How to Improve the Citation Impact of a Paper: Choice of Journal, Co-authors and Institutional Addresses.

José A.N.F. Gomes[1] and Elizabeth S. Vieira[2]

[1]jfgomes@fc.up.pt, [2]elizabeth.vieira@fc.up.pt
REQUIMTE/Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto (Portugal)

Abstract
This paper attempts at correlating the citation impact of Physics and Chemistry documents in the Web of Science with some parameters of the source, namely the impact factor of the journal, the number of co-authors and the number of institutional addresses. After a comparison of the distribution functions of the 5-year citations of the 2004 Physics and Chemistry documents, a detailed study of the correlation between the citation counts and the three parameters above is presented. There is a striking similarity between the properties found for Physics and Chemistry but for the occurrence of a small but meaningful number of Physics documents with an extremely high number of co-authors and institutional addresses. The average number of citations may double as the number of co-authors increases within its more common range.

Introduction
Counting citations is a bibliometric tool used widely to measure the impact of a research publication and this impact is frequently taken as an indicator of its quality. Citation counts were used for the first time by Gross & Gross (1927) and since then citation analyses have been conducted for assessment of national science policies and disciplinary development, for evaluating countries, institutions, books, journals and individual scientists (Lewison, 1998; Nicolaisen, 2002; Tijssen, van Leeuven & van Raan, 2002; Ventura & Mombrú, 2006). Despite its limitations (Adler, Ewing & Taylor, 2008; Cheek, Garnham & Quan, 2006; MarcRoberts & MarcRoberts, 1996), the application of citation analysis and other bibliometric indicators is increasingly used by academic and research institutions in several countries for promotion, tenure, salary raise and grants decisions (Borgman &Furner, 2002; Weingart, 2005).

For the analysis of citations, the sources most used are the Web of Science from Thomson Reuters and the Scopus bibliographic databases. Both give the number of citations obtained by a particular document or set of documents over the years. Their coverage is different and this should be well understood if the citation metrics extracted from one or the other are to be compared. These differences and their impact have been discussed in a number of publications (Bakkalbasi, Bauer & Glover, 2006; Jacsó, 2005; Meho & Yang, 2007; Vieira & Gomes, 2009).

Several studies have been made to determine how the impact of a published document can be influenced by features such as international cooperation (Glänzel & Lange, 2002; Glänzel & Thus, 2004), self-citations (Glänzel & Thus, 2004) and number of references (Persson, Glänzel & Danell, 2004). Rousseau (2000), based on evaluation data from Limburgs Universitair Centrum, Universitaire Campus (Belgium), investigates the citation output of single authored articles versus the output of articles with two or more co-authors. This is done for citable articles published in journals covered by the Journal Citation Reports (JCR) of Thomson Reuters and for all other articles published in non-JCR journals. In this study he observed that multi-authored articles have usually higher citations frequencies than single-authored ones, but this relation does not appear to hold in all cases. Glänzel & Lange (2002) studied the observed, expected and relative citation rate for domestic papers, bilateral

Proceedings of ISSI 2009, pages 31–39.
Edited by B. Larsen and J. Leta.

international papers (papers with exactly one other partner country) and multinational papers (publications with co-authors addresses from at least three different countries) for the most active 30 countries in Biomedical Research and Chemistry. They concluded that international collaboration results in greater visibility and higher citation impact. They observed that for Biomedical Research multinational papers have more impact than bilateral and domestic publications. In Chemistry they observed that the expected citation rate does not always increase with the growing number of countries involved. Glänzel & Thus (2004) analyzed bibliographic data extracted from the Web of Science in order to determine the effect of co-authorship on foreign citations and self-citations to conclude that co-authorship has a strong effect on foreign citations but only a moderate influence on the number of self-citations. Persson, Glänzel & Danell (2004) made a study based in all papers indexed in the 1980-2000 annual volumes of the Science Citation Index of Thomson Reuters to analyze the relationship between the number of co-authors and the length of reference lists for all fields combined, for national and international papers. The same study was made to analyze the relationship between the number of co-authors and the mean number of citations using a three year citation window. The results showed a positive relationship between the number of co-authors and the length of reference lists both in national and international papers. The same behavior was observed for the relationship between co-authors and mean number of citations. Figg, Dunn & Liewehr (2006) counted the number of original research articles published in six leading journals - Cell, Science, Nature, New England Journal of Medicine, The Lancet and Journal of the American Medical Association - for the years 1975, 1985 and 1995 and determined the number of authors and the number of separate institutions in order to establish the relationship between the number of times an article was cited and the number of authors and institutions. For these journals they found that the number of times an article was cited correlated well with the number of authors and with the number of institutions. Radicchi, Fortunato & Castellano (2008) studied the distribution of citations obtained by a single publication in several disciplines. They showed that the probability that an article is cited c times has a large variations among different fields, but all distributions may be rescaled on a universal curve when the relative indicator $c_f=c/c_0$ is considered, where c_0 is the average number of citations per paper for the discipline. The same behavior is observed when the citation distributions of articles published in the same field but in different years are compared.

In the next section below, the methodology used to set up a database of the 2004 web of Science documents for Physics and Chemistry is described. The citation distribution for the two disciplines are presented and discussed in some detail. Then, we consider the distribution of the three parameters that were selected, number of co-authors, number of institutional addresses and the journal impact factor. The correlation between each of these variables and the mean citation rate of the documents with a given value for the parameter selected is presented. In the last section, a set of conclusions are proposed.

Methodology

The study presented here is based on the analysis of 206093 documents published in 2004 and indexed in the Web of Science (WoS) for the fields of Physics (88026 documents) and Chemistry (118067 documents). A document is considered as belonging to Physics (or Chemistry) when it is published in one of the journals that the Essential Science Indicators considers in the field of Physics (or Chemistry). This includes 338 journals for Physics and 574 journals for Chemistry. Complete information about these documents was extracted from WoS to build a database with the number of co-authors, number of institutional addresses, the journal impact factor and the number of citations received. For the citation counts of each and

every document in WoS, a five year citation window (from publication to the present[1]) was considered.

To study the relation between the number of citations per paper and the number of its co-authors, we selected all documents produced by a single author, by two authors, by three authors and so on, until the maximum number of authors in the particular field. (Documents where the author's field was identified as anonymous were eliminated.) A similar procedure was used to assess the relation between the citation rate and the number of addresses in a document.

The study of the relation between the number of citations per paper and the impact factor of the journal where it was published was made by taking the impact factor of each journal from the JCR for 2006. We considered the impact factor of journals in the year 2006 because our study was based on the citations received in 2004 through 2008 by documents published in 2004. The citations received by these documents in the five year window should correlate better with the impact factor calculated for 2006 than that for 2005 due to the methodology used by Thomson Reuters.

Results and discussion

The distribution of the citations in 2004-08 of the documents published in 2004 was found by counting the citations of each individual document and then aggregating the documents with the same number of citations. In this way, we found the number of documents with 0, 1, 2, 3, etc citations. The average number of citations per document is 8.404 for Physics and 9.546 for Chemistry. Figure 1 shows the distributions of citations with inserts to appreciate the long tail.

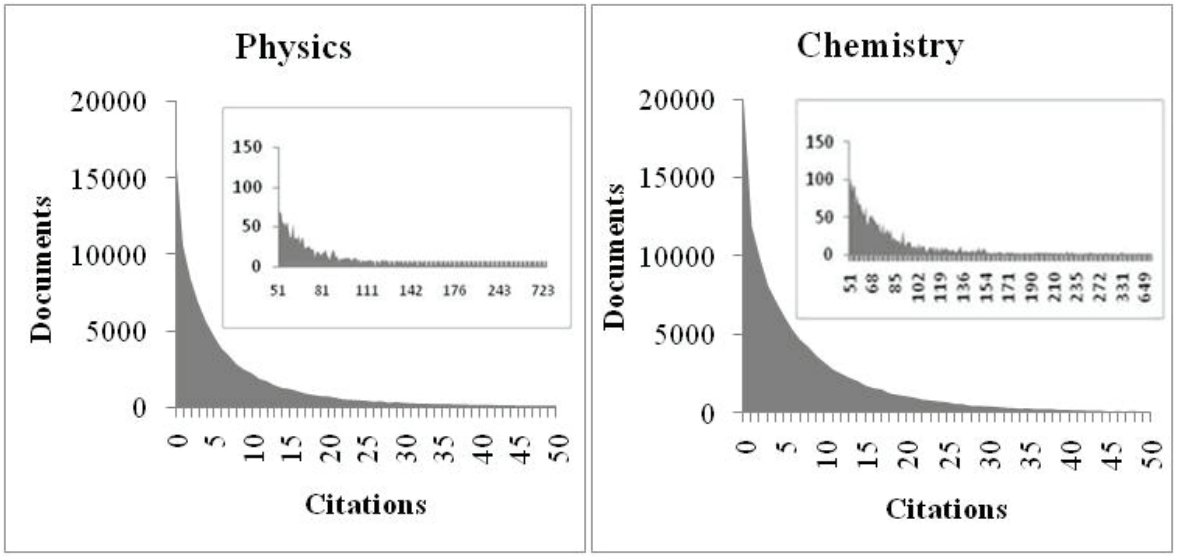

Figure 1. Distribution of citation count of documents dated 2004 in the Web of Science.

The frequency of zero citations is 19.2% for Physics and 17.2% for Chemistry, while a single citation was found for 12.0% of the Physics and 10.3% of the Chemistry documents. This means that the citation distribution for Chemistry will show a slightly longer tail in agreement with the larger average value. The differences between the two disciplines are very small indeed. The standard deviation of the distributions for Physics and Chemistry are both about 20, that of Physics being slightly larger (by 0.26) due to the occurrence of a number of

[1] The data were extracted from the WoS in 9 to 13 of January of 2009 so that the citation period for papers published in 2004 is very close to a five year window.

documents with more than a thousand citations. For Chemistry, the 3 documents with more than a thousand citations contribute about ¼ of the variance; for Physics, the three most cited documents contribute about ½ of the variance. Above 250 citations, we find 62 documents in Chemistry with an average of 3.2 authors very close to the general average of 3.84; for Physics, we find just 31 documents but these have an average of 19.9 authors compared with the general average of 3.94. This makes it clear that, while the most cited Chemistry papers do not achieve this by increasing the number of authors, the top Physics papers have a unusually high number of authors.

Number of co-authors

Figure 2 shows the distribution of the number of co-authors in all documents in WoS for Physics and for Chemistry.

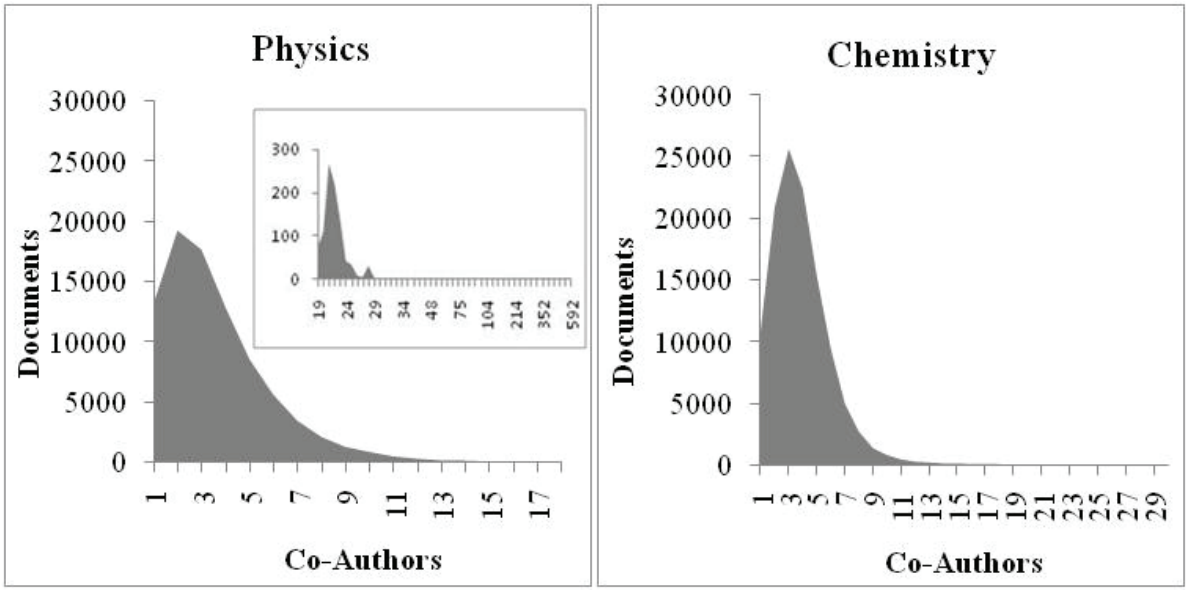

Figure 2. Distribution of co-authors counts of documents dated 2004 in the WoS.

The average number of authors per document is 3.94 in Physics and 3.84 in Chemistry. A major difference between the two disciplines is the relative large number of Physics papers with more than 17 authors as seen in the inserted plot and a number of papers with a very large number of authors. In fact, the maximum number of co-authors identified per document was 592 in Physics and 30 in Chemistry. Comparing in detail the two distributions, we see that the mode is 2 authors in Physics and 3 authors in Chemistry. If we consider just the peak in the range 1-18 authors, then the means are 3.67 authors for Physics and 3.83 authors for Chemistry. In summary, we can say that Physics publications tend to have a slightly smaller number of authors with the exception of just 1.1% of them with more than 18 authors. For Chemistry, there is just 0.1% of the publications with 19-30 authors.

Figure 3 shows the plot of the average impact of the documents against the number of co-authors for Physics and for Chemistry. The observed standard deviation is shown as an error bar.

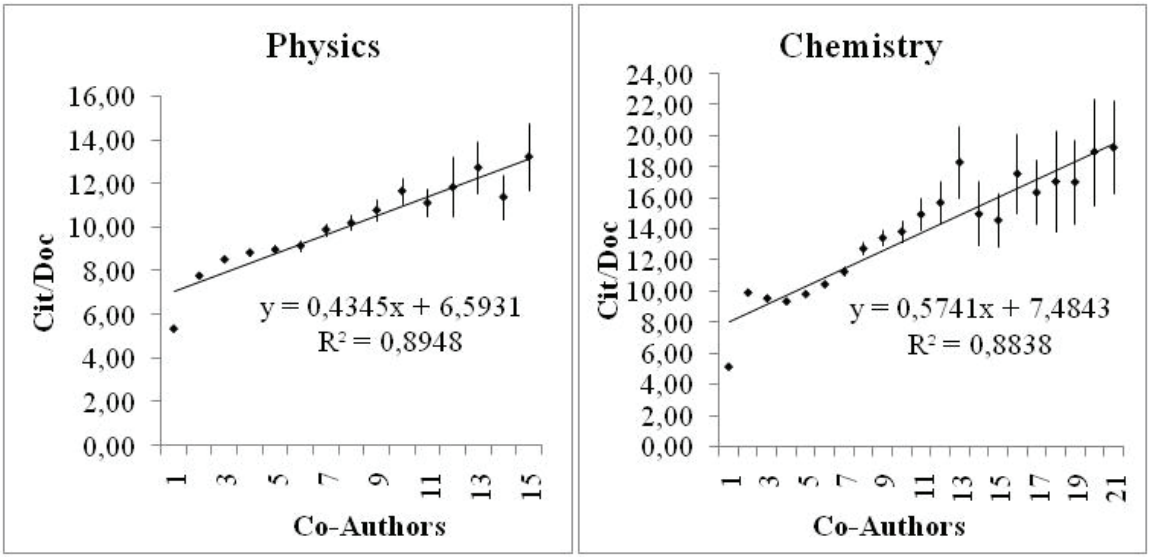

Figure 3. Mean citation rate (and its standard deviation) against the number of co-authors for documents published in 2004 in Physics and in Chemistry.

The mean citation rate correlates well with the number of authors, having a similar effect in both disciplines: the number of expected citations doubles as we go from 1 to 18 authors. In reality, the average citation rate of single-authored documents is well below the regression line and is almost the same for Physics (5.34) and Chemistry (5.14). For 2-authored documents Chemistry documents have a higher impact of 9.90 compared with 7.76 for physics. The increase in the standard deviation as the number of co-authors grows is a direct consequence of the smaller number of documents in this region. Glänzel, W. & Thus, B. (2004) explained this increase of the citation rate by the increment of self-citations.

Number of addresses

The second parameter to be considered is the number of addresses associated with each document. Figure 4 shows the distribution of number of addresses per document in Physics and in Chemistry.

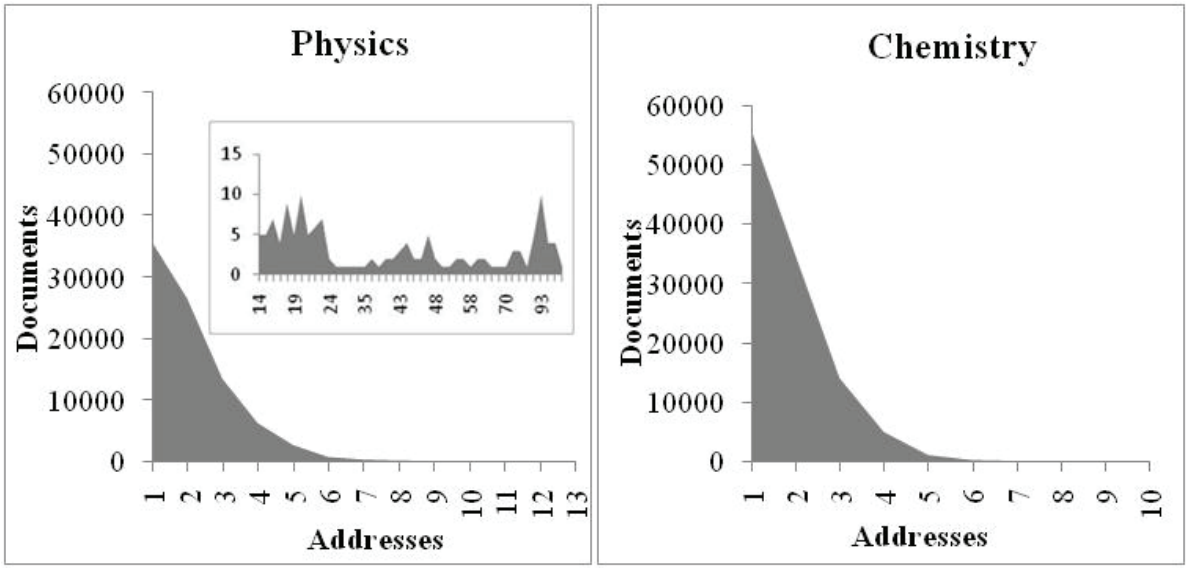

Figure 4. The distribution of the number of addresses in documents dated 2004 in the WoS.

For the set of Physics documents examined, the maximum number of addressees identified per document was 114 while for chemistry the maximum number was 10. The mode is one address in both disciplines but the average number of addresses is 2.11 for Physics and 1.75 for Chemistry. Again, this average suffers the influence of the very long tail of the Physics distribution, being reduced to 2.04 if we consider just the peak between 1 to 13 addresses, a value still larger than the Chemistry average. The number of addresses in the Physics decreases steadily from 1 to 13; the inserted plot shows the occurrences of documents with more than 13 addresses. In Chemistry, the number of documents decreases monotonously from 1 to 10 addresses.

Figure 5 shows the plot of the average impact of the documents against the number of addresses for Physics and for Chemistry. The observed standard deviation is shown as an error bar.

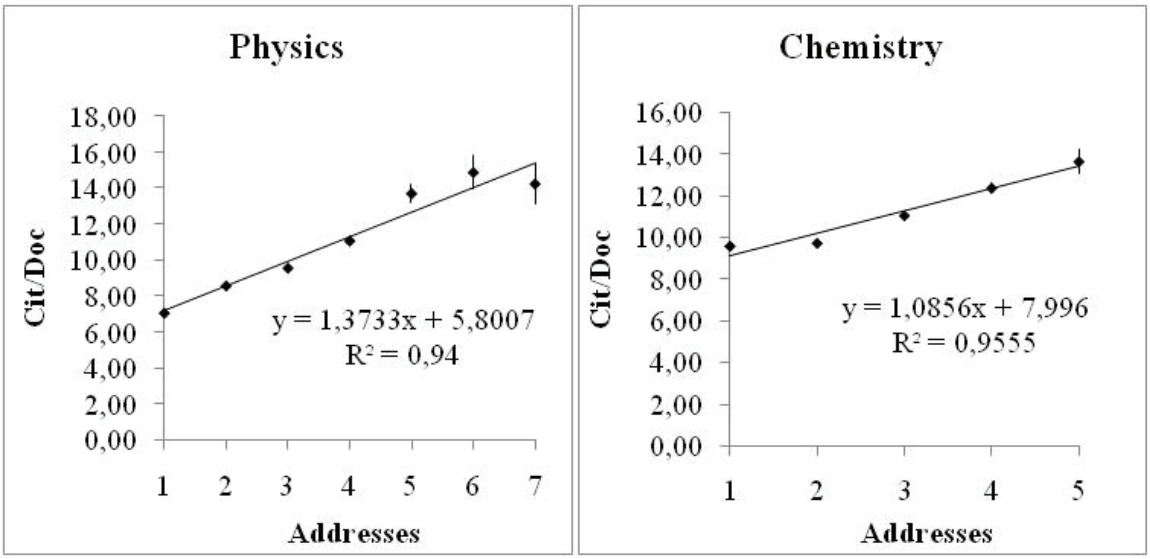

Figure 5. Mean citation rate (and its standard deviation) against the number of addresses for documents published in 2004 in Physics and in Chemistry.

The correlation is very good, showing a marked effect of the number of addresses in the citation rate. The impact growth is larger for Physics: As we go from a single address to the bottom of the peak, we get a 50% improvement in the impact for Chemistry (5 addresses) while it doubles for Physics (7 addresses).

Impact Factor

A third parameter studied was the impact factor of each journal. Given the way in which Thomson Reuters calculates the impact factor of a journal, there must be a strong correlation with our citation count. Figure 5 shows the variation of the mean citation rate in the 5-year window considered in this paper with the impact factor of the journal.

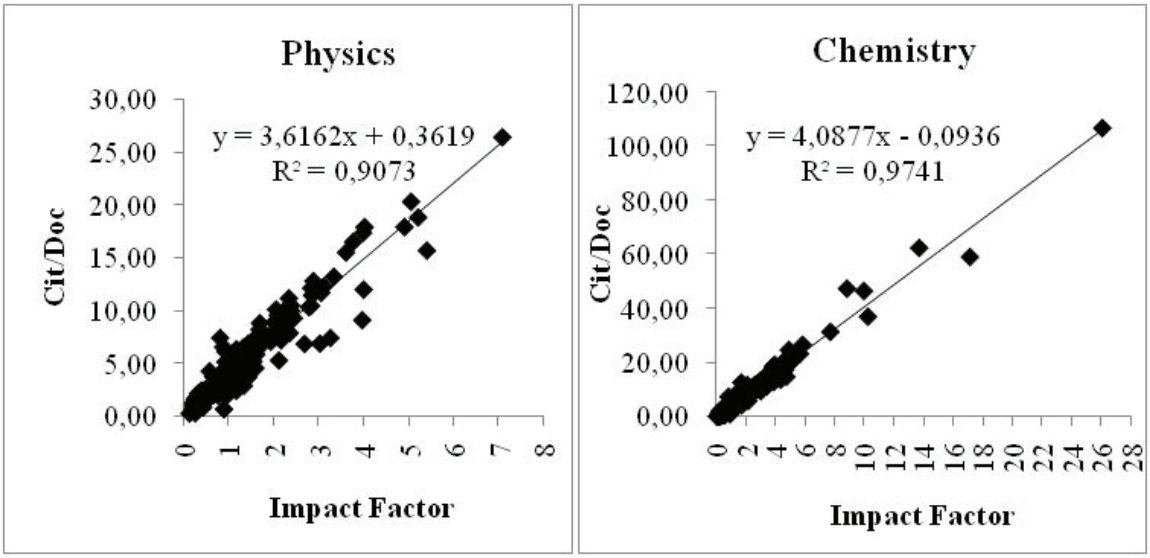

Figure 6. Mean citation rate dependence on the impact factor of the Physics and Chemistry journals.

The correlation between the mean citation rate and the impact factor is good for Physics and Chemistry, the correlation being better for Chemistry. We considered values of the impact factor for which there are more than 50 documents from one or more journals. For Chemistry, the impact factor of journals varies between 0,05 and 26,05 while for Physics the impact factor varies between 0,12 and 7,07. About 5% of the documents in Physics and 3% in Chemistry were published in journals that are not in the 2006 JCR. The slopes of the regression line must be explained by the width of the window used in this study. The difference between the two disciplines is probably due to the different time gap between publication and citation. Another difference is that we calculated the citation rate for all documents while Thomson Reuters counts all citations but divides by articles and review articles only.

Conclusions

This paper considers all documents referenced in 2004 in the Web of Science for Physics and Chemistry, more than 200000 in total, and seeks to clarify the possible correlations between the citations of these papers in the five year period 2004-2008 and some simple document properties like the number of co-authors, the number of addresses and the impact factor of the journal where it was published. A few conclusions may be drawn from this study.

a) The first striking finding is the similarity between Physics and Chemistry in what concerns publication cultures. Chemistry is more prolific with 57% of the total the number of documents against 43% for Physics.

b) The average number of citations per document is also higher in Chemistry but Physics trails by just 12% (8.40 vs. 9.55).

c) The big difference between Physics and Chemistry in the co-author distribution comes from the occurrence of a few documents with a large number of authors. The number of authors in Chemistry documents never exceeded 30 and the distribution peaks at 3 authors with a slightly asymmetric tail to the larger numbers. The distribution for Physics is somewhat similar, 2 being the most frequent number of authors, but 1.1% of the documents have 20 or more authors with a secondary peak at 23 authors. With

more than 100 authors, we found 17 documents with a number of citations that goes up to a few hundred.

d) The mean citation rate correlates well with the number of co-authors under the main peak of the distribution. The average number of citations may double as we go from 1 to 18 co-authors.

e) For documents with a single author the impact obtained is similar for both disciplines and it grows faster in Chemistry as the number of authors increases.

f) The distribution of the number of addresses reproduces to a certain extent that of the number of authors as should be expected. While the number of addresses in Chemistry documents never exceeded 10, it went as high as 114 in Physics with no clear pattern in this high end tail and some of these documents performed rather well in collecting citations.

g) There is a strong correlation between the mean citation rate and the number of addresses in the peak region up to ca. 6 addresses for Physics and Chemistry.

h) The increase in the mean citation rate with number of addresses is stronger in Physics than in Chemistry.

i) There is a good correlation between the mean citation rate and the journal impact factor for Physics and Chemistry.

Acknowledgements: The financial support of Fundação para a Ciência e a Tecnologia (Lisbon) is acknowledged.

References

Adler, R., Ewing, J & Taylor., P. (2008). Citation Statistics. *IMU Report,* Retrieved 9 January 2009 from:http://www.mathunion.org/Publications/Report/CitationStatistics

Bakkalbasi, N, Bauer, K., Glover, J & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*
2006. Retrieved 12 January 2009 from: http://www.bio-diglib.com/content/pdf/1742-5581-3-7.pdf

Borgman, C. L. & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology,* 36, 3-72.

Cheek, J., Garnham, B. & Quan, J. (2006). What's in a number? Issues in providing evidence of impact and quality of research(ers). *Qualitative Health Research*, 16, 423–435.

Figg, W. D., Dunn, L., Liewehr, D. J., teinberg, S. M., Thurman P. W., Barrett, J. C. & Birkinshaw, J. (2006). Scientific collaboration results in higher citations rates of published articles. *Pharmacotherapy*, 26, 6, 759-767.

Glanzel, W. & Large, C. (2002). A distributional approach to multinationality measures of international scientific collaboration. *Scientometrics,* 54, 1, 75-89.

Glanzel, W. & Thus, B. (2004). Does co-authorships inflate the share of self-citations? *Scientometrics*, 61, 3, 395-404.

Gross, P.L.K. & Gross, E.M. (1927). College libraries and chemical education. *Science*, 66, 385-389.

Jacsó, P. (2005). As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89, 9, 1537-1547;

Lewison, G. (1998). Gastroenterology research in the United Kingdom: funding sources and impact. *Gut*, 43, 288-293.

MacRoberts, M. H. & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics*, 36, 3, 435–444.

Meho, L. I., Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58, 13, 2105-2115

Nicolaisen, J. (2002). The J-shaped distribution of citedness. *Journal of Documentation*, 58, 383-395.

Persson, O., Glanzel, W. & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60, 3, 421-432.

Radicchi, F., Fortunato, S. & Castellano C. (2008). Universality of citation distributions: Towards an objective measure of scientific impact. In M. E. Fisher (Ed.) *Proceedings of the National Academy of Sciences of the United States of America*, 105, 45, 17268-17272.

Rousseau, R. (2000). Are multi-authored articles cited more than single-authored ones? Are collaborations with authors from others countries more cited than collaborations within the country? A case study. In Havemann, F. & Wagner-Döbler, R. & Kretschmer, H. (Ed.), *Proceedings of the second Berlin workshop on scientometrics and informetrics*. (pp. 173-176.)

Tijssen, R.J.W., Van Leeuwen, T.N. & Van Raan, A.F.J. (2002). Mapping the Scientific Performance of German Medical Research. An International Comparative Bibliometric Study, Schattauer, Stuttgart.

Ventura, O. & Mombrú, A. W. (2006). Use of bibliometric information to assist research policy making. A comparison of publication and citation profiles of Full and Associate Professors at a School of Chemistry in Uruguay. *Scientometric*s, 69, 2, 287-313.

Vieira, E.S. & Gomes, J.A.N.F., A comparison of Scopus and Web of Science for a typical university. *Scientometrics*, accept for publication in December, 2008.

Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics,* 62, 1, 117-131.