Does It Matter Who Contributes? – A Study on Featured Articles in the German Wikipedia

Klaus Stein   and   Claudia Hess

Laboratory for Semantic Information Technology
University of Bamberg

{klaus.stein,claudia.hess}@wiai.uni-bamberg.de

ABSTRACT

The considerable high quality of Wikipedia articles is often accredited to the large number of users who contribute to Wikipedia's encyclopedia articles, who watch articles and correct errors immediately. In this paper, we are in particular interested in a certain type of Wikipedia articles, namely, the featured articles – articles marked by a community's vote as being of outstanding quality. The German Wikipedia has the nice property that it has two types of featured articles: *excellent* and *worth reading*. We explore on the German Wikipedia whether only the mere number of contributors makes the difference or whether the high quality of featured articles results from having experienced authors contributing with a reputation for high quality contributions. Our results indicate that it *does* matter *who* contributes.

Categories and Subject Descriptors:
    H.m [Miscellaneous]

General Terms: Human Factors

Keywords: wiki, Wikipedia, measures of quality and reputation, statistical analysis of Wikipedia, collaborative working

1. INTRODUCTION

The online encyclopedia Wikipedia is the most prominent example of a collaborative hypertext authoring environment. It is characterized by a particular interaction paradigm, namely, that everybody can contribute, even without previously registering. As no one is personally responsible for a certain article, the question arises: are wiki articles of a high quality? In general, the quality is quite high. Giles [4] finds in his comparison of Wikipedia and Encyclopedia Britannica the analyzed articles to be of equal quality. This is often attributed to Wikipedia's "self-healing" effect: the large number of contributors who constantly care for articles detect errors very fast and correct them immediately. Wikipedia uses the reviewing capabilities of its contributors

for identifying articles as featured articles, i. e., as articles of an outstanding accuracy and quality. Articles are marked as candidates and after a lead time for further improvement the community of Wikipedia contributors decides in a peer-reviewing process and a vote on the future status of an article as featured article [9]. Typically a large number of users is interested in the candidate article and participates in the discussions. In this paper, we are interested in the contributors to such potentially featured articles. The question is: Is it sufficient to have many people working on an article or does it matter who contributes, e.g., authors with a reputation for high quality writing?

2. QUALITY AND REPUTATION

The best articles in Wikipedia are marked as featured articles. They are promoted on the main page of Wikipedia. The German Wikipedia has two types of featured articles: *excellent* (in German exzellent) and *worth reading* (in German lesenswert). The label '*excellent*' is the highest award and given only to articles of outstanding quality. Articles labeled as *worth reading* may not yet be complete with respect to some minor topics. Rateike et al. [7] addressed the question whether there is any formal pattern that featured articles follow. And if so, is it sufficient to apply this pattern in order to be promoted to featured article status, regardless of the content? Having compared featured articles with respect to length, number of external links, layout, content embedded, and the bibliography, they concluded that there is no such pattern.

Instead of looking at the formal characteristics of featured articles, we can directly look at the contributors. For instance, do featured articles differ from normal articles with respect to the number of contributors? According to Braendle [2], the more relevant a topic, hence the more people are willing to contribute, the more likely it is to have a high quality article. However, the investigation by Rateike et al. [7] showed that featured articles are not necessarily written by a huge number of people, but that it is important that some contributor(s) feels personally responsible for the article.

The number of past contributions is often considered as a measure of reputation of some author [3]. More elaborated measures of reputation consider the quality of the articles that the user contributes to [6]. Korfiatis et al. [5] introduce a measure that is based on the succeeding edits by other users: subsequent edits and roll-backs to earlier versions are considered as disapproval whereas keeping the edits is taken as approval. Adler and de Alfaro [1] refine this approach.

3. STUDY

The following research is done on an XML-dump of the German Wikipedia metadata from 2007-04-02. It contains 976,016 regular articles, from which 945,520 pages have at least one author. Category "exzellent" (*excellent*) has 1032 pages, "lesenswert" (*worth reading*) 1889 pages (we skipped one doublette). We refer to the remaining pages as *others*. All wiki pages are collaboratively written by 137,238 known (i. e. non-anonymous) authors. 14,494 (11%) edited at least one *excellent* and 21867 (16%) edited at least one *worth reading* article, 109971 (82%) edited only other articles. Please note that with the term "edited an *excellent* article" we do *not* mean that the article was in the *excellent* category at the time the editing was done but it has this status now. This applies to all descriptions in the following sections. As we need distinct and identifiable authors, we excluded all anonymous edits from all pages, so the numbers in the following paragraphs count only non-anonymous authors. This is important for the interpretation of the given tables: an average of 15 edits per page is 15 edits by logged-in authors. All anonymous edits are ignored.

Obviously, featured pages receive far more edits by far more different authors than others (Table 1). This is due to the process of awarding the labels 'excellent' and 'worth reading'. The candidates are promoted on special pages, and therefore, a larger number of users gets attracted to these pages and, consequently, contributes to them. Moreover, some users may likely feel responsible for these pages to become featured pages and may improve them by contributing with many edits.

We therefore ask whether the high quality of the featured articles results from the large number of contributors or whether it is important that experienced users with a high responsibility for quality participate. We compute the reputation of authors based on their contribution to *excellent* pages. We then compute the rating of a page based on the reputation of the contributing authors. If this author reputation has a direct influence on the quality of the articles, then the articles labeled as *worth reading* should have a higher rating than the *other* articles.

We use two approaches to measure an author's contribution: page-based and edit-based. The **page-based author contribution** (pb) measures an author's contribution based on the number of pages that he/she edited (regardless of the number of edits). The **edit-based author contribution** (eb) measures an author's contribution based on the number of edits that he/she made whereby each revision counts, i. e. 5 edits of the same page count 5. We define the page-based and edit-based reputation of an author a:

$$\text{rep}_{\text{pb}}(a) = \frac{\# \text{ of edits of } a \text{ on } excellent \text{ pages}}{\text{total } \# \text{ of edits by } a},$$

$$\text{rep}_{\text{eb}}(a) = \frac{\# \text{ of } excellent \text{ pages } a \text{ edited}}{\text{total } \# \text{ of pages } a \text{ edited}}.$$

The reputation of an author is determined by the percentage of *excellent* pages she edited (pb), or edits on *excellent* pages (eb), respectively. We do not claim that $\text{rep}_{\text{pb}}(a)$ or $\text{rep}_{\text{eb}}(a)$ correspond to the social reputation an author gains in the Wikipedia community. We choose these measures because they are appropriate for this study. Measures that take the evolution of the content into account, such as the measure by [1], would be better suited for measuring social reputation. The measure used in this study would be vulnerable to attacks in the sense that a user can inflate the personal reputation by making many small edits on excellent pages. Here, we consider past interactions on a fixed dump. Users would not have any motivation to manipulate the reputation measure used.

The rating of a page p is computed based on the contributing authors' reputation. It is the average of the contribution-based reputations of its authors, either on a per-author-basis (ab) or on a per-edit-basis (eb):

$$\text{rating}_{\text{ab}}(p) = \frac{\sum_{a \in \text{authors}(p)} \text{rep}_{\text{pb}}(a)}{|\text{authors}(p)|},$$

$$\text{rating}_{\text{eb}}(p) = \frac{\sum_{e \in \text{edits}(p)} \text{rep}_{\text{eb}}(\text{author}(e))}{|\text{edits}(p)|}.$$

3.1 Results

The average ratings of pages of the three categories *excellent*, *worth reading* and *others* are shown in table 2. As expected, *excellent* pages gain highest ratings by this measure. The interesting observation is that the *worth reading* pages get a much higher rating than pages in the category *others*: authors who contribute to *excellent* pages largely also contribute with a higher percentage to *worth reading* pages than to *others*.

Candidates for the categories *worth reading* and *excellent* gain more attention than *other* pages. As it can be seen in table 1, the number of edits is considerably higher and a larger number of authors contributes.[1] So one could claim that if a page is nominated for promotion to *excellent* or *worth reading* status, a certain group of authors drops in and does a lot of editing. This would mean that some authors concentrate on pages that are candidates for featured articles. This would explain the results shown in table 2.

To analyze whether this really holds true, we pruned the edit history of all pages to a maximum of the first 14 edits (14 because this is the average number of edits for *other* pages). So here only the first 14 (non-anonymous) contributions are counted, and the (hypothetic) group of users who aim to improve a candidate page during the reviewing process before obtaining the label *worth reading* or *excellent* is cut away. Table 3 and 4 show that *worth reading* pages keep their higher ranking compared to *others*. The positive effect on the ratings of *worth reading* articles holds even with a totally pruned revision history which considers only the first non-anonymous author editing a page (Table 5 and 6).

Table 7 gives the average edit-based and page-based authors' reputation on the full and the pruned history. The edit-based and page-based reputation is highly correlated despite the very different edit rates of *excellent* to *other* articles.[2] So both measures give a similar author reputation ranking with low variance. The absolute reputation values decrease when the pruned history is taken, because the high impact of the large number of edits on featured articles is discarded.

---

[1]We cannot clearly determine whether a page receives many edits before being proposed to obtain the label *worth reading* or *excellent* or afterwards, because the date of announcement is not provided in the metadata.

[2]For history length of 1 there is no difference between both measures.

Category	1) $|\operatorname{edits}(p)|$	$|\operatorname{authors}(p)|$	2) $\overline{\operatorname{rating}_{\mathrm{eb}}^{*/*}}(p)$	$\overline{\operatorname{rating}_{\mathrm{ab}}^{*/*}}(p)$	3) $\overline{\operatorname{rating}_{\mathrm{eb}}^{14/*}}(p)$	$\overline{\operatorname{rating}_{\mathrm{ab}}^{14/*}}(p)$
excellent	304.4	84.3	0.1120	0.0421	0.0890	0.0328
worth reading	252.1	76.8	0.0486	0.0182	0.0416	0.0181
others	14.2	7.9	0.0174	0.0095	0.0177	0.0097
all	15.0	8.1	0.0176	0.0096	0.0178	0.0097

Category	4) $\overline{\operatorname{rating}_{\mathrm{eb}}^{14/14}}(p)$	$\overline{\operatorname{rating}_{\mathrm{ab}}^{14/14}}(p)$	5) $\overline{\operatorname{rating}_{\mathrm{eb}}^{1/*}}(p)$	$\overline{\operatorname{rating}_{\mathrm{ab}}^{1/*}}(p)$	6) $\overline{\operatorname{rating}_{\mathrm{eb}}^{1/1}}(p)$	$\overline{\operatorname{rating}_{\mathrm{ab}}^{1/1}}(p)$
excellent	0.0360	0.0226	0.0897	0.0404	0.0451	0.0451
worth reading	0.0093	0.0062	0.0409	0.0197	0.0077	0.0077
others	0.0019	0.0013	0.0181	0.0099	0.0010	0.0010
all	0.0019	0.0014	0.0182	0.0100	0.0011	0.0011

Table 1: Average number of edits/authors per page.
Table 2–6: Average page ratings.
$\overline{\operatorname{rating}^{m/n}}(p)$ **gives the average rating of the pages. The page rating is based on the first (non-anonymous)** m **edits of the page and the authors' is based on the first** n **edits on all pages.** $* =$ **all edits.**

$\overline{\operatorname{rep}_{\mathrm{eb}}}(a)$	$\overline{\operatorname{rep}_{\mathrm{pb}}}(a)$	$\overline{\operatorname{rep}_{\mathrm{eb}}^{14}}(a)$	$\overline{\operatorname{rep}_{\mathrm{pb}}^{14}}(a)$	$\overline{\operatorname{rep}_{\mathrm{eb}}^{1}}(a)$	$\overline{\operatorname{rep}_{\mathrm{pb}}^{1}}(a)$
0.01454	0.01435	0.00131	0.00127	0.00076	0.00076
$r = 0.9674$		$r = 0.9702$		$r = 1.0$	

Table 7: average authors' reputation based on all, the first 14 and the first edit of each page. The third row gives the correlation coefficient between the edit- and page-based reputation of each author.

4. DISCUSSION

4.1 It matters who contributes

The results shown in section 3.1 indicate that there is a relation between the quality of an article and its authors: *worth reading* articles have a higher rating than *other* articles, i.e. the non-featured articles. This means that the average reputation of the authors who edit *worth reading* articles is higher than the reputation of authors contributing to *other* articles; or the other way around: the authors who wrote the *excellent* articles wrote the *worth reading* ones, too. This result holds true for the page-based author reputation, which takes into account only the number of articles of each category edited by the author, as well as for the edit-based author reputation, where the amount of contribution (number of edits) to the articles of each category matters. The effect is larger for the edit-based measure but evident in both.

4.2 Are there "featured article specialists"?

Do some authors write better articles than others and these articles simply get awarded *worth reading* or *excellent*? This is a possible explanation, but there are several other effects. As mentioned above, there could be a group of specialists which is only interested in improving the quality of articles to become *worth reading* or *excellent*. Those authors would gain high reputation by our measure from the *excellent* articles and this high reputation would increase the rating of the articles labeled as *worth reading*. We know that there are much more edits on *excellent* (average of 304 edits) and *worth reading* (252 edits) articles than on *others* (14 edits). The featured article specialists would drop in when an article is proposed as a candidate for *excellent* or *worth reading*, and do a lot of editing to improve it.

Under this assumption, there should be a difference be-

tween the authors starting to write an article and the authors improving it later on when nominated as featured article. While the first few authors should have average reputation (as they edit various articles), the featured article specialist should have gained a high reputation based on their edits of *excellent* candidates which acquire *excellent* status afterwards. This difference does not show up in our data. By pruning the edit history of each page to the first 14 edits, the featured article specialists who drop in to support the final steps should be cut away. The rating of the *worth reading* articles should go down towards the rating of *others* because the support by the high reputation of featured article specialists vanishes. Table 3, however, shows that the *worth reading* pages keep their high rankings. And even if the authors' reputation is computed merely based on the 14 first edits, the *worth reading* articles keep their higher ratings compared to the *others* as table 4 shows.[3] This even holds for very radical pruning by only using the very first (non anonymous) edit of each page (see table 5 and 6).

These numbers indicate that pages edited in the very beginning by authors with high reputation have a higher chance to get featured in the future. Recall that anonymous edits are skipped in our study, so that pages can have many anonymous edits before the first edit by a known author. We do not believe that this has a large effect on our data, but we will investigate these anonymous edits in future work.

4.3 Bias due to self-voting?

An additional bias is that articles are awarded the labels *excellent* and *worth reading* in a voting by the community of Wikipedians. So it might be the case that a certain group of authors votes for their own articles, i.e., for the articles that they have started writing and/or that they have edited a lot, to become *worth reading* or *excellent*, respectively, regardless of their quality. We consider this bias to be small because featured articles have to fulfill a set of quality criteria to gain this status and these criteria are enforced with high emphasis by the community. So one cannot easily push a low quality article to be voted as featured.

4.4 Edit- vs. page-based author reputation

To explain the difference between the edit-based and page-based measures we give a small example. We have a look at

---

[3]The absolute values, especially of the eb ratings, decrease because of the much lower reputations.

three pages, an *excellent* page with 300 edits, a *worth reading* page with 200 edits and an *other* page with 10 edits. An author a_s who contributes once to each of the three articles has $\mathrm{rep}_{\mathrm{eb}}(a_s) = \mathrm{rep}_{\mathrm{pb}}(a_s) = 0.33$ while an author a_t editing the *excellent* article 98 times and one time the other two has $\mathrm{rep}_{\mathrm{eb}}(a_t) = 0.98$ and $\mathrm{rep}_{\mathrm{pb}}(a_t) = 0.33$. As table 7 shows, the edit-based and page-based reputation values are nearly equal, so German Wikipedia authors seem to behave more like a_s rather than a_t, and the editing styles do not seem to differ much. If there were groups of authors with very different editing styles regarding *excellent* vs. *other* articles (one group of authors behaves like a_s with rather spreading contribution and another group behaves like a_t with many edits on few *excellent* articles and few edits on *other* articles), this would show up in a reduced correlation between $\mathrm{rep}_{\mathrm{eb}}(a)$ and $\mathrm{rep}_{\mathrm{pb}}(a)$. As shown by table 7, the correlation between both measures is rather high, which disproves this.

Note that most differences in edit styles cannot be distinguished by these measures. For example, an author a_E who edits *only* excellent articles has $\mathrm{rep}_{\mathrm{eb}}(a_E) = \mathrm{rep}_{\mathrm{pb}}(a_E) = 1$. Page-based and edit-based reputations are also identical for an author a_{EL} who contributes equally to *excellent* and *worth reading* articles but not to *others*, e.g. 10 edits on the *excellent* article and 10 on the *worth reading* give $\mathrm{rep}_{\mathrm{eb}}(a_{EL}) = \mathrm{rep}_{\mathrm{pb}}(a_{EL}) = 0.5$. So the proposed "featured article specialists" would not show up – which does not matter as they would be caught by the pruned history approach.

4.5 Edit- vs. author-based page ratings

Tables 2 to 6 show that there are nearly no differences when edit-based and author-based page ratings are compared with respect to the relative values of *excellent*, *worth reading* and *other* pages.

Let's have a look at our example from section 4.4: the *worth reading* page p_{L1} with 200 edits received 100 of these edits from one single author a^+ with $\mathrm{rep}_{\mathrm{pb}}(a^+) = \mathrm{rep}_{\mathrm{eb}}(a^+) = 0.2$ and the other 100 edits from 100 different authors $a_1, \ldots, a_i, \ldots, a_{100}$ with $\mathrm{rep}_{\mathrm{pb}}(a_i) = \mathrm{rep}_{\mathrm{eb}}(a_i) = 0$. This gives the page ratings $\mathrm{rating}_{\mathrm{eb}}(p_{L1}) = 0.1$ and $\mathrm{rating}_{\mathrm{ab}}(p_{L1}) = 0.00198$. Another *worth reading* page p_{L2} with 200 edits received 100 of these from 100 different authors $a_1^+, \ldots, a_i^+, \ldots, a_{100}^+$ with $\mathrm{rep}_{\mathrm{pb}}(a_i^+) = \mathrm{rep}_{\mathrm{eb}}(a_i^+) = 0.2$ and the other 100 edits from 100 different authors $a_1 \ldots a_i \ldots a_{100}$ with $\mathrm{rep}_{\mathrm{pb}}(a_i) = \mathrm{rep}_{\mathrm{eb}}(a_i) = 0$. For this page p_{L1} we get the page ratings $\mathrm{rating}_{\mathrm{eb}}(p_{L1}) = \mathrm{rating}_{\mathrm{ab}}(p_{L1}) = 0.1$

This shows that if the high edit-based rating of a page is based on a large number of edits from a single author with high reputation, then the author-based rating of the same page is close to 0. Table 2 shows some decrease of page rankings from edit-based to author-based by a factor of 2.67 for featured, and 1.83 for *other* articles. This implies that the number of edits by authors with high reputation is above average, but it also shows that the higher ratings of *worth reading* articles are the result of consolidated work of many authors with high reputation (otherwise the difference between $\mathrm{rating}_{\mathrm{eb}}(p)$ and $\mathrm{rating}_{\mathrm{ab}}(p)$ would be much larger).

5. CONCLUSION

We started this paper with the question whether it is sufficient to have many peoples working on an article in order to be promoted to featured article status or whether it does really matter that users with a reputation for high quality writing contribute. In a study on a dump of the German Wikipedia, we explored this question by descriptive statistical measures. Using very simple measures for author reputation and page rating we could show a connection between the authors of *excellent* and *worth reading* articles. Restricting these measures to the first few edits a page receives gives evidence that it is unlikely that this connection is simply a side effect of the reviewing process that pages pass in order to obtain the label *excellent* or *worth reading*, respectively. For the German Wikipedia, we can say that it seems to matter who contributes.

The next step will be to have a closer look on the role that anonymous authors play in article creation and to analyze their distribution in the edit history of featured and non-featured articles. The temporal structure of the edit history will be an important source of information, especially if we are able to identify distinct phases in the development of a featured article. And finally, our quantitative measures should be supported by a qualitative analysis to explain the effects we found in our data.

6. REFERENCES

[1] B. T. Adler and L. de Alfaro. A content-driven reputation system for the wikipedia. In *Proceedings of WWW 2007*, Banff, Alberta, Canada, May 8–12 2007.

[2] A. Braendle. Zu wenige Köche verderben den Brei. Eine Inhaltsanalyse der Wikipedia aus Perspektive der journalistischen Qualität, des Netzeffekts und der Ökonomie der Aufmerksamkeit. Master's thesis, University of Zurich, 2005.

[3] A. Ciffolilli. Phantom authority, self-selective recruitment and retention of members in virtual communities: The case of wikipedia. *First Monday*, 8(12), December 2003. available at `http://firstmonday.org/issues/issue8_12/ciffolilli/index.html`.

[4] J. Giles. Internet encyclopaedias go head to head. *Nature*, 438:900–901, December 2005.

[5] N. Korfiatis, M. Poulos, and G. Bokos. Evaluating authoritative sources using social networks: an insight from wikipedia. *Online Information Review*, 30(3):252–262, 2006.

[6] D. McGuinness, H. Zeng, P. Pinheiro da Silva, L. Ding, D. Narayanan, and M. Bhaowal. Investigations into trust for collaborative information repositories: A wikipedia case study. In *Workshop on the Models of Trust for the Web (MTW06)*, May 2006. Edinburgh, Scotland.

[7] V. Rateike, J.-M. Rösner, L. Denks, and C. Eberts. Die Entwicklung exzellenter Artikel in der deutschsprachigen Wikipedia - Keine Qualität ohne Küchenchef? `http://141.13.22.238/mediawiki/index.php/Die_Entwicklung_exzellenter_Artikel_in_der_deutschsprachigen_Wikipedia`, retrieved on April 24, 2007, January 2007.

[8] B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser. Assessing information quality of a community-based encyclopedia. In *Proceedings of the International Conference on Information Quality - ICIQ 2005*, pages 442–454, 2005.

[9] F. B. Viégas, M. Wattenberg, and M. M. McKeon. The hidden order of wikipedia. In *Proceedings of HCII 2007*, 2007.