Response to Comment on "Quantifying long-term scientific impact"
Dashun Wang et al.
*Science* 345, 149 (2014);
DOI: 10.1126/science.1248961

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of July 13, 2014):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/345/6193/149.3.full.html

A list of selected additional articles on the Science Web sites related to this article can be found at:
http://www.sciencemag.org/content/345/6193/149.3.full.html#related

This article cites 2 articles, 2 of which can be accessed free:
http://www.sciencemag.org/content/345/6193/149.3.full.html#ref-list-1

This article appears in the following subject collections:
Technical Comments
http://www.sciencemag.org/cgi/collection/tech_comment
Response to Comment on “Quantifying long-term scientific impact”

Dashun Wang,1 Chaoming Song,2 Hua-Wei Shen,3,4 Albert-László Barabási3,5

Wang, Mei, and Hicks claim that they observed large mean prediction errors when using our model. We find that their claims are a simple consequence of overfitting, which can be avoided by standard regularization methods. Here, we show that our model provides an effective means to identify papers that may be subject to overfitting, and the model, with or without prior treatment, outperforms the proposed naïve approach.

Response 2: This is the result of a few papers with large errors, caused by the overfitting mentioned in response 1. The WSB model offers a way to detect papers that may be subject to overfitting. Upon filtering these papers, the large errors reported in (1) disappear and the WSB model outperforms the naïve approach.

To be specific, the analysis by Wang, Mei, and Hicks documents an interesting practice of evaluating citation predictions. Although the WSB model makes probabilistic predictions on future citations for a given paper, represented by a citation envelope, (1) instead uses MAPE (mean prediction errors) to evaluate predictive power, despite the well-known fact that citations follow a fat-tailed distribution. Indeed, “outliers” can dominate mean prediction errors. For example, when overfitting occurs, the predicted citation counts can be as large as 10^100. Therefore, a few papers with large errors skew the large prediction errors reported in (1) (Fig. 1A, inset). The existence of such erroneous behavior raises an important question: When can we trust citations predicted by the WSB model? The answer lies in the citation envelope that we proposed in (2). Indeed, the WSB model not only predicts the future citations but, equally important, it provides the citation envelope, capturing the uncertainties of the predictions. Denoting with σ2 the variance of its predictions (size of the envelope), we computed the variance relative to its most likely citations, that is, σ2/σ2med. When this ratio is small—that is, the uncertainty of the most likely citations is small—one expects the average prediction error to be small, indicative of high accuracy in the predicted outcome. We compared the predicted and the real citations for cases where σ2/σ2med ≤ 1, accounting for 87.2% of the papers (Fig. 1B). That is, we imposed a cutoff where predicted citations are comparable to their uncertainties, which represents an indication of potential overfitting. Comparing Fig. 1B with the inset of Fig. 1A, we found that the overfitting cases are effectively filtered out and that the WSB model clearly outperforms the naïve approach (MAPE 0.25 versus 0.31). Therefore, when it is confident, the WSB model predicts accurately the most likely citation counts. Alternatively, when overfitting occurs, the WSB model through its citation envelope indicates that it is not confident, providing guidance to treat these predicted citations with care.

Claim 3: The performance of the WSB model with prior is still worse than the naïve approach.

Response 3: Wang, Mei, and Hicks’s experiments were based on an incorrect set of prior parameters. The WSB model with prior consistently outperforms the naïve approach.

We repeated the analysis on the same corpus as the one used by Wang, Mei, and Hicks (the same as in Response 1). We find that for both 5 and 10 years training periods, the WSB model with prior also yield much lower MAPE (0.204 versus 0.305 for 10 years and 0.396 versus 0.504 for 5 years). We also plotted the quantities shown in the figure 1 of (1), finding substantial differences in both parameter correlations (Fig. 1A) and predicted citations (Fig. 1C). Wang, Mei, and Hicks also used other evaluation metrics, such as Spearman correlation and percentage of correctly identified top 10% papers. We therefore evaluated these two metrics as well, finding that the WSB model with prior consistently outperforms the naïve approach for both metrics using 5 and 10 years of training. We repeated our analysis on other sets of papers as well, such as papers published in the 1970s, finding consistent results. The results in (1) were based on an incorrect set of prior parameters. Indeed, they applied priors learned on review papers in Reviews of Modern Physics to predict research papers published by Physical Review. It is reasonable to assume that the WSB model with prior outperforms the naïve approach for a wide range of prior parameters, as long as they are within a reasonable range, because the naïve approach reduces to a special case of the WSB model with prior with a trivial prior parameter β → 0 (4).

Taken together, the large prediction errors reported by Wang, Mei, and Hicks are a consequence of overfitting, which can be avoided using standard regularization methods. They misinterpreted the true message of the paper. In our view, the main message of (2) lies in the uncovered regularity of a complex evolving system that previously was perceived as noisy and unpredictable. Indeed, the proposed model is a minimal citation model that captures all quantifiable mechanisms known to date to affect citation histories. Hence, it represents a fundamental building block on which one needs to add industrial-quality implementation protocols, which could lead to a more robust and accurate citation prediction tool. Therefore, judging our results on the quality of the implementation is like judging the laws of thermodynamics on the performance of the cars a particular company can build.

---

Dashun Wang,1 Chaoming Song,2 Hua-Wei Shen,3,4 Albert-László Barabási3,5

1IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA. 2Department of Physics, University of Miami, Coral Gables, FL 33146, USA. 3Center for Complex Network Research and Departments of Physics, Computer Science and Biology, Northeastern University, Boston, MA 02115, USA. 4Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China. *Corresponding author. E-mail: alb@neu.edu
Fig. 1. (A) Correlation between parameters $\lambda$ and $\mu$ using the WSB model with prior, indicating that applying a prior resolves the issue of overfitting completely. The $\mu$ parameters obtained by Wang, Mei, and Hicks are smaller than ours. We suspect that this is due to unit conversion. That is, they might have used years instead of days as a time unit. (Inset) Same correlation by using the WSB model, documenting the existence of outliers also observed in (1) due to overfitting. (B) Comparison between the predicted and the real citations for the WSB model in cases where $s_{\text{pred}} < 1$, accounting for 87.2% of the papers. The lack of outliers demonstrates that the citation envelope provides an effective means to filter out cases that are affected by overfitting. (C) Comparison between predicted and real citations for the WSB model with prior. Comparing (B) and (C), we find that applying prior improves the predictive power of the WSB model. (D) Comparison between the predicted and the real citations for the naïve approach, documenting that it systematically underestimates the future citations. We used 10 years of training for results in (A) to (D). All conclusions continue to hold if we use 5 years of training.