

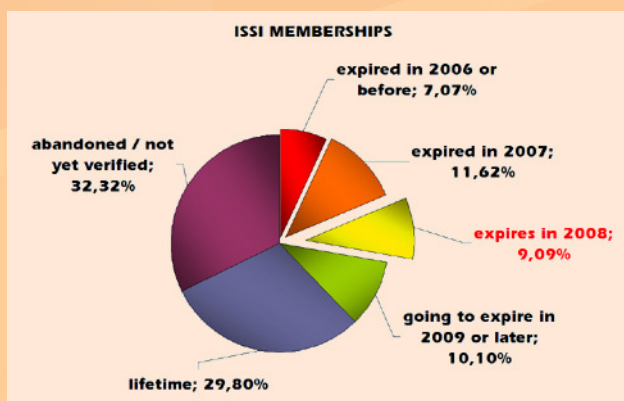
## EDITORIAL

### ■ Time to prolong your ISSI membership...



Another year has passed. Before we invite you from the freezing cold of the Northern hemisphere (given that you are not sitting somewhere close to or South of the Equator at the moment, too) to the tropical summer of the ISSI 2009 conference in Brazil, we are grabbing the opportunity to insert a little statistics here. On this year's 67 pages the ISSI e-Newsletter has published 4 editorials, 6 conference calls, 9 articles, 8 short communications, 2 conference reports and more than 100 figures & photos. We had 16 contributors (Wolfgang Glänzel, Juan Gorraiz, B.M. Gupta, Laura Himanen, Hildrun Kretschmer, Birger Larsen, Jacqueline Leta, Loet Leydesdorff, Bluma C. Peritz, Olle Persson, Ronald Rousseau, Edgar Schiebel, Balázs Schlemmer, András Schubert, Dietmar Wolfram, Ping Zhou) from all over the world. Thanks to their excellent contributions, our Newsletter has started to attract quite a number of citations.

At the end of the year we have another usual announcement: Those members, whose membership expire this calendar year, are kindly



requested to update their membership in time. This is also necessary to guarantee continuous access to the Newsletter. You can find the expiration date on your white or green membership card. Nonetheless, if you have any question concerning your membership status, please, contact us by email. For detailed information consult <http://www.issi-society.info/membership.html>. Owners of a gold membership card are, of course, subscribed for lifetime.

With very best wishes for 2009 we remain yours sincerely,

**Wolfgang Glänzel** *secretary-treasurer ISSI*  
**Balázs Schlemmer** *ISSI administration, technical editor*

## CONTENTS

Editorial .....	<b>54</b>
Conference Call <b>ISSI 2009</b> Conference, Rio ....	<b>55</b>
Summary of <b>Current</b> Funded <b>Metrics-Related Projects</b> (D. Wolfram) .....	<b>57</b>
Conference Report: <b>10th STI Conference</b> , Wien (E. Schiebel & J. Gorraiz) .....	<b>59</b>
The <b>Role of Classics</b> in Social Sciences (O. Persson) .....	<b>63</b>
What are Your <b>Best Papers?</b> (W. Glänzel) .....	<b>64</b>

## EDITORIAL BOARD

### Editor in chief:

Wolfgang Glänzel (*KU Leuven, BEL*)

### Editors:

Judit Bar-Ilan (*Bar-Ilan University, Ramat Gan, ISR*)

Sujit Bhattacharya (*NISTADS, New Delhi, IND*)

María Bordons (*CINDOC, Madrid, ESP*)

Jacqueline Leta (*UFRJ, Rio de Janeiro, BRA*)

Olle Persson (*Umeå University, SWE*)

Ronald Rousseau (*KHBO Oostende, BEL*)

Dietmar Wolfram (*University of Wisconsin-Milwaukee, USA*)

### Technical Editor:

Balázs Schlemmer (*KU Leuven, BEL*)

### Published By:

International Society for Scientometrics and Informetrics

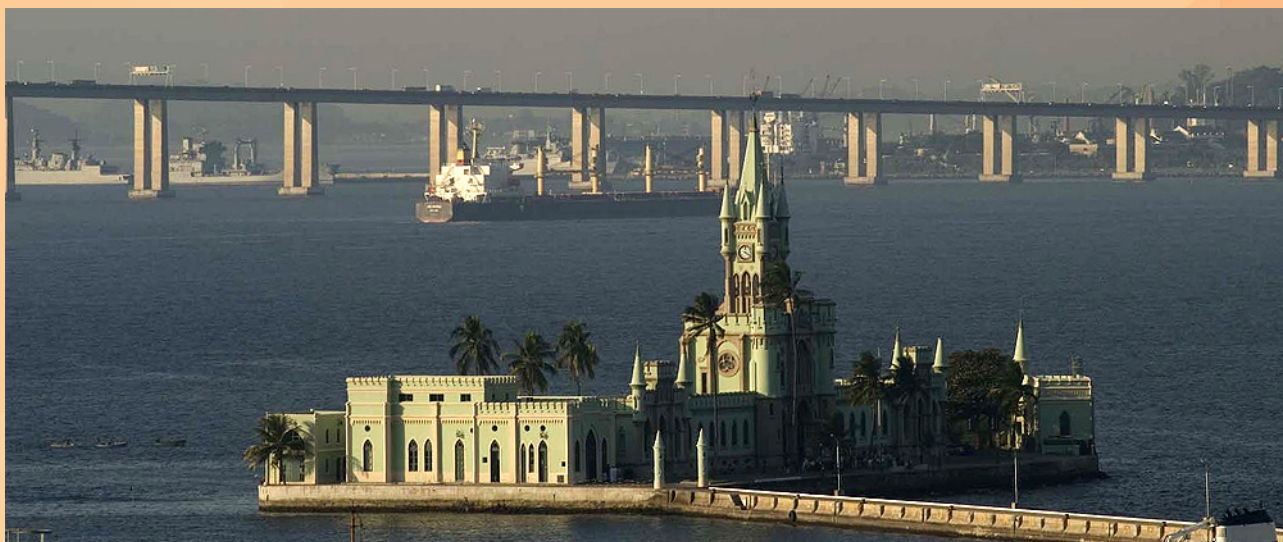


# CONFERENCE CALL

## 12<sup>th</sup> International Conference of the International Society for Scientometrics and Informetrics

On behalf of organizing committee, we are pleased to invite you to participate in the XII International Conference of the International Society for Scientometrics and Informetrics (ISSI), to be held in Rio de Janeiro, Brazil, on July 14-17, 2009. The ISSI Conference will be organized by the Latin American and Caribbean Center on Health Sciences Information of the Pan American Health Organization/World Health Organization (BIREME/PAHO/WHO) and by the Federal University of Rio de Janeiro (UFRJ), been co-organized by a list of public and private partner organizations.

As for paper submission, researchers and students can already send their contributions. Authors have to use the submission form template in one of the three paper formats: full paper, research in progress and poster. Contributions, in English, will be accepted as far as **January 19, 2009**. Authors are recommended to indicate the main topic(s) of their papers, that include: General works on Informetrics, Theory, Methods and techniques, Citation analysis, Indicators, Webometrics, Journals, Open access and electronic publications, Productivity and publications, Collaboration, Research policy, Patent analysis and Bibliographic databases. Doctoral forum and workshop submissions are also opened.





This is the first time that the Society is holding a conference in South America. Thus, among the Conference goals, we do hope it brings a key contribution to the socialization of Scientometrics and Informetrics in the region. The venue is a unit of Federal University of Rio de Janeiro (UFRJ), the oldest Brazilian university. The unit, Forum of Science and Culture, is one of the most charming spaces of UFRJ. It is a building with a neoclassical architecture, located in a fascinating neighborhood very close to Sugar Loaf Mountain.

The organizing committee is been working actively to provide a comprehensive conference program as well as to offer some socio-cultural events in the "Cidade Maravilhosa" ('The Marvelous City'), the nickname for Rio de Janeiro. "Rio" is famous worldwide for its natural beauty, stunning landscape and friendly atmosphere. Among the several options that one has to get to know the city, there are: the Statue of Christ the Redeemer and the Sugar Loaf Mountain. From both places, people can enjoy a breathtaking view of almost the entire city. Many other amazing places are spread in the city, especially the beaches of Copacabana and Ipanema. All these fabulous places are quite close to the Conference venue. Within this atmosphere, we hope to make ISSI 2009 a memorable and rewarding experience for all participants.

More information on the ISSI 2009 – including Committees, Conference Venue, Accommodation and Registration Fees – is available at <http://www.issi2009.org/php/index.php?lang=en>

For questions, please, send us an e-mail by the website "contact".

Hope to meet you in Rio.

**Jacqueline Leta & Birger Larsen**



*Photos: courtesy of Alex Salim. All the photos of the conference call have been published with the permission of the copyright owner.*



ISSI e-Newsletter (ISSN 1998-5460) is published by ISSI (<http://www.issi-society.info/>).

Contributors to the newsletter should contact the editorial board **by e-mail**.

Wolfgang Glänzel: [wolfgang.glanzel@econ.kuleuven.be](mailto:wolfgang.glanzel@econ.kuleuven.be)

Judit Bar-Ilan: [barilaj@mail.biu.ac.il](mailto:barilaj@mail.biu.ac.il)

Sujit Bhattacharya: [sujit\\_academic@yahoo.com](mailto:sujit_academic@yahoo.com)

María Bordons: [mbordons@cindoc.csic.es](mailto:mbordons@cindoc.csic.es)

Jacqueline Leta: [jleta@bioqmed.ufrj.br](mailto:jleta@bioqmed.ufrj.br)

Olle Persson: [olle.persson@soc.umu.se](mailto:olle.persson@soc.umu.se)

Ronald Rousseau: [ronald.rousseau@khbo.be](mailto:ronald.rousseau@khbo.be)

Dietmar Wolfram: [dwolfram@uwm.edu](mailto:dwolfram@uwm.edu)

Balázs Schlemmer: [balazs.schlemmer@gmail.com](mailto:balazs.schlemmer@gmail.com)

Accepted contributions are **moderated** by the board. **Guidelines** for contributors can be found at <http://www.issi-society.info/editorial.html>. Opinions expressed by contributors to the Newsletter do not necessarily reflect the official position of ISSI. Although all published material is expected to conform to ethical standards, **no responsibility** is assumed by ISSI and the Editorial Board for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material therein.

# CURRENT FUNDED METRICS-RELATED RESEARCH PROJECTS

Compiled by Dietmar Wolfram  
School of Information Studies, University of Wisconsin-Milwaukee

Increasing levels of research funding are becoming available through public and private sources to support research in informetrics, scientometrics and allied areas. To broaden awareness of current projects, I'd like to periodically compile summaries of these projects. The summaries will be useful to inform ISSI readers of up-and-coming research being undertaken by colleagues around the world. Also, it will make readers more aware of the range of funding options that are available to support metrics research. Although much of this information is publicly available through funding agency websites or other published venues, these sources are scattered, not universally known, and not comprehensive in their coverage.

Several projects are highlighted in this initial compilation. If you would like to have your currently funded projects included in future summaries, please send me an e-mail ([dwolfram@uwm.edu](mailto:dwolfram@uwm.edu)) with the following information:

- the title of the project
- the investigators and their affiliations
- the funding agency supporting the project
- a brief paragraph summarizing the nature of the project (no more than 150 words)

**Title:** Innovative Scientometric Methods for a Continuous Monitoring of Research Activities in Social Sciences

**Investigators:** Werner Dees, Alexander Botte  
(German Institute for International Educational Research, Frankfurt a.M./Germany)

**Funding Agency:** German Research Foundation  
(Deutsche Forschungsgemeinschaft DFG)

**Abstract:** The project focuses on new or newly applied forms of scientometric analysis that can complement citation-based indicators of research performance, since international instruments for analysing scientific publications include German educational science only marginally, and German databases inadequately support the demands of scientometric analysis. The scientometric project intends to develop and test methodological approaches and technical procedures for an indicator-based monitoring of educational research. It aims at improving the instrumental conditions to evaluation procedures, quality assurance, and transparency of research activities in the field of educational research (which can methodologically be transferred to other social sciences).

**Title:** EERQI - European Educational Research Quality Indicators

**Investigators:** Ingrid Gogolin, University of Hamburg, Germany; Stefan Gradmann, Humboldt University, Germany, plus many partners

**Funding Agency:** EU-funded collaborative project in the 7th Framework Programme for Research in the Socio-economic Sciences and Humanities Theme (SSH).

**Abstract:** The goals of the EERQI project are to reinforce and enhance the worldwide visibility and competitiveness of European educational research. More specifically, the project aims to

- develop new indicators and methodologies to determine quality of educational research publications,
- propose a prototype framework for establishing such indicators and methodologies,
- make this framework operational on a multilingual basis (starting with English, German, French and Swedish),
- produce a search and query engine for resource harvesting and text analysis,
- test transferability of the EERQI indicators into other fields in social sciences and the humanities,
- develop a sustainability plan for quality assessment of European educational research publications.

The project will improve the current standards of research quality indicators, especially for the field of Social Sciences and the Humanities.

**Title:** III-COR:Small: Information Genealogy

**Investigators:** Thorsten Joachims, Cornell University, NY, USA

**Funding Agency:** National Science Foundation, USA

**Abstract:** This project addresses the task of automatically detecting the influence structure and flow of ideas in document corpora that have grown over time (e.g. scientific literature, political debates, news, email, wikis, blogs) in order to trace the origin and development of ideas over time. The intellectual merit of this project lies in the development of statistically well-founded methods for discovering and analyzing the influence structure in evolving archives. In particular, this project will focus on the development of statistical tests for two relations that are central to understanding the structure of an archive and how ideas developed – namely originality and influence. The ability to detect influence and origin of ideas will be of



substantial help in understanding, exploring, interpreting, visualizing, and aggregating the rapidly growing body of historical text available online. The project will also evaluate in how these methods augment traditional citation analysis in hyperlinked collections, and whether they allow similar functionality even in non-hyperlinked archives. The new capability will benefit a number of widely used applications, including search engines for internet content.

**Title:** An Integrated Approach to Web Data Mining for Business Intelligence: Web Content, Web Hyperlinks, and Web Traffic Combined

**Investigators:** Liwen Vaughan, University of Western Ontario, Canada

**Funding Agency:** Social Sciences and Humanities Research Council of Canada

**Abstract:** The Web is becoming an increasingly important place not only to conduct business (e.g. e-commerce) but also to collect business information. Web data mining is a growing area of research in which large scale collection and analysis of Web data is used to discover regularities and patterns. This use of Web data to discover useful business information is called Web data mining for business intelligence. Research on Web data mining has mainly used three types of data sources: Web content, Web hyperlinks, and Web usage. Earlier studies often focused on each data source separately. The proposed program of research will investigate how to use the three data sources together to achieve results better than that from a single data source.

**Title:** Libcitations: A Measure for Comparative Assessment of Book Publications in the Humanities and Social Sciences

**Investigators:** Howard D. White, Drexel University, Philadelphia, PA; Sebastian K. Boell, Hairong Yu, Mari Davis, Concepción S. Wilson, Fletcher Cole University of New South Wales, Sydney, Australia

**Funding Agency:** John Metcalfe Memorial Fund at the University of New South Wales

**Abstract:** Bibliometric measures for evaluating research units in the book-oriented humanities and social sciences are underdeveloped relative to those available for journal-oriented science and technology. We therefore present a new measure designed for book-oriented fields: the "libcitation count." This is a count of the libraries holding a given book, as reported in a national or international union catalog. As librarians decide what to acquire for the audiences they serve, they jointly constitute an instrument for gauging the cultural impact of books. Their decisions are informed by knowledge not only of audiences but also of the book world, e.g., the reputations of authors and the prestige of publishers. From libcitation

counts, measures can be derived for comparing research units. Our work has implications for programs such as Excellence in Research for Australia and the Research Assessment Exercise in the United Kingdom. It also has implications for data mining in OCLC's WorldCat.

**Title:** Citation-based All-author Knowledge Network Analysis

**Investigators:** Dangzhi Zhao, University of Alberta, Canada

**Funding Agency:** Social Sciences and Humanities Research Council of Canada

**Abstract:** Supported by a 3-year, Standard Research Grant from the Social Sciences and Humanities Research Council of Canada (SSHRC), this research project aims to improve our understanding of, and confidence in, a main method available today for studying knowledge networks: citation-based author network analysis. It will focus on designing and testing new citation-based measures to see if and how they can improve the analysis and visual representations of author knowledge networks, and consequently the discovery of new knowledge and the management of existing knowledge resources. To this end, it will also design and test effective methods for building full datasets that enable these new measures, and will design, develop and test software programs for data collection and analysis that are necessary for social scientists to utilize these new methods in their research. The research project is scheduled to run from mid-2008 to mid-2011.

**Title:** Designing Metrics for Impacts and Social Benefits of the Stem Cell Network

**Investigators:** Tania Bubela, University of Alberta (PI); Andreas Strotmann, University of Alberta, Canada

**Funding Agency:** The Stem Cell Network (stemcellnetwork.ca) - a Network of Centres of Excellence Canada

**Abstract:** NCE funding for the Stem Cell Network (SCN) is directed at translation of SC research into clinical applications, commercial products, or public policy. We address the fundamental issue for policy makers of designing metrics and tools to measure benefits arising from such research activities. Using the SCN as a case study, we develop objective metrics and associated software to: (1) assess social and other benefits of SC research and innovation; and (2) assess the impact of Canadian science and technology policy on knowledge flows and scientific collaboration networks. Key milestones are: 1) A prototype software environment for the creation, visualization and analysis of social and scholarly network graphs; 2) Assessment and development of new metrics using bibliometric and scientometric methods, and; 3) Evaluation of the impact of Canadian science and technology policy on knowledge flows and scientific collaboration networks using the SCN as a case study.





# 10<sup>TH</sup> INTERNATIONAL CONFERENCE ON SCIENCE AND TECH- NOLOGY INDICATORS

"Excellence and Emergence – A new  
Challenge for the Combination of  
Quantitative and Qualitative Approaches"  
Vienna, Austria, September 17-20, 2008



Edgar Schiebel



Juan Gorraiz

The 10th International Conference on Science and Technology Indicators took place at the University of Vienna 17-20 September. It was jointly organised by the "Austrian Research Centers GmbH - ARC" and the University of Vienna and was dedicated to "Excellence and Emergence - A new Challenge for the Combination of Quantitative and Qualitative Approaches".

Despite the cold weather, that obliged us to make some last minute modifications, Vienna confirmed its attractiveness and charm as conference place. So finally more than 250 participants registered!

Two pre-conference workshops focused on specific themes and were very well attended. The first one, which discussed the quality and impact of Social Sciences and was entered by renowned specialists in this area (like Anthony Van Raan, Linda Butler, Reinhard Zintl and Hans-Peter Daniel among others), pointed out the necessity to develop new bibliometric strategies in these fields.



The second one, which gave some detailed introduction into the "BibTechMon" methodology as a powerful tool for science mapping, was held by its creator and developer, Edgar Schiebel.

The welcome reception, generously sponsored by Thomson Reuters, took place at the University's Main Ceremonial Chamber, adorned with the reproductions of the famous Klimt "faculty paintings". Typical Austrian food was served while Conference Chair Edgar Schiebel delivered the welcome speech. The ceremony continued accompanied by the sound of saxophone professional Martin Haslinger.

On Thursday, the Rector of the University of Vienna, Georg Winckler, the Director of Austrian





Research Centers, Heinrich Garn, and important representatives of the involved Austrian ministries opened the conference and welcomed all the participants.

Thereafter two brilliant keynotes were presented: "Measuring Science" by

Anthony Van Raan and "Measuring Technology and Innovation" by Hariold Grupp. Unfortunately a third one had to be cancelled due to the sudden and unexpected illness of the speaker, Tibor Braun.

During the next days, 9 selected oral contributions were presented in 3 plenary sessions ("Selected Topics", "Science Policy" and "New Indicators") at any time followed by 63 presentations grouped in 18 sessions distributed in three parallel sessions each.



The conference in Vienna raised the issues of excellence and emergence in science and opened the floor for the discussion about the following 7 subjects:

- Quantitative and qualitative approaches: a special focus in evaluation of the academic performance;
- S&T indicators for the identification of emerging fields;
- Disciplinary relevance of bibliometric indicators: Science and Technology, Social Sciences and Humanities;
- Interactions between Open Access initiatives and Scientometrics;
- Visualisation and Science Mapping: tools, methods and applications;
- Accuracy and reliability of data sources for scientometric studies;
- Management and measurement of bibliometric data within scientific organisations.

Following the tradition of the conference, the authors sent extended abstracts comprised of up to 800 words.

More than 160 contributions were received after the call for papers. Figure 1 shows the distribution of the submitted contributions according to these 7 topics.

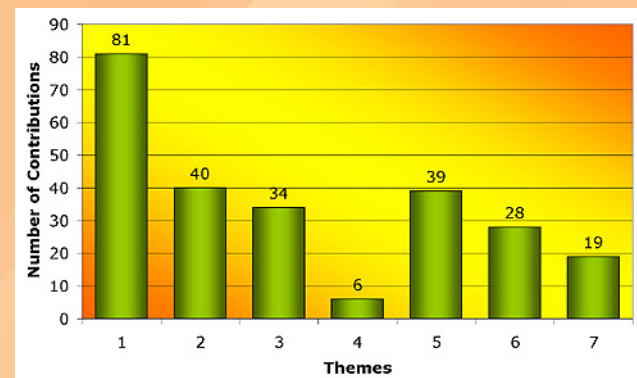


Figure 1 Number of submitted contributions by topic

All contributions were evaluated independently by three reviewers of the International and Local Committees giving quality scores scaling from 1 to 9. As can be seen in Figure 2, the distribution had a bias to higher scores illustrating the quality of the conference content.

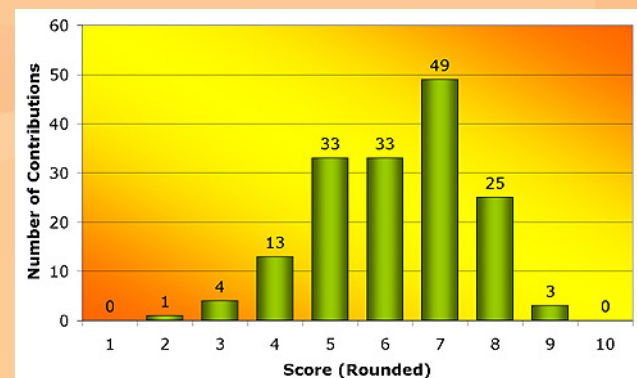


Figure 2 Distribution of scores given by Reviewers

Finally, 137 contributions were accepted for presentation at the conference. Henk Moed, as Programme Co-Chair, was helping us by selecting the oral presentations and organising them into different thematic sessions. We also used text mining, co-word analysis and science mapping to cluster accepted papers and posters to formulate the final programme.

Most of the presentations aroused big interest, so that lively discussions unfolded after them.

For example many of them were dealing with comparisons of established bibliometric data sources (WoS) with newer consolidated ones



(SCOPUS or Google Scholar), their differences and advantages. So a complete session was devoted to "new databases and their indicators". Additionally, there was also an invited presentation: Carmen López Illescas introduced us into the Journal Rank Portal "SCImago", based on Scopus data. This portal experienced a very positive resonance featuring similar characteristics as JCR.



Special attention was dedicated to the 62 approved posters distributed in two main groups: "Research performance, policy and new indices" and "Field studies, data sources and visualisation". They were on show permanently during all of the conference. The three most highly voted were awarded at the closing ceremony on Saturday with a symbolic certification and a succulent "Sacher-Torte". The winners were Elea Giménez-Toledo et al., José Luis Ortega et al. and Ulrich Schmoch et al.

The social events were also very well attended and obviously won the favor of all participants. The first one, on Thursday evening, was a visit to one of the most charismatic Viennese "Heurigen" (wine taverns), "Mayer am Pfarrplatz". This romantic old house has remained completely unchanged since Beethoven

lived there in 1817. In its old courtyard and many cosy rooms, the participants had the possibility to degustate Austrian wine and foods, complemented by the original "Viennese Schrammel-musik" (music for strings).

The second social event, on Friday, was the Gala Dinner at the Main Reading Room of the Vienna University Library, which was specially adapted for this purpose. After the address welcome by the Head of the Library, Maria Seissl, the 10th anniversary of the Conference was commemorated by its founder, Anthony van Raan, with a most entertaining speech. The warm and inviting atmosphere created by this scenario and the beautiful voice of Soprano Diana Higbee singing Mozart arias, converted this dinner in an unforgettable soiree.



By organising the 10th International Conference on Science and Technology Indicators in Vienna we intended to promote scientometric and bibliometric activities in Austria. Alone the number of Austrian participants (see Table 1), the largest so far in a conference in this subject, has recompensed our efforts.

The conference also experienced large and positive response in the press media. Several articles and interviews were published in some



Country	Participants
Austria	33
Netherlands, Spain	23
Germany	19
Belgium	18
France	9
USA	8
Denmark, United Kingdom	7
Czech Republic, Russia	6
Finland, India, Japan, Norway, Switzerland	5
Canada, China, Croatia, Italy, Korea, Turkey	4
Australia, Hungary, Iran, Portugal, Sweden, Taiwan	3
Bangladesh, Estonia, Gambia, Israel, Latvia, Sierra Leone	2
Algeria, Bulgaria, Georgia, Indonesia, Luxembourg, Poland, Saudi Arabia, South Africa, Sri Lanka	1

Table 1 Number of Participants per Country

of the most renowned Austrian newspapers emphasizing the importance of Science and Technology Indicators for the academic life and recalling their indispensability in measuring science and technology.



A "Book of Abstracts" containing all extended abstracts of oral presentations and posters grouped according to their thematic sessions was published and given to all participants. Its contents can be downloaded for free from <http://www.sti2008.at/contributions.htm>.

More photos and a video clip will be soon available at the website of the Conference (<http://sti2008.at>).

Video clips of all the presentations held at plenary sessions can be requested via e-Mail: [juan.gorraiz@univie.ac.at](mailto:juan.gorraiz@univie.ac.at).

On behalf of the Local Organising Committee and Organisers we want to thank Anthony van Raan and Henk Moed for their cooperation and support, all the contributors for their submissions and fruitful discussions, the members of the International Committee for reviewing as well as the sponsors for their generous financial support, and, finally, all the University of Vienna, specially the Event Management Department (Falk Pastner and Nikolaus Ortner) for their superb cooperation.





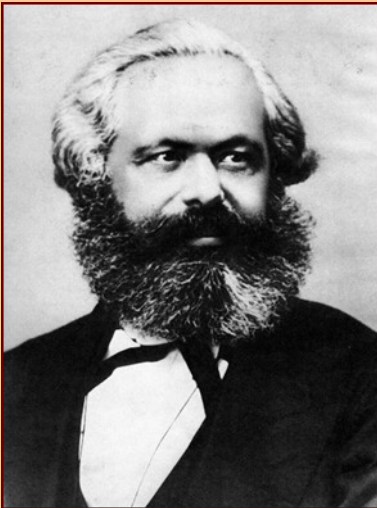


# THE ROLE OF CLASSICS IN SOCIAL SCIENCE

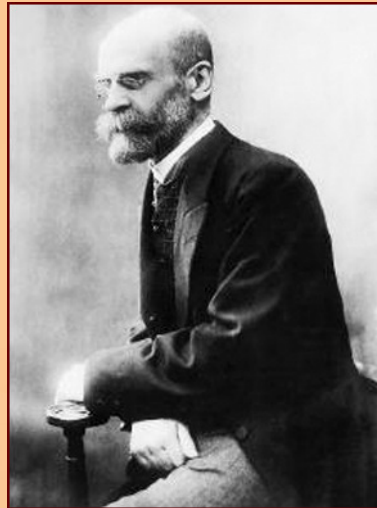
Olle Persson

Dept. Sociology, Umeå university, Sweden

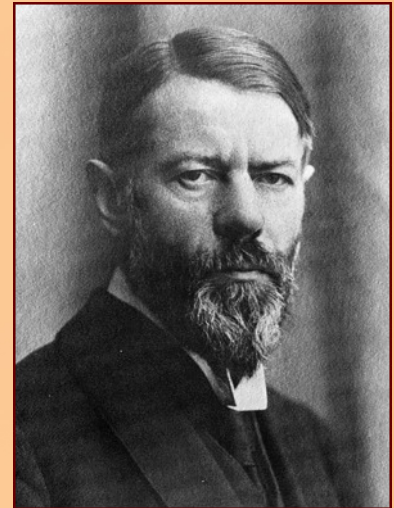
The influence of classical writers in social science can be studied quite easily, since now Web of Science dates back to 1945, and also offers the possibility of analyzing the records online. In social science, particularly sociology and political science, Karl Marx, Max Weber and Émile Durkheim represent lines of thought that still influence current research. If we search for them in the titles of journals in the SSCI part of Web of Science, we can note the strong impact of Marx from the 1960ies up to early 1990ies. Before and after that period the influence of Weber and Durkheim seems to be as strong, or weak, as that of Marx. What we all have witnessed in academia over the last 40 years is quite well concurrent with these trends.



Karl Marx

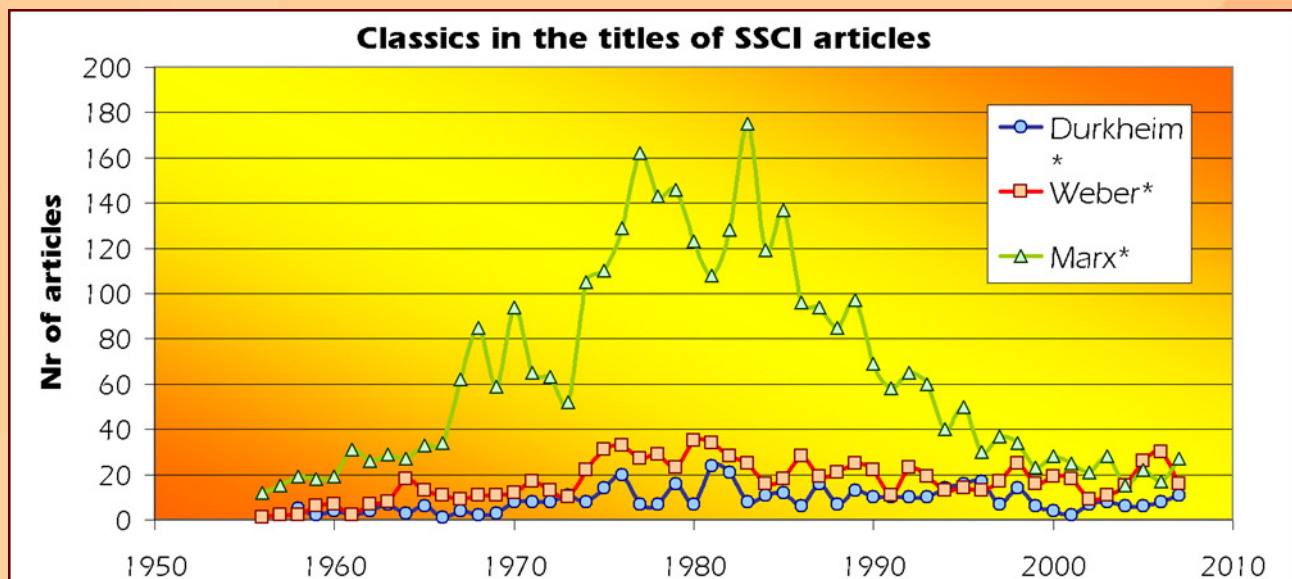


Émile Durkheim



Max Weber

The long period of dominance of Marx doesn't mean anything more than that his thoughts were up to discussion, but nothing in terms of pro or con. Still, scepticism is probably the major reason for the drop. A cited reference search of Marx K, Weber M and Durkheim E would also be possible. Those of you who would like to take up that somewhat more demanding challenge, please do.





# WHAT ARE YOUR BEST PAPERS?



Wolfgang Glänzel

SOOI, KU Leuven, Belgium

IRPS, HAS, Budapest, Hungary

At a first glance, this seems to be kind of rhetorical question but it often becomes a real one. We have to think about this issue, at the latest, when we are to be evaluated for promotion, funding or other purposes. Every now and then we are asked to complete application forms in which, among others, a couple of extraordinary publications should be indicated. Since our answer is used in an evaluative context, we are aware that we have to select these publications very carefully. We are also aware that our own favourites do not necessarily receive the same attention and appreciation from our colleagues. The fact that our favourites are sometimes less cited than our less important work is a permanent source of frustration. Finally, we accept that reviewers are probably not going to read our publications but rather look at impact factors and citation rates and so we are inclined to go for visibility and to list the most cited publications rather than our true favourites.

So far so good. If the number of documents to be listed is given, the problem is solved. Sometimes, however, the decision of how many publications may be chosen is left to the applicant. The most difficult decision in the selection process is finding a suitable threshold or borderline. The h-core (cf. Burrell, 2007) would provide an interesting option as the h-index (Hirsch, 2005) can be used as an ideal truncation point for the tail of

a distribution (cf. Glänzel, 2008). On the other hand, many authors in most areas of the sciences and social have already individual h-indexes of 20, 30 or even more according to the Web of Science. This is in part due to the recent extension of the database but also a consequence of what we called *inflationary bibliometric values* (Persson et al, 2004). Anyhow, a compilation of 20 or 30 papers or more already forms an impressive publication list by itself. Therefore, self-adjusting citation classes reducing this amount would be a real alternative. Such zones or classes have been suggested, for instance, by Glänzel and Schubert (1988) and Zitt et al. (2006). The model of *characteristic scores and scales* (CSS) suggested by Glänzel and Schubert was created to gauge individual performance of a subset or sample against the reference standard set by the underlying population. However, it turns out that this method can be used to measure individual observations on basis of the sample standard itself. In other words, CSS can be used to gauge citation rates of individual papers published by an author against the citation standard of the same author. This means that citation classes based on the author's citation distribution can be used to identify papers with outstanding citation impact with respect to the author's own standard. The number of outstanding papers is thus determined by nothing else but the author's own citation distribution. This method directly provides lists of most cited papers with variable length.

## ■ Characteristic scores and scales for compiling lists of most cited papers

In a nutshell, this method can be summarised as originated from iteratively truncating samples at their mean value and recalculating the mean of the truncated sample until the procedure is stopped or no new scores are obtained. In the present case, we proceed from the citation distribution of the papers by a given author. We calculate truncated samples, which are called *characteristic scores* and are denoted by  $\beta_k$  ( $k \geq 0$ ), in an iterative procedure. We start with  $\beta_0$ , which equals 0 by definition, and  $\beta_1$  being the mean citation rate of the author's papers. Now we discard all papers cited less than the mean. The mean citation rate of the remaining papers is denoted by  $\beta_2$ . Again, those papers cited less



than  $\beta_2$  are removed. The mean citation rate of the rest is  $\beta_3$ . This procedure is repeated till no new scores are obtained (see Glänzel and Schubert, 1988). We denote the greatest obtained index by  $k_{\max}$ . Thus we can define the following zones or classes.  $[0, \beta_1)$  is the class of 'poorly cited' papers,  $[\beta_1, \beta_2)$  contains 'fairly cited' papers,  $[\beta_2, \beta_3)$  and  $[\beta_3, \infty)$  are the two classes of highly cited papers called 'remarkably cited' for  $k=2$  and 'outstandingly cited' papers for  $k=3$ , respectively. If  $k > k_{\max}$  then the corresponding class is empty by definition.

For our purpose we choose the class of *outstandingly cited* papers with  $k = 3$ . In order to illustrate this selection model I have selected three authors representing the group of scientists with about 25 or more years professional experience in three different fields, *mathematics*, *chemistry* and *social sciences*. For this exercise I denote them by A, B and C, respectively. Note that there is no concordance between author codes and the above-mentioned order of subject areas. I have only counted citations for articles, letters, notes, reviews and proceeding papers published in journals. Citation ranks for the 15 most cited papers are presented in Table 1. In addition, the h-index and the mean citation rate are given in the table.

The thresholds  $\beta_k$  defined the following zones,  $[\beta_0, \infty)$  contains all papers,  $[\beta_1, \infty)$  all papers with attribute *fairly*, *remarkably* and *outstandingly* cited,  $[\beta_2, \infty)$  papers that are *remarkably* and *outstandingly* cited and, finally,  $[\beta_3, \infty)$  forms the group of *outstandingly* cited papers. Zones above score  $\beta_3$  usually contain very few papers, or are even empty. The scores  $\beta_k$  for the three authors and the number of papers in the corresponding zones are presented in Table 2. The score and scale for  $k = 3$  is emphasised since this might, from the quantitative viewpoint, form an appropriate set of individual top papers. For those who like to list some more highly-cited papers,  $k = 2$  might provide an alternative threshold; the  $v_2$ -lists contain just a 50% to 70% of the corresponding  $h$  core (cf. Table 1 and 2). However, citation rates of the papers in the zone  $[\beta_2, \beta_3)$  can hardly be considered exceptional if compared with the corresponding individual standards given in Table 1 and 2.

Rank	Citations		
	A	B	C
1	143	86	127
2	67	61	104
3	63	45	102
4	58	36	80
5	43	34	69
6	42	29	66
7	41	28	59
8	40	28	58
9	37	25	53
10	37	22	51
11	37	21	51
12	35	20	50
13	34	17	48
14	34	17	48
15	30	16	47
...	...	...	...
$n$	110	73	161
$h$	25	16	32
$x$	16.25	9.64	19.45

Table 1 The 15 most cited papers of three selected authors A, B and C ( $n$ : number of publications,  $h$ : h-index,  $x$ : mean citation rate)

$k$	$\beta_k$			Papers		
	A	B	C	A	B	C
0	0.0	0.0	0.0	110	73	161
1	16.3	9.6	19.5	43	22	59
2	32.1	26.2	39.5	14	8	22
3	50.8	43.4	59.5	4	3	6
4	82.8	64.0	91.3	1	1	3
5	n/a	n/a	111.0	0	0	1

Table 2 Thresholds and number of papers by authors A, B and C above the thresholds  $\beta_k$  ( $k = 3$  corresponds to the set of outstandingly cited authors)

It should be noticed that the method introduced above provides acceptable results if the distribution is skewed enough. This is usually the case if  $x \leq h$ . The citation distributions of authors A, B and C meet this criterion. Otherwise, if  $x > h$ , the identification of top papers on basis of the suggested method becomes rather difficult. I use the citation distribution of a chemist's publication output just as an example. Let D denote this author. He has 58 papers that received 1954 citations in total. The mean citation rate amounts to 33.86 and the h-index is 21. According to this admittedly unusual distribution 13.8% of all papers are whether less cited than three times and more cited than 47 times each. The most cited papers received 793 citations. In this case  $[\beta_3, \infty)$  is already empty and  $[\beta_2, \infty)$  contains only one paper, namely the most cited one. This example raises the question of how the relation between *characteristic scores and scales* and *h-related indexes* determine the high end of citation distributions. This question will be answered in the following section.

## ■ H-index and A-index as Characteristic Score

In order to solve the task formulated in the previous section, we have to use a more formal definition of CSS. According to Glänzel and Schubert (1988) we first put  $\beta_0 = 0$  and  $v_0 = n$  to obtain the *characteristic scores and scales* of an underlying citation distribution.  $\beta_1$  is then defined as the sample mean

$$1) \quad \beta_1 = \sum_{i=1}^n \frac{X_i}{n} = \sum_{i=1}^n \frac{X_i^*}{v_0},$$

where the citations

$$\{X_i\}_{i=1}^n$$

received by each paper by the author in question are then ranked in descending order  $X_1^* \geq X_2^* \geq \dots \geq X_n^*$ .  $X_1^*$  denotes the citation rate of most frequently cited paper of the author and  $X_n^*$  consequently the number of citations the least cited paper has received.

The value  $v_1$  is defined by the following inequality:

$$2) \quad X_{v_1}^* \geq \beta_1 \text{ and } X_{v_1+1}^* < \beta_1.$$

This procedure is repeated recurrently, particularly,

$$3) \quad \beta_k = \sum_{i=1}^{v_{k-1}} \frac{X_i^*}{v_{k-1}};$$

and  $v_k$  is chosen so that

$$4) \quad X_{v_k}^* \geq \beta_k \text{ and } X_{v_k+1}^* < \beta_k, k \geq 2.$$

The properties  $\beta_0 \leq \beta_1 \leq \dots$  and  $v_0 \geq v_1 \geq \dots$  are obvious from the definition. The interval  $[\beta_k, \infty)$  than contains  $v_k$  papers by definition.

In order to analyse the relation between  $\beta_k$  and  $v_k$  values and h-related indexes one has only modify the initial condition in Eq. (1). Instead of  $\beta_1 = \sum X_i^* / v_1$ , choosing the initial condition  $\beta_1 = v_1$  immediately results in the definition of the h-index since  $X_{v_1}^* \geq v_1$  and  $X_{v_1+1}^* < v_1$  is a version of its definition (cf. Glänzel, 2006). Consequently, Eq. (2) yields Jin's A-index (Jin, 2006), which is actually the average citation rate of the h-core. In particular, one obtains

$$(2^*) \quad A = \beta_2 = \sum_{i=1}^{v_1} \frac{X_i^*}{v_1} \text{ with } \beta_1 = v_1 = h.$$

This is an interesting result that illustrates that both self-adjusting approaches to the statistics of distribution tails, i.e., Hirsch's approach and the *characteristic scores and scales*, result in consistent solutions. Therefore the question arises of both solutions could be also made truly congruent. In order to analyse this, one has to distinguish the following three cases.

- (i)  $h = x$ ,
- (ii)  $h > x$ ,
- (iii)  $h < x$ ,

where  $h$  is the h-index and  $x$  is the mean citation rate.

Case (i): Since  $h$  is always an integer while  $x$  can be an arbitrary rational number, we assume here  $h = \lceil x \rceil$ , where  $\lceil \cdot \rceil$  is called the *ceiling function* giving the smallest integer  $\geq x$ . In this case both the Hirsch and the CSS approach coincide and yield the same results, i.e.,  $\beta_1 = h$  and  $\beta_2 = A$ .

Case (ii): The question arises of whether the h-core could be extended by removing uncited and poorly cited articles so that  $\beta_1 = h = \sum X_i^* / v'_0$  is obtained, where  $v'_0 < n$  is a number that can be used to truncate the original ranked sample to be in line with case (i). This means that the elements  $X_{v'_0+1}^* \geq \dots \geq X_n^*$  have to be removed. Since  $h > x$ , there is always an approximate solution, that is,  $\beta_1 \sim h$ . There for we call the core extension *regular* and the value  $\theta := n - v'_0$  indicates the extent of modification.

Case (iii): In this case the mean exceeds the h-index. Regular core extension is obviously not possible; removing elements from the low end of the ranked sample would further increase the mean value. Therefore we can sequentially add uncited papers until we reach the solution  $\beta_1 = h = \sum X_i^* / v'_0$ . Here  $v'_0 > n$  and  $X_{n+1}^* = \dots = X_{v'_0}^* = 0$ . Again, there is always an approximate solution with  $\beta_1 \sim h$ . This extension is not regular; this is indicated by negative values of  $\theta = n - v'_0$ .

I just mention in passing that in the above examples the following  $\theta$  values are obtained. For author A index  $\theta$  amounts to 45, for B we have  $\theta = 31$ , and for author C the extension index  $\theta$  even amounts to 75. In the case of the somewhat unusual citation distribution of author D one obtains the non-regular extension with  $\theta = -36$ . I admit this is a merely theoretical treatment but it nevertheless emphasises some com-



mon properties of the two approaches and shows that there might emerge problems in cases when the *h*-core contains 'poorly' cited papers according to the author's own standard.

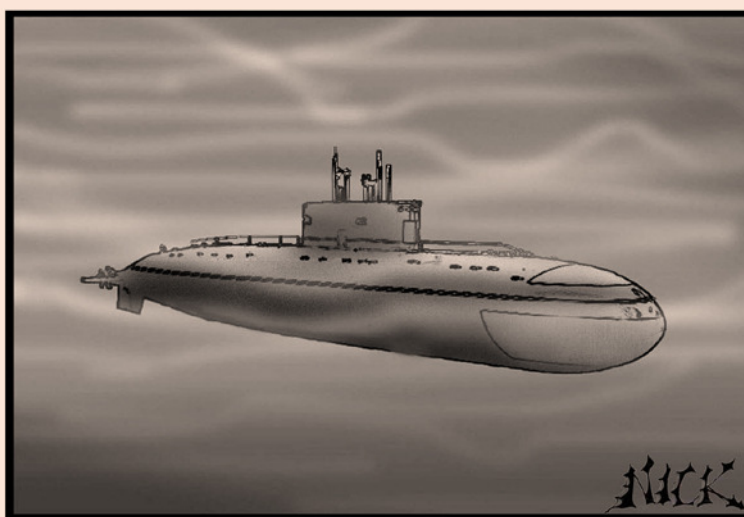
### ■ Concluding remarks

Except for distributions that are not in line with the common citation model, the method of *characteristic scores and scales* (CSS) gives an interesting solution for how to select top papers from an individual citation distribution on the basis of the individual's own standard. Furthermore, the relation between *characteristic scores* and the *h*- and *A*-index was shown and analysed. Both methods provide excellent tools for studying the tail behaviour of bibliometric distributions. One question, of course, remains after all: What are actually your *best* papers?

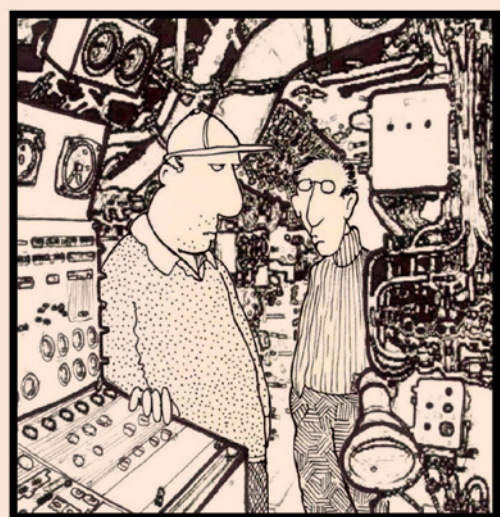
### ■ References

- Burrell, Q.L. (2007) On the *h*-index, the size of the Hirsch core and Jin's *A*-index. *Journal of Informetrics*, 1 (3), 170-177.
- Glänzel, W., Schubert, A. (1988), Characteristic scores and scales in assessing citation impact. *Journal of Information Science*, 14 (2), 123-127.
- Glänzel, W. (2007), Characteristic scores and scales. A bibliometric analysis of subject characteristics based on long-term citation observation, *Journal of Informetrics*, 1 (1), 92-102.
- Glänzel, W. (2008), On some new bibliometric applications of statistics related to the *h*-index. *Scientometrics*, 77 (1), 187-196.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*. 102(46), 16569-16572. (Also available in preprint form as ar Xiv: physics/0508113, accessible at <http://xxx.arxiv.org/abs/physics/0508025>).
- Jin, B. H. (2006). *h*-Index: An evaluation indicator proposed by scientist. *Science Focus*. 1(1), 8-9. (In Chinese)
- Persson, O., Glänzel, 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.
- Zitt M, Bassecoulard E, Filiatreau G, Ramanana-Rahary S. (2007), Revisiting country and institution indicators from citation distributions: Profile performance measures. *Proceedings of ISSI 2007: 11th International Conference of the International Society for Scientometrics and Informetrics*, 797-802.

## CARTOON



Four hundred nautical miles North of Antarctica, and over 100 feet beneath the chill night-time seas, the chartered Kilo Class Type 636 hunter-killer submarine knifed its way silently through some of the most desolate waters on earth. With the two diesel engines shut off, and running on battery at only one-quarter revolutions, the submarine had become the most noiseless vessel ever to slip through the deep.



The crew were weary with tension but quietly patient. Would they be found? Time alone would tell. It can be bloody hard to avoid the relatives at Christmas.

© Nick Kim (Nearing Zero). Reproduced with the permission of the author.