

Wikipedia, Elder or Teen?

A Look at Growth, Stagnation and Decline Patterns Across 50 Language Communities

Marc Miquel-Ribé
Eurecat - Centre Tecnològic de
Catalunya
Barcelona, Catalunya
mmiquelr@tecnocampus.cat

Cristian Consonni
Eurecat - Centre Tecnològic de
Catalunya
Barcelona, Catalunya
cristian.consonni@acm.org

David Laniado
Eurecat - Centre Tecnològic de
Catalunya
Barcelona, Catalunya
david.laniado@eurecat.org

ABSTRACT

Wikipedia is an undeniably successful project, with unprecedented numbers of online volunteer contributors. However, researchers observed that the number of active editors for the largest Wikipedias started to decline in 2007, after rapid initial growth. Years after those announcements, researchers and community activists still need to understand community growth. We studied the growth, decline, and stagnation patterns of 50 Wikipedia language editions, and we found that half of them are still growing.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Collaborative content creation**.

KEYWORDS

Wikipedia, online communities, editor engagement, retiring, growth, decline, online collaboration

ACM Reference Format:

Marc Miquel-Ribé, Cristian Consonni, and David Laniado. 2021. Wikipedia, Elder or Teen?: A Look at Growth, Stagnation and Decline Patterns Across 50 Language Communities. In *Proceedings of Wiki Workshop 2022 (Wiki Workshop '22)*. ACM, New York, NY, USA, 2 pages. <https://doi.org/XXXXXXX.XXXXXX>

1 INTRODUCTION

In 2005, along with growing popularity on the Internet, the number of registrations on Wikipedia accelerated until obtaining a critical mass of participants in several languages. Nonetheless, in 2007, English Wikipedia peaked in the number of active editors - reaching over 60,000 - and started declining in the following year (e.g., 43,000 in 2010, 37,000 in 2013, 36,000 in 2016).¹ Academic studies presented the overall decline in active editors on the English Wikipedia as a consequence of the trade-off between massive participation and the need to manage content quality, that led to a more closed system calcified against changes - especially those proposed

¹ *Wikimedia Statistics*, all time number of active editors, <https://w.wiki/4wEb> [accessed 2022-03-10]

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

Wiki Workshop '22, April 25, 2022, London, UK (virtual event)

© 2021 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM... \$15.00
<https://doi.org/XXXXXXX.XXXXXX>

by newcomers - in the form of policies, among other aspects [1, 3, 7]. As early as 2009, the Wikimedia Movement started a strategy process, which included a task force dedicated to debate about the state of what they called Community Health.² This new term was coined to discuss aspects related to burn-out and editing fatigue, as well as the potential impact of reverts and community norms on the decline in the number of contributors. Ever since then, the Wikimedia Foundation has developed numerous initiatives to improve the user experience, dedicating special efforts to newcomers over the past years (2014 Growth,³ 2017-2018 New Editor Experiences,⁴ 2018-2022 Growth Team⁵) to understand the first days of an editor and personalize the newcomer experience, and to provide for safety to all editors.⁶ However, for large languages like English or German Wikipedia, the number of active editors remains stagnant. For this reason, while the discourse of stagnant and declining Wikipedia language communities is commonly generalized to all Wikipedia language editions, we wonder if this is really the case. We propose the objective of assessing the growth, stagnation, decline patterns in the history of Wikipedia language communities.

2 APPROACH

To pursue the objective of assessing growth, stagnation, or decline patterns in Wikipedia communities, we inspected the temporal evolution of the number of active editors over time, comparing the trends obtained for different language editions and performing clustering to identify general patterns. We computed the monthly number of active editors for each of the 308 Wikipedia language editions, and we focused on communities with a minimum of 100 active editors in August 2021. There are 50 such communities. An active editor is any editor who makes at least 5 edits in one month. We have used the threshold of 100 active editors, considering that communities with a lower number may not be consolidated.

We created a timeline for each language edition representing its monthly number of active editors. To be able to group communities exhibiting similar temporal patterns, we applied k-means clustering to the time series, and we used dynamic time warping to measure similarity between the temporal sequences focusing on

² Meta contributors, 'Community health', *Meta, discussion about Wikimedia projects*, https://meta.wikimedia.org/w/index.php?title=Community_health&oldid=17219685

³ MediaWiki contributors, 'Growth/Growth 2014', *MediaWiki*, https://www.mediawiki.org/w/index.php?title=Growth/Growth_2014&oldid=4775153

⁴ MediaWiki contributors, 'New Editor Experiences', *MediaWiki*, https://www.mediawiki.org/w/index.php?title=New_Editor_Experiences&oldid=5094738

⁵ MediaWiki contributors, 'Growth', *MediaWiki*, <https://www.mediawiki.org/w/index.php?title=Growth&oldid=5050955>

⁶ Meta contributors, 'Universal Code of Conduct', *Meta, discussion about Wikimedia projects*, https://meta.wikimedia.org/w/index.php?title=Universal_Code_of_Conduct&oldid=22953587

the general pattern [5]. In this way, we obtain a general idea of the most common patterns that exist among the selected communities.

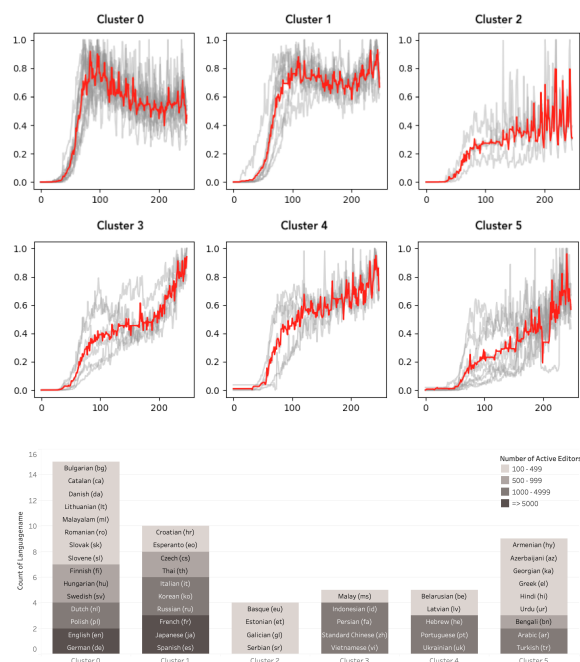


Figure 1: Each cluster represents a group of language editions exhibiting a similar temporal trend, according to an execution of the k-means clustering algorithm. Time is expressed on the x-axis in the number of months since the creation of a language edition. Gray lines represent the time series of the individual language editions, and red lines represent the average over each cluster. The language editions belonging to each cluster are reported below, with background color indicating the size range in the number of active editors.

2.1 Results

Dataset and Code. The dataset and code used in this paper are available at: <https://zenodo.org/record/6344487>. The latest version is available on GitHub at: <https://github.com/WikiCommunityHealth/time-clustering>.

We ran the k-means algorithm with parameters sigma $\sigma = 6$, and learning rate $l_r = 0.1$; we obtained 6 clusters, shown in Figure 1.

We observe that many language editions of different size belong to the first cluster (cluster 0), characterized by a first phase of growth of about 7 years (which correspond to 84 months) and then a decline period until stabilization with more or less accentuated oscillations around a lower number of editors. This roughly corresponds to the decline observed for English (included in cluster 0, together with German, Dutch and Polish among other European languages) and other major language editions since 2007 [4, 6].

The second cluster (cluster 1) also comprises some of the largest language editions, including French, Japanese, and Spanish. The

trend is similar to that of the previous cluster, with the difference that after the rise and peak, instead of decline, we see in general a more stable stagnation pattern. We may also observe a tendency to decline in the first years of stagnation, and a smooth rise again in the last years.

The remaining clusters exhibit a different pattern, with a common tendency to keep growing, although at different rates after the initial rise. Cluster 2 includes smaller European language communities, characterized in the second phase by stronger oscillations around a smooth growing trend. Cluster 3 represents some Asian language communities that interestingly exhibit a decline/stagnation period followed by a strongly growing pattern. Communities in Cluster 4, that includes Portuguese, Hebrew, and languages from post-Soviet countries, see a first rapid growth period, with a different duration for different communities, followed by a less skewed but still growing trend. Finally, Cluster 5 groups together communities of different sizes, characterized by a more or less stable growth trend.

3 CONCLUSIONS

Our results suggest that, among the largest 50 language editions in the number of monthly active editors, only half of them exhibit a pattern of decline or stagnation, while the others are still growing in the size of their editor community. This represents a significant breakthrough, given that it was widely assumed that communities were in decline for not being able to maintain their number of active editors [2] - possibly because of a focus on the English Wikipedia and other major language editions and a lack of analysis of smaller or younger communities.

Further research could shed light on the factors associated with the different patterns observed, including external factors, e.g., Internet access, geopolitical context, size and demographic composition of a language’s speaker community, language status (official language or not), and internal factors, i.e., community dynamics, calcification of policies, social interactions, technical complexity, platform usability, among others.

ACKNOWLEDGMENTS

This research was funded by the Wikimedia Foundation grant “Community Health Metrics: Understanding Editor Drop-off.”⁷

REFERENCES

- [1] Aaron Halfaker, R Stuart Geiger, Jonathan T Morgan, and John Riedl. 2013. The rise and decline of an open collaboration system: How Wikipedia’s reaction to popularity is causing its decline. *American Behavioral Scientist* 57, 5 (2013), 664–688.
- [2] Benjamin Mako Hill and Aaron Shaw. 2020. Wikipedia and the End of Open Collaboration. *Wikipedia* 20 (2020).
- [3] Dariusz Jemielniak. 2014. *Common Knowledge?: An Ethnography of Wikipedia*. Stanford University Press. <https://doi.org/doi:10.1515/9780804791205>
- [4] José Felipe Ortega Soto et al. 2009. Wikipedia: A quantitative analysis. (2009).
- [5] François Petitjean, Alain Ketterlin, and Pierre Gançarski. 2011. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition* 44, 3 (2011), 678–693. <https://doi.org/10.1016/j.patcog.2010.09.013>
- [6] Bongwon Suh, Gregorio Convertino, Ed H Chi, and Peter Pirolli. 2009. The singularity is not near: slowing growth of Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*. 1–10.
- [7] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. 2018. Revisiting “The rise and decline” in a population of peer production projects. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–7.

⁷Grant page: <https://w.wiki/4wEm>