Comment on
The Anatomy of Censorship and Propaganda : Evidence from Russian Wikipedia

by V. Avetian, R. Durante, U. Matter and E. Zhuravskaya

Jérôme Hergueux

French National Center for Scientific Research (CNRS)
Center for Law and Economics at ETH Zurich
Berkman Klein Center for Internet & Society at Harvard University

Digital Economics conference, Toulouse School of Economics (Jan. 10, 2025)

What the paper does

- A carefully executed quantitative study of the **Russian governmnent**'s attempts to **censor or influence** its **domestic online information ecosystem**

- Seeks to identify **"what types of information dictators prioritize when it comes to censorship and propaganda"**

  - Answer = culture, history and politics

- Target = **Wikipedia** : **strategic informational resource** (as is + other uses) ; **resilient to censorship or manipulation** (duplicated as Ruwiki, then modified)

- **Great descriptive paper**

  + Reads like a CS paper (descriptive studies are underrated in economics)

  - Relatively ad-hoc (what theoretical implications, or generalizability ?)

Few tweaks

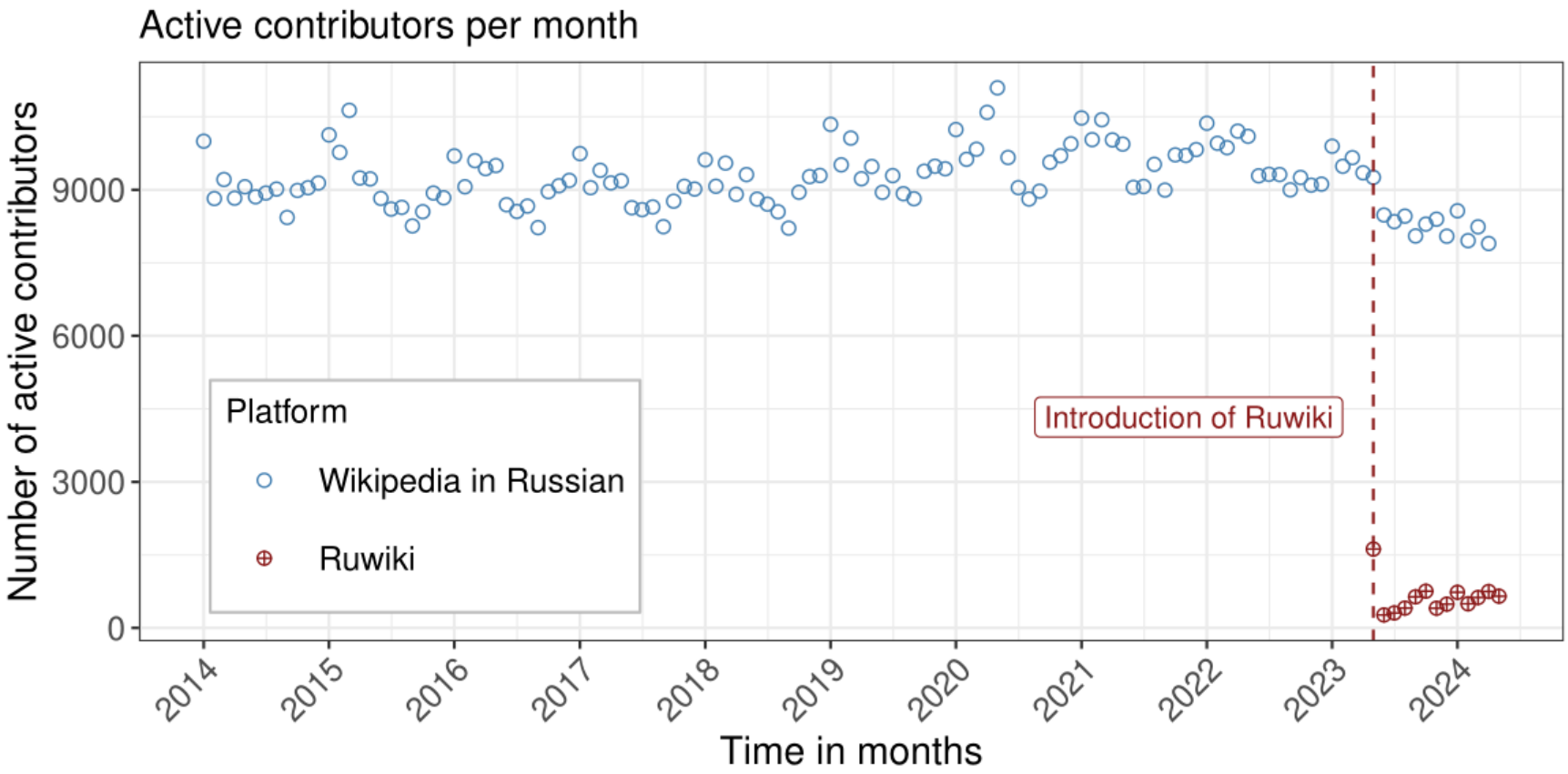
- **Well** (but fast) **written** : no hide and seek (and no literature)
- A few **typos** (Wikipedia is not best described as a crowd-sourcing platform, MediaWiki does not host Wikipedia…)
- Analysis of **deleted articles** (thematic censorship) :

  - Details on **human coding procedures** ?

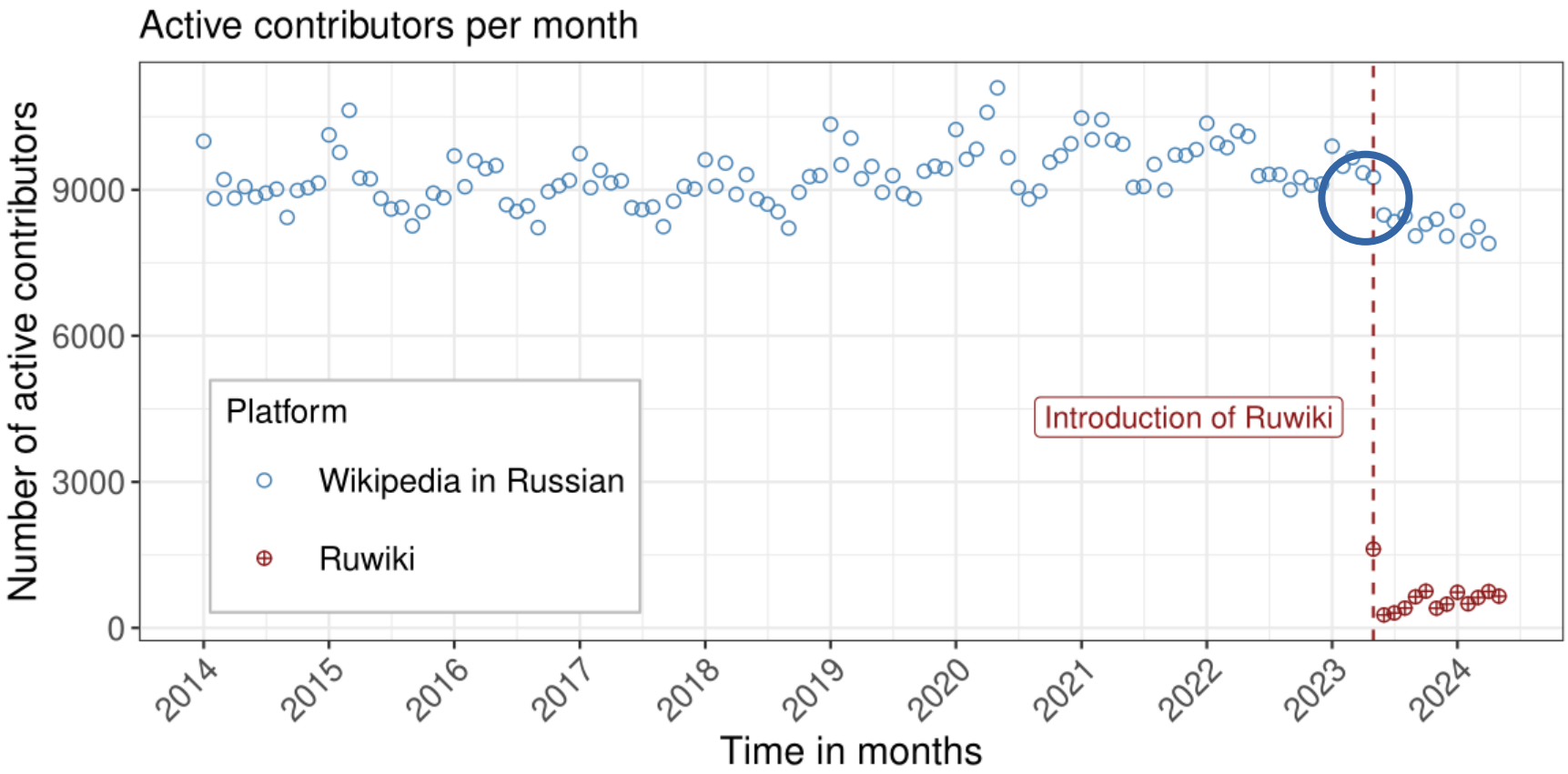
  - Why not rely on, e.g., **Wikipedia category tags** to extract the most common themes across deleted articles ?

- Section 1 (documents editing activity across Wikipedia and Ruwiki) feels like a validity check… or the **beginning of another paper** !

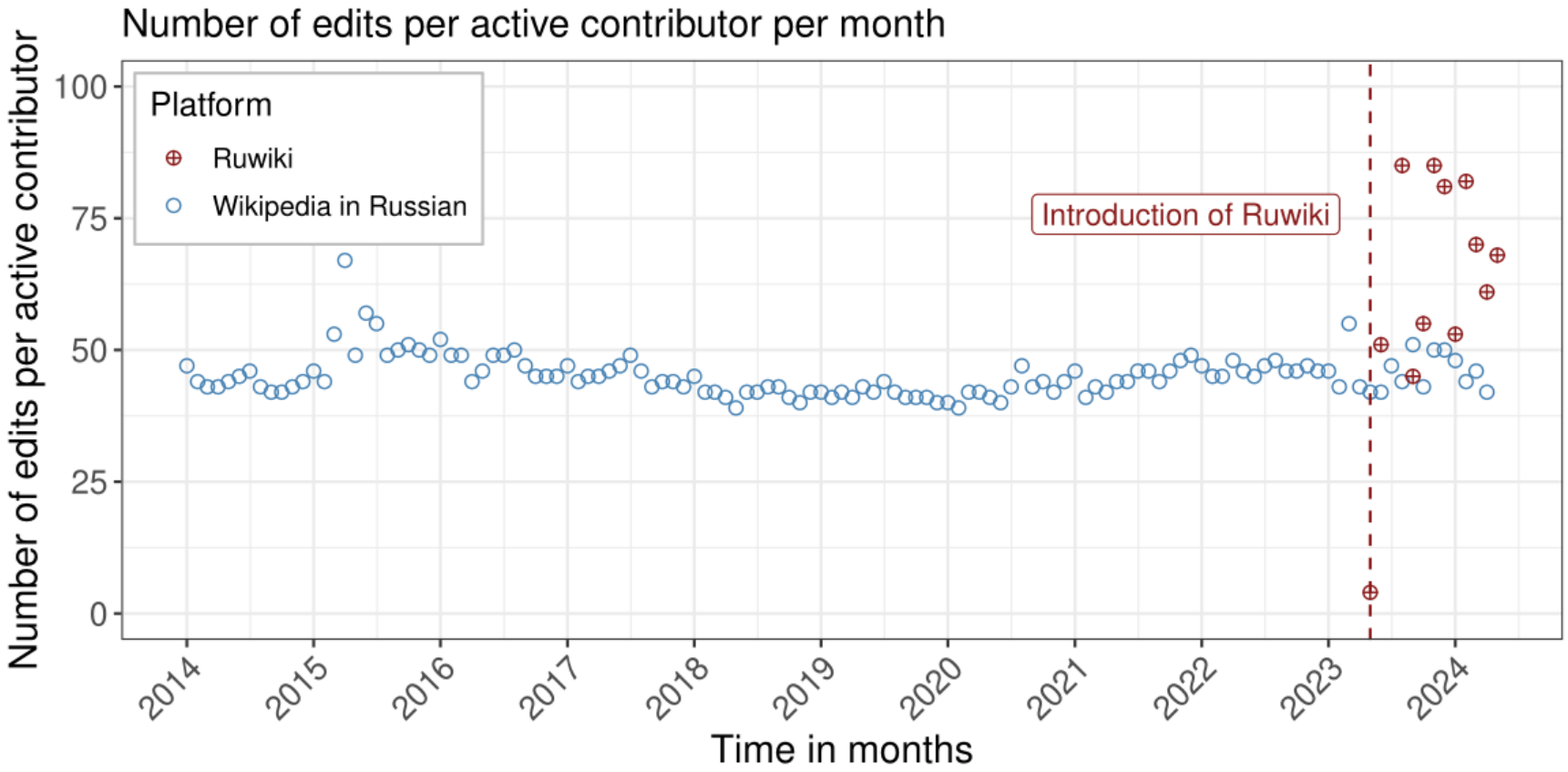
Moving forward?



Moving forward?



Moving forward?



No (de)motivating effect of Ruwiki's introduction on existing Wp contributors ?
(Would be interesting to zoom in…)

Moving forward?

- Paper identifies **2,385 Wp articles deleted** from and **134,060 Wp articles modified** in Ruwiki

- From sample of modified articles : **28,118 Wp articles were 'heavily edited'**

- Leverage this set of target Wp articles : shift focus from description of Ruwiki deletions /edits to **impact of Ruwiki 'treatment' on Wp as a distributed information production ecosystem**

- Two potential treatment times ?

  - May 2023 : introduction of Ruwiki (anticipated?)

  - August 2023 : Ruwiki opens up to public contributions

Moving forward?

Cooperation and Quality in Wikipedia

Dennis Wilkinson Bernardo Huberman
Information Dynamics Laboratory, Hewlett-Packard Labs
{dennis.wilkinson,bernardo.huberman}@hp.com

Abstract

The rise of the Internet has enabled collaboration and cooperation on an unprecedentedly large scale. The online encyclopedia Wikipedia, which presently comprises 7.2 million articles created by 7.04 million distinct editors, provides a consummate example. We examined all 50 million edits made to the 1.5 million English-language Wikipedia articles and found that the high-quality articles are distinguished by a marked increase in number of edits, number of editors, and intensity of cooperative behavior, as compared to other articles of similar visibility and age. This is significant because in other domains, fruitful cooperation has proven to be difficult to sustain as the size of the collaboration increases. Furthermore, in spite of the vagaries of human behavior, we show that Wikipedia articles accrete edits according to a simple stochastic mechanism in which edits beget edits. Topics of high interest or relevance are thus naturally brought to the forefront of quality.

Introduction

The online encyclopedia Wikipedia[1] provides an unprecedented example of large-scale, worldwide collaboration. Its 7.2 million articles have been generated from 282 million edits by 7.04 million distinct contributors[2], as of this writing. Wikipedia's exponential [33] growth since its inception in January 2001 has been enabled by the wiki interface [18], which allows any user to easily modify any article or to create new articles. This arrangement virtually eliminates the barrier to contribution, paving the way for intense activity at uncertain cost to article quality and value.

While Wikipedia's overall quality is difficult to measure in comprehensive way, its content has unquestionably been deemed useful and relevant by the user community at large. Its website is the 10th most visited on the Internet[3], serving an average of 18925 requests per second[4].

In light of its popular success, the question of which Wikipedia articles are high-quality, and how these articles

MIS Quarterly

RESEARCH NOTE

DO EXPERTS OR CROWD-BASED MODELS PRODUCE MORE BIAS? EVIDENCE FROM ENCYCLOPEDIA BRITANNICA AND WIKIPEDIA[1]

Shane Greenstein and Feng Zhu
Harvard Business School, Morgan Hall 439/431
Boston, MA 02163 U.S.A. {sgreenstein@hbs.edu} {fzhu@hbs.edu}

Organizations today can use both crowds and experts to produce knowledge. While prior work compares the accuracy of crowd-produced and expert-produced knowledge, we compare bias in these two models in the context of contested knowledge, which involves subjective, unverifiable, or controversial information. Using data from Encyclopedia Britannica, authored by experts, and Wikipedia, an encyclopedia produced by an online community, we compare the slant and bias of pairs of articles on identical topics of U.S. politics. Our slant measure is less (more) than zero when an article leans toward Democratic (Republican) viewpoints, while bias is the absolute value of the slant. We find that Wikipedia articles are more slanted toward Democratic views than are Britannica articles, as well as more biased. The difference in bias between a pair of articles decreases with more revisions. The bias on a per word basis hardly differs between the sources because Wikipedia articles tend to be longer than Britannica articles. These results highlight the pros and cons of each knowledge production model, help identify the scope of the empirical generalization of prior studies comparing the information quality of the two production models, and offer implications for organizations managing crowd-based knowledge production.

- Wikipedia achieves NPOV by **aggregating / synthetizing many different viewpoints**
- Works best for **highly controversial AND popular topics**

Moving forward?

Home > Management Science > Vol. 67, No. 5 >

Ideology and Composition Among an Online Crowd: Evidence from Wikipedians

Shane Greenstein, Grace Gu, Feng Zhu

Published Online: 24 Sep 2020 | https://doi.org/10.1287/mnsc.2020.3661

Abstract

Online communities bring together participants from diverse backgrounds and often face challenges in aggregating their opinions. We infer lessons from the experience of individual contributors to Wikipedia articles about U.S. politics. We identify two factors that cause a tendency toward moderation in collective opinion: Either biased contributors contribute less, which shifts the composition of participants, or biased contributors moderate their own views. Our findings show that shifts in the composition of participants account for 80%–90% of the moderation in content. Contributors tend to contribute to articles with slants that are opposite their own views. Evidence suggests that encountering extreme contributors with an opposite slant plays an important role in triggering the composition shift and changing views. These findings suggest that collective intelligence becomes more trustworthy when mechanisms encourage confrontation between distinct viewpoints. They also suggest, cautiously, that managers who aspire to produce content "from all sides" should let the most biased contributors leave the collective conversation if they can be replaced with more moderate voices.

- In this convergence process :
  - (i) many radical editors **leave the conversation**
  - (ii) remaining editors **see their viewpoints converge**

Moving forward?

- Wikipedia is difficult/costly to manipulate directly :

  Can governments rely on indirect attack strategies to **affect distributed / participatory information systems**' ability to **attract contributors**, develop **high quality content**, and **retain users** ? (OpenStreetMap *vs.* Google story)

- Here : how did the **introduction of Ruwiki** impact Wikipedia as an **information ecosystem ?**

  1. Supply side : did Ruwiki impact **Wp's ability to reach NPOV ?**

  (i) Input : total **number of edits to target Wp articles ; number + ideological composition of editors** contributing to these target articles (deleted or modified on Ruwiki)

  (ii) Output : evolution of political / ideological **slant of target Wp articles** (deleted or modified on Ruwiki)

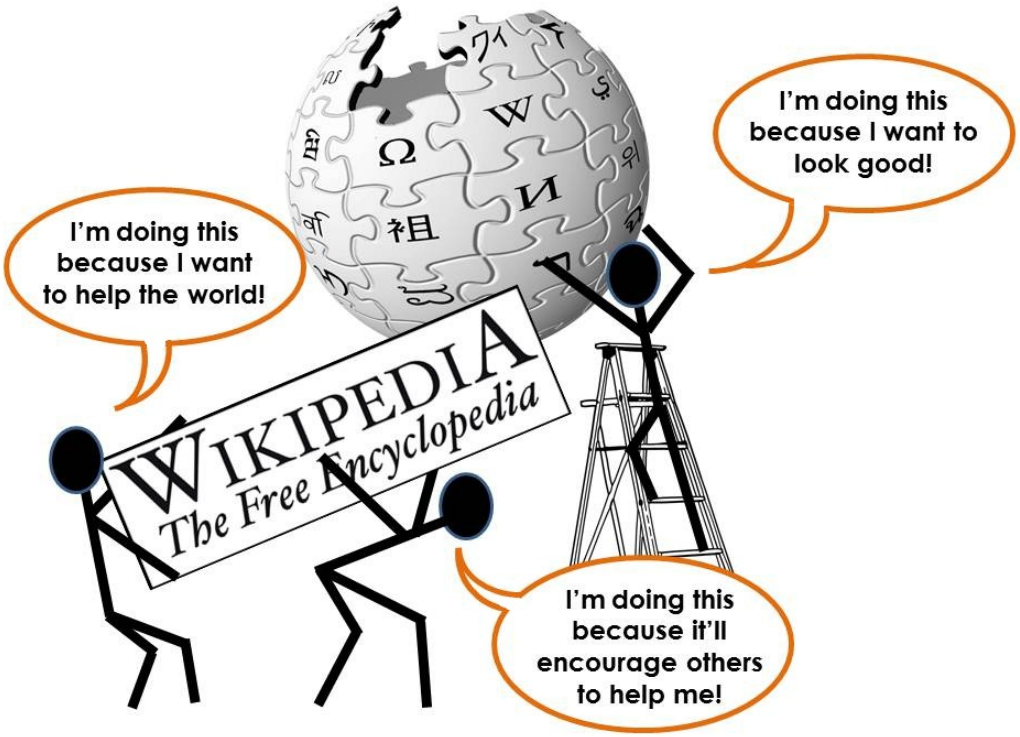
  (iii) Production cost : frequence of **'edit wars' between editors** ?

  2. Demand side : did Ruwiki affect the **demand for information** ?

  (i) Wp page views

Thank you :)

jerome.hergueux@gess.ethz.ch



« Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That's what we're doing. »