

# Exploration into the evolution and historical roots of citation analysis by referenced publication year spectroscopy

Jianhua Hou<sup>1</sup>

Received: 19 August 2016 / Published online: 3 January 2017  
© Akadémiai Kiadó, Budapest, Hungary 2016

**Abstract** In this study, we investigated the evolution process and historical roots of citation analysis study by reference publication year spectroscopy (RPYS), which is an advanced research method recently introduced in the field of Scientometrics. Through analyzing the publication year and citation frequency of cited references in a knowledge domain, RPYS can identify the citation peak of such cited references. We collected 2543 articles including 56,392 references regarding citation analysis in SCI-E and SSCI databases between 1970 and July 2016 as our data source. Based on the RPYS method by the CRExplorer program, the results showed that the peak citation publication years by chronological order occurred in 1955, 1963, 1973, 1979, 1981, 1990, 2005 and 2008 in the field of citation analysis study. According to the overall distribution of peak citation publication years, the RPYS for citation analysis was divided into five time periods for the convenience of comparison in this paper: that is, before 1900, 1901–1950, 1951–1970, 1971–2000, and 2001–2016. Pre 1950, especially during the 1900s–1950s, before the citation analysis method was introduced, there were three rather high peak citation publication years, Lotka’s law and Bradford’s law laid the knowledge foundation for citation analysis. 1950s–1970s was the forming period of citation analysis, among the three citation peaks in this period, Garfield (*Science* 122(3159):1127–1128, 1955), Price (*Von Der Studierstube Zur* 7(3–6):443–458, 1963, *Science* 149:510–515, 1965), and Kessler (1963) established an important knowledge base for the formation of citation analysis study. 1971–2000 was a developing period of citation analysis. Document Co-citation Analysis and Author Co-citation Analysis methods laid the foundation of the development of citation analysis study. Since 2000, citation analysis study has been expanding rapidly. By the number of published papers and number of cited references with highly cited frequency, Garfield E., White H.D., Small H., MacRoberts M.H., Price D.D. have played an important role in promoting the evolution of citation analysis study.

---

✉ Jianhua Hou  
houjianhua@dlu.edu.cn

<sup>1</sup> Research Center of Science Technology and Society, Dalian University, Dalian 116622, China

**Keywords** Referenced publication year spectroscopy (RPYS) · Citation analysis · Citation peak · Historical root · CRExplorer

## Introduction

Citation analysis is a quantitative research method that analyzes the citation of scientific journals, papers and authors to examine their quantitative characteristics and inherent laws. Since its establishment, citation analysis has been the subject of unceasing argument concerning the rationality, legitimacy and knowledge content it reflects. Yet citation analysis has been widely used by researchers as the main research method for bibliometrics and scientometrics. During this process, it has built its own theoretical system, although there are some problems in application (Macroberts and Macroberts 1989, 2009). In Garfield (1955), proposed a method to search for scientific literature through citations in his *Citation Indexes for Science*. In 1963, he further expanded upon the comprehensive citation index database, the Science Citation Index (SCI), which was regarded as the establishment of the citation analysis method and laid an important foundation for citation analysis study. However, as a field involving cross-over studies, citation analysis has a far-reaching theoretical history; to be specific, researches in and before the 1950s had already made significant strides in the establishment, development and subsequent evolution of citation analysis.

Scientometrics researchers usually believe that citations reflect the research's influence instead of its quality (Leydesdorff et al. 2014). Generally speaking, literature citations can be classified into short-term citation and long-term citation. Short-term citation can reveal the research fronts within a scholarly field, while long-term citation can convert knowledge into conceptual symbols (e.g. Kuhn, paradigm theory). Short-term citation indexes (such as impact factor) are more often used to predict the long-term changes in academic fields. In general, we analyze the knowledge base, research evolution and research fronts of academic fields through citations, but we don't make in-depth analysis on publication year, type, discipline, citation motivation and journals (publishing house) of cited references. These still influence the references' citation and contribution to the academic field in different ways. Therefore, by analyzing the citation publication year, we can further explore the evolution process and transformation of citation analysis study in a given field.

This research utilized the latest analysis method of scientometrics—referenced publication year spectroscopy (RPYS) to deepen our understanding of citation analysis study. As an effective method for tracking the circulation of scientific knowledge, citation analysis has become increasingly popular among researchers in different disciplines. By analyzing the publication year of references with high citation frequency, this research revealed the milestones in citation analysis and also identified the earlier literature of this field as a knowledge base, which provided a new perspective to analyze the historical evolution of citation analysis study.

## Related work

New research in academic fields is typically based on previous research in the scientific community, therefore, citations and the number of cited references can usually illustrate how knowledge accumulates, integrates, transfers and generates new themes or new

discoveries (Comins and Hussey 2015a, b). Although there are many different discussions on the theory of scientific development, the knowledge base of existing research cannot push the advancement of other academic and scientific fields. However, it plays an important role in promoting the progress of its own field (Popper 1961; Kuhn 1962). Price (1965) discussed the relationship between knowledge base and research front, revealing that scientists tended to cite recently published articles by the concept of immediate index. He pointed out that research front refers to scientists actively citing recently published articles and knowledge base refers to the evolving network and co-citation track formed by scientific literature and cited by frontier literature. The knowledge base of a research field has, and will continue to, evolve along with its research front.

At present, in the research of scientometrics and bibliometrics, measuring indicators are usually employed to analyze the co-citation frequency or citation frequency of the literature, or analyze the sudden increase or sharp reduction of co-citation frequency at a specific time to study the mapping relation between knowledge base and research front, and further to discover the interactive evolution law between them and predict emerging trends (Garfield 2006; Chen 2006; Chen et al. 2010). These measuring indicators usually analyze indicators of “short-term citation”, while rarely analyzing indicators of “long-term citation”, moreover, they lack analysis of research evolution and historical root of academic fields from the view of publication year distribution of “long-term citation”. Marx et al. (2014) proposed a new analytical method, using RPYS to reveal the times cited of important historical literature and the publication years in the field, in this way, further probe the historical root of research in academic fields and analyze the effect of citation on current research. This analytical method mainly analyzes the total citation frequency of references published in different years within a specific research field.

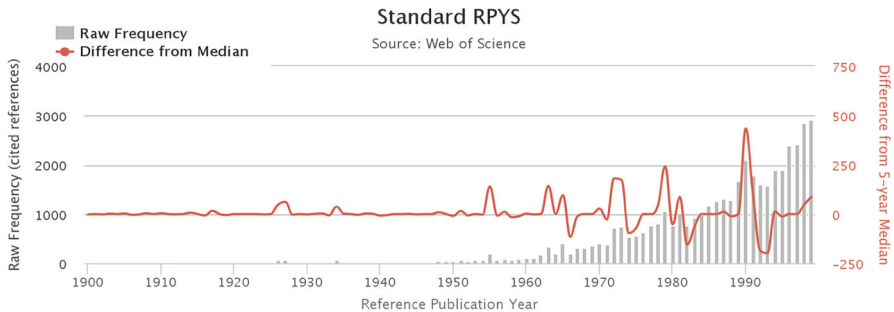
Through RPYS analysis, we focused on analyzing indicators of “long-term citation”. Using the research field “citation analysis” as an example, the paper not only concerned publication year of cited references with high citation frequency in citation analysis study, but also paid attention to early cited references with high citation frequency, mainly analyzing how this early literature laid a knowledge base for citation analysis and promoted its development.

## Methods and data

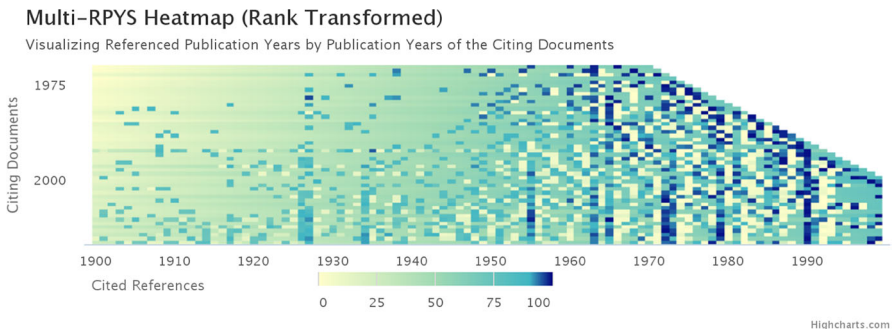
RPYS is to analyze the cited frequency and publication year of a piece of literature, so as to reveal the historical root and important articles. The tool used to map RPYS was recently proposed by Bornmann and Marx (2013), Marx et al. (2014); combined with the newly-developed software for quantitative analysis of referenced publication year and cited frequency, it can assist in identifying the historical root of an academic field or research subject. In short, this software can sort the cited frequency of references in a specific research domain in accordance with the distribution of their publication years, thus demonstrating the dynamic distribution of the cited frequency of important references in line with their publication years. This method is now widely applied to analyze the historical root of the evolution of academic fields or research subjects, for example, physics by Barth et al. (2014), information science by Leydesdorff et al. (2014), biology by Marx and Bornmann (2014), and global positioning system (GPS) by Comins and Hussey (2015a, b).

The first demonstration of RPYS by Marx et al. (2014) was mainly based on SciSearch of American Chemical Society. In order to develop software for analyzing data downloaded from WOS, Lutz Bornmann cooperated with Loet Leydesdorff, who expanded his existing software package to form the Bibliometric Coupling2. In order to improve the analysis performance of RPYS, Andreas Thor further developed the program RPYS.exe into CRExplorer, and the new analysis tool Cited References Explorer (CRExplorer) was integrated by Thor et al. (2016). In addition to working out RPYS, CRExplorer can also analyze the referenced data in the spectrum and handle the disambiguation of references. Comins and Hussey (2015a, b) further developed software Multi-RPYS (available at <http://comins.leydesdorff.net>), and released RPYS i/o in 2016 for online analysis, which covered both Standard RPYS and Multi-RPYS. For Standard RPYS, the distribution of publication years of highly cited references and curve of difference from 5-year median can better reflect the citation peaks (Fig. 1); for Multi-RPYS, it can work out multi-level RPYS, demonstrate the “heat degree” of multi-subject literature in peak publication year, and reveal the relationship between cited references and research fronts (Fig. 2; Comins and Leydesdorff 2016).

The data in this research is from Web of Science of Thomson Reuters. Searching for the topic “citation analysis” from 1970 to July 2016 (46 years) in SCI-E and SSCI, 2543 articles were downloaded, covering 56,392 references (55,583 references after removing duplicate by CRExplorer).



**Fig. 1** Standard RPYS of citation analysis domain



**Fig. 2** Multi-RPYS of citation analysis domain

Next, we used CRExplorer to analyze the cited references. The research attempted to reveal the evolution process of citation analysis study through topic, publication year and citation frequency of highly cited literature. The data processing and analysis results are as follows:

1. Searching for and downloading all articles about citation analysis in Web of Science by “topic = citation analysis”;
2. Extracting all references with CRExplorer and identifying the publication year of the most important literature in the citation analysis field;
3. Removing duplicates of the references (CRExplorer—“Standardization”—“Cluster equivalent Cited References”/“Merge Cited References of the Same Cluster”). Cleaning and standardizing the variety of one reference, also merging cited frequencies of the same references.
4. Using CRExplorer to identify the most important historical articles in the citation analysis field;
5. Establishing distribution of cited frequency through RPYS and determining distribution of literature with the highest citation frequency.

## Result

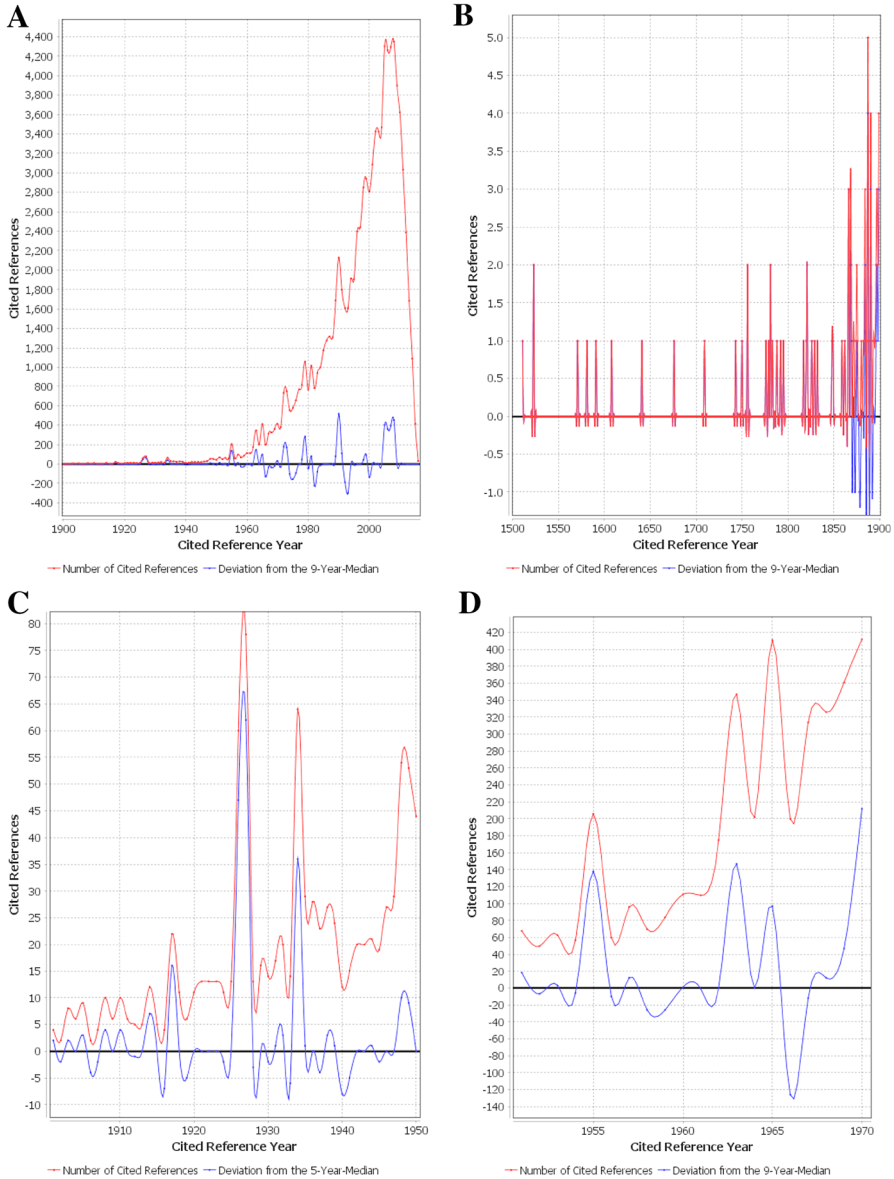
We used CRExplorer for RPYS to conduct an overall RPYS analysis on the downloaded literature of citation analysis studies in the previous section with analysis time from 1511 to 2016. In the analysis results, the distribution of consecutive citation publication years from 1900 to the present were selected (Fig. 3a). As suggested by the curve of median deviation value, throughout the history of citation analysis, there are a total of 8 relatively significant peaks among the referenced publication years. In particular, the most obvious is 1990, followed by 2008, 2005, and 1979. In the order of citation publication year, all peaks of citation publication year occurred in 1955, 1963, 1973, 1979, 1981, 1990, 2005, and 2008, respectively. Information about the citations for each year is shown in Table 1.

According to the statistical table, with the peak citation publication year changing, the number of cited references has increased year by year, and the total cited frequency has also increased. However, the cited frequency per paper ( $N_{c-y}$ ) shows an overall decline, indicating that the occurrence of cited frequency peaks is not only caused by the increased number of cited references. Meanwhile, statistics show that the cited frequency of the most frequently cited literature in the peak citation publication years is often dozens of times more than the cited frequency per paper in the same year, and, specifically, Hirsch (2005), P NATL ACAD SCI USA was nearly 160 times more than the cited frequency per paper in that year. The increase in publications leads to an increase in citation frequency over time. However, the increase in cited frequency is not only caused by the increase in citation counts, because there is usually a relatively small amount of milestone, high-impact citation, the cited frequency of which is so high that pulls up the total cited frequency in that year.

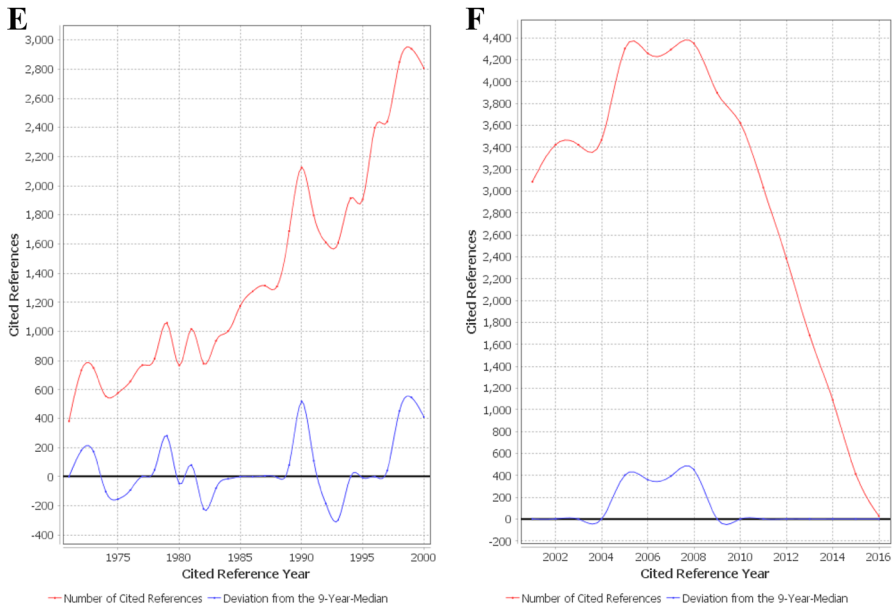
According to the overall distribution of peak citation publication years, for the convenience of comparison in this paper, the RPYS for citation analysis was divided into five time periods, that is, pre 1900, 1901–1950, 1951–1970, 1971–2000, and 2001–2016, and the RPYS was plotted for each time period (Fig. 3b–f).

**Pre 1950: the budding period of citation analysis**

It is generally accepted that the *Science Citation Index* published by Garfield (1955) and the Citation Index Database established in 1963 marked the initiation of the citation analysis method. However, according to the RPYS before 1950, a lot of literature had already laid the knowledge foundation of citation analysis. Before 1900, few references



**Fig. 3** RPYS of citation analysis. **a** 1900–2016. **b** 1500–1900. **c** 1901–1950. **d** 1951–1970. **e** 1971–2000. **f** 2001–2016



**Fig. 3** continued

**Table 1** Information about the citations for each peak citation publication year

Reference Published Year	Number of References	Total of Reference Frequencies	Average Cites per Paper ( $N_{c-v}$ )	Top Cited in Published Year (P)	Cited Number of P ( $N_c$ )	$N_c/N_{c-v}$
1955	50	205	4.1	Garfield (1955)	140	34.15
1963	116	347	2.99	Kessler (1963)	82	27.42
1973	309	749	2.42	Small (1973)	190	78.51
1979	495	1059	2.14	Garfield (1979)	175	81.78
1981	502	1016	2.02	White and Griffith (1981)	137	67.82
1990	1086	2126	1.96	Mccain (1990)	123	62.76
2005	2338	4300	1.84	Hirsch (2005)	293	159.24
2008	2482	4349	1.75	Bormmann and Daniel (2008)	78	44.57

were cited. The earliest cited reference was published in 1511 (CONSTANTINOPLE, 1511, HAGADOT HA TALMUD). Three citation peaks occurred since 1850, i.e., 3 references were cited 5 times in 1887, 3 references were cited 4 times in 1890, and 3 references were cited 3 times in 1868 (Fig. 3b).

The period of 1900–1950 was crucial in the budding of citation analysis, including a total of 665 citations (658 after duplicate removal), and 3 significant peak citation publication years. As suggested by Fig. 3c, these peaks appeared in 1927 (1926), 1934, and

**Table 2** Information about the highly cited literature in peak citation publication years (1900–1950)

No.	Cited References	Published Year	Cited Count	Proportion in Published year (%)	Authors
1	Gross (1927)	1927	56	71.79	Gross P. L.
2	Lotka (1926)	1926	53	88.33	Lotka A. J.
3	Bradford (1934)	1934	37	57.81	Bradford S. C.
4	Cole and Eales (1917)	1917	12	54.54	Cole F. J.
5	Bradford (1948)	1948	14	25.92	Bradford S. C.

1917. In addition, the peak of citation publication also appeared in 1948. Table 2 lists the information about the highly cited literature in the peak citation publication years from 1900 to 1950.

The first peak citation year in this time period is 1926–1927, with two citation peaks in total, in which *The frequency distribution of scientific productivity* published by Lotka A. J. in America's famous academic journals J WASHINGTON ACAD SC was cited 53 times, accounting for the total citation rate of 88.33%. The paper aimed at finding out technical workers' production capacity and their contributions to scientific and technological progress and social development, which is considered to be the first time to reveal the law of scientific productivity using the relationship between the author and the number of published papers (Lotka 1926). In addition, *College Libraries and Chemical Education* published by Gross P. L. in Science in 1927 was cited 56 times in total, accounting for the total citation rate of 71.79%. Gross (1927) took citation data in Journal of the American Chemical Society as an evaluation indicator, obtaining valuable chemical journal lists. The second peak citation publication year was 1934 with two papers cited 37 times in total. *Sources of Information on Specific Subjects*, published by Bradford (1934) in ENGINEERING-LONDON, was cited 6 times. The paper presented literature's law of scattering, namely Bradford's Law, which has important guiding significance in determining the core journal, literature retrieval and making literature acquisition strategies. Cole and Eales (1917), SCI PROGR, V11, P578 constituted the third citation peak of this period and was cited 12 times, accounting for 54.54% of the total frequency of the cited references in that year. Moreover, 1948 is another peak citation publication year of this period, the characteristic of citation of this year is that cited references increased to a total 36 articles with a total citation frequency of 54, but citation frequency of each individual article was 1.5, indicating that research in this year showed a tendency for diversity, including *Documentation* published by British philologists Bradford S. in London Crosby Lockwood Press, *The self-fulfilling prophecy* published by Merton R. K. in Antioch Review and *A mathematical theory of communication* by Shannon C. E.

The basic research on citation analysis before 1950 mainly concentrated in Gross's taking citation count as a journal evaluation index, Lotka's distribution of scientific productivity, and Bradford's law of scattering of literature. Lotka's Law, Bradford's Law, and Zipf's Law are also known as the three laws of bibliometrics. However, the law of word frequency distribution proposed by the American scholar G.K. Zipf in the 1940s did not generate a citation peak among the research of citation analysis, and his *Human Behavior and the Principle of Least Effort* published in 1949 was cited 9 times. Therefore, in the budding period of citation analysis pre 1950, Lotka's Law and Bradford's Law made important impacts, but Zipf's law of word frequency distribution did not play a key role in the formation of citation analysis.



**Table 3** Information on the highly cited literature in peak reference citation publication years (1951–1970)

No.	Cited References	Published Year	Cited Count	Proportion in Published Year (%)	Proportion of Total Cited Frequency (%)
1	Kessler (1963)	1963	82	23.63	55.61
2	Price (1963)	1963	81	23.34	
3	Garfield and Sher (1963)	1963	30	8.64	
5	Garfield (1955)	1955	140	67.96	67.96
6	Price (1965)	1965	129	31.39	51.58
7	Kaplan (1965)	1965	32	7.79	
8	Hagstrom and Selvin (1965)	1965	15	3.65	
9	Merton (1965)	1965	11	2.68	
10	Lipetz (1965)	1965	11	2.68	
11	Zadeh (1965)	1965	7	1.70	
12	Kessler (1963)	1965	7	1.70	

Citations with cited counts accounting for more than 50% of the total cited frequency in each year were selected

### 1951–1970: the forming period of citation analysis

1951–1970 was an important period for the formation of citation analysis, including a total of 2075 citations (2036 after duplicate removal). As suggested by Fig. 3c, there were three obvious citation peaks; 1955, 1963, and 1965. Table 3 lists the information on the highly cited literature in the 3 citation peaks from 1951 to 1970.

The first citation publication peak in this period is 1963, when Kessler M.M published *Bibliographic coupling between scientific papers* in *American Documentation*, proposing the analysis method of bibliographic coupling, and Price D. J. D published his *LITTLE SCI BIG SCI*, proposing the exponential growth law of scientific development, which provided an important scientific theoretical basis for the establishment of citation analysis. The two papers were cited 163 times in total, accounting for 46.97% of the total cited frequency in that year. 1955 is the second citation peak, when 50 articles were cited 205 times. Among them, *Science Citation Index* published by Garfield in *Science* was cited 140 times, accounting for 67.96% of the total cited count. The paper first proposed a method of searching scientific literature through the Citation Index, providing a new approach for retrieval and marking the establishment of the citation analysis method. 1965 is the third peak citation publication year in this period, presenting 164 references, which were cited 411 times. Among them, the most frequently cited was Price's *Network of scientific papers* published in *Science*, which was cited 129 times, accounting for 31.39% of the total citations in that year.

Among the three citation peaks in this period, Garfield (1955), Price (1963, 1965), and Kessler (1963) established an important knowledge basis for the formation of citation analysis, providing both methodological and theoretical foundations. Furthermore, in 1961, Garfield founded the Science Citation Index (SCI) database in Philadelphia, which provided the basis for citation analysis, as a key point in the formation period of citation analysis.

## 1971–2000: the development period of citation analysis

The research of citation analysis entered an important development period in the 1970s. Document Citation Analysis and Author Citation Analysis methods laid the knowledge base for the development of citation analysis study during this period. This period presented a total of 24,284 citations (23,846 after duplicate removal). There were five significant citation peaks, 1990, 1979, 1973, 1998, and 1981. Table 4 lists the information on the highly cited literature in the five citation peaks from 1971 to 2000.

We detected the evolution of citation analysis study through RPYS, the first important step of this period occurred in 1973. *Co-citation in the scientific literature: A new measure of the relationship between two documents* by Small H. published in the *Journal of the American Society for Information Science* was cited 187 times accounting for 26.88% of total times cited. In this paper Small H. proposed the Document Citation Analysis Method. It was a milestone in the history of citation analysis study.

Citation peaks occurred frequently from 1976 to 1979, especially in 1979, *Theory and Application of Citation Index in Science and Technology and Human Science* published by Garfield E. was cited 176 times accounting for 16.84% of total times cited in that year. Garfield E. systematically consolidated the theories and methods of citation index and proved that science citation index was a new tool that could evaluate the performance of scientific research and promote the progress of science.

1981 could also be viewed as a year with a relatively high reference citation peak, which marked by *Author co-citation: A literature measure of intellectual structure* published by White H.D. and Griffith B.C. in the *Journal of the American Society for Information Science*. This article put forward the method of author co-citation analysis for the first time and was cited 137 times, accounting for 13.48% of total times cited that year. It was also a milestone research paper in citation analysis history.

**Table 4** Information on the highly cited literature in peak reference citation publication years (1971–2000)

No.	Cited References	Published Year	Cited Count	Proportion in Published Year (%)	Proportion of Total Cited Frequency (%)
1	Small (1973)	1973	187	26.88	26.88
2	Garfield (1979)	1979	175	16.84	24.60
3	Garfield (1979)	1979	91	8.76	
4	White and Griffith (1981)	1981	137	13.48	22.93
5	Smith (1981)	1981	96	9.45	
6	Mccain (1990)	1990	123	5.79	10.31
7	Borgman et al. (1990)	1990	61	2.87	
8	Griliches (1990)	1990	35	1.65	
9	White and McCain (1998)	1998	132	4.63	10.80
10	Brin and Page (1998)	1998	58	2.04	
11	Redner (1998)	1998	32	1.12	
12	Ingwersen (1998)	1998	32	1.12	
13	Cronin et al. (1998)	1998	28	0.98	
14	Kostoff (1998)	1998	26	0.91	

For the 1970s–1980s, the references with a total cited count of more than 20% of the total citation in that year were selected; for the 1990s, the references with a total cited count of more than 10% of the total citation in that year were selected

1990 is the reference citation publication peak year with the largest citation count in this period, and the most frequently cited article was Mccain K.W.’s *Mapping authors in intellectual space: A technical overview*, which was published in *Journal of the American Society for Information Science* and was cited 123 times, accounting for 5.79% of the total citation frequency in that year. The paper systematically introduced the analysis methods and processes of total cited count by authors, and reviewed the progress of related research since the analysis method of total cited count by authors was proposed in 1981. In this paper, she proposed a method of cluster analysis based on authors co-citation, factor analysis and knowledge mapping. The second citation peak occurred in 1998. There were 1609 articles cited 2849 times in total. Of which *Visualizing a discipline: An author co-citation analysis of information science 1972–1995* was published by White H.D. and Mccain K.W. in *Journal of the American Society for Information Science* and was cited 132 times, accounting for 4.63% of total times cited in that year. This paper mainly used Author Citation Analysis for information visualization of information science. Both White in 1981 and Mccain in 1990 made detailed and systematic introductions of the analysis method of total cited count by authors. Both papers employed the method of author citation analysis to visualize the field and plot the map of scientific knowledge based on ACA, pioneering the study of mapping of scientific knowledge.

1971–2000 became an important development period for citation analysis, mainly because two crucial methods of citation analysis, i.e., Document Citation Analysis and Author Citation Analysis, were formed after citation analysis was proposed in the 1960s. According to the peak citation publication years in this period, it is these methods that helped expand the milestone citation publication years. In 1973, Small put forward the analysis method of total cited count by references; in 1981, White proposed the analysis of total cited count by authors; in 1990, Mccain systematically introduced and reviewed the analysis method of total cited count by authors; in 1998, White and Mccain proposed an analysis method of visualized literature information based on ACA (the map of scientific knowledge).

### After 2001: the rapid expansion period of citation analysis

This period included a total of 42,756 citations (42,251 after duplicate removal). The cited frequency of references in the recent decade has not been stable. To date, two obvious citation peaks have occurred in this period 2008 and 2005, which is considerably noteworthy. Table 5 lists the information on the highly cited literature in the two citation peaks from 2001–2016.

**Table 5** Information on the highly referenced publication years (2001–2016, top 3 in cited frequency each peak year)

No.	Cited References	Published Year	Cited Count	Proportion in Published Year (%)	Author
1	Hirsch (2005)	2005	293	6.81	Hirsch J. E.
2	Moed (2005)	2005	109	2.53	Moed H. F.
3	Boyack et al. (2005)	2005	43	1.00	Boyack K. W.
4	Bornmann and Daniel (2008)	2008	78	1.79	Bornmann L.
5	Radicchi et al. (2008)	2008	54	1.24	Radicchi F.
6	Bar-Ilan (2008)	2008	42	0.97	Bar-Ilan J.

From 2001 to the present, two significant citation publication peaks were formed (citations in recent years were not stable), i.e., 2008 and 2005. In particular, there were 2482 references in 2008, with a total citation count of 4349 times. And the most frequently cited paper was Bornmann L's *What do citation counts measure? A review of studies on citing behavior*, which was published in *Journal of Documentation* and cited 78 times, only accounting for 1.79% of the total cited frequency in that year. The paper carried out an empirical study on citing behavior. The second citation peak appeared in 2005, with the two most frequently cited papers, Hirsch J.E.'s *An index to quantify an individual's scientific research output* published in *Proceedings of the National Academy of Science* and Moed H.F.'s *Citation Analysis in Research Evaluation*. Among them, H index proposed by Hirsch has become one of the most popular methods in the field of information science research in recent years.

Since the 2000s, citation analysis has entered a period of rapid expansion. The number of published papers, citation count, and citation frequencies all increased sharply. Based on the existing theories of citation analysis, document co-citation, and author co-citation, citation analysis began to evolve in diversified directions, including scientific evaluation index, citation network, information visualization, and citing behaviors.

### Highly cited authors of citation analysis study

Based on the above analysis, we further explored the highly cited literature in the citation analysis field, and presented the notable examples, researchers, and published journals throughout the evolution of the field. Among the data in this study, there were a total of 56,393 references published between 1970 and 2016, cited 93,452 times in total and cited 1.66 times per paper. Among them, there were 11 papers with a cited frequency of more than 100 times, 35 with a cited frequency of more than 50, and their combined cited frequency was 3519 times, accounting for 3.77% of the total citation frequency. Hirsch's paper (2005) was cited 293 times, ranking first. Garfield (1972) was cited 274 times, ranking second. In terms of the most cited authors, Garfield contributed the most frequently cited references (cited frequency of over 50 times; Table 6), including 6 papers cited more than 50 times. Publication years of these references included 1955, 1972, 1979, and 2006, basically throughout the entire lifespan of citation research. Both White H.D. and Small H. contributed 3 references that had been cited more than 50 times. Specifically, White H.D.'s contributions were mainly in the 1980s and 1990s, Small H.'s contributions were mainly in the 1970s, and Price D.J.D.'s contributions were mainly in the 1960s. Contributions of MacRoberts M.H. (Problems in Citation Analysis) and Seglen P.O. (Journal Impact Factors and Evaluation of Research) were mainly in the 1990s.

**Table 6** Contribution rankings of authors of references cited more than 50 times

No. of Papers	Authors	Total Citations	Published Year
5	Garfield E.	772	1955, 1972, 1979 (twice), 2006
3	Small H.	341	1973, 1974, 1978
3	White H. D.	320	1981, 1989, 1998
2	MacRoberts M. H.	240	1989, 1996
2	Price D. J. D.	210	1963, 1965
2	Seglen P. O.	180	1992, 1997

In addition to these highly cited authors, the highly cited literature was generally the most highly cited literature in each peak citation publication year from 1955 to 2008. In addition to the above-mentioned highly cited references, the papers with a cited frequency of more than 50 times also included those highly cited since 2000. This revealed several basic aspects of the research hot topics in the citation analysis field. The first aspect was the citation analysis applied in research evaluation (the highly cited literature on the journal impact factors and evaluation of research proposed by Seglen P.O. had emerged in the late 1990s), mainly including the research on H index proposed by Hirsch, research on G index proposed by Egghe L, and research on citation analysis applied in research evaluation proposed by Moed, forming the first citation peak in 2005. The second aspect was the research related to citation behavior. This research was proposed in Bornmann L's study on citation behavior in 2008, forming the second citation peak. The third aspect was the information visualization based on citation analysis. The techniques of citation analysis and information visualization were mainstreamed by Chaomei Chen. However, no citation publication peak has been formed in this research hotspot yet.

## Conclusion and discussion

In the analysis method of RPYS, by analyzing the cited frequency of references and the distribution of publication years, the historical origin and evolutionary path of the knowledge domain or research topics can be analyzed. This is especially true for the early influential literature in a specific field. Its basic function is to identify the peak years of the citation publication in the development of an academic field, to reveal the contributions made by the publications in peak years to the evolution of the field, and to explore the theoretical basis and historical origin of the field or research topics. In this study, the CRExplorer, software was employed to investigate the evolution and historical origin of the citation analysis field, and the following major conclusions were drawn.

1. By using long-term citation analysis, the overall evolution (1900–2016) of citation analysis study was explored. It was found that there were three important historical stages in the creation and development of citation analysis. The first stage: 1950s–1960s, the formation period, marked by Garfield's method of Science Citation Index in 1955, Price's law of exponential growth of scientific development in 1963, and Price's network analysis of scientific literature in 1965. The second stage: 1970s the late 1990s, the development period, marked by three milestones, i.e., Small's analysis method of reference co-citation in 1973, analysis method of author co-citation by White and Griffith (1981) and McCain (1990), and White and McCain's map of scientific knowledge based on ACA in 1998. The third stage: 2000 and later, the rapid expansion period of citation analysis, characterized by diversified research subjects, including three major hot topics: Scientific research evaluation index, citation behavior, and information visualization.
2. Through the RPYS analysis, the study of the historical origins of research on citation analysis was greatly advanced. It was also found that the earliest cited references in this field were published in 1511. Additionally, it was discovered that 1900–1950 was the budding period of citation analysis. There have been a large number of basic studies on citation analysis. Among them, Lotka's Law and Bradford's Law became the main theoretical basis for the creation of citation analysis.

3. In terms of the total times cited counts of references, before 1950s, the total times cited count of references in each year was nearly 25 (Lotka 1926, cited 26 times); during the 1950s, the total number of cited references at the citation peak in 1955 was nearly 50, cited 205 times in total. During the 1960s, the number of references at the citation peak in 1963 was nearly 100, cited 347 times in total. During the 1970s, there were 309 cited references, cited 749 times in total. From the 1980s, to 2008, there were 2482 references among the citation peaks, cited 4349 times in total. For the citation peaks within each period of citation analysis, the number and citation frequency of references gradually increased, generally presenting multiple growth trends every ten years. After the year 2000, citation analysis was characterized by a rapid and vigorous development period.
4. By exploring the historical origins of citation analysis, it was found that all citation peaks throughout the history of this field were indicative of milestones or ground breaks in research. Before 1950, the citation peaks mainly represented research of Lotka's Law and Bradford's Law, and the citation peak in 1955 was marked by the research of the method of Science Citation Index proposed by Garfield. The year 1963 was marked by the research of Price's law of accelerated growth of scientific development. 1973 was marked by that of by Small's analysis method of Document Co-Citation Analysis. In 1981, White first put forward the analysis of Author Co-Citation Analysis. In 1990, McCain systematically introduced and reviewed the analysis method of author co-citation. The peak in 1998 was represented by research into White and McCain's map of scientific knowledge based on author co-citation. 2005 was marked by the research of H index for scientific research evaluation. 2008 was marked by researching Bornmann's analysis of citing behavior.
5. By analyzing the highly cited authors of citation analysis and their published volumes, Hirsch (2005) and Garfield (1972) ranked the first and second most highly cited authors. Among the highly cited literature, Garfield E contributed 5 papers, both White H.D. and Small H. contributed 3 papers, and Price DJD contributed 2 papers. In terms of the overall published volumes of the authors, Garfield E contributed a total of 230, Leydesdorff Loet a total of 166, Bornmann L a total of 125, Egghe L a total of 114, Glanzel W a total of 108, Small H a total of 93, and Kostoff RN a total of 81. These researchers played important roles in the study of citation analysis and knowledge dissemination.

Since the mid 1950s, the method of citation analysis formally proposed by Garfield achieved rapid development and dissemination. In the early 1960s, Garfield founded SCI, providing the essential data basis for citation analysis, and Price's *LITTLE SCI BIG SCI* laid an important theoretical foundation for the development of citation analysis. 1970–1990s was a crucial development period for citation analysis: Small and Marshakova introduced the analysis method of Document Co-Citation Analysis in 1973, White et al. proposed Author Co-Citation Analysis in 1981, and in the 1990s, White and McCain et al. further developed the analysis method of ACA and introduced the Map Of Scientific Knowledge on this basis. Since the year 2000, citation analysis has witnessed a diversifying development trend particularly in research topics based on citation analysis. These topics include research evaluation index, information visualization, and citation behavior, and have been or are forming peak citation publication years, currently becoming a research hotspot.

**Acknowledgements** This research was supported by the National Natural Science Foundation of China under Grant 71103022.

## References

- Bar-Ilan, J. (2008). Which h-index?—A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257–271.
- Barth, A., Marx, W., Bornmann, L., & Mutz, R. (2014). On the origins and the historical roots of the Higgs boson research from a bibliometric perspective. *European Physical Journal Plus*, 129(6), 1–13.
- Borgman, C. L., Gallagher, A. L., & Krieger, D. (1990). Childrens use of an interactive catalog of science materials. *Proceedings of the ASIS Annual Meetings*, 27, 55–68.
- Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80.
- Bornmann, L., & Marx, W. (2013). The proposal of a broadening of perspective in evaluative bibliometrics by complementing the times cited with a cited reference analysis. *Journal of Informetrics*, 7(1), 84–88.
- Boyack, K. W., Klavans, R., & Borner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351–374.
- Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering*, 137, 85–86.
- Bradford, S. C. (1948). *Documentation*. London: Crosby Lockwood & Son.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117.
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57, 359–377.
- Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis. *Journal of the American Society for Information Science and Technology*, 61(7), 1386–1409.
- Cole, F. J., & Eales, N. B. (1917). The history of comparative anatomy: Part I.—A statistical analysis of the literature. *Science Progress*, 11(44), 578–596.
- Comins, J. A., & Hussey, T. W. (2015a). Compressing multiple scales of impact detection by Reference Publication Year Spectroscopy. *Journal of Informetrics*, 9(3), 449–454.
- Comins, J. A., & Hussey, T. W. (2015b). Detecting seminal research contributions to the development and use of the global positioning system by reference publication year spectroscopy. *Scientometrics*, 104(2), 575–580.
- Comins, J.A., & Leydesdorff, L. (2016). Identification of long-term concept-symbols among citations: Can documents be clustered in terms of common intellectual histories? <http://arxiv.org/abs/1601.00288>.
- Cronin, B., Snyder, H. W., & Rosenbaum, H. (1998). Invoked on the web. *Journal of the American Society for Information Science*, 49(14), 1319–1328.
- Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, 122(3159), 1127–1128.
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178(4060), 471–479.
- Garfield, E. (1979). *Citation indexing: Its theory and application in science technology and humanities*. New York: Wiley.
- Garfield, E. (2006). Citation indexes for science. A new dimension in documentation through association of ideas. *International Journal of Epidemiology*, 35(5), 1123–1127.
- Garfield, E., & Sher, I. H. (1963). New factors in evaluation of scientific literature through citation indexing. *American Documentation*, 14(3), 195–201.
- Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, 28(4), 1661–1707.
- Gross, P. L. K., & Gross, E. M. (1927). College libraries and chemical education. *Science*, 66, 386–389.
- Hagstrom, W. O., & Selvin H. C. (1965). 2 Dimensions of cohesiveness in small-groups. *Sociometry*, 28(1), 30–43.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572.
- Ingwersen, P. (1998). The calculation of Web impact factors. *Journal of Documentation*, 54(2), 236–243.
- Kaplan, N. (1965). The norms of citation behavior: Prolegomena to the footnote. *American Documentation*, 16(3), 179–184.
- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25.

- Kostoff, R. N. (1998). The use and misuse of citation analysis in research evaluation—comments on theories of citation? *Scientometrics*, 43(1), 27–43.
- Kuhn, T. S. (1962). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.
- Leydesdorff, L., Bornmann, L., Marx, W., & Milojević, S. (2014). Referenced publication years spectroscopy applied to iMetrics: Scientometrics. *Journal of Informetrics*, 8(1), 162–174.
- Lipetz, B. A. (1965). Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators. *American Documentation*, 16(2), 81–90.
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of Washington Academy of Sciences*, 16(12), 317–323.
- Macroberts, M. H., & Macroberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science and Technology*, 40(40), 342–349.
- Macroberts, M. H., & Macroberts, B. R. (2009). Problems of citation analysis: A study of uncited and seldom cited influences. *Journal of the American Society for Information Science and Technology*, 61(1), 1–12.
- Marx, W., & Bornmann, L. (2014). Tracing the origin of a scientific legend by reference publication year spectroscopy (RPYS): The legend of the Darwin finches. *Scientometrics*, 99, 839–844.
- Marx, W., Bornmann, L., Barth, A., & Leydesdorff, L. (2014). Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology*, 65(4), 751–764.
- Mccain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*, 41(6), 433–443.
- Merton, R. K. (1965). Nurse-physician collaboration toward improved patient care. Observations. *Papers. National Conference for Professional Nurses and Physicians*, 2, 51–52.
- Moed, H. F. (2005). Citation analysis in research evaluation. *Information Science and Knowledge Management*, 57(1), 13–18.
- Popper, K. R. (1961). *The logic of scientific discovery* (2nd ed.). New York, NY: Basic Books.
- Price, D. J. D. (1963). Little science, big science. *Von Der Studierstube Zur*, 7(3–6), 443–458.
- Price, D. J. D. (1965). Networks of scientific papers. *Science*, 149, 510–515.
- Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105(45), 17268–17272.
- Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, 4(2), 131–134.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science and Technology*, 24(24), 265–269.
- Smith, L. C. (1981). Citation analysis. *Library Trends*, 30(1), 83–106.
- Thor, A., Marx, W., Leydesdorff, L., & Bornmann, L. (2016). Introducing cited references explorer: A program for reference publication year spectroscopy with cited references disambiguation. *Journal of Informetrics*, 10(1), 1–31.
- White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32, 163–171.
- White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science*, 49(4), 327–355.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.