Journal of Documentation

Identifying landmark publications in the long run using field-normalized citation data

Lutz Bornmann, Adam Ye, Fred Ye,

Received 20 July 2017
Revised 6 December 2017
Accepted 17 December 2017

Identifying landmark publications in the long run using field-normalized citation data

Lutz Bornmann
Max Planck Society, Munich, Germany
Adam Ye
Center for Bioinformatics, Peking University, Beijing, China, and
Fred Ye
School of Information Management, Nanjing University, Nanjing, China

Abstract

Purpose – The purpose of this paper is to propose an approach for identifying landmark papers in the long run. These publications reach a very high level of citation impact and are able to remain on this level across many citing years. In recent years, several studies have been published which deal with the citation history of publications and try to identify landmark publications.

Design/methodology/approach – In contrast to other studies published hitherto, this study is based on a broad data set with papers published between 1980 and 1990 for identifying the landmark papers. The authors analyzed the citation histories of about five million papers across 25 years.

Findings – The results of this study reveal that 1,013 papers (less than 0.02 percent) are "outstandingly cited" in the long run. The cluster analyses of the papers show that they received the high impact level very soon after publication and remained on this level over decades. Only a slight impact decline is visible over the years.

Originality/value – For practical reasons, approaches for identifying landmark papers should be as simple as possible. The approach proposed in this study is based on standard methods in bibliometrics.

Keywords Bibliometrics, CSS method, Dynamically normalized impact score,
Highly influential publication, Landmark publication, Outstandingly cited publication

Paper type Research paper

1. Introduction

Besides peer review, research evaluation is frequently based on indicators (Bornmann, 2011; Moed and Halevi, 2015). The most prominent group of indicators comprises bibliometric indicators with the basic indicators being number of publications and their citation impact – as a proxy of quality (Wilsdon *et al.*, 2015). "Bibliometric assessment of research performance is based on one central assumption: scientists, who have to say something important, do publish their findings vigorously in the open, international journal literature" (van Raan, 2008, p. 463). Also, citations are rooted in the basic requirements of scientific activity: "It is one of the basic rules of scientific research that a piece of written research, in order to warrant publication, needs to be adequately situated within the existing research literature" (Woelert, 2013, p. 350).

Since bibliometric indicators are frequently used in evaluative contexts and evaluations focus as a rule on the last few years, bibliometric reports often refer to recent publication years. Also, indicators are used which are generated on the base of short time frames: the frequently applied Journal Impact Factor only considers the publications from two years and the citations from the following year. However, the focus of research evaluation on short and recent time frames obscures the fact that scientific progress is usually oriented toward discoveries which prove successful in the long run. "Scientific progress is driven by important, infrequent discoveries that cannot be readily identified and quantified, which makes research assessment very difficult" (Rodríguez-Navarro, 2016, p. 731). According

Journal of Documentation
Vol. 74 No. 2, 2018
pp. 278-288

0022-0418
DOI 10.1108/JD-07-2017-0108

to van Raan (1996), "quality is a measure of the extent to which a group or an individual scientist contributes to the progress of our knowledge. In other words, the capacity to solve problems, to provide new insight into 'reality', or to make new technology possible" (p. 398).

In this study, we follow activities like those of Redner (2005), Ponomarev *et al.* (2014), and Winnink *et al.* (2016) who tried to identify breakthrough-class discoveries in science using bibliographic information. In a further empirical study using the method of group-based trajectory modeling, Baumgartner and Leydesdorff (2014) distinguish between sticky and transient knowledge claims. Sticky knowledge claims continue to be cited more than ten years after publication and transient knowledge claims show a decay pattern after reaching a peak within a few years. Colavizza and Franceschet (2016) analyzed the citation histories of papers over many years based on normalized citation data and used spectral clustering to group the histories: "Citation histories extend citation counts by adding a temporal dimension, providing a more informative and less immediate indication of the impact of a publication. While citation counts are snapshots of publication impact at a given time, citation histories move publication impact over time and map a publication's ageing process" (p. 1037).

Using a comprehensive data set in this study, we identify all papers which can be labeled as sticky knowledge claims or landmark papers, respectively. This group of papers is defined as highly influential publications in the corresponding fields. With all articles published between 1980 and 1990, our study is based on a much broader database than the studies by Baumgartner and Leydesdorff (2014), Colavizza and Franceschet (2016), and others.

2. Methods

2.1 Data

Table I shows the number of papers from 1980 to 1990 which have been considered in this study. The bibliometric data are from an in-house database developed and maintained by the Max Planck Digital Library (Munich). The in-house database is based on the Web of Science (Clarivate Analytics, formerly the IP & Science business of Thomson Reuters). From the in-house database, we selected only papers with the document type "article" to have comparable citable units. The citation impact for each paper refers to the period from its publication year until the end of 2015. As we are interested in identifying landmark papers based on citations, we ignored all non-cited articles (about 1 percent). In order to compare the citation impact of the papers over comparable time periods, we used citation windows of 25 years. For example, the citations refer to the years 1981-2005 for the papers from 1980 and 1982-2006 for the papers from 1981.

Publication year	Total number of articles	Non-cited articles	Cited articles
1980	402,417	6,594	395,823
1981	423,754	6,381	417,373
1982	438,201	6,244	431,957
1983	464,131	5,868	458,263
1984	479,977	5,422	474,555
1985	495,496	4,705	490,791
1986	508,608	3,856	504,752
1987	524,467	3,077	521,390
1988	539,656	2,192	537,464
1989	558,316	1,037	557,279
1990	536,566	0	536,566
Total	5,371,589	45,376	5,326,213

Table I. Numbers of articles published between 1980 and 1990

2.2 Field normalization of citation impact

This study uses standard impact scores in bibliometrics, namely field- and time-normalized citation impact scores (in a dynamical variant) (Vinkler, 2010). These scores are defined as:

$$\mathrm{DNIC}_{it} = \frac{C_{it}}{E_{ft}}, f = f(i) \tag{1}$$

$$E_{ft} = \frac{1}{N_{ft}} \sum_{i|f=f(i)} C_{it} \tag{2}$$

where $i = 1,2, \ldots$ are publications, $t = 1,2, \ldots$ are citing years, and $f = 1,2, \ldots$ are fields. Here, field delineations based on OECD minor codes[1] are used. C_{it} denotes citations received by publication i in year t, and E_{ft} denotes mean (received) citations of all publications in field f and year t (i.e. E_{ft} is the expected value). N_{ft} is the number of cited publications in field f and year t (N_{ft} is based on publications with at least one citation), and $f = f(i)$ means a certain field of a given publication. The indicator follows one of the most frequently used approaches in bibliometrics with both field- and time-normalized citations (Waltman, 2016). The difference from this standard approach in bibliometrics is that the calculation is based on annual citations (dynamically), but not on all citations between publication year and a fixed time point later on.

If $C_{it} = 0$, then $\mathrm{DNIC}_{it} = 0$. If $\mathrm{DNIC}_{it} > 1$, the citation impact of the publication is higher than the average in the corresponding field and (cited as well as citing) publication years. If $\mathrm{DNIC}_{it} < 1$, the impact is lower than the average. In practical terms, citation counts C_{it} and expected values E_{kt} frequently vary. The DNIC distribution of many papers changes from year to year.

2.3 Classifying of publications using the characteristic scores and scales (CSS) method

Glänzel and Schubert (1988) introduced the characteristic scores and scales (CSS) method for grouping ranked observations into rank-specific categories (see also Glänzel, 2007, 2010, 2011). Consider a set of n papers. The observed citations X_i received by paper i are ranked in descending order, $X_1^* \geqslant X_2^* \geqslant \cdots \geqslant X_n^*$, where X_1^* and X_n^* denote the citations of the most and least frequently cited papers, respectively. Set the initial values $\beta_0 = 0$ and $v_0 = n$, where n is the number of papers. β_1 is defined as the mean citations; v_1 is defined by the comparison $X_{v_1}^* \geqslant \beta_1$ and $X_{v_1+1}^* < \beta_1$. This comparison is repeated, yielding:

$$\beta_k = \sum_{i=1}^{v_{k-1}} \frac{X_i^*}{v_{k-1}} \text{ with } X_{v_k}^* \geqslant \beta_k \text{ and } X_{v_k+1}^* < \beta_k, \text{ for } k \geqslant 2 \tag{3}$$

Thus, we obtain series $\beta_0 \leqslant \beta_1 \leqslant, \ldots$ and $v_0 \geqslant v_1 \geqslant, \ldots$. The kth class is defined by the pair of threshold values $[\beta_{k-1}, \beta_k]$; the number of papers belonging to this class amounts to v_{k-1}-v_k.

The CSS method can be used to classify the papers within certain fields into four impact classes: "poorly cited," "fairly cited," "remarkably cited," and "outstandingly cited." Then, for example, the share of outstandingly cited papers can be determined for a set which includes papers from different fields (e.g. all papers published by a university). Bornmann and Glänzel (2017) proposed using the CSS method to classify the universities in a specific ranking (e.g. the Leiden ranking) into performance classes (e.g. based on the number of highly cited papers). Then, the universities can be separated into low and (very) high performers.

In this study, we use the CSS method for classifying the papers from 1980 to 1990 into four citation impact classes based on DNIC_{it}: Consider the set $\{\mathrm{DNIC}_{it}\}$ of n papers published in various fields (field delineations are based on OECD minor codes).

The observed citations C_{ij} received by paper i in a given year j are used to calculate DNIC_{it} values which are ranked in descending order $\left(\mathrm{DNIC}_1^* \geqslant \mathrm{DNIC}_2^* \geqslant \cdots \geqslant \mathrm{DNIC}_n^*\right)_t$. The comparison between DNIC and β is defined by:

$$\beta_{kt} = \sum_{i=1}^{v_{k-1}} \frac{\mathrm{DNIC}_{it}^*}{v_{k-1}},\ \mathrm{DNIC}_{v_k t}^* \geqslant \beta_{kt} \ \text{ and } \ \mathrm{DNIC}_{v_k t+1}^* < \beta_{kt} \tag{4}$$

Then, the pair of threshold values $[\beta_{k-1}, \beta_k]$ forms the impact class. We used the CSS method to categorize the annual citation impact of the papers – the annual DNIC scores – into impact classes. The values of the annual DNIC scores are kept with min $k \geqslant 2, 3, \ldots$, respectively, in every year after the publication year. Since the values $k \geqslant 2, 3, \ldots$ are used to identify highly cited papers (Glänzel, 2011), we set $k \geqslant 2$ as "fairly cited" papers, $k \geqslant 3$ as "remarkably cited" papers, and $k \geqslant 4$ as "outstandingly cited" papers. Furthermore, we identify with $k \geqslant 5$ and $k \geqslant 6$ two additional top-groups within the group of "outstandingly cited" papers.

2.4 Cluster analysis

Units in data can be grouped by using different methods of cluster analysis. According to Kaufman and Rousseeuw (1990), "cluster analysis is the art of finding groups in data" (p. 1). In this study, we cluster the bibliometric data by the K-means clustering algorithm, which is a popular method of vector quantization. Given a set of n observations $(x_1, x_2, \ldots, x_n)$, where each observation is a d-dimensional real vector $(x_i = (x_{i1}, x_{i2}, \ldots, x_{id}))$, K-means clustering aims to determine K centers. The clustering procedure partitions the n observations into K ($\leqslant n$) non-overlapping sets $S = \{S_1, S_2, ..., S_K\}$ according to their minimum distance to each center. K is predetermined, and the objective is to minimize the within-cluster sum of squares (WCSS):

$$\underset{\text{partition}}{\arg\min}_S \sum_{k=1}^{K} \sum_{x_i \in S_k} \left\| x_i - \mu_k \right\|^2 = \underset{S}{\arg\min} \sum_{k=1}^{K} \sum_{x_i \in S_k} \sum_{j=1}^{d} \left(x_{ij} - \mu_{kj} \right)^2 \tag{5}$$

where μ_k is the mean of the points in S_k:

$$\mu_k = \frac{1}{|S_k|} \sum_{x_i \in S_k} x_i$$

arg min over partition S denotes the process of finding out the partition S which minimizes the latter formula. According to the rule $f(K) < 0.85$ (Pham *et al.*, 2005), the values 2 or 3 can be selected for K provided there are around 1,000 observations.

3. Results

We used the annually classified papers from 1980 to 1990 to identify the landmark papers. These are papers which receive a very high impact across decades. Furthermore, we clustered the data to identify patterns in impact distributions across the citing years.

3.1 Distributions of outstandingly cited papers

In order to illustrate the distribution of annual DNIC scores, we selected randomly from every publication year one outstandingly cited publication ($k \geqslant 4$). The distributions of the 11 random papers are shown in Figure 1. For comparison, the figure also includes the distributions of bare citation counts of the same papers. As the comparison shows, distributions of citation counts with increasing or constant impact over the years for most of the papers are visible for the raw citation counts, but not for the annual DNIC scores. This result indicates that the impact normalization method used in this study works effectively.

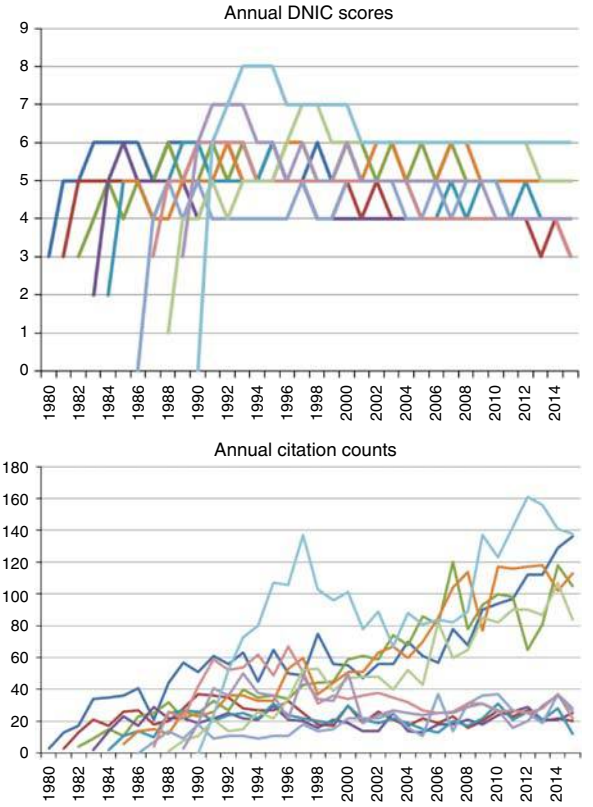

Figure 1.
Annual DNIC scores for eleven randomly selected, outstandingly cited publications ($k \geq 4$)

Note: For comparison, the raw citation counts of the same papers are shown

3.2 Identifying landmark papers

Table II shows the number of articles in different citation impact classes – measured across 25 years. In total, 16,966 papers can be named as "fairly cited" ($k \geq 2$) since their impact was constantly above the average in the corresponding field. 4,256 papers are labeled as "remarkably cited" with $k \geq 3$. We defined three classes of papers which have been outstandingly cited ($k \geq 4$, $k \geq 5$, and $k \geq 6$). Table III lists the top group of outstandingly cited papers with $k \geq 6$ ($n = 40$). For example, one of the listed papers (Bland and Altman, 1986), deals with methods for comparing different measurement techniques. Since the use of correlation coefficients in these comparisons leads to misleading results, an alternative approach, based on graphical techniques and simple calculations, is introduced in this paper. The paper had received 27,585 citations by February 21, 2017.

Further analyses revealed that most of the papers in Table II have been published in 1981 ($n = 8$) and 1990 ($n = 7$). Five papers appeared in the *Journal of the American Chemical Society* and three in *The Lancet*. In total, 80 percent of the papers had at least one author from the USA, 15 percent at least one author from the UK. Many journals which published papers in Table II have been assigned to the subject categories "Chemistry, Multidisciplinary" ($n = 6$, 12 percent of the papers), "Medicine, General & Internal" ($n = 5$), and "Biochemistry & Molecular Biology" ($n = 5$).

Table II.
Number of articles in different citation impact classes across the citing years (in absolute numbers and percent)

	Fairly cited $k \geq 2$		Remarkably cited $k \geq 3$		$k \geq 4$		Outstandingly cited $k \geq 5$		$k \geq 6$	
Publication year	Absolute	In percent	Absolute	In percent	Absolute	In percent	Absolute	In percent	Absolute	In percent
1980	1,085	0.274	292	0.074	79	0.02	18	0.00455	3	0.00076
1981	1,112	0.266	299	0.072	88	0.021	25	0.00599	9	0.00216
1982	1,167	0.27	295	0.068	70	0.016	14	0.00324	4	0.00093
1983	1,297	0.283	320	0.07	79	0.017	12	0.00262	1	0.00022
1984	1,374	0.29	348	0.073	81	0.017	17	0.00358	3	0.00063
1985	1,535	0.313	403	0.082	93	0.019	16	0.00326	3	0.00061
1986	1,558	0.309	387	0.077	108	0.021	22	0.00436	3	0.00059
1987	1,714	0.329	401	0.077	99	0.019	18	0.00345	5	0.00096
1988	1,857	0.346	471	0.088	75	0.014	15	0.00279	3	0.00056
1989	2,100	0.377	513	0.092	113	0.02	23	0.00413	4	0.00072
1990	2,167	0.404	527	0.098	128	0.024	33	0.00615	8	0.00149
Total	16,966	0.319	4,256	0.08	1,013	0.019	213	0.004	46	0.00086

Notes: Since papers are assigned to more than one subject category in our data set, some papers are counted multiply. For example, the number of outstandingly cited papers with $k \geq 6$ reduces to 40 if doublets are excluded

First author	Year	Title	Journal
Biersack, J.P.	1980	A Monte-Carlo computer program for the transport of energetic ions in amorphous targets	*Nuclear Instruments & Methods*
Binkley, J.S.	1980	Self-consistent molecular-orbital methods 0.21. Small split-valence basis-sets for 1st-row elements	*Journal of the American Chemical Society*
Bower, G.H.	1981	Mood and memory	*American Psychologist*
Beasley, R.P.	1981	Hepatocellular-carcinoma and hepatitis-b virus – a prospective-study of 22,707 men in Taiwan	*The Lancet*
Hamill, O.P.	1981	Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches	*Pflugers Archiv – European Journal of Physiology*
Vale, W.	1981	Characterization of a 41-residue ovine hypothalamic peptide that stimulates secretion of corticotrophin and beta-endorphin	*Science*
Anderson, S.	1981	Sequence and organization of the human mitochondrial genome	*Nature*
Burnette, W.N.	1981	Western blotting – electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radio-iodinated protein-a	*Analytical Biochemistry*
Dziewonski, A.M.	1981	Preliminary reference earth model	*Physics of the Earth and Planetary Interiors*
Guth, A.H.	1981	Inflationary universe – a possible solution to the horizon and flatness problems	*Physical Review D*
Bandura, A.	1982	Self-efficacy mechanism in human agency	*American Psychologist*
Bennett, J.M.	1982	Proposals for the classification of the myelodysplastic syndromes	*British Journal of Haematology*
Kyte, J.	1982	A simple method for displaying the hydropathic character of a protein	*Journal of Molecular Biology*
Ando, T.	1982	Electronic-properties of two-dimensional systems	*Reviews of Modern Physics*
Hanahan, D.	1983	Studies on transformation of Escherichia coli with plasmids	*Journal of Molecular Biology*
Church, G.M.	1984	Genomic sequencing	*PNAS*
Weiner, S.J.	1984	A new force-field for molecular mechanical simulation of nucleic-acids and proteins	*Journal of the American Chemical Society*
Shechtman, D.	1984	Metallic phase with long-range orientational order and no translational symmetry	*Physical Review Letters*
Grynkiewicz, G.	1985	A new generation of ca-2+ indicators with greatly improved fluorescence properties	*Journal of Biological Chemistry*
Yanischperron, C.	1985	Improved m13 phage cloning vectors and host strains – nucleotide-sequences of the m13mp18 and puc19 vectors	*Gene*
Dewar, M.J.S.	1985	The development and use of quantum-mechanical molecular-models 0.76. Am1 – a new general-purpose quantum-mechanical molecular-model	*Journal of the American Chemical Society*
Bland, J.M.	1986	Statistical methods for assessing agreement between two methods of clinical measurement	*The Lancet*
Mosmann, T.R.	1986	2 types of murine helper t-cell clone 0.1. Definition according to profiles of lymphokine activities and secreted proteins	*Journal of Immunology*
Bednorz, J.G.	1986	Possible high-tc superconductivity in the ba-la-cu-o system	*Zeitschrift für Physik B – Condensed Matter*
Engle, R.F.	1987	Cointegration and error correction – representation, estimation, and testing	*Econometrica*

Table III.
40 outstandingly cited articles with $k \geqslant 6$

(*continued*)

First author	Year	Title	Journal
Chomczynski, P.	1987	Single-step method of RNA isolation by acid guanidinium thiocyanate phenol chloroform extraction	*Analytical Biochemistry*
Alley, M.C.	1988	Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay	*Cancer Research*
Arnett, F.C.	1988	The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid-arthritis	*Arthritis and Rheumatism*
Vogelstein, B.	1988	Genetic alterations during colorectal-tumor development	*New England Journal of Medicine*
Sikorski, R.S.	1989	A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae	*Genetics*
Anders, E.	1989	Abundances of the elements – meteoritic and solar	*Geochimica et Cosmochimica Acta*
Stewart, J.J.P.	1989	Optimization of parameters for semiempirical methods 0.1: Method	*Journal of Computational Chemistry*
Bain, C.D.	1989	Formation of monolayer films by the spontaneous assembly of organic thiols from solution onto gold	*Journal of the American Chemical Society*
Johansen, S.	1990	Maximum-likelihood-estimation and inference on cointegration – with applications to the demand for money	*Oxford Bulletin of Economics and Statistics*
Bentler, P.M.	1990	Comparative fit indexes in structural models	*Psychological Bulletin*
Macmahon, S.	1990	Blood-pressure, stroke, and coronary heart-disease 0.1. Prolonged differences in blood-pressure – prospective observational studies corrected for the regression dilution bias	*The Lancet*
Levy, D.	1990	Prognostic implications of echocardiographically determined left-ventricular mass in the Framingham heart study	*New England Journal of Medicine*
Hunter, C.A.	1990	The nature of pi-pi interactions	*Journal of the American Chemical Society*
Sheldrick, G.M.	1990	Phase annealing in shelx-90 - direct methods for larger structures	*Acta Crystallographica Section A*
Gelfand, A.E.	1990	Sampling-based approaches to calculating marginal densities	*Journal of the American Statistical Association*

Table III.

3.3 Cluster analysis

We used the rule $f(K) < 0.85$ (Pham *et al.*, 2005) and selected the two and three clusters solutions for the 1,013 outstandingly cited papers (with $k \geqslant 6$). Both solutions lead to distributions including a monotonic decline as shown in Figure 2. The distributions can be named as "middle-of-the-roads" following Colavizza and Franceschet (2016): "publications which start fast or slow, reach a moderate peak and keep improving the ratio of received citations, or at least keep being relevant over prolonged amounts of time by manifesting a slow decline or a plateau" (p. 1043).

Figure 2 demonstrates that the grouped papers have similar distributions, but on a different level. For example, the three clusters solution separates the papers into three citation impact groups where 10.6 percent of the papers constantly performed on a very high level.

4. Discussion

The standard distribution of citation impact over citing years starts with a rapid increase and slow decrease, which generally happens on a relatively low impact level. Instead,

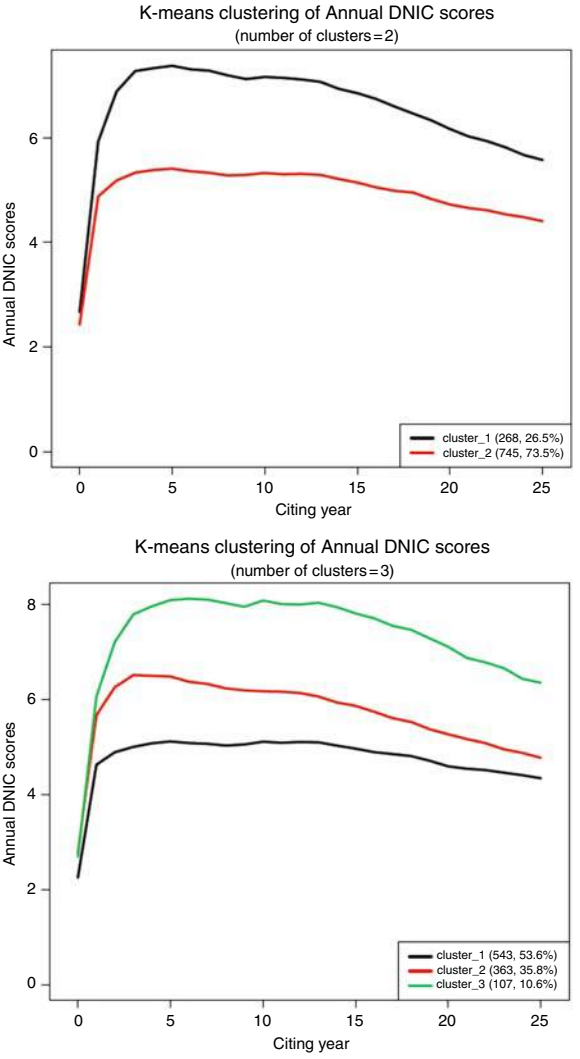

Figure 2.
Solutions with two and three clusters of 1,031 outstandingly cited publications

landmark publications reach a very high level of citation impact and are able to remain on this level across many citing years.

In recent years, several studies have been published which deal with the citation history of publications and try to identify landmark publications. In contrast to other papers published hitherto, this study is based on a broad data set with more than five million papers published between 1980 and 1990 for identifying the landmark papers. We analyzed the citation histories of the papers across 25 years. We used two techniques for the impact comparisons: first, field-normalized impact scores allowed the comparison of papers across different fields. Second, the CSS method was used to classify the papers as "poorly cited," "fairly cited," "remarkably cited," and "outstandingly cited."

The results of the study reveal that 1,013 papers (less than 0.02 percent of the papers in our set) are "outstandingly cited" in the long run (with $k \geq 4$ maintained across 25 years).

These papers can be named as landmark papers within the pool of more than five million papers published between 1980 and 1990. Further analyses of 40 exceptionally cited papers (within the group of "outstandingly cited papers; the top group with $k \geqslant 6$) show that 80 percent of the papers have at least one US address; many papers in this group can be assigned to chemistry, internal medicine, biochemistry, and molecular biology. The cluster analyses of the 1,013 "outstandingly cited" papers revealed that the papers received the high impact level very soon after publication (and remained on this level over decades). Only a slight impact decline is visible over the years.

In this study, we proposed an approach for identifying landmark papers in the long run. We identified a small proportion of papers in the data set which can be called "landmark papers." Since these papers have been constantly cited on a very high level over decades, they seem to be very important for scientific progress in the corresponding fields (and could be analyzed in future studies).

We deemed it important in the current emphasis on research evaluation in many countries worldwide to develop approaches for citation analyses which focus on classic papers within fields. Research evaluation tends to focus on the last few years, which might fit with institutional evaluation cycles, but not with scientific progress in fields. The review of the literature and responses to a call for evidence by the Department for Business; Energy; Industrial Strategy (2016) show that "there remains a concern that the REF [Research Excellence Framework] does influence the way researchers design and conduct their work, in ways that sometimes compromises and obstructs long-term, high-risk research endeavours. Such distortions could be of real significance" (p. 14).

We would like to encourage other researchers to follow our approach and to use (and also improve) the methods for identifying landmark papers in the long run for research evaluation purposes. Landmark papers are the shoulders supporting scientific progress (Merton, 1965).

Note

1. The OECD field definitions can be found at www.oecd.org/science/inno/38235147.pdf. We used the two digit level scheme.

References

Baumgartner, S.E. and Leydesdorff, L. (2014), "Group-based trajectory modeling (GBTM) of citations in scholarly literature: dynamic qualities of 'transient' and 'sticky knowledge claims'", *Journal of the Association for Information Science and Technology*, Vol. 65 No. 4, pp. 797-811, doi: 10.1002/asi.23009.

Bland, J.M. and Altman, D.G. (1986), "Statistical methods for assessing agreement between two methods of clinical measurement", *Lancet*, Vol. 1 No. 8476, pp. 307-310.

Bornmann, L. (2011), "Scientific peer review", *Annual Review of Information Science and Technology*, Vol. 45, pp. 199-245.

Bornmann, L. and Glänzel, W. (2017), "Applying the CSS method to bibliometric indicators used in (university) rankings", *Scientometrics*, Vol. 110 No. 2, pp. 1077-1079, doi: 10.1007/s11192-016-2198-5.

Colavizza, G. and Franceschet, M. (2016), "Clustering citation histories in the physical review", *Journal of Informetrics*, Vol. 10 No. 4, pp. 1037-1051, doi: 10.1016/j.joi.2016.07.009.

Department for Business, Energy & Industrial Strategy (2016), *Building on Success and Learning from Experience An Independent Review of the Research Excellence Framework*, Department for Business, Energy & Industrial Strategy, London.

Glänzel, W. (2007), "Characteristic scores and scales – a bibliometric analysis of subject characteristics based on long-term citation observation", *Journal of Informetrics*, Vol. 1 No. 1, pp. 92-102, doi: 10.1016/j.joi.2006.10.001.

Glänzel, W. (2010), "The role of the h-index and the characteristic scores and scales in testing the tail properties of scientometric distributions", *Scientometrics*, Vol. 83 No. 3, pp. 697-709, doi: 10.1007/s11192-009-0124-9.

Glänzel, W. (2011), "The application of characteristic scores and scales to the evaluation and ranking of scientific journals", *Journal of Information Science*, Vol. 37 No. 1, pp. 40-48, doi: 10.1177/0165551510392316.

Glänzel, W. and Schubert, A. (1988), "Characteristic scores and scales in assessing citation impact", *Journal of Information Science*, Vol. 14 No. 2, pp. 123-127.

Kaufman, L. and Rousseeuw, P.J. (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York, NY.

Merton, R.K. (1965), *On the Shoulders of Giants*, Free Press, New York, NY.

Moed, H.F. and Halevi, G. (2015), "Multidimensional assessment of scholarly research impact", *Journal of the Association for Information Science and Technology*, Vol. 66 No. 10, pp. 1988-2002, doi: 10.1002/asi.23314.

Pham, D.T., Dimov, S.S. and Nguyen, C.D. (2005), "Selection of K in K-means clustering", *Proceedings of the Institution of Mechanical Engineers Part C-Journal of Mechanical Engineering Science*, Vol. 219 No. 1, pp. 103-119, doi: 10.1243/095440605x8298.

Ponomarev, I.V., Williams, D.E., Hackett, C.J., Schnell, J.D. and Haak, L.L. (2014), "Predicting highly cited papers: a method for early detection of candidate breakthroughs", *Technological Forecasting and Social Change*, Vol. 81, pp. 49-55, doi: 10.1016/j.techfore.2012.09.017.

Redner, S. (2005), "Citation statistics from 110 years of physical review", *Physics Today*, Vol. 58 No. 6, pp. 49-54, doi: 10.1063/1.1996475.

Rodríguez-Navarro, A. (2016), "Research assessment based on infrequent achievements: a comparison of the United States and Europe in terms of highly cited papers and Nobel prizes", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 3, pp. 731-740, doi: 10.1002/asi.23412.

van Raan, A.F.J. (1996), "Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises", *Scientometrics*, Vol. 36 No. 3, pp. 397-420.

van Raan, A.F.J. (2008), "Bibliometric statistical properties of the 100 largest European research universities: prevalent scaling rules in the science system", *Journal of the American Society for Information Science and Technology*, Vol. 59 No. 3, pp. 461-475, doi: 10.1002/asi.20761.

Vinkler, P. (2010), *The Evaluation of Research by Scientometric Indicators*, Chandos Publishing, Oxford.

Waltman, L. (2016), "A review of the literature on citation impact indicators", *Journal of Informetrics*, Vol. 10 No. 2, pp. 365-391.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J. and Johnson, B. (2015), *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*, Higher Education Funding Council for England (HEFCE), Bristol.

Winnink, J.J., Tijssen, R.J.W. and van Raan, A.F.J. (2016), "Breakout discoveries in science: what do they have in common?", in Ràfols, I., Molas-Gallart, J., Castro-Martínez, E. and Woolley, R. (Eds), *Proceedings of the 21 ST International Conference on Science and Technology Indicator*, Universitat Politècnica de València, València, pp. 1069-1075.

Woelert, P. (2013), "The 'economy of memory': publications, citations, and the paradox of effective research governance", *Minerva*, Vol. 51 No. 3, pp. 341-362, doi: 10.1007/s11024-013-9232-z.

Corresponding author
Lutz Bornmann can be contacted at: bornmann@gv.mpg.de