UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Quantifying Engagement with Citations on Wikipedia

Piccardi, T.; Redi, M.; Colavizza, G.; West, R.

DOI
10.1145/3366423.3380300

Publication date
2020

Document Version
Final published version

Published in
The Web Conference 2020

License
CC BY

Link to publication

Citation for published version (APA):
Piccardi, T., Redi, M., Colavizza, G., & West, R. (2020). Quantifying Engagement with Citations on Wikipedia. In *The Web Conference 2020: Proceedings of The World Wide Web Conference WWW 2020 : April 20–24, 2020 Taipei, Taiwan* (pp. 2365-2376). International World Wide Web Conference Committee. https://doi.org/10.1145/3366423.3380300

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:15 Mar 2026

Quantifying Engagement with Citations on Wikipedia

Tiziano Piccardi
EPFL
tiziano.piccardi@epf\/l.ch

Miriam Redi
Wikimedia Foundation
miriam@wikimedia.org

Giovanni Colavizza
University of Amsterdam
g.colavizza@uva.nl

Robert West*
EPFL
robert.west@epf\/l.ch

ABSTRACT

Wikipedia is one of the most visited sites on the Web and a common source of information for many users. As an encyclopedia, Wikipedia was not conceived as a source of original information, but as a gateway to secondary sources: according to Wikipedia's guidelines, facts must be backed up by reliable sources that reflect the full spectrum of views on the topic. Although citations lie at the heart of Wikipedia, little is known about how users interact with them. To close this gap, we built client-side instrumentation for logging all interactions with links leading from English Wikipedia articles to cited references during one month, and conducted the first analysis of readers' interactions with citations. We find that overall engagement with citations is low: about one in 300 page views results in a reference click (0.29% overall; 0.56% on desktop; 0.13% on mobile). Matched observational studies of the factors associated with reference clicking reveal that clicks occur more frequently on shorter pages and on pages of lower quality, suggesting that references are consulted more commonly when Wikipedia itself does not contain the information sought by the user. Moreover, we observe that recent content, open access sources, and references about life events (births, deaths, marriages, etc.) are particularly popular. Taken together, our findings deepen our understanding of Wikipedia's role in a global information economy where reliability is ever less certain, and source attribution ever more vital.

1 INTRODUCTION

Wikipedia is the largest encyclopedia ever built, established through the collaborative effort of a large editor base, self-governed through agreed policies and guidelines [7, 16]. Thanks to the tenacious work of the editor community, Wikipedia's content is generally up to date and of high quality [25, 45], and is relied upon as a source of neutral, unbiased information [35].

Wikipedia's inline references, or citations,[1] are a key mechanism for monitoring and maintaining its high quality. Wikipedia's core content policies require that "people using the encyclopedia can check that the information comes from a reliable source",[2] and citations are the main way to connect a statement to its sources. A clearly distinctive feature of Wikipedia is the fact that many citations are actionable: they are often equipped with hyperlinks to the cited material available on the Web.

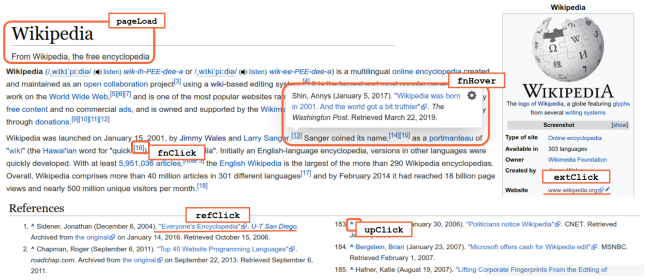


Figure 1: Examples of the 6 types of interactions with pages and citations that we record on English Wikipedia using Wikimedia's EventLogging tool.

As a result, Wikipedia's role on the Web has been defined as the "bridge to the next layer of academic resources" [19], and the "gateway through which millions of people now seek access to knowledge" [11]. Nevertheless, a question remains open: to which extent do Wikipedia readers actually cross the bridge and access the broader knowledge referenced in the encyclopedia?

Given the collaborative and open nature of Wikipedia, being able to quantify readers' engagement with the content and its supporting sources is of crucial importance for the constant betterment of the encyclopedia and its role in fostering a self-critical society. By understanding readers' interactions with citations, we can better assess the role of Wikipedia editors and policies in maintaining a high quality of information, measure public demand for secondary sources, and provide insights and potential recommendations to increase the public's interest in references.

This paper takes a step in this direction, by addressing, for the first time, the problem of quantifying and studying Wikipedia readers' engagement with citations. More specifically, we ask the following research questions,

RQ1 To what extent do users engage with citations when reading Wikipedia? (Sec. 4)

RQ2 What features of a page predict whether a reader will interact with a citation on the page? (Sec. 5)

*Robert West is a Wikimedia Foundation Research Fellow.

[1] We use the terms "reference" and "citation" largely interchangeably.

[2] https://en.wikipedia.org/wiki/Wikipedia:Verifiability, https://en.wikipedia.org/wiki/Wikipedia:Reliable_sources

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.
WWW '20, April 20–24, 2020, Taipei, Taiwan
© 2020 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.
ACM ISBN 978-1-4503-7023-3/20/04.
https://doi.org/10.1145/3366423.3380300

RQ3 What features of a citation predict whether a reader will interact with it? (Sec. 6)

In order to answer these questions, we collect a large dataset comprising all citation-related events (96M) on the English Wikipedia for two months (October 2018, April 2019), including reference clicks, reference hovers, and downwards and upwards footnote click, as visualized in Fig. 1. By analyzing this dataset,[3] we make the following main contributions:

- We quantify users' engagement with citations and find that it is a relatively rare event (RQ1, Sec. 4): 93% of the links in citations are never clicked over a one-month period, and the fraction of page views that involve a click on a citation link is 0.29%.
- We gain insights into factors associated with seeking additional information via citation interactions, both at the page level (RQ2, Sec. 5) and at the link level (RQ3, Sec. 6). Through matched observational studies, we show that articles that are of higher quality, and thus also longer and more popular, are associated with a lower propensity of users to interact with citations. Using a logistic regression model trained on linguistic features, we show that more frequently clicked citation links tend to relate to social or life events.

We thus conclude that readers are more likely to use Wikipedia as a *gateway* on topics where Wikipedia is still wanting and where articles are of low quality and not sufficiently informative; and that Wikipedia tends to be the *final destination* in the large majority of cases where the information it contains is of sufficiently high quality.

Our work provides the first study aimed at understanding if and how users engage with citations on Wikipedia, thus paving the way for a broader and deeper understanding of Wikipedia's role in the global information ecosystem.

2 RELATED WORK

This paper is related to research on a number of different themes.

Characterizing Wikipedia readers. A substantial amount of prior work has focused on understanding user engagement with Wikipedia from the point of view of the *editor* community [2, 41, 43, 68]. Studies on the behavior of Wikipedia *readers* have mostly considered interest in contents [31, 49, 63], content popularity [8, 47, 56], or event timing [37]. More recently, a study explored the question why users read Wikipedia, by combining multiple-choice surveys with log-based analyses of user activity [53]. A similar design was used to study 14 languages other than English [32]. Little is known, however, on how users engage with Wikipedia's citations of external sources; ours is the first study on this subject.

Navigation in Wikipedia. Wikipedia citations are part of the hyperlink network connecting Wikipedia and the Web. Understanding citation usage can yield useful insights for improving this network [29, 66]. The analysis, modeling, and prediction of human navigation *inside* Wikipedia has been considered in previous studies [13, 18, 21, 30, 52, 62], largely relying on traces from the navigation games Wikispeedia [50, 64, 65] and WikiGame [12, 26, 54]. For our study, we collect instead a new, fine-grained dataset of user interactions with Wikipedia references to *external* content.

Science in Wikipedia. A sizeable portion of citations on Wikipedia refer to scientific literature [40]. Consequently, Wikipedia is a fundamental gateway to scientific results and enables the public understanding of science [33, 34, 38, 51, 61]. The chance of a scientific reference being cited on Wikipedia varies with the impact factor of the publication venue and its open-access availability [58]. Being cited on Wikipedia can thus be considered an indicator of impact [27]. Despite the indirect influence that Wikipedia has on scientific progress [59], Wikipedia is in turn rarely acknowledged in the scientific literature [23, 60].

Improving Wikipedia. Wikipedia content quality relies on the work of editors and their gradual improvement of articles [9, 46]. Automated or semi-automated tools [17, 39, 44] can help improve user experience [29, 69], content variety [42, 67], and quality [1, 20, 28]. The reliability of Wikipedia can also be improved automatically, e.g., by finding potential citations [15] and Wikipedia statements in need of evidence [48]. The insights from our work can help improve Wikipedia via new citations with which users would be more likely to interact.

Quantifying Web user engagement. User engagement is crucial for the success of Web services, and numerous researchers have focused on quantifying how Web users engage with online content, e.g., in computational advertising [6, 70], social media [5, 10, 22], or information retrieval [24, 55]. Also, while the body of work focusing on understanding readers' and editors' engagement with content *within* Wikipedia has been growing in the recent years [36], we study here for the first time how Wikipedia readers engage with the broader *outside* knowledge linked from the online encyclopedia.

3 CITATION DATA COLLECTION

To study readers' engagement with citations, we collected data capturing where readers navigate and how they interact with citations in English Wikipedia.

3.1 Background: Citations in Wikipedia

Articles in Wikipedia are written by editors in wikicode, a markup language that is then translated to HTML by MediaWiki, the software that powers the website. There are different ways to add citations to sources in the text, summarized below. In all cases, the full reference descriptions are rendered as footnotes at the bottom of the page (in a dedicated section called *References*) with an automatically assigned footnote number that is added as a link anchor (e.g., "[1]") in the text of the article wherever the reference is cited (Fig. 1). Most references in the *References* section consist of text including the title of the source, the authors' names, the year of publication, and the source's publisher. For 80% of Wikipedia references, the source title is actionable via a clickable link to the source. Also, when reading a page, hovering over a reference's footnote number with the mouse cursor will display a *reference tooltip*,[4] a pop-up containing the reference text and a clickable link (when present), e.g.,

[3] Notebooks with code at https://github.com/epfl-dlab/wikipedia-citation-engagement

[4] https://www.mediawiki.org/wiki/Reference_Tooltips

Daniel Nasaw (July 24, 2012). "Meet the 'bots' that edit Wikipedia". BBC News.

When readers click on the reference's footnote number, they are sent to the reference description at the page bottom, from where they can jump back to the locations where the reference is cited by clicking on a small icon (e.g., ^).

The most common method to add a reference to an article, also recommended by the Wikipedia guidelines, is via an inline citation using a `<ref/>` tag directly in the context where the reference is first cited. In the tag, the editors can specify the reference details (text and links) by using a predefined template or plain wikicode. In addition to this standard method, some references are added automatically by templates included in the page, such us the geolocations present in the infobox. It is worth noting that a reference can be cited multiple times by assigning it a name and appending the tag to every sentence that should link to it. Given the numerous ways to use the `<ref/>` tag, and in order to have an accurate view of the article, we parsed pages from wikicode to HTML and extracted the information from the HTML code.

3.2 Logging citation and page load events

We make use of Wikimedia's *EventLogging* tool,[5] an extension of the MediaWiki software that performs client-side logging of specific types of events. We detect 5 main types of citation-related events and 1 page load event. In terms of citations, we capture the mouse events that involve any kind of reader interaction with the references (see Fig. 1 for a visual explanation):

refClick: a click on a hyperlink in an article's reference section.
extClick: a click on an external link outside the reference section.
fnHover: a hover over a footnote number in the text, logged when the reference tooltip is visible for more than 1 second.
fnClick: a click on a footnote number, which takes the user to the reference section at the bottom of the page.
upClick: the inverse of *fnClick:* a click on a reference's up arrow icon that takes the reader back to the part of text where the reference is cited.
pageLoad: in addition to the above citation-related events, this event is triggered whenever a Wikipedia article is loaded.

The EventLogging platform manages a so-called *session token,* a cookie-based identifier that allows us to group events that happened within the same browser tab. We henceforth refer to event sequences that occur with the same session token as **sessions.**

We collected 4 contiguous weeks[6] of Wikipedia mobile and desktop traffic data of citation-related events. We repeated the 4-week data collection over two periods: from September 26 to October 25, 2018, and from March 24 to April 21, 2019. In both cases, we collected all citation-related events (*extClick*, *refClick*, *fnHover*, *fnClick*, *upClick*) and (due to computational infrastructure constraints) sampled *pageLoad* events at the session level at a rate of 33%.

To ensure that the logs reflect reader, rather than editor, behavior, we exclusively retained data from users who in the 4 weeks of data collection acted only as anonymous readers, discarding all events generated by Wikipedia editors (logged-in users or users with anonymous edits) and by bots (which can be filtered out using a detector provided by the EventLogging tool).

Throughout the paper, we will mostly focus on the data from the second data collection period (April 2019) and only use the October 2018 data for a longitudinal study measuring the impact of article quality on readers' engagement with citations.

[5] https://www.mediawiki.org/wiki/Extension:EventLogging/Guide
[6] We collected exactly 4 weeks to reduce potential seasonal effects due to uneven day-of-the-week frequencies.

3.3 Definition of engagement metrics

Two key metrics in our analysis will be the citation click-through rate (CTR) and the footnote hover rate.

For each page p and each session s, let $C(p,s)$ be the indicator function that is 1 if at least one reference was clicked on page p during session s by the respective user (*refClick* event), and 0 otherwise. Analogously, let $H(p,s)$ indicate if the user hovered over at least one footnote (*fnHover* event). Furthermore, let $N(p)$ be the number of sessions during which p was loaded (*pageLoad* event)

Global click-through rate. The global CTR measures overall reader engagement via reference clicks across Wikipedia. It is defined as the fraction of page views on which at least one reference click occurred (treating all views of the same page in the same session as one single event):

$$\text{gCTR} = \frac{\sum_p \sum_s C(p,s)}{\sum_p N(p)}, \tag{1}$$

where p ranges over the set of pages that contain at least one reference with a hyperlink.

Page-specific click-through rate. The page-specific CTR for page p is defined as the probability of observing at least one click on a reference in p during a session in which p was viewed:

$$\text{pCTR}(p) = \frac{\sum_s C(p,s)}{N(p)}. \tag{2}$$

Finally, we denote the average page-specific CTR over a set P of pages by

$$\text{pCTR}(P) = \frac{1}{|P|} \sum_{p \in P} \text{pCTR}(p). \tag{3}$$

Note that pCTR(P) corresponds to a macro average where every page gets the same weight, whereas gCTR corresponds to a micro average where pages are weighted in proportion to the number of sessions in which they were viewed.

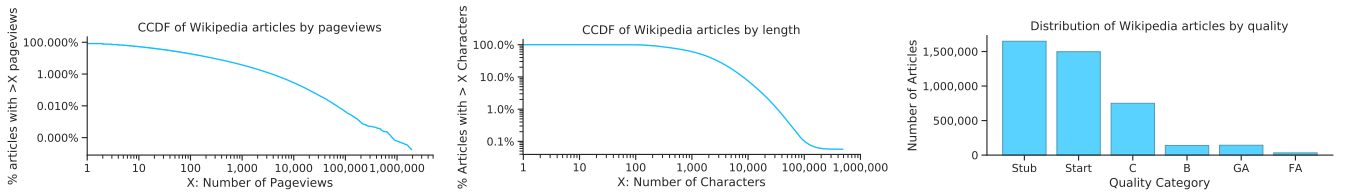
Footnote hover rates. In analogy to the above definitions, but when replacing the click indicator $C(p,s)$ with the hover indicator $H(p,s)$, we obtain the global and page-specific footnote hover rates:

$$\text{gHR} = \frac{\sum_p \sum_s H(p,s)}{\sum_p N(p)}, \qquad \text{pHR}(p) = \frac{\sum_s H(p,s)}{N(p)}. \tag{4}$$

3.4 Capturing event context

Each event is characterized by a set of features that capture information about three aspects of the event: the session in which the event happened, the page, and the reference.

Session: We collect the unique *session token* (cf. Sec. 3.2) that identifies the browser tab in which the event occurred.
Pages: At the article level, we store *title, page id, text length of wikicode in characters, number of references,* and *popularity* (number



(a) (b) (c)

Figure 2: Distribution of Wikipedia articles by (a) popularity (number of pageviews), (b) page length (number of characters in wikicode), and (c) quality (increasing from left to right; "GA" for "Good Article", "FA" for "Featured Article") (Sec. 3.5).

of *pageLoad* events during the data collection period). We also use the ORES *drafttopic* classifier [3] to label each Wikipedia article with a vector of *topics*, whose elements reflect the probability of the page to belong to one the 44 topics from the highest level of the WikiProjects taxonomy.[7] We further use the ORES *articlequality* model [20] to label articles with a *quality* level, which can take the following values (from low to high quality): "Stub", "Start", "C-class", "B-class", "Good Article", "Featured Article".

References: For each reference clicked or hovered, we record its *URL*, the *text in the reference*, the *text of the sentence* in which the reference is cited, and the *relative position* (character offset from the start in plain text, divided by page length) in the page where the reference is cited. Since we associate references to their contexts, references to the same source appearing on different pages are treated as distinct.

Wikipedia is dynamic by nature: articles are continuously updated, and their changes are tracked through revisions. To account for the evolution of articles over the 4 weeks of data collection, we aggregate individual revision-level metrics at the article level. To compute article-specific characteristics such as article length or number of references, we calculate their average over all revisions from the logging period. To quantify the amount of reader engagement with a given article (e.g., page loads, reference clicks), we sum all events recorded at each revision of the article.

3.5 General statistics of English Wikipedia

By the end of the data collection, English Wikipedia contained 5.8M articles, 5.4M (95%) of which were loaded at least once in our data sample, in a total of 7.4M revisions. Out of these articles, 3.9M (73%) contain at least one citation, linking to a total of 24M distinct URLs.

Over the 4 weeks of data collection, we collected (at a 33% sampling rate) 1.5B *pageLoad* events (62% from the mobile site and the rest from the desktop site). In Fig. 2a we report the (complementary cumulative) popularity distribution for the Wikipedia pages that were viewed at least once during the data collection period. The distribution is heavily skewed, with approximately 83% of the articles loaded fewer than 100 times in the 33% random sample (cf. Sec. 3.2), or fewer than 300 times when extrapolating to all data.

We observe a similar uneven distribution of page length (Fig. 2b), with the majority of articles being very short.

Fig. 2c shows that the distribution of article quality levels is also heavily skewed toward low quality levels: most articles are identified as "Stub" or "Start", and fewer than 300K articles are marked as "Good" or "Featured" articles.

[7] https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Council/Directory

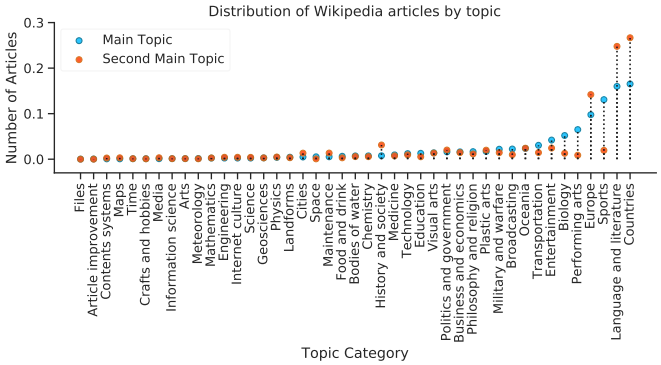


Figure 3: Distribution of most and second most prominent Wikipedia article topics (Sec. 3.5).

Finally (Fig. 3), we find that a majority of articles are about geography or "Language and literature" (the latter including biographies), followed by topics related to sports and science.

4 RQ1: PREVALENCE OF CITATION INTERACTIONS

After these preliminaries, we are now ready to address our first research question, which asks to what extent Wikipedia readers engage with citations.

4.1 Distribution of interaction types

We start by analyzing the relative frequency of the different citation events, as defined in Sec. 3.2. Over the month of data collection, we captured a total of 96M citation events. Fig. 4 shows how these events distribute over the 5 event types, broken down by device type (mobile vs. desktop). We observe that most interactions with citations happen on desktop rather than mobile devices, despite the fact that the majority of page loads (62%) are made from mobile.

The interactions also distribute differently across types for mobile vs. desktop. The by far prevailing event on desktop is hovering over a footnote (*fnHover*) in order to display the reference text. Hovering requires a mouse, which is not available on most mobile devices, which in turn explains the low incidence of *fnHover* on

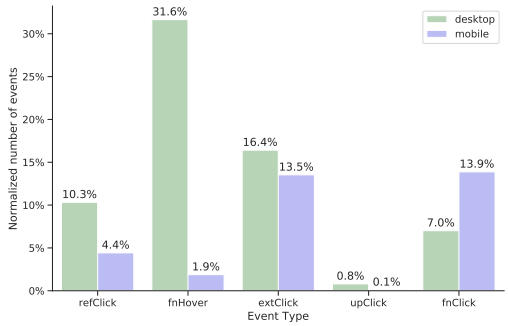


Figure 4: Relative frequency of citation-related events (Sec. 3.2), split into desktop (green, left bars) and mobile (blue, right bars) in April 2019 (Sec. 4.1).

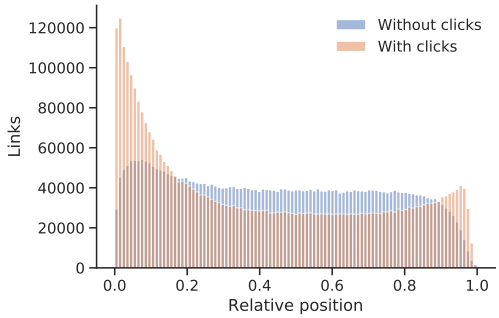


Figure 5: Relative position in page of clicked vs. unclicked references, for references with hyperlinks (Sec. 4.3).

mobile. In order to reveal the reference text behind a footnote, mobile users instead need to click on the footnote, which presumably explains why *fnClick* is the most common event on mobile.

Clicking external links outside of the *References* section at the bottom of the page (*extClick*) is the second most common event on both desktop and mobile, followed by clicks on citations from the *References* section (*refClick*). Finally, the *upClick* action, which lets users jump back from the *References* section to the locations where the citation is used in the main text, is almost never used.

4.2 Citation click-through rates

We now focus on the two prevalent interactions with citations, hovering over footnotes (*fnHover*) and leaving Wikipedia by clicking on citation links (*refClick*). (We do not dwell on *extClick* events, as they do not concern citations but other external links; cf. Sec. 3.2.)

First, we observe that, out of the 24M distinct URLs that are cited across all articles in English Wikipedia, 93% of the URLs are never clicked during our month of data collection.

Next, we note that the global click-through rate (CTR) across all pages with at least one citation (gCTR, Eq. 1) is 0.29%; i.e., clicks on references happen on fewer than 1 in 300 page loads. Breaking the analysis up by device type, we observe again substantial differences between desktop and mobile: on desktop the global CTR is 0.56%, over 4 times as high as on mobile, where it is only 0.13%.

The average page-specific CTR (pCTR, Eq. 3) is higher, at 1.1% for desktop and 0.52% for mobile. This is due to the fact that there are many rarely viewed pages (cf. Fig. 2a) with a noisy, high CTR. After excluding pages with fewer than 100 page views, the global CTR is 0.67% on desktop, and 0.21% on mobile.

Engagement via footnote hovering is slightly higher, at a global footnote hover rate (gHR, Eq. 4) of 1.4%. The average page-specific footnote hover rate (pHR, Eq. 4) is 0.68% when including all pages with at least one clickable reference, and 1.1% when excluding pages with fewer than 100 page views.[8]

Given these numbers, we conclude that readers' engagement with citations is overall low.

[8] As mentioned in Sec. 4.1, hovering is not available on most mobile devices, so the hovering numbers pertain to desktop devices only.

4.3 Positional bias

Previous work has shown that users are more likely to click Wikipedia-internal links that appear at the top of a page [42]. To verify whether this also holds true for references, we sample one random page load with citation interactions per session and randomly sample one clicked and one unclicked reference for this page load. We then compute each reference's relative position in the page as the offset from the top of the page divided by the page length (in characters). Fig. 5, which shows the distribution of the relative position for clicked and unclicked references, reveals that users are more likely to click on references toward the top and (less extremely so) the bottom of the page.

4.4 Top clicked domains

Next, we investigate what are the most frequent domains at which users arrive upon clicking a citation.

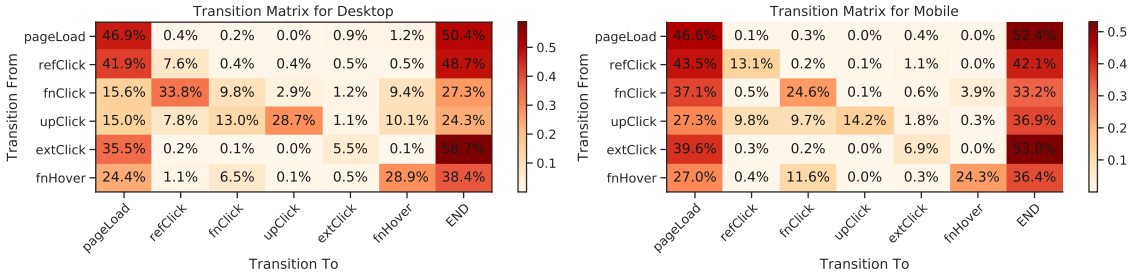
Initially, we found that the most frequently clicked domain is `archive.org` (Internet Archive), with 882K *refClick* events. Such URLs are usually snapshots of old Web pages archived by the Internet Archive's Wayback Machine. To handle such cases, we extract the original source domains from wrapping `archive.org` URLs.

In Fig. 7 we report the top 15 domains by number of *refClick* events. The most clicked domain is `google.com`. Drilling deeper, we checked the main subdomains contributing to this statistic, finding that a significant proportion of clicks goes to `books.google.com`, which is providing partial access to printed sources. The second most clicked domain is `doi.org`, the domain for all scholarly articles, reports, and datasets recorded with a Digital Object Identifier (DOI), followed by (mostly liberal) newspapers (*The New York Times*, *The Guardian*, etc.) and broadcasting channels (BBC).

4.5 Markovian analysis of citation interactions

Whereas the above analyses involved individual events, we now begin to look at *sessions:* sequences of events that occurred in the same browser tab (as indicated by the session token; Sec. 3.2). Every session starts with a *pageLoad* event, and we append a special *END*event after the last actual event in each session.

By counting event transitions within sessions, we construct the first-order Markov chain that specifies the probability $P(j|i)$ of observing event j right after event i, where i and j can take values from



(a) Reader behavior on desktop devices. **(b) Reader behavior on mobile devices.**

Figure 6: Transition matrices of first-order Markov chains for (a) desktop devices and (b) mobile devices, aggregating reader behavior with respect to citation events when navigating a Wikipedia article with references (Sec. 4.5).

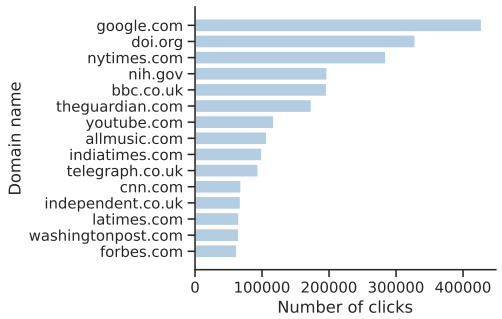


Figure 7: Top 15 domain names appearing in English Wikipedia references (Sec. 4.4), sorted by number of clicks received during April 2019.

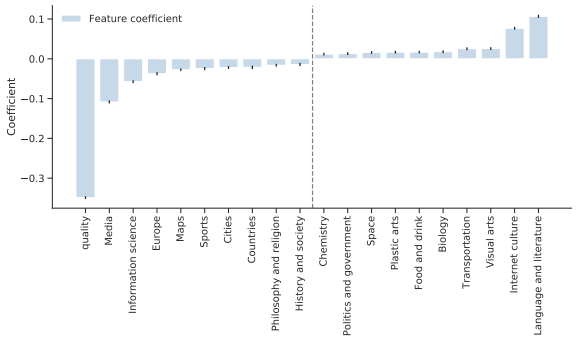


Figure 8: Contribution of features to logistic regression model predicting if *refClick* event will eventually occur after page load (Sec. 5.1). Top 10 positive and negative coefficients shown, with 95% CIs.

the event set introduced in Sec. 3.2 (*pageLoad*, *refClick*, *extClick*, *fnClick*, *upClick*, *fnHover*) plus the special *END*event.

The transition probabilities are reported in Fig. 6. We observe that most reading sessions are made up of page views only: on both desktop and mobile, after loading a page, readers tend to end the session (with a probability of around 50%) or load another page in the same tab (47%). All citation-related events have a very low probability (at most 1.2%) of occurring right after loading a page.

On desktop, reference clicks become much more likely after footnote clicks (34%), and footnote clicks in turn become much more likely after footnote hovers (6.5%), hinting at a common 3-step motif (*fnHover*, *fnClick*, *refClick*), where the reader engages ever more deeply with the citation. Note, however, that this is not true for mobile devices, where, even after readers clicked on a footnote, the probability of also clicking on the citation stays low (0.5%).

Finally, reference clicks (*refClick*) are also common immediately after other reference clicks (8% on desktop, 13% on mobile). Note that for external links outside of the *References* section (*extClick*) we see a different picture: such external clicks are only rarely followed by interactions with citations (*fnHover*, *fnClick*, *refClick*), and in the majority of cases (59% on desktop, 53% on mobile) they conclude the session, suggesting that Wikipedia is in these cases commonly used as a gateway to external websites.

5 RQ2: PAGE-LEVEL ANALYSIS OF CITATION INTERACTIONS

We now proceed to our second research question, which asks what features of a Wikipedia page predict whether readers will engage with the references it contains.

5.1 Predictors of reference clicks

As a first step, we perform a regression analysis. We train a logistic regression classifier for predicting whether a given *pageLoad* event will eventually be followed by a *refClick* event. To assemble the training set, we first find sessions with at least one (positive) *pageLoad* followed by a *refClick* and at least one (negative) *pageLoad* not followed by a *refClick*, and make sure to include at most one such pair per session in order to avoid over-representing power users with extensive sessions. The dataset totals 938K pairs, which we split into 80% for training and 20% for testing.

As predictors we use the article's *topic* vector (with entries from [0, 1]; Sec. 3.4) and the *quality* label (Sec. 3.4), which we also normalize to a score in the range [0, 1] using the mapping from a previous study [20]. We did not use the number of references and the length of the page, as they are important features in the quality model and

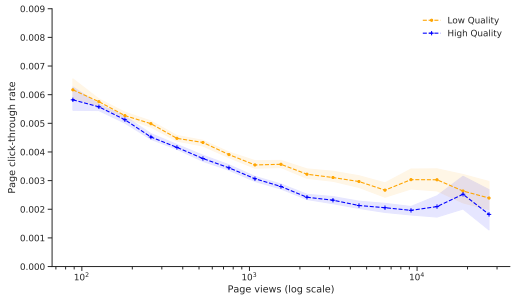


Figure 9: Comparison of page-specific click-through rate for low- (yellow) vs. high-quality (blue) articles, as function of popularity (Sec. 5.2). Error bands: bootstrapped 95% CIs.

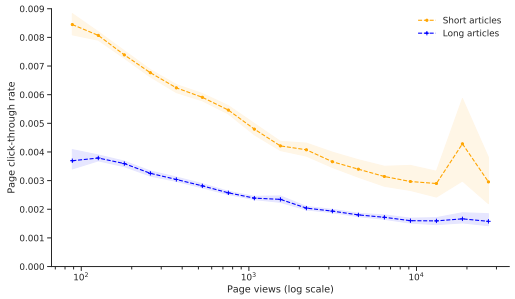


Figure 10: Comparison of page-specific click-through rate for short (yellow) vs. long (blue) articles, as function of popularity (Sec. 5.3). Error bands: bootstrapped 95% CIs.

would cause collinearity issues due to their high correlation with quality (Pearson's correlation 0.81 and 0.75, respectively).

The resulting regression model has an area under the ROC curve (AUC) of 0.6 on the testing set. A summary of the 10 most predictive positive and negative coefficients is given in Fig. 8. By far the most important predictor—with a large negative weight—is the article's quality. Moreover, some topics are positive predictors (e.g., "Language and literature", which also includes all biographies, as well as "Internet culture"), while others are negative predictors (e.g., "Media", "Information science").

Given the importance of the quality feature in this first analysis, we now move to investigating its role in a more controlled study.

5.2 Effects of page quality

To come closer to a causal understanding of the impact of an article's quality on readers' clicking citations in the article, we perform a matched observational study. The ideal goal would be to compare the page-specific CTR (Eq. 2) for pairs of articles—one of high, the other of low quality—that are identical in all other aspects.

Propensity score. Finding such exact matches is unrealistic in practice, so we resort to propensity score matching [4], which provides a viable solution. The propensity score specifies the probability of being treated as a function of the observed (pre-treatment) covariates. Crucially, data points with equal propensity scores have the same distribution over the observed covariates, so matching treated to untreated points based on propensity scores will balance the distribution of observed covariates across treatment groups.

In our setting, we define being of high quality as the treatment and estimate propensity scores via a logistic regression that uses topics, length, number of citations, and popularity as observed covariates in order to predict quality as the binary treatment variable. We consider as low-quality all articles tagged as *Stub* or *Start* (74% of the total; Fig. 2c), and as high-quality the rest. Articles without a *refClick* or fewer than 100 *pageLoad* events are discarded in order to avoid noisy estimates of the page-specific CTR. This leaves us with 854K articles.

Matching. We compute a matching (comprising 198K pairs) that minimizes the total absolute difference of within-pair propensity scores, under the constraint that the length of matched pages should not differ by more than 10%. This constraint is necessary to ascertain balance on the page length feature because page length is so highly correlated with quality (Pearson correlation 0.81; cf. Sec. 5.1). After matching, we manually verify that all observed covariates, including page length, are balanced across groups.

Results. Fig. 9 visualizes the average page-specific CTR for articles of low (yellow) and high (blue) quality as a function of article popularity. We can observe that the CTR of low-quality articles significantly surpasses that of high-quality articles across all levels of popularity. In interpreting this result, it is important to recall that page length is one of the most important features in ORES [20], the quality-scoring model we use here. As we control for page length, the gap observed in Fig. 9 may be attributed to the remaining features used by ORES, such as the presence of an infobox, the number of images, and the number of sections and subsections.

We hence dedicate our next, final page-level analysis to estimating the impact of page length alone on page-specific CTR.

5.3 Effects of page length

In order to measure the effect of page length on CTR, we take a two-pronged approach, first via a cross-sectional study using propensity scores, and second via a longitudinal study.

Cross-sectional study. First, we conduct a matched study based on propensity scores analogous to Sec. 5.2, but now with page length as the treatment variable (using the longest and the shortest 40% of articles as treatment groups), and all other features (except quality) as observed covariates. Matching yields 683K pairs, and we again manually verify covariate balance across treatment groups.

The average page-specific CTR of short articles (0.68%) is more than double that of long articles (0.27%; $p \ll 0.001$ in a two-tailed Mann–Whitney U test). Moreover, as seen in Fig. 10, this relative difference obtains across all levels of article popularity.

Longitudinal study. While in the above cross-sectional study propensity score matching ensures that the covariates of long vs. short articles are indistinguishable at the aggregate treatment group level, it does not necessarily do so at the pair level. Also, we did not include as observed covariates features describing the users who read the respective articles, and it might indeed be the case that users with a liking for short, niche articles also have a higher probability of clicking citations. In order to mitigate the danger of such remaining potential confounds and achieve even finer control,

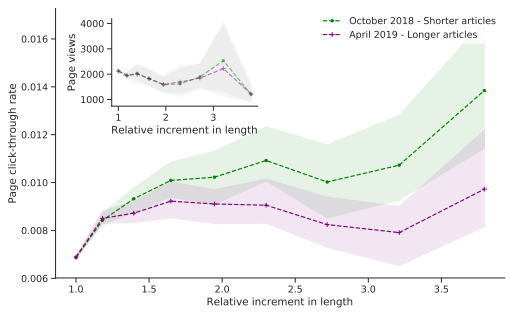


Figure 11: Comparison of page-specific click-through rate of shorter (green) vs. longer (purple) revisions of identical articles, as function of length ratio (Sec. 5.3). Inset: popularity as function of length ratio. Error bands: bootstrapped 95% CIs.

we now conduct a longitudinal study to assess how a variation in length of the *same* article impacts its CTR.

To do so, we select all articles that grew in length between October 2018 and April 2019, our two data collection periods (Sec. 3.2). To control for the effect of page popularity, which was observed to negatively correlate with CTR (Fig. 9 and 10), we assign a popularity level to each article by binning page view counts into deciles and discard articles whose popularity level has changed between the two periods. This way, we obtain a set of 120K articles with matched long and short revisions.

By grouping these articles by the length ratio of their two revisions and plotting this ratio against the CTR for the long (purple) vs. short (green) versions (Fig. 11), we provide a further strong indicator that page length causally decreases the prevalence of citation clicking. According to a Mann–Whitney U test, the CTR difference between long and short revisions is statistically significant with $p < 0.05$ starting from a length increase of 17%, and with $p < 0.01$ from 31%. In addition, to verify that the effect is not confounded by a concomitant change in article popularity, the inset plot in Fig. 11 shows that the popularity indeed stays constant between revisions.

6 RQ3: LINK-LEVEL ANALYSIS OF CITATION INTERACTIONS

Our final research question asks which features of a specific reference predict if readers will engage with it. Note that this is different from RQ2 (Sec. 5), where we operated at the page level and did not differentiate between different references on the same page.

6.1 Predictors of reference clicks

We begin with a regression analysis to detect which features predict whether a reference will be clicked. We selected all the references with external links, and we carefully rule out a host of confounds by sampling pairs of clicked and unclicked references from the same page view, thus controlling for situational features such as the page, user, information need, etc. As we saw in Fig. 5, references at the top and bottom of pages are *a priori* more likely to be clicked. Thus, to exclude position as a confound and maximize the probability that the user saw both references in a pair, we pick as the unclicked reference in a pair the one that appears closest in the page to the clicked reference. To make sure we sample references associated with a sentence, we discard all footnotes in tables, infoboxes, and images, and keep only those within the article text. Finally, we again sample only one pair per session in order to avoid over-representing readers who are more prone to click on references. This process yields 1.8M reference pairs.

	Positive contribution				Negative contribution			
	In sentence		In reference		In sentence		In reference	
	Word	Coeff.	Word	Coeff.	Word	Coeff.	Word	Coeff.
All topics	greatest	0.36	know	0.25	debut	-0.25	awards	-0.33
	born	0.28	pmc	0.24	moved	-0.16	deadline	-0.32
	died	0.23	2019	0.21	worked	-0.16	billboard	-0.17
	website	0.23	website	0.21	awarded	-0.16	register	-0.17
	ranked	0.23	dies	0.20	joined	-0.13	link	-0.16
	known	0.20	former	0.19	began	-0.13	isbn	-0.15
	professional	0.19	family	0.16	appeared	-0.12	board	-0.14
	relationship	0.19	behind	0.15	score	-0.11	variety	-0.14
	rating	0.18	allmusic	0.15	festival	-0.11	next	-0.14
	article	0.18	story	0.15	attended	-0.11	archive	-0.13
STEM	online	0.25	definition	0.30	requirements	-0.17	oclc	-0.26
	tests	0.23	2019	0.24	run	-0.17	best	-0.23
	2019	0.23	free	0.22	rather	-0.16	jstor	-0.22
	short	0.17	pmc	0.21	another	-0.15	evaluation	-0.16
	known	0.17	website	0.20	said	-0.15	wiley	-0.16
	algorithms	0.16	pdf	0.19	launched	-0.15	london	-0.15
	published	0.16	overview	0.17	less	-0.14	isbn	-0.14
	defined	0.15	methods	0.15	make	-0.12	internet	-0.14
	programming	0.15	introduction	0.14	better	-0.12	industrial	-0.14
	digital	0.15	years	0.13	popular	-0.12	source	-0.14
Culture	article	0.30	daughter	0.36	indicating	-0.42	awards	-0.36
	born	0.28	obituary	0.31	premiered	-0.28	award	-0.33
	greatest	0.27	know	0.31	chart	-0.21	deadline	-0.28
	professional	0.27	instagram	0.29	debut	-0.21	cast	-0.22
	died	0.26	boy	0.28	moved	-0.20	global	-0.21
	known	0.25	sex	0.25	began	-0.17	next	-0.19
	ranked	0.24	wife	0.24	earned	-0.16	isbn	-0.18
	relationship	0.23	former	0.24	recorded	-0.16	drama	-0.18
	website	0.23	historic	0.24	alongside	-0.16	standard	-0.18
	sexual	0.23	2019	0.23	worked	-0.16	tour	-0.18
History and Society	born	0.29	definition	0.43	came	-0.20	jstor	-0.25
	website	0.21	overview	0.22	award	-0.16	record	-0.21
	2019	0.21	best	0.19	transportation	-0.13	link	-0.20
	died	0.20	2019	0.19	protection	-0.12	2002	-0.17
	currently	0.19	website	0.19	member	-0.12	election	-0.16
	known	0.17	statistics	0.17	began	-0.11	1998	-0.15
	referred	0.17	death	0.16	originally	-0.11	ed	-0.15
	customers	0.16	last	0.16	specific	-0.11	isbn	-0.15
	study	0.16	ship	0.15	awarded	-0.10	announces	-0.14
	activities	0.15	top	0.15	addition	-0.10	board	-0.12
Geography	politician	0.50	woman	0.34	debut	-0.45	crime	-0.28
	born	0.26	know	0.27	missing	-0.22	awards	-0.28
	magazine	0.25	dies	0.26	career	-0.21	register	-0.24
	believed	0.23	family	0.23	timmothy	-0.20	link	-0.24
	married	0.23	website	0.20	executive	-0.19	interview	-0.19
	ranked	0.22	mail	0.19	episode	-0.17	2000	-0.17
	video	0.22	father	0.18	months	-0.17	culture	-0.17
	directed	0.18	son	0.18	close	-0.15	htm	-0.16
	crime	0.18	boy	0.18	case	-0.15	music	-0.15
	natural	0.18	biography	0.17	appointed	-0.15	paris	-0.15

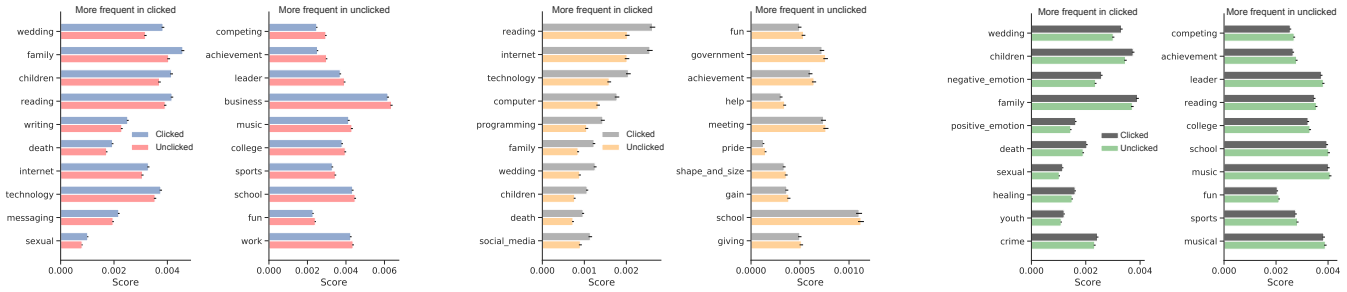
Table 1: Top positive and negative predictors (words) of reference clicks (Sec. 6.1), for different article topics. Words are organized based on where they appear: in the sentence annotated by the reference, or in the reference text.

As predictors we use the words in the sentence that cites the respective reference, as well as the words in the reference text (cf. Sec. 3.1), represented as binary indicators specifying for each of the 1K most frequent words whether the word appears in the sentence.[9] Using these features as predictors, we train a logistic regression to predict the binary click indicator.

We perform this analysis on the full above-described dataset, as well as on subsets consisting only of page views from each of 4 broad categories (derived by aggregating the 44 WikiProjects categories from Sec. 3.4): "Culture" (1.3M pairs), "STEM" (436K), "Geography" (530K), and "History and Society" (467K). The model achieves a testing AUC of around 0.55 across these 5 settings.

The words with the largest and smallest coefficients are displayed in Table 1, where we observe that, for all article topics except for

[9]Stop words were removed, and numbers (except for 4-digit numbers that potentially represent years) were converted to a special number token.



(a) Click event (sentence text) **(b) Click event (reference text)** **(c) Hover event (sentence text)**

Figure 12: *Empath* [14] topics most strongly (anti-)associated with citation events (cf. Sec. 6.2 for description). Reference text not studied for hover event (Sec. 6.3) because unlikely to be visible to user before hovering.

"STEM", many positive features are related to social and life events and relationships ("dies", "obituary", "married", "wife", "relationship", "sex", "daughter", "family", etc.). Another common pattern across topics is that "2019" is strongly related with clicking, and that career-related references ("awards", "debut", etc.) are less likely to be clicked. We shall further discuss these observations in Sec. 7.

On STEM-related pages, open-access references seem to receive more clicks than others, with words like "free" and "pdf" among the top predictors, whereas words related to traditionally closed-access libraries such as JSTOR appear among the negative predictors, in line with previous findings [58].

6.2 Topical correlates of reference clicks

For a higher-level view, we perform a topical analysis of citing sentences and reference texts, separately for the clicked vs. the unclicked references from the paired dataset of Sec. 6.1.

To extract topics, we use *Empath* [14], which comes with a pre-trained model for labeling input text with a distribution over 200 wide-ranging topics. After applying the model to each data point, we compute the average topic distribution for clicked and unclicked references, respectively, and sort topics by the signed difference between their probability for clicked vs. unclicked references.

The topics with the largest positive and negative differences are listed in Fig. 12a and 12b for citing sentences and reference texts, respectively. The results corroborate those from Sec. 6.1, with human factors (wedding, family, sex, death) being more prominent among clicked references, whereas career-related topics such as competitions or achievements receive less attention. Among the most prominent topics for reference texts (Fig. 12b), topics related to technology and the Internet also emerge.

6.3 Predictors of footnote hovering

The analyses of Sec. 6.1 and 6.2 considered engagement via reference clicks. As we observed in Fig. 4, on desktop devices, hovering over a footnote to reveal the reference text in a tooltip is an even more common way to interact with references. We hence replicated the above analyses with the *fnHover* instead of the *refClick* event (8.7M reference pairs), with the only difference that we excluded words from reference texts as features, since the user is unlikely to have seen those words before hovering over the footnote.

Positive		Negative	
Word	Coeff.	Word	Coeff.
killer	0.16	oclc	-0.22
greatest	0.16	jason	-0.16
critic	0.15	episode	-0.15
things	0.15	die	-0.15
daughter	0.15	dictionary	-0.13
reveals	0.14	spanish	-0.12
baby	0.14	isbn	-0.12
instagram	0.13	le	-0.11
wife	0.13	board	-0.11
sheet	0.13	channel	-0.11

Table 2: Top 10 positive and negative predictors (words) of reference click following footnote hover (Sec. 6.4).

The results echo those of Sec. 6.1 and 6.2, so for space reasons we do not discuss the regression analysis for footnote hovering (cf. Sec. 6.1) and focus on the topical analysis instead (cf. Sec. 6.2). Inspecting Fig. 12c, we observe that we see a stronger tendency of *fnHover* events, compared to *refClick* events, to be elicited by words that are related to both positive and negative emotions.

6.4 Predictors of reference clicks after hovering

Once a user hovers over a (*fnHover*), the text of the corresponding reference is revealed in a so-called reference tooltip (Fig. 1). At this point, the user has the choice to either click through to the citation URL (*refClick*) or to stay on the article page. In the final analysis of the paper, we are interested in understanding what words in the reference text influence the user when making this decision.

We create a dataset by selecting the page loads with at least two footnote hover events, where one converted to a *refClick* (positive), whereas the other did not (negative). As in the previous studies, we selected at most one random pair per session, giving rise to a dataset of 440K pairs of hover events.

Similar to the study in Sec. 6.1, we represent reference texts as 1K-dimensional word indicator vectors and use them as predictors in a logistic regression to predict *refClick* events (testing AUC 0.54).

The strongest coefficients are summarized in Table 2, painting a picture consistent with the previous analyses: readers, after seeing a reference preview via the tooltip, are more likely to click on the cited link when the reference text mentions social and life aspects ("wife", "baby", "instagram", etc.). The strongest negative coefficients suggest that readers tend to not click through to dictionary entries, book

catalogs (ISBN, OCLC), and information in languages other than English: manual inspection revealed that "spanish" is mainly due to the note "In Spanish", "le" is the French article common in French newspaper names (e.g., *Le Monde*), and "die" is a German article.

7 DISCUSSION AND CONCLUSIONS

Our analysis provides important insights regarding the role of Wikipedia as a gateway to information on the Web. We found that **in most cases Wikipedia is the final destination** of a reader's journey: fewer than 1 in 300 page views lead to a citation click. In our analysis, we focused on the fraction of users who engage with references, and characterized **how Wikipedia is used as a gateway to external knowledge.** Our findings suggest the following.

- **We engage with citations in Wikipedia when articles do not satisfy our information need.** Sec. 5 showed that readers are more likely to click citations on shorter and lower-quality articles. Although this result seemed counter-intuitive at first, since higher-quality articles actually contain *more* references that could potentially be clicked, it is in line with the finding that citations to sources reporting atomic facts that are typically available in Wikipedia articles (e.g., awards, career paths), are also generally less engaging (Sec. 6). Collectively, these results suggest that readers are inclined to seek content beyond Wikipedia when the encyclopedia itself does not satisfy their information needs.
- **Citations on less engaging articles are more engaging.** In all of Sec. 5 we found that citation click-through rates decrease with the popularity of an article. While this may follow from the previous point because long, high-quality articles tend to be more popular, it may also suggest that less popular articles are visited with a specific information need in mind. Previous work indeed suggests that popular articles are more likely to be viewed by users who are randomly exploring the encyclopedia [53].
- **We engage with content about people's lives.** We clearly saw that readers' interest is particularly high in references about people and their social and private lives (Sec. 6). This is especially true for hovers, a less cognitively demanding form of engagement with citations. Hover events are also more likely to be elicited by words that are related to emotions, both positive and negative.
- **Recent content is more engaging.** We found that references about recent events (whose text includes "2019") are more engaging, both in terms of hovering and clicking.
- **Open content is more engaging.** Finally, we saw that references in Wikipedia pages about science and technology, especially if they point to a open-access sources (e.g., having "free" or "pdf" in the reference text), are also more likely to be clicked.

Theoretical implications. Our findings furnish novel insights about Web users and their information needs through the lens of the largest online encyclopedia. For the first time, by characterizing Wikipedia citation engagement, we are able to quantify the value of Wikipedia as a gateway to the broader Web. Our findings enable researchers to develop novel theories about readers' information needs and the possible barriers separating knowledge within and outside of the encyclopedia. Our research can also guide the broader community of Web contributors in prioritizing efforts towards improving information reliability: we found that people especially rely on cited sources when seeking information about recent events and biographies, which suggests that Web content in these areas should be especially well curated and verified. Finally, the fact that readers engage more with freely accessible sources highlights the importance of open access and open science initiatives.

Practical implications. Quantifying Wikipedia article completeness has proven to be a non-trivial task [45]. The notion that article completeness is highly related to readers' engagement with Wikipedia references opens up ideas for novel applications to help satisfy Web users' information needs, including models that quantify lack of information in an article by incorporating signals related to reference click-through rate. Our findings will also help prioritize areas of content to be checked for citation quality by Wikipedia editors: in areas of content where Wikipedia acts as a major gateway, the quality and reliability of sources that readers visit become even more crucial. Finally, the data we collected could empower a model that, given a sentence missing a citation (i.e., with a *citation needed* tag), could quantify how likely readers are to be interested in accessing the corresponding information and thereby help Wikipedia editors prioritize the backlog of unsolved missing-reference cases.

Limitations and future work. The overall low AUC (0.54 to 0.6) of the regression models (Sec. 5–6) emphasizes the inherent unpredictability of reader behavior. While the significantly above-chance performance renders the models useful for analyzing the impact of various predictors, their performance is currently too low to make them useful as practical predictive tools. Future work should hence invest in more powerful sequence models to improve accuracy.

By focusing on English Wikipedia only, the present analysis provides a limited view of the broader Wikipedia project, which is available in almost 300 languages and accessed by users all over the world. In our future work, we therefore plan to replicate this study for other language editions. So far, we also omitted any user characteristics from our study, such as more global behavioral traits beyond the page-view level, as well as geographic information, which are known to play an important role in user behavior [32, 57]. Future work should incorporate such signals.

We will also investigate reader intents more closely. While click and hover logs reflect the extent to which readers are interested in knowing more about a given topic, they cannot tell us about the specific circumstances that led the user to engage by clicking or hovering, nor about the level of satisfaction achieved by following up on a reference. In the future, we plan to better understand these aspects via qualitative methods such as surveys and interviews.

Further, whereas our analysis focused on links in the *References* section of articles, future work should also study other types of external links (cf. Fig. 1) in satisfying readers' information needs.

Finally, as exogenous events strongly affect Wikipedia users' information needs [53], future work should go beyond studying Wikipedia as an isolated platform and analyze how citation interaction patterns are warped by breaking news and events with uncertain information. This will sharpen our picture of Wikipedia as a gateway to global information.

ACKNOWLEDGMENTS

We thank Leila Zia, Michele Catasta, Dario Taraborelli for early contributions; Bahodir Mansurov, WMF Analytics for help with event logging; James Evans for good discussions; Microsoft, Google, Facebook, SNSF for supporting West's lab.

REFERENCES

[1] B. Thomas Adler, Krishnendu Chatterjee, Luca de Alfaro, Marco Faella, Ian Pye, and Vishwanath Raman. 2008. Assigning trust to Wikipedia content. In *Proc. 4th International Symposium on Wikis.*

[2] Ofer Arazy, Hila Liifshitz-Assaf, Oded Nov, Johannes Daxenberger, Martina Balestra, and Coye Cheshire. 2017. On the "How" and "Why" of emergent role behaviors in Wikipedia. In *Proc. ACM Conference on Computer Supported Cooperative Work and Social Computing.*

[3] Sumit Asthana and Aaron Halfaker. 2018. With few eyes, all hoaxes are deep. *Proc. ACM Conference on Computer Supported Cooperative Work and Social Computing.*

[4] Peter C Austin. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research* 46, 3 (2011), 399–424.

[5] Saeideh Bakhshi, David A Shamma, and Eric Gilbert. 2014. Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proc. SIGCHI conference on human factors in computing systems.*

[6] Nicola Barbieri, Fabrizio Silvestri, and Mounia Lalmas. 2016. Improving post-click user engagement on native ads via survival analysis. In *Proc. International Conference on World Wide Web.*

[7] Ivan Beschastnikh. 2008. Wikipedian Self-Governance in action: Motivating the policy lens. In *Proc. International AAAI Conference on Web and Social Media.*

[8] Xiaoxi Chelsy Xie, Isaac Johnson, and Anne Gomez. 2019. Detecting and gauging impact on Wikipedia page views. In *Proc. International Conference on World Wide Web.*

[9] Chih-Chun Chen and Camille Roth. 2012. {{Citation needed}}: the dynamics of referencing in Wikipedia. In *Proc. Annual International Symposium on Wikis and Open Collaboration.*

[10] Jörg Claussen, Tobias Kretschmer, and Philip Mayrhofer. 2013. The effects of rewarding user engagement: The case of Facebook apps. *Information Systems Research* 24, 1 (2013), 186–200.

[11] William Cronon. 2012. Scholarly authority in a Wikified world. *Perspectives in History* (2012).

[12] Alexander Dallmann, Thomas Niebler, Florian Lemmerich, and Andreas Hotho. 2016. Extracting semantics from random walks on Wikipedia: comparing learning and counting methods. In *Proc. The Workshops of the Tenth International AAAI Conference on Web and Social Media.*

[13] Dimitar Dimitrov and Florian Lemmerich. 2019. Different topic, different traffic: How search and navigation interplay on Wikipedia. *The Journal of Web Science* 6 (2019).

[14] Ethan Fast, Binbin Chen, and Michael S. Bernstein. 2016. Empath: Understanding topic signals in large-scale text. In *Proc. CHI Conference on Human Factors in Computing Systems.*

[15] Besnik Fetahu, Katja Markert, Wolfgang Nejdl, and Avishek Anand. 2016. Finding news citations for Wikipedia. In *Proc. Conference on Information and Knowledge Management.*

[16] Andrea Forte, Vanesa Larco, and Amy Bruckman. 2009. Decentralization in Wikipedia governance. *Journal of Management Information Systems* 26, 1 (2009), 49–72.

[17] R. Stuart Geiger and Aaron Halfaker. 2013. When the levee breaks: without bots, what happens to Wikipedia's quality control processes?. In *Proc. International Symposium on Open Collaboration.*

[18] Patrick Gildersleve and Taha Yasseri. 2018. Inspiration, captivation, and misdirection: Emergent properties in networks of online navigation. In *Complex Networks IX.* 271–282.

[19] Casper Grathwohl. 2011. Wikipedia comes of age. *Chronicle of Higher Education* 57 (2011).

[20] Aaron Halfaker. 2017. Interpolating quality dynamics in Wikipedia and demonstrating the Keilana effect. In *Proc. International Symposium on Open Collaboration.*

[21] Denis Helic, Markus Strohmaier, Michael Granitzer, and Reinhold Scherer. 2013. Models of human navigation in information networks based on decentralized search. In *Proc. ACM Conference on Hypertext and Social Media.*

[22] Yuheng Hu, Shelly Farnham, and Kartik Talamadupula. 2015. Predicting user engagement on twitter with real-world events. In *Proc. International AAAI Conference on Web and Social Media.*

[23] Dariusz Jemielniak and Eduard Aibar. 2016. Bridging the gap between Wikipedia and academia. *Journal of the Association for Information Science and Technology* 67, 7 (2016), 1773–1776.

[24] Yushi Jing, David Liu, Dmitry Kislyuk, Andrew Zhai, Jiajing Xu, Jeff Donahue, and Sarah Tavel. 2015. Visual search at Pinterest. In *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

[25] Brian Keegan, Darren Gergle, and Noshir Contractor. 2011. Hot off the wiki: dynamics, practices, and structures in Wikipedia's coverage of the Tōhoku catastrophes. In *Proc. WikiSym - International Symposium on Wikis and Open Collaboration.*

[26] Tobias Koopmann, Alexander Dallmann, Lena Hettinger, Thomas Niebler, and Andreas Hotho. 2019. On the right track! Analysing and predicting navigation success in Wikipedia. In *Proc. ACM Conference on Hypertext and Social Media.*

[27] Kayvan Kousha and Mike Thelwall. 2017. Are Wikipedia citations important evidence of the impact of scholarly articles and books? *Journal of the Association for Information Science and Technology* 68, 3 (2017), 762–779.

[28] Srijan Kumar, Robert West, and Jure Leskovec. 2016. Disinformation on the Web: Impact, characteristics, and detection of Wikipedia hoaxes. In *Proc. International Conference on World Wide Web.*

[29] Daniel Lamprecht, Dimitar Dimitrov, Denis Helic, and Markus Strohmaier. 2016. Evaluating and improving navigability of Wikipedia: a comparative study of eight language editions. In *Proc. International Symposium on Open Collaboration.*

[30] Daniel Lamprecht, Kristina Lerman, Denis Helic, and Markus Strohmaier. 2017. How the structure of Wikipedia articles influences user navigation. *New Review of Hypermedia and Multimedia* 23, 1 (2017), 29–50.

[31] Janette Lehmann, Claudia Müller-Birn, David Laniado, Mounia Lalmas, and Andreas Kaltenbrunner. 2014. Reader preferences and behavior on Wikipedia. In *Proceedings of the 25th ACM conference on Hypertext and social media.*

[32] Florian Lemmerich, Diego Sáez-Trumper, Robert West, and Leila Zia. 2019. Why the world reads Wikipedia: beyond English speakers. In *Proc. ACM International Conference on Web Search and Data Mining.*

[33] Włodzimierz Lewoniewski, Krzysztof Węcel, and Witold Abramowicz. 2017. Analysis of references across Wikipedia languages. *Information and Software Technologies* 756 (2017), 561–573.

[34] Lauren A Maggio, John M Willinsky, Ryan M Steinberg, Daniel Mietchen, Joseph L Wass, and Ting Dong. 2019. Wikipedia as a gateway to biomedical research: The relative distribution and use of citations in the English Wikipedia. *PLOS ONE* 12, 12 (2019).

[35] Mostafa Mesgari, Chitu Okoli, Mohamad Mehdi, Finn Nielsen, and Arto Lanamäki. 2015. "The sum of all human knowledge": A systematic review of scholarly research on the content of Wikipedia. *Journal of the Association for Information Science and Technology* 66, 2 (2015), 219–245.

[36] Marc Miquel-Ribé. 2015. User engagement on Wikipedia, a review of studies of readers and rditors. In *Proc. International AAAI Conference on Web and Social Media.*

[37] Helen Susannah Moat, Chester Curme, Adam Avakian, Dror Y. Kenett, H. Eugene Stanley, and Tobias Preis. 2013. Quantifying Wikipedia usage patterns before stock market moves. *Scientific Reports* 3, 1 (2013).

[38] Finn Årup Nielsen. 2007. Scientific Citations in Wikipedia. *First Monday* 12 (2007).

[39] Finn Årup Nielsen. 2012. Wikipedia Research and Tools: Review and Comments. *Social Science Research Network (SSRN) - Electronic Journal* (2012).

[40] Finn Årup Nielsen, Daniel Mietchen, and Egon Willighagen. 2017. Scholia, Scientometrics and Wikidata. In *The Semantic Web: ESWC 2017 Satellite Events.* Vol. 10577. 237–259.

[41] Nov Oded. 2007. What motivates Wikipedians? *Commun. ACM* 50, 11 (2007), 60–64.

[42] Ashwin Paranjape, Robert West, Leila Zia, and Jure Leskovec. 2016. Improving website hyperlink structure using server logs. In *Proc. International Conference on Web Search and Data Mining.*

[43] Svitlana Petrasova, Nina Khairova, Włodzimierz Lewoniewski, Orken Mamyrbayev, and Kuralay Mukhsina. 2018. Similar text fragments extraction for identifying common Wikipedia Communities. *Data* 3, 4 (2018).

[44] Tiziano Piccardi, Michele Catasta, Leila Zia, and Robert West. 2018. Structuring Wikipedia articles with section recommendations. In *Proc. ACM SIGIR Conference on Research & Development in Information Retrieval.*

[45] Alessandro Piscopo and Elena Simperl. 2019. What we talk about when we talk about Wikidata quality: a literature survey. In *Proc. International Symposium on Open Collaboration.*

[46] Reid Priedhorsky, Jilin Chen, Shyong Tony K. Lam, Katherine Panciera, Loren Terveen, and John Riedl. 2007. Creating, destroying, and restoring value in wikipedia. In *Proc. ACM Conference on supporting group work.*

[47] Jacob Ratkiewicz, Santo Fortunato, Alessandro Flammini, Filippo Menczer, and Alessandro Vespignani. 2010. Characterizing and Modeling the Dynamics of Online Popularity. *Physical Review Letters* 105, 15 (2010).

[48] Miriam Redi, Besnik Fetahu, Jonathan Morgan, and Dario Taraborelli. 2019. Citation Needed: A Taxonomy and algorithmic assessment of Wikipedia's verifiability. In *Proc. International Conference on World Wide Web.*

[49] Flavia Salutari, Diego Da Hora, Gilles Dubuc, and Dario Rossi. 2019. A Large-scale study of Wikipedia users' quality of experience. In *Proc. International Conference on World Wide Web.*

[50] Aju Thalappillil Scaria, Rose Marie Philip, Robert West, and Jure Leskovec. 2014. The last click: why users give up information network navigation. In *Proc. International Conference on Web Search and Data Mining*.
[51] Thomas Shafee, Gwinyai Masukume, Lisa Kipersztok, Diptanshu Das, Mikael Häggström, and James Heilman. 2017. Evolution of Wikipedia's medical content: past, present and future. *Journal of Epidemiology and Community Health* (Aug. 2017), jech–2016–208601.
[52] Philipp Singer, Denis Helic, Behnam Taraghi, and Markus Strohmaier. 2014. Detecting memory and structure in human navigation patterns using Markov chain models of varying order. *PLoS ONE* 9, 7 (2014), e102070.
[53] Philipp Singer, Florian Lemmerich, Robert West, Leila Zia, Ellery Wulczyn, Markus Strohmaier, and Jure Leskovec. 2017. Why we read Wikipedia. In *Proc. International Conference on World Wide Web*.
[54] Philipp Singer, Thomas Niebler, Markus Strohmaier, and Andreas Hotho. 2013. Computing semantic relatedness from human navigational paths: A case study on Wikipedia. *International Journal on Semantic Web and Information Systems* 9, 4 (2013), 41–70.
[55] Yang Song, Xiaolin Shi, and Xin Fu. 2013. Evaluating and predicting user engagement change with degraded search relevance. In *Proc. International Conference on World Wide Web*.
[56] A. Spoerri. 2007. What is popular on Wikipedia and why? *First Monday* 12, 4 (2007). https://firstmonday.org/ojs/index.php/fm/article/view/1765/1645
[57] Nathan TeBlunthuis, Tilman Bayer, and Olga Vasileva. 2019. Dwelling on Wikipedia: investigating time spent by global encyclopedia readers. In *Proc. International Symposium on Open Collaboration*.
[58] Misha Teplitskiy, Grace Lu, and Eamon Duede. 2017. Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology* 68, 9 (2017), 2116–2127.
[59] Neil Thompson and Douglas Hanley. 2018. Science is shaped by Wikipedia: evidence from a randomized control trial. *MIT Sloan Research Paper* 5238, 17 (2018).
[60] Robert Tomaszewski and Karen I. MacDonald. 2016. A Study of citations to Wikipedia in scholarly publications. *Science & Technology Libraries* 35, 3 (2016), 246–261.
[61] Daniel Torres-Salinas, Esteban Romero-Frías, and Wenceslao Arroyo-Machado. 2019. Mapping the backbone of the humanities through the eyes of Wikipedia. *Journal of Informetrics* 13, 3 (2019), 793–803.
[62] Christoph Trattner, Denis Helic, Philipp Singer, and Markus Strohmaier. 2012. Exploring the differences and similarities between hierarchical decentralized search and human navigation in information networks. In *Proc. International Conference on Knowledge Management and Knowledge Technologies*.
[63] Vivienne Waller. 2011. The search queries that took Australian Internet users to Wikipedia. *Information Research* 16, 3 (2011).
[64] Robert West and Jure Leskovec. 2012. Automatic versus human navigation in information networks. In *International AAAI Conference on Web and Social Media*.
[65] Robert West and Jure Leskovec. 2012. Human wayfinding in information networks. In *Proc. International Conference on World Wide Web*. Lyon, France.
[66] Robert West, Ashwin Paranjape, and Jure Leskovec. 2015. Mining missing hyperlinks from human navigation traces: A case study of Wikipedia. In *Proc. International Conference on World Wide Web*.
[67] Ellery Wulczyn, Robert West, Leila Zia, and Jure Leskovec. 2016. Growing Wikipedia across languages via recommendation. In *Proc. International Conference on World Wide Web*.
[68] Taha Yasseri, Robert Sumi, András Rung, András Kornai, and János Kertész. 2012. Dynamics of conflicts in Wikipedia. *PLoS ONE* 7, 6 (June 2012), e38869.
[69] Ramtin Yazdanian, Leila Zia, Jonathan Morgan, Bahodir Mansurov, and Robert West. 2019. Eliciting new Wikipedia users' interests via automatically mined questionnaires: For a warm welcome, not a cold start. *International AAAI Conference on Web and Social Media*.
[70] Xing Yi. 2015. Dwell time based advertising in a scrollable content stream. US Patent App. 13/975,157.