RIGORS AND ERRORS CONCERNING BIBLIOMETRIC ANALYSIS FOR YOUNG PHD STUDENTS

Dana Stana

University of Pitești, Romania, e-mail: dana.stana@upit.ro

Abstract. *A bibliometric analysis takes into account such criteria as qualitative and academic research, and transferability of the results. The PhD student was compared with the reference group, the activity or event, and the case study method was developed. The quality and validity of the study is however omitted, and the quantitative and qualitative analysis in various growth strategies calls for further research.*

Keywords: *bibliometric, case study, research, PhD student, internationalization*

1. INTRODUCTION

The aim of bibliometric analysis is to understand the mechanism of scientific research. One can get information not from the actual data, but from the relationships between the set of data and their interpretation.

If, in this context, we consider that an analysis is based on quality and academic criteria, we then think of biblioscientometry, where the role is involved of those data sets capable of addressing specific problems. The most widely used indicators are the statistical ones and those of citations. If the former type is obtained based on empirical statistical data, the latter type of indicators concern the relevance and importance of the author who makes the citation and the time that passed from the date of the publication until the moment of citation.

2. METHODS OF MEASURING SCIENTIFIC RESEARCH

Bibliometry deals with the **quantitative** measurement of scientific research, and provides the image of how much influence or interest a certain researcher represents in that field of research. The indicators are calculated by means of bibliographic databases, evidently admitting that bibliometric indicators differ from one database to another, and that some bibliometric indicators are specific to a particular database.

The bibliometric indicators for ***publications*** are:

- **impact factor** (equal to the number of citations of the paper or article published, divided by the number of the citable articles over a period of time);
- **relative influence factor** (represents the influence score of the papers in a journal and the reference influence score of the journal);
- **Index Copernicus Value** (is an international platform specialized in the promotion and scientific results, which allows collaboration between researchers and publishers of scientific journals).

The bibliometric indicators for ***papers / articles*** are:

- **number of citations (i.e. how many times a paper / an article has been cited by another researcher, which characterizes scientific performance);**
- **Hirsch index (or h-index), which represents** the number of papers ***n*** that have *n* or more citations.

In addition to *bibliometry* there are also:

- *almetrics* (how far and how widely spread on the web the content of a paper was);
- *webometrics* (the indicators underlying the volume of web content, and the web visibility and impact – i.e. how many times the link was mentioned on the web).

Scientometry deals with analyzing the **qualitative** aspects of generating, disseminating and using information, and its main target is the contribution to a better understanding of the mechanism of scientific research. To do that, probability calculus and mathematical statistics are used.

As far as the qualitative methods are concerned, the **case study** method is known as the most widespread, followed by **historical research**.

The *case study* type of research is used to answer questions such as: *why?* and *how?*, and build a research theory based on in-depth analysis, highlighting the institutional framework, the applicability scope of the study, the research objectives, the sampling area, the method used, the number of cases, the data source, the theoretical development and the opportunities. The case study methods that are common are the *interview method*, the *observation method* and the *questionnaire method*.

Since 2011 up to the present accredited universities completed a questionnaire (http://chestionar.uefiscdi.ro/public5/index.php?page=punivlist) where the criteria used in the evaluation were mentioned. Under criterion no. 2 (c) relating to scientific research for Standard C1.1, the universities mentioned "Papers indexed as *ISI Web of Knowledge*: Total number of papers published in *ISI Web of Knowledge* by the teaching staff and the scientific researchers who are employed as tenured staff by the university, and also by the persons in a relationship of fixed-term employment contract with the HEI – doctoral students, postdoctoral students, associate academic staff".

From this point of view, we can notice the error having to do with the impact factor for journals, which is essentially an indicator of citations (and so it is by no means fair to be used as an indicator of quality). Unfortunately, researchers, too, have their work assessed in terms of quality with the help of this indicator, which is wrong: we must not equal fame and quality.

The ISI impact factor very well illustrates the impact of the literature, while it cannot indicate the level of scientific quality. On the other hand, an ISI journal does not imply that it is a highly appreciated journal in its field.

A young doctoral student who published five papers / articles that are cited 60 times each cannot be compared with

an elderly teacher with 15 papers / articles to his credit, which are cited 12 times. The former will have an ***h*** factor of 5, while the latter will have 15, which does not mean that the latter is better.

The ***h*** Index highlights researchers who can boast outstanding contributions in their field, and yet have not necessarily earned a reputation in the scientific community, at home or abroad. For instance, Manuel Cardona from the Max Planck Institute for Solid Research in Stuttgart, Germany, has an h index = 86, and Philip Warren Anderson (1977 Nobel Laureate in Physics) h = 91. It should be stressed that the Nobel Prize is awarded for what an individual has achieved rather than what he has published.

Analyzing the ***h*** indicator to assess the results of a researcher, we find that there are both advantages and disadvantages in using it, and so both rigors and errors can be produced rigor concerning the analysis on the extracted data.

Asserting that there are criteria to measure the value and performance of a researcher, Jorge Hirsch comes with the following arguments in favour of using his index:

1. ***total number of papers / articles*** (Na)

Advantage: it measures productivity

Disadvantage: fails to measure the value or impact of the papers / articles

2. ***total number of citations*** (Nc)

Advantage: it measures the total impact

Disadvantage: a) it incorrectly gives priority weight to reviews (overall assessments) over the original contributions in the research articles

b) it is difficult to assess the total impact, due to a "small number of articles with many citations" that are not representative of the researcher, because he/she is coauthors the papers in question with several other researchers

3. ***citations per papers / articles*** (e.g. the ratio of Nc and Na)

Advantage: it allows comparing the output of the researchers at various ages

Disadvantage: it rewards low productivity and penalizes high productivity

4. ***number of "significant papers" defined as number of widely or usually cited papers /articles cited***

Advantage: it eliminates the disadvantages of the criterion

Disadvantage: the threshold for "numerous" citations is arbitrary and it should be adjusted for different age levels

5. ***citations for usually cited papers / articles***

Advantage: it rectifies many of the disadvantages of the criteria

Disadvantage: this is not one number, i.e. the number of most cited papers may favour or disfavor a researcher.

The individual *rankings* are aggregated to give the ranking of the department and of the research team.

It is common knowledge that research was introduced as part of the professional or job evaluation methodology of each university teacher, and it is also common knowledge that there are academic classifications worldwide.

Publication of articles in prestigious scientific journals and recognition of their value through the citations they receive has, of course, both rigors and errors. If publication is the basic criterion in evaluating academics in Romania (by including such an assessment in the minimum standards required and mandatory for conferring teaching titles in higher education), and if this represents the main method of encouraging research, citations are however appraised, in assessment, in a differentiated manner, and in some cases this can be exaggerated, since publishing papers / articles counts more than publishing books. For example, a paper / article indexed in *Thomson* will get 20 points in evaluation, i.e. the same as would be have been given for publishing a critical didactic volume, being the sole author.

Citations are rewarded with 2 additional points compared with the publication of a paper or study in the volume of a national conference, which is evaluated by only 5 points.

In terms of the number of citations designed to determine the ranking of universities, there are two types of rankings:

1. **The Times Higher Education rankings**, with data extracted from the *Thomson Reuters* bibliometric platform
2. **The QS World University rankings**, with data extracted from *Scopus*

The Times Higher Education rankings are based on the following criteria:

1. education – accounted for 30%;
2. international visibility – accounting for 7,5%;
3. revenue and funds from industry attracted by research – accounting for 2,5%;
4. research – with a share of 30%;
5. citations – again with a share of 30% .

The QS World University Rankings analyzes six categories of criteria:

1. academic reputation – accounting for 40%,
2. reputation among employers – accounting for 10%,
3. the ratio of teachers to students – accounting for 20%;
4. citations per faculty according to the Scopus bibliometric platform – accounting for 20%;
5. proportion of international students – accounting for 5%;
6. proportion of international teachers – accounting for 5%.

Also taking account of the fact that bibliometric performance becomes the main focus rather than scientific discovery, scientific work is reduced to *writing* rather than *doing research*. And PhD students are the target in this endeavour, being used to carry it out. Moreover, as collective signature of authors is a wide practice, the teachers tend to misappropriate the work of their PhD students by putting their name on the papers or articles, especially as the "survey" articles dealing with existing literature are cited more than the original productions.

In order to increase **academic productivity**, it is recommended to attend the grand *Polach* of references, because that service will be returned.

However, universities make a big mistake in comparing different areas in an attempt to foresee the publishing results.

The ISI database, which is used as a reference point for the value of journals, does cover most fields of science, yet not all of them. The areas covered are: mathematics, physics, chemistry, computer science, biology, geography, geology, environmental and earth sciences, sciences of education, psychology, economics, business administration, finance, accounting, statistics and economic informatics, economics and international business, management, marketing, agricultural science and forestry, medicine, veterinary medicine, engineering (sciences), physical education and sports, philosophy, history, theology, arts, architecture, urban

planning, sociology, social work, political science, international relations, European studies, administrative studies, sciences of communication. The areas that are not covered are: Romanian philology, Romanian cultural studies, law (except for American law), military science and information science (apart from the technical aspects related to other areas).

To describe a qualitative phenomenon it is hard to designate a quantitative indicator, especially that indicators do not tell us, when it comes to groups, what scientific value a group produces. In order to overcome that challenge three minimal indicators or denominators were found, which have three different meanings:

• *Minimum Presence* (*MP*) is the proportion of people in the group who ever published a full scientific paper present in the ISI database;

• *Minimum Activity* (*MA*) is the proportion of people in the group who published a full scientific paper that is present in the ISI database for the last 5 years;

• *Minimum Visibility* (*MV*) is the proportion of people in the group who make known to the public, on the web, references to their own publications that they deem most important, and which can be found by everyone and read, yet with no limitation concerning their nature and the way in which they were published. The Scientific Efficiency Index, the number of papers from a population compared to the number of the population, is universally accepted to compare the level of scientific activity of nations.

Apart from research, most universities commonly have other purposes, as well:

- *the practical dimension* (meaning the commercial exploitation of their scientific authority in the form of consultancy services);
- *the educational dimension.*

Applying indicators MP and MV to the population of graduates of doctoral programs, immediately or five years after obtaining the PhD degree means an important evaluation. The result shows an elite of minimal scientific performance, who, during say the 2000-2005 period, would be quoted between 0.02% and 0.05%, without however specifying whether that elite lives in the midst of an ignorant or well-educated population.

The analysis of **research performance** for institutions, nations and journals can be conducted by using the ESI indicators (or the Essential Science Indicators), which can make rankings by activity area and can determine the results of research and the impact in specific areas of research (minimum number of citations for a paper or article to position itself in the 1% or 50% top worldwide, by area and by year). The 50% top rankings use technology transfer. In any case, it represents an image of a top of a selected number of journals that do not reflect the relevance of the research as it can be noted that top-ranking journals in the international databases, and the number of articles that disseminate the research results is rather low, even tending to zero. Doctoral researches are in the trend internationally (they are finally published), so they are relevant from that angle, both theoretically and practically, and their relevance becomes useful when users are able to use the resources the research conveys, i.e. the part that is available to them and they can develop if they have it. Hence research is useful when one knows how to use it and when one is able to develop one's ability to use those resources.

The number of conjectures has lately been lower, because one has to achieve the required load indices, even by publishing in related fields.

Fig. 1. Top Papers by Territories

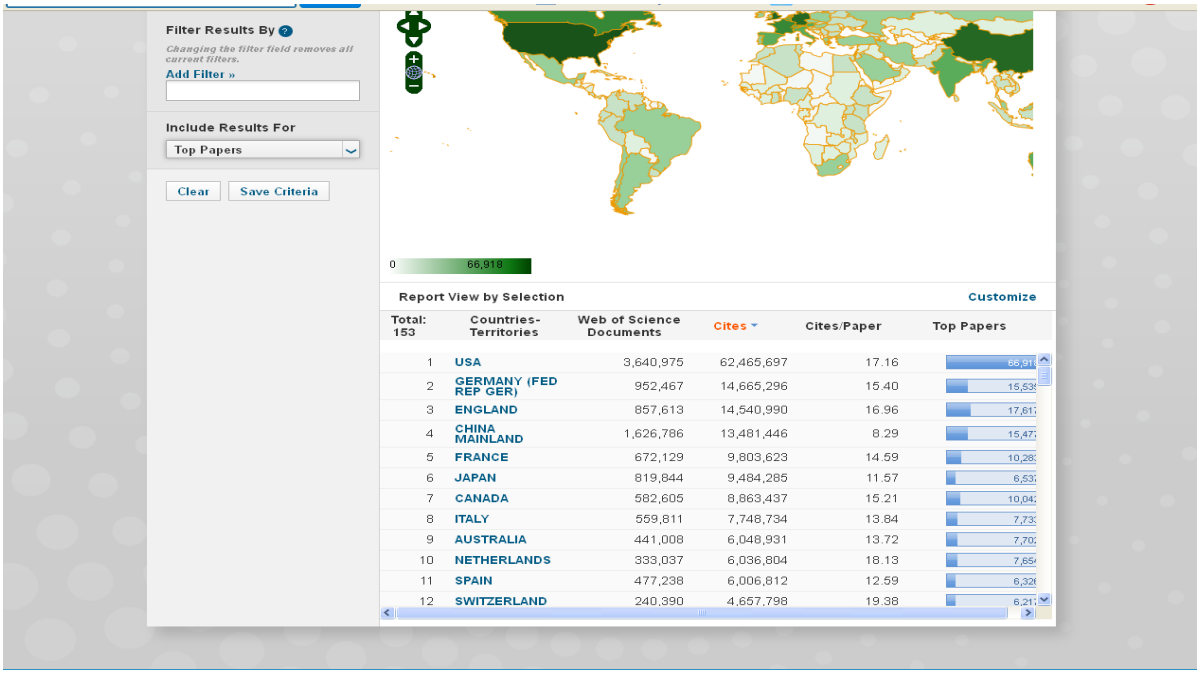

Fig. 2. Top citations by country (1)

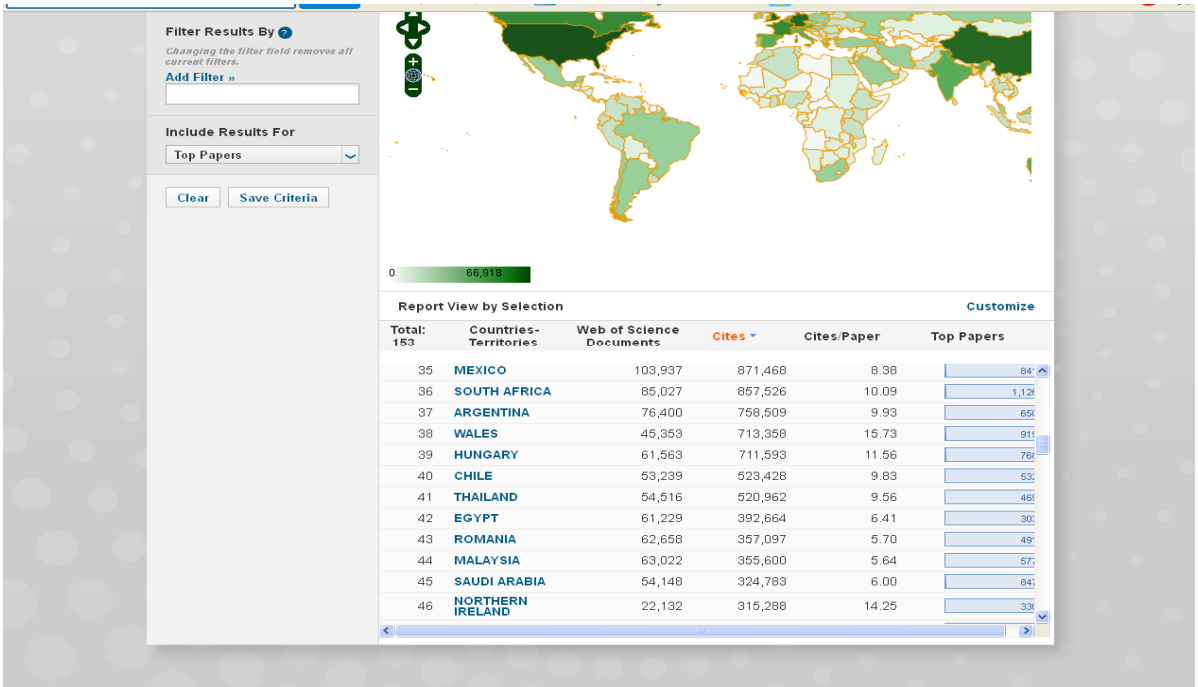

Fig. 3. Top citations by country (2)

Thomson Reuters measures the degree of redundancy (many indexes contain about the same information, only the calculation methods differ). We have indices that are integers, indices that are rational numbers, etc., yet we do not have an index that could represent these factors thus rendering evaluation by one digit.

In the United States no indicator is calculated, instead the application is sent to the three specialists in the domain, who are to analyze the expected promotion.

Asked what would the criteria for access to the Romanian Academy be, researcher Ionel Haiduc said that "that individual must have such a reputation that no additional criteria should be required any longer".

Indeed, how could one expect proper research being conducted when the Romanian state fails to value increased investment in education and research? In step with the process of internationalization and globalization, structural reform is needed in the sphere of science, a modern approach to research, and certainly attracting new investment. Experimental facilities should be at least sufficient. For example, chemistry is, an experimental science *par excellence*, and studies should contain the characterization of the samples via every method, and if there are no proper laboratories, how could one possibly get such results? The solution may lie in this: there are a few research centres in

this country, only there should be collaboration between them.

Finally, by using the *Thomson Reuters* and *Scopus* bibliometric platforms, one can present a comparative analysis of academic evaluation from the angle of an ANELIS user. The presentation below examines the work of a scientist over a certain period of time, in the two databases.

Individual Analysis

I. in *Thomson Reuters* for the period 1990-2015:
- 254 recorded publications
- 3,305 citations, out of which 2,801 are self-citations
- 1,965 citing articles
- 13.01 average of citations / article
- 29 Hirsch Index

II. in *Scopus* for the period 1972-2015
- 691 publications and mentions.

In keeping with the year of publication and the number of citations, the data are highly different, and so we cannot really say which is the better database. If we search for a well-known author, e.g. Nenițescu, in the same databases and for the same periods of time, we will find that the name appears 2 times in *Thomson Reuters* (thus, without taking into account that he is an older author) and 39 times in *Scopus* (in comparison with a PhD student, the work is not to be ignored, as the earlier period is included).

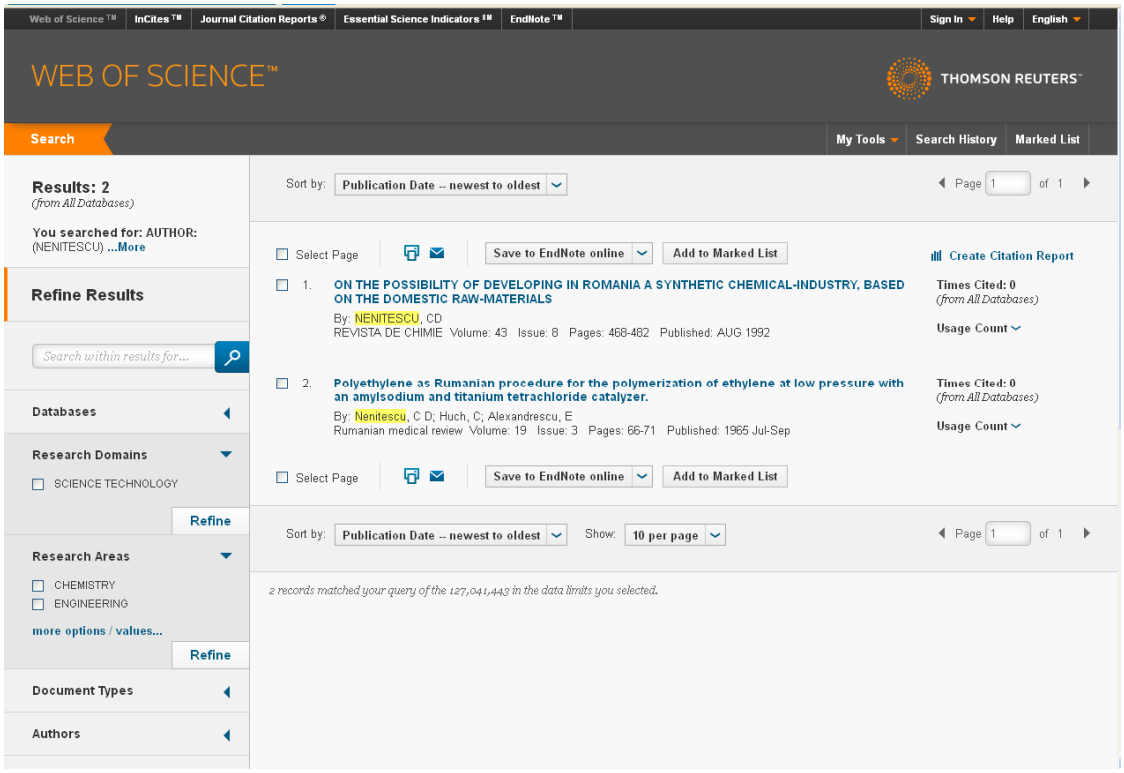

Fig. 4. Search results in Web of Science

Institutional analysis (e.g. Babeş-Bolyai University in Cluj-Napoca, Romania):

In *Thomson Reuters*
- 1,105 mentions (address)
- 56 subject (in the title)

In *Scopus*
- 9,106 mentions (address)
- 65 subject (in the title)

Thematic analysis "crown ether", which won the Nobel Prize
- Thomson Reuters – 30,850 mentions
- Scopus – 58,386 mentions

As a modern instrument for impact analysis concerning researchers, one can talk about the ***Publish or Perish*** software application, which allows importing results from both the *Web of Science* and *Scopus* so that a comparative bibliometric analysis can be made a in a single search box.

PUBLISH OR PERISH, available free of charge at http://www.harzing.com/pop.htm, calculates:

• the research impact and visibility in the Internet by *Google Scholar*

• the bibliometric quantitative indicators: *total number of papers; total number of citations; the average number of citations per paper / article; the average number of citations per author; the total author's papers; the average number of citations per year; the importance of age in the rate of citation; an analysis of the number of authors per paper, etc.*

3. CONCLUSIONS

Each database has its merits and its usefulness. The qualitative value of research cannot be synthesized by a figure or a number. In other words, there is still no unanimously accepted indicator reflecting the scientific value of the results of reasearch. There are both rigors and errors that variously focus on time, impact, etc., etc. On the other hand, academic productivity, too, can be influenced by quite numerous factors.

Biblioscientometry is not sufficient to reveal the quantitative and qualitative aspects of scientific research. There should be a guiding, orienting aspect, which must not necessarily be mandatoryin assessment so as to make visible the results of the investigations. Originally conceived as a selection method, in an attempt to get rid of ballast, scientometry began to act as a method oriented against the initial task. There are omissions for each particular analysis, meaning that research results have to be published. Some even make their own ISI journals because they do not have access to the group of the field in question.

One must conclude that both research and research evaluation should be carefully rethought.

REFERENCES

[1]. Frangopol, P. T., (2005). Indexul Hirsch – un nou indicator scientometric pentru evaluarea rezultatelor unui cercetător ştiinţific, *Revista de Politica Ştiinţei şi Scientometrie*, vol III, no. 4. pp. 75-78;

[2]. Prodan, V., (1999). Interpretarea bibliometrică a unei bibliografii de specialitate, in *ABIR. Buletin Asociaţia Bibliotecarilor din Învăţământ România*, Bucureşti, vol 10, no. 4. pp. 21-24;

[3]. Ursache, L., (2015). *Bibliometria pe înţelesul tuturor*, Galaţi, available on line at http://www.arthra.ugal.ro/xmlui/bitstream/handle/123456789/3221/2014_bibliometria.pdf?sequence=3;

[4] *** (2011). *Ordin privind aprobarea standardelor minimale necesare şi obligatorii pentru conferirea titlurilor didactice […]* – published in *Monitorul Oficial,* no. 540/29.07.2011

[5] *** http://www.timeshighereducation.co.uk/world-university-rankings/2013-14/world-ranking/methodology;

[6] *** http://www.topuniversities.com/university-rankings/world-university-rankings/2013, accessed 4.03.2014;

[7] Haiduc, I. (2015). *Abordare comparativă a evaluării academice pe platforme bibliometrice*, available on line at www.e-nformation.ro.