Social Network Analysis and Mining (2018) 8:36
https://doi.org/10.1007/s13278-018-0512-3

ORIGINAL ARTICLE

Social-collaborative determinants of content quality in online knowledge production systems: comparing Wikipedia and Stack Overflow

Sorin Adam Matei[1] · Amani Abu Jabal[2] · Elisa Bertino[2]

Received: 20 December 2017 / Revised: 18 April 2018 / Accepted: 18 April 2018

Abstract
Online knowledge production sites, such as Wikipedia and Stack Overflow, are dominated by small groups of contributors. How does this affect knowledge quality and production? Does the persistent presence of some key contributors among the most productive members improve the quality of the knowledge, considered in the aggregate? The paper addresses these issues by correlating week-by-week value changes in contribution unevenness, elite resilience (stickiness), and content quality. The goal is to detect if and how changes in social structural variables may influence the quality of the knowledge produced by two representative online knowledge production sites: Wikipedia and Stack Overflow. Regression analysis shows that on Stack Overflow both unevenness and elite stickiness have a curvilinear effect on quality. Quality is optimized at specific levels of elite stickiness and unevenness. At the same time, on Wikipedia, quality increases linearly with a decline in entropy, overall, and with an increase in stickiness in the maturation phase, after an entropy elite stickiness, quality of content peak is reached.

Keywords Wikipedia · Stack Overflow · Unevenness · Elite stickiness · Quality of Content

1 Introduction

We live in a world dominated by social media and user-generated content. Many human interactions, from entertainment to learning and work, are the product of people collaborating with each other directly. It is not surprising that knowledge itself has been affected by social production processes. The wild success of question and answer (Q&A) sites, of bulletin boards, of social news sites like Reddit, of Wikipedia and more recently of blockchain based knowledge sites like Everipedia have moved the burden of informing the world from formal institutions to voluntary groups and what is commonly called "peer-production".

If we are to differentiate between modes of knowledge production, two types can be distinguished: just-in-case and just-in-time. The first is associated with general-purpose reference sites (e.g., http://Wikipedia.org, http://Everipedia.org, http://Scholarpedia.org, or http://Citizendium.org). The second is associated with specialized, immediate-needed information identification and dissemination platforms. Prominent among these are the Question and Answer sites (e.g., http://StackOverflow.com, http://Answers.Yahoo.com, or http://Quora.com). A critical question, in both types of knowledge production sites, is how quality knowledge is produced. Quality, of course, may depend on many different factors, such as the skills of the contributors or topic maturity. However, more complex factors may influence quality. In this paper, we focus on the social structures of knowledge production sites and their temporal evolution. We ask if social processes make a difference across levels of quality and editorial activity or not.

The paper explores primarily the role of the highly productive individuals who generate a majority of the content. We consider their specific weight in the production system via entropy measurements of the entire space in which they

✉ Amani Abu Jabal
aabujaba@purdue.edu

Sorin Adam Matei
smatei@purdue.edu

Elisa Bertino
bertino@purdue.edu

1 Brian Lamb School of Communication, Purdue University, 100 N. University St., West Lafayette, IN 47907, USA

2 Department of Computer Science, Purdue University, 305 N. University St., West Lafayette, IN 47907, USA

Published online: 05 May 2018

work and through their own temporal resilience. We propose that by their focused and continuous work these individuals may influence the quality of the content. Yet, the effect might be conditioned by amount and timing of effort concentration. Furthermore, we explore the role played by high productivity and temporally resilient individuals in generating some of the highest quality and most engaging content on the site. In other words, a subsidiary goal is to better understand if highly productive individuals may make a difference in the production of high involvement–high quality content or not. This content refers to subsamples of high quality and highly edited articles or posts.

Given that we compare two types of production systems, just-in-time and just-in-case, we expect both commonalities and differences across modes of production and types of content. Commonalities may be explained by the voluntary and open nature of the collaborative process in both of modes of production. Decisions to contribute, interactions, and evaluations are voluntary in both systems. Furthermore, as demonstrated by previous research (Palloff and Pratt 2010; Matei and Bruno 2015), which showed a highly skewed distribution of contributions across contributors, both types of knowledge production have a particular social structure of collaboration. There is a "pecking order", which divides the production groups into a few top and a huge majority of bottom contributors. Our previous research, which analyzed contribution volume by the amount of effort using a different metric than the one used here, suggests that on Wikipedia the top 1% contributors may be responsible for about 80% of the editorial activity (Matei and Britt 2017).

The top 1% group is not a mere statistical construct. Its members tend to be present in the elite for longer periods of time. We refer to this behavior as elite stickiness. Skewed contributions by elites that are stable in time create a very specific production regime, in which individuals establish practices, norms, and interaction patterns. These factors are vital for production, and when considering the issue of quality, they become important predictive factors.

At the same time, the nature of the production systems, especially their just-in-time vs. just-in-case orientation, might interact with social structural factors. For example, elite stickiness may be initially stronger in just-in-time spaces, due to the fact that work is more closely normed and with higher entry barriers. Users have to follow more rigorous rules, and success depends on a longer period of socialization. This creates a certain barrier to entry among the active users, which fosters a process of selection and higher skewness of stickiness of contributors. Yet, due to the sequential and isolated nature of the contributions in over longer periods of time just-in-time (Q&A) production might decline both in terms of effort distribution and elite temporal stickiness. On the other hand, wiki-sites (just-in-case spaces), which demand continuous and close interaction between members, might lead in time to increases in both uneven distribution of effort and elite stickiness.

As uneven distribution and elite stickiness vary in time, content quality might change as well. The core argument is that if groups have a core group of highly productive contributors who become "stickier" in time, this will translate in the long run into higher quality content (Stvilia et al 2005; Wu et al 2012). Stable groups acquire better work practices, can check and revise content faster and more reliably. They develop know-how and forms of institutional memory, which make production more efficient and self-correcting (Stvilia et al 2008). Yet, there may be boundaries around production concentration and elite stickiness. Too heavy domination by a small group or not sufficient leadership can both negatively affect content quality (Bruno 2010) through group bias, group think, or unchecked assumptions.

Furthermore, by its very characteristics such as currency, significance, popularity, or controversial nature some content attracts more attention. In certain situations, collaborative practices may lead to high-quality content, as defined by depth of sourcing, structural and organizational efficiency, or narrative fluency. In other situations, especially for controversial articles, we might end up with a high volume of editing, although quality might not be quite optimal. In both situations we expect the collaborative process and the role of highly productive editors to intervene in the production process. Specifically, as previous literature suggests (Kane 2011), high production contributors might have a stronger impact on the quality of contribution.

Thus, the main goal of this paper is to determine if collaboration unevenness and elite stickiness impact quality positively or negatively, if there are any boundaries around this positive effect, and if the effect varies across just-in-case and just-in-time sites or within each across high and low levels of editorial activity or levels of quality.

To address our goal, we explore the following questions:

1. How do contribution inequality and elite stickiness evolve on Wikipedia and Stack Overflow?
2. Is there a relationship between contribution inequality, stickiness, and quality of content on Wikipedia and Stack Overflow?
3. Are there variations in the role of elite stickiness and contribution inequality across types of content (high vs. low quality or editorial activity)?
4. What do the results tell us about the differences between just-in-case and just-in-time knowledge systems?

We provide answers to the questions based on an extensive data analysis of Wikipedia and Stack Overflow, which compares temporal co-evolution of content quality with user-contribution inequality. We look both at the entire dataset and at subsets of high quality and high editorial activity articles. Contribution inequality is represented by an entropy-based model that considers user effort. Overall, the study uses a combination of metrics to triangulate both social processes and content quality.

In what follows we discuss related work (Sect. 2) and present dataset summaries and the methodology including contribution unevenness, elite stickiness, and content quality (Sect. 3). Section 4 presents the main analytic results. Section 5 discusses the implications of the findings and outlines future work.

2 Related work

2.1 Social structure and knowledge production quality

The quality of online knowledge production depends on many factors. One of the most important is the social organization of the knowledge production process. It is not at all indifferent for high-quality knowledge production if a group is loosely or tightly organized or if the composition of the most productive group is stable or highly variable. Previous research has shown that the evolution of social structures (Kittur et al 2009) and stable contribution elites (i.e., high volume contributors), both on Wikipedia and Stack Overflow (Liu et al 2005; Jurczyk and Agichtein 2007; Pal et al 2012), may impact the outcome of the production itself. The high involvement of a given set of individuals who specialized in certain knowledge domains was identified as one of the most important factors in the success of wiki groups (Kane 2011). Similarly, high reputation and high activity individuals are more likely to ask good questions and provide good answers on Stack Overflow (Baltadzhieva and Chrupała 2015). High reputation individuals on Stack Overflow are also more likely to anchor the activity, having higher degree of pagerank centrality in the network of participation (Movshovitz-Attias et al 2013).

In general, online groups that are moderately dominated by certain individuals and who are stable in time tend to produce better knowledge, as Bruno (2010) showed for knowledge production in education wikis. On Wikipedia, a group of highly active users, who are involved in a variety of activities is uniquely responsible for generating good quality content (Liu and Ram 2011).

At the same time, active participation is not sufficient for generating high-quality content (Pal et al 2012). Timing of participation is equally important. Articles increase in quality as more persistent changes are made to them (Wöhner and Peters 2009) by a group of "all around" editors (Liu and Ram 2011). Kittur and Kraut (2008) studied the impact of coordination methods between contributors on content quality, also highlighting the importance of interactional stability in wiki spaces. Also, Kittur et al (2009) analyzed the role of uneven distribution of effort on productivity across thousands of articles on wiki spaces. This work drew attention to the core issue of coordination via concentration of effort among a few selected editors.

Stack Overflow research similarly highlighted the importance of a core group of contributors to the questions and answers posted on the site (Movshovitz-Attias et al 2013). They looked at the ratio of answers vs. questions and integration of the editorial work via network analysis. They concluded that high-impact users can be predicted by early work and that high reputation and contribution are correlated with centrality in the production process.

Bruno (2010), following Kuk (2006), showed experimentally that in educational wiki groups collaborative unevenness may also impact other outcome variables, such as learning. Observing groups of students tasked to engage in collaborative research via a wiki to learn about the traditions of their campus, Bruno (2010) found that collaborative unevenness co-varies with learning. High and low levels of distribution of contributions across learning groups lead to sub-optimal levels of learning. In other words, learning was maximized at a certain level of collaborative inequality. This suggests that other processes, such as those that shape the quality of the project, might be associated curvilinearly with uneven participation and elite stickiness.

2.2 Previous work by the authors

Matei and Britt (2017) have analyzed the temporal evolution of Wikipedia's social structures and the relationship between collaborative evenness, elite stickiness and quality at the global level. The current work expands our previous work (Matei et al 2017), deepening the analysis by considering patterns of association between social structural factors specific and types of content, specifically high vs. low. Specifically, we investigate how the relationship between contribution elites and the quality of user-generated content evolves over the time across content subsets (high vs. average quality and editorial activity) of production contexts. This in depth exploration gives us a better view on what makes knowledge production systems truly productive.

The current paper as well as Matei et al (2017) also relates to our previous work on social structural differentiation in social media (Matei and Britt 2017), in which we determined that elite stickiness is the product of contribution unevenness and that Wikipedia evolved through discrete phases moved ahead by specific "evolutionary motors". The current work expands our previous research on structural differentiation by comparing platforms (Wikipedia vs. Stack Overflow), differentiated scores, and by proposing a modified method for calculating effort and entropy. Thus, in the current paper, we report more specific and comprehensive results from a different angle, while enriching the previous work in this area.

To the best of our knowledge, there is no comparative study which evaluates the social structure and quality relationship from a temporal perspective, both for Wikipedia and Stack Overflow.

2.3 Quality in knowledge production systems

Detecting the measurable effect of a given production system on quality demands careful and systematic operationalization of quality. This is particularly important when considering two different production sites, such as Wikipedia and Stack Overflow. Our research utilizes existing and new work to tackle this problem. This subsection discusses previous work that supports our work on quality detection in the two knowledge production spaces.

Wikipedia is an open-source repository of reference knowledge. Articles are written just-in-case a reader needs quick reference information about a topic. Wikipedia is at the same time a fully editable platform. Most content can be freely edited. Due to its immense growth and success, Wikipedia has developed several methods to evaluate the quality of its articles. First, there are user-driven approaches. One approach allows human editors to label articles as "featured". The criteria used in making quality decisions are accuracy, neutrality, completeness and writing style.[1] This mechanism is laborious and can be scaled only for a small number of articles. As a result, several methods for automatic quality analysis of Wikipedia articles have been proposed (Stvilia et al 2005; Zeng et al 2006; Cross 2006; Dondio and Barrett 2007; Blumenstock 2008). Cross (2006) proposed an approach which colors the article portions based on the time of the inserted text; hence the text which remains after multiple edits is considered reliable. Also, Zeng et al (2006) devised a quality model for article edits based on a Bayesian network of the reputation of authors. The reputation of authors Zeng et al (2006) determines the quality of the content. Edit quality takes into account the number of the modified words, the reputation of the editor, and the quality score of the previous edit. Other approaches to assess article quality (Stvilia et al 2005; Dondio and Barrett 2007; Blumenstock 2008) use a combination of metrics (e.g., the number of words, characters, sentences, internal and external links). In October 2016, the in-house Wikipedia research team released (Halfaker and Sarabadani 2016) a content-based machine learning quality measurement procedure. This procedure is derived from an algorithm by Warncke-Wang et al (2015) and comes with a web-service API (referred as Objective Revision Evaluation System[2]). Both provide a quality score for each article edit. In our work, we utilized this dataset to assess the evolution of quality in tandem with social processes.

Due to the huge amount of user-generated content on Stack Overflow (and other just-in-time sites), it is important to provide an effective quality control mechanism of such content to recognize useful content and expert users. This problem has been investigated thoroughly by past research. Two research strategies have been developed: one focuses on the quality of answers (Shah and Pomerantz 2010; Burel et al 2012), and the other on the quality of questions (Anderson et al 2012; Ravi et al 2014; Arora et al 2015). Regarding the first category, Shah and Pomerantz (2010) proposed a regression model for evaluating the quality of answers on the Yahoo! Questions. The model takes into account a combination of content-based features (e.g., the length of the content of answers, references within answers) and community feedback based features (e.g., the number of comments, ranks of answers). A logistic regression model was trained on these features to predict a model for the answer scores. Question quality was explored by a content-only approach (i.e., combinations of textual and topic modeling features) (Ravi et al 2014). On the other hand, Stack Overflow developed its scoring system for the posts based on community members' feedback. In particular, Stack Overflow allows users to provide their feedback on the questions, answers, comments by either voting up or down.[3] In our work, we rely on this scoring system to predict the quality of the content (posts).

3 Dataset and methodology

3.1 Datasets

We focus on two prototypical sites for the two modes of production: Wikipedia for just-in-case production, and Stack Overflow (the software programmers' community in the Stack Exchange[4] network) for just-in-time production.

On Wikipedia (and in just-in-case wiki-like sites), groups of individuals come together around a topic, building it up into an integrated information stack complete with references, links, summaries, visual illustrations, and data. Editorial interaction is loose and free-flow. Work is performed on the same material, which grows by accretion and iterative editing. Contributions are ambiguously normed (Matei and Dobrescu 2010), but they are mutually editable, which creates an ad hoc process of editorial supervision.

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

[2] https://ores.wikimedia.org/

[3] http://stackoverflow.com/help/privileges/.

[4] http://stackexchange.com/sites.

Table 1 Wikipedia dataset statistics

# of registered users	234,371,732
# of pages	39,450,659
# of revisions	525,034,797
Revisions per page	Avg.: 13.3, Median: 2.0, Max: 1,175,197
Unique users per page	Avg.: 5.9, Median: 2.0, Max: 108,852
Edited pages per week	Avg.: 291,961.2, Median: 364,286, Max: 1,004,511
Revisions per week	Avg.: 630,292.4, Median: 800,418, Max:1,866,275
Unique users per week	avg.: 33,594.5, Median: 43,551, Max: 70,552
Dataset period	Jan 2001 August 2016

Table 2 Stack Overflow dataset statistics

# of registered users	2,939,880
# of questions	12,209,179
# of answers	19,646,266
# of comments	50,703,120
Questions per user	Avg.: 4.2, Median: 1.0, Max: 2097
Answers per user	Avg.: 6.7, Median: 1.0, Max: 33303
Comments per user	Avg.: 17.2, Median: 1.0, Max: 72889
Users per week	Avg.: 42,187.3, Median: 44,400, Max: 82,134
Questions per week	Avg.: 28,552.5, Median: 31,570, Max: 54,663
Answers per week	Avg.: 46,276.3, Median: 54,340, Max: 77,808
Comments per week	Avg.: 118,822.4, Median: 134,260, Max: 219,985
Dataset period	July 2008 July 2016

To obtain the Wikipedia data for entropy and unevenness, we processed the Wikipedia archived[5] files released in September 2016. The dataset contains around 40 million articles where each article has a sequence of historical edits. The total number of edits for Wikipedia articles exceeds 500 million, and these edits were performed by more than 234 million registered users (i.e., we ignored any contribution performed by anonymous users, which represent a minority of edits).

On Stack Overflow (and in other just-in-time sites), interaction takes place in a more tightly scripted manner—typically according to a question–answer pattern. First of all, information is solicited by a specific individual, who expects a specific answer. Answers or comments are provided in return by site members. Although answers, questions or comments are editable, generally, information is created in discrete units and is kept as such.

To obtain the Stack Overflow data, we processed the archived files from the Stack Exchange platform[6] released in

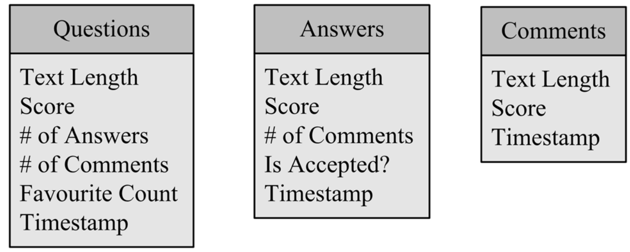


Fig. 1 Parameters describing the different types of posts in Stack Overflow

July 2016. The dataset has a total 82.6 million posts consisting of 12.2 million questions, 19.7 million answers, and 50.7 million comments which were contributed by 2.9 million users (by the Stack Overflow site's rules, all contributions are performed only by registered users).

Tables 1 and 2 show the statistics about the Wikipedia and Stack Overflow datasets, respectively.

3.2 Processing methods

For our analysis, we processed each dataset to quantify user contribution and content quality at the global level, using weeks and months as analysis periods. The two datasets were processed using a Java program executed on high-performance computing clusters. The program was run on a cluster of two nodes, each with 16 cores and 64 GB memory.

3.2.1 User contribution

To quantify the contribution of each user in each period, we developed two methods (i.e., one for Wikipedia and another for Stack Overflow) which are theoretically comparable. While different in computation details, the two methods converge in that they aim at capturing the amount of contribution for each user after considering the different weights in the collaborative process of each contribution.

For Wikipedia, we examined each article edit (abbreviated as u) and evaluated the amount of user contribution by considering the number of characters added (abbreviated as A), deleted (abbreviated as D), or modified (abbreviated as M) compared with the preceding edit (abbreviated as v). The number of modified character is calculated using the edit distance (Adler et al 2008) to measure the total amount of relative change in text position and structure. As a result, the user contribution is formally defined through the contribution delta formula:

$$d(u, v) = \max(A, D) - 0.5 \times \min(A, D) + M. \tag{1}$$

On Wikipedia, we have multiple edits for each article; hence we can measure the user contribution by simply computing the difference between two subsequent edits using the delta

5 https://dumps.wikimedia.org/enwiki/20160901/.

6 https://archive.org/download/stackexchange.

formula (Eq. 1). By contrast, in Stack Overflow, collaboration is characterized by different types of contribution on a single topic, namely: question, answer, comment. Hence, we need a different method to measure the user contribution, which takes into account post types. Our approach considers a set of aggregated parameters for each post type (see Fig. 1). Before calculating the user contribution, the parameter values for each post are first standardized (i.e., normalized) by using z-scores. After standardizing all values, we calculated the weight for each parameter using factor analysis. Subsequently, the weighted linear combination of the parameter values was used to estimate the significance of each post. The significance of a post referred to the importance and value of the post. As answers are related to questions and comments are related to questions or answers, the significance of an answer or comment is weighted by the significance of its corresponding predecessor (e.g., the significance of an answer is weighted by the significance value of its corresponding question). Moreover, because significance decreases over time, the contribution (question, answer, or comment) is weighted by a temporal decay factor. In our analysis, we used the half-time decay formula to estimate the temporal value of the post. Subsequently, user contribution is a summation of the significance of all his or her posts as shown in equation Eq. (2).

$$\begin{aligned}\mathrm{UserContribution}(u) &= \sum_{i=1}^{N_Q} \mathit{Significance}(Q_i)\\ &+ \sum_{i=1}^{N_A} \mathrm{Significance}(A_i)\\ &+ \sum_{i=1}^{N_C} \mathrm{Significance}(C_i)\end{aligned} \tag{2}$$

3.2.2 Content quality

For Wikipedia, content quality is derived from the Wikimedia Foundation dataset of article quality (Halfaker and Sarabadani 2016). The dataset predicts the quality of each article created on Wikipedia since 2001 at a monthly level. Prediction is performed via trained machine learning. Objective quality features, such as article length, the number of references, the number of headings, information richness, and the number of functional links, are used to predict quality. Quality values go up to 5. The highest score value indicates the best-quality article. For Stack Overflow, the quality of the post is directly evaluated by all those involved in the knowledge production or consumption. Evaluation is done by a simple voting system. Higher quality is implied by higher votes. Answers and comments are evaluated by the number of votes, while the questions are assessed by both the number of votes and favorite counts.

3.3 Analysis measures

3.3.1 Social unevenness, elite stickiness

Once the amount of content contribution is defined and measured for each intervention, for each user and each site, we calculated contribution evenness at a weekly level and elite stickiness, also at week level.

Content unevenness at weekly level is calculated both for Wikipedia and for Stack Overflow using entropy applied for each site-specific metric for contribution, as defined in (Eq. 1) for Wikipedia and (Eq. 2) for Stack Overflow.

$$H(X) = -\sum_{x} p(x)\log_2 p(x) \tag{3}$$

We chose the entropy measure for social unevenness (using the canonic entropy formula (Shannon and Weaver 1998) shown in Eq. (3)) since it was shown that entropy rather than other measures (e.g., Gini coefficient) is an effective measure for inequality and structuration (Bailey 1985). The entropy is maximized when all contributors contribute evenly, and it is minimized when one contributor is responsible for all the content. As entropy values increase with the size of the population, we normalize each weekly entropy value by the logarithm of the total number of users, which is the maximum possible entropy. This brings all entropy within a 0–1 metric, making the weekly values comparable in time.

Elite stickiness, both for Wikipedia and Stack Overflow, also varies between 0 and 1 on a weekly basis. It simply measures the degree to which members of the top 1% contributing group in an immediately previous period were in the top 1% contributing group during the current period. A score of 1 indicates that elite stickiness is 100%.

3.3.2 Cumulative average of content quality

Content quality at weekly level is aggregated for all interventions for both Wikipedia and Stack Overflow using the cumulative average. This means that for each week, the quality value reflects the average of the quality for all the work done up to that point. For Wikipedia, the cumulative average was calculated in a similar manner. However, the data were initially calculated monthly due to the fact that the data are provided at monthly level; then the values were interpolated linearly accordingly to the weekly level. At the same time, entropy and stickiness were averaged at weekly levels, to keep the data within the same temporal scale.

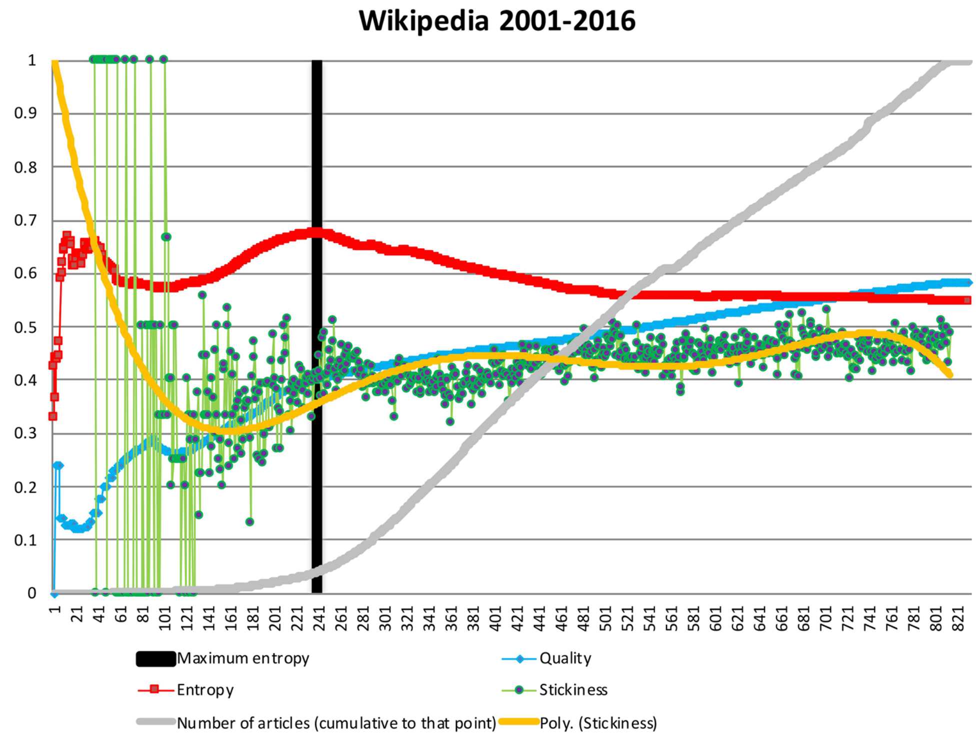


Fig. 2 Evolution diagram for Wikipedia 2001–2016. Blue line—quality, red line—entropy, green dotted line—stickiness, grey—number of articles, yellow—polynomial trend line for stickiness. The numbers on the *x*-axis indicate the ordinal number of the weeks

3.3.3 Subsetting for follow-up analysis

One of the research questions of this study addresses the differential effect of social production factors on quality within specific groups of articles. We focused on representative samples of articles from Wikipedia and Stack Overflow that meet two criteria: high vs. average quality and high vs. average editorial contributions. Specifically, we selected four random samples of 16,000 Wikipedia articles and Stack Overflow questions, each. In total, we looked at 74,000 Wikipedia articles and 74,000 Stack Overflow questions.

"High quality" Wikipedia articles were at least one standard deviation above the overall quality mean, while "average quality" articles had quality values of of standard deviation around the mean. The 16,000 articles for each group were randomly selected from all articles that meet one of the two criteria. Similarly, high editorial activity articles had to meet a threshold of at least 1 standard deviation above the mean number of edits for all Wikipedia articles, while "average editorial activity" articles were randomly selected from the group of articles that meet the threshold range of 1 standard deviation around the mean.

For Stack Overflow, similarly, we used the "quality" measure described in Sect. 3.2.2 and number of answer and comments of the questions. We used the same standard deviation threshold to select, in the same way, random samples of high and average quality and editorial activity posts (questions).

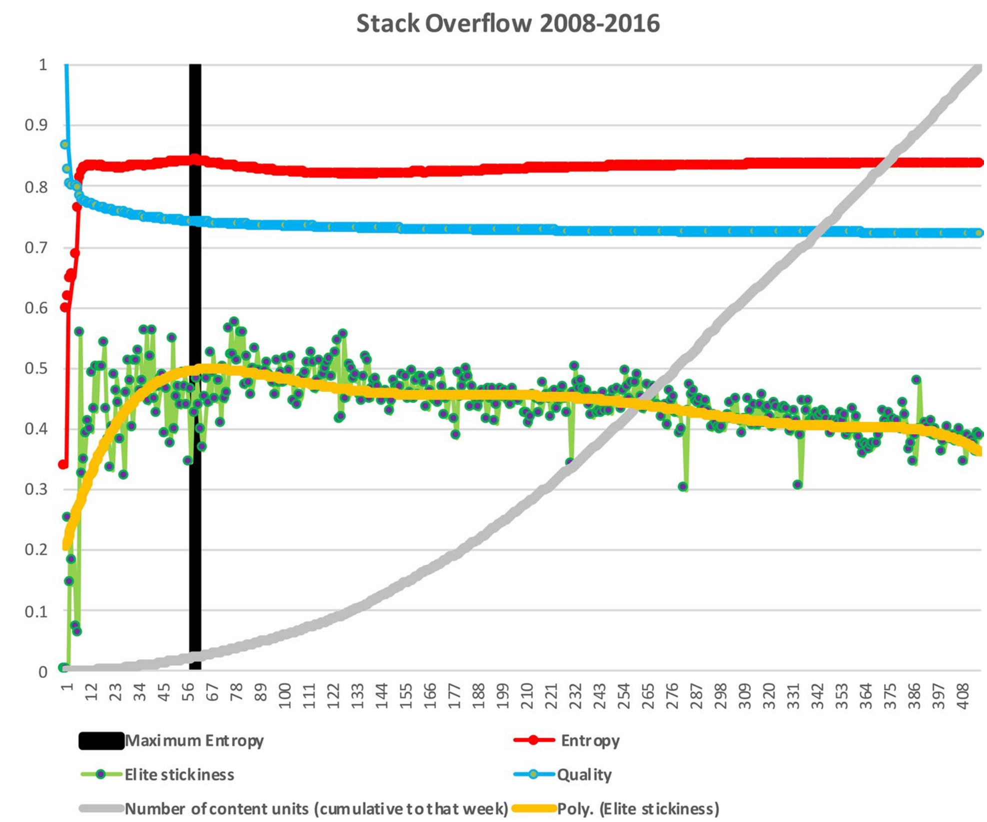
Because the subsets of high quality and high editorial activity articles were obtained by random sampling of articles from various time periods, elite stickiness was consistently zero. In other words, because the editorial activity does organically overlap across articles, the likelihood that we will capture the same individuals across the entire period was very small, almost zero. Thus, elite stickiness is not used in this context. For all four samples we only calculated quality and entropy.

4 Analysis results

4.1 Overall data analysis

We start the analysis by tracking the evolution of the core variables for Wikipedia and for Stack Overflow for the entire period during which the two projects were in existence. For Wikipedia, this included the period January 2001 August 2016 and for Stack Overflow July 2008 July 2016. See Figs. 2 and 3, in which the data are displayed by week number, not by calendar date. The average normalized entropy for Wikipedia is 0.59 and for Stack Overflow is 0.83. More directly, the top 1% Wikipedia users are, overall, responsible for 87% of the content, while on the top Stack Overflow contributors are responsible for 45% of the content. The findings for Wikipedia show an increase compared to 77% reported in Matei and Britt (2017) due to

Fig. 3 Evolution Diagram for Stack Overflow 2008–2016. Blue line—quality, red line—entropy, green dotted line—stickiness, grey—number of articles, yellow—polynomial trend line for stickiness. The numbers on the *x*-axis indicate the ordinal number of the weeks



the fact that the data reported in this paper include 5 extra years, during which the content has become even more unevenly distributed.

Upon mapping Wikipedia entropy (contribution unevenness, see formula 2), elite stickiness, the number of articles (normalized to 1) and quality, we obtain the evolution diagram shown in Fig. 2, which covers the years 2001–2016. Numbers on *x*-axis represent week numbers since the founding of the site. This, as all other temporal charts start at one, representing the first week in the existence of the project. As we can see from the diagram, entropy and stickiness tend, after a period of wild variation, to stabilize. This holds true even and especially when content creation (purple line) follows an explosive, exponential growth trajectory after week 270. At the same time, quality (blue line) increases at a steady pace.

As we can also see from Fig. 2, the first 3 years of editorial activity on Wikipedia were characterized by intense and wide variations in all three dimensions designated by the blue (quality), red (entropy) and green (stickiness) curves. The yellow curve is the best fit polynomial trend line for stickiness. Entropy and stickiness, especially, present a cyclical evolution, with ups and downs determined by changes in the history of the site. However, after a point of maximum entropy in week 239, entropy follows a smooth declining slope, while quality and stickiness increase. Simultaneously, content production increases at a sustained faster pace.

To evaluate the impact of stickiness and entropy on quality we performed a first linear regression analysis on differentiated values (which eliminates temporal autocorrelations) at a monthly level for all time periods. The results indicated only a strong and negative correlation between entropy and quality. As Wikipedia became more top-heavy and entropy declined, quality increased (beta = − 3.025, $p < .01$). Elite stickiness, however, did not impact quality (beta = .009, $p = .17$). The R-square was a significant .73, suggesting that over three-quarter in the variability of the dependent variable (quality) was explained by entropy.

Focusing on this last period (132 monthly data points), after entropy reaches a maximum (week 239), we performed a second analysis to regress quality on entropy and stickiness again. For this interval, there was a very slight positive effect for elite stickiness and none for entropy. Stickiness increase is positively correlated with a quality increase (beta = 0.0016, $p < .05$, R-square = .02). Given our theoretical concerns about the possibility that entropy and stickiness may, in fact, be curvilinearly associated with quality, we ran a final model in which we introduced both linear and quadratic terms for both entropy and stickiness.

The results uncovered, indeed, a curvilinear effect. The quadratic model that we ran has a healthy R-square value of .83, which shows that this is a model superior to the one previously tested and should be retained as the best fit. We also detected a clear effect for entropy, in both the quadratic and

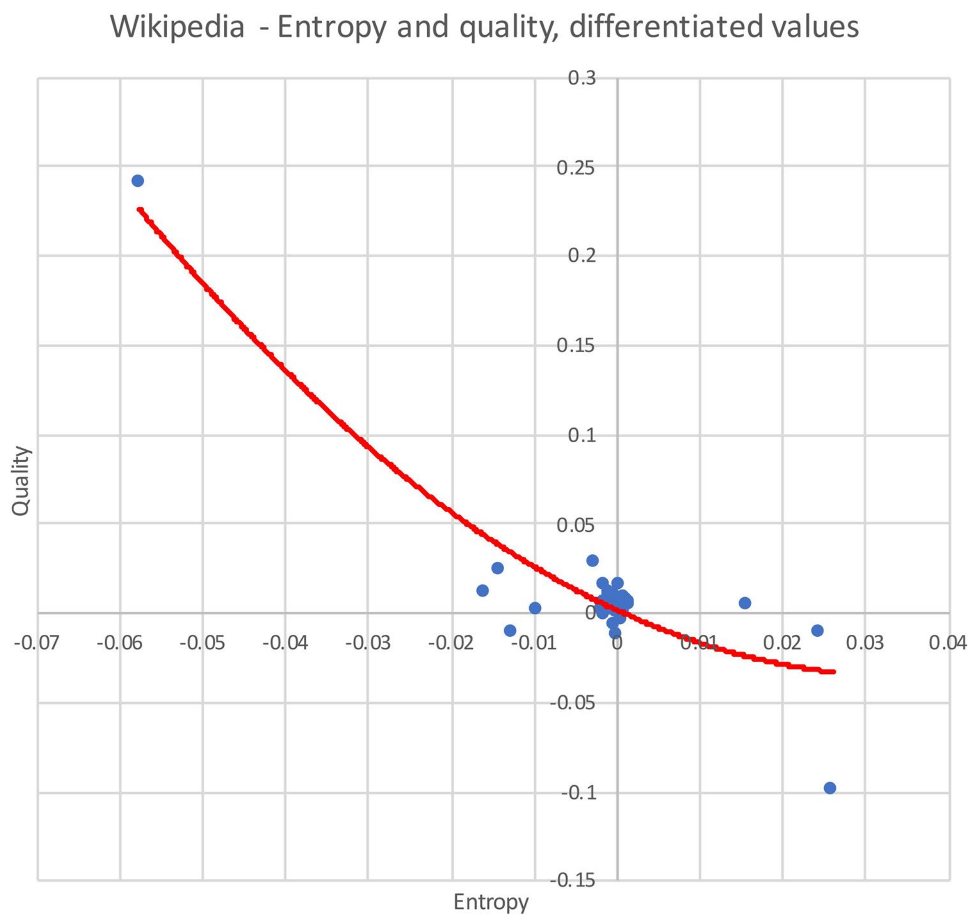


Fig. 4 Wikipedia—entropy and quality, differentiated values

linear terms. Furthermore, while the linear term for entropy on Wikipedia was negative (− .22, $p < .01$), the quadratic term was positive (.08). This indicated that higher period-to-period decline in entropy is associated with higher quality.

Yet, the decline has greater effects at the negative end of the spectrum (to the left of the vertical axis, Fig. 4). In other words, there is a stronger effect of the decline in entropy on quality when entropy is in decline from period to period. When entropy increases from period to period, the effect on quality levels off. Even upon removing the outlier in the upper left quadrant, although the relationship becomes linear, the trend does not change, while the beta remains negative and commensurate with that detected previously (− .24) and significant ($p < .01$) (Fig. 5).

Furthermore, the results indicated a small positive effect for elite stickiness on quality (beta = .016, $p < .01$) for the period following maximum entropy. Stickier elites lead to better content. The quadratic term for stickiness was not significant, indicating that stickiness is only linearly and positively associated with quality. As stickiness increases, quality increases uniformly, not curvilinearly.

The analysis performed on the Stack Overflow dataset presents some similarity, but also some notable differences to the Wikipedia analysis (see Fig. 3). Again, the chart x-axis

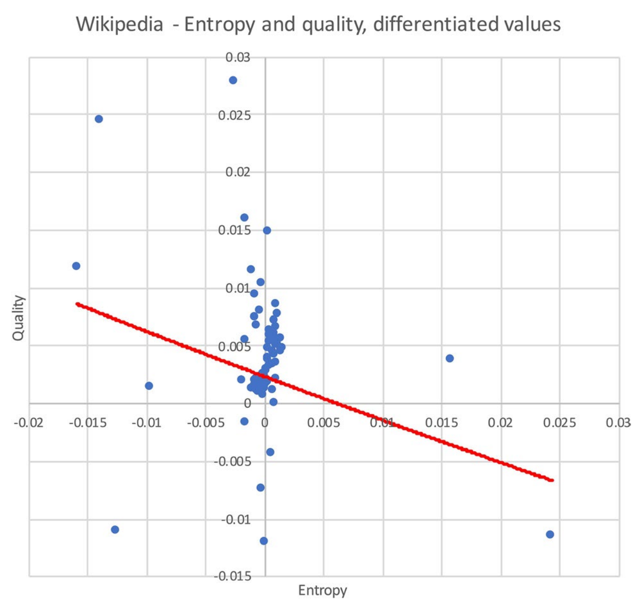


Fig. 5 Wikipedia—entropy and quality, differentiated values after removing outlier

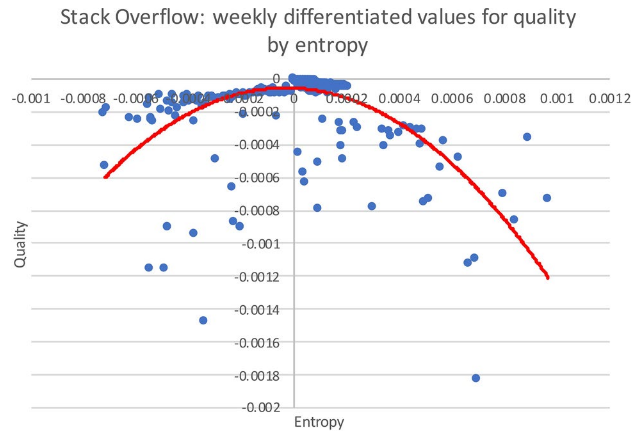


Fig. 6 Stack Overflow—weekly differentiated values for quality by entropy

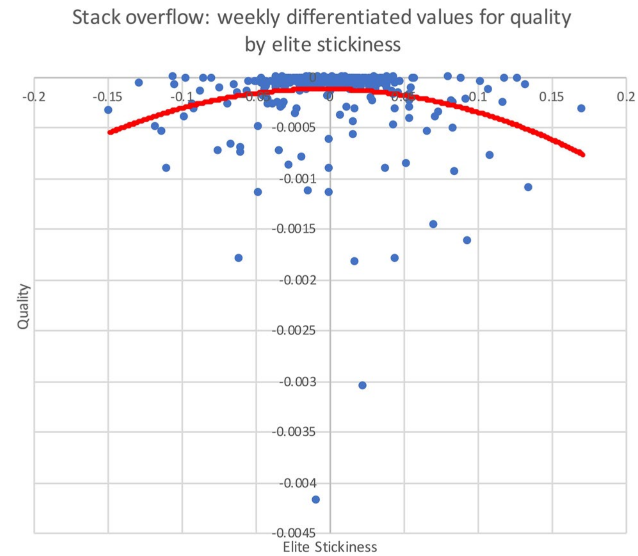


Fig. 7 Stack Overflow—weekly differentiated values for quality by elite stickiness

starts at one, the first week of the project existence in 2008 and covers over 400 weeks, until 2016. First, at a purely descriptive level, we notice that entropy fluctuates in the first 3 years far less on Stack Overflow than on Wikipedia. Thus, it stabilizes earlier and remains in a steady state for most of the analyzed period. At the same time, entropy is consistently higher and there is a slight but steady upward drift in entropy, suggesting a decrease in unevenness. Stickiness, in turn, declines in Stack Overflow throughout the analyzed period, from over 50% in the first 2 years, to under 40% toward the end. The trend is opposed to that observed on Wikipedia, where stickiness tends to increase from under 40% to as high as 50% toward the end of the analyzed period (2016).

Linear regression analysis of differentiated values, which controls for autocorrelation, was performed for Stack Overflow as well, both for the entire period and for the interval after entropy reaches a maximum (week 59). The results for the whole period indicate a negative effect for entropy. Considering all 8 years of Stack Overflows life, an increase in weekly differentiated entropy values leads to a decline in differentiated weekly quality values, and vice versa (beta $= -.05$, $p < .01$, R-square $= .87$). There is no effect for elite stickiness. Furthermore, focusing on the period after entropy reaches a maximum (week 59), entropy remains significant, but somewhat surprisingly changes direction, from negative to positive. As entropy increases, after this period, quality increases (beta $= .22$, $p < .01$, R-square $= .56$).

Due to the theoretical reasons regarding curvilinearity stated above and to the reversal in effect for entropy in the previous analysis, we tested for the possibility of a curvilinear association between both stickiness and entropy. A quadratic model, which took into account linear and quadratic terms for both entropy and stickiness, was applied to the Stack Overflow dataset. Results indicate that both terms (linear and quadratic) for each variable are significant and negative. The R-square value improves significantly to 0.92. This suggests that the better explanatory model should consider curvilinearity.

Starting with entropy, quality declines both when entropy increases and decreases in the extreme from week to week (see Fig. 6). This is also true for stickiness (see Fig. 7). Faster decline or increase in stickiness change quality in a negative direction. In other words, there seems to be a range of values where both entropy and stickiness optimize quality. Thus, on Stack Overflow our theoretical curvilinear model maps quite well.

4.2 High vs. average quality and editorial activity samples analysis

Overall, the co-evolution of social structural processes with quality across high vs. average quality or editorial activity articles indicates very different processes across just-in-time vs. just-in-case production systems.

Analyzing the subsets of high vs. low quality and editorial activity articles and posts, we notice several things (see Figs. 8a–d, 9a–d in which the information is displayed by week number, as before). While entropy of contribution is high and stationary on Stack Overflow, on Wikipedia it presents a clear phased development. In other words, the distribution of effort across types of content on Stack Overflow is undifferentiated. On Wikipedia, on the other hand, unevenness (entropy) follows a phased trajectory (Fig. 8a–d), which explains the differences between the early and later segments of the curves. After a period of low entropy and high variability, entropy increases, to reach a maximum.

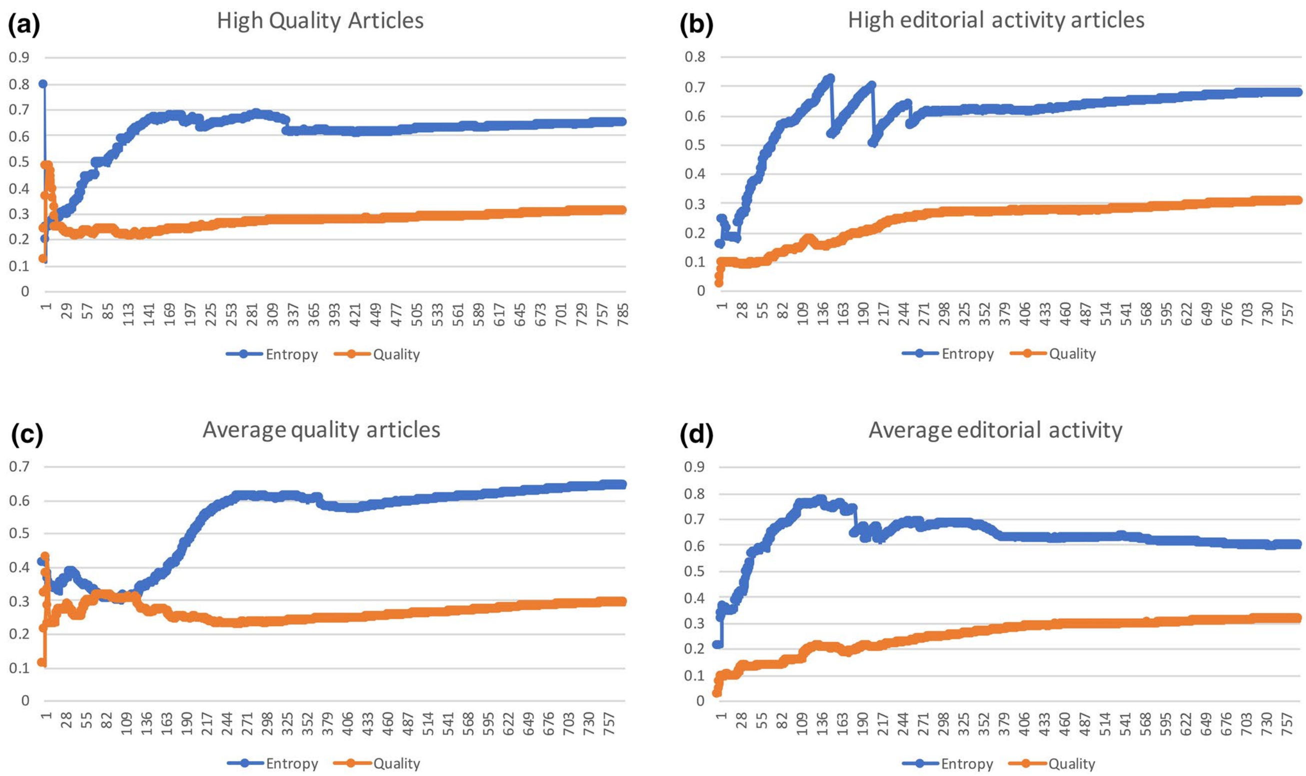


Fig. 8 Wikipedia overall relationships: entropy (blue) vs. quality (orange) in Wikipedia subsets

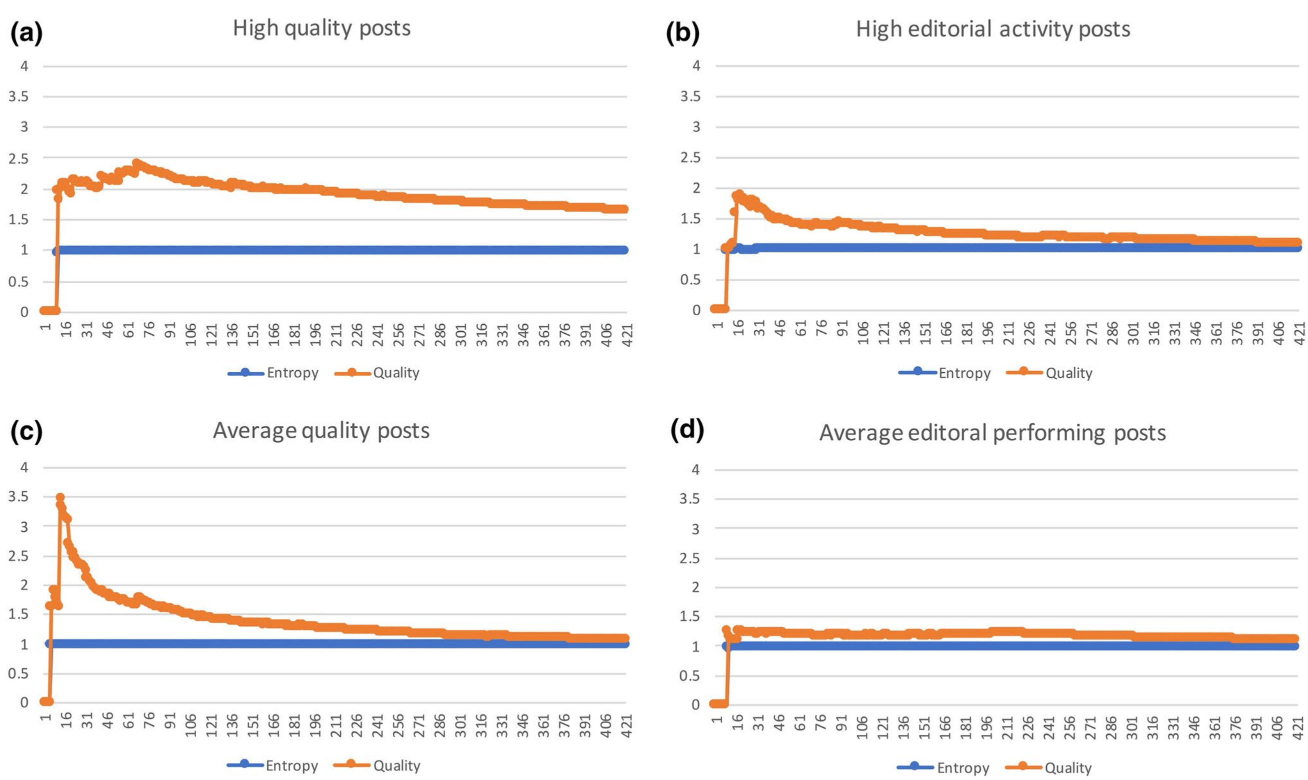


Fig. 9 Stack overflow overall relationships: entropy (blue) vs. quality (orange) in stack overflow subsets

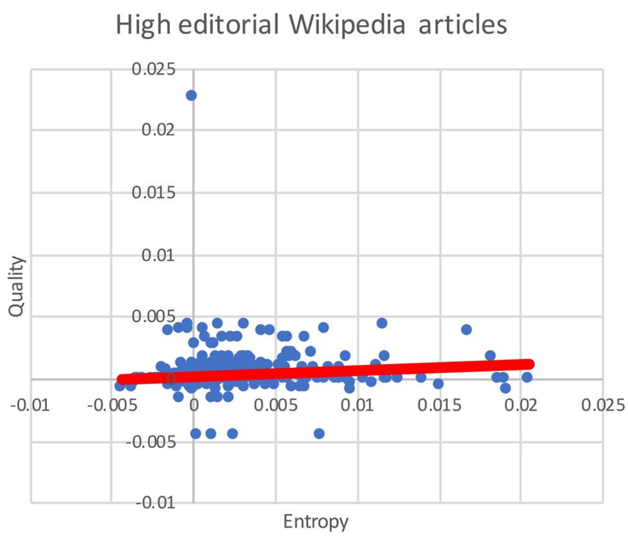


Fig. 10 High editorial activity Wikipedia articles

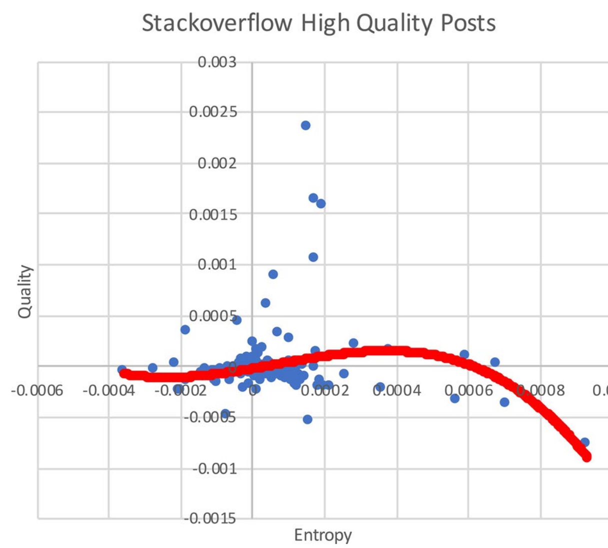


Fig. 11 High-quality stack overflow posts

Finally, entropy stabilizes on a flatter trajectory, which drifts, from case to case, slightly higher or lower. Quality, at the same time, after a temporary fluctuation for average quality articles, increases uniformly across all groups. The apparent discontinuities during the first phase of development on Wikipedia, especially in Fig. 8b and to a lesser extent in Fig. 8d, are the product of random sampling of articles. More interesting, while quality on Wikipedia has increased for the most part, especially in the last few years, on Stack Overflow we notice a decline in the metric-based quality of the articles (Fig. 9a-d). On average, Wikipedia articles increase in objective indicators of quality, including sourcing and structure. On Stack Overflow, the average score of questions and answer declines to a certain extent, suggesting a regression to the mean and possible more random distribution of quality across content.

Thus, the descriptive data indicate different association processes across the to knowledge production spaces which demand asking both broader and specific questions. Broadly the questions are: "How do changes in entropy co-evolve with changes in quality?" and "What is the significance of these differences?" These are to be addressed by asking the more specific and operational question if marginal week-to-week increases or decreases in entropy are connected with marginal increases or decreases in quality.

4.2.1 Wikipedia subsets

High vs. average quality articles

Regression of the differentiated data sampled from Wikipedia did not detect any relationship between entropy and quality in high quality or average quality articles. Thus, increase or decrease in unevenness on a weekly basis does not impact quality in these subsets of articles.

High vs. average editorial activity

However, for high editorial activity articles quality increases match increases in entropy (beta = .221, $p < .05$) (see Fig. 10). This is not true for average editorial quality articles. In other words, for articles that are edited a lot marginal increases in collaboration evenness contribute to marginal increases in quality. This should be taken with caution, though, since the amount of variance explained is very small (r-square = .012).

4.2.2 Stack overflow subsets

High vs. average quality posts

For Stack Overflow high-quality posts, however, we found a curvilinear cubic relationship, albeit one that explains small amounts of variance in the dependent variable (r-square = .08) (see Fig. 11). Extreme marginal increases or decreases in entropy lead to declines in quality. Quality is only maximized for mid-range decreases in entropy. Most important, the most pronounced decline takes place when entropy increases marginally.

There is no relationship between entropy and quality for average quality posts.

High vs. average editorial activity

Regarding high editorial activity on Stack Overflow posts, we detected a small (beta = − .14, $p < .05$) but significant relationship between marginal increase in entropy and marginal decrease in quality (see Fig. 12). In other words, concentration of effort from week to week leads to marginal increases in quality. Like before, the amount of variance explained is very small (r-square = .02).

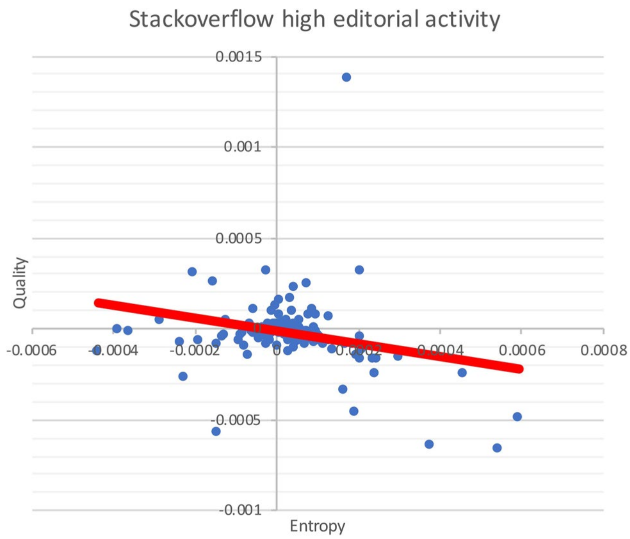


Fig. 12 Stack overflow high editorial activity posts

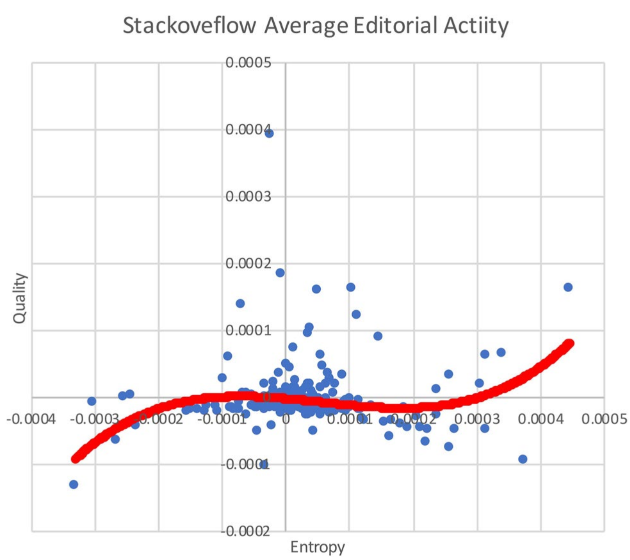


Fig. 13 Stack overflow average editorial activity posts

This is not the case, however, for average editorial activity Stack Overflow posts, which seem to increase in quality as entropy increases. There is a cubic curvilinear relationship between entropy on quality in average editorial activity posts (beta = .415, $p < .01$) (see Fig. 13). The increase in quality in these articles takes place especially when increases in entropy are toward the higher end of the spectrum

5 Conclusions and future work

Our study aimed at assessing how the presence of small contribution elites (the top few percent contributors) on social media knowledge sites may impact the quality of the content. We considered both the degree of contribution unevenness, measured through entropy at period-level (weeks or months), and the degree to which the members of the contribution elite (the top 1% contributors) are stable in time (elite stickiness). Our main argument is that both content production concentration and elite stickiness may, indeed, influence quality content. We also advanced the proposition that both unevenness and entropy may be curvilinearly associated with quality. Finally, in addition to analyzing the overall association between unevenness (entropy) and quality, we looked at specific sub-categories of article and posts. We analyzed four samples of 16,000 articles or posts for high vs. average quality, and high vs. average editorial activity extracted from Wikipedia and Stack Overflow.

A first thing that can be noticed from Figs. 2 and 3, which display the evolution of the two study sites in terms of entropy, elite stickiness, and quality is the phased development of each site. After a period of fluctuation, entropy reaches on both sites a point of local maximum (week 238 for Wikipedia and week 60 for Stack Overflow). Especially on Wikipedia, this initial period is one of periodic shifts, indicating movements in collaborative dynamics. A small but noticeable decline in entropy during weeks 40–60 on Wikipedia corresponds to a significant increase in quality, as expected. The next period (weeks 60-238) sees a reversal of the process, which ultimately stabilizes after week 238 into a steady and clear decline in entropy and an increase in quality and stickiness (see trends after black line marking maximum entropy in Fig. 2).

Although less obvious on Stack Overflow, phases are still present. Entropy and stickiness peak around week 60, after which they diverge: entropy goes up while stickiness goes down. More important, quality levels off with a slight downward drift. Stack Overflow reaches a quasi-steady state, which indicates both competition and a regression to the mean in terms of quality.

The goal of the study was to go beyond descriptives and to analyze the association between stickiness, entropy, and quality.

Let us start with the proposition that there might be curvilinear relationships between entropy, stickiness, and quality. The ultimate goal is to ascertain if too much or too little unevenness or elite stickiness may affect the global quality of the content differently than when present in moderation.

We found this to be particularly true for Stack Overflow. On Wikipedia, however, only entropy had a weak curvilinear behavior, which can be in fact seen as a simple negative correlation, while elite stickiness tended to be linearly associated with quality. Furthermore, on Wikipedia, the curvilinear association for entropy was a mere variation of a linear association, in that it only captured a slow decline in the effect for entropy on quality at the positive end of the spectrum (period-to-period increases in entropy). In fact, if ignoring the outlier in the upper left quadrant of Fig. 4, the linear association becomes stronger. For Stack Overflow, we have full blown effects both for entropy and elite stickiness because the content is optimized for moderate period-to-period value changes in both variables.

More pointedly, and addressing the question in the title of our study, sticky elites are thus important factors in determining the quality of the content, although in a differential manner. For Wikipedia, sticky elites lead to better content linearly, while on Stack Overflow, there is a limit to what elite stickiness can do to improve content quality. In practical terms, stickiness is to be encouraged without restriction on wikis or other just-in-case knowledge production systems and moderately encouraged on just-in-time knowledge production systems.

Theoretically, the findings tell an interesting story about the way in which social dynamics influence content quality. Elite stickiness can only have positive effects in just-in-case (Wikipedia) knowledge production. As individuals become more vested and more involved with the site, content quality increases uniformly. "Too much stickiness" concentration of effort, however, has a limited effect. While a decline in entropy and increase in concentration may positively affect quality, this slows down after a while and plateaus. Concentration seems to have a self-limiting effect.

On just-in-time systems (Stack Overflow), content quality is optimized within certain bounds of entropy and elite stickiness. Such systems are more sensitive to too much or too little elite concentration of effort or temporal concentration. The difference between Stack Overflow and Wikipedia might be driven by the fact that while Wikipedia is a truly collaborative space, where deep and direct collaboration is needed, Stack Overflow is a more competitive, individual-effort driven site. On Wikipedia, collaboration requires a deep and continuous involvement of the elites, who keep the knowledge production going. As Fig. 2 shows, as a general trend across all 15 years of data, entropy ends up at a lower level than the one present at the beginning, while stickiness, after a period of decline, increases. Accompanying these trends, quality keeps increasing at a steady pace.

On the other hand, Stack Overflow is dominated by an ethos of competition, where individuals craft both questions and answers individually, for which they get specific scores that measure the quality of their work and reward participation. As all members compete to become high scores, those that are a part of the contribution elite are often pushed aside by newcomers, or they give up along the way at a higher rate than on Wikipedia. Elites thus decline in stickiness on Stack Overflow and, although ever so slowly, the site increases in entropy, as well. Quality on Stack Overflow seems to decline, too, as a consequence, albeit in small increments. Thus, Stack Overflow becomes over time more and more decentralized, with the cumulative assessments of participation and quality decreasing from the higher levels of the initial periods.

As a broader conclusion, we may say that a just-in-time knowledge production system, like Stack Overflow, is a divergent social system, while a just-in-case production systems, such as Wikipedia, is a convergent system. What converges or diverges is quality and social structure.

At the same time, as shown in Figs. 2 and 3, the quality on Wikipedia follows an increasingly steep slope, while Stack Overflow, slowly drifts down. The quality decline on Stack Overflow indicates that the site needs to trade off competition (and increased entropy) on lower average quality scores across the entire site. On the other hand, an increasingly concentrated site, such as Wikipedia, generates higher quality content.

At the same time, focusing on subsets of high vs. average articles or posts along two dimensions, quality and editorial volume, we found that high-quality articles on Wikipedia do not have their own "structural differentiation" signature (Matei and Britt 2017). Entropy and quality are not associated in this case. It does appear, however, that high editorial activity articles increase in activity if the contributions tend to be more evenly distributed. This should be put in the context of the finding that overall, higher entropy reduces quality. Thus, it appears that the minority of the articles that continue being edited due to currency or controversy do not benefit from constant changes. The more people change them, the more chaotic the process become and the lower the quality. This indicates that the much vaunted idea of the "wise crowds" is in fact highly problematic for controversial articles on Wikipedia.

Continuing the discussion of the relationship between entropy and quality across high vs. average quality and editorial activity on Stack overflow, we notice the the curvilinear patterns of association between entropy and quality across high quality and high editorial activity posts are very similar to those discovered in the entire Stack overflow space. Specifically, entropy is again curvilinearly associated with quality in high-quality posts, while average posts do not indicate such association.

Across high editorial activity posts we discover, however, a negative association between entropy and quality. As editorial unevenness decreases from week to week, quality increases and vice versa. In other words, questions or answers that are edited more, become better. This is quite different from the process we identified in Wikipedia articles, which decline with more editorial activity. The difference is intriguing, yet not surprising. Stack Overflow, as a more competitive environment, in which reputation is actively sought out (MacLeod 2014). It is also a space where topics are "owned" by rapid and continuous response to challenges, which raises the bar for newcomers (Slegers 2015). With more editorial activity quality increases only if fewer individuals make more changes. In other words, Stack Overflow is a more tightly curated space.

Of course, these conclusions are limited by our specific approach, which considered unevenness strictly in terms of quantitative measures of contribution. We did not operationalize the intrinsic value of the contributions by the importance of the topics. Neither did we attempt to measure interactions at the micro-level, which may better explain some patterns in contributions among both elite and non-elite members. We believe, however, that our study sheds important light on how, considering macro processes, some significant trends of association between input (work concentration and elite stickiness) may impact quality.

Future work includes validating curvilinear effects across a variety of other just-in-time (e.g., Quora) and just-in-case sites (e.g., Non-English Wikipedia and Wikia). Further work may also focus on improving the content quality measurement of Stack Overflow, which could weigh the votes by the weight of the participants in terms of their content scores. Finally, entropy can be assessed not only in terms of number, but also of kinds and relevance of contributions.

Acknowledgements The work reported in this paper has been partially supported by NSF under Grants IIS-1636891 and ACI-1547358, and by the US Army Research Laboratory and the UK Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the US Government, the UK Ministry of Defence or the UK Government. The US and UK Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

Adler BT, de Alfaro L, Pye I, Raman V (2008) Measuring author contributions to the Wikipedia. In: Proceedings of the 4th international symposium on Wikis, ACM, p 15

Anderson A, Huttenlocher D, Kleinberg J, Leskovec J (2012) Discovering value from community activity on focused question answering sites: a case study of stack overflow. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 850–858

Arora P, Ganguly D, Jones GJ (2015) The good, the bad and their kins: Identifying questions with negative scores in stackoverflow. In: Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015, ACM, pp 1232–1239

Bailey KD (1985) Entropy measures of inequality. Sociol Inq 55(2):200–211

Baltadzhieva A, Chrupała G (2015) Predicting the quality of questions on stackoverflow. In: Proceedings of the international conference recent advances in natural language processing, pp 32–40

Blumenstock JE (2008) Size matters: word count as a measure of quality on Wikipedia. In: Proceedings of the 17th international conference on World Wide Web, ACM, pp 1095–1096

Bruno R (2010) A democracy of unequals (Ph.D. dissertation)

Burel G, He Y, Alani H (2012) Automatic identification of best answers in online enquiry communities. The semantic web: research and applications, pp 514–529

Cross T (2006) Puppy smoothies: improving the reliability of open, collaborative wikis. First Monday 11(9)

Dondio P, Barrett S (2007) Computational trust in Web content quality: a comparative evalution on the Wikipedia project. Informatica, pp 151–160. http://link.galegroup.com/apps/doc/A168662927/AONE?sid=googlescholar

Halfaker A, Sarabadani A (2016) Monthly Wikipedia article quality predictions. https://goo.gl/vyMQtr

Jurczyk P, Agichtein E (2007) Discovering authorities in question answer communities by using link analysis. In: Proceedings of the sixteenth ACM conference on conference on information and knowledge management, ACM, pp 919–922

Kane GC (2011) A multimethod study of information quality in Wiki collaboration. ACM Trans Manag Inf Syst 2(1):4:1–4:16. https://doi.org/10.1145/1929916.1929920

Kittur A, Kraut RE (2008) Harnessing the wisdom of crowds in Wikipedia: quality through coordination. In: Proceedings of the 2008 ACM conference on computer supported cooperative work, ACM, pp 37–46

Kittur A, Lee B, Kraut RE (2009) Coordination in collective intelligence: the role of team structure and task interdependence. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, pp 1495–1504

Kuk G (2006) Strategic interaction and knowledge sharing in the KDE developer mailing list. Manag Sci 52(7):1031–1042

Liu J, Ram S (2011) Who does what: collaboration patterns in the Wikipedia and their impact on article quality. ACM Trans Manag Inf Syst 2(2):11:1–11:23. https://doi.org/10.1145/1985347.1985352

Liu X, Croft WB, Koll M (2005) Finding experts in community-based question-answering services. In: Proceedings of the 14th ACM international conference on Information and knowledge management, ACM, pp 315–316

MacLeod L (2014) Reputation on stack exchange: tag, you're it! In: 2014 28th international conference on advanced information networking and applications workshops, pp 670–674. https://doi.org/10.1109/WAINA.2014.108

Matei SA, Jabal AA, Bertino E (2017) Do sticky elites produce online knowledge of higher quality? In: Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017, ACM, New York, NY, USA, ASONAM '17, pp 72–79. https://doi.org/10.1145/3110025.3110040. http://doi.acm.org/10.1145/3110025.3110040

Matei SA, Britt BC (2017) Structural differentiation in social media: adhocracy, entropy, and the "1 % effect", 1st edn. Springer, Berlin

Matei SA, Bruno RJ (2015) Pareto's 80/20 law and social differentiation: a social entropy perspective. Public Relat Rev 41(2):178–186

Matei SA, Dobrescu C (2010) Wikipedia's neutral point of view: settling conflict through ambiguity. Inf Soc 27(1):40–51

Movshovitz-Attias D, Movshovitz-Attias Y, Steenkiste P, Faloutsos C (2013) Analysis of the reputation system and user contributions on a question answering website: Stackoverflow. In: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining, ACM, pp 886–893

Pal A, Chang S, Konstan JA (2012) Evolution of experts in question answering communities. In: ICWSM

Palloff RM, Pratt K (2010) Collaborating online: learning together in community, vol 32. Wiley, Oxford

Ravi S, Pang B, Rastogi V, Kumar R (2014) Great question! question quality in community Q&A. In: ICWSM

Shah C, Pomerantz J (2010) Evaluating and predicting answer quality in community QA. In: Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval, ACM, pp 411–418

Shannon CE, Weaver W (1998) The mathematical theory of communication. University of Illinois Press, Illinois

Slegers J (2015) The decline of stack overflow. https://hackernoon.com/the-decline-of-stack-overflow-7cb69faa575d

Stvilia B, Twidale MB, Smith LC, Gasser L (2008) Information quality work organization in Wikipedia. J Assoc Inf Sci Technol 59(6):983–1001

Stvilia B, Twidale MB, Smith LC, Gasser L (2005) Assessing information quality of a community-based encyclopedia. In: IQ

Warncke-Wang M, Ayukaev VR, Hecht B, Terveen LG (2015) The success and failure of quality improvement projects in peer production communities. In: Proceedings of the 18th ACM conference on computer supported cooperative work and social computing, ACM, pp 743–756

Wöhner T, Peters R (2009) 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, WikiSym '09, pp 16:1–16:10. https://doi.org/10.1145/1641309.1641333. http://doi.acm.org/10.1145/1641309.1641333

Wu G, Harrigan M, Cunningham P (2012) Classifying Wikipedia articles using network motif counts and ratios. In: Proceedings of the eighth annual international symposium on wikis and open collaboration, ACM, New York, NY, USA, WikiSym '12, pp 12:1–12:10. https://doi.org/10.1145/2462932.2462948. http://doi.acm.org/10.1145/2462932.2462948

Zeng H, Alhossaini MA, Ding L, Fikes R, McGuinness DL (2006) Computing trust from revision history. Tech. rep, Stanford Univ Ca Knowledge Systems LAB