Participation Inequality in Wikis: A Temporal Analysis Using WikiChron

Abel Serrano
Facultad de Informática. Universidad Complutense de Madrid
Madrid, Spain
abeserra@ucm.es

Javier Arroyo
Facultad de Informática. Universidad Complutense de Madrid
Madrid, Spain
javier.arroyo@fdi.ucm.es

Samer Hassan
Berkman Klein Center at Harvard University &
Facultad de Informática at Universidad Complutense de Madrid
Madrid, Spain
samer@fdi.ucm.es

ABSTRACT

It is widely accepted that peer production communities show a high level of inequality in the level of participation. Typically, we can observe a power law in the distribution of contributions. However, we argue that participation inequality, and specifically its evolution over time, has been understudied. Previous research has concentrated on large successful projects (e.g. Wikipedia), leaving a gap regarding small or mid-size communities. The wiki ecosystem is highly diverse, and so it may be the participation distribution in the communities and its evolution. The aim of this work is twofold: (1) To show a novel webtool, WikiChron, for the analysis of wiki evolution, with a focus on inequality metrics; (2) To provide relevant cases of comparative analysis that show the diversity of the wiki ecosystem, with examples that are counter-intuitive or contradict existing literature, while reflecting the limits of single-metric studies. We aim to open lines of wiki research which require the use of multiple metrics to study participation inequality.

CCS CONCEPTS

• **Information systems** → **Wikis**; • **Human-centered computing** → **Wikis**; *Collaborative and social computing design and evaluation methods*;

KEYWORDS

online communities, contribution, evolution, inequality, participation, peer production, temporal analysis, wikis, Wikia

1 INTRODUCTION

There is a general acceptance in the literature that participation in peer production communities shows strong levels of inequality. This is shown in general in online communities with user-generated contents [7, 10], where we can observe a power law in the distribution of contributions, typically expressed in the mnemonic 1-9-90 rule. That is, the idea that a 1% of the community population are core contributors concentrating the majority of workload; a 9% are occasional contributors, with irregular and sporadic contributions; and a 90% are "readers", sometimes framed as "active audience", sometimes as the pejorative "lurkers", which do not directly contribute, but consume the created content and may contribute indirectly, e.g. to the visibility of the project [16].

We observe similar levels of participation inequality in a wide range of peer production communities [9]. It is clear in large projects like Wikipedia [1, 12], OpenStreetMap [17], or Wikia's largest wikis [14].

However, there are still multiple aspects of participation inequality that deserve further exploration. For instance, since participation inequality studies have concentrated on large successful projects such as those mentioned above, there is a gap regarding small or mid-size communities. Moreover, the wiki ecosystem is highly diverse, and there may be interesting clusters or trends, which correspond to certain types of wikis, which remain hidden when considering only the largest and most successful ones.

Particularly interesting research questions may be related to how participation inequality evolves over time, taking into account different types of wikis, or different stages of growth, such as those defined in [2]. For instance, would the topic influence their evolution? E.g. do fandom wiki communities (the most popular ones in Wikia) have similar inequality rates to hobbyist wiki communities? Ceteris paribus, would the cultural differences impact participation inequality? In fact, some studies already show differences by culture in peer production [4, 11]. Would the wiki features/aims affect its participation distribution? For instance, there are wikis strongly event-dependant (e.g. about tv-show), or wikis mostly used for Q&A, would those factors affect their inequality levels?

There are several temporal analyses of participation inequality, but again focused on large communities. In addition, they systematically use the Gini coefficient, which is arguably the most popular statistical measure of inequality in distributions. The Gini coefficient was originally proposed (and is typically used) as wealth or income inequality measure, but it generally measures statistical dispersion or inequality. Regarding participation inequality in

wikis, Ortega et al. [12] showed that the the top-ten Wikipedias ere highly unequal with an accumulated Gini coefficient above 90% in 2006. While the evolution of the monthly Gini coefficient reached a stable phase around 85% after 20 months of the creation of each Wikipedia. Matei and Bruno [8] analyzed the weekly evolution of the cumulative entropy, another inequality measure, of English Wikipedia until 2010. They also appreciated a first unsteady period, followed by a stable phase. Interestingly, they appreciate a slight decrease of the unequality in the last 4 years, suggesting that it may be caused by resilient growth. Recently, Yang et al. [17] studied the yearly evolution of 4 countries in OpenStreetMap from 2007 to 2014. They showed that the Gini coefficients for the 4 countries eventually reach a high level (95%) even if they follow different trajectories.

The case of Kittur et al. [6] is slightly different as it focused on more diverse communities. Their study showed the evolution during the first three years of life of the average Gini coefficient of two sets of Wikia wikis: highly active wikis (those which eventually accumulate at least 5000 edits) and those that did not reach that level of activity. The inequality reported is much higher over time in the group of highly active wikis. While the different behaviors are sound since they suggest that activity and inequality are related, the aggregation of diverse wikis in two broad types prevents fine grained conclusions.

Following this tradition of participation inequality research in wikis, we will perform an exploration of different wikis from Wikia, highlighting cases in which the participation inequality may contradict existing literature or open questions that require further research. In order to perform this study, we have developed a new webtool for the analysis and visualization of wiki evolution, WikiChron [13], which is explained in section 2. Afterwards, in section 4 we will provide a succinct view of three interesting cases of wiki inequality explored with our tool, ending with some concluding remarks in section 5.

2 WIKICHRON: PLOTTING WIKI EVOLUTION

We have conducted a thorough study of the available tools that aim to visualize wiki analytics. However, most of them are discontinued or outdated projects. Others are, either limited to a certain small set of wikis only, such as the WikiStats project[1]; or they have a poor support of metrics and do not allow visualization of plots of several wikis at the same time, like WikiApiary[2] or StatMediawiki[3].

For the sake of the attainment of this and further research, we have developed a free/open source web application called WikiChron[4]. Figure 1 shows a caption of it. This tool is focused on facilitating the exploration and comparison of different metrics on a selection of wikis. In particular, WikiChron is specialized in plotting time series graphs and combine the plots in the same graph. Then, the evolution of the metrics is intuitively presented and wikis can be easily compared.

---

[1]The Wikimedia Statistics project: **https://stats.wikimedia.org/**
[2]Wikiapiary wiki: **https://wikiapiary.com**
[3]StatMediawiki: **http://osl2.uca.es/statmediawiki/**
[4]WikiChron is publicly available at **http://wikichron.science/**

We can classify the currently available metrics that WikiChron allows to visualize in three main groups, all displayed as monthly time series:

- A rich set of basic metrics, e.g. number of edits, new users, new pages.
- Higher-level ratios, e.g. edits per user, edits per page, percentage of anonymous edits.
- Metrics regarding the distribution of participation, e.g. Gini coefficient, Ratio 10:90, covered in the next section.

We plan to keep adding more metrics that can provide insights on other aspects of wikis and the communities around them.

Most of these metrics can be calculated both in a monthly basis, and using cumulative values up to each month. Additionally, the tool has an option to select the graphs time axis, in natural calendar dates or in relative dates from the birth of the wiki. A slider for slicing a date range and a couple of wikis and metrics selectors allow quick filtering of the data being shown. Figure 1 shows a caption of WikiChron. The tool also provides handy features such as high interactivity in the graphs, and export capabilities in PNG, thanks to its plotting library.

3 METRICS OF DISTRIBUTION OF PARTICIPATION

In order to evaluate the distribution of participation in a community, we have implemented a diverse group of metrics.

The metrics presented in this section are coefficients and ratios. They are represented as a monthly time series using the cumulative contributions until every given month from the birth of the wiki. In other words, for every month of the wiki, we take into account all the contributions made until that month to perform the computation of the metric. Besides, we have established a minimum number of contributors required for these metrics to be estimated, which is different from each metric depending on the 'subtlety' of the metric.

It is important to clarify that by contributions we refer to edits in articles and not in other pages. By users we refer to both anonymous and registered users. Furthermore, we assume that two contributions with the same IP address come from the same person; in this way, compensating that some persons can anonymously contribute from different IPs.

3.1 Gini coefficient

First, we resort to a general inequality metric. The Gini coefficient is a very common measure used to measure the inequality in a population, and in our case it is used to measure the participation inequality. It takes a real value from 0 to 1, being 0 a perfectly equal community (e.g. all community members perform the exact amount of contributions) and 1 an absolutely unequal community (e.g. one person assumes all the workload while the others do nothing). We have implemented the Gini coefficient with a correction factor for small data sets, following [3].

Let n be the number of contributors and y_i the contribution of individual i, with y_i, i = 1 to n, indexed in non-decreasing order
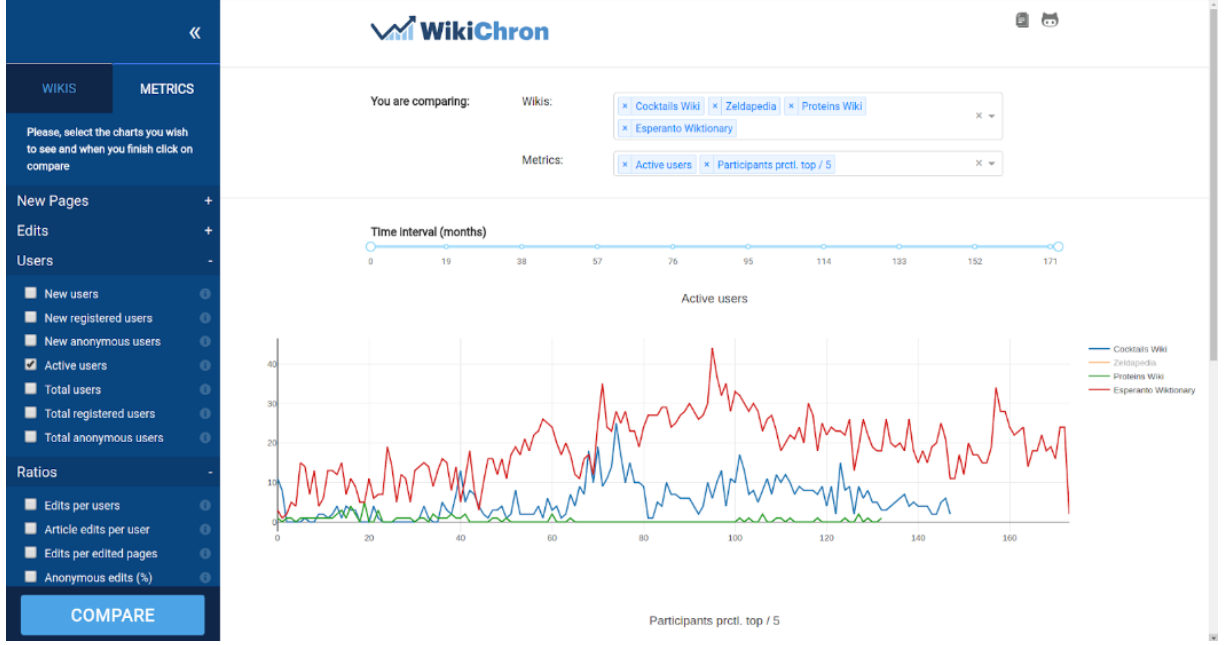
Figure 1: A caption of WikiChron.

$(y_i \leq y_{i+1})$ then

$$G = \frac{1}{n-1}\left(n+1-2*\left(\frac{\sum_{i=1}^{n}(n+1-i)*y_i}{\sum_{i=1}^{n}y_i}\right)\right). \qquad (1)$$

3.2 Ratio 10:90

Second, we aim to compare the participation from top contributors and the rest of the community. In order to do this, we have created a metric called *Ratio 10:90*, which calculates the quotient between the number of contributions from to the top 10% of contributors divided by the number contributions made by the other 90%. The ratio is 1 if the total contributions by the top contributors are equal to the contributions made by the rest; less than zero if the top 10% contributes less than the rest; and greater than 1 if the top top 10% contributes more than the other ninety percent, which is often the case in collaborative communities.

Let n be the number of contributors and x_i the contribution of individual i with $i = 1, ..., n$ indexed in non-increasing order $(x_i \geq x_{i+1})$:

$$R_{10:90} = \frac{\sum_{i=1}^{k}x_i}{\sum_{i=k+1}^{n}x_i}, \qquad (2)$$

where k is $\lceil 0.10 * n \rceil$.

3.3 Ratios between top contributors

We have implemented a series of ratios to show the relation within the top contributors of the wiki, so we can observe if the work is highly concentrated in a very small group of contributors or, conversely, it is more spread across a larger group. The tool provides different ratios to focus on different parts of the tail of top contributors and to better observe the evolution of its stratification.

More precisely, the tool provides 3 ratios that calculate the quotient between the top 1st contributor, namely, the user who has made the largest number of edits, and other of the most active contributors. In particular, we make this calculation between the top 1st contributor and the contributor positioned at 5th, 10th and 20th percentile, from the list of contributors ordered in descending order of contribution size. Given the sorted list of contributors, the contributor at a given percentile, say 5th, has contributed less than the 5% of contributors that precede him or her, but more than the rest, i.e. the 95th percent of contributors after him or her.

Besides, the tool provides two additional similar ratios. One computes the quotient between the user in the position of the 5th percentile and the user in the 10th percentile. And the other computes the quotient between the 10th percentile user and 20th percentile user.

The higher the values of these metrics the longer the right tail of the participation distribution, i.e. the tail of the top contributors. It is expected to have high values of these metrics for communities with a long tailed distribution of participation between the core contributors, and low values within communities with a rather flat distribution between core users.

Let n be the number of contributors and x_i, the contribution of individual i, with $i = 1, ..., n$, indexed in non-increasing order $(x_i \geq x_{i+1})$:

$$P_{j/k} = \frac{x_{\lfloor j*0.01 \rfloor}}{x_{\lfloor k*0.01 \rfloor}}, \qquad (3)$$

being $j \in \{1, 5, 10\}$ the dividend percentile and $k \in \{5, 10, 20\}$ the divider percentile, with $j > k$.

4 TEMPORAL ANALYSIS OF PARTICIPATION INEQUALITY IN SOME WIKIA WIKIS

Below we present some of the findings we have accomplished in our exploratory research using WikiChron. In these three cases, we aim to present a diverse representation of growth curves and trajectories of participation distributions from the wikisphere. These findings show that the relationship between growth and inequality is more complex than we usually assume, and that in some cases may result counter-intuitive or challenge some of our preconceptions. From our experience, complexity naturally arises if we carefully observe diverse wikis in terms of topics, community sizes, languages, etc. We hope that the cases presented below serve to stimulate reflection and further studies about growth and inequality in wikis.

It is important to remark that in our analysis we only used WikiChron in order to prove that is a suitable tool for exploratory research. The graphs included have been also created by WikiChron.

Finally, it is worth clarifying that the monthly time series that we use in the plot represent the cumulative version of the metrics, unless stated otherwise. For example, when plotting the number of edits we refer to the cumulative edits from the wiki creation until that month. However, if we plot only the edits made during each month we would refer to the time series as *monthly* edits. This holds also true for ratios, coefficients, etc.

4.1 Opposite inequality trends in hobbies wikis

We will analyze two wikis concerning hobbies, namely the Cocktails Wiki[5] and the Home Wiki[6]. The topic of the first one is more restricted around cocktail recipes, while the second comprises decoration, gardening, improvements, etc. We will compare their evolution in their first 118 months, almost 10 years, totaling for 1700 articles the Cocktails wiki and 900 articles the Home wiki, approximately.

Interestingly, at the end of this period the size of both communities (including both registered and anonymous users) was around 440 people. However, the distribution of the contributions was different and the inequality evolution shows opposite trajectories. Both the Gini coefficient and the ratio 10:90 reflect that the growth from the Cocktails wiki happens while the wiki inequality is growing; however, in the case of the Home wiki it is decreasing. We show the later metric because its value is more understandable and its changes better reflect inequality changes (Fig. 2). The cross between both trajectories in the month 80 can be appreciated in other metrics around the same time.

Since the Cocktails wiki inequality is higher, it would be easy to reach to the quick conclusion that it is more "elitist" than the Home wiki. However, a closer look into the distribution evolution reveals that the elitism of the Cocktails wiki is rather inclusive as there is a larger pool of active core users. Meanwhile, the Home wiki contributions come from a smaller elite, together with casual contributors. Figure 3 shows the ratio between the 10th and 20th percentiles of the edits per contributor distribution. In the case of the Cocktails wiki the trend of the ratio is higher and more pronounced from month 64 and at the end of the period its value is 54, i.e. the contributor in the 10th percentile contributes 54 more
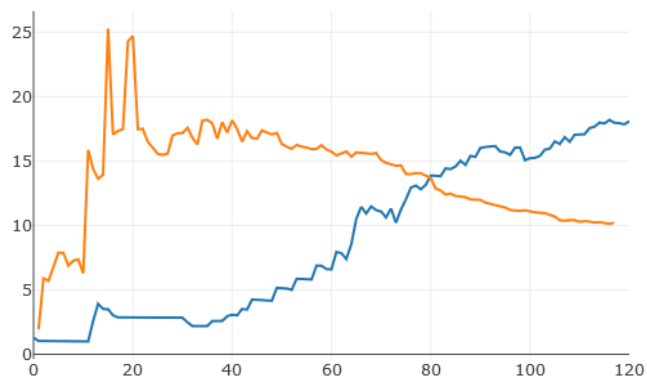
---

[5] `http://cocktails.wikia.com`
[6] `http://home.wikia.com`

Figure 2: Ratio 10:90 in the Cocktails (blue) and Home (orange) wikis.
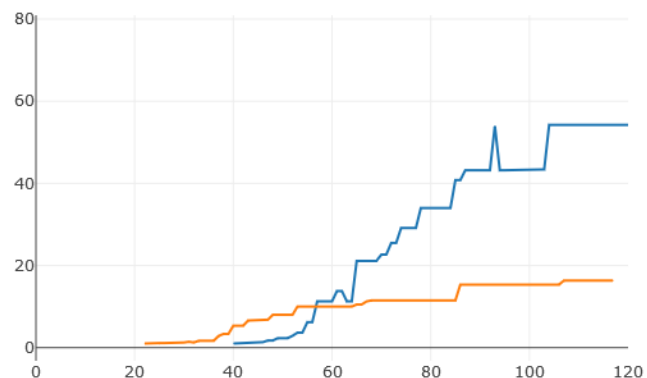

Figure 3: *Contributor percentile 10th-20th* metric in the Cocktails (blue) and Home (orange) wikis.

times than the contributor in the 20th percentile The plot of the ratio between the 10th and 5th percentile is similar, which together with the constant growth of the community further supports the hypothesis of inclusive elitism.

Both communities grow continuously, but Cocktails wiki slightly grows faster in the last period. However, the key point is that from month 87 the edit activity in the cocktails community notably increases and also regularly involves more users. This growth in terms of community, activity and larger elite is resilient and leads the project towards a more mature stage, also reflected when visiting the wiki.

4.2 Examples of horizontal Q&A and fan communities

The Wikianswers Pets wiki[7] was created in 2007, when Q&A popular competitors included Answers.com and Yahoo! Answers. According to the plots, the first 18 months the wiki remained virtually inactive (only 5 articles and 12 edits in articles), but during the next 18 months it blossomed reaching 1802 articles, 1872 users, and 7239
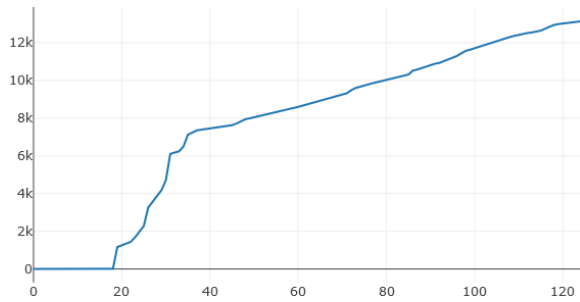
---

[7] `http://pets.answers.wikia.com`

Figure 4: Article edits in Pet Answers wiki.

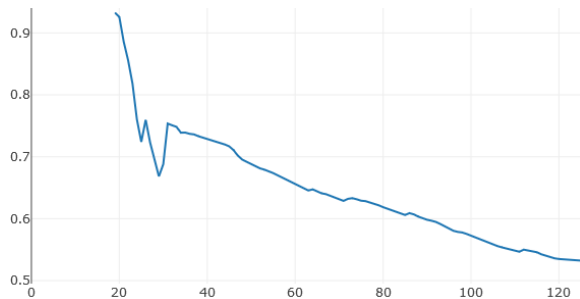Figure 6: Ratio 10:90 in Good luck Charlie wiki.

Figure 5: Gini coefficient in Pet Answers wiki.

edits. From this time, the wiki enters in a maturity period of steady growth in terms of edits, articles and users of more than 7 years. We show just the edits time series in Fig. 4 for the sake of brevity.

The most relevant feature of this community is its surprisingly participatory and horizontal dynamics. For example the monthly time series of article edits per user rarely takes a value over 2, the percentage of anonymous edits steadily increases after the blossoming period and reaches 51.5% of total edits at the end of the studied period.

The community still has an active core users, but their contribution is less significant than in the prototypical wiki. This is reflected by the surprisingly low values of the *ratio top 10% Vs rest* and Gini coefficient, which are 1.3 and 0.53, respectively, at the end of the studied period. Figure 5 shows the decreasing evolution of the Gini coefficient.

An inspection of Wikianswers Pets reveals that it is a useful resource where questions and sound answers are regularly posted, even if the experience as a Q&A system feels slightly outdated in 2018.

In fact, this wiki may be a counter-example to previous findings by Khansa et al. [5] that suggest Q&A systems active participation rely on incentives, past behavior and tenure.

We can also find examples of successful fan wikis with a rather high degree of collaboration. This is the case of the wiki about Good luck Charlie[8], an American sitcom aired on Disney Channel from April 2010 to February 2014. From February, the wiki is still growing in terms of content and contributors, even if at a slower
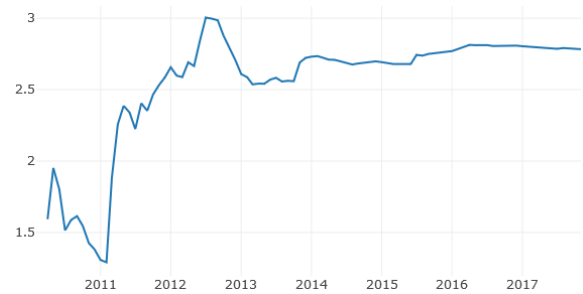
[8] http://goodluckcharlie.wikia.com

pace. It has 23.9k edits and 5.7k contributors. The participation of anonymous users reached a 39.4% in March 2013, while the show was aired. However, it has decreased to 26.2% meaning that the wiki is maintained now mainly by registered users. Despite the activity and success of the wiki, the Gini coefficient is relatively low and stable around 0.76. More significantly, Fig. 6 shows that the evolution of the 10:90 ratio never raised over 3, not even during the moments when the wiki grew faster. This means that the contribution of the top 10% of core users has been never greater than the 75% of the total contributions, which from our experience using this ratio is a remarkable achievement in terms of fan collaboration in wikis.

4.3 The limits of Gini coefficient for comparative studies

In this section, we will compare the first 99 months of activity of the LEGO (the popular construction toys) wikis in French[9] and German[10]. After such period both communities seem healthy and active. However, the level of edit activity (Fig. 7) reached by the French version is almost 4 times higher, and its number of articles is 2 times higher. On the other hand, the German community size almost doubles the French one (Fig. 8). A look at the Gini evolution in Fig. 9 shows that both wikis reach a stability level in week 42 around 0.99 for the French version and 0.96 for the German one. Thus, we could conclude that contributions are similarly distributed in both communities, being the difference in the community size and the contributions volume.

However, a look at the 10:90 ratio in Fig. 10 refutes that conclusion. The top 10% of core contributors in the German Wikipedia constantly accounts for between 25 times more contributions than the rest of the contributors, starting in month 40. However, this ratio rockets in the French case, up to close to 150 in the same period. That is, the French top 10% concentrates most of the contributions: specifically, they contribute 150 times more than the rest of the community, i.e. astonishingly 90% of the community performs less than 1% of the contributions.

A close look at the ratio between the top percentile contributors in both wikis reveals more subtleties. In the French wiki, the group of strong core contributors represents less than the 5% of the total

[9] http://fr.lego.wikia.com
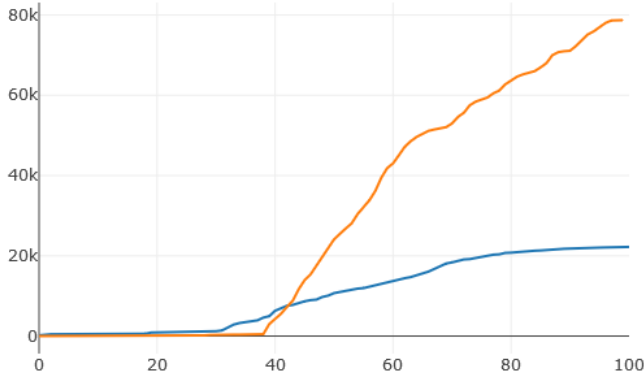[10] http://de.lego.wikia.com

Figure 7: Article edits in the German (blue) and French (orange) LEGO wikis.
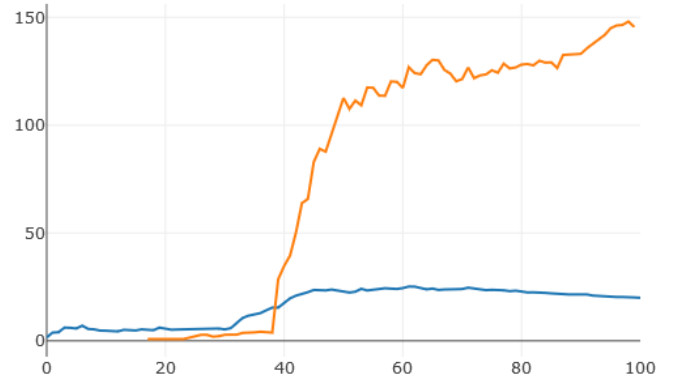

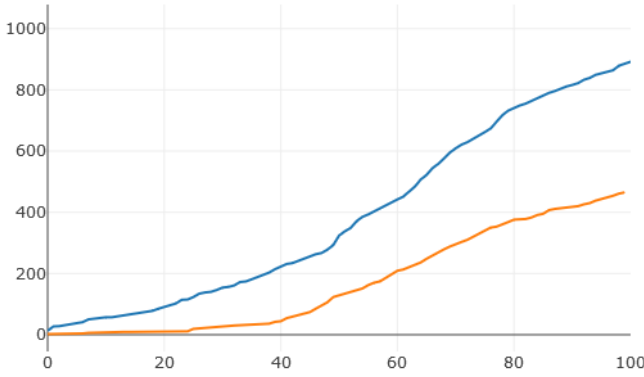Figure 10: Ratio 10:90 in the German (blue) and French (orange) LEGO wikis.

Figure 8: Users in the German (blue) and French (orange) LEGO wikis.
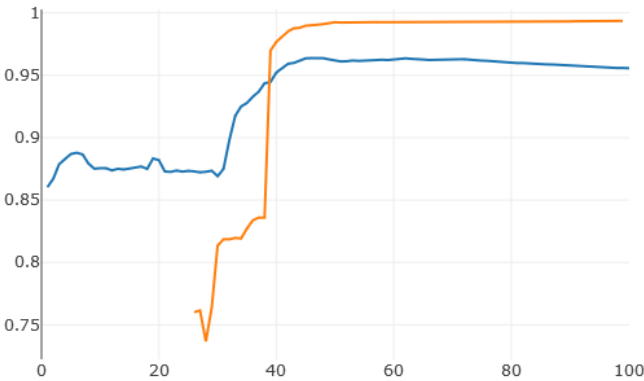

Figure 9: Gini coefficient in the German (blue) and French (orange) LEGO wikis.

population: the ratio between the top contributor and the contributor in the 5th percentile is 3.2k for the French wiki and "only" 780 for the German one.

In conclusion, a superficial analysis of the final Gini coefficient in both wikis, which is equivalent, would have given misleading findings. The combination of using time series and multiple metrics reveals that the participation distribution in both wikis is in fact very different.

5 CONCLUSIONS

We have illustrated that the distribution of participation in peer production communities exhibit a wide spectrum of diversity. Even if the levels of inequality are typically high, it can show up in multiple ways, and evolve very differently throughout the project life.

So far the field has identified structural factors that allow wiki growth, but there is still further research needed. From our approach, participation inequality is understudied, and in it we may find key aspects about resilience and survivability of peer production communities.

In order to deepen into this topic, it is required to use multiple metrics and not just confine ourselves to one. Furthermore, as this work shows, the use of one or two metrics may result in misleading conclusions. Specifically, the Gini coefficient may conceal more complex realities, hiding relevant aspects in the evolution of inequality. This adds up to objections presented in other works [15]. Thus, we suggest to use it in combination with other metrics. And in that sense, we consider WikiChron and its metrics to be highly appropriate to facilitate the studies on participation inequality, and in general, on wiki evolution.

We are aware that metrics alone cannot explain the health or state of a community; and yet, we can infer behavior and explore community features, to provide insights on e.g. the community growth. In order to reach a deeper understanding, it may be required to combine these quantitative metrics with information on structural factors such as governance and coordination [1], and even with the evolution of the Internet popularity of the topic the wiki is about. These quantitative studies ideally should be complemented with qualitative studies when possible.

Still, we strongly believe WikiChron may provide additional insights to both, quantitative and qualitative researchers that can

preliminarily explore the set of wikis they want to analyze. For example, from the quantitative point of view, WikiChron could help to guide studies about different growth patterns in wikis or even about predicting the development of certain communities based on their characteristics and its past behavior. On the other hand, qualitative researchers may find it a useful tool to guide research about differences in fan wikis from different cultures, or to study whether the adoption of some norms affected the evolution of quantitative aspects in the community. Even wiki admins can use WikiChron to understand the past of similar communities in order to try to stimulate participation or detect stagnation signs.

We are deeply committed in the value of WikiChron, and thus will further develop it adding new metrics and features. Hopefully, that will enable interventions to, if not reduce inequality, approach challenges like under-participation, resiliency and sustainability of the wiki sphere.

6 ACKNOWLEDGMENTS

REFERENCES

[1] O. Arazy and O. Nov. 2010. Determinants of wikipedia quality: the roles of global and local contribution inequality. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work.* ACM, 233–236.

[2] K. Crowston, N. Jullien, and F. Ortega. 2013. Is wikipedia inefficient? modelling effort and participation in wikipedia. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on.* IEEE, 3197–3206.

[3] George Deltas. 2003. The Small-Sample Bias of the Gini Coefficient: Results and Implications for Empirical Research. *The MIT Press* (2003).

[4] N. Hara, P. Shachaf, and K. F. Hew. 2010. Cross-cultural analysis of the Wikipedia community. *Journal of the American Society for Information Science and Technology* 61, 10 (2010), 2097–2108.

[5] L. Khansa, X. Ma, D. Liginlal, and S. S. Kim. 2015. Understanding membersâĂŹ active participation in online question-and-answer communities: A theory and empirical analysis. *Journal of Management Information Systems* 32, 2 (2015), 162–203.

[6] A. Kittur and R. E. Kraut. 2010. Beyond Wikipedia: Coordination and Conflict in Online Production Groups. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10).* ACM, New York, NY, USA, 215–224.

[7] R. E. Kraut, P. Resnick, S. Kiesler, M. Burke, Y. Chen, A. Kittur, J. Konstan, Y. Ren, and J. Riedl. 2012. *Building successful online communities: Evidence-based social design.* Mit Press.

[8] S. A. Matei and R. J. Bruno. 2015. Pareto's 80/20 law and social differentiation: A social entropy perspective. *Public Relations Review* 41, 2 (2015), 178–186.

[9] M. F. Morell. 2010. Participation in online creation communities: Ecosystemic participation. In *Conference Proceedings of JITP 2010: The Politics of Open Source*, Vol. 1. 270–295.

[10] J. Nielsen. 2006. Participation inequality: Encouraging more users to contribute. *Jakob Nielsen's Alertbox* (2006).

[11] N. Oliveira, N. Andrade, and K. Reinecke. 2016. Participation differences in Q&A sites across countries: opportunities for cultural adaptation. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction.* ACM, 6.

[12] F. Ortega, J. M. Gonzalez-Barahona, and G. Robles. 2008. On the Inequality of Contributions to Wikipedia. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008).* 304–304.

[13] A. Serrano, J. Arroyo, and S. Hassan. 2018. Webtool for the analysis and visualization of the evolution of wiki online communities. In *Proceedings of the 26th European Conference on Information Systems (ECIS).*

[14] A. Shaw and B. M. Hill. 2014. Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication* 64, 2 (2014), 215–238.

[15] J. Stuckman and J. Purtilo. 2011. Analyzing the wikisphere: Methodology and data to support quantitative wiki research. *Journal of the American Society for Information Science and Technology* 62, 8 (2011), 1564–1576.

[16] N. Sun, P. P.-L. Rau, and L. Ma. 2014. Understanding lurkers in online communities: A literature review. *Computers in Human Behavior* 38 (2014), 110–117.

[17] A. Yang, H. Fan, N. Jing, Y. Sun, and A. Zipf. 2016. Temporal analysis on contribution inequality in OpenStreetMap: A comparative study for four countries. *ISPRS International Journal of Geo-Information* 5, 1 (2016), 5.