Effects of stigmergic and explicit coordination on Wikipedia article quality

Kevin Crowston
Syracuse University
School of Information Studies
crowston@syr.edu

Amira Rezgui
Syracuse University
School of Information Studies
arezgui@syr.edu

Abstract

Prior research on Wikipedia has noted the importance of both explicit coordination of edits (i.e., through the article Talk page) and stigmergic coordination (i.e., through the article itself). Using a panel data set of article quality and edits for 23 articles over time, we examine the impact of different kinds of edits on article quality. We find that stigmergically-coordinated edits seem to have the biggest effect on quality, but that explicit coordination of major edits also predicts article quality. The findings have implications for both research on coordination in Wikipedia and for supporting editors.

1. Introduction

Wikipedia has become one of the most important information sources on the Internet. Accordingly, there is great interest in how Wikipedia editors work together to create high-quality articles, both as a practical issue and as a research problem. Writing articles collaboratively in Wikipedia requires coordination among editors. Each new editor working on an article in Wikipedia has the potential to contribute new knowledge with which to extend an article, insight into how the article should be written and vigilance to discover errors in fact, grammar or judgment. But for these contributions to be productive, the editors need to manage the interdependent aspects of the article, such as its content, structure and style.

In this article, we explore two different modes of coordination, explicit coordination through posts to article Talk pages and non-explicit (i.e., stigmegic) coordination, involving only edits to the article itself. We present evidence that non-explicitly coordinated edits are the majority of edits to the Wikipedia articles in our sample, but that explicitly coordinated edits are also predictive of article quality. These findings have implications for understanding and supporting Wikipedia editors.

2. Literature review

Interest in Wikipedia article quality has sparked several lines of research to determine how to assess quality and to identify factors that lead to quality articles.

Quality assessment. An initial question was simply how to assess article quality. Early efforts involved hand-coding articles, either by experts (e.g., [1, 2, 3]) or by editors themselves [4]. For example, for the English Wikipedia 1.0 project, an assessment scale[1] was created that ranges from "Featured Article" (FA) at the top, through B and C in the middle to "Stub" or "Start" at the bottom. However, manual rating is labourious, hard to scale to the number of articles on Wikipedia and requires frequent re-evaluations to keep up-to-date [4].

Various proxies for quality have been identified, e.g., measures as simple as the length of the article [5], or based more detailed article features (e.g., [6, 7]). Other studies have considered characteristics of the group of contributors to the article and their interactions (e.g., [8, 9]). Fortunately for our study, research on quality assessment has matured to the point where quality can be assessed automatically with reasonable agreement with manual ratings [6, 4].

Input factors. A second important question was to identify factors that lead to a high-quality article. A simple approach simply considers the volume of inputs, e.g., "featured" articles have been found to have more editors and more edits than non-featured [10]. Of course, it is well known that not all edits are a contribution to the article: some are deliberate vandalism and others are simply not helpful.

Editor experience. Other research considers differences among contributing editors, such as editor experience. For instance, studies have found a positive impact on quality of having contributions from editors who have experience from contributing to articles in different categories and also with formal administrative

---

[1] https://en.wikipedia.org/wiki/Wikipedia:
Content_assessment

positions in Wikipedia [2]. Another important factor is the group's experience working together. For instance, reference [11] found a positive impact of editors' prior interactions, as well as their number (though interaction is not defined in the article). Reference [12] found positive impacts of the editors' social capital as derived from internal bonding (i.e., having worked together before) as well as the previously mentioned external bridging (i.e., working on diverse other articles) and team size.

Editor contributions. A second line of work considers in more detail the type of work done by the editors. One approach is to cluster editors by their contributions to identify different roles (e.g., substantive expert vs. fact updater [13]). Interestingly, this work found a positive impact of article edits and negative impact of Talk edits [13] and in contrast to prior work, a negative impact of number of editors. Finally, they identified different impacts from different kinds of editors, though the improvement from adding these to the model was minimal [13]. Reference [14] identified 6 roles and 5 different patterns of contributions of these roles to articles, finding, for example, that articles where contributions from "all-around contributors" dominated tended to be of higher quality, though some of these results may reflect the article's stage of development.

A key value of Wikipedia is that editors are involved in process not just of identifying problems but also in fixing them [15]. This emphasis on action suggest that talking alone will not effective in improving the encyclopedia (indeed, might even be negative as found by [13]). As another example of the importance of editing, in defining roles on Wikipedia, eference [16] find that editors in most roles, regardless of their focus, contribute the majority of their edits to content. Even "social networkers" [16], those contributors whose focus is on "community aspects of Wikipedia", still make about 20% of their edits to content on average, with the cut-off defined as "less than 45% of total edits to content pages" [16].

Process and coordination. A related line of research examined more specifically the process editors follow. For example, reference [15] described various processes adopted for assuring information quality. A particular concern is how editors coordinate their actions, which is the focus of this paper. We discuss three approaches to coordination. First, past work suggests that article Talk pages are an important forum for *explicit coordination*. For example, [17] manually coded discussion on Talk pages and found that they were used for (among other things) planning and evaluating edits. Reference [1] found that the length of the Talk page was predictive of article quality as assessed by raters. Reference [18]

found that Talk discussions were particularly important early in the life of a project.

While Talk pages are a venue for explicit coordination, coordination can also happen through the work contributed to an article, as the work done provides a structure for others to contribute [18]. One perspective on this work focuses on possibilities of *implicit coordination* [18], the second form of coordination, coordination guided by shared understandings. For example, editors may share a vision of what an article should cover that guides their decisions about coverage, perhaps based on earlier discussions, earlier collaboration, or a common point of view. Reference [18] found that such implicit coordination was most effective when a small group of editors made the majority of contributions (i.e., number of editors had a negative impact on quality), as only the smaller group needed to come to agreement and develop a shared mental model of the article.

However, the reliance on shared mental models poses limits on the viability of implicit coordination. For example, in self-organized groups such as a Wikipedia article editors, there is no authority to impose a particular way of working. Such distributed teams pose particular problems, as they often lack face-to-face meetings at which to develop shared understandings. As a result, distributed work is characterized by numerous discontinuities [19], that is, a lack of coherence in some aspects of the work setting (e.g., organizational membership, business function, task, language or culture) that hinders members trying to make sense of the task and communication with others, or that produces unintended information filtering or misunderstandings [20]. These interpretative difficulties, in turn, make it hard for group members to develop the shared mental models necessary for implicit coordination.

More recent research has identified a third possible mode of coordination: that the information needed to coordinate work can be communicated through the outcome of the work itself [21, 22, 23, 24], a mode of coordination analogous to the biological process of stigmergy [25]. Heylighen defines stigmergy thusly: "A process is stigmergic if the work... done by one agent provides a stimulus (stigma) that entices other agents to continue the job" [26]. Accordingly, *stigmergic coordination* can be defined as coordination based on signals from the shared work rather than on shared understandings or explicit communication.

We note that the Wikipedia infrastructure provides direct support for stigmergic coordination. To facilitate tracking modifications and edits, the Wikimedia systems enables a logged-in user to set a watchlist. A watchlist

is a page that generates a list of recent changes made to the pages being watched. In this way, an editor can keep track of what's happening to these pages and so react to these changes. Reference [27] found that the majority of edits to two example articles were not associated with discussion on the article Talk page, suggesting the possibility of stigmergic coordination. For example, minor fixes and vandalism fixing did not seem to require discussion. However, they also found that Talk posts did seem to be related to article quality, suggesting the continued importance of explicit coordination of at least some edits.

Limits of past studies. A limitation of most studies reviewed is that they rely on cross-sectional data, making it impossible to assess causality (e.g., do more editors result in higher quality or do higher quality articles attract more editors). In contrast, references [13, 18] measured quality at two or more points in time and controlled for prior quality. Both found a negative impact of prior quality on quality, which the authors interpreted as an indication of floor and ceiling effects (e.g., it is more difficult to improve articles that are already of high quality) and regression to the mean. The authors also note negative impacts of time on quality that they attribute to increases in assessment standards over time.

Summary. From this review, we draw out several predictions that we will test in this paper. Given the observed impact of the number of edits, we expect that edits made to the article will generally be associated with increased article quality, both edits that are discussed on the Talk page (i.e., explicitly coordinated, **H2**) and those that are not (i.e., implicitly or stigmergically coordinated, **H1**). An open question that we will examine is which approach increases quality more. Of course, vandalism is expected to decrease quality (**H3**). Finally, we suggest that talk without editing will not improve quality (**H4**). To address the problems with causality, we will examine these relationships over time. More specifically:

H1 More good faith article edits without Talk will increase article quality

H2 More good faith article edits with Talk will increase article quality

H3 Damaging edits of all sorts will reduce article quality

H4 Talk-only edits will have no effect on article quality

Of particular interest is the relative importance of H1 and H2 for article quality (i.e., the relative importance of stigmergic and explicit coordination) as the literature does not provide a reason to expect one to be more important than the other.

3. Method

The design of the study presented in this article is a panel study comparing the effect of edits on article quality in a sample of articles measured over time. We describe the pattern of editing observed in articles to determine the extent to which coordination appears to be done stigmergically and the impact of different kinds of edits on quality.

3.1. Data

We selected a stratified set of 23 Wikipedia articles of different quality levels from the English-language version of Wikipedia, chosen as it has the largest number of articles and the software we used was originally developed to process this language. The number of articles is limited because extracting data was time consuming.

For each article, we used the web service ORES (Objective Revision Evaluation Service[2]) to retrieve a measure of the article quality as of the start of each year. Quality is measured on a five-point scale that "roughly corresponds to the English Wikipedia 1.0 assessment rating scale"[3]. For a few articles for a few years, quality scores were not available; these years were omitted from the data for these articles. An advantage of this approach is that it applies a consistent metric to all versions of an article, eliminating effects of changing standards.

Work done in Wikipedia is recorded in the revision history of the Wikipedia page. The revision history shows nearly every version of the page (in extremely rare cases, a revision can be deleted, e.g., if an edit added libelous content), with a time stamp (date and time of the edit), the editor's ID (or IP address for anonymous edits), an optional flag for minor changes applied by the editor (to be applied when "only superficial differences exist between the current and previous versions"[4]), the size of the change in bytes and an optional comment given by the editor.

For each edit, we obtained an estimate of the quality of the edit, also from ORES, which generates a score of edit quality as either damaging or good-faith. For good-faith edits we recorded the editor's use of the "minor edit" flag. The result is that edits are sorted into three types: good-faith major edits, good-faith minor edits and damaging edits. We similarly classified edits

---

[2] http://ores.wmflabs.org/
[3] https://www.mediawiki.org/wiki/ORES#Article_quality
[4] https://en.wikipedia.org/wiki/Help:Minor_edit

to each article's Talk page to capture editors' discussion and explicit coordination of work on the article.

We used the Wikipedia API to collect the data for our study because it provides precise information about the edits that can be easily connected to other data, such as the edit quality from ORES score. We wrote a program to extract data using the API and to parse the revision history for each article and the associated Talk page to identify the individual edits. For each edit, the program retrieved the edit quality from the ORES Web service. We obtained edits to the 23 articles and associated Talk pages from as early as 2001 through 2015.

3.2. Data processing

We wanted to explore the extent to which editors explicitly coordinated their work versus relying on stigmergic coordination. We therefore looked for cases where edits to an article were accompanied by edits to the associated Talk page. Talk pages were identified in prior research as a forum for coordination, so we took these discussions as evidence of the possibility of explicit coordination.

Edits from the revision history were grouped into sessions [28], defined as a set of consecutive edits made by the same editor to the same article or related Talk page, where consecutive edits are separated by less than 720 minutes. The intuition is that editors may have different work habits, e.g., saving frequently vs. infrequently, but grouping temporally-adjacent edits together captures work that was done in one sitting. The initial work on sessions [28] used a gap of 60 minutes to define sessions. However, their data seem to suggest an inter-session gap somewhere between an hour and a day. We therefore chose a gap of 720 minutes as a more conservative definition of a session.

In forming sessions, we included edits to the article and to the associated Talk page. As a result, sessions can be composed of edits to the article only, edits to the Talk page only or a mix of article and Talk edits. To test our hypotheses, we wanted counts of edits done, not just sessions. We therefore labelled each of the edits in a session the same as the session it is in (that is, an edit in a Talk-only session is labelled a Talk-only edit). As a result, there are 9 kinds of edits: good-faith major, good-faith minor or damaging that are article-only, Talk-only or article and Talk. We identify as possible cases of stigmergic collaboration edits made in an article by an editor who does not contribute to the discussion on the article Talk page during the same editing session. The opposite situation suggests that the editor collaborates by editing both the article and the Talk page in the same session.

Finally, data were aggregated per year as counts of the number of each kind of session and of each kind of edit for each article. The result of data processing was a data file with one line for each article for each year including the article quality at the start of the year and the count of different sessions and edits, 296 observations in total. Because articles were of different ages, there are different numbers of years represented for different articles.

3.3. Analysis

We tested the hypotheses by regressing article quality on the log-transformed counts of different kinds of edits (i.e., a subset of the variables in Table 1) made in the previous year (i.e., lagged by one year). The data form a panel, with multiple units (the articles) observed for multiple years. Panel data need special treatment in regression, as there can be both auto-correlation across time, heteroskedasticity (i.e., differences among units) and correlation across the units. Diagnostic tests showed that the data were both auto-correlated and heteroskedastic. We therefore performed the regression with panel-adjusted standard errors using the `xtpcse` command in Stata 15. As we had different numbers of years for different pages, observations were included pairwise, rather than examining only complete years. We note that some of independent variables are highly correlated, raising concerns about multi-colinearity. This problem is due in part to the use of panel data, which is addressed by the use of panel-adjusted standard errors. Further, the regression results suggest that multi-colinearity is not a problem. Multi-colinearity tends to inflate standard errors, making estimates non-significant, but the computed standard errors seem reasonable and the two most highly correlated variables are still both signficiant.

As is not unexpected, the count variables were quite skewed. Normality of dependent variables is not an assumption of regression, so strictly speaking, skewed data are not a problem for data analysis. However, we are using linear regression, and the skewed data suggest the possibility of a non-linear relationship. To make the relationships more linear and so more amenable to analysis, we log-transformed the count variables (more specifically, $log10(count + 1)$, since some counts were zero).

4. Findings

We start by describing our data set. Figure 1 presents descriptive statistics for the data while Figure 4 shows the correlations. Note that the "activities" variables are the sum of the edits of that sort, hence

Variable	Raw				Log transformed			
	Mean	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**
quality score	2.90	1.44	0.09	4.85				
sessions article-only	188.72	227.21	0	1269	1.880	0.706	0	3.104
total edits article-only	317.28	440.59	0	2842	2.033	0.765	0	3.454
major-good-faith article-only	145.82	226.92	0	1649	1.676	0.764	0	3.217
minor-good-faith article-only	91.60	123.12	0	786	1.554	0.702	0	2.896
damage article-only	79.86	128.59	0	667	1.303	0.824	0	2.825
sessions article and Talk	7.92	15.38	0	142	0.574	0.545	0	2.155
total edits article and Talk	55.11	151.04	0	1332	0.941	0.852	0	3.125
major-good-faith article and Talk	38.89	103.24	0	938	0.848	0.801	0	2.973
minor-good-faith article and Talk	9.64	37.29	0	350	0.420	0.569	0	2.545
damage article and Talk	6.58	25.63	0	368	0.348	0.540	0	2.567
sessions Talk-only	22.09	37.65	0	294	0.882	0.690	0	2.470
total edits Talk-only	31.71	64.12	0	605	0.958	0.743	0	2.782
major-good-faith Talk-only	21.91	45.80	0	457	0.841	0.693	0	2.661
minor-good-faith Talk-only	7.69	15.06	0	123	0.563	0.547	0	2.093
damage Talk-only	2.11	6.11	0	56	0.227	0.379	0	1.756

Figure 1. Descriptive statistics for raw and transformed data. Variables are measured per article per year.

the high correlations between these variables and their components.

Figure 6 shows for each article the plot over time of the quality level and log-transformed counts of different types of edits (on the right axis). The figure also shows

Figure 2. Counts of article-only, talk-only and article+talk edits for Belarus, an FA-quality article

Figure 3. Counts of article-only, talk-only and article+talk edits for Moncton, an B-quality article

the articles included in the sample. Note that in most cases, there are the most article-only edits and fewest talk-only edits. The relative counts of different classes of edits are shown for two representative articles in Figures 2 and 3. Finally, Figure 5 show the distribution of the fraction of edits of the three types across all projects. Note that article-only edits are the majority.

We next turn to the test of our hypotheses. The regression results are shown in Figure 7. The log-transformed counts of edits do a good job predicting the quality score: along with the auto-correlation, they predict 31% of the variance and the regression is significant ($R^2 = 0.315$, $\chi^2(9) = 141$, $p = 0.000$). The auto-correlation coefficient is $\rho = 0.502$.

As for the specific hypotheses, first, the regression shows that the counts of good-faith edits of articles (both major and minor) that are not accompanied by edits of the corresponding Talk page are predictive of higher article quality. For major edits, $\beta = 0.462$, $z = 2.57$, $p = 0.010$; for minor, $\beta = 0.839$, $z = 3.32$, $p = 0.001$. Accordingly, **H1** is supported. As the counts were log-transformed, the regression indicates that a doubling of the number of good-faith major edits (i.e., an increase of 0.3 on a log scale) predicts an increase in quality of about 1/8 of a quality point; and for minor edits, an increase of 1/4 point.

Second, the count of major good-faith edits that occur in a session that includes edits to both the article and the Talk page (an article-Talk session) is also predictive of higher article quality, but with a smaller beta weight, $\beta = 0.279$, $z = 2.70$, $p = 0.007$. A doubling of the number of these edits would increase
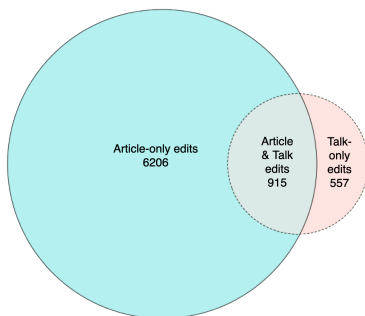
Variable	quality_score	session article-only	total edits article-only	major-good-faith article-only	minor-good-faith article-only	damage article-only	session article and Talk	total edits article and Talk	major-good-faith article and Talk	minor-good-faith article and Talk	damage article and Talk	session Talk-only	total edits Talk-only	major-good-faith Talk-only	minor-good-faith Talk-only	damage Talk-only
quality score	1															
session article-only	0.4375	1														
total edits article-only	0.3763	0.9783	1													
major-good-faith article-only	0.3334	0.882	0.9356	1												
minor-good-faith article-only	0.3813	0.9645	0.9662	0.8531	1											
damage article-only	0.3358	0.872	0.8502	0.6242	0.8476	1										
session article and Talk	0.2408	0.7619	0.8002	0.8297	0.7901	0.5211	1									
total edits article and Talk	0.1583	0.667	0.7302	0.7738	0.7207	0.4464	0.9249	1								
major-good-faith article and Talk	0.1766	0.6339	0.6972	0.7703	0.6656	0.3923	0.8916	0.9805	1							
minor-good-faith article and Talk	0.0731	0.5167	0.5585	0.5517	0.6035	0.3622	0.827	0.7956	0.6928	1						
damage article and Talk	0.1152	0.6254	0.6822	0.6546	0.6878	0.5238	0.6557	0.786	0.7418	0.4427	1					
session Talk-only	0.3828	0.6268	0.643	0.6823	0.658	0.369	0.7411	0.6646	0.669	0.5153	0.4718	1				
total edits Talk-only	0.3149	0.6017	0.6217	0.6722	0.6313	0.3394	0.7699	0.7169	0.7301	0.5408	0.4969	0.9592	1			
major-good-faith Talk-only	0.2897	0.6172	0.6369	0.6904	0.6381	0.3528	0.797	0.7496	0.7674	0.5511	0.5246	0.9142	0.9863	1		
minor-good-faith Talk-only	0.3492	0.5691	0.5851	0.6178	0.6097	0.3308	0.7067	0.6484	0.6495	0.5327	0.43	0.9643	0.9521	0.8937	1	
damage Talk-only	0.272	0.2853	0.3081	0.3564	0.3384	0.1025	0.3629	0.3056	0.3086	0.2307	0.222	0.8363	0.755	0.6515	0.8277	1

Figure 4. Correlation among variables

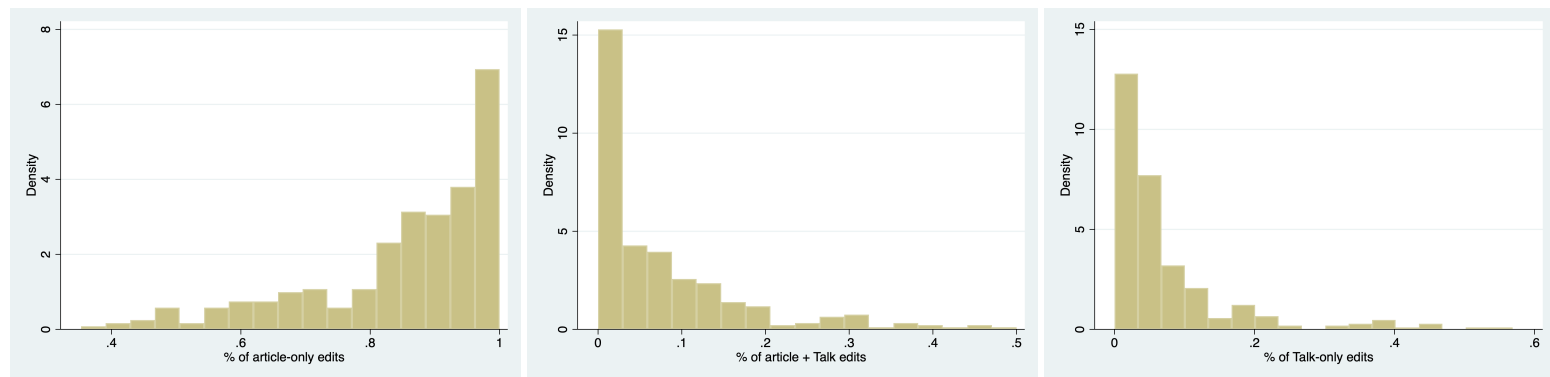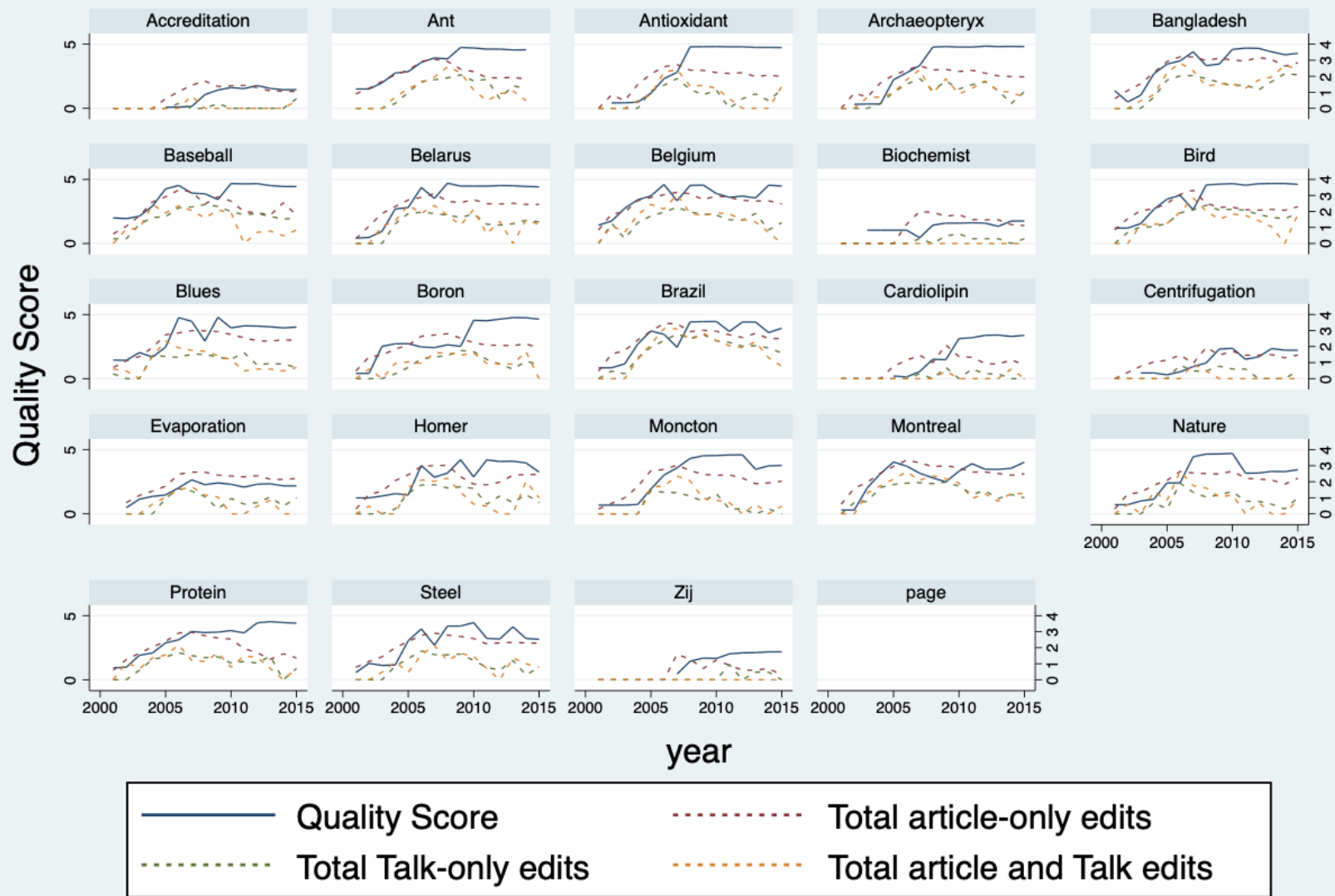**Figure 5. Histograms showing the distribution of article-only, article + Talk and Talk-only edits**

.

Figure 6. Quality level and number of different types of edits (log-transformed, right axis) over time per article

```
Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)
Group variable:   page_f                       Number of obs     =        296
Time variable:    year                         Number of groups  =         23
Panels:           correlated (unbalanced)      Obs per group:
Autocorrelation:  common AR(1)                                min =          8
Sigma computed by pairwise selection                         avg =  12.869565
                                                             max =         14
Estimated covariances      =        276        R-squared         =     0.3148
Estimated autocorrelations =          1        Wald chi2(9)      =     141.28
Estimated coefficients     =         10        Prob > chi2       =     0.0000
------------------------------------------------------------------------------
                           |            Panel-corrected
            quality_score |    Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------------+--------------------------------------------------
log_majorgoofdfaitharticle |
                       L1. |   .4620493   .1799076    2.57   0.010    .1094368    .8146618
                           |
 log_minorgoodfaitharticle |
                       L1. |   .8386761   .2527187    3.32   0.001    .3433564    1.333996
                           |
         log_damagearticle |
                       L1. |  -.1569096   .1313717   -1.19   0.232   -.4143934    .1005742
                           |
log_majorgoofdfaithartiTalk |
                       L1. |    .279232    .103406    2.70   0.007     .07656    .4819041
                           |
log_minorgoodfaithartiTalk |
                       L1. |  -.1974714   .1228967   -1.61   0.108   -.4383444    .0434017
                           |
        log_damageartiTalk |
                       L1. |  -.3092111   .1426771   -2.17   0.030   -.5888532   -.029569
                           |
log_majorgoofdfaithTalkonly |
                       L1. |   .1432899   .1635786    0.88   0.381   -.1773182    .4638981
                           |
log_minorgoodfaithTalkonly |
                       L1. |   .1339502   .1997313    0.67   0.502    -.257516    .5254165
                           |
        log_damageTalkonly |
                       L1. |  -.0649848   .1858443   -0.35   0.727   -.4292328    .2992633
                           |
                     _cons |      .9792   .2156469    4.54   0.000    .5565398    1.40186
---------------------------+--------------------------------------------------
                       rho |   .5019161
------------------------------------------------------------------------------
```

Figure 7. Results of panel corrected standard error regression. L1 means a one-period lagged variable. log_ are the log-transformed counts of edits.

quality only by 1/10 of a point. On the other hand, the number of minor-good faith edits that occur in an article-Talk session is not predictive of quality. In fact, the beta weight is negative, though not significant in the regression, $\beta = -0.197$, $z = -1.61$, $p = 0.108$. In other words, **H2** is only partly supported.

Third, the number of damaging edits in article-Talk sessions also predicts lower quality, $\beta = -0.309$, $z = -2.172$, $p = 0.030$. However, surprisingly the coefficient for the counts of other damaging edits (article-only and Talk-only) are not significant, meaning that **H3** is only partly supported.

Finally, none of the counts of Talk-only edits are significantly predictive of article quality. Accordingly **H4** is supported.

5. Discussion

As discussed above, article Talk pages have been identified as a venue in which Wikipedia editors coordinate the development of the article. We therefore interpret edits made to the article without edits to the corresponding Talk page as evidence of editing activity without explicit coordination. The regression results suggest that such stigmergic editing, major and especially minor, is quite predictive of higher article quality, emphasizing the importance of these independent contributions to articles. Nevertheless, the data also show that explicit coordination also plays an important role, but only for major edits. It may be that discussion of minor edits is an unnecessary distraction. The lack of impact of damaging edits to the article may be an indication that these edits are quickly fixed, and so do not affect quality, while damaging edits that occur in the context of explicit discussion pose more of a challenge. Finally, the lack of impact of Talk-only contributions supports our view of the action-oriented nature of Wikipedia editing: Talk without editing does not seem to improve quality, while editing without Talking clearly does. An interesting finding of this study is that stigmergic edits seem to be more predictive of quality than edits with explicit discussion.

A frequent concern in studies of editing and quality is the possibility of reverse causation, namely that good articles attract editors, rather than editors improving the quality of articles. The use of a panel design helps mitigate this concern, as we are examining quality after the edits, and allowing auto-correlation, meaning that we examining the effect of edits taking into consideration last year's quality as one of the predictors. In contrast though to previous findings, we find a positive coefficient, suggesting that quality is generally increasing over time.

6. Conclusion

In summary, the data provide evidence for the complementary roles of different modes of coordination in Wikipedia. Minor edits improve article quality seemingly without the need for discussion, while explicit coordination is needed for at least some major edits (though not all).

One implication of our results is that it may be helpful to further support the process of stigmergic coordination. A particular need is to make the goal of changes both visible and comprehensible. The Wikipedia infrastructure accomplishes the first through features such as the change log and watch lists. However, the only support for the second is the comment field on edits, which is not systematically used. Further, a single edit may mix multiple changes, e.g., adding a new section while also fixing typos or other problems. To improve understandability of edits, Wikipedia editors might want to consider the notion of an "atomic commit" from open source software development, the idea that a single save to the system should make one change. However, without system support to track different kinds of changes, the discipline needed to keep changes separate may be too taxing.

The work presented here offers several opportunities for improvement. First, we analyzed only a handful of articles; a larger dataset might make the results more precise. Second, the data used here are highly aggregated, simply annual counts of edits. It would be interesting to examine the impact on article quality while preserving more of the nuance of the data. A particular omission is consideration of how editors actually interact, which was the focus of prior research that examined stigmergic coordination [27]. Future work might examine these interaction in more detail to document the processes of coordination and their impact on article quality. Third, alternative analysis approaches could be considered, e.g., to address multi-colinearity. Finally, the literature review has identified a range of other factors that affect article quality that could be added to create a more complete model.

References

[1] O. Arazy and O. Nov, "Determinants of Wikipedia quality: The roles of global and local contribution inequality," in *Proceedings of the Conference on Computer-supported Cooperative Work*, pp. 233–236, 2010.

[2] O. Arazy, O. Nov, R. Patterson, and L. Yeo, "Information quality in Wikipedia: The effects of group composition and task conflict," *Journal of Management Information Systems*, vol. 27, no. 4, pp. 71–98, 2011.

[3] E. Yaari, S. Baruchson-Arbib, and J. Bar-Ilan, "Information quality assessment of community

generated content: A user study of Wikipedia," *Journal of Information Science*, vol. 37, no. 5, pp. 487–498, 2011.

[4] A. Halfaker, "Interpolating quality dynamics in Wikipedia and demonstrating the Keilana effect," in *Proceedings of the International Symposium on Open Collaboration*, p. 19, 2017.

[5] J. E. Blumenstock, "Size matters: word count as a measure of quality on Wikipedia," in *Proceedings of the International Conference on World Wide Web*, pp. 1095–1096, 2008.

[6] M. Warncke-Wang, D. Cosley, and J. Riedl, "Tell me more: An actionable quality model for Wikipedia," in *Proceedings of the International Symposium on Open Collaboration*, p. 8, 2013.

[7] D. Hasan Dalip, M. André Gonçalves, M. Cristo, and P. Calado, "Automatic quality assessment of content created collaboratively by web communities: A case study of Wikipedia," in *Proceedings of the Joint Conference on Digital Libraries*, JCDL '09, (New York, NY, USA), pp. 295–304, ACM, 2009.

[8] B. de La Robertie, Y. Pitarch, and O. Teste, "Measuring article quality in Wikipedia using the collaboration network," in *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, pp. 464–471, 2015.

[9] X. Li, J. Tang, T. Wang, Z. Luo, and M. de Rijke, "Automatically assessing Wikipedia article quality by exploiting article-editor networks," in *Advances in Information Retrieval* (A. Hanbury, G. Kazai, A. Rauber, and N. Fuhr, eds.), (Cham), pp. 574–580, Springer International Publishing, 2015.

[10] D. M. Wilkinson and B. A. Huberman, "Cooperation and quality in Wikipedia," in *Proceedings of the International Symposium on Wikis*, pp. 157–164, 2007.

[11] K. Carillo and C. Okoli, "Generating quality open content: A functional group perspective based on the time, interaction, and performance theory," *Information & Management*, vol. 48, no. 6, pp. 208–219, 2011.

[12] J. Liu and S. Ram, "Using big data and network analysis to understand Wikipedia article quality," *Data & Knowledge Engineering*, vol. 115, pp. 80–93, 2018.

[13] D. Yang, A. Halfaker, R. Kraut, and E. Hovy, "Who did what: Editor role identification in Wikipedia," in *Proceedings of the Conference on Web and Social Media*, 2016.

[14] J. Liu and S. Ram, "Who does what: Collaboration patterns in the Wikipedia and their impact on article quality," *ACM Transactions on Management Information Systems (TMIS)*, vol. 2, no. 2, p. 11, 2011.

[15] B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser, "Information quality work organization in Wikipedia," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 6, pp. 983–1001.

[16] H. T. Welser, D. Cosley, G. Kossinets, A. Lin, F. Dokshin, G. Gay, and M. Smith, "Finding social roles in Wikipedia," in *Proceedings of the iConference*, iConference '11, (New York, NY, USA), pp. 122–129, ACM, 2011.

[17] F. B. Viegas, M. Wattenberg, J. Kriss, and F. V. Ham, "Talk before you type: Coordination in Wikipedia," in *Proceedings of the Hawai'i International Conference on System Sciences*, pp. 78–78, Jan 2007.

[18] A. Kittur and R. E. Kraut, "Harnessing the wisdom of crowds in Wikipedia: Quality through coordination," in *Proceedings of the Conference on Computer Supported Cooperative Work*, CSCW '08, pp. 37–46, 2008.

[19] M. B. Watson-Manheim, K. M. Chudoba, and K. Crowston, "Perceived discontinuities and constructed continuities in virtual work," *Information Systems Journal*, vol. 22, no. 1, pp. 29–52, 2012.

[20] D. J. Armstrong and P. Cole, "Managing distance and differences in geographically distributed work groups," in *Distributed Work* (P. Hinds and S. Kiesler, eds.), pp. 167–186, Cambridge, MA: MIT Press, 2002.

[21] F. Bolici, J. Howison, and K. Crowston, "Stigmergic coordination in FLOSS development teams: Integrating explicit and implicit mechanisms," *Cognitive Systems Research*, vol. 38, pp. 14–22, 2016.

[22] L. R. Christensen, "Practices of stigmergy in architectural work," in *Proceedings of the Conference on Supporting Group Work*, GROUP '07, pp. 11–20, 2007.

[23] L. R. Christensen, "Practices of stigmergy in the building process," *Computer Supported Cooperative Work (CSCW)*, vol. 23, no. 1, pp. 1–19, 2014.

[24] J. Secretan, "Stigmergic dimensions of online creative interaction," *Cognitive Systems Research*, vol. 21, pp. 65–74, 2013.

[25] P.-P. Grassé, "La reconstruction du nid et les coordinations inter-individuelles chez Bellicositermes natalensis et Cubitermes sp. La théorie de la stigmergie: Essai d'interprétation du comportement des termites constructeurs," *Insectes sociaux*, vol. 6, no. 1, pp. 41–80, 1959.

[26] F. Heylighen, "Why is open access development so successful? stigmergic organization and the economics of information," in *Open Source Jahrbuch 2007* (B. Lutterbeck, M. Brwolff, and R. A. Gehring, eds.), Berlin: Lehmanns Media, 2007.

[27] A. Rezgui and K. Crowston, "Stigmergic coordination in Wikipedia," in *Proceedings of OpenSym*, 2018.

[28] R. S. Geiger and A. Halfaker, "Using edit sessions to measure participation in Wikipedia," in *Proceedings of the Conference on Computer-supported Cooperative Work*, CSCW '13, pp. 861–870, 2013.