On The Inequality of Contributions to Wikipedia

Felipe Ortega, Jesus M. Gonzalez-Barahona and Gregorio Robles
Libresoft Group, Universidad Rey Juan Carlos
Tulipan, s/n 28933. Mostoles.
Madrid. SPAIN.
Email: {jfelipe,jgb,grex}@gsyc.es

***Abstract*—Wikipedia is one of the most successful examples of massive collaborative content development. However, many of the mechanisms and procedures that it uses are still unknown in detail. For instance, how equal (or unequal) are the contributions to it has been discussed in the last years, with no conclusive results. In this paper, we study exactly that aspect by using Lorenz curves and Gini coefficients, very well known instruments to economists. We analyze the trends in the inequality of distributions for the ten biggest language editions of Wikipedia, and their evolution over time. As a result, we have found large differences in the number of contributions by different authors (something also observed in free, open source software development), and a trend to stable patterns of inequality in the long run.**

I. Introduction

Nowadays, Wikipedia is one of the most successful examples of large-scale collaborative content development. In just some few years, it has been able to attract an impressive number of contributors who create content, update it, modify it, and try to adapt it to the emerging policies that the project is establishing. Probably one of the key factors in its success is the low entry barrier for new authors. Thanks to the easy-to-use and friendly user interface that Wikipedia provides to authors, they can start to contribute even upon arrival to a Wikipedia page for the first time.

However, the fact that anyone can contribute to the Wikipedia does not imply that every Wikipedia visitor becomes an author (actually, only a very small fraction of visitors contribute). And the fact that it is easy to contribute does not mean that all authors contribute the same way, neither with the same intensity. On the contrary, the number of contributions is known to vary much from author to author [1]. But the pattern of these differences, or, in other words, how equal or unequal the contributions from different authors are, has not yet been characterized in detail.

Indeed, in the last years there has been a great controversy about this point. For instance, on September 4th, 2006, Aaron Swartz included a quote from Wikipedia's founder, Jimmy Wales, in his blog [2] where Wales argued that the majority of the total number of contributions to the Wikipedia came from a small group of authors. Swartz used a different metric, counting the number of characters in each contribution (rather than the number of contributions), and searched for the text blocks that remained in the final version of the article. He then applied this metric to several articles picked up at random, and showed that less frequent contributors were actually providing much of the articles contents. Despite that, as articles continue to evolve and change over time, we will inevitably miss some contribution effort if we focus only in the final revision of an article. Perhaps some contributions are removed later in the article's life not remaining in the final version, but we should take it into account if we want to have a complete picture of the mechanisms found in the Wikipedia.

To analyze what is actually happening in this respect, in this paper we show the results of calculating the level of inequality in the contributions to several language editions of the Wikipedia, using the Gini coefficient (a well-known metric to measure inequality distributions introduced by Conrado Gini). This coefficient was originally thought to analyze the distribution of wealth among a population, but is used in other domains as well to provide a quantitative measure of the inequality of a distribution under study. By comparing the Gini coefficients of the various language editions of the Wikipedia, and its evolution over time, we can learn not only about how unequal the contributions by different authors are, but also if it is converging (or not) to some value.

More in detail, we have considered the top-ten Wikipedia language editions, according to their total number of articles (which will allow comparison of results among different language editions, as well as the search for peculiarities and similarities). We calculated the sum of the number of contributions made by authors for each language edition, and then, obtained the Gini coefficient of this overall effort. This coefficient will give information about the inequality in contributions from the whole group of authors, for every language edition considered in our study. But, taking into account that the activity of authors may change over time, we have also calculated the Gini coefficient for the contributions to each edition on a monthly basis as well.

As a result, we have obtained a good picture of the evolution of inequality of contributions over time for all the studied editions. This picture shows interesting patterns, such as a tendency to converge to a stable Gini coefficient after some oscillations during the early life of the editions, despite the tremendous growth in number of authors and articles. In addition to this, we also found relevant differences among the editions under study.

In the following sections, we first revisit some previous research about the contribution of users to Wikipedia. Then, we present our methodology for measuring the inequality of contributions to the Wikipedia, for the top-ten language editions, followed by results and discussion on them. Finally,

1530-1605/08 $25.00 © 2008 IEEE

we present the most important conclusions of our study, and future work we should undertake to address additional challenges.

II. Previous Research

The Gini coefficient is a normalized measure of inequality [3]. To calculate it, first the Lorenz curve for contributions has to be created. The Lorenz curve is a graphical representation of the cumulative sum of contributions where we sort contributors on the horizontal axis by their amount of contribution. Then, we plot in the vertical axis the accumulated contribution. An equal distribution would result in a straight diagonal line that divides the first quadrant in two equal parts. The Gini coefficient is the area between the diagonal (corresponding to a Lorenz curve for a community with an equal distribution) and the actual Lorenz curve, normalized to 1. Consequently, values of the Gini coefficient close to 0 correspond to equal or almost equal distributions, while values close to 1 are good indicators of high inequalities.

Besides the initial study presented by Conrado Gini in [3], the Gini coefficient has been employed frequently in research works measuring inequalities. Some examples are [4] and [5] regarding Economics, and [6], and [7] in the Health Sciences field. The Gini coefficient has also been used before for research of inequalities in fields that may be related to the Wikipedia, in the sense that participation is open (at least ideally) to anyone. Such is the case of participation in libre (free, open source) software projects [8], where inequality has been the focus of many studies [9], [10].

Some previous research on the contribution of users to the Wikipedia has also been published in the last years. For example, Jakob Voss analyzes the German Wikipedia in [11]. He found evidence about the number of distinct authors per article following a power law, while the number of distinct articles per author follows Lotka's Law. Viegas et al. propose in [12] an alternative to visually study the evolution over time of the contributions to Wikipedia articles, developing their own tool, which they have called *History Flow*. In a more recent work, Viegas et al. [13] apply their software tool to describe the collaborative processes that drive the creation of articles, employing the *talk pages* associated to every Wikipedia page. They also demonstrate how the Wikipedia is growing with an exponential pattern by studying the history graphs of some popular articles. We also find previous research works, for example [14], [15], [16] and [17], analysing the quality of contents in Wikipedia's articles. Wilkinson et al. further characterize in [18] the collaborative creation of contents undertaken by the Wikipedia community of users. In addition to that, [19], [20], [21] and [22] present methodologies and tools to analyse Wikipedia's content semantics.

We can find in [23] a first approximation to the question of who is really authoring the majority of contributions to Wikipedia. In this paper, the authors analyze the quality of contributions using the percentage of aggregated contents that remain in later revisions of a certain article, much in the same way as the Aaron Swartz study cited above. They infer that two main groups of contributors are responsible for most of the high quality contributions. On one side, what they identify as high quality contents come from *zealots*, registered users with a high interest in obtaining reputation, and with a high level of participation. On the other side, we have the *good Samaritans*, anonymous users sporadically contributing to the Wikipedia. The authors show that there is a strong correlation between the quality of contributed contents and the level of contributions made by individual authors. Regarding *zealots*, the greater the number of contributions per user is, the better the quality of the content is. If we turn to *good Samaritans*, high quality contributions come from users with fewer contributions, and that quality decreases as the number of contributions per user raises.

Finally, a recent paper by Kittur et al. [1] presents a new approach to analyze the Wikipedia community of authors. Their hypothesis is that after an initial period in which contributions usually came from registered and very active users, nowadays most of the contributions to the Wikipedia come from users with a smaller participation level. Hence, these authors argue that the main force behind the current growth of Wikipedia is what they call *the wisdom of the crowd*, the larger group integrated by infrequent contributors, versus *the power of the few*, the smaller group of very active users [1].

III. Measuring the Inequality of Contributions to Wikipedia

In this section, we present our methodology for measuring the inequality of contributions made by the community of users of Wikipedia for each language edition. We then present our results, and finally present a discussion about these results.

A. Methodology

In a first step, we retrieved the database dumps of the top-ten Wikipedia language editions, taking the total number of articles in them as the selecting parameter. These dumps were the most up-to-date versions available on 30th November, 2006. At that time, the top-ten language editions were English, German, French, Polish, Japanese, Dutch, Italian, Portuguese, Swedish and Spanish. Various versions of these dump files are publicly available, and can be downloaded from the Wikimedia Foundation Downloads web page[1]. We then obtained the complete dump for each language edition, with the whole revision history for every article. We employed WikiXRay, a Python tool implemented by the authors of this paper for automating the process, to parse these dumps and generate graphics and quantitative results with the measures of the inequality of contributions. This software tool is publicly available at BerliOS[2], under the GNU GPL license.

Figure 1 shows, for each language edition in this study, the total number of articles and total number of registered authors found in the dumps. Not surprisingly, the English language edition leads the top-ten list in both aspects, followed by all other languages at a great distance. It is noteworthy to find

[1] http://download.wikimedia.org

[2] http://developer.berlios.de/projects/wikixray/

the Spanish edition in the tenth position of the top-ten list according to its number of articles, despite being the fourth in number of authors. The fraction of articles per author could be an interesting measure, especially if we compare it with the number of native speakers that each language has in the world. In this sense, it seems that those languages with a high number of speakers (like English, Spanish, German or French) have an average of articles per author that ranges from almost 2 (English) to 8 (French). All other languages have ratios close to or over values of 10, being Sweden the most significant case where almost 20 articles per author exist.

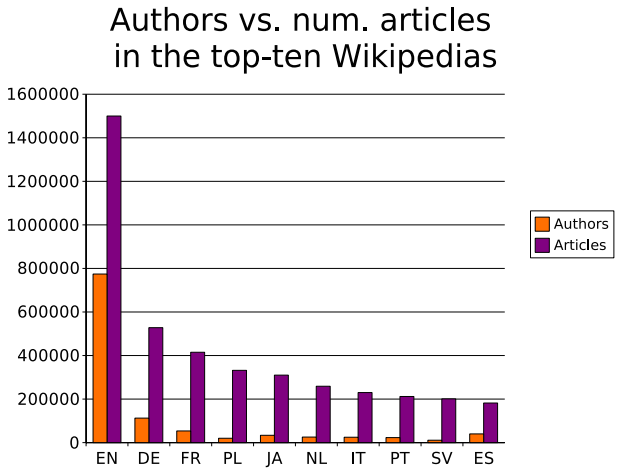

Fig. 1. Number of articles and registered authors in the top-ten Wikipedias.

Following this, we have computed the Gini coefficient first for the aggregate number of contributions by registered authors to each language edition. Non-registered author were filtered out, due to the impossibility of discriminating properly among them, as the only identifier stored for them is their IP address (and there exists the possibilities of NAT and proxies masquerading different authors under the same IP). Then, we computed the Gini coefficient for each language edition on a monthly basis to reflect the evolution over time of the inequality of user contributions.

B. Results

In this subsection, we present the most relevant results obtained from our analysis of the inequality of contributions in the Wikipedia.

1) Gini Coefficient for the Aggregate Number of Contributions: Table I contains the Gini coefficients for the total number of contributions performed by authors for each language edition.

As we can see, we found very similar Gini coefficients for all the languages, ranging from a minimum of 0.9246 for the Japanese edition to a maximum of 0.9665 in the Swedish edition. We can derive from these numbers that there is a high level of inequality in the total number of contributions to each Wikipedia language edition.

TABLE I
NUMBER OF ARTICLES, REGISTERED AUTHORS AND GINI COEFFICIENTS FOR THE TOP-TEN WIKIPEDIAS

Lang. Edition	Numb. of Articles	Numb. of authors	Gini Coeff.
English	1.5M	774,301	0.93076
German	528K	112,624	0.93785
French	415K	53,678	0.950071
Polish	332K	20,153	0.957015
Japanese	310K	33,756	0.9246
Dutch	259K	25,545	0.963844
Italian	230K	24,851	0.957104
Portuguese	212K	23,109	0.962223
Swedish	201K	11,135	0.959677
Spanish	182K	40,042	0.950077

Figure 2 represents how we calculate the Gini coefficient, taking the German language edition data as an example. As explained before, first we have to depict the Lorenz curve, which gives a graphical representation of the cumulative distribution function of the contributors. Perfect distribution among authors is hence given by a 45 degrees line. The Gini coefficient is calculated as the area between the two curves, giving a numerical value that is relative to the distance between the actual distribution and perfect equality.

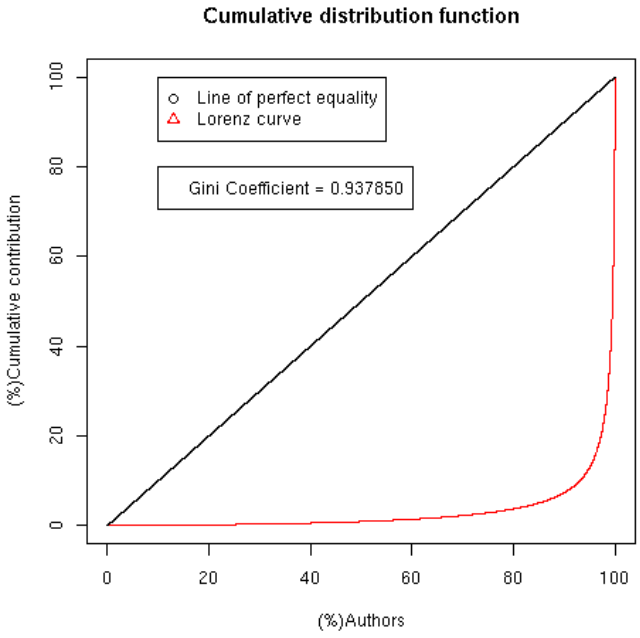

Fig. 2. Gini coefficient and Lorenz curve for the total number of contributions made by registered authors (German Wikipedia).

The next figure (Figure 3) presents the Lorenz curve for all the languages under study. As we summarized in Table I, all of them present similar behaviors, with approximately 90% of the users authoring less than 10% of the contributions. We can see that we have got a similar pattern as in the

case of libre software projects, where a small amount of very active contributors is responsible for the vast majority of the work [24]. This behavior in libre software projects is not only specific of source code production, but can generally be found on other elements found in software, such as documentation and translation tasks [25].

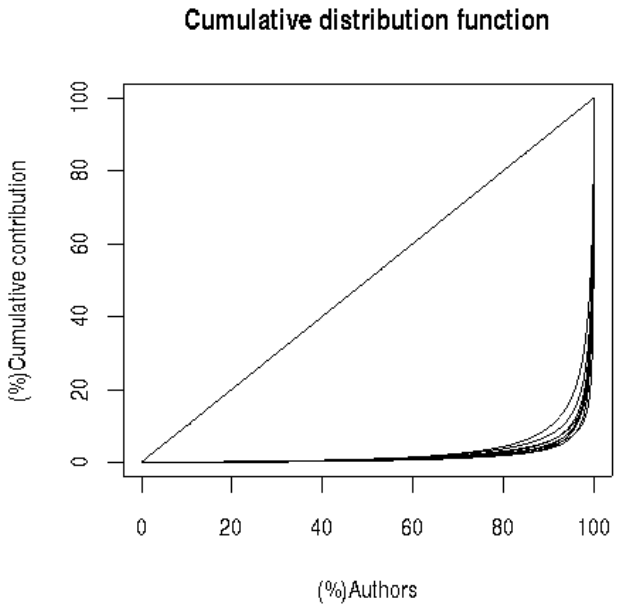

Fig. 3. Lorenz curves for the total number of contributions made by registered authors (Top-ten language editions).

Table II shows the Gini coefficients and the age of the top-ten Wikipedias. We can see that the English and German language editions have similar Gini coefficients, although the English edition is almost two years older than the German one. Language editions with an age of 49 months present similar levels of inequality, with the exception of the German edition, possibly because it has the higher number of authors of this group.

TABLE II
GINI COEFFICIENTS AND AGE OF THE TOP-TEN WIKIPEDIAS

Lang. Edition	Gini Coeff.	Age (months)
English	0.93076	71
German	0.93785	49
French	0.950071	49
Polish	0.957015	49
Japanese	0.9246	50
Dutch	0.963844	49
Italian	0.957104	47
Portuguese	0.962223	47
Swedish	0.959677	49
Spanish	0.950077	49

Figure 4 presents Gini coefficients for the aggregate number of contributions against the logarithm of the total number of articles. Those languages with fewer number of articles tend to have higher values of the Gini coefficient for the total number of contributions, and languages with a higher number of articles tend to present lower values of the Gini coefficient. This probably indicates that, as the number of articles rises, a higher number of potential authors find topics to which they feel like contributing. This would distribute the total number of contributions among a larger group of authors, thus decreasing the Gini coefficient. The Japanese language edition (with a Gini coefficient of 0.9246) is the clear exception to this general tendency, and further analysis should be conducted to explain this fact. On the other hand, Figure 5 shows the logarithm of the total number of authors against the Gini coefficient for the total number of contributions in each edition. Besides the Japanese edition (with a low Gini coefficient), we can observe that the Gini coefficient tends to increase as the number of authors decrease. These results suggest that in communities with a small number of authors, contributions tend to be concentrated by the group of most active ones (therefore rising the Gini coefficient). This result is in concordance with the evidence of the *the wisdom of the crowd* trend gaining in importance over the *the power of the few* as discussed in [1].

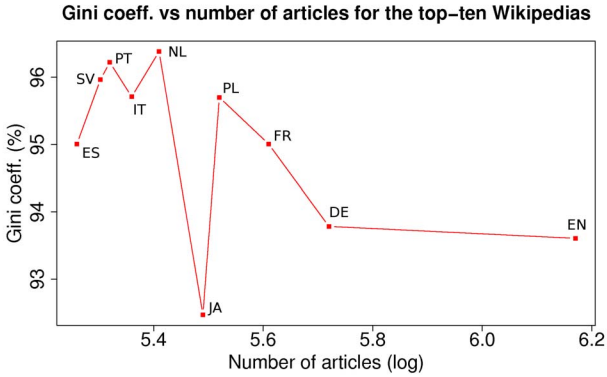

Fig. 4. Gini coefficients against total number of articles (Top-ten Wikipedias).

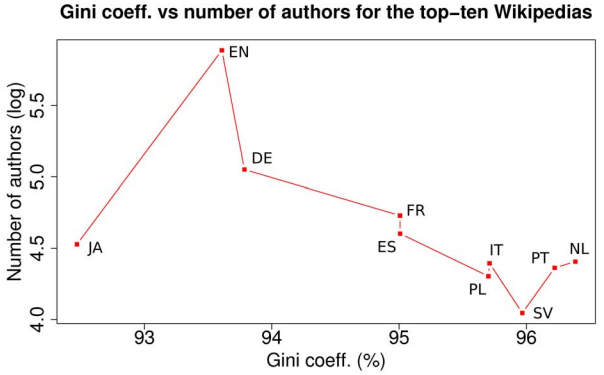

Fig. 5. Total number of registered authors against Gini coefficients (Top-ten Wikipedias).

Finally, Figure 6 presents the histogram and the probability density function for the distribution of the number of contributions made by authors to the English Wikipedia. The resulting graph is strongly skewed to the left, indicating a great number of authors with few contributions. However, though there are many more authors with a lower activity level, the Gini coefficient for the total number of contributions tells us that contributions coming form the most active users overwhelm contributions coming from the crowd of sporadic authors.

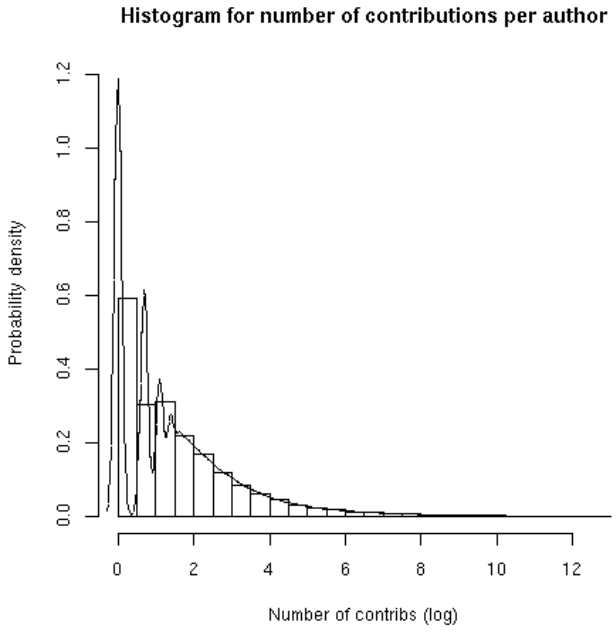

Fig. 6. Histogram and probability density function for contribution of authors (English Wikipedia).

2) Gini Coefficients per month: Although we have computed the Gini coefficient for the aggregate number of contributions for each language edition, one question still remains unanswered: is this coefficient representative enough to get an accurate picture of the inequality of contributions? This question rises because the aggregate effect of the total number of contributions could hide details about how the level of inequality has evolved over time. If we consider a hypothetical language edition with a very low level of inequality in contributions in its early history, and a very strong inequality in contributions in the last year, the aggregated behavior could result to be very similar to another hypothetical language edition with monthly Gini coefficients that have remained more stable.

So, to complete this quantitative analysis of inequality in the Wikipedia, we will present some results of the evolution of the Gini coefficients over time, obtained on a monthly basis, for each language edition. Figures 7, 8, 9, 10, 11 and 12 show the evolution over time of the monthly Gini coefficient for the English, German, Dutch, Japanese, Swedish and Spanish language editions respectively. The first period corresponds to the first month of existence of each language edition. Hence, if the English Wikipedia appeared January 2001, that month will correspond to period 0 in its graph, February 2001 will be period 1, and so on.

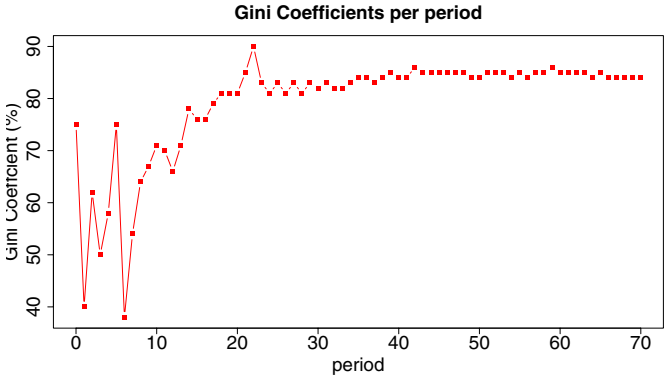

Fig. 7. Evolution in time of the Gini coefficients for the contributions of registered authors in the English Wikipedia.

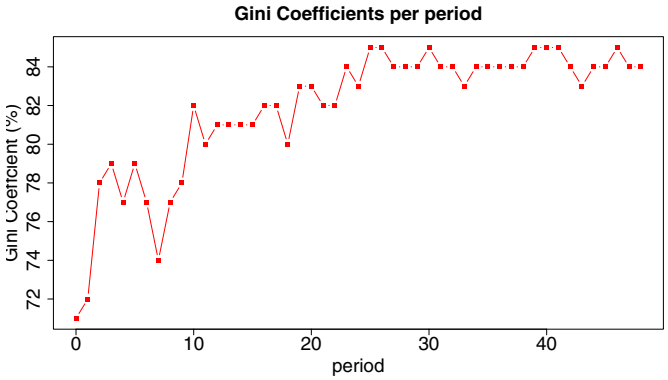

Fig. 8. Evolution in time of the Gini coefficients for the contributions of registered authors in the German Wikipedia.

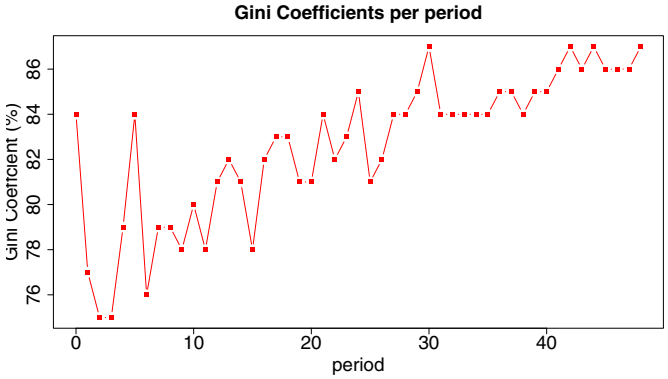

Fig. 9. Evolution in time of the Gini coefficients for the contributions of registered authors in the Dutch Wikipedia.

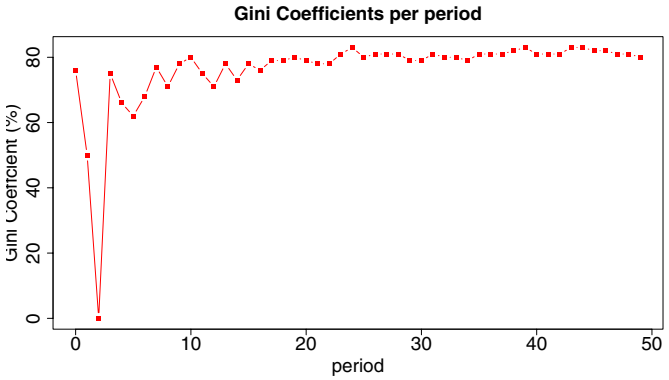

Fig. 10. Evolution in time of the Gini coefficients for the contributions of registered authors in the Japanese Wikipedia.

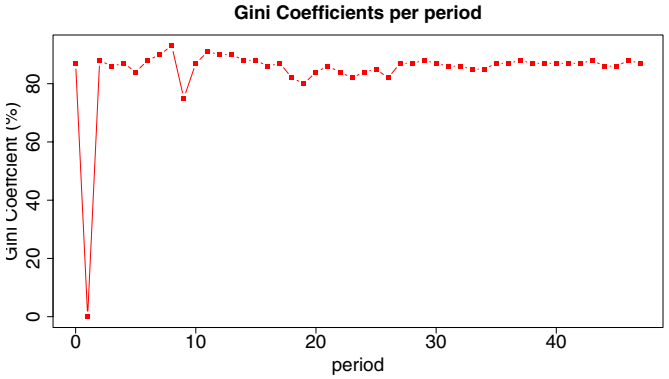

Fig. 11. Evolution in time of the Gini coefficients for the contributions of registered authors in the Swedish Wikipedia.

The first remark on these results is that in all graphs the values of the monthly Gini coefficient are below the ones in Table I with the total number of contributions. Values in the evolutionary plots tend to stabilize in the 80s far below the mid-90 values for the global measurement. This means that the inequality on a monthly basis is smaller than considering all the life of the language edition, a consequence mainly of avoiding the long tail of small contributors that we can find in the global one. The reason for this is that authors with no contributions are not considered for the calculation of the Gini coefficient and thus many authors considered for the accumulated one are filtered out in the ones calculated on a monthly basis.

Beyond this, it is significant that all languages present a similar pattern with an unstable behavior at first that may take up to 20 months and a relative stable phase from then on. So, as the popularity of Wikipedia grew, we can see how the monthly Gini values began to increase towards a value around 85%. This refutes the idea of the *the wisdom of the crowd* proposed in [1] as in the stable phase contributions remain highly skewed and a few authors (not having to be the same) participate in a large amount of contributions. If we assume that the number of contributions and contributors

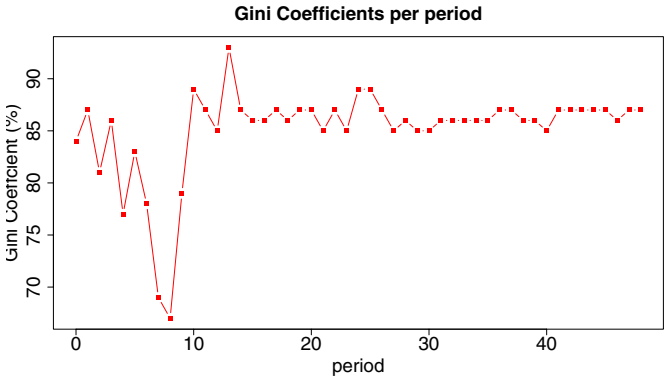

Fig. 12. Evolution in time of the Gini coefficients for the contributions of registered authors in the Spanish Wikipedia.

to all Wikipedia languages has increased over time [13], we find that these coefficients may characterize mature Wikipedia environments.

IV. Conclusions, Discussion and Future Work

We have presented a quantitative analysis of the level of inequality found in the contributions received by the top-ten Wikipedia language editions, according to their total number of articles. We have found a great level of inequality for every language edition processed, with less than 10% of the total number of authors being responsible for more than the 90% of the total number of contributions. Furthermore, we found that the evolution in time of the Gini coefficient has remained very stable over the recent history of every language edition, with a typical value between the 80% and 85% limits. This could give us a hint about the inequality value that should be expected for a well-established Wikipedia language edition.

From our analysis, we can infer several interesting conclusions from the analysis of the inequalities in contribution found in the Wikipedia. The first one is that contributions to all language editions of Wikipedia are strongly unequal. A small percentage of the total number of authors (less than 10% of the total number of authors in all the editions we considered) are responsible for the majority of the total number of contributions to the encyclopedia. These results sustain the hypothesis that, if we focus on the number of contributions, we should be able to identify a "core" group of authors that make most of the contributions to the project, as it can be done for large libre software projects [9].

This is an important conclusion, because as mentioned in the *Previous Research* section, we can find some previous results in [1] presenting the hypothesis that there is an increasing number of authors with a little number of contributions that is supposedly taking over the main role of contributing to Wikipedia. The Gini coefficients obtained in our analysis seem to help us sustain the hypothesis of a large number of contributions coming from a small group of very active users. We observed this behaviour not only for the accumulated number of contributions to Wikipedia, but also on a monthly

basis.

Moreover, the level of inequality was not correlated, in many cases, with other parameters like the number of articles or the number of authors. For example, the Japanese Wikipedia, which is in the 4th place by its number of articles, and in the 5th place attending its total number of authors, presented the most equal pattern regarding the aggregate number of contributions. Other language editions, like German and especially English, with a higher number of authors, showed a higher level of inequality. This reflects that it is not the number of authors producing contents or the total number of articles being modified what determines the level of inequality in a certain community of users. Other behavioral and social issues could be affecting how equal or unequal those contributions are.

The second remarkable conclusion is that the evolution over time of this inequality level has remained almost stable over the months in every language edition. We found that the typical value of the Gini coefficient per month was situated between the 80% and 85% limits in every case, at least for the last two years period. In the ancient history of these language editions, we found a very variable behavioral pattern at the very beginning of the graphs. But later, we found a common growing trend of the inequality level in every language edition, just until it reaches the *standard* margin mentioned above. This is also very interesting, because it probably points out a stability value towards which every edition of Wikipedia tends to evolve, thus possibly giving us a hint about the level of maturity achieved by a certain language edition.

Thus, we have found not only that the inequality of the aggregated number of contributions is very high, but also that this level of inequality has remained somewhat constant in the recent history of every language edition. We should remark that we need both data to sustain these hypotheses, because as we have shown previously, the Gini coefficient of the aggregate number of contributions is merely a useful, but not definitive, indicator of the real behavior of the communities of users we analyzed.

Our future lines of research include considering the size of the contributions made by each author in this analysis, (possibly in number of characters and in number of words), thus reflecting the real amount of data added or deleted from the Wikipedia by an author. Hence, we would be able to offer another interesting point of view to get insight into the inner behavioral patterns of the Wikipedia.

REFERENCES

[1] A. Kittur, E. Chi, B. A. Pendleton, B. Suh, and T. Mytkowicz, "Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie," in *Proceedings of the 25th Annual ACM Conference on Human Factors in Computing Systems (CHI 2007)*, Apr./May 2007.

[2] A. Swartz. (2006, Sep.) Who writes wikipedia. [Online]. Available: http://www.aaronsw.com/weblog/whowriteswikipedia

[3] C. Gini, "On the measure of concentration with especial reference to income and wealth," in *Cowless Comission*, 1936.

[4] R. Dorfman, "A formula for the gini coefficient," *The Review of Economics and Statistics*, no. 61, pp. 146–149, Mar. 1979.

[5] J. A. Mills *et al.*, "Statistical inference via bootstrapping for measures of inequality," *Epidemiological Bulletin*, vol. 22, no. 12, Mar./Apr. 1997.

[6] C. Castillo-Salgado *et al.*, "Measuring health inequalities: Gini coefficient and concentration index," *Epidemiological Bulletin*, vol. 22, no. 1, Mar. 2001.

[7] A. Wagstaff *et al.*, "On the measurements of inequalities in health," *Soc. Sci. Med.*, vol. 33, no. 5, pp. 545–577, 1991.

[8] G. von Krogh, S. Spaeth, and K. R. Lakhani, "Community, joining, and specialization in Open Source Software innovation: A case study," *MIT Sloan Working Paper No. 4413-03*, 2003.

[9] A. Mockus, R. T. Fielding, and J. D. Herbsleb, "Two case studies of Open Source software development: Apache and Mozilla," *ACM Transactions on Software Engineering and Methodology*, vol. 11, no. 3, pp. 309–346, 2002.

[10] S. Koch and G. Schneider, "Effort, cooperation and coordination in an open source software project: GNOME," *Information Systems Journal*, vol. 12, no. 1, pp. 27–42, 2002.

[11] J. Voss, "Measuring wikipedia," in *Proceedings of the ISSI 2005*, Jul. 2005.

[12] F. B. Viegas, M. Wattenberg, and K. Dave, "Studying cooperation and conflict between authors with history flow visualizations," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, Apr. 2004, pp. 575–582.

[13] F. B. Viegas, M. Wattenberg, J. Kriss, and F. van Ham, "Talk before you type: Coordination in wikipedia," in *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. Computer Society Press, Jan. 2007, p. p. 78a.

[14] B. Stvilia, M. Twidale, L. Gasser, and L. Smith, "Information quality discussions in wikipedia," UIUCLIS, Technical Report ISRN 2005/2+CSCW, 2005. [Online]. Available: http://mailer.fsu.edu/bstvilia/papers/qualWiki.pdf

[15] ——, "Information quality in a community-based encyclopedia," in *Knowledge Management: Nurturing Culture, Innovation, and Technology - Proceedings of the 2005 International Conference on Knowledge Management*, S. Hawamdeh, Ed., Charlotte, NC: World Scientific Publishing Company, 2005, pp. 101–113.

[16] B. Stvilia, M. Twidale, L. Smith, and L. Gasser, "Assessing information quality of a community-based encyclopedia," in *Proceedings of the International Conference on Information Quality - ICIQ 2005*, Cambridge, MA, USA, 2005, pp. 442–454.

[17] ——. (2006) Information quality work organization in wikipedia. (Under review). [Online]. Available: http://mailer.fsu.edu/bstvilia/papers/stvilia_wikipedia_infoWork_p.pdf

[18] D. M. Wilkinson and B. A. Huberman, "Assessing the value of coooperation in wikipedia," 2007. [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0702140

[19] M. Völkel, M. Krötzsch, D. Vrandecic, H. Haller, and R. Studer, "Semantic wikipedia," in *WWW '06: Proceedings of the 15th international conference on World Wide Web*. New York, NY, USA: ACM Press, 2006, pp. 585–594.

[20] M. Krotzsch, D. Vrandecic, and M. Volkel, "Wikipedia and the semantic web - the missing links," 2005. [Online]. Available: citeseer.ist.psu.edu/krotzsch05wikipedia.html

[21] M. Strube and S. P. Ponzetto, "Wikirelate! computing semantic relatedness using wikipedia," in *Proceedings of the 21th National Conference on Artificial Intelligence (AAAI06)*, Boston, Massachusetts, USA, Jul. 16–20, 2006.

[22] T. Holloway, M. Bozicevic, and K. Borner, "Analyzing and visualizing the semantic coverage of wikipedia and its authors," 2005. [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0512085

[23] D. Anthony, S. W. Smith, and T. Williamson, "Explaining quality in internet collective goods: Zealots and good samaritans. the case of wikipedia," Nov. 2005.

[24] R. A. Ghosh and V. V. Prakash, "The orbiten free software survey," *First Monday*, vol. 5, no. 7, May 2000, http://www.firstmonday.dk/issues/issue5_7/ghosh/.

[25] G. Robles, J. M. Gonzalez-Barahona, and J.-J. Merelo, "Beyond executable source code: The importance of other source artifacts in software development (a case study)," *Journal of Systems and Software*, vol. 79, no. 9, pp. 1233–1248, September 2006.