CrossMark

A citation-based cross-disciplinary study on literature aging: part I—the synchronous approach

Lin Zhang[1,2] (ORCID) · **Wolfgang Glänzel**[2,3]

Abstract We provide a view of the literature aging features in the sciences and social sciences, at different aggregation levels, of major fields, subfields, journals and individual papers and from different perspectives. Against to the wide belief that scientific literature may become more rapidly obsolete, we found that, in general, the share of more recent references were distinctly lower in 2014 than that in 1992, which holds for all aggregation levels. As exceptions, the subfields related to *Chemistry* and the subfield *energy and fuels*, have shown a clear trend to cite more recent literature than older articles. *Particle and nuclear physics* and *astronomy and astrophysics*, the two subfields which strongly rely on e-print archives, have shown a 'polarization' tendency of reference distribution. Furthermore, we stress that it is very important to measure the Price Index at the paper level to account for differences between the documents published in the same journals and (sub-)fields.

Keywords Literature aging · Price Index · References · Citations

Introduction

References in scholarly literature play a very important role in acknowledging the use of scientific information. During the last twenty years, the volume of information available for the academic public has drastically increased. In addition, co-authorship and citation networks have considerably grown. From the bibliometric viewpoint, this resulted in

✉ Lin Zhang
zhanglin_1117@126.com

1 Department of Management and Economics, North China University of Water Resources and Electric Power, Zhengzhou, China

2 Centre for R&D Monitoring (ECOOM) and Department of MSI, KU Leuven, Louvain, Belgium

3 Department of Science Policy & Scientometrics, Library of the Hungarian Academy of Sciences, Budapest, Hungary

"inflationary values" (Persson et al. 2004). The fast development of electronic database and Web sources allow researchers not only to track easily the latest advances in their research field and to easily access relevant recent information, but also to trace back the older sources. The age of cited literature can thus indicate the up-to-date characteristics of the information resources used in scientific communication.

The synchronous (alias retrospective approach) is an intuitive and the most convenient way to approach ageing-related aspects of information. We say ageing related because this covers a broad range of aspects: since ageing and obsolescence of information has subject-specific characteristics, certain subject characteristics can be measured by the specific disciplinary ageing behavior. The synchronous approach thereby proceeds from the present (i.e. recently published literature) and looks back to the preceding information, usually on the basis of cited references. This can be done even if cited documents are analysed as, for instance, in Thomson Reuters (now Clarivate Analytics) journal half-life: Both cited and citing half-life are synchronous measures since the perspective is always the most recent year. Differences and consequences of synchronous and diachronous approaches has been discussed, for instance by Glänzel (2004). However, most bibliometric measures are designed from the synchronous viewpoint, we will then follow this approach and base the first part of our citation-based cross-disciplinary study on the analysis of the age structure of cited literature.

It was widely believed that given the accelerated pace of scientific development, scientific literature would become more rapidly obsolete (Price 1965; Line 1970). Nevertheless, Egghe et al. (1995) studied the influence of production on utilization functions and showed that, in the synchronous case, the larger the increase in production, the larger the obsolescence. In the diachronous case the opposite relation holds: the larger the increase in production the smaller the obsolescence rate. Glänzel and Schoepflin (1999) found that communication processes in theoretical topics and fundamental research are comparatively slow. Some topics in parasitology, zoology, botany, entomology and other fields with historical components are characterized by a higher age of cited literature. By contrast, obsolescence of literature in the life sciences and in less fundamental research topics in physics and chemistry is relatively fast with respect of both references and citations.

It was also reported that the age distribution of cited references is highly skewed, that is the vast majority of references are made to recent material (e.g., van Raan 2000). However, Lariviere et al. (2008) studied the evolution of the aging phenomenon of cited literature over a period of more than 100 years of scientific activity, and found that researchers were relying on an increasingly old body of literature since the mid-1960s. By contrast, Evans (2008) showed that as more journal issues came on line, cited articles tended to be more recent, and he speculated that the shift from browsing print collections to searching online collections facilitated avoidance of older literature. Alarran and Ruiz-Castillo (2011) found that the distributions of references made and citations received share some basic features in subject fields: Their study was based on 3.7 million articles published between 1998 and 2002 in the 22 science fields according to the Essential Science Indicators classification scheme. A recent study by Google Scholar's team (Verstak et al. 2014) found an apparently increasing fraction of citations to older publications during 1990–2013, and suggested that citing older documents is possible thanks to accessibility improvements to scientific knowledge, including digitization of old documents, proliferation of repositories and search engines. Martin-Martin et al. (2016) confirmed the findings by Verstak et al. (2014), and further concluded that the growing trend of citing old documents is partly due to Google Scholar search services.

Although the above-mentioned studies have provided interesting and insightful results, some conclusions seem rather conflicting. In this study, we will continue and expand the explorations on the obsolescence of literature along the following research questions.

1. As a consequence of the fast development of electronic data sources and open access et al., might researchers incline to cite more recent literatures?
2. How far do subject matters influence the age of references and how is the evolution reflected by different subject fields and subfields?
3. Do characteristics of ageing differ between individual paper level and different aggregation levels (journals, subfields and major fields)? Does measuring Price Index at the paper level account for differences between the documents published in the same journals and (sub-)fields?

Finally we would like to stress that the main objective of the present study is not to simply compare patterns across fields or journals, to identify anomalies or just to find some interesting cases, but to point to the necessity of disaggregation to depict the phenomenon of "hardness" and "obsolescence" in an appropriate manner. Analysis at higher aggregation levels, such as subject fields, alone might not account for evolutional aspects of substructures, as has shown, for instance, with regard to the changing "hardness" of the field of scientometrics (Schoepflin and Glänzel 2001). This is also one of the reasons why Henk Moed has introduced the Price Index for individual publications (Moed 1989). Quite similarly to the hardness and ageing, we have observed similar phenomena with respect to *interdisciplinarity*, where a breakdown to individual articles enhances the discriminatory power, improves the distinction between inter- and multidisciplinarity and therefore becomes an inevitable methodological approach (Zhang et al. 2016). In this context we still mention that to cover the journal level, we only selected several periodicals, that we considered typical of their filed, as pars pro toto; the processing, comparative analysis and discussion of all journals covered by the database is certainly possible, but beyond the scope of the study.

Data and methods

In the present study, only journal publications which contain at least one reference are considered. More than 749,000 source documents indexed in the 1992 volume, and 1,733,000 source documents indexed in the 2014 volume of the Web of Science Core Collection (WoS) have been analyzed. The total number of references processed amounts to about 79,000,000.

The age of each individual reference was determined as the difference between the publication year of the reference and that of the citing document accordingly. References to unpublished papers ("in press", "forthcoming", etc.) have been considered of age zero. All source documents are further assigned to different journals, subject subfields and fields. The subfields and fields classification are obtained from the Leuven-Budapest (ECOOM) subject-classification scheme (Glänzel and Schubert 2003). The 68 subfields and 16 major-fields in the ECOOM classification provide two hierarchical levels for the reference analysis, and facilitate the further comparison between subfields and major-fields.

The analysis is based on the following indicators, which are applied at different aggregation levels: major-fields, subfields, journals, and individual papers.

1. *The mean reference age* the mean age of cited references counted by years.

2. *The median reference age* the median value of the age of cited references.
3. *The Price Index* the share of the references not older than five years as introduced by Price (1970).
4. *The mean reference rate* the mean number of cited references. Here we exclude papers that have not any references.

Results

General reference analysis for source documents in 1992 and 2014

The reference-based indicators of source documents indexed in 1992 and 2014 WoS volumes are listed in Table 1. The number of publications has more than doubled from 1992 to 2014, while the number of references grew by factor four (cf., Persson et al. 2004). *The mean reference rate* has increased from 20.9 in 1992, to 36.6 in 2014. The growth rate of mean reference numbers thus amounts to 75%. As already mentioned above, here we excluded publications without references. The *median reference age* in 1992 and 2014 amounts to 6 and 8, respectively, while that of *the mean reference age* is 10.4 and 11.6. Note that *the mean reference age* is more influenced by extremes. i.e., very old papers, than *the median age*. In general, source documents in 2014 have cited more references than their counterparts in 1992, and both the mean and median age of references have increased. To obtain statistically reliable results, we only include papers citing at least 10 references for the calculation of *the mean reference age*, *the median reference age* and the *Price Index* for each individual paper. In 1992, 66% publications have at least 10 references each, and in 2014, 87% publications can meet the reference selection threshold.

Figure 1 displays the cumulative percentage of all references cited by source documents in 1992 and 2014. The line of 2014 is distinctly below the line of 1992 when references are younger than 10 years old, and the two lines approach coincidence for older references. In general, the fraction of younger references is clearly lower in 2014 than that in 1992.

Unlike Fig. 1, which displays the cumulative distribution of cited references in the overall database, Fig. 2 shows the distribution of the *median references age* of *individual publications* in 1992 and 2014. Two salient features can be readily observed in Fig. 2: (1) a large share of publications have a median reference age between 4 and 8 years. (2) Compared with publications in 1992, the reference age curve of 2014 volume has a clear shift to the right. This means papers in 2014 have generally cited older references than papers in 1992. The peak of reference-age distribution in 1992 volume is around 5–6 years, while in 2014 it shifted towards an age of 6–7 years. In terms of very old and very recent references, we also find clear different trends in the two WoS volumes under study. Publications in 2014 tend to cite more 'very old' and fewer 'very recent' references.

Table 1 Reference-based indicators of source documents indexed in 1992 and 2014. Data sourced from Thomson Reuters Web of Science Core Collection

Year	Papers	References	MRR	Papers (10+)	MdnRA	MRA
1992	749,554	15,645,556	20.9	491,120	6	10.4
2014	1,733,087	63,387,562	36.6	1,503,952	8	11.6

MRR mean reference rate, *MdnRA* median reference age, *MRA* mean reference age, Papers (10+)—papers with at least 10 references
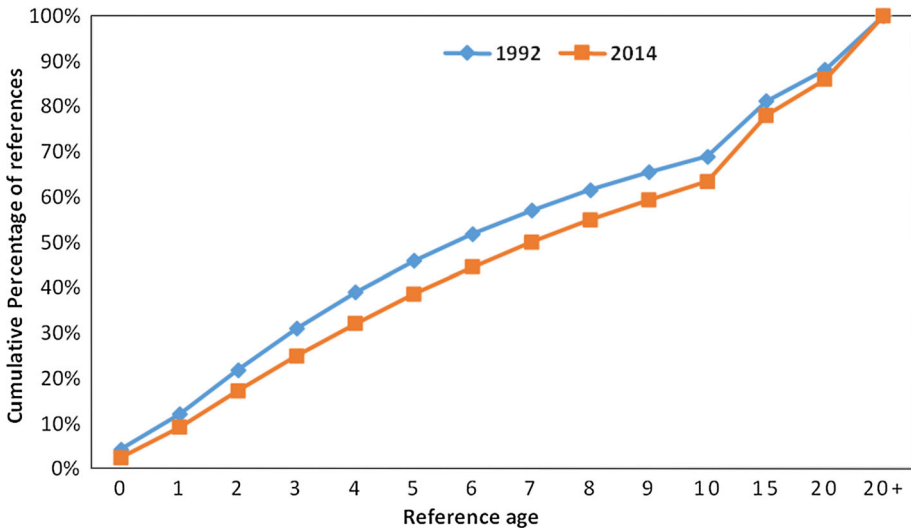
Fig. 1 Cumulative percentage of cited reference in 1992 and 2014. Data sourced from Thomson Reuters Web of Science Core Collection
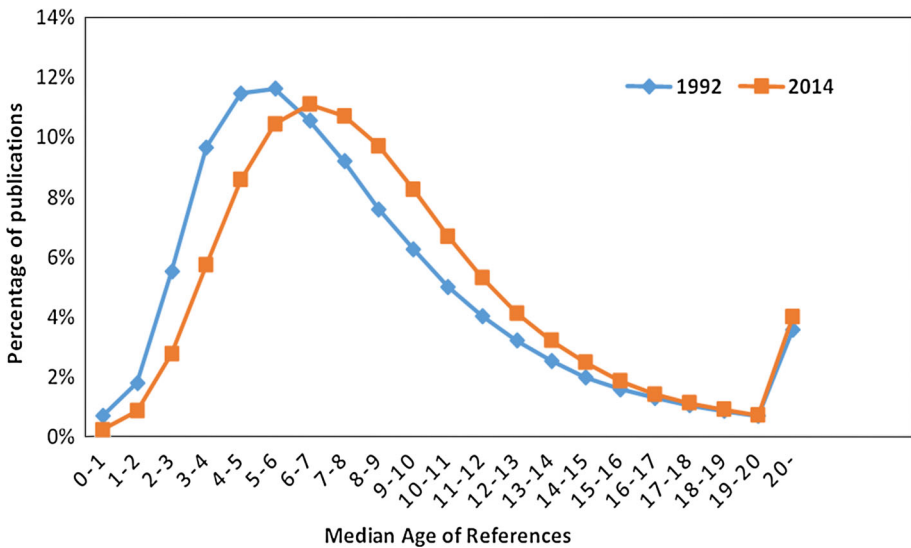

Fig. 2 Distribution of the median reference age of publications (1992 vs. 2014). Data sourced from Thomson Reuters Web of Science Core Collection

As already mentioned in the *Methods* part, the *Price Index* is calculated as the share of the references not older than five years in all references. Figure 2 shows the distribution of the Price Index of individual publications in 1992 and 2014. The distribution of Price Index over papers has its peak between 0.2 and 0.4 for both years, which means, papers with 20–40% of their references not older than five years were most frequent. Compared to 1992, a higher share of documents published in 2014 have a low Price Index (0–0.2), and a

lower share of publications was found with relatively high Price Index (above 0.4), that is, the distribution for 2014 is more polarized. Furthermore, the fraction of publications with extremely high Price Index also decreased in 2014. For example, we found 3355 publications in 1992 (0.68%), and 3170 publications in 2014 (0.21%) having Price Index of 1.0 (Fig. 3).

Reference analysis by major fields

Unlike in the study by Lariviere et al. (2008), the reference analysis by major fields in our study is not only conducted at this aggregation level (considering all the papers for each level at once), but also derived from reference-based indicators at the individual-paper level.

Tables 2 and 3 display the indicators for 15 major fields in the sciences and social sciences according to the ECOOM classification based on the aggregation level and the level of individual papers. To obtain statistically meaningful results, we only include publications with at least 10 references for analysis on paper level. Publications in the arts and humanities are excluded, since the role of non-serial literature is considerable in the references and the publication years of references are frequently missing.

Table 2 suggests that the Price Index decreases in all fields in 2014, with the only exception of *Chemistry*. *Chemistry* is also the only field with a decreasing mean age of references. In Table 3, '(Price > 0.5)%' means the percentage of publications with a Price Index larger than 0.5. This indicator can measure the relative frequency of 'harder papers' (according to Price) in each field. All fields have a decreased share of publications with high Price Index (above 0.5) in 2014. Compared to other fields, the percentage of publications with Price Index greater than 0.5 in *Chemistry* is relatively stable (24% in 1992 vs. 23% in 2014). The rank of each field by Price Index is indicated by a number in brackets for each year. Some changes in ranking between 1992 and 2014 are eye-catching. For
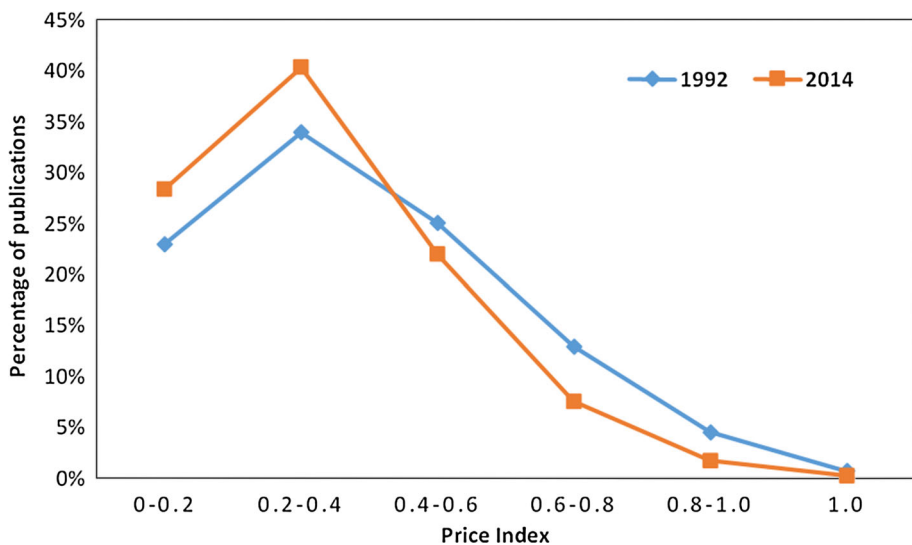

Fig. 3 Distribution of Price Index of individual publications (1992 vs. 2014). Data sourced from Thomson Reuters Web of Science Core Collection

Table 2 Reference-based indicators on aggregation level for 15 major fields in 1992 and 2014 (papers with at least 1 reference are included). Data sourced from Thomson Reuters Web of Science Core Collection

Field	1992			2014		
	Papers	Price Index (%)	MRA	Papers	Price Index (%)	MRA
Multidisciplinary sciences	15,738	(1) 50.9	8.10	57,707	(4) 34.3	10.07
Clinical and Experimental Medicine I	81,795	(2) 45.4	7.59	194,732	(1) 37.9	8.68
Biosciences	65,630	(3) 44.2	8.05	133,809	(6) 33.6	9.98
Physics	89,964	(4) 42.2	9.49	193,129	(3) 34.7	11.55
Engineering	46,330	(5) 39.0	9.42	186,105	(8) 33.6	10.56
Biomedical Research	50,075	(6) 38.7	8.97	112,097	(5) 33.7	9.83
Social Sciences II	46,860	(7) 37.4	16.08	77,039	(14) 24.6	17.68
Clinical and Experimental Medicine II	90,823	(8) 37.0	9.64	307,763	(7) 33.6	9.89
Social Sciences I	39,478	(9) 35.5	12.81	87,072	(13) 26.0	13.49
Chemistry	107,826	(10) 35.2	10.78	314,296	(2) 35.7	10.47
Neuroscience and Behavior	37,948	(11) 35.1	9.96	86,942	(11) 27.9	11.49
Geosciences and Space Sciences	28,882	(12) 34.6	10.79	86,931	(12) 27.5	13.76
Biology	53,828	(13) 33.6	11.12	145,157	(9) 29.2	12.85
Agriculture and Environment	27,891	(14) 30.9	11.09	120,737	(10) 28.9	11.34
Mathematics	22,775	(15) 28.9	12.69	66,245	(15) 24.4	15.19

MRA mean reference age

Table 3 Reference-based indicators on paper level for 15 major fields in 1992 and 2014 (only papers with at least 10 references are included). Data sourced from Thomson Reuters Web of Science Core Collection

Field	1992		2014	
	Papers	(Price > 0.5) (%)	Papers	(Price > 0.5) (%)
Multidisciplinary sciences	8806	(1) 51.2	53,837	(7) 19.7
Clinical and Experimental Medicine I	59,847	(2) 40.5	161,084	(1) 27.7
Biosciences	57,375	(3) 38.9	124,347	(8) 18.2
Physics	64,689	(4) 38.0	182,193	(3) 22.2
Engineering	28,122	(5) 30.3	173,237	(4) 21.0
Biomedical Research	40,983	(6) 28.7	102,848	(6) 19.7
Clinical and Experimental Medicine II	63,411	(7) 27.1	252,077	(5) 20.9
Social Sciences II	16,219	(8) 26.9	52,837	(13) 11.0
Social Sciences I	15,617	(9) 24.7	63,028	(11) 11.3
Chemistry	80,481	(10) 24.0	300,825	(2) 23.0
Neuroscience and Behavior	29,566	(11) 23.3	80,627	(14) 10.4
Geosciences and Space Sciences	22,302	(12) 21.8	82,405	(12) 11.1
Biology	46,380	(13) 21.0	135,574	(9) 13.0
Agriculture and Environment	21,939	(14) 16.5	115,632	(10) 12.7
Mathematics	13,798	(15) 14.4	60,258	(15) 8.5

example, *Chemistry* raised its position from the rank 10 in 1992, to rank 2 in 2014 in both Tables 2 and 3. It is also interesting to observe that although the average reference ages in *Social Sciences (I and II)* are rather high, the Price Index and the share of publications with high Price Index are considerable. *Clinical and Experimental Medicine I (General and Internal Medicine)* shows a significant and stable feature of 'hard science' for both years. By contrast, *Mathematics* having fewest 'hard papers' ranks at the end of the list. This is basically in line with the slow ageing of mathematics (cf. Glänzel and Schoepflin 1999).

The full name of some major fields in Tables 2 and 3 are: *Clinical and Experimental Medicine I (General and Internal Medicine)*, *Clinical and Experimental Medicine II (Non-Internal Medicine Specialties)*, *Biosciences (General, Cellular and Subcellular Biology; Genetics)*, *Biology (Organismic and Supraorganismic Level)*, *Social Sciences I (General, Gerional and Community Issues)*, *Social Sciences II (Economical and Political Issues)*.

Reference analysis by subject subfields

The reference analysis by the 15 major ECOOM fields provides a general view of reference distributions, and now a question arises that whether the similar features can be observed within the 65 subfields in the ECOOM classification system. Here again the subfields in the arts and humanities have been excluded. In what follows, we would also like to answer the question of whether the reference characteristics could be different obtained from individual publications at various levels of aggregation.

"Appendix" shows two tables displaying the reference-based indicators for 65 subfields ranked by *Price Index* in 1992 and 2014. The *Price Index* here is based on the aggregation level of subfields. The reference rate of each subfield has increased in 2014, and the growth ranges between 27.3 references (for *Economics, business and management*) and 4.6 references (for *Cardiovascular and respiratory medicine*). *Pure and applied ecology*, *neurosciences and psychopharmacology* are among the subfields with highest reference rates, and *pure mathematics*, as expected, has cited the fewest references for both years. *Astronomy and astrophysics*, with 53.0 cited items on average in 2014, has raised its position from rank 8 (in 1992) to rank 2 (in 2014) according to the reference rates. However, different from the observations by Lariviere et al. (2008), where the authors found that starting in 1991– the year when arXiv was launched—the average and median age of Astronomy and astrophysics started to decrease, we did not observe the same tendency in our data. We may not directly compare our results with those by Lariviere et al. (2008), since their analysis was based on a chronological dataset ending in 2004, while we only use the data from two publication year 1992 and 2014. However, what particular impact e-print servers might have on subfield specific citation behavior, still requires more evidence.

Compared to 1992, most of the subfields have increased their median-age of cited references in 2014, only with a few exceptions (marked in Table 6 in "Appendix" with a down arrow indicating decline, or a horizontal arrow indicating no change). It is quite interesting that the four subfields with descending median reference age are all assigned to the major field of '*Chemistry*', respectively, *applied chemistry and chemical engineering*, *polymer science*, *materials science*, *multidisciplinary chemistry*. Another three subfields in *Chemistry* (*analytical, inorganic and nuclear chemistry*, *organic and medicinal chemistry*, *physical chemistry*) have stagnating median reference age. The similar observation also holds for the evolution of the mean reference age.

Along with the general rise in the median and mean age of references, a clear decline of the Price Index is observed for most of the subfields. The relative importance of recent cited literature has remarkably decreased, although the decline varies between subfields.

Multidisciplinary sciences and *cell biology* are found with largest reduction in the share of recent references. Again, exceptions regarding subfields apply to *Chemistry*: *physical chemistry*, *applied chemistry and chemical engineering*, *polymer science*, *multidisciplinary chemistry*, but also to the subfield *energy and fuels*.

The share of 'immediate references' with age zero (e.g., the references are published in the same year as the source document), also decreased for most subfields (48 out of 65 subfields). In turn, an increasing share of citations to older literature (10 or 15 years old papers) can be found in most subfields (57 out of 65 subfields). It is interesting to observe that, all subfields related to *Chemistry*, again together with the subfield *energy and fuels*, have cited old literature to a lesser extent. It is worth mentioning that, *particle and nuclear physics* and *astronomy and astrophysics*, the two subfields which strongly rely on e-print archives, have shown the 'polarization' tendency of reference distribution. The share of immediate references with age zero in *particle and nuclear physics* has increased from 6.5% (in 1992) to 9.1% (in 2014), and the same rate in *astronomy and astr*ophysics is raised from 3.9 to 4.9%. The rises led them to the top two subfields with largest share of 'immediate references' in 2014. In the meantime, growing shares of citations to old literature (over 10 years old) are also found in both two subfields, respectively, from 31 to 40% in *particle and nuclear physics*, and from 34 to 40% in *astronomy and astrophysics*. Although the e-print archive may provide faster access to the immediate research, we can't make a simple conclusion that the scholarly communication in the two subfields are eventually sped up. The growth of 'immediate reference' doesn't change the general tendency as found in most subfields: a decline of the Price Index and a rise in the median and mean age of references are still observed for these two subfields.

Now we would like to answer another question: Do characteristics of ageing differ between individual paper level and at higher aggregation levels? We select a number of subfields and subdivide them into two groups based on their Price Index at this aggregation level. *Energy and fuels*, *Particle and nuclear physics*, *Physical chemistry* and *Cell biology* are representative for the first group with relatively high Price Index, and a second set of four subfields, namely, *Animal sciences*, *Pure mathematics*, *History*, *Politics and Law* and *Education and Information* belong to the group with lower Price Index. The values of Price Index for each subfield are shown in Tables 5 and 6 in the "Appendix". In addition to the aggregated Price Index, we also calculated the Price Index for all individual papers within each subfield. Only papers with at least ten references are included for the calculation of individual paper level. The distribution of the Price Index for individual publications in the selected subfields are shown in Figs. 4 and 5. The Y axis indicates the percentage of the number of publications in each sub-filed.

The four subfields shown in Fig. 4 are among the fields with greatest values of aggregated *Price Index* in 2014. However, the distribution and evolution of Price Index based on individual papers have shown remarkable differences. The four subfields in Fig. 5 show a more skewed distribution towards the left than those given in Fig. 4. The most extreme case was found in *Animal sciences*, where half of the publications have a low Price Index (0–0.2) in 2014. The striking rise of low *Price Index* (0–0.2) in the publications in *History, Politics and Law*, and the remarkable drop of high *Price Index* (0.6–1) in *Education and Information* are worth mentioning.

Reference analysis by journals

To obtain statistically reliable results and to facilitate the direct comparison between the indicators on journal and paper level, we only include journals having at least 50
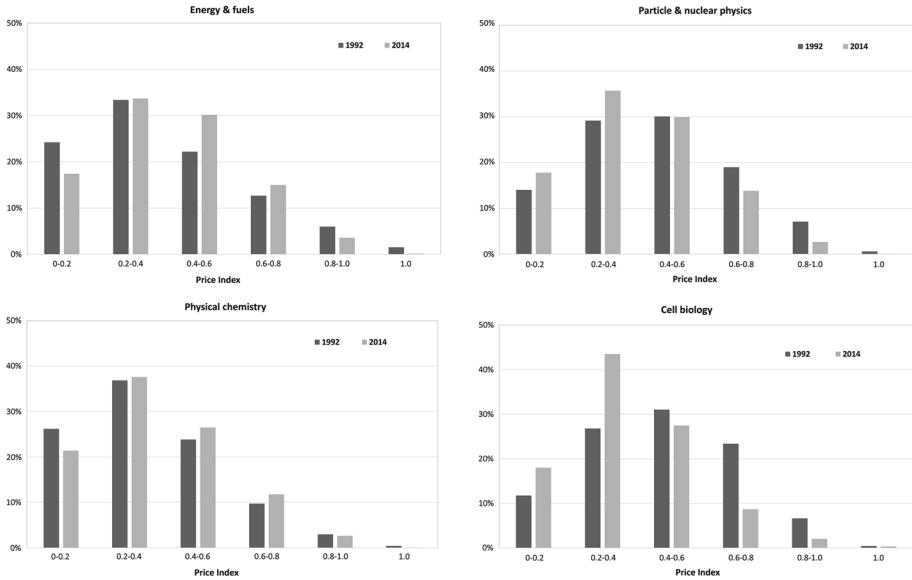
Fig. 4 Distribution of Price Index for individual publications in four subfields of the sciences (1992 vs. 2014). Data sourced from Thomson Reuters Web of Science Core Collection
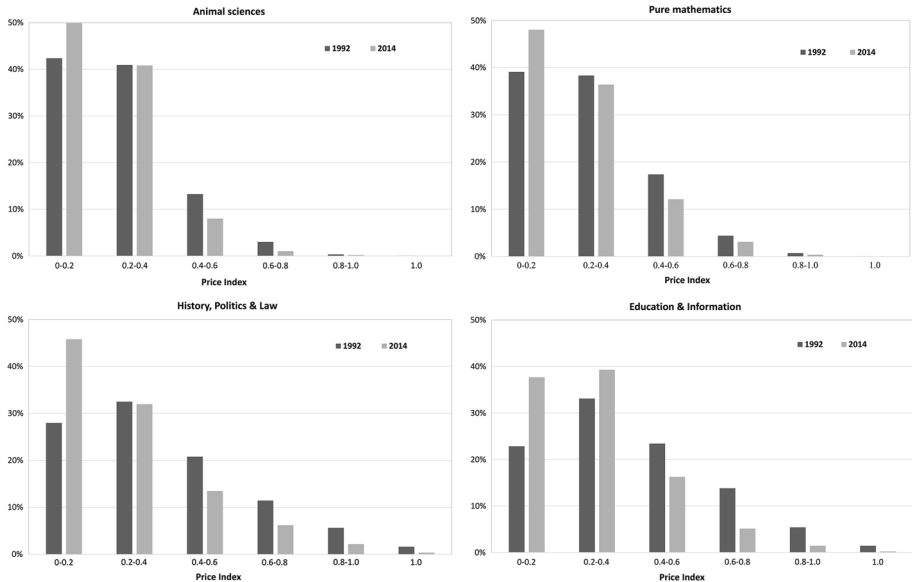

Fig. 5 Distribution of Price Index for individual publications in four subfields of the sciences and social sciences (1992 vs. 2014). Data sourced from Thomson Reuters Web of Science Core Collection

publications in each of the two years, and each publication should have at least 10 references. This results in 2317 journals in 1992, and 6707 journals in 2014. The overlap comprises 1858 journals. Journals with unique assignment to the arts and humanities are excluded for the following analysis.

The aggregated Price Index of different journals ranged between 0.792 and 0.056 in 2014, and 0.896 to 0.100 in 1992. Table 4 presents the reference-based indicators for the five journals with the highest aggregated Price Index in 2014. '(Price > 0.5) %' means the percentage of publications with a Price Index larger than 0.5. The most interesting case is observed in *Water Environment Research*: compared to other journals, the share of papers with high Price Index is rather low, and the mean age of references is quite large. To explore the reasons behind, we looked into the reference-based indicator of the individual papers. Figure 6 displays the distribution of Price Index for individual publications in *Water Environment Research*, and that of another journal (*Washington Quarterly*, listed in Table 4) for comparison. The Y axis indicates the percentage of the number of publications in each journal. Although both have relatively high Price Index at the journal aggregation level, the distribution of Price Index of individual publications in the two journals are differently shaped. While the distribution of papers in *Washington Quarterly* skewed to the right (higher Price Index), the distribution of *Water Environment Research*, is polarized (U-shaped). Furthermore, most publications with high Price Index have long reference lists and most documents with a low Price Index tend to have rather short reference lists. This finally results in the high value of aggregated Price Index at the journal level, due to the large number of recent references. These observations apply to both years 1992 and 2014. We have also found counterexamples to such phenomena, for instance, the aggregated Price Index for the journal of *Nuclear data sheets* is low in 2014 (0.133), but we found 35.6% publications in this journal with Price Index greater than 0.4. The share of publications with high Price Index in *Nuclear data sheets* is even higher than that in *Water Environment Research*, although the aggregated Price Index of the latter journal is five times larger.

In terms of the evolution, most journals have a decreasing integrated Price Index by citing more old references, while only a quarter of journals have an increased value of integrated Price Index in the year of 2014. *Bioresource Technology*, *American Review of Respiratory Disease*, *New England Journal of Medicine*, *Mikrochimica Acta*, *Journal of Power Sources* are the five journals with largest growth in Price Index. By contrast, e.g., *Strategic Management Journal*, has a striking decrease of Price Index (from 0.80 in 1992 to 0.29 in 2014). The same trend is also found for the individual papers of this journal, where the share of publications with higher Price Index (over 0.5) dramatically shrunk from 93.3% (in 1992) to 17.6% (in 2014).

Discussion and conclusions

Our results provide a view of the literature aging features in the sciences and social sciences, at different aggregation levels, of major fields, subfields, journals and individual papers and from different perspectives. Some fundamental characteristics are shared by the

Table 4 Reference-based indicators for 5 journals with the highest *Price Index* (in 2014). Data sourced from Thomson Reuters Web of Science Core Collection

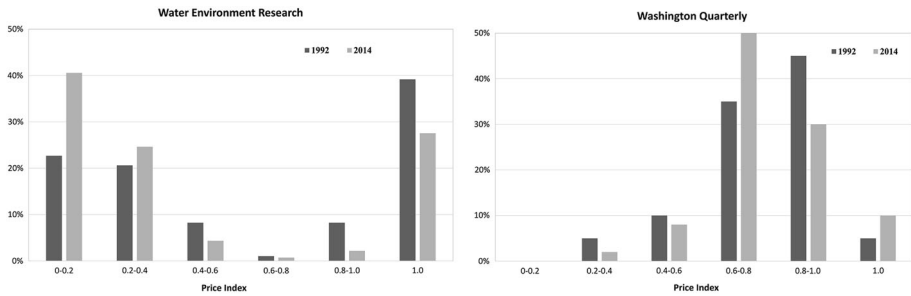| Journal	Papers	References	MRA	Price Index	(Price > 0.5) (%)
CURR OPIN CLIN NUTR	78	3193	3.3	0.79	96.2
WASH QUART	52	1823	4.4	0.78	94.2
IEEE NETWORK	69	1023	4.0	0.73	89.9
SURVIVAL	84	2707	5.5	0.70	84.5
WATER ENVIRON RES	138	9415	9.2	0.68	31.2

Fig. 6 Distribution of Price Index for individual publications in Water Environment Research, and Washington Quarterly (1992 vs. 2014). Data sourced from Thomson Reuters Web of Science Core Collection

reference distributions in most fields and subfields, although exceptions could always be found. In general, the share of more recent references was distinctly lower in 2014 than that in 1992, which holds for all aggregation levels. Publications in 2014 tend to cite more 'very old' sources, and less 'very recent' literature. At the journal level, we found decreasing Price Index by citing more old references, while only a quarter of journals have increased values of Price Index in 2014.

As exceptions, the subfields related to *Chemistry* and the subfield *energy and fuels*, have shown a clear trend to cite more recent literature than older articles. *Particle and nuclear physics* and *astronomy and astrophysics*, which strongly rely on e-print archives, have shown the 'polarization' tendency of reference distribution: towards citing more 'immediate literature (aged 0)' and more 'very old' publications (over 10 years old). Our results are partly consistent with Verstak et al. (2014), where the authors found that the share of reference lists in articles indexed by Google Scholar Metrics going to articles 10 or more years old has increased since 1990 in all Scholar Metrics subject areas except *Engineering* and *Chemistry*.

Although based on different data sets, periods and measures, our findings are by and large in line with the findings by Lariviere et al. (2008) and Martin-Martin et al. (2016). We could reinforce their previous findings by further explorations at different aggregation levels. However, our results are in direct contrast with the findings with Evans (2008), where the author concluded that as more journal issues became online published, the cited articles tended to be more recent.

Price (1970) introduced the '*Price Index*' as a measure of the 'hardness' of science. From a study of 162 journals, he concluded that Price Index was able to distinguish between hard science, soft science and non-science. *Physics* and *biochemistry* journals topped the scale with values between 60 and 70%. If 'Price Index' still holds as a measure of the 'hardness' of science, but according to our results, most of the subjects in sciences and social sciences seems to become 'softer'. But drawing final conclusions from these observations might be premature: Electronic publishing, archiving and access might already have thrown traditional communication off balance. Journals make papers available *Online First* and immediately citable before those are assigned to journal issues. On the other hand, repositories, self-archiving, digitalization and web services make also old literature, that might otherwise gather dust in libraries, available to the broad public and thus to the interested researchers. Also new retrieval opportunities and techniques might play an important part as these new trends may have made it easier for researchers to find the most relevant literature regardless

of whether it is the more recent or not (Verstak et al. 2014). There are some further explanations for the general tendency towards citing older papers. Simkin and Roychowdhury (2005) assumed that scientists retrieve source information from reference lists used in another papers instead of reading the original ones. The authors used the term 'non-reader' to describe a citer who copies. And finally citation as a part of the reward system in science (according to Merton 1968) might also contribute to re-citation of classics and "honorific" citations of the well-received work of high-profile authors.

The future will show if these seemingly contradicting effects will result in "polarization" or a general shift towards older or newer references, respectively. Only systematic comparative analysis can help answer the question of whether this affects the traditional standards of the corresponding indicators or of there is an ongoing hardening or softening process indeed. Since changing communication patterns might also affect publication types, the other strong indicator of "hardness", the share of references to periodicals (cf. Glänzel and Schoepflin 1999) could also change in a similar way.

We can conclude that it is very important to measure the Price Index at the paper level to account for differences between the documents published in the same journals and (sub-)fields (cf. Moed 1989). As observed in Fig. 6, references and citation distribution at the individual paper level can be strikingly different even among journals having quite similar aggregated Price Index.

Appendix

See Tables 5 and 6.

Table 5 Reference-based indicators for 65 subfields ranked by Price Index (1992). Data sourced from Thomson Reuters Web of Science Core Collection

Subfield	Papers	MRR	Price Index	MdnRA	MRA
Multidisciplinary sciences	15,738	17.69	0.509	4	8.10
Immunology	15,225	29.54	0.504	4	6.64
Cell biology	11,460	35.09	0.500	5	6.79
Applied physics	26,333	15.48	0.472	5	8.43
Hematology and oncology	18,277	29.22	0.462	5	7.13
General and internal medicine	27,371	17.17	0.458	5	8.19
Biochemistry/biophysics/molecular biology	44,326	31.71	0.451	5	7.72
Particle and nuclear physics	7197	23.94	0.450	5	8.28
Genetics and developmental biology	12,877	30.22	0.449	5	8.09
Multidisciplinary physics	16,870	20.61	0.438	6	9.51
Agricultural science and technology	451	15.76	0.438	5	8.35
Experimental/laboratory medicine	7562	25.22	0.434	5	8.17
Electrical and electronic engineering	16,161	14.07	0.431	5	8.42

Table 5 continued

Subfield	Papers	MRR	Price Index	MdnRA	MRA
Endocrinology and metabolism	8386	32.28	0.419	5	7.90
Microbiology	19,541	27.51	0.418	6	8.60
Physics of solids, fluids and plasmas	23,508	18.75	0.416	6	9.60
Computer science/information technology	15,188	16.51	0.409	6	8.67
Cardiovascular and respiratory medicine	15,398	24.87	0.400	6	8.38
Radiology and nuclear medicine	8044	18.47	0.400	6	8.83
Education and information	11,665	14.54	0.399	6	10.45
Astronomy and astrophysics	7321	30.27	0.395	6	9.17
Neurosciences and psychopharmacology	18,685	35.42	0.393	6	8.74
Pharmacology and toxicology	24,857	26.11	0.392	6	8.61
Energy and fuels	8942	16.75	0.391	6	9.58
Paramedicine	23,877	21.93	0.391	6	9.05
Dermatology/urogenital system	8855	19.82	0.390	6	9.26
History, politics and law	34,563	12.96	0.384	7	18.74
Mathematical and theoretical physics	2152	19.94	0.382	6	10.36
Anatomy and pathology	7456	27.32	0.379	6	9.81
Materials science	22,772	16.82	0.378	7	10.41
Atomic, molecular and chemical physics	7874	27.07	0.375	7	10.29
Multidisciplinary biology	5922	27.32	0.372	7	11.06
Analytical, inorganic and nuclear chemistry	22,672	22.99	0.371	7	10.10
Age and gender related medicine	14,109	19.77	0.365	6	9.26
Environmental science and technology	10,567	22.16	0.363	7	9.74
Organic and medicinal chemistry	14,588	24.27	0.361	7	10.54
Rheumatology/orthopedics	3792	21.82	0.359	6	10.10
Economics, business and management	13,565	17.80	0.355	7	11.03
Physiology	10,030	31.63	0.355	7	9.62
Biomaterials and bioengineering	2460	21.01	0.352	7	9.76
Meteorology/atmospheric and aerospace science and technology	4656	22.65	0.351	7	9.97
Physical chemistry	21,272	24.63	0.351	7	10.86
Surgery	13,609	19.14	0.349	7	10.47
Classical physics	14,444	17.06	0.345	7	11.16
Psychiatry and neurology	17,069	24.30	0.345	7	10.54
General, regional and community Issues	28,327	15.52	0.338	7	13.80
Multidisciplinary chemistry	20,991	23.50	0.335	7	11.14
Geosciences and technology	14,269	29.40	0.326	8	11.62
General and traditional engineering	12,701	15.76	0.325	7	11.05
Plant sciences	11,218	26.42	0.325	7	12.32
Ophthalmology/otolaryngology	6706	18.59	0.324	7	11.21
Polymer science	7753	21.06	0.317	8	11.55
Mineralogy and petrology	2775	24.77	0.314	8	11.51
Applied chemistry and chemical engineering	9570	16.90	0.307	8	11.94
Dentistry	3461	20.89	0.305	8	11.14

Table 5 continued

Subfield	Papers	MRR	Price Index	MdnRA	MRA
Veterinary sciences	6543	19.95	0.300	8	10.66
Applied mathematics	14,069	15.90	0.298	8	11.86
Hydrology/oceanography	2529	28.98	0.297	8	11.14
Aquatic sciences	5480	31.60	0.294	8	12.37
Psychology and behavioral sciences	20,824	24.78	0.292	8	11.68
Food and animal science and technology	9038	22.88	0.285	8	11.48
Pure and applied ecology	4285	35.79	0.276	8	11.84
Pure mathematics	10,530	12.79	0.269	10	14.16
Plant and soil science and technology	10,156	22.23	0.267	9	12.30
Animal sciences	9334	27.22	0.244	9	14.44

Table 6 Reference-based indicators for 65 subfields ranked by Price Index (2014). Data sourced from Thomson Reuters Web of Science Core Collection

Subfield	Papers	MRR	Price Index	MdnRA	MRA
Energy and fuels	38,068	33.61	0.416	6	9.44
Particle and nuclear physics	13,940	42.55	0.400	7	11.68
Multidisciplinary chemistry	64,071	45.25	0.398	6 ↓	9.55
Hematology and oncology	55,049	35.97	0.398	6	8.06
Applied physics	74,041	31.21	0.383	6	9.74
Cardiovascular and respiratory medicine	43,983	29.51	0.380	6 →	8.83
Immunology	27,185	41.67	0.379	6	8.62
General and internal medicine	57,265	27.79	0.374	6	8.96
Physical chemistry	78,367	40.97	0.372	7 →	10.42
Materials science	105,855	34.31	0.369	6↓	10.22
Astronomy and astrophysics	18,523	52.96	0.363	7	11.39
Cell biology	31,004	49.22	0.359	7	8.81
Experimental/laboratory medicine	27,750	38.94	0.357	7	9.06
Dermatology/urogenital system	24,466	25.15	0.352	7	9.67
Paramedicine	110,707	31.71	0.351	7	9.35
Physics of solids, fluids and plasmas	43,544	36.56	0.350	7	11.31
Electrical and electronic engineering	57,683	27.44	0.350	7	9.77
Microbiology	64,208	42.73	0.349	7	10.00
Endocrinology and metabolism	18,888	45.19	0.344	7	9.38
Analytical, inorganic and nuclear chemistry	44,739	38.50	0.343	7 →	10.69
Biomaterials and bioengineering	16,597	39.40	0.343	7 →	9.51
Organic and medicinal chemistry	35,365	44.76	0.343	7 →	10.69
Radiology and nuclear medicine	22,473	31.10	0.343	7	9.67
Multidisciplinary sciences	57,708	42.13	0.343	7	10.07

Table 6 continued

Subfield	Papers	MRR	Price Index	MdnRA	MRA
Pharmacology and toxicology	49,125	45.70	0.342	7	9.61
Polymer science	18,719	40.82	0.339	7↓	10.32
Multidisciplinary physics	23,136	33.14	0.338	8	12.82
Biochemistry/biophysics/molecular biology	78,820	46.97	0.338	7	9.72
Age and gender related medicine	40,669	29.30	0.333	7	9.76
Genetics and developmental biology	34,347	48.24	0.328	7	10.65
Applied chemistry and chemical engineering	41,055	36.29	0.328	7↓	10.92
Surgery	47,128	24.49	0.326	7 →	10.39
Anatomy and pathology	12,373	35.96	0.324	7	11.06
Classical physics	50,202	29.66	0.320	8	12.28
Computer science/information technology	60,802	34.01	0.320	7	10.34
Psychiatry and neurology	52,992	35.19	0.317	7 →	10.42
Rheumatology/orthopedics	20,525	27.28	0.317	7	10.63
Atomic, molecular and chemical physics	17,583	44.57	0.314	8	12.72
Environmental science and technology	60,451	43.86	0.313	7 →	10.64
Neurosciences and psychopharmacology	46,124	51.90	0.307	8	10.49
Agricultural science and technology	12,791	35.05	0.305	8	11.25
Meteorology/atmospheric and aerospace science and technology	15,774	41.30	0.299	8	11.50
Ophthalmology/otolaryngology	16,970	28.20	0.297	8	11.60
Multidisciplinary biology	11,100	45.03	0.296	8	12.16
General and traditional engineering	56,676	30.62	0.288	8	12.00
Mathematical and theoretical physics	9768	34.02	0.278	9	14.52
Physiology	10,576	51.30	0.275	9	11.78
Dentistry	10,110	31.81	0.272	8 →	11.43
Education and information	30,560	31.26	0.270	8	12.05
Pure and applied ecology	18,982	53.53	0.269	9	12.64
Plant sciences	23,843	47.66	0.268	9	14.18
History, politics and law	45,618	31.64	0.265	10	21.96
Food and animal science and technology	34,395	38.31	0.265	9	11.78
Veterinary sciences	14,691	31.56	0.261	9	12.23
General, regional and community Issues	57,878	35.27	0.255	9	14.25
Psychology and behavioral sciences	47,985	49.30	0.252	9	12.46
Applied mathematics	51,669	28.60	0.252	10	14.36
Plant and soil science and technology	19,409	41.75	0.238	9 →	13.13
Aquatic sciences	12,637	51.84	0.238	10	14.28
Geosciences and technology	45,302	51.09	0.237	10	15.32
Hydrology/oceanography	7669	46.38	0.235	10	13.70
Mineralogy and petrology	6999	37.06	0.234	10	14.80
Pure mathematics	26,339	21.53	0.233	12	18.18
Economics, business and management	33,851	45.06	0.229	10	13.48
Animal sciences	19,752	44.77	0.212	11	19.91

→: Unchanged; ↓: descend

References

Alarran, P., & Ruiz-Castillo, J. (2011). References mode and citations received by scientific articles. *Journal of the American Society for Information Science and Technology, 62*(1), 40–49.

Egghe, L., Rao, R., & Rousseau, R. (1995). On the influence of production on utilization functions: Obsolescence or increased use? *Scientometrics, 34*(2), 285–315.

Evans, J. A. (2008). Electronic publication and the narrowing of science and scholarship. *Science, 321*(5887), 395–399.

Glänzel, W. (2004). Towards a model for diachronous and synchronous citation analyses. *Scientometrics, 60*(3), 511–522.

Glänzel, W., & Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing and Management, 35*(1), 31–44.

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics, 56*(3), 357–367.

Lariviere, V., Archambault, E., & Gingras, Y. (2008). Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). *Journal of the American Society for Information Science and Technology, 59*(2), 288–296.

Line, M. B. (1970). The 'half-life' of periodical literature: Apparent and real obsolescence. *Journal of Documentation, 26*(1), 46–54.

Martin-Martin, A., Orduna-Malea, E., Ayllon, J. M., et al. (2016). Back to the past: On the shoulders of an academic search engine giant. *Scientometrics, 107*(3), 1477–1487.

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56–63.

Moed, H. F. (1989). Bibliometric measurement of research performance and Price's theory of differences among the sciences. *Scientometrics, 15*(5–6), 473–483.

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.

Price, D. J. D. (1965). Networks of scientific papers. *Science, 149*(3683), 510–515.

Price, D. J. D. (1970). Citation measures of hard science, soft science, technology, and nonscience. In C. E. Nelson & D. K. Pollock (Eds.), *Communication among scientists and engineers* (pp. 3–22). Lexington, MA: Heath.

Schoepflin, U., & Glänzel, W. (2001). Two decades of "Scientometrics". *Scientometrics, 50*(2), 301–312.

Simkin, M. V., & Roychowdhury, V. P. (2005). Stochastic modeling of citation slips. *Scientometrics, 62*(3), 367–384.

van Raan, A. F. J. (2000). On growth, ageing, and fractal differentiation of science. *Scientometrics, 47*(2), 347–362.

Verstak, A., Acharya, A., Suzuki, H., et al. (2014). *On the shoulders of giants: The growing impact of older article*s. http://arxiv.org/abs/1411.0275. Accessed July 22, 2015.

Zhang, L., Rousseau, R., & Glänzel, W. (2016). Diversity of references as an indicator for interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the American Society for Information Science and Technology, 67*(5), 1257–1265.