11-29-2018

Webtool for the Analysis and Visualization of the Evolution of Wiki Online Communities

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

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

Samer Hassan
Harvard University, shassan@cyber.harvard.edu

WEBTOOL FOR THE ANALYSIS AND VISUALIZATION OF THE EVOLUTION OF WIKI ONLINE COMMUNITIES

Research in Progress

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

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

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

Abstract

Commons-based peer production is a mode of social innovation where communities co-create openly-accessible resources exploiting the network effect of the Internet. Traditionally, research has focused on the successful cases of FLOSS and Wikipedia, whereas other communities such as wikis have not been studied sufficiently in-depth. Moreover, there is a clear lack of analytical tools to facilitate their study. In this paper, we describe WikiChron, a new web-tool for the analysis and visualization of the evolution of wiki online communities. Such tool will facilitate the generation and validation of hypotheses around multiple research questions, for example, those around curves of growth or those about factors of rise and decline of wikis . This work presents two case studies: the analysis of growth curves in similar wikis, and the different evolution of coordination and communication in wiki projects. In sum, we argue WikiChron fills a gap in the software tools available for wiki research.

Keywords: commons-based peer production, online communities, analytics, knowledge production, wikis, temporal analysis, software tool

Acknowledgments

This work was partially supported by the project P2P Models funded by the European Research Council ERC-2017-STG (grant no.: 759207) and by the project COLOSAAL funded by the Spanish Ministry of Economy (grant id.: TIN-2014-57028-R).

1 Introduction

Commons-based peer production (CBPP) is an increasingly significant model of social innovation based on collaborative production, where communities rely on digital technologies to co-create knowledge and solutions for a variety of social needs exploiting the network effect of the Internet. This typically implies the creation of shared resources, which are freely accessible and re-usable by anyone (Benkler 2002, Fuster Morell 2010). Popular examples are Free/Libre/Open Source Software (FLOSS) and Wikipedia, although CBPP is progressively expanding into many other fields, from information (Wikihow, OpenStreetMap), to resource sharing (FreeCycle, BeWelcome) and manufacturing (Arduino, FabLabs, Thingiverse).

These communities have been praised as a better or even more virtuous mode of organization (Benkler and Nissenbaum, 2006). However, researchers in the last decade have been trying to understand the underlying problems that may exist under the hood, and across different fields. For instance:

- Organization Theory: Do they create horizontal organizations (Benkler 2006) or have hidden oligarchies (Shaw and Hill 2014)?
- Political Science: Are their fluid structures deconstructing hierarchical power structures (Bauwens 2005) or is their informality enabling undemocratic forms of power (Freeman 1972, Reagle 2012)?
- Economics: Are models of commons-oriented production inherently superior to classical models (Ostrom 2015), and if that is the case, why the rate of success is not higher (Grudin and Poole 2010, Schweik et al 2009)?

In order to respond to these and other questions, abundant analysis of CBPP communities have been performed, especially focused on its most classical examples: FLOSS and Wikipedia. Beyond their maturity and fame, there are also obvious reasons that facilitate analysis: FLOSS rely on large repositories of projects where it is easy to compare them, while Wikipedia acts also as a repository of multiple wikis in different languages (or subcommunities if WikiProjects are analyzed).

However, there are serious limitations of this classical research. In FLOSS we usually find one-man projects with hacker-type of contributors, which may not be representative of other types of communities (Lakhani and Wolf 2005) and show inherent biases, such its deep gender gap (Reagle 2012). In the case of Wikipedia communities, they are useful to understand successful projects, but are not suitable to understand the dynamics in not necessarily mature collaboration projects, or even failed/abandoned ones (Grudin and Poole 2010). They suffer inherent specificities that make them difficult to generalize to the whole CBPP field. Furthermore, most of the existing software tools to analyze CBPP communities are also focused on either FLOSS (Robles et al 2004, Gonzalez-Barahona et al 2015) or Wikipedia (Milne and Witten 2013). Hence, to answer the questions that CBPP poses, it is mandatory to study other kinds of communities, and, use new tools allowing their exploration.

Specifically, it is considered of interest to analyze other *wikis* that may or may not be as successful as Wikipedia, in order to see, first, whether the conclusions drawn from Wikipedia can be applied to other wikis. But also, to analyze the evolution of a diversity of wiki projects, and their curves of growth, to find factors that may help to detect the stagnation of a wiki project or its blossoming. Note the term "wiki" refers to a software designed to facilitate the collaborative, asynchronous creation and distribution of textual content. It also refers to the communities that use wiki software and to the products created by these groups (Leuf and Cunningham, 2001). The object of our study will be wiki repositories that host wikis where online communities produce open contents. Among those wiki repositories, Wikia (recently renamed "Fandom") is the largest one, accounting for 385,000 wikis[1] and having been object of previous research (Shaw and Hill 2014, Zhu et al 2014).

This paper will describe a new software tool to visualize the evolution of wiki communities beyond Wikipedia. More precisely, the tool currently includes wikis from Wikia due to the high number and diversity of communities that it hosts. Such tool, named WikiChron, is focused on enabling the generation and validation of research hypotheses, by providing a wide range of analytics and metrics.

[1] http://fandom.wikia.com/explore

This will enable the exploration of a diversity of questions, including the following: what causes the stagnation of growth? Is it possible its early detection? Which are the common factors of successful wikis and failed ones? Can we provide early warnings of abandonment?

Thus, WikiChron main aim is to analyze the evolution of collaborative production of knowledge, stimulating the exploration of under-researched areas, and expand the insights available about these communities. Moreover, to maximize its use, the tool prioritizes both openness, since it is released as free/open source, and usability, with a clean interface design.

The article is structured as follows. Section 2 provides a state of the art of the existing tools for analyzing and visualizing online collaborative production. Section 3 describes WikiChron, while Section 4 provides a variety of phenomena that may be discovered and analyzed with the tool in order to reason hypotheses about their cause. The paper closes with some concluding remarks in Section 5.

2 Tools for Analysing and Visualizing Wiki Production Beyond Wikipedia

We have conducted an extensive review of both the literature and available software tools for wiki research. To the best of our knowledge, the first works on wiki viability beyond Wikipedia are the works of Roth (2007) and Roth et al. (2008). We can also highlight Kittur and Kraut (2010) that focused on the topic of coordination and conflict and by Stuckman and Purtilo (2011) that focused on characterizing the distribution of effort across contributors and pages. More recently, Shaw and Hill (2014) proves in wikis from Wikia that peer production entails oligarchic organizational forms. These works are specially relevant for our case, because they acknowledge that studying only the case of Wikipedia limits the ability to generalize research conclusions to wiki production in general (and further, CBPP). However, only the work by Stuckman and Purtilo produced a tool that could be reused, *WikiCrawler*, used for crawling and parsing wikis, but it is not available online and, after contacting the authors, they acknowledged it has not been maintained since, so it is hence outdated.

The *Wikipapers* wiki collects a list of tools for both visualization[2] and data processing[3]. However, few of them seem to focus on the matter of exploring the evolution of the collaborative production of wikis through metrics or analytics; *StatMediawiki* and *Wikievidents* seem the most relevant cases. *StatMediawiki*[4] is already deployed in a website with some wikis as example. It has some deficiencies in terms of interactivity, time frame selection and variety of metrics. Metrics are not displayed as time series, but grouped by months, days, weeks and hours, which make it difficult to see evolution and maturity of wikis through time. *Wikievidens*[5] is on "Alpha stage" (a preliminary and unstable version) since 2012, currently outdated and abandoned since then. The tool is a desktop application written in Python, so users must download and run it. Besides, the tool failed in our local tests with recent wiki datasets. In addition to these tools, there was a Ruby library named *Wiki Explorator*[6], but we could not find its code or any relevant information anywhere, so we presume the project has been abandoned.

More interesting is the *Wikiapiary*[7] project that lists and monitors thousands of wikis running MediaWiki. However, the metrics displayed are simple (counts of edits, users, pages, etc.) and it aims to be a dashboard for the wiki admins and not a tool for wiki comparison. Similarly, Wikimedia Foundation recently released a web with interactive plots for their own projects that shows statistics of content view and content creations of their projects[8]. Again the tool is closer in spirit to WikiChron, but is a dashboard not useful for wiki comparison and the metrics displayed are mainly counts.

[2] List of visualization tools. http://wikipapers.referata.com/wiki/List_of_visualization_tools (visited on 11/24/2017)

[3] List of data processing tools. http://wikipapers.referata.com/wiki/List_of_data_processing_tools (visited on 11/24/2017).

[4] http://osl2.uca.es/statmediawiki/

[5] https://code.google.com/archive/p/wikievidens/

[6] http://wikipapers.referata.com/wiki/Wiki_Explorator

[7] https://wikiapiary.com/

[8] https://stats.wikimedia.org/v2/

Summing up, as researchers interested in the analysis of the evolution of wikis, we have not found any tool that may serve for this purpose. Instead of facing a discretional and laborious examination of several wikis using ad hoc code, we aim to fill this hole by providing the scientific and wiki communities an accessible tool that makes it possible to guide our research and others'.

3 WikiChron Analytical Webtool

With WikiChron we aim to fill the gap we found in the tools for analyzing, visualizing and comparing wikis evolution, as seen in the previous section. WikiChron aims to facilitate the task of observing and analyzing the temporal evolution of different MediaWiki wikis in parallel. A critical design decision was to use web technologies for the implementation, in order to maximize usability, and facilitate sharing and collaboration. Thus, it uses state-of-the-art technologies, and its architecture is designed to be extensible and maintainable.

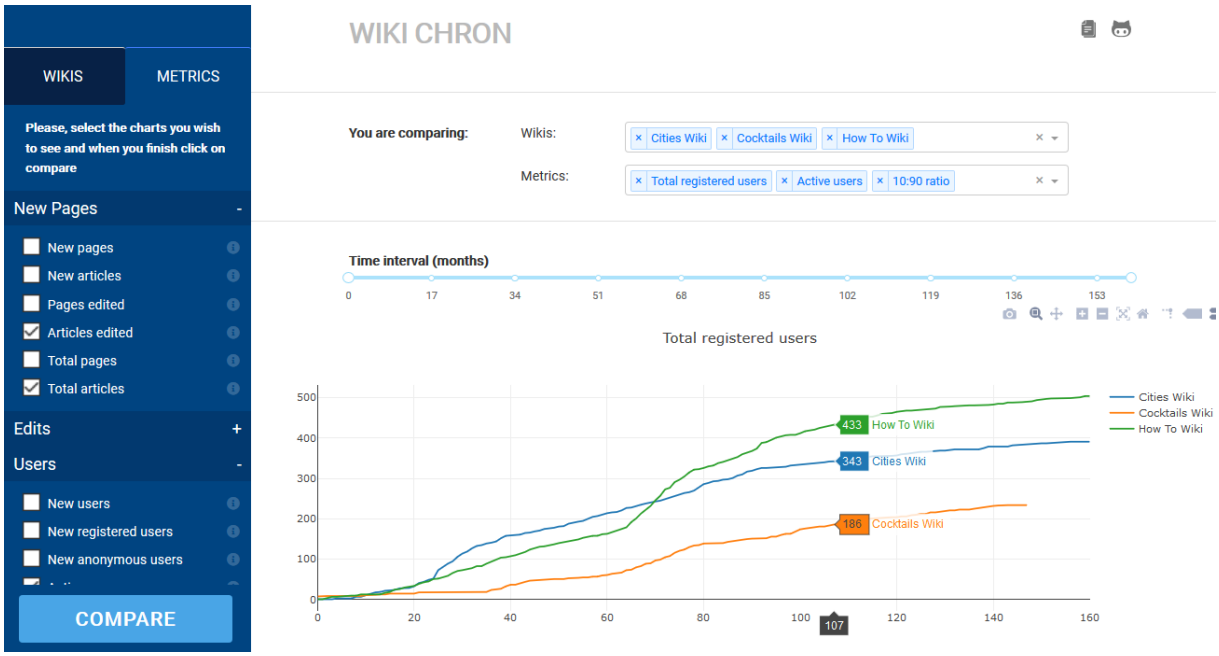

Figure 1. Capture of WikiChron comparing the evolution of registered users in three wikis.

The tool is written in the Python 3 programming language, a popular option in data analysis, and it uses Plotly[9] for the plotting and graphs interaction, in particular it makes use of the Plotly's Dash framework[10] to build an entire interactive web application. For the data backend processing and the statistics computations, it uses the Pandas[11] Python Data Analysis Library.

As of today, this tool supports a rich set of basic metrics, displayed as monthly time series, such as: new pages, new registered users, new anonymous users, edits in articles, edits in article talk, edits in user talk, the number of active users, etc. Ratios such as edits per user or edits per page and participation metrics such as the Gini coefficient can also be plotted. Since WikiChron focuses in collaboration, metrics about edits refer to 'human' edits (unless otherwise stated) as bots edits are filtered. Activity from registered and anonymous user is also differenced: registered users are grouped by name and anonymous users by IP address.

Most of these metrics can be shown as a time series of monthly values, or as a time series of accumulated values up to each month, being the latter more useful to analyze the so-called curves of

[9] https://plot.ly/d3-js-for-python-and-pandas-charts/
[10] https://plot.ly/products/dash/
[11] http://pandas.pydata.org/

growth. The tool briefly explains each metric. However, it provides a link to the documentation[12] that further explains the terms used and the assumptions made for the computation of the metrics.

In WikiChron, metrics can be selected for any subset of wikis available and all those selected wikis will be plotted together in a plot for each metric, facilitating comparison. The time axis of the plots displayed can be shown in either natural dates or months from the wiki creation date. The first option is useful to show how the contextual events affect the activity of one wiki. This can be useful, for example, in the case of fan wikis of franchise movies that are affected by the release of a new instalment. The second option is useful when the interest lies on the comparison of the evolution of different wikis. In this case, we are not interested so much in the events happening at particular dates, but in comparing how wikis evolve as they grow old.

By design, the tool allows researchers to interact directly with the graphs: zooming, filtering, moving around, adding and removing graphs dynamically, and similar functions. The aim is a smooth behaviour that feels intuitive for the user. The tool allows users to set up their analysis and, once they observe the results, to modify it by adding/removing wikis and/or metrics. Since user experience is a priority, most interactions imply appropriate response times, with a good performance.

The tool is publicly deployed for anyone's use in http://wikichron.science/. The tool provides a wide diversity of preloaded wikis of different characteristics, available to be analyzed with one click. Currently all of them are from Wikia, but other MediaWiki wiki data could be added.

WikiChron is free/open source, with its source code available in https://github.com/Grasia/WikiChron, and anyone can deploy it and inspect its code. Wikis can be downloaded and parsed using the scripts available[13]. The scripts download xml files, known as "data dump" files, which contain the history of all revisions within a wiki (i.e. the history of changes, their authors, timestamps, etc). That dump can be theoretically generated using the Special:Export page available in every Wikia wiki. However, there is a known bug in Wikia that leaves autogenerated dumps uncompleted starting from a certain amount of pages. To workaround this, we used a script that programmatically requests the revision history for every page in a wiki and automatically builds the dump. The dump is then processed in order to generate a clean and meaningful data file. Such processing is performed using a Python script that parses the dump and outputs a preprocessed CSV file. Lastly, WikiChron will load the input data from this CSV file in order to generate the requested graphs.

As it can be seen, we aim for clarity and transparency because we want our tool to be reliable and useful for online collaboration researchers.

4 Validating and Generating Research Hypotheses with WikiChron

In this section, we will illustrate the use of WikiChron in two case studies: the analysis of growth curves in similar wikis, and the different evolution of coordination and communication in wiki projects. Phenomena explored in other articles can also be easily observed with the help of our tool, and are not exposed here due to lack of space. For instance, Ceroni et al. (2014) shows that the number of monthly edits concomitantly peaks with the number of edited pages. Besides, it argues that when it does not happen it is because of the occurrence of a significant external event which concerns a subset of wiki articles, drawing more attention than usual towards those pages only. This can be easily observed with our tool. WikiChron can also be used to explore controversial phenomena, e.g. the effect of preventing anonymous users from editing the wiki. This decision is supposed to be helpful to avoid vandalism and spam, but may also harm the growth of the wiki and the community around it. Showing up next, There are a few examples of what can be studied with the help of WikiChron.

[12] https://github.com/Grasia/WikiChron/wiki

[13] https://github.com/Grasia/wiki-scripts

4.1 Analysis of growth curves in wikis

The tool allows to compare the different growth curves of wikis (especially those of similar age and size) in order to observe differences or structural similarities, and attempt to ascertain the reasons behind. In some cases the similarities can be striking. For example, we will analyze the case of two small but healthy communities around two anime series: the English community of fans of Strike Witches (ESW) and the Spanish community of fans of Shaman King (SSK). Both communities have around 7.6K edits in articles, and they have the same age since both were created in February 2010.

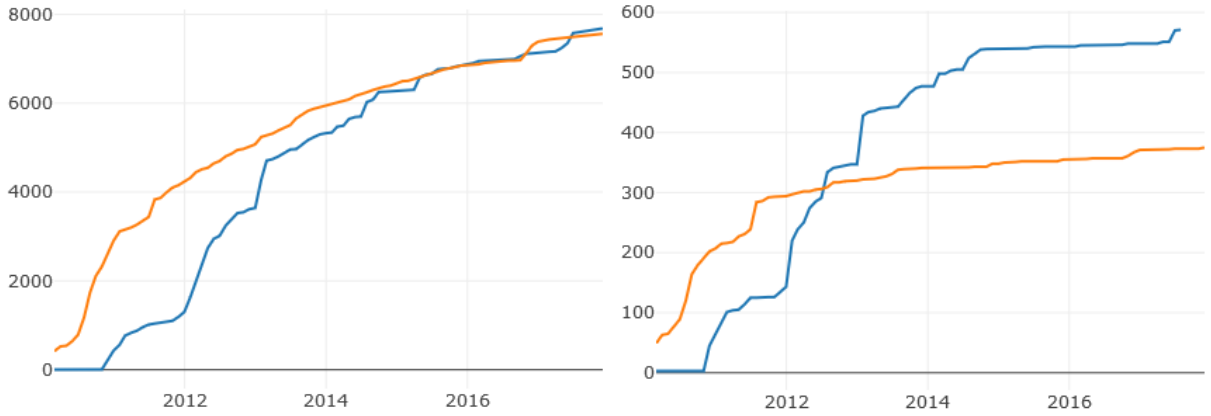

Figure 2. Total edits in articles (left) and total number of articles (right) of SSK (in blue) and ESW (in orange).

The difference in the number of articles could be caused by their different ways of organizing content (374 in ESW and 570 in SSK). It is interesting to observe that not only the number of edits in articles is similar but also the trend in the accumulated article edits, in both wikis, has been the same in the last 2 years. During that period, both wikis exhibit a growth in the number of articles close to none, so edits are mainly in old articles. The trend of the total number of edits per page is stable in the last years for both wikis, a value around 4.5 in ESW and around 5 in SSK. Interestingly, both wikis seem to be in a similar stage of growth in the last 3 years, even though their growth was different in the beginning; the ESW "took of" before than SSK, being a critical point for both wikis when they reached around 35 registered users (an approximate critical mass effect). The stage of growth is similar even if the number of registered users in ESW is increasing faster than in SSK in the last 2 years.

We can also compare the evolution of SSK with the English wiki of Shaman King (ESK), which is 2 years and a half older (i.e. comparing Spanish and English versions of wikis on the same topic). If we analyze the curves of growth in terms of edits in articles, registered and anonymous users, we can see that the ESK develops faster because it is the result of a bigger community. A closer look reveals that community members are slightly more active in average (ESK has around 12 total edits per user, while SSK has 14), but the size of the community seems the key driver. Interestingly, the increase of the rate of growth of the ESK community happens at an early stage (from month 13) when both wikis are similar in terms of articles and registered users. However, the number of anonymous users is very superior in the case of ESK which suggests that the number of casual participants is greater. We can consider that both wikis are successful, but ESK seems to be of use for a greater community than in the case of SSK, and hence it has more potential to develop, reaching a further stage of growth.

These explorations may provide insights on the differences of subcultures on equivalent topics in different languages, and, more generally, concerning the causes of the differences in the rate of growth of wikis that may be similar (or may have been at some point) in terms of articles and number of users.

4.2 The evolution of coordination and communication

Kittur and Kraut (2010) study the communication and coordination mechanisms in a diverse sample of wiki projects and argue that the conclusions that can be drawn are similar to those already known from

Wikipedia. Namely, that the percentage of edits in "article talk" (the discussion section of an article) remains relatively constant over time (between 5% and 7% of all edits), while the percentage of "user talk" (the discussion section of a user page) usually increases along time. The first one is believed to be proportional to the number of articles, while the latter scales super-linearly with the number of users, since adding a contributor increases the number of possible communication pairs exponentially. However, our tool allows us to find some counterexamples that contradict it.

For instance, we can explore the wiki of Good luck Charlie, an American sitcom aired on Disney Channel from April 2010 to February 2014. In this case, we can find that during the first two and a half years, the user talk was the communication mechanism used, with marginal activity on article talk. However, from that time on, the coordination abruptly moves to article talk pages and abandons user talk pages. It is important to note that at the time of the change (August 2012) the wiki already had 701 registered users and 399 articles. Communication mainly happens in activity burst periods. However, it was close to none during the first year of the wiki even if the number of users was over 40 since the 7th month, and it notably decreased when the show ended.

Another counterexample is the wiki of Harry Potter, a mature 12-year old wiki of currently over 23k registered users and over 24k articles. This wiki experienced a notable increase in the user talk edits between April 2009 and June 2013, but the activity notably decreases after that period and remains approximately constant since then, regardless of the fact that the number of registered users increased at a constant rate in the last four years. Interestingly, the decrease of activity in user talk pages coincides with the restriction of editions to anonymous users in the wiki. On the other hand, the article talk does not increase linearly with articles, but it moves concomitantly with the number article edits in the wiki, as Kittur and Kraut (2010) state.

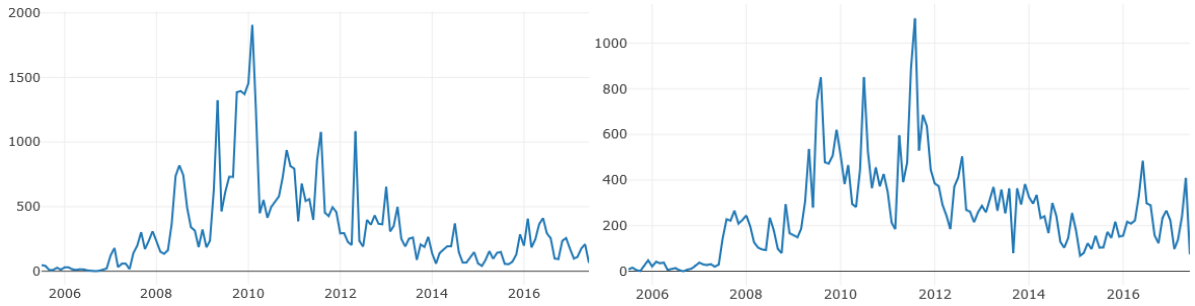

Figure 3. Number of edits in user talk (left) and article talk (right) of the Harry Potter wiki

5 Conclusions and future work

We believe that the described tool will serve well to stimulate research on online communities focused on knowledge production. Besides, it is expected to help to produce a better understanding of the events and factors that affect a community, by facilitating the quantitative analysis of the evolution of a set of wikis. The tool will help researchers to early validate incipient hypotheses by finding examples that confirm them, or counterexamples that contradict them. In this way, hypotheses may be refined or rejected before a confirmatory statistical analysis is designed to test them.

WikiChron has no access barriers as it is available in a public website with a diverse set of wikis ready to be analyzed. It can be used for social scientists, wiki administrators or researchers with no knowledge or time to use other analysis tools that require more effort on their side. Furthermore, the tool is being developed as free/open source and "in the open" to ensure transparency. We encourage researchers and programmers to try it and contribute to its further development and expansion.

WikiChron future improvements focus on a better user experience and richer capabilities. The tool will support the grouping of metrics and selected wikis within tabs, allowing to have several comparisons opened at the same time. In addition, it will allow to save and share opened "sessions". In this way, researchers will share their analysis not just using screenshots, but giving a direct link to the actual analysis. We also want to allow users to add a wiki to the corpus of wikis available in the tool.

Besides improvements in terms of interface, user experience and documentation, our road map will focus on the following strategic aspects: *higher-level metrics* and *curated datasets*.

The *metrics* already defined in the tool mainly focus on quantitative aspects of knowledge production. While the quantitative analysis of the production quality is an elusive aspect, Blumenstock (2008) shows that article size in terms of word count is a good proxy of its quality in Wikipedia. Other measurable quantitative aspects are considered proxy indicators of the resulting quality, and can be incorporated into our tool. More precisely we will add metrics to represent the following aspects:

- Contribution by user type. One criterium could be the Wikimedia definition of active user (people making at least 5 edits within a single month-long period), but this definition may not generalize well to smaller wikis. Moreover, we could characterize users in terms of the 1-9-90 profiles usually found in CBPP (Nielsen 2006) and study the evolution of core and casual contributors over time. Going beyond this quantitative approach, it would be possible to define roles such as the stable set of emergent roles that can be typically found in wikis according to Arazy et al. (2016), e.g. copy-editors, content shapers, layout shapers, watchdogs…
- Contribution by type. Fong and Biuk-Aghai (2010) propose a metric of contribution significance that depends on both the volume and type of an edit (e.g. adding content, inserting references, correcting spelling, moving content around, etc.). Later, Fong and Biuk-Aghai (2011) explore the visualization of such a metric. Similarly, we aim to represent the evolution of different contribution types in WikiChron in a meaningful way.
- Track the article richness and comprehensibility considering aspects such as cross–linking, media objects, layout of the page (headings, subheadings, etc.), Flesh Reading Ease scale of the articles, as it is analyzed in works such as Ehman et al. (2008).

Regarding *curated datasets*, we would like to make available through our tool different data sets of wikis that we or others use in published articles. The availability of curated datasets is key to facilitate the replicability of previous research and to deepen into the analysis of data sets already studied. It is not only a "good science" practice, but also a way to stimulate research.

In order to go beyond the Wikia universe, we aim to generalize our parser to other flavours of MediaWiki hosting services (the underlying software of both Wikipedia and Wikia). In this way, WikiChron could be used to contrast in Wikia findings found in Wikipedia and vice versa. As the number of wikis increases, the tool will need to incorporate mechanisms to search, filter and retrieve wikis by that previously defined characterization.

Finally, we are also users of our tool and want to use it to guide our research. For example, in the study on the growth and stagnation in wikis and the different ways of growing that wikis have. Hopefully to obtain theoretically-grounded empirical indicators of the "health" of a wiki, which would help wiki admins to assess the wiki state and guide them into taking corrective measures.

References

Arazy, O. Daxenberger, J. Hifshitz-Assaf, H. Nov O. and I. Gurevych (2016). "Turbulent stability of emergent roles: The dualistic nature of self-organizing knowledge co-production". *Information Systems Research* 27 (4), 792- 812.

Bauwens, M. (2005). The political economy of peer production. *CTheory*, 12-1.

Benkler, Y. (2002). Coase's Penguin, or, Linux and "The Nature of the Firm". *Yale Law Journal*, 369-446.

Benkler, Y. (2006). The wealth of networks: How social production transforms markets and freedom. Yale University Press.

Benkler, Y. and H. Nissenbaum (2006). Commons-based peer production and virtue. *Journal of Political Philosophy*, 14(4), 394-419.

Blumenstock, J.E. (2008). "Size matters: word count as a measure of quality on wikipedia". In: *Proceedings of the 17th international conference on World Wide Web* (WWW '08). ACM, pp. 1095-1096.

Ceroni, A., Georgescu, M., Gadiraju U., Naini K.D. and M. Fisichella (2014). "Information evolution in Wikipedia". In: *Proceedings of The 10th International Symposium on Wikis and Open Collaboration*. ACM. p. 24.

Ehmann K., Large,A. and J. Beheshti (2008). "Collaboration in context: Comparing article evolution among subject disciplines in Wikipedia." *First Monday* 13(10), 6.

Fong, P. K-F. and R. P. Biuk-Aghai (2010). What did they do? Deriving high-level edit histories in Wikis. In: *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*. ACM, a. 2.

Fong, P. K-F. and R. P. Biuk-Aghai (2011). "Visualizing author contribution statistics in Wikis using an edit significance metric." In: *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*. ACM, pp. 197-198.

Freeman, J. (1972). The tyranny of structurelessness. *Berkeley Journal of Sociology*, 151-164.

Fuster Morell, M. (2010). Governance of online creation communities: Provision of infrastructure for the building of digital commons (Doctoral dissertation).

González-Barahona, J. M., Robles, G. and Izquierdo-Cortazar, D. (2015). "The metrics grimoire database collection". In: *Proceedings of the 12th Working Conference on Mining Software Repositories*, pp. 478-481.

Grudin, J. and Poole, E. S. (2010). "Wikis at work: success factors and challenges for sustainability of enterprise Wikis". In *Proceedings of the 6th international symposium on Wikis and open collaboration*, p. 5.

Kittur, A. 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*. ACM, pp 215–224.

Lakhani, K. R. and Wolf, R. G. (2005). Why hackers do what they do: Understanding motivation and effort in free/open source software projects. *Perspectives on free and open source software*, 1, 3-22.

Leuf, B. and Cunningham, W. (2001). The Wiki way: quick collaboration on the Web.

Milne, D. and Witten, I. H. (2013). An open-source toolkit for mining Wikipedia. *Artificial Intelligence*, 194, 222-239.

Nielsen, J. (2006). *The 90-9-1 rule for participation inequality in social media and online communities*. URL: https://www. nngroup. com/articles/participation-inequality/

Ostrom, E. (2015). *Governing the commons*. Cambridge University Press.

Reagle, J. (2012). "Free as in sexist?" Free culture and the gender gap. *First Monday*, 18(1).

Robles, G., Koch, S., González-Barahona, J. M. and Carlos, J. (2004). "Remote Analysis and Measurement of Libre Software Systems By Means of the CVSAnalY tool". In: *Proceedings of the 2nd ICSE Workshop on Remote Analysis and Measurement of Software Systems* (RAMSS). pp. 51-56.

Roth, C. (2007). "Viable wikis: Struggle for life in the wikisphere." In: *Proceedings of the 3rd International Symposium on Wikis and Open Collaboration.* ACM. pp. 119–124.

Roth, C., Taraborelli, D. and N. Gilbert (2008). "Measuring wiki viability: An empirical assessment of the social dynamics of a large sample of wikis". In: *Proceedings of the 4th International Symposium on Wikis and Open Collaboration*. ACM, a. 27.

Schweik, C. M., English, R. and Haire, S. (2009). Factors leading to success or abandonment of open source commons: An empirical analysis of Sourceforge. net projects. *South African Computer Journal*, 2009(43), 58-65.

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

Stuckman, J. 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), 1564-1576.

Zhu, H., Kraut, R. E. and Kittur, A. (2014). "The impact of membership overlap on the survival of online communities". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 281-290.