

Information Fortification: An Online Citation Behavior

Andrea Forte, Nazanin Andalibi, Tim Gorichanaz,
Meen Chul Kim, Thomas Park
Drexel University
{aforte, na477, gorichanaz, mk3266}@drexel.edu
parkov@gmail.com

Aaron Halfaker
Wikimedia Foundation
ahalfaker@wikimedia.org

ABSTRACT

In this multi-method study, we examine citation activity on English-language Wikipedia to understand how information claims are supported in a non-scientific open collaboration context. We draw on three data sources—edit logs, interview data, and document analysis—to present an integrated interpretation of citation activity and found pervasive themes related to controversy and conflict. Based on this analysis, we present and discuss *information fortification* as a concept that explains online citation activity that arises from both naturally occurring and manufactured forms of controversy. This analysis challenges a workshop position paper from Group 2005 by Forte and Bruckman, which draws on Latour’s sociology of science and citation to explain citation in Wikipedia with a focus on credibility seeking. We discuss how information fortification differs from theories of citation that have arisen from bibliometrics scholarship and are based on scientific citation practices.

Author Keywords

Citation, Bibliometrics, Open Collaboration, Wikipedia

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI):
Miscellaneous

INTRODUCTION

When groups of people work together to create information, they must systematically document its veracity in a way that can be interpreted and assessed by others. In the sciences and humanities, this has traditionally been done via citation systems, using techniques like bibliographies and footnotes. Online, such citation practices have been adapted to the context of open collaboration, blogging, citizen journalism, and other activities where participants are not necessarily trained as scientists or academics.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Request permissions from Permissions@acm.org.

GROUP '18, January 7–10, 2018, Sanibel Island, FL, USA

© 2018 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-5562-9/18/01...\$15.00

<https://doi.org/10.1145/3148330.3148347>

Understanding online citation behaviors and tools is important first because it can strengthen scholarly citation analysis: Analyses of citation data generated in online contexts should not be uncritically informed by assumptions derived from historic scientific practice, but should incorporate an understanding of contemporary online practice. Second, it is important to understand these online behaviors because citation tools themselves can be designed to better suit the motivations and goals of participants in online projects.

In this paper, we examine the history of citation as a shared information practice and reexamine the act of citation in the context of open collaboration on English-language Wikipedia. We use analysis of citation-related edits on Wikipedia and interviews with Wikipedia editors to deconstruct the act of “citing” on the open web. We demonstrate how the tools and context of online participation create a new kind of citation data that challenge some traditional assumptions.

A BRIEF HISTORY OF CITATION TOOLS AND NORMS

The academic practice of citing written works dates back hundreds of years to the exchange of letters among Renaissance scholars, but much of the contemporary interest in citations is owing to a more recent technological renaissance that enabled citation analysis and visualization at unprecedented scale. The intricacies of citation networks have attracted journals full of primarily quantitative analyses of citation patterns and trends that reveal what topics and authors attract increasing, decreasing, or enduring scientific interest. Yet, the meaning of citation and the motivations of citers is not often examined in connection with the interlinked artifacts they produce. Why scientists cite and how they make sense of their own citations is a less-traveled empirical path.

Historian of science Alex Csiszar tracks the rise of the modern scientific publishing apparatus in 19th century Western Europe [6, 7]. Csiszar’s account describes how scientific discourse through the early nineteenth century largely unfurled through ad hoc personal relationships, written letters and meetings that were punctuated by monographs [7]. During the 19th century, the proceedings of academic societies such as Britain’s Royal Society came to be published in volumes [6]. As the emerging global scientific community sought political legitimacy, peer

review was established in order to improve the public's view of science. Publication and citation became more clearly connected to prestige and stature in the scientific community as evidenced by use of publication count in *Philosophical Transactions* in order to determine fellowship in the Royal Society.

As Csiszar explained: "*these events were prompted by, and helped to consolidate, a monumental shift whereby scientists increasingly perceived the social and intellectual life of science to be lodged in the pages of the specialized scientific literature, and especially in the expansive terrain of the scientific periodical. This shift was qualitative, rather than quantitative, and it concerned the ascendancy of the journal as the primary media type for representing authoritative scientific knowledge*" [7, p. 400].

In the wake of this transformation, and particularly since literature came to be practically equated to knowledge, there grew a need to tame the unwieldy and ever-growing scientific literature. Tools like bibliographic artifacts and lists of references were developed as means of accessing knowledge. Following references became a way that scientists kept abreast of the knowledge in their fields [25].

In the 20th century, citations became a form of data and a kind of scientific enterprise in their own right. Not only did cited scientific material become more easily and broadly accessed, but the use of citations in the 20th century as data for both self-assessment and wayfinding in scientific disciplines helped engender new conceptions of science.

In *The Citation Process*, Blaise Cronin described how science was recast from a "storybook science" in which passions and predilections of the scientist were subordinate to the greater institutional imperatives of procedure and objective truth, to a more pragmatic view of science as a kind of game in which scientists jockey for visibility and recognition [5]. Similarly, Latour and Woolgar observed the role of citations in scientific practice as taking two forms: 1) as a tool that allows scientists to strengthen and defend knowledge claims and 2) as a currency, the accrual of which lends credibility to scientists and their work [17]. In Latour and Woolgar's work as in others, the academe is characterized as a marketplace of credibility, in which the value of having one's work cited is manifest in the award of grants, allocation of space, and other tangible benefits. Latour's description of knowledge construction [18] involves the gradual phasing out of citation as knowledge propositions pass from being regarded as tentative to being regarded as common knowledge in a process they describe as stylization, similar to Merton's notion of obliteration by incorporation (OBI), which was popularized by Garfield [11]. Simply, when a knowledge proposition is tentative, anyone who refers to the claim is required to make liberal use of citations to position and attribute it appropriately; however, if a proposition is accepted as fact, it eventually can be claimed without citation.

Henry Small, an early citation analysis specialist, was among the first to develop a theory of citation. For Small, citations are "concept symbols," not for the document being cited, but for the ideas therein [27]. In Small's framework, citations can be metonymical (having physically shared characteristics, e.g., quotes or the same words), or metaphorical (no obvious shared characteristics, e.g., turning some finding into a more general statement). Small considers metonymical citations to be "more faithful" to the originating idea. In either case, what is essential to citation for Small is a reference to ideas, and thus citations are not merely technical devices for making attributions, but rather they manifest the social process of idea exchange in the development of knowledge.

The functions of citation in facilitating idea exchange were explored by Moravcsik and Murugesan [23] who identified four dimensions of citation function, each representing two categories: 1. Conceptual or Operational, 2. Organic or Perfunctory, 3. Evolutionary or Juxtapositional, and 4. Confirmative or Negational. They further found that 36% of citations in physics papers were redundant, for example by supporting claims already made in other cited papers, and mainly served the purpose of keeping everyone happy and distributing credit over a larger number of people in the "game."

In the intervening decades, the "game" has changed radically—a proliferation of citation databases and rankings expanded the playing field. Proposals emerged for alternative metrics that circumvent some of the constraints and game-ability of traditional citation metrics [24]. Moreover, genres and practices of writing and publication have been altered by tools that support not only writing, and not only citation, but collaborative writing and citing at a massive scale. Among the foremost examples of transformation of publishing and writing practices is the open collaboration project, Wikipedia.

Citation Tools and Practices in Wikipedia

There exists a large literature that explains how, unlike many traditional models of information production that rely on expert gatekeepers and peer review processes, Wikipedia relies on an open, participatory model to maintain its quality [28] and coverage [16]. In this section, we will highlight features of Wikipedia's culture, policies and toolkit that characterize and influence the role of citation.

Wikipedia is built on the collaborative writing platform, MediaWiki. The platform started out with few distinctions from other early wikis like the original wikiwikiweb or usemodwiki but with a few modifications to serve the needs of an encyclopedia writing community. Features like edit logging, recent changes lists, discussion pages, and watchlists were quickly developed to allow the community to write together, come to consensus, and keep a watchful eye on the content to maintain its quality.

Citation tools were notably missing from early Mediawiki development efforts. Over the years, citation became a well-documented and central practice, yet Mediawiki was not designed to support the creation of reference data. Instead of being stored as objects in a database that can be linked with articles, citations are embedded in the text of each revision of each article using markup and are rendered as a bibliography when the article is displayed in a browser (See Figure 1).

As Wikipedia grew in scale and popularity, verification of claims quickly became a critical practice inscribed in policies like Wikipedia:Verifiability [32] and standards for identifying reliable sources [31]. Around 2006, about five years after its creation, the project entered a period of rapid content growth, during which commitments to quality were strengthened and policies (the official rules of Wikipedia) began to "calcify," which lead to formalization and automated enforcement of quality-related policies [12]. Yet the tools to support citation practices remained largely external to the infrastructure that support collaborative writing. Third-party tools like *ProveIt* [20] soon materialized to make citation simpler for editors and, eventually, support for citation was integrated into the Wikipedia visual editor via *RefToolBar* in 2010 and the tool *Citoid* [30], which is still in beta testing for English Wikipedia as of this writing. Automated tools like *Citation bot* have also been deployed to automatically standardize and structure citations (User:Citation bot).

As the Wikipedia community evolved citation norms and tools to ensure quality, discussion of Wikipedia citation patterns and practices appeared in some research papers. Notably, Sundin used ethnographic methods to capture the conservatism of Wikipedia's citation policies in his description of knowledge stabilization in Swedish language Wikipedia articles [29]. In Sundin's words, Wikipedia authorship is a process of "recycling" knowledge from established publications in a process through which Wikipedia authors become "janitors" of what is known. Conversely, Ford et al. [8] suggested that, despite policies that state a preference for peer-reviewed, established sources, citations that appear in Wikipedia often include links to "alternative" news or primary sources; however, their analysis omitted references that did not include a URL, which may have excluded citations to preferred secondary and tertiary sources such as textbooks. In an earlier quantitative examination of Wikipedia citation edit histories, Chen and Roth described patterns of citation over time, observing that articles did not experience consistent attention to citations, but experienced periods of citation-related activity once they had matured; they also noted that experienced Wikipedians were more likely to add citations than novices [4].

Wikipedia articles and authors differ in terms of goals, tools, and practices from the academic publications and their authors who have informed so much of our understanding of citation practices. If our understanding of

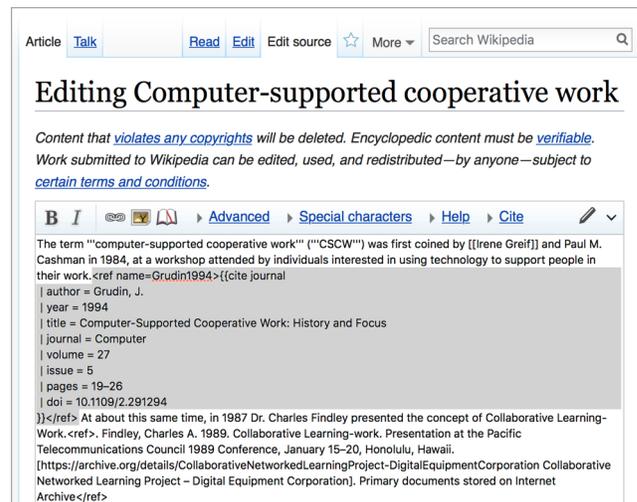


Figure 1: Example of reference syntax (highlighted) when editing a Wikipedia article.

citation is grounded in centuries-old traditions of academic—and specifically scientific—practice, can it be applied to understand the practices of people who contribute to online information sources like Wikipedia? In reviewing literature about traditional scientific citation practice, we saw that citation not only plays the role of justifying claims, but as a way of accruing and conferring credit in the scientific community. Is citation in Wikipedia embedded in a similar system of accountability and prestige? What motivates acts of citation in Wikipedia and how do editors make sense of their own citation practices? What roles do citations play in the online encyclopedia?

STUDY DESIGN AND METHODS

To answer these questions, we use a sociotechnical perspective and mixed methods approach to examine the interplay between tools, norms, and individual motivations as people adapted citation to meet the demands of open collaboration on English-language Wikipedia (All further references to "Wikipedia" are to the English language site unless otherwise noted).

First, we obtained all of the reference-related changes made to the Wikipedia articles through March 2015 using standard wiki markup tags. Using these data, we were able to examine citation-related editing activity over time. Second, we used Wikipedia citation data to identify Wikipedia editors who added citations to the site and recruited 9 Wikipedia editors who were prolific citers to participate in interviews about their practices. Third, we manually inspected revisions of thirty-five random articles to develop an artifact-based understanding of how citations are used in the development of articles. We explain how we obtained each of these data and the kind of analysis we conducted to understand different features of citation.

Reference Revision Data and Quantitative Analysis

To explore reference-related activity on Wikipedia and identify likely interviewees, we parsed all reference-related

revisions to English language Wikipedia between 2001 and March 2015 that used the standard <ref> markup tag. From these data, we were able to ascertain which references were edited in a given revision, whether reference-related text was inserted or removed, and what username (or IP address) was associated with the revision. To understand citation activity at the article level, we compared articles that had undergone peer review as part of achieving Featured Article status as well as those that hadn't. Featured Articles, which appear on the Wikipedia main page, are considered to be among the highest quality articles on the site. We randomly sampled 50 articles that had never been nominated for featured article status, as well as 50 that had been chosen as featured articles and examined their citation activity over time.

Interview Data and Thematic Analysis

To understand the meaning and role of citation to the people doing the citing, three of the authors conducted 10 semi-structured interviews with Wikipedia editors using phone or Skype and, in one case, email. Using the reference revision dataset described above, we identified Wikipedia editors who had a record of prolific citation activity. Those whose email addresses were available publicly online or who had enabled email communication on Wikipedia were invited to participate. In all, we attempted to reach sixty-six people, although we do not know how many of these received our messages. The 10 individuals who were interviewed included 8 men and 2 women. All participants had logged thousands of citation-related edits, so it is unsurprising that all were experienced editors. Each recounted approximately ten years of editing experience.

As we conducted interviews, we discussed their content and continued recruitment until we reached data saturation, meaning all the interviewees reported hearing similar themes in each interview. Interview recordings were transcribed with the exception of one interview during which the recording failed and only interviewer notes were retained and one which was conducted via email by participant request. In the interviews, we started off by asking Wikipedians about how they got started editing Wikipedia and how their perceptions and contributions changed over time. Then we prompted them to review a few specific citations they had added in the recent past and asked them to share why they added that citation, as well as details about how they came across the article, and how they chose that particular citation. We also asked them about citation-related tools they use, citation habits outside of Wikipedia, personal and perceived rules about citations on Wikipedia and off, removing citations or having their citations removed, interactions with other Wikipedians and perceptions of others' citation behavior.

These interviews were semi-structured and retrospective, and the resulting data are subject to limitations such as fallible memories and uneven coverage of data points, depending on the idiosyncratic experiences of the interviewees.

Article Data and Document Analysis

We collected a sample of thirty-five Wikipedia articles and used an inductive approach based on holistic reading analysis techniques [21] to understand reference editing activity as a feature of the article. To collect this sample, we used the *random article* feature of Wikipedia. If *random article* returned an article without references, we discarded that article until thirty-five articles with references had been identified. Unsurprisingly, since only about 6,000 out of over 5.5 million English language articles have been featured articles (approximately 0.1%), none of the 35 randomly selected articles had been nominated for featured article status. The articles age averaged 7 years at the time of the analysis (max: 13; min: 2).

We reviewed the edit history of each of these articles with attention to reference-related changes as well as talk page content. The goal was to develop a holistic understanding of how Wikipedia articles evolve, inspired by Joseph Campbell's notion of the *monomyth*—a kind of general template that describes commonalities among myths [14]. Campbell explored the myths and folklore of diverse cultures and discerned recurring typical themes in human mythology. In this spirit, we sought to construct a narrative that characterizes the evolution of typical Wikipedia articles with respect to citation activity. Van Manen advocates a similar sort of anecdote construction for the conveying and understanding of possible, plausible experiences in phenomenology of practice [21].

Integrated Analysis

These methods seek to integrate social-scientific and humanistic approaches. There is some precedent for this in HCI research. For example, Blythe and Cairns conducted a qualitative social-science analysis (grounded theory) alongside a critical-hermeneutic analysis of the same dataset [3]. The grounded theory approach gave them a sense of what was going on macroscopically, and the hermeneutic approach allowed them to dig deeper into an apparent anomaly. The concept of defamiliarization is also an important humanistic feature of our work [2]. Defamiliarization is a matter of becoming naive toward a familiar phenomenon in attempt to uncover aspects of it that have been taken for granted. In our case, the authors sought to rediscover citation (an activity we practice ourselves regularly) in a new context with different goals and meanings.

As Bardzell and Bardzell argue, adopting social scientific and humanistic approaches in concert can further inquiry into aspects of life “*that are so hard to pin down adequately with any single method or mentality*” [1] (p. 57). In our work, we took a critical-thinking approach to integrating the findings from our various data, keeping in mind the epistemological commitments of each, in order to accomplish theoretical work.

In our findings and discussion sections, we use the term *sourcing* to refer to the act of substantiating a statement with a source, a *reference* to refer to the text that refers to

the source, *referencing* to refer to the technical act of inserting the reference, and *citation* to refer to the activity as a whole. In revision history data, we have access to *references* and *referencing* acts. In document analysis we have access to both *references* and, to a limited extent, *sourcing* as we examine the citation activity in context. In interviews, we receive explanations of both *sourcing* and *referencing* activity.

FINDINGS

The data and analyses described above revealed different aspects of citation activity; in this section, we draw on all three data sources to present an integrated interpretation of citation on Wikipedia that is reinforced by each of our analyses. Each data type yielded different insights related to themes of controversy and conflict.

Contrary to Forte and Bruckman's insights in [10], citing in Wikipedia seems to bear only a tenuous connection to the marketplace of credibility described by Latour and Woolgar [17]. In fact, citation behavior on Wikipedia appears to be much more closely related to Latour and Woolgar's militaristic descriptions of citation.

In interviews and document analysis, we observed that citation activity is frequently an outgrowth of conflict, either due to *naturally arising controversy* that yield article improvement activity or as a result of review processes, which we cast here as a form of *manufactured controversy*, designed to generate debate and scrutiny to yield improvements. In the findings sections, we describe how natural and manufactured controversy give rise to citation activity in Wikipedia. In the discussion section, we use these findings to generate the concept of information fortification.

Natural Controversy

Our interviews and document analysis demonstrated how citation activity was often a result of shared norms and observation of policy, but this activity increased in the presence of real or perceived controversy that arose spontaneously as article co-editors worked and interacted.

In interviews with editors whose contributions to Wikipedia includes a large number of reference-related edits, we most often heard general explanations of citation activity that implicitly or explicitly invoked Wikipedia policy. Citation-related policies were summarized by one participant as encapsulating a defensive editing mindset: "*The basic idea is that everything that is disputable is most likely to be challenged and has to have a reference with a reliable source*" (P2). The same participant also explained that on Wikipedia, statements that might constitute common knowledge when writing in a "normal way" still require a verifiable source on Wikipedia:

...on Wikipedia you have to—if you take for example, a football player who was born in Leimen, usually on Wikipedia, you have to source this. But if everyone knows because the guy is very popular,

then you would not put any reference on this in, let's say, work or normal way, because it's widely known that he was born there and so, this is not something that is disputed (P2).

Another participant also explained that anything might be considered contentious:

The rules strictly say that you only need to cite something if it's contentious, if someone is likely to bother you about it or disagree with it. But the reality is that you have to cite everything, because someone will decide that it's contentious... And it gets very weary, because I mean, many things, you'll know yourself, you don't need to cite them, they're just—people know they're true (P4).

When asked to describe the rationale for adding specific references, participants sometimes described local conditions such as whether a statement was likely to be challenged or removed. When asked to explain why they had added references in recent edits, one participant noted that it was "*because somebody had challenged what I'd put in, so I went back and added the citations as needed*" (P7). In another case, a participant explained that they had noticed a quotation (which had been added by another editor) had been removed from an article (by a second editor) because it was unsourced, so the participant (the third editor in this case) restored the content and inserted a reference to an appropriate source in order to protect the passage from deletion:

Somebody—an history editor—deleted a quotation because it did not have a citation. So, I looked through and found the citation and found a better citation for his previous quote, so I changed two of his footnotes. I restored one. And I do the search, in this case, through Google Scholar. Google Scholar picks up the quotations real easy (P6).

Google Scholar came up repeatedly as an easy way to locate appropriate sources to support encyclopedic claims.

Our manual inspection of the reference-related edit histories of 35 articles revealed that historically many pages were created with no references. In fact, none of the 13 articles that were created before 2007 included references at the time they were created. The 13 articles created from 2007 to 2009 sometimes included references at inception. The remaining 9 articles created in 2010 and later all included references from the beginning. In the case of articles that were created without references, reference sections were often later created by a bot, and eventually populated by a human who, based on the interviews described above, did not want their edits or other content removed. Over the life of early articles, we noted that references were often inserted to support later-added material, but much of the original article content can remain unsubstantiated unless it is challenged.

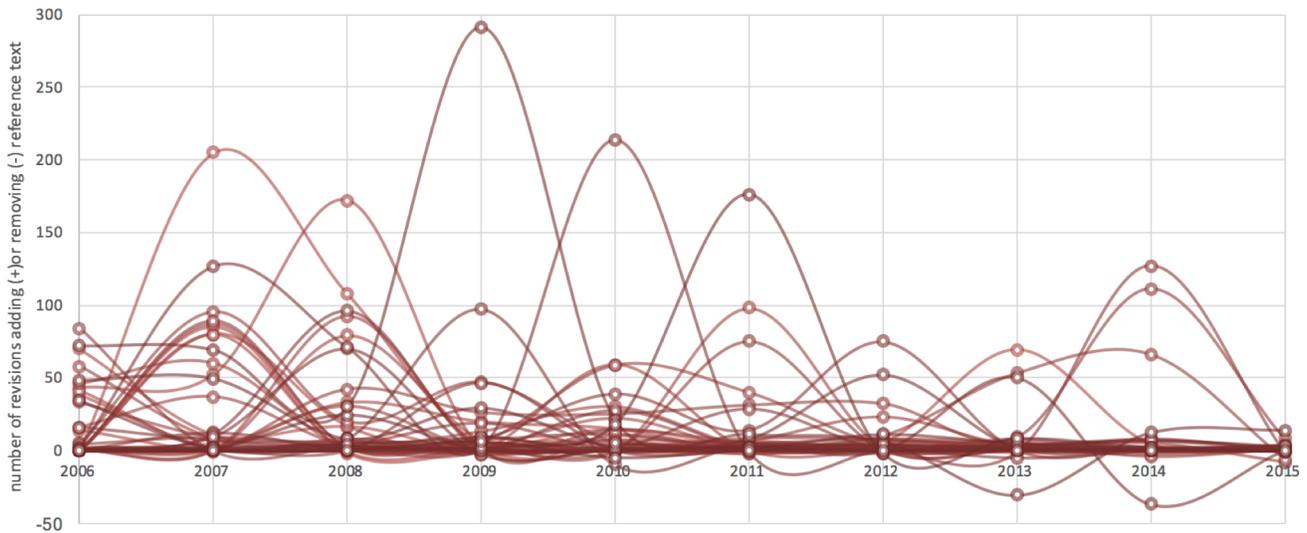


Figure 2: Revisions in which reference text was added to or removed from 50 FEATURED ARTICLES by year

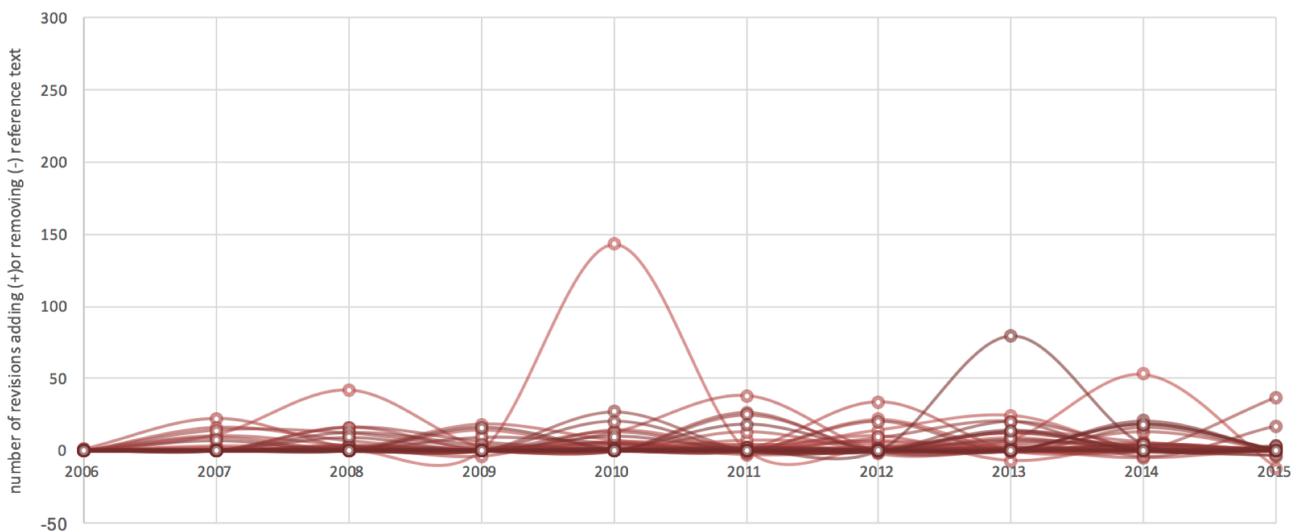


Figure 3: Revisions in which reference text was added to or removed from 50 NON-FEATURED ARTICLES by year

By examining revision data, we observed that reference-related editing activity on individual articles ebbed and flowed over years with distinct peaks and valleys (See Figures 2 and 3). As expected, non-featured articles exhibited lower peaks than featured articles. We also observed that real-world events sparked the reference activity peaks in non-featured articles, for example, the two highest peaks in non-featured article reference activity were when a volcano erupted and an actor won an award. This is consistent with observations that breaking news events precipitate fast-paced editing and dense co-author networks [15] in which well-sourced claims would be critical to the survival of text.

Manufactured controversy

Controversy doesn't always arise naturally either because of one editor challenging another or because material is

deemed likely to be challenged, sometimes formal processes are initiated to ritualize the process of challenging content. These rituals are commonly referred to (both inside Wikipedia and in academic contexts) as “peer review.” In Wikipedia, there are two primary forms of invoking a review of claims, one is decentralized through the insertion of a `{{citation needed}}` template and the other is by engaging a more centralized review process, the most rigorous of which is the nomination for Featured Article (FA) status.

Decentralized

By inserting the text `{{citation needed}}` into an article, editors can tag an unsupported claim and challenge the community to find a reliable source for it. In the ideal case, this attracts the attention of other editors or readers and triggers citation activity to help verify the claim. Precisely

this kind of activity is described by one interviewee who recounted one occasion on which they found and added references to deter the removal of content from an article:

someone... put a lot of 'citation needed' text on this article or removed stuff which was not referenced. And then I thought, 'okay, it's a shame that all this stuff has to be removed because it's not referenced,' and then I did some work and Googled for some references and added them (P2).

Another editor explained that if you don't know what to cite but, "*you've got a good interesting segment, you should put in a 'citation needed' flag*" (P6). Yet another interviewee's behaviors suggest that this is not always effective because, although they enjoy citation-related editing, they do not go out of their way to address claims tagged with {{citation needed}}:

I find bad citations, or dead links, or that sort of thing all the time, but unless I feel like I can correct it pretty easily, I generally just leave that. I'd leave all the ones that say "citation needed." They pretty much just sit there for me (P5).

The same participant went on to explain that:

There seems to be an awful lot of calls for citations—citation needed—and I don't know what happens with people... it seems like people create the article and then go away. It's like somehow when somebody else goes through it and calls for a citation or marks as a citation needed that the [original] author never comes back (P5).

Aside from randomly encountering content marked with {{citation needed}}, Wikipedia editors may use a tool called a "watchlist"¹ to keep an eye on the content they have created or have a special interest in. The excerpts above suggest that these techniques may be too haphazard or underutilized to yield desired citation activity. New tools may be needed help facilitate decentralized content review.

Centralized

Peer review is a process for subjecting content to scrutiny and inviting critique. In Wikipedia, peer review happens most centrally as an encyclopedia-wide process that culminates in Featured Article status. It also happens nested within topic-specific and thematic "wikiprojects" that maintain their own standards and review processes. On Wikipedia, much of the peer review process involves verifying whether an article is sufficiently well sourced, and standards have been refined over time:

I could remember my first few featured articles, I had to return to a few years later and bring them up to scratch with extra citations because a lot of the stuff was un-cited, or not specifically cited. We had a different view then, this was 2007. But round

about the time, as in 2007, 2008, the idea of inline citations as a necessity became, well, it was a big issue. And it became general. And I've always followed that since. I've always been careful to cite everything. (P3).

In fact, during the featured article process, editors also commonly use {{citation needed}} tags to organize article improvement efforts. Featured Articles are considered to be the highest quality articles, but other levels of review also exist, including Good Article, A-class, B-class, and others. Initiation of a review process serves as an invitation to the community to examine an article critically and apply relevant standards. One interviewee explained how review processes helped them organize their hobby of reference cleaning and correction:

I have a bit of a peculiar hobby of liking to look at articles, maybe that are going through a good article review, and if the references aren't perfectly formatted, I like to fix them (P1).

These kinds of helpful copyedits are distinct from sourcing activity, in that they only serve to improve the reference text itself. They account for some of the reference-related editing activity we see in graphs of Featured article reference-related revisions (see Figure 2). We did not categorize revisions to differentiate them qualitatively.

In the 100 revision histories we examined, featured article process shows up as a clear antecedent to reference-related editing activity. In figure two, the peaks and valleys associated with reference-related editing in featured articles is more pronounced than in non-featured articles. Most often, peaks directly preceded the featured article process, although in some cases peaks preceded featured article status by years. In the most extreme case, the article on Hillsboro, Oregon, peaked at 95 reference-related revisions in 2007 but was not nominated for featured status until 2009 and was not featured until seven years after the peak in 2014. In this case, other centralized quality improvement processes were responsible for the peak, namely efforts to prepare the article for "good article" status in 2007. In only one case, peak citation activity occurred after FA status: The Simpsons was featured in 2007, but in 2009 the show became the longest-running television show and the format changes and in that year citation-related revisions to the article reached a peak.

Quality and Quantity

A few themes recurred that were not related to conflict but are important because they explain important features of citation activity and can help with interpretation of revision data. All of our participants discussed the quality of sources in terms of publisher or source type (blog, news article, journal article, etc.) and mentioned the quality of the reference itself in terms of formatting, completeness, and/or consistency with other references in an article. The value of a complete and well formatted reference is important to note when interpreting data about the number of reference-

¹ <https://en.wikipedia.org/wiki/Help:Watchlist>

related edits since they may include many revisions that do not add a new source although they add new reference text.

More than once, participants mentioned the desirability of economizing references. One participant described someone with whom they stopped co-editing because: “*she used to add references to statements that were already referenced and the references she added weren’t necessary at all. I said ‘Hey, these references aren’t very necessary’ and she wasn’t very happy about that at all*” (P8). Participant 3 offered a similar opinion:

I'm dead against that, when people offer a fairly simple fact, and because that fact is in eight different books, they feel they've got to cite it to the eight different books, and I tell people not to do that. I say, ‘No, one citation is enough, or two if you really want.’ (P3).

A preference for few but high-quality references may also yield reference-related deletion of text.

DISCUSSION

Our findings characterize much of the citation activity in Wikipedia as an effort to protect encyclopedic content in response to real and perceived conflict. When we embarked on this project, we were familiar with citation as a knowledge construction activity. Our assumptions about the rhetorical and functional value of citation in the sciences quickly gave way as we perceived Wikipedians’ tendency to describe citation as a defensive act. We use the term *information fortification* to describe citation activity designed to preserve the visible and online status of information claims in Wikipedia and which may occur online more generally in blogs, discussion threads, and elsewhere. Information fortification differs in several important dimensions from well-known theories of citation.

Information Fortification and Other Theories of Citation

Although theories of citation activity in the sciences account for some motivations similar to information fortification, the functions and meanings of citation in science typically include a layer of political and social relationships among the citers and cited that is absent in information fortification. Recall that Latour and Woolgar included two roles for citation: 1) as a tool that allows scientists to strengthen and defend knowledge claims and 2) as a currency, the accrual of which lends credibility to scientists and their work [17]. Although the first role is largely similar to information fortification, the “marketplace of credibility” that sustains practices like redundant citation (the 36% of physics citations sampled by Moravcsik and Murugesan [23]) does not appear to be present in the context of open collaboration. Or at least not among editors who cite prolifically. Participants noted that they discouraged others from citing redundantly and even removed such references. No participant mentioned the need to add citations to an article for reasons like signaling membership in a theoretical camp, to appease likely reviewers, or to prove their own expertise.

Another point where information fortification departs from commonly described citation practices in the sciences is in which statements need to be sourced. Widely known Latourian and Mertonian accounts of citation in the sciences incorporate concepts of erosion [18] or obliteration by incorporation [11]. Both of these ideas suggest that the construction of fact entails a diminishing need to provide a citation for a claim. For example, a new idea in a scientific field requires citation to substantiate the claim, but a well-accepted “fact” no longer requires citation. Information fortification obligates citation for claims regardless of how well accepted they might be, in part because the potential grounds for controversy is not bounded by a particular field of expertise. Part of what distinguishes Wikipedia as an open collaboration project is its low barrier to entry and exit—anyone who wants to contribute to the project is welcome to edit regardless of their past experience or training. This creates a review context distinct from scientific fields where “peers” are generally scientists who share at least some general disciplinary training if not deep expertise in the topic at hand. Writing an article about cell biology sourced to satisfy the information needs of all internet users is an entirely different matter than writing one to satisfy the needs of a biologist. Our interviewees described writing in a context where anyone might challenge anything at any time.

We note that an additional reason for citation that did not surface explicitly in interviews, is the need to support the information needs of readers not only as potential reviewers but also as encyclopedia users who come there to learn. In order to learn about a topic or to find information sources about a topic, people use the site to get started or as a pointer to other information sources [9, 13]. For this reason, the quality of information sources used on the site is important not only to substantiate claims in the article, but also as sources of further reading and consultation.

The Effectiveness of Stigmergic Citation

Stigmergy describes a form of indirect coordination in which one actor leaves a trace of activity in the environment that is interpreted by and prompts the activity of the next actor. We heard from interviewees that the addition of a {{citation needed}} template was used to support distributed, stigmergic verification of information claims in the encyclopedia. This tool was used both in a distributed way by editors as they wrote articles, and as a part of organized review processes.

Centralized processes like featured article review by definition produce high-quality, well-sourced articles, but the effectiveness of decentralized, stigmergic citation efforts is less clear. How appeals for citation using {{citation needed}} get resolved not remains unexamined. When Wikipedians (or bots) assess the quality of articles, C-class and above are expected to have at least some references; however, fewer than 8% of English Wikipedia’s 5.5 million articles have been rated at C-class or above [33]. That means the vast majority of articles have not undergone a centralized review process.

Recently, the tool CitationHunt² was deployed to encourage the addition of references to resolve instances of {{citation needed}} and reduce the number of unsupported claims. CitationHunt prompts users with an unsourced statement in Wikipedia and offers a link to edit the article where the statement appears to add a reference and remove the citation needed template. When Wikipedia editors described sourcing statements that had been originally written by others, they often described using a web search—often Google or Google Scholar—to quickly find a source that supported the claim in need of verification. This is an example of the kind of new tool we speculated may be helpful in guiding editors to pages where {{citation needed}} is used.

The question remains whether CitationHunt or the use of the {{citation needed}} template in general yields references to sources of comparable quality as those added by authors of articles who include references as they compose an article. In light of filter bubbles and the vast number of resources available to information, these practices raise the question of whether information statements written first and sourced later are more subject to confirmation bias. The question of neutrality has long been discussed in Wikipedia, where a foundational policy is NPOV (neutral point-of-view), which requires Wikipedia articles to report multiple interpretations or controversy about a concept without supporting a particular perspective. Many scholars have raised the question of whether neutrality is possible, not just in Wikipedia [22, 26], but in the production of knowledge at all [19]. The question becomes newly relevant in the context of sourcing information claims. Can citation be a neutral activity? According to Latour and Woolgar, the construction of scientific fact is a process that depends on persuasive acts of literary inscription [17]. Citation in Wikipedia may be mobilized by different interests than in the sciences, but do the biases of the cited sources remain? If Wikipedia's sources express implicit biases, can these biases be neutralized through a collaborative process of consensus building and meta-analysis that aspires to neutral representation of facts? The questions raised by this work will inform our future studies of citation activity online.

SUMMARY

Information fortification is a form of citation activity distinct from the scientific practice in which citation practices originally emerged. The fortification mindset reflects the precarious status of information statements that are asserted on Wikipedia. Social functions of citation like signaling theoretical positioning or credibility exchange were absent. Unlike scientists, Wikipedia editors do not themselves take responsibility for creating new knowledge about the world, instead they do the work of assembling information. In this context, references to sources defend information that has been integrated and curated from

others who would challenge or undo that work of information assembly. The expansive grounds on which statements might be challenged in Wikipedia is a characteristic of the online open collaboration context.

ACKNOWLEDGEMENTS

This work was supported by a National Science Foundation grant (#1253302).

REFERENCES

1. Jeffrey Bardzell and Shaowen Bardzell. 2015. Humanistic Hci. *Synthesis Lectures on Human-Centered Informatics*, 8, 4, 1-185.
2. Genevieve Bell, Mark Blythe and Phoebe Sengers. 2005. Making by making strange: Defamiliarization and the design of domestic technologies. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12, 2 (2005), 149-173.
3. Mark Blythe and Paul Cairns. Year. Critical methods and user generated content: the iPhone on YouTube. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
4. Chih-Chun Chen and Camille Roth. 2012. {{Citation needed}}: the dynamics of referencing in Wikipedia. In *Proceedings of Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration* (Linz, Austria).
5. Blaise Cronin. 1984. *The citation process. The role and significance of citations in scientific communication*.
6. Alex Csiszar. 2015. *Objectivities in Print*. Springer, City.
7. Alex Csiszar. 2010. Seriality and the search for order: Scientific print and its problems during the late nineteenth century. *History of science*, 48,399-434.
8. Heather Ford, Shilad Sen, David R Musicant and Nathaniel Miller. Year. Getting to the source: where does Wikipedia get its information from? In *Proceedings of Proceedings of the 9th international symposium on open collaboration*.
9. Andrea Forte and Amy Bruckman. 2009. Citing, writing and participatory media: wikis as learning environments in the high school classroom. *International Journal of Learning and Media*, 1, 4 (2009), 23-44.
10. Andrea Forte and Amy Bruckman. 2005. Why do people write for Wikipedia? Incentives to contribute to open-content publishing. Workshop Position Paper, ACM Conference on Groupwork (GROUP 2005).
11. Eugene Garfield. 1975. The Obliteration Phenomenon. *Current Contents*, 51-52 (1975).
12. Aaron Halfaker, R Stuart Geiger, Jonathan T Morgan and John Riedl. 2013. The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 57, 5, 664-688.

² <https://tools.wmflabs.org/citationhunt/>

13. Alison J. Head and Michael B. Eisenberg. 2010. How today's college students use Wikipedia for course-related research. *First Monday*.
14. Campbell Joseph. 1949. *The hero with a thousand faces*. Princeton, NJ: Princeton University.
15. Brian Keegan, Darren Gergle and Noshir Contractor. 2013. Hot off the wiki: Structures and dynamics of Wikipedia's coverage of breaking news events. *American Behavioral Scientist*, 57, 5, 595-622.
16. Aniket Kittur, Ed H. Chi and Bongwon Suh. 2009. What's in Wikipedia?: mapping topics and conflict using socially annotated category structure. In *Proceedings of Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA, 2009). ACM.
17. B. Latour and S. Woolgar. 1986. *Laboratory Life: the Construction of Scientific Facts*. Princeton University Press.
18. Bruno Latour. 1988. *Science in Action: How to Follow Scientists and Engineers through Society*. Harvard University Press.
19. Richard Lewontin. 1996. *Biology as ideology: The doctrine of DNA*. House of Anansi.
20. Kurt Luther, Matthew Flaschen, Andrea Forte, Christopher Jordan and Amy Bruckman. Year. Provelt: A New Tool for Supporting Citation in MediaWiki. In *Proceedings of ACM Symposium on Wikis and Open Collaboration* (WikiSym '09).
21. Max van Manen. 2014. *Phenomenology of Practice*. Left Coast Press, Walnut Creek, CA.
22. Paolo Massa and Federico Scrinzi. 2012. Manypedia: comparing language points of view of Wikipedia communities. In *Proceedings of Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration* (Linz, Austria, 2012).
23. Michael Moravcsik and Poovanalingam Murugesan. 1979. Citation patterns in scientific revolutions. *Scientometrics*, 1, 2 (1979), 161-169.
24. Jason Priem, Dario Taraborelli, Paul Groth and Cameron Neylon. 2010. Altmetrics: A manifesto (2010).
25. W Boyd Rayward. 1997. The origins of information science and the International Institute of Bibliography/International Federation for Information and Documentation (FID). *JASIS*, 48, 4, 289-300.
26. Joseph Reagle. 2005. Is the Wikipedia Neutral?.
27. Henry G Small. 1978. Cited documents as concept symbols. *Social studies of science*, 8, 3 (1978), 327-340.
28. Besiki Stvilia, Michael B Twidale, Linda C Smith and Les Gasser. 2008. Information quality work organization in Wikipedia. *Journal of the Association for Information Science and Technology*, 59, 6, 983-1001.
29. Olof Sundin. 2011. Janitors of knowledge: constructing knowledge in the everyday life of Wikipedia editors. *Journal of Documentation*, 67, 5, 840-862.
30. Wikipedia. Citoid. <https://www.mediawiki.org/wiki/Citoid>. Accessed: June 30, 2017.
31. Wikipedia. Identifying Reliable Sources. https://en.wikipedia.org/wiki/Wikipedia:Identifying_reliable_sources. Accessed: 6-30-17.
32. Wikipedia. Wikipedia: Verifiability. <http://en.wikipedia.org/wiki/Wikipedia:Verifiability>. Accessed: 6-30-17.
33. WP 1.0 bot. Wikipedia Articles by Quality and Importance. https://en.wikipedia.org/wiki/User:WP_1.0_bot/Tables/OverallArticles. Accessed: 06/30/17.