Association for Information Systems
AIS Electronic Library (AISeL)

Wirtschaftsinformatik Proceedings 2015 | Wirtschaftsinformatik

3-5-2015

Good Authors = Good Articles? - How Wikis Work

Thomas Wöhner

Sebastian Köhler

Ralf Peters

Good Authors = Good Articles? – How Wikis Work

Thomas Wöhner, Sebastian Köhler, Ralf Peters

Martin-Luther-Universität Halle-Wittenberg, 06099 Halle (Saale)
{thomas.woehner,sebastian.koehler,ralf.peters}@wiwi.uni-halle.de

Abstract. Wikis are websites to develop content collaboratively. The question arises to what extent the reputation of participants influences the quality of wiki sites. We analyze the impact of author reputation using the example of Wikipedia. We extend previous research by considering a set of different reputation metrics and a new model for aggregating reputation values. Since anonymous authors tend to have a lower reputation, we also quantify the level of participation of anonymous authors as an indicator for the reputation of the crowd. Our analysis finds out that reputation matters, but strongly depends on the used reputation metric and therefore on the corresponding author characteristics. The study shows that the experience of authors in the development of high-quality articles is highly relevant whereas the number of edits and the quality of contributions are of lower importance. Finally, our investigation proves the open editing model and the self-healing mechanism of wikis.

Keywords: Wiki, Quality, Reputation, Wisdom of the Crowd

1 Introduction

Since the collapse of the dot-com bubble in the year 2000 Web 2.0 has been becoming increasingly popular. The main characteristic of this specific part of the World Wide Web is the particular importance of user-generated content and user interaction [1]. Typical types of Web 2.0 applications are social network sites, wikis, blogs, and photo- and video-sharing portals [1]. The relevance of Web 2.0 can be seen in the user statistics of the World Wide Web. Thus, four (Facebook, YouTube, Wikipedia and Twitter) of the ten worldwide most visited websites belong to Web 2.0 [2].

An important factor of success of Web 2.0 is the so called wisdom of the crowd. This term refers to the phenomenon that under particular conditions a group of people is able to reach better decisions than individual experts. The reason for this phenomenon is the aggregation of different skills and knowledge [3]. Since Web 2.0 applications involve a high number of users to create content, the wisdom of the crowd is particularly relevant.

This paper focuses on wikis as a very popular Web 2.0 application. Wikis are websites that allow viewers to edit content directly within the web browser. In comparison to HTML, Wiki syntax is less complex, so that users are able to contribute to wiki sites without having high technical knowledge [4]. The most famous wiki is the free online-encyclopedia Wikipedia. Wikipedia consists of more than 32 million articles

12th International Conference on Wirtschaftsinformatik,
March 4-6 2015, Osnabrück, Germany

Wöhner, T.; Köhler, S.; Peters, R. (2015): Good Authors = Good Articles? - How Wikis Work, in: Thomas. O.; Teuteberg, F. (Hrsg.): Proceedings der 12. Internationalen Tagung Wirtschaftsinformatik (WI 2015), Osnabrück, S. 872-886

worldwide in 287 different languages. The largest Wikipedia is the English version with about 4.5 million articles followed by the Swedish, Dutch and the German Wikipedia each with about 1.8 million articles [5]. The entire Wikipedia attracts about 500 million unique visitors monthly as of March 2014 [6]. Due to the high popularity and the large amount of data, Wikipedia is a particularly suitable example to investigate user behavior in wikis. For this reason, similar to most of the previous research on wikis, our study also focuses on Wikipedia.

Despite the high relevance, the phenomenon of the wisdom of the crowd in wikis is not yet explored fully. It is still unexplained which characteristics of the crowd lead to a high content quality. Some research has investigated the impact of the number of authors on the quality [7–9]. The studies show that with an increasing number of authors the quality grows. However, the impact of the reputation of authors is poorly understood and only one initial study was published by Stein and Hess [10].

In this paper, we extend the previous research and investigate the relevance of author reputation for the wisdom of the crowd in more detail. In our study, we apply a set of different reputation metrics to evaluate which editing patterns lead to a high quality of articles. The analysis helps to understand the editing process in wikis more detailed. Indeed, some publications [11, 12] discuss approaches that use author reputation to assess the quality of articles, although it has not been fully explored how reputation matters. These publications [11, 12] do not specify any applicable reputation metrics for the quality assessment. This research gap is addressed by our analysis.

The paper is structured as follows. Chapter 2 describes the related work and details present research gaps. Chapter 3 explains the metrics we use for reputation assessment and introduces models to compute aggregated reputation values of a given article. Chapter 4 describes our evaluation method and discusses the results of the study in detail. The paper concludes with a summary and an outlook on future research directions.

2 Related Work

Relating to the present research question, there are two relevant research directions: research on reputation assessment in wikis and research on the phenomenon of the wisdom of the crowd.

Research on reputation assessment provides metrics to assess the reputation of Wikipedia authors automatically. The approaches introduced by Adler and Alfaro [13] and Javanmardi et al. [14] are based on the quality of the authors' contributions which is measured by their persistence. It is assumed that a long lifetime of contributions indicates the acceptance of the edits within the Wikipedia community. Wöhner et al. [15] analyze a set of seven reputation metrics and evaluate their performance.

Besides the research work that focuses on reputation assessment, some studies on the editing process in Wikipedia also consider the reputation of authors and, in doing so, they provide metrics for reputation assessment. In this context, the participation in the development of featured articles [10], the participation in discussions on talk pages [16, 17] as well as the experience of authors [16] are suggested as features to assess the author reputation.

Regarding the wisdom of the crowd, the first publication was published by Galton [18]. The paper describes an experiment in which a crowd of laymen estimates the weight of an ox. The results of the study show that on average the crowd could assess the weight more precisely than individual experts like butchers or cattle-dealers. [3] introduces the term of the wisdom of the crowd and investigates the decision types and criteria required to benefit from that phenomenon. The main criteria are diversity and independence of people, decentralized decision making and suitable aggregation of individual decisions. Wikipedia meets the criteria that have been defined by Surowiecki [3], but the quality of Wikipedia articles differs considerably. So the question arises how the composition of the crowd influences the wisdom of the crowd.

A large number of research papers focus on the quality of Wikipedia articles [8, 9, 11, 12, 19–23]. Some of these papers provide first findings related to the phenomenon of the wisdom of the crowd in wikis. Thus, [8, 23] as well as [9] prove that with an increasing number of authors or edits the quality of articles tends to grow. This connection can be explained by the aggregation of different knowledge and various skills.

The most related work to ours is the study of Stein and Hess [10] in which the relation between the reputation of authors and the article quality has initially been investigated. The study concludes that reputation matters for the article quality. However, as the main limitation, [10] build their reputation assessment only on the participation of authors in the development of featured articles. The study was published in an early stage of Wikipedia research, so that recently introduced reputation metrics could not be applied and it is unknown in how far these metrics are relevant for the article quality. Furthermore, Stein and Hess [10] only focus on registered authors and disregard anonymous contributions. Finally, their findings are based on a simple comparison of means of reputation scores for high-quality and low-quality articles. Concluding from these issues, the impact of reputation on the wisdom of the crowd is not fully understood.

Considering the research gaps discussed above, in this paper we extend the previous research in three ways. Firstly, we apply a set of different reputation metrics from recent research. Secondly, since anonymous authors tend to have a low reputation [15], we also investigate the participation of anonymous authors to determine the impact of reputation. Finally, instead of a simple comparison of mean values, we introduce and use a new model to aggregate the reputation values of registered authors of a given article. This model weights the authors according to their impact on the article. For the statistical analysis we employ a logistic regression and calculate the effect sizes by means of Cohen’s d.

3 Reputation Assessment

In this section, we describe the reputation metrics being used in our analysis. Subsequently, we present our models to aggregate the reputation values of authors of a given article. Finally, the section introduces our new metrics to measure the level of participation of anonymous authors.

3.1 Reputation Metrics

Common reputation metrics differ in the considered editing patterns that represent a high reputation. The following three main sources of author reputation can be distinguished: the quality of contributions, the participation in the development of high-quality articles and the experience of the author. To provide a comprehensive analysis, our chosen set of reputation metrics comprises metrics from each category.

Reputation metrics based on the quality of contributions are published in [13], [14] and [15]. Since the computation is less complex in comparison to the other approaches, we choose the efficiency R^{eff} from Wöhner et al. [15]. For the calculation of R^{eff}, persistent and transient contributions are distinguished. Persistent contributions survive a time period of at least 14 days and are therefore assessed to be of high quality. Transient contributions are discarded within 14 days and thus judged as low-quality contributions. The threshold of 14 days was determined empirically [15]. Based on the definition of persistent and transient contributions, the efficiency R^{eff} is calculated as follows:

$$R^{eff} = \frac{A^{pc}}{A^{total}} \tag{1}$$

A^{pc} defines the amount of persistent contributions and A^{total} the amount of all contributions of a given author. The amount of contributions is measured by the number of words that were inserted or deleted. Hence, R^{eff} refers to the fraction of the persistent contribution of a given author. In [15] the reputation metric is evaluated by means of a classification of blocked and regular users. By using R^{eff}, a high accuracy of about 85% is achieved.

Our second reputation metric

$$R^{qh} = \frac{N^{qh}}{N^{total}} \tag{2}$$

measures the participation of an author in the development of high-quality articles. The metric was introduced in [10]. It is calculated as the quotient of the number of edits on high-quality articles N^{qh} and the total number of edits N^{total} by a given author. As examples for high-quality articles [10] employed *featured articles*. In addition, in this study we consider *good articles* as a further example. Featured as well as good articles are of particularly high quality and were labeled by a community-based quality evaluation [24]. In contrast to featured articles, some minor inconsistencies in the quality are tolerated for good articles [25].

The last metric in our set [26, 27]

$$R^e \tag{3}$$

denotes the number of edits of a given author. The metric is also currently being used in Wikipedia to quantify the experience of authors, for example to assign the administrator status [24].

3.2 Aggregation of Reputation Values

To measure the impact of the reputation on the article quality, the reputation values of the authors of a given article have to be aggregated. The comparison of the aggregated reputation values of low-quality and high-quality articles shows whether the author reputation matters or not. In this study, we use two aggregation models.

The simplest approach for aggregation is the average-based aggregation model $R^r{}_{avg}$ that is calculated as follows:

$$R^r{}_{avg} = \frac{1}{n}\sum_{i=1}^{n} R^r \tag{4}$$

In this equation R^r refers to the reputation value of each registered author $i = 1 .. n$ of an article, where r denotes the considered reputation metric (*eff* for R^{eff}, *qh* for R^{qh} and e for R^e). The average-based aggregation is the only approach which was applied by Stein and Hess [10]. According to this calculation rule, authors are equally weighted, independent of the amount and the persistence of their contributions to the given article. In more detail, if an article is vandalized and even the vandalism is reverted completely, the reputation of the vandal is considered within the average-based aggregation model. Therefore, articles that are often vandalized tend to have a lower aggregated reputation score.

To deal with these shortcomings, we introduce a second aggregation model in our study that takes into account the persistence of contributions. The persistence-based aggregation model $R^r{}_{pers}$ weights the authors according to the number of words A^{lv} of the latest article version that were inserted by an author. The calculation of $R^r{}_{pers}$ is performed as follows:

$$R^r{}_{pers} = \sum_{i=1}^{n} R^r \frac{A^{lv}}{A^{lv}{}_{reg}} \tag{5}$$

$A^{lv}{}_{reg}$ refers to the total number of words of the latest article version inserted by registered authors. According to this calculation rule, only authors that have contributed to the latest article version are considered, whereas authors whose contributions have been rejected are disregarded. The persistence-based aggregation model particularly considers the main authors of the given article.

3.3 Participation of Anonymous Authors

Anonymous authors tend to have a lower reputation than registered authors [9]. Therefore, the level of participation of anonymous authors also reflects the reputation of the crowd. In the case of an anonymous edit, the IP address of the author is saved in the editing history instead of a username. IP addresses are usually assigned dynamically so that users have varying IP addresses at different times. Hence, anonymous authors cannot be identified clearly and the reputation metrics being discussed in the previous section are not applicable. To overcome this problem, by analogy with the aggregation models for the reputation of registered authors, we introduce two metrics to quantify the participation of anonymous authors:

$$R^{ano}{}_{e} = \frac{N^{e}{}_{ano}}{N^{e}{}_{total}} \tag{6}$$

$$R^{ano}{}_{pers} = \frac{A^{lv}{}_{ano}}{A^{lv}{}_{total}} \tag{7}$$

$R^{ano}{}_{e}$ weights the anonymous edits equally, independent from the acceptance and the amount of the contribution. It is computed by the quotient of the number of edits $N^{e}{}_{ano}$ performed by anonymous authors and the total number of edits $N^{e}{}_{total}$ of the given article. The persistence-based metric $R^{ano}{}_{pers}$ is calculated by the quotient of the number of words $A^{lv}{}_{ano}$ of the latest article version inserted by anonymous authors and the total number of words of the latest version $A^{lv}{}_{total}$. Since the metrics (6) and (7) refer to the total group of anonymous authors, the metrics are calculated exactly even though the authors are identified by varying IP addresses. In comparison to the aggregated reputation values for registered authors, the reputation values for anonymous authors have to be read reversely. High values for $R^{ano}{}_{e}$ or $R^{ano}{}_{pers}$ reflect a low reputation of the crowd, since anonymous authors are more present.

4 Impact of Reputation

In this chapter we first describe our research method and the data being used. Subsequently we present the results of our analysis, discuss the findings and draw some implications.

4.1 Research Method and Data

The present research is based on the data of the German Wikipedia as of 01/21/2008. Overall our data set includes 1,023,507 articles and 26,392,081 article versions. The data contain the complete editing history and provides all information required for the calculation of our reputation metrics. Our data set is also used in [15]. We decided to use this data set instead of a newer one since in May 2008 the German Wikipedia modifies the editing model by introducing *Flagged revisions*. According to this approach a contribution is only accepted if it is verified by an experienced author (called *Sichter)* to avoid obvious vandalism. Since we are interested in the impact of author reputation in pure wiki systems, the used data set should not be affected by significant modifications of the wiki principle. Therefore, data should be captured before May 2008.

Wikipedia is divided into different namespaces. Our study is restricted to the main namespace which includes all encyclopedia articles. We disregard other Wikipedia pages such as discussion pages and user pages because of their special editing process. In order to compute our metrics, we developed a set of Java tools. The calculation of the persistence-based aggregation models requires information about the authorship of each word of the latest article versions. Since this information is not provided by Wikipedia, we developed an algorithm that deduces the ownership of words from the editing history. In the case of reverted article versions, the algorithm assigns the authorship to the author who inserted the text firstly.

To determine the impact of the reputation on the wisdom of the crowd, we analyze the differences of low-quality and high-quality articles regarding the author reputation by using statistical techniques. For this purpose we employ a logistic regression [28] and calculate the effect size by means of Cohen's d values [29]. Like previous research [9, 10, 12, 22], we identify the article quality with the aid of quality ratings by the Wikipedia community. The community-based rating ensures that the judgment is not biased by the subjectivity of a single reviewer. As examples for high-quality articles we selected featured and good articles [24]. Our data sample includes 1,211 featured and 2,184 good articles, so that the group of high-quality articles contains 3,395 articles overall. As representatives for low-quality articles we randomly selected 3,395 non-labeled articles. Hence, high- and low-quality articles are equally weighted. Because of the lack of ratings, the quality of non-labeled articles is in general unknown, but, as defined by Wikipedia, featured and good articles meet the highest quality standards [24]. Therefore, we assume that non-labeled articles are of lower quality than featured and good articles. Nevertheless, for the calculation of the author reputation we consider all edits of the user, not only the edits on the selected articles. Thus, our study considers 37,158 authors and 17,630,621 article versions.

One could claim that the analysis is affected by the selection of featured and good articles as examples for high-quality articles. If an article is nominated as candidate for a featured or good article, the article is listed on central Wikipedia sites and therefore gets more attention than other articles. In order to take care of this effect, we cut the editing histories of high-quality articles immediately before the articles were nominated for the status of featured or good article. Hence, in the considered time interval the articles are not listed on central sites and therefore are not distinguishable from other articles with regard to their visibility.

The evaluation is performed in two steps. Firstly, we investigate the reputation values of registered authors. Subsequently, we study the participation of anonymous authors in the development of low-quality and high-quality articles.

4.2 Results and Discussion

Impact of the Reputation of Registered Authors. The box plots in Fig. 1 show the distribution of the aggregated reputation values in both article categories using the different aggregation models and the different reputation metrics. In addition, Table 1 lists the mean values in the two article groups as well as the output of the logistic regression which includes the calculated significance levels and the odds ratio as a measure for the effect size. The clearer the odds ratio differs from 1, the larger the effect is. An odds ratio less than 1 means that the metric influences the quality of articles negatively. Contrarily, an odds ratio of more than 1 refers to a positive effect. However, the odds ratio has to be read relating to the range of values and the distribution of the respective metric. Since in our scenario the metrics differ clearly according to these features, the odds radio is difficult to interpret. For that reason we calculated Cohen's d [29] as a further metric to quantify the effect size. Cohen's d refers to the quotient of the difference of the mean values in both article groups and the standard deviation. The value is independent from the range of values and allows a simple comparison of the metrics to be investigated.

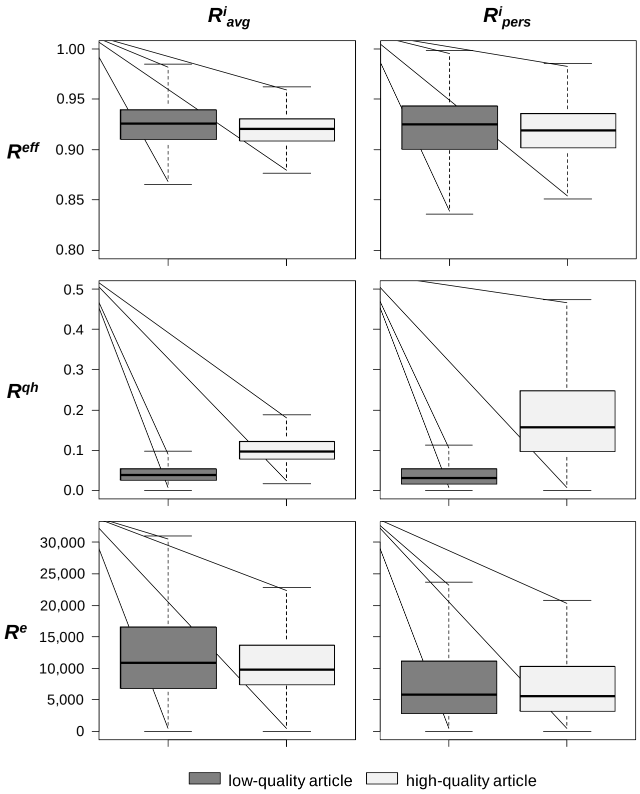

Fig. 1. Distribution of Reputation Values

As a main result, the evaluation challenges previous research and shows that the impact of reputation on the article quality in fact matters, but it strongly depends on the reputation metric in use.

Our study confirms the findings of [10] who have already proven the impact of R^{qh} on the article quality. In Addition to [10], we substantiate the results by a logistic regression and effect sizes. Furthermore, we consider the persistence-based aggregation model to ensure that authors are weighted according to their influence on the article. Cohen [29] suggests that Cohen's d values of about 0.2 can be taken as a small effect, values of about 0.5 reflect a medium effect, and Cohen's d higher than 0.8 means a large effect. Following this suggestion, the Cohens's d values of 1.442 and 1.660 are considerably high and indicate a large effect of R^{qh} on the article quality for both aggregation models.

However, in our setting the use of R^{qh} leads to a bias. In case of high-quality articles, the calculations of the reputation metric and the aggregated reputation values are based on the same data set. Since authors of high-quality articles have edited at least one high-quality article, they tend to have a higher reputation. It is obvious that this effect leads to higher aggregated reputation values for the group of high-quality articles. To quantify this bias we performed a second experiment. We split the group of high-quality articles randomly into two parts of the same size. We used the first part (1698 articles) for training to calculate the reputation of the authors. In this context, reputation refers to the percentage of the edits of a given author on the articles of the training sample. We applied the second part (1697 articles) as test sample to compute the aggregated reputation values. Therefore, we ensure that the author reputation and the aggregated reputation values of an article are calculated independently. Even in this scenario the difference of R^{qh} between high- and low-quality articles is highly significant ($\alpha = 0.01$). Cohen's d amounts to 1.060 for the average-based aggregation model and 0.962 for the persistence-based aggregation model. Therefore, the effect is smaller, but even if author reputation and aggregated reputation values are calculated on different samples, a large effect size is measured.

Regarding the experience-based reputation metric R^{e} our analysis leads to surprising results. The connection between reputation and article quality is in a different way than excepted. As the mean values show, low-quality articles involve a higher percentage of experienced authors than high-quality articles. We are not able to explain this phenomenon exactly. However, we do not believe that experienced authors are not valuable for Wikipedia. But a high percentage of experienced authors means also that there is a low percentage of authors with less edits. Whereas experienced authors with a high number of edits might be generalists, some authors with a lower number of edits might be specialists focusing on a particular subject. An explanation of the lower percentage of experienced authors for high-quality articles could therefore be that high-quality articles are written by a higher number of specialists. Following this argument, a low average experience of the authors can be read as a high reputation of the crowd. However, the low Cohen's d values indicate that this effect is only slightly relevant.

Table 1. Statistics on Reputation Values

Reputation Metric	Aggregation Model	Mean		Logistic Regression		Effect Size
		Low-quality Articles	High-quality Articles	Significance Level	Odds Ratio	
R^{eff}	R^{eff}_{avg}	91.4%	91.8%	**	0.971	0.063
	R^{eff}_{pers}	90.8%	91.5%	–	0.999	0.097
R^{qh}	R^{qh}_{avg}	4.4%	10.6%	***	2.024	1.660
	R^{qh}_{pers}	4.5%	19.0%	***	1.317	1.442
R^{e}	R^{e}_{avg}	14967	11845	***	0.999	0.234
	R^{e}_{pers}	11666	7885	***	0.999	0.235

Significance Level: – No Significance; * $\alpha = 0.1$; ** $\alpha = 0.05$; *** $\alpha = 0.01$

Applying the efficiency R^{eff}, we identified the lowest relevance in our analysis. The difference is only significant applying the average-based aggregation model. Using the persistence-based aggregation, which considers the authors according to their relevance, the difference in the reputation values is not significant. According to the Cohen's d values there is only a very small effect for both aggregation models. That means that the differences in the reputation values are marginal. The low relevance of the efficiency R^{eff} is also indicated by the similar means of the aggregated reputation values in both article groups and the odds ratio of almost 1 (see Table 1).

Impact of the Participation of Anonymous Authors. Fig. 2 presents the distributions of our metrics to quantify the level of participation of anonymous authors. In Table 2 the mean values for the low- and high-quality articles as well as the calculated output of the logistic regression and effect sizes are displayed.

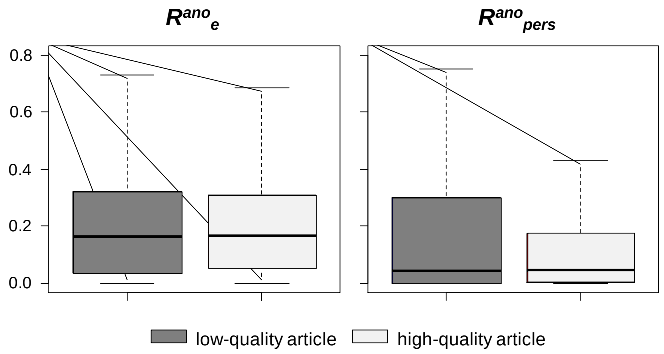

Fig. 2. Distribution of the Level of Participation of Anonymous Authors

The evaluation shows that the portion of anonymous authors has a moderate impact on the wisdom of the crowd. However, a statistical relevance is only measured using the persistence-based aggregation model. Thus, for $R^{ano}{}_{pers}$ Cohen's d amounts to 0.273 and the effect size can be judged as small. Regarding the average-based aggregation model $R^{ano}{}_{e}$, the Cohen's d value is almost zero. The strong impact of the aggregation model on the measured effects can also be seen clearly by the mean values in both article groups. While the mean value of $R^{ano}{}_{e}$ for both article groups is about 0.2, $R^{ano}{}_{pers}$ varies noticeably. As mentioned above, the persistence-based aggregation model weights the authors according to the acceptance and the amount of their contribution. Therefore, the results demonstrate that anonymous authors contribute to low- and high-quality articles in the same intensity, but their contributions are less accepted on high-quality articles.

Table 2. Statistics on the Participation of Anonymous Authors

Calculation Model	Mean		Logistic Regression		Effect Size
	Low-quality Articles	High-quality Articles	Significance Level	Odds Ratio	
$R^{ano}{}_{e}$	20.3%	19.7%	–	0.998	0.032
$R^{ano}{}_{pers}$	19.5%	13.0%	***	0.989	0.273
Significance Level: – No Significance; * α = 0.1; ** α = 0.05; *** α = 0.01					

Discussion. Our analysis proves that author reputation matters, but the impact on the article quality strongly depends on the definition of reputation. Since the different reputation metrics regard certain characteristics of authors, our study demonstrates that some characteristics are relevant for the article quality whereas other characteristics have no direct impact. A comparison of the relevance of the different reputation metrics leads to a better understanding of the editing process of high-quality articles. Thus, the high relevance of the participation of authors in the development of high-quality articles as reputation metric (R^{qh}) shows that a specific group of authors is particular important to achieve a high quality level. In other words, if an author is involved in the development of high-quality articles, the participation of the author in the editing process of another article increases the probability that this article also achieves a high quality level. As we have already discussed above, the lower average experience of authors (R^{e}) of high-quality articles might indicate that high quality also requires some specialists that focus on certain topics. If reputation is measured by the quality of contributions (R^{eff}), the reputation has a marginal impact on the article quality. This means that a high-quality article is not simply defined by the average quality of contributions, but by the number and composition of different contributions. Furthermore, the study indicates that contributions of anonymous authors with a tendentially lower quality of contributions are more often rejected on high-quality articles than on low-quality articles and therefore the percentage of anonymous authors of the last article versions (R^{ano}_{pers}) also indicates the quality of the article.

It is not the goal of this paper to judge different reputation metrics. Rather, we suggest that reputation metrics should be chosen carefully depending on the purpose of the reputation system. Thus, if reputation is used for coloring questionable sections of articles [30], reputation metrics based on the quality of contributions might be most suitable. If the purpose of the reputation system is the motivation of users to contribute to Wikipedia frequently, an experience-based reputation metric may lead to the best results. However, to assess the quality of articles [11, 12], the participation of authors in the development of high-quality articles seems to be a better reputation source.

As a final finding, our analysis proves the *self-healing mechanism* in wikis. The average-based aggregation model regards all authors that have contributed to an article, whereas the persistence-based aggregation model only considers authors of accepted contributions that survive up to the latest article version. In general, with the exception of the efficiency, the study shows that the aggregated reputation values indicate a higher reputation applying the persistence-based aggregation model in comparison to the average-based aggregation model. This phenomenon can be seen clearly especially in the group of high-quality articles. Thus, employing the reputation metric R^{qh} for example, the aggregated reputation value R^{qh}_{pers} amounts to 0.190 whereas R^{qh}_{avg} amounts to 0.106 (see Table 1). A higher reputation of the crowd applying the persistence-based aggregation model means that high-reputation authors are more represented in the latest article version. Consequently, the contributions of some low-reputation authors must have been rejected.

5 Conclusion

In this paper, we analyzed the impact of author reputation on the wisdom of the crowd in wikis. Our study is based on the example of Wikipedia. In particular, we extend previous research by applying different reputation metrics and introducing a new model to aggregate the reputation of authors of a given article. Our new aggregation model weights authors on the basis of the amount and the acceptance of their contributions. Moreover, we analyzed the participation of anonymous authors. Since anonymous authors are judged to have a lower reputation than registered authors, the participation of anonymous authors also reflects the reputation of the crowd.

To determine the impact of author reputation on the wisdom of the crowd, we analyzed differences in the measured reputation values between low- and high-quality articles. We identified the quality of articles with the aid of user ratings by the Wikipedia community. Thus, we used featured and good articles as examples for high-quality articles and non-labeled articles as representatives for low-quality articles. We assessed the impact of reputation by a logistic regression and Cohen's d as a measure for the effect size.

Our study challenges the conclusion of previous research that reputation matters for the quality of articles. We found out, that this claim is not generalizable. The impact strongly depends on the definition of reputation and therefore on the corresponding author characteristics. Not all editing patterns that are judged to be valuable have a direct impact on the article quality. Our study proves the results from [10] in which the participation of authors in the development of high-quality articles was used as basis for reputation measurement. Applying this definition, reputation of authors is highly relevant for the quality of articles.

Surprisingly our analysis shows that authors of high-quality articles have a slightly lower experience on average than authors of low-quality articles. This effect might be explained by the trend that high-quality articles involve some specialized authors that focus on certain subjects. Furthermore, reputation of authors is often calculated on the basis of the persistence of the contributions of an author [13–15]. Applying this widely used definition of reputation, the study points out that reputation does not matter. Our analysis of anonymous authors shows that a high-quality articles have are lower percentage of anonymous authors. The results demonstrate that contributions of anonymous authors are more often rejected on high-quality articles than on low-quality articles.

Our findings can be used for the implementation of approaches [11, 12] that assess the quality of articles on the basis of author reputation. According to our investigation, the reputation metrics for this purpose have to be selected carefully. Thus, reputation metrics based on the experience in the development of high-quality articles seem to be the most suitable reputation metrics for such approaches. Finally, we prove the effectiveness of the self-healing mechanism in wikis that ensures that contributions of low-reputation authors are revised more likely.

The study reveals that some authors are particularly relevant to achieve a high article quality. A detailed investigation of this group of authors in future work may identify tasks and editing patterns of these authors. Such an analysis may derive new reputation metrics that can be used for an automatic assessment of the article quality. As a limitation of this paper, the analysis is based on data of Wikipedia. Further research is needed to confirm our findings for other wiki systems.

References

1. O'Reilly, T.: What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software (2005)
2. Alexa.com: Top Sites. the top 500 sites on the web, http://www.alexa.com/topsites (Accessed: 08.08.2014)
3. Surowiecki, J.: The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations. Doubleday Books (2004)
4. Leuf, B., Cunningham, W.: The Wiki way. Quick collaboration on the Web. Addison-Wesley, Boston (2004)
5. Wikimedia: List of Wikipedias, http://meta.wikimedia.org/wiki/List_of_Wikipedias (Accessed: 08.08.2014)
6. Wikimedia: Wikimedia Report Card, http://reportcard.wmflabs.org/ (Accessed: 08.08.2014)
7. Ofer Arazy, Wayne Morgan, Raymond A. Patterson: Wisdom of the crowds: Decentralized knowledge construction. In: Proceedings at the 16th Annual Workshop on Information Technologies & Systems (WITS) (2006)
8. Lih, A.: Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. In: Proceedings of the 5th International Symposium on Online Journalism, pp. 16–17 (2004)
9. Wöhner, T., Peters, R.: Assessing the Quality of Wikipedia Articles with Lifecycle based Metrics. In: Proceedings of the 5th International Symposium on Wikis and Open Collaboration, ACM, New York, NY, USA (2009)
10. Stein, K., Hess, C.: Does It Matter Who Contributes: A Study on Featured Articles in the German Wikipedia. In: Proceedings of the 18th Conference on Hypertext and Hypermedia, pp. 171–174. ACM, New York, NY, USA (2007)
11. Hu, M., Lim, E.-P., Sun, A., Lauw, H.W., Vuong, B.-Q.: Measuring article quality in wikipedia: models and evaluation. In: Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, pp. 243–252. ACM, New York, NY, USA (2007)
12. Zeng, H., Alhossaini, M., Ding, L., Fikes, R., Mcguinness, D.L.: Computing Trust from Revision History. In: Proceedings of the 2006 International Conference on Privacy, Security and Trust (2006)
13. Adler, B.T., Alfaro, L.: A Content-Driven Reputation System for the Wikipedia. In: Proceedings of the 16th international conference on World Wide Web, pp. 261–270 (2007)
14. Javanmardi, S., Lopes, C.V., Baldi, P.: Modeling User Reputation in Wikis. Statistical Analysis and Data Mining 3, 126–139 (2010)
15. Wöhner, T., Köhler, S., Peters, R.: Automatic Reputation Assessment in Wikipedia. In: Dennis F. Galletta, Ting-Peng Liang (eds.) Proceedings of the International Conference on Information Systems, ICIS 2011. Association for Information Systems (2011)
16. Halfaker, A., Kittur, A., Kraut, R., Riedl, J.: A jury of your peers: quality, experience and ownership in Wikipedia. In: Proceedings of the 5th International Symposium on Wikis and Open Collaboration, ACM, New York, NY, USA (2009)
17. Kittur, A., Kraut, R.E.: Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination. In: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, pp. 37–46. ACM, New York, NY, USA (2008)
18. Galton, F.: Vox Populi. Nature 75, 450–451 (1907)

19. Blumenstock, J.E.: Size Matters: Word Count as a Measure of Quality on Wikipedia. In: Proceedings of the 17th international conference on World Wide Web, pp. 1095–1096. ACM, New York, NY, USA (2008)
20. Dalip, D.H., Gonçalves, M.A., Cristo, M., Calado, P.: Automatic Assessment of Document Quality in Web Collaborative Digital Libraries. J. Data and Information Quality 2, 14 (2011)
21. Dondio, P., Barrett, S.: Computational Trust in Web Content Quality: A Comparative Evalutation on the Wikipedia Project. Informatica (Slovenia) 31, 151–160 (2007)
22. Javanmardi, S., Lopes, C.: Statistical measure of quality in Wikipedia. In: Proceedings of the First Workshop on Social Media Analytics, pp. 132–138. ACM, New York, NY, USA (2010)
23. Wilkinson, D.M., Huberman, B.A.: Assessing the value of coorperation in Wikipedia. First Monday 12 (2007)
24. Wikipedia: Wikipedia: about, http://en.wikipedia.org/wiki/Wikipedia:About (Accessed: 08.08.2014)
25. Wikipedia: Compare Criteria Good v. Featured, http://en.wikipedia.org/wiki/Wikipedia:Compare_Criteria_Good_v._Featured (Accessed: 08.08.2014)
26. Anthony, D., Smith, S.W., Williamson, T.: Explaining Quality in Internet Collective Goods: Zealots and Good Samaritans in the Case of Wikipedia (2007)
27. Panciera, K.A., Halfaker, A., Terveen, L.G.: Wikipedians are born, not made: a study of power editors on Wikipedia. In: Proceedings of the ACM 2009 international conference on Supporting group work, pp. 51–60, ACM, New York, NY, USA (2009)
28. Izenman, A.J.: Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning Springer Publishing Company, Incorporated (2008)
29. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences Lawrence Erlbaum Associates (1988)
30. Adler, B.T., Chatterjee, K., Alfaro, L. de, Faella, M., Pye, I., Raman, V.: Assigning trust to Wikipedia content. In: ACM (ed.) Procceedings of the 2008 International Symposium on Wikis (WikiSym) Porto, Portugal (2008)