

On bibliometrics in academic promotions: a case study in computer science and engineering in Italy*

Camil Demetrescu · Irene Finocchi · Andrea Ribichini ·
Marco Schaerf

Abstract Due to its quantitative nature, bibliometrics is becoming increasingly popular among policy makers for academic hiring and career promotions. In this article, we quantitatively assess the impact that the granularity level in the classification of scientific areas would entail on research evaluation based on bibliometric indicators. We use as a case study the Italian national habilitation system (ASN), which classifies faculty members according to their academic discipline and relies on journal counts, citations, and h-indices as a basis for promoting tenure track researchers to associate professors and associate to full professors. The assessment checks whether the individual indicators of a researcher are above a certain threshold, e.g., the median over the population of researchers working in the same discipline.

Our investigation focuses on two related, rather broad disciplines: computer science and computer engineering. We show that the ASN practice of using the same thresholds for all members of a scientific discipline can favor certain sub-communities that are characterized by higher bibliometric indicators, and disfavor others. We report evidence that up to 30% of Italian faculty members of certain sub-communities would see their indicators drop below the threshold, thus becoming not eligible for promotion, if the ASN were conducted on a more accurate, fine-grained classification. Conversely, in the same scenario, up to 11% of faculty members, in different sub-communities, would see their indicators rise above the threshold, granting them eligibility. Our data set includes 1,685 authors, 89,185 distinct publications, and 262,286 author-publication pairs.

Keywords Bibliometrics, academic recruitment, research productivity, citations, h-index

1 Introduction

In many countries, academic recruitment and promotions are based on some form of habilitation. The current Italian system (ASN), established by Law 240 in 2010, is based on a qualitative evaluation of the candidate's curriculum vitae and, in the case of STEM disciplines, on bibliometrics. Key indicators are the number of published journal articles, the number of received citations, and the h-index, each over some

* Work partially supported by MIUR, the Italian Ministry of Education, University and Research, under PRIN Project n. 20174LF3T8 AHeAD (Efficient Algorithms for HArnessing Networked Data).

Camil Demetrescu
Department of Computer, Control, and Management Engineering "Antonio Ruberti", Sapienza University of Rome, Via Ariosto 25, 00185 Rome, Italy.
E-mail: demetres@diag.uniroma1.it

Irene Finocchi
Department of Statistical Sciences, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy.
E-mail: irene.finocchi@uniroma1.it

Andrea Ribichini
Department of Computer, Control, and Management Engineering "Antonio Ruberti", Sapienza University of Rome, Via Ariosto 25, 00185 Rome, Italy.
E-mail: ribichini@diag.uniroma1.it

Marco Schaerf
Department of Computer, Control, and Management Engineering "Antonio Ruberti", Sapienza University of Rome, Via Ariosto 25, 00185 Rome, Italy
E-mail: marco.schaerf@uniroma1.it

given period of time depending on the promotion level. In general, habilitation is not granted unless at least two out of three indicators are above certain thresholds specific for each recruitment field, which has its own dedicated evaluation committee. For instance, an associate professor who applies to full professorship in organic chemistry must have at least 13 journal publications in the last 5 years, at least 329 citations received in the last 10 years, and at least an h-index of 11 computed over manuscripts published in the last 10 years¹. In the case of industrial chemistry, the thresholds rise to 15, 400, and 12, respectively. The candidate's indicators to be compared with the thresholds must be measured on either Scopus or Web of Science, whichever is more convenient for them.

In the first round of the ASN, which was conducted in 2012-2013, the thresholds were computed as medians over the population of associate professors for tenure-track candidates, and medians over the population of full professors for associate professor candidates. In later rounds (2016-2018 and 2018-2020) the medians were replaced by threshold values computed by an undisclosed algorithm defined by the Italian Ministry of Education, University, and Research.

Currently, ASN indicators are used in Italy not only for hiring and promotions but also as a requirement for being part of hiring committees and for the evaluation of doctoral programs.

Finally, we point out that, akin to other habilitation systems (see, e.g., Marini [2018] for an overview of the Spanish case), research output is not the only evaluation parameter, grants, teaching activities and third mission (i.e., knowledge transfer, academic spin-offs) being also taken into account by ASN. We maintain, however, that research productivity remains by far the predominant factor in the evaluation.

Contributions of the article. In this article, we explore whether computing thresholds over the entire academic population of individual recruitment fields as done by the ASN might introduce a bias in academic promotions. In other words, we try to shed light on the following question: *to what extent does the level of granularity offered by Italian recruitment fields provide fair bibliometric indicator thresholds for habilitation?* As a case study, we consider different research areas within two closely-related STEM disciplines, computer science and computer engineering, and characterize the bibliometric indicators of professors working in these areas.

As a result, we show that certain sub-communities benefit from the fact that thresholds are computed at the recruitment field level, and not at a finer-grained level that considers prominent sub-areas of computing, while others are hindered. Our data show that candidates working, e.g., in theoretical computer science would be more likely to be eligible for habilitation if their bibliometric indicators were compared with those of their peers working in the same area, rather than in the whole recruitment field. Conversely, an author working, e.g., in multimedia can afford to have smaller indicators than their peers working in the same area and still be eligible for habilitation in the current system. We argue that the "right" level of granularity should be carefully established to avoid bias that favors certain sub-communities and disfavors others. While this is beyond the scope of this article, we believe that the key would be to ensure that the considered areas are bibliometrically balanced.

Due to its use in the ASN and its high coverage of computer science and engineering publications, our work is based prominently on Scopus, integrated with a variety of other sources. We have considered 1,685 academic authors with an Italian affiliation, 89,185 distinct publications, and 262,286 author-publication pairs. Due to the lack of information on how thresholds are currently computed in the ASN, we base our work on medians, mimicking the approach used in the first ASN round held in 2012-2013. We believe that our findings would lead to similar conclusions by uniformly adopting other threshold definitions (e.g., other percentiles).

Structure of the article. The remainder of this article is organized as follows. Section 2 discusses related work. Section 3 presents our data sets and the way they were retrieved and pre-processed. In Section 4 we describe our methodology, discussing the main issues we had to face in setting up the experiments. The main findings of our investigation are presented in Section 5, while Section 6 draws some general conclusions.

2 Related work

This section reviews the scientific literature most relevant to our work. Section 2.1 targets research assessment at the level of individual researchers, while Section 2.2 focuses specifically on the recruitment process in Italian universities. A discussion of the validity of the habilitation procedure adopted in Italy is outside the scope of this paper. However, we summarize here the relevant literature.

¹ https://abilitazione.miur.it/public/documenti/2018/Tabelle_Valori_Soglia_ALLEGATI_DM_589_2018.pdf

2.1 Research assessment at the individual level

Becher [1989] is considered foundational in regarding universities as made up of different disciplinary communities, each with its common set of practices, a concept that informs most of the subsequent literature on research assessment. A second edition (Becher and Trowler [2001]), and a third book in the series (Trowler et al. [2012]) have progressively acknowledged the evolution of academic disciplines into more complex and dynamic entities, characterized by internal hierarchies, shifting boundaries and an ever increasing number of subdisciplines (see Trowler [2014], and references therein, for an overarching retrospective on the three books, and the intervening changes in higher education).

Evaluation of the scientific output of individual researchers has been the focus of much attention in recent years, serving as a guide both for fund allocation and for granting promotions and tenure (see, e.g., Bornmann et al. [2008b], Demetrescu et al. [2019], Todeschini and Baccini [2016], and the references therein).

Since its initial definition by Hirsch [2005], the h-index has been a popular tool for research evaluation, combining both qualitative and quantitative factors in a single easy-to-compute number: an author has h-index h if he/she has h publications with at least h citations each and each of the remaining publications has at most h citations. Kelly and Jennions [2006], while recognizing that high h-index values single out highly influential scientists, argue that modest differences are not so significant. Moreover, they point out that biases in h-index values exist, at least for ecologists and evolutionary biologists, according to geographic location, sub-field of study, and gender (female researchers tend to publish fewer papers, while not being less established as scholars).

Kreiman and Maunsell [2011], while not proposing any specific metric, list nine criteria that a good index of scientific output should satisfy. Such an index should be “quantitative, based on robust data, rapidly updated and retrospective, presented with confidence intervals, normalized by the number of contributors, career stage and discipline, impractical to manipulate, and focused on quality over quantity”.

Abramo and D’Angelo [2014] propose the Fractional Scientific Strength (FSS), a productivity indicator that can be applied at all academic levels, from individual researchers to whole universities. FSS involves a normalization relative to both input resources (e.g., salary) and average citations for any specific research field. Furthermore, Abramo and D’Angelo [2016] criticize the Mean Normalized Citation Score (Waltman et al. [2011]), which measures the average number of citations, normalized by scientific field, of the publications of an individual or an institution, and similar size-independent indicators, for not being true performance measures. They propose to adopt new indicators which, like the FSS, take into account the amount of input resources.

Empirical results on the h-index and its most important variants are presented by Bornmann et al. [2008a]. Garner et al. [2017] also review several bibliometric indices, designed to overcome some of the shortcomings of the h-index. These include the g-index (Egghe [2006]), intended to give more weight to highly cited papers, the e-index (Zhang [2009]), meant to take into account the total number of citations of the papers contributing to the h-index, and the contemporary h-index (Sidiropoulos et al. [2007]), which gives more relevance to recent papers. The intrinsic limitations of each examined index are pointed out, and it is emphasized that each may assume widely different values depending on the data source used for its computation (i.e., Scopus, WoS, or Google Scholar).

Braithwaite et al. [2019] present a systematic review of the literature on research assessment, highlighting pros and cons of both citation-based metrics and alternative approaches such as online article views, downloads, bookmarks, PageRank algorithms, and attention by social media. They propose a comprehensive evaluation framework that integrates citation-based and non-citation-based methods, while at the same time remarking that, in spite of its limitations, the role of expert judgment should not be discounted, and that metrics should be adopted to complement the peer-review process rather than to supplant it.

2.2 Analysis of the Italian academic recruitment process

An in-depth analysis of Italy’s last recruitment campaign (2008), before the introduction of the current habilitation system, is presented in Abramo et al. [2014], comparing the academic performance of competition winners, over the years 2009-2011, against that of the losers and the pre-existing staff of equal academic rank. It is shown that newly hired professors generally performed better than pre-existing staff, but non-winner candidates tend to be even more productive over the subsequent triennium. The determinants of the results emerging from this study were statistically analyzed in Abramo et al. [2015], suggesting that a candidate’s success seems strongly correlated with the number of years of service in the same university as

the president of the selection committee, as well as with joint research. The authors themselves, however, advise caution in interpreting these results, pointing out that their analysis does not take into account dimensions of a candidate’s work such as teaching skills and ability to attract funds.

Italy’s national habilitation system was presented in 2010, with the official aim of introducing objective bibliometric indicators into the academic recruiting process. Since its inception, the system has been heavily scrutinized by several authors.

Abramo and D’Angelo [2015] suggest that candidates already hired, either at the assistant or associate professor level, obtained an habilitation more easily than candidates external to academia. The same is true for researchers already belonging to the academic discipline for which they are requesting the habilitation, compared to applicants belonging to different disciplines. Possible discriminations seem more frequent than possible favoritisms, and especially so for habilitation to full professorship.

Marzolla [2015] presents a thorough quantitative analysis, highlighting several points of interest. While over-median candidates were on average more likely to obtain the qualification, it is also true that in half of the scientific disciplines more than half of the applicants that satisfied the quantitative requirements did not eventually obtain the habilitation. Moreover, there are disciplines where the thresholds for associate professor are higher than those for full professor, making it potentially harder to pass the test for the lower academic rank. For the so-called “non-bibliometric” disciplines (for which only publication counting metrics are used), medians equal to zero are not rare: this implies that some indicators are zero for at least half of tenured professors, making them not very useful for assessing the scientific profile of applicants.

Pautasso [2015] reports on the number of applications from male and female scholars to the first ASN campaign, and their respective success rates. Applications from female scholars turned out to be 36% of the total, higher than the overall percentage of female professors (27%), but lower than the proportion of female researchers (45% with a permanent position and 51% with a temporary one). Success rates of male and female scholars are found to be comparable. Furthermore, no gender bias emerges from the composition of the judging committees. It is observed that the quantitative nature of the indicators employed by the habilitation procedure may have contributed to these results. [We remark, however, that, in spite of the general fairness of the ASN in this respect, gender discrimination still persists in academic careers in Italy, at least when it comes to actual promotions \(not just habilitations\) to full professorship, as reported in Marini and Meschitti \[2018\].](#)

In Marzolla [2016] it is investigated whether the ASN complies with the ten best practices for the use of bibliometric indicators in the evaluation of researchers and organizations, as they appear in Hicks et al. [2015], showing that the ASN conforms to only half of them. It is argued that the quantitative indicators used in the ASN are not differentiated enough to cope with the variability of practices and goals across diverse fields of study. Moreover, it may be not possible for researchers without access to WoS or Scopus to independently verify data and analyses, whose values may be anyway updated without notice. Not last, the set of publications used to compute the medians has not been made public, making it impossible to verify that the thresholds are correct. The author has also analyzed a sample of the reports concerning candidates who were denied habilitation, confirming that the perceived low quality of some of them was indeed justified.

Marini [2017] focuses on results concerning habilitation to full professorship in two bibliometric areas (physics and engineering) and two non-bibliometric ones (law and economics). He reports that the best predictors towards obtaining an habilitation are quality of the scientific output, current position in university ranks, and younger age. This suggests that traditional career-advancement patterns, based on seniority, may be losing ground in favor of quicker careers for the more productive researchers.

3 Bibliometric datasets

In this section, we describe the datasets used in our experimental study. In addition to relevant statistics, we provide details on data sources as well as on data cleaning and pre-processing tasks. We distinguish between four different data types related to faculty, publications, research areas, and publication venues. All datasets have been downloaded in August 2019.

For brevity, throughout the article, we will use the acronyms CS (Computer Science) and CE (Computer Engineering) to refer to the two recruitment fields addressed in our study.

	RE (Researcher)	AP (Associate Professor)	FP (Full Professor)	Total
CS (Computer Science)	355	342	215	912
CE (Computer Engineering)	262	304	207	773
Total	617	646	422	1685

Table 1 CS and CE faculty members in Italy, classified by academic rank. Only professors with at least one Scopus profile are counted.

	RE (Researcher)	AP (Associate Professor)	FP (Full Professor)	Total
CS (Computer Science)	95 (27%)	87 (25%)	43 (20%)	225 (25%)
CE (Computer Engineering)	35 (13%)	51 (17%)	27 (13%)	113 (15%)
Total	130 (21%)	138 (21%)	70 (17%)	338 (20%)

Table 2 CS and CE female faculty members in Italy, classified by academic rank. Only professors with at least one Scopus profile are counted.

		Conferences				Journals				Others				All types			
		5 years	10 years	15 years	All	5 years	10 years	15 years	All	5 years	10 years	15 years	All	5 years	10 years	15 years	All
CS	FP	3769	7415	10806	21990	2527	4304	6061	12892	612	1103	1559	3274	6908	12822	18426	38156
	AP	5758	10816	14531	31105	3644	5896	7705	17245	831	1381	1662	3874	10233	18093	23898	52224
	RE	5115	8336	10263	23714	2862	4111	4851	11824	570	881	1005	2456	8547	13328	16119	37994
	All	14642	26567	35600	76809	9033	14311	18617	41961	2013	3365	4226	9604	25688	44243	58443	128374
CE	FP	5088	10181	14847	30116	2835	4774	6465	14074	718	1419	1922	4059	8641	16374	23234	48249
	AP	6403	12079	16366	34848	3538	5654	7092	16284	956	1678	2040	4674	10897	19411	25498	55806
	RE	3959	6734	8147	18840	2327	3238	3613	9178	417	670	752	1839	6703	10642	12512	29857
	All	15450	28994	39360	83804	8700	13666	17170	39536	2091	3767	4714	10572	26241	46427	61244	133912
All	FP	8857	17596	25653	52106	5362	9078	12526	26966	1330	2522	3481	7333	15549	29196	41660	86405
	AP	12161	22895	30897	65953	7182	11550	14797	33529	1787	3059	3702	8548	21130	37504	49396	108030
	RE	9074	15070	18410	42554	5189	7349	8464	21002	987	1551	1757	4295	15250	23970	28631	67851
	All	30092	55561	74960	160613	17733	27977	35787	81497	4104	7132	8940	20176	51929	90670	119687	262286

Table 3 Statistics on Scopus publication records with at least one coauthor with an Italian affiliation according to recruitment field (CS, CE), academic rank (FP, AP, RE), publication type (conference, journal, others), and publication date (last 5 years, last 10 years, last 15 years, any date).

3.1 Faculty members

Using an online service provided by the Italian Ministry of Education, University and Research (available at <https://cercauniversita.cineca.it/php5/docenti/cerca.php>), we have obtained the official list of the current 1,698 CS and CE professors, partitioned into full professors (FP), associate professors (AP), and researchers at the assistant professor level (RE). In the case of authors with the same first and last names, we have resorted to Scopus profiles using affiliation and research areas as disambiguation criteria. Even after manual inspection, we were unable to find the Scopus profile of 13 professors, that we have thus discarded without further consideration. In a few cases, we have instead identified in Scopus multiple profiles for the same author, merging them. In the end, we have obtained a list of 1,685 professors, associated with possibly multiple Scopus profiles, that have been used in our experimental analysis. The detailed statistics by academic rank and academic discipline are given in Table 1. CS is slightly larger than CE in terms of the number of Italian authors but overall the sizes of the two communities appear to be similar. **Female faculty members are significantly under-represented, as shown in Table 2: this is true for all the academic ranks, with the smallest percentage among full professors. This confirms the findings in Marini and Meschitti [2018]. There is a non-negligible difference between CS and CE.**

3.2 Publication metadata

Publication metadata have been retrieved from Scopus, considering the entire scientific production of each CS/CE professor associated with at least one Scopus profile. For each publication, we have extracted several Scopus fields including publication title, year, venue, type (journal, conference, etc.), number of citations, and Italian CS/CE faculty coauthors. Statistics on the numbers of Scopus publication records are provided in Table 3, which distinguishes between different time periods (last 5, 10, or 15 years) and different academic ranks, as required to compute indicators used in the Italian national habilitation system. In total, we have processed 89,185 unique scientific publications. As the same publication may be co-authored by different Italian faculty members, overall we have considered 262,286 (author, publication) pairs.

Area	Description
AI	Artificial Intelligence; Genetic and Evolutionary Computation
APP	Applied Computing; Bioinformatics, Computational Biology; Computers and Society; Economics and Computation
ARCH	Computer Architecture; Embedded Systems; High-Performance Computing; Microarchitecture; Design Automation (Signal processing)
BROAD	Multidisciplinary
CYBER	Security, Audit, and Control
DATA	Information Retrieval; Knowledge Discovery and Data Mining; Management Information Systems; Management of Data; Spatial Information; Hypertext, Hypermedia, and Web
EDU	Computer Science Education; Information Technology Education
MULTIM	Computer-Human Interaction; Computer Graphics; Multimedia Systems; Accessibility and Computing
NET	Data Communication; Design of Communication; Mobility of Systems, Users, Data & Comp
SOFT	Programming Languages; Software Engineering
SYSTEM	Measurement and Evaluation; Operating Systems; Simulation (Parallel and Distributed)
THEORY	Algorithms & Computation Theory; Logic and Computation; Symbolic & Algebraic Manipulation

Table 4 ACM SIGs grouped into 12 research areas.

Figure 1 shows the distributions of three different habilitation indicators for both CS and CE authors, which are sorted on the x -axis by decreasing order of the observed value. Each chart addresses a different indicator: number of journals published in the last 10 years (top chart), number of citations to articles published in the last 15 years (middle chart), and h-index based on publications of the last 15 years (bottom chart). The curve trend and skewness for the two academic disciplines are in line with each other, except for a few outliers at the left tail of the distribution. Most notably, there is a CE associate professor with 14,478 citations, and another one with h-index 47: these values result in large maxima in citations and h-index graphs. We also note that the slightly larger size of the CS community results in a longer right tail that gracefully scales down to 0.

We have integrated Scopus data with publication records extracted from the DBLP Computer Science Bibliography (see <https://dblp.uni-trier.de/>), which provides open bibliographic information on major computer science journals and proceedings and currently indexes 4,853,582 publications. DBLP records allowed us to obtain complete information about publications that appeared on series such as Lecture Notes in Computer Science or CEUR Workshop Proceedings, for which the available Scopus APIs generally provide only the series name, without full details on the specific publication venue. We then used record linkage techniques, whose description is beyond the scope of this article, to match Scopus and DBLP records, and we merged information extracted from the two data sources for each publication.

3.3 Research areas

As a rationale to distinguish between different research areas within the computer science and computer engineering fields, we have exploited the Special Interest Groups of the Association for Computing Machinery (ACM SIGs). SIGs represent major areas of computing and manage conferences, publications, and activities focused on specific computing sub-disciplines. There are currently 37 SIGs, that cover a broad spectrum of topics including artificial intelligence, software engineering, programming languages, graphics and interactive techniques, data management, networks and mobile communications, education, and theory. The full list is available at <https://www.acm.org/special-interest-groups/sigs-by-knowledge-area>. We have extended this list with two additional topics (Signal processing and Parallel and distributed computing) that appear to be relevant according to other technical organizations, including the IEEE Computer Society, but are not well represented in the current SIGs list.

Using the information at the first level of the 2012 version of the ACM Computing Classification System (see <https://www.acm.org/publications/class-2012>), we have eventually grouped the 39 extended SIGs into 12 broader research areas, which are shown in Table 4.

3.4 Publication venues

In order to be able to associate each researcher with one (or more) research area, we concentrated on the publication venues and assigned to each publication venue a set of related research areas. In Computer Science, conferences are as important as journals (or maybe more) as venues for publishing high quality research. For this reason, the research assessment criteria used in the Italian national habilitation system

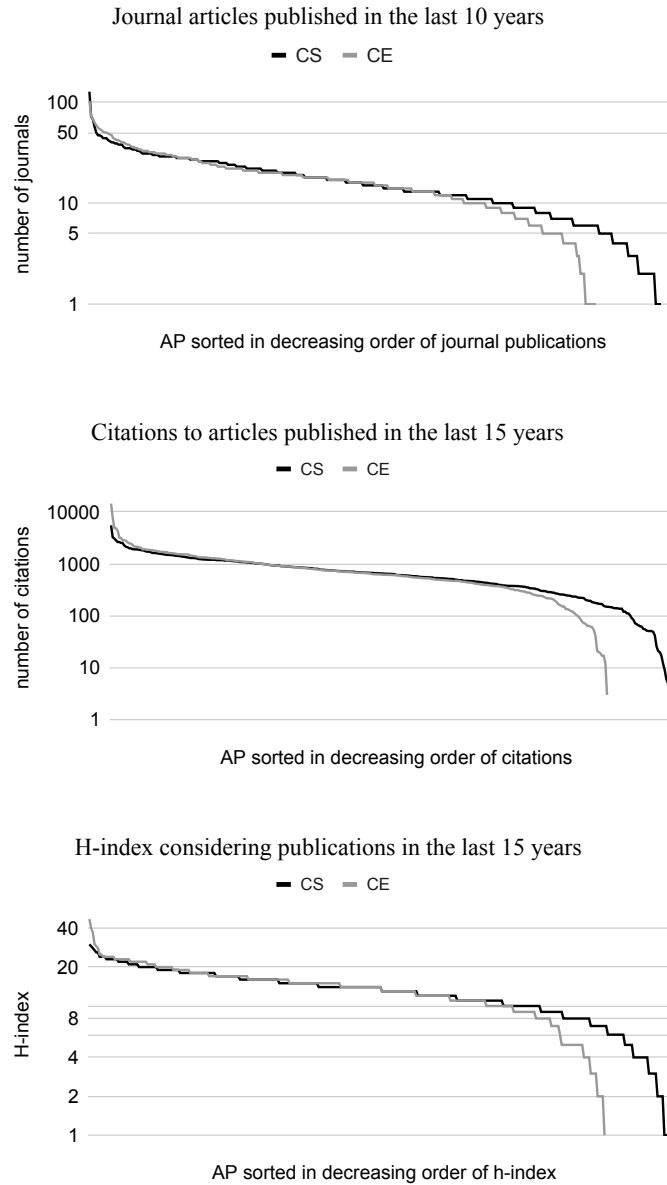


Fig. 1 Distribution of different habilitation indicators (journals, citations, h-index) for CS and CE faculty members.

for CS and CE take both venue types into consideration, yet requiring to distinguish between conference and journal publications. In order to decide which publication venues we should include, we first considered using the Scopus category 1700 (Computer Science). However, this includes far too many publication venues and uses a very large variety of indexing methods, even for publications that appeared on related venues (e.g., different editions of the same conference). Therefore, we resorted to prominent lists of conferences and journals, that we later linked to Scopus data entries:

- Concerning conferences, we have used the 2018 GII-GRIN-SCIE conference list, developed as a joint initiative sponsored by GII (Group of Italian Professors of Computer Engineering), GRIN (Group of Italian Professors of Computer Science), and SCIE (Spanish Computer Science Society). The list, available at <http://www.consortio-cini.it/gii-grin-scie-rating.html>, provides a unified rating of computer science conferences according to three classes: class 1 includes 82 top-notch conferences, class 2 includes 175 conferences of very good quality and class 3 consists of 366 conferences of good quality. This results in 623 rated conferences. Additionally, there are 2,172 conferences whose evaluation is still work-in-progress.

- Concerning journals, we have used a list released in 2012 by the Groups of CS and CE Experts in charge of the Italian National Evaluation Exercise (VQR 2004–2010). The list includes 3265 journals, rated according to four different classes.

Similarly to DBLP data, we have exploited a record linkage algorithm to match Scopus records with conferences and journals in the two aforementioned lists. The goal of linkage operations between different data sources will be addressed in Section 4.

We remark that, by exploiting sponsorship information provided on the Web pages of the ACM SIGs, many major conferences and journals can be naturally assigned with one or more SIGs of interest. For instance, the ASPLOS Conference on Architectural Support for Programming Languages and Operating Systems is sponsored both by SIGARCH, the Special Interest Group on Computer Architecture, and by SIGPLAN, the Special Interest Group on Programming Languages.

4 Approach

In this section, we discuss two main issues that we had to face in producing the experimental data and the way we addressed them: (1) automatically associating CS/CE professors with their (main) research area(s), and (2) computing area thresholds to be compared with values obtained over the population of the entire recruitment field. In order to automatically find the main research areas of a CS/CE professor, we have exploited a two-step approach, first tagging a large subset of the author’s Scopus publication records with one or more research areas, and then extracting the predominant areas among those in which the author has at least one publication.

4.1 Tagging conferences and journals with their research areas

Before tagging individual publications in the Scopus dataset, we have focused on associating research areas to conference and journal venues in the lists described in Section 3. The whole area classification process has been further divided into three steps:

- We first associated each of the 257 class 1 and class 2 conferences in the GII-GRIN-SCIE list with one or more of the 39 extended SIGs. While this has been done manually, a large number of such high-quality conferences appear on the Web pages of the relevant ACM SIGs and can be then classified accordingly. In a limited number of cases, we had to inspect the call for papers of an event to find its relevant SIGs. Using the manual classification as our ground truth, we have then automatically propagated the classification to a broader set of conferences and journals, exploiting keywords and a variety of string similarity measures. In the end, we have classified all the class 1, 2, and 3 conferences, a subset of the work-in-progress conferences from the GII-GRIN-SCIE list most popular among authors, and a subset of journals from the VQR list including 88% of class 1 journals and 63% of class 2 journals.
- After SIGGROUP tagging, we have extracted the research area of each tagged conference and journal using the SIGs-area mapping described in Table 4. Notice that publications labeled with more than one SIG might be associated with multiple research areas: for instance, the ASPLOS conference that we have taken as an example in Section 3 falls within both the ARCH and the SOFT areas. In other cases, when multiple SIGs collapse to the same area, the area classification becomes instead unique: as an example, the IEEE Wireless Communications and Networking Conference is tagged as SIGCOMM and SIGMOBILE, which both fall into the NET research area.
- Through a record linkage mechanism, we have finally associated the tagged conference and journal titles to the specific records extracted from the authors’ profiles in Scopus. In some cases – and especially for conferences – this is a non-trivial task, due to the large variety of source titles available in Scopus records: for instance, different editions of the same conference can have rather different names and even different IDs in Scopus records but should be all associated to the same conference name. Exploiting information obtained from DBLP allowed us to increase significantly the success rate of this linkage task.

At the end of the process, we could tag on average over 50% of the publications for all roles in both CS and CE. We also obtained, for each author, a 12-dimensional *frequency vector* containing a counter, for each research area, of the number of publications of that author in the area.

Academic discipline	CE						CS					
Role	FP		AP		RE		FP		AP		RE	
Faculty members	207		304		262		215		342		355	
w/main area(s)	206	99.5%	303	99.7%	259	98.8%	215	100%	340	99.4%	350	98.6%
w/unique main area	121	58.5%	239	78.6%	245	93.5%	126	58.6%	253	74.0%	319	90.0%

Table 5 Numbers of faculty members with one or more main areas of interest.

4.2 Defining the research areas of authors

We computed the authors’ main areas from frequency vectors, extracting areas whose counters are at least 50% of the maximum in the frequency vector, provided that such counters are larger than the 25% of the total number of classified author’s publications (main area threshold). We empirically determined the 25% and 50% thresholds by choosing the most effective setup based on sampled manual inspection of the profile of faculty members. Consider the following three scenarios:

- An author with 108 classified publications whose frequency vector is:

AI	APP	ARCH	BROAD	CYBER	DATA	EDU	MULTIM	NET	SOFT	SYST	THEORY
1	0	12	0	0	0	0	0	0	15	1	79

The main area is THEORY, which has the maximum counter corresponding to 73% of 108. No other counter is larger than half of the maximum.

- An author with 91 classified publications whose frequency vector is:

AI	APP	ARCH	BROAD	CYBER	DATA	EDU	MULTIM	NET	SOFT	SYST	THEORY
14	2	1	1	4	28	0	0	32	0	2	7

The main areas are NET and DATA: NET has the maximum counter, and DATA is 87% of the maximum. Both counters are larger than 25% of 91.

- An author with 10 classified publications whose frequency vector is:

AI	APP	ARCH	BROAD	CYBER	DATA	EDU	MULTIM	NET	SOFT	SYST	THEORY
2	2	0	0	0	2	0	0	0	0	2	2

We assign no main areas: five areas have the maximum value but none of them is larger than 25% of 10. These cases are quite rare, and mostly involve young researchers with a small number of publications.

Extracting the predominant areas as described above made it possible to assign main areas to 1,673 out of 1,685 professors (i.e., over 99%). To increase our confidence in the results, we have sampled 15% of the CS and CE professors, manually checking the validity of their main research areas also with the help of area experts. A detailed breakdown is reported in Table 5. For example, we identified working areas for 206 out of 207 FP in CE, and 121 out of them (i.e., 58.5%) have a unique area. We observe that senior faculty members with a unique main area tend to be fewer than juniors with a unique main area. This is to be expected as researchers in the early stages of their careers tend to focus on specific topics. Conversely, senior researchers are more likely to diversify their scientific interests over the years. Authors with three or more areas are very rare. Differences between CS and CE are negligible.

4.3 Computing area-based bibliometric indicators

Bibliometric area thresholds have been computed for each research area following the same criteria used for the whole recruitment field in the ASN. For each author, we have computed the number of journal publications in the last 5 or 10 years, the number of citations attracted by conference papers, journal articles, book chapters, etc., published in the last 10 or 15 years, and the h-index on the same subset of publications. For each research area r , we have then considered the subsets of CS/CE professors, partitioned by academic rank, whose main area list includes r and have computed the medians of the bibliometric indicators of authors in each subset, considering the appropriate temporal period (as observed in Section 1, differentiating between the last 5, 10, and 15 years is needed when examining either FP or AP promotions). Notice that professors with multiple research areas are considered in multiple communities and that, in line with the national habilitation process, we do not distinguish between publications in different areas to compute the author’s indicators: the author’s entire scientific production is always used to compute the author’s indicators, that could be then exploited for median computation in different areas.

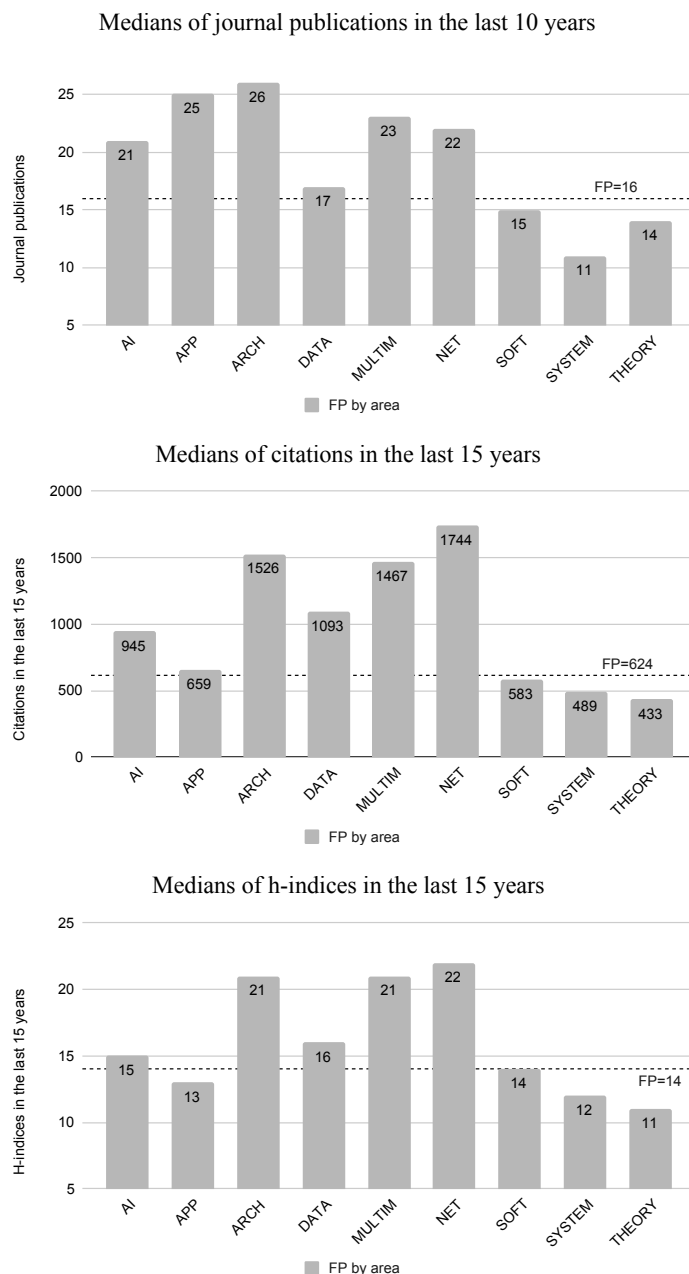


Fig. 2 CS medians of journal publications (top), citations (middle), and h-indices (bottom) of FPs per area compared to the medians of FPs across all areas (dashed lines). Figures for journal publications refer to the last 10 years, the others refer to the last 15 years as prescribed by the ASN.

5 Experimental results

In this section, we quantitatively show that, if habilitation thresholds are not computed at the proper level of granularity, sub-communities characterized by higher indicators can be favored, while others can be penalized. In the long term, this might introduce a bias even on the choice of research topics, favoring some research areas with respect to others, which would exacerbate even further the bibliometric differences between areas. We address a few main research questions:

- **RQ1:** how do medians computed on individual sub-areas of CS and CE compare to the medians computed across the overall academic disciplines?
- **RQ2:** to what extent do the results of RQ1 impact on individual researchers? How many authors would be favored by evaluating them on a finer-grained area classification? How many of them would be disfavored?

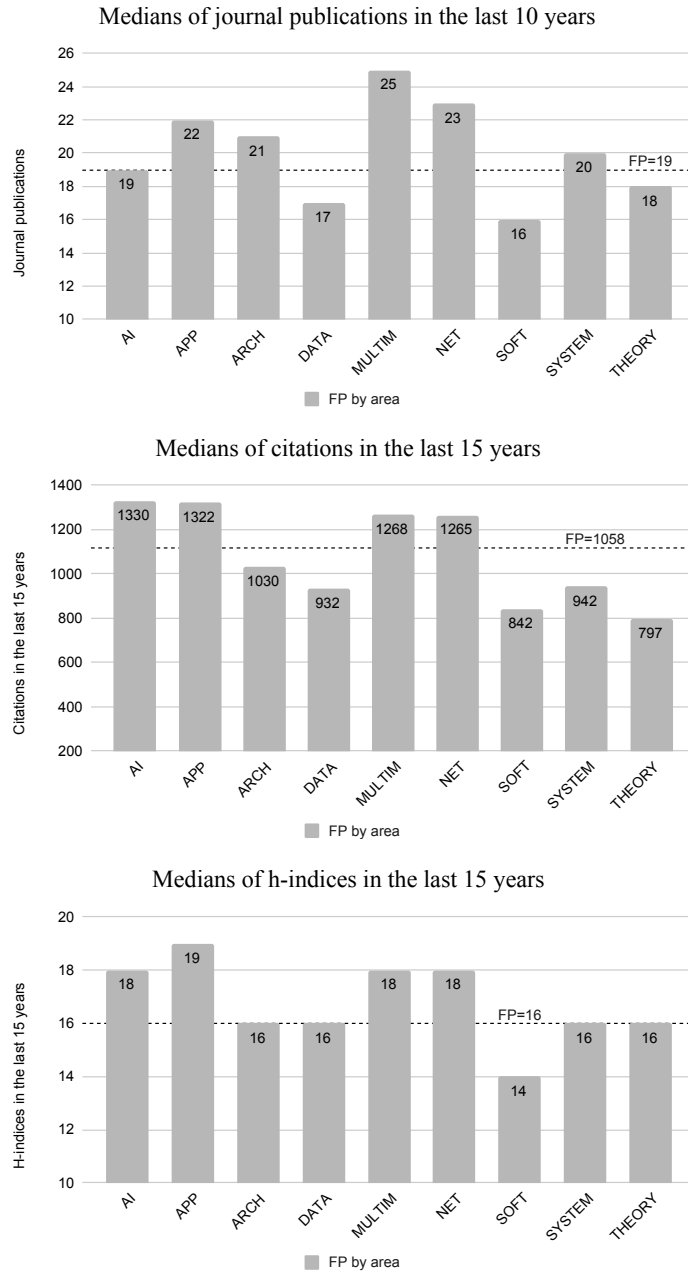


Fig. 3 CE medians of journal publications (top), citations (middle), and h-indices (bottom) of FPs per area compared to the medians of FPs across all areas (dashed lines). Figures for journal publications refer to the last 10 years, the others refer to the last 15 years as prescribed by the ASN.

– **RQ3:** what is the evolution of research areas over a time period of a few years? Do newly hired researchers choose to work in areas that are more promising in terms of ASN career promotions?

All charts shown in this section report data in the time intervals prescribed by the Italian national habilitation system for the promotion from associate to full professorship. Results related to promotions from assistant to associate professorship are similar.

5.1 RQ1: Comparison of medians

Figure 2 and Figure 3 summarize our findings, reporting histograms that show for each academic discipline (CS, CE) and each bibliometric indicator (number of journal publications, number of citations, and h-index) its *median* computed over the population of researchers classified in each of the 12 sub-areas we considered.

		AI	APP	ARCH	DATA	MULTIM	NET	SOFT	SYSTEM	THEORY
Above	CS	3	2	3	3	3	3			
	CE	2	3	1		3	3		1	
Below	CS		1					2	3	3
	CE			1	2			3	1	2
Above - Below		5	4	3	1	6	6	-5	-3	-5

Fig. 4 Number of area-based indicators above and below the CS/CE medians. The last row shows the difference between the number of indicators above threshold and below threshold for each area. Areas sorted in decreasing order by the last row values are {MULTIM, NET} AI APP ARCH DATA SYSTEM {SOFT, THEORY}.

We also report, as horizontal dashed lines, the medians of full professors computed over the entire dataset of authors, rather than on sub-areas. For instance, the median of the number of journal articles published in the last 10 years by CS full professors working in the ARCH sub-area is 26, while the median computed over all CS full professors for the same indicator is 16 (top chart of Figure 2). Hence, an area-based evaluation would disfavor an associate professor working in ARCH, substantially raising the bar as far as their number of journal publications is concerned. Conversely, a CS associate professor working in THEORY would need 433 citations in the last 15 years to be above the bar and not 624 as it would be required using the ASN approach (middle chart of Figure 2).

We note that AI, APP, ARCH, MULTIMEDIA, and NET are almost never below the FP threshold in all charts of Figure 2 and Figure 3. Hence, our data suggests that associate professors publishing in those areas are likely to be favored by the current national habilitation system. Conversely, SOFT, SYSTEM, and THEORY are almost never above the FP threshold, which implies that they may be disfavored by the current system. DATA exhibits a somehow opposite behavior in CS and CE, with DATA favored in CS and generally disfavored in CE in the current ASN system. These results are summarized in Figure 4.

5.2 RQ2: Favored and disfavored authors

Figure 5 shows the impact on individual associate professors of using a finer-grained, area-based, classification of disciplines compared to the one currently used by the Italian national habilitation system. We note that the results are in line with the conclusions we drew by analyzing Figure 2 and Figure 3. For instance, 11% of CS associate professors whose main research areas include THEORY would become above the threshold if the ASN evaluation was carried out in an area-based system. In the same scenario, 30% of CS associate professors in MULTIM would drop below the threshold and most likely would not be eligible for habilitation as full professor. In our analysis, we have considered only the areas with a population of at least 10 individuals. For instance, we have excluded EDU and CYBER as fewer than 10 associate professors in Italy in either CS or CE work primarily in these areas. [Although the absolute number of females is significantly smaller than males in CS, and even smaller in CE as shown in Table 2, we noticed no clear bias as far as favored and disfavored authors are concerned.](#)

5.3 RQ3: Evolution of the research areas

Figure 6 shows the change in the number of researchers in the considered areas in the period from the end of 2012 till August 2019. Overall, there has been an increase in the number of researchers/professors in CS and CE during this period, but the increase is not uniform across the various areas. For example, THEORY is the only area that sees a reduction in the number of active researchers, while the largest overall increase is in AI and the highest relative increase is in CYBER. Even though these growing trends reflect the increased relevance of AI and CYBER in the last years, they are also probably influenced by the perception that researchers active in areas such as THEORY are at a disadvantage in ASN evaluations.

5.4 Caveats

There are some caveats regarding our work that should be taken into account. First, we classify authors based on the subset of their publications that we could successfully link to a classified conference or journal. On average, we cover more than 50% of publications per author and the tagging algorithm guarantees that the top conferences and journals of each research area are classified. However, though unlikely, it may happen that an author has top publications in a research area r and many secondary publications in a

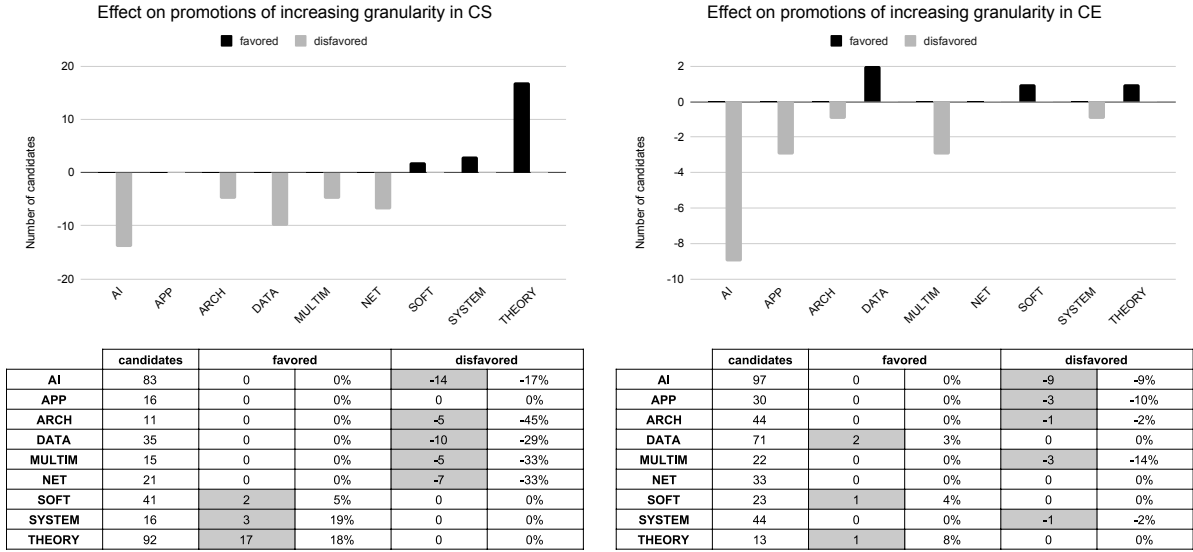


Fig. 5 Effect on promotions deriving by switching from an evaluation of Italian associate professors based on overall disciplines such as Computer Science and Computer Engineering to a finer-grained, area-based evaluation.

	Overall Total	Retired before August 2019	Hired after 2012	Total in 2012	Total in 2019	% increase
AI	368	33	86	282	335	18.79%
APP	71	6	19	52	65	25.00%
ARCH	101	7	20	81	94	16.05%
BROAD	4	0	1	3	4	33.33%
CYBER	19	1	7	12	18	50.00%
DATA	180	10	35	145	170	17.24%
EDU	9	0	1	8	9	12.50%
MISC	0	0	0	0	0	0.00%
MULTIM	88	10	23	65	78	20.00%
NET	101	11	26	75	90	20.00%
SOFT	124	12	28	96	112	16.67%
SYSTEM	111	7	26	85	104	22.35%
THEORY	259	38	23	236	221	-6.36%

Fig. 6 Change in the number of researchers in each area in the period 2012-2019 (pre-ASN and post-ASN).

different area r' , publications in r are classified, while those in r' are not. In that case, only the activities in area r could be reflected when defining the author's research area, while r' could be neglected due to missing publication tags. For this reason, our investigation does not focus on single researchers, but rather on average over a population of at least dozens of individuals. We also remark that the aforementioned scenario will not happen if an author has a good publication record in both r and r' , since representative publications in both areas would be classified and considered.

Second, errors in Scopus or in the GII-GRIN-SCIE classification may lead to inaccuracies. For instance, some conference publications are archived as journals in Scopus or do not appear in the GII-GRIN-SCIE, hence they hinder our classification. This may be correct for certain conferences whose papers are all published as articles in a companion journal (e.g., SIGGRAPH). As the evaluation, at least in Italy, is done with the Scopus data, we argue that the same errors are likely to appear in the actual national habilitation process as well.

Third, errors may be introduced by our manual classification that assigns conferences and journals to sub-areas. We checked correctness on a random sample of roughly 15% of authors, analyzing their main areas of interest derived by our classification and found no evident errors.

Finally, the most complex step in our study was linking Scopus source titles (names of publication venues) to the conferences in the GRIN-GII-SCIE list and the journals in the VQR list. For instance, there are a plethora of different cases in which the source title may appear. Two common patterns are "Proceedings of X" vs. "X, proceedings of", but differences may be far more subtle. For instance, the same conference may have changed name along the way, e.g., being "promoted" from workshop to symposium or congress, merged with another conference, and so on. In some cases, record linkage was difficult even for a human expert, requiring way more information than the one found in the raw source titles in Scopus.

We have developed an algorithm that breaks source titles into parts and checks for matches between the subparts. However, the details and the evaluation of precision and recall are outside the scope of this article. Even in this case, validation was made manually on a sample of 10% of publications, and the number of errors turned out to be less than 1%. This is due to our design choice to favor precision over recall and to avoid false positives in the entity resolution algorithm. Similar issues had to be faced when mapping authors' names to Scopus profiles, and were solved by exploiting techniques described in Demetrescu et al. [2018].

6 Concluding remarks

In this article, we have addressed the problem of quantitatively assessing the impact that a finer-grained classification of scientific areas would entail on the evaluation of researchers based on bibliometric indicators. Using as a case study the Italian national habilitation system, which classifies faculty members according to their academic discipline, we have focused on two tightly related disciplines: Computer Science and Computer Engineering. We have considered 12 common sub-areas, classifying researchers based on their productivity in each of them. Our investigation shows that the bibliometric indicators of some sub-areas tend to be significantly higher, even within the same academic discipline. As a consequence, evaluations conducted by comparing researchers in the same overall academic discipline are likely to favor them.

In the long term, the bibliometric distance between diverse research sub-communities may even introduce a bias on the authors' decisions of which discipline to focus on, in order to maximize their chances of being hired or promoted, further exacerbating the disparity and possibly causing a loss of interest for potentially relevant research topics. We argue that, in the case of academic disciplines with unbalanced bibliometric indicators across sub-areas, the evaluation should be carried out at a finer level of granularity, by comparing researchers with peers in the same sub-area.

Empirical studies that, similarly to our work, show substantial bibliometric differences between papers in different sub-areas of the same discipline have recently been presented, e.g., by Simko [2015] and by Gorraiz et al. [2016] focusing on biology and geography, respectively. The results show very heterogeneous publication strategies and different bibliometrics indicators, even within the same discipline. For instance, the investigation in Simko [2015] appears to confirm the anecdotal evidence, well-known among plant researchers, that research performed on certain plant species is much more likely to be cited: this corresponds to an extremely fine granularity level. We finally notice that in the first rounds of the ASN (2012-2013) only one scientific discipline (experimental physics) was divided into two subareas to accommodate for the multimodality of the distribution, partitioning professors of this discipline into two separate sets².

References

- Giovanni Abramo and Ciriaco Andrea D'Angelo. How do you define and measure research productivity? *Scientometrics*, 101(2):1129–1144, Nov 2014. ISSN 1588-2861. doi: 10.1007/s11192-014-1269-8. URL <https://doi.org/10.1007/s11192-014-1269-8>.
- Giovanni Abramo and Ciriaco Andrea D'Angelo. An assessment of the first “scientific habilitation” for university appointments in Italy. *Economia Politica*, 32(3):329–357, Dec 2015. ISSN 1973-820X. doi: 10.1007/s40888-015-0016-9. URL <https://doi.org/10.1007/s40888-015-0016-9>.
- Giovanni Abramo and Ciriaco Andrea D'Angelo. A farewell to the mncs and like size-independent indicators. *Journal of Informetrics*, 10(2):646 – 651, 2016. ISSN 1751-1577. doi: <https://doi.org/10.1016/j.joi.2016.04.006>. URL <http://www.sciencedirect.com/science/article/pii/S1751157716300955>.
- Giovanni Abramo, Ciriaco Andrea D'Angelo, and Francesco Rosati. Career advancement and scientific performance in universities. *Scientometrics*, 98(2):891–907, Feb 2014. ISSN 1588-2861. doi: 10.1007/s11192-013-1075-8. URL <https://doi.org/10.1007/s11192-013-1075-8>.
- Giovanni Abramo, Ciriaco Andrea D'Angelo, and Francesco Rosati. The determinants of academic career advancement: Evidence from Italy. *Science and Public Policy*, 42(6):761–774, 02 2015. ISSN 0302-3427. doi: 10.1093/scipol/scu086. URL <https://doi.org/10.1093/scipol/scu086>.
- T. Becher. *Academic tribes and territories: intellectual enquiry and the cultures of disciplines*. Buckingham: Open University Press, 1989.

² We refer to the Web page <https://www.anvur.it/attivita/asn/asn-2012-2013/indicatori-e-relative-mediane/> and to the document https://www.anvur.it/wp-content/uploads/2012/08/documento_accompagnamento_all_2.pdf (both in Italian) for details.

- T. Becher and P. R. Trowler. *Academic tribes and territories: intellectual enquiry and the cultures of disciplines*. Buckingham: Open University Press/SRHE, 2nd ed. edition, 2001. ISBN 033520628X.
- Lutz Bornmann, Rüdiger Mutz, and Hans-Dieter Daniel. Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the Association for Information Science and Technology*, 59(5):830–837, 2008a.
- Lutz Bornmann, Rüdiger Mutz, and Hans-Dieter Daniel. Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics*, 8:93–102, 2008b.
- Jeffrey Braithwaite, Jessica Herkes, Kate Churrua, Janet C Long, Chiara Pomare, Claire Boyling, Mia Bierbaum, Robyn Clay-Williams, Frances Rapport, Patti Shih, Anne Hogden, Louise A Ellis, Kristiana Ludlow, Elizabeth Austin, Rebecca Seah, Elise McPherson, Peter D Hibbert, and Johanna Westbrook. Comprehensive researcher achievement model (cram): a framework for measuring researcher achievement, impact and influence derived from a systematic literature review of metrics and models. *BMJ Open*, 9(3), 2019. ISSN 2044-6055. doi: 10.1136/bmjopen-2018-025320. URL <https://bmjopen.bmj.com/content/9/3/e025320>.
- Camil Demetrescu, Andrea Ribichini, and Marco Schaerf. Accuracy of author names in bibliographic data sources: an italian case study. *Scientometrics*, 117(3):1777–1791, 2018.
- Camil Demetrescu, Francesco Lupia, Angelo Mendicelli, Andrea Ribichini, Francesco Scarcello, and Marco Schaerf. On the Shapley value and its application to the Italian VQR research assessment exercise. *J. Informetrics*, 13(1):87–104, 2019.
- Leo Egghe. An improvement of the h -index: The g -index. *ISSI Newsletter*, 2, 05 2006.
- Rebecca Garner, Joshua Hirsch, Felipe Albuquerque, and Kyle Fargen. Bibliometric indices: Defining academic productivity and citation rates of researchers, departments and journals. *Journal of neuroint-erventional surgery*, 10, 08 2017. doi: 10.1136/neurintsurg-2017-013265.
- Juan Gorraiz, Christian Gumpenberger, and Thomas Glade. On the bibliometric coordinates of four different research fields in geography. *Scientometrics*, 107(2):873–897, 2016.
- Diana Hicks, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548):429–431, Apr 2015. doi: 10.1038/520429a.
- J. E. Hirsch. An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences*, 102(46):16569–16572, 2005. ISSN 0027-8424. doi: 10.1073/pnas.0507655102. URL <https://www.pnas.org/content/102/46/16569>.
- Clint D. Kelly and Michael D. Jennions. The h index and career assessment by numbers. *Trends in Ecology & Evolution*, 21(4):167 – 170, 2006. ISSN 0169-5347. doi: <https://doi.org/10.1016/j.tree.2006.01.005>. URL <http://www.sciencedirect.com/science/article/pii/S0169534706000255>.
- Gabriel Kreiman and John Maunsell. Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience*, 5:48, 2011. ISSN 1662-5188. doi: 10.3389/fncom.2011.00048. URL <https://www.frontiersin.org/article/10.3389/fncom.2011.00048>.
- Giulio Marini. New promotion patterns in italian universities: Less seniority and more productivity? Data from ASN. *Higher Education*, 73(2):189–205, Feb 2017. ISSN 1573-174X. doi: 10.1007/s10734-016-0008-x. URL <https://doi.org/10.1007/s10734-016-0008-x>.
- Giulio Marini. Tools of individual evaluation and prestige recognition in spain: how sexenio ‘mints the golden coin of authority’. *European Journal of Higher Education*, 8(2):201–214, 2018. doi: 10.1080/21568235.2018.1428649. URL <https://doi.org/10.1080/21568235.2018.1428649>.
- Giulio Marini and Viviana Meschitti. The trench warfare of gender discrimination: Evidence from academic promotions to full professor in italy. *Scientometrics*, 115(2):989–1006, May 2018. ISSN 0138-9130. doi: 10.1007/s11192-018-2696-8. URL <https://doi.org/10.1007/s11192-018-2696-8>.
- Moreno Marzolla. Quantitative analysis of the Italian National Scientific Qualification. *Journal of Informetrics*, 9(2):285 – 316, 2015. ISSN 1751-1577. doi: <https://doi.org/10.1016/j.joi.2015.02.006>. URL <http://www.sciencedirect.com/science/article/pii/S1751157715000243>.
- Moreno Marzolla. Assessing evaluation procedures for individual researchers: The case of the Italian National Scientific Qualification. *Journal of Informetrics*, 10(2):408 – 438, 2016. ISSN 1751-1577. doi: <https://doi.org/10.1016/j.joi.2016.01.009>. URL <http://www.sciencedirect.com/science/article/pii/S1751157715300961>.
- Marco Pautasso. The Italian University Habilitation and the Challenge of Increasing the Representation of Women in Academia. *Challenges*, 6(1):26–41, 2015. ISSN 2078-1547. doi: 10.3390/challe6010026. URL <https://www.mdpi.com/2078-1547/6/1/26>.
- Antonis Sidiropoulos, Dimitrios Katsaros, and Yannis Manolopoulos. Generalized hirsch h -index for disclosing latent facts in citation networks. *Scientometrics*, 72:253–280, 08 2007. doi: 10.1007/s11192-007-1722-z.

- Ivan Simko. Analysis of bibliometric indicators to determine citation bias. *Palgrave Communications*, 1: 15011, 2015.
- Roberto Todeschini and Alberto Baccini. *Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research*. Wiley, 2016. ISBN 978-3-527-33704-0.
- Paul Trowler. Academic tribes and territories: The theoretical trajectory. *Osterreichische Zeitschrift fur Geschichtswissenschaften*, 25:17–26, 01 2014.
- Paul Trowler, Murray Saunders, and Veronica Bamber. *Tribes and territories in the 21st century: Rethinking the significance of disciplines in higher education*. Routledge, 2012.
- Ludo Waltman, Nees Jan van Eck, Thed N. van Leeuwen, Martijn S. Visser, and Anthony F.J. van Raan. Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1):37 – 47, 2011. ISSN 1751-1577. doi: <https://doi.org/10.1016/j.joi.2010.08.001>. URL <http://www.sciencedirect.com/science/article/pii/S1751157710000817>.
- Chun-Ting Zhang. The e-index, complementing the h-index for excess citations. *PLOS ONE*, 4(5):1–4, 05 2009. doi: [10.1371/journal.pone.0005429](https://doi.org/10.1371/journal.pone.0005429). URL <https://doi.org/10.1371/journal.pone.0005429>.