Digital divides in the social construction of history: Editor representation in Wikipedia articles on African independence processes

Schlögl, Stephan; Bürger, Moritz; Schmid-Petri, Hannah

Erstveröffentlichung / Primary Publication
Konferenzbeitrag / conference paper

Digital divides in the social construction of history: Editor representation in Wikipedia articles on African independence processes

Stephan Schlögl, Moritz Bürger, Hannah Schmid-Petri

University of Passau

Zusammenfassung

*Ziel des vorliegenden Beitrages ist es, verschiedene Formen digitaler Ungleichheit im Rahmen von Kollaborationsprozessen online zu untersuchen. Konkret analysieren wir digitale Ungleichheiten hinsichtlich der geografischen Herkunft von Editor*innen der Wikipedia. Der aktuelle Forschungsstand zeigt, dass Teilnehmer*innen aus Industrieländern hier besonders stark vertreten sind, während in anderen Regionen (insbesondere in sogenannten Entwicklungsländern) kaum Menschen teilnehmen. Die vorliegende Studie untersucht, wie sich die Autor*innengeographie in der Bearbeitung (Partizipation, Einfluss und Erfolg) von Artikeln zu den Unabhängigkeitsprozessen ehemaliger Kolonien in Afrika widerspiegelt. Die Analyse basiert auf insgesamt 354 Wikipedia-Artikeln. Anhand der Geolokalisierung von 75% der beteiligten Editor*innen (N = 23.408) zeigen wir, dass die meisten Bearbeitungen von Editor*innen aus Frankreich vorgenommen werden. Dieses Ungleichgewicht zeigt sich auch gemessen am gesamten Textanteil im Laufe der Zeit. Auf der Ebene der einzelnen Nutzer*innen lässt sich jedoch feststellen, dass Editor*innen aus Frankreich nur geringfügig erfolgreicher sind als jene aus den afrikanischen Nachfolgestaaten, wenn es um die längerfristige Sichtbarkeit ihrer Beiträge geht.*

Keywords: Online Partizipation, Digitale Ungleichheit, Wikipedia, Geographie des Wissens, Kolonialgeschichte

Summary

The aim of the present paper is to examine different forms of digital divides that may occur during online collaboration. Specifically, we analyse digital inequalities regarding the geographical origin of editors active on Wikipedia. Existing evidence suggests a strong geographic bias among Wikipedia editors, as industrialized countries are strongly represented while others (especially developing countries) are virtually absent. The present study examines how editor geography is reflected in the editing of articles (participation, impact and success) about the independence of former French colonies in Africa. The analysis is based on 354 Wikipedia articles; by geolocating 75% of the editors (N = 23,408), we show that the majority of edits are made by users located in France. This imbalance is also reflected in the overall share of text they contribute over time. However, when looking at the individual user level, we find that editors from France are only slightly more successful in maintaining their contributions visible to the reader, than editors from African successor states.

Keywords: online participation, digital inequality, Wikipedia, geographies of knowledge, colonial history

1 Introduction

Wikipedia has become the key online source of encyclopaedic knowledge (Alexa, 2018). Its declared goal is to "bring about a world in which every single human being can freely share in the sum of all knowledge" (Wikimedia, n.d.a). For that reason, Wikipedia has (at least theoretically) the potential to democratize global participation in digital media as well as to diminish existing social inequalities by closing social gaps regarding the generation and representation of knowledge (Hargittai & Hsieh, 2013).

However, previous studies have shown that only a small proportion of Wikipedia users produce the content available on the platform (e.g. Ortega, Gonzalez-Barahona & Robles, 2008; Shaw & Hargittai, 2018). As a result, contributors and contributions are subject to numerous digital inequalities for example regarding gender (women are underrepresented on Wikipedia) or regarding the representation of geographical regions (e.g. Graham, Straumann & Hogan, 2015). The danger of these kinds of digital divides is that certain worldviews, opinions, or interpretations of events may be excluded from Wikipedia, as the available information is dominated by a particular and/or small group of editors.

Social inequalities and the influence of digital media on social stratification have already gained great scholarly attention (for an overview see Robinson et al., 2015). But previous research (e.g. Hargittai, 2002) mainly focuses on disparities in the access to digital tools and often neglects communication processes that occur within the digital realm – although differences in the usage may also lead to divides regarding online participation and representation (Barzilai-Nahon, 2006; Hargittai & Hsieh, 2013; Hargittai & Walejko, 2008; Min, 2010; Warschauer, 2004) and chasms regarding the engagement and efficacy in online environments may reflect social inequalities that exist offline (DiMaggio & Garip, 2012).

The aim of the present paper is to examine different forms of digital divides that occur on Wikipedia in the context of potentially conflicting historical events. To this end, we exemplarily study the depiction of processes of independence of former French colonies in Africa on the French language version of Wikipedia. Rather than investigating what is said about the former colonies, our study focuses on who said it—that is, the geographical distribution of editors who participate in describing the processes of independence, and the extent to which they succeed in asserting their version of history.

We contribute to the field in two ways. First, as mentioned above, only few studies so far exist that shed light on processes regarding the engagement in and efficacy of online participation which can lead to or foster digital inequalities. Our study helps to gain a deeper knowledge in and disentangle processes connected to digital divides that go beyond the analysis of mere participation. Second, existing Wikipedia-research addresses editor geography only in general settings (e.g. Graham, Straumann & Hogan, 2015), with little, and mostly qualitative research studying this aspect in the context of concrete geopolitical conflicts (e.g. Bilić & Bulian, 2014; Kumar, 2017; Luyt, 2017), where the issue is of particular interest. Our comprehensive quantitative analysis of Wikipedia accounts of a post-colonial conflict serves to consolidate these results. In line with previous empirical studies, we argue that an editors' geographical location is an important aspect when it comes to assess equal representation of knowledge on Wikipedia (Graham, Straumann & Hogan, 2015), as their editing decisions are not only influenced by the encyclopaedia's norms, but also by editors' territorial identities that, at least, potentially inform their viewpoints on geopolitical conflicts (Bilić & Bulian, 2014).

2 Digital divides in the geographical representation and the definition of historical events on Wikipedia

In their work on the social construction of reality, Berger and Luckman (1966) proposed an empirical research agenda for a "sociology of knowledge" to study processes by which any body of knowledge comes to be socially established as reality (Berger & Luckman, 1966, p. 15). A sociology of this kind does not assess the validity of knowledge but inquires into the "conditions of knowledge itself" (Berger & Luckman, 1966, p. 24). This involves studying those "individuals who serve as definers of reality" (Berger & Luckman, 1966, p. 134). In much the same way, Pocock (1998) states that the history of a country is always socially constructed and is embedded in comprehensive narratives. This is especially relevant when examining key historical events such as independence, which concern "the narrative and myth

of how the society is said to have come into being and acquired the capacity for autonomy" (Pocock, 1998, p. 219) and must therefore be understood as a means of identity creation (Pocock, 1998, p. 225). In the present case, then, the task is to unveil the social organization and interactions, which ultimately define and frame key historical events.

Berger and Luckman's (1966) approach proposes a shift of emphasis from what is the definition of reality to who defines that reality – in our case those who edit the Wikipedia. Wikipedia editors are powerful actors in this respect as they shape what may be perceived as canonical knowledge on the most popular global platform for encyclopaedic knowledge (Alexa, 2018). As one of its key principles, the Wikimedia foundation promotes the concept of knowledge equity—that is, the commitment to "counteract structural inequalities to ensure a just representation of knowledge and people" (Wikimedia, 2017). On that basis, representation of diverse editorial views is clearly an important norm both to users and to the Wikipedia itself, notwithstanding the empirical reality which is often characterized by power relations between editors that shape editorial processes (see e. g. Lerner & Lomi, 2017 and Lerner & Lomi, 2020 for the effects of past interactions between editors; Ford & Wajcman, 2017 for the impact of the editorial community's culture and policies).

In addressing this question, we rely on the various forms of different digital divides identified by previous research.

The first possible divide refers to mechanisms that exclude potential participants from the use of digital media, whether intentionally or not (Hargittai & Hsieh, 2013). Transferred to the context of Wikipedia, this form of divide includes anything that prevents people from participating in the process of knowledge production on the platform; ranging from a lack of internet access (which is related to the first-level digital divide) to a hostile discussion culture (Ford & Wajcman, 2017) or other participation barriers which belong per definition to the second-level divide.

A first indication of these kind of inequalities on Wikipedia is that contributions are very unevenly distributed. For example, Mattei and Britt (2017) showed that the top 1% of all editors are responsible for 77% of all editor activity while the bottom 20% are responsible for only 0.02%.

Several studies have also identified signs of gaps in the participation of editors from different geographical regions; among these, Graham (2015) found that participation on Wikipedia as measured by number of edits is highly skewed towards high-income countries. For example, while the United States, the UK and Germany each account for several million edits each quarter year, users from many countries in Africa and the Middle East are responsible for only a few thousand edits over that period (see also Graham, 2014; Graham & Hogan, 2014). All these findings relate to first-level digital divides, which encompass any mechanism that prevent potential participants from using digital media and lead to our first research question:

RQ1. Editors in which geographic regions participate in the editing of articles on processes of independence?

The second-level divide refers to differences related to skills, engagement, and efficacy regarding the use of digital media (Hargittai & Hsieh, 2013). In the case of Wikipedia, we argue that mechanisms of second-level digital divides include anything that impedes productive and lasting contributions from socially determinate groups (Hargittai & Walejko, 2008). For example, in his qualitative study of controversies regarding the designation of the river Ganga/Ganges-article, Kumar (2017) suggested that Wikipedia's policies tend to favour a Eurocentric point of view, so undermining neutral and impartial knowledge production (see also Luyt, 2017). Bilic and Bulian (2014, p. 42) identified "social, cultural and political aspects that drive the editing process and distort the objective, neutral content production". Zhou, Cristea and Roberts reported that war-related Wikipedia articles tend to exhibit a more negative sentiment if written in the official languages of the belligerents, and that "people speaking different languages have different focuses and interests about the same war-related topic" (2015, p. 167). Against this backdrop and the enormous historical relevance of independence for both colony and colonizer, we are interested in two indicators with regards to the efficacy of editors in different geographical regions. First, the impact of editors on the encyclopaedia's content can be determined by how much text they contribute to articles and for how long these contributions are visible to readers. This results in our second research question:

RQ2. Editors in which geographic regions have the strongest impact on the depiction of processes of

independence on Wikipedia?

Second, an editor's success can be understood as their effective impact on an article's content relative to their potential impact in terms of their contribution, resulting in our third research question:

RQ3. Do editors in different geographic regions differ with regard to their individual success in the editorial process?

3 Methods and measurement

3.1 Case selection and sampling

Our sample comprises French Wikipedia articles about the independence processes in France's former African colonies, which include 17 countries (Truhart, 1996, see Online Appendix). We decided to study these processes, first, because the territories of contemporary states can be clearly traced back to former French colonies and second, because all of the selected countries directly achieved independence, while other territories received legal status of League of Nations mandate first. Additionally, the French Wikipedia is among the larger language versions; as of September 2019, it is the fifth largest worldwide and the third largest among former European colonial powers involved in independence struggles (Wikimedia, 2019). As French is still widely spoken in the former colonies and in some cases remains an official language (Central Intelligence Agency, n. d.), we could expect to find sufficient levels of editorial activity for the purposes of our study.

Having decided to include all articles about actors, processes, or documents related to a country's independence, we first identified one article for each country that covered its independence process in the greatest detail and length. From there, we checked all outgoing links for further relevant articles. We identified 354 relevant articles in this way between October 29th 2018 and January 3rd 2019 with an overwhelming majority of articles on topics regarding the independence of Algeria (212, note that articles may relate to the independence of more than one country), Tunisia (81) and Morocco (30). On February 8th 2019, we retrieved the complete revision history for all articles in our sample. We sampled the content of each edit and extracted the usernames of all editors, from the article's creation to its current version. In total, this process yielded 122,931 edits by 23,408 editors.

3.2 Identification and geolocalization of editors

Of the contributing editors, about 70% (n = 16,236) were unregistered users (the average monthly share of unregistered editors is 75% for the entire French Wikipedia between June 2001 January 2019, according to the data provided by Wikimedia (n. d.b)). As Wikipedia saves these editors' IP address as their username, it was possible to locate them by using an IP registry (Maxmind, 2019) and to detect their location automatically. The remaining editors (n = 7,172) were registered users, which made the geocoding task more dificult. To locate these editors, we performed two further coding steps. First, we identified the templates most frequently used by contributing editors (e.g. 'Utilisateur: habite' = 'this user lives in…') on their user pages and classified this information automatically where available. Second, we identified the most active of the remaining registered editors (n = 480). These contributors were collectively responsible for 80% of all edits and/or for 80% of content. We checked manually for any geographical information on their user pages (residence, country of origin, and nationality in the order of coding preference). From the population of all contributing editors (N = 23,408), this combined approach enabled us to locate 75% (n = 17,471). These geocoded editors formed the basis of our analysis.

3.3 Attribution of text to individual editors

Using an implementation of Myers's (1986) diff-algorithm (Google, 2019), we compared each article revision to its predecessor, enabling us to attribute new text in a given revision to its author and their geographical location. By doing so, we were able to calculate an editor's share of the text for every version of every article. We estimate the accuracy of this approach to text attribution as 88.9% (p = 0.05).

3.4 Measuring editorial impact and individual success

Measuring the impact of editors or groups of editors on Wikipedia articles is a complex task, as articles are never finished but remain in an ongoing process of revision. For that reason, it is essential to take account of the temporal dimension. We computed the impact score (I), as the sum of products of all text portions contributed by each editor to an article and the duration of their visibility to Wikipedia users (see Formula 1).

Figure 1 shows the application of this approach to an

article about the Moroccan Manifest of Independence of 1944 by specified user groups (French, former colonies, third states). A similar visual approach (slightly different in scope) was proposed by Viégas, Wattenberg and Kushal (2004). The Impact (I) of a given group corresponds to the respectively coloured area. In this case, editors from former colonies can be seen to have had the largest impact on the article, accounting for 61% of the coloured areas as compared to 13% by French contributors.

However, it remains in question whether users differ in terms of their individual success rates when controlling for the participation bias, where 'success' refers to the likelihood that a given contribution will remain visible for a period of time. An accurate measure of an individual editor's success must therefore take account of the extent of their positive text contributions relative to what remains visible. On that basis, we defined an editor's success rate (S) as the observed individual impact (I) divided by potential impact—that is, the impact achieved if all of an editor's added text remained visible until the last revision in our data set (see Formula 2). The success rate varies between close to 0 and 1, where 1 indicates that all text contributed at a given point in time remained to the last revision. On the other hand, a low rate indicates either that little of an author's contribution remains or that it remained for a relatively small period of time (or some combination of both). As the success rate could not be estimated for editors who deleted text without adding any text of their own, they were excluded from the analysis.

4 Results

The first aim of this study was to examine participation by analysing the geographic locations of editors contributing to Wikipedia articles on processes of independence. In total, editors in 112 different countries contributed to the articles in our sample. Of these, 70% were located in France; 11% were located in former colonies, and 19% were located in third countries (N = 17,471). Among former colonies, the largest group of editors came from Algeria (n = 1,125), followed by Morocco (n = 390) and Tunisia (n = 281). Figure 2 shows detailed results for the location of participating editors on a world map.

A similar pattern was identified in relation to editor activity, with 67% of all edits performed in France, 13% in former French colonies and 20% in third countries (N = 64,741). Again, the most active former colony was Algeria, although France and Algeria's relatively higher activity reflects the large number of editors in those countries. French editors performed 3.5 edits on average (SD = 19.53) as compared to 4.5 edits by editors in the former colonies (SD = 45.59). Among the ten most active editors, four were based in France, four in former colonies and two in third countries. These ten editors alone were responsible for 17% of all geolocated edits.

Regarding our first research question, we can conclude that, in absolute numbers, editors located in France predominate, indicating some degree of divide in the editing of articles on the processes of independence of former French colonies. However, among the ten most active editors, the ratio of French editors and editors from former colonies is relatively balanced.

In relation to our second and third research question, we drew on our operationalization of editorial impact and individual success to determine whether editors from different geographical backgrounds differ in terms of efficacy. As shown in Table 1, 61% of all geolocated text over time is authored by editors in France, as compared to only 21% by editors in third states and 18% by editors based in successor states of former colonies. A comparison of these results with those presented above reveals similar distributions of editors, edits and impact, although editors in former colonies have a relatively higher impact and editors in France a relatively lower impact in relation to their respective participation rates. Nonetheless, the overall distribution gives rise to the suspicion that the French dominance in terms of impact merely reflects the higher participation rates of editors from France rather than any difference in editorial success. In fact, we found a highly significant rank correlation between number of edits and impact for individual editors (Kendall's tau = .33, z = 65.9, $p < .0001$) regardless of user group, indicating that the most active editors are among those with the highest impact scores (and vice versa).

Turning to our third research question we calculated the success rates for all editors who added text to at least one of the articles in our sample. The average success rate for all geolocated editors was .33; while the average for French users hits that overall average exactly, editors from former colonies fall slightly below at .30. Although this difference is very small, a two-sample t-test revealed that it is statistically significant ($p < .01$); in general, this means that users in former colonies are on average slightly less successful than their counterparts in France in terms of

maintaining a visible contribution.

In sum, the data indicate that editors from France have a greater impact and a marginally higher success rate than their counterparts in former colonies.

5 Discussion

In this study we examined to which extent digital inequalities may be present in the social dynamics of content creation on Wikipedia as exemplified by the geographical representation of editors and their contribution to articles on processes of independence. We integrated theoretical approaches dealing with digital disparities (Hargittai & Hsieh, 2013; Robinson et al., 2015) with Berger and Luckman's (1966) argument that all knowledge is socially constructed and that it is therefore important to study the processes and conditions of how available knowledge is established and who is able to define the shared reality.

For that reason, we focused on the editors' geographical distribution. As existing research has shown (Graham, Straumann & Hogan, 2015), editors from industrialized world regions such as Europe are by far the most active, while those from developing countries barely participate in the production of knowledge on Wikipedia.

The present study reassesses those findings in the context of the shared history of France and its former colonies' successor states. We went on to explore who has the greatest impact on Wikipedia articles over time and to determine whether groups of editors differ in terms of editorial success. These distinctions allowed us to gain a deeper knowledge of different forms of the second-level digital divide, namely of how exactly the engagement and efficacy related to online communication varies among different geographical groups.

With regard to the participation of different geographical regions, we find a clear majority of editors in France as compared to editors in former colonies or third states. These findings align with existing research, showing that participation rates in African countries are extremely low and that content related to those countries is edited mainly by users residing elsewhere (Graham & Dittus, 2018). We interpret these findings as evidence of participation divides. From the normative standpoint of Wikimedia's goals, unequal participation is problematic, as presence on the platform is the first prerequisite for being able to contribute. Editors from diverse geographical regions seem ideally positioned to construct the perceived reality of processes of independence of former French colonies. While our findings do not clarify which factors hinder participation, past research has advanced a range of explanations. The Wikimedia Foundation may be able to influence some of these, such as technical complexity of the editing interface and a hostile editor community (Ford & Wajcman, 2017), but it is powerless when it comes to address other factors such as broadband penetration rate, leisure time, access to technology, and education (Graham & Hogan, 2014; Robinson et al., 2015). However, the few editors stemming from former colonies who participate in the editing process are very active. It is plausible to assume that these few very active editors from former colonies make their geographical location especially salient to demonstrate and emphasize their competence regarding the topic. Further research should explore more deeply how these editors could be characterized, perhaps they could serve as multipliers or 'opinion-leaders' to convince others to participate.

We then developed a simple metric that measures an editor's impact on an article's text over time, and we aggregated this for editor groups. Using the vocabulary of Berger and Luckman (1966), this provided further confirmation that the definers of the history of independence of former French colonies in Africa are mainly based in France. However, as these results correspond roughly to the distributions of editors and edits, we contend that the unequal impact of different user groups can perhaps again be traced to the participation bias, in the sense that more active user groups account for a higher share of text over time. This aligns with our finding of a significant positive correlation between the number of edits and editor impact.

To disentangle these data, we defined an individual success rate for editors as the ratio (expressed as a percentage) of their contributions' observed impact to their potential impact if none were ever deleted. Despite their low participation rates, editors from former colonies are relatively successful, but their counterparts in France are, on average, slightly more successful. However, our findings regarding the second-level divide are less clear than for mere participation, given that the marginal differences in success rates provide only weak evidence of systemic patterns of inequality among user groups.

The study has some conceptual limitations; the most obvious of these is that we did not analyse the content of the included articles. The fact that a majority of the editors were from France provides no information about the quality of the displayed content. It may well be that the French editors describe the history of independence in a fair and balanced way. Even if, as argued here, understanding a constructed reality depends on understanding the social groups who construct that reality, further research should link the definers and their definitions. In the present context, that means relating the participation of distinct geographical groups and associated participation rates to Wikipedia content. In this regard, the geographical origin of the sources used to substantiate certain views on historical events comprise an additional dimension of analysis for future research.

The present research does not illuminate the underlying mechanisms that might explain differences in participation and contribution, such as why specific groups of editors are less successful in certain situations. Future research should look more closely into these dynamics. Additionally, we faced methodological limitations, with regard to geolocalization. We were able to geolocate only 74% of all editors and 52% of edits. While the latter is at the level of previous research (Graham, Strauman and Hogan, 2015), clearly, a better geocoding rate would improve the quality of our results.

In sum, the present findings confirm the need for the Wikimedia Foundation, as an organization that actively seeks diversity in its community of editors, to take account of external and internal exclusion issues. As we have shown, there is still a long way to go in recruiting new community members, but there is also much to be gained. Some of the aspects that contribute to geographic participation bias, cannot readily be influenced by the Wikimedia Foundation. Similarly, internal exclusion factors are complicated, subtle and difficult to tackle or even identify. However, providing a space in which editors are afforded equal opportunities for successful participation is a sustainable investment in developing a more diverse editor community.

Online Appendix

https://www.uni-passau.de/fileadmin/dokumente/fakultaeten/phil/lehrstuehle/schmid-petri/Online_Participation_and_Social_Construction_of_history_appendix.pdf

References

Alexa. (2018). *The top 500 sites on the web.* https://www.alexa.com/topsites

Barzilai-Nahon, K. (2006). Gaps and bits: Conceptualizing measurements for digital divide/s. *Information Society*, 22(5), 269-278.

Berger, P. L., & Luckmann, T. ([1990], 1966). *The social construction of reality: A treatise in the sociology of knowledge* (Repr). Anchor Books.

Bilic, P., & Bulian, L. (2014). Lost in Translation: Contexts, Computing, Disputing on Wikipedia. In *iConference 2014 Proceedings* (p. 32–44). https://doi:10.9776/14027

Central Intelligence Agency (n. d.). *The World Factbook*. https://www.cia.gov/library/publications/resources/the-world-factbook/

DiMaggio, P. & Garip, F. (2012). Network effects and social inequality. *Annual Review of Sociology*, 38, 93-118.

Ford, H. (2011). The missing Wikipedians. In G. Lovink & N. Tkacz (Eds.), *Critical Point of view: A wikipedia Reader* (pp. 258–268). Amsterdam: Institute of Network Cultures.

Ford, H., & Wajcman, J. (2017). 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. *Social Studies of Science*, 47(4), 511–527. https://doi.org/10.1177/0306312717692172

Google (2019). Diff Match Patch Is a High-Performance Library in Multiple Languages That Manipulates Plain Text. Python. Google. Retrieved from https://github.com/google/diff-match-patch

Graham, M. (2011). Wiki Space. Palimpsests and the Politics of Exclusion. In G. Lovink & N. Tkacz (Eds.), *Critical Point of view: A wikipedia Reader* (pp. 269–282). Institute of Network Cultures.

Graham, M. (2014). Inequitable Distributions in Internet Geographies: The Global South Is Gaining Access, but Lags in Local Content. *Innovations: Technology, Governance, Globalization*, 9(3-4), 3–19. https://doi.org/10.1162/inov_a_00212

Graham, M. (2015). Information Geographies and Geographies of Information. *New Geographies*, 7, 159–165.

Graham, M., & Dittus, M. (2018). *The Uneven Geography of Wikipedia.* https://geography.oii.ox.ac.uk/the-uneven-geography-of-wikipedia/

Graham, M., & Hogan, B. (2014). *Uneven openess: Barriers to MENA representation on Wikipedia.* Oxford Internet Institute Report. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2430912

Graham, M., Straumann, R. K., & Hogan, B. (2015). Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia. *Annals of the Association of American Geographers*, 105(6), 1158–1178. https://doi.org/10.1080/00045608.2015.1072791

Hargittai, E. (2002). Second-Level Digital Divide: Differences in People's Online Skills. *First Monday*, 7(4), 1-13.

Hargittai, E. & Hsieh, Y. P. (2013). Digital inequality. In W. H. Dutton (Ed.), The Oxford Handbook of Internet Studies. Oxford Handbooks online: 10.1093/oxfordhb/9780199589074.013.0007

Hargittai, E. & Walejko, G. (2008). The participation divide: content creation and sharing in the digital age. *Information, Communication and Society*, 11(2), 239-256.

Kumar, S. (2017). A river by any other name: Ganga/Ganges and the postcolonial politics of knowledge on Wikipedia. *Information, Communication & Society*, 20(6), 809–824. https://doi.org/10.1080/1369118X.2017.1293709

Lerner, J., & Lomi, A. (2017). The Third Man: hierarchy formation in Wikipedia. *Applied Network Science*, 2, Article 24. https://doi.org/10.1007/s41109-017-0043-2

Lerner, J., & Lomi, A. (2020). The free encyclopedia that anyone can dispute: An analysis of the micro-structural dynamics of positive and negative relations in the production of contentious Wikipedia articles. *Social Networks*, 60, 11-25. https://doi.org/10.1016/j.socnet.2018.12.003

Luyt, B. (2017). "A wound that has been festering since 2007": The Burma/Myanmar naming controversy and the problem of rarely challenged assumptions on Wikipedia. *Journal of Documentation*, 73(4), 689–699. https://doi.org/10.1108/JD-09-2016-0109

Maxmind. (2019). *GeoLite2 IP City Database*. Release: 30.04.2019. Maxmind Inc. https://dev.maxmind.com/geoip/geoip2/geolite2/

Min, S.-J. (2010). From the digital divide to the democratic divide: Internet skills, political interest, and the second-level digital divide in political internet use. *Journal of Information Technology & Politics*, 7(1), 22-35.

Myers, E. W. (1986). AnO(ND) difference algorithm and its variations. *Algorithmica*, 1(1-4), 251–266. https://doi.org/10.1007/BF01840446

Ortega, F., Gonzalez-Barahona, J. M., & Robles, G. (2008, January). On the Inequality of Contributions to Wikipedia. *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008). IEEE.* https://doi.org/10.1109/HICSS.2008.333

Pocock, J. G. A. (1998). The Politics of History: The Subaltern and the Subversive. *Journal of Political Philosophy*, 6(3), 219–234. https://doi.org/10.1111/1467-9760.00053

Polikanov, D. & Abramova, I. (2003). Africa and ICT: A chance for breakthrough? *Information, Communication & Society*, 6(1), 42-56.

Robinson, L., Cotten, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., Schulz, J., Hale, T. M. & Stern, M. (2015). Digital inequalities and why they matter. Information, *Communication & Society*, 18(5), 569-582.

Shaw, A. & Hargittai, E. (2018). The pipeline of online participation inequalities: The case of Wikipedia editing. *Journal of Communication*, 68, 143-168.

Truhart, P. (1996). Lexikon der Historischen Staatennamen: Von den Ursprüngen bis zur Gegenwart [Historical Dictionary of States: States and State-like Communities from Their Origins to the Present]. De Gruyter.

Viégas, F. B., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors

with history flow visualizations. In E. Dykstra-Erickson (Ed.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 575–582). New York, NY: ACM. https://doi.org/10.1145/985692.985765

Warschauer, M. (2004). Technology and social inclusion: rethinking the digital divide. MIT Press.

Wikimedia (n.d.a). *Wikimedia. Landing page.* Retrieved from https://www.wikimedia.org/

Wikimedia (2017). *Wikimedia Movement Strategy 2017.* https://meta.wikimedia.org/wiki/Vision

Wikimedia (2019). *List of Wikipedias.* https://meta.wikimedia.org/wiki/List_of_Wikipedias

Wikimedia (n. d.b). *Wikimedia Statistics.* https://stats.wikimedia.org/#/fr.wikipedia.org/contributing/editors/normal|table|2001-01-01~2019-03-01|editor_type~anonymous*user|monthly

Zhou, Y., Cristea, A., & Roberts, Z. (2015). Is wikipedia really neutral? A sentiment perspective study of war-related wikipedia articles since 1945. *In Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation.*

$$I = \sum_{i=1}^{j} c_i \cdot \left(t_{i+1} - t_i \right)$$

Formula 1: Impact score (I), where c is the number of characters contributed to a given revision; t is the revision's Unix time stamp (describing the number of seconds elapsed since January 1st 1970 00:00:00 UTC) and j is the total number of observed revisions.

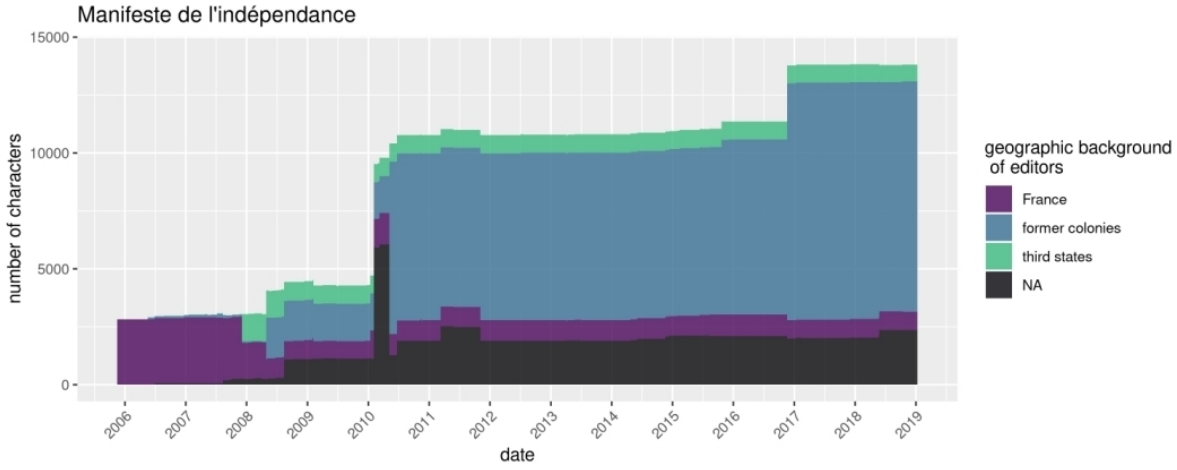

Figure 1: Share of text by user group over time for the article 'Manifeste de l'indépendance'.

$$S = \frac{\sum_{i=1}^{j} c_i \cdot \left(t_{i+1} - t_i \right)}{\sum_{i=1}^{j} c'_i \cdot \left(t_j - t_i \right)}$$

Formula 2: Success rate (S), where c' corresponds to the number of additional characters introduced in a given revision.

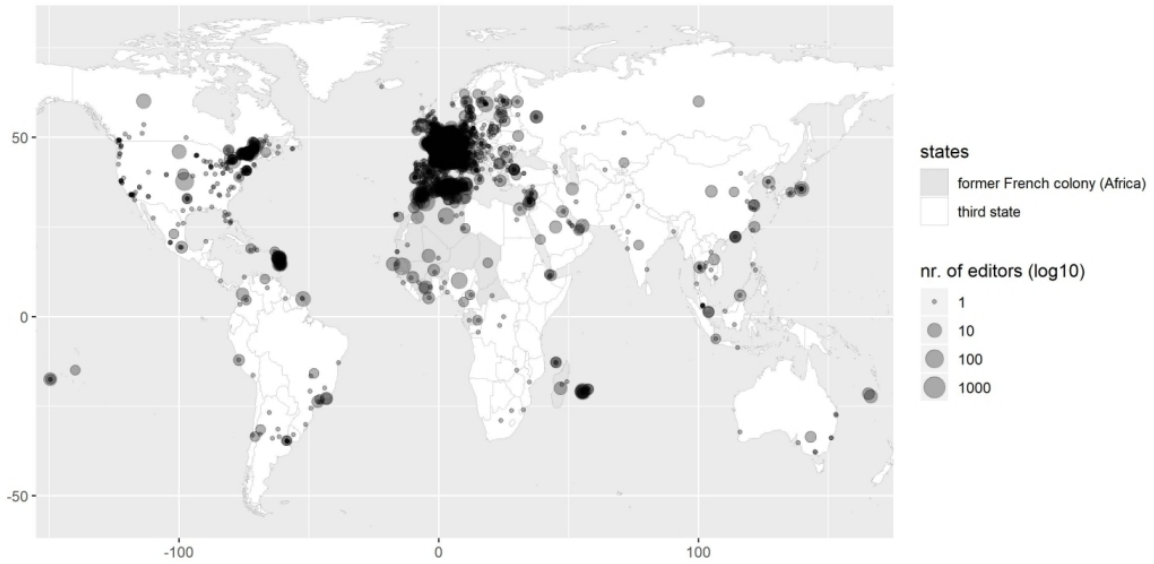

Figure 2: Locations of participating editors (N = 17,471). (Circles show editor locations. Circle size indicates the number of editors at that location on a log transformed scale. Where editors could only be located at state level, a dot is placed at the country's geographical centre. Former French colonies are shown in darker grey.)

	Editors	**Edits**	**Impact**
France	70 %	67 %	61 %
Former colonies	11 %	13 %	18 %
Third states	19 %	20 %	21 %

Table 1: Share of editors, edits and impact for geolocated users (N = 17,471)