

RESEARCH ARTICLE

Content Volatility of Scientific Topics in Wikipedia: A Cautionary Tale

Adam M. Wilson^{1,2*}, Gene E. Likens^{2,3}

1 Geography Department, University at Buffalo, Buffalo, NY, United States of America, **2** Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut, United States of America, **3** Cary Institute of Ecosystem Studies, Millbrook, New York, United States of America

* adamw@buffalo.edu

Abstract

Wikipedia has quickly become one of the most frequently accessed encyclopedic references, despite the ease with which content can be changed and the potential for ‘edit wars’ surrounding controversial topics. Little is known about how this potential for controversy affects the accuracy and stability of information on scientific topics, especially those with associated political controversy. Here we present an analysis of the Wikipedia edit histories for seven scientific articles and show that topics we consider politically but not scientifically “controversial” (such as evolution and global warming) experience more frequent edits with more words changed per day than pages we consider “noncontroversial” (such as the standard model in physics or heliocentrism). For example, over the period we analyzed, the global warming page was edited on average (geometric mean \pm SD) 1.9 ± 2.7 times resulting in 110.9 ± 10.3 words changed per day, while the standard model in physics was only edited 0.2 ± 1.4 times resulting in 9.4 ± 5.0 words changed per day. The high rate of change observed in these pages makes it difficult for experts to monitor accuracy and contribute time-consuming corrections, to the possible detriment of scientific accuracy. As our society turns to Wikipedia as a primary source of scientific information, it is vital we read it critically and with the understanding that the content is dynamic and vulnerable to vandalism and other shenanigans.



OPEN ACCESS

Citation: Wilson AM, Likens GE (2015) Content Volatility of Scientific Topics in Wikipedia: A Cautionary Tale. PLoS ONE 10(8): e0134454. doi:10.1371/journal.pone.0134454

Editor: Hussein Suleman, University of Cape Town, SOUTH AFRICA

Received: November 12, 2014

Accepted: June 29, 2015

Published: August 14, 2015

Copyright: © 2015 Wilson, Likens. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are available via Figshare (<http://dx.doi.org/10.6084/m9.figshare.1397533>).

Funding: The authors have no support or funding to report.

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

Introduction

Wikipedia.org is now over ten years old and has become the sixth most popular website globally [1]. As of August 2014, over thirty thousand active editors were making three million edits per month to maintain and add topics to the four million articles in the English language version [2]. The Wikipedia infrastructure has made it possible for contributors to assemble quickly the world’s largest and most up-to-date encyclopedia; the project has, in fact, been praised for accuracy that is ‘close’ to traditional reference sources such as the *Encyclopedia Britannica* [3–7]. There is some evidence that as Wikipedia matures, its scientific content increasingly references articles in established scientific journals and that the citation frequency of those journals is in general agreement with citation patterns in the scientific literature [8]. Scientists and

students have also begun to embrace Wikipedia as a resource for public engagement and to improve scientific literacy[9]. For example, Bond argues that because Wikipedia has become such an important reference for so many people (ranging from students to scientists), that ornithologists (and other scientists) should ‘embrace and contribute’ to its content as an outlet for “science education, science communication, and public service” [10]. Despite these encouraging signs, however, there is reason to be concerned about the dynamic information content in Wikipedia.

One of Wikipedia’s fundamental principles is the “bold, revert, discuss” cycle, in which volunteer editors are encouraged to make ‘bold’ edits including adding, removing, rearranging, or rewriting the content in the article [11]. When someone else is monitoring the article, they are welcome to ‘revert’ any edit and return the article to its previous form. Following the reversion, the editors are encouraged to discuss the topic on the “talk” page that accompanies each article. In some ways, this mirrors the scientific process of peer review in which peers must critically evaluate new ideas, with the important exceptions that publication precedes review, there is no outright option for “permanent” rejection, and that the motivation, commitment, and qualifications of Wikipedia’s editors are typically unknown (especially for anonymous edits). To encourage participation, Wikipedia advises, “When in doubt, edit!” and then wait for someone else to revert your change. Neil Waters has written a related discussion of how the “democratization” of access to information is intertwined with the democratization of the information itself [12].

In addition to generating articles of unknown quality, this process can lead to ‘edit-wars’ between editors[13]. The problem is so severe for some articles that Wikipedia has guidelines to manage edit warring, including the three revert rule, which prevents editors from reverting more than three times on the same page in a 24-hour period. In general, though, overt edit warring is thought to be relatively uncommon; Sumi, et. al reported that “less than 1% of articles” are likely to have serious conflict[14]. However, little is known about how this potential for controversy affects the accuracy and stability of content on scientific topics, especially those with associated political controversy.

Following a long-standing research interest and expertise in acid rain [15], we noticed that some corrections we or others made on the acid rain article had been changed by major edits to introduce (or re-introduce) balderdash and factual errors into the content [16]. An illustrative example of tempestuous edits to the English language Wikipedia acid rain entry begins on November 30, 2011 [17]. At 10:20am, an anonymous editor (identified only by an IP address), removed the introductory paragraph which defined acid rain and replaced it with a statement calling acid rain “a load of bullshit.” This change was quickly reverted, but the next day the paragraph was again deleted and replaced by “Acid rain is a popular term referring to the deposition of wet poo and cats.” Five minutes later this edit was reverted and repeated again, and then reverted again. The following day (December 2, 2011) another sentence was changed from “During the 1990s, research continued.” to “During the 1990s, research *on elves* continued [emphasis added],” which remained for over seven hours. Later that day the sentence “Acid Rain [sic] killed bugs bunny” was briefly added. Fifteen minutes later the section title “Chemistry in cloud droplets” was changed to “Blowjobs.”

Hopefully readers accessing the article during those tumultuous edits would recognize the jabberwocky via tone and absurdity. Algorithms that automatically detect and revert malicious edits like those described above are available and are continuously improving [18], but will probably never identify all cases of vandalism. For example, less than a month later, the sentences, “Acid rain does not directly affect human health. The acid in the rainwater is too dilute to have direct adverse effects” were briefly changed to “Acid rain directly affect [sic] human health. The acid in the water is too concentrated to have indirect adverse effects [19].” Similar

edits have been made nearly every day over the history of the page despite the fact that it has typically had “protected” status to prevent edits by anonymous users [20].

After looking through this history and feeling frustrated by attempts at correction that were later changed, we decided to look at the issue more systematically. Given the political controversy that has surrounded the issue of acid rain, in this study we explore whether the frequency and magnitude of edits to the acid rain article are typical of scientific topics in general, or limited to controversial topics [21].

Material & Methods

We compared three topics we consider to be politically (though not scientifically) controversial (acid rain, global warming, and evolution) and four we consider to be politically uncontroversial (heliocentrism, general relativity, continental drift, and the standard model in physics). To quantify the comparison, we downloaded the complete revision history of each article from 2003-06-12 (when the most recent article, Heliocentrism, originated) through 2012-07-31 (the most recent full archive available when this analysis was run) using Wikipedia’s export API (http://www.mediawiki.org/wiki/API:Main_page). The data were obtained in compliance with Wikipedia’s Terms and Conditions of Use (http://wikimediafoundation.org/wiki/Terms_of_Use), which allows sharing and adapting the data with attribution. We then calculated three metrics from each article’s history: 1) daily edit rate (excluding successive edits by the same user, $n = 23,156$), 2) mean edit size (the total number of words inserted, deleted, or changed) on days with at least one edit ($n = 8,525$), and 3) mean number of page ‘views’ per day (which includes requests by computer programs, only available after 2008-01-01). We also calculated two-tailed nonparametric rank-based multiple contrasts to estimate the significance of pairwise differences in edit rates and words changed for each article with no distributional assumptions about the data [22]. The data and code used to run this analysis are available at <http://dx.doi.org/10.6084/m9.figshare.1397533>.

Results & Discussion

The geometric means (\pm SD) ranged from 0.2 ± 1.4 edits per day and 9.4 ± 5.0 words changed per day for the standard model to 1.9 ± 2.7 edits per day and 110.9 ± 10.3 words changed per day for global warming (Table 1). The mean (\pm SD) number of page views ranged from $15\,549 \pm 6\,897$ on the global warming page to $1\,026 \pm 564$ on the heliocentrism page. We found that the edit rate of the acid rain article was less than the edit rate of the global warming ($p < 0.0001$) and evolution ($p < 0.0001$) articles, but significantly more ($\alpha = 0.05$) than each of the ‘noncontroversial’ topics (Table 1). Furthermore, the three ‘controversial’ topics each had greater mean edit rates than each of the ‘noncontroversial’ topics ($p < 0.05$). Similarly, the mean edit size for each ‘controversial’ topic was larger than each of the ‘non-controversial’ topics ($p < 0.05$). While this analysis was intentionally limited in scope and it is difficult to assess causality (the “controversial” pages were also viewed more often, Table 1), edit rates can be much higher for “controversial” scientific topics. This finding is especially troubling in combination with the knowledge that Wikipedia content *about* scientists is not a good proxy for academic notability [23].

So what should be done? In the future, it may be possible to automatically identify and flag pages with significant controversy [24] and quantify user reputation [18], both of which could be made visible to help readers critically evaluate the content of a page. For now, however, these results reinforce the position that Wikipedia should not be used in academic citations without very careful consideration and scrutiny [12]. Wikipedia acknowledges this and reports that, “while some articles are of the highest quality of scholarship, others are admittedly complete rubbish [25].” Furthermore, Wikipedia’s policy on academic use is clear that “Wikipedia

Table 1. Statistics summarizing the view and edit history of selected Wikipedia articles.

Wikipedia Article	Mean Daily Page Views \pm SD ^a	Maximum daily edits ^b	Edits per day geometric mean \pm SD (n) ^{b,c}	Words changed per Day geometric mean \pm SD (n) ^{b,d}
Acid_rain	2 954 \pm 1 310	26	0.5 \pm 2.0 (3307)	36.2 \pm 10.2 (1103)
Global_warming	15 549 \pm 6 897	231	1.9 \pm 2.7 (3307)	110.9 \pm 10.3 (2211)
Evolution	6 260 \pm 2 450	89	1.3 \pm 2.5 (3307)	142.3 \pm 22.9 (1867)
Continental_drift	1 335 \pm 641	19	0.3 \pm 1.7 (3307)	23.6 \pm 7.8 (844)
Heliocentrism	1 026 \pm 564	20	0.3 \pm 1.6 (3307)	25.2 \pm 8.6 (818)
General_relativity	2 060 \pm 1 443	37	0.4 \pm 1.7 (3307)	19.7 \pm 7.8 (1107)
Standard_model	1 202 \pm 2 792	25	0.2 \pm 1.4 (3307)	9.4 \pm 5.0 (575)

^a “Mean Daily Page Views” from <http://toolserver.org/~emw/wikistats/> were only available after 2008-01-01 and include programmatic page requests.

^b include data from 2003-06-12 (when the most recent article, Heliocentrism, originated) through 2012-07-31, when this analysis was run.

^c Mean daily edit count excludes successive edits by the same user (n = 23,156).

^d Mean count of words changed (inserted, deleted, or changed, n = 8,525). Due to the heavily right-skewed distributions, geometric means and standard deviations are shown. The number of observations (n) is constant for mean edits per day because all days were included, while only days with at least one edit were used to calculate the mean words changed.

doi:10.1371/journal.pone.0134454.t001

is not considered a credible or authoritative source . . . any encyclopedia is a *starting point* for research, not an *ending point* [26].” What is needed is a wider appreciation of how to best leverage the vast quantity of information in Wikipedia to take advantage of its strengths (vast coverage and frequent updates) and avoid its weaknesses (potential for errors, conflict between editors, and content stability). Users should be aware that content in Wikipedia can be extremely dynamic; two students could obtain, within seconds, diametrically different information on a controversial scientific topic. Educators should ensure that students understand the limitations and appropriate uses of Wikipedia, especially for controversial scientific issues.

Acknowledgments

We thank Margaret Rubega, Kent Holsinger, John A Silander, and Gregory Anderson at the University of Connecticut in Storrs for helpful comments.

Author Contributions

Conceived and designed the experiments: AMW GEL. Analyzed the data: AMW. Wrote the paper: AMW GEL.

References

1. Alexa. Alexa Top 500 Global Sites index [Internet]. 15 Dec 2013. Available: Alexa.com/topsites. Accessed 15 December 2013
2. Wikimedia. Wikimedia Report Card—October 2014 [Internet]. 1 Oct 2014. Available: <http://reportcard.wmflabs.org/>. Accessed 1 October 2014
3. Voss J. Measuring wikipedia. International Conference of the International Society for Scientometrics and Informetrics. Stockholm, Sweden: International Society for Scientometrics and Informetrics; 2005. pp. 221–231.
4. Giles J. Internet encyclopaedias go head to head. *Nature*. 2005; 438: 900–901. doi: [10.1038/438900a](https://doi.org/10.1038/438900a) PMID: [16355180](https://pubmed.ncbi.nlm.nih.gov/16355180/)
5. West K, Williamson J. Wikipedia: friend or foe? *Reference Services Review*. 2009; 37: 260–271. doi: [10.1108/00907320910982758](https://doi.org/10.1108/00907320910982758)
6. Apic G, Betts MJ, Russell RB. Content Disputes in Wikipedia Reflect Geopolitical Instability. *PLoS ONE*. 2011; 6: e20902. doi: [10.1371/journal.pone.0020902](https://doi.org/10.1371/journal.pone.0020902) PMID: [21731630](https://pubmed.ncbi.nlm.nih.gov/21731630/)

7. Yasseri T, Sumi R, Rung A, Kornai A, Kertész J. Dynamics of Conflicts in Wikipedia. *PLoS ONE*. 2012; 7: e38869. doi: [10.1371/journal.pone.0038869](https://doi.org/10.1371/journal.pone.0038869) PMID: [22745683](https://pubmed.ncbi.nlm.nih.gov/22745683/)
8. Nielsen FA. Scientific citations in Wikipedia. *First Monday*. 2007; 12. Available: http://firstmonday.org/issues/issue12_8/nielsen/index.html
9. Bateman A, Logan DW. Time to underpin Wikipedia wisdom. *Nature*. 2010; 468: 765. doi: [10.1038/468765c](https://doi.org/10.1038/468765c)
10. Bond AL. Why ornithologists should embrace and contribute to Wikipedia. *Ibis*. 2011; 153: 640–641. doi: [10.1111/j.1474-919X.2011.01135.x](https://doi.org/10.1111/j.1474-919X.2011.01135.x)
11. Wikipedia. Wikipedia:BOLD, revert, discuss cycle. In: Wikipedia, the free encyclopedia [Internet]. 2013. Available: http://en.wikipedia.org/wiki/Wikipedia:BOLD,_revert,_discuss_cycle. Accessed 15 December 2013
12. Waters NL. Why you can't cite Wikipedia in my class. *Communications of the ACM*. 2007; 50: 15–17. doi: [10.1145/1284621.1284635](https://doi.org/10.1145/1284621.1284635)
13. Kittur A, Suh B, Pendleton BA, Chi EH. He says, she says: conflict and coordination in Wikipedia. *Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA; 2007. pp. 453–462.
14. Sumi R, Yasseri T, Rung A, Kornai A, Kertesz J. Edit Wars in Wikipedia. *Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on and 2011 IEEE Third International Conference on Social Computing (SocialCom)*. IEEE; 2011. pp. 724–727. doi: [10.1109/PASSAT/SocialCom.2011.47](https://doi.org/10.1109/PASSAT/SocialCom.2011.47)
15. Likens GE, Bormann FH, Johnson NM. Acid rain. *Environment: Science and Policy for Sustainable Development*. 1972; 14: 33–40.
16. Wikipedia. Acid rain. In: Wikipedia, the free encyclopedia [Internet]. 14 Dec 2013. Available: http://en.wikipedia.org/w/index.php?title=Acid_rain&oldid=586087973. Accessed 18 December 2013
17. Wikipedia. Acid rain: Difference between revisions. In: Wikipedia, the free encyclopedia [Internet]. 2011. Available: http://en.wikipedia.org/w/index.php?title=Acid_rain&oldid=463272000. Accessed 15 December 2013
18. Adler BT, Alfaro L de, Mola-Velasco SM, Rosso P, West AG. Wikipedia Vandalism Detection: Combining Natural Language, Metadata, and Reputation Features. In: Gelbukh A, editor. *Computational Linguistics and Intelligent Text Processing*. Springer Berlin Heidelberg; 2011. pp. 277–288. Available: http://link.springer.com/chapter/10.1007/978-3-642-19437-5_23
19. Wikipedia. Acid rain: Difference between revisions (2). In: Wikipedia, the free encyclopedia [Internet]. 2011. Available: http://en.wikipedia.org/w/index.php?title=Acid_rain&oldid=467148215. Accessed 15 December 2013
20. Acid Rain Protection log. In: Wikipedia [Internet]. 2013. Available: http://en.wikipedia.org/w/index.php?title=Special:Log&type=protect&page=Acid_rain. Accessed 18 December 2013.
21. Oreskes N, Conway EM. *Merchants of doubt: how a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. New York, NY, USA: Bloomsbury Press; 2010.
22. Konietzschke F, Hothorn LA, Brunner B. Rank-based multiple test procedures and simultaneous confidence intervals. *Electron J Statist*. 2012; 6: 738–759. doi: [10.1214/12-EJS691](https://doi.org/10.1214/12-EJS691)
23. Samoilenko A, Yasseri T. The distorted mirror of Wikipedia: a quantitative analysis of Wikipedia coverage of academics. *EPJ Data Sci*. 2014; 3: 1–11. doi: [10.1140/epjds20](https://doi.org/10.1140/epjds20)
24. Borra E, Weltevrede E, Ciuccarelli P, Kaltenbrunner A, Laniado D, Magni G, et al. *Contropedia—the Analysis and Visualization of Controversies in Wikipedia Articles*. *Proceedings of The International Symposium on Open Collaboration*. New York, NY, USA: ACM; 2014. pp. 34:1–34:1. doi: [10.1145/2641580.2641622](https://doi.org/10.1145/2641580.2641622)
25. Wikipedia. Wikipedia: Ten things you may not know about Wikipedia [Internet]. 2013. Available: http://en.wikipedia.org/wiki/Wikipedia:Ten_things_you_may_not_know_about_Wikipedia#We_do_not_expect_you_to_trust_us. Accessed 15 December 2013
26. Wikipedia. Academic use. In: Wikipedia [Internet]. 2013. Available: http://en.wikipedia.org/wiki/Wikipedia:Academic_use. Accessed 15 December 2013