

TECHNICAL COMMENT

SCIENCE COMMUNICATION

Comment on “Quantifying long-term scientific impact”

Jian Wang,^{1,2*} Yajun Mei,³ Diana Hicks⁴

Wang *et al.* (Reports, 4 October 2013, p. 127) claimed high prediction power for their model of citation dynamics. We replicate their analysis but find discouraging results: 14.75% papers are estimated with unreasonably large μ (>5) and λ (>10) and correspondingly enormous prediction errors. The prediction power is even worse than simply using short-term citations to approximate long-term citations.

Wang *et al.* (1) proposed an elegant model for citation dynamics and have quickly drawn attention, including a feature in *Nature News* (2), because high prediction power was reported: “With $T_{\text{Train}} = 5$, only 6.5% of the papers left the prediction envelope 30 years later...” Unfortunately, this is an odd way of evaluating the prediction power because the width of the prediction envelope is ignored. Figure 4A in (1) shows that the predicted number of citations in year 30 for the middle paper is between 100 and more than 300; such a wide range can hardly be said to constitute a “good” prediction, even if no paper left this envelope.

To examine the prediction power of their model (hereafter, WSB), we replicated their analysis using 5-year citation history to predict future citations. We used 1973 *Physical Review* papers published in 1980 that acquired at least 10 citations in the first 5 years. We first tried a Newton method to solve Eqs. S25 and S26 in the supplementary materials for (1), but results were sensitive to initial choices of the parameters. Thus, we adjusted the optimization method and maximized their objective function (Eq. S22) directly. Plugging Eq. S25 into Eq. S22 yields an objective function only depending on μ and σ . Then, we simply search within a region of μ and σ to find the maximization solution. One important task is to specify the searching region, and the trickiest part is the upper boundary for μ . As the impact time, $T_i^* \approx \exp(\mu)$, “represents the characteristic time it takes for a paper to collect the bulk of its citations,” we feel safe to assume $\mu \leq 5$ ($T_i^* \leq 148$). Therefore, our final searching region is $\mu \in [-1.00, 5.00]$ and $\sigma \in [0.10, 10.00]$, with a grid of 0.01.

Our results show that no papers hit the boundaries of σ or the lower boundary of μ , but 14.75%

of the papers hit the upper boundary of μ , indicating that their best-fitting μ might be even larger than 5. Note that large μ^* is accompanied by large λ^* (Fig. 1A), and the ultimate impact, $C^\infty = m(e^{\lambda^*} - 1)$, “represents the total number of citations a paper acquires during its lifetime.” If $\lambda > 10$, then $C^\infty > 660,764$, which is very unrealistic. When $\mu^* = 5$, all papers except one have λ^* larger than 10, and their citations are enormously overestimated (Fig. 1B).

One possible explanation for the inclination to pick large μ^* and λ^* is the limitation of using only 5-year citation history. WSB exhibits an S-shaped curve, with acceleration followed by deceleration in the citation accumulation process. Many papers are still in the initial acceleration stage in the fifth year. Therefore, model estimation based on this short period may mistakenly pick a very large μ^* and predict exponential growth all the way out to 148 years.

How good is the prediction? We use a naïve approach as a benchmark: $C^{30} = C^5$, i.e., assuming that papers receive no further citations after the training period. We compare the true and predicted number of citations in year 30 (C^{30}) and report three evaluation statistics: (i) mean absolute percentage error (MAPE), (ii) Spearman correlation, and (iii) percentage of correctly identified top 10% highly cited papers (3). We report these statistics on two sets of papers: all papers and papers with $\mu^* < 5$, i.e., excluding “misbehaving” papers unsuitable for WSB (Table 1). Even after excluding misbehaving papers, MAPE of WSB is 1.98×10^{57} , Spearman correlation is 0.58, and only 31.95% of the top papers are correctly identified. The naïve approach outperforms WSB, with a much lower MAPE (0.56), a higher Spearman correlation (0.74), and a higher percentage of correctly identified top papers (58.29%).

If we use 10-year citation history to predict C^{30} , the percentage of misbehaving papers decreases to 1.57%. After excluding misbehaving papers, MAPE decreases to 0.38, and the Spearman correlation and the percentage of correctly identified top papers increase to 0.90 and 71.28%. However, the naïve approach (i.e., $C^{30} = C^{10}$) still outperforms WSB.

In response to an earlier version of this Comment, the authors attributed our observed poor performance of WSB to overfitting and claimed that it can be addressed by a conjugate prior method proposed in their latest technique report (hereafter, WSB-with-prior) (4). Indeed, fitting WSB on N papers involves estimating $3N$ parameters (each paper with three individual parameters: λ , μ , and σ). The WSB-with-prior method incorporates a conjugate prior on λ —i.e., the N individual λ 's follow a gamma distribution, $\Gamma(\alpha, \beta)$, thereby reducing the number of estimated parameters from $3N$ to $2N + 2$: one α , one β , N μ 's, and N σ 's. This overfitting issue and the conjugate prior method were not mentioned in Wang *et al.* (1). Nevertheless, we implemented WSB-with-prior. Unfortunately, its performance is still inadequate.

We first fit WSB-with-prior using the 5-year citation history, but it failed to find finite values of α and β for optimal solutions. This reveals another issue of their methods (with or without prior): The maximum likelihood estimator does not always exist when the parameter space is not compact. Even if we adopt the four sets of α and β values reported by Shen *et al.* (4) and report their best evaluation statistics, the naïve approach still outperforms (Table 1).

With $T_{\text{Train}} = 10$, we found a finite solution: $\alpha^* = 4.04$ and $\beta^* = 3.83$. Compared with the naïve approach, WSB has a smaller MAPE (0.27 versus 0.34), a slightly higher Spearman correlation (0.9100 versus 0.9052), but a lower percentage of correctly identified top papers (74.75% versus 76.41%). Compared with the naïve approach, the improvement of WSB-with-prior is marginal given its model/computational complexity.

Although ever-fancier statistical methods (e.g., regularization) can be incorporated to fit the parameters in WSB, the results will always work poorly on extreme values. Indeed, WSB is inadequate to any type of data that has extreme values. However, scientific outcomes are uncertain and skewed, and there are always high-impact outliers that represent the greatest discoveries and cannot be dismissed (5, 6). Although we find WSB elegant, we do not believe it to be very useful, and it would be going too far to market this model as a prediction tool for policy-making.

REFERENCES

1. D. Wang, C. Song, A.-L. Barabási, *Science* **342**, 127–132 (2013).
2. R. Van Noorden, *Nature News* (3 October 2013).
3. J. Wang, *Scientometrics* **94**, 851–872 (2013).
4. H. Shen, C. Song, D. Wang, A.-L. Barabási, Modeling and predicting popularity dynamics via reinforced Poisson processes. Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, Québec, Canada, 27 to 31 July 2014; preprint available at <http://arxiv.org/abs/1401.0778>.
5. W. Glänzel, B. Schlemmer, B. Thijs, *Scientometrics* **58**, 571–586 (2003).
6. A. van Raan, *Scientometrics* **59**, 467–472 (2004).

ACKNOWLEDGMENTS

Data are retrieved from a bibliometrics database developed by the Competence Center for Bibliometrics for the German Science System (K.B.) and derived from Thomson Reuters Web of Science. K.B. is funded by the German Federal Ministry of Education and Research (BMBF), project no. 01PQ08004A.

20 November 2013; accepted 3 June 2014
10.1126/science.1248770

¹Center for R&D Monitoring (ECOOM) and Department of Managerial Economics, Strategy and Innovation, University of Leuven, Waaistraat 6, Bus 3536, 3000 Leuven, Belgium.

²Institute for Research Information and Quality Assurance (IFQ), Schuetzenstrasse 6a, 10117 Berlin, Germany.

³H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 755 Ferst Drive Northwest, Atlanta, GA 30332–0205, USA. ⁴School of Public Policy, Georgia Institute of Technology, 685 Cherry Street, Atlanta, GA 30332–0345, USA.

*Corresponding author. E-mail: jian.wang@kuleuven.be

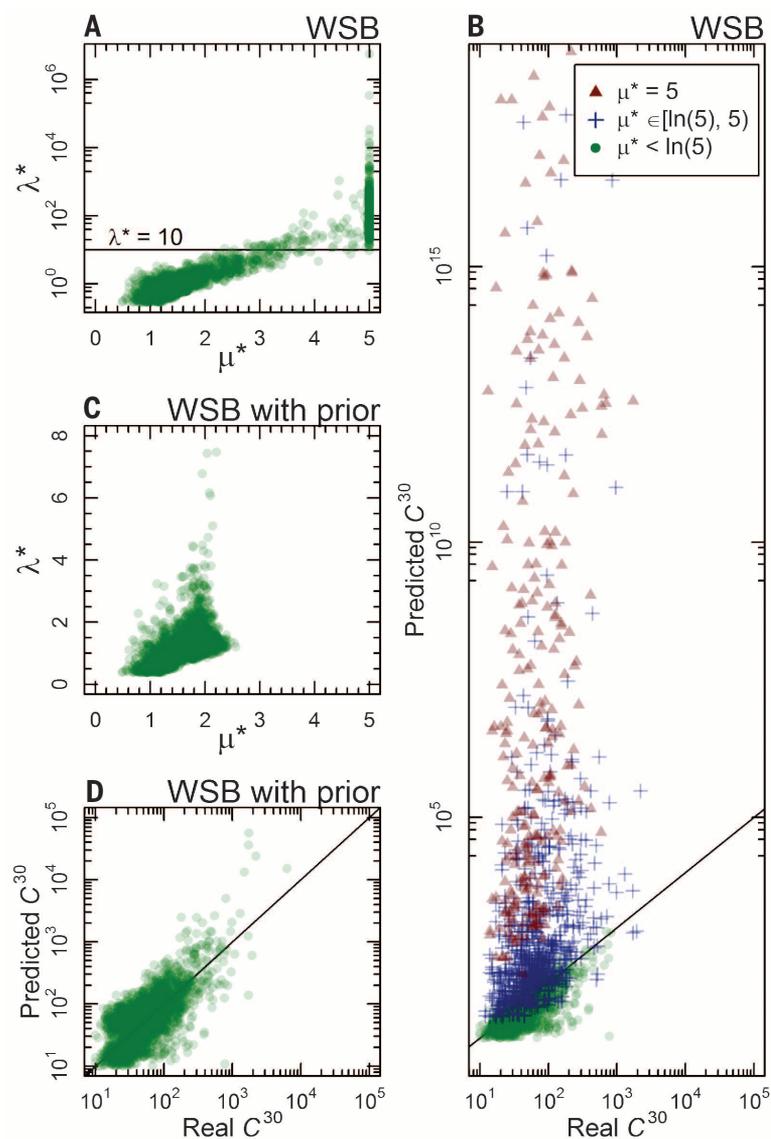


Fig. 1. Prediction power evaluation. $T_{\text{Train}} = 5$. **(A)** Scatterplot of the optimal λ^* and μ^* estimated for WSB. **(B)** Scatterplot of WSB-predicted C^{30} and true C^{30} ; citations are enormously overestimated, with very large μ^* . Many observations are outside the vertical-axis boundary. **(C)** Scatterplot of the optimal λ^* and μ^* estimated for WSB-with-prior. **(D)** Scatterplot of WSB-with-prior predicted C^{30} and true C^{30} . With $T_{\text{Train}} = 5$, WSB-with-prior failed to find finite values of α and β for optimal solutions, so we adopt the four sets of α and β values reported by Shen *et al.* (4). (C) and (D) use $(\alpha = 4.759, \beta = 4.440)$, the one with the smallest MAPE.

Table 1. Prediction power evaluation.

	WSB	Naïve	WSB-with-prior ‡	WSB	Naïve
$T_{Train} = 5$		All papers ($N = 1973$)		$\mu^* < 5$ ($N = 1682$)	
Mean absolute percentage error	$6.71 \times 10^{303\dagger}$	0.56	0.75	1.98×10^{57}	0.54
Spearman correlation	0.51	0.74	0.64	0.58	0.76
Percentage of correctly identified top 10% papers	23.74	58.29	59.60	31.95	57.99
$T_{Train} = 10$		All papers ($N = 1973$)		$\mu^* < 5$ ($N = 1942$)	
Mean absolute percentage error	1.91×10^7	0.34	0.27	0.38	0.34
Spearman correlation	0.90	0.91	0.91	0.90	0.91
Percentage of correctly identified top 10% papers	67.68	76.41	74.75	71.28	75.90

†Conservative estimations. The largest number our software can handle is 1.797693×10^{308} , so numbers larger than this threshold will be treated as $+\infty$. In our calculations, we truncate these larger numbers and only record them as 1.797693×10^{308} , so the actual prediction error is even larger than reported here. ‡With $T_{Train} = 5$, WSB-with-prior failed to find finite values of α and β for optimal solutions, so we adopt the four sets of α and β values reported by Shen *et al.* (4) and report their best evaluation statistics. These four sets of (α, β) are (4.237, 4.061), (4.759, 4.440), (6.130, 4.924), and (10.706, 5.379). Among them, the smallest MAPE (0.75) is yielded by (4.759, 4.440), the highest Spearman correlation (0.64) is yielded by (4.237, 4.061), and the highest percentage of correctly identified top 10% papers (59.60%) is yielded by (6.130, 4.924).

Comment on "Quantifying long-term scientific impact"

Jian Wang, Yajun Mei and Diana Hicks

Science **345** (6193), 149.
DOI: 10.1126/science.1248770

ARTICLE TOOLS	http://science.sciencemag.org/content/345/6193/149.2
RELATED CONTENT	http://science.sciencemag.org/content/sci/345/6193/149.3.full http://science.sciencemag.org/content/sci/342/6154/127.full
REFERENCES	This article cites 5 articles, 1 of which you can access for free http://science.sciencemag.org/content/345/6193/149.2#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2014, American Association for the Advancement of Science