



How offline meetings affect online activities: the case of Wikipedia

Nicole Schwitter^{1,2*}

*Correspondence:

nicole.schwitter@uni-mannheim.de

¹MZES, University of Mannheim,
68131, Mannheim, Germany

²Department of Sociology,
University of Warwick, Coventry,
CV4 7AL, United Kingdom

Abstract

As open-source communities and other peer production projects face the challenge of sustaining themselves over time, the role of offline gatherings in fostering community resilience becomes a vital question. This study centres on the potential of offline social interactions as a means to mitigate declines within such communities, using Wikipedia as a case example. Using a comprehensive dataset spanning informal meetups within the German-language Wikipedia community from 2001 to 2020 and combining it with large scale online activity data, this study investigates the relationship between offline gatherings and online contribution behaviour. Results show that attending meetups has a positive effect on contributing towards Wikipedia both in the short and long term. By shedding light on the significance of offline social interactions in bolstering online communities, this study offers valuable insights into potential strategies for sustaining collaborative online environments amidst broader declines and further shows how offline and online activities intertwine.

Keywords: Online community; Wikipedia; Peer production; Offline meetings; Community resilience; Online and offline integration; Difference in differences; Matching; Web data; Editathons

1 Introduction

Peer production is one of the most significant innovations emerging from Internet-mediated social practice [1]. With remarkable sustainability and effort, unpaid volunteers have collaborated in a decentralised and participatory fashion to create software (like the GNU/Linux operating system), maps (like OpenStreetMap), or encyclopaedias (like Wikipedia). Successful collaborative online communities require an active, contributing user base. However, online peer production communities often rely on a small set of core contributors who have contributed large parts [see e.g. for Wikipedia 2] and retaining users over time, particularly beyond an initial trial phase, poses to be a challenge for many open-source projects and virtual communities in general [see e.g. 3–7].

In this article, I will focus on the encyclopaedia Wikipedia as an example of an active peer production project and an entity of sociological interest [see also 8]. Wikipedia was launched in 2001 and has since become one of the top destinations for information to many people, it lies at the backbone of many technologies and has become a key figure in the internet landscape in general. Its scope and sustainability are remarkable: as of 2024,

© The Author(s) 2024. **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/>.

Wikipedia is available in over 300 different languages and the English-language version of Wikipedia alone features over 6 million articles of, overall, good quality [9]. All of this is based on the crowd-knowledge of engaged volunteers in a collaborative effort to organise and present human knowledge, giving their time without receiving anything in return. While many people contribute to Wikipedia, the project also struggles with a decline in activity and with the retention of new users [see e.g. 10]. The level of activity and the size of the active user base also varies greatly between the different language versions of Wikipedia [2, 11].

While Wikipedia is primarily an online community, it also features a notable offline component: particularly the German-language Wikipedia—one of the largest and most active language versions—is characterised by regular local offline meetups which give editors the chance to get to know each other personally [12–14]. Such offline meetups are important to the community: long-term member and active Wikipedian Richter [15, 148] states that personal acquaintances are central in a project that is based on anonymous contributions. 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 [15, 132–136]. Offline meetings thus offer new avenues of interaction as well as ways for editors to immerse themselves into the project.

Several studies have acknowledged the occurrence of offline interactions within a variety of online communities and discussed the interplay between the offline and the online [e.g. 16–25]. Offline meetings can play a part in complementing the low social presence inherent in most computer-mediated environments [26]. There is supporting evidence that stronger ties develop through offline gatherings in online communities; for example, Angelopoulos and Merali [16] find enhanced sociability of users of an online community after meeting offline; Lin [19] finds that offline interactions are important for the sustained success of online communities; Koh et al. [22] state that offline activities increase the solidarity and cohesiveness of virtual communities and strengthen links between members: they expect that offline meetings facilitate virtual community activism and lead to a higher sense of virtual community, expecting positive relationships between offline activities and different dimensions of immersion into a virtual community. They find some support in their analysis on Korean virtual communities. However, not all studies on online communities come to the conclusion that offline meetups are positive: McCully et al. [23] state that meetups of a collaborative writing community strengthen online relationships but decrease the amount of participation, leading to a counter-intuitive impact on community sustainability.

Taken together, how online communities are affected by offline interactions seems to depend on the context and the specifics of the community. Wikipedia as an open collaboration project which aims at providing free knowledge features notable differences to most other communities: Wikipedians join the community to create an encyclopedia and not primarily to socialise. Against this background, this study will focus on the effect of meetup participation on editing activity to shed light on the question of whether offline interactions can help to stop the decline of peer production projects, taking the German-language version of Wikipedia as a case example.

2 Theory and related work

Understanding who Wikipedians are, why they join the platform, and why they keep editing has been of interest to many scholars [see for a recent review 27]. Considering that

Wikipedia is a prime example of a collective good, the incentives to contribute are low. Even though the provision of articles is costly for the authors as they need to invest time in researching and writing, Wikipedia exists and its scope is remarkable. Many individuals have opted against free-riding and instead participate, sometimes staying active for years to come after joining the community.

Crowston and Fagnot [28] have established a theoretical framework drawing on different theories to explain contributions towards user-generated content at different stages, particularly incorporating theories of helping behaviour [29] and social movement theory [30]. They differentiate between three separate sets (initial, sustained, and meta) of motivations for participation, arguing explicitly that the motivations to make a first contribution are not the same as the motivations to make additional contributions. This study focuses on the (motivational) effect of meetups on sustained contribution of Wikipedians.

Crowston and Fagnot [28] understand contributions to projects such as Wikipedia as a form of voluntary participation in a voluntary organisation; Konieczny [31] has also previously framed Wikipedia as a social movement. Crowston and Fagnot [28] argue that three precursor conditions must be fulfilled so that helping behaviour can arise in the context of user-generated content projects: individuals must recognise the need to help, individuals must have the capabilities and feel the obligation and/or responsibility to help, and individuals must weigh their obligation and capability of helping favourably against its (social and tangible) costs.

In this study, I focus on sustained contribution; given this focus, it can be assumed that contributors of Wikipedia are aware of the project from their initial encounter. However, Crowston and Fagnot [28] argue that a continuing contributor must also perceive the need for further contributions. They also expect that sustained contributors who report higher domain expertise will contribute more, as well as those who viewed contributing positively (whether due to learning more, finding it fun, or receiving positive feedback). More relevant for this study, Crowston and Fagnot [28] expect that feelings of social obligations play a role in deciding whether users become sustained contributors [32]. Drawing on the literature on social movements, defined as an organised effort by a group of people to bring about societal change, they suggest that projects become better at retaining participants if they develop characteristics of such movements. Klandermans [30] suggests four different areas of motivation for participation in a social movement: collective motives, identification with the group or a subgroup, reward motives, and social motives [see also 28, 33]. Further, the fact that social networks matter for social movements is well documented and they are shown to be important in explaining their success [34, 35].

Collective motivations come from the individual's evaluation of the group's goals or ideology. This becomes relevant as many social movements share an ideology and are organised around this shared ideology. In the case of Wikipedia, the project aims at promoting free knowledge and many contributors express feelings of agreement with this goal [32, 36–40]. Crowston and Fagnot [28] further argue in their theory that participants who report a higher level of identification are more likely to sustain contributions and contribute more [see also 41]. Group identification can lead to feelings of obligation to the group which then provide motivations for sustained contribution. Feeling part of a group is essential to transform interests into collective actions [42] and generally increase commitment to it [43]; commitment and self-identification can then motivate to contribute to the collective good of the group [44] and work as motivating drivers of work more gen-

erally [45, 46], also in online communities [47]. In the context of Wikipedia, research has shown that experienced editors tend to see Wikipedia as a community of editors instead of just a collection of articles, and are more likely to report social and community factors as important motivations for editing compared to newcomers [36, 39, 48].

Following this line of argumentation, I argue that offline meetings increase commitment to and identification with the project and its user base. Face-to-face meetings allow new friendships to develop and discussions about Wikipedia to take place. They offer an additional venue for interaction and thus strengthen existing online weak ties, increasing the identification and commitment to Wikipedia, which in turn increases contributions. This leads to the following hypothesis:

Hypothesis 1: Attendees of offline meetups increase their contributions after a meetup.

Wikipedia meetups come as either social meetups or are more strongly work-related. Some meetups are specifically designed so people can learn more about how to edit Wikipedia, about gaps and missing content on Wikipedia, or about specific information which is currently lacking in the encyclopaedia. New ideas and knowledge are generated at such work meetups which—following the framework of Crowston and Fagnot [28]—increase users' capabilities to contribute as well as their awareness of the need to contribute. I thus also expect the following:

Hypothesis 2: Attendees of work-related offline meetups increase their contributions more than attendees of social meetups after attending.

In the context of Wikipedia, a number of studies have shown the benefit of positive social feedback such as symbolic recognitions in the form of barnstars [49, 50] or careful and polite newcomer socialisation [51, 52]. Only few previous studies have, however, explored the role of face-to-face meetings and their effect on users' editing behaviour. Stegbauer [13] collected data on 240 different meetups with 750 different attendees. Contrary to his expectations, he finds that around 60 per cent of users, who have taken part in meetings, decreased their activity on Wikipedia, measured in the number of edits in the month after the meetup compared to activity in the month before (calculating bivariate associations). More detailed analysis shows that users who later become administrators increase their activity while only those who do not become administrators decrease it. Stegbauer [13, Chap. 15] also mentions as anecdotal evidence that users report that meetings are integral in deciding on new administrators and in suggesting who should be nominated. The findings of Stegbauer [13] thus seem mixed: meetups can either lead to a withdrawal from the community, but they can also fuel further engagement, culminating in the promotion to an admin. Analysing interview data from an editathon—a typical work-related meetup, Littlejohn et al. [53] highlight that the personal relationships made at these events are important to some of the participants when continuing to edit Wikipedia, and Farzan et al. [54] find that Wikipedians joining the site as part of an editathon stay slightly more active (in the short time frame observed). Several studies have focused on the Wiki Education Project, which does not promote offline meetups of existing editors, but is designed as a

online program where college instructors assign students in their classes to edit Wikipedia articles collaboratively in an offline environment. This socialisation program which includes tutorials, trainings, constructive guidance from Wikipedia staff, and cohort support, leads to generally positive outcomes: students in the Wiki Education Project make more edits, perform at higher quality, and are more likely to continue editing Wikipedia even a year after the program compared to editors which joined Wikipedia the conventional way [55]. Students in college classes which promoted a collaborative work environment had even greater gains than those who promoted independent work.

In the following, I will focus on informal meetups within the German-language Wikipedia and test how meetup participation affects sustained contributions in the form of quantity of edits on Wikipedia.

3 Methods and data

This study aims to identify a (causal) effect of meetup attendance on productivity. This section will describe the data used, particularly focusing on meetups and productivity, and describe the analytical, quasi-experimental approach.

3.1 Online activity

All online actions contributors undertake on Wikipedia are logged. The activity of Wikipedians is measured via edits recorded in the meta data dump¹ using the stub history meta files. These retain information on which user edits which article to what extent and at what point in time without including the actual text data.

To measure activity, I counted all edits undertaken by users who are registered on Wikipedia. To assess the effect of meetups on editing behaviour, three different time frames are being analysed: a very short time frame of one week (7 days), a medium length time frame of one month (28 days), as well as a long term time frame of one year (364 days). In practice, this means that all edits in the week up to the meeting have been counted, all edits in the month up to the meeting, and all edits in the year up to the meeting, as well as the number of edits in the corresponding time frame after the meetup. Activity on the day of the meetup was excluded. These time frames are chosen as the weekly and monthly changes in editing behaviour have been used in previous research [13, 54]. They represent a short and a medium time frame. The yearly time frame allows for a long-term perspective which previous research has usually not considered.

As Wikipedia is separated into different namespaces with different functions, two measures of activity are differentiated in the following: total activity across all namespaces and mainspace activity. The latter refers to productive edits made only on articles, while the former also includes activity on discussion pages, meta pages and others.

As this study is concerned with sustained contributions and not with newcomers, users who have never made an edit before taking part in a meetup will be excluded from the following analyses (i.e. users must have made at least one edit in any namespace before taking part in the meeting to be included in the analyses). As the range of activity levels on Wikipedia is wide and the distribution of the number of edits is thus very skewed [see for research on this e.g. 2, 11], I run a robustness check where I excluded the most active 2 per cent of users.

¹Data dumps provided by the Wikimedia Foundation offer well-structured data exports of different facets of Wikipedia. See for the German-language version of Wikipedia <https://dumps.wikimedia.org/dewiki>.

3.2 Offline activity: meetings and meetup attendees

This article makes use of the offline meetup data described in Schwitter [12]. Most offline gatherings between Wikipedians are organised on Wikipedia itself on dedicated organisational pages. The dataset collected by Schwitter [12] covers (almost) all offline meetings between 2001 and 2020; excluded are only those meetings which took place in community hubs and where the same small group of users repeatedly attended very regular meetings (and which lack a rigorously maintained list of attendees which is typical for other meetings). For all recorded meetings, the dataset contains information on which user signed up to attend which meeting at what point in time. Overall, the dataset contains information on 4418 meetups organised within the German-language version of Wikipedia. For the following analyses, I exclude ten very large meetings (the Wikimánias and WikiConventions). The list of attendees of these very large meetings tend to be much more unreliable as the option of registering is often also used to just signal interest.

The 4408 meetings in my data thus contain all smaller-scale gatherings organised within the German-language version of Wikipedia. The first meeting recorded took place on October 28th 2003 with five attendees in Munich, the last ones took place on March 13th 2020—just before social distancing restrictions due to the outbreak of the Coronavirus pandemic were introduced—with three attendees in Cologne and with four attendees in Leipzig. This data includes both, events with a social orientation, where people primarily meet to get to know each other, as well as more work-oriented meetings, where Wikipedians get together to improve and directly work on Wikipedia. Thus, social meetups are meetups that have an inherently social component, such as the classic informal meetup (*Stammtisch*), parties and celebrations, yearly meetings, hiking trips, barbecues and similar. Work meetups are generally organised with the intention to improve Wikipedia (such as editathons, open editing events, workshops, photo tours, etc.). In the data, 77% of the meetups are classified as mainly social while the other 23% are considered work meetings (however, this distinction is not clear-cut and there is a subjective component).

Who are the editors who took part in these meetings? 4013 users have joined at least one offline meetup and 2167 users partook in more than one Wikipedia meeting. If a user took part in meetings, they on average joined 9.2 meetings in total (mean; median of 2, standard deviation 21.1) with a minimum of 1 and a maximum of 289 meetups.

Before ever taking part in a meetup, users had been active on Wikipedia on average for 921.24 days (days since their first edit; median 489.37, standard deviation 1125.16, minimum -3824.06, maximum 5968.19). This means some users partook in a meeting about 10.5 years before they made their first edit in the German-language version of Wikipedia, while others had already been on Wikipedia for over 16 years before meeting other Wikipedians face-to-face. Users who have made their first edit after attending a meeting will be excluded from this study as this would explain initial contribution instead of sustained contribution (out of 4013 users, 3724 users have made an edit before attending their first meeting). In total, there are 36364 observations of users partaking in meetings (excluding observations where a user joined a meeting before their first edit). Users might have partaken in meetings before making their first edit for a number of different reasons: users might be active in other language versions or sister projects of Wikipedia and have only signed up to the German-language Wikipedia to read more about a meetup; non-registered people might accompany an already active user (as friend, spouse, etc.) but then also decide to sign up; (joke) accounts which do not belong to humans (but instead

to e.g. dogs); users partook in meetings advertised elsewhere (e.g. editathons organised via a library or other institutions) and then decided to sign-up; or it might also refer to an error in the data.

3.3 Control group and control variables

This study is concerned with the (causal) effect of attending offline meetings on the contributing behaviour of single actors. To allow the identification of a *treatment effect*, a control group is created to follow a quasi-experimental approach. However, it needs to be noted that reverse causality and possible self-selection biases can still threaten causal inferences.

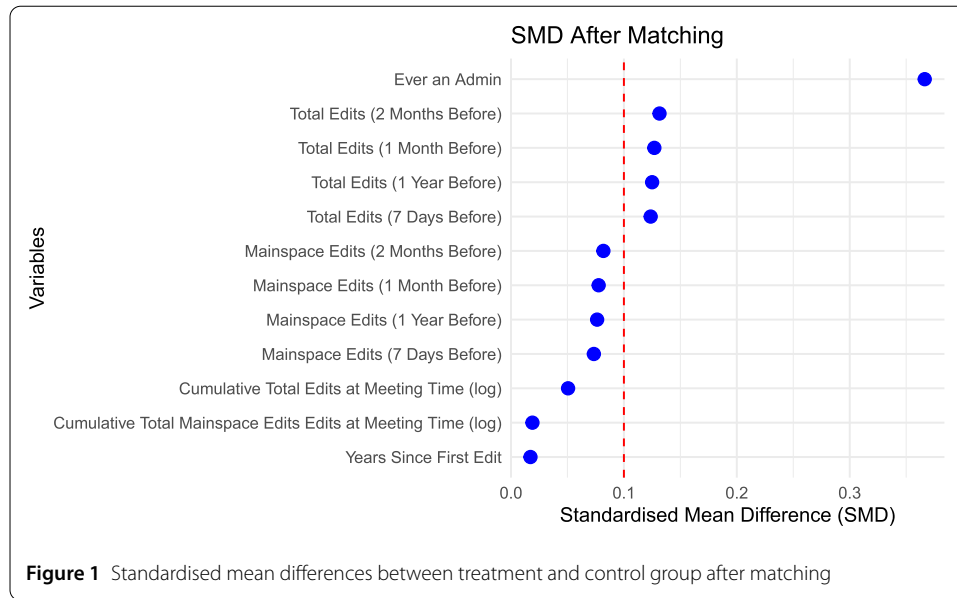
To identify effects, meetup attendees are compared with comparable control users. Comparing the meetup attendees with the full population of Wikipedia users is computationally too expensive as there are over four million registered users. The goal of the matching approach was thus to subset the data in such a way that differences in pre-meeting variables between treatment (meetup attendees) and selected control group (non-attendees) are minimised. The control group of comparable users is constructed by matching meetup attendees with the most similar others in the population of German-language Wikipedia users using covariate matching [56].² The population of potential matchable non-attendees only includes users who did not take part in any meetups (and were never recorded to do so in the data). The matchable non-attendee was found by comparing users based on the following five features:

1. days since registration,
2. sum of activity (number of edits) in the article mainspace of Wikipedia since registration up until the day of the meetup,
3. sum of activity (number of edits) anywhere but the article mainspace since registration up until the day of the meetup,
4. recent activity in the article mainspace of Wikipedia before the day of the meetup (number of edits in the last seven days, the last month, the last two months, and last year), and
5. recent activity anywhere but the article mainspace before the day of the meetup (number of edits in the last seven days, the last month, the last two months, and last year).

Each of these five features was assigned an equal weight of 20 per cent (for the recent activity, each of the four time periods was weighted 5 per cent). The most similar other user was identified and selected as a control non-attendee. Users were compared using a distance measure based on ordinary least squares between Wikipedian X who attended a meetup and all those who have never attended a meetup and are not already a matched non-attendee for another user at that specific meetup. In case multiple users have an identical minimal distance to Wikipedian X , one of them was chosen randomly. A meetup attendee can have different matches for different meetups.

The effectiveness of the matching process was evaluated using standardised mean differences (SMD) across covariates, where values below 0.1 generally indicate negligible differences between groups and suggest adequate balance. Figure 1 presents the SMD for each

²In contrast to the popular propensity score matching [see e.g. 57], covariate matching is computationally less expensive: in the case of Wikipedia, there are many registered users who need to be matched to the meeting attendees at over 4000 different points in time. This leads to a very large pool of observations for which a propensity score would need to be calculated.



covariate after matching (see also Table S1 in the appendix which provides descriptive information on all independent and dependent variables included in the models for the treatment and the control group separately). Most values fall below the 0.1 threshold after matching, indicating that the matching procedure was relatively successful in balancing the covariates between the treatment and control groups. However, there is a very notable difference in the proportion of administrators per group (29 per cent of users in the treatment group had been administrators and 14 per cent in the control group).

Given that the matching was not done on exact matches and some differences remain, further control variables are included to control for the lack of balance. I control for the previous total level of activity up to the time of the meetup and tenure which is measured as years passed since a user's very first edit. I also control for the year of the meetup as fixed effects. I have also measured whether a user has ever been, is or will ever be an administrator of Wikipedia; an additional interaction is included to test whether this can influence the effect of a meetup [as shown by 13]. To further test for differing treatment effects, I include an indicator (and interaction) on whether the meetup is of a work or social nature. A differentiation will further be made between the very first meetup of a user and all other meetups when assessing contribution behaviour to explore whether the first meetup has a particularly strong effect on creating an identity as a Wikipediaian; this also allows for better comparison with other studies which have just assessed effects of one/the first meeting [13, 54].

3.4 Statistical approach

To answer the research question on the effect of meetup participation on contributing behaviour, the treatment group of attendees was assigned a control group of comparable non-attendees. Making use of the control group allows for a quasi-experimental design. A difference-in-differences (DiD) approach will be used to assess the effect of meetups on productive behaviour: changes in behaviour before and after the meetup will be compared across the actual attendees (= treatment group) and the matched non-attendees who have not attended the corresponding (any) meetup (= control group). A DiD estimate is the

difference between the change in outcomes before (pre) and after (post) a treatment in a treatment versus a control group [58, 59]:

$$(\bar{y}_{TREAT}^{POST} - \bar{y}_{TREAT}^{PRE}) - (\bar{y}_{CONTROL}^{POST} - \bar{y}_{CONTROL}^{PRE}).$$

This measure corresponds to the estimated coefficient on the interaction between a treatment group dummy ($treat_i$), indicating whether user i belongs to the treatment group, and a post-treatment period dummy ($post_t$), indicating whether the observation occurs after the treatment, in a regression:

$$y_{it} = \beta_1 + \beta_2(treat_i) + \beta_3(post_t) + \beta_4(treat_i * post_t) + \epsilon_{it}.$$

In this specification:

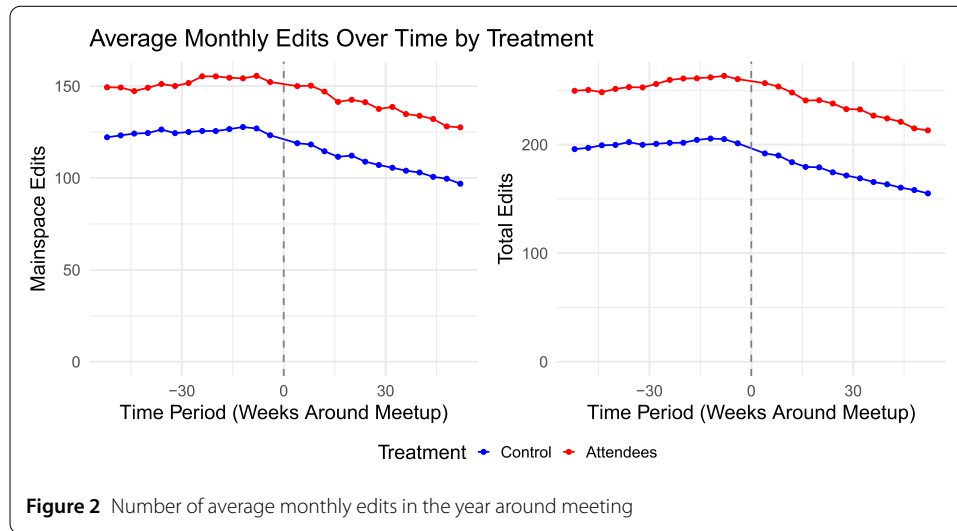
- y_{it} represents the outcome variable for individual i at time t ,
- $treat_i$ is a dummy variable that equals 1 if individual i belongs to the treatment group and 0 otherwise,
- $post_t$ is a dummy variable that equals 1 for the post-treatment period and 0 for the pre-treatment period,
- $treat_i \times post_t$ is the interaction term capturing the effect of the treatment on the treated group after the intervention,
- ϵ_{it} is the error term.

This formulation allows β_4 to capture the difference-in-differences estimate, which reflects the impact of the treatment on the outcome. To test for heterogeneity in the treatment effect (i.e. the effect of adminship and the type of meeting), triple interaction terms are included in the model which interact the treatment group dummy, the post-treatment dummy, and the variables of interest.

The DiD approach requires that the parallel trend assumption is fulfilled, meaning that in the absence of treatment, the treatment and control groups would have followed similar trends over time. Figure 2 presents the pre- and post- treatment trends for both groups and both dependent measures, i.e. number of edits within the German-language version of Wikipedia, in the article mainspace (left) and across all namespaces (right). Average number of edits for each month are shown. Visual inspection of the plot shows that the trends for the outcome variable were largely parallel before the introduction of the treatment. The treatment group had always been more active, on average, than the control group, but the number of edits were developing very similarly in the year leading up to the meetups.

As the data exhibits a multilevel structure [60] with attendance at meetups being nested in users, multilevel models with random intercepts for each user are estimated.

To assess the changes in productivity, I break the process into two separate parts, somewhat similarly to the approach followed in hurdle models [61]. First, I focus on all those users who have not made an edit in the week/month/year before the meetup and dichotomously model the decision of whether they make any edits after the meetup. A multilevel linear probability model (LPM) is used to model this decision [62, 63]. While LPMs exhibit some shortcomings, the computation and the interpretation of a DiD estimates and interactions effects are more straight-forward in linear models [64]. Non-linear DiD methods have been suggested but come with their own non-negligible complexities and challenges [65, 66]. The essential issue of heteroscedasticity is addressed using robust standard errors, employing the original form of the sandwich estimator [67].



In the second step, only users who have made an edit in the specified time frame before the meetup are included and the change in the extent of editing is analysed using a multilevel negative binomial model. The negative binomial model is particularly well-suited because the dependent variables—the number of edits—are count variables that are overdispersed, meaning the variance exceeds the mean. The model deals effectively with skewed data [68]. The coefficients from the model are interpreted as changes in the log of the expected number of edits.

This hurdle model like approach allows for the separation of the two distinct behaviours of users: first, whether users who have not edited in a certain timeframe and thus had some sort of break from Wikipedia resume editing after the meetup, and second, how the extent of editing activity changes for users who were already active. The first binary component examines the decision to re-start participation after a period of inactivity, which is an important indicator of the meetups' effectiveness in re-engaging users who had temporarily disengaged from the platform. The second part focuses on the intensity of activity among those who were already active before the meetup, allowing to assess how the meetup influenced the volume of contributions. The approach thus separates the decision to re-engage from the level of ongoing activity.

In addition to the DiD model, I follow the lagged dependent variable (LDV) approach to bound the causal effect (58, 243–247; 69, 70). In these model formulations, I control for the lagged level of activity in the period (in the week, the month, the year) before. Ding and Li [70] show that the DiD and LDV approaches share a bracketing property and can be used to calculate bounds on the causal effect. The model results for the LDV models as well the DiD models excluding outliers can be found in the Additional file 1. The main effect of meetups remains relatively stable across all model specifications.

4 Results

How do users' contributing activity on Wikipedia change after taking part in a meetup? This will be explored in the following: first, I will compare the before and after levels of productivity around a user's first meetup, and then discuss changes in activity across all meetups. After showing bivariate relationships, I will present multilevel models.

Table 1 Changes in editing activity after first meetup compared to activity before

Topic	Group	7 days	28 days	364 days
Changes in editing behaviour after first meetup: mainspace	Treatment	0.32 (51.87) 0	4.18 (172.95) 0	103.80 (8.44)* 0
	Control	-1407 / 905 -1.85 (39.48)***	-2916 / 4385 -8.95 (128.51)***	-18294 / 57852 -89.09 (1262.75)***
		0	0	0
		-819 / 487	-1973 / 1877	-15624 / 23170
	Difference Treatment - Control	2.17*	13.13***	192.86***
Changes in editing behaviour after first meetup: total	Treatment	2.19 (69.0)+ 0	5.57 (210.32) 0	210.7 (2765.83)*** 0
	Control	-1424 / 1470 -3.80 (48.54)***	-3100 / 4341 -16.60 (171.71)***	-19263 / 57456 -150.1 (1801.07)***
		0	0	0
		-818 / 637	-2725 / 1919	-35483 / 24173
	Difference Treatment - Control	6.00***	22.17***	360.75***

Given are mean (standard deviation), median, minimum / maximum for treatment group (n=3724) and control group (n=3718; some control users acted as controls for multiple meetup attendees). + denotes significance on 10 per cent level, * on the 5 per cent level, ** on 1 per cent level, *** on 0.1 per cent level.

4.1 Bivariate results

Based on first-time meetup goers, attendees increase their activity after taking part in the meetup both in the article mainspace as well as across all namespaces. All changes are positive, and while not all changes are significantly different from zero (which would reflect no change compared to their activity before the meetup), all changes are significantly different from the changes observed in the control group which decreased their activity (see Table 1). However, it needs to be noted that there is a very large variance in the changes in the number of edits, reflecting some skewness in the data which make simple, bivariate comparisons less straightforward.

Looking at all meetups, meetup attendees make slightly more edits in the week after the meetup (both across all namespaces as well as the mainspace), but again reduce their editing activity in the longer time frame. However, it is important to compare these negative numbers to those of the matched non-attendees: the control group also significantly reduces their activity in all time frames after the meetup. Compared to the control group, the reduction in the treatment group is smaller, but only significantly so in the short- and medium time frame when focusing on all namespaces (see Table 2).

4.2 Multivariate results

For each of the three time frames analysed, a multilevel LPM is estimated on those users who have not made an edit in the corresponding time frame before. This means, the probability is estimated that someone who has not edited in the week (month/year) before the meetup makes an edit in the week (month/year) after the meetup. In a second step, only users who have made an edit before will be analysed. Using a negative binomial model with random effects for users, it will be checked to what extent users who have shown some activity before the meetup changed the extent of their editing behaviour. In each of the models, meetup attendees will be compared with the matched control users consisting of similarly active non-attendees.

Table 2 Changes in editing activity after meetups compared to activity before

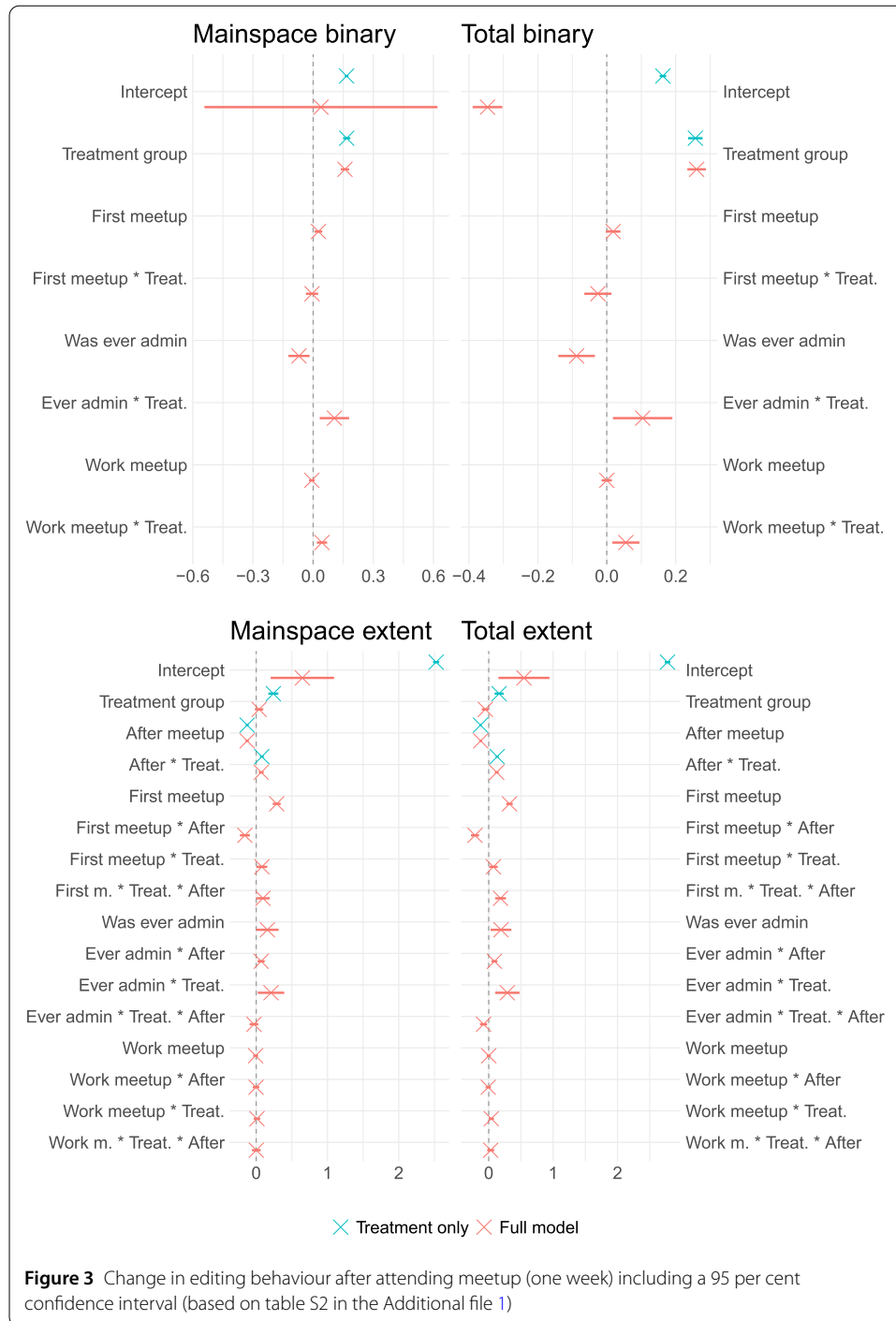
Topic	Group	7 days	28 days	364 days
Changes in editing behaviour after any meetup: mainspace	Treatment	0.26 (78.08)	-2.33 (274.76)	-211 (3010.81)***
		0	0	-36
	Control	-2956 / 5199	-18339 / 7918	-302092 / 73055
		-0.057 (141.17)	-4.09 (269.49)**	-236.5 (1741.17)***
		0	0	-24
Difference Treatment - Control	-2419 / 25166	-8110 / 38803	-68763 / 53296	
	0.32	1.76	25.53	
Changes in editing behaviour after any meetup: total	Treatment	1.17 (93.68)*	-3.88 (313.71)*	-361.6 (3482.71)***
		0	0	-75
	Control	-3145 / 5243	-18175 / 8501	-302571 / 73605
		-0.79 (146.06)	-8.87 (300.68)***	-389.7 (2411.32)***
		0	0	-45
Difference Treatment - Control	-2420 / 25218	8119 / 38841	-68841 / 53355	
	1.96*	4.99*	28.10	

Given are mean (standard deviation), median, minimum / maximum for treatment group (n=37216) and control group (n=36533; some control users acted as controls for multiple meetup attendees). + denotes significance on 10 per cent level, * on the 5 per cent level, ** on 1 per cent level, *** on 0.1 per cent level.

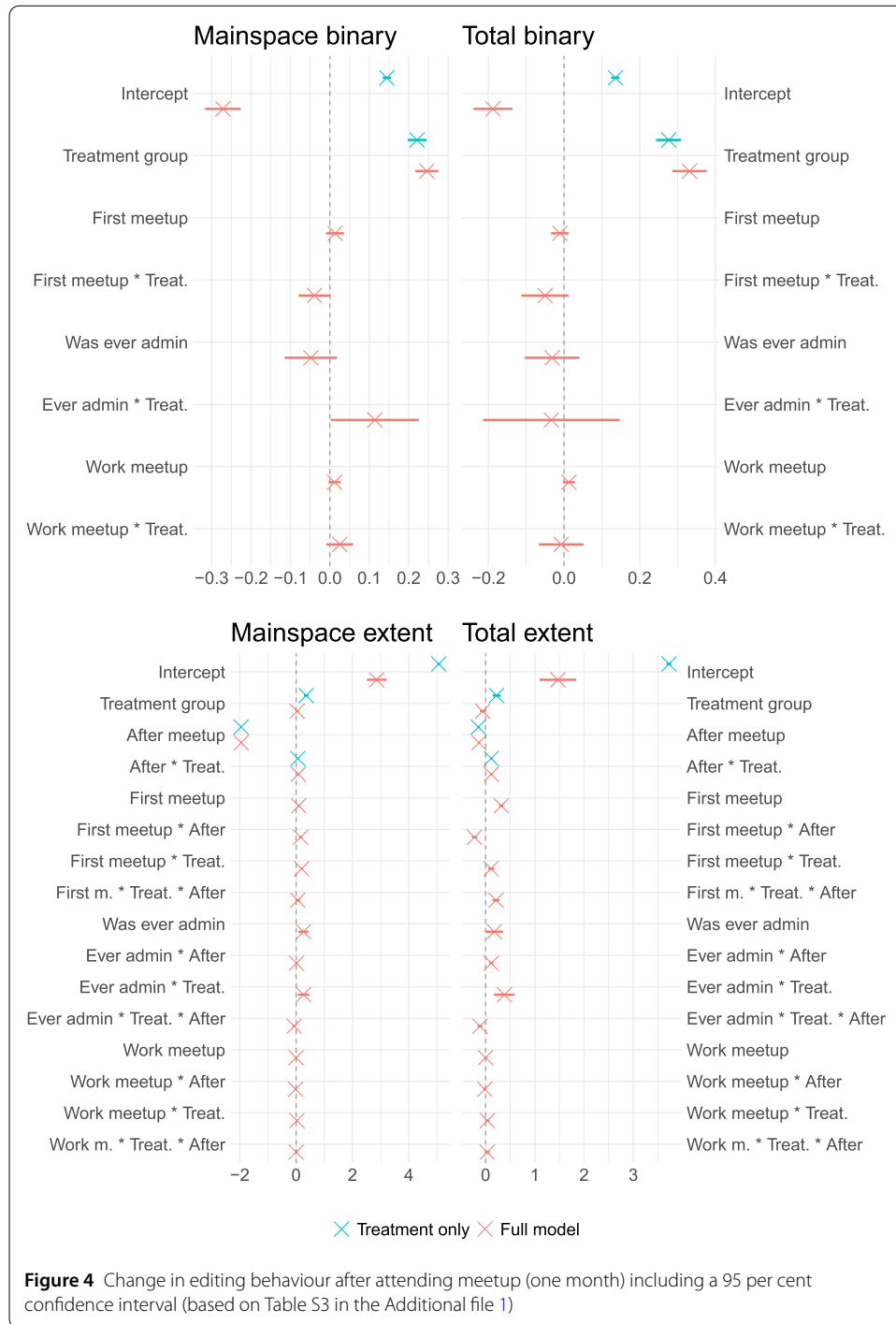
Two models are presented in the form of coefficient plots (result tables can be found in the Additional file 1). In the first one, only the single DiD treatment effect is estimated (model in green). The second model includes additional control variables and assesses whether the treatment effect depends on the type of meeting attended and whether the user is an administrator (model in red). Estimation results are shown in coefficient plots, differentiating binary (top) and continuous (bottom) models, as well as models concerning only the article namespace (left) and those looking at activity across all namespaces (right). The continuous models only include editors which have made at least one edit in the corresponding time frame before the meeting. The effect of control variables is not depicted but can be found in the corresponding tables in the Additional file 1.

Short term effect: one week Figure 3 displays the short term effect of meetups on editing behaviour of Wikipedians. The binary models show the estimated effects for a user who has not edited in the seven days before the meetup. The results suggest that a user is significantly more likely to contribute towards Wikipedia in the seven days after a meetup if they went to the meetup, i.e. they are in the treatment group instead of the control group. The probability for a user to make an edit in the article namespace in the week after the meetup if they have not edited in the week before lies at 16.7 per cent if they are in the control group, and rises to 33.4 per cent if they actually took part in the meetup. Across all namespaces, the probability to edit increases from 16.2 per cent to 42.0 per cent. These differences of 16.8 and 25.7 per cent respectively reflect the average treatment effect on the treated (ATT; model with only treatment effect).

Users in the control group make on average 11.9 per cent fewer edits in the mainspace after the meetup, as indicated by the negative binomial coefficient of -0.126, while attendees make, on average, 21.4 per cent more edits (0.194); across all namespaces, users in the control group make on average 12.0 per cent fewer edits (-0.128), while the number of contributions of attendees increases (17.8 per cent, coefficient of 0.164).



Further, users taking part in a meeting become more likely to start editing both across all namespaces and in the article namespace when taking part in a work-related meeting, but the extent of editing is not affected. Users attending their first meetup on average increase their editing behaviour even more after their first meetup than after any other meetup. This effect cannot be found for the article mainspace. This might suggest it is not the number of actual productive edits which increases, but edits in other namespaces which potentially refer to the meeting itself or discussions with others. While administrators



tend to make more edits across all namespaces—whether they have attended a meetup or not—they tend to increase their activity less after a meetup.

Medium term effect: one month Looking at a longer time frame, I again find positive effects of the treatment (see Fig. 4; focusing on the treatment model): considering users who have not edited in the month before a meetup, they become more likely to edit after the meetup if they have taken part (i.e. are in the treatment group). The predicted probability

to contribute in the month after the meetup increases from 14.4 per cent to 36.6 per cent in the mainspace, if the user is in the treatment instead of the control group. The probability to edit any site in the German-language version of Wikipedia increases from 13.6 per cent to 41.2 per cent.

Comparing the month before with the month after the meeting, users in the control group make on average 85.7 per cent (coefficient of -1.95) fewer edits in the mainspace (13.6 per cent fewer edits across all namespaces, coefficient of -0.15), while meetup attendees make only 78.4 per cent fewer edits in the mainspace (21.0 per cent more edits across all namespaces; coefficients of -1.53 and 0.19).

Focusing on the full model, there is a positive effect of the first meetup on the extent of edits when looking at all namespaces. Further, users who had ever been administrators are more likely to re-engage after a meeting if they had not been active in the month before, but, in line with the shorter term model, they tend to increase their activity less after a meetup than other users. The models do not show notable differences between work and social meetings.

Long term effect: one year Lastly, how does the activity one year after a meetup compare to the activity one year before? This is the longest trend that will be analysed; the estimated effects are shown in Fig. 5. In general, it is only seldom the case that users who have not made an edit in the past year will do so in the next. In fact, it was never the case when focusing on total edits so that no model was estimated for this specific case.

Looking at the mainspace model (including only the treatment effect), the baseline probability to edit in the next year is 5.7 per cent if the user did not take part in a meetup, and rises to 26.7 per cent if they did so. Again, there is a positive effect of taking part in a meetup, even in this very long term. Given the long time frame, it is important to note that other things could have also changed in the year (e.g., a user could have become administrator, taken part in more meetings, started working on a new subtopic, become retired and thus gained more time, etc.).

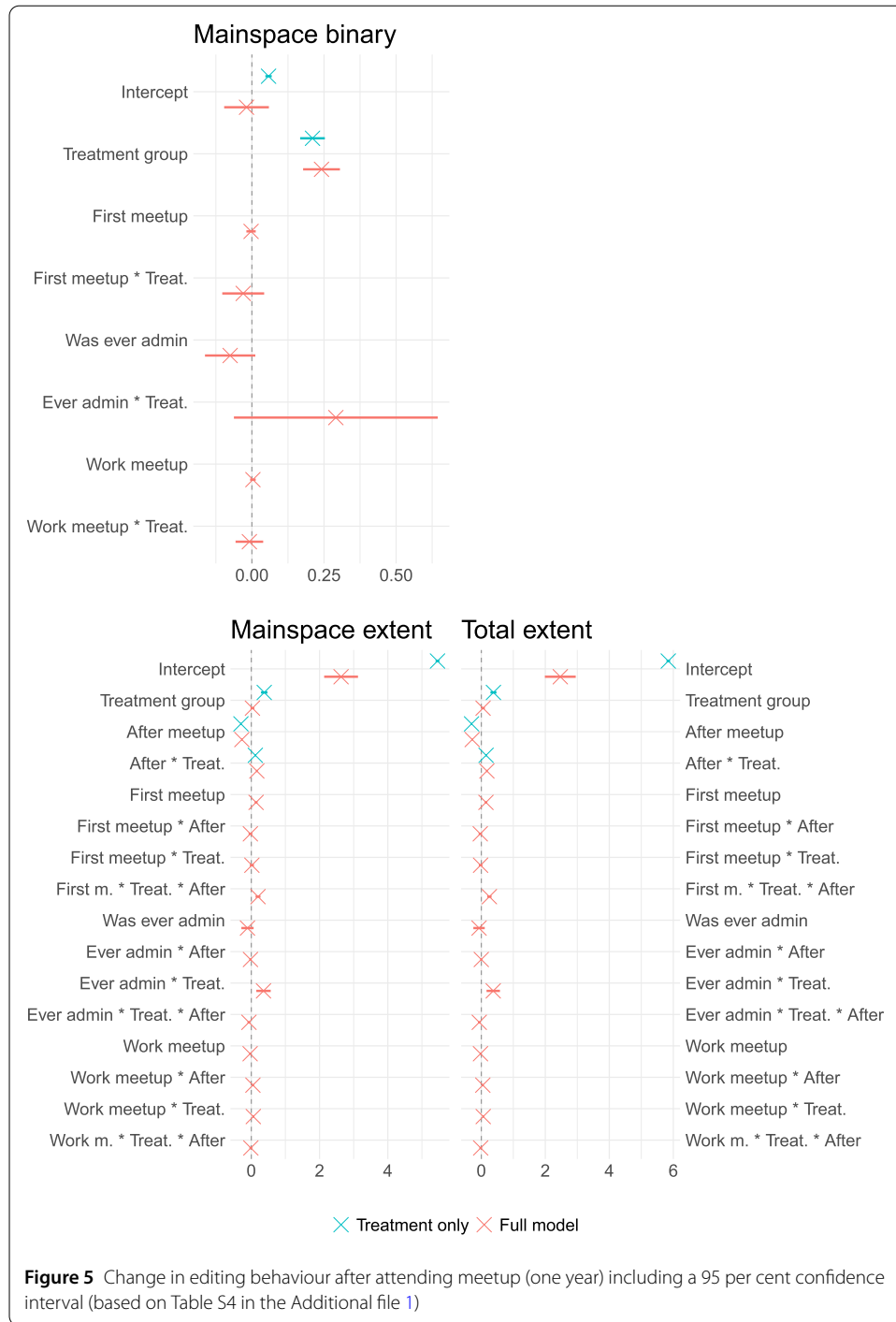
While few people start editing who have not edited before, many more change their editing behaviour which feed into the bottom models in Fig. 5: again (focusing on the model which includes only the treatment effect), users who took part in a meetup edit more. While members of the control group make on average 25.9 per cent fewer edits in the mainspace (and also 25.9 per cent fewer edits across all namespaces; coefficients of -0.30) in the year after the meetup, meetup attendees make 22.1 per cent more edits in the mainspace and 24.6 per cent more edits across all namespaces (coefficients of 0.20 and 0.22, respectively).

Also, there is a positive effect of the first meetup in the models analysing the extent of changes in editing behaviour. Even when looking at this long term trend, there is a negative effect of adminship, but no effect of whether the meeting was work-oriented or of a social nature.

In all binary models, the effect of being in the treatment group is positive and highly significant. This also holds for the DiD interaction of time and treatment in the models which focus on the extent of changes.

5 Discussion and conclusion

This paper has investigated to what extent offline meetups influence the productive online behaviour of Wikipedians. It was analysed how users change their editing behaviour after



a meetup in comparison to before. Comparing meetup attendees with a control group composed of comparable non-attendees, I found that across all time frames observed (one week, one month, one year), attending an offline meetup has a positive statistical effect on the contribution behaviour of users, supporting hypothesis 1. It is not necessarily the case that users increase their contributions after a meetup in comparison to before the meetup—while, bivariately, this holds true on average for the very first meetup a user attends, it is not the case when looking at all meetups—their reduction in contri-

butions is less than the reduction users of the control group experience. The difference-in-differences design revealed that even though there is a general trend to decrease editing activity across time, this decrease is significantly smaller for the treatment group of meetup attendees. Generally, users attending a meetup are much more likely to start contributing again after a meetup if they have not been editing articles before compared to non-attendees. This finding thus provides some support for the theoretical framework presented by Crowston and Fagnot [28]: after attending an offline meetup, which reflects an increased commitment to the project and the people, users increase their contributions and effort spent. Making users identify with the community—and one of the ways for identification being offline interactions—is important for sustained contribution to the online public good [30].

Concerning the comparison of work-related meetings with social meetings, there is only little support for hypothesis 2: attending a work-related offline meetup has a stronger, positive effect on editing behaviour than social meetings in the short time frame of one week. However, this is only in terms of re-engaging with the platform and not in terms of the extent of editing. When looking at the longer time frame of one month or one year, there are no significant differences between users attending a work or a social meeting. It is important to note that my analyses excluded activity on the day of the meetup: while work meetings like editathons generally lead to the production of edits and thus an increase in activity for users on that day [as users work collaboratively on Wikipedia during the meeting, see e.g. 53], my analyses show that this does not translate to sustained differences. Following the framework of Crowston and Fagnot [28], a work meeting might not be improving a person's capabilities to contribute towards user-generated content, but rather highlight potential gaps which need contribution. This could then explain that short-term effort is directed towards filling those gaps in the near future (i.e. the days after the meetup). It might well be that users substitute their usual editing activity instead of increasing it to address newly identified gaps. On the other hand, the positive impact of social meetups on editor productivity, which are equal to the effect of work meetups in the longer term, highlights the value in community-building.

In summary, there are positive effects for the community of Wikipedians after face-to-face meetings. In comparison to a matched control group of users, users who attended a meeting become or remain more active in the project. In contrast to other online communities [like those studied by 20, 23], offline gatherings do seem to support the community. These results suggest that users might feel more attached to the project after taking part in meetings and that the development of offline social capital is advantageous. The general decline found in the activity levels of users holds across open-source communities and peer production projects. This study again reiterated this finding for the German-language version of Wikipedia. Based on the framework of Crowston and Fagnot [28], it might well be that users do not see a need to contribute after the initial edits they made and that these edits have not led to any identification with the project. Given that the difference in the number of edits between users attending a meetup and those not attending one is minimal, and that relatively few users attend meetings at all, the overall benefit of offline meetings appears limited.

This study has several limitations which also offer future research opportunities. This study has compared before and after levels of activity in different time frames around a meetup. It is important to highlight that it is not possible to draw causal claims with the

observational digital trace data used in this study as there might be self-selection biases at play. As users are not and cannot be randomised into attending meetings, other unobserved factors may drive the relationship between meetup attendance and editing behaviour. For instance, users who are more interested in Wikipedia or have more free time may be more inclined to attend meetings. However, these users may also just edit more due to their inherent interest in the platform and the fact that they have the time to do so. While the most similar non-attendees were selected as control users in this study, matching could not overcome all differences. Future research might focus on profiling users which attend such offline events to better understand the composition and motivation of this group of people.

Beyond this, other statistical approaches could be feasible and might lead to more fine-grained results concerning the short and long term effects of meetups. For example, the collection of daily activity rates could be used in an interrupted time series approach [71]. This would allow assessing whether a meetup works as a sort of shock. The timing of meetups itself could also be considered to allow for more advanced panel data analysis with the meetup as an event to assess dynamic treatment effects [see also 59]; the different treatment times were not considered in this study beyond the inclusion of a year fixed effect. Further research could also explore the effect of multiple meetings on one person, i.e. the idea that a treated observation can become re-treated and whether there are potential diminishing returns of attending meetups. Relatedly, when matching attendees to comparable non-attendees, the matched users were always users who never partook in meetings. Another approach would be to match users who have had an identical treatment history, i.e. attended the same number of previous meetings (at the same time) [see for such discussions in the recent experimental methodological literature e.g. 72]. However, when following this approach, the pool of potential matchable users would have shrunk enormously, as those going to more than one meetup could only be compared with the pool of other meetup goers. Such a matching procedure allows to better assess the effects of additional meetups on a user.

In my data, I observe activity data of all users ever active on Wikipedia in its twenty years of existence. Across the years observed, user data is retained, and Wikipedia accounts are generally not deleted. From a technical perspective, there is thus no dropout from this longitudinal dataset. However, over time, many users stop being active: they stop contributing towards the online encyclopaedia, or they might also stop attending offline meetups. In my analyses, I have considered long-term changes in behaviour and I considered who started to contribute after meetings, however, I have not specifically addressed a withdrawal from online and/or offline components of the project. In this regard, future research can focus on leavers of the website and contrast them with those who remain active. In line with other research on Wikipedia, I had to exclude anonymous editing as anonymous edits cannot be ascribed to a specific user. However, it might be the case that users edit anonymously, then attend a meeting, and then sign up. Such dynamics are impossible to capture with the data at hand.

Further, newcomers—defined as users not having made an edit before taking part in a meetup—have been excluded from this study. The analysis could be extended to specifically focus on newcomers [like the study of 54], comparing the Wikipedia trajectories of those starting with an offline encounter with a comparable control group of users who have registered at the same point in time but who have not taken part in a meetup. Such

an analysis can provide insights into explaining initial editing. Also, I have only focused on the change in the number of edits—further studies could look at the quality of edits. This would allow us to find out whether the quality of edits increases after a collaborative editing event such as an editathon.

Lastly, there might also be some biases in the data. Meetup attendance in the dataset of Schwitter [12] was collected from the registration of interest written before the meetup took place. It is, of course, not mandatory to sign up for a meeting before attending; it is also not mandatory to attend after registering or deregister if not attending. It can thus be expected that there are some errors in the attendance lists.

This paper provides the first large-scale analysis of the effect of informal meetups within a community focused on a peer production project. The study took the German-language version of Wikipedia as an example case—extending this line of research across other projects is an important next step. Contrary to findings in other online communities, this research revealed a positive influence of meetups on individual contribution behaviour within Wikipedia. Rather than undermining the community, offline gatherings serve to bolster the remarkable success of this online public good through the cultivation of offline social capital. Future research should employ qualitative or mixed-method approaches to study why and through what mechanisms meeting attendance translates into sustained contribution, which will deepen our understanding of collaborative online environments.

Abbreviations

DiD, Difference-in-differences; LDV, Lagged dependent variable; LPM, Linear probability model.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-024-00506-w>.

Additional file 1. An online appendix is uploaded separately. (PDF 270 kB)

Acknowledgements

The author wants to thank Ulf Liebe, Andreas Murr, Wojtek Przepiorka, and Richard Lampard for helpful feedback to previous versions of this manuscript, Achim Edelmann for technical advice regarding data preparation and the Scientific Computing Team at the University of Leipzig as well as the IT Team at the Mannheim Centre for European Social Research (MZES) for computing support. The publication of this article was funded by the Mannheim Centre for European Social Research (MZES) which the author gratefully acknowledges.

Author contributions

NS confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Wikimedia Foundation: https://meta.wikimedia.org/wiki/Grants:Project/nschwitter/The_Role_of_Offline_Ties_of_Wikipedians. The funder had no role in the conceptualisation, design, data collection, analysis, decision to publish, or preparation of the manuscript. The publication of this article was funded by the Mannheim Centre for European Social Research (MZES).

Data Availability

The datasets generated and/or analysed during the current study are available in the OpenScienceFramework repository: <https://doi.org/10.17605/OSF.IO/PS3TF>.

Declarations

Competing interests

The project received funding by the Wikimedia Foundation.

Received: 4 July 2024 Accepted: 29 October 2024 Published online: 05 November 2024

References

1. Benkler Y, Shaw A, Hill BM (2015) Peer production: A form of collective intelligence. In: Malone T, Bernstein M (eds) *Handbook of collective intelligence*. MIT Press, Cambridge
2. Ortega F, Gonzalez-Barahona JM, Robles G (2008) On the inequality of contributions to Wikipedia. In: *Proceedings of the 41st annual Hawaii international conference on system sciences*. IEEE, New York. <https://doi.org/10.1109/hicss.2008.333>
3. Jiang L, Mirkovski K, Wall JD, Wagner C, Lowry PB (2018) Proposing the core contributor withdrawal theory (CCWT) to understand core contributor withdrawal from online peer-production communities. *Internet Res* 28(4):988–1028. <https://doi.org/10.1108/intr-05-2017-0215>
4. Steinmacher I, Wiese I, Chaves AP, Gerosa MA (2013) Why do newcomers abandon open source software projects? In: *6th international workshop on cooperative and human aspects of software engineering*. IEEE, New York <https://doi.org/10.1109/chase.2013.6614728>
5. McInnis BJ, Murnane EL, Epstein D, Cosley D, Leshed G (2016) One and done. In: *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. ACM Press, New York. <https://doi.org/10.1145/2818048.2820075>
6. Chen IYL (2007) The factors influencing members' continuance intentions in professional virtual communities: A longitudinal study. *J Inf Sci* 33(4):451–467. <https://doi.org/10.1177/0165551506075323>
7. TeBlunthuis N, Shaw A, Hill BM (2018) Revisiting “the rise and decline” in a population of peer production projects. In: *Proceedings of the 2018 conference on human factors in computing systems*. ACM Press, New York. <https://doi.org/10.1145/3173574.3173929>
8. Adams J, Brückner H (2015) Wikipedia, sociology, and the promise and pitfalls of big data. *Big Data Soc* 2(2):2053951715614332. <https://doi.org/10.1177/2053951715614332>
9. Giles J (2005) Internet encyclopaedias go head to head. *Nature* 438(7070):900–901. <https://doi.org/10.1038/438900a>
10. Halfaker A, Geiger RS, Morgan JT, Riedl J (2012) The rise and decline of an open collaboration system. *Am Behav Sci* 57(5):664–688. <https://doi.org/10.1177/0002764212469365>
11. Ortega F, Gonzalez-Barahona JM, Robles G (2008) Quantitative analysis of the top ten wikipedias. In: *International conference on software and data technologies: communications in computer and information science*. Springer, Berlin and Heidelberg, pp 257–268. https://doi.org/10.1007/978-3-540-88655-6_19
12. Schwitzer N (2023) Bridging the offline and the online: twenty years of offline meeting data of the German-language Wikipedia. *J Comput Soc Sci*. <https://doi.org/10.31235/osf.io/g96tk>
13. Stegbauer C (2009) *Wikipedia*. Springer, Wiesbaden
14. Schwitzer N (2022) *The role of offline ties in online communities: the case of Wikipedia*. PhD thesis, University of Warwick
15. Richter P (2020) *Die Wikipedia-Story*. Campus Verlag, Frankfurt am Main
16. Angelopoulos S, Merali Y (2015) Bridging the divide between virtual and embodied spaces: exploring the effect of offline interactions on the sociability of participants of topic-specific online communities. In: *48th Hawaii international conference on system sciences*. <https://doi.org/10.1109/hicss.2015.239>
17. Sessions LF (2010) How offline gatherings affect online communities. *Inf Commun Soc* 13(3):375–395. <https://doi.org/10.1080/13691180903468954>
18. Ganglbauer E, Fitzpatrick G, Subasi Ö, Güldenpfennig F (2014) Think globally, act locally. In: *Proceedings of the 17th ACM conference on computer supported cooperative work & social computing*. ACM Press, New York. <https://doi.org/10.1145/2531602.2531664>
19. Lin H-F (2007) The role of online and offline features in sustaining virtual communities: an empirical study. *Internet Res* 17(2):119–138. <https://doi.org/10.1108/10662240710736997>
20. Shen C, Cage C (2013) Exodus to the real world? assessing the impact of offline meetups on community participation and social capital. *New Media Soc* 17(3):394–414. <https://doi.org/10.1177/1461444813504275>
21. Xie B (2008) The mutual shaping of online and offline social relationships. *Inf Res: Int Electr J* 13(3)
22. Koh J, Kim Y-G, Kim Y-G (2003) Sense of virtual community: A conceptual framework and empirical validation. *Int J Electron Commer* 8(2):75–94. <https://doi.org/10.1080/10864415.2003.11044295>
23. McCully W, Lampe C, Sarkar C, Velasquez A, Sreevivasan A (2011) Online and offline interactions in online communities. In: *Proceedings of the 7th international symposium on wikis and open collaboration*, pp 39–48
24. Hristova D, Quattrone G, Mashhadi A, Capra L (2013) The life of the party: impact of social mapping in OpenStreetMap. In: *Proceedings of the international AAAI conference on web and social media* 7(1), pp 234–243. <https://doi.org/10.1609/icwsm.v7i1.14416>
25. Schwitzer N (2024) Offline connections, online votes: the role of offline ties in an online public election. *New Media Soc* 14614448241274456. <https://doi.org/10.1177/14614448241274456>
26. Lombard M, Ditton T (1997) At the heart of it all: the concept of presence. *J Comput-Mediat Commun* 3(2):321
27. Ren Y, Zhang H, Kraut RE (2023) How did they build the free encyclopedia? a literature review of collaboration and coordination among Wikipedia editors. *ACM Trans Comput-Hum Interact* 31(1):7. <https://doi.org/10.1145/3617369>
28. Crowston K, Fagnot I (2018) Stages of motivation for contributing user-generated content: A theory and empirical test. *Int J Hum-Comput Stud* 109:89–101. <https://doi.org/10.1016/j.ijhcs.2017.08.005>
29. Schwartz SH, Howard JA (1982) Helping and cooperation: A self-based motivational model. In: Derlega VJ, Grzelak J (eds) *Cooperation and helping behavior*. Elsevier, Amsterdam, pp 327–353. <https://doi.org/10.1016/b978-0-12-210820-4.50019-8>
30. Klandermans B (1997) *The social psychology of protest*. Blackwell Publishers, Oxford
31. Konieczny P (2009) Wikipedia: community or social movement? *Interface: J Soc Mov* 1(2):212–232
32. Schroer J, Hertel G (2009) Voluntary engagement in an open web-based encyclopedia: wikipedians and why they do it. *Media Psychol* 12(1):96–120. <https://doi.org/10.1080/15213260802669466>
33. Simon B, Loewy M, Stürmer S, Weber U, Freytag P, Habig C, Kampmeier C, Spahlinger P (1998) Collective identification and social movement participation. *J Pers Soc Psychol* 74(3):646–658. <https://doi.org/10.1037/0022-3514.74.3.646>
34. Diani M, McAdam D (2003) *Social movements and networks*. Oxford University Press, Oxford. <https://doi.org/10.1093/0199251789.001.0001>

35. Passy F, Giugni M (2001) Social networks and individual perceptions: explaining differential participation in social movements. *Sociol Forum* 16(1):123–153. <https://doi.org/10.1023/a:1007613403970>
36. Bryant SL, Forte A, Bruckman A (2005) Becoming Wikipedian. In: Proceedings of the 2005 international conference on supporting group work. ACM Press, New York. <https://doi.org/10.1145/1099203.1099205>
37. Kuznetsov S (2006) Motivations of contributors to Wikipedia. *Comput Soc* 36(2). <https://doi.org/10.1145/1215942.1215943>
38. Algan Y, Benkler Y, Morell MF, Hergueux J (2013) Cooperation in a peer production economy experimental evidence from Wikipedia. *SSRN Electron J*. <https://doi.org/10.2139/ssrn.2843518>
39. Anthony D, Smith SW, Williamson T (2009) Reputation and reliability in collective goods: the case of the online encyclopedia Wikipedia. *Ration Soc* 21(3):283–306. <https://doi.org/10.1177/1043463109336804>
40. Balestra M, Arazy O, Cheshire C, Nov O (2016) Motivational determinants of participation trajectories in Wikipedia. In: Tenth international AAAI conference on web and social media
41. González-Anta B, Orengo V, Zornoza A, Peñarroja V, Martínez-Tur V (2021) Understanding the sense of community and continuance intention in virtual communities: the role of commitment and type of community. *Soc Sci Comput Rev* 39(3):335–352. <https://doi.org/10.1177/0894439319859590>
42. Gotham KF (1999) Political opportunity, community identity, and the emergence of a local anti-expressway movement. *Soc Probl* 46(3):332–354. <https://doi.org/10.2307/3097104>
43. Ren Y, Harper FM, Drenner S, Terveen L, Kiesler S, Riedl J, Kraut RE (2010) Increasing commitment to online communities: designing from theory. In: Proceedings of the ACM 2011 conference on computer supported cooperative work. ACM Press, New York
44. Kramer RM, Brewer MB (1984) Effects of group identity on resource use in a simulated commons dilemma. *J Pers Soc Psychol* 46(5):1044–1057. <https://doi.org/10.1037/0022-3514.46.5.1044>
45. Ellemers N, Gilder DD, Haslam SA (2004) Motivating individuals and groups at work: A social identity perspective on leadership and group performance. *Acad Manag Rev* 29(3):459–478. <https://doi.org/10.2307/20159054>
46. Johnson RE, Chang C-HD, Yang L-Q (2010) Commitment and motivation at work: the relevance of employee identity and regulatory focus. *Acad Manag Rev* 35(2):226–245. <https://doi.org/10.5465/amr.2010.48463332>
47. Bateman PJ, Gray PH, Butler BS (2011) Research note - the impact of community commitment on participation in online communities. *Inf Syst Res* 22(4):841–854. <https://doi.org/10.1287/isre.1090.0265>
48. Baytiyeh H, Pfaffman J (2009) Why be a Wikipedian. In: Proceedings of the 9th international conference on computer supported collaborative learning - volume 1. CSCW'09. International Society of the Learning Sciences, Rhodes, pp 434–443
49. Gallus J (2017) Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia. *Manag Sci* 63(12):3999–4015. <https://doi.org/10.1287/mnsc.2016.2540>
50. Restivo M, van de Rijt A (2014) No praise without effort: experimental evidence on how rewards affect Wikipedia's contributor community. *Inf Commun Soc* 17(4):451–462. <https://doi.org/10.1080/1369118X.2014.888459>
51. Morgan JT, Bouterse S, Walls H, Stierch S (2013) Tea and sympathy: crafting positive new user experiences on Wikipedia. In: Proceedings of the 2013 conference on computer supported cooperative work. CSCW '13. Association for Computing Machinery, New York, pp 839–848. <https://doi.org/10.1145/2441776.2441871>
52. Morgan JT, 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. ACM Press, New York. <https://doi.org/10.1145/3233391.3233544>
53. Littlejohn A, Hood N, Rehm M, McGill L, Rienties B, Highton M (2019) Learning to become an online editor: the editathon as a learning environment. *Interact Learn Environ* 29(8):1258–1271. <https://doi.org/10.1080/10494820.2019.1625557>
54. Farzan R, Savage S, Saviaga CF (2016) Bring on board new enthusiasts! A case study of impact of Wikipedia art+feminism edit-A-thon events on newcomers. In: Lecture notes in computer science. Springer, New York, pp 24–40. https://doi.org/10.1007/978-3-319-47880-7_2
55. Li A, Yao Z, Yang D, Kulkarni C, Farzan R, Kraut RE (2020) Successful online socialization: lessons from the Wikipedia education program. In: Proceedings of the ACM on human-computer interaction 4(CSCW1), pp 1–24. <https://doi.org/10.1145/3392857>
56. Zhao Z (2004) Using matching to estimate treatment effects: data requirements, matching metrics, and Monte Carlo evidence. *Rev Econ Stat* 86(1):91–107. <https://doi.org/10.1162/003465304323023705>
57. Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55. <https://doi.org/10.1093/biomet/70.1.41>
58. Angrist JD, Pischke J-S (2009) Mostly harmless econometrics. Princeton University Press, Princeton
59. Goodman-Bacon A (2021) Difference-in-differences with variation in treatment timing. *J Econom* 225(2):254–277. <https://doi.org/10.1016/j.jeconom.2021.03.014>
60. Raudenbush SW, Bryk AS (2001) Hierarchical linear models: applications and data analysis methods. SAGE Publications Ltd, Los Angeles
61. Cragg JG (1971) Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* 39(5):829–844. <https://doi.org/10.2307/1909582>
62. Wooldridge JM (2010) Econometric analysis of cross section and panel data. MIT Press, Cambridge
63. Mood C (2009) Logistic regression: why we cannot do what we think we can do, and what we can do about it. *Eur Sociol Rev* 26(1):67–82. <https://doi.org/10.1093/esr/jcp006>
64. Ai C, Norton EC (2003) Interaction terms in logit and probit models. *Econ Lett* 80(1):123–129. [https://doi.org/10.1016/s0165-1765\(03\)00032-6](https://doi.org/10.1016/s0165-1765(03)00032-6)
65. Athey S, Imbens GW (2006) Identification and inference in nonlinear difference-in-differences models. *Econometrica* 74(2):431–497. <https://doi.org/10.1111/j.1468-0262.2006.00668.x>
66. Blundell R, Dias MC (2009) Alternative approaches to evaluation in empirical microeconomics. *J Hum Resour* 44(3):565–640
67. Liang K-Y, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73(1):13–22. <https://doi.org/10.1093/biomet/73.1.13>

68. Green JA (2021) Too many zeros and/or highly skewed? A tutorial on modelling health behaviour as count data with Poisson and negative binomial regression. *Health Psychol Behav Med* 9(1):436–455. <https://doi.org/10.1080/21642850.2021.1920416>
69. Keele L, Cubbison W, White I (2021) Suppressing black votes: A historical case study of voting restrictions in Louisiana. *Am Polit Sci Rev* 115(2):694–700. <https://doi.org/10.1017/s0003055421000034>
70. Ding P, Li F (2019) A bracketing relationship between difference-in-differences and lagged-dependent-variable adjustment. *Polit Anal* 27(4):605–615. <https://doi.org/10.1017/pan.2019.25>
71. Bernal JL, Cummins S, Gasparrini A (2017) Interrupted time series regression for the evaluation of public health interventions: A tutorial. *Int J Epidemiol* 46(1):348–355. <https://doi.org/10.1093/ije/dyw098>
72. Imai K, Kim IS, Wang EH (2021) Matching methods for causal inference with time-series cross-sectional data. *Am J Polit Sci* 1(19). <https://doi.org/10.1111/ajps.12685>

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
