January 31, 2017

A Scholarly Paper Index With Strong Truth-Seeking Incentives

Alexandros Salazar Cardozo

A SCHOLARLY PAPER INDEX WITH STRONG TRUTH-SEEKING INCENTIVES

ABSTRACT

A system and method are disclosed to automatically estimate the quality of individual papers published by a scholar in a web search engine. The index of search engine consists of a graph model which has three main elements - papers, scholars and publications. Paper edges such as cites-presumes true, replicates, fails-to-replicate and meta-analyzes are the features where most of the power resides. This graph-dependent method provides truth-seeking incentives for scholars to obtain higher or lower rating for their published papers based on the information estimated from the paper edges. The credibility metrics are preserved and updated live, as they propagate through the network. Scores are given to individual scholars, papers, and publications based on their credibility metrics. Journalists and laypeople may have an easy, minimum-expertise way of assessing the importance of a particular scholar or a publication.

BACKGROUND

Replication of scientific results is often difficult or problematic. Occasionally, flawed results are used as the foundation of research without being properly replicated and confirmed, leading to wasted effort and institutional inertia. Factors contributing to this include little incentive to replicate results, absence of a method to immediately know whether a result has been replicated, previous results or replication failure entirely being ignored by subsequent replication attempts. Beyond replication, reviewing publications is also a time-consuming effort that is done on a volunteer-basis with a limited compensation or prestige attached, despite how crucial it is to the proper functioning of science. Currently, the best proxy of all the work is the number of citations an article receives. However, exciting results that fail to replicate can still have high citations due to strong initial interest, which makes this metric highly unreliable. Existing methods do not provide an automated way to gauge the

quality of individual papers with truth, importance etc. Finally, a few sites list preprints for discussion and validation, but papers there are often seized on by media outlets which is even worse, since they are explicitly not vetted. Thus there is need for a better method to estimate the quality of individual papers with truth seeking incentives effectively.

DESCRIPTION

This disclosure presents a system and a method to automatically estimate the quality of individual papers published by a scholar in a web search engine. The index of search engine consist of a graph model (depicted in FIG. 1) has three main elements - papers, scholars and publications. The graphs semantics can be derived from the edges. For instance, scholar edges of paper can be of two types - author and reviewer. The paper edges of publication can be of one type as to where the paper is published. Paper-paper edges are where most of the power resides. They can be different types of paper edges such as cites - presumes true, replicates, fails-to-replicate and meta-analyzes. This graph structure can provide truth-seeking incentives for scholars effectively.

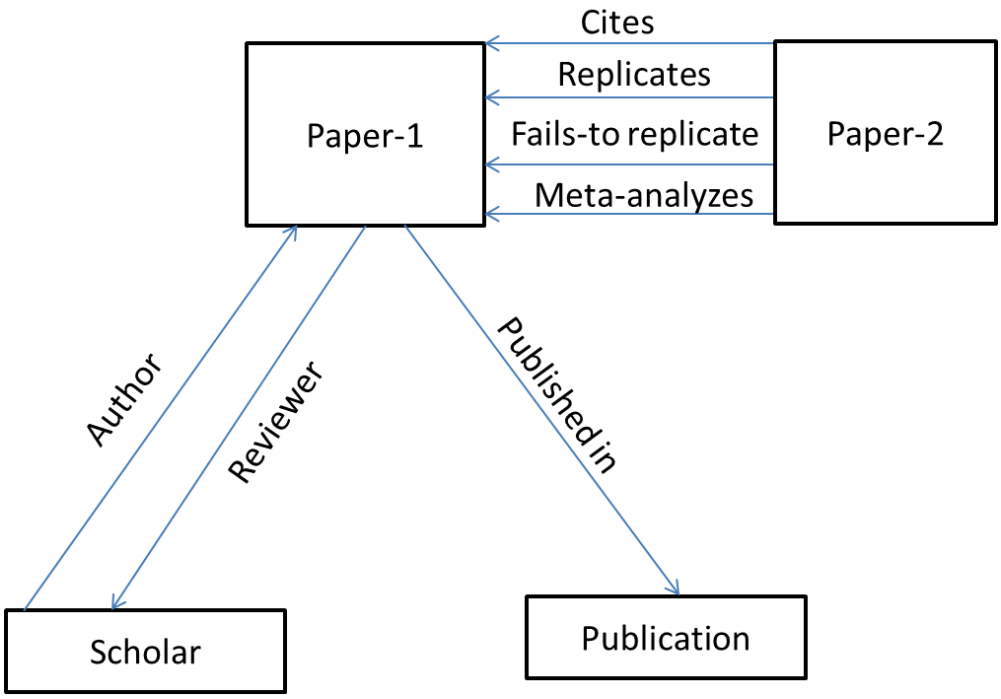

FIG. 1: Graphical model of a scholarly paper index

A truth-incentive method illustrated in FIG. 2 can be implemented from the basic data obtained using the graphical model of FIG. 1. The scholar can get higher or lower rating for their published papers based on the information estimated from the paper edges. A well-designed implementation can answer questions as follows:

a. What papers are important or presumed true by a lot of other papers but not actually replicated?

b. How credible should a paper be considered?

c. What other papers by the same author are cast in doubt if a given paper fails to replicate?

d. What scholars approve papers that get replicated or scholars with an eye for quality?

e. What scholars approve papers that often do not get replicated or lack review?

f. What journals publish oft-replicated papers?

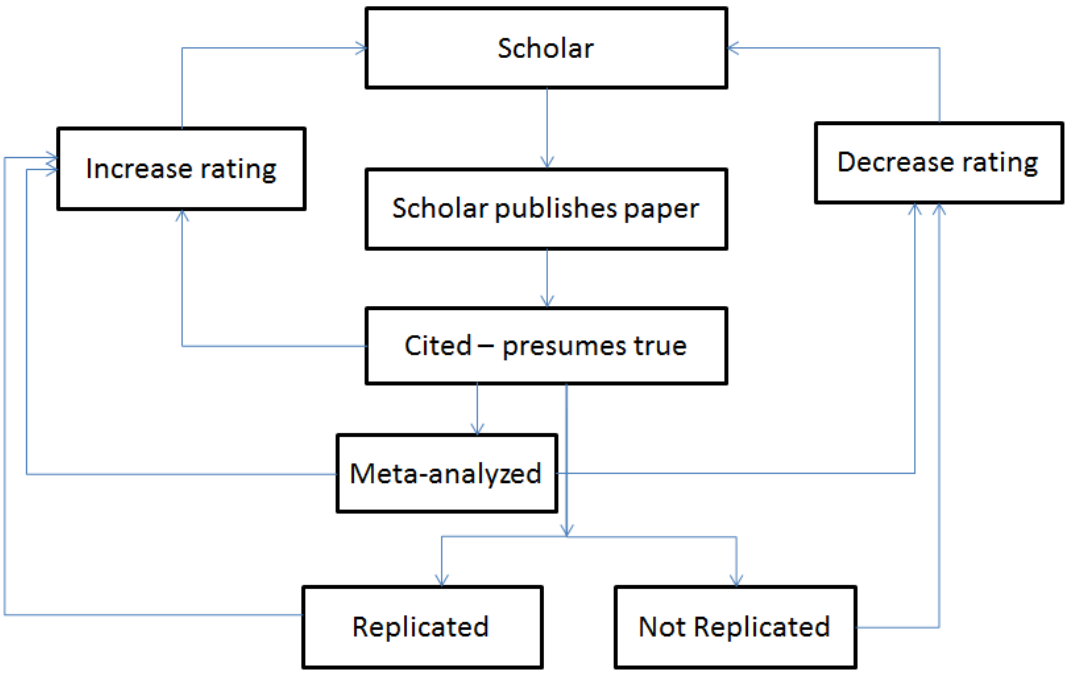

FIG. 2: Truth-incentive method for scholars

Credibility metrics can be preserved and updated live, as they propagate through the network. Events that can impact credibility can include a scholar who was working on an

unpopular interpretation and is vindicated. The quality of their work should be reevaluated accordingly, across all papers published. If a scholar is found to have committed fraud, the quality of the work should be put in question accordingly. Well-replicated studies can be shielded by impact of later fraud by their replication. This would also help to distinguish sloppy science from outright fraud. For example, if a scholar turns in a bad paper, but has a long history of creating well-replicated results, the presumption may be that it was an error.

Scores can then be given to individual scholars, papers, and publications based on their credibility metrics as follows:

Scholar

- Credibility in publication - should correlate to likelihood that a result is true, given that the scholar was involved.
- Credibility in review - should correlate to likelihood that a result may be replicated, given that the scholar reviewed and approved it.
- Credibility in replication – should correlate to the likelihood that if this scholar failed to replicate a result, further scholars may also fail to replicate it.
- Influence papers – the papers published by the scholar have high influence.
- Foundational value - the papers published by the scholar have high foundational value as defined under “Paper”.

Paper

- Credibility – should correlate to the likelihood the results hold.
- Influence - should correlate to the number of other papers that cite this paper
- Foundational value – should correlate to the number of papers that presume this paper's results to be true and build on them.

- Potential foundational value – should correlate to the number of authors of this paper who have a history of contributing foundational work. This metric is only relevant if the paper has not yet been live for long.

Publisher

- Credibility – should correlate to the likelihood that a paper will not be retracted or fail replication, given that it is published by this publisher.
- Foundational value – should correlate to the number of papers published by this journal that have high foundational value.
- Influence – should correlate to the number of papers published by this journal that have high influence.
- Nurturing – should correlate to the number of scholars published in this journal before they were high scoring on certain metrics who later went on to become high scoring.

The system may run simulations of various gaming strategies to see if scholars or publications can inflate certain key numbers without adhering to the spirit of what those numbers represent. In addition, careful thought must be given to the appropriate UX for the index, as it can have multiple uses. For example, for prior art search, it needs to clearly and elegantly indicate the importance of various papers, as well as their replication status and possible importance. For scholar information, it needs to indicate along what axes a scholar is successful, and along what axes less. For journal information, it needs to indicate what journals are good for submissions by younger scholars, or experienced scholars, etc.

Current approaches are either ad-hoc such as citation counts, citation-based such as citation tracking software which uses the h-index, or informal such as repositories of electronic preprints. The disclosed approach offers a structured way of evaluating research and researchers via properties that are inherent to the research itself but go beyond simple citation.

Alternatively, similar information could be gathered by applying machine learning algorithms to all the articles the users have at their disposal, and create the graph automatically, instead of manually. However, it would then be implemented in slightly different ways at multiple different publications, precluding truly global and meaningful comparisons. For any implementations produced, competitors might try to create a similar system, but change the importance attached to various elements such as replication vs. reviewer quality in evaluating papers.

This graph-dependent approach provides truth-seeking incentives such as review contributions that can be quantified, and qualified, which means they can be appreciated. Changing appraisal of work changes the appraisal of the people and journals who produced it, as is fair; if a result is very popular at first (high influence), but then fails to be replicated, its importance may be reduced, as will that of the scholars that reviewed and the journal that published it. Journalists and laypeople may have an easy, minimum-expertise way of assessing the importance of a particular result at any particular point in time.

Machine learning algorithms can be used on the high-confidence results to drastically improve all search and knowledge services. Sorting on search can be improved on factual questions by returning results that match the more trusted scholarly results, greatly improving the reliability of the service and the trust of the users. Users can start answering questions with more complex answers. For example, when a user searches for "Does sugar cause cancer" the software can return an answer based on scholarly results, and not just hits on click-bait pages. The user can track news stories relevant to papers and get intelligent answers to their questions based on a combination of trusted scholarly results and the newly incorporated scholarly index data. For instance, if a new paper comes out extolling the virtues of cold baths and a user asks "Do cold baths lead to better health", instead of returning links, user may view a card that says "There is no established evidence of that, but a paper came out

yesterday on the subject." Thus this tool may be helpful towards taking the world's *knowledge* and making it universally accessible and useful.

The main users of this service, beyond daily search users, may be scholars, for obvious reasons. Beyond that, this tool may help science journalists avoid being easily fooled by exciting press releases. The tool helps users to better inform themselves and their readers about the likelihood an exciting result is true. The scholars may also write follow-up articles on successful or unsuccessful replications. The system and method would help keep track of results over a long period of time, as well as make decisions based on those results. Moreover, better information about publications and scholars may help agencies, both governmental and non-governmental to decide how to allocate funding effectively.