The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies

Connor McMahon[1*], Isaac Johnson[2*], and Brent Hecht[2]
*indicates co-First Authors; [1]GroupLens Research, University of Minnesota;
[2]People, Space, and Algorithms (PSA) Computing Group, Northwestern University
mcmah250@umn.edu, isaacj@u.northwestern.edu, bhecht@northwestern.edu

Abstract

While Wikipedia is a subject of great interest in the computing literature, very little work has considered Wikipedia's important relationships with other information technologies like search engines. In this paper, we report the results of two deception studies whose goal was to better understand the critical relationship between Wikipedia and Google. These studies silently removed Wikipedia content from Google search results and examined the effect of doing so on participants' interactions with both websites. Our findings demonstrate and characterize an extensive interdependence between Wikipedia and Google. Google becomes a worse search engine for many queries when it cannot surface Wikipedia content (e.g. click-through rates on results pages drop significantly) and the importance of Wikipedia content is likely greater than many improvements to search algorithms. Our results also highlight Google's critical role in providing readership to Wikipedia. However, we also found evidence that this mutually beneficial relationship is in jeopardy: changes Google has made to its search results that involve directly surfacing Wikipedia content are significantly reducing traffic to Wikipedia. Overall, our findings argue that researchers and practitioners should give deeper consideration to the interdependence between peer production communities and the information technologies that use and surface their content.

Introduction

As one of the most prominent peer production communities in the world, Wikipedia has been the subject of extensive research in computational social science, human-computer interaction (HCI), and related domains. As a result of this work, we now have a detailed understanding of sociotechnical designs that lead to successful peer production outcomes (e.g., Kittur and Kraut 2008; Zhu et al. 2013). We have also uncovered aspects of Wikipedia that are more problematic (e.g., Menking and Erickson 2015; Warncke-Wang et al. 2015; Johnson et al. 2016).

However, the literature on Wikipedia has largely considered Wikipedia in isolation, outside of the context of its broader information technology ecosystem. This ecosystem contains potentially critical relationships that could affect Wikipedia as much as or more than any changes to internal sociotechnical design. For example, in order for a Wikipedia page to be edited, it needs to be visited, and search engines may be a prominent mediator of Wikipedia visitation patterns. If this is the case, search engines may also play a critical role in helping Wikipedia work towards the goal of "[providing] every single person on the planet [with] access to the sum of all human knowledge" (Wales 2004).

Also remaining largely unexamined are the inverse relationships, or the contributions that Wikipedia makes *to* other widely-used information technologies. For instance, Wikipedia data helped to seed Google's Knowledge Graph, which powers Google's ability to provide direct answers to certain queries (e.g., "What is the second-largest metropolitan area in Québec?") (Singhal 2012), and Wikipedia has similar relationships with systems like Wolfram Alpha (Taraborelli 2015), Amazon Echo (Amazon.com, Inc. 2016) and Apple's Siri (Lardinois 2016). Even more fundamentally, the mere presence of Wikipedia articles may help search engines respond with highly relevant links to a large variety of queries for which relevant links would have been otherwise hard to identify (or may not exist).

This research seeks to begin the process of understanding Wikipedia and peer production communities in the context of their broader information technology landscapes. We do so by focusing on the relationship between Wikipedia and Google, with anecdotal reports indicating that this relationship might be particularly strong and important for both parties (e.g. Devlin 2015; Mcgee 2012).

To initiate the work of understanding the relationship between Wikipedia and Google, we executed two controlled deception studies that examined search behavior and Wikipedia use both in naturalistic settings (Study 1) and in an online laboratory context (Study 2). In both studies, participants installed a browser extension we built that, behind-the-scenes, removes all or part of the Wikipedia content that is present on Google search results pages (SERPs). When a participant was in the most extreme Wikipedia removal condition, they experienced search results as if Wikipedia did not exist (at least at the surface level; see below). We then tracked participants' interactions with Google and Wikipedia in various Wikipedia-removal and control conditions during the study period, which for the in-the-wild study (Study 1) lasted 2-3 weeks. Through these studies and extensive semi-structured exit interviews, we gathered quantitative and qualitative data that sheds important new light on the Wikipedia-Google relationship (and likely extends to Wikipedia's relationship with other technologies).

At a high level, our results show that Wikipedia both contributes to and benefits from its relationship with Google, and that these contributions and benefits are substantial. Specifically, in the first attempt to robustly quantify the importance of Wikipedia pages to search performance, we find that Wikipedia makes Google a significantly better search engine for a large number of queries, increasing relative SERP click-through rate (CTR) by about 80% for these queries. Conversely, we are also able to confirm reports that Google is the primary source of traffic to Wikipedia. Overall, our results make it clear that Wikipedia and Google help each other achieve their fundamental goals.

Our results also indicate, however, that this mutually beneficial relationship has been placed at risk by the overhaul of Google's SERPs connected with Google's rollout of its "Knowledge Graph" technology (Singhal 2012). As shown in Figure 1, SERPs that are visibly enhanced by the Knowledge Graph include "knowledge panels" and/or "rich answers" (among other content-bearing search assets) in addition to the traditional "ten blue links" to relevant web pages. Knowledge panels generally contain summaries, facts, and concepts related to the query, while rich answers attempt to directly respond to queries that resemble questions. Our results suggest that, as has been hypothesized by leaders of the Wikipedia community (e.g., DeMers 2015; Taraborelli 2015; Wales 2016), this content may be fully satisfying the information needs of searchers in some cases, eliminating the need to click on Wikipedia links.

An irony inherent in the above finding is that not only was the lower-level semantic infrastructure of the Knowledge Graph seeded with Wikipedia content (Singhal 2012), but content directly from Wikipedia comprises large portions of many user-facing Knowledge Graph assets. Indeed, 60% of knowledge panels encountered by our participants had content directly attributable to Wikipedia, as in the case in Figure 1. The same was true of 25% of rich answers.

A long-term reduction of traffic to Wikipedia due to the Knowledge Graph (and perhaps due to Google's direct sur-

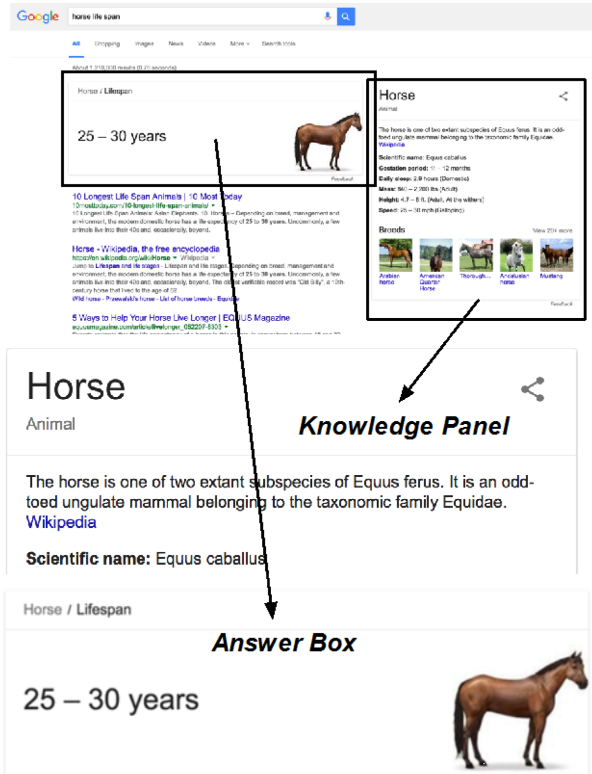

Figure 1: A "Knowledge Panel" and "Rich Answer"

facing of Wikipedia content) could have substantial negative effects on Wikipedia's peer production model. In particular, some have argued that this reduction in traffic could lead to a "death spiral", in which a decrease in visitors leads to a decline in both overall edits and new editors, not to mention much-needed donations (Dewey 2016; Taraborelli 2015). This would then lead to a reduction in the quality and quantity of the very content that Google and other information technologies are using to power their semantic technologies and are surfacing directly to users. If substantiated, this trend would represent a new and concerning interaction between peer production communities and information technologies that likely would generalize beyond the relationship between Wikipedia and Google. Moreover, if it exists, this interaction will likely only grow in the near future, especially as Siri, Amazon Echo, Cortana, and other information technologies follow Google's lead and address more information needs directly rather than pointing people to webpages (often using Wikipedia content to do so).

Related Work

A New Wikipedia Research Agenda

The primary motivation for this work arises out of a call for a "new [Wikimedia] research agenda" made by Dario Taraborelli (2015), the Head of Research at the Wikimedia Foundation (the foundation that operates Wikipedia). Our research seeks to address two specific questions raised in Taraborelli's agenda: (1) What is the role of Wikipedia in a world in which other information technologies (e.g. the Knowledge Graph[1]) are increasingly able to directly address people's information needs? and (2) What is the effect of the "paradox of reuse" of Wikimedia content, in which Wikipedia and other Wikimedia sites (e.g., Wikidata) power these increasingly powerful technologies, which in turn reduce the need to visit Wikipedia.

Others have begun to raise separate concerns about the relationship between Wikipedia and the Knowledge Graph. Most notably, Ford and Graham (2015) have shown that the reduction of an entire concept to a single knowledge panel snippet (see Figure 1) can be problematic for controversial concepts (e.g. Taiwan is labeled as an independent nation, even though only a minority of nations recognize it).

Google and Wikipedia

Wikipedia has been suffering a decline in overall traffic in recent months (a point to which we return later), and discussions about the cause of this decline (e.g., DeMers 2015; Dewey 2016; Wales 2016) also strongly motivated this research. Indeed, as part of this discussion, Jimmy Wales, the co-founder of Wikipedia, has written that the Knowledge Graph (and related technologies) could present a "long-term issue" for Wikipedia (DeMers 2015; Wales 2016).

One recent outcome of this discussion are data points dating back to August 2015 in which the Wikimedia Foundation found that the share of traffic to Wikipedia from Google has been trending slightly up rather than down (Keyes 2015). However, this has not ended the debate, as the report had certain disclosed limitations and, more importantly, like all observational data, this data is subject to uncontrolled exogenous influences. Our research adds a key data point to this debate by presenting the results of a controlled study. Specifically, our findings point to what happens to Wikipedia visitation rates when a Knowledge Graph asset that would have been shown is silently removed.

In addition to its high-level use of Wikipedia in the form of links to Wikipedia pages and its inclusion of Wikipedia content in Knowledge Graph assets, Google also uses Wikipedia to power some of its lower-level artificial intelligence systems. Wikipedia has proven as valuable to the AI literature as it has to HCI and computational social science (see (Hovy, Navigli, and Ponzetto 2013) for an overview), and Google has demonstrated this utility in practice. We know, for instance, that Wikipedia helps Google rank search results (even those that have nothing to do with Wikipedia) (Devlin 2015; Hodson 2016). We also know that Wikipedia helped to seed Google's Knowledge Graph (Singhal 2012), meaning that Wikipedia may assist Google in doing inference to address complex questions and queries, regardless of whether the ultimate result comes from Wikipedia.

While the literature in AI and related fields has described the value of Wikipedia to research systems, this paper provides the first quantification of the contributions Wikipedia makes to a popular, highly-profitable intelligent technology. In this study, we do not focus on the lower-level value Wikipedia brings to Google but instead on the user-facing contributions Wikipedia makes to Google's effectiveness. In other words, we ask, "How good of a search engine would Google be if it could not surface Wikipedia data directly?" Measuring the contributions Wikipedia makes to Google at a lower level would be impossible to accomplish without having access to Google's proprietary systems. As such, our assessment of the benefit Google receives from Wikipedia *should be viewed as a lower-bound on the actual benefit.*

Search Relevance and User Experience

Our approach in this paper is motivated by the work of Song et al. (2013), who studied the impact of intentionally degraded search performance on short- and long-term user engagement with search engines. As is common in the search literature (De Rijke 2016), Song et al. made heavy use of the click-through rate (CTR) metric, or the ratio of queries that had at least one clicked search result. This is a critically important engagement statistic for search engines (De Rijke 2016): if users rarely click on links returned by a search engine, it is unlikely that they are satisfied with that search engine. As such, we adopt CTR as a core dependent variable.

Song et al.'s work highlights another way in which our results should be viewed as a lower-bound on Wikipedia's value to Google. Namely, Song et al. found that degraded search performance can result in long-term decreases in searches per day and queries per session. This suggests that if our study adopted a longitudinal perspective, the observed benefit of Wikipedia to Google would only grow.

Algorithmic Auditing

This work is also inspired by the rapidly-growing algorithmic auditing literature, which seeks to understand "from the outside" the behavior of important, closed algorithms whose detailed functionality is not made public (e.g. Ananny, Karahalios, and Wilson 2015). For instance, in one of the more well-known pieces of work in this literature, Eslami et al. (2015) developed the FeedVis system to expose the algorithmic curation that occurs in the Facebook News Feed. This tool enabled Eslami and colleagues to run a study with

[1] Google unveiled the "Knowledge Vault" in 2014 as a potential successor to Knowledge Graph, but it is unclear whether it has been put into production and Google still refers to this technology as the Knowledge Graph in its public-facing materials. Regardless of the name of the technology, the surfacing of Wikipedia content for the end user remains the same.

40 Facebook users to assess their perceptions of the curation. Along the same lines, Kay et al. (2015) found that workers on Mechanical Turk rated image search results for employment-related queries higher when they conformed to career gender stereotypes (among related findings). In many ways, key parts of this paper can be thought of as algorithmic auditing research that targets Google's reliance on Wikipedia rather than, for instance, the Facebook newsfeed algorithm or gender dynamics in image search.

Research Questions

The above related work highlights the importance of better understanding the relationship between Wikipedia and Google. It also highlights the complexity of the Wikipedia-Google relationship, which contains many low- and high-level components. To begin the investigation into this relationship, we posed three straightforward research questions targeted at the key issues raised in related work:

RQ1: How much does Google rely on Wikipedia search results to address information needs?

RQ2: How much does Wikipedia rely on Google as a source of its readership?

RQ3: Are "Knowledge Graph" assets on Google's SERPs reducing traffic to Wikipedia?

We designed our two studies to target these three questions. Below, we discuss our methods in more detail, beginning by providing some context for the methodological challenges associated with this work.

Methods

Through its association with algorithmic auditing, this research inherits the substantial methodological challenges common in that literature. Namely, like most auditing work, our research involves studying a black-box system (i.e. Google's search algorithm) that we must manipulate "from the outside" (Ananny, Karahalios, and Wilson 2015). Additionally, like work that has audited Facebook (Eslami et al. 2015; Eslami et al. 2016), Google Image Search (Kay, Matuszek, and Munson 2015), TaskRabbit (Thebault-Spieker, Terveen, and Hecht 2015), UberX (Ge et al. 2016; Thebault-Spieker, Terveen, and Hecht 2017) and other private systems, we cannot study our system of interest (Google Search) at close to the same scale that would be possible internally. These systems have millions of users, but auditing researchers must typically work at scales more common in social science and HCI research (e.g. 40 participants (Eslami et al. 2015; Eslami et al. 2016; Thebault-Spieker, Terveen, and Hecht 2015) or typical Mechanical Turk scales (Kay et al. 2015)).

As such, to address our research questions, we sought methodological inspiration from the algorithmic auditing literature and adopted a broad methodological framework that is frequently utilized in that domain. Namely, like Eslami et al. (2015; 2016), we developed a tool that allows us to make manipulations externally and ran a smaller-scale study that allowed us to gain both quantitative and qualitative understanding of the phenomena of interest. We also complemented our smaller-scale study with a second, larger study (though it was still very small relative to Google's user base) using Mechanical Turk (similar to Kay et al. (2015)).

Interestingly, while search research is often done at large-scales, there is also a strong and active tradition of smaller-scale studies like ours. For instance, Pan et al. (2007) ran a highly-influential eye-tracking study that, examining 22 participants from their local university, sought to better understand user trust in Google's search ranking algorithm. In general, these smaller-studies typically address newer questions or questions that cannot be investigated at a larger scale, for instance because outcome metrics are too difficult to measure at scale or because they would involve too much degradation of the customer experience (as would be the case for our work). We use this smaller-scale search literature to inform our lower-level methodological choices.

Below, we expand on our methodology in more detail, beginning by describing our browser plug-in. We then discuss our study design and statistical approaches.

Browser Extension

We developed a custom browser extension to serve as the primary apparatus for this research. The extension, which is designed for Google Chrome (desktop only; see below), silently removes Wikipedia links and Knowledge Graph assets from Google search results, and does so to varying degrees according to our experimental design. In order to understand the effects of these changes, the extension also logs a limited set of interactions with the modified results pages as well as basic information about visits to Wikipedia.

The extension implements three experimental conditions that vary the extent and type of information removed from Google SERPs. These conditions are described below:

- The *WikipediaKnowledgeGraphAssetsRemoved* condition removes all visible Knowledge Graph assets that are directly or possibly attributable to Wikipedia. We discuss how we defined these assets below.
- The *AllWikipediaRemoved* condition is a superset of the *WikipediaKnowledgeGraphAssetsRemoved* condition. In addition to removing all Wikipedia-sourced Knowledge Graph assets, it also removes all standard search results that are links to Wikipedia content.
- The *Control* condition is a baseline condition. The only modification is a 300-millisecond page load delay to control for the delay caused by the above conditions.

It is important to note that these extension conditions interact with the SERP conditions that are defined by the results returned by Google for a given query. For instance, even if a participant was in the *WikipediaKnowledgeGraphAssetsRemoved* condition, the extension would not modify a SERP if it did not have Wikipedia-based Knowledge Graph assets. The variation in SERP conditions was both an opportunity and something for which we needed to control. With regard to the latter, we took care in our analyses to not rely solely on the extension condition but to control for which Wikipedia assets and links were actually present when comparing between conditions (e.g. as in Table 1). With regard to the former, we were able to leverage the variation in SERP conditions to draw conclusions about how the removal of Wikipedia links alone impacts search behavior even though this was not its own extension condition.

Determining which Knowledge Graph assets have Wikipedia content is not straightforward. To address this challenge, we issued the top two queries in each category listed in Google's Trending Topics in early 2016 (Google 2016). We then identified Knowledge Graph assets that were directly attributable to Wikipedia (e.g., the summary about horses in Figure 1). As there have been claims that Google may not be properly attributing Wikipedia content (especially when minor inference may be involved) (e.g., Dewey 2016; Taraborelli 2015), we also removed the page elements of information that could have been derived from Wikipedia, e.g., by downloading the Wikipedia database dump or by scraping Wikipedia infoboxes. Our removal of this information gives us greater confidence that no Wikipedia information appeared in the experimental conditions.

Prior to launching both studies, we took care to ensure that our extension was performing as expected, even with any changes to Google's returned SERP HTML. To do so, we developed a set of 135 queries – three queries per Google Trending Topic – and validated manually that the extension was removing the correct parts of each SERP under both the *WikipediaKnowledgeGraphAssetsRemoved* and the *AllWikipediaRemoved* condition.

Further details about the implementation of the browser extension can be obtained by downloading its source code and corresponding documentation.

Study 1: In-the-Wild Experiment

The goal of our first experiment was to gain an understanding of the relationship between Google Search and Wikipedia with respect to people's natural information seeking behavior and, critically, their organic information needs. As such, we designed the experiment as an *in-situ* deception study in which participants installed our extension on their personal computers for a period of 2-3 weeks. We then observed how participants' search and Wikipedia behavior changed across our experimental conditions. Below, we discuss this design and its benefits and tradeoffs in more detail.

We recruited 22 people in our university's community for this first experiment (following Pan et al. (2007)). Each participant installed our browser extension on their computer. Participants were not told of the purpose of the extension, just that the extension would log their Google search behavior in support of search research. At the end of the study, all participants indicated that they had not inferred the manipulations that were being performed. Due to the sensitive nature of the logging and deception aspect of the study, we developed a protocol with our IRB that involved taking great care in installing the extension, collecting only the minimum amount of data to address our research questions, and thoroughly debriefing each participant in-person.

The average age of participants was 23, and 18 of the subjects were students in a field of science, engineering, or medicine. Sixteen of the participants were male, and eight were female. Participants were paid $15 for completing the study. Participants were recruited based on the criteria that Google Chrome is their predominant web browser and Google Search is their predominant search engine.

Following common practice in algorithmic auditing, we designed this study as a mixed methods study. While our quantitative findings emerge from controlled experimentation and corresponding statistical tests (see below), our qualitative findings are based on semi-structured exit interviews that we conducted with each participant. Each interview revolved around 20 inquiries on topics related to our research questions. Example interview questions included "What are some benefits you experience from Google's knowledge cards?" [after showing participants an example knowledge card] and "Where do you think the information found in the knowledge cards and answer boxes comes from?" During this interview, the participants were also asked to investigate a fixed set of information needs designed to probe how they interacted with Knowledge Graph assets (a think-aloud protocol was used). A single coder reviewed the recordings of these interviews and identified statements that related to each of our three research questions.

During the study period, participants were randomly assigned to a condition at the beginning of each search session. Following standard practice, a new session was established if more than 30 minutes had passed since the last query (e.g., Song, Shi, and Fu 2013; Wulczyn 2016). Randomly reassigning the experimental condition at each new session was done to ensure greater parity of data between each condition. For ethical reasons, we also sought to avoid a design in which a small group of participants suffered from potentially degraded search performance for an extended period.

Similarly, out of concern for participants' privacy, when participants visited a URL under Google.com, we tracked very little information about SERPs other than the query, details about the presence of Wikipedia content and Knowledge Graph assets, and whether a link was clicked. As such, when we report the share of Google queries that contained Wikipedia content below, it represents a conservative lower bound: this statistic reflects the share of total

Google queries, not Google web search queries (i.e. the denominator likely includes a small percentage of image or map queries).

Study 1 Basic Descriptive Results

Overall, our extension logged 4,092 queries from all Google platforms (2,703 of these queries were unique). We recorded 298 separate visits to Wikipedia pages. Five participants did not visit Wikipedia at all, and nine visited it less than 10 times. For the analyses below, we verified that no participant (i.e. an outlier) changed our results in any way that would affect our overall conclusions.

Study 2: Focused Knowledge Graph Experiment

As is described in more detail below, we designed our second study specifically to further investigate our third research question (about the Knowledge Graph's effect of Wikipedia traffic). For this study, we recruited 224 participants through Amazon Mechanical Turk. Participants were paid $3.25 for an approximately 20-minute task (above the local minimum wage).

Each participant was randomly assigned either to the *Control* or *WikipediaKnowledgeGraphAssetsRemoved* condition for the duration of the experiment. Participants were presented with seven information needs and asked to address each need using Google Search. Each individual's first information need was self-generated (e.g., "a question that you have been meaning to ask someone," survey language drawn from Oeldorf-Hirsch et al. (2014)). The remaining information needs were pre-defined and their order was randomized. Two of these six pre-defined information needs were consistent across all participants as quality assurance checks, with one chosen to be highly likely to elicit Wikipedia Knowledge Graph assets and one chosen to be highly unlikely to do so.

We designed the other four pre-defined information needs to afford as much ecological validity as possible within Study 2's controlled setting. To do so, we leveraged prior work in the human-centered search literature, specifically the approach of Pan et al. (2007) that involved using information needs in the following four categories: Travel, Current Events, Movies, and Celebrities. We updated Pan et al.'s information needs to be relevant in the year 2016 and outside the original study context (Cornell University).

We also collected qualitative data in this study, but of a more structured variety. In particular, in addition to tracking where participants got their information to address their information needs, we also asked them where they *believed* they got their information. Participants' responses to this question were coded by two researchers who achieved a sufficient Cohen's Kappa (0.91) on 100 responses and coded the remainder of responses individually. As we discuss below, this data provided a key insight into Knowledge Graph-related risks to Wikipedia beyond those that can be measured by tracking clicks.

For this study, we again worked with our IRB to develop a detailed plan to protect participants. We informed participants up-front that their queries were being logged and provided them with instructions for uninstalling the extension. We also automatically disabled logging and search modification by the extension after the study duration had expired in case a participant failed to uninstall the extension.

Study 2 Basic Descriptive Results

Overall, our extension logged 1690 queries in this study, 1069 (63.3%) of which were determined to have Wikipedia content (either links or Knowledge Graph assets). We recorded 178 separate visits to Wikipedia.

Metrics and Statistical Methods

The two primary dependent variables considered in both experiments are SERP click-through rate (CTR) and Wikipedia Visitation Rate (WVR). As discussed above, CTR is an important search engagement metric that captures the percentage of SERPs on which a user clicked on a link. Our Wikipedia visitation rate (WVR) metric can be understood as "SERP click-through rate for Wikipedia". We calculated this by simply examining whether a visit to Wikipedia occurred within 10 seconds of a SERP visit. This allowed us to capture clicks to Wikipedia links outside the "ten blue links" (the standard search results on a SERP).

Like is the case with some studies of query logs (e.g. Silverstein et al. 1999), our research questions target the overall outcomes of search behavior, regardless of the behavior of individual users. In other words, the primary unit of interest is the query (i.e. what percentage of queries end up at Wikipedia?), not the participant (i.e. what percentage of people end up at Wikipedia?). However, it is also interesting to consider our results from a user-by-user perspective, especially with respect to what our results may mean for donations and new editors. In our results section below, we present the output of statistical tests that take both of these perspectives.

To examine the overall relationship between Wikipedia and Google at the query-level, it is possible to employ straightforward logistic regressions when comparing across conditions. However, for the user-by-user perspective, it is necessary to control for the user who issues each query, which requires mixed-effects modeling. Specifically, we employ mixed-effects logistic regression models that use a random intercept to account for repeated measures for individuals. As we will see below, regardless of the statistical perspective we take, the results point to the same high-level findings about Google-Wikipedia relationship.

Finally, it is important to note that because we did not replace a link from the second page of search results when we removed a Wikipedia link (this likely would have affected page load times significantly), the reduced number of links on a given SERP in the manipulation may lead to a small reduction in CTR. However, the tenth link gets very few clicks (i.e. has a CTR of 0.5-3%) (Petrescu 2014), so the smaller number of results is very unlikely to have played a significant role in any CTR differences between conditions.

Results

We present our results by research question. In each section, we draw on our quantitative and qualitative results.

RQ1: Google's Dependence on Wikipedia

The results of Study 1 present clear evidence that Wikipedia contributes substantially to Google's success. Put simply, our data suggest that without being able to surface Wikipedia links and content, Google is simply a worse search engine for many queries. Google's dependence on Wikipedia was clear in Study 1's descriptive statistics as well as in our CTR comparisons across conditions that more directly targeted our research question.

With respect to descriptive statistics, we saw that Google returned Wikipedia content (links and/or assets) for 28% of queries in Study 1 (The equivalent number in Study 2 was 36.2%). Moreover, as noted above, this represents a lower-bound on the share of web search queries that contained Wikipedia content. Although our data collection restrictions did not permit us to compare Wikipedia's 28% share of queries to that of other websites, it is likely that few – if any – content providers approached this degree of prominence.

The more surprising Study 1 result, however, came when we examined the effects on CTR between the conditions when Wikipedia links were present on SERPs. When these links were made visible to participants, the click-through rate was 26.1% (n = 618). However, when the browser extension silently removed the Wikipedia links, we saw the click-through-rate drop to 14.0% (n = 387). In other words, *Wikipedia links increase SERP CTR by roughly 80% when they are present*, with CTR being a critical metric on which search performance is evaluated. Our fixed- and mixed-effects models (Table 1) indicate that this difference was significant with fixed effects ($p < 0.01$) and marginally significant when controlling for participant ($p = 0.065$).

If this 80% effect size holds across the entire population of users and queries, it would be a somewhat remarkable testament to Wikipedia's value to search engines. For comparison (with respect to queries), improvements to search algorithms generally result in much smaller CTR effects. For instance, when *intentionally* trying to reduce search performance by using an inferior results ranking algorithm, Song et al. (2013) only saw a 1% CTR decrease. In other words, our results suggest that the mere presence of Wikipedia links may have an effect approximately 80 times larger than the difference between a good ranker algorithm and bad one (for many queries).

Coefficient	Fixed Effects	Mixed Effects
Intercept	-0.874**	-1.885**
LinksRemoved	-0.796**	-0.385†
KGAssetsOrigPresent	-0.564**	-0.407†
KGAssetsRemoved	0.458*	0.493†

*Table 1. Model coefficients when evaluating the effect of removing Wikipedia links on click-through rate. The primary row of interest is LinksRemoved (shaded). The bottom two variables are dummy variables to ensure that we were controlling for the experimental condition and Knowledge Graph-related SERP properties. KGAssetsOrigPresent indicates whether the SERP originally contained Knowledge Graph assets and KGAssetsRemoved indicates if those assets were subsequently removed by our extension; ** $p < 0.01$, * $p < 0.05$, $^{\dagger} p < 0.10$.*

Our qualitative data also supports Wikipedia's import to Google. Examining our interview dataset for comments related to the role of Wikipedia in participants' Google search sessions, we observed that a number of participants mentioned that this role is quite large. For instance, P10 reported, "I think for almost all the searches which I do, the first link I visit is Wikipedia". Similarly, P6 said "For more than 50% of my Google Searches, Wikipedia is the link I go to first, and am satisfied with it."

RQ2: Wikipedia's Dependence on Google

In order to evaluate Wikipedia's dependence on Google, we simply calculated the percentage of overall visits to Wikipedia pages in the in-the-wild experiment that could be attributed to Google. We included in this count both visits to Wikipedia pages that were the direct result of links on Google SERPs and visits to Wikipedia pages that came as a result of traffic from Google (i.e. pages that were visited after someone arrived on Wikipedia due to a Google search result). Following the procedure of Song et al. (2013), any Wikipedia page that was accessed within 30 minutes of a visit to Wikipedia due to a Google SERP was included in this second class of visits.

We found that 84.5% of the 298 visits to Wikipedia pages that we observed were attributable to Google. That is, only 15.5% came from other sources or from direct visits. These quantitative findings from our browser extension logs are substantiated by information from our exit interviews. When asked if they visit Wikipedia directly or through search engines, only four (of 22) respondents said that they visit Wikipedia directly, and even these participants said they usually visit through search engines. The other 18 participants reported that they *always* visited through search engines. One subject even stated that she "only [goes] to Wikipedia because it is at the top" of the search results (P8).

It is interesting to consider the implications of this result for Wikipedia's peer production content production model.

A person must visit Wikipedia to become an editor, or to make an edit more generally. Our findings suggest that Google may play a fundamental role in the success of this model by being the source of the majority of Wikipedia page views. While it is unlikely that power editors (who contribute most of Wikipedia's content) visit most of the pages they edit through Google, these findings suggest that they may have originally become engaged with Wikipedia because of Google. This result strongly advocates for further research to investigate the percentage of overall edits and editors in Wikipedia that can be attributed to lead from Google.

Critically, our findings also indicate that Google contributes to a vital component of the Wikipedia ecosystem that receives far too little attention in the literature (Lund and Venäläinen 2016): funding through donations. Wikipedia gets almost 60% of its revenue through donation banner ads (Wikimedia Foundation 2015), which require page views. If our results generalize to all Wikipedia visitors, a large portion of these donations may originate with a Google search query.

RQ3: Effect of the Knowledge Graph

To address our question about the effect of Knowledge Graph assets on traffic to Wikipedia, we first examined the Wikipedia Visitation Rate (WVR) across the *WikipediaKnowledgeGraphAssetsRemoved* and *Control* conditions in Study 1 (when both Wikipedia links and Knowledge Graph assets were present on [unmodified] SERPs). We saw a trend that supported the hypothesis that Knowledge Graph assets were reducing the WVR. However, due to the organic nature of search behavior in our first study, the sample sizes considered in this comparison were relatively low and both our models were inconclusive.

We ran our second study with the goal of investigating whether we could verify the trends that we observed in our in-the-wild study more robustly. In our analysis of our Study 2 results, we were careful to only consider SERPs that had at least one Wikipedia link and at least one Wikipedia-based Knowledge Graph asset (the assets were removed in the manipulation and allowed to surface in the control).

The results of our second study strongly confirm the trends we identified in Study 1: across our four categories of fixed information needs in Study 2 (n = 438 queries), the WVR jumped from 11.1% (n = 207) to 20.5% (n = 231) when the browser extension removed the Wikipedia-based Knowledge Graph content. As can be seen in Table 2, this difference was significant in both our fixed- and mixed-effects models (p < 0.01). In other words, *when we silently deleted the Wikipedia content that Google surfaces directly to users, participants clicked on many more Wikipedia links*. We also made the same comparison across all seven information needs and identified a very similar outcome (n = 670 queries; effects significant in both models).

Our Study 2 results are reinforced by qualitative data gathered during Study 1's exit interviews. In particular, a number of participants mentioned that not having to visit Wikipedia was a major benefit of Google's knowledge panels. For instance, P4 noted that "[With the knowledge panels], I don't have to look at Wikipedia. Even though Wikipedia is great, I still have to sift to find out dates, or just an important summary". P15, closely echoing our quantitative results, remarked that "If I Google something, or a question, and a knowledge card doesn't pop up, then that is when I will typically click on Wikipedia". Six other participants noted the quick access to this set of facts as helpful, with one participant claiming that "Even though it's not that hard to go to Wikipedia, all of the information I would want to know is listed right here, so it is more convenient" (P19).

Coefficient	**Fixed Effects**	**Mixed Effects**
Intercept	-2.031**	-2.975**
KGAssetsRemoved	0.820**	1.072**

*Table 2. Model coefficients when evaluating the effect of removing Knowledge Graph assets (highlighted in gray) on WVR (n = 438). As this was a direct comparison across conditions, no additional controls were appropriate. ** indicates p < 0.01*

Examining the qualitative responses to our provenance-related question in Study 2, we identified that the problem for Wikipedia may be even more serious. Not only do people not visit Wikipedia at the same rate when Wikipedia is surfaced through Knowledge Graph assets, they attribute the information they find on SERPs *to Google directly*. After coding responses to the question "Where did you find this information?" (n = 689), we saw a 5.7-fold increase in the percentage of participants who reported finding the information that satisfied their information *from Google* when Knowledge Graph assets were present compared to when we removed them (52% of participants compared to 9%). We saw an expectedly inverse effect for people who reported that Wikipedia satisfied their information need: 22% indicated Wikipedia with Knowledge Graph assets present, but 34% did when we removed the assets. We discuss the implications of this finding in more detail below.

Discussion

The results above suggest a relationship between Wikipedia and Google that is mutually beneficial, with each making large contributions to the other's core goals. However, our results also highlight that this beneficial equilibrium may be in danger due to Knowledge Graph-related changes. Below, we discuss the implications of these findings in more detail.

Broadening Peer Production Research

One key implication of the above work is that research on peer production, especially Wikipedia, should more often take an information technologies ecosystem perspective. We saw above that Wikipedia is tremendously reliant on other information technologies – namely Google – to provide it with traffic, from which Wikipedia derives edits, new editors, donations, and, critically, its readership. As a result, it is likely that any changes in how and when Google refers people to Wikipedia would have substantially larger effects than improvements in Wikipedia's communication and collaboration infrastructures and practices, which have been the predominant focus in the Wikipedia literature. Additionally, recognizing the value that Wikipedia's peer-produced content brings to Google is also a subject whose scale and potential impact deserves substantial further investigation.

Wikipedia in a Knowledge Graph World

Google and Wikipedia have had discussions about how to adapt their relationship to a "semantic search" era in which Knowledge Graph-derived results will be more common and "ten blue links" will be less common (Taraborelli 2015). Our results related to the impact of the Knowledge Graph on Wikipedia visitation rates suggest that this era may be a difficult one for Wikipedia.

Fortunately, our other results point to potential means by which any "death spiral" caused by the Knowledge Graph may be addressed. In particular, our findings for the first time put numbers around the important role that Wikipedia links still play in helping Google meet its core goal of connecting people with the information they want. Combined with Wikipedia's lower-level contributions to Google's ranking and inference algorithms (among other algorithms), the Wikipedia community should know that it is bringing value "to the table" in its discussions with Google. This may help Wikipedia achieve design changes in Google and other information technologies that could help keep Wikipedia healthy over the long term. The specific character of these design changes could be a fruitful area of future work. For instance, one could imagine incorporating edit buttons (and donation buttons) directly in knowledge panels.

The Knowledge Graph and Attribution

While this paper is focused on the traffic-related concerns surrounding the changing relationship between Wikipedia and Google, there are other concerns as well. Most notably, as indicated above, some have expressed worry that Google is not properly attributing Wikipedia content in Knowledge Graph assets. Our work provides a few important data points associated with this discussion.

Examining all the (unmodified) SERPs from Study 1, we found that only 7% of rich answers encountered by our participants were un-cited (and 25% even had citations to Wikipedia articles). However, we did find support for attribution concerns with respect to facts in knowledge panels. These were almost never cited, and one likely cause is that this information is coming (in part) from Wikidata, Wikipedia's structured data cousin. Wikidata has an even more permissive licensing agreement that does not require attribution.

This creates an interesting within-Wikimedia tension: the Knowledge Graph is hurting Wikipedia and Wikidata is helping the Knowledge Graph. Moreover, Wikidata may itself be suffering from the "paradox of reuse", especially because Google is not providing searchers with links back to Wikidata so that they can contribute to the project. Examining these complex relationships in more detail is an important subject of future work.

Finally, our results from Study 2 suggest a new attribution-related concern related to the re-use of peer produced content: incorrect *perceived* attributions. In this study, we saw that more than five times as many people said they got their information from Google when the Wikipedia-based Knowledge Graph assets were present than when they were removed. It is unclear whether the lack of attributions or the style of attribution (e.g. the attributions are currently in relatively small font) resulted in so many people saying they got their information "from Google" rather than from the likely original source. However, this is certainly a fruitful direction of future research.

Limitations

Although we followed the long-standing scientific practice of recruiting members of a university community as participants for our in-the-wild experiment, our population may have unrepresentative search and Wikipedia use behavior and was too small to examine heterogeneity among the participants. Additionally, none of the queries measured were on mobile devices, which now make up over half of all of Google's queries in the US (Dischler 2015), and we did not examine the effect of "Wikipedia clones" (i.e. mirrors), although we saw no evidence that these affected our results.

Our follow-up study with a larger (though still biased) population through Amazon Mechanical Turk addressed some of these concerns. However, this research should ideally be replicated and extended using a participant pool several orders of magnitude larger than in our studies. Unfortunately, doing so will likely be very difficult. The best institution to scale up our work is an operator of a search engine itself. However, degrading search results intentionally as we have done here has short- and long-term implications (Song, Shi, and Fu 2013) and there may be reasons information technology operators may not want to advertise their reliance on other institutions (e.g. the Wikimedia Foundation).

Lastly, as noted above, future work should seek to take a longitudinal perspective on our research questions. While difficult, doing so would afford examination of the benefits Wikipedia provides to long-term search engagement. Such a study would also shed light on the amount of traffic that would eventually find its way back to Wikipedia if Google stopped being a reliable intermediary.

Conclusion

In this paper, we have provided evidence that demonstrates and characterizes the mutually beneficial relationship between Google and Wikipedia. We have also identified a concerning trend: Google's changes to its search results pages that surface Knowledge Graph assets may be reducing Google's benefit to Wikipedia, a problem that could lead to serious follow-on effects for both organizations. More generally, this research demonstrates the value of considering peer production communities like Wikipedia in the broader information technology ecosystem in which they exist.

Acknowledgements

We would like to thank Loren Terveen, Darren Gergle, GroupLens Research, our participants and our anonymous reviewers for their support in the preparation of this manuscript. This research was funded in part by NSF CAREER IIS-1707296 and by NSF IIS-1707319.

References

Amazon.com, Inc. 2016. "Amazon.com Help: Ask Alexa."

Ananny, Mike, Karrie Karahalios, and Christo Wilson. 2015. "ICWSM Workshop: Auditing Algorithms From the Outside: Methods and Implications." In *ICSWSM '15*.

De Rijke, Maarten. 2016. "SIGIR 2016 Tutorial on Online Learning to Rank." In *ACM SIGIR '16*.

DeMers, Jayson. 2015. "Is The Google Knowledge Graph Killing Wikipedia? - Forbes." *Forbes*.

Devlin, Mark. 2015. "Blog: Google and Wikipedia: Best Friends Forever" *Newslines Blog*.

Dewey, Caitlin. 2016. "You Probably Haven't Even Noticed Google's Sketchy Quest to Control the World's Knowledge." *The Washington Post*.

Dischler, Jerry. 2015. "Building for the next Moment." *Inside AdWords*. May 5.

Eslami, Motahhare et al. 2016. "First I 'like' It, Then I Hide It: Folk Theories of Social Feeds." In *ACM SIGCHI '16*.

Eslami, Motahhare et al. 2015. "'I Always Assumed That I Wasn't Really That close to [Her]': Reasoning about Invisible Algorithms in News Feeds." In *ACM SIGCHI '15.*

Ford, H., and Graham, M. 2016. Semantic Cities: Coded Geopolitics and the Rise of the Semantic Web. In *Code and the City*. eds. Kitchin, R., and Perng, S-Y. London: Routledge.

Ge, Yanbo, et al. 2016. "Racial and Gender Discrimination in Transportation Network Companies." Working Paper 22776. *National Bureau of Economic Research.*

Google. 2016. "Google Trends Report." https://www.google.com/trends/topcharts#date=201602.

Hodson, Hal. 2016. "Google Wants to Rank Websites Based on Facts Not Links". *New Scientist.*

Hovy, Eduard, Roberto Navigli, and Simone Paolo Ponzetto. 2013. "Collaboratively Built Semi-Structured Content and Artificial Intelligence: The Story so Far." *JAIR*, 194 (January): 2–27.

Johnson, Isaac, et. al. 2016. "Not at Home on the Range: Peer Production and the Urban/Rural Divide." In *CHI '16*.

Kay, Matthew, Cynthia Matuszek, and Sean A. Munson. 2015. "Unequal Representation and Gender Stereotypes in Image Search Results for Occupations." In *ACM SIGCHI '15*.

Keyes, Oliver S.B. 2015. "Google Referrer Data to Wikipedia." The Wikimedia Foundation.

Kittur, Aniket, and Robert E. Kraut. 2008. "Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination." In *ACM CSCW '08.*

Lardinois, Frederic. 2016. "Apple Updates Siri With Twitter, Wikipedia, Bing Integration, New Commands And Male And Female Voices." *TechCrunch*.

Lund, Arwid, and Juhana Venäläinen. 2016. "Monetary Materialities of Peer-Produced Knowledge: The Case of Wikipedia and Its Tensions with Paid Labour." *tripleC.* 14 (1): 78–98.

McGee, Matt. 2012. "Wikipedia Appears On Google's Page One 46 Percent Of The Time, Compared to 31 Percent On Bing [STUDY]." *Search Engine Land*.

Menking, Amanda, and Ingrid Erickson. 2015. "The Heart Work of Wikipedia: Gendered, Emotional Labor in the World's Largest Online Encyclopedia." In *ACM SIGCH '15.*

Oeldorf-Hirsch, Anne, et al. 2014. "To Search or to Ask: The Routing of Information Needs Between Traditional Search Engines and Social Networks." In *ACM CSCW '14.*

Pan, Bing et al. "In Google We Trust: Users' Decisions on Rank, Position, and Relevance." *JCMC* 12 (3): 801–823.

Petrescu, Philip. 2014. "Google Organic Click-Through Rates in 2014." *Moz*.

Singhal, Amit. 2012. "Introducing the Knowledge Graph: Things, Not Strings." *Google: Official Blog*.

Silverstein, Craig, et al. "Analysis of a Very Large Web Search Engine Query Log." In *ACM SIGIR Forum*, 33:6–12. ACM, 1999.

Song, Yang, Xiaolin Shi, and Xin Fu. 2013. "Evaluating and Predicting User Engagement Change with Degraded Search Relevance." *WWW '13*.

Taraborelli, Dario. 2015. "The Sum of All Human Knowledge in the Age of Machines: A New Research Agenda for Wikimedia." *ICWSM-15 Workshop on Wikipedia, a Social Pedia.*

Thebault-Spieker, Jacob, Loren G. Terveen, and Brent Hecht. 2015. "Avoiding the South Side and the Suburbs: The Geography of Mobile Crowdsourcing Markets." In *ACM CSCW '15.*

Thebault-Spieker, Loren Terveen, and Brent Hecht. 2017. "Towards a Geographic Understanding of the Sharing Economy: Systemic Biases in UberX and TaskRabbit." *ACM TOCHI*.

Wales, Jimmy. 2004. "Wikipedia Founder Jimmy Wales Responds - Slashdot." *Slashdot.*

Wales, Jimmy. 2016. "User talk:Jimbo Wales." *Wikipedia, the Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=User_talk:Jimbo_Wales&oldid=722157311.

Warncke-Wang, Morten, et al. "Misalignment Between Supply and Demand of Quality Content in Peer Production Communities." *ICWSM 2015.*

Wikimedia Foundation. "2014-2015 Fundraising Report."

Wulczyn, Ellery. (2016). Wikipedia Navigation Vectors. figshare. doi:10.6084/m9.figshare.3146878

Zhu, Haiyi et al. 2013. "Effects of Peer Feedback on Contribution: A Field Experiment in Wikipedia." In *ACM SIGCHI '13.*