

The Dynamics of Peer-Produced Political Information During the 2016 U.S. Presidential Campaign

Brian C. Keegan

Department of Information Science
University of Colorado Boulder
Boulder, CO 80304 USA

Abstract

Wikipedia plays a crucial role for online information seeking and its editors have a remarkable capacity to rapidly revise its content in response to current events. How did the production and consumption of political information on Wikipedia mirror the dynamics of the 2016 U.S. Presidential campaign? Drawing on systems justification theory and methods for measuring the enthusiasm gap among voters, this paper quantitatively analyzes the candidates' biographical and related articles and their editors. Information production and consumption patterns match major events over the course of the campaign, but Trump-related articles show consistently higher levels of engagement than Clinton-related articles. Analysis of the editors' participation and backgrounds show analogous shifts in the composition and durability of the collaborations around each candidate. The implications for using Wikipedia to monitor political engagement are discussed.

Introduction and Background

How was information about political candidates produced and consumed on Wikipedia during and following the 2016 U.S. presidential campaign? This study examines the dynamics of Trump's and Clinton's biographical and related articles and the users who revised them during the 2016 U.S. presidential campaign. Data on 375,315 revisions to Clinton's biographical and related articles and 366,268 revisions for Trump were retrieved from the English Wikipedia and analyzed along with the contribution histories of the 2,211 most active editors. The results find evidence of a early, significant, and sustained gap in enthusiasm favoring Trump's biographical and related articles over Clinton's articles. All articles showed peaks of activity corresponding to major events over the course of the campaign, but Trump's articles consistently had significantly more revisions, page views, editors, and content than Clinton's articles throughout the campaign. The composition of Trump's article collaborations saw a greater influx of users even though there was substantial overlap in the set of editors contributing to both. Engagement with this political content was likewise disruptive: active editors who began contributing during the campaign had significant changes in their editing behaviors compared to their editing beforehand.

Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Wikipedia's editors and content are extremely responsive to current events (Keegan, Gergle, and Contractor 2011; Keegan 2013). Wikipedia's coverage of candidates, elections, and officeholders show a high level of accuracy when a relevant article exists, but there are many gaps and omissions, especially for older and less prominent topics (Brown 2011). Although the supply and demand for quality content on Wikipedia is misaligned (Warncke-Wang et al. 2015), political campaigns have realized Wikipedia's influence and proactively edit articles in advance of elections and announcements (Göbel and Munzert 2017). The relative quantity and timing of how political information is produced and consumed over the course of a campaign will reveal emergent patterns for subsequent inductive analysis.

RQ1: How does Wikipedia's production and consumption of political information vary during campaigns?

Predicting election outcomes with social media activity data is fraught (Gayo-Avello 2013; Jungherr et al. 2016), but Wikipedia data can forecast overall turnout and changes in vote share for parties (Yasseri and Bright 2016). Wikipedians who self-identify with political ideologies do not exhibit polarized editing behaviors of avoiding collaboration with opponents or preferentially collaborating with allies (Neff et al. 2013). In democratic systems, the success of political candidates clearly hinges on voter participation: campaigns must motivate voters to cast ballots rather than staying home. But minor deviations in the same voters' turnout across different elections — also known as the *enthusiasm gap* — can produce large changes in first-past-the-post electoral systems (Hill 2014). Wikipedia may reproduce these enthusiasm gaps through the relative differences in the ability for each candidate to mobilize user engagement with its campaign content: a candidate motivating their "own" users to engage with their Wikipedia content relatively more than their opponent's editors may reflect differences in enthusiasm among the general electorate.

RQ2: Who are the Wikipedia editors revising information about candidates during campaigns?

Label	Description	Date
A	Clinton announces candidacy	12 April 2015
B	Trump announces candidacy	13 June 2015
C	Super Tuesday primary elections	3 March 2016
D	“Acela” primary elections	29 April 2016
E	Trump secures nomination	27 May 2016
F	Clinton secures nomination	8 June 2016
G	Republican National Convention	11 July 2016
H	Democratic National Convention	28 July 2016
I	Access Hollywood controversy	7 October 2016
J	Election day	8 November 2016
K	Inauguration	20 January 2016

Table 1: Annotated events in Figures 1 and 2.

Data

We analyze both the *biographical articles* about the candidates (“Hillary Clinton” and “Donald Trump”) as well as the *related articles* that are members or children of the Wikipedia categories for each candidate: 1,336 for Clinton and 949 for Trump. The time range spanning 1 January 2015 through 9 November 2017 was selected for detailed analysis, although the revision data goes back more than a decade beforehand for all articles. The revision history for each English Wikipedia biographical and related article was retrieved from the Wikipedia’s API using the “Revisions” endpoint.¹ This generated a corpus of 375,315 revisions for Clinton’s biographical and related articles and 366,268 revisions for Trump. 57,944 (15.4%) of Clinton’s revisions and 77,110 (21.1%) of Trump’s revisions occurred over the course of the campaign. Page view activity is broken down by user type and platform, but we report on the aggregated “all-agent” and “all-access” statistics. We extracted a sub-sample of 2,211 *active editors* who made at least five unique revisions, contributed to at least three pages, and were active for more than one day from the Wikipedia API using the “Usercontribs” endpoint.²

Results

Biographical Articles

RQ1 asked “How does Wikipedia’s production and consumption of political information vary during campaigns?” Focusing on each candidate’s biographical article, we analyzed data about changes in the revisions, page size, page views, and page protections over the course of the campaign.

Revisions. Over the history of their articles (through 9 November 2017), Donald Trump’s article received an average of 4.2 ± 9.6 (max. 155) revisions per day while Hillary Clinton’s article received an average of 4.2 ± 5.8 (max. 172) revisions per day. Given the skewed distributions of daily activity and heteroskedastic variance in the samples examined throughout this paper, we employ a non-parametric Kruskal-Wallis H -test to test the null hypothesis that the medians of Trump’s and Clinton’s daily revision activity distributions are identical. The H -test for daily revisions was

¹<https://www.mediawiki.org/wiki/API:Revisions>

²<https://www.mediawiki.org/wiki/API:Usercontribs>

$105.66, p < 0.001$). Over the course of the campaign (1 June 2015 through 8 November 2016), Trump’s article received an average of 16.5 ± 15.9 (max. 121) revisions per day while Clinton’s article received an average of 4.2 ± 5.9 (max. 56) revisions per day ($H = 295.96, p < 0.001$).

Figure 1 visualizes the daily revisions to each candidate’s biographical article over the course of the campaign. Ten major events are annotated with descriptions given in Table 1. Three additional bursts of activity are annotated with stars (*) that do not correspond to major exogenous events, but endogenous bursts of revisions from single users making many sequential changes.

Page size. Over the history of each candidate’s biographical articles, Clinton’s article had a median size of 167 kB compared to Trump’s median size of 52 kB ($H = 510.35, p < 0.001$). While Clinton’s article was significantly larger than Trump’s preceding the 2016 campaign (241 kB vs. 106 kB), Trump’s article more than tripled in size over the course of the campaign, growing to 342 kB on the day of the election compared to Clinton’s 280 kB. Clinton’s article had a median size of 272 kB, which was significantly smaller than Trump’s median size of 286 kB during the campaign ($H = 45.08, p < 0.001$).

Editors. Clinton’s article had more cumulative unique editors (3,652) than Trump (3,432) when he announced his campaign (point *B*). Between the start of Trump’s campaign and Election day, the cumulative number of unique editors on Trump’s article grew to 4,773 editors (39.6% increase) compared to Clinton’s article growing to 4,145 editors (13.8% increase). The composition and differences in these editor sets are explored in more detail in a later section.

Page views. Clinton’s article (and redirects) received a median of 41,787 page views compared to Trump’s 159,283 page views ($H = 460.6, p < 0.001$). Clinton received 19,535,002 page views and Trump received 73,116,431 over the course of the campaign. Figure 2 visualizes the number of daily page views to the Clinton and Trump articles. The same 10 events are annotated with descriptions given in Table 1. There were only 3 date ranges when Clinton’s page view activity surpasses Trump’s: mid-October 2015, early June 2016 after she clinched the Democratic nomination, and during the Democratic National Convention in July.

Related Articles

Did the successful candidate mobilize greater information production and consumption activity for the candidates’ *related articles*? There is a common pattern of Trump-related articles having higher levels of activity than Clinton-related articles by Election Day (point *J* in Figure 3).

New article creation. Most related articles predate the start of the campaign, but many were created after the campaign began as Wikipedia editors fill in additional details about the campaign, people, events, and controversies beyond the scope of the candidates’ biographies. In the period starting after 1 January 2015, 84 Clinton-related articles and 692 Trump-related articles were created.

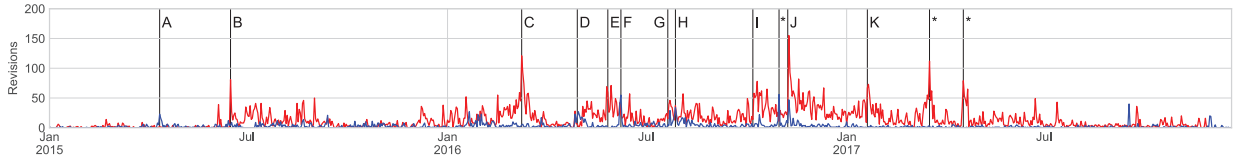


Figure 1: The daily revisions to the Clinton (blue) and Trump (red) biographical articles.

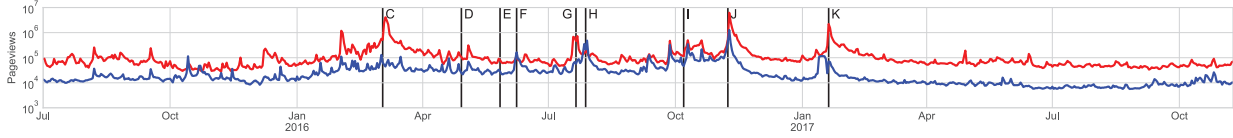


Figure 2: The daily page views to the Clinton (blue) and Trump (red) biographical articles.

Revisions. Among all (new) related articles between 1 January 2015 and 7 November 2016, Clinton’s articles received 43,010 (3,602) revisions and Trump’s articles received 55,232 (26,557) revisions. Over the course of the campaign itself, from Trump’s 2015 announcement through 7 November 2016 (to exclude the bursts of activity surrounding from the day of the election), Clinton’s all (new) related articles had accumulated an average of 28.2 (41.1) revisions compared to Trump’s 55.0 (38.2) revisions ($H_{all} = 267.0, p_{all} < 0.001, H_{new} = 82.6, p_{new} < 0.001$).

Editors. Trump’s related articles attracted significantly more unique editors over the course of the campaign than Clinton’s related articles. Over the course of the campaign itself, Clinton’s all (new) related articles had accumulated an average of 14.1 (49.2) editors compared to Trump’s 18.5 (50.1) editors ($H_{all} = 129.6, p_{all} < 0.001, H_{new} = 4.3, p_{new} = 0.037$).

Size. Trump’s related articles accumulated significantly more content than Clinton’s related articles over the course of the campaign. Over the course of the campaign itself, Clinton’s all (new) related articles had accumulated an average of 1,573 (4,360) bytes of content compared to Trump’s 3,406 (3,070) bytes ($H_{all} = 0.5, p_{all} = 0.48, H_{new} = 81.9, p_{new} < 0.001$).

Page views. Trump’s related articles attracted significantly more page views over the course of the campaign than Clinton’s related articles. Among all (new) related articles from 1 July 2015 through 7 November 2016, Clinton’s articles received 153.5 million (2.47 million) page views and Trump’s articles accumulated 147.8 million (17.3 million) page views. Over the course of the campaign, Clinton’s all (new) related articles had accumulated an average of 123,519 (30,498) total page views compared to Trump’s 163,686 (26,015) total page views ($H_{all} = 352.3, p_{all} < 0.001, H_{new} = 81.0, p_{new} < 0.001$).

Editor dynamics

RQ2 asked “Who are the editors revision information about candidates during campaigns?” The composition of collab-

	<i>Before</i>	<i>After</i>	<i>H-test</i>
Comments	1852 ± 5294	5399 ± 11952	215.4 *
Dates active	153 ± 181	264 ± 214	246.8 *
Entropy	4.7 ± 1.8	5.9 ± 1.8	198.5 *
Latency (days)	0.39 ± 5.4	0.11 ± 1.5	24.0 *
Pages created	183 ± 1437	497 ± 1622	77.4 *
Pages edited	944 ± 4098	3096 ± 8033	217.5 *
Namespaces	6.5 ± 3.6	8.4 ± 4.2	103.8 *
Revisions	2187 ± 5772	5980 ± 12479	226.3 *
Rev. size (bytes)	69.7 ± 384.8	60.4 ± 125.1	0.7

Table 2: Average user contribution behavior before and after first revision to campaign article (*, $p < 0.001$)

orators contributing to Trump’s biographical article change dramatically at three distinct points in time. The first discontinuity was in 2011 (annotated with a * in Figure 4) during Trump’s discredited accusations about Obama’s birth certificate. The fraction of revisions from Trump and Clinton related articles increased substantially reflecting an influx of editors to his biographical article. The second discontinuity happened following the announcement of Trump’s 2016 candidacy (point A). The shift in the composition of the editors on his biographical article accelerated as contributors to Trump related, Clinton related, and Clinton’s biographical article. The third discontinuity corresponds with Trump’s victory (point J) as the contributions from these other sets of editors stabilized above 60% of the total revisions.

Predecessor and successor collaborations. How did active editors’ behavior change after their first revision to a campaign article? The contributions histories from 1 January 2014 through 9 November 2017 for 1,075 active users who made their first revision to a candidate’s biographical or related article after 1 January 2015 were retrieved and analyzed to compare their contribution history *after* their first revision to a Clinton or Trump article to their contributions *before* this first revision. This analysis uses an active editor’s first contribution to these candidates’ biographical or related articles as a discontinuity to test the changes in behavior before and after this expression of interest in editing political content during a campaign.

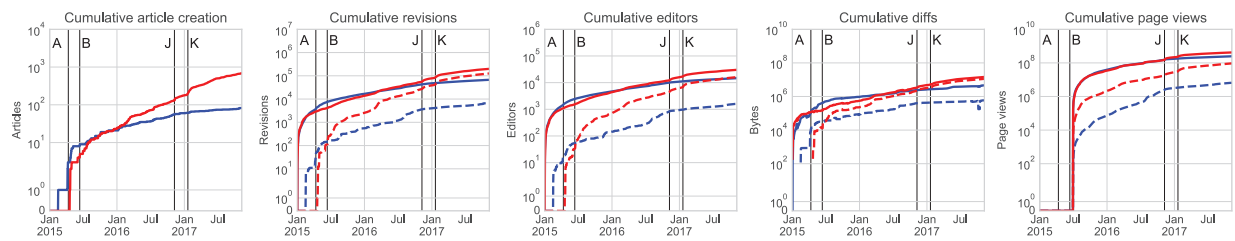


Figure 3: Cumulative activity on related articles for Clinton (blue) and Trump (red). Solid lines for *all* child articles and dashed lines for *new* articles created after 1 January 2015. Page view data is only available after 1 July 2015.

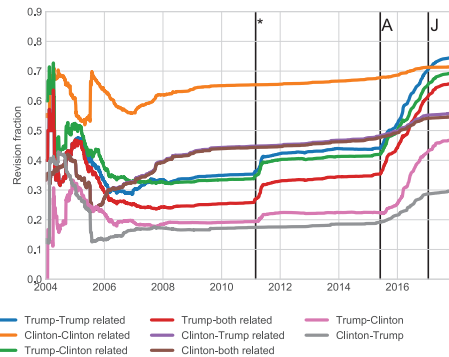


Figure 4: Revision fraction on biographical articles among editors sets.

Table 2 summarizes the average active editors’ contribution behavior before and after their first revision to a candidate’s biographical or related article. Following their first “political” contribution, active editors make significantly more comments, are active on more days, increase the entropy of revisions made across articles, reduce the latency between successive edits, create more pages, edit more pages, contribute in more namespaces, and make more revisions in the period afterwards. The the average size of their individual revisions does not change significantly.

Discussion

How was information about political candidates produced and consumed on Wikipedia during and following the 2016 U.S. presidential campaign? We explored this research question through three levels of analysis: the dynamics of Clinton and Trump’s biographical articles, the dynamics of their related articles, and the dynamics of the editors who contributed to them during the campaign. The biographical articles showed high levels of information production and consumption activity responded to major events over the course of the 2016 campaign with a focus on announcements, primary elections, and the party conventions. The gap in information production and consumption on the candidates’ biographical articles likewise extended to the related articles about each candidate. Over the course of the campaign between January 2015 and November 2016, Wikipedians created more new articles, made more revisions, generated larger

collaborations, made larger articles, and viewed more articles about Trump than Clinton. A similar and significant gap in information production and consumption favoring Trump over Clinton also unfolded for new articles created during the campaign. The composition of the editors on the candidates’ biographical and related articles likewise showed substantial changes over the course of the campaign. The contributions from these overlapping editors made up a majority of the revisions made to these articles by the end of the campaign (Figure 4). Finally, there were significant differences (Table 2) among active editors who began editing during the campaign between their contribution behavior before and after their first edit to a candidate’s biographical or related article.

References

- Brown, A. R. 2011. Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage. *PS: Political Science & Politics* 44(2):339–343.
- Gayo-Avello, D. 2013. A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data. *Social Science Computer Review* 31(6):649–679.
- Göbel, S., and Munzert, S. 2017. Political Advertising on the Wikipedia Marketplace of Information. *Social Science Computer Review* 1–19.
- Hill, S. J. 2014. A behavioral measure of the enthusiasm gap in American elections. *Electoral Studies* 36:28–38.
- Jungherr, A.; Schoen, H.; Posegga, O.; and Jürgens, P. 2016. Digital Trace Data in the Study of Public Opinion An Indicator of Attention Toward Politics Rather Than Political Support. *Social Science Computer Review* 35(3):336–356.
- Keegan, B.; Gergle, D.; and Contractor, N. 2011. Hot off the Wiki: Dynamics, Practices, and Structures in Wikipedia’s Coverage of the Tōhoku Catastrophes. In *Proc. WikiSym ’11*. ACM.
- Keegan, B. 2013. A History of Newswork on Wikipedia. In *Proc. WikiSym ’13*. ACM.
- Neff, J. J.; Laniado, D.; Kappler, K. E.; Volkovich, Y.; Aragón, P.; and Kaltenbrunner, A. 2013. Jointly They Edit: Examining the Impact of Community Identification on Political Interaction in Wikipedia. *PLOS ONE* 8(4).
- Warncke-Wang, M.; Ranjan, V.; Terveen, L. G.; and Hecht, B. J. 2015. Misalignment Between Supply and Demand of Quality Content in Peer Production Communities. In *Proc. ICWSM ’15*. AAAI.
- Yasseri, T., and Bright, J. 2016. Wikipedia traffic data and electoral prediction: Towards theoretically informed models. *EPJ Data Science* 5(1):1–15.