

## Wikipedia: Wikipedia Signpost/2019-06-30/Recent research

< Wikipedia: Wikipedia Signpost | 2019-06-30

Recent research

# What do editors do after being blocked?; the top mathematicians, universities and cancers according to Wikipedia

By FULBERT and Tilman Bayer

**Contribute** — Share this



A monthly overview of recent academic research about Wikipedia and other Wikimedia projects, also published as the Wikimedia Research Newsletter.

### "Trajectories of Blocked Community Members"

Reviewed by FULBERT

In Chang and Danescu-Niculescu-Mizil's study "Trajectories of Blocked Community Members: Departure", [1] Redemption, Recidivism and presented at The Web Conference last month, the authors explored the effectiveness of temporary editor blocks imposed as acts of moderation for inappropriate conduct on Wikipedia. researchers wanted to understand how effective these blocks were as a tool intended to reform editors whose online conduct was problematic enough to warrant discipline, but not so egregious

PDF DOWNLOAD (HTTPS://EN.WIKIP EDIA.ORG/API/RES T V1/PAGE/PDF/WI KIPEDIA%3AWIKIPE DIA%20SIGNPOS T%2F2019-06-30%2 FRECENT%20RESE ARCH) E-MAIL (MAILT O:?SUBJECT=WHA T%20DO%20EDITO RS%20DO%20AFTE R%20BEING%20BL OCKED%3F%3B%2 0THE%20TOP%20M ATHEMATICIANS%2 C%20UNIVERSITIE S%20AND%20CAN CERS%20ACCORDI NG%20TO%20WIKI PEDIA&BODY=WHA T%20DO%20EDITO RS%20DO%20AFTE R%20BEING%20BL OCKED%3F%3B%2 0THE%20TOP%20M ATHEMATICIANS%2 C%20UNIVERSITIE S%20AND%20CAN CERS%20ACCORDI NG%20TO%20WIKI PEDIA%0AHTTPS://

as to impose an indefinite block, i.e. where there was hope that this action may help reform the editors who demonstrate evidence they can be productive ongoing contributors to Wikipedia.

The researchers limited their study to the first block received by users, beginning with a total sample of 104,245 blocks over the history of English Wikipedia. They focused on four types of blocks that enforced community norms — personal attacks and incivility, harassment, edit warring, and disruptive editing — arriving at a dataset comprising 6,026 blocked users. They were particularly interested in editor departure, which they defined as the time of the last comment a user makes on community talk pages, in part because they focused their attention on community engagement. Likewise, they explored recidivism, which is when a previously-blocked user is blocked again for any additional breach of community rules, compared to those who reform and are never blocked again.

They found that "users are much more likely to depart if they were blocked before, especially in the first months of their life in the community (p. 4)". Indeed, users are more likely to depart during the month of their block, and this is compounded by the finding that users who were blocked a first time had an increased probability of being blocked again. These findings were impacted by the length of time one was an active member of the user community, where the greater the length of one's involvement, the less likely blocks were to occur, especially through recidivism. The authors then explored user activity level and activity spread, learning that recidivist users engage in interaction with members of the community with more depth than breadth, involving a smaller set of users while continuing more prolonged engagement. Finally, the researchers explored blocked users' perception of fairness, focused through blockage appeals. Those who apologized for their infractions were less likely to suffer from recidivism, while those who directly questioned their blocks or claimed a lack of fairness instead faced a moderate increase in recidivism.

- EN.WIKIPEDIA.OR G/WIKI/WIKIPEDIA: WIKIPEDIA\_SIGNP OST/2019-06-30/RE CENT\_RESEARCH)
- MASTODON (HTT) PS://FEDI-SHARE.T OOLFORGE.ORG/?J SON=%22WHAT+D O+EDITORS+DO+A FTER+BEING+BLO CKED%3F%3B+THE +TOP+MATHEMATI CIANS%2C+UNIVE RSITIES+AND+CAN CERS+ACCORDING +TO+WIKIPEDIA+H TTPS%3A%2F%2FE N.WIKIPEDIA.OR G%2FWIKI%2FWIKI PEDIA%3AWIKIPED IA SIGNPOST%2F2 019-06-30%2FRECE NT RESEARCH+%2 3WIKIPEDIASIGNP OST+%23WIKIPEDI A%22)
- INKEDIN (HTTP S://LINKEDIN.COM/SHAREARTICLE?MINI=TRUE&URL=HTT PS://EN.WIKIPEDIA.ORG/WIKI/WIKIPEDIA:WIKIPEDIA\_SIGN POST/2019-06-30/RECENT\_RESEARCH)
- FACEBOOK (HTT) PS://FACEBOOK.CO M/SHARER.PHP?U= HTTPS%3A%2F%2F **EN.WIKIPEDIA.OR** G%2FWIKI%2FWIKI PEDIA%3AWIKIPED IA SIGNPOST%2F2 019-06-30%2FRECE NT RESEARCH&T= WHAT+DO+EDITOR S+DO+AFTER+BEIN G+BLOCKED%3F% 3B+THE+TOP+MAT **HEMATICIANS%2C+** UNIVERSITIES+AN D+CANCERS+ACC

The authors concluded that working toward a "more nuanced approach to moderation that more broadly accounts for the tradeoffs between possible outcomes and that considers how the affected individuals might perceive the moderators and their actions (p. 10)" is something that should be considered when moderators (administrators) consider blocking users for transgressions to community standards.

#### **Conferences and events**

See the research events page on Meta-wiki for upcoming conferences and events, including submission deadlines, and the page of the monthly Wikimedia Research Showcase for videos and slides of past presentations.

#### Other recent publications

Other recent publications that could not be covered in time for this issue include the items listed below. Contributions, whether reviewing or summarizing newly published research, are always welcome.

Compiled by Tilman Bayer

### Hilbert and Newton top the "Wikipedia Network of Mathematicians"

From the abstract:<sup>[2]</sup>

"We look at the network of mathematicians defined by the hyperlinks between their biographies on Wikipedia. [...] We illustrate how such Wikipedia data can be used by performing a centrality analysis. These measures show that Hilbert and Newton are the most important mathematicians."

ORDING+TO+WIKIP EDIA) X X (TWITTER) (HT TPS://TWITTER.CO M/SHARE?VIA=WIKI SIGNPOST&URL=H TTPS://EN.WIKIPEDI A.ORG/WIKI/WIKIPE DIA:WIKIPEDIA SIG NPOST/2019-06-30/ RECENT RESEARC H&TEXT=WHAT+DO +EDITORS+DO+AFT ER+BEING+BLOCK ED%3F%3B+THE+T **OP+MATHEMATICIA** NS%2C+UNIVERSIT IES+AND+CANCER S+ACCORDING+TO +WIKIPEDIA&RELAT ED=WIKISIGNPOST, WIKIPEDIA) REDDIT (HTTP) S://WWW.REDDIT.C OM/SUBMIT?URL=H TTPS%3A%2F%2FE N.WIKIPEDIA.OR G%2FWIKI%2FWIKI PEDIA%3AWIKIPED IA SIGNPOST%2F2 019-06-30%2FRECE NT RESEARCH&TI TLE=WHAT+DO+ED ITORS+DO+AFTER+ BEING+BLOCKED% 3F%3B+THE+TOP+

MATHEMATICIAN

S%2C+UNIVERSITI ES+AND+CANCERS

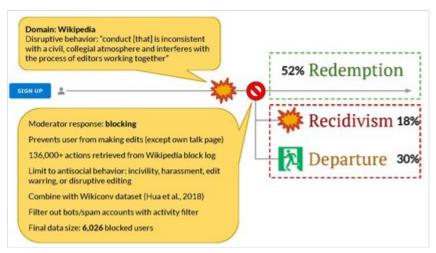
+ACCORDING+TO+

WIKIPEDIA)

Oxford, Cambridge and Harvard top the "Wikipedia Ranking of World Universities"

From the abstract and paper:[3]

"We present Wikipedia Ranking of World Universities (WRWU) based on analysis of networks of 24 Wikipedia editions collected in May 2017. With PageRank and CheiRank algorithms we determine ranking universities averaged over cultural views of these editions. The comparison with the Shanghai ranking gives overlap of 60% for 100 universities top showing that WRWU gives more significance to their historical development. [... Our] approach also determines the influence of specific universities on world countries. also We compare different cultural views of Wikipedia editions on significance influence and universities." " The top 3 places of WRWU are occupied by Oxford, Cambridge and Harvard while for ARWU [the Shanghai ranking] it is Harvard, Stanford and Cambridge."



Slides from a Wikimedia Research Showcase presentation about the paper



Linguistic indicators of how fair the blocked user perceived the sanction to have been (from showcase presentation slides)

An earlier, related preprint [4] by some of the same authors was covered by <u>Technology Review</u> under the headline "Wikipedia-Mining Algorithm Reveals World's Most Influential Universities (http://www.technologyreview.com/view/544266/wikipedia-mining-algorithm-reveals-worlds-most-influential-universities/)".

### C is the top programming language and Cambridge the top university for computer science, according to (some) Wikipedia network ranking

From the abstract and paper: [5]

We construct 10 different networks from Wikipedia entries (articles) related to the chosen domain. The goal of the experiment is to extract domain knowledge in terms of identifying entries that are centrally positioned and entries that are strongly related. [...] the first task is the construction of a web scraping program which would extract hyperlinks from a Wikipedia entry's text. The hyperlinks are extracted using a Python package for HTML parsing called Beautiful Soup which parses the HTML structure of a given HTML document into a parse tree. By navigating the tree we locate the tag ID which corresponds to article content ("mw-content-text") and proceed to extract the hyperlinks which themselves are found within paragraph () tags and finally inside link (<a>) tags in that section of the page. "

The various resulting rankings place e.g. "<u>University of Cambridge</u>" 2nd (after "human") by <u>degree centrality</u> in the article network for "<u>computer science</u>", and "<u>C (programming language</u>)" 2nd by <u>closeness centrality</u> in the article network for "programming language" (after "programming language" itself).

#### "Top 100 historical figures of Wikipedia"

From the abstract: [6]

"The top 100 historical figures of Wikipedia are determined on the basis of statistical methods and mathematical algorithms like PageRank, CheiRank and 2DRank applied to networks of Wikipedia in up to 24 languages. ... This short popular note presents overview of results and methods of different [research] groups..."

#### "Wikipedia network analysis of cancer interactions and world influence"

From the abstract:<sup>[7]</sup>

"We apply the <u>Google matrix</u> algorithms for analysis of interactions and influence of 37 cancer types, 203 cancer drugs and 195 world countries using the network of 5 416 537 English Wikipedia articles with all their directed hyperlinks. The PageRank algorithm provides the importance order of cancers which has 60% and 70% overlaps with the top 10 cancers extracted from World Health Organization GLOBOCAN 2018 and Global Burden of Diseases Study 2017, respectively. [...] We argue that this analysis of knowledge accumulated in Wikipedia provides useful complementary global information about interdependencies between cancers, drugs and world countries."

The resulting ranking is lead by lung cancer, breast cancer and leukemia.

See also earlier coverage of a related paper by the same authors: "World Influence of Infectious Diseases from Wikipedia Network Analysis"

### Facebook AI researchers construct the "Wizard of Wikipedia: Knowledge-Powered Conversational agents"

From the abstract: [8]

"We collect and release a large dataset with conversations directly grounded with knowledge retrieved from Wikipedia. We then design architectures capable of retrieving knowledge, reading and conditioning on it, and finally generating natural responses"

*See also OpenReview discussion (https://openreview.net/forum?id=r1173iRqKm)* 

### "The dynamics of Wikipedia article revisions: an analysis of revision activities and patterns"

From the abstract: [9]

"We identify four revision patterns: 1) revision actions at the sentence and link levels appear in similar paces; 2) the numbers of revision actions at sentence and link levels comparatively evenly grow with the article's age prior to the last time period; 3) the paces of media and reference-level actions tend to be lagged behind sentence and link-level actions; 4) before being promoted to the GA or FA rank, articles nominated to the GA or FA rank exhibit a significant rising pattern in amounts of revisions and revision actions."

### "Interactive Quality Analytics of User-generated Content: An Integrated Toolkit for the Case of Wikipedia"

From the abstract: [10]

"Great efforts have been invested in algorithmic methods for automatic classification of Wikipedia articles (as featured or non-featured) and for quality flaw detection. Instead, our contribution is an interactive tool that combines automatic classification methods and human interaction in a toolkit, whereby experts can experiment with new quality metrics and share them with authors that need to identify weaknesses to improve a particular article. A design study shows that experts are able to effectively create complex quality metrics in a visual analytics environment. In turn, a user study evidences that regular users can identify flaws, as well as high-quality content based on the inspection of automatic quality scores."

### "The Impact of Topic Characteristics and Threat on Willingness to Engage with Wikipedia Articles: Insights from Laboratory Experiments"

From the abstract and paper: [11]

"We presented the introduction parts of 20 Wikipedia articles and asked participants to rate each article with respect to familiarity and controversiality. In addition, we experimentally manipulated participants' level of mortality salience in terms of the amount of threat they experienced when reading the article. Participants also indicated their willingness to engage with a particular article. The results revealed that willingness to engage with a Wikipedia article was predicted by both topic familiarity and controversiality of a given article. Although mortality salience increased accessibility of death-related thoughts, it did not result in any changes in people's willingness to work with the articles. [...] Sixty-one participants reported that they read Wikipedia articles at least once a week. They indicated an average reading time of M = 1.57 h per week (SD = 1.11)."

#### "With Few Eyes, All Hoaxes Are Deep"

From the abstract: [12]

"we describe a long-standing set of inefficiencies that have plagued new page patrolling by drawing a contrast to the more efficient, distributed processes for counter-vandalism. Further, to address this issue, we demonstrate an effective automated topic model based on a labeling strategy that leverages a folksonomy developed by subject specific working groups in Wikipedia (WikiProject tags) and a flexible ontology (WikiProjects Directory) to arrive at a hierarchical and uniform label set. We are able to attain very high fitness measures [...] and real-time performance using word2vec-based features. Finally, we present a proposal for how incorporating this model into current tools will shift the dynamics of new article review positively."

### "Finding Prerequisite Relations using the Wikipedia Clickstream"

From the abstract: [13]

"Where should the learner start? What should the learner know before tackling a new course? [...] we present a new method for identifying prerequisite relations based on naturally occurring data, namely the navigation patterns of users on the Wikipedia online encyclopedia. Our supervised learning approach shows that the navigation network structure can be used to identify dependencies among concepts in several domains."

### "Query for Architecture, Click through Military: Comparing the Roles of Search and Navigation on Wikipedia"

From the abstract: [14]

"...we study large-scale article access data of the English Wikipedia in order to compare articles with respect to the two main paradigms of information seeking, i.e., search by formulating a query, and navigation by following hyperlinks. To this end, we propose and employ two main metrics, namely (i) searchshare -- the relative amount of views an article received by search --, and (ii) resistance -- the ability of an article to relay traffic to other Wikipedia articles -- to

characterize articles. We demonstrate how articles in distinct topical categories differ substantially in terms of these properties. For example, architecture-related articles are often accessed through search and are simultaneously a 'dead end' for traffic, whereas historical articles about military events are mainly navigated."

### "How Article Topic, Quality and Dwell Time Predict Banner Donation on Wikipedia" From the abstract: [15]

"In this work, using a dataset of aggregated donation information from Wikimedia's 2015 fundraising campaign, representing nearly 1 million pages from English and French language versions of Wikipedia, we explore the relationship between the properties of contents of a page and the number of donations on this page. Our results suggest the existence of a reciprocity mechanism, meaning that articles that provide more utility value attract a higher rate of donation. We discuss these and other findings focusing on the impact they may have on the design of banner-based fundraising campaigns."

#### References

- 1. Chang, Jonathan; Danescu-Niculescu-Mizil, Cristian (2019). "Trajectories of Blocked Community Members: Redemption, Recidivism and Departure" (http://dl.acm.org/citation.cfm?doid=3308558.3313638). The World Wide Web Conference. WWW '19. San Francisco, CA, USA: ACM Press. pp. 184–195. arXiv:1902.08628 (https://arxiv.org/abs/1902.08628). Bibcode:2019arXiv190208628C (https://ui.adsabs.harvard.edu/abs/2019arXiv190208628C). doi:10.1145/3308558.3313638 (https://doi.org/10.1145%2F3308558.3313638). ISBN 9781450366748. S2CID 67855784 (https://api.semanticscholar.org/CorpusID:67855784).
- 2. Chen, Bingsheng; Lin, Zhengyu; Evans, Tim S. (2019-02-20). "Analysis of the Wikipedia Network of Mathematicians". <a href="mailto:arXiv:1902.07622">arXiv:1902.07622</a> (https://arxiv.org/abs/1902.07622) [cs.DL (https://arxiv.org/archive/cs.DL)].
- 3. Coquidé, Célestin; Lages, José; Shepelyansky, Dima L. (2019-01-09). "World influence and interactions of universities from Wikipedia networks". *The European Physical Journal B.* **92** (1): 3. arXiv:1809.00332 (https://arxiv.org/abs/1809.00332). Bibcode:2019EPJB...92....3C (https://ui.adsabs.harvard.edu/abs/2019EPJB...92....3C). doi:10.1140/epjb/e2018-90532-7 (https://doi.org/10.1140%2Fepjb%2Fe2018-90532-7). ISSN 1434-6036 (https://search.worldcat.org/issn/1434-6036). S2CID 52154548 (https://api.semanticscholar.org/CorpusID:52154548). [a], supplementary page: http://perso.utinam.cnrs.fr/~lages/datasets/WRWU17/

- 4. Lages, José; Patt, Antoine; Shepelyansky, Dima L. (2015-11-29). "Wikipedia Ranking of World Universities". <a href="mailto:arXiv:1511.09021">arXiv:1511.09021</a> (https://arxiv.org/abs/1511.09021) [cs.SI (https://arxiv.org/archive/cs.SI)].
- Matas, N.; Martincic-Ipsic, S.; Mestrovic, A. (May 2015). "Extracting domain knowledge by complex networks analysis of Wikipedia entries". 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). pp. 1622–1627. doi:10.1109/MIPRO.2015.7160531 (https://doi.org/10.1109%2FMIPRO.2015.7160531).
- 6. Shepelyansky, Dima (2015). <u>"Top 100 historical figures of Wikipedia" (https://hal.archives-ouvertes.fr/hal-01184245/document)</u>. {{cite journal}}: Cite journal requires | journal = (help)
- 7. Rollin, Guillaume; Lages, José; Shepelyansky, Dima (2019-01-23). "Wikipedia network analysis of cancer interactions and world influence". p. 527879. <a href="mailto:bioRxiv">bioRxiv</a> 10.1101/527879 (https://doi.org/10.1101%2F527879).
- 8. Dinan, Emily; Roller, Stephen; Shuster, Kurt; Fan, Angela; Auli, Michael; Weston, Jason (2018-11-03). "Wizard of Wikipedia: Knowledge-Powered Conversational agents". <a href="https://arxiv.org/abs/1811.01241">arXiv:1811.01241</a> (https://arxiv.org/abs/1811.01241) [cs.CL (https://arxiv.org/archive/cs. CL)].
- 9. Ma, Zhongming; Tao, Jie; Hu, Jing (2017-01-01). "The dynamics of Wikipedia article revisions: an analysis of revision activities and patterns". *International Journal of Data Mining, Modelling and Management.* **9** (4): 298–314. <a href="https://doi.org/10.1504%2FIJDMMM.2017.088415">doi:10.1504/IJDMMM.2017.088415</a> (https://doi.org/10.1504%2FIJDMMM.2017.088415). ISSN 1759-1163 (https://search.worldcat.org/issn/1759-1163).
- 10. Sciascio, Cecilia Di; Strohmaier, David; Errecalde, Marcelo; Veas, Eduardo (March 2019). "Interactive Quality Analytics of User-generated Content: An Integrated Toolkit for the Case of Wikipedia". *ACM Trans. Interact. Intell. Syst.* **9** (2–3): 13–1–13:42. doi:10.1145/3150973 (https://doi.org/10.1145%2F3150973). ISSN 2160-6455 (https://search.worldcat.org/issn/2160-6455).
- 11. Yenikent, Seren; Holtz, Peter; Kimmerle, Joachim (2017). "The Impact of Topic Characteristics and Threat on Willingness to Engage with Wikipedia Articles: Insights from Laboratory Experiments" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5681955). Frontiers in Psychology. 8: 1960. doi:10.3389/fpsyg.2017.01960 (https://doi.org/10.3389/fpsyg.2017.01960). ISSN 1664-1078 (https://search.worldcat.org/issn/1664-1078). PMC 5681955 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5681955). PMID 29163323 (https://pubmed.ncbi.nlm.nih.gov/29163323).

- Asthana, Sumit; Halfaker, Aaron (November 2018). "With Few Eyes, All Hoaxes Are Deep". *Proc. ACM Hum.-Comput. Interact.* 2 (CSCW): 1–18. doi:10.1145/3274290 (https://doi.org/10.1145%2F3274290). ISSN 2573-0142 (https://search.worldcat.org/issn/2573-0142). S2CID 53223190 (https://api.semanticscholar.org/CorpusID:53223190).
- 13. Mohsen Sayyadiharikandeh, Jonathan Gordon, José Luis Ambite, Kristina Lerman: "Finding Prerequisite Relations using the Wikipedia Clickstream" https://www.isi.edu/~ambite/papers/www-wiki-2019.pdf
- Dimitrov, Dimitar; Lemmerich, Florian; Flöck, Fabian; Strohmaier, Markus (2018). "Query for Architecture, Click through Military: Comparing the Roles of Search and Navigation on Wikipedia". WebSci '18 Proceedings of the 10th ACM Conference on Web Science. Association for Computing Machinery. pp. 371–380. arXiv:1805.04022 (https://arxiv.org/abs/1805.04022). Bibcode:2018arXiv180504022D (https://ui.adsabs.harvard.edu/abs/2018arXiv180504022D). doi:10.1145/3201064.3201092 (https://doi.org/10.1145%2F3201064.3201092). ISBN 978-1-4503-5563-6. S2CID 13705681 (https://api.semanticscholar.org/CorpusID:13705681).
- 15. Kocielnik, Rafal; Keyes, Os; Morgan, Jonathan T.; Taraborelli, Dario; McDonald, David W.; Hsieh, Gary (November 2018). "Reciprocity and Donation: How Article Topic, Quality and Dwell Time Predict Banner Donation on Wikipedia". *Proc. ACM Hum.-Comput. Interact.* 2 (CSCW): 91–1–91:20. <a href="doi:10.1145/3274360">doi:10.1145/3274360</a> (https://doi.org/10.1145%2F3274360). ISSN 2573-0142 (https://search.worldcat.org/issn/2573-0142). S2CID 53240544 (https://api.semanticscholar.org/CorpusID:53240544).

← PREVIOUS "Recent research"

NEXT "Recent research" →

+ Add a comment (https://en.wikipedia.org/w/index.php?editintro=Wikipedia:Wikipedia\_Signpost/Templates/Comment-editnotice&title=Wikipedia\_talk:Wikipedia\_Signpost/2019-06-30/Recent\_research&action=edit&preload=Wikipedia:Signpost/Tem

### plates/Signpost-article-comments-end/preload)

THESE COMMENTS ARE AUTOMATICALLY <u>TRANSCLUDED</u> FROM THIS ARTICLE'S <u>TALK PAGE</u>. TO FOLLOW COMMENTS, <u>ADD THE PAGE TO YOUR WATCHLIST (HTTPS://EN.WIKIPEDIA.ORG/W/INDEX.PHP?TITLE=WIKIPEDIA</u> <u>TALK:WIKIPEDIA\_SIGNPOST/2019-06-30/RECENT\_RESEARCH&ACTION=WATCH)</u>. IF YOUR COMMENT HAS NOT APPEARED HERE, YOU CAN TRY.

That last one about the concept of reciprosity is an interesting take. I had some ideas of my own on the concept (armchair theorist), but having them factually confirmed, denied, or fed more information is nice. Elfabet (talk) 19:05, 1 July 2019 (UTC)

#### IN THIS ISSUE



#### 30 JUNE 2019 (ALL COMMENTS)

- Discussion report
- News and notes
- In the media
- On the bright side
- Traffic report
- Technology report
- Gallery
- Special report
- Recent research
- From the archives
- Opinion
- In focus
- Community view
- Op-Ed

#### IT'S YOUR SIGNPOST. YOU CAN HELP US.

<u>Home</u> <u>About</u> <u>Archives</u> <u>Newsroom</u> <u>Subscribe</u> Suggestions (https://en.wikipedia.org/wiki/Wikipedia:Wikipedia\_Signpost/Newsroom/Suggestions)</u>

Retrieved from "https://en.wikipedia.org/w/index.php?title=Wikipedia:Wikipedia\_Signpost/2019-06-30/Recent\_research&oldid=1193878633"