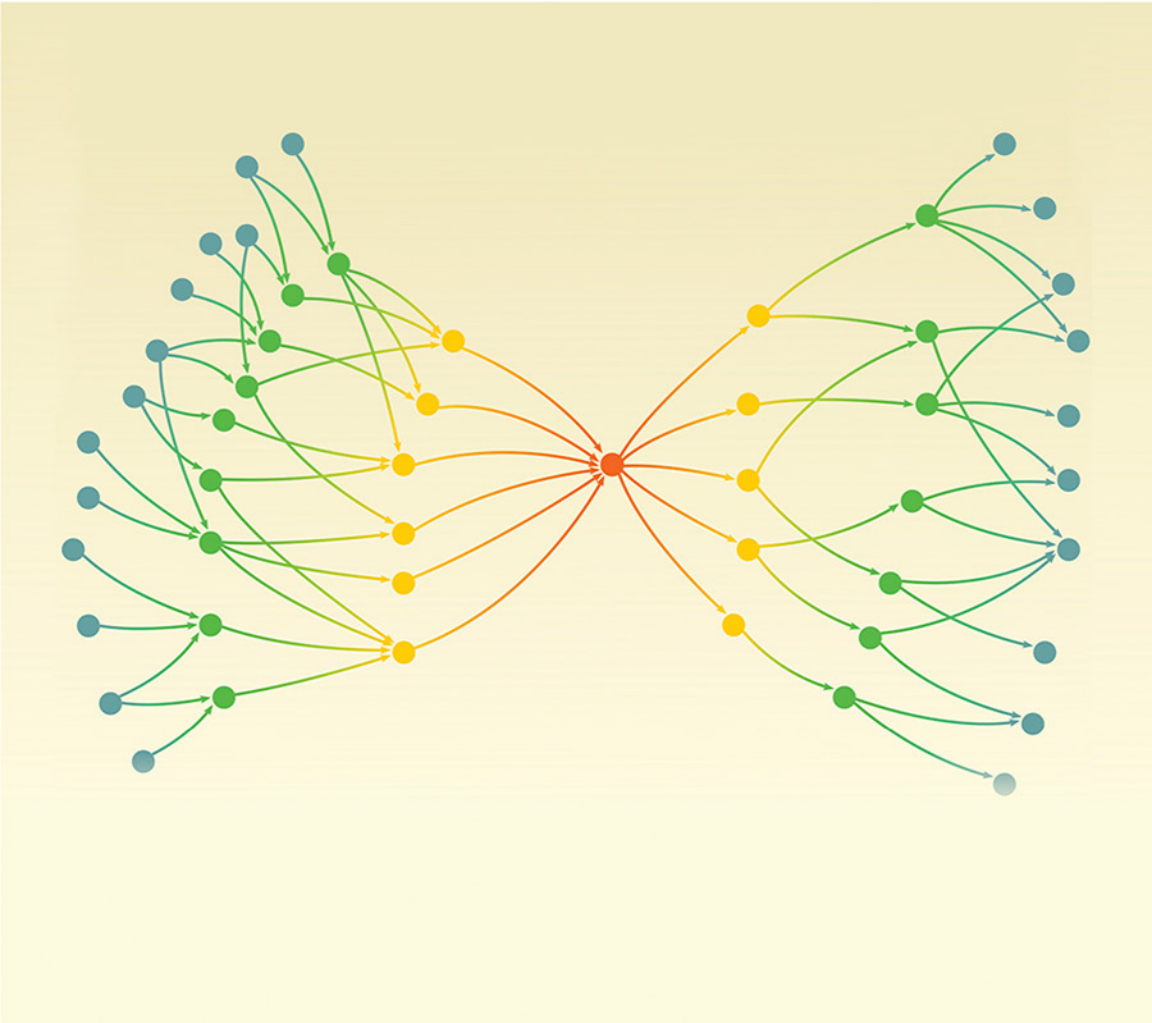
CP
CHANDOS
PUBLISHING

Becoming Metric-Wise

A Bibliometric Guide for Researchers

Ronald Rousseau, Leo Egghe, Raf Guns



BECOMING METRIC-WISE

This page intentionally left blank

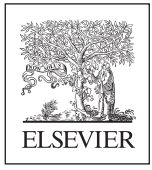
BECOMING METRIC-WISE

A Bibliometric Guide for Researchers

RONALD ROUSSEAU

LEO EGGHE

RAF GUNS



CP
CHANDOS
PUBLISHING
An imprint of Elsevier

Chandos Publishing is an imprint of Elsevier
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, OX5 1GB, United Kingdom

Copyright © 2018 Elsevier Ltd. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability 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 herein.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-08-102474-4 (print)
ISBN: 978-0-08-102475-1 (online)

For information on all Chandos Publishing publications
visit our website at https://www.elsevier.com/books-and-journals

Publisher: Glyn Jones
Acquisition Editor: Glyn Jones
Editorial Project Manager: Charlotte Rowley
Production Project Manager: Debasish Ghosh
Cover Designer: Mark Rogers, Xiaojun Hu and Raf Guns

Typeset by MPS Limited, Chennai, India

CONTENTS

BIOGRAPHIES

Ronald Rousseau was born on August 14, 1949 in Antwerp, Belgium. He is married to Myrian Salembier with whom he has three children.

Rousseau obtained a Doctoral Degree in Mathematics (University of Leuven, 1977), the Habilitation degree in Mathematics (University of Leuven, 1983), and a doctorate in Library and Information Science (University of Antwerp, 1992). He received the Prize of the Belgian Academy of Sciences (1979) for his mathematical work.

He was a full-time mathematics professor at the engineering college KHBO (Oostende) and, over the years, taught different courses for the Education in Library and Information Science at the University of Antwerp (applied bibliometrics—knowledge representation—scientific communication, and research evaluation).

Together with Leo Egghe he wrote *Introduction to Informetrics* (1990), published by Elsevier. In 2001, they received the Derek De Solla Price Award for their work in scientometrics. He is a regular visitor of Chinese universities (with more than 50 Chinese co-authors) and became an Honorary Professor of Zhejiang University and of Henan Normal University. He is the former president of the International Society of Scientometrics and Informetrics (ISSI).

Rousseau is (co-)author of some 500 publications, most of them in journals covered by the Web of Science, but also in local journals (India, China) and journals for mathematics teachers (written in Dutch). His main interest is citation analysis and diversity measurement.

Leo Egghe was born on February 10, 1952 in Beveren, Belgium. He is married to Maria De Bock with whom he has three children.

Egghe has a Doctoral Degree in Mathematics (University of Antwerp, 1978), a Habilitation in Mathematics (University of Antwerp, 1985) and a Ph.D. in Information Science (City University London, 1989).

Egghe was Chief Librarian at the University of Hasselt in the period 1979–2017 and Professor in Library and Information Science at the University of Antwerp in the period 1983–2016 (teaching Information Retrieval and Informetrics). He was also visiting professor in several institutes in the world, among which the Indian Statistical Institute in

Bangalore, India several times. He coordinated development projects in information science in Kenya in the 1990s. Since 2017 he is retired.

Together with Rousseau he organized the first ISSI Conference at LUC, Belgium in 1987, and wrote *Introduction to Informetrics*, published by Elsevier. In 2001 they received the Derek De Solla Price Award for their work in scientometrics.

Egghe is (co-)author of almost 300 scientific publications, most of them in journals covered by the Web of Science. Among these are five books. He invented the g-index in 2006 (a generalization of the h-index) in an article in the journal *Scientometrics*, which is, at this date (2017) the most cited article in this journal. His main interest is the mathematical development of the theory of the Law of Lotka which resulted in a book on this Law in 2005, published by Elsevier.

In 2007, Egghe became the Founding Editor-in-Chief of the *Journal of Informetrics* (Elsevier).

Raf Guns was born on December 15, 1979 in Turnhout, Belgium. He is married to Caroline Van de Velde with whom he has three children.

Guns has a Master's Degree in Germanic Philology (University of Leuven, 2001) and in Library and Information Science (University of Antwerp, 2005). He obtained his Ph.D. in Library and Information Science (University of Antwerp, 2012) under the supervision of Ronald Rousseau.

He has worked for the Library and Information Science program at the University of Antwerp in various functions, including as coordinator, teaching assistant, doctor-assistant, and lecturer in ICT and knowledge organization. Since October 2015, he works as coordinator of the Antwerp branch of the Flemish Center for R&D Monitoring (ECOOM) and as a research administrator at the University of Antwerp's department of research and innovation.

Guns is (co-)author of about 35 scientific publications, mostly in journals that are indexed in Web of Science. In 2013, he was awarded the Emerald/EFMD Outstanding Doctoral Research Award. Since 2015, he has served as an editorial board member of the *Journal of Informetrics*.

PREFACE

This book has two origins, the first one being *Introduction to Informetrics* (1990), a book written by Leo Egghe and Ronald Rousseau, two of the authors of this new book, and a course, entitled *Scientific Communication and Research Evaluation*, taught by these two academics for many years as part of the education in Library and Information Science at Antwerp University (Belgium).

Introduction to Informetrics consisted of four parts: Statistics, aspects of operations research, citation analysis, and informetric models. Over the years, colleagues asked us to update its contents. As this book has a second origin, namely, our course, it is not really an update. Since none of the authors is a specialist in statistics, we decided to keep this part to the bare minimum. Techniques originating in operations research have not become mainstream in our field so that we do not cover this part anymore. The part on citation analysis has been kept and expanded while the part on bibliometric laws is reduced to one chapter, covering its basic aspects. A full account on the so-called "Lotkaian" aspects of the bibliometric laws has been written by Leo Egghe (2005). Besides this, we added—with respect to *Introduction to Informetrics*—chapters on scientific research and communication, publishing, research evaluation and networks. In view of recent developments in informetrics, a basic introduction to research evaluation and to networks was felt to be essential. Moreover, a timeline is provided as a (short) historical overview of the field.

For this new book, "new blood" in the person of a former doctoral student, Raf Guns, was more than welcome. His expertise is in networks and research evaluation, especially in the social sciences and humanities. Raf Guns is currently a coordinator at the Centre for R&D Monitoring at the University of Antwerp.

Quantitative methods to identify statistical regularities, as well as qualitative approaches to explain local deviations from global patterns are both essential for the study of the science of science. This way of thinking perfectly coincides with the ideas underlying the purpose of this book—namely to introduce the basics of bibliometrics and its applications to students and to a global audience of researchers from different backgrounds. We hope that this book is of help to educators preparing

a course related to quantitative aspects of the information sciences. Yet, we hope in addition that researchers, scholars, and doctoral students interested in indicators and science policy will find many parts of this book interesting and useful.

The book is divided into 11 chapters and contains a reference list, preface, an appendix providing the names of the recipients of the Derek J. de Solla Price award, and some indexes. The reference list is, of course, not meant to be exhaustive. Moreover, we apologize to those colleagues who feel they have been unjustly omitted.

We thank our institutes, the former KHBO (Oostende), the University of Hasselt, the University of Antwerp, and KU Leuven for their interest and support. We thank in particular the teaching and supporting staff from the education in Library and Information Science at Antwerp University who, over a period of more than 30 years, created a pleasant working environment, sometimes in difficult external circumstances.

Ronald Rousseau, Leo Egghe and Raf Guns

ACRONYMS

A&HCI	Arts & Humanities Citation Index
ABP	article-based publishing
ABS	Association of Business Schools
AC-IE	Angewandte Chemie – International Edition
AF	audience factor
AIF	average impact factor
AIS	article influence score
ALM	article level metrics
APC	article processing charge
AR	acceptance rate
ARWU	Academic Ranking of World Universities
BCI	Book Citation Index
BOAI	Budapest Open Access Initiative
BRICK	Brazil–Russia–India–China–Korea
BRICKS	Brazil–Russia–India–China–Korea–South Africa
CAB(I)	originally, Commonwealth Agricultural Bureaux (International)
CERN	originally, Conseil Européen pour la Recherche Nucléaire
CPCI-S	Conference Proceedings Citation Index – Science
CPCI-SSH	Conference Proceedings Citation Index – Social Science & Humanities
CpP	citations per publication
CRIS	Current Research Information System
CRIStin	Current Research Information System in Norway
CRL	Center for Research libraries
CSS	Characteristic scores and scales
CWTS	Centrum voor Wetenschap en Technologie Studies = Centre for Science and Technology Studies
DCP	database citation potential (part of SNIP)
DEA	Data Envelopment Analysis
DIF	diachronous impact factor
DOAJ	Directory of Open Access Journals
DOI	digital object identifier
DORA	Declaration on Research Assessment
EIC	editor-in-chief
ESI	Essential Science Indicators
ESCI	Emerging Sources Citation Index
FTE	full-time equivalent
GDP	gross domestic product
GERD	gross domestic expenditure on R&D
GIF	global impact factor
GS	Google Scholar
HEEACT	Higher Education Evaluation & Accreditation Council of Taiwan
HTML	HyperText Markup Language
I3	integrated impact indicator

IC	Index Chemicus
ICMJE	International Committee of Medical Journal Editors
ICT	information and communications technology
IMRaD	Introduction, Methods, Results and Discussion
IPP	Information Production Process
IREG	International Rankings Expert Group
ISBN	International Standard Book Number
ISI	Institute for Scientific Information
ISO	is not an acronym; it refers to the International Organization for Standardization
ISSI	International Society for Scientometrics and Informetrics
ISSN	International Standard Serial Number
JASIS(T)	Journal of the American Association of Information Science (and Technology) later: Journal of the Association of Information Science and Technology
JCR	Journal Citation Reports
JIF	journal impact factor
JUF	journal usage factor
MIF	median impact factor
MNCS	mean normalized citation score
MOCR	Mean Observed Citation Rate
MOOCs	Massive Open Online Courses
NIF	Normalized Impact Factor
NPG	Nature Publishing Group
OA	Open Access
OAI	Open Archives Initiative
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
OCLC	Online Computer Library Center
ODLIS	Online Dictionary for Library and Information Science
ORCID	Open Researcher and Contributor IDentifier
p-c matrix	publication-citation matrix
PAC	Probably Approximately Correct
PDF	Portable Document Format
PLoS	Public Library of Science
QS	Quacquarelli Symonds
RDCP	Relative Database Citation Potential (part of SNIP)
RQI	Review Quality Instrument
RR	rejection rate
SAIF	synchronous author impact factor
SCI(E)	Science Citation Index (Expanded)
SIR	SCImago Institutions Ranking
SJR	SCImago Journal Ranking
SNA	social network analysis
SNIP	Source Normalized Impact per Paper
SSCI	Social Science Citation Index
SSRN	Social Science Research Network
THE	Times Higher Education

UDC	Universal Decimal Classification
UK	United Kingdom
US(A)	United States (of America)
VABB (-SHW)	Vlaams Academisch Bibliografisch Bestand (voor de Sociale en Humane Wetenschappen) = Flemish Academic Bibliography (for the Social Sciences and Humanities)
WIF	Web Impact Factor
WoS	Web of Science

This page intentionally left blank

CHAPTER 1

Introduction

The term *metrics* has become one of the main topics of interest in science in general. The DORA declaration (DORA, 2012), the Leiden Manifesto (Hicks et al., 2015) and the Metric Tide report (Wilsdon et al., 2015) are just the tip of the iceberg of the discussions going on in top journals such as *Nature* and *Science* as well as in evaluation committees all over the world.

For this reason, all scientists should become knowledgeable about indicators used to evaluate them. They should know the publication-citation context, the correct mathematical formulae of indicators being used by evaluating committees, their consequences and how such indicators can be misused. In one word: they should become metric-wise (Rousseau & Rousseau, 2015, 2017). This is exactly the purpose of this book: To make scientists metric-wise. In this way readers will become aware of the evaluation techniques applied to their scientific output, making them stronger when being evaluated for funding, hiring and tenure.

1.1 METRICS IN THE INFORMATION SCIENCES

Information science is defined in the Online Dictionary for Library and Information Science (ODLIS) (Reitz, s.a.) as "The systematic study and analysis of the sources, development, collection, organization, dissemination, evaluation, use, and management of information in all its forms, including the channels (formal and informal) and technology used in its communication." Already in the 1930s, Otlet (1934) defined bibliometrics as the measurement of all aspects related to books and documents. Otlet, a Belgian documentalist, and one of the developers of the Universal Decimal Classification, introduced the idea of bibliometrics in his main work: *Traité de Documentation* (Treatise on Documentation). In this book, a general overview of the information sciences at that time, he covers a plethora of topics related to books, libraries and documentation. Among these he introduced the term "bibliometrics" or more precisely—as the book was written in French—"bibliométrie". This new

Becoming Metric-Wise
DOI: http://dx.doi.org/10.1016/B978-0-08-102474-4.00001-7
© 2018 Elsevier Ltd.
All rights reserved.

science makes use of measurements of library-related objects and facts, and leads to law-like relations. According to Otlet, bibliometrics must include single objects as well as groups of objects (and hence makes use of statistics). Relations between key aspects must give rise to indices. Moreover, sociometric aspects must be taken into account. As an example, Otlet mentions that one should measure how often a book or an author is read. Mathematics must find its place in this new field. Otlet observed that in his time scientific fields tended to use more and more mathematics (physics, chemistry, biology, sociology, economics) and regretted its total absence in library science. Clearly, his ideas preceded similar ones published by Pritchard (1969) by more than 35 years.

Indeed, unaware of Otlet's work, Pritchard (1969) wrote: "*Bibliometrics is the application of mathematics and statistical methods to books and other media of communication. It is the metrology of the information transfer process; its purpose is analysis and control of the process. In short: bibliometrics is the scientific study of recorded discourse*". Pritchard coined the term bibliometrics—or at least he thought he did—to replace the term "statistical bibliography" which had been in some use, but which was not popular, not very descriptive, and could be confused with bibliographies on statistics. Around 1948, Ranganathan proposed the term librametrics (*librametry*), but this proposal has never caught on outside India.

A related term, namely *scientometrics* (*naukometria* in Russian), a term proposed by Nalimov (Nalimov & Mul'chenko, 1969), is defined as the study of the quantitative aspects of science as a discipline or economic activity. It includes research evaluation and aspects of research policy. Yet, we note that nowadays the terms bibliometrics and scientometrics are used by many colleagues without any differentiation. As such, not trying to win a lost battle, we too will treat these two terms as synonyms.

The revolution in information and communications technology (ICT), including the rise of the Internet, changed the way scientists performed research and communicated their results. Consequently in 1979 the new term *informetrics* was proposed by Blackert and Siegel (1979) on the one hand, and by Nacke (1979) on the other. The term gained popularity by the organization of biennial international conferences of which the first was organized in 1987 in Belgium (Egghe & Rousseau, 1988, 1990). The foundation of the International Society for Scientometrics and Informetrics (ISSI) in 1993 under the impetus of Hildrun Kretschmer increased the popularity of the term.

According to Tague-Sutcliffe (1992) and Ingwersen & Björneborn (2004), informetrics is defined as the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists. This definition has been formulated to stress that informetrics is broader than bibliometrics and other metrics that existed at that time. Yet, we think that nowadays this stress is not necessary anymore so we define informetrics as:

The study of the quantitative aspects of information in any form and in any social group.

Although "any social group" implies that informetrics also covers nonscientific information, in practice most informetric studies focus on scientific and scholarly information and its context (producers, consumers, contents, etc.). In other words, most informetric research is also scientometrics or webmetric (digital) research. This will also be the focus of the present book. One may say that informetrics is situated on the intersection between applied mathematics and social sciences.

In the networked world in which we live nowadays, informetrics becomes more like webmetrics (Almind & Ingwersen, 1997; Barabási, 2003; Thelwall, 2004). Here webmetrics is defined as the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Internet drawing on bibliometric and informetric approaches. We admit that the difference between bibliometrics and webmetrics is not always clear, but essentially we would say that using web sources is not webmetrics, but studying their use is.

On October 20, 2010 Jason Priem, Dario Taraborelli, Paul Groth, and Cameron Neylon published a manifesto (Priem et al., 2010) in which they stated that besides classical bibliometric metrics and usage statistics one needs alternative metrics—altmetrics—when evaluating scientists, results of scientific investigations and groups of researchers. They claimed that indicators should evolve with time and hence classical approaches using peer review, counting citations, judging journals by impact factors and so forth must be "extended" by modern, e-based approaches. Concretely, they pointed out that peer review and all citation-based approaches are too slow, and the journal impact factor (JIF) is for most purposes unacceptable.

Nowadays scientists apply Web 2.0 techniques to discuss problems and disseminate results. Written text is often accompanied by data sets, computer code and designs. Via Twitter and other social media scientific

knowledge is distributed in snippets or nano publications. Finally, researchers often place new findings on their personal websites or in blogs and annotate others' work.

Altmetrics is a term referring to all metric techniques measuring new forms of performing, discussing or communicating science, especially through social media. It captures different forms of engagement with an article, a scientist or theory. A clear advantage of these techniques is the fact that they react on the spot and hence are able to map new tendencies and reactions. Not surprisingly journals such as *Nature* immediately reacted to this phenomenon (Piwowar, 2013).

This said, we think that the term altmetrics is a bad choice; its pronunciation resembles "old" metrics, what is alternative today will certainly not be alternative in 10 years, and finally altmetrics is just a special form of informetrics. Perhaps altmetrics should better be called social influmetrics (Rousseau & Ye, 2013) or social media metrics (Haustein, Costas, & Lariviére, 2015). Contemporary methods to describe and evaluate science should include new ways of science communication and the social implications of communicating and performing scientific results. Hence a multimetric approach is called for (Rousseau & Ye, 2013). Aspects of impact captured by altmetrics include (Lin & Fenner, 2013; Taylor, 2013):

> Viewed—*HTML or PDF views on a website, often this is a publisher's website but other websites may also provide view data.*
> Discussed—*in science blogs, Wikipedia, Twitter, Facebook, and similar social media.*
> Saved—*Mendeley, CiteULike and other social bookmark sites.*
> Recommended *(formal endorsement)—a metric used for example by F1000Prime.*
> Cited—*altmetrics also adopts citations in secondary and other knowledge sources, such as number of times a paper has been referenced by Wikipedia.*

If one has access to the data, one can make a distinction between "viewed by other scholars" and "viewed by the public." A similar distinction could be made for the other actions. Sometimes altmetrics is described as the metrics of the computerization of the research process. We do not agree with this as the step of including the computerization of the research process already happened when the term informetrics was introduced. Clearly altmetrics is a subfield of informetrics.

In standard citation studies, the citing population coincides with the cited population in the sense that authors cite authors. This clearly is usually not the case in altmetric studies. A tweeter can be just that. If