Describing Citations as a Function of Time

Xiaojun Hu[1], Xian Li[1], Ronald Rousseau[2,3†]

[1]Medical Information Center, and Department of Neurology of Affiliated Hospital 2, Zhejiang University School of Medicine, Hangzhou 310058, China
[2]Facultair Onderzoekscentrum ECOOM, MSI, KU Leuven, Naamsestraat 61, Leuven B-3000, Belgium
[3]Faculty of Social Sciences, University of Antwerp, Antwerp B-2020, Belgium

Abstract

Purpose: Providing an overview of types of citation curves.

Design/methodology/approach: The terms citation curves or citation graphs are made explicit.

Findings: A framework for the study of diachronous (and synchronous) citation curves is proposed.

Research limitations: No new practical applications are given.

Practical implications: This short note about citation curves will help readers to make the optimal choice for their applications.

Originality/value: A new scheme for the meaning of the term "citation curve" is designed.

Keywords Citation curves; Citation graphs; Diachronous and synchronous functions; Size-dependent and size-independent indicators; Knowledge domains; Growth

1 Introduction

According to (Rousseau et al., 2018), citation analysis is defined as follows:

Citation analysis is a subfield of informetrics. In this subfield scientists study frequencies (numbers) and patterns related to giving (reference behavior) and receiving (being cited) citations. Citation studies are performed on the level of documents, authors, universities, countries, and any unit that might be of interest.

A citation is an acknowledgment received by a publication from another one. These citations can be counted and represented as functions of a time variable, leading to time-dependent citation curves as graphical representations. Such curves play an important role as visualization tools in citation studies. In recent

---

† Corresponding author: Ronald Rousseau (E-mail: ronald.rousseau@kuleuven.be).

investigations, we realized that there does not exist an overview of the different ways in which citation functions and their corresponding graphs can be represented (at least we could not find such an overview suited to our purposes). Providing such a systematic overview is the main purpose of this article.

Before continuing we like to make a cautionary note. The term "the citation distribution of papers" as used in the informetric literature as well as elsewhere, may refer to two totally different situations. The first distribution describes the (relative) number of articles in the set under study, typically a journal, with $n = 0,1,2,\ldots$ received citations. This leads to a function G(n), where n is a natural number. In most studies, this number of citations is a cumulative number, collected over some period. It is then, moreover, possible to study this distribution in a dynamic way, i.e., over different periods as was done e.g. in (Stringer et al., 2008).

The second distribution describes the total number of citations given or received (see further for the difference between a diachronous and a synchronous study) by one article or a set of articles over the years. This leads to a function f(t), with t denoting time. Remarkably, G(n) as well as f(t) has been modeled by a (discrete) lognormal distribution (Matricciani, 1991; Stringer et al. 2008).

As mentioned above, there are at least two different points of view when considering time-dependent citation curves: a diachronous (prospective) and a synchronous (retrospective) one. These types of curves differ as follows (Rousseau et al., 2018). In a synchronous study, one considers one object (database, journal, research group, maybe even one article or one patent) and collects the age of the references used by this object. In other words: one studies the distribution over time, of cited publications. In this case, the arrow of time goes from the present to the past. In a diachronous study, one considers a set of publications (possibly just one) published in the same year and collects the number of citations received over the years. The time arrow goes from the publication year in the direction of the future. It is important to note here that in this case citations are received from different populations, as databases may change their coverage and the population of scientists certainly changes over time. Nakamoto (1988) provides a nice comparison of a synchronous and a diachronous view of data originating from the Science Citation Index (diachronous for 1961 publications and synchronous for 1982 citations, showing a high degree of symmetry). We finally mention that many citation curves are a mixture of a diachronous and a synchronous viewpoint, see (Liu & Rousseau, 2008; Rousseau et al., 2018). In these studies, we described time series of the form: a number of citations divided by a number of publications. In the present work, we focus on the basic case, i.e., with one fixed publication year. We refer to (Adams, 2018) for a warning about possible misunderstandings when the time series used to construct a citation curve is not precisely defined.

Describing Citations as a Function of Time   Xiaojun Hu et al.

Note

An important aspect of the basic diachronous case (but also of other diachronous series) is how to deal with the changing user population. Although in this contribution, citations will also be studied as fractions with respect to all possible ones (see further for details) the result is not intended for comparisons with e.g., citations in other fields. Hence these fractions are not calculated for normalization purposes (as is necessary for comparisons) but to study how the shape of a citation curve may change, taking changing user populations, i.e., citing or possibly citing scientists, into account.

Formulated as a research question we come to:

Which types of basic diachronous (synchronous) citation curves may be useful in practical applications?

The present article further consists of a section on diachronous citation curves for absolute (size-dependent) data and for data relative with respect to the database (size-independent), followed by a section on diachronous citation curves, relative with respect to the article's knowledge base. We briefly redo this analysis for the synchronous case. In the following sections, we give a short overview about how to determine an article's knowledge base and how citation curves have been modeled. We end with some remarks about the use of citation curves and a conclusion.

2   First types of diachronous citation curves: size-dependent and size-independent values

We consider a fixed article A published in the year Y, which will be referred to as year 0. The number of citations received by this article in the year t, with t = 0, 1, 2, … is denoted as c(t). Citations are collected from a given, but of course, growing, database DB. We note that this database DB may be a virtual one, in the sense that it is a union of several "concrete", i.e., really existing databases. This happens, e.g., when data are collected in the Web of Science (WoS) and in Scopus. The number of received citations is collected on the first day of the year. Hence c(0) is by definition equal to zero, and c(1) denotes the number of citations received during the publication year.

Collecting (absolute) citation data year by year leads to a graph of the function c(t). Adding these yearly data leads to cumulative data $C(t) = \Sigma_{j=0}^{t} c(j)$, and its corresponding graph. By definition C(0) = 0 and C(1) = c(1). The graph of C(t) is by definition non-decreasing. We note that cumulative data are often described in the literature as growth curves. These cases can be found in Table 1, in the "size-dependent" row.

Next, we consider a size-independent case, namely data relative to the whole database (see Table 1). If d(t) denotes the number of publications published in the

year t and included in database DB, then we may consider the relative number $c_d(t) = c(t)/d(t)$. This ratio denotes the relative number of "new" publications, and, when multiplied by 100, the percentage of articles added to the database in the year t that cite article A. This number is, of course, always very small. Note that $c_d(0)$ is set equal to zero. The ratios $c_d(t)$ lead to a yearly relative citation graph.

If now $D(t) = \Sigma_{j=0}^{t} d(j)$ denotes the total number of publications included in the database since the year Y, then $C_D(t) = C(t)/D(t)$ gives the relative number of publications in the database that cite article A, since the year Y. These ratios lead to yet another citation curve, namely a cumulative one.

Now we make two observations. Unless an article has universal appeal or the used database covers a small set of publications related to the same field of science, it makes little sense to compare to the whole database: comparing to the knowledge domain (field, discipline) to which article A belongs is practically a better alternative. This will be studied further on. Next, we would like to make another observation: it makes more sense to restrict publications in the database to "normal articles" or "normal articles and reviews" than to include all types of publications. We will not return to this second observation and keep the term "publications" but this possible restriction to certain types of publications is always assumed.

Table 1.   Types of basic diachronous citation curves.

	Non-cumulative	Cumulative
Size dependent	c(t)	$C(t) = \Sigma_{j=0}^{t} c(j)$
Size independent	With respect to the knowledge domain $c_k(t) = c(t)/k(t)$	With respect to the knowledge domain $C_K(t) = C(t)/K(t)$
	With respect to the database $c_d(t) = c(t)/d(t)$	With respect to the database $C_D(t) = C(t)/D(t)$

3   A second type of size-independent citation curves: relative values with respect to an article's knowledge domain

3.1   Basic observations

The main problem in this step is to define the knowledge domain of article A. Once the knowledge domain KD is precisely defined then we may replace the term "database" in the previous section by the term "knowledge domain" to obtain the corresponding definitions and citation graphs. Of course, we assume that articles belonging to this knowledge domain are also included in database DB.

If k(t) denotes the number of publications published in the year t, belonging to the knowledge domain KD, then we consider the relative number $c_k(t) = c(t)/k(t)$. This ratio denotes the relative number or when multiplied by 100, the percentage of "new" publications in the knowledge domain KD that cite article A. Again this

Describing Citations as a Function of Time   Xiaojun Hu et al.

Note

number is usually much smaller than one, but larger than or equal to $c_d(t)$. The ratios $c_k(t)$ lead to a yearly relative citation graph with respect to the knowledge domain to which A belongs.

If now $K(t) = \Sigma_{j=0}^t k(j)$ denotes the total number of publications belonging to knowledge domain KD since the year Y, then $C_K(t) = C(t)/K(t)$ gives the relative number of publications in knowledge domain KD that cite article A, since the year Y. These ratios lead to yet another size-independent cumulative citation curve.

Probably, using relative values with respect to an article's knowledge domain is the best approach in most practical situations.

Although not the essential part of this article, we still provide a short overview in the next section, of the construction/definition of knowledge domains, because they are essential for the construction of relative citation graphs.

3.2   What are knowledge domains?

Now we come to a difficult question: how to define the knowledge domain KD? Answering this question is important for many scientometric studies, interdisciplinarity being one of them (where the term discipline is used as a synonym for knowledge domain). We provide some suggestions, admitting that many more exist or could be proposed. For a recent and detailed description of the delineation of knowledge domains we refer to (Zitt et al., 2019).

To determine the knowledge domain of an article we consider three alternatives: using ready-made classifications, using an algorithmic approach and determining the knowledge domain of an article through the domains to which its authors belong.

Traditionally, when scientists use the WoS as the database, WoS Subject Categories are used as knowledge domains. This choice has several disadvantages. When using WoS Subject Categories the knowledge domain of an article is determined not by its contents, but by the journal in which it is published. Moreover, a journal may belong to several WoS Subject categories. Finally, it is well-known that many Subject Categories are not correctly defined, the field of information science being a case in point (Leydesdorff & Bornmann, 2016) and Multidisciplinary Sciences is a subject category, but not a knowledge domain.

Using ESI (Essential Science Indicators) categories as knowledge domains is slightly better, as there is no overlap between ESI categories (Liu & Rousseau, 2010). Still, also when applying ESI categories, an article's knowledge domain is determined by a journal and not by its actual scientific content. These are two examples, among several more (Zitt et al., 2019) of ready-made classifications.

A more refined, publication-based, algorithmic approach is used by CWTS (Leiden, the Netherlands). Waltman and van Eck (2012) constructed a classification

system based on publications in the Web of Science database for the period 2001–2010. All publications of document types article, letter, and review in the sciences and the social sciences were included. Publications in the arts and humanities were not included. Their general methodology allowed for L hierarchical levels but in (Waltman & van Eck, 2012) they provided an example with three levels (taking L = 3): a first level of 20 broad disciplines, a second level of 672 research areas, and a third level of 22,412 small subfields (micro-level fields). For further details and its relation with the Leiden rankings we refer to (Waltman & van Eck, 2012) and the website of the Leiden Ranking (https://www.leidenranking.com/). Using these small non-overlapping, clusters as "knowledge domains" could yield a much finer result than with the classical Web of Science categories. Yet, we do not know any example where these micro-level fields have been used in studies of interdisciplinarity. We also like to point out that these micro-level fields change each year, which might be a reason why they have not yet been used for studies related to interdisciplinarity. An updated version of the original 2012 version has been used, e.g. in (Milanez et al., 2016). The authors of this article were able to retrieve and delineate the real nuclei and the peripheral research areas related to nanocellulose studies.

Somewhat similarly, knowledge domains can also be defined as topics, found through a topic detection algorithm as discussed e.g. in (Yan, 2015). Recently, Sjögårde and Ahlgren (2018) experimented with another algorithmic classification, also based on community detection techniques, leading to an algorithmically constructed publication-level classification of research publications (ACPLC). They propose using synthesis papers and their reference articles to construct a baseline classification. In this way, several ACPLCs of different granularity were constructed. Each ACPLC is compared to the baseline classification and the best performing ACPLC is identified. They found that in this approach class size variation is moderate, and only a small proportion of the publications belong to very small classes. These investigations were continued in (Ahlgren et al., 2019).

Finally, one may define the knowledge domain of a publication as the knowledge domain (or knowledge domains), counted fractionally, to which the authors of this publication, or alternatively the authors of the references, belong. Here one may use departmental affiliations (but these are administrative units, not knowledge domains) or one may try to define the knowledge domain of a scientist algorithmically, e.g., as the knowledge domain in which they publish the most (noting that this is just the most elementary way of proceeding and that nowadays applying a deep learning algorithm would be more advisable). Yet, in some countries such as Italy, professors must classify themselves in one, and only one, of 370 scientific knowledge domains (referred to as scientific disciplinary sectors), see (Abramo et al., 2018). This yields another interesting definition of the knowledge domain to which a publication belongs.

Describing Citations as a Function of Time   Xiaojun Hu et al.

Note

4   Synchronous citation curves: size-dependent and size-independent values

Although this section is largely a copy of sections 2 and 3.1 it may be useful to have a description of the synchronous and the diachronous approach together in the same paper. Moreover, this was required by a reviewer.

We consider an article or more appropriate, a set S of articles dealing with the same topic and published in the year Y, which will be referred to as year 0. The number of references in this set of articles published in the year Y-t, with t = 0, 1, 2, … is denoted as r(t). Note that r(0) is the number of references published in the year Y, the same publication year as the articles in the set S. The graph r(t), t = 0, 1, … is the synchronous citation curve of the set S. In practice such a graph is cut off after a given number of years, so that the occasional reference to an article published 73 years ago does not play a role. Adding these yearly data leads to cumulative data $R(t) = \Sigma_{j=0}^{t} r(j)$, and its corresponding graph. The graph of R(t) is by definition non-decreasing. We note that when studying the references of a set S the same article may be included in several reference lists. Depending on the purpose of the investigation (focusing on publication years or on actually used articles) these articles may be counted once or as often as they occur.

Next, we consider the size-independent case, namely data relative to the whole database (DB) or to the knowledge domain (KD). If d(t) denotes the number of publications published in the year Y-t and included in database DB (or belonging to knowledge domain KD), then we consider the relative number $r_d(t)$ = r(t)/d(t). This ratio denotes the relative number of used publications occurring in the set S. The ratios $r_d(t)$ lead to yearly relative reference graphs (going back in time).

If now $D(t) = \Sigma_{j=0}^{t} d(j)$ denotes the total number of publications included in the database DB (or belonging to the knowledge domain KD) since the year Y-t, then $R_D(t)$ = R(t)/D(t) gives the relative number of publications in the database that are referred to in the set S. These ratios lead to cumulative citation curves. Table 2 gives an overview of the synchronous case.

Table 2.   Types of basic synchronous citation curves.

	Non-cumulative	Cumulative
Size dependent	r(t)	$R(t) = \Sigma_{j=0}^{t} r(j)$
Size independent	With respect to the knowledge domain $r_k(t)$ = r(t)/k(t)	With respect to the knowledge domain $R_K(t)$ = R(t)/K(t)
	With respect to the database $r_d(t)$ = r(t)/d(t)	With respect to the database $R_D(t)$ = R(t)/D(t)

5   Modeling citation curves as a function of time

5.1   The Avramescu model

Besides using observed citation data in informetric studies, one may first model the observed data and then continue investigations using this model. We mention some articles studying publications, but also include some studying journals. This makes no difference when it comes to absolute citations as a function of time. Avramescu (1979) proposed the following equation for the diachronous absolute citation distribution of individual papers or journals:

$$c_A(t) = C_1(e^{-bt} - e^{-mt})$$

Clearly, the function $c_A(t)$ has three parameters: C_1, the number of citations in the publication year, m and b, with $m > b$ (where the parameter b is allowed to be negative). If the parameters b and m are positive, one obtains a curve which quickly reaches a top and then decreases slowly with limit zero. This form corresponds roughly to the so-called basic journal citation model as studied in (Rousseau et al., 2001). If m is small (but positive) and the parameter b is negative one obtains an exponentially increasing function, which may be used to model delayed recognition, among others. Yet, based on theoretical and practical grounds, Avramescu's model has been rejected by Egghe and Rao (1992a) because its aging function $a(t) = c(t+1)/c(t)$ does not have a minimum.

5.2   The lognormal distribution describes two totally different citation curves

Matricciani (1991) was probably the first to observe that in the synchronous case (age of references in papers) citation data can be described by a lognormal distribution. This fact has been confirmed by Egghe and Rao (1992a), be it that they published only data for references in books. Recently, the lognormal distributions for diachronous as well as for synchronous citation curves have been brought together in an overarching model, the so-called Wang-Song-Barabási model (Wang et al., 2013; Yin & Wang, 2017).

Yet, as mentioned before, when studying journal citations a totally different citation curve can be used. For these curves, the horizontal axis does not describe time (age) but the number of articles with a given number of citations. Time is kept fixed for one such curve. Stringer et al. (2008, 2010) showed how also these graphs follow a (discretized) lognormal distribution. Studying these functions over time these authors observed that such lognormal distributions moved over time (the top is situated at a higher number of received citations) until they reach a stable state.

Describing Citations as a Function of Time   Xiaojun Hu et al.

Note

If one wants to find a best-fitting lognormal curve one must first make a decision about how to handle the zero case (as log(0) = - ∞). There are three simple options: removing zeros, adding 0.5 to all observed data, or adding 1 to all data. For a general (not just for the lognormal) best solution we refer to a recent working paper by Bellégo and Pape (2019).

Besides the continuous lognormal distribution, one may also model citation data using the discretized lognormal distribution, as done by Stringer et al. (2010) and illustrated also by Thelwall (2016). Yet, this author points out that in most practical cases it is more reasonable to use the continuous distribution approximation in order to mathematically analyze citation indicators.

6   Some observations

In this section, we present some observations related to how citation curves have been used in the literature. The cases listed are just illustrations and the authors are well aware of the fact that many more examples can be found.

6.1   Growth

Cumulative publication and citation curves for authors, research groups, journals and so on, never decrease. As a consequence growth has been studied by many colleagues. For an older article including several growth models, we refer to (Egghe & Rao, 1992b). Especially when studying the relative growth or decline of a country's publication output, it is necessary to consider relative values with respect to the database used. This was already observed by Leydesdorff (1988), among others.

6.2   Influence of growth on obsolescence

Citation and publication curves, play a role in aging studies, i.e., work on obsolescence, in which the decrease in the use and hence the utility of the scientific literature is studied (Egghe et al., 1995; Sun et al., 2016). If the literature would not grow than the diachronous citation curve would have its "pure" form. Yet, because of the growth of literature, the observed citation curve is not the same as this "pure" form. This leads to the question: what is the influence of growth on obsolescence? Using an exponential model for the aging curve as well as for the growth function, Egghe et al. (1995) found that growth of the literature increases synchronous obsolescence, but decreases the diachronous one.

6.3   A new finding based on a relative citation curve with respect to the knowledge domain

In (Hu & Li, 2019) the authors define the notion of a rejuvenated article as one that received, relatively speaking, i.e., with respect to the knowledge domain, more

citations in the second period of its lifetime than in the first. Such an article must, moreover be at least 20 years old and may not have suffered delayed recognition. The present article is meant as a preparation for a more precisely formulated follow-up of (Hu & Li, 2019).

7 Conclusion

Once one has a time series it can be represented as a function or a graph with time as its parameter. In the present work we only considered the basic diachronous and synchronous case (size-dependent and size-independent), i.e., only one, fixed, publication year is taken into account. The word "size-independent" refers either to the whole database or to the knowledge domain to which the target article(s) belong(s). A short overview of how to delineate an article's knowledge domain is included. We hope that this short note about citation curves helps readers to make the best choice for their applications.

Acknowledgments

Author contributions

Conceptualization: Ronald Rousseau (ronald.rousseau@kuleuven.be); funding acquisition: Xiaojun Hu (xjhu@zju.edu.cn); methodology: Xiaojun Hu, Xian Li (lx2009yet@126.com), and Ronald Rousseau; original draft: Ronald Rousseau; writing and reviewing final version: Xiaojun Hu, Xian Li, and Ronald Rousseau.

References

Abramo, G., D'Angelo, C.A., & Zhang, L. (2018). A comparison of two approaches for measuring interdisciplinary research output: The disciplinary diversity of authors vs the disciplinary diversity of the reference list. Journal of Informetrics, 12(4), 1182–1193.

Adams, J. (2018). Information and misinformation in bibliometric time-trend analysis. Journal of Informetrics, 12(4), 1063–1071.

Ahlgren P., Chen, YW., Colliander,C., & van Eck, N.J. (2019). Community detection using citation relations and textual similarities in a large set of PubMed publications. In: Catalano, G., Daraio, C., Gregori, M., Moed, H.F., & Ruocco, G., (Eds.). In Proceedings of the 17th International Conference on Scientometrics & Informetrics (ISSI2019). (pp. 1380–1391). Rome: Efesto.

Avramescu, A. (1979). Actuality and obsolescence of scientific literature. Journal of the American Society for Information Science, 30(5), 296–303.

Describing Citations as a Function of Time   Xiaojun Hu et al.

Note

Bellégo, C., & Pape, L.D. (2019). Dealing with the logs of zeros in regression models. Série des Documents de Travail, 2019-13. Available at SSRN: https://ssrn.com/abstract=3444996 or http://dx.doi.org/10.2139/ssrn.3444996

Egghe, L., & Rao, I.K.R. (1992a). Citation age data and the obsolescence function: Fits and explanations. Information Processing & Management, 28(2), 201–217.

Egghe, L., & Rao, I.K.R. (1992b). Classification of growth models based on growth rates and its applications. Scientometrics, 25(1), 5–46.

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

Hu, X.J., & Li, X. (2019). Do "rejuvenated" articles exist? In Proceedings of the 15th International Conference on Webometrics, Informetrics, & Scientometrics (WIS) & the 20th COLLNET Meeting, (pp. 23–29). Dalian: Dalian University of Technology.

Leydesdorff, L. (1988). Problems with the 'measurement' of national scientific performance. Science and Public Policy, 15(3), 149–152.

Leydesdorff, L., & Bornmann, L. (2016). The operationalization of "Fields" as WoS Subject Categories (WCs) in evaluative bibliometrics: The cases of "Library and Information Science" and "Science & Technology Studies". Journal of the Association for Information Science and Technology, 67(3), 707–714.

Liu, Y.X., & Rousseau, R. (2008). Definitions of time series in citation analysis with special attention to the h-index. Journal of Informetrics, 2(3), 202–210.

Liu, Y.X., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion. Journal of the American Society for Information Science and Technology, 61(2), 340–351.

Matricciani, E. (1991). The probability distribution of the age of references in engineering papers. IEEE Transactions on Professional Communication, 34(1), 7–12.

Milanez, D.H., Noyons, E., & de Faria, L.I.L. (2016). A delineating procedure to retrieve relevant publication data in research areas: the case of nanocellulose. Scientometrics, 107(2), 627–643.

Nakamoto, H. (1988). Synchronous and diachronous citation distributions. In: L. Egghe & R. Rousseau. (Eds.), Informetrics 87/88 (pp. 157–163). Amsterdam: Elsevier Science Publishers.

Rousseau, R., Egghe, L., & Guns, R. (2018). Becoming Metric-Wise. A bibliometric guide for researchers. Kidlington (UK): Chandos-Elsevier.

Rousseau, R., Jin, B.H., Yang, N., & Liu, X. (2001). Observations concerning the two- and three-year synchronous impact factor, based on the Chinese Science Citation Index. Journal of Documentation, 57(3), 349–357.

Sjögårde, P., & Ahlgren, P. (2018). Granularity of algorithmically constructed publication-level classifications of research publications: Identification of topics. Journal of Informetrics, 12(1), 133–152.

Stringer, M.J., Sales-Pardo, M., & Amaral, L.A.N. (2008). Effectiveness of journal ranking schemes as a tool for locating information. PLoS One, 3(2): e1683.

Stringer, M.J., Sales-Pardo, M., & Amaral, L.A.N. (2010). Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. Journal of the American Society for Information Science and Technology, 61(7), 1377–1385.

Note

Sun, J.J., Min, C., & Li, J. (2016). A vector for measuring obsolescence of scientific articles. Scientometrics, 107(2), 745–757.

Thelwall, M. (2016). Citation count distributions for large monodisciplinary journals. Journal of Informetrics, 10(3), 863–874.

Waltman, L., & van Eck, N.J. (2012). A new methodology for constructing a publication-level classification system of science. Journal of the American Society for Information Science and Technology, 63(12), 2378–2392.

Wang, D., Song, C., & Barabási, A.L. (2013). Quantifying long-term scientific impact. Science, 342(6154), 127–132.

Yan, E.J. (2015). Research dynamics, impact, and dissemination: A topic-level analysis. Journal of the Association for Information Science and Technology, 66(11), 2357–2372.

Yin, Y., & Wang, D.S. (2017). The time dimension of science: Connecting the past to the future. Journal of Informetrics, 11(2), 608–621.

Zitt, M., Lelu, A., Cadot, M., & Cabanac, G. (2019). Bibliometric delineation of scientific fields. In: W. Glänzel, H.F. Moed, U. Schmoch, M. Thelwall (Eds.), Springer Handbook of Science and Technology Indicators (pp. 25–68). Cham: Springer.