How Citation Boosts Promote Scientific Paradigm Shifts and Nobel Prizes

Amin Mazloumian1, Young-Ho Eom2, Dirk Helbing1, Sergi Lozano1, Santo Fortunato2*

1 ETH Zürich, Zürich, Switzerland, 2 Complex Networks and Systems Lagrange Laboratory, Institute for Scientific Interchange (ISI), Torino, Italy

Abstract

Nobel Prizes are commonly seen to be among the most prestigious achievements of our times. Based on mining several million citations, we quantitatively analyze the processes driving paradigm shifts in science. We find that groundbreaking discoveries of Nobel Prize Laureates and other famous scientists are not only acknowledged by many citations of their landmark papers. Surprisingly, they also boost the citation rates of their previous publications. Given that innovations must outcompete the rich-gets-richer effect for scientific citations, it turns out that they can make their way only through citation cascades. A quantitative analysis reveals how and why they happen. Science appears to behave like a self-organized critical system, in which citation cascades of all sizes occur, from continuous scientific progress all the way up to scientific revolutions, which change the way we see our world. Measuring the “boosting effect” of landmark papers, our analysis reveals how new ideas and new players can make their way and finally triumph in a world dominated by established paradigms. The underlying “boost factor” is also useful to discover scientific breakthroughs and talents much earlier than through classical citation analysis, which by now has become a widespread method to measure scientific excellence, influencing scientific careers and the distribution of research funds. Our findings reveal patterns of collective social behavior, which are also interesting from an attention economics perspective. Understanding the origin of scientific authority may therefore ultimately help to explain how social influence comes about and why the value of goods depends so strongly on the attention they attract.


Editor: Yami Moreno, University of Zaragoza, Spain
Received January 5, 2011; Accepted March 14, 2011; Published May 4, 2011

Copyright: © 2011 Mazloumian et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: A.M., S.L. and D.H. were partially supported by the Future and Emerging Technologies programme FP7-COSI-ICT of the European Commission through the project QLectives (grant no.: 231200). Y.-H. E and S.F. gratefully acknowledge ICTeCollective, grant 238597 of the European Commission. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding was received for this study.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: fortunato@isi.it

Introduction

Ground-breaking papers are extreme events [1] in science. They can transform the way in which researchers do science in terms of the subjects they choose, the methods they use, and the way they present their results. The related spreading of ideas has been described as an epidemic percolation process in a social network [2]. However, the impact of most innovations is limited. There are only a few ideas, which gain attention all over the world and across disciplinary boundaries [3]. Typical examples are elementary particle physics, the theory of evolution, superconductivity, neural networks, chaos theory, systems biology, nanoscience, or network theory.

It is still a puzzle, however, how a new idea and its proponent can be successful, given that they must beat the rich-gets-richer dynamics of already established ideas and scientists. According to the Matthew effect [4–7], famous scientists receive an amount of credit that may sometimes appear disproportionate to their actual contributions, to the detriment of younger or less known scholars. This implies a great authority of a small number of scientists, which is reflected by the big attention received by their work and ideas, and of the scholars working with them [8].

Therefore, how can a previously unknown scientist establish at all a high scientific reputation and authority, if those who get a lot of citations receive even more over time? Here we shed light on this puzzle. The following results for 124 Nobel Prize Laureates in chemistry, economics, medicine and physics suggest that innovators can gain reputation and innovations can successfully spread, mainly because a scientist’s body of work overall enjoys a greater

Results

We evaluated data for 124 Nobel Prize Laureates that were awarded in the last two decades (1990–2009), which include an impressive number of about 2 million citations. For all of them and other internationally established experts as well, we find peaks in the changes of their citation rates (Figs. 2 and 3). Moreover, it is always possible to attribute to these peaks
Similarly, the citation rate (see Materials and Methods), which reduces the influence of random variations in the citation paper. We found that the occurrence of this boosting effect is much higher citation rate. Thus, a sizeable part of previous scientific work has reached a big impact after the publication of the landmark paper in 1989, a number of other papers also enjoyed a literally exploded in the first years after its appearance. However, indicates that the number of citations of the landmark paper has landmark paper (blue solid line) and by six older papers. The diagram shows typical plots of the boost factors of four fields: Medicine, Physics, Chemistry and Economy. After discarding those with no citations, we ended up with 1361 scientists. In Fig. 6 we draw on a bidimensional plane each scientist of our random sample (empty circles), together with the Nobel Prize Laureates considered (full circles). The two dimensions are the value of the boost factor and the average number of citations of a scientist. A cluster analysis separates the populations in the proportions of 79% to 21%. The separation is significant but there is an overlap of the two datasets, mainly because of two reasons. First, by picking a large number of scientists at random, as we did, there is a finite probability to choose also outstanding scholars. We have verified that this is the case. Therefore, some of the empty circles deserve to sit on the top-right part of the diagram, like many Nobel Prize Laureates.

To gain a better understanding of our findings, Figs. 4 and 5 present a statistical analysis of the boosts observed for Nobel Prize Laureates. Figure 4 demonstrates that pronounced peaks are indeed related to highly cited papers. Furthermore, Fig. 5 analyzes the size distribution of peaks. The distribution looks like a power law for all choices of the parameters w and k (at least within the relevant range of small values). This suggests that the bursts are produced by citation cascades as they would occur in a self-organized critical system [22]. In fact, power laws were found to result from human interactions also in other contexts [23–25].

The mechanism underlying citation cascades is the discovery of new ideas, which colleagues refer to in the references of their papers. Moreover, according to the rich-get-richer effect, successful papers are more often cited, also to raise their own success. Innovations may even cause scientists to change their research direction or approach. Apparently, such feedback effects can create citation cascades, which are ultimately triggered by landmark papers.

Finally, it is important to check whether the boost factor is able to distinguish exceptional scientists from average ones. Since any criteria used to define “normal scientists” may be questioned, we have assembled a set of scientists taken at random. Scientists were chosen among those who published at least one paper in the year 2000. We selected 400 names for each of four fields: Medicine, Physics, Chemistry and Economy. After discarding those with no citations, we ended up with 1361 scientists. In Fig. 6 we draw on a bidimensional plane each scientist of our random sample (empty circles), together with the Nobel Prize Laureates considered (full circles). The two dimensions are the value of the boost factor and the average number of citations of a scientist. A cluster analysis separates the populations in the proportions of 79% to 21%. The separation is significant but there is an overlap of the two datasets, mainly because of two reasons. First, by picking a large number of scientists at random, as we did, there is a finite probability to choose also outstanding scholars. We have verified that this is the case. Therefore, some of the empty circles deserve to sit on the top-right part of the diagram, like many Nobel Prize Laureates.
Figure 2. Typical time evolutions of the boost factor. Temporal dependence of $R_w'(t)$ for Nobel Laureates [here for (a) Mario R. Capecchi (Medicine, 2007), (b) John C. Mather (Physics, 2006), (c) Roger Y. Tsien (Chemistry, 2008) and (d) Roger B. Myerson (Economics, 2007)]. Sharp peaks indicate citation boosts in favor of older papers, triggered by the publication and recognition of a landmark paper. Insets: The peaks even persist (though somewhat smaller), if in the determination of the citation counts $C_p(t)$, the landmark paper is skipped (which is defined as the paper that produces the largest reduction in the peak size, when excluded from the computation of the boost factor). We conclude that the observed citation boosts are mostly due to a collective effect involving several publications rather than due to the high citation rate of the landmark paper itself.

doi:10.1371/journal.pone.0018975.g002

Figure 3. Dynamics of the boost factor $R_w'(t)$ versus traditional citation variables. Each panel displays the time histories of four variables: the boost factor $R_w'(t)$, the average number of citations per paper $<c(t)>$, the cumulative number of citations $C(t)$, and the $h$-index earned until year $t$ [21]. The panels refer to the same Nobel Laureates as displayed in Fig. 2. The classical indices have relatively smooth profiles, i.e. they are not very sensitive to extreme events in the life of a scientist like the publication of landmark papers. An advantage of the boost factor is that its peaks allow one to identify scientific breakthroughs earlier.

doi:10.1371/journal.pone.0018975.g003
The second reason is that we are considering scholars from different disciplines, which generally have different citation frequencies. This affects particularly the average number of citations of a scientist, but also the value of the boost factor. In this way, the position in the diagram is affected by the specific research topic, and the distribution of the points in the diagram of Fig. 6 is a superposition of field-specific distributions. Nevertheless, the two datasets, though overlapping, are clearly distinct. Adding further dimensions could considerably improve the result. In this respect, the boost factor can be used together with other measures to better specify the performance of scientists.

Discussion

In summary, groundbreaking scientific papers have a boosting effect on previous publications of their authors, bringing them to the attention of the scientific community and establishing their “authority”. We have provided the first quantitative characterization of this phenomenon by introducing a new variable, the “boost factor”, which is sensitive to sudden changes in the citation rates. The fact that landmark papers trigger the collective discovery of older papers amplifies their impact and tends to generate pronounced spikes long before the paper receives full recognition. The boosting factor can therefore serve to discover new breakthroughs and talents more quickly than classical citation indices. It may also help to assemble good research teams, which have a pivotal role in modern science [27–29].

The power law behavior observed in the distribution of peak sizes suggests that science progresses through phase transitions [30] with citation avalanches on all scales—from small cascades to the collective discovery of older papers.
Figure 6. Two-dimensional representation of our collection of Nobel Prize Laureates and a set of 1361 scientists, which were randomly selected. On the x-axis we report the average number of citations of a scientist, on the y-axis his/her boost factor. It can be seen that, on average, Nobel Prize winners clearly perform better. However a Nobel Prize is not solely determined by the average number of citations and the boost factor, but also by further factors. These may be the degree of innovation or quality, which are hard to quantify.

Citation Boosts and Nobel Prizes

The long history of research on these subjects, such phenomena are still not fully understood. There is evidence, however, that the power of a person or the value of a company increase with the level of attention they enjoy. Consequently, our study of scientific impact is likely to shed new light on these scientific puzzles as well.

Materials and Methods

The basic goal is to improve the signal-to-noise ratio in the citation rates, in order to detect sudden changes in them. An effective method to reduce the influence of papers with largely fluctuating citation rates is to weight highly cited papers more. This can be achieved by raising the number of cites to the power k, where k>1. Therefore, our formula to compute $R'_w(t)$ looks as follows:

$$R'_w(t) = \frac{\sum_p \sum_{i=0}^{t+w} (c_{p,i})^k}{\sum_p \sum_{i=1}^{t-w+1} (c_{p,i})^k}.$$  

Here, $c_{p,i}$ is the number of cites received by paper p in year $i$.

The sum over $p$ includes all papers published before the year $t$; $w$ is the time window selected to compute the boosting effect. For $k=1$ we recover the original definition of $R_w(t)$ (see main text).

For the analysis presented in the paper we have used $k=4$ and $w=5$, but our conclusions are not very sensitive to the choice of smaller values of $k$ and $w$.

Acknowledgments

We acknowledge the use of ISI Web of Science data of Thomson Reuters for our citation analysis.

Author Contributions

Conceived and designed the experiments: AM YHE DH SL SF. Performed the experiments: AM YHE SL. Analyzed the data: AM YHE SL. Wrote the paper: SF DH.

References