DISA

Dipartimento di Informatica
e Studi Aziendali

2009/3

Coordination, Division of Labor, and Open Content Communities: Template Messages in Wiki-Based Collections

Loris Gaio, Matthijs den Besten, Alessandro Rossi, Jean-Michel Dalle

UNIVERSITÀ DEGLI STUDI
DI TRENTO

DISA Working Paper

DISA

Dipartimento di Informatica
e Studi Aziendali

2009/3

Coordination, Division of Labor, and Open Content Communities: Template Messages in Wiki-Based Collections

Loris Gaio, Matthijs den Besten, Alessandro Rossi, Jean-Michel Dalle

UNIVERSITÀ DEGLI STUDI
DI TRENTO

DISA WORKING PAPER

DISA Working Papers
The series of DISA Working Papers is published by the Department of Computer and Management Sciences (Dipartimento di Informatica e Studi Aziendali DISA) of the University of Trento, Italy.

Guidelines for authors
Papers may be written in English or Italian but authors should provide title, abstract, and keywords in both languages. Manuscripts should be submitted (in pdf format) by the corresponding author to the appropriate Associate Editor, who will ask a member of DISA for a short written review within two weeks. The revised version of the manuscript, together with the author's response to the reviewer, should again be sent to the Associate Editor for his consideration. Finally the Associate Editor sends all the material (original and final version, review and response, plus his own recommendation) to the Editor, who authorizes the publication and assigns it a serial number.

The Managing Editor and the Technical Officer ensure that all published papers are uploaded in the international RepEc public-action database. On the other hand, it is up to the corresponding author to make direct contact with the Departmental Secretary regarding the offprint order and the research fund which it should refer to.

Ricardo Alberto MARQUES PEREIRA
Dipartimento di Informatica e Studi Aziendali
Università degli Studi di Trento
Via Inama 5, TN 38100 Trento ITALIA
Tel +39-0461-882147 Fax +39-0461-882124
E-mail: ricalb.marper@unitn.it

Coordination, Division of Labor, and Open Content Communities: Template Messages in Wiki-Based Collections

Loris Gaio
University of Trento
Department of Computer and Management Science
Via Inama 5
I-38100, Trento, Italy
+39 0461 882142
lgaio@cs.unitn.it

Matthijs den Besten
Ecole Polytechnique
Centre de Recherche en Gestion
ENSTA PREG CRG 32
boulevard Victor
75739 Paris Cedex 15, France
+33 1 45 52 64 27
matthijs.denbesten@polytechnique.edu

Alessandro Rossi
University of Trento
Department of Computer and Management Science
Via Inama 5
I-38100, Trento, Italy
+39 0461 882101
arossi@cs.unitn.it

Jean-Michel Dalle
Université Pierre et Marie Curie
4 place Jussieu
75252 Paris cedex 05
France
+33 1 44 18 07 15
jean-michel.dalle@upmc.fr

Abstract

In this paper we investigate how in commons based peer production a large community of contributors coordinates its efforts towards the production of high quality open content. We carry out our empirical analysis at the level of articles and focus on the dynamics surrounding their production. That is, we focus on the continuous process of revision and update due to the spontaneous and largely uncoordinated sequence of contributions by a multiplicity of individuals. We argue that this loosely regulated process, according to which any user can make changes to any entry, while allowing highly creative contributions, has to come into terms with potential issues with respect to the quality and consistency of the output. In this respect, we focus on emergent, bottom up organizational practice arising within the Wikipedia community, namely the use of template messages, which seems to act as an effective and parsimonious coordination device in emphasizing quality concerns (in terms of accuracy, consistency, completeness, fragmentation, and so on) or in highlighting the existence of other particular issues which are to be addressed. We focus on the template "NPOV" which signals breaches on the fundamental policy of neutrality of Wikipedia articles and we show how and to what extent imposing such template on a page affects the production process and changes the nature and division of labor among participants. We find that intensity of editing increases immediately after the "NPOV" template appears. Moreover, articles that are treated most successfully, in the sense that "NPOV" disappears again relatively soon, are those articles which receive the attention of a limited group of editors. In this dimension at least the distribution of tasks in Wikipedia looks quite similar to what is know about the distribution in the FLOSS development process.

JEL Codes: M15 (IT Management), L86 (Information and Internet Services)

Keywords: Commons based peer production, Wikipedia, wiki, survival analysis, quality, bug fixing, template messages, coordination.

1. Introduction

The emergence of online communities has fostered a new interest in distributed problem solving. However, most of this interest has been focused on how to extract information or solutions from various and notably peripheral problem-solvers [1]. This approach, if albeit promising, still leaves aside the coordination efforts that are crucial to all production activities of online communities. Indeed, what is new and amazing with online peer production [2] stems not only from the attraction of motivated participants and of the solutions they are able to provide to distributed problems, but also from the 'spontaneous' coordination [3,4] of those efforts in order to give birth, for instance, to efficient software products composed of millions of lines of code or to a widely used encyclopaedia with millions of pages such as Wikipedia.

In this respect, recent works have suggested that signs that spontaneously tend to characterize ongoing collective endeavors and their organization could serve as *coordination signals* [5,6]. Moving one step further along this road, we are led to consider how online communities can consciously make use of such coordination mechanisms, in a sense adopting a managerial perspective. Wikipedia has precisely implemented such a mechanism through the use of template messages.

By studying how these tags can influence the coordination of work efforts within online communities, we could be able to gain further insights about how distributed problem solving works and can be made more efficient within online communities by *signaling problems* in a way that is adapted to the self-organized nature of those communities and more precisely to their stigmergic aspects. Essentially, the attention of potential contributors is attracted by these signs and their problem-solving efforts are therefore *oriented* in a direction that is particularly appropriate to collective peer production. Orienting efforts can typically be instrumental to help the community switch from an exploration toward an exploitation mode [7] or else, as we will suggest in this paper, to foster the resolution and settlement of online disputes.

We build from previous empirical research in the field, which has started to shed light on the role of institutions and organizational practices in channeling the largely unstructured efforts of voluntary contributors [12-13-19]. According to this line of research, peer production within wiki platforms makes extensive use of template messages – standard info-boxes placed on top of a given page – as coordination tool which ease the contribution to the production process of the various participants. In Wikipedia, for instance, there is an overwhelming number of templates, a.k.a. tags, which are used as a means to facilitate various goals and activities, such as to flag particular anomalies and dysfunctions of pages (e.g., violations of common policies or guidelines), and to ask for specific actions (e.g., cleaning, improvements in the organization of the text, and so on).

In what follows, we concentrate on template messages signaling breaches of important policies or guidelines – consensual standards and advisory statements which every editor should bear in mind when editing an article in the collection – and, drawing a parallel with open source software methodologies, we treat placing/removing such templates from the text of an article a process similar to filing /closing a bug report in software development.

Previous research on bug fixing in open source development [8-10-11] has shed light on various important organizational issues, e.g., which are the main drivers of coders' attention, which elements account for quick fixing of bugs, and so on. Conversely, the same topic has received much less coverage in the realm of open content peer production. We believe that the analysis of bug fixing activities is crucial in improving our understanding how and to what extent it is possible to reconcile the apparent contrast between spontaneous collaborative authorship and quality assurance of a wiki collection.

In particular, in this paper we aim at understanding through descriptive and survival analysis that are the variables in an article production process which account for the emergence of a bug or have influence on how the bug is fixed. In particular, we model the dynamic of tagging as a survival process, linking the probability of entry/exit of a page into the "pathological state" to various explanatory variables. According to this framework, we perform survival on the duration of the pathological state, exploring how different variables affects the treatment or the persistence of such undesirable features.

2. Data

Data source and data extraction. We retrieved the November 2006 .xml meta-history dump of the English version of Wikipedia, available at: http://downloads.wikimedia.org. We subsequently produced an .xml sub-archive made from all article pages tagged at least once in their lifetime with the {{NPOV}} template. There is a large family of template messages which are used to signal the breach of the neutrality policy in Wikipedia. Table 1 shows the frequencies of the various existing NPOV template messages. For a data consistency rationale we limited the analysis to strict {{NPOV}} templates (which accounts for around 80 per cent of all instances), while disregarding all remaining NPOV template messages (around 14 percent is related to NPOV messages place at section level and 6 per cent are represented of a large number of variations of marginal use).

In order to avoid some inconsistencies on the original .xml archive of Wikipedia (due to some older conversion scripts which have been in place until February 2002, some older articles have incomplete histories which suffer from missing initial revisions), we filtered out around 700 articles with starting date older than March, 1st 2002.

After this selection, we ended up with 6042 article pages for the analysis. While some studies on the English Wikipedia have shown that actual changes in a given article page are sometime the result of longer discussions occurring at the level of the corresponding talk page [18, 26], the use of talk pages as a means to anticipate and discuss actual changes is not investigated here and our analysis relies solely on data collected from article pages.

Table 1. The NPOV template message family

tag	1. # article pages	2. # article revisions
[{{NPOV}}]	6815	160772
[{{NPOV-section}}]	941	37452
[{{msg:NPOV}}]	196	3700
[{{Long NPOV}}]	143	5672
[{{SectNPOV}}]	134	8404
[{{sectNPOV}}]	106	6302
[other 129 tags]	260	37745
TOTAL	8595	260047

For every article page we extracted the following data at the single revision level: the user–id of the editor (IP address in case of anonymous edits), date and time of the edit, comments made by the editor and the full (wiki markup) text of the revision.

Preparation of the dataset for the analysis. De-wikification of the text and categorization of registered users (in terms of administrators, bots, registered and anonymous users) have been performed according to previous literature [12,13].

Finally, readability and similarity metrics were computed according to [13].

3. Methodology

Our main purpose is to characterize the existing differences in the production process of an article when a NPOV template is present or absent. Accordingly, we analyze the dynamics surrounding the birth of an article page, the emergence of neutrality concerns and their resolution.

We designate the period of time which goes from the article page inception to the appearance of the NPOV template as “regime 1” and we label as “regime 2” the subsequent period which subsists until the template is removed. As a matter of fact if one tracks down the appearance of the NPOV template in the revision history of an article page, one frequently observes repeated cycles of appearance-disappearance.

This dynamics is due to several reasons: first, it is not uncommon for the article page to experience problems of neutrality at different periods over time. Using a medical analogy, if we consider the act of tagging the page with the NPOV template as a marker of a dysfunction, editing out the template might correspond to an indefinite remission (a cure) or a temporary remission until the illness shows up again. A second reason according to which a NPOV might repeatedly be placed/removed in an article is due to disagreements between the Wikipedia editors over the NPOV status of the page. Finally, a third motive can be attributed to the effects of vandal edits that when substituting a non-negligible part of the of the article with other text might wipe out in the process also wiki-code present in the preamble (as in case the template messages). In this respect, both disputes and vandal edits produce as the result a process of cycling of the template that is repeatedly placed and removed in subsequent revisions of an article page.

Previous work has highlighted the short life span for vandal edits in wiki collections [24]. While this generally reassures us of the limited impact of these malicious activities on the quality of the whole archive, at the same time we still feel that when studying the process of development of articles one has to carefully evaluate whether vandal edits might introduce distortions in the interpretation of the data.

The above mentioned elements suggest that a careful operationalization of the "regime 2" is essential in order to inform the subsequent analyses. In this respect, taking as regime 2 the period which goes from the first appearance to the first removal of the NPOV template, might introduce a "shortening bias" due to vandal editing. Consequently, three alternative definitions of regime 2 have been implemented, which differs on how the regime end in computed:

"strict r2": the regime ends as soon as the NPOV is removed from the article page. In this case any vandal edit which wipes out the NPOV template or any dispute over the NPOV status result in the regime ending;

- "robust r2": regime 2 ends as soon as the removal of a NPOV template lasts at least one full day. This choice is consistent with the relatively shorter life time of a vandalism and also correct for fast-paced disputes over the NPOV status, while still considering slow-paced disagreements as partial remissions;

- "complete r2" consider the regime ended the last time in which the NPOV template is removed. In this case the regime cover the complete history of the dysfunction, including all remission periods.

In the following we report on three series of analysis: some descriptive statistics comparing the two regimes, an analysis of the speed of textual changes made in revisions of the two regimes and a survival analysis on regime 2, describing the covariates accounting for neutrality concerns resolution.

4. Results

4.1. Descriptive Statistics

Table 2 details the summary statistics for the duration of the articles in the sample computed on the whole lifetime and on regimes 1 and 2. Recalling that our sample is made by articles with starting date no earlier than March 1st, 2002 and the Wikipedia meta-history dump has information until November 2006, we observe that lifetime durations are well distributed over the entire possible span. There are two other prominent facts which arises from Table 1: usually an article develops neutrality issues in its maturity and the treatment for these issues is relatively fast (with a median value from 7.25 to 16.69 days according to the various definitions of r2). For all definitions of r2 it is clear that the distribution have a tail with some article durations inflating the mean values with respect to median ones.

Table 2. Summary statistics for durations (days).

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
all	2.19	397.80	700.40	747.30	1055.00	1710.00
r1	0.00	63.84	304.06	421.16	691.45	1691.72
strict r2	0.00	0.28	7.25	42.59	45.32	855.28
robust r2	0.00	0.51	10.03	48.66	53.62	884.32
complete r2	0.00	0.92	16.69	71.86	79.24	884.32

Table 3 offers some summary statistics computed over the whole lifetime and on regimes 1 and 2 on the number of revisions performed by human editors only (bot activity is not considered). Considerations similar to the one given for durations still holds here. In particular relatively few revision are usually required to fix the neutrality concerns for the majority of articles, while the tail also suggest that for a minority of them the process can take up a higher number of interventions.

Table 3. Summary statistics for revisions (human editors only).

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
all	2.00	32.00	90.00	292.50	275.00	15120.00
r1	0.00	10.00	31.00	109.90	97.00	7107.00
strict r2	1.00	1.00	3.00	10.23	9.00	977.00
robust r2	1.00	1.00	4.00	14.77	12.00	977.00
compl. r2	1.00	1.00	5.00	50.65	18.00	12350.00

In a similar vein, Table 4 offers some summary statistics on the number of human editors. A more interesting statistics in offered in Table 5, which is computed taking the revisions/editor ratio. The sensible difference between r1 and r2 (all three variants) is here represented by relative increase of participation compared to contribution in regime 2. This is particularly true of strict r2, suggesting that the phenomenon seems to be particularly active during the early stages after the emergence of the NPOV template.

Table 4. Summary statistics for number of editors (human editors only).

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
all	1.00	16.00	39.00	108.00	106.00	4411.00
r1	0.00	5.00	14.00	43.97	40.00	2418.00
strict r2	1.00	1.00	2.00	4.76	5.00	163.00
robust r2	1.00	1.00	3.00	6.27	6.00	281.00
compl. r2	1.00	1.00	3.00	16.62	9.00	2167.00

Table 5. Summary statistics for the ratio human revisions/human editors.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
all	1.00	1.63	2.10	2.62	2.86	82.60
r1	1.00	1.48	2.00	2.68	2.83	120.70
strict r2	1.00	1.00	1.00	1.71	1.81	29.00
robust r2	1.00	1.00	1.25	1.86	2.00	28.33
compl. r2	1.00	1.00	1.35	2.01	2.10	27.00

All statistics from the Tables 2-5 pool together data from articles for which at November 2006 regime 2 resulted still ongoing or already ended.

Table 6. Summary statistics for regime 2, distinguishing between articles still in regime 2 or not at the moment of data collection.

	Min.	1st Qu.	Median	Mean	3rd Qu	Max
	All					
duration	0.00	0.28	7.25	42.59	45.32	855.28
revisions	1.00	1.00	3.00	10.23	9.00	977.00
editors	1.00	1.00	2.00	4.76	5.00	163.00
revs/editors	1.00	1.00	1.00	1.71	1.81	29.00
	regime 2 ended only (around 90%)					
duration	0.00	0.16	4.54	33.33	32.33	832.97
revisions	1.00	1.00	3.00	9.61	8.00	977.00
editors	1.00	1.00	2.00	4.44	4.00	119.00
revs/editors	1.00	1.00	1.00	1.71	1.80	29.00
	regime 2 still ongoing only ((around 10%)					
duration	0.22	29.67	71.64	122.47	158.47	855.28
revisions	1.00	2.00	6.00	15.53	16.00	555.00
editors	1.00	2.00	4.00	7.55	9.00	163.00
revs/editors	1.00	1.00	1.33	1.69	1.85	17.34

Table 6 distinguishes between these two sets (analysis is limited to “strict r2”), and allow to observe that statistics for the articles for which regime 2 is ended are even more extreme with respect to regime 1. On the other side, statistics for articles still in regime 2 suggest that for a minority of entries (10%) the resolution seems to be much harder.

Finally, Table 7 and 8 present per time unit statistics obtained dividing, respectively, revisions and number of editors by the corresponding duration of the regime. Both metrics shows a considerable increase in participation per time unit to the editing process from regime 1 to regime 2. The shift is even larger for already ended regime 2 articles, while still ongoing regime 2 articles have measures of activities similar to regime 1.

Table 7. Summary statistics for revisions/days.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
all	0.00	0.06	0.15	0.45	0.38	130.40
r1	0.00	0.07	0.16	31.81	0.58	2979.00
strict r2 (all)	0.00	0.14	0.69	134.30	8.77	43480.00
strict r2 (ended)	0.00	0.19	1.02	149.90	13.24	43480.00
strict r2 (ongoing)	0.00	0.04	0.09	0.31	0.24	10.31

Table 8. Summary statistics for editors/days.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
all	0.01	0.05	0.08	0.16	0.15	23.36
R1	0.01	0.05	0.08	21.19	0.21	2882.01
strict r2 (all)	0.01	0.10	0.44	124.91	5.94	43480.01
strict r2 (ended)	0.01	0.13	0.62	139.41	9.12	43480.01
strict r2 (ongoing)	0.01	0.04	0.07	0.18	0.15	4.50

This similarity between the pattern of activity in ongoing articles and the pattern of activity in the default regime suggests that the effect of NPOV diminishes over time or that NPOV is ignored if it cannot be addressed in an obvious way.

Figure 1 shows the frequencies of wikipedians participation in the editing process for r1 and for strict r2 in terms of number of revisions and number of editors involved. Pie categories distinguish between administrators, bots, registered and anonymous users. The Figure suggests that in r2 registered users activity increases at the expenses of anonymous participations, both in terms of number of revisions and of number of editors.

Figure 1. Pie charts detailing the frequencies of participation of the various categories of editors in terms of number of revisions (upper side) and of number of editors (lower side) in r1 (left side) and in strict r2 (right side)

Finally, Table 9 summarizes some metrics which we take as proxies of quality of the textual entry.

Table 9. Statistics for quality proxies of the text at various points of the editing process.

	Min.	1st Qu.	Median	Mean	3rdQu.	Max
	Flesh readability index					
beg r1	20.02	30.93	38.74	39.83	47.41	77.40
beg r2	20.07	30.53	37.33	38.08	44.47	76.47
end strict r2	20.02	30.13	37.02	37.45	43.74	78.10
end comp. r2	20.04	30.34	37.06	37.48	43.74	78.10
	# internal references					
beg r1	0.00	2.00	11.00	20.54	23.00	993.00
beg r2	0.00	16.00	36.00	64.20	78.50	1697.00
end strict r2	0.00	20.00	40.00	69.72	83.00	1697.00
end comp. r2	0.00	21.00	43.00	75.83	92.25	1697.00
	# external references					
beg r1	0.00	0.00	0.00	1.92	2.00	140.00
beg r2	0.00	1.00	3.00	7.85	8.00	289.00
end strict r2	0.00	1.00	3.00	8.57	9.00	294.00
end comp. r2	0.00	1.00	4.00	10.32	10.00	366.00
	# endnotes					
beg r1	0.00	0.00	0.00	1.59	0.00	435.00
beg r2	0.00	0.00	0.00	4.90	3.00	1412.00
end strict r2	0.00	0.00	0.00	5.46	4.00	1412.00
end comp. r2	0.00	0.00	0.00	5.74	4.00	1412.00

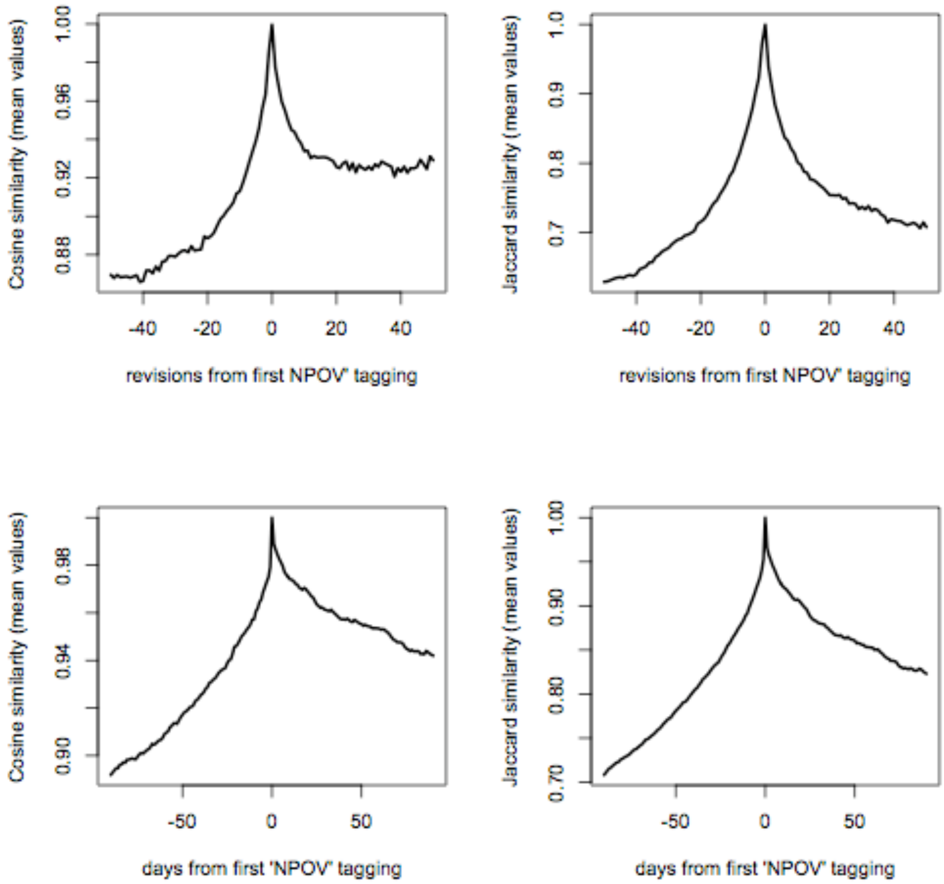
We measured readability (using the Flesh readability index) of the text at the first revision, at the beginning of r2 and at the end of r2 (measured according to the strict and the complete definition) as well as counting the presence of internal and external references and endnotes.

Table 9 suggests that the resolution of NPOV concerns does not come with any increases in readability and increases in all other item counts in very marginal, with the exception of the number of internal references.

4.2. Similarity analysis

Figure 2 makes use of some metrics taken from the computational linguistic field in order to measure to what extent the text of articles are updated over time before and after the emergence of the NPOV template.

Figure 2. Similarity plots (left side=Cosine metric, right side=Jaccard metric) centered over the first emergence of the NPOV template, computed over the revision domain (upper side) and the time domain (lower side).



In the plots, the *x* axis is centered over the revision which corresponds to the beginning of r2. Cosine and Jaccard similarity measures are then computed for all previous/subsequent revisions (upper side) and the revision at the beginning of r2; or in the time domain (lower side). Values are then averaged. Overall the plots show asymmetric speeds of change for the text of articles in r1 and r2. In particular in r2 (we linit our analysis to "strict r2"), it is possible to observe a considerable halt in textual changes starting from around the 10^{th} revision or after around 10 days. These in turns corresponds, respectively, to about the 3^{rd} quartile of the number of revisions in "strict r2" and about the 65 percentile of the of the durations of "strict r2", thus suggesting that the slowing down seems to be taking place for the minority of articles for which the resolution of NPOV concerns seems to be more problematic. This, combined with the observations previously made on Table 7 seems to suggest that the tail of "difficult NPOV articles" seems to gather less attention and activities from wikipedians, again hinting to the presence of an underlying principle of "economy of attention" at work behind the process and affecting the behaviors of participants.

4.3. Survival Analysis

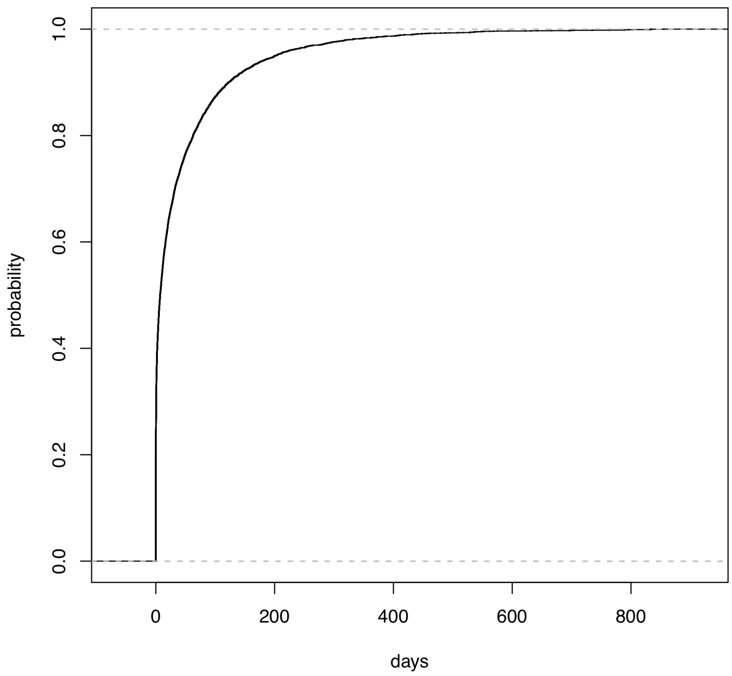
We applied survival analysis to study the dynamics of NPOV regime 2 remission, and to identify variables directly influencing the resolution of neutrality issues.

For any of the alternative regime 2 definitions (strict, complete, robust), some observations are right censored, meaning that in some cases the template has never been removed from the article page; in details, out of a sample of 6041 pages, a subset of 628, 823 and 727 are right censored (still open) respectively for the strict, complete and robust regimes.

A first inspection of survival durations allows us to speculate around the nature of distribution of regime times; in this respect, Figure 3, which depicts the empirical cumulate density function for durations computed according to "strict r2" rule, confirms the clue for a parametric distribution of regime durations.

A selection of distribution function for right censored durations has been performed, in order to spot the most appropriate model for a parametric survival analysis. For this purpose, different distributions have been used as candidate for maximum likelihood (MLE) fitting and parameter estimation.

Figure 3. Empirical CDF of durations for "strict r2" rule



Results of the procedure for identifying the most suitable distribution are summarized in Table 10.

Table 10. Distribution Fitting for three regimes

	estimates					
Model	**a**	**b**	**LL**	X_L	**p-value**	**AIC**
Strict Regime						
Exponential	43.39	—	-13084.8	3108.4	<.0001	13085
Gamma	0.474	43.39	-11530.6	665.8	<.0001	11532
Lognormal	2.420	2.646	-11197.7	166.6	<.0001	11199
Weibull	0.597	26.35	-11114.4	—	—	11116
Complete Regime						
Exponential	72.47	—	-12899.9	3522.8	<.0001	12900
Gamma	0.455	72.47	-11138.5	301.2	<.0001	11140
Lognormal	2.868	3.187	-10987.9	301.6	<.0001	10989
Weibull	0.580	43.83	-10837.1	—	—	10839
Robust Regime						
Exponential	51.09	—	-12702.7	2968.2	<.0001	12703
Gamma	0.481	51.09	-11218.6	446.6	<.0001	11220
Lognormal	2.605	2.767	-10995.3	255.4	<.0001	10987
Weibull	0.601	31.47	-10867.6	—	—	10869

In order to identify the most promising candidate, alternative distributions have been sorted according to AIC measure (Akaike, 1974) and the *XL* statistics, that is 2*(LL_i-LL_{i+1}) have been computed to test the null hypothesis *H*0 that data follows a given distribution, versus the alternative *H*1 that the underlining distribution follows the next candidate; then *HL* is compared to *p*-value for the usual chi-square.

The final outcome of this procedure permits us to identify the Weibull (extreme value) distribution is the best choice for data fitting for all regime alternatives, and thus it seems to represent the best candidate for three regimes for the estimation of parametric survival models.

Subsequently, to fit survival models a set of accelerated failure time models (AFT) has been used, in order to appraise the effect of covariates on survival time. An AFT model assumes that the effect of a covariate is to multiply the predicted event time by some constant, acting multiplicatively on the failure time by a scale factor. The effect of a predictor (covariate) is to alter the rate at which a page proceeds along the time axis (i.e., to accelerate the time to failure).

In order to perform the survival analysis the following variables to be used as covariates in parametric regression models have been computed:

1. revsAdm{R1,R2}: average number of edits by administrator in regime 1 and 2;
2. revsAno{R1,R2}: average number of revisions edited by anonymous contributors;
3. revsReg{R1,R2}: average number of edits, for registered contributors.
4. singleReg{R1,R2}: total number of registered contributors in regime 1 and 2;
5. shareAdm{R1,R2}: share of administrators respect to registered contributors;

6. shareAno{R1,R2}: the same ratio, for anonymous versus registered contributors;
7. delta readability: difference in readability measure from the first revision tagged NPOV until first succeeding edit not tagged;
8. tag/untag: a dummy variable showing that the user who put the tag also removed it from the page.

Consequently, a set of parametric AFT casual models has been fitted in order to explain the impact of the covariates depicted above, for the three alternative definitions of regime 2.

Table 11. Regression parameters and diagnostics.

Model	Strict	Complete	Robust
(intercept)	1.7539*** (0.0796)	3.6964*** (0.1080)	2.2499*** (0.0775)
revsAdm R1	-0.0781*** (0.0082)	-0.0940*** (0.0113)	-0.0875*** (0.0092)
revsAno R1	-0.0101*** (0.0036)	-0.0141*** (0.0048)	-0.0117*** (0.0037)
revsReg R1	-0.0414*** (0.0050)	-0.0167** (0.0068)	-0.0332*** (0.0052)
singleReg R1	-0.0148*** (0.0005)	-0.0125*** (0.0006)	-0.0141*** (0.0005)
shareAdm R1	0.1309** (0.0794)	0.1877* (0.0854)	0.1928** (0.0784)
shareAno R1	-0.3070*** (0. 0398)	-0.3142*** (0.0437)	-0.2928*** (0.0389)
revsAdm R2	0.0683*** (0.0239)	0.1501*** (0.0244)	0.0947*** (0.0220)
revsAno R2	0.1213*** (0.0327)	0.2678*** (0.0343)	0.0791*** (0.0258)
revsReg R2	0.0626*** (0.0180)	0.0780*** (0.0160)	0.0332** (0.0144)
singleReg R2	0.4339*** (0.0167)	0.0341*** (0.0032)	0.3006*** (0.0126)
shareAdm R2	0.8335*** (0.0932)	0.1151*** (0.0960)	0.5491*** (0.0887)
shareAno R2	0.5327*** (0.0616)	0.4971*** (0.0669)	0.4594*** (0.0577)
delta readability	-0.0430*** (0.0017)	-0.0048** (0.0019)	-0.0463*** (0.0017)
tag/untag	-0.8099*** (0.0686)	-1.3148*** (0.0701)	-0.8020*** (0.0676)
Log(scale)	0.8837 (0.0110)	0.8980 (0.0116)	0.8535 (0.0112)
Log likelihood	-18714	-22015	-19942
Likelihood Ratio	2739	1468.6	2597.3

Significance levels: *** = 0.01; ** = 0.05; * = 0.1

For sake of compactness we present the coefficient estimates and their standard errors in parentheses, along with some diagnostics, which are summarized in Table 11 for all three alternatives.

The model diagnostics (LL, LR) show that strict and robust models fits better to data than the complete alternative. Moreover, the scale effect is quite unimportant, being its log near to one.

The vast majority of covariates are significant at the 1% level, for all regressions. As for casual relations, the hypothesis of direct relationship between effort in regime 2 and remission time is confirmed, since all parameters regarding average revisions and number of single contributors are positive.

On the contrary, efforts and editors prior to NPOV tagging bring a negative role in regime termination, apart from administrators who seem to play a quite different role. Overall, the models seem to suggest that both the level of effort on a page (in terms number of edits) and the number of participants in the editing process seem to anticipate the emergence of concerns. At this point of the analysis it is still difficult to judge whether this shortening is more due to a variant of the Linus' law (more eyeballs resulting in the anticipatory detection of a defect) or rather due to diminishing returns related with increases in the number of contributors.

The positive effect of the relative number of edits by administrators upon regime duration could suggest that high disputes inside this class of user could signal a potential problem in the neutrality of the page long before its tagging, and could pose a serious friction in the process of regime termination.

Finally, both the difference in readability measures along the regime, and the dummy variable related to tag marking/removing are negatively correlated with the duration of the regime. The first covariate sheds a light on some particular way of resolving disputes, arising for those pages which problems limit to linguistic or composition issues, such as ambiguities or other misfits, which can be quickly solved by writing means.

The latter one underlines the importance of those contributors who are systematically involved in the dynamics of page revisions, and directly supervise the process of distributed knowledge creation by using the tagging/untagging mechanism in a deliberate way.

5. Conclusion

The success that Wikipedia has had in compiling a large quantity of articles on a wide variety of topics is a great example of online peer production and it shows that the coordination of large distributed efforts can be feasible. However, for all the praise Wikipedia receives, there is comparatively little information on the mechanisms that it has adopted to make it all work (but see [29]). We believe that the practice of adding banner to pages, which are standardised by selecting from a variety of template messages, constitutes a key mechanism for coordination in Wikipedia. For, often template messages contain very specific information about what the person who added the message likes or dislikes about a page.

The adoption of templates thus results in the identification of often recurring problems within articles. Furthermore, the placement of these template messages creates an explicit task, which everyone who notices it is invited to address.

Among the goals that Wikipedia sets itself, the preservation of a "neutral point of view" is probably one of the most daunting. We have made a first attempt to empirically describe how the use of templates to voice complains about lack of neutrality by means of banner message on article pages helps achieve this goal. In particular, we have focused on the template that seems to be most widely adopted for this purpose, NPOV.

Our findings suggest that "NPOV" does indeed serve as a signal to initiate a set of actions. Crowston and others found that the most common form of task-assignment in open source software development teams is self-assignment [4]. In the context of our investigations, it is not very easy to distinguish self-assignments from other assignments among the NPOV-tasks set by the NPOV-banners. Still, we have been able to discover differences in the dynamics of the groups of editors that deal with a page, which may be associated with the benefits and drawbacks of self-assignment. To begin with, we found that the label "NPOV" tends stick longer, that is, the neutrality problem remains unresolved, for articles where many people contribute. Yet, articles that have been edited by just a few people before the "NPOV" first appeared also tend to take longer to be untagged – reflecting, perhaps, a tendency of people to guard heavily the content of text for which they feel a sense of ownership. Then again, maintaining a sense of ownership also has positive effects at times as we also found that an article that was tagged and untagged by the same person tended to be untagged faster.

Of course, a major drawback of our findings so far is that we haven't controlled for the level of exposure of the article, nor did we control for the inherent complexity of maintaining neutrality that may be different from topic to topic. Besides, so far we have had little success in characterizing the actions triggered by NPOV themselves. A previous study on the "unsimple" template in Simple Wikipedia has found out a clear effect both in terms of the type of contributions in terms of size and their effect in terms of readability. Conversely, here the issue of neutrality has less obvious effects on surface characteristics of articles and the potentially much greater diversity among Wikipedia editors as compared to the small group of Simple Wikipedia afficianados are elements complicating our picture which surely is worth of further scrutiny.

References

[1] K. R. Lakhani, L. B. Jeppesen, P. A. Lohse, and J. A. Panetta. The value of openness in scientific problem solving. Harvard Business School, Working paper, 2007.

[2] Y. Benkler. The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press, New Haven, Connecticut, 2006.

[3] T. W. Malone and K. Crowston. The interdisciplinary study of coordination. *ACM Computing Surveys*, 26(1):87–119, 1994.

[4] K. Crowston, K. Wei, Q. Li, U. Y. Eseryel, and J. Howison. Coordination of free/libre open source software development. In *Twenty-Sixth International Conference on Information Systems*, 2005.

[5] J.-M. Dalle and P. David. Simulating code growth in libre (open-source) mode. In N. Curien and E. Brousseau, editors, *The Economics of the Internet*. Cambridge University Press, Cambridge, UK, 2007.

[6] M. den Besten, J.-M. Dalle, and F. Galia. The allocation of collaborative efforts in open-source software. *Information Economics and Policy*, 20(4):316–322, 2008.

[7] A. Rossi, L. Gaio, J.-M. Dalle, and M. den Besten. Novelty creation within organizations: the case of collaborative/user generated content. In *Academy of Management*, Annaheim, CA, 2008.

[8] Crowston, K. and Scozzi, B. 2004. Coordination practices for bug fixing within FLOSS development teams. In: Proc. 1st CSAC. Porto, Portugal.

[9] Cuozzo, S. 2008. Urban Myths: What Wikipedia gets wrong about NYC. New York Post. http://www.nypost.com/seven/08242008/postopinion/opedcolumnists/urban_myths_125773.htm?page=0. Accessed 21 January 2009.

[10] Dalle, J.-M. and den Besten, M. 2007. Different bug fixing regimes? A preliminary case for superbugs. In: Proc of the 3rd Int Conf on Open Source Systems. Limerick, Ireland.

[11] Dalle, J.-M., den Besten and M., Masmoudi, H. 2008. Channelling Firefox developers: Mom and dad aren't happy yet. In: Open Source Development, Communities and Quality, B. Russo, E. Damiani, S. Hissam, B. Lundell, G. Succi, Eds., IFIP International Federation for Information Processing;275. Boston, Springer:265–271.

[12] den Besten and M., Dalle, J.-M. 2008. Keep it Simple: A Companion for Simple Wikipedia? Ind. Innovat. 15(2):169–178.

[13] den Besten, M., Rossi, A., Gaio, L., Loubser, M. and Dalle, J.-M. 2008. Mining for practices in community collections: finds from Simple Wikipedia. In: Open Source Development, Communities and Quality, B. Russo, E. Damiani, S. Hissam, B. Lundell, G. Succi, Eds., IFIP International Federation for Information Processing;275. Boston, Springer:105–120.

[14] Denning, P., Horning, J., Parnas, D. and Weinstein, L. 2005. Wikipedia risks. CACM 12(48):152.

[15] Giles, J. 2005. Internet encyclopaedias go head to head. Nat. 438(7070):900–901.

[16] Ito, J. 2004. Wikipedia attacked by ignorant reporter. Joi Ito's Web. http://joi.ito.com/weblog/2004/08/29/wikipedia-attac.html. Accessed 3 February 2009.

[17] Keen, A. 2007. The Cult of the Amateur: How Today's Internet is Killing Our Culture. Doubleday Business, New York.

[18] Kittur, A. and Kraut, R.E. 2008. Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination. CSCW 2008: Proceedings of the ACM Conference on Computer-Supported Cooperative Work. New York: ACM Press.

[19] Loubser, M. and den Besten, M. 2008 Wikipedia Admins and Templates: The Organizational Capabilities of a Peer Production Effort. SSRN. http://ssrn.com/abstract=1116171. Accessed 1 December 2008.

[20] J. Mateos Garcia and W. E. Steinmueller. Applying the open source development model to knowledge work. INK Open Source Research Working Paper 2, SPRU - Science and Technolgy Policy Research, University of Sussex, UK, January 2003.

[21] Potthast, M., Stein, B. and Gerling, R. 2008. Automatic Vandalism Detection in Wikipedia. In: Advances in Information Retrieval, C. Macdonald, I. Ounis, V. Plachouras, I. Ruthven, R.W. White, Eds., Boston, Springer:663–668.

[22] Sanger, L. 2001. Wikipedia is wide open. Why is it growing so fast? Why isn't it full of nonsense?. Kuro5hin. http://www.kuro5hin.org/story/2001/9/24/43858/2479. Accessed 10 November 2008.

[23] Wikipedia:Simple English Wikipedia. 2008. In: Simple Wikipedia, the free encyclopedia. http://simple.wikipedia.org/wiki/Wikipedia:Simple_English_Wikipedia. Accessed 10 November 2008.

[24] Wikipedia:How to write Simple English articles. 2008. In: Simple Wikipedia, the free encyclopedia.

http://simple.wikipedia.org/wiki/Wikipedia:How_to_write_Simple_English_articles. Accessed 10 November 2008.

[25] Viégas, F., Wattenberg, M. and Dave, K. 2004. Studying cooperation and conflict between authors with history flow visualizations. In: Proc of the SIGCHI Conf on Human factors in computing systems 2004, New York, ACM Press: 575–582.

[26] Viégas, F., Wattenberg, M., Kriss, F. and van Ham, F. 2007. Talk before you Type: Coordination in Wikipedia. In: 40th Annual Hawaii International Conference on System Sciences (HICSS'07).

[27] H. Aikake. A new look at statistical model identification. *IEEE Transactions on Automatic control*, 1974.

[28] Collett, D. 2003, *Modelling Survival Data in Medical Research* (2nd ed.), CRC press.

[29] 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*, 2008.