Do Sticky Elites Produce Online Knowledge of Higher Quality?

Sorin Adam Matei [#1], Amani Abu Jabal [*2], Elisa Bertino [*3]

[#] *Brian Lamb School of Communication, Purdue University, West Lafayette, USA*

[*] *Dept. of Computer Science, Purdue University, West Lafayette, USA*

{[1]smatei,[2]aabujaba,[3]bertino}@purdue.edu

Abstract— Online knowledge production sites, such as Wikipedia or Stack Overflow, are dominated by small groups of contributors. How does this impact the knowledge production and its quality? Does the presence of some key contributors among the most productive members improve or not the quality of the knowledge, considered in the aggregate? The present paper considers 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 online knowledge production sites. The paper addresses such question by an extensive data analysis carried out on the datasets of two representative sites: Wikipedia and Stack Overflow. Results from the analysis show 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 decline on entropy, overall, and with increase in stickiness in the maturation phase, after an entropy peak is reached.

Keywords—Wikipedia, Stack Overflow, Unevenness, Elite Stickiness, Quality of Content

I. INTRODUCTION

Vast groups of widely dispersed online individuals have reshaped the lifecycle of knowledge production and consumption. Two types of knowledge production modalities have emerged: just-in-case and just-in-time. The first is associated with general-purpose reference sites (e.g., Wikipedia[1], Scholarpedia[2], and Citizendium[3]). The second is associated with specialized and immediate-need information identification and dissemination. Prominent among these are the Question and Answer sites (e.g., Stack Overflow[4], Yahoo! Answers[5], and Quora[6]). A critical question, in both types of knowledge production sites, is the quality of the knowledge produced, as good-quality knowledge is a key element for a wide adoption of a given production site. Quality, of course, may depend on many different factors, such as the skills of the contributors or topic maturity and currency. However more complex factors may influence quality. In this paper, we specifically address the question whether quality is correlated with the temporal dynamics of the social structures that power the online knowledge production systems.

With respect to the temporal dynamics of social structures, we expect some commonalities between just-in-case and just-in-time knowledge production sites, which may be explained by the voluntary and open nature of the collaborative process on both of them. On each type of site, decisions to contribute, interactions, and evaluations are voluntary. Furthermore, as demonstrated by previous research [14] [22], which showed a highly skewed distribution of contributions across contributors, both types of knowledge production have a definite social structure of collaboration. There is a "pecking order," which divides the production groups into top and bottom contributors. Previous research, which analyzed volume by amount of effort using a different metric than the one used here suggests that on Wikipedia the top 1% contributors may be responsible for more than 80% of the content [23] [22].

We claim, based on the results reported in this paper, that both just-in-case and just-in-time knowledge production sites are characterized by high rates of uneven contributions. They also have clear-cut leadership groups that tend to persist over time – we refer to this behavior as *elite stickiness*. Moreover, we claim that, at least initially, elite stickiness may be 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 on Question and Answer sites, just-in-time work may become less top heavy and, most important, over time their level of elite stickiness could decline.

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.

Furthermore, as uneven distribution and elite stickiness vary in time, content quality changes, as well. For theoretical and empirical reasons, presented below, there are reasons to believe that the relationship between unevenness and content quality is curvilinear. Uneven contribution rates might maximize quality within certain values. This might also be true for elite stickiness.

[1] http://www.wikipedia.org

[2] http://www.scholarpedia.org

[3] http://citizendium.org/

[4] https://stackoverflow.com/

[5] https://answers.yahoo.com/

[6] https://www.quora.com/

The argument on which we build these claims is that if groups have a core group of highly productive contributors and if they in time become "stickier," this will translate in the longer run into higher quality content [3][22]. Yet, there may be boundaries around this effect. Too heavy domination by a small group or not sufficient leadership can both negatively affect content quality [15].

Thus, the main goal of this paper is to determine the over-time evolution and association between collaboration unevenness, elite stickiness and quality in two representative just-in-case and just-in-time sites and to determine if increases/decreases in elite stickiness and unevenness of contributions are associated with 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. 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. The latter is represented by an entropy-based mathematical model which considers user effort. Our study uses a combination of metrics to better triangulate both social processes and content quality.

The paper is organized as follows. Section II discusses related work. Section III presents dataset summaries and the methodology including contribution unevenness, elite stickiness, and content quality. Section IV presents the main analytic results. Section V discusses the implications of the findings and outlines future work.

II. Related Work

Wikipedia is an open-source repository of reference knowledge. Articles are written just-in-case a reader needs a quick reference guide for 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 several 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 [26]. 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 [3] [5] [7] [9][10]. Cross [7] 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. In addition, Zeng et al. [5] devised a quality model for article edits based on a Bayesian network of the reputation of authors. The reputation of authors [5] 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 were developed [3] [9] [10] using a combination of metrics (e.g., the number of words, characters, sentences, internal and external links). In October 2016, a dataset [25] was released which captures the quality of all Wikipedia articles overtime. The data is generated by a content-based machine learning classification procedure derived from an algorithm by Warncke-Wang et al. [24]. A web-service API (referred as *Objective Revision Evaluation System* [27]) has also been developed, which provides a score for article edit quality. 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 in Stack Overflow (and other just-in-time sites), it is important to provide an effective quality control mechanism of such content in order 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 [13] [17], and the other on question quality [18] [20] [21]. Regarding the first category, Shah et al. [13] 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., 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 modelling features) [20]. On the other hand, Stack Overflow developed its own 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[7] or down[8]. In our work, we rely on this scoring system to predict the quality of the content (posts).

In parallel to the research focusing on the assessment of content quality on either just-in-case or just-in-time sites, there has been some work on identifying the evolution of social dynamics and structures, including contribution elites (i.e., high volume contributors), both on Wikipedia [22] and Stack Overflow [4] [8] [19]. For example, Matei and Bruno [22] have investigated the optimal level of inequality and leadership involvement on Wikipedia by the social entropy. Aditya et al. [19] have developed a machine learning method to distinguish experts in Stack Overflow and shown that using the temporal data the method finds experts more accurately. Approaches for connecting collaboration patterns and distribution of effort with content have also been proposed especially in knowledge production sites. Kittur et al. [12] 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.

[7] http://stackoverflow.com/help/privileges/vote-up

[8] http://stackoverflow.com/help/privileges/vote-down

Bruno [15], following Kuk [6], showed in experimental work with wiki groups that 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 [15] 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 in a curvilinear manner with uneven participation and elite stickiness.

A previous paper by Matei et al. [28] has analyzed the temporal evolution of Wikipedia social structures and elite emergence. The paper found that social structuration leads to "stickier elites." As groups become top-heavy or more dispersed in collaboration, elite stickiness (resilience in time) goes up and down. The previous study used an entropy based approach to detect social structuration. In this respect the research reported in [28] overlaps with the current study, providing some of its theoretical and methodological support. At the same time, compared with the previous paper, the current one presents several major differences. First, the previous paper only focused on Wikipedia and did not consider the temporal dynamics in just-in-time knowledge production sites, such as Stack Overflow; the current paper considers both just-in-case and just-in-time knowledge production sites in a comparative manner. Second, in the current paper the social evolution analysis uses a modified method for calculating effort and entropy. Third, the previous paper utilizes an abbreviated dataset, which stops in 2010. The current analysis includes 6 more years (2010-2016). Finally, the pervious paper utilized lagged regression, while the current one uses ordinary least square regressions. The current methodology looks at week-to-week net increase or decrease in the variables (which eliminates autocorrelations), while the previous methodology utilized absolute values. Thus, in the current paper, we report more complete results from a different angle. However, the new analysis confirms the results from our previous work.

In the current paper, we investigate how the relationship between contribution elites and the quality of user generated content evolves over the time and how they impact quality in just-case (Wikipedia) and just-in-time (Stack Overflow) sites. To the best of our knowledge, there is no comparative study which evaluates the social structure - quality relationship from a temporal perspective, both for Wikipedia and Stack Overflow.

III. Dataset and Methodology

A. Datasets

We focus on two prototypical sites for the two modes of production: Wikipedia, for just-in-case and Stack Overflow (the software programmers' community in the Stack Exchange[9] 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 [16], but they are mutually editable, which creates an ad-hoc process of editorial supervision.

To obtain the Wikipedia data for entropy and unevenness, we processed Wikipedia archived[10] 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 (we ignored any contribution performed by anonymous users, which represent of minority of edits).

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 201**6**

Table 2: StackOverflow Dataset Statistics

# of 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.: 28552.5, 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

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,

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

[10] https://dumps.wikimedia.org/enwiki/20160901/

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 Stack Exchange platform[11] released on 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.

B. 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 at Purdue Research Center. The program was run on a cluster of two nodes, each with 16 cores and 64GB memory.

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 [11] 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:

$$d(u,v) = \max(A,D) - 0.5 * \min(A,D) + M$$

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. 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 type. Our approach considers a set of aggregated parameters for each post type (see Figure 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 significance of each post was computed as a weighted linear combination of the parameter values composing 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. Moreover, because significance decreases over time, contribution (question, answer, or comment) importance 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 the following formula.

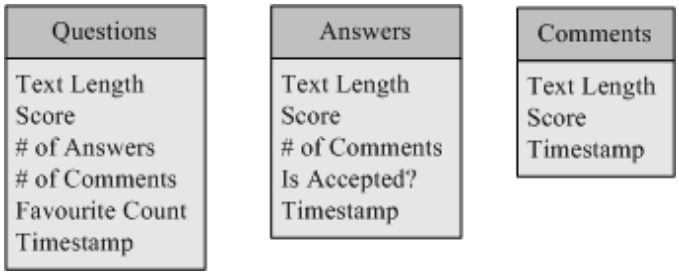

Figure 1: Parameters Describing the Different Types of Posts in Stack Overflow

$$UserContribution(u) = \sum_{i=1}^{N_Q} Significance(Q_i) + \sum_{i=1}^{N_A} Significance(A_i) + \sum_{i}^{N_C} Significance(C_i)$$

2) Content Quality

For Wikipedia, content quality is derived from the Wikimedia Foundation dataset of article quality [25]. The dataset predicts the quality of each article created on Wikipedia since 2001 at monthly level. Prediction is done via trained machine learning. Objective quality features such as length, number of references, number of headings, information richness, and 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. Hgher quality is implied by higher votes. Answers and comments are evaluated by the number of votes, while the questions are evaluated by both the number of votes and favorite counts.

C. Analysis Measures

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 weekly level and elite stickiness, also at week level.

[11] https://archive.org/download/stackexchange

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 above. The canonic entropy formula [1] is used:

$$H(X) = -\sum_{i=1}^{m} p(x) log_2 p(x)$$

We chose the entropy measure for social unevenness since it was shown that entropy rather than other measures (e.g., Gini coefficient) is an effective measure for inequality [2]. The entropy is maximized when all contributors contribute evenly and 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%.

2) Cumulative Average of Content Quality

Content quality at weekly level is aggregated for all interventions for both Wikipedia and Stack Overflow using cumulative average. For Wikipedia, cumulative average was first calculated monthly due to the fact that the data is provided at monthly level; then the values were interpolated linearly accordingly to the weekly level. Thus, entropy and stickiness were then averaged at weekly levels, to keep the data within the same temporal scale.

IV. Analysis Results

We start the analysis by tracking the evolution of our analysis measures both for Wikipedia and for Stack Overflow, for the entire period during which the two projects were in existence and we had recent data available for. For Wikipedia, this included the period January 2001 – August 2016 and for Stack Overflow July 2008 – July 2016. See Figures 2 and 4. 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.

Upon mapping Wikipedia entropy (contribution unevenness), elite stickiness, number of articles (normalized to 1) and quality, we obtain the evolution diagram shown in Figure 2. 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 follows an explosive, exponential growth trajectory (after week 270. At the same time, quality increases at a steady pace.

As we can also see from Figure 2, the first three years of Wikipedia editing 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 ever faster pace.

To evaluate the impact of stickiness and entropy on quality we performed a first linear regression analysis on differentiated values (which eliminates autocorrelations) at 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.

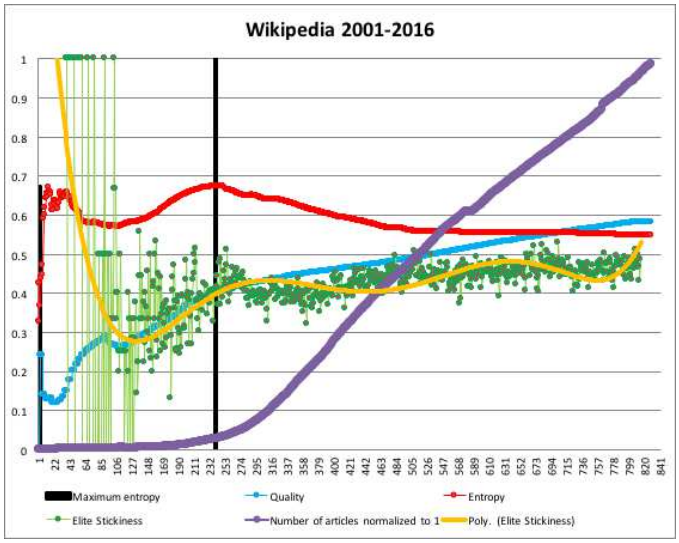

Figure 2: Evolution Diagram for Wikipedia 2001 - 2016

Focusing on this last period (132 monthly data points), after entropy reaches a maximum (week 239), we performed a second analysis to regress again quality on entropy and stickiness. For this interval, there was a very slight positive effect for elite stickiness and none for entropy. Stickiness increase is positively correlated with 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 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, Figure 3). Or, in other words, there is a stronger effect for 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.

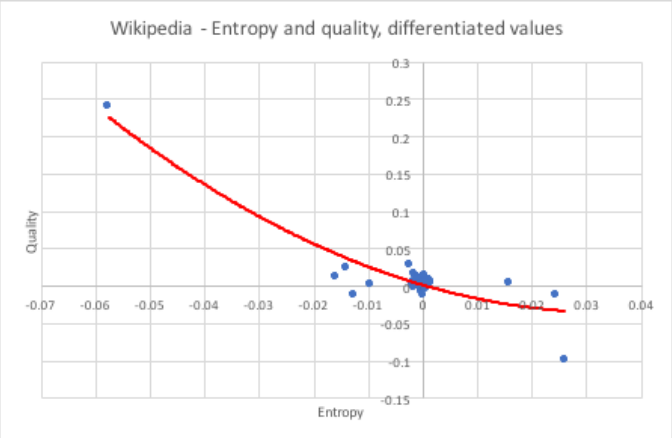

Figure 3: Wikipedia - Entropy and Quality, Differentiated Values

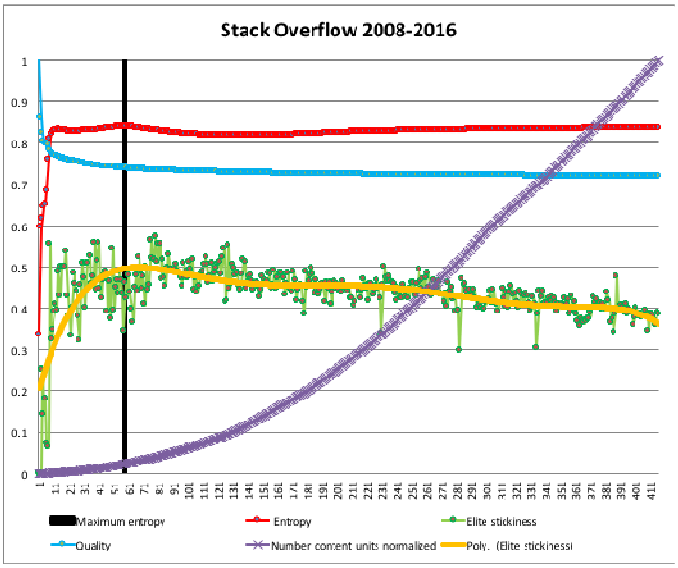

Figure 4: Evolution Diagram for Stack Overflow 2008 - 2016

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 Figure 4). First, at a purely descriptive level we notice that entropy fluctuates in the first two 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, 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 two 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).

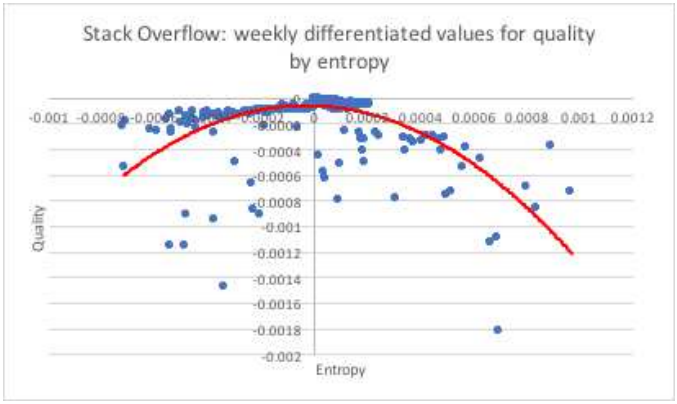

Figure 5: Stack Overflow- Weekly Differentiated Values for Quality by Entropy

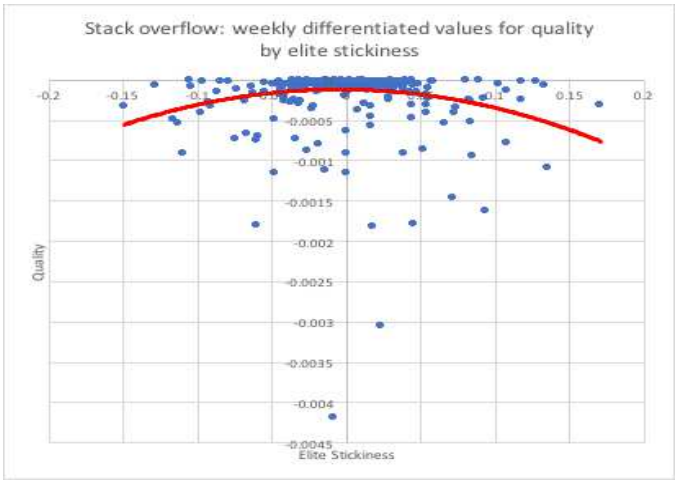

Figure 6: Stack Overflow- Weekly Differentiated Values for Quality by Elite Stickiness

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 entire period indicate a negative effect for entropy. Considering all 8 years of Stack Overflow's 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 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 take into account curvilinearity.

Starting with entropy, quality declines both when entropy increases and decreases in the extreme from week to week (see Figure 5). This is also true for stickiness (see Figure 6). 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.

V. Conclusions and Future Work

Our study aimed at assessing how the presence of small contribution elites (the top 1% contributors) in 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. In other words, too much or too little unevenness or elite stickiness may affect the global quality of the content. We found this to be particularly true for Stack Overflow. On Wikipedia, however, only entropy had a curvilinear behavior, while elite stickiness tended to be linearly associated with quality in the maturation period. 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). On the other hand, increase in stickiness is positively associated with content quality only for the period after entropy peaked.

For Stack Overflow, we have full blown effects both for entropy and elite stickiness, content being optimized for moderate period-to-period value changes in both variables.

Overall, the findings tell an interesting story about the way in which social dynamics influence content quality. Wikipedia seems to be a space on which elite stickiness can only have positive effects. As individuals become more vested and more involved with the site, content quality increases uniformly. As far as Wikipedia is concerned, there could not be “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 Stack Overflow, there are clear indications that quality follows our initial theoretical model, which predicted optimization of content quality within certain bounds of entropy and elite stickiness. This means that Stack Overflow is much 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 true collaborative space, where deep and direct collaboration is needed, Stack Overflow is a more competitive, individual-effort driven site. On Wikipedia, collaboration needs deep and continuous involvement of the elites, who keep the knowledge production going. As Figure 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 tire and 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. This shows that over time, Stack Overflow becomes 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 is quality and social structure.

At the same time, as shown in Figures 1 and 2, quality on Wikipedia follows an increasingly steep slope, while Stack Overflow, slowly drifts down. 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 (and cohesive site), such as Wikipedia, generates higher quality content.

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 an important light on how, considering macro processes, some significant trends of association between input (work concentration and elite stickiness) may impact quality.

Future work needs to be performed to validate curvilinear effects across a variety of other just-in-time and just-in-case sites. These should include a variety of Question-and-Answer sites, such as Quora, or the many other specialized sites federated by the Stack Exchange platform. Similarly, we need to look at Wikipedia versions in non-English languages and at the Wikia (commercial wiki) sites. Further work may also focus on improving the content quality measurement of Stack Overflow content, which could weigh the votes by the weight of the participants in terms of their own 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 U.S. Army Research Laboratory and the U.K. 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 U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

REFERENCES

[1] Claude E. Shannon. "W. Weaver The mathematical theory of communication." Urbana: University of Illinois Press 29 (1949).

[2] Kenneth D. Bailey. "Entropy measures of inequality." Sociological Inquiry 55, no. 2 (1985): 200-211.

[3] Besiki Stvilia, Michael B. Twidale, Linda C. Smith, and Les Gasser. "Assessing Information Quality of a Community-Based Encyclopedia." In Prooceedings of the10th International Conference on Information Quality (ICQ), pp. 442-454. MIT, 2005.

[4] Xiaoyong Liu, W. Bruce Croft, and Matthew Koll. "Finding experts in community-based question-answering services." In Proceedings of the 14th ACM international conference on Information and knowledge management, pp. 315-316. ACM, 2005.

[5] Honglei Zeng, Maher A. Alhossaini, Li Ding, Richard Fikes, and Deborah L. McGuinness. “Computing trust from revision history.” In Proceedings of the International Conference on Privacy, Security and Trust. ACM, 2006.

[6] George Kuk. "Strategic interaction and knowledge sharing in the KDE developer mailing list." Management Science 52, no. 7 (2006): 1031-1042.

[7] Tom Cross. "Puppy smoothies: Improving the reliability of open, collaborative wikis." First Monday 11, no. 9 (2006).

[8] Pawel Jurczyk, and Eugene Agichtein. "Discovering authorities in question answer communities by using link analysis." In Proceedings of the 16th ACM conference on Conference on information and knowledge management (CIKM), pp. 919-922. ACM, 2007.

[9] Pierpaolo Dondio and Stephen Barrett. " Computational Trust in Web Content Quality: A Comparative Evalutation on the Wikipedia Project." Informatica (Slovenia) 31(2): 151-160 (2007).

[10] Joshua E. Blumenstock. "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, 2008.

[11] B. Thomas Adler, Luca De Alfaro, Ian Pye, and Vishwanath Raman, “Measuring author contribution to the Wikipedia,” In Proceedings of the 4th International Symposium on Wikis (p. 15). ACM., 2008.

[12] Aniket Kittur, Bryant Lee, and Robert E. Kraut. "Coordination in collective intelligence: the role of team structure and task interdependence." In Proceedings of the SIGCHI conference on human factors in computing systems, pp. 1495-1504. ACM, 2009.

[13] Chirag Shah and Jefferey Pomerantz. "Evaluating and predicting answer quality in community QA." In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, pp. 411-418. ACM, 2010.

[14] Rena M. Palloff, and Keith Pratt. Collaborating online: Learning together in community. Vol. 32. John Wiley & Sons, 2010.

[15] Robert Bruno, “A democracy of unequals,” Ph. D. Dissertation, Purdue, West Lafayette, IN, 2010.

[16] Sorin A. Matei and Caius Dobrescu, “Wikipedia’s ‘Neutral Point of View:’Settling conflict through ambiguity,” The Information Society, vol. 27, pp. 1–12, 2011.

[17] Grégoire Burel, Yulan He, and Harith Alani. "Automatic identification of best answers in online enquiry communities." In Extended Semantic Web Conference, pp. 514-529. Springer Berlin Heidelberg, 2012.

[18] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. "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, pp. 850-858. ACM, 2012.

[19] Aditya Pal, Shuo Chang, and Joseph A. Konstan. "Evolution of Experts in Question Answering Communities." In Proceedings of the International AAAI Conference On Web and Social Media (ICWSM). 2012.

[20] Sujith Ravi, Bo Pang, Vibhor Rastogi, and Ravi Kumar. "Great Question! Question Quality in Community Q&A." In Proceedings of the International AAAI Conference On Web and Social Media (ICWSM), pp. 426-435. 2014.

[21] Piyush Arora, Debasis Ganguly, and Gareth JF Jones. "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, pp. 1232-1239. ACM, 2015.

[22] Sorin A. Matei, and Robert J. Bruno. "Pareto's 80/20 law and social differentiation: A social entropy perspective." Public Relations Review 41, no. 2 (2015): 178-186.

[23] Sorin A. Matei, Elisa Bertino, Michael Zhu, Chuanhai Liu, Luo Si, and Brian Britt. “A Research Agenda for the Study of Entropic Social Structural Evolution, Functional Roles, Adhocratic Leadership Styles, and Credibility in Online Organizations and Knowledge Markets.” In Roles, Trust, and Reputation in Social Media Knowledge Markets, pp. 3-33. Springer International Publishing, 2015.

[24] Warncke-Wang, Morten, Vladislav R. Ayukaev, Brent Hecht, and Loren G. Terveen. "The success and failure of quality improvement projects in peer production communities." In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 743-756. ACM, 2015.

[25] Aaron Halfaker and Amir Sarabadani.: Monthly Wikipedia article quality predictions. figshare. https://doi.org/10.6084/m9.figshare.3859800.v3. (2016).

[26] Wikipedia Featured Article, https://en.wikipedia.org/wiki/Wikipedia:Featured_articles. Accessed on March 2017.

[27] The Objective Revision Evaluation Service , https://ores.wikimedia.org/. Accessed on March 2017.

[28] S. Matei, W. Wei, B. Britt, C. Liu, Y. M. Zhu, E. Bertino, and A. Abu Jabal. “Social differentiation and online collaboration: The dynamics of elite emergence on Wikipedia,” submitted for publication.