



Bridging the offline and online: 20 years of offline meeting data of the German-language Wikipedia

Nicole Schwitter¹ 

Received: 31 January 2023 / Accepted: 10 September 2023 / Published online: 26 September 2023
© The Author(s) 2023

Abstract

Wikipedia is one of the most visited websites worldwide. Thousands of volunteers are contributing to it daily, making it an example of how productive non-market collaboration on a very wide scale is not only viable but also sustainable. Wikipedia's freely available data on the online actions conducted make it a popular source of data, particularly for computer scientists and computational social scientists. This data brief will present the *dewiki meetup dataset* which covers the offline component of the German-language version of the online encyclopaedia Wikipedia: informal offline gatherings between Wikipedia contributors. These gatherings are organised online and information about who is attending them, where they take place and what has happened at these meetings is shared publicly. The *dewiki meetup dataset* covers almost 20 years of offline activity of the German-language Wikipedia, containing 4418 meetups that have been organised with information on attendees, apologies, date and place of meeting, and minutes recorded. It is a valuable source of data for social science research: it captures the development of the offline network over time of one of the largest and most sustainable online public goods and communities. The data can easily be merged with online activity data on Wikipedia which allows us to bridge the gap between offline and online behaviour.

Keywords Wikipedia · Face-to-face meetings · Dataset · Big data · Network data

Introduction

The digital revolution and the widespread use of the internet have influenced and changed many realms of empirical social science research. Novel digital sources of data are becoming popular to gain new insights into old and new questions of the social sciences [7, 16, 21, 22, 38]. In today's internet landscape, the online encyclopaedia Wikipedia takes on a key role: it is one of the most visited websites

✉ Nicole Schwitter
nicole.schwitter.4@warwick.ac.uk

¹ Department of Sociology, University of Warwick, Coventry CV4 7AL, UK

worldwide, is at the backbone of many modern technologies and functions as an important source of information for the general public. It has also been met with strong interest by the scientific community as the encyclopaedia and its specific software structure have created a rich and freely accessible data source, offering the opportunity to study large-scale, self-organising collaboration networks. In addition to the online activity of Wikipedians, the platform also features a notable offline component which has been largely neglected in previous research. Wikipedia is characterised by regular local offline meetups which give editors a time and place to get to know each other personally. In the typical spirit of Wikipedia, these meetings are organised publicly and are well documented with lists of attendees and minutes. This data brief presents the newly published *dewiki meetup dataset* containing 4418 meetups that have been organised on the German-language version of Wikipedia. The dataset contains information on the date and place of a meetup, of the users attending it, of those excusing their absence, of the minutes recorded, and the source information. The dataset thus provides a rich and granular view of the offline component of Wikipedia. It does not only shed light on the magnitude of meetups but it also paints a narrative of their social dynamics.

The dataset covers almost 20 years of offline activity of the German-language Wikipedia, one of the largest and most active language versions (featuring, as of January 2023, over 2.7 million content sites, having been edited a total of over 225 million times).¹ From its launch in 2001 to March 2020 when face-to-face meetings came to a halt due to the outbreak of the Coronavirus pandemic, almost all meetups were collected and can now be easily merged with online activity data. These data allow to shed light on a part of Wikipedia which has previously been overlooked. Wikipedia has been a popular source of data for scholars of various disciplines: from computer science to linguistics and the social sciences, researchers have used data from Wikipedia to answer a multitude of scientific questions. For example, Wikipedia's structured and large body of text in multiple languages has been used to develop, improve, and showcase new algorithms, particularly in the field of natural language processing and supervised machine learning in general [see e.g. 8, 9, 14, 17]. Other studies are more focussed on Wikipedia as an online phenomenon itself and do not consider it as only a data-generating platform. Such research includes studies about the (determinants of) quality and credibility of article content [see, e.g. 1, 10, 18, 20, 24, 25, 34, 37, 42, 46], or tries to better understand the community of Wikipedians as contributors towards an online public good with remarkable sustainability [see, e.g. 11, 26, 30–32, 44]. In a broader societal context, Wikipedia data have also been used to predict real-world events such as election outcomes [47], movie success [27], or stock market moves [28], and cultural differences have been investigated by making use of different language versions [4, 12]. The rich source of data and its publicness also allow us to investigate fundamental questions of the social sciences, such as the role of networks when explaining norm enforcement [33] or how group size affects contributions to a public good [48]. The *dewiki meetup dataset* is a valuable source of data for social science research: it captures

¹ See for numbers about any language version of Wikipedia: <https://stats.wikimedia.org/>.

the development of the offline network over time of one of the most sustainable, online public goods and the community producing it. It allows to extend previous research of Wikipedia with the important dimension of offline interactions.

This article is structured as follows: in the next section, Wikipedia and the community of Wikipedians will be described in more detail. After this, it will be described how offline meetings have been collected and how the *dewiki meetup dataset* can be combined with online data to unzip its full potential. The collected data will then be described and novel research opportunities for computational social scientists will be highlighted thereafter. In the last section, conclusions will be drawn and access and limitations regarding the data discussed.

Wikipedia: more than just an encyclopaedia

There are specific challenges to using digital trace data for research. It is important to know the underlying data-generating process and context, meaning the dynamics of the site where the data were produced; understanding this context further helps to contextualise and understand potential ethical, legal, and methodological issues which come with the data. This section will introduce Wikipedia as a community creating an encyclopaedia.

Wikipedia was founded in January 2001 and has been dedicated to the building of free encyclopaedias in all languages of the world. The goal of Wikipedia was and is to engage ordinary, uncredentialed people using a wiki format. The focus on creating an encyclopaedia provides the common task and the open content license works as a motivating force for people to work for the good of the world. By now, Wikipedia exists in almost 300 language versions. The German-language Wikipedia is the second-oldest language version, founded in March 2001, and, by now, it is now one of the largest ones. While Wikipedia can be read and edited freely and anonymously, the German-language Wikipedia features over 3 million registered users, of which over 17,000 are currently actively contributing.

With the encyclopaedia, a community has grown. Wikipedia features not only encyclopaedic contents, but also offers spaces for interaction between contributors: articles are generally the result of a collaborative effort of multiple users [see, e.g. 2], contributors take care of the enforcement of norms by making use of positive and negative sanctions—for example, through awards [see, e.g. 36] and through reverting unsuitable article edits [see e.g. 11, 33]—both the article pages as well as the profile pages of registered users feature designated discussion pages where Wikipedians discourse [see e.g. 19], and users can take part in democratically organised polls and election processes where new guidelines are set or new administrators selected [see e.g. 3, 15]. Offline meetups between Wikipedians are one further avenue for interaction and are important to the community: as long-term member and active Wikipedian Richter [35: 132–136] writes, such face-to-face meetings allow users to connect to others and help in times of conflict; they can fulfil a Wikipedian's needs for social contacts, community and personal exchange, in the same way as other local associations. Richter [35: 148] further states that personal acquaintances

are central to a project that is based on anonymous contributions. These ties allow to create a net of trust, making the collaborative labour easier.

Much research has focussed on the online component of Wikipedia with the online activity data generally coming in a readily available, process-generated, structured format through the Wikimedia data dumps² or, at least, being available online in a consistent format, making it straightforward to web-scrape [see e.g. the election data in 23]. A systematic study of offline gatherings proves to be more difficult with the user-generated data being inconsistent and messy. Data on (almost) all meetups organised in the German-language Wikipedia were collected and published in the *dewiki meetup dataset* to open up this avenue of research. The collection process will be outlined in the next section.

Collection of meetup data

Offline meetups between Wikipedians are generally organised online. Meeting data are thus publicly available on Wikipedia; however, the data are lacking a clear and consistent structure as they are user-written. The process of collecting these data is explained in detail in the following.

Scraped pages

The starting point of the meetup collection was an overview list of meetings between Wikipedians.³ This list includes the links to over one hundred subpages relating to regions and cities where meetups between German-speaking Wikipedians are organised and archived. These regionally based meetups are most often informal meetings with the point of socialising in public spaces (the so-called *Stammtische*). In addition, all editathons and open editing events were collected.⁴ These are events where (potentially new) editors of Wikipedia meet to edit and improve a specific topic or type of content. They generally include basic editing training for new editors and are often combined with a social meetup. There are both online and in-person editathons; however, virtual editathons were not collected as part of this dataset.

Furthermore, all events listed on two overview event sites were collected.⁵ These events include activities such as attending and looking after stalls representing Wikipedia at fairs, partaking in workshops about photography, and similar events. This also includes events organised as part of the *GLAM* initiative (GLAM stands for Galleries, Libraries, Archives, Museums)⁶ in which cultural institutions are

² For the German-language Wikipedia, see <https://dumps.wikimedia.org/dewiki>.

³ See https://de.wikipedia.org/wiki/Kategorie:Wikipedia:Treffen_der_Wikipedianer.

⁴ Listed here <https://de.wikipedia.org/wiki/Wikipedia:Edit-a-thon> and https://de.wikipedia.org/wiki/Wikipedia:Offenes_Editieren.

⁵ See here <https://de.wikipedia.org/wiki/Kategorie:Wikipedia:Archiv/Veranstaltungen> and here https://de.wikipedia.org/wiki/Kategorie:Wikipedia:Versicherte_Veranstaltung.

⁶ See <https://de.wikipedia.org/wiki/Wikipedia:GLAM>.

supported through collaborative projects with experienced Wikipedia editors. Similar to these are the so-called *KulTouren*⁷ which were also collected: these are smaller scale events where Wikipedians visit exhibitions or take part in excursions. Lastly, all WikiProjects⁸ and task forces⁹ were checked for meetings. WikiProjects and task forces are central places for discussing specific content; they are used to communicate, collect sources, and provide summaries on specific topics. They form a sort of virtual gathering place for Wikipedia editors interested to work on a specific cluster of topics.

Throughout the scraping of all these pages, a snowballing approach was followed: when one meetup page was linked to other ones, these were additionally collected until no new pages were found. Still, there is no guarantee that all pages with meetups were visited and scraped. The approach of data collection has been outlined in a Wikimedia Learning Pattern—a Wikimedia specific guideline that provides advice, shares insights and is easily accessible to people and researchers in the Wikimedia community—including a code example to foster future data collection across different language versions.¹⁰

Excluded pages and excluded meetings

Some pages and meetings were excluded from the data and/or the data collection process. First, all meetups that took place only virtually were skipped. In the observation period, only a small number of editathons have taken place online (this rapidly changed after the outbreak of the Coronavirus pandemic). Next, portals¹¹ were not checked for meetups unless they are covering regional entities (portals about cities or regions). Portals are somewhat similar to WikiProjects and task forces but are directed towards readers instead of editors. Portals provide well-maintained introductory landing pages into the encyclopaedia; they provide an overview of the most important articles in a certain topic area. As these are thus not places where authors are gathering, it was not expected that any meetings are organised in the context of portals.

This dataset is further restricted to meetings organised on the German-language version of Wikipedia. Some meetings which are also directed towards German-speaking Wikipedians are organised on other platforms maintained by Wikimedia, such as commons¹² or meta.¹³ However, anything that is not organised on the German-language Wikipedia was excluded from data collection.

⁷ See https://de.wikipedia.org/wiki/Wikipedia:Wikipedianische_KulTour.

⁸ See <https://de.wikipedia.org/wiki/Wikipedia:WikiProjekte>.

⁹ Called *Redaktionen* in the German-language Wikipedia, see https://de.wikipedia.org/wiki/Wikipedia:Redaktionen#Liste_der_Redaktionen.

¹⁰ See https://meta.wikimedia.org/wiki/Learning_patterns/Collecting_data_on_offline_meetups.

¹¹ See <https://de.wikipedia.org/wiki/Wikipedia:Portale>.

¹² See <https://commons.wikimedia.org>.

¹³ See <https://meta.wikimedia.org>.

In addition, very regular meetings taking place in community spaces were excluded from the dataset. Such community spaces exist in some cities in the German-speaking region and offer headquarters for staff members and engaged volunteers. They are places of often extraordinarily high Wikipedia activity: in some spaces, multiple meetings and open editing events per week take place. These meetings are often attended by the same very small, core group of editors. In many cases, the attendees stop recording their attendance. As these community spaces exhibit a very different dynamic to other meetups and as it was often impossible to reliably collect data on the attendees, the following very regular meetings were excluded from the dataset/data collection:

- Berlin, Tempelufer bureau: exclusion of general open editing events, exclusion of open editing events for women.
- Berlin, WikiBär: exclusion of general open editing events, exclusion of open editing events for women.
- Berlin, WikiWedding: exclusion of open Sundays, office hours.
- Hamburg, Kontor: weekly events are not organised on Wikipedia and were thus not collected in the first place.
- Hanover: exclusion of Tuesday editing events, open editing events, office hours.
- Cologne, Lokal K: weekly events are not organised on Wikipedia and were thus not collected in the first place.
- Munich, WikiMUC: all regular events are excluded, such as open evenings, introductory workshops to Wikipedia, board game Fridays, monthly meetings of another organisation, monthly work meetings,¹⁴ cleaning events and other internal office events.
- Stuttgart, Stadtbibliothek: exclusion of monthly open editing events.
- Ulm, Verschwörhaus: exclusion of monthly open editing events.
- Vienna: exclusion of Wikipedia Tuesdays.

Other events taking place in these amenities were not excluded. These non-excluded events are those that are still considered regular meetups by the community (for example all events organised in the *technik.cafe* in Lörrach), are irregular ones taking place in the community spaces (such as specially organised workshops), and include other irregular events which make use of the location but are organised externally, such as meetings by project teams.¹⁵

The exclusion of certain meetups is not ideal, but nevertheless necessary as those excluded do not allow for a reliable collection of data, specifically on the attendees.

¹⁴ See <https://de.wikipedia.org/wiki/Kategorie:Wikipedia:WikiMUC/Arbeitstreffen>.

¹⁵ Community spaces might be better understood in the context of movement organisers [45].

Collection of information: automatic scraping and manual extraction

The data collection aimed to collect information on at least the date, place/venue, and attendees of all offline gatherings, except those excluded, since the launch of Wikipedia until March 2020. In most cases, the data collected also included apologies and minutes about the meetup.

The way meetings were organised is depicted in Fig. 1; the example is taken from the Rhine-Hessian regional organisation page.

It varied how meetings were organised and archived. Generally, there were the following two approaches:

1. An organised archive of all meetings with a consistent structure was created.¹⁶ In these cases, every meetup was recorded, and data on—at least—the attendees, date and place/venue of the meeting are available. In terms of data collection, this was the best-case scenario as it allowed writing an automated script for the archive.
2. Meetings were not archived at all. The organisational pages were used to organise only the most recent meeting.¹⁷ Due to Wikipedia's technical structure, it is still possible to retrieve information about past meetings using the version history. In these cases, it was necessary to scan through the complete version history to find past meetings before they had been deleted in favour of the next meeting.

These are two ideal types. In reality, they occurred in different sub- and hybrid forms. In the best version of case 1, all meetings were archived on a single page and all meetings were recorded following a consistent structure. In a less ideal version of case 1, all meetings were recorded in archives, but there were separate archives for single years and the structure and format of the archives varied between years.¹⁸ In some cases, organisational pages did not maintain an archive, but at least provided an overview of all meetings and linked to the respective pages in the version history.¹⁹ In other cases, only some meetings were archived; for example, in the case of Berlin, only meetings up to 2010 and then again in 2016 were archived, but not in all other years. In Cologne, most meetings were recorded, but some were simply left out. This was the most unfortunate case as the skipped meetings could only be noticed when the rhythm of meetings was broken (i.e. they seemed to have monthly meetings but there was a month without one) and then the version history needed to be checked manually. Overall, when possible, an automatic scraper was written to extract the information. Still, most meeting data were collected manually.

¹⁶ See for example Zurich <https://de.wikipedia.org/wiki/Wikipedia:Zürich/Archiv>.

¹⁷ See for example Potsdam <https://de.wikipedia.org/wiki/Wikipedia:Potsdam>.

¹⁸ See for example Hanover https://de.wikipedia.org/wiki/Wikipedia:Treffen_der_Wikipedianer/Hannover.

¹⁹ See for example Mainz <https://de.wikipedia.org/wiki/Wikipedia:Mainz>.

1st Meeting [edit source]

1st Meeting on Saturday, March 15th 2008 6pm onwards in Alzey, in the Greek restaurant across the (main) railway station.

Attendees [edit source]

1. [Kandschwar](#) 20:10, Feb. 15th 2008 (CET)
2. [kloppo76](#) 09:56, Feb. 15th 2008 (CET)
3. [Smalltown Boy](#)
4. [Symposiarch](#) 13:55, March 13th 2008 (CET)

Cannot come but would have liked to [edit source]

1. [Stefan](#) unfortunately I cannot make it
2. [Zwengelmann](#) 17:45, March 10th 2008 (CET) 6pm is unfortunately too late for me. I cannot even stay for an hour then.

As only people who arrive by car have signed up so far, I would like to know which time you would suggest, Zwengelmann. Maybe we can start earlier if none of the attendees minds. [kandschwar](#) 18:38, March 10th 2008 (CET)

Topics [edit source]

- Round of introduction

Results [edit source]

- With having all people show up that signed up, the first rhine hessian meet up was a total success. Even if Smalltown Boy needed to go a bit earlier and kloppo76 came a bit later. Besides this, it was a total success. Repetition is not impossible.
- Next meeting maybe in
 - Kirchheim-Bolanden or
 - Gau-Odernheim and hike to the [wild tulips](#) (date will be announced on short notice when the tulips bloom)

Fig. 1 Translation of exemplar organisational page, see <https://de.wikipedia.org/wiki/Wikipedia:Rheinhessen/Archiv>

Combining offline and online data and the challenge of usernames

The offline meetup data collected can be combined with online activity data available on Wikipedia. Structured data about the online activity are readily available: all online actions contributors undertake on Wikipedia are logged. Any changes made—whether it is the creation of a completely new article, the addition of a word, the deletion of a source, or the restructuring of a sentence—are registered in the revision history. The revision history of an article allows to trace the contributions and reverts of authors. It is possible to obtain information on all revisions of all pages which allows for an extremely detailed analysis of contributing behaviour. For example, it could be traced how users have co-authored specific articles, which sentences stemming from which users have been kept in the document for how long, etc. It is also possible to just extract the metadata which provides information on which users have changed what article and to what extent for computationally less expensive analyses.

To combine the *dewiki meetup dataset* with the rich body of online data, it needs to be merged on the basis of usernames. This comes with one major challenge: Wikipedians can request name changes, which are usually granted, or sign up under

a new name with a new account, which allows for greater anonymity.²⁰ Research focussing on online activity only is generally focussed on Wikipedia user accounts which have unique identifiers, but when working with meetup data, the interest lies in the people behind those accounts. As people changed their names or signed up with new accounts, it might well be that they have previously attended a meeting and signed up for it with their past name. The author aimed at consolidating these name changes.

All Wikipedia users and the redirection links linking to them were collected using a MediaWiki-supported API (application programming interface) call.²¹ In addition, all renames as logged in the renaming logbook²² were web-scraped with an automated browser using *RSelenium* [13]. This allowed the author to create a list of all current users and their redirections and previous usernames, in as far as they requested an official rename or linked to other accounts using redirection lists. In cases where users created a new account, potentially to gain more anonymity, it is impossible to link them to their previous name. In some rare cases concerning users that took part in meetings or other pages scraped during this project (such as in elections), previous usernames were explicitly mentioned and discussed. In these cases, those changes in usernames were also noted.

In this step of data preparation, substantial effort was spent on guaranteeing the matching of usernames from different sources. In the end, the author created a list of 1,751,808 different usernames (and variants of their encoding and spelling) belonging to 1,149,511 unique IDs (in the ideal case, this would be reflective of 1,149,511 different people). This information is saved within the *dewiki meetup dataset*, as file *nametoid*. This list can be merged with the usernames to replace usernames with IDs.

Collected data

The *dewiki meetup dataset* includes 4418 meetings. Table 1 provides a glimpse into the dataset and its most important variables. Full documentation of all variables is available with the dataset. The aim of the data collection was to collect the place/venue, date, type of meetup, attendees, apologies (if available), and minutes (if available) of (almost) all offline meetings organised on the German-language version of Wikipedia since the launch of Wikipedia until March 2020.

Meetings take place in a specific venue and place. The venue is recorded and the longitude and latitude of the place were retrieved using the OpenStreetMap API via *geo_osm()* of the *tidygeocoder* R package [5]. The date meetings took place is recorded as well.

Further, meetings are of different types. While a categorisation of the meeting types can depend on the research question at hand, the dataset contains a categorisation into

²⁰ See https://de.wikipedia.org/wiki/Wikipedia:Benutzernamen_ändern.

²¹ See <https://www.mediawiki.org/wiki/API:Allusers>.

²² See <https://de.wikipedia.org/wiki/Spezial:Logbuch/renameuser>.

Table 1 Glimpse of the *dewiki meetup dataset*

venue	meeting type	social vs work	organiser	dt	minutes	lat	long	attendees list	apologies list	sources
Restaurant Outback Lodge, Lagerplatz 4, 8400 Winterthur, Schweiz	Stammtisch	Social	Zürich	2020-02-22	Nach einem intensiven, arbeitsreichen Tag [...]	47.49545	8.718302	Horgner, Fundriver, [...]		https://de.wikipedia.org/wiki/Zürich
Restaurant St. Gallerhof, Konradstrasse 2, 8005 Zürich, Schweiz	Stammtisch	Social	Zürich	2020-01-23	Das frostige Winterwetter draussen übertrug [...]	47.38003	8.537334	Namrood, Sartta98, [...]		https://de.wikipedia.org/wiki/Zürich
Stapferhaus, Bahnhofstrasse 49, 5600 Lenzburg, Schweiz	Neujahrsapéro	Social	Zürich, Wikimédia CH	2020-01-12	Wie jedes Jahr lud Wikimédia CH [...]	47.39043	8.169974	Lantus, Bobo11, [...]		https://de.wikipedia.org/wiki/Zürich
Restaurant St. Gallerhof, Konradstrasse 2, 8005 Zürich, Schweiz	Stammtisch	Social	Zürich	2019-11-30	Obwohl oder vielleicht auch weil [...]	47.38003	8.537334	Gestumbli, M1712, [...]		https://de.wikipedia.org/wiki/Zürich
Bistrowagen ICE 109 Wuppertal Basel	Stammtisch	Social	Zürich	2019-10-06	Während wir den [...]	NA	NA	Fundriver, Gestumbli, [...]		https://de.wikipedia.org/wiki/Zürich
...										
...										
...										

work-oriented meetings and more socially oriented meetings. Meeting types such as informal meetups (*Stammtische*), parties and celebrations, yearly meetings, biking, walking or hiking tours, funerals, visits to the cinema or a festival, barbecues, outings for breakfast, or other spontaneous meetings can be considered social meetings. Meetings classified as work meetings cover all editathons or open editing events, field trips and guided tours (which tend to be to places which are of interest to the Wikipedia community), administrator conventions, workshops, photo tours, meetups of juries to judge on Wikipedian competitions, supporting Wikipedia booths at events and fairs, or other meetings directly oriented towards a Wikipedian initiative such as GLAM or KulTour, as well as meetings of authors collaborating within the so-called Wikipedia portals and task forces, and any office hours (of Wikimedia or sub-projects). In addition, meetups simply considered as “meeting” without further detail are also considered as work meetups as per default, meetings generally exhibit an underlying motivation of working on Wikipedia and improving it. When working with this dataset, it is important to reflect on the different types of meetings as well as their respective sizes, and to understand their implications for individual users and for the relationships that emerge between users. For example, if the focus of one’s research is on examining close face-to-face relationships, it might be sensible to exclude very large gatherings. This is in contrast to investigations centred around users which feel comfortable travelling to large, international meetups; however, with these meetings, the assumption that all attendees interact with every other one does not hold.

The attendees of meetings were collected from protocols that were published after the meeting took place, if available. If not available, attendance was recorded from the list of registration. Only users which were sure to attend were considered attendees. If available, this information was further double-checked with photo evidence in which users could be tagged, as well as protocols of later minutes if they referred to previous attendees. The usernames that editors signed up with were collected in the list of attendees. The list of apologies reflects the same for apologies sent. Users excusing their absence were thus collected in a separate column from the attendees; users, which initially wanted to join the meetup but did not attend, were added to this list of apologies. Further, the minutes recorded after the meeting were collected. They are generally in German and offer discussions and descriptions of who attended and what happened at meetings. Lastly, the source links to where the information was collected from are supplied.

Description of the *dewiki meetup dataset*

The *dewiki meetup dataset* includes 4418 meetings. Basic information about the meetups will be given in the following.

Temporal distribution

The collected data cover the time between 2001 and March 2020. The first meeting recorded took place on October 28th 2003 with five attendees in Munich, the

last ones on March 13th 2020 (2 days after the World Health Organisation declared COVID-19 a global pandemic) with three attendees in Cologne and with four attendees in Leipzig (a lot of Wikipedians interested in these meetings did send their apologies closer to the day due to the epidemiological situation).

77.0% of those 4418 meetups are classified as mainly social, while the other 23.0% are considered work meetings using the aforementioned categorisation. The distribution of meetups over time is pictured in Fig. 2. Please note that meetups happening in 2020 are not plotted to allow for better comparability across years as data collection did not collect any meetups after March 2020 (sparse meetups resumed in the summer of 2020). For 2020, a total of 67 meetups are in the dataset with 38 being social in nature (56.7%).

As seen in Fig. 2, the number of meetups increased steadily in the first years after the launch of the German-language Wikipedia until 2009. Numbers have remained on a relatively stable level since then, roughly counting around 300 meetups every year. The proportion of work meetups has been increasing over the years. While the meetups in 2020 are not plotted and data were only collected until March, the number of meetups is expected to have reached a new low for the year.

Spatial distribution

Meetups organised on the German-language version of Wikipedia take place primarily in German-speaking countries. The spatial global distribution of meetups is plotted in Figs. 3 and 4 is restricted to meetups in Germany, Switzerland, Austria, and Liechtenstein (German-speaking area, GSA).

The large majority of meetups, 88.8% (3922), took place in Germany, 5.5% (244) in Austria, 4.3% (188) in Switzerland and 0.023% (1) in Liechtenstein. Even though this captures around 99% of the meetups, the remaining per cent took place in 20 different countries around the globe: Australia (5), Belgium (2), Canada (1), China (1) Czech Republic (4), Finland (6), France (3), Hungary (1), Italy (5), Japan (8), Mexico (1), the Netherlands (2), Poland (10), Slovakia (1), Slovenia (1), South Africa (1), Majorca in Spain (1), Sweden (2), the United Kingdom (7), and Ukraine (1). Meetings can include the more global Wikimánias and WikiConventions, as long as a German-speaking group of people organised themselves on the German-language Wikipedia. Figure 4, restricted to the German-speaking area, highlights how meetings take place all over the area, but most often happen in urban centres and highly populated areas such as Vienna, Berlin, Dresden, Hamburg, Munich, and the North Rhine-Westphalia region. In addition, the map reveals that meetings are also rather frequent in cities which house dedicated community spaces (as these often grew from an active offline community).

Meetup network

Through offline meetups, a network develops: users attending the same meetings get to know each other and develop a tie. In a first step, this leads to an affiliation network of users belonging to meetups [see 43: Chapter 8]. This is a non-dyadic,

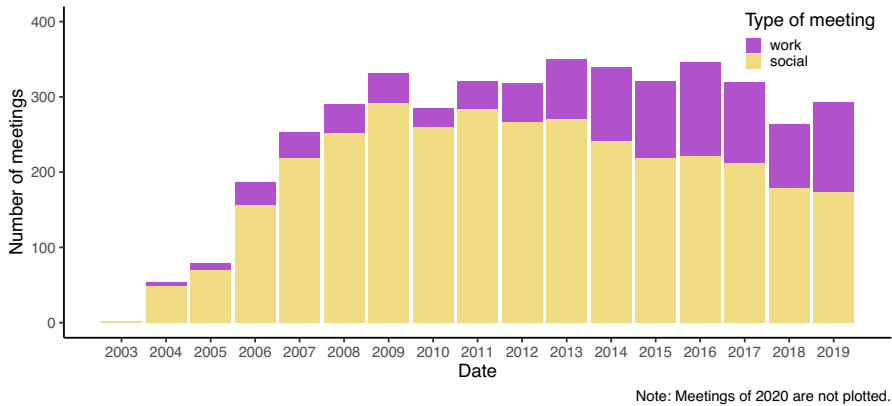


Fig. 2 Temporal distribution of meetups

two-mode network, also known as a membership network. Figure 5 shows the network at different points in time; the plot uses the same colour scheme as the ones in Figs. 3 and 4, thus capturing geographical clusters. The network started in October 2003 with the first meetup in Munich. It then consisted of one cluster with a meeting and its five attendees. By the end of the year, one other meetup took place, giving an additional user the chance to join the meetup scene. By the end of 2004, there were 258 nodes, consisting of 202 users who attended 56 different meetings and belonging to one large component. The colour scheme highlights how three of those meetings took place in Austria and that there was an active meetup scene in for example Bavaria and Berlin. In 2005, the network grew to include 547 nodes (413 users who attended 144 different meetings), belonging to one large and one small component. This means that, even though different meetups happen all across the German-speaking area, they are not just visited by local Wikipedians; the large, interconnected component suggests that at least some Wikipedians must take part in meetups in different places, thus connecting potential local components to a large one. Some geographical clustering is, however, observable, highlighting that users tend to attend multiple meetings from one region (as they might live close by). At the end of data collection in 2020, the meeting-to-user network consists of 8540 nodes (with 4122 users who have attended 4418 meetings). The network features five clusters; most nodes belong to one large component, while the four other components are small, single meetups with eight, five, and two (2x) attendees, respectively.

Using bipartite network projection, this meeting-to-user network can then be transformed into a user-to-user network, connecting those users that have met. In the meetup network as of 2020 (network connecting users with the meetups they attended), there are 8540 vertices sharing 38,314 edges (density of 0.0011). In the user network (network connecting users with other users who have attended the same meetup), there are 4122 vertices with 164,199 edges (density of 0.019). The global, unweighted transitivity (global clustering coefficient) of the user network is 0.41, giving the ratio of triangles and connected triples in the graph. On the level of nodes, the transitivity is on average 0.80 (median 0.92, standard deviation 0.24)

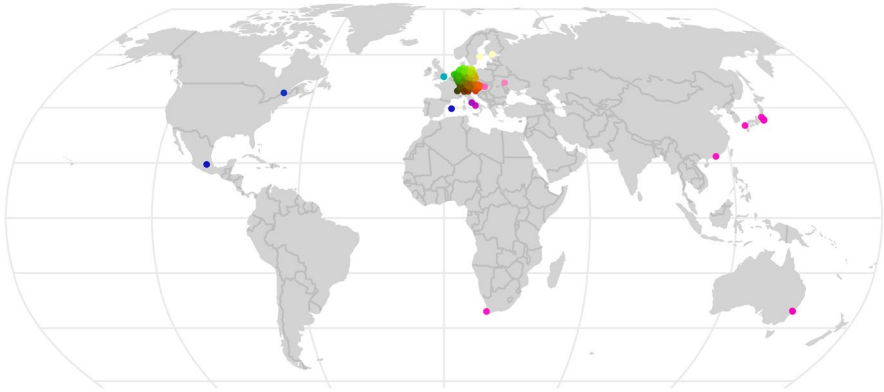
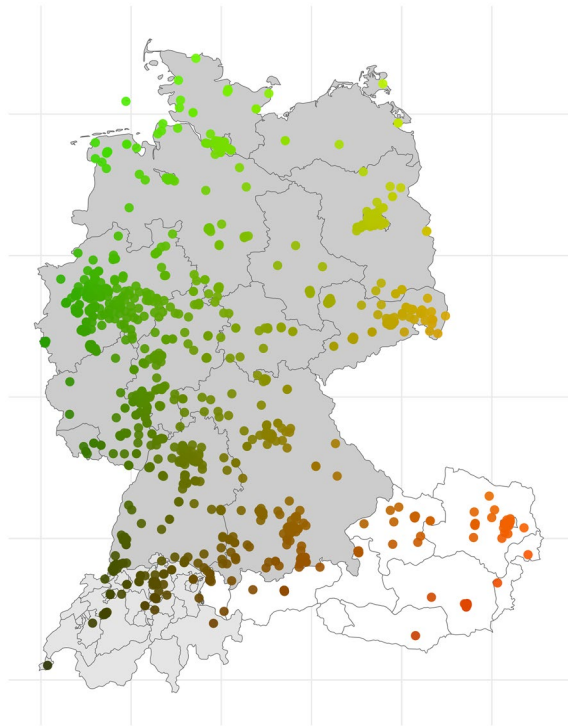


Fig. 3 Spatial distributions of meetups (world). The points are coloured according to their longitude and latitude with meetups being close to each other being similar in colour (normalised to the GSA region)

Fig. 4 Spatial distributions of meetups (GSA region). The points are coloured according to their longitude and latitude with meetups being close to each other being similar in colour (normalised to the GSA region)



with the theoretical minimum and maximum of 0 and 1, respectively, realised. The diameter of the user network, i.e. the longest path in the network, is 7, with an average distance of 2.76.

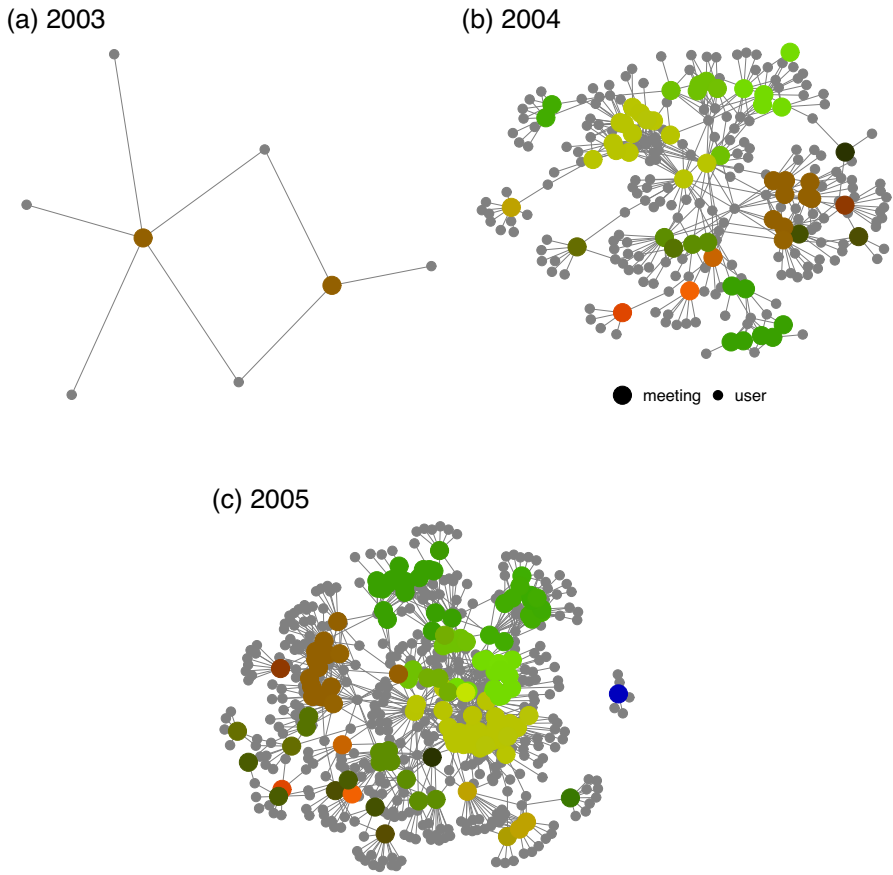
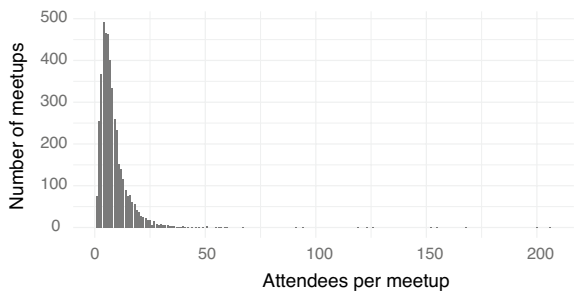


Fig. 5 Meeting-to-user network in **a** 2003, **b** 2004, and **c** 2005. The points are coloured according to their longitude and latitude with meetups being close to each other being similar in colour (normalised to the DACH region)

Fig. 6 Attendees per meetup



The average number of attendees per meetup is 8.71 (mean; median of 7, standard deviation 9.28) with a minimum of 1—meaning there were meetups where users were alone—and a maximum of 206—referring to very large, (supra-)national meetings; the distribution of attendees per meetup is displayed in Fig. 6.

Whilst the majority of Wikipedians never attended a meeting, those who did attended, on average, 9.30 (mean; median of 2, standard deviation 21.42) with a minimum of 1 and a maximum of 294 meetups. The degree of Wikipedians relates to the number of other users they have met through meetups. The average degree in the user network is 79.67 (mean; median is 24, standard deviation 136.74) with a minimum of 1 and a maximum of 1377. The strength of ties can further be measured by the number of times two users have met. The mean of the number of times users have met is 2.04 (median of 1, standard deviation 3.37), with a minimum of 1 and a maximum of 161.

These descriptions just form the start of more in-depth analyses using the *dewiki meetup dataset*. The 20 years it covers allow for longitudinal analysis across the years. In addition, the data can be linked with online activity as outlined above.

Novel research opportunities

The *dewiki meetup dataset* opens the door to numerous novel research opportunities in the (computational) social sciences. It provides data on a (nearly) complete social network: almost all edges between all users are recorded. The potential of the *dewiki meetup dataset* is unleashed when it is merged with data on the online behaviour. It can then be used to bridge the gap between offline and online actions. Online relationships and interactions have previously been shown to be relevant in explaining the online behaviour of Wikipedians—for example, users that are invited to an editor support forum—called the “Teahouse”—are more likely to remain active on Wikipedia [29], users are more likely to support a candidate in an election to become administrator when they have previously interacted on Wikipedia talk pages [23], and the density of a user’s collaboration network is important in explaining their norm-relevant behaviour [33]. Wikipedia is more than an encyclopaedia; it is an online community where interactions matter—the *dewiki meetup dataset* now allows to assess to what extent offline interactions matter as well.

In its first applications, the author has used the data to discuss the meetups’ (causal) effects on contributions to the collective good [for an early use of a smaller part of the meetup data on the same topic, see also 41: Chapter 15], to test Coleman’s mechanism of norm enforcement [6] using offline network density, and to explore how offline social ties matter in the context of explaining participation in governance activities [39]. It was found, for example, that attending offline meetups exhibits a positive statistical effect on the contribution behaviour of users. While it is not necessarily the case that users increase their contributions after a meetup in comparison to before the meetup, their reduction in contributions is less than the reduction similar control users, who do not attend meetups, experience. Further, when testing Coleman’s mechanism of norm enforcement, the author found that meetup attendees often behave significantly different than non-attendees; however,

the network density did not seem relevant in contrast to the proposition made by Coleman. Regarding elections, there are significant relationships between users' offline meeting behaviour and their online voting behaviour: for example, the larger the proportion of voters a candidate running for administrator has met, the more likely they are to win, and the higher the proportion of other voters a user has met, the more likely they are to vote themselves (this also holds true for the direction of votes: the more pro-voters a user knows, the more likely they are to also vote supportively, and the more anti-voters they have met, the less likely they are to vote supportively). Users are also more likely to vote if they have met the candidate, and they tend to support those more central in the meetup network. Given that elections on Wikipedia resemble public assembly elections, such findings can be extrapolated to hold significance within the broader realm of political science.

Future research can now build upon this and make use of the dataset collected to explore more complex and nuanced questions. Researchers can delve deeper into core concepts of the social sciences like network structures, information diffusion, representation and bias, or community norms, while also addressing questions that are more specific to the context of Wikipedia and other online communities. For example, the dataset can be used to study the structure and patterns of interactions among Wikipedia editors. The network of face-to-face interactions can be analysed to identify the most central and influential editors and the formation of cliques or subgroups. Researchers could also use the dataset to study the patterns of interactions and collaborations among editors before and after meetups, and analyse whether these meetings lead to the formation of strong, cohesive social networks or whether they result in more diffuse networks. As the data cover a time span of 20 years and dozens of regional communities, it can also be studied how these patterns change over time and depend on the local context.

Further, the offline connections can be understood as a form of social capital which offers additional resources. Against this background, it is also of interest to understand how offline meetings work as places of socialisation where community norms are potentially learnt: combining the *dewiki meetup dataset* with data about norm-relevant online behaviour, it can be analysed to what extent behaviour before and after meetups is norm-conforming and who is rewarding and punishing whom. Focussing on user pages and adopting a psychological perspective, the meetup data can also be used to understand how offline meetings shape the formation and expression of editors' identities and group membership, and how they influence the way Wikipedians interact with and perceive other editors. Further, the dataset can help to gain insight into social influence and persuasion; researchers can use the data to analyse how offline interactions influence the persuasion and attitude change of editors in specific contexts, like different discussions and voting situations (for example, on articles for deletion, winner of contests, or positions of power within Wikipedia). Within the context of media studies, it is also interesting to explore how Wikipedia as a product and as an important source of information is affected by meetings, for example regarding the inclusion of underrepresented topics and perspectives. The data can further be used by researchers in the field of cultural studies to conduct cultural and linguistic analysis by asking how these often very local meetings contribute to the representation of different regional customs.

To extend previous work, future research can aim to assess the impact of face-to-face interactions on article quality. While previous research focussed on the quantity of edits, researchers could analyse whether articles that were edited by Wikipedians who attended offline meetups have higher levels of accuracy, completeness, or readability than articles that were edited by others. More generally, it can be studied if and how a user's way of contributing towards Wikipedia changes or improves after attending a meetup. Future research could also focus on explaining meetup participation in the first place and explore how people who attend one or two meetups are different from people who attend twenty or fifty meetups. Furthermore, one could use the dataset to study whether face-to-face interactions have an impact on the location of Wikipedia articles. Researchers could analyse whether articles that were edited by editors who attended face-to-face meetings are more likely to be about places and locations that are close to the physical location of the meeting.

Overall, the *dewiki meetup dataset* presents a unique opportunity for researchers to study the dynamics of online collaboration and the role of social networks in the functioning of online communities. Wikipedia is one of the most impressive and sustainable online public goods. As it is inherently open and further explicitly lists the possibility that the setting up of a registered user account allows the linking of different edits and actions on Wikipedia to a single account/person and that anyone might investigate and analyse these data in any way,²³ it is a rewarding and accessible case for researchers from various disciplines.

Discussion and concluding remarks

Data access and future data improvement

The *dewiki meetup dataset* can be accessed via the Open Science Framework: <https://doi.org/10.17605/OSF.IO/EHA4R> [40]. It is shared via a Creative Commons Attribution-ShareAlike 4.0 International license (CC BY-SA 4.0), allowing others to both share and adapt the dataset in any format and for any purpose. Appropriate credit must be given and any adaptations must be shared alike. The strengths of the dataset also come with some privacy concerns that must be kept in mind. For example, geographical information is recoverable from the dataset and can be linked to users. It is important to use the data fairly and reasonably. The *dewiki meetup dataset* covers almost all meetings organised on the German-language Wikipedia up to the outbreak of the COVID-19 pandemic (except those outlined in the section on excluded meetings). The data are shared in a way that allows for updating and expanding the dataset. For one, the data collection can be continued on to the future. For example, the major changes related to the COVID-19 outbreak offer possibilities for (natural) experimental research designs including and beyond offline meetups. Further, it would be of utmost interest to expand such analyses to other language versions of Wikipedia. This is not only interesting to the Wikimedia Foundation to

²³ See https://de.wikipedia.org/wiki/Hilfe:Benutzerkonto_anlegen.

better organise and evaluate their projects but allows the comparisons of different cultures with different user bases. In addition, more cumulative effort can be spent on consolidating name changes (*namestoid*).

Limitations

The dataset presented does not and cannot include all offline interactions which have occurred between Wikipedians. While the data collection process aimed to comprehensively capture all offline meetings, it is important to recognise that certain types of meetups were deliberately excluded from the dataset and the collection process (such as meetings in community spaces). Further, Wikipedia editors may meet each other in person through various means not documented within the online platform. These exclusions can influence the overall representativeness of the dataset, as for example the exclusion of frequent community meetings can lead to an underrepresentation of certain interaction patterns.

The documented interactions represent a distinct category of edges: this dataset covers the majority of Wikipedia meetings which are publicly organised on the German-language platform and the users which have signed up. It is important to acknowledge that the edges recorded in this dataset thus provide insights into a particular facet of offline interactions among Wikipedians and not the entirety of these interactions. Additional data collection (which goes beyond what is documented publicly on Wikipedia) would be needed to, for example, study the role of meetings in community spaces.

Conclusion

Wikipedia is an online free-content encyclopaedia and one of the largest and most successful examples of online peer-production. Wikipedia celebrated its twentieth birthday in 2021 and has developed into a key figure in the internet landscape. Known by most internet users, it provides the backbone of many information technologies, gives the answers to Alexia, Google, and Siri, and is also a phenomenon that has attracted considerable attention from researchers. Not only does Wikipedia provide a rich and valuable source of data, but it is also a peculiar community to research in itself—a fact often unknown to the end-user. The dataset introduced in this data brief focuses on one of the most unknown facts of the Wikipedia community: Wikipedians meet offline, often, regularly, and across the globe. In the typical spirit of Wikipedia, these meetings are organised publicly and are well documented with lists of attendees, minutes, and photo evidence. The dataset covers 20 years of offline activity which can be combined with online activity logged on Wikipedia—this offers valuable and new avenues for computational social scientists to bridge the gap between the offline and the online and to gain a deeper understanding of the Wikipedia community.

Acknowledgements The author wants to thank Ulf Liebe, Andreas Murr, Richard Lampard and Wojtek Przepiorka, as well as the attendees of the WikiWorkshop 2023 for helpful feedback and guidance.

Funding The author gratefully acknowledges support from the Wikimedia Foundation.

Data availability The dataset presented in this study is available in the Open Science Framework repository: <https://doi.org/10.17605/OSF.IO/EHA4R>.

Declarations

Conflict of interest Part of this work was supported by the Wikimedia Foundation (project grant https://meta.wikimedia.org/wiki/Grants:Project/nschwiter/The_Role_of_Offline_Ties_of_Wikipedians). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Arazy, O., Nov, O., Patterson, R., & Yeo, L. (2011). Information quality in Wikipedia: The effects of group composition and task conflict. *Journal of Management Information Systems*, 27(4), 71–98. <https://doi.org/10.2753/mis0742-1222270403>
2. Brandes, U., Kenis, P., Lerner, J., & Van Raaij, D. (2009). Network analysis of collaboration structure in Wikipedia. In *Proceedings of the 18th International Conference on World Wide Web* (pp. 731–740). New York: ACM Press.
3. Burke, M. & Kraut, R. (2008). Mopping up: Modeling Wikipedia promotion decisions. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (pp. 27–36). New York: ACM Press.
4. Callahan, E. S., & Herring, S. C. (2011). Cultural bias in Wikipedia content on famous persons. *Journal of the American Society for Information Science and Technology*, 62(10), 1899–1915. <https://doi.org/10.1002/asi.21577>
5. Cambon, J., Hernangómez, D., Belanger, C., & Possenriede, D. (2021). tidygeocoder: An R package for geocoding. *Journal of Open Source Software*, 6(65), 3544. <https://doi.org/10.21105/joss.03544>
6. Coleman, J. S. (1990). *Foundations of Social Theory*. Harvard University Press.
7. Edelman, A., Wolff, T., Montagne, D., & Bail, C. A. (2020). Computational social science and sociology. *Annual Review of Sociology*, 46(1), 61–81. <https://doi.org/10.1146/annurev-soc-121919-054621>
8. Eskander, R., Muresan, S., & Collins, M. (2020). Unsupervised cross-lingual part-of-speech tagging for truly low-resource scenarios. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (pp. 4820–4831). Online: Association for Computational Linguistics.
9. Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science Advances*. <https://doi.org/10.1126/sciadv.aq1360>
10. Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070), 900–901. <https://doi.org/10.1038/438900a>
11. Halfaker, A., Kittur, A., & Riedl, J. (2011). Don't bite the newbies. In *Proceedings of the 7th international symposium on wikis and open collaboration*. New York: ACM Press. <https://doi.org/10.1145/2038558.2038585>
12. Hara, N., Shachaf, P., & Hew, K. F. (2010). Cross-cultural analysis of the Wikipedia community. *Journal of the American Society for Information Science and Technology*, 61(10), 2097–2108. <https://doi.org/10.1002/asi.21373>

13. Harrison, J. (2020). *Package ‘RSelenium’*. R Package.
14. Hu, X., Zhang, X., Lu, C., Park, E. K., & Zhou, X. (2009). Exploiting Wikipedia as external knowledge for document clustering. In *Proceedings of the 15th international conference on knowledge discovery and data mining*. New York: ACM Press. <https://doi.org/10.1145/1557019.1557066>
15. Jankowski-Lorek, M., Ostrowski, L., Turek, P., & Wierzbicki, A. (2013). Modeling Wikipedia admin elections using multidimensional behavioral social networks. *Social Network Analysis and Mining*, 3, 787–801.
16. Keuschnigg, M., Lovsjö, N., & Hedström, P. (2017). Analytical sociology and computational social science. *Journal of Computational Social Science*, 1(1), 3–14. <https://doi.org/10.1007/s42001-017-0006-5>
17. Kim, S., Toutanova, K., & Yu, H. (2012). Multilingual named entity recognition using parallel data and metadata from Wikipedia. In *Proceedings of the 50th annual meeting of the association for computational linguistics* (pp. 694–702). Online: Association for Computational Linguistics.
18. de La Robertie, B., Pitarch, Y., & Teste, O. (2015). Measuring article quality in Wikipedia using the collaboration network. In *Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining*. New York: ACM Press. <https://doi.org/10.1145/2808797.2808895>
19. Laniado, D., Tasso, R., Volkovich, Y., & Kaltenbrunner, A. (2011). When the Wikipedians talk: Network and tree structure of Wikipedia discussion pages. In *Fifth international AAAI conference on e-blogs and social media*. Washington: AAAI.
20. Lavsa, S. M., Corman, S. L., Culley, C. M., & Pummer, T. L. (2011). Reliability of Wikipedia as a medication information source for pharmacy students. *Currents in Pharmacy Teaching and Learning*, 3(2), 154–158. <https://doi.org/10.1016/j.cptl.2011.01.007>
21. Lazer, D. M., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Alstynne, M. V. (2009). Computational social science. *Science*, 323(5915), 721–723. <https://doi.org/10.1126/science.1167742>
22. Lazer, D. M., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science*, 369(6507), 1060–1062. <https://doi.org/10.1126/science.aaz8170>
23. Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Signed networks in social media. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1361–1370). New York: ACM Press.
24. Liu, J., & Ram, S. (2018). Using big data and network analysis to understand Wikipedia article quality. *Data & Knowledge Engineering*, 115, 80–93. <https://doi.org/10.1016/j.datak.2018.02.004>
25. Luyt, B., Kwek, W. T., Sim, J. W., & York, P. (2007). Evaluating the comprehensiveness of Wikipedia: The case of biochemistry. In *International conference on Asian digital libraries: Looking back 10 years and forging new frontiers* (pp. 512–513). Springer.
26. Merz, M. (2019). *Die Wikipedia-community*. Springer.
27. Mestyan, M., Yasseri, T., & Kertész, J. (2013). Early prediction of movie box office success based on Wikipedia activity big data. *PLoS One*, 8(8), e71226. <https://doi.org/10.1371/journal.pone.0071226>
28. Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T. (2013). Quantifying Wikipedia usage patterns before stock market moves. *Science and Reports*, 3(1), 1801.
29. Morgan, J. T. & Halfaker, A. (2018). Evaluating the impact of the Wikipedia teahouse on newcomer socialization and retention. In *Proceedings of the 14th international symposium on open collaboration* (pp. 1–7). New York: ACM Press.
30. Ortega, F., Gonzalez-Barahona, J. M., & Robles, G. (2008). On the inequality of contributions to Wikipedia. In *Proceedings of the 41st annual Hawaii international conference on system sciences*. New York: IEEE. <https://doi.org/10.1109/hicss.2008.333>
31. Ortega, F., Gonzalez-Barahona, J. M., & Robles, G. (2008). Quantitative analysis of the top ten Wikipedias. In *International conference on software and data technologies: Communications in computer and information science* (pp. 257–268). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-88655-6_19

32. Panciera, K., Halfaker, A., & Terveen, L. (2009). Wikipedians are born, not made. In *Proceedings of the ACM 2009 international conference on supporting group work*. New York: ACM Press. <https://doi.org/10.1145/1531674.1531682>
33. Piskorski, M. J., & Gorbatâi, A. (2017). Testing Coleman's social-norm enforcement mechanism: Evidence from Wikipedia. *American Journal of Sociology*, 122(4), 1183–1222. <https://doi.org/10.1086/689816>
34. Rajagopalan, M. S., Khanna, V. K., Leiter, Y., Stott, M., Showalter, T. N., Dicker, A. P., & Lawrence, Y. R. (2011). Patient-oriented cancer information on the internet: A comparison of Wikipedia and a professionally maintained database. *Journal of Oncology Practice*, 7(5), 319–323. <https://doi.org/10.1200/jop.2010.000209>
35. Richter, P. (2020). *Die Wikipedia-story*. Campus Verlag.
36. van de Rijt, A., Kang, S. M., Restivo, M., & Patil, A. (2014). Field experiments of success-breeds-success dynamics. *Proceedings of the National Academy of Sciences*, 111(19), 6934–6939. <https://doi.org/10.1073/pnas.1316836111>
37. Rosenzweig, R. (2006). Can history be open source? Wikipedia and the future of the past. *Journal of American History*, 93(1), 117–146. <https://doi.org/10.2307/4486062>
38. Salganik, M. J. (2018). *Bit by bit*. Princeton University Press.
39. Schwitter, N. (2022). *The Role of Offline Ties in Online Communities: The Case of Wikipedia*. Ph.D. Thesis. <http://wrap.warwick.ac.uk/175384>. Accessed 23 Sep 2023.
40. Schwitter, N. (2023). *Dataset: Offline meetups on the German-language Wikipedia*. <https://doi.org/10.17605/OSF.IO/EHA4R>. Accessed 23 Sep 2023.
41. Stegbauer, C. (2009). *Wikipedia*. Springer.
42. Stein, K. & Hess, C. (2007). Does it matter who contributes. In *Proceedings of the 18th conference on hypertext and hypermedia*. <https://doi.org/10.1145/1286240.1286290>
43. Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications (structural analysis in the social sciences)*. Cambridge University Press.
44. Welser, H. T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., & Smith, M. (2011). Finding social roles in Wikipedia. In *Proceedings of the 2011 iConference*. New York: ACM Press. <https://doi.org/10.1145/1940761.1940778>
45. Wikimedia Foundation (2019). *Movement Organizers Study*. Tech. rep. Wikimedia Foundation. https://meta.wikimedia.org/wiki/File:Wikimedia_Movement_Organizers_Study.pdf. Accessed 27 Jan 2023
46. Wilkinson, D. M. & Huberman, B. A. (2007). Cooperation and quality in Wikipedia. In *Proceedings of the 2007 International Symposium on Wikis*. New York: ACM Press. <https://doi.org/10.1145/1296951.1296968>
47. Yasseri, T., & Bright, J. (2016). Wikipedia traffic data and electoral prediction: Towards theoretically informed models. *EPI Data Science*. <https://doi.org/10.1140/epjds/s13688-016-0083-3>
48. Zhang, X., & Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review*, 101(4), 1601–1615. <https://doi.org/10.1257/aer.101.4.1601>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.