Citation analysis

Have you ever wanted to know how much impact an article or paper has had upon its discipline? Citation analysis is one tool which you can use to get some indication of how significant a piece of work has been. This is done by analysing the number of times it has been cited by other scholars.

In this resource, we will explore what citation analysis is, demonstrate how it is used to measure impact, and how you can conduct a basic analysis. We will also highlight how citation analysis differs depending on the field of academic study.

# What is citation analysis?

A citation is where you refer to a particular piece of work within the main body of your text.

Below you can see the heading of a paper by the Nobel Prize-winning physicist Ernest Rutherford of the University of Manchester, in which he put forward a new and highly influential model of the atom.

"The scattering of α and β particles by matter and the structure of the atom"

Rutherford, E. Philosophical magazine (Abingdon, England), 2012, Vol.92 (4), p.379-398

This article appeared in the scientific journal called the Philosophical Magazine in 1911. Other scientists then read the article, and subsequently carried out and wrote up research which was influenced by Rutherford's work. We know this because they included that 1911 article in their list of references. It is very unlikely they would have included it if it was not relevant to their research.

## **The** assumptions

Below you can see two example citations to Ernest Rutherford’s paper.

Moseley, H.G., 1914: The high-frequency spectra of the elements. Part II. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 27(160), pp.703-713.

Noyes, W.A., 1917: A kinetic hypothesis to explain the function of electrons in the chemical combination of atoms. Proceedings of the National Academy of Sciences, 3(5), pp.356-360.

This leads us to the main basic assumptions of citation analysis (although it must be noted most citation analyses do limit the risk of these assumptions with all sorts of caveats and qualifiers).

The assumptions are:

1. That citation of an earlier article in a later article is an indicator that the earlier article had some positive influence on the later article, and...
2. that the more citations an article receives, the more influential it is.

These assumptions are open to question, and the possibility of questioning them lies behind the caveats and qualifiers that appear in many published citation analyses.

# Sources of citation data

Fortunately, it's not necessary to read through all the lists of references in all the articles ever published in order to find out which articles have cited which other articles.

In the 1950s, the researcher Eugene Garfield came up with the idea of compiling a reference work in which one could look up a journal article and find details of all the subsequent articles which had cited it. The work which he created has now become the electronic database Web of Science.

In the first years of this century, the publishers Elsevier followed suit with a similar database, called Scopus. Other sources of citation data are also available, such as the free package Publish or Perish, which is based on citations from Google Scholar.

The [Library's Research Metrics team](https://www.library.manchester.ac.uk/services/research/metrics/) can provide you with guidance and support in the effective use of specific tools listed below:

* [Scopus](https://www.librarysearch.manchester.ac.uk/permalink/44MAN_INST/bofker/alma9928963144401631): Citation data is embedded throughout Scopus at the journal, document and author levels. This can be a good place to look for information quickly – Scopus user guide.
* [SciVal](https://www.scival.com/home): Uses data from Scopus but allows you to compare and benchmark institutions, research areas and researchers. It allows for a more detailed analysis and contains reporting modules.
* [Web of Science](https://www.librarysearch.manchester.ac.uk/permalink/44MAN_INST/bofker/alma992975775680101631): The Core Collection provides indexing back to 1900 across the world’s highest-quality and most impactful publications.
* [Research Metrics team](https://www.library.manchester.ac.uk/services/research/metrics/index.htm): Guidance on how to use all the above is available from the Library's Research Metrics team

# Impact

It is worth remembering that bibliometrics were originally designed by librarians, primarily as a tool to help them make decisions around which content to purchase.

During the 1960s, researchers began to realise that they could also be used as a way in which to identify developments in various sciences – also known as “the science of science”. Following the shift to the online availability of citation data in the 2000s (through databases such as Web of Science and Scopus), the metrics generated began to be used by universities as a means of demonstrating research impact and to support research assessment exercises.

More recently, there has been a shift towards ensuring that citation data, considers qualitative as well as quantitative data, via the 'responsible metrics movement'.

## What can citations tell us about individual articles, journal publications, individual researchers, HE institutions?

* **Articles** – High levels of citations indicate important / influential research. They often take on a snowball effect of influence due to the citation count.
* **Journals** – Articles published in prestige journals generally receive high numbers of citations e.g. Nature. This in turn means certain journals attract high quality submissions because of their prestige.

Individual authors can leverage this prestige for career recognition and advancement, or even fame and fortune. All universities can (and do) make use of citation metrics for rankings purposes, and to demonstrate the impact of their research.

# Summary

Citation analysis is a way of measuring the impact of an author, article or publication by counting the number of times these have been cited by other works. In the next section, we will explore why we conduct citation analysis and how the number of citations varies according to subject area.

Why do we use citations to measure impact?

# Introduction

In the previous section you looked the basic assumptions of citation analysis, but these assumptions are not always a true reflection. In this section you will explore how citations vary between different subject areas and how this affects the way in which we analyse them.

# Academic comparison

Here we will explore the citations received by two well-known academics from two separate fields. Learn more about the academics below.

### Brian Cox:

* Articles published: 379
* Citations received: 13,251
* Citations received per article: 34.96

Formally the keyboard player on D:Ream's number one hit single Things Can Only Get Better, Brian is now a Professor of Physics at the University of Manchester. These days, he is better known as the presenter of BBC2’s Stargazing Live plus a number of science documentaries, and as co-presenter of Radio 4's Infinite Monkey Cage with comedian Robin Ince.

### Tristram Hunt:

* Articles published: 9
* Citations received: 13
* Citations received per article: 1.44

Tristram Hunt has presented television programmes on the English Civil War, the work of Sir Isaac Newton, the Protestant Revolution and the Joy of Motoring, and is Labour MP for Stoke-on-Trent Central and a former Shadow Education Secretary. He is also a Senior Lecturer in History at Queen Mary University of London, and his breadth of activities may make him the humanities world's answer to Brian Cox.

## How can you compare citations?

**Is it possible to come up with a fair comparative citation statistic?**Citation analysis allows us to use statistics to present citation figures that are comparable. This section will look at how this is done.

## How you compare citations

Going back to the fields of Physics and History; below we can see a breakdown using percentiles which makes the data more comparable.

Using percentiles, we can compare citation levels in the two fields. A percentile is a group of the top X% of something, on some scale. In this case, we are talking about academic papers on a scale of numbers of citations. The first percentile is the group of papers which are in the top 1% for number of citations. In this example, for an article to be in the top 1% (or first percentile) of physics papers, it needs to have been cited 108 times. Whereas in history, it would only need 26 citations to be in the top 1%.

Using these figures, you can measure the importance of a piece of work in each field using our basic assumptions of citation analysis.

### Academic comparison

History and Physics

Below you can see a comparison table which can be used to compare the citation levels of articles across various fields.

|  |
| --- |
| Thresholds for percentile ranges for 2010 publications |
| Field | Top 1% | Top 5% | Top 10% | Top 25% | Top 40% | Top 50% | Top 60% |
| Computer Networks and Communications | 16 | 6 | 3 | 1 | 0 | 0 | 0 |
| Control and Systems Engineering | 22 | 8 | 5 | 2 | 1 | 0 | 0 |
| Electrical and Electronic Engineering | 26 | 11 | 6 | 2 | 1 | 0 | 0 |
| Hardware and Architecture | 17 | 7 | 4 | 1 | 0 | 0 | 0 |
| Materials Chemistry | 40 | 19 | 13 | 7 | 4 | 3 | 2 |
| Renewable Energy, Sustainability and the Environment | 48 | 24 | 16 | 8 | 5 | 3 | 1 |
| Signal Processing | 22 | 10 | 6 | 2 | 1 | 0 | 0 |
| Theoretical Computer Sciences | 16 | 7 | 4 | 1 | 0 | 0 | 0 |

You can use the completed comparison table below to compare citation levels. For example, the ‘Distribution network capacity assessment’ article has received more than three times as many citations as the ‘Adaptive nonlinear manifolds’ article. Despite this, we know that they are both popular and influential in their fields, as they both fall into the top 1% of citations. Being able to make comparisons across subject areas is key to citation analysis.

All articles in the table were published in 2010.

|  |  |  |  |
| --- | --- | --- | --- |
| Article title | Field | Number of citations | Percentile range (Top%) |
| A Negative Imaginary Lemma and the stability of interconnections of linear negative systems | Control and Systems Engineering | 21 | 5% |
| A passivity-based approach to reset control systems stability | Control and Systems Engineering | 22 | 1% |
| Adaptive nonlinear manifolds and their applications to pattern recognition | Theoretical Computer Sciences | 16 | 1% |
| An 80x80 general-purpose digital vision chip in 0.18?m CMOS technology | Hardware and Architecture | 8 | 5% |
| Cyclopentadithiopjene based polymers – A comparison of optical, electrochemical and organic field-effective transistor characteristics | Materials Chemistry | 19 | 5% |
| Design of a multiresonant beam for broadband piezoelectric energy harvesting | Signal Processing | 13 | 5% |
| Distribution network capacity assessment: Variable DG and active networks | Electrical and Electronic Engineering | 59 | 1% |
| Life cycle analysis for future photovoltaic systems using hybrid solar cells | Renewable Energy, Sustainability and the Environment | 59 | 1% |

## Responsible metrics

Citations and percentiles are examples of metrics used in research assessment. We’ve already highlighted some of the limitations of research metrics. There are clear benefits to being able to measure and quantify aspects of research activity, but there are inherent risks in the inferences and conclusions we might draw from quantitative data. Citations and other metrics are part of the broader picture of research assessment, but other aspects such as contributions to open science, mentoring etc are more difficult to measure.

"Some of the most precious qualities of academic culture resist simple quantification”

(Wilson, 2015) from article [We need a measured approach to metrics](https://www.nature.com/articles/523129a#:~:text=Some%20of%20the%20most%20precious,distorting%20behaviour%20and%20determining%20careers) (https://doi.org/10.1038/523129a)

The University of Manchester is a signatory of the [San Francisco Declaration on Research Assessment (DORA)](https://sfdora.org/read/), which means that as an institution we’ve made a number of commitments in terms of how we’ll use research metrics.

A crucial one is the commitment to not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of an individual piece of research, or in recruitment or promotion decisions – simply having had a paper published in a ‘high profile’ journal shouldn’t automatically be an assumption of the quality of the research or inform decisions about the researcher’s career progression.

"...research metrics can provide crucial information that would be difficult to gather or understand by means of individual expertise. But this quantitative information must not be allowed to morph from an instrument into the goal." [The Leiden Manifesto for Research Metrics](http://www.leidenmanifesto.org/)

The Library has contributed to a University of Manchester project to increase awareness of the importance of using research metrics responsibly, and to further embed responsible principles into the University’s research practice.

"The University uses metrics to inform its assessment of research performance, whilst recognising that metrics should be used only in an appropriate and responsible way" [University of Manchester responsible metrics statement](https://www.library.manchester.ac.uk/services/research/open-research/responsible-metrics/)

You can read the full University of Manchester’s current position on the [responsible use of metrics on our website](https://www.library.manchester.ac.uk/services/research/open-research/responsible-metrics/).

## Summary and further support

Citation analysis can allow you to interpret citation data in a way which enables you to make comparisons between different subject disciplines. It also allows you to get an indication of the impact an article or author has had on subsequent works.

Remember, however, that when conducting citation analysis, we are working with a set of assumptions that are open to questioning, particularly an assumption that the more citations an article receives, the more influential it has been.

# Related resources

1. The [Research Metrics web pages](https://www.library.manchester.ac.uk/services/research/metrics/) provide expert advice on matters relating to the citation performance of research outputs.
2. Eugene Garfield is the creator of the Science Citation Index (the first citation database) and pioneer of citation analysis. His [website contains many useful links](http://garfield.library.upenn.edu/category.html), in particular the links to [Essays by Category](http://garfield.library.upenn.edu/cronin/citationprocess.pdf) and the links under the heading My Colleagues & Other Authors. These give open access to a large number of full-text publications on the theory of citation analysis, including such works such as Blaise Cronin's book '[The Citation Process](http://garfield.library.upenn.edu/cronin/citationprocess.pdf)' and Paul Wouters's doctoral thesis [The Citation Culture](http://garfield.library.upenn.edu/wouters/wouters.pdf).
3. [The Metrics Toolkit](https://www.metrics-toolkit.org/metrics/) helps scholars and evaluators understand and use citations, web metrics, and altmetrics responsibly in the evaluation of research. This includes information on how each metric is calculated, where you can find it, and how each should (and should not) be applied.