The State of Wikimedia Research: 2013-2014

Benjamin Mako Hill
Tilman Bayer
Aaron Shaw
Wikimania 2014, London
August 8, 2014

revision: 72fa64c (2014/08/08)

I've been doing this for many years. I started in 2008 and have done this almost every single year since.

This began as an excuse for me to make sure I was up to date on Wikimedia Research.

"This talk will try to [provide] a quick tour – a literature review in the scholarly parlance – of the last year's academic landscape around Wikimedia and its projects geared at non-academic editors and readers. It will try to categorize, distill, and describe, from a birds eye view, the academic landscape as it is shaping up around our project."

– **From my Wikimania 2008 Submission**

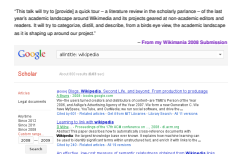
"This talk will try to [provide] a quick tour – a literature review in the scholarly parlance – of the last year's academic landscape around Wikimedia and its projects geared at non-academic editors and readers. It will try to categorize, distill, and describe, from a birds eye view, the academic landscape as it is shaping up around our project."

– From my Wikimania 2008 Submission

Back in Wikimania 2008, I set out to run a session at Wikimania that would provide a comprehensive literature review of articles in Wikipedia published in the last year.

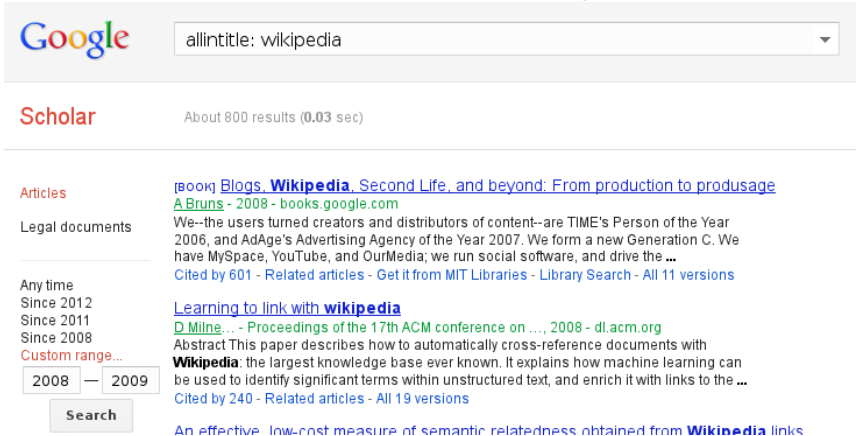
> *"This talk will try to [provide] a quick tour – a literature review in the scholarly parlance – of the last year's academic landscape around Wikimedia and its projects geared at non-academic editors and readers. It will try to categorize, distill, and describe, from a birds eye view, the academic landscape as it is shaping up around our project."*
>
> *– **From my Wikimania 2008 Submission***

Then, about two weeks before Wikimania, I did the scholar search so I could build the literature.



"This talk will try to [provide] a quick tour – a literature review in the scholarly parlance – of the last year's academic landscape around Wikimedia and its projects geared at non-academic editors and readers. It will try to categorize, distill, and describe, from a birds eye view, the academic landscape as it is shaping up around our project."

– **From my Wikimania 2008 Submission**



I tried to import the whole list into Zotero and managed to get banned for abusing the Google Scholar because they thought that no human being could realistically consume the amount of material published on Wikipedia that year.

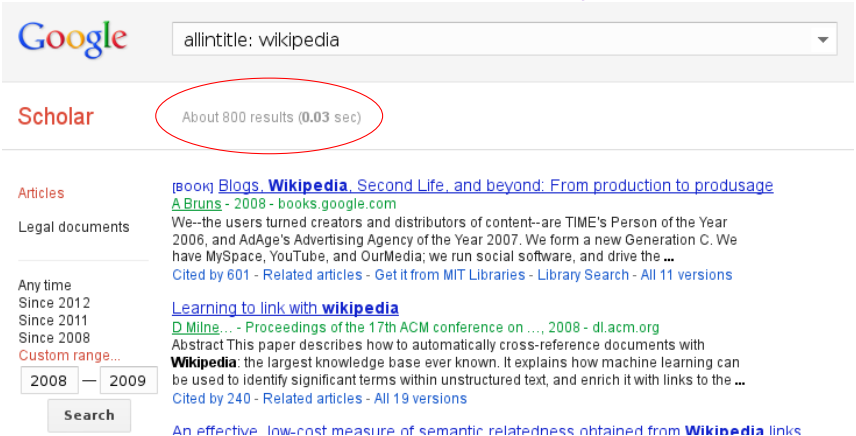
So anyway, I had a 45 minute talk so it worked out to 3.45 seconds to per paper...

And believe it or not, this year is even bigger.

And my talk is even shorter.

"This talk will try to [provide] a quick tour – a literature review in the scholarly parlance – of the last year's academic landscape around Wikimedia and its projects geared at non-academic editors and readers. It will try to categorize, distill, and describe, from a birds eye view, the academic landscape as it is shaping up around our project."

– From my Wikimania 2008 Submission

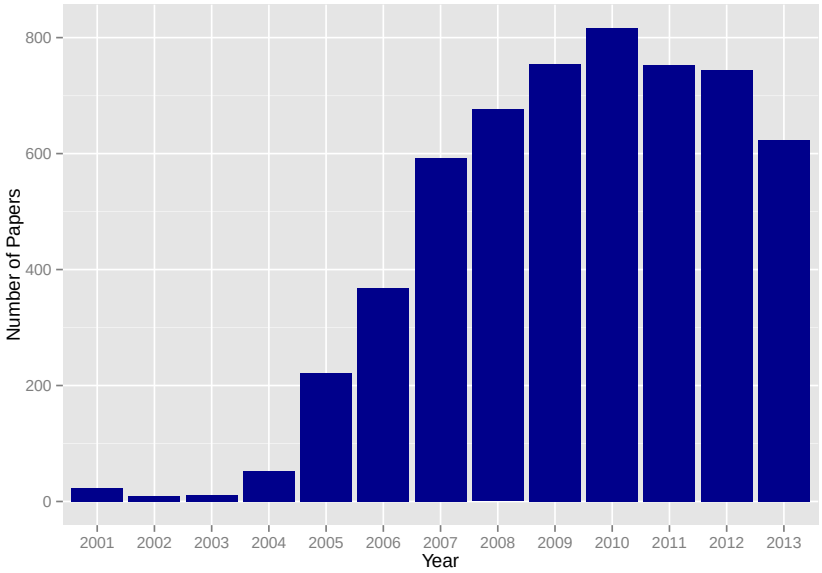


I tried to import the whole list into Zotero and managed to get banned for abusing the Google Scholar because they thought that no human being could realistically consume the amount of material published on Wikipedia that year.

So anyway, I had a 45 minute talk so it worked out to 3.45 seconds to per paper...

And believe it or not, this year is even bigger.

And my talk is even shorter.



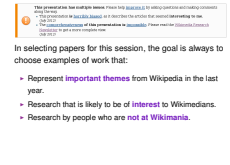
Number of citation, per year, with the term "wikipedia" in the title.

(Source: Google scholar results. Accessed: 2013-08-06)

Academics have written **a lot** of papers about Wikipedia. There are more than 500 papers published about Wikipedia each year and although we've reached and moved past a peak it seems, it's not slowing by much.

- **2968** Wikipedia-related publications in the Scopus database as of November 2013
- **160** recent publications reviewed or mentioned in the 12 issues of the Wikimedia Research Newsletter August 2013-July 2014.

- **2968** Wikipedia-related publications in the Scopus database as of November 2013
- **160** recent publications reviewed or mentioned in the 12 issues of the Wikimedia Research Newsletter August 2013-July 2014.



This presentation has multiple issues. Please help improve it by asking questions and making comments along the way.

- This presentation **is** horribly biased, as it describes the articles that seemed **interesting to me.** *(July 2012)*
- The **comprehensiveness of this presentation is** impossible. Please read the Wikimedia Research Newsletter to get a more complete view. *(July 2012)*

In selecting papers for this session, the goal is always to choose examples of work that:

- Represent **important themes** from Wikipedia in the last year.
- Research that is likely to be of **interest** to Wikimedians.
- Research by people who are **not at Wikimania**.

This is my disclaimer slide...

Within these goals, the selections are **incomplete**, and **wrong**.

Event Prediction

Event Prediction

Mako

This was the year that studies of readership of Wikipedia really blossomed. People figured out how to use the view data. Much of what they used it for was prediction.

Wikipedia Viewership and Flu Prediction

McIver, David J., and John S. Brownstein. “**Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time**.” PLoS Comput Biol 10, no. 4 (April 17, 2014): e1003581. doi:10.1371/journal.pcbi.1003581.

Wikipedia Viewership and Flu Prediction

McIver, David J., and John S. Brownstein. “Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time.” PLoS Comput Biol 10, no. 4 (April 17, 2014): e1003581. doi:10.1371/journal.pcbi.1003581.

Wikipedia Viewership and Flu Prediction: Motivation

- **Google Flu Trends** uses search engine queries to try to predict influenza epidemics more quickly than traditional methods.
- ..but it has been criticized as being biased (e.g., by media coverage).
- WP is freely available and viewership data is free, unlike Google which is proprietary.

Wikipedia Viewership and Flu Prediction: Motivation
- Google Flu Trends uses search engine queries to try to predict influenza epidemics more quickly than traditional methods.
- ..but it has been criticized as being biased (e.g., by media coverage).
- WP is freely available and viewership data is free, unlike Google which is proprietary.

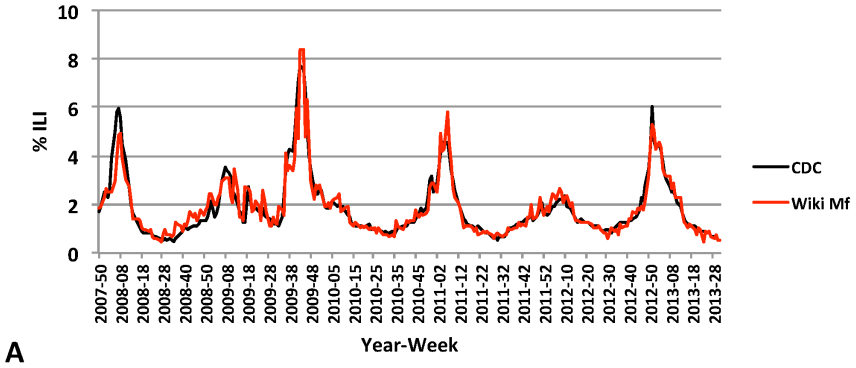
2009 H1N1 Swine Flu broke GFT.

Wikipedia Viewership and Flu Prediction: Methods

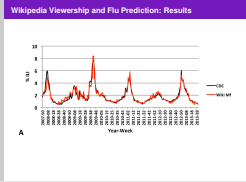
- Measure traffic to flu related articles on Wikipedia
- Compare to the “gold standard” data from the Center for Disease Control (CDC)

Wikipedia Viewership and Flu Prediction: Methods
- Measure traffic to flu related articles on Wikipedia
- Compare to the “gold standard” data from the Center for Disease Control (CDC)

Wikipedia Viewership and Flu Prediction: Results



A



- Wikipedia better than Google at predicting peak flu weeks.
- Wikipedia better at predicting relative influenza rates.

Other things people have tried to predict include...

- Global disease forecasting
- Box office revenue based on films
- Election results in Iran, Germany and the UK
- Breaking news stories
- Trending topics, general zeitgeist. [1] [2] [3] [4]

Other things people have tried to predict include...
- Global disease forecasting
- Box office revenue based on films
- Election results in Iran, Germany and the UK
- Breaking news stories
- Trending topics, general zeitgeist. [1] [2] [3] [4]

Wikimedia as a Corpus

Aaron

From the Bar-Ilan lit review: 48% of them about Wikipedia per se, 52% are just using Wikipedia (e.g. as a text corpus)
Wiktionary, non Wikipedia projects.

Wiktionary as a source of data, not the substance/object of analysis. Projects as amazing multilingual corpuses of natural language.

Multi-Lingual Dictionary from Wiktionary: Methods

Ács, Judit. “**Pivot-Based Multilingual Dictionary Building Using Wiktionary**.” In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14). Rekyjavik, Iceland, 2014. http://www.lrec-conf.org/proceedings/lrec2014/pdf/864_Paper.pdf.

Multi-Lingual Dictionary from Wiktionary: Methods

Ács, Judit. “**Pivot-Based Multilingual Dictionary Building Using Wiktionary**.” In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14). Rekyjavik, Iceland, 2014. http://www.lrec-conf.org/proceedings/lrec2014/pdf/864_Paper.pdf.

Methods: This paper uses a machine classifier to take advantage of “pivots” — words with common translations across more than one language — to automatically construct a multilingual dictionary via triangulation! They then manually evaluate the precision of this multilingual dictionary on a small subset of words.

Multi-Lingual Dictionary from Wiktionary: Results

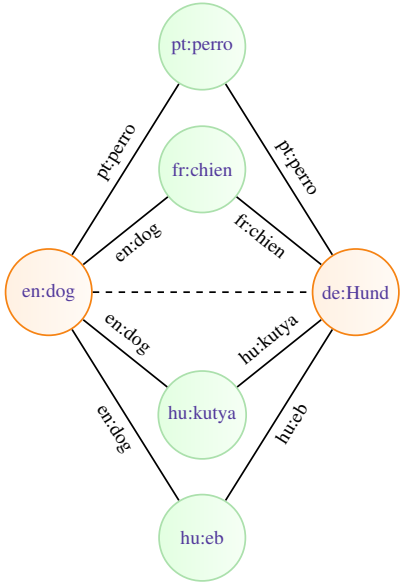
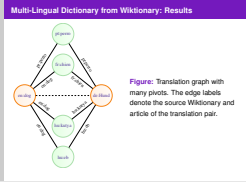


Figure: Translation graph with many pivots. The edge labels denote the source Wiktionary and article of the translation pair.



Finding: So, kind of incredibly, this sort of works. The author succeeds in constructing the multilingual dictionary, but finds that problems like polysemy (one word meaning multiple things) limit the precision of the resulting output. Using the multilingual pivots offered by Wiktionary, however, performs much better using “triangles” between three languages alone.

Content Quality

Content Quality

Mako

Mostly focused on sub-areas. There was one this year we considered discussion on hematology. Or information on infectious diseases.

Comparing Wikipedia Quality to Britannica

Nifrário Rodrigues, Fernando Silvério. “**Colaboração Em Massa Ou Amadorismo Em Massa? Um Estudo Comparativo Da Qualidade Da Informação Científica Produzida Utilizando Os Conceitos E Ferramentas Wiki**.” Universidade de Évora, 2012. http://massamateurism.blogspot.co.uk/p/synopsis.html.

Comparing Wikipedia Quality to Britannica

Nifrário Rodrigues, Fernando Silvério. “**Colaboração Em Massa Ou Amadorismo Em Massa? Um Estudo Comparativo Da Qualidade Da Informação Científica Produzida Utilizando Os Conceitos E Ferramentas Wiki**.” Universidade de Évora, 2012. http://massamateurism.blogspot.co.uk/p/synopsis.html.

A Portuguese-language dissertation at the Universidade de Évora, titled "Colaboração em Massa ou Amadorismo em Massa?" ("Mass collaboration or mass amateurism?")

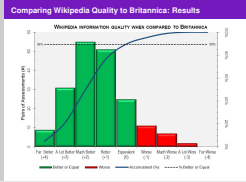
Comparing Wikipedia Quality to Britannica: Methods

- Random sample of 245 article pairs from both encyclopedias.
- Graded by an expert in its subject area using a five-point scale.
- Experts asked, “to concentrate only on some [...] intrinsic aspects of the articles’ quality, namely accuracy and objectivity, and discard the contextual, representational and accessibility aspects.”
- Experts were mostly university teachers.

Comparing Wikipedia Quality to Britannica: Methods
- Random sample of 245 article pairs from both encyclopedias.
- Graded by an expert in its subject area using a five-point scale.
- Experts asked, “to concentrate only on some [...] intrinsic aspects of the articles’ quality, namely accuracy and objectivity, and discard the contextual, representational and accessibility aspects.”
- Experts were mostly university teachers.

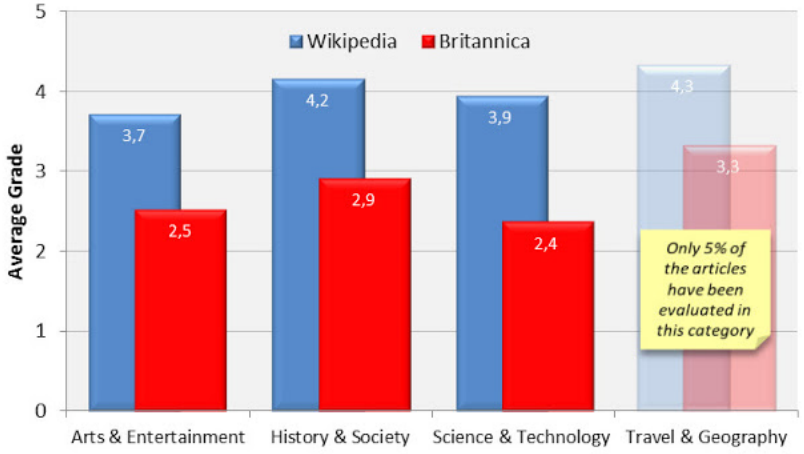
Compared the quality of English Wikipedia with that of Encyclopedia Britannica

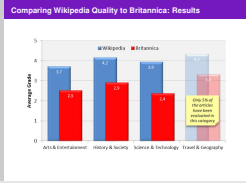
Comparing Wikipedia Quality to Britannica: Results



They rated "**90%** of the Wikipedia articles ... as having **equivalent or better quality** than their Britannica counterparts".

Comparing Wikipedia Quality to Britannica: Results





Controversy and Conflict

Controversy and Conflict

Tilman

Osman, Kim. “**The Role of Conflict in Determining Consensus on Quality in Wikipedia Articles**.” In Proceedings of the 9th International Symposium on Open Collaboration, 12:1–12:6. WikiSym ’13. New York, NY, USA: ACM, 2013. doi:10.1145/2491055.2491067.

Conflict, Consensus and Quality in Wikipedia

Osman, Kim. “The Role of Conflict in Determining Consensus on Quality in Wikipedia Articles.” In Proceedings of the 9th International Symposium on Open Collaboration, 12:1–12:6. WikiSym ’13. New York, NY, USA: ACM, 2013. doi:10.1145/2491055.2491067.

Conflict, Consensus and Quality in Wikipedia: Methods

- Literature review of research on online communities
- Discourse analysis of [[Talk:Australia]] (+ 17 archive pages), coding 147 threads (156,112 words) in a grounded theory approach.

Conflict, Consensus and Quality in Wikipedia: Methods

- Literature review of research on online communities
- Discourse analysis of [[Talk:Australia]] (+ 17 archive pages), coding 147 threads (156,112 words) in a grounded theory approach.

Very simplified: Grounded theory is an approach in social sciences where one starts from empirical data first and develops hypotheses by coding...

Conflict, Consensus and Quality in Wikipedia: Results

"**Conflict** was significantly more prevalent .. than **collaboration**."

- However: Personal attacks are rare (as opposed to disagreement).
- "The four main themes that emerged as cause for debate among the editorial community were **sources**, **wording**, **structure** and **content accuracy**".
- In e.g. sourcing debates, "conflict ... had a role in developing a mechanism to ensure the accuracy of information by prompting participants to properly source and reference material."
- Conflict is not always bad. "**Generative friction**" benefits Wikipedia quality.

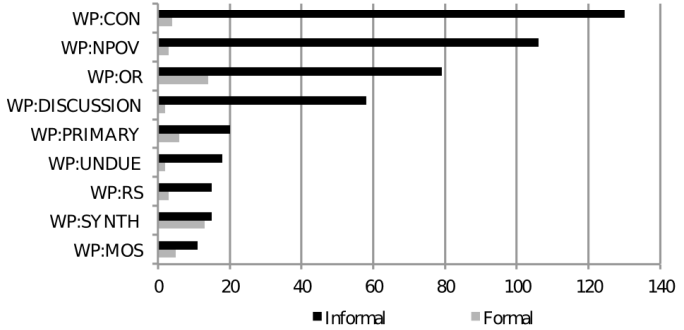
Conflict, Consensus and Quality in Wikipedia: Results

"Conflict was significantly more prevalent .. than collaboration."

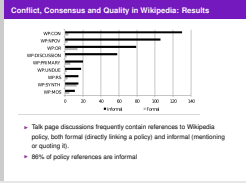
- However: Personal attacks are rare (as opposed to disagreement).
- "The four main themes that emerged as cause for debate among the editorial community were sources, wording, structure and content accuracy".
- In e.g. sourcing debates, "conflict ... had a role in developing a mechanism to ensure the accuracy of information by prompting participants to properly source and reference material."
- Conflict is not always bad. "Generative friction" benefits Wikipedia quality.

i.e. editors generally play the ball, not the man.

Conflict, Consensus and Quality in Wikipedia: Results



- Talk page discussions frequently contain references to Wikipedia policy, both formal (directly linking a policy) and informal (mentioning or quoting it).
- 86% of policy references are informal



Bear in mind that while it was a sizable corpus, it still pertained to only one article on the English Wikipedia.

Osman: "More than being a set of isolated rules for the community, policies are part of the fabric of the culture of the talk pages. They regulate both behavior and the production process and manage conflict so that it remains a generative friction", e.g. by discouraging personal attacks.

WikiProjects

WikiProjects

Another area of research focuses on understanding wikis and other peer production communities as organization. Some of the most interesting work in this area compares many projects in order to better understand the characteristics that might lead them to grow and attract contributors.

Critical Mass in WikiProjects

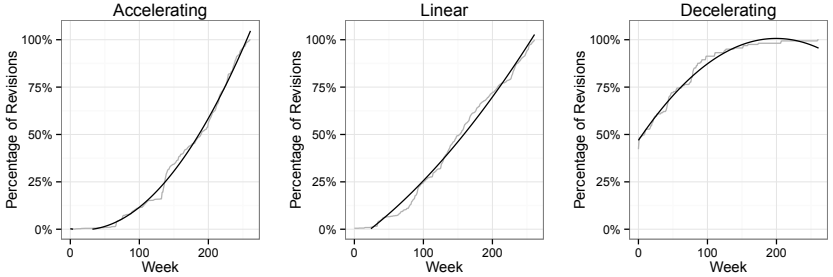
Solomon, Jacob, and Rick Wash. “**Critical Mass of What? Exploring Community Growth in WikiProjects**.” In Eighth International AAAI Conference on Weblogs and Social Media, 2014. [1].

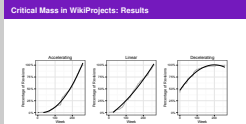
Critical Mass in WikiProjects

Solomon, Jacob, and Rick Wash. “Critical Mass of What? Exploring Community Growth in WikiProjects.” In Eighth International AAAI Conference on Weblogs and Social Media, 2014. [1].

Methods: Analyzed data on 1069 EN:WP WikiProjects. Fit models to describe the projects’ growth curves in terms of editors and project edits. They then use these growth curves to estimate the relationship between adding editors or edits (at an early stage) on subsequent project growth.

Critical Mass in WikiProjects: Results





Results: The authors find that projects with more contributors are more likely to experience subsequent growth in contributions and contributors.They also find that contributions from both "power users" and more casual, one-off contributors predicts subsequent growth. These findings (like others in this area) have important implications for project leaders and designers.

Vandalism

Vandalism

Edit Patterns and Vandalism Detection

Sethi, Deepika. **A Large Scale Study of Edit Patterns in Wikipedia and its Applications to Vandalism Detection**. M. Sc. thesis, University of Georgia. Submitted December 2012

Edit Patterns and Vandalism Detection

Sethi, Deepika. **A Large Scale Study of Edit Patterns in Wikipedia and its Applications to Vandalism Detection**. M. Sc. thesis, University of Georgia. Submitted December 2012

Edit Patterns and Vandalism Detection: Methods

Used the **PAN Wikipedia vandalism corpus 2010** of 32,452 edits, classified as vandalism (2,391 edits) or non-vandalism by Mechanical Turk workers.
Identified vandalism based on:

- Editor country
- Time of day
- **Article content domain** using 12 “classes” from DBpedia (e.g., 1. Person, 2. Work, 3. Sports, 4. Places, 5. Food ...)
- Content of edits

Edit Patterns and Vandalism Detection: Methods

Used the PAN Wikipedia vandalism corpus 2010 of 32,452 edits, classified as vandalism (2,391 edits) or non-vandalism by Mechanical Turk workers.
Identified vandalism based on:

- Editor country
- Time of day
- Article content domain using 12 “classes” from DBpedia (e.g., 1. Person, 2. Work, 3. Sports, 4. Places, 5. Food ...)
- Content of edits

Corpus was created to train vandalism detectors and formed the basis of several competitions.

Edit Patterns and Vandalism Detection: Results

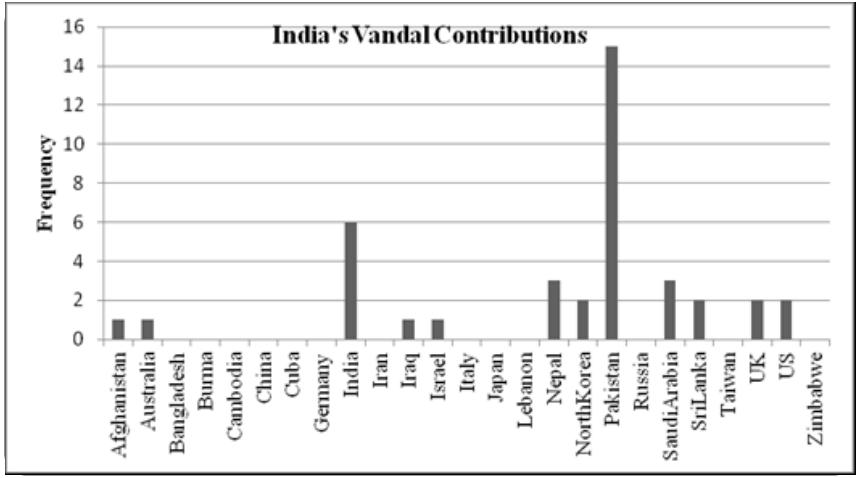
"**Vandalism occurs the most during office hours** while **non-vandalism occurs the most during late evenings**."

Edit Patterns and Vandalism Detection: Results

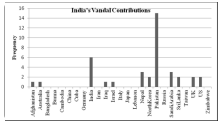
"Vandalism occurs the most during office hours while non-vandalism occurs the most during late evenings."

Edit Patterns and Vandalism Detection: Results

"Hostilities among the countries are one major cause of vandalism."



Edit Patterns and Vandalism Detection: Results
"Hostilities among the countries are one major cause of vandalism."



Country where the vandal is based..

Methodology: Get Geolocation of IP editors to articles about countries Result: image : Figure 20: India's Vandal Contributions (caption: Where Indians vandalize most) [[India]] most frequently vandalized from India, Bahrain, Bangladesh, Pakistan, ... [[Pakistan]] most frequently vandalized from India, Pakistan, Poland, SwedenOther example: [[Taiwan]] from Taiwan, China, Germany ... / [[China]] from Ireland, Germany, Poland, ...

Common Vandalism Words

Ball, chicken, British, woman, hole, handicap, meat, kiss, play, old, love, death, course, kick, American, bomb

Edit Patterns and Vandalism Detection: Results
Common Vandalism Words
Ball, chicken, British, woman, hole, handicap, meat, kiss, play, old, love, death, course, kick, American, bomb

Methodology: Word most frequently occurring in vandalized versions

Result: Ball, chicken, British, woman, hole, handicap, meat, kiss, play, old, love, death, course, kick, American, bomb, ...

Editor Motivation

Editor Motivation

When do barnstars increase edits?

Restivo, Michael, and Arnout van de Rijt. “**No Praise without Effort: Experimental Evidence on How Rewards Affect Wikipedia’s Contributor Community**.” Information, Communication & Society 0, no. 0 (0): 1–12. doi:10.1080/1369118X.2014.888459.

When do barnstars increase edits?

Restivo, Michael, and Arnout van de Rijt. “No Praise without Effort: Experimental Evidence on How Rewards Affect Wikipedia’s Contributor Community.” Information, Communication & Society 0, no. 0 (0): 1–12. doi:10.1080/1369118X.2014.888459.

When do barnstars increase edits? Methods

Prior work by the authors showed that randomly given barnstars to very active editors (top 1% in a month) resulted in more edits and a positive feedback loop in English Wikipedia.

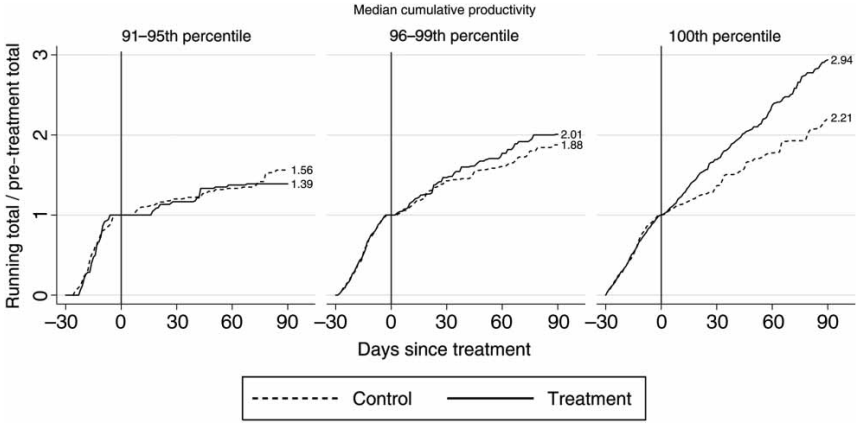
- Like before, gives barnstars to **randomly** selected users – but this time to editors of varying activity level (e.g., 91-95$^{\text{th}}$, 96-99$^{\text{th}}$, and 100th percentile of editing in the month). Plus a “control” group of other editors who do not receive the award.
- Follow post-award activity on Wikipedia.

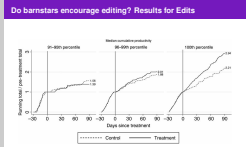
When do barnstars increase edits? Methods

Prior work by the authors showed that randomly given barnstars to very active editors (top 1% in a month) resulted in more edits and a positive feedback loop in English Wikipedia.

- Like before, gives barnstars to randomly selected users – but this time to editors of varying activity level (e.g., 91-95th, 96-99th, and 100th percentile of editing in the month). Plus a “control” group of other editors who do not receive the award.
- Follow post-award activity on Wikipedia.

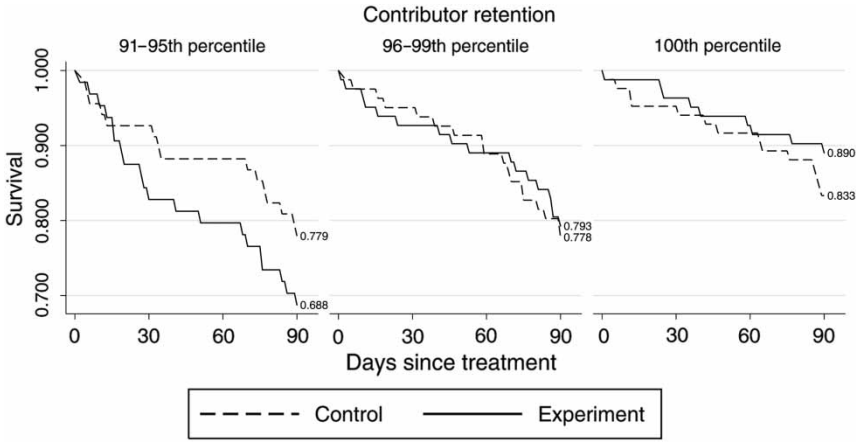
Do barnstars encourage editing? Results for Edits

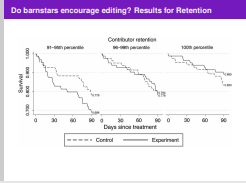




Positive effect only in the case of the very active – top 1% of editors. No significant difference in the other two groups.

Do barnstars encourage editing? Results for Retention

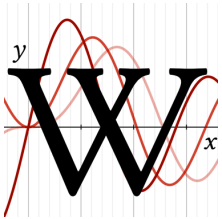




Lower retention among award recipients in the less active group! No significant difference in the other two groups.

More Resources

- **Wikimedia Research Newsletter**
  [[:meta:Research:Newsletter]]
- **WikiSym** (Later this month in Berlin!)
- **WikiPapers Repository** [http://wikipapers.referata.com]
- **Much More**



More Resources
- Wikimedia Research Newsletter
  [[:meta:Research:Newsletter]]
- WikiSym (Later this month in Berlin!)
- WikiPapers Repository [http://wikipapers.referata.com]
- Much More

Those are my six exemplary studies from the past year.
There has been just tons and tons of work in this area. Trying to talk about this in 20 minutes strikes me as increasingly crazy every year I try to do it.
The most important source, now going for a couple years, is the Wikimedia Research Newsletter which is published monthly in the (English) Signpost and syndicated on the Wikimedia Research.
But there are other resources as well. And I encourage you to get involved.

Meta-Analyses

- Okoli et al., “The sum of all human knowledge”: a systematic review of scholarly research on the content of Wikipedia.
- Bar-Ilan and Aharony, Twelve years of Wikipedia research.
- Taraborelli. Keynote on Wikipedia Research. OpenSym 2013. Hong Kong.
- Benkler, Shaw, and Hill, Peer Production: A Modality of Collective Intelligence.

Meta-Analyses

- Okoli et al., “The sum of all human knowledge”: a systematic review of scholarly research on the content of Wikipedia.
- Bar-Ilan and Aharony, Twelve years of Wikipedia research.
- Taraborelli. Keynote on Wikipedia Research. OpenSym 2013. Hong Kong.
- Benkler, Shaw, and Hill, Peer Production: A Modality of Collective Intelligence.