 macro-categories. A researcher is described by three vectors of 333 components, each vector component representing a given SA.

3 The index used in our work

To measure the interdisciplinarity of authors research outputs we used the complement of the Herfindahl-Hirschman (CHH) index, also known as Simpson diversity index. The Herfindahl-Hirschman index is a very popular concentration index in economics, used to measure how the market share is distributed among companies. In our context, the market-share distribution among companies translates into the subject areas related to a researcher, computed as a vector containing the percentage of publications categorized under each subject area, as explained in Section 2.

Let us call V a vector containing percentage of publications per SA and let us assume its values are normalised to range between 0 and 1. It is easy to see that less concentration means more multidisciplinary. Let us define $CHH(V)$ as the complement of the Herfindahl-Hirschman index, computed on a vector V whose components represent N subject areas, as:

$$CHH(V) = 1 - \sum_{i=1}^N v_i^2, \quad v_i \in V \quad (1)$$

The higher the value, the less concentrated are the subject areas, hence more multidisciplinary is achieved. The value returned by Equation 1) takes values in the range $0 - \frac{1}{N}$.

4 Experiments on IIT and INFN dataset

The first experiments relate two sets of researchers. The first set contains 64 researchers affiliated with the Italian Institute of Technology (IIT), expected to show an high multidisciplinary degree: their scientific output often combines, for example, robotics with life sciences, medicinal chemistry with biology, and so on. The

second set contains 56 researchers affiliated with the National Institute for Nuclear Physics (INFN), whose work has inherently a narrower scope: hence, less interdisciplinarity is expected. They are senior researchers, i.e. with a number of publications at least larger than 20. We are interested in verifying if the CHH index can actually distinguish two sets of researchers with values of multidisciplinary expected to be very different, as for our IIT and INFN sets: the index should show higher values on the IIT set compared to the INFN set. The aim of our tests is to check the index discriminating power between the two groups of researchers, when considering i) the subject areas of the authors documents, ii) the bibliographical references of authors documents, iii) or the citing documents of authors documents. The computed index is the CHH index on the vectors of percentages of publications falling in each Scopus subject area, for the 120 researchers, for the three cases DOC, BIB and CIT described in Section 2. Only documents published within the years 2010-2018 were considered, to avoid a potential correlation between multidisciplinary and the age of a researcher. Each of the 120 researchers published documents before 2010: these documents were simply discarded.

Figure 1 shows the results obtained after computing the CHH index, multiplied by 100 for convenience. In the plots, the IIT researchers are indicated in black colour, while the INFN researchers in grey. Figure 1a shows the index values unsorted and grouped by researchers affiliation, for the three cases. In the plots, we notice that the range of index values for INFN researchers is much broader in the BIB and CIT case than in the DOC case. The mean value for the INFN researchers in the DOC case is much more different from the BIB and CIT case. Instead, the mean values for IIT researchers fall in a similar range in each of the three cases. Table 1 shows in details the range of indexes and the mean/median values for the IIT and INFN researchers for the three cases. In Figure 1b, the index values are shown sorted, for each case. Ideally, all the IIT researchers should stay above the INFN researchers, due to the fact that they are more multidisciplinary. We notice that the two sets are split up to a good extent. While an ideal situation occurs in the CIT case, in which all the INFN researchers stay below the IIT sets (see Figure 1b), unfortunately it does not hold in DOC and BIB: there are researchers who cross the boundary of their own group, *overlapping* with the others. Interestingly, the same researcher is common to both DOC and BIB. The actual scientific production of these "overlapping" authors should be further investigated: their research could be not so monothematic or multidisciplinary as we thought initially. The limit between what can be multidisciplinary and what cannot be is unlikely a rigid and exact boundary: it is a promising sign, however, that researchers of a set overlap with researchers of the other set only near the boundary region. If we select a few researchers less multidisciplinary and less "monodisciplinary" (i.e., at the boundary of the black and grey "regions" of Figure 1b), we note that most of them are the same researchers across the three cases: they have approximately the same ranking position in each case. Then, we extended our analysis to study how the relative ranking positions of the researchers change throughout DOC, BIB and CIT. We measured that 45.83% of the researchers remained in a range of ± 5 positions, comparing the ranking of DOC w.r.t. BIB. The percentage goes up to 60.83% for BIB w.r.t. CIT. If the range considered is ± 10 positions, the percentage is 62.5% and 80.83%, considering DOC w.r.t. BIB and BIB w.r.t. CIT. Also, the experiments show

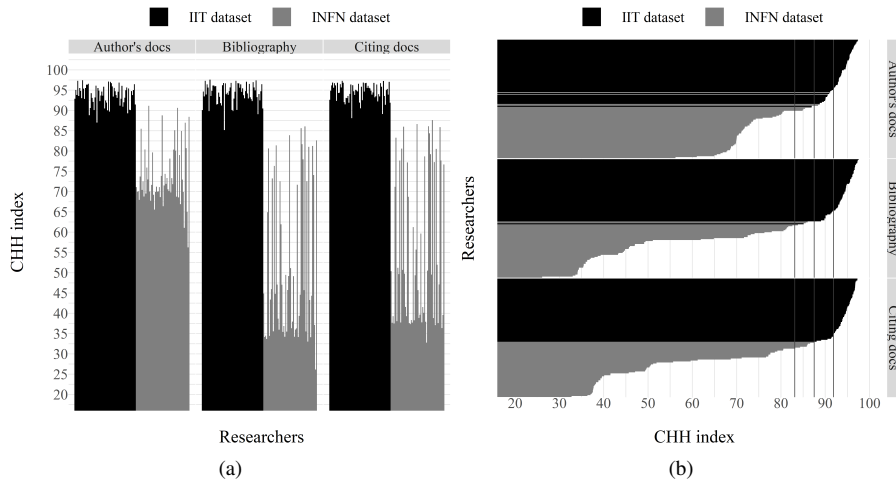


Fig. 1: (a) values of the complement of the Herfindahl-Hirschman (CHH) index computed on the subject areas of the authors documents (DOC), the bibliographical references of authors documents (BIB) and the citing documents of authors documents (CIT) on 120 IIT and INFN researchers; (b) sorted values of CHH index on the same sets. CHH index values are multiplied by 100 for convenience.

Case	Researchers	Min	Max	Median	Mean
DOC	IIT	0.8697	0.9742	0.9378	0.9371
DOC	INFN	0.5622	0.9113	0.7132	0.7362
BIB	IIT	0.8513	0.9756	0.9457	0.9413
BIB	INFN	0.2611	0.8605	0.4480	0.5165
CIT	IIT	0.8807	0.9727	0.9485	0.9452
CIT	INFN	0.3276	0.8760	0.4957	0.5508

Table 1: Min/max, median and mean values of the complement of the Herfindahl-Hirschman index for the cases DOC, BIB and CIT described in Section 2 and the IIT/INFN sets of researchers.

that the ranking changes less on the INFN set, although the size of the sets is not enough for drawing precise conclusions. Overall, only about 10% of the researchers maintain the same position in the ranking across the cases. By and large, it appears that researchers keep roughly their position (about 85% of the researchers stay in a range of ± 15 positions) in the three cases. We computed also the Kendall rank correlation among the vectors of percentages in the three cases. On the whole set of 120 researchers, Kendalls correlation is 0.78 and 0.81, for DOC w.r.t. BIB and DOC w.r.t. CIT, respectively. Considering the IIT set, the correlation decreases to 0.61 and 0.67, while on the INFN set it goes down to 0.54 and 0.58. This means that the order of the researchers generated by the index based on the three different vectors (DOC, BIB,

SSD	Topic	Number of researchers
FIS/02	Theoretical Physics, mathematical models and methods	251
MAT/05	Mathematical Analysis	589
MED/50	Applied medical techniques	80
ING-INF/05	Information processing systems	637
MED/09	Internal Medicine	626
CHIM/03	General and Inorganic Chemistry	450
BIO/10	Biochemistry	684

Table 2: The table describes the 7 Italian Scientific-Disciplinary Sectors (SSD) and the number of researchers categorised in each SSD.

CIT) varies largely within each set, but the overall order keeps the separation of the two sets, hence we measure higher correlation when considering all the researchers together. This happens because Kendalls correlation counts how many concordant or discordant couples are present: an increase in correlation means that the percentage of concordant couples (couples of researchers for whom the order does not change) is higher considering the whole sets of researchers together.

We observe that DOC, BIB and CIT may measure different kinds of multidisciplinary: the citations indicate the visibility of an authors work among various disciplines rather than the multiple disciplines a work is based on. Integration of different disciplines is probably better shown by the references cited by an article in its bibliography (Porter et al. 2007): an authors own publications would instead indicate how diverse the individual production is. From our study, the use of bibliographical references (the BIB case) showed a better separation between IIT and INFN than the DOC case: the mean/median are further apart and less researchers end up in the wrong region. The use of the bibliographical references gives good clues about multidisciplinary and also allows for measuring younger researchers whose scientific output is still limited or who are not cited many times: the bibliography is normally composed by a large number of entries, making it suitable in these situations, too. It would also be interesting to check if a correlation exists between long reference lists (i.e., knowledge base) and being cited by a more multidisciplinary audience (examining the CIT case).

5 Experiments on seven Italian Disciplinary Sectors

Section 4 shows that the CHH index computed on bibliographical references' data discriminates well the two sets of researchers with lower and higher multidisciplinary. In this section, we extend the analysis to a larger set of 3,317 Italian researchers. In the Italian system, academic research areas are categorized into officially pre-assigned Scientific Disciplinary Sectors (SSDs), as defined by the Attachment A of the Ministerial decree n. 855 of October 30, 2015. The full list is also available on Wikipedia Italy (permalink <https://w.wiki/Byp>). There are 367 SSDs, divided into 88 macro-sectors distributed over 14 main disciplinary areas. Every Italian researcher in the Italian academic system belongs to a SSD. Multidisciplinary can

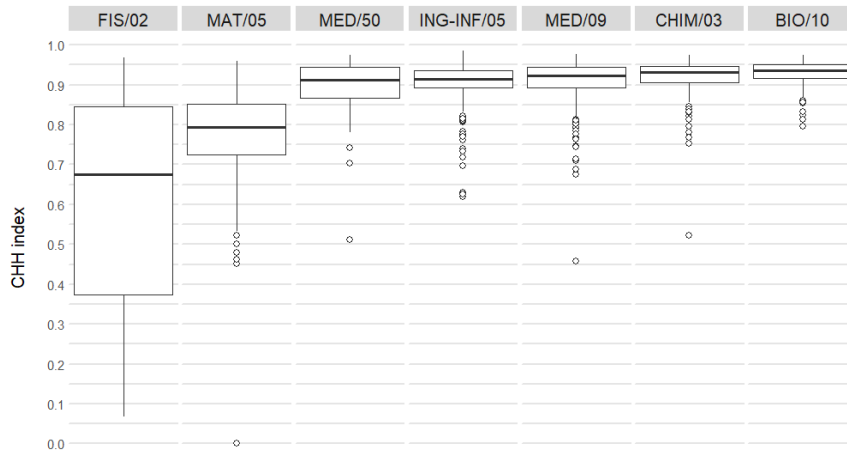


Fig. 2: The boxplot shows the distribution of the values of the CHH index (in the range 0-1) for each of the seven Scientific-Disciplinary Sectors (SSD) analysed. FIS/02 (and to a lesser extent MAT/05) shows a large spread of values.

be analysed then within a particular SSD classification, available for all the authors. [Abramo et al. \(2012, 2018\)](#) measured multidisciplinary through author's affiliations and SSDs. In our work, instead, we measure multidisciplinary through the computation of the CHH index. The whole population of researchers in seven SSD is selected, as detailed in Table 2. Similarly to what described in Section 2, the subject areas of each author's publications were retrieved from Elsevier Scopus to build the corresponding 3,317 vectors (V) of subject areas, computed on the bibliographical references of the authors' documents (BIB case).

In general, when evaluating researchers, it would be important to define a comparison set made of researchers who showcase similar features, inherent to their own actual scientific production. Researchers belonging to the same SSD can be fairly compared if they have similar multidisciplinary values and their scientific production spans similar areas. For each SSD, the CHH index on the subject areas of the bibliographical references' data is computed. Figure 2 shows the results. The boxplots of the SSDs are sorted in increasing order of the median value of the index. Some SSDs, such as ING-INF/05, MED/09, CHIM/03 and BIO/10, show a low level of dispersion of the CHH index. On the contrary, the dispersion of FIS/02 and MAT/05 is considerable. This means that within "Theoretical Physics (FIS/02)" and the "Mathematical Analysis (MAT/05)" disciplinary sectors, there are researchers with varying values of the CHH index. The scientific production of FIS/02 and MAT/05 researchers seems more multidisciplinary or at least their scientific production spans a broader set of disciplines compared to the other disciplinary areas considered.

We analyse the profile of FIS/02 to understand in more details the behaviour of the index in these two SSDs. We choose to run in R-language the well-known k -

means clustering method¹ to see how the set of FIS/02 researchers gets split into two clusters ($k = 2$) according to their vectors of subject areas' distributions (BIB case). The two panels in Figure 3 illustrate the discipline distributions for the researchers grouped in the two clusters identified by k -means. For each of the two clusters, the average percentage of publications falling in each subject area is shown. The first group (upper panel in Figure 3) contains researchers with a single main subject area related to PHYS (Physics), with code X3106; the second group (lower panel in Figure 3) contains researchers whose activity spans more subject areas, linked to both PHYS and MATH (Mathematics). From these results, the whole set of researchers in FIS/02 is split in two clusters that show different publication profiles. The cluster in the lower panel is more multidisciplinary than the other: this explains the fact that the CHH index values related to FIS/02 in Figure 2 span a wide range of values: within FIS/02, there are multidisciplinary and less multidisciplinary researchers. Similar reasoning and considerations apply to MAT/05, too. A hypothetical fair comparison among researchers within FIS/02 should account for the different aspects of their scientific production, as highlighted by the distribution of SAs. The CHH index and the inspection of the subject areas involved can possibly point out differences in scientific production, useful for evaluation purposes and for building a more homogeneous set of researchers sharing similar publication characteristics.

6 Conclusions and further extensions

6.1 Possible extensions using Topic Clusters

As explained in Section 4, the CHH index well indicates researchers whose research is very "monodisciplinary". However, from the box plots in Figure 2, it appears that for some SSDs, the index values have median value close to the maximum: the values *saturate* quickly, even if a researcher is characterized by a relatively small number of disciplines. This holds especially for ING-INF/05, MED/09, CHIM/03 and BIO/10. We could compute the CHH index on a different distribution of research areas. Moreover, given that the SAs are assigned by Scopus to journals and not to the actual publications of a researcher, we work under the (ambiguous) assumption that an article inherits the SAs from the source where it is published. Developments of this work will study the behaviour of the CHH index computed on author's Topics or Topic Clusters, a science mapping developed by Elsevier for its product SciVal. In short, SciVal Topics' calculation relies on Scopus publications: clusters of nodes are found on a graph, generated using direct-citation analysis among all existing documents. Clusters of node define a Topic. On the same graph, Topic Clusters are formed by aggregating nearby Topics, to define a broader area of research. More than 100,000 Topics are clustered in around 1,600 Topic Cluster: a publication belongs to one Topic (and one Topic Cluster) only. More details about the generation of topics over the graph can be found in [Waltman and Van Eck \(2012\)](#); [Klavans and Boyack \(2017\)](#); [Small et al. \(2014\)](#). We have recently started to evaluate the CHH index computed on vectors

¹ Details about the k -means algorithm are in [MacQueen et al. \(1967\)](#). We ran the standard R library k -means implementation based on the method in [Hartigan and Wong \(1979\)](#).

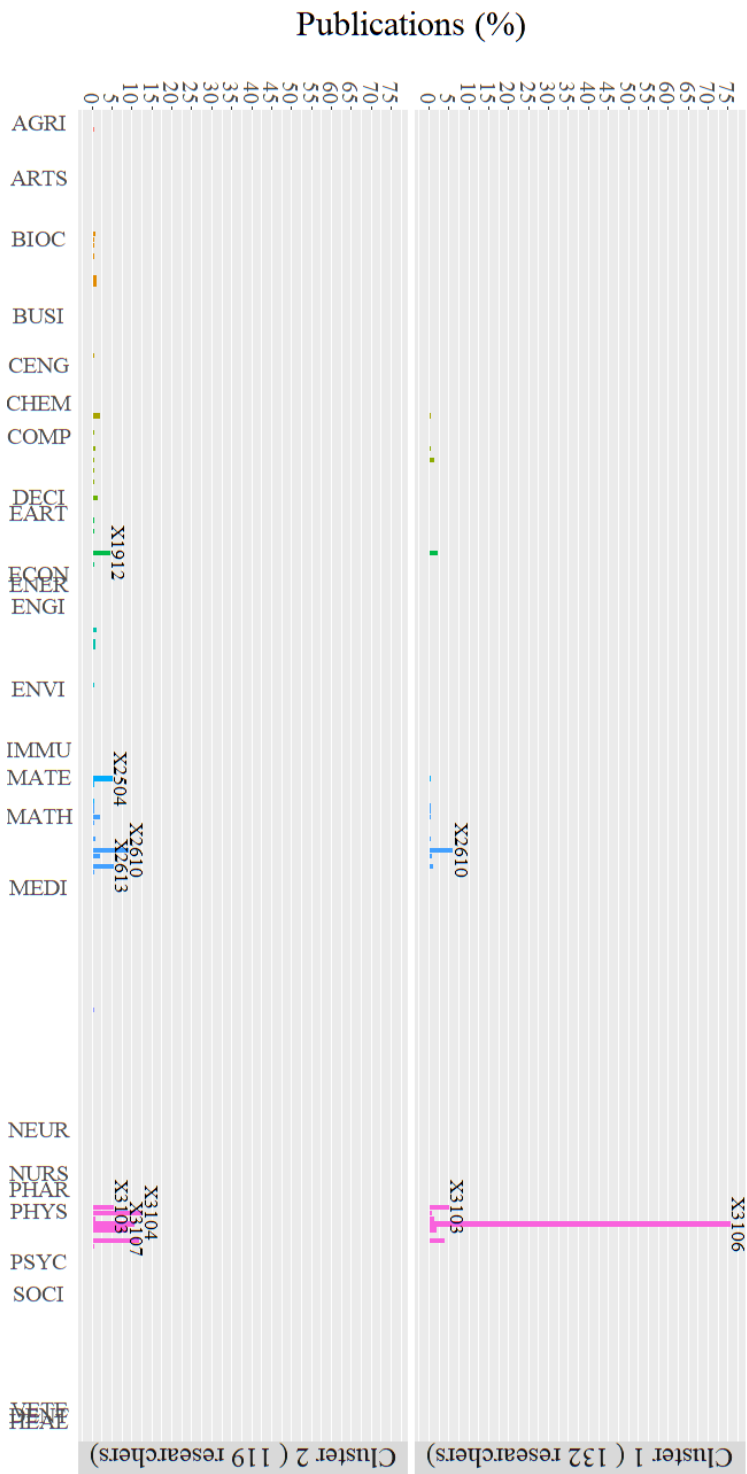


Fig. 3: Distribution of the publications by discipline for FIS/02 researchers grouped by *k*-means in two clusters. In Elsevier Scopus, SAs are coded with identifiers starting with a 'X' character (full list on the Scopus website, e.g. at https://service.elsevier.com/app/answers/detail/a_id/15181/): the grouping of SAs in the 26 macro-categories is shown on the x-axis. Upper panel: distribution for 132 FIS/02 researchers, impacting relevantly on one main subject area (X3106). Lower panel: distribution for 119 FIS/02 researchers, impacting on many subject areas (X3104, X3107, X2610, ...), hence showing more multidisciplinary than the distribution of the upper panel.

representing Topic Clusters. Some preliminary analysis conducted on Topic Clusters seems promising, in terms of distribution of index values and better representing the actual scientific production of a researcher. Future extensions of this work will investigate deeper Topic Clusters, analysing and comparing also other interdisciplinarity indices.

6.2 Conclusions

In literature, the CHH index has been often used to measure multidisciplinary, although, to the best of our knowledge, not on Scopus subject areas (SAs). In our paper, first we analysed its performance on data related to IIT and INFN researchers. The two sets were chosen so that the former would potentially show a more multidisciplinary scientific production, while the latter less multidisciplinary. We considered three cases of subject areas (SAs) from Scopus: the SAs of the documents written by an author, those of the bibliographical references of an authors documents and those of the citing documents of an authors documents. The two sets are differentiated to a good extent by the CHH index. Using the SAs of the bibliography of the publications of a researcher seems the best way forward: the distinction between the two groups of researchers is higher than the one obtained using the SAs of the researchers articles. Moreover, the larger number of bibliographical references would make the method suitable for researchers with a low number of publications and few citations.

The analysis was extended to a large sample of 3,317 Italian researchers, working in seven scientific fields to analyse the range of the multidisciplinary index inherent to a specific area. The scientific fields are defined through a categorization of the Italian academic system, in which research areas translate into officially pre-assigned Scientific Disciplinary Sectors (Settore Scientifico Disciplinare, SSD). In two SSDs, the CHH index values span a broad range of values: we noticed that the researchers are grouped in the same SSD, in spite of showing a scientific production quite different in terms of subject areas. The use of the CHH index could help in defining a proper comparison set when evaluating researchers, embracing also researchers from institutions for which the official disciplinary classification is not available. The evaluations we have carried out in this paper could be further extended to check the robustness of the whole disciplinary classification of the Italian academic system.

This work is an extended version of the paper by [Moschini et al. \(2019\)](#) presented at the 17th International Conference on Scientometrics and Informetrics, 2-5 September 2019, Rome.

References

- Abramo, G., D'Angelo, C. A., and Di Costa, F. Identifying interdisciplinarity through the disciplinary classification of coauthors of scientific publications. *Journal of the American Society for Information Science and Technology*, 63(11):2206–2222, 2012.
- Abramo, G., D'Angelo, C. A., and Zhang, L. A comparison of two approaches for measuring interdisciplinary research output: The disciplinary diversity of authors vs the disciplinary diversity of the reference list. *Journal of Informetrics*, 12(4):1182–1193, 2018.

- Boyack, K. W. and Klavans, R. Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12):2389–2404, 2010.
- Choi, B. C. and Pak, A. W. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clinical and investigative medicine*, 29(6):351–364, 2006.
- Elsevier. Elsevier Developers. <https://dev.elsevier.com/>, 2019. Accessed: 2019-11-11.
- Hartigan, J. A. and Wong, M. A. Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.
- Klavans, R. and Boyack, K. W. Research portfolio analysis and Topic prominence. *Journal of Informetrics*, 11(4):1158–1174, 2017.
- Leydesdorff, L. Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, 58(9):1303–1319, 2007.
- Leydesdorff, L. and Rafols, I. Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics*, 5(1):87–100, 2011.
- Leydesdorff, L., Rafols, I., and Chen, C. Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal–journal citations. *Journal of the American society for Information science and Technology*, 64(12):2573–2586, 2013.
- MacQueen, J. et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.
- Moschini, U., Fenioldi, E., Daraio, C., Ruocco, G., and Molinari, E. A comparison of three individual multidisciplinary indices based on the diversity of the Scopus subject areas, of the bibliography and of the citing papers of an author's documents. In *Proceedings of International Conference on Scientometrics & Informetrics (ISSI 2019)*, volume 1, pages 1226–1231, 2019.
- Porter, A. and Rafols, I. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81(3):719–745, 2009.
- Porter, A., Cohen, A., David Roessner, J., and Perreault, M. Measuring researcher interdisciplinarity. *Scientometrics*, 72(1):117–147, 2007.
- Rafols, I. and Meyer, M. Diversity measures and network centralities as indicators of interdisciplinarity: case studies in bionanoscience. In *Proceedings of International Conference on Scientometrics & Informetrics (ISSI)*, volume 2, pages 631–637, 2007.
- Rao, C. R. Diversity: its measurement, decomposition apportionment and analysis. *The Indian Journal of Statistics, Series A (1961-2002)*, 1(44):1–22, 1982.
- Rhoades, S. A. The Herfindahl–Hirschman index. *Federal Reserve Bulletin*, 79:188–189, 1993.
- Rousseau, R. The repeat rate: from Hirschman to Stirling. *Scientometrics*, 116(1):645–653, 2018.
- Rousseau, R., Zhang, L., and Hu, X. Knowledge integration: Its meaning and measurement. In *Springer handbook of science and technology indicators*, pages 69–94. Springer, 2019.
- Simpson, E. H. Measurement of diversity. *Nature*, 163(4148):688, 1949.
- Small, H., Boyack, K. W., and Klavans, R. Identifying emerging topics in science and technology. *Research Policy*, 43(8):1450–1467, 2014.
- Stirling, A. A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15):707–719, 2007.
- Stokols, D., Fuqua, J., Gress, J., Harvey, R., Phillips, K., Baezconde-Garbanati, L., Unger, J., Palmer, P., Clark, M. A., Colby, S. M., et al. Evaluating transdisciplinary science. *Nicotine & Tobacco Research*, 5(Suppl.1):S21–S39, 2003.
- Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols, I., and Börner, K. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of informetrics*, 5(1):14–26, 2011.
- Waltman, L. and Van Eck, N. J. A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, 63(12):2378–2392, 2012.
- Wang, J., Thijs, B., and Glänzel, W. Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. *PLoS one*, 10(5), 2015.