ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Zagonari, Fabio

Working Paper

Scientific production and productivity in curriculum vitae characterisation: Simple and nested H indices that support cross-disciplinary comparisons

Quaderni - Working Paper DSE, No. 1100

Provided in Cooperation with:
University of Bologna, Department of Economics

Suggested Citation: Zagonari, Fabio (2017) : Scientific production and productivity in curriculum vitae characterisation: Simple and nested H indices that support cross-disciplinary comparisons, Quaderni - Working Paper DSE, No. 1100, Alma Mater Studiorum - Università di Bologna, Dipartimento di Scienze Economiche (DSE), Bologna, https://doi.org/10.6092/unibo/amsacta/5601

This Version is available at:
https://hdl.handle.net/10419/177600

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

https://creativecommons.org/licenses/by-nc/3.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Scientific production and productivity in curriculum vitae characterisation: simple and nested *H* indices that support cross-disciplinary comparisons

Fabio Zagonari

Quaderni - Working Paper DSE N°1100

Scientific production and productivity in curriculum vitae characterisation: simple and nested *H* indices that support cross-disciplinary comparisons

Fabio Zagonari
Dipartimento di Scienze Economiche, Università di Bologna, via Angherà 22, 47921 Rimini (Italy)
Phone: 0039 0541 434135 Fax: 0039 0541 434120 E-mail: fabio.zagonari@unibo.it

Abstract

In this study, I developed operational versions of Hirsch's *H* index that can be applied to each researcher's curriculum vitae (CV) to allow cross-disciplinary comparisons. The revised indices account for anomalies that potentially arise from tactical or opportunistic citation and publication behaviours by authors and editors, and can be calculated from readily available information. I split the original *H* index into nested indices to isolate networking activity, distinguish scientific production and productivity, and used nested Gini indices to identify intentional and successful inter-topical and inter-disciplinary research. I applied the *most popular* normalisations (i.e., per author and per year) using simple methodologies (i.e., least-squares linear and cubic interpolation fitting, whole-career vs. sub-periods, two-dimensional graphs) to solve *empirical* problems (e.g., sensitivity to citations, the "fashion" effect, attribution to disciplines, life cycle of articles) as well as *open* questions (e.g., the attribution of an article to a given discipline) associated with the original *H* index. I provided *three* numerical examples based on a representative heterodox, a representative orthodox multi-disciplinary, and a representative orthodox uni-disciplinary CV: the *first* CV includes 17 Scopus publications, and shows a highly heterodox (i.e., 5.8%), but no interdisciplinary research career, with a tiny networking component (i.e., 0.9%); the *second* CV includes 24 Scopus publications, and shows a slightly heterodox (i.e., 0.3%), but highly interdisciplinary (i.e., 53.9%) research career, with a small networking component (i.e., 14.3%); the *third* CV includes 16 Scopus publications, and shows slightly heterodox (i.e., 0.1%) and no interdisciplinary research career, with no networking component.

Keywords

Bibliometric ranking, scientific productivity, scientific production, *H* index, cross-discipline research

JEL code

Y91

1. Introduction

All *theoretical* definitions of scientific activity depend on social institutions, which change in time and space. For example, astrology used to be a science, but is no longer, whereas sociology was not considered to be a science, but now is. However, regardless of a discipline's classification, researchers in a discipline and their managers seek ways to evaluate research activity, and this is most commonly done based on a researcher's publication record (hereafter, their curriculum vitae [CV]). As a result, many theoretical and numerical methods (ranging from simple counts of the total number of publications to complex indices) have been proposed to make the evaluation process more objective and more effective. Unfortunately, there are significant problems with each of these methods, some of which relate to a lack of practical simplicity, and some of which relate to how researchers can "game the system" and artificially improve their rating.

In the present paper, my goal was to review and improve these methods for scientific researchers. To do so, I will first constrain my context by proposing the following *operational* definition of scientific activity: *disseminating original scientific knowledge*. In this definition, *disseminating* refers to publication of research results in a form that is available to the research community, with the goal of directly or indirectly benefiting society, despite the many examples of misuse of scientific knowledge; *original* is (necessarily) based on direct assessments by experts, although peer review might be affected by subjective evaluations, conflicts of interest, and biases against innovative (heterodox) scientific ideas and approaches; and *scientific* rests on publications worldwide considered as scientific, with reference, but without loss of generality, to the Scopus dataset (https://www.scopus.com), although other databases could instead be used in such an analysis. In particular, I will refer to the following publication features:

- I will focus on English, to emphasize *international* dissemination. Note that citing publications in languages other than English are also included in this analysis.
- I will focus on full-length peer-reviewed articles to rely on a prior scrutiny of their *originality* by peer reviewers.
- I will focus on net citations, after eliminating self-citations and cross-citations by all *co-authors* and *colleagues*, by deleting records in which the same author appears in both the citing publication and the cited article (although this will exclude some legitimate self-citations, it also mitigates the problem of excessive citation of one's own papers) as well as records in which the same affiliation appears in the citing publication and cited article (although this will exclude some legitimate citations of the work of colleagues that provide important context, it also mitigates the problem of excessive cross-citation). Here, I define cross-citations as situations in which co-authors reciprocally cite each other's work. This will mitigate "apostle" effects (i.e., inflating citations by relying on temporal linkages) and network effects (i.e., boosting citations by relying on personal linkages). Note that *co-authors* refer to any kind of publication (e.g., books, symposium proceedings, research notes) and *colleagues* refer to all researchers affiliated at any time with the author whose CV is being studied.

This *first* step will be to differentiate networking from scientific activity (Cainelli et al., 2012, 2015; Schubert, 2012) by highlighting anomalies in a CV that potentially arise from the questionable practice of unjustified cross-citations by authors. Indeed, networking might be beneficial if it increases the cultural exchanges between organizations (De Stefano & Zacccarin, 2016), but it is detrimental if it affects researchers' attitudes towards scientific interests and publishing (McCarty & Jawitz, 2013). In other words, each organization should be in a position to transparently evaluate the networking activity of each researcher.

Note that deleting cross-citations could discriminate against heterodox scientists, who are typically few, familiar to each other, and likely to be co-authors (e.g., post-Keynesian or Marxist economists). However, I will deal with this issue by characterizing the CV in terms of publication inequality: here, the Gini index is applied to measure dispersions of publications across different journals and different disciplines. Moreover, instead of "purifying" or "cleaning" the individual

scientific activity by estimating the individual researcher's network from the database of all scientists (e.g., De Stefano et al., 2013; Petersen, 2015), I deleted cross-citations from the dataset in order to keep calculations simple for each researcher. Finally, to be conservative, I will retain records in which the citing and cited authors are, or have been, members of the same editorial board, the same family, the same PhD course, the same workshop, or similar relationships. In other words, consistently with the Scopus dataset, I will consider citations from indirect co-authors and colleagues (i.e., the co-authors of co-authors who are not themselves the researcher's co-authors; similarly, the colleagues of colleagues who are not themselves the researcher's colleagues). However, authors working on a given topic within the same organization are likely to be co-authors, at least occasionally. The focus on full-length articles in peer-reviewed journals (as opposed to notes, comments, or letters) will minimize the probability of considering spontaneous or induced citations due to casual errors (i.e., mistakes) or deliberate errors (i.e., conscious deception) by identifying citations used in new (i.e., not yet published) research and distinguishing them from the existing literature.

Second, the literature on the evaluation of researchers' activity consists of many theoretical papers (e.g., Bouyssou & Marchant, 2016a). Instead, I will develop a practical index (e.g., Abramo et al., 2013a). Alternatively, one could rely on expert panels (e.g., Engels et al., 2013), despite their subjectivity, or on factor analysis (e.g., Schreiber et al., 2011), despite its methodological issues. Many alternative operational bibliometric indices have been suggested in the literature, but very often without significant positive correlations between the metrics (Wainer & Vieira, 2013). For example, indices discussed in recent papers include the *c* index (Yan et al., 2013), *z* index (Petersen & Succi, 2013), *g* index (Adachi & Kongo, 2015; Bartolucci, 2015; Bertoli-Barsotti, 2016; De Visscher, 2011; Prathap, 2014; Rousseau, 2015; Schreiber, 2013a,b), and generalized Hirsch (*H*) index (Gagolewski & Mesiar, 2012). Here, I will refer to the most popular *H* index (Bornmann et al., 2011; Bertoli-Barsotti & Lando, 2015). In particular, I will calculate alternative nested (as clarified in Section 2) versions of the *H* index based on common information available in the Scopus dataset, by referring to insights suggested by the authors of other indices and by attempting to solve the *empirical* shortcomings (e.g., sensitivity to citations, the "fashion" effect, attribution to disciplines, life cycle of articles; Dienes, 2015) as well as the *open* questions (e.g., the attribution of an article to a given discipline; Gagolewski, 2013) of the original *H* index. Note that the focus on articles reduces the dependence of my results on the dataset used because most databases include all full-length journal articles by an author.

This *second* step will let me characterize each researcher's CV in terms of the publication inequality in percentages (i.e., the *actual* dispersion of publications in different journals and disciplines with respect to the *potential* distribution of publications in all distinguished journals and disciplines) as a result of heterodoxy (i.e., research that challenges the prevailing dogma) vs. orthodoxy (i.e., research that confirms the prevailing dogma). This will also account for the intra- vs. inter-disciplinary nature of the publication without requiring complicated statistical analyses of bibliometric datasets. To do so, the proposed approach reveals the following potentially questionable practices:

- Publication of low-quality manuscripts, because of an author's intra-discipline reputation, with the *ex ante* aim of a journal's editors (often unachieved, due to a lack of *ex post* citations) of increasing the journal's impact factor.
- Many publications of low-quality manuscripts in the same journal as a result of personal relationships between an editor and an author.
- Many citations of papers published by the same journal as a pre-requisite for acceptance of a manuscript for publication, in order to improve the impact factor of the journal.

In addition, an inter-disciplinary CV requires longer to become familiar to a well-known part of the literature, so that fewer articles are published, whereas a heterodox CV is likely to be cited only within the same journals (i.e., not by orthodox journals), resulting in fewer citations. In other words, each organization should be in a position to transparently favour or deprecate inter-disciplinary or heterodox articles or CVs. Note that I will *not* consider the importance of citations based on

algorithms for ranking Web pages such as PageRank (Senanayake et al., 2015; Yates & Dixon, 2015). Moreover, I will *not* use the impact factors of the cited journals (Abbas, 2012; Moskovkin et al., 2014), because (among other reasons) these continuously changing figures are not included in individual datasets such as Scopus. Finally, I will *not* compare alternative datasets in terms of their reliability and stability (e.g., Harzing & Alakangas, 2016; Wildgaard, 2015).

Third, many normalizations have been suggested for use in cross-disciplinary comparisons of individual scientific activity (Bornmann & Marx, 2015; Harzing & Alakangas, 2016; Harzing et al., 2014):

- Number of articles per author
- Number of citations per author or per year for the cited article
- Numbers of citations in the citing article
- Average numbers of citations or publications in each discipline
- Weighted numbers of citations or publications according to an author's position in the list of authors for an article

Here, I will apply the two most popular normalizations (i.e., numbers of citations per author and per year in the cited article). These choices are popular because they are supported by easily available data that requires little or no transformation.

This *third* step will let me distinguish scientific production and productivity (as defined in Section 2) by applying different indices for these different goals. Note that I will *not* use the average numbers of citations in each discipline (Radicchi & Castellano, 2012), since this continuously changing figure is not provided by the available bibliometric datasets. Moreover, I will *not* apply the number of citations in the citing publication (Bouyssou & Marchant, 2016b), since it has been empirically shown to be ineffective in cross-disciplinary comparisons (Radicchi & Castellano, 2012). Finally, I will *not* use the weighted number of citations or publications according to an author's position in the list of authors for an article (Abramo et al., 2013b; Liu & Fang, 2012), since different practices prevail in different universities, disciplines and countries.

In summary, the purpose of this paper is to identify *simple* (i.e., based on information readily available in Scopus or other datasets) *nested* versions of the *H* index that use feasible normalizations to avoid discipline discrimination. To do so, the index consistently combines insights suggested in the literature, disentangles *scientific production and productivity* from networking activity, and solves the *empirical* shortcomings and *open* questions for other indices. As a result, it characterizes each researcher's CV (i.e., heterodox vs. orthodox and intra- vs. inter-disciplinary research) to avoid curriculum discrimination by highlighting possibly questionable tactical or opportunistic behaviours in publication and citation by authors and journal editors [1]. To illustrate this approach, the methodology will be applied to *three* representative CVs.

Note that I will disregard individual characteristics that change over time, such as the researcher's age or position (Cainelli et al., 2012, 2015; Morichika & Shibayama, 2015), and individual characteristics that are fixed in time, such as gender or race (Abatemarco & Dell'Anno, 2013; Hopkins et al., 2013; Sotudeh & Khoshian, 2014). In this paper, my interest is in *indices for*, rather than *determinants of*, scientific activity (Penner et al., 2013). Moreover, I will omit papers that use indices to compare countries (e.g., Sangwal, 2013), institutions (e.g., Abramo et al., 2013c), or journals (e.g., Ko & Park, 2013; Tsai, 2014). Finally, I will disregard indices that are unsuitable for cross-disciplinary comparisons, such as the *P* top 10% of Bornmann et al. (2012) and the *P100* of Prathap (2012) and Schreiber (2014a,b,c).

Specifically, the present work provides two main contributions to the existing literature. First (Section 4), it provides a simple methodology (i.e., least-squares linear and cubic interpolation fitting, whole career vs. sub-periods, normalisations per author and per year) that allows each researcher to calculate a more suitable version of their *H* index (i.e., with production distinguished from productivity) and that can be used for both senior and junior researchers. This approach also

[1] It's important to note that although this approach can identify *potentially* questionable behaviour, human judgment is still required to determine whether the behaviour is, in fact, legitimate.

solves the abovementioned key *empirical* issues and *open* questions associated with the original *H* index and reveals some potentially questionable practices in citations by authors and journal editors. Second (Section 5), it uses a simple methodology (i.e., two-dimensional graphs) that lets each researcher visually characterize the diversity of their CV. This approach also tackles issues of discrimination against heterodox and inter-disciplinary CVs (i.e., a smaller number of citations and of publications, respectively), and highlights CV anomalies that may arise from questionable publication and citation practices by editors, and in publication decisions by authors and editors.

2. Methodology

In this section, I will clarify the meaning of *nested* versions of the *H* index to estimate individual scientific *production and productivity*.

In this context, I treat the production of articles up to a given point in time as a total (stock) variable, analogous to the total GWh of hydropower produced since the beginning of this year. In contrast, I treat productivity as a marginal (flow) variable that can be used to measure the sensitivity (or dynamics) of the total production to a changed or potentially changing factor, analogous to the GWh of power produced last month by each turbine or the potential (marginal) change in GWh due to addition of a new turbine. The integral (summation) of a series of marginal variables (i.e., productivity) results in the total value of that variable (i.e., production); conversely, the derivative of the total variable with respect to a factor (e.g., with respect to time) of a total variable (i.e., production) amounts to the marginal variable (i.e., productivity). In particular, I will use production to estimate the total scientific activity, but will use productivity to evaluate changes over time in scientific activity, where the sum of productivities for each part of the overall period sum up to production during the whole period. Indeed, production and productivity refer to different goals: the decision to recruit a junior researcher as an Assistant or Associate Professor based on scientific productivity is different from the assessment of a senior researcher for promotion to an endowed chair or from the awarding of *ad honorem* degrees based on scientific production.

Therefore, I will use the normalization "per author" to assess production over the total career of a researcher (i.e., based on the number of years from the first publication to the present), but will use the normalization "per author per year" (i.e., based on the number of years since publication for each article) to evaluate productivity in specific sub-periods. In particular, I will refer to the last 22 years to obtain a reliable dataset on production (as suggested by the Scopus Web site, many inconsistencies might arise for publications and citations before 1995). In contrast, I will refer to the last 10 years to reduce the impacts on productivity of recent changes in editorial strategies (such as increased availability of articles online) that have led to more publications and citations. Indeed, shorter periods are more likely to miss the citation cycle of articles in some disciplines, since the length of the cycle often increases with increasing originality of an article; for example, the most innovative articles might be ignored for some time after publication, but then cited for a long time. See Liu & Yang (2014) for an application of the *H* index to show a scientist's dynamic research trajectory and scientific performance during different periods.

Note that I will rely on the numbers of citations per author and per year in the cited articles, although these normalizations have only been supported by statistical analyses of small samples (Harzing & Alakangas, 2016; Ryan, 2016).

In the context of nested indices, the Hirsch index (*H*), which is based on the number of publications and citations in different journals and disciplines, can be coupled with the Gini index (*G*), which accounts for differences between journals and disciplines (i.e., dispersion of articles among journals and disciplines). I will define both indices analytically in Section 5. In the Scopus dataset, each publication is attached to a source (i.e., a journal or review) and classified into one or more subject, discipline, and topic areas (i.e., 4 subjects, 27 disciplines within those subjects, and 306 topics within those disciplines). Consistently with the Scopus dataset, I will assume that each article is linked to the most representative discipline for its source. In future analyses, a more complex classification (i.e., topics) might be possible that associates a publication with two or more different

groups of topics, although this would unreasonably enlarge the interdisciplinary features of CVs. Of course, the opposite problem applies for a less complex classification (i.e., subjects); this level artificially decreases the degree of interdisciplinarity. In other words, I will assume that a journal cannot be attached to two different disciplines. In future analyses, a multi-discipline classification for journals such as *Ecological Economics* that clearly span two or more disciplines should be considered, although this will require formal definition of the distance between disciplines so that a standard that is objective (quantifiable) as possible (i.e., there is some subjectivity involved) can be applied to the vectors (i.e., 27 relative weights for the 27 disciplines) that characterise each journal. After all, there are few journals in which all published articles are attributed to a single discipline (i.e., 100% uni-disciplinary), even though most journals can be said to be mainly focused on a given discipline (i.e., its prevailing discipline). Next, I will assume that each journal represents a single topic within a discipline; in other words, I will assume that a journal cannot be attached to two different topics. Although this assumption is an obvious simplification, some combinations of fields and methodologies are only accepted by a few journals. Consistently with the Scopus dataset, each article could be linked to the most representative topic for its source. In future analyses, a multi-topic classification for journals should be considered, although this will require formal definition of the distance between topics so that a standard that is objective (quantifiable) as possible (i.e., there is some subjectivity involved) can be applied to the vectors (i.e., 306 relative weights for the 306 topics) that characterise each journal.

Therefore, to characterize CVs, I will apply a differential approach. This is similar to the approach used by Blagus et al. (2015), who applied it to alternative versions of the *H* index. In the present approach, both levels and differences are meaningful for both the Hirsch and Gini indices, for deleting records at each step, and for constructing a system of nested indicators. This is true even though *G* indices are based on the topic and discipline of each publication, whereas *H* indices are based on the total number of citations (i.e., gross citations) and the number of citations after deleting records based on the abovementioned criteria (i.e., net citations), together with the topic and discipline of a given publication.

Note that I will use similar inequality constraints for both indices: *G* indices for disciplines are less than or equal to *G* indices for journals, and *H* indices for disciplines are less than or equal to *H* indices for journals. Moreover, I will present Gini indices as percentages, but Hirsch indices as levels. Indeed, there is a maximum level of inequality that can be used to normalize the *G* indices: $(N-1)/N$, with N being the total number of publications by an author. In contrast, there is no maximum level for *H* indices. Finally, I will develop the analysis by building on the most popular bibliometric index (i.e., the *H* index for total scientific production). Table A1 summarizes the indices applied in this analysis. Section 4 describes calculations of alternative *H* indices to estimate scientific production and productivity (e.g., horizontal analysis), whereas Section 5 characterizes alternative CVs based on both *G* and *H* indices (e.g., vertical analysis).

3. Data

The Scopus data set includes the following variables for both cited and citing articles:

- Year;
- Author;
- Affiliation: institute/university, city, country;
- Source: journal/review title;
- Subjects: health, life, physical, social sciences;
- Disciplines: 5 in health sciences (medicine, veterinary, nursing, dentistry, health professions), 5 in life sciences (pharmacology & toxicology, biological, neurology, agricultural, immunology), 9 in physical sciences (chemistry, physics & astronomy, mathematics, Earth & planetary, energy, environmental, materials, engineering, computing & information), and 8 in social sciences (psychology, economics & econometrics & finance, arts & humanities, business & management & accounting, decision, politics, architecture, sociology).

Note that in this study, I used all information included in the Scopus dataset, together with the provided calculations of alternative *H* indices that exclude cross-citations and select different time periods. In particular, the Scopus dataset includes 17,325,760 and 13,238,483 authors in 1996-2015 and 2006-2015, respectively; 23,953,840 and 14,905,298 are the total number of English articles published in 1996-2015 and 2006-2015, respectively, with the following percentages for the 4 subjects: health 23, life 24, physical 43, and social 10 in 1996-2015 and health 24, life 22, physical 43, and social 11 in 2006-2015; means/standard deviations of *H* indices that include cross-citations in 1996-2015 and 2006-2015 are 2.70/4.89 and 2.41/3.93, respectively; means/standard deviations of *H* indices that exclude cross-citations in 1996-2015 and 2006-2015 are 2.45/4.37 and 2.16/3.42, respectively.

4. Scientific production and productivity

In this section, I will apply least-squares linear and cubic interpolation fitting for the relationship between the net citations and the number of publications to calculate both scientific production and productivity (see Table A2, A3 and A4 in the Appendix for the three analysed CVs). Note that a quadratic interpolation would be inappropriate because it would produce an increasing curve for the less frequently cited articles. In particular, linear interpolation will emphasize net citations above the minimum required by the original *H* index (i.e., the linear index values are likely to be larger than the original *H* index values), whereas cubic interpolation will de-emphasize them. See Martinez et al. (2014) for an application of the *H* index to highly cited papers. For this reason, I will name the four cases that I study *total production* (i.e., a linear interpolation applied to production), *total core production* (i.e., a cubic interpolation applied to production), *average productivity* (i.e., a linear interpolation applied to productivity), and *average core productivity* (i.e., a cubic interpolation applied to productivity). See Glanzel (2012) for the notion of *core* documents. Thus, once focused on articles in English and normalised for the number of authors, the *H* index for total production (H_{ltn22}, where l = linear, t = total, n = net citations, and 22 = the time from the start of the Scopus dataset to the present) will be calculated by applying the following formula:

$$x \text{ such that } y = a_0 - a_1 x = x$$

with positive parameters a_0 and a_1 coming from a linear regression of the total number of net citations y over the total number of publications x in the last 22 years. The *H* index for total core production (H_{ctn22}, where c = cubic, t = total, n = net, and 22 = the time from the start of the Scopus dataset to the present) will be calculated as follows:

$$x \text{ such that } y = a_0 - a_1 x + a_2 x^2 - a_3 x^3 = x$$

with positive parameters a_0, a_1, a_2 and a_3 arising from a cubic regression of the total number of net citations y over the total number of publications x in the last 22 years. Note that an interpolation based on x such that $y = 1/(a_0 - a_1 x) = x$ (i.e., two parameters) would produce similar results, but it is not implemented by the most popular software. The *H* index for average productivity (H_{lyn10}, where l = linear, y = citations divided by number of years since publication, n = net citations, and 10 = the past 10 years in the Scopus dataset) will be calculated by applying the following formula:

$$x \text{ such that } y = a_0 - a_1 x = x$$

with positive parameters a_0 and a_1 coming from a linear regression of the number net citations per year y over the number of publications per year x in the last 10 years. The *H* index for average core productivity (H_{cyn10}, where c = cubic, y = citations divided by number of years since publication, n = net citations, and 10 = the past 10 years in the Scopus dataset) will be calculated as follows:

$$x \text{ such that } y = a_0 - a_1 x + a_2 x^2 - a_3 x^3 = x$$

with positive parameters a_0, a_1, a_2 and a_3 arising from a cubic regression of the number net citations per year y over the number of publications per year x in the last 10 years. Note that the obtained *H* indices are continuous and do not change abruptly when the number of citations of a single article changes (i.e., they solve the problem of the discontinuity that could potentially be created by an additional citation received by the marginal article) (Thelwall & Wilson, 2014), because they account for the citations received by the entire set of published articles (Anania & Caruso, 2013).

Figure 1. Least-squares linear interpolation fitting for the *H* index per author (i.e., scientific production): H_{ltn22} = 6.29 (R^2 = 0.66). The increasing linear line represents *y* = *x*; l = linear interpolation, t = total number of publications and citations, n = net citations, and 22 = 22 years.

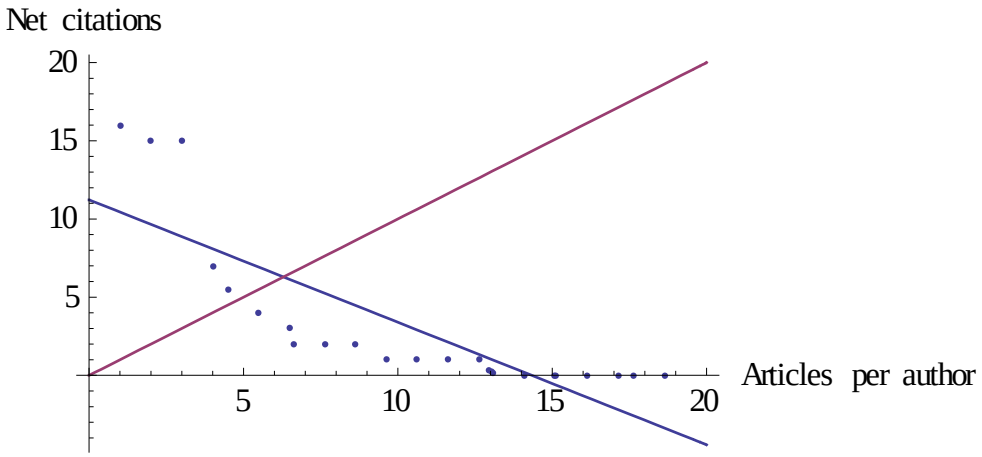


Figure 2. Least-squares cubic interpolation fitting for the *H* index per author (i.e., scientific core production): H_{ctn22}= 5.28 (R^2 = 0.95). The increasing linear line represents *y* = *x*; c = cubic interpolation, t = total number of publications and citations, n = net citations, and 22 = 22 years.

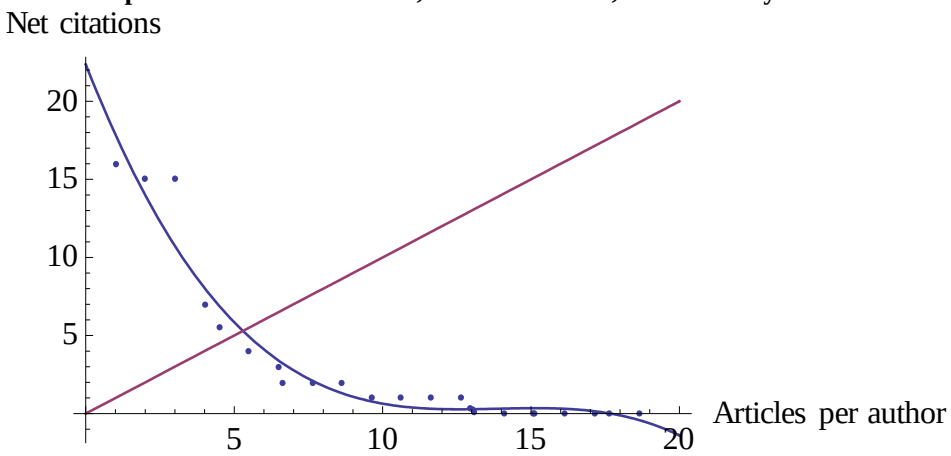


Figure 3. Least-squares linear interpolation fitting for the *H* index per author per year (i.e., scientific productivity): H_{lyn10} = 2.56 (R^2 = 0.76). The increasing linear line represents *y* = *x*; l = linear interpolation, *y* = publications and citations per year, n = net citations, and 10 = 10 years.

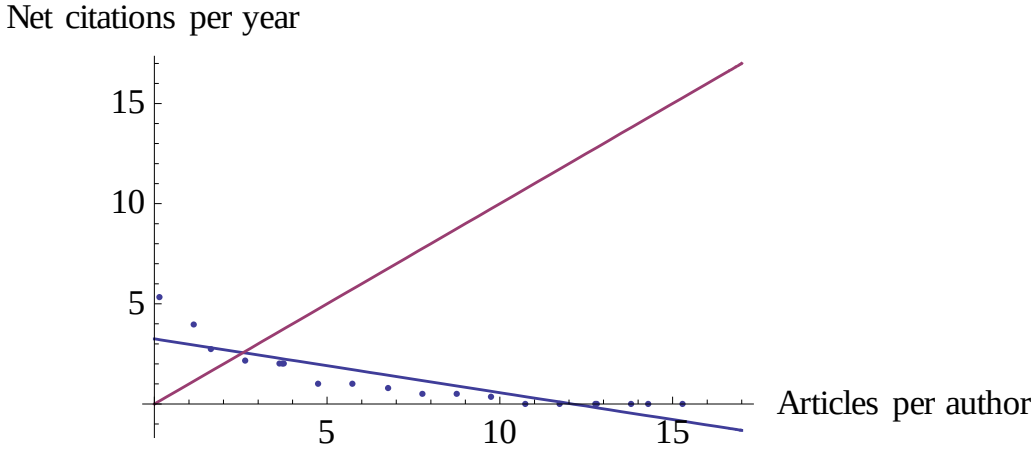
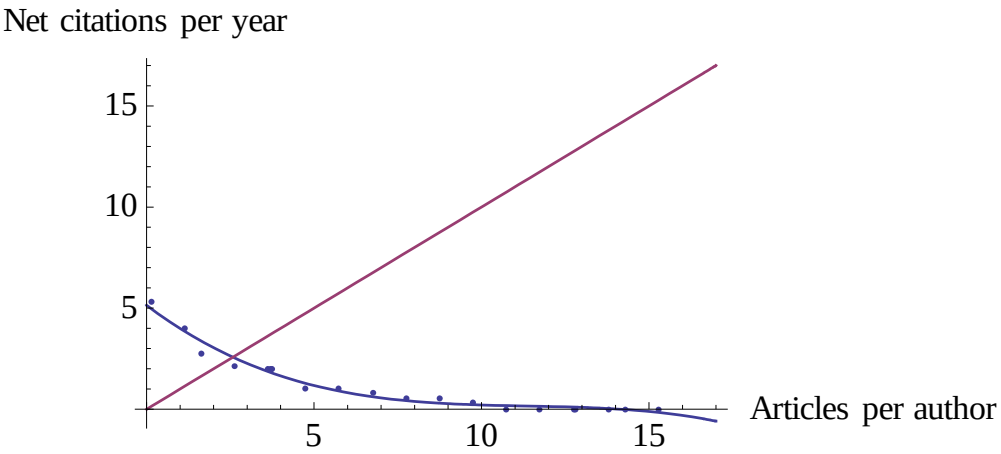


Figure 4. Least-squares cubic interpolation fitting for the *H* index per author per year (i.e., scientific core productivity): H_{cyn10} = 2.57 (R^2 = 0.97). The increasing linear line represents *y* = *x*; c = cubic interpolation, *y* = publications and citations per year, n = net citations, and 10 = 10 years.



Figures 1 and 2 show the *H* indices for the total and core scientific production based on linear and cubic interpolation, respectively (i.e., interceptions between the 45° lines and the linear and cubic interpolations, respectively). Note that the cubic interpolation de-emphasizes the most frequently cited articles. Figures 3 and 4 show the *H* indices for the average and core scientific productivity based on linear and cubic interpolation, respectively (i.e., interceptions between the 45° lines and the linear and cubic interpolations, respectively). Note that the cubic interpolation emphasizes the most persistently successful articles. Table 1 summarizes the values of the *H* indices for the *second* analysed CV. Note that the application of relative weights (e.g., the harmonic mean in Yang et al., 2013) to emphasize the author who serves as the author for correspondence or the order of author names (i.e., larger weights to authors listed first) would not significantly change these results: indeed, the vast majority of articles have a single author for all three analysed CVs.

Table 1. The estimated values of the *H* indices per author for the *second* analysed CV. All values are for the net production or productivity (i.e., after removal of cross-citations): c = cubic interpolation; l = linear interpolation; n = net number of citations; t = total articles; y = citations are divided by the number of years since publication; 22 = author's whole career; 10 = the last 10 years.

H	10 years (2007-2016)	*H*	22 years (1995-2016)
H_{ltn10}	6.02	H_{ltn22}	6.29 (Figure 1)
H_{ctn10}	4.93	H_{ctn22}	5.28 (Figure 2)
H_{lyn10}	2.56 (Figure 3)	H_{lyn22}	2.43
H_{cyn10}	2.57 (Figure 4)	H_{cyn22}	2.60

However, a linear interpolation gives too much weight to fashionable articles (i.e., articles with many citations in a few years), whereas a cubic interpolation disregards them by giving more weight to articles with few citations in many years. Consequently, a linear interpolation (i.e., H_{ltn22} or H_{lyn22}) seems to be most meaningful for total scientific production, whereas a cubic interpolation (i.e., H_{ctn10} or H_{cyn10}) seems to be most representative for the average core scientific productivity. Note that summary statistics of the Scopus dataset let me calculate both parameters (i.e., α and λ) that characterize the gamma distributions for the H indices. Here, the gamma distribution was chosen because, within non-negative distributions (i.e., H indices show non-negative values) and asymmetric distributions (i.e., scientists are more likely to achieve tiny or small than large or huge H indices), it can account for qualitatively different frequencies that depend on alternative values of its two parameters. For example, the *second* analysed CV turns out to be within the best 0.53% of scientists over the whole career (based on H_{ltn22} = 6.29), by normalising for an average of 4 authors per article, and within the 0.0008% best scientists in the last 10 years (based on H_{cyn10} = 2.57), by normalising for an average publication period of 5 years. In addition, similar calculations for the *first* and *third* analysed CVs suggest that *H* indexes per year should be used to compare careers of

senior researchers. Indeed, H_{ltn22} for the first CV is slightly larger than H_{ltn22} for the third CV (i.e., 10.00 > 8.34), whereas H_{lyn22} for the first CV is considerably smaller than H_{lyn22} for the third CV (i.e., 1.52 < 4.17).

These results are *internally* consistent, since the indexes achieved by the second analysed author are smaller for 22 years than for the last 10 years, when the majority of articles have been published (i.e., 18.6 and 15.3 English articles per author in the last 22 and 10 years, respectively), with H_{lyn22} = 2.43 being smaller than H_{cyn10} = 2.57. These results are also *externally* consistent. For example, calculations for the third CV (chosen for the reasons explained in Section 6) show 11.2 English articles per author and H_{lyn22} = 4.17, with his likely best index being 7.5 times better than the index achieved by the second analysed author, once normalised for publication years. Moreover, the present results are clearly better than those calculated using previous versions of the *H* index. The traditional *H* index for the *second* representative author for 22 years is 6, versus 11 for the third CV for 96 years, which is only 1.83 times the achieved index. Finally, these results can be easily interpreted. Indeed, if 2.57 articles are cited 2.57 times per year per author, there is the following implications: for four authors, the same *H* index could be achieved only if 10.3 articles were cited 10.3 times per year, which, over a period of 10 years, means that the 10.3 articles would each be cited 103 times. These results can also be simply justified. Indeed, it is difficult to support the belief that citation of one article with 10 authors only one time will directly or indirectly benefit science or society to the same extent as 10 articles with a single author, each cited 10 times.

5. CV characterization

To extend this analysis, I will now apply nested Gini (*G*) and Hirsch (*H*) indices for CV characterization, with the *G* indices based on the number of publications and the *H* indices based on both gross and net citations. First, I will disentangle networking from scientific activities. Then, I will apply the *G* and *H* indices to distinguish heterodox from orthodox CVs (Sutter, 2012) and multiple-topic and multiple-discipline CVs from single-topic and single-discipline CVs (De-Moya-Anegon et al., 2014). Finally, I will identify unintentional and unsuccessful (i.e., overall cited less than the expected or desired number of citations) inter-topical and inter-disciplinary CVs. Note that I will refer to the author's whole research career (i.e., total scientific production) in this analysis, although similar reasoning could be applied to shorter periods (i.e., scientific productivity). In particular, by applying linear interpolation to the gross number of citations (i.e., citations of English articles, including cross-citations) to calculate the linear total gross *H* index (i.e., H_{ltg22}), and by calculating the percentage difference between H_{ltg22} and H_{ltn22} (i.e., $[H_{ltg22} - H_{ltn22}]/H_{ltg22}$), this approach provides a measure of the relative importance of networking activity in a CV.

In the numerical example based on the *first* analysed CV (see Table A2 in the Appendix), H_{ltg22} = 10.00 versus a value of 7 for the traditional gross *H* index calculated by Scopus and H_{latg22} = 10.00 with the linear interpolation applied to the Scopus gross data. This compares with H_{ltn22} = 9.91, versus 7 for the traditional net *H* index calculated by Scopus and H_{latn22} = 10.00 using the linear interpolation applied to the Scopus net data. On this basis, networking activity is very small.

In the numerical example based on the *second* analysed CV (see Table A3 in the Appendix), H_{ltg22} = 7.34 versus a value of 6 for the traditional gross *H* index calculated by Scopus and H_{latg22} = 8.42 with the linear interpolation applied to the Scopus gross data. This compares with H_{ltn22} = 6.29, versus 6 for the traditional net *H* index calculated by Scopus and H_{latn22} = 7.62 using the linear interpolation applied to the Scopus net data. On this basis, networking activity is small (i.e., 8.6%). Indeed, its estimation calculated by comparing the gross and net H-indexes based on the Scopus dataset is 9.1% and 10.5% in 1996-2015 and 2006-2015, respectively.

In the numerical example based on the *third* analysed CV (see Table A4 in the Appendix), H_{ltg22} = 8.34 versus a value of 11 for the traditional gross *H* index calculated by Scopus and H_{latg22} = 11.52 with the linear interpolation applied to the Scopus gross data. This compares with H_{ltn22} = 8.34, versus 11 for the traditional net *H* index calculated by Scopus and H_{latn22} = 11.52 using the linear interpolation applied to the Scopus net data. On this basis, networking activity is null.

Second, I calculated H_{ljn22} as a linear interpolation of points where citations are in journals other than the journal that published the cited article (i.e., a multi-topical measure), I computed H_{ldn22} as a linear interpolation of points where citations are in disciplines other than the discipline of the cited article (i.e., a multi-disciplinary measure), and calculated the G indices for journals (G_j) and disciplines (G_d) by applying the following formulas:

$$G_j = [1 / (2 \times 0.5\, N^2)\, \Sigma_i\, \Sigma_k\, |j_i - j_k|]/[(N-1)/N]$$
$$G_d = [1 / (2 \times 0.5\, N^2)\, \Sigma_i\, \Sigma_k\, |d_i - d_k||]/[(N-1)/N]$$

Where N is the total number of articles in the author's career, and $(N-1)/N$ is the maximum value for both G_d (i.e., an inter-disciplinary measure) and G_j (i.e., an inter-topical measure); j represents the journal title; $j_i - j_k = 0$ if articles i and k appear in the same journal and $j_i - j_k = 1$ otherwise; d represents the discipline name; and $d_i - d_k = 0$ if articles i and k belong to the same discipline and $d_i - d_k = 1$ otherwise. See Egghe (2013a) for an analysis of H indices based on citations by different citers.

Note that this classification cannot be criticised as ambiguous (i.e., either the journal is the same or is different), although it could be disapproved because it overestimates a CV differentiation (e.g., an heterodox Post-Keynesian economist publishes in very few journals such as Cambridge Journal of Economics or Journal of Post Keynesian Economics or Review of Political Economy, but not in a single journal). However, heterogeneity of CVs will be estimated by comparing percentages.

Figures 5, 6 and 7 are based on all H values calculated for the *first, second and third* analysed CV, respectively, using the indices developed in this paper. Moreover, I define a CV as (intra-disciplinary and intra-topical) heterodox (e.g., Figure 5) if G_d is 0 (i.e., the author publishes in a single discipline), G_j is small (i.e., the author publishes in few journals), and H_{ljn22} and H_{ldn22} are equal and small (i.e., the vast majority of citations are in the same journals). In contrast, I define a CV as (intra-disciplinary and inter-topical) orthodox (e.g., Figure 7) if G_d is small (i.e., the author publishes in few disciplines), G_j is large (i.e., the author publishes in many journals), H_{ljn22} is considerably larger than H_{ldn22}, and H_{ldn22} is small (i.e., the vast majority of citations are in the same discipline). Finally, I define a CV as (inter-disciplinary and inter-topical) orthodox (e.g., Figure 6) if G_d is large (i.e., the author publishes in many disciplines), G_j is large (i.e., the author publishes in many journals), H_{ljn22} is slightly larger than H_{ldn22}, and H_{ldn22} is large (i.e., the vast majority of citations are in different disciplines). In other words, an orthodox CV can be either intra- or inter-disciplinary.

Figure 5. Characterisation of a representative intra-disciplinary and intra-topical heterodox CV. Black area (0.9%) = networking; red area (5.8%) = heterodox; yellow area (30.1%) = intra-disciplinary orthodox; blue area (0%) = inter-disciplinary orthodox. G_d = 0, G_j = 0.33, H_{ldn22} = 2.11, H_{ljn22} = 9.04, H_{ltn22} = 9.91, H_{ltg22} = 10.00.

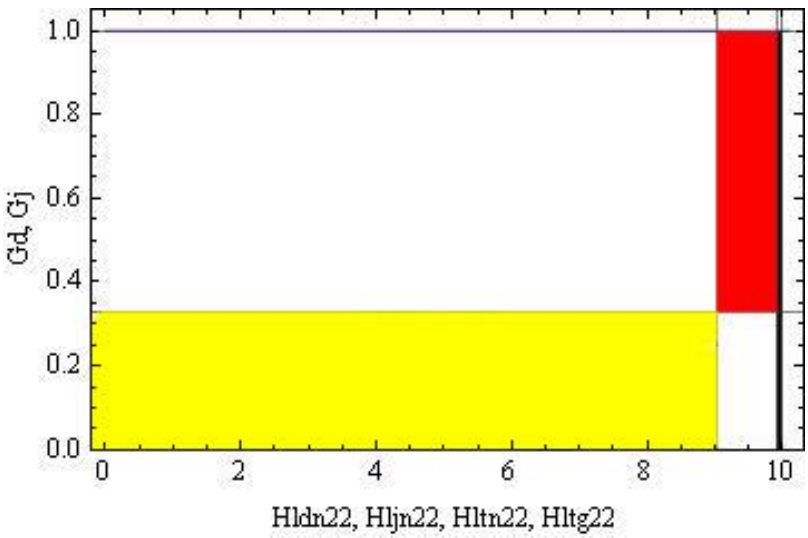


Figure 6. Characterisation of a representative inter-disciplinary and inter-topical orthodox CV. Black area (8.6%) = networking; red area (0.3%) = heterodox; yellow area (3.3%) = intra-disciplinary orthodox; blue area (48.9%) = inter-disciplinary orthodox. G_d = 0.72, G_j = 0.97, H_{ldn22} = 4.71, H_{ljn22} = 5.74, H_{ltn22} = 6.29, H_{ltg22} = 7.34.

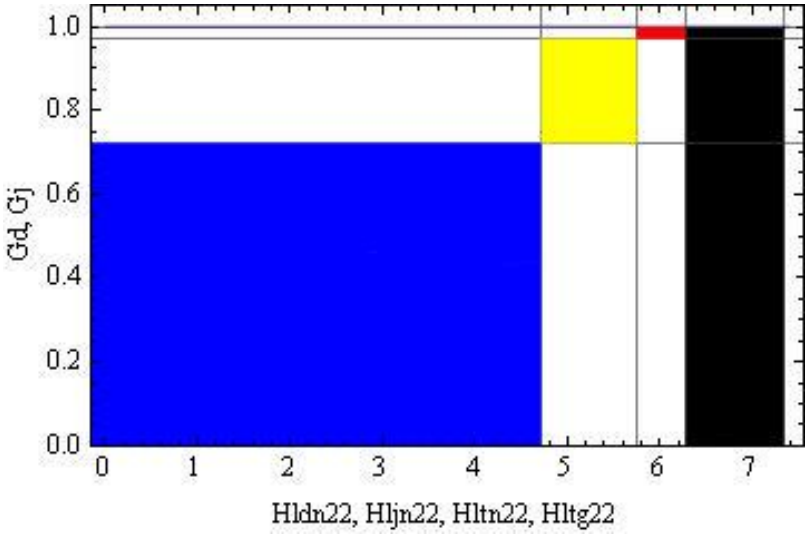
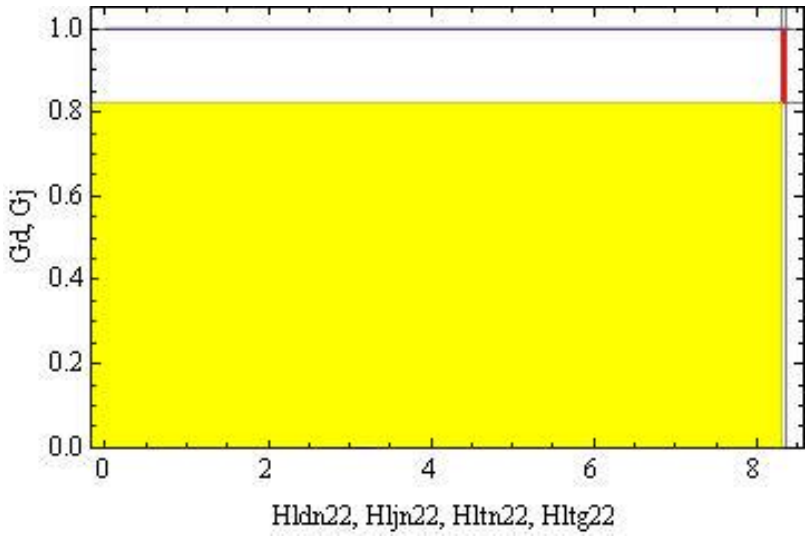


Figure 7. Characterisation of a representative intra-disciplinary and inter-topical orthodox CV. Black area (0%) = networking; red area (0.1%) = heterodox; yellow area (81.9%) = intra-disciplinary orthodox; blue area (0%) = inter-disciplinary orthodox. G_d = 0, G_j = 0.82, H_{ldn22} = 8.19, H_{ljn22} = 8.33, H_{ltn22} = 8.34, H_{ltg22} = 8.34.



Note that in Figures 5 and 7, H_{ldn22} is greater than 0. Indeed, in a heterodox CV, H_{ljn22} = 0 only if each article is cited by an article in the same journal, whereas a more likely citation by an article in a different journal from a small group of journals is excluded. Similarly, in an intra-disciplinary orthodox CV, H_{ldn22} = 0 only if each article is cited by an article in the same discipline, whereas a less likely citation by an article in a different discipline is excluded.

Third, I define a CV as *unintentional* inter-disciplinary or inter-topical CV if the author publishes in few disciplines or journals, but is nonetheless cited by many different disciplines or journals. This is represented by a decrease in G_d and G_j for a given H_{ldn22} and H_{ljn22} (i.e., a decrease in the inequality of publications for a given number of citations). Moreover, I define a CV as an *unsuccessful* intra-disciplinary and intra-topical CV if the author publishes in few journals, and is also cited by few different journals. I define a CV as an *unsuccessful* intra-disciplinary and inter-topical CV, if the author publishes in many journals, but is nonetheless cited by few different journals. This is represented by a decrease of H_{ldn22} and H_{ljn22} for a given G_d and G_j (i.e., a decrease in the number of citations at a given inequality of publications). This case could depict *an intra-discipline reputation* if G_j is large while H_{ljn22} is small; that is, the author publishes in many journals *possibly* because

editors expect many citations of papers in their journal and consequently an increase in its impact factor, but this does not happen because papers are published without suitable scrutiny to ensure their quality. Similarly, I define an *unsuccessful* inter-disciplinary and inter-topical CV, if the author publishes in many disciplines, but it is cited by journals in a few different disciplines. This is represented by a reduction of H_{ldn22} and H_{ljn22} for a given G_d and G_j (i.e., a decrease in the number of citations at a given inequality of publications). This case could depict *an intra-topical reputation* if G_d is large while H_{ldn22} is small (i.e., the author publishes in a few journals *possibly* because the author knows the editors). Finally, I calculated the areas as percentages using the following equations:

$$\text{Black area (networking)} = [(H_{ltg22} - H_{ltn22})] / (H_{ltg22})$$
$$\text{Red area (intra-disciplinary intra-topical heterodoxy)} = [(H_{ltn22} - H_{ljn22}) \times (1- G_j)] / (H_{ltn22})$$
$$\text{Yellow area (intra-disciplinary inter-topical orthodoxy)} = [(H_{ljn22} - H_{ldn22}) \times (G_j - G_d)] / (H_{ltn22})$$
$$\text{Blue area (inter-disciplinary inter-topical orthodoxy)} = [H_{ldn22} \times G_d] / (H_{ltn22})$$

Where the percentages do not sum to 100%, this is because the analysis only considered intentional and successful inter-topical and inter-disciplinary features.

In the numerical example based on the *first* analysed CV (Table A2 in the Appendix), $G_j = 0.33$ and $G_d = 0$, $H_{ljn22} = 9.04$ (i.e., 91% of the overall H_{ltn22}), $H_{ldn22} = 2.11$ (i.e., 21% of the overall H_{ltn22}), and the blue area is 0%, the yellow area is 23.0%, the red area is 5.8%, and the black area is 0.9%. These figures show that the *first* analysed CV is uni-disciplinary (large G_d), heterodox (small G_j), and intra-disciplinary (small H_{ldn22}), and intra-topical (H_{ldn22} is distant from H_{ljn22}).

In the numerical example based on the *second* analysed CV (Table A3 in the Appendix), $G_j = 0.97$ and $G_d = 0.72$, $H_{ljn22} = 5.74$ (i.e., 91% of the overall H_{ltn22}), $H_{ldn22} = 4.71$ (i.e., 75% of the overall H_{ltn22}), and the blue area is 53.9%, the yellow area is 4.1%, the red area is 0.3%, and the black area is 14.3%. These figures show that the *second* analysed CV is multi-disciplinary (large G_d), multi-topical (large G_j), and inter-disciplinary (large H_{ldn22}), but not inter-topical (H_{ldn22} is close to H_{ljn22}). Its inter-topical characteristic is accounted for by its inter-disciplinary characteristic. It is also non-heterodox (H_{ljn22} is close to H_{ltn22}).

In the numerical example based on the *third* analysed CV (Table A4 in the Appendix), $G_j = 0.82$ and $G_d = 0$, $H_{ljn22} = 8.33$ (i.e., 99% of the overall H_{ltn22}), $H_{ldn22} = 8.19$ (i.e., 98% of the overall H_{ltn22}), and the blue area is 0%, the yellow area is 81.9%, the red area is 0%, and the black area is 0%. These figures show that the *third* analysed CV is uni-disciplinary (G_d at 0), multi-topical (large G_j), but not inter-topical (H_{ldn22} is close to H_{ljn22}). It is also non-heterodox (H_{ljn22} is close to H_{ltn22}).

In other words, although one could rank researchers according to H_{ltn22} or H_{cyn10}, as suggested in the previous section, one should *also* characterise each CV by comparing the areas of the four colours or calculating the ratios based on the areas of the four colours. In particular, if an organization is interested in encouraging its authors to develop an inter-disciplinary CV, it should look for CVs with large blue areas, whereas if it is more interested in heterodox CVs, it should look for CVs with large red areas. This could reduce the risk of discrimination against heterodox or inter-disciplinary CVs, for which smaller values of H_{ltn22} are likely to be observed, by introducing some form of compensation (e.g., a smaller H_{ltn22} with large red or blue areas could be preferred to a larger H_{ltn22} with small red or blue areas). This is why only intentional and successful inter-disciplinary and heterodoxy criteria are considered in this study.

6. Discussion

Although comparisons across disciplines are crucial for rewarding the successful careers of senior researchers, they might be irrelevant for the recruitment of junior researchers, since these authors are more likely to compete for a position within a given discipline. That is, if the goal is to recruit a good scientist in a given discipline, you do not need to compare them with scientists in other disciplines, because (for example) only economists will apply for a position in economics.

However, senior and junior researchers should be ranked according to their scientific production and productivity (i.e., H_{ltn22} and H_{cyn10}), respectively, which requires disentangling of *networking* from research activity (i.e., H_{ltg22} vs. H_{ltn22}) so that these activities can be separately and positively evaluated. The research activity should then be evaluated by potentially favouring or contrasting CVs according to basic *intentional* characteristics such as heterodox vs. orthodox publications and intra- vs. inter-discipline publications (as proportions), and these should be distinguished from *unsuccessful* or *unintentional* characteristics.

Section 4 suggested the use of alternative nested *H* indices based on linear and cubic interpolation of normalized numbers of publications and citations, whereas Section 5 presented two-dimensional graphs based on alternative nested *H* and *G* indices. This would reduce incentives to engage in tactical or opportunistic behaviours in publication and citation by authors and journal editors, and should reduce discrimination against heterodox and inter-disciplinary CVs that would be characterized by few citations and few publications. Table 2 summarizes suggested warning indicators that could be used to identify potentially questionable practices by editors and authors [2].

In other words, the suggested methodology provides support for organizations that are interested in supporting networking and favouring orthodox and intra-disciplinary researchers. Indeed, departments are usually ranked according to publications of its members in few and specialized journals.

Table 2. Summary of potentially questionable practices by editors and authors: sources, behaviour types, observations, and warning indicators.

	Single author	Many authors
Single editor	Intra-journal personal relationship Opportunistic behaviour by the editor and the author Many publications in the same journal (Tall but not wide red area)	Bargaining power of the editor Tactical behaviour by the editor Many citations in the same journal (Large red area, but orthodox topics)
Many editors	Inter-disciplinary reputation of the author Tactical behaviour by editors Many publications in many journals (Tall but not wide yellow area)	
Single author	Personal relationship Opportunistic behaviour by authors Many citations in many journals (Large black area)	

The main *strengths* of the new approach are that:

- All main proposals for modifying the original *H* index have been included (Dienes, 2015), including the elimination of self- and cross-citations, an increased weighting to highly cited articles, a focus on peer-reviewed scientific journals, the use of fractional citations to account for the number of authors (i.e., awarding authors a fraction of a point instead of a full point for multi-author articles), an increased sensitivity to variability of the overall citation profile, and the consideration of the life cycle of an article.
- Discrimination against inter-disciplinary and heterodox CVs can be reduced by mitigating the bias created by conventional rankings, without relying on the application of advanced methodologies to complex datasets, as in the case of applying empirically based scaling factors to different disciplines (Ruocco & Daraio, 2013), comparisons with the performance of other researchers in the same field (Nair, 2015), or comparison with the average number of citations per paper in a given discipline (Radicchi & Castellano, 2012)
- All main questions left open by the original description of the *H* index have been tackled (Gagolewski, 2013), including the attribution of an article to a given discipline, since this is

[2] It's important to note that if an indicator suggests the *possibility* of questionable behaviour, this does not indicate the *certainty* of such behaviour. Instead, the actual publications and citations should be carefully examined; there are many legitimate reasons for publishing many papers in the same journal (e.g., because it has the most suitable audience for a research result), for citing a colleague's work (e.g., because that work is more relevant to the author's paper), and so on.

done by the author. This is done while retaining the practicality and simplicity that made the original *H* metric so appealing to a large audience.

- Indicators are distinguished according to the goals being pursued by amending well-established procedures such as years from publication rather than academic age (i.e., the duration of a researcher's career at the time of the analysis; Egghe, 2013b), and they can be applied at different levels of aggregation (e.g., at department or university levels).
- Indicators are based on information that is available at an individual level, including citations that would be disregarded by the original *H* index (Anania & Caruso, 2013), and they can be easily computed.
- Rankings can also be obtained when the publication period is prior to the citation period under consideration (e.g., neglecting citations older than 22 years rather than articles published more than 22 years ago). Indeed, Albert Einstein was chosen as a reference third CV to show how this feature of the proposed model works, though also because his scientific career is among the most remarkable of the last century (i.e., it represents an absolute maximum) (Simkin and Roychowdhury, 2013), and consistent with the information set required for the application of this new ranking approach (i.e., published articles and recorded citations).

The main *weaknesses* of this approach are that:

- Results depend on the dataset used, and many alternatives could be applied (Harzing & Alakangas, 2016). However, the Scopus dataset for the last 22 years is both authoritative and comprehensive, and the same criticism can be raised for other datasets.
- The focus is on past (retrospective) real performance rather than on future expected (prospective) performance (Carrasco & Ruiz-Castillo, 2014; Chang et al., 2011), and it does not account for the life cycles of articles. However, using impact factors would require a reliance on debatable information, such as the 2-year vs. 5-year impact factors described by Sangwal (2013), from a dispersed and always in-progress dataset, as in the case of the temporal evolution of impact factors that is discussed by Finardi (2013). In addition, there are potentially opposite interpretations. For example, the presence of few citations in journals with a high impact factor could be a negative feature, because it would represent the lack of ability to exploit an important audience.
- Insights are not based on axiomatization, in which many alternatives could be suggested (Kongo, 2014). However, the formulas are easy to implement and straightforward to interpret.
- Characterization of CVs depended on the simplifying assumption that a journal could not belong to two or more disciplines (Radicchi & Castellano, 2012). Although factor analysis could be used to univocally sort journals into single hypothetical disciplines in terms of estimated correlations, this is unrealistic in practice because researchers may be unable to perform this analysis without support from suitable software. However, accounting for multidisciplinary journals remains a challenge for future research

7. Conclusions

To allow cross-disciplinary comparisons, traditional bibliometric indices of a researcher's publication history must be modified. In the present study, I modified one of these indices (Hirsch's *H*) by accounting for the feasible normalizations that have been suggested in the literature. To allocate public funds among researchers, scientific activity must be prioritized, and potentially questionable behaviours by authors and journal editors must be identified so that it can be accounted for in researcher evaluations. In the present analysis, I purified the traditional *H* index by eliminating information other than scientific activity. To be widely used, bibliometric indices must be easily calculated and highly relevant to the goals of the organization that is using them to evaluate researchers. The present study used information that is common to any bibliometric dataset to solve problems that have affected many previous bibliometric indices. Because ongoing updates would be required as new papers by an author are published as well as new citations of published articles are recorded, software should be developed to help authors and their managers rapidly

recalculate the revised indices described in this paper, as was noted by Bornmann & Ozimek (2012). In other words, this study provides a simple methodology based on insights from the literature that allows researchers and their managers to easily characterize a researcher's CV. I do not use the adjective *fair* to describe this methodology (unlike Gagolewski, 2013), since that implies a value scale, but would instead describe it as a way to define the *suitability* for a given goal, which is a more objective criterion.

Needless to say, the findings presented in this paper require further investigation. In particular, an econometric analysis based on a sufficiently large sample seems to be essential, both in terms of time (i.e., at least 10 years to consider the life cycle of articles) and in terms of authors (i.e., at least 10 000 CVs from all 27 disciplines to adequately represent the 10 000 000 scientists who have published at least one English article in the last 10 years that were included in Scopus). This would let us test whether the proposed methodology solves or mitigates the problems that I have highlighted with previous bibliometric indices. I am currently working with colleagues to design and perform such a test. In contrast, a statistical analysis based on a similarly large sample could be used to characterise the average author, but this seems unnecessary; the micro-scale approach adopted in this study does not need to be scaled up.

Acknowledgments

I would like to thank Alberto Zigoni, Senior Consultant, Research Intelligence, Elsevier, for providing summary statistics of the Scopus dataset.

Appendix

Table A1. Summary of the *H* indices used in this study, and the associated calculated values for the *second* analysed CV.

All authors (24 articles)		Production per author (18.6 articles)		Productivity per author per year (18.6 articles)
H_{latg22} = 8.42	→	H_{ltg22} = 7.34		
H_{catg22} = 6.84	→	H_{ctg22} = 5.75		
↓		↓		
H_{latn22} = 7.62	→	H_{ltn22} = 6.29, H_{ltn10} = 6.02	→	H_{lyn22} = 2.43, H_{lyn10} = 2.56
H_{catn22} = 6.67	→	H_{ctn22} = 5.28, H_{ctn10} = 4.93	→	H_{cyn22} = 2.60, H_{cyn10} = 2.57
		↓		
		H_{ljn22} = 5.74		
		H_{cjn22} = 4.97		
		↓		
		H_{ldn22} = 4.71		
		H_{cdn22} = 4.31		

Notation: a = all authors; c = cubic interpolation; d = citations are in disciplines other than the discipline that published the cited article; g = gross number of citations; j = citations are in journals other than the journal that published the cited article; l = linear interpolation; n = net number of citations; t = total articles; y = citations are divided by the number of years since publication; 22 = author's whole career (based on the consistent duration from 1995 to the present defined for the Scopus dataset); 10 = the last 10 years. Both the gross and the net citation *H* indices calculated based on data from www.scopus.com equalled 6.

Table A2. Descriptive statistics and Scopus categories for the *first* analysed CV.

	Year	Disc	A.N	Number of citations			
				Gross	Net		
				Tot	Tot	By ≠ journals	By ≠ disciplines
Cambridge Journal of Economics	2012	Eco	1	7	7	5	0
Cambridge Journal of Economics	2012	Eco	1	7	7	5	0
Cambridge Journal of Economics	2005	Eco	1	36	36	29	1
Journal of Post-Keynesian Economics	2001	Eco	1	6	6	4	0
Cambridge Journal of Economics	1994	Eco	1	1	1	1	0
Structural Change and Economic Dynamics	1990	Eco	1	2	2	1	0
Cambridge Journal of Economics	1989	Eco	1	19	19	8	2
Cambridge Journal of Economics	1989	Eco	1	0	0	0	0
Cambridge Journal of Economics	1988	Eco	1	38	35	33	5
Cambridge Journal of Economics	1988	Eco	1	12	11	7	3
Cambridge Journal of Economics	1986	Eco	1	0	0	0	0
Cambridge Journal of Economics	1983	Eco	1	9	9	6	1
Review of Economic Studies	1981	Eco	1	1	1	1	0
Cambridge Journal of Economics	1977	Eco	1	11	11	8	0
Quarterly Journal of Economics	1966	Eco	1	45	43	35	1
Review of Economic Studies	1964	Eco	1	0	0	0	0
Oxford Economic Papers	1960	Eco	1	7	7	5	0
	Total			201	195	148	13

Notation: Disc = Discipline; A.N = Number of authors; Eco = Economics.

Table A3. Descriptive statistics and Scopus categories for the *second* analysed CV.

	Year	Disc	A.N	Number of citations			
				Gross	Net		
				Tot	Tot	By ≠ journals	By ≠ disciplines
Applied Mathematical Modelling	**2016**	Mat	1	0	0	0	0
Applied Soft Computing Journal	**2016**	Com	1	3	1	0	0
Environmental & Resource Economics	**2016**	*Eco*	1	0	0	0	0
Science of the Total Environment	**2016**	Env	1	6	4	1	1
Sustainability (Switzerland)	**2016**	Env	1	1	1	0	0
Sustainability (Switzerland)	**2016**	Env	17	2	2	1	0
Sustainability (Switzerland)	**2016**	Env	21	0	0	0	0
Journal of Happiness Studies	**2015**	*Hum*	1	0	0	0	0
Sustainability (Switzerland)	**2015**	Env	16	4	4	2	1
Sustainability Science	**2015**	Env	1	1	0	0	0
Coastal Engineering	**2014**	Eng	8	27	16	8	4
Journal of Hydrology	**2014**	Env	2	0	0	0	0
Environmental Modelling and Assessment	**2013**	Env	1	0	0	0	0
Environmental Modelling and Software	**2013**	Com	2	11	11	6	6
Natural Hazards	**2013**	Env	1	2	2	2	2
Environmental Management	**2011**	Env	1	4	3	3	3
Journal of Happiness Studies	**2011**	*Hum*	1	3	2	2	2
Water Resources Management	**2010**	Env	1	15	15	13	11
International Journal of Hospitality Management	**2009**	*Man*	1	16	16	14	12
Journal of Environmental Management	**2008**	Env	1	8	7	6	6
Papers in Regional Science	2003	*Eco*	3	2	1	1	1
Environment and Development Economics	1998	*Eco*	1	1	1	1	1
Journal of Environmental Economics and Management	1998	*Eco*	1	15	15	14	2
Economic Journal	1995	*Eco*	1	1	1	1	1
	Total			122	102	75	53

Notation: Disc = Discipline; A.N = Number of authors; Com = Computer Sciences; Eco = Economics, Econometrics and Finance; Eng = Engineering; Env = Environmental Sciences; Hum = Arts & Humanities; Man = Business, Management & Accounting; Mat = Mathematics. Boldfaced values represent the most recent 10 years. Disciplines in italics are in the social sciences subject, whereas other disciplines are in the physical sciences subject.

Table A4. Descriptive statistics and Scopus categories for the *third* analysed CV.

	Year	Disc	A.N	Number of citations			
				Gross	Net		
				Tot	Tot	By ≠ journals	By ≠ disciplines
Physical Review	1953	Phy	1	7	7	7	2
Science	1951	Phy	1	4	4	4	1
Science	1949	Phy	8	1	1	1	0
Reviews of Modern Physics	1948	Phy	1	56	56	56	13
Reviews of Modern Physics	1946	Phy	2	85	85	84	12
Reviews of Modern Physics	1945	Phy	2	249	249	244	75
Science	1940	Phy	1	27	27	27	17
Journal of the Franklin Institute	1937	Phy	2	254	251	246	76
Science	1936	Phy	1	305	305	299	92
Journal of the Franklin Institute	1936	Phy	1	101	101	99	31
Physical Review	1936	Phy	2	29	29	28	9
Physical Review	1935	Phy	3	6806	6805	6663	2058
Physical Review	1935	Phy	2	319	318	311	96
Physical Review	1931	Phy	3	28	28	26	6
Nature	1923	Phy	1	10	10	10	3
Nature	1921	Phy	1	10	10	10	3
	Total			8291	8286	8115	2495

Notation: Disc = Discipline; A.N = Number of authors; Phy = Physics.

References

Abatemarco, A., Dell'Anno, R. (2013) A bibliometric evaluation of the research outputs of Italian economists, Economia Politica 30: 97-125

Abbas, A.M. (2012) Bounds and inequalities relating H-index, g-index e-index and generalized impact factor: an improvement over existing models, PLoS ONE 7: e33699

Abramo, G. et al. (2013a) Assessing the accuracy of the h- and g-indexes for measuring researchers' productivity, Journal of the Association for Information Science and Technology 64: 1224-1234

Abramo, G. et al. (2013b) The importance of accounting of the number of co-authors and their order when assessing research performance at the individual level in life sciences, Journal of Informetrics 7: 198-208

Abramo, G. et al. (2013c) The suitability of h and g indexes for measuring the research performance of institutions, Scientometrics 97: 555-570

Adachi, T., Kongo, T. (2015) Further axiomatizations of Egghe's g-index, Journal of Informetrics 9: 839-844

Anania, G., Caruso, A. (2013) Two simple new bibliometric indexes to better evaluate research in disciplines where publications typically receive less citations, Scientometrics 96: 617-631

Bartolucci (2015) A comparison between the g-index and the H-index based on concentration, Journal of the Association for Information Science and Technology 66: 2708-2710

Bertoli-Barsotti (2016) Normalizing the g-index, Scientometrics 106: 645-655

Bertoli-Barsotti, L., Lando, L. (2015) On a formula of the H-index, Journal of Informetrics 9: 762-776

Blagus, R. et al. (2015) Comparisons of bibliometric measures for assessing relative importance of researchers, Scientometrics 105: 1743-1762

Bornmann, L. et al. (2011) A multi-level meta-analysis of studies reporting correlations between the H-index and 37 different H-index variants, Journal of Informetrics 5: 346-359

Bornmann, L. et al. (2012) The new excellence indicator in the world report of the SCImago institutions ranking 2011, Journal of Informetrics 6: 333-335

Bornmann, L., Marx, W. (2015) Methods for the generation of normalized citation impact scores in bibliometrics: which method best reflects the judgements of experts?, Journal of Informetrics 9: 408-418

Bornmann, L., Ozimek, A. (2012) Stata commands for importing bibliometric data and processing author address information, Journal of Informetrics 6: 502-512

Bouyssou, D. & Marchant, T. (2016a) An axiomatic approach to bibliometric rankings and indices, Journal of Informetrics 8: 449-477

Bouyssou, D. & Marchant, T. (2016b) Ranking authors using fractional counting of citations: an axiomatic approach, Journal of Informetrics 10: 183-199

Cainelli, G. et al. (2012) Co-authorship and productivity among Italian economists, Applied Economic Letters 19: 1609-1613

Cainelli, G. et al. (2015) The strength of strong ties: How co-authorship affect productivity of academic economists? Scientometrics 102: 673-699

Carrasco, R., Ruiz-Castillo, J. (2014) The evolution of scientific productivity of highly productive economists, Economic Enquiry 52: 1-16

Chang, C.-L. et al. (2011) What makes a great journal great in sciences? Which came first, the chicken or the egg? Scientometrics 87: 17-40

De-Moya-Anegon, F. et al. (2014) How to interpret the position of private sector institutions in bibliometric rankings of research institutions, Scientometrics 98: 283-298

De Stefano, D. et al. (2013) The use of different data sources in the analysis of co-authorship networks and scientific performance, Social Networks 35: 370-381

De Stefano, D., Zacccarin, S. (2016) Co-authorship networks and scientific performance: an empirical analysis using the generalised extreme value distribution, Journal of Applied Statistics 43: 262-279

De Visscher, A. (2011) What does the g-index really measure?, Journal of the Association for Information Science and Technology 64: 2290-2293

Dienes, K.R. (2015) Completing h, Journal of Informetrics 9: 385-397

Egghe, L. (2013a) A rationale for the relation between the citer H-index and the classical H-index of a researcher, Scientometrics 94: 873-876

Egghe, L. (2013b) On the correlation of the H-index for career length, Scientometrics 96: 563-571

Engels, T.C.E. et al. (2013) Group size, H-index, and efficiency in publishing in top journals explain expert panel assessments of research group quality and productivity, Research Evaluation 22: 224-236

Finardi, U. (2013) Correlation between journal impact factor and citation performance: an experimental study, Journal of Informetrics 7: 357-370

Gagolewski, M. (2013) Scientific impact assessment cannot be fair, Journal of Informetrics 7: 792-802

Gagolewski, M., Mesiar, R. (2012) Aggregating different paper quality measures with a generalized H-index, Journal of Informetrics 5: 566-579

Glanzel, W. (2012) The role of core documents in bibliometric network analysis and their relation with h-type indices, Scientometrics 93: 113-123

Harzing, A.-W., Alakangas, S. (2016) Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison, Scientometrics 106: 787-804

Harzing, A.-W., et al. (2014) hIa: an individual annual H-index to accommodate disciplinary and career length difference, Scientometrics 99: 811-821

Hopkins, A.L. et al. (2013) Disparities in publication patterns by gender, race and ethnicity based on a survey of a random sample of authors, Scientometrics 96: 515-534

Ko, Y.M., Park, J.Y. (2013) An index for evaluating journals in a small domestic citation index database whose citation rate is generally low: a test based on the Korea Citation Index (KCI) database, Journal of Informetrics 7: 404-411

Kongo, T. (2014) An alternative axiomatization of the Hirsch index, Journal of Informetrics 8: 252-258

Liu, X.Z, Fang, H. (2012) Fairly sharing the credit of multi-authored papers and its application in the modification of h- index and g-index, Scientometrics 91: 37-49

Liu, Y., Yang, Y. (2014) Empirical study of L-sequence: the basic H-index sequence for cumulative publications with consideration of the yearly citation performance, Journal of Informetrics 8: 478-485

Martinez, M.A. et al. (2014) H-classics: characterising the concept of citation classics through H-index, Scientometrics 98: 1971-1983

McCarty, C., Jawitz, J.W. (2013) Attitudes about publishing and normal science advancement, Journal of Informetrics 7: 850-858

Morichika, N., Shibayama, S. (2015) Impact of inbreeding on scientific productivity: a case study of a Japanese university department, Research Evaluation 24: 146-157

Moskovkin, V.M. et al. (2014) Construction if scoring rule within the framework of new generation of metric citations, Research Journal of Applied Sciences 9: 1167-1170

Nair, M.S. (2015) M-score: A context-specific score to assess scientific productivity with OEVGSI grading, Current Science 109: 596-600

Penner, O. et al. (2013) On the predictability of future impact in science, Scientific Reports 3: 3052

Petersen, A.M. (2015) Quantifying the impact of weak, strong, and super ties in scientific careers, PNAS 10: E4671-E4680

Petersen, A.M., Succi, S. (2013) The Z-index: a geometric representation of productivity and impact accounts for information in the entire rank-citation profile, Journal of Informetrics 7: 823-832

Prathap, G. (2012) The quality-quantity-quasity and energy-exergy-entropy exegesis of expected value calculation of citation performance, Scientometrics 91: 269-275
Prathap, G. (2014) Measures of impact, consistency, and the h- and g-indexes, Journal of the Association for Information Science and Technology 65: 1076-1078
Radicchi, F., Castellano, C. (2012) Testing the fairness of citation indicators for comparison across scientific domains: the case of fractional citation counts, Journal of Informetrics 6: 121-130
Rousseau, R. (2015) Egghe's g-index is not a proper concentration measure, Journal of the Association for Information Science and Technology 66: 1518-1519
Ruocco, G., Daraio, C. (2013) An empirical approach to compare the performance of heterogeneous academic fields, Scientometrics 97: 601-625
Ryan, J.C. (2016) A validation of the individual annual h-index (hIa): application of the hIa to a qualitatively and quantitatively different sample, Scientometrics 109: 577-590
Sangwal, K. (2013) Citation and impact factor distributions of scientific journals published in individual countries, Journal of Informetrics 7: 487-504
Schreiber, M. (2013a) Do we need the g-index?, Journal of the Association for Information Science and Technology 64: 2396-2399
Schreiber, M. (2013b) How to derive an advantage from the arbitrariness of the g-index, Journal of Informetrics 7: 555-561
Schreiber, M. (2014a) How to improve the outcome of performance evaluations in terms of percentiles for citation frequencies of my papers, Journal of Informetrics 8: 873-879
Schreiber, M. (2014b Is the new citation-rank approach P100' in bibliometrics really new? Journal of Informetrics 8: 997-1004
Schreiber, M. (2014c) Examples of counterintuitive behavior of the new citation-rank indicator P100 for bibliometric evaluations, Journal of Informetrics 8: 738-748
Schreiber, M. et al. (2011) Categorizing H-index variants, Research Evaluation 20: 397-409
Schubert, A. (2012) A Hirsch-type index of co-author partnership ability, Scientometrics 91: 303-308
Senanayake, U. et al. (2015) The pagerank-index: going beyond citation counts in quantifying scientific impact of researchers, PLoS ONE 10: e0134794
Simkin, M.V., Roychowdhury, V.P. (2013) A mathematical theory of fame, Journal of Statistical Physics 151: 319-328
Sotudeh, H., Khoshian, N. (2014) Gender, web presence and scientific productivity in nanoscience and nanotechnology, Scientometrics 99: 717-736
Sutter, D. (2012) Different but equal? On the contribution of dissident economists, American Journal of Economics and Sociology 71: 1143-1156
Thelwall, M., Wilson, P. (2014) Regression for citation data: an evaluation of different methods, Journal of Informetrics 8: 738-748
Tsai, C.-F. (2014) Citation impact analysis of top ranked computer science journals and their rankings, Journal of Informetrics 8: 963-971
Wainer, J., Vieira, P. (2013) Correlations between bibliometrics and peer evaluation for all disciplines: the evaluation of Brazilian scientists, Scientometrics 96: 395-410
Wildgaard, L. (2015) A comparison of 17 author-level bibliometric indicators for researchers in astronomy, environmental science, philosophy and public health in Web of Science and Google Scholar, Scientometrics 104: 873-906
Yan, X. et al. (2013) C-index: a weighted network node centrality measure for collaboration competence, Journal of Informetrics 7: 223-239
Yang, J. et al. (2013) A bibliometric analysis of academic publication and NIH funding, Journal of Informetrics 7: 318-324
Yates, E.J., Dixon, L.C. (2015) PageRank as a method to rank biomedical literature by importance, Source Code for Biology and Medicine 10: 16

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it