

Citation inflation and its remedies

James D. Wells

Scholarbox H1 / May 2, 2016

The recent announcement from the ATLAS and CMS particle physics experiments at CERN that they might be seeing a signal of a 750 GeV resonance (i.e., new particle) decaying into two photons is very exciting. Much is being written on this potential signal for new physics. However, there is another phenomenon that is not new but is equally visible in the wake of this development. That phenomenon is citation inflation.

Citation inflation is when authors write a paper and reference many more papers than need be referenced, and often well beyond those they have even read or looked at. In the old days a reference to a paper was listed because the author(s) directly used a result for their present study, or the reference was acknowledged to be first to recognize some specific finding in the research field. Today references are added in the dozens in nonspecific contexts. For example, you may read in a paper a sentence like this: "Other studies [1-78] have addressed the possible interpretations of anomalous $g-2$." And then in the references section of the paper there are 78 papers listed, numbered 1 through 78.

When Einstein wrote his theory of Brownian motion article he cited only two authors, himself and Lectures on Mechanics by Kirchhoff. Today, reference lists in papers much shorter than Einstein's can extend into the hundreds of publications.

Origin of citation inflation

What is the reason for this inflation of citations? For one, science has progressed. We have many more theoretical physicists in the world than when Einstein was working, and many more publications. Perhaps the ratio of citations in individual papers today to those of Einstein's time is consistent with the ratio of total number of papers today vs. then. However, even if this were so, it is unambiguous that the referencing today includes carpet bombing of marginally relevant papers compared to the referencing of yore.

A second and more insidious reason for this dramatic increase in referencing is that it is completely free to reference as many papers as you like. There is no down side to reference an even marginally relevant paper, but potential downsides if you do not — you may get an angry email asking why you are not citing their paper(s) even if they are only tertiarily relevant to your study. There is no reason to deny such giving in to such demands since there is no penalty today for citing an almost arbitrary number of papers.

Perhaps we as a community do not wish to rectify this problem. Citation inflation is occurring, yes, but with online articles it arguably does not matter that it takes up a lot more space at the end of an article, and maybe readers want to see all the papers that are even remotely relevant to the subject.

However, there are at least two reasons why we may wish to bring this citation inflation under control. One, it becomes harder to evaluate the quality of papers upon entering a subject. In the limit that every paper is cited that is merely related to the subject at hand, citation rates for a paper further lose their correlation with quality. Second, it obfuscates the questions of prior art. The huge citation rates tend to obscure the people who first made significant observations.

What to do about it?

Perhaps these reasons are not strong enough to do something about citation inflation. However, if we do want to do something about it, we somehow have to introduce a penalty for over-citing. Two ideas come to mind. The first approach impacts one's career metrics. The count for citations in your paper could be normalized to the number of references you have in the paper. For example, if you reference 50 people in your paper and your paper receives 100 citations, you get a "normalized citation metric" of 2 ($100/50$). Likewise if you reference 50 papers and your paper receives 10 citations you get a citation metric of 0.2 ($10/50$). This actually correlates quite well with the purposes of controlling citations and the identification of original papers. For example, very mature fields and review papers always have more papers that one really must reference. Yet, these are the most likely papers to not have much original thought in them. Therefore, the proposed "normalized citation metric" has additional value beyond stabilizing citation rates.

Another penalty that could be introduced is a "readability penalty". Somehow your paper should become unreadable if you write something like, "And others have worked on this [1-78]." How to accomplish this? One way is to change the style rules of the articles. An effective style rule against such over-citing is that all citations must have author and year in the text itself and then the reference page at the end is in alphabetical order to find the details of the reference. For example, in such a style you would have to write "And others have worked on this (Weinberg 1964; Glashow 1962; Jarlskog & Yndurain 1972;)." If you wanted to write out the citations for all 78 articles your paper would become totally unreadable. Authors are then forced to give up precious in-line reading space only to references that really deserve to be there. Some journals already have this style mandate, but it was formed well before the onset of citation inflation. Perhaps all journals should consider going to it.