Date submitted: 15/07/2010

Measuring the visibility of the University's scientific production using GoogleScholar, "Publish or Perish" software and Scientometrics

Prof.dr. **Angela Repanovici**
Transilvania University of Brasov
Brasov, Romania

Meeting: 155. **Science and Technology Libraries**

WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY
10-15 August 2010, Gothenburg, Sweden
http://www.ifla.org/en/ifla76

Abstract:

Open access to scientific information can transform the world information. The research presented here measures the visibility and the impact of the university's scientific production of the Transilvania University of Brasov using the scientific methods of scientometry. These methods provide a means for determining the international value of an university and the statistical evaluation of an individual's scientific research results. In this paper, we define the scientific production and productivity, and present the main indicators for the measurement of the scientific activity. The impact of the research is measured and analyzed through citation analysis. The number of citations suggests the quality of the scientific information. Google Scholar, a freely available scientometric database, indexes academic papers from open access repositories and commercial sources, and also identifies referenced citations. The free Publish or Perish software can be used as an analysis instrument for the impact of the research. We present an exploratory study made at the Transilvania University of Brasov to evaluate the research output of the faculty. We analyzed their 2008 research performances as documented in their annual evaluation that states the number of papers, books, and research contracts. Using Publish or Perish, we calculated the H-index, G-index, HC-index and HI norm, of the 60 more productive professors. We present correlation indicators and discuss the importance of open access tools and repositories for increasing the impact of scientific research.

Key words: H-index, *scientific research, citations, scientometric indicators*, *open access, institutional repository*

The Evolution of Scientometric Methods to Measure Scientific Performance

The statistical analysis of scientific literature began in the first quarter of the 20^{th} century by comparing the scientific productivity of several countries, based on the works published. The interest was very low. The appearance in 1963 of the *Science Citation Index* database produced by the Institute for Scientific Information (ISI) constituted a turning point for scientists and managers all over the world. The database provided an instrument of quantitative evaluation for studies regarding the development of science. The ISI started the database by collecting information from 2300 journals. Today, it is known as *ISI Web of Science* and indexes close to 5000 periodicals from almost all scientific fields representing 90% of most valuable publications for the progress of modern science and technology. These represent the 'mainstream journals'. The number of scientific journals that appear globally is estimated to be around 150,000. Besides the ordinary bibliographical data, *ISI Web of Science* processes the references of all works published in the analyzed journals, thus offering a unique possibility to follow the dissemination of scientific information and in this way to highlight relational structures (Frangopol P., 2005).

The term, bibliometrics, is defined as 'the application of mathematical and statistical methods to books and other means of communication, which are mainly in charge of the management of libraries and documentation centers', while scientometry refers to 'those quantitative methods which are used in the analysis of science regarded as a process of information'. Although bibliometric and scientometric methods are similar, scientometry analyses the quantitative aspects of generation, dissemination and utilization of scientific information in order to contribute to the understanding of the mechanism of scientific research. The number of scientific works published by a scientist is not enough to obtain significant statistical data in the scientometric evaluation of a scientific community. These communities include research groups, departments of universities, institutions, corporations, societies, countries, geopolitical regions, scientific fields or subfields. The primary scientific data of any scientometric investigation are represented by all the authors, their works, their bibliographical and the citations they receive. The set of data produced by a community represent can vary and thus the evaluation indicators can as well. At the national level, the data and resulting indicators suggest ways to compare scientific impact, output and productivity with other countries and to justify federal spending on scientific research.

For examples, the United States National Science Board (NSB) publishes annually the *Science and Engineering Indicators*. The NSB president sent the first report to the President in 1969. In 1972, President Nixon forwarded it to the U.S. Congress, together with a letter:

> „We present the first results of a new effort initiated with the purpose of development of some indicators regarding the state of science, as an institution, in the USA. If such indicators could be developed in the following years, they could help us improve the allocation and management of resources used for science and technology in order to guide the research of our nation to the ways which will reward our society the most."

Two years later, NSB sent the second report and mentioned in the cover letter that the development of such indicators must not be seen as an academic effort, but as being needed to highlight the research efforts of the U.S. in fields of research and their importance to economic growth, future wealth of Americans and the maintenance of a strong security. Support for a strong effort in fundamental research provides new knowledge that is essential for the scientific and technological progress (Courtial, 1990). Thus, a new disciplinem, sciencometrics, was politically tested in an original way. *Science and Engineering Indicators* have been published biannually since 1972 and are not only a source of reference for global science but also a politic instrument mainly for the observation and evaluation of the quantity and quality of American science.

The 2002 edition of *Science and Engineering Indicators* comprises all that is known about the global scientific community. It introduced new indicators such as international collaborations, enrollment of post graduates in England, return rate of post graduate to country of origin, the interest of certain countries for mathematics in pre-university education,and the wages of the teachers. For Romania, information can be found regarding the number of Romanian students in U.S., the scientific production of Romania related to its population, how many articles each Romanian university or institute published and more. The scientometric methodology used by U.S. was adopted as working instrument in the evaluation of the development of states by the the World Bank, the International Monetary Fund, the European Union, and the Organization for Economic Cooperation and Development. Scientific production is an important indicator of a country's human development index.

Measuring Quality of Scientific Production

The measurement of the quality of science is difficult. The determination is currently based on certain criteria such as the number of Nobel Prize winners per country, per university, and per number of inhabitants, which gives small countries (e.g. Switzerland, Sweden) an advantage over larger countries (e.g. U.S., Japan, Russia). Two more criteria are the publication of articles in top journals of scientific world, *Science* (USA) and *Nature* (England), and the number of patents recorded in a country per number of inhabitants. These criteria move Jiao Tong University (Shanghai, China) to top the list of the 500 leading world universities classified by scientific performances. No Romanian university appears on the list.

Quality appears in the countries that have created a proper, free, academic intellectual environment (Frangopol P., 2005). However, it takes time to develop this environment of science and technology. Quality is often not revealed immediately. Consider Einstein's work published in 1905 (at the age of 26), who substantiated the relativity theory, or those of the Italian Enrico Fermi (born in 1901), who at the same age, foresaw the use of the neutrons for the decay of the atoms, idea that lead him to the construction of the first nuclear reactor in the world at Chicago (1942). These opened a new era in the history of science and universal technology. However, the meaning and value of these pieces of work were not immediately recognized upon publications. The identification of the quality of some need an historical perspective in the appreciation of their value.

Quality is often equated with publication the most read and cited scientific journals. Publication in these journals is controlled by accredited readers and presents a certain guarantee of quality of a scientific work. The ISI maintains what many consider the premiere jouranl list. Within this list exists a classification of scientific journals, according to the 'impact coefficient', a size which represents the ratio between the total number of citations of that journal in a two years period and the total number of scientific works published in the journal in the same period. Obviously, the more prestigious or important a journal is in a certain field of science, the more used and cited it will be and it will have a bigger impact coefficient. The works published in other journals than those from the mainstream are rarely cited and often lost for science.

Utilizing Scientometry

The evaluation of the performance of scientific research is the most important application of scientometry. Recently, interest grows in the use of scientometry techniques for the measurement the efficiency and productivity of the research. These techniques are not substitutes for expert reviews and evaluations, but are complimentary.

The common scientometry measurements include these four.

H-index: The H-index was introduced by Hirsch (2005) and simultaneously measures the quality and the sustainability of the impact of a researcher's publication. It is based on the quantity (number of papers) and quality (impact, or citations to these papers) as well as the distribution of the citations received by the researcher's publications. There are many tools available for the computation of H-index today. They are integrated into the databases *Scopus* and *ISI Web of Science.*

G-index: In order to overcome the deficiencies of the H-index, Egghe proposed the G-index to measure the productivity of the researchers based on their publications (Egghe L., 2006).

HC-index: This is the contemporary H-index which adds a weight related to the age of the article to each cited article of a researcher through less parameterised weight of the articles which have more years. For example, citations received by one article published during the current year is weighted 4 times more that those the citations received by an article published 4 years. In the same way, an article published 6 years ago counts 0,67 times, etc.

HI norm: The index represents a modification of the H-index formulated by the developers of *Publish or Perish*. In order to calculate this index, the number of citations to each article is normalised through the division of the number of citations received by an article with the number of authors of the article. The index is calculated as an H-index normalised by the count of the citations. This approach takes into consideration the co-authors. The index offers a better approximation of the author's impact.

The citation remains the best way of recognizing value as it represents conceptual association of scientific ideas, connections between current research and previous activity, relationships between specialized research within a certain field classic research, and maps identifying significant fields of research. Although the number of publications indicates a measure of productivity, this number does not say anything about the quality of the research. The number of citations is a better indicator of quality. A caveat is that being co-author to articles that are highly cited could 'exaggerate' the reputation of a scientist. For researchers, citations are intrinsically related to the reward system of the science. They recognize a scientist's knowledge and indicate the use of information.

The primary measurements used in academic evaluations include these:

- The total number of publications (Np):
  - Advantage: it measures the productivity.
  - Disadvantage: it does not measure the importance or impact of the publications.
- The total number of citations (Nc, tot):
  - -Advantage: it measures the total impact.
  - -Disadvantage: it can be hard to find all; it can be biased by a small number of 'big hits' that are not representative for the person if he is co-author of those essays; Nc, tot gives excessive weight to review articles that are frequently cited in comparison with the number of citations per publication.
- Citations per publication, the ratio between Nc, tot and Np:
  - -Advantage: it allows comparisons between publications of different ages.
  - -Disadvantage: it is hard to find; it rewards low productivity; it penalizes high productivity and personal contributions to research.

Utilzing Open Access Scientometric Tools

Tools that are freely available can be used to measure scientific research performance. As mentioned earlier, Google Scholar is a freely available databases of scientific references with links to full text of articles when available. *Publish or Perish* is a software application intended for the analysis of the citations. This software is available from Professor Anne Wil Harzing, a specialist in international management at Melbourne University from Australia (Harzing A.-W., 1997-2009).

This software uses Google Scholar to obtain the references, the sources which cite them and then it analyses those presenting the following statistics:

- Total number of papers
- The total number of citations
- The medium number of citations per article
- The medium number of citations per author

- The total number of essays of the author
- The medium number of citations per year
- The H-index, G-index, Hc-index and HI norm

Publish or Perish is designed to assist scientists to present personal cases of the impact of their research. It is not intended as a mechanical evaluation but should be integrated with care. For example, a professor has numerous citations to his publications obviously has a significant impact in the field. However, the reverse is not necessarily true. If a professor has few citations s, the causes may be the lack of impact of the field of research, or the fact that he works in a small field, publishes in a language other than English, or publishes only in books and proceedings.

So, some consideration is needed about which scientometric tool to use depending on the field of research. Google Scholar has better indexing of proceedings and non-English language material than *ISI Web of Science*, still does not perform so well tracking citations from books and chapters of books. The natural sciences and those related to health are well covered in *ISI Web of Science* given its journal coverage, and as a result Google Scholar fewer citations in these fields. In general, the measurement of the citations in Social Sciences or Humanities s will be underestimated given the problem of book citations. Also, fields dependant on conference proceedings will be undercounted.

The use of Google Scholar can be beneficial in the following fields:

- Business, Administration, Finance & Economics;
- Engineering, Computer Science & Mathematics;
- Social Sciences, Arts & Humanities.

If possible, results from *Scopus*, *ISI Web of Science* and GoogleScholar should be compared. Coast and accessibility may prevent this, hence the utility of *Publish or Perish* and Google Scholar.

Measuring Research Performance at Transilvania University of Brasov

We explored the scientific performane of the faculty at the Transilvania Universityof Brasov using Google Scholar and *Publish or Persish*. We compared the resulitng H-index, G-index, Hc-index and HI norm with those found in *ISI Web of Science.*

The Methodology

We compiled the 2008 research output of the academic community using the faculty's annual reviews. These include the number of publications including books, articles and report, and the number of research contracts. From these, we identified 60 faculty members with more than 500 research points. These ranged from P1= 3599.96 to P60=503.29. In the annual number of points, an important weight is held by the research contracts and their value. Thus, the number of research points does not reflect the quality of the research but the quantity.

Using the *Publish or Perish* software, we searched these 60 faculty members and compiled the results. Figure 1 illustrates the results of an author search.

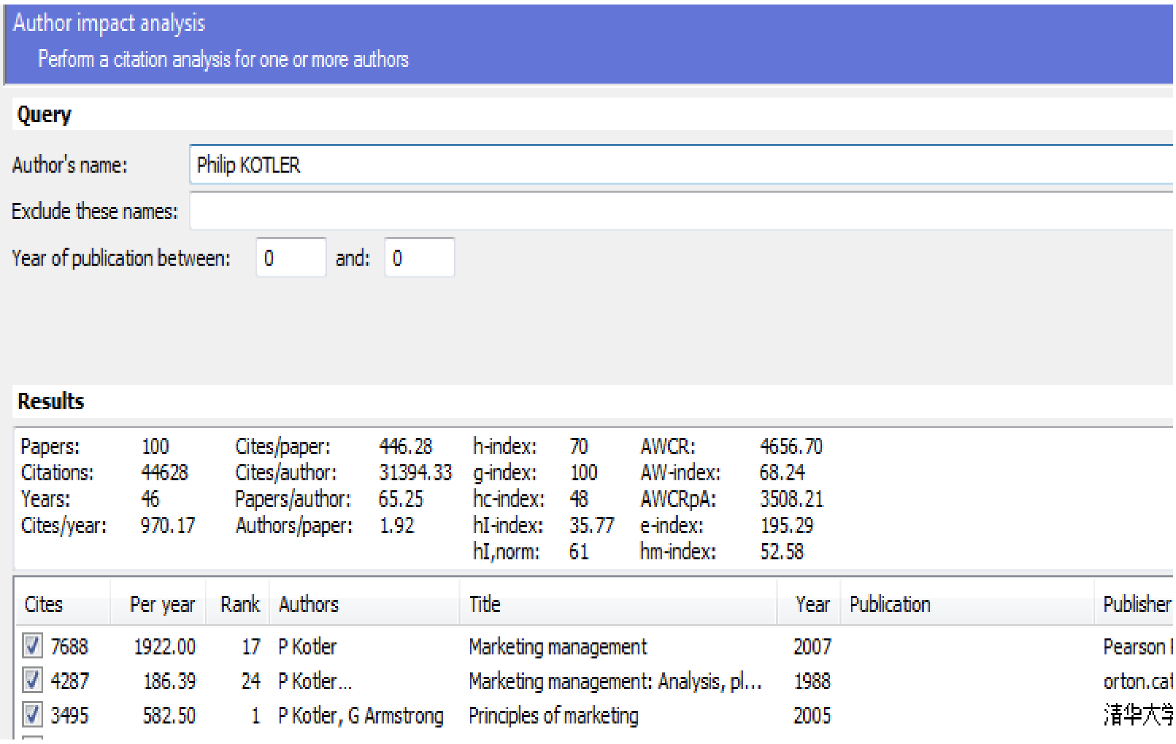

Figure 1: Author Impact Analysis using *Publish or Perish*

In order to verify the validity of the H-index generated by the *Publish or Perish* software ,we searched *ISI Web Science* for each of the 60 professors, compiling the number of works indexed in this data base and the number of citations.

The Results

Validating the *Publish or Perish* H-Index

Validating the H-index generated by *Publish or Perish* was important. We did two correlations: one to the number of papers and one to the number of cited papers. The correlation index was calculated by the formula:

$$r = \frac{\operatorname{cov}(X,Y)}{\sigma_x \sigma_y} \qquad \operatorname{cov}(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n} \quad (1)$$

For the case of the number of indexed papers the correlation index, r = 0,353285, which indicates a weak intensity connection.

TABELA DE REGRESIE	
SUMMARY OUTPUT	

Regression Statistics	
Multiple R	0,353285
R Square	0,12481
Adjusted R Square	0,109721
Standard Error	6,479329
Observations	60

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regression	1	347,2448	347,2448	8,271338	0,005625
Residual	58	2434,939	41,9817		
Total	59	2782,183			

	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95,0%*	*pper 95,0%*
Intercept	-0,3408	1,115879	-0,30541	0,761149	-2,57447	1,892876094	-2,57447	1,892876
X Variable 1	2,022982	0,703403	2,875993	0,005625	0,614968	3,430995362	0,614968	3,430995

Figure 2: The Regression Table Correlating the *Publish or Perish* H-index with the Number of Papers Indexed in *ISI Web of Science*

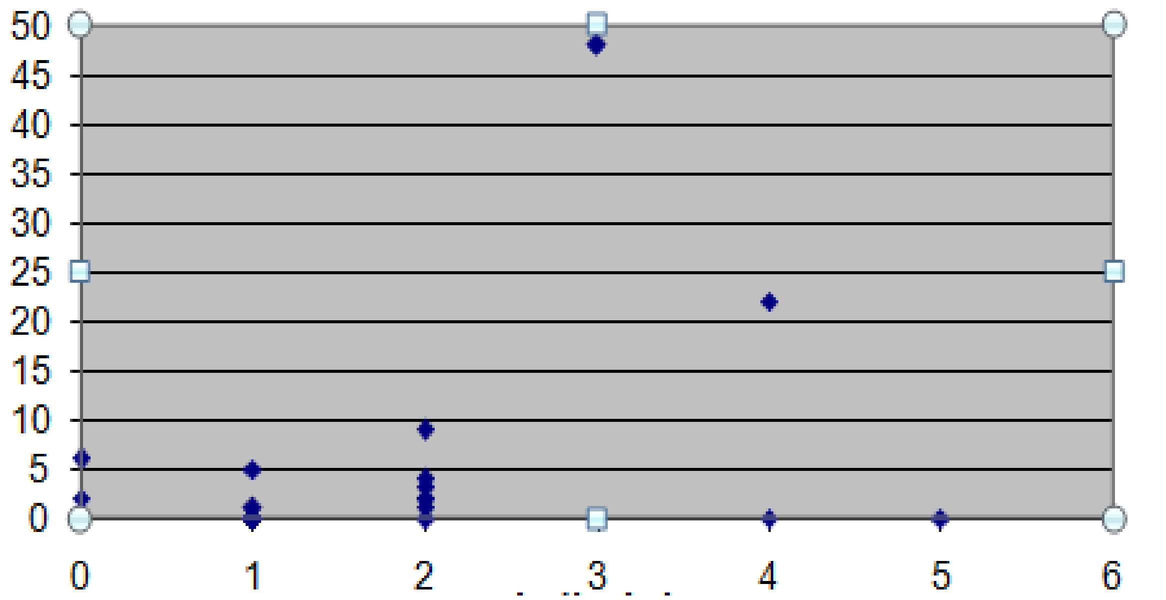

Figure 3: The correlation between the H-index and the number of papers indexed in Web of Science

The graphic analysis reveals a dispersed cloud of points that suggest a weak intensity connection between the two indicators.

For the case of the number of cited papers, the correlation index, r = 0,471483, indicates a weak intensity connection.

TABELA DE REGRESIE								
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,471483							
R Square	0,222296							
Adjusted R Square	0,208887							
Standard Error	7,401524							
Observations	60							
ANOVA								
	df	*SS*	*MS*	*F*	*gnificance F*			
Regression	1	908,2117	908,2117	16,57848	0,000143			
Residual	58	3177,388	54,78256					
Total	59	4085,6						
	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95,0%*	*pper 95,0%*
Intercept	3,364761	1,2747	2,639649	0,010643	0,813172	5,916351133	0,813172	5,916351
X Variable 1	3,271656	0,803517	4,071668	0,000143	1,663241	4,880070465	1,663241	4,88007

Figure 4: The regression table for the correlation between the H-index and the number of citations indexed in Web of Science

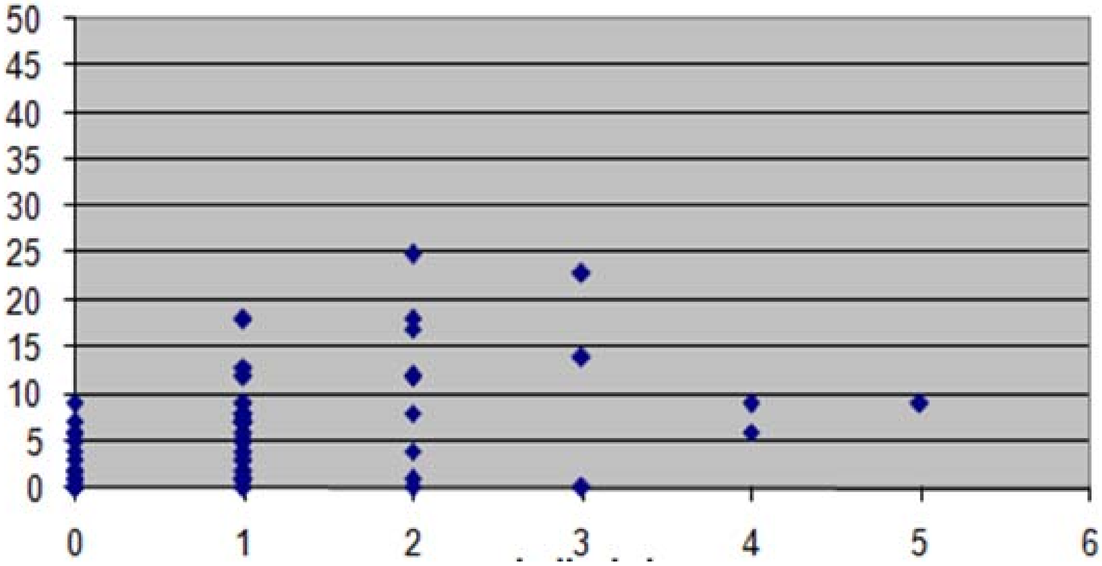

Figure 5: The correlation between the H-index and the number of citations indexed in Web of Science

The graphic analysis reveals a dispersed cloud of points which suggests a weak intensity connection between the two indicators.

H-index compared to Research Points

We compared the number of research points to the individual's H-index and found no correlation.

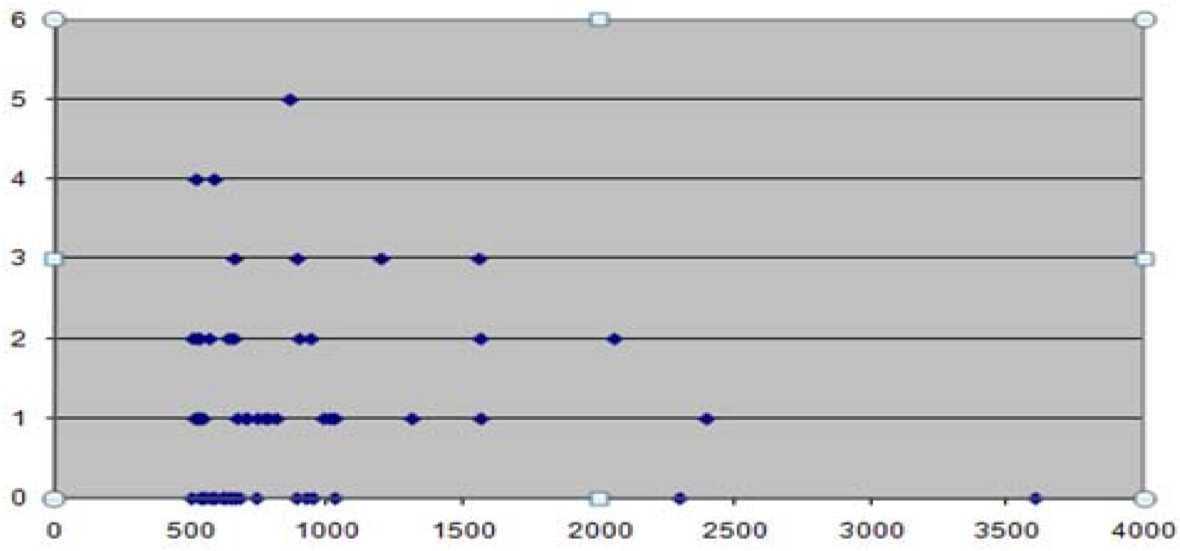

Figure 6: The correlation between the H-index and the number of research points

In the research points calculated by Transilvania University, the weight of the articles published is not relevant. The highest number of points is obtained thanks to the research projects, in accordance to the amount financed in the project.

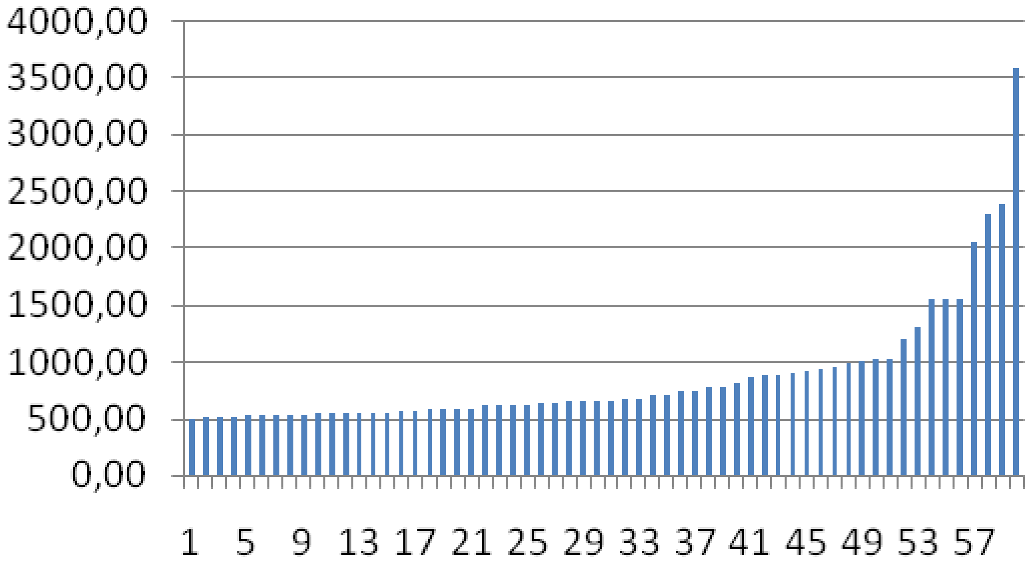

Figure 7: Number of research points for the 60 professors with the most research points.

Figure 8 shows the H-index for those same 60 professors. There is no correlation between the H-index and the research points even though although funded projects should disseminate the results through scientific publications. The professors with more than 500 points do not have a corresponding H-index. The greatest value of the H-index for this group is 5 (one professor), followed by 4 (two professors), 3 (four professors) 2 (nine professors). Most of these professors have an H-index lower than 1.

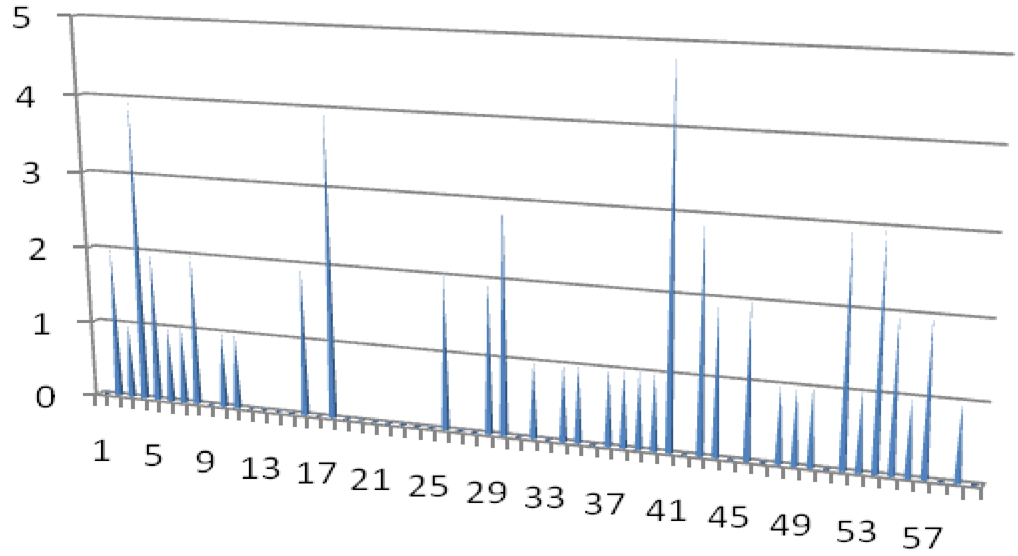

Figure 8: The H-index for the first 60 professors at the achievement of the 2008 research points

One professor has 52 articles indexed in the *ISI Web of Science* with 48 citations. His H-index is 3, G- index is 4, and research points are over 1000. Most professors have fewer papers indexed in *ISI Web of Science* with limited citations.

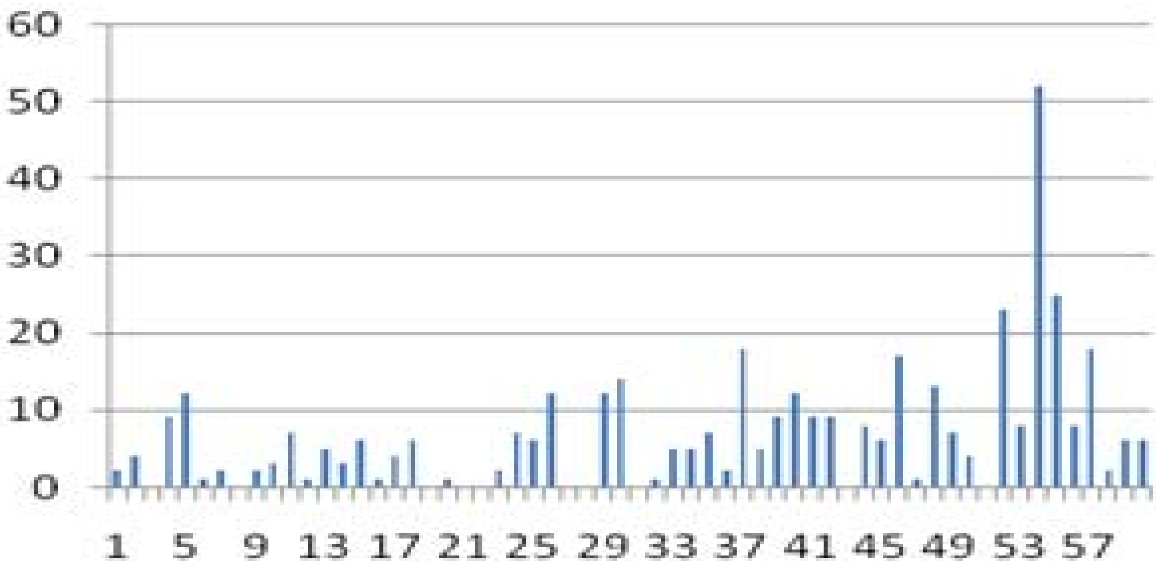

Figure 9: The number of articles indexed in *ISI Web of Science* by the top 60 professors

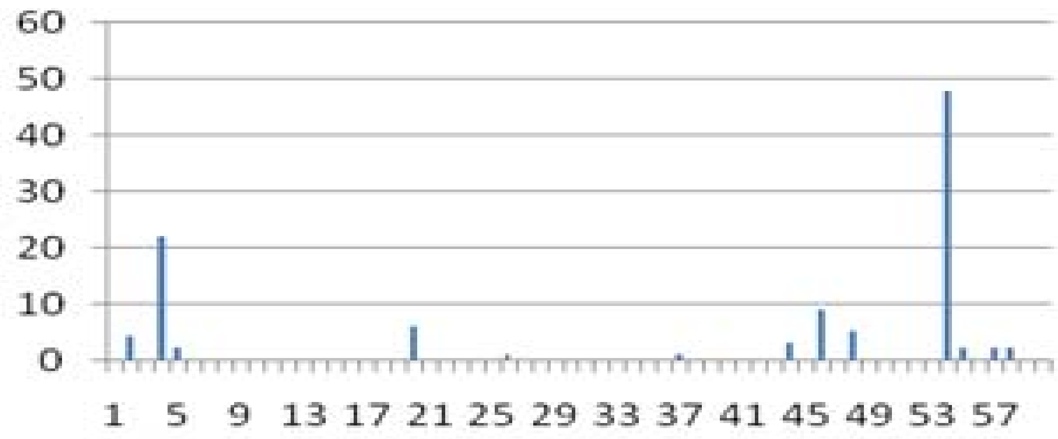

Figure 10: The number of citations to papers indexed in *ISI Web of science* by the top 60 professor

The quality of the scientific papers measured by the number of the citations is the highest for professor P5 who has 9 papers indexed in *ISI Web of Science* with 22 citations. His H-index is 4 and the G-index 6, some of the higher values.

Conclusions

Authors' impact analysis and citations are not an assessment tool in Romanian universities. But it is commonly accepted that increasing research impact through more citations is one qualitative indicator. *ISI Web of Science* and Google Scholar are scientometric databases that can generate an individual's H-index. The academic community is not familiar with these instruments and their potential role in describing the impact of one's science. In 2010, we have free access to *ISI Web of Science* through the Romanian Ministry of Research. This access allows us to compare this resource with the freely available Google Scholar and *Publish or Perish* software. Google Scholar offers a better alternative for every professor to be informed about their impact.

This exploratory study was presented to academic community to illustrate the utility of these tools and to inform faculty of challegnes with assessing their impact. We showed that professors with high H-index calculated using *Publish or Perish* are those with papers indexed in *ISI Web of Science*. This is not correlated with the number of research points garnered by faculty. We suggested to the faculty that all professors have to be analysed with the same indicator. Further, we recommended that Google Scholar and the H-index obtained using *Publish or Perish* offer tools for assessing scientific research in university and evaluating professors.

Within the European scientific community, the Frame Programme 7 calls for publishing research in open access. In the future, it is recommended to take into account this instrument in the academic evaluation and give additional weight to publications that are accessible globally. Affordable and accessible tools for doing this are now available. The *Publish or Perish* software is an easy to use instrument for analyzing the impact of research. It calculates the impact of the researchers using the very visible resource, Google Scholar. The data source provides a more comprehensive coverage of citations than *ISI Web of Science,* including citations in books, conference proceedings, working papers and non-ISI indexed journals. Our study is the first step in implementing this type of assessment, affirming the open access principle and creating the first digital repository in Romania.

We live in an era full of the unpredictable, new models for promoting scientific development. It is exciting but daunting to academic institutions not yet involved in projects for promoting their own research. Reality changes, and the need for information by different academic communities is constantly increasing and diversifying. We recommend that libraries and academic institution promote the use of tools such *as Publish or Perish* software and Google Scholar as means to increase access to scientific information. We also urge the creation of repositories within the research institutions adding to the visibility of researcher's work and creating a central national repository.

Bibliography

1. Borgman, C. (1990). *Scholarly Communication and Bibliometrics.* Newbury Park: Sage.
2. Bormann, L. D.-D. (2007). What do we know about the H-index? *Journal of the American Society for Information Science and Technology* , 58(9): 1381-1385.
3. Brookes, B. (1990). Biblio-, sciento-, infor-metrics??? What are we talking about. *Informetrics 89/90,Elsevier Science Publishers B.V.* , 31-43.
4. Brown, T. (2002). *Definiția scientometriei.* București: Ed-astra.
5. Carriso Sainero, G. (2000). Toward a Concept of Bibliometrics. *Journal of Spanish Research on Information Science 1 (2)* , 5986.
6. Courtial, J.-P. (1990). *Introduction a la Scientometrie, De la bibliometrie a la veille tehnologique.* Paris: Anthropos-Economica.
7. Egghe, L. (2006). An improvement of the H-index: the g-index. *ISSI Newsletter* , 2(1): 8-9.
8. Egghe, L. R. (in press). An H-index weighted by citation impact. *Information Processing & Management* , in press.
9. Frangopol, P. T. (2005, Marrch 25). Revista internațională Scientometrics. *România Liberă* , p. 2.
10. Haiduc, I. (2001). Imaginea externă a științei românești. *Revista de politica științei și scientometrie* , 2-15.
11. Haitun, S. (1983). *Scientometrics: State and Perspectives.* Moskow: Nauka.
12. Harzing, A.-W. (1997-2009). *HARZING HOME*. Cited in August 13, 2009, de pe HARZING COM: http://www.harzing.com/index.htm

13. Harzing, A.W.K.; Wal, R. van der (2009) A Google Scholar H-index for Journals: An Alternative Metric to Measure Journal Impact in Economics & Business?, *Journal of the American Society for Information Science and Technology,* vol. 60, no. 1, pp 41-46.

14. International journal of Scientometrics, B. a. (2006, 2 21). *CYBERMetrics*. Preluat pe August 14, 2009, de pe http://www.cindoc.csic.es/cybermetrics/map.html

15. Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original H-index. *ISSI Newsletter* , 2(3): 4-6.

16. Munteanu, R., & Apetroae, M. (2007). Factorii de impact și ierarhizarea revistelor științifice. *Revista de Politica Științei și Scientometrie* , 97-104.

17. Nalimov, V. M. (1969). Scientometrics. *Nauka, Moskow* .

18. Repanovici, A.(2009) „Marketing Research about Attitudes, Difficulties and interest of academic Community about Institutional Repository, PLENARY LECTURE." *Advances in Marketing, Management and Finances, Proceedings of the 3rd International Conference in Management, Marketing and Finances, (MMF'09), Houston,USA,April 30-May 2, 2009, ISSN 1790-2769, ISBN 978-960-474-073-4,pag.88-95.* Houston,USA: WSEAS.

19. Repanovici, A.(2010): Measuring the visibility of the universities' scientific production using scientometric methods, PLENARY LECTURE,,*The 6th WSEAS/IASME International Conference on EDUCATIONAL TECHNOLOGIES, (EDUTE '10)* ,pag.56-62, Sousse, Tunis ,WSEAS. http://www.wseas.us/conferences/2010/tunisia/edute/Plenary1.htm

20. *Shanghai Jiao Tong University*(2009), Cited in August 12, 2009, from Institute of higher Education: http://www.arwu.org/rank2008/ARWU2008_TopEuro(EN).htm