



Universiteit
Leiden
The Netherlands

Measuring Scholarly Discourse Change with Respect to Citations - A Nobel Prize Case Study

Cox J.; Kohler C.; Scerri A.; Harper C.; Groth P.; Daniel R. Jr.

Citation

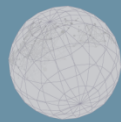
Measuring Scholarly Discourse Change with Respect to Citations - A Nobel Prize Case Study. (2018). Measuring Scholarly Discourse Change with Respect to Citations - A Nobel Prize Case Study. *Sti 2018 Conference Proceedings*, 922-924. Retrieved from <https://hdl.handle.net/1887/65351>

Version: Not Applicable (or Unknown)

License:

Downloaded from: <https://hdl.handle.net/1887/65351>

Note: To cite this publication please use the final published version (if applicable).



STI 2018 Leiden

23rd International Conference on Science and Technology Indicators
"Science, Technology and Innovation Indicators in Transition"

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at <https://hdl.handle.net/1887/64521>

ISBN: 978-90-9031204-0

© of the text: the authors

© 2018 Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands



This ARTICLE is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Measuring Scholarly Discourse Change with Respect to Citations – A Nobel Prize Case Study

Jessica Cox, Curt Kohler, Anthony Scerri, Corey Harper, Paul Groth, Ron Daniel, Jr.

{j.cox; c.kohler; a.scerri; c.harper; p.groth; r.daniel}@elsevier.com

Elsevier Labs, 1600 John F. Kennedy Boulevard, Suite 1800, Philadelphia, PA 19103-2822

Introduction

Understanding how discourse around a scientific publication changes over time may help us understand the way knowledge diffuses and help us better characterize and detect research fronts (Chen & Lobo, 2006). To do so, we need methods to determine the change of discourse over time, with respect to the overall citation structure of the published literature. These methods need to scale to accommodate the breadth of the scientific literature. Recent work has pointed to the use of citation context - the text around a citation - as a mechanism for better understanding the discourse around the cited publication (Jha, Jbara, Qazvinian, Radev 2017). In this work, we perform analyses that use citation context to understand temporal change in discourse around publications.

We investigate the change in discourse around papers that act as a proxy for the discoveries that won their author the Nobel Prize. These papers provide a useful starting point for exploration of this method as they have a definitive time point that could correspond to change in discourse. They have also been previously used for time series investigations within scientometrics (Bjork, Offer, Söderberg, 2014).

Our hypothesis is citation behavior and terminology used will differ before and after the Nobel prize is awarded. We expect to see differences in terminology around the citation, differences in sections where the citation occurs and diversity in subject area to change after the prize is awarded.

Data

We use a combination of datasets to obtain the foundation for our study. We started with a dataset of papers that are a proxy for winning the Nobel Prize from 1995-2017¹. The dataset includes prize discipline, the reason for prize, the Scopus identifier of referenced publications for the particular prize, as well as the year of the prize. This database contained 160 papers. We used Scopus (a comprehensive citation database) to collect all bibliographic information about the roughly 335 thousand papers that cited one or more of the Nobel prize papers. We first reduced this list to the 39,096 papers published by Elsevier for which we have the full text in XML format. From those papers, we extracted 75,432 sentences that contained a citation to one or more of the Nobel papers. In addition to the citing sentence, we extracted

¹ We use papers published after 1995 because our archive of full-text articles extends that far.

the heading for the section that contained it. The data is available at (Cox, Kohler, Groth 2018)

Method

We used the above dataset as input to our analysis. We studied discourse change at the disciplinary level. We aggregated citations contexts in the following disciplines as shown in Table 1.

Table 1. Citation contexts per discipline.

Chemistry	20,411
Physics	20,358
Medicine	20,856
Economics	13,771

For each discipline, we then aggregated citation contexts from papers before and after prize was award to the reference paper. Using these temporal aggregations, we computed the term frequency based on bi- and tri-grams of the contexts to characterize the discourse. Additionally, we computed in which sections of papers the citations occurred again to see if there was a shift in where authors discussed the target papers. We also used a curated vocabulary of innovation terms (e.g. seminal, ground-breaking, discovery) as another approach to detect discourse change.

Initial Results

Our first hypothesis is that the number of citations to the papers will increase after the awarding of the prize. Figure 1 shows the distribution of citation contexts before and after the award. The raw counts did increase. We need to normalize these against the number of years for each time period before the counts are truly comparable as expected.

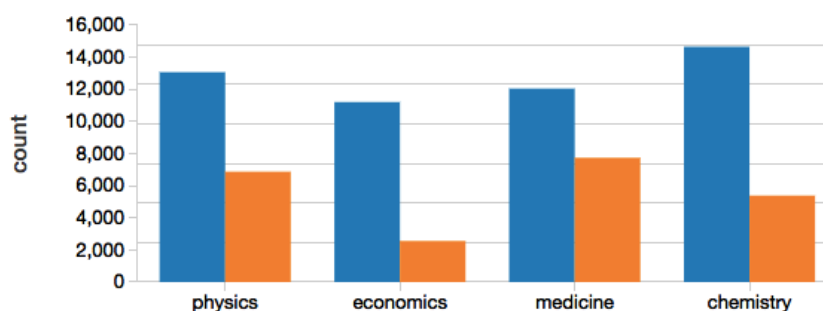


Figure 1: Counts of citation contexts before (in orange) and after (in blue) a paper winning the Nobel Prize.

Our second and third hypotheses were that we would see differences in the frequencies of words within the context. We did not see large vocabulary shifts using word frequency as the domain language (e.g. words such as graphene in physics) dominated the vocabulary. However, using the specific innovation vocabulary, we did see some indicators of shifts in discourse. For example, in physics, medicine and chemistry, the word "discovery" increased by a few percentage points. Medicine also saw increases in the use of "development" and "standard". Economics saw small increases in mentions of "model", "seminal" and "standard".

Our fourth hypothesis was that the citations would move to different parts of the paper. The variability in section headings made the analysis difficult and we have not yet obtained good enough aggregated sections headings to provide descriptions by discipline. However, looking at individual prize-winning papers we did see such shifts. For example, for Yamanaka's stem cell paper, the number of citations appearing in the introduction of the citing paper increased 10% after the award of the prize, while those appearing in the main text of the paper decreased by 6%.

Conclusion

Overall, these initial results show the applicability of temporal analysis of citation contexts to studying the change in scholarly discourse. We aim to broaden our work on this cases study with better aggregations and analysis. We also see the need to expand temporal analysis to include finer grain temporal slicing and to cover the entire literature. We also intend to compare the behavior towards the Nobel papers with a representative sample of peer papers.

References

- Bjork, S. & Offer, A. & Söderberg, G. (2014). Time series citation data: The Nobel Prize in economics. *Scientometrics* 98(1). 10.1007/s11192-013-0989-5.
- Chen, C., & Lobo, N. (2006). Analyzing and Visualizing the Dynamics of Scientific Frontiers and Knowledge Diffusion. *Encyclopedia of Human-Computer Interaction*, 24-30.
- Cox, J. & Kohler, C & Groth, P (2018). Citation Contexts for Nobel Prize Wining Papers. Mendeley Data, v1. <http://dx.doi.org/10.17632/g75gcpp49k.1>
- Jha, R. & Jbara, A. & Qazvinian, V., & Radev, D. (2017). NLP-driven citation analysis for scientometrics. *Natural Language Engineering*, 23(1), 93-130. 10.1017/S135132