APPENDIX

B

Methodology and indicators

Methodology

Our methodology is based on theoretical principles and best practices developed in the field of quantitative science and technology studies, particularly in science and technology indicators research. The *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (Moed, Glänzel, and Schmoch, 2004)[1] gives a good overview of this field. It is based on the pioneering work of Derek de Solla Price (1978),[2] Eugene Garfield (1979),[3] and Francis Narin (1976)[4] in the United States, Christopher Freeman, Ben Martin, and John Irvine in the United Kingdom (1981, 1987),[5] and researchers in several European institutions including the Centre for Science and Technology Studies at Leiden University, the Netherlands, and the Library of the Academy of Sciences in Budapest, Hungary.

[1] Moed H., Glänzel W., & Schmoch U. (2004). *Handbook of Quantitative Science and Technology Research*, Kluwer: Dordrecht.

[2] de Solla Price, D.J. (1977–1978). "Foreword," *Essays of an Information Scientist*, Vol. 3, v–ix.

[3] Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, 1 (4), 359-375.

[4] Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory with application to literature of physics. *Information Processing & Management* 12 (5): 297–312.

[5] Irvine, J., Martin, B. R., Abraham, J. & Peacock, T. (1987). Assessing basic research: Reappraisal and update of an evaluation of four radio astronomy observatories. *Research Policy*, 16(2-4), 213–227.

The analyses of bibliometric data in this book are based on recognized advanced indicators (e.g., the concept of field-weighted citation impact). Our base assumption is that such indicators are useful and valid, although imperfect and partial measures, in the sense that their numerical values are determined by research performance and related concepts, but also by other influencing factors that may cause systematic biases. In the past decade or so, the field of indicators research has developed best practices that state how indicator results should be interpreted and which influencing factors should be considered. Our methodology builds on these practices.

Identifying the relevant documents

For this book, a static version of the Scopus database covering 2000–2019 inclusive was aggregated by country, region, and subject. Subjects were defined by All Science Journal Classification subject areas (see Appendix D for more details).

- The search string[6] used to search nano-publications in Scopus is:

(TITLE-ABS-KEY (nano*) AND NOT TITLE-ABS-KEY (nano2 OR nano3 OR nano4 OR nano5 OR nanosecon* OR "NANO SECON*" OR "NANO GRAM*" OR nanogram* OR nanomol* OR nanophtalm* OR nanomeli* OR nanogeterotroph* OR nanoplankton* OR nanokelvin* OR nanocur)) AND DOCTYPE (ar OR re) AND PUBYEAR >1999 AND PUBYEAR <2020

Between 2000 and 2019, about 75,000 articles or publications in the field of "graphene" have been covered, and more than 54,000 articles have not been covered. "Two-dimensional materials" has 12,000 articles in the reporting period, of which 6,800 articles are not covered. Although the search string has some limitations in recalling all nano-related research outputs, based on the total recorded outputs of 1.4 million publications, the analysis we presented in the book is reliable and solid.

[6] Jiancheng Guan, Nan Ma, China's emerging presence in nanoscience and nanotechnology: A comparative bibliometric study of several nanoscience 'giants', Research Policy, Volume 36, Issue 6,2007, Pages 880–886, ISSN 0048-7333.

Publication types used in the analysis

Throughout this book, analyses only include these publication types that are indexed in Scopus:

- Articles
- Reviews

Counting

All analyses use whole counting rather than fractional counting. For example, if a paper has been coauthored by one author from China and one author from the United States, that paper counts toward the publication count of China as well as the publication count of the United States. Total counts for each country are the unique count of publications.

Bibliometric indicators

Author refers to any individual listed in the author byline of a Scopus-indexed publication.

Citation is a formal reference to earlier work made in an article or patent, frequently to other journal articles. A citation is used to credit the originator of an idea or finding and is usually employed to indicate that the earlier work supports the claims of the work citing it. The number of citations received by an article from subsequently published articles is a proxy of the importance of the reported research.

Cross-sector collaboration (academic–corporate/academic–medical collaboration) is defined as a publication in which either *corporate* or *medical* entities are included in the author affiliation byline in addition to authors affiliated to academic entities. Academic–corporate and academic–medical collaboration are terms used in this book to denote, respectively, the count of collaborations (i.e., research collaborations) between authors from academic and corporate sectors and between authors from academic and medical sectors.

Compound annual growth rate is defined as the year-over-year constant growth rate over a specified period. Starting with the first value in any series and applying this rate for each of the time intervals yields the amount in the final value of the series.

The following definition holds:

$$\mathrm{CAGR}(t_0, t_1) = \left(\frac{V(t_1)}{V(t_0)}\right)^{\frac{1}{t_1 - t_0}} - 1$$

where $V(t_0)$ is the start value and $V(t_1)$ is the final value of a time series and difference $t_1 - t_0$ defines the length of the time interval.

Field-weighted citation impact[7] is an indicator of mean citation impact and compares the actual number of citations received by an article with the expected number of citations for articles of the same document type (article, review, or conference proceeding paper), publication year, and subject field. When an article is classified in two or more subject fields, the harmonic mean of the actual and expected citation rates is used. The indicator is therefore always defined with reference to a global baseline of 1.0 and intrinsically accounts for differences in citation accrual over time, differences in citation rates for different document types (reviews typically attract more citations than research articles, for example), as well as subject-specific differences in citation frequencies overall and over time and document types. It is one of the most sophisticated indicators in the modern bibliometric toolkit (see also **Citation**).

International collaboration (i.e., research collaboration) in this book is indicated by articles with at least two different countries listed in the authorship byline.

Institutional collaboration (i.e., research collaboration) in this book is indicated by articles with a single institute listed in the authorship byline.

[7] Amrita Purkayastha, Eleonora Palmaro, Holly J. Falk-Krzesinski, Jeroen Baas, Comparison of two article-level, field-independent citation metrics: Field-Weighted Citation Impact (FWCI) and Relative Citation Ratio (RCR), Journal of Informetrics, Volume 13, Issue 2, 2019, Pages 635–642, ISSN 1751-577, https://doi.org/10.1016/j.joi.2019.03.012.

Output/publication output for an institute or country is the count of articles with at least one author from that institution or country, respectively (according to the affiliation listed in the authorship byline). All analyses use both full and fractional counting. A **full publication** is a publication to which at least one author affiliated with the institution has contributed.

National collaboration (i.e., research collaboration) in this book is indicated by articles with at least two different institutes from the same country listed in the authorship byline.

Single-author publication is a publication in which only a single author is listed in the authorship byline.

Top 1% highly cited publications are those among the top 1% based on citations of all articles published and cited in a given period. An institution's number or share of highly cited articles is treated as indicative of the excellence of their research.

Topics (as pertaining to Topics of Prominence), refer to nearly 96,000 research topics created using citation patterns of Scopus-indexed publications. The methodology for using citation patterns to define research topics was developed through an Elsevier collaboration with research partners. The advantage of taking a citation-based approach to identify research topics is that one need not rely on identifying all relevant keywords to define a research area. Rather, the research area is delineated by citation patterns in the topic, in which research that appears in the same citation network is clustered together in the same topic. This approach provides a more nuanced definition of the research topic.

Topic clusters are a higher-level aggregation of these research topics based on the same direct citation algorithm that creates the topics. Although topics are easy to understand for the subject experts, they are more difficult for subject generalists. To aid discovery and understanding of the topics, we have taken the topics and aggregated them to around 1500 topic clusters. When the strength of the citation links between topics reaches a threshold, a topic cluster is formed. More details on topics are available at https://service.elsevier.com/app/answers/detail/a_id/28428/.

The **Prominence score** is an indication of the momentum related to a particular topic. The Prominence score is calculated by considering three metrics:

- Citation Count in year n to papers published in n and $n-1$
- Scopus Views Count in year n to papers published in n and $n-1$
- Average CiteScore for year n

The equation for Prominence score for each topic j in year n is:

$$P_j = \frac{0.495\left(C_j - \text{mean}\left(C_j\right)\right)}{\text{stdev}\left(C_j\right)} + \frac{0.391\left(V_j - \text{mean}\left(V_j\right)\right)}{\text{stdev}\left(V_j\right)} + \frac{0.114\left(CS_j - \text{mean}\left(CS_j\right)\right)}{\text{stdev}\left(CS_j\right)}$$

Values that appear in this equation are log-transformations of the raw values in which:

$$C_j = \ln\left(c_j + 1\right)$$
$$V_j = \ln\left(v_j + 1\right)$$
$$CS_j = \ln\left(cs_j + 1\right)$$

C_j = citations in year n of publications in cluster J published in year n and $n-1$

V_j = views in year n of publications in cluster J published in year n and $n-1$

CS_j = weight average Citescore of journals in year n containing publications in cluster J.

Prominence, as defined here, is a linear combination of citations, views, and journal impact for a given topic, in which each factor is normalized by the topic standard deviation.[8]

Topic Prominence score has been shown to correlate with topic level funding per author in a US sample of funded grants. The analysis found that on average, the higher the Prominence score, the more money per US author was available for research on that topic.

[8] https://www.sciencedirect.com/science/article/pii/S1751157717302110.

Patent indicators

Technology Relevance is based on forward citations. Technology Relevance measures whether a patent has been cited more often than have other patents from the same technology field and year, while also considering that international patent offices follow different citation rules. The total number of patent citations received depends on the technology field of the patented invention and also the time that has passed since the patent was published. Patents only recently published tend to have received much fewer citations than older patents. The time-dependency of citations is corrected by dividing the number of citations received by a patent by the average number of citations received by all patents published in the same year.

Technology Relevance also considers that international patent offices follow different citation rules. Therefore, the number of patent citations is corrected for age, patent office citation practice, and technology field. It is a relative measure comparing one patent with other patents. A value of two means that the patent is twice as relevant for subsequent developments as an average patent in the same technology field and of the same age.

Market Coverage: The total size of the worldwide markets in which patent protection exists. The more patents a patentee (in this case an institution or a country the patent owner is affiliated with) owns in important markets, the more valuable the patents are estimated to be. This is because innovators spend more effort and resources on protection in multiple (global) markets via patents if they believe an invention is more valuable.

Market Coverage is calculated based on granted and pending patents; hence, valid patents per country are adjusted for each market's size, as opposed to simple country counts. The size of each market is estimated using the sum of countries' gross national income (GNI) relative to the US GNI as the largest global economy. A Market Coverage of 2 means that the protected markets are, in total, twice as large as the US market alone.

Competitive Impact: The economic value of a patent as measured by its Technology Relevance and Market Coverage. Competitive Impact is stated relative to the other patents in the same field (e.g., a value of 3 means that the patent is three times as important as the average patent in the field).

Patent Asset Index: The Patent Asset Index[9] of a patent portfolio is defined as the aggregate strength of all patents the portfolio contains. The strength of each individual patent is measured by its Competitive Impact. The PatentSight Competitive Impact consists of two dimensions: PatentSight Technology Relevance and PatentSight Market Coverage.

[9] Ernst, H., & Omland, N. (2011). The patent asset index - A new approach to benchmark patent portfolios. World Patent Information, 33(1), 34-41. https://doi.org/10.1016/j.wpi.2010.08.008.