

Mass collaboration or curatorship? The functioning of Wikipedia needs both

Federico Pilati

*Department of Sociology and Social Research, University of Milan-Bicocca,
Milan, Italy*

Pier Luigi Sacco

*Department of Philosophical, Pedagogical and Economic-Quantitative Sciences,
Gabriele d'Annunzio University of Chieti and Pescara, Chieti, Italy, and*

Oriol Artime

Department of Complex Systems, University of Barcelona, Barcelona, Spain

Abstract

Purpose – Our paper aims to address an open question in the scientific literature about Wikipedia: the structural logic behind two seemingly contradictory trends – peer-leveraged crowdsourcing and gatekeeping of core participation.

Design/methodology/approach – Using the complete dataset of the English, Spanish, and Italian versions of Wikipedia (2001–2020), we analyzed metrics such as the number of articles, creations, or edits performed by users. We calculated their distributions, adapted the Gini index to measure participation inequalities and employed network science methods to understand user-edit interactions.

Findings – Our analysis confirms significant disparities in content generation and engagement, emphasizing content editing. However, we demonstrate that these differences coexist with extensive collaboration. Specifically, our findings reveal that disparities in participation levels and collaborative editing complement each other. Curatorial leadership by a central group of contributors is extremely collaborative, while occasional contributors intervene flexibly in specific contexts.

Originality/value – High participation levels do not necessarily equate to high engagement. Our analysis shows that the inequalities in agency characterizing Wikipedia's creation and editing process stem from a high barrier to curatorship. This inequality fosters the self-selection of highly committed editors, whose sustained efforts underpin large-scale collaboration. Such collaboration, in turn, encourages new contributors to make occasional yet valuable interventions. This interplay between sustained curatorial efforts and flexible participation is a cornerstone of Wikipedia's success.

Keywords Wikipedia, Digital knowledge production, Digital participation

Paper type Research paper

1. Introduction

In the early 2000s, the birth of Web 2.0 and the new, surprising possibilities for collaboration on a massive scale that it enabled, shocked the public and favored the proliferation of utopian views of the role that the Internet was going to play in future economic value creation (Benkler, 2006; Williams and Tapscott, 2011), societal organization (Castells and Cardoso, 2005), and knowledge generation (Beninger, 2009), among others.

In this context, Wikipedia (and its logic based on wikis) was among the main supporting sources of evidence that optimistic manifestos of online participatory cultures drew upon (Van Dijck and Nieborg, 2009; Jankowski, 2023). As Wikipedia has steadily grown to



become the world's largest knowledge ecosystem to date, it has become clear that its societal value is related to the egalitarian promise of universal accessibility that the platform gave life (Rijshouwer *et al.*, 2021). On equally important grounds, Wikipedia represents a new socio-technical environment that pursues a truly specific goal in the context of a knowledge society: making the decentralized authorship and curatorship of massive, rapidly evolving knowledge corpora not only possible, but effective in terms of accuracy (Niederer and Van Dijck, 2010).

Over the years, an ever-growing number of research has demonstrated that Wikipedia's accuracy largely depends on the massive participation and diversity of its contributors (Benkler, 2006; Benkler *et al.*, 2015). On the other hand, the sociological literature highlighted that Wikipedia is based upon a much more hierarchical structure than it might be expected on the basis of its foundational rhetorics (Konieczny, 2009; Shaw and Hill, 2014). How is it possible thus to make sense of the apparently contradictory principles of massive participation and oligarchical control in Wikipedia?

Following previous multilingual studies (Ingawale *et al.*, 2013), in this paper we offer a compact, up-to-date, quantitative characterization of the structure of knowledge production for the whole first twenty years of the history of Wikipedia for its English, Spanish and Italian versions. Discussing our analysis, we propose that the inequalities in the content editing agency of participants and the presence of a core group of users who have a major role in the curation of the encyclopedia (i.e. the "Wikipedians") are key structural elements in the maintenance of the collaborative nature of Wikipedia.

Indeed, the emergent properties of Wikipedia as a knowledge ecosystem seems not to be a straightforward reflection of its foundational philosophy, but may rather be the result of a complex negotiation between different but *de facto* complementary organizational principles. In a nutshell, what characterizes Wikipedia's decentralized knowledge production is not just moderated peer interaction, but a subtle and tacit mix of openness and curatorship.

2. Literature review

The tension between open participation and oligarchical control in Wikipedia is a central aspect that requires careful examination. As it is well known from the analysis of large-scale social systems, there is a subtle and often counter-intuitive relationship between the rules and principles that are adopted at the micro-scale and the emergent properties at the macro-scale (Gershenson and Fernández, 2012). Enabling more egalitarian access at the individual level need not result in a system that favors egalitarian participation and agency, and vice versa (e.g. Bernelius and Vaattovaara, 2016). The fact that the explosion of digital culture has been largely driven by a strong demand of more active participation (Gere, 2009) does not guarantee in any way that the digital platforms that have evolved in response to such social expectations and demands actually cater to them (Tarnoff, 2022), irrespectively of the intentions of their designers. We are still learning how large-scale digital platforms really work, and what kind of behavioral, social, cultural and economic effects they bring about in the medium-long term (Miller, 2020). And no comprehensive theory is as yet available that allows us to make sensible predictions in this regard.

Wikipedia is no exception, and therefore the fact that it has been engineered to favor decentralized knowledge production does not amount to saying that such decentralization principle necessarily causes the emergence of highly pluralistic knowledge production at the macro-scale, or that the homeostatic forces that are at work in regulating the system need to function accordingly (Reagle and Koerner, 2020). However, Wikipedia is more than a repository of knowledge open to contributions from anyone; it serves as the most visible manifestation of a thriving online community (Pfister, 2011; Pentzold, 2011; Halfaker *et al.*, 2013; O'sullivan 2016). Consequently, the study of users actively contributing to Wikipedia's content has commanded significant attention, both within computer science, the humanities, and the social sciences (Loveland and Reagle, 2013; Schroeder and Taylor, 2015).

Over time, research has unveiled the existence of two opposing forces shaping Wikipedia. On one hand, the accuracy of its content appears to be an outcome of the *Wisdom of the Crowd* effect. The sheer multitude of contributors on a given topic gives rise to a unique form of collective intelligence, rendering Wikipedia’s content robust and trustworthy (Benkler, 2006; Benkler *et al.*, 2015). Conversely, scholars probing Wikipedia’s internal dynamics have unearthed traces of the *Iron Law of Oligarchy* (Michels, 1911). This phenomenon elucidates how the initial generation of “Wikipedians”, once established, sets a formidable barrier for new members to penetrate the core circle of content creators and editors (Konieczny, 2009; Shaw and Hill, 2014).

Drawing from these foundations, an ever-growing body of research, employing varied approaches and methodologies, has embarked on unraveling the intricate mechanisms governing Wikipedia’s functionality (Benkler *et al.*, 2015; O’sullivan 2016). This expanding corpus of literature hints that the accuracy of Wikipedia’s content is not solely dependent on normative devices such as guidelines (Viegas *et al.*, 2007; Forte *et al.*, 2009; Logan *et al.*, 2010), correction protocols and bots (Butler *et al.*, 2008; Yasseri *et al.*, 2012; Geiger, 2017; Yang and Colavizza, 2024), or vigilant and automated patrolling systems addressing errors (Hobbs and Roberts, 2018). It is equally reliant on a distinctive participatory culture motivating users to engage collaboratively in the publication and editing of content (Wilkinson and Huberman, 2007; Kittur *et al.*, 2009; Antin and Cheshire, 2010; Geiger and Ribes, 2011; Filder and Lavbič, 2017; Zha *et al.*, 2016).

Given this extensive body of literature, we posit the following hypotheses to delve into the intricate tension between mass collaboration and curatorship within Wikipedia:

- HP1. Inequalities in participation levels and heightened collaboration are not antagonistic but mutually reinforcing facets of Wikipedia’s structure.
- HP2. The core group of highly engaged Wikipedia users, who dominate the curatorial landscape, play a pivotal role in preserving the collaborative nature of Wikipedia’s content editing process.

3. Materials and methods

We work with the entire dataset of the English, Spanish and Italian versions of Wikipedia, from their launch (e.g. 2001 for the English version, which is the oldest) to the end of 2020. The dataset has been collected by adapting a script made available by Consonni (2019) and contains information on the date and time of creation of an article, and on the actual creator. Moreover, all the history of the edits of the articles is accessible – in particular, the information on who edited an article and when (see Table 1).

To characterize general patterns in Wikipedia’s content creation and editing we used a number of metrics, which we detail as follows.

On the one hand, to quantify the probability of observing a certain metric (e.g. the number of articles creations or edits performed by a user) we calculated its distribution on a logarithmic scale (due to the large variability observed in the data). These probabilities are computed by

Table 1. Description of the dataset

Wikipedia’s version	# of articles	# of edited elements	# of hyperlinks
English (EN)	~ 1.3 • 10 ^{^7}	~ 5.5 • 10 ^{^8}	~ 1.7 • 10 ^{^8}
Spanish (ES)	~ 3 • 10 ^{^6}	~ 7.7 • 10 ^{^7}	~ 4 • 10 ^{^7}
Italian (IT)	~ 2.1 • 10 ^{^7}	~ 6.8 • 10 ^{^7}	~ 4 • 10 ^{^7}
Source(s): Authors’ own work			

logarithmically binning in intervals the range of interest and counting the number of occurrences in each interval, and normalizing by the total number of occurrences and the range of the interval.

On the other hand, we employed the Gini index – an indicator largely used in economics and the social sciences to measure inequality in the distribution of resources, among other topics. We adapted and computed the Gini index as the relative mean absolute difference between all pairs of the elements of a set of measures $\{x_i\}$, that is,

$$G = \frac{1}{2N^2 \langle x \rangle} \sum_{i=1}^N \sum_{j=1}^N |x_i - x_j|$$

where N is the set size and $\langle x \rangle$ is the mean value. The Gini index is defined in the range $[0, 1]$, where 0 represents sets with perfect equality (all values are the same) and 1 represents sets that are maximally unequal (all but one of the values are 0).

Finally, to understand which types of users edit which kind of articles, we employed bipartite network analysis of user-article edit relationships. We clustered users by the number of unique articles edited (i.e. their unweighted degree in the user-article network). For each user cluster, we then calculated the average number of unique editors per article they edited. Conversely, we also clustered articles by the number of unique editors and calculated the average number of unique articles edited by the users contributing to each article cluster.

4. Results

The statistics of content creation in Wikipedia indicate very heterogeneous user behaviors. Although the expected participation pattern is that most users create and edit a small amount of content, this does not rule out the possibility that there is also a significant proportion of users who create and edit a lot of content. This is indeed confirmed by the fat-tailed functional shape of the probability distributions of several quantities both at the users and articles levels, which broadly span several orders of magnitude, in agreement with other distributions related to online activity (Barabási, 2005; Radicchi, 2009; Karsai *et al.*, 2018). Such quantities, shown in Figure 1, are the number of edits per article (Figure 1a), the number of article creations or edits

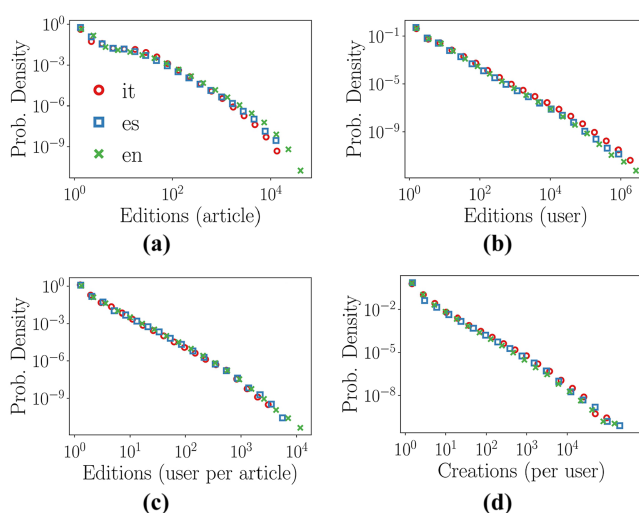


Figure 1. Probability density functions for the creation and editing patterns. Source: Authors' own work

per user (Figure 1b–d), and the number of edits made by a single user on a single article (Figure 1c).

If, as shown in Figure 1, there are huge differences both in the number of edits and creations per user, as well as in the number of edits per article by a single user, there must also be a significant inequality in active participation. Such inequality can be directly observed in Figure 2, which shows on a logarithmic scale the Gini index (amounting to ~0.95) calculated for the number of edits performed by users. When mapped against perfect equality (gray line in the graph) we see how small is the actual number of users who are actively involved in significant, stable editing activity, compared to potential ones.

The functioning of Wikipedia then admits the coexistence of a hard core of very active users, and the rare and occasional participation of less committed users in the editing process. But how this heterogeneous and implicitly hierarchical variety of modes of participation is conducive to the “wisdom of the crowd” that many scholars identify as the main driver of the accuracy of Wikipedia’s articles? To understand this aspect, we study the collaborative editing of articles (i.e. the number of single users who participate in their content curation through individual edits of a specific kind of article). We then need to measure the number of unique users as a function of the number of edits of an article. We show the result in Figure 3. What we see is that for low to intermediate values of the number of edits, the most likely outcome is that there are as many unique users as the number of edits. In other words, for significantly but not heavily edited articles, different individuals all contribute just one or few edits. Therefore, in the low-medium range of editing intensity of an article, we find a strong collaborative effort. Things tend to change as the number of edits of a given article grows. For frequently edited articles, it is more likely to observe repeated edits from individual users. The same applies to articles with very few edits.

However, in general, Figure 3 shows that for all the three languages considered it is highly unlikely to observe frequently edited articles by very few users only. These results then show that, in terms of edits distribution, crowdsourcing plays a very important role in the functioning of Wikipedia when it comes to checking and correcting the accuracy of most

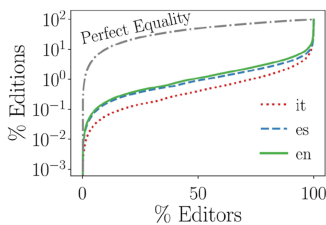


Figure 2. Participation inequality (Gini index calculated on edits per users). Source: Authors’ own work

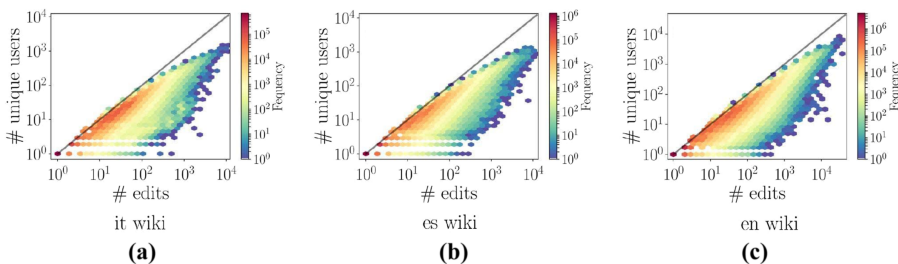


Figure 3. Heat map for the number of unique users participating in the editing of single articles. Source: Authors’ own work

of the articles. This applies to both the “average” Wikipedia article (i.e. those articles with low to medium number of edits), as well as, albeit to a little lesser extent, to the most curated entries.

Finally, we address the question of which type of users edit which type of articles. To this purpose, we first cluster the users by the number of different articles that they have edited, n_d . In other words, we are grouping editors according to their unweighted degree (number of connections) as if the system “users-articles” is seen as a bipartite network aggregated over time. In parallel, we compute for each article the number of single users that have edited them, n_A . Then, for each class of users – i.e. those that have edited $n_d = 1, 2, 3, \dots$ articles –, we take the average of n_A over the specific articles that the users belonging to that class have edited. Thus, we can know, on average, what kinds of articles are edited by users with a low vs. high editing activity. This information is shown in the upper part of Figure 4, where we observe similar patterns across the three considered language versions of Wikipedia. In particular, we observe that editors who have edited a few articles tend to be active in articles that have already been frequently edited. Moreover, such two quantities correlate negatively: the more active an editor, the more their contribution to less edited articles, on average.

To complement this picture, we switch from a user-based perspective to an article-based one. We now cluster the articles according to the number of unique users that have edited them – the quantity n_A defined above. In parallel, for each user we compute the number of distinct articles they have edited, n_d . Hence, for all the articles of each class – those sharing the same $n_A = 1, 2, 3, \dots$ –, we select the set of users who have edited them and compute the average of n_d over that set. This way, we gain insights on the editing behavior of the users with a certain frequency of editing activity of a certain type of articles, e.g. articles with a low vs. high level of collaborative editing. We report this information at the bottom of Figure 4.

Pulling together the analysis in Figure 4 with the patterns of participation described in Figure 3, it is possible to highlight a central implication of our results: as the maximum effort of collaboration (i.e. as many unique users as edits) is mainly present in low to intermediate values of the number of edits, and as in such range the edits are made mostly by very active users (i.e. with most probability the core circle of “Wikipedians”), we conclude that a relatively small group of very active users shapes the construction of the very same articles where we find the highest collaborative effort.

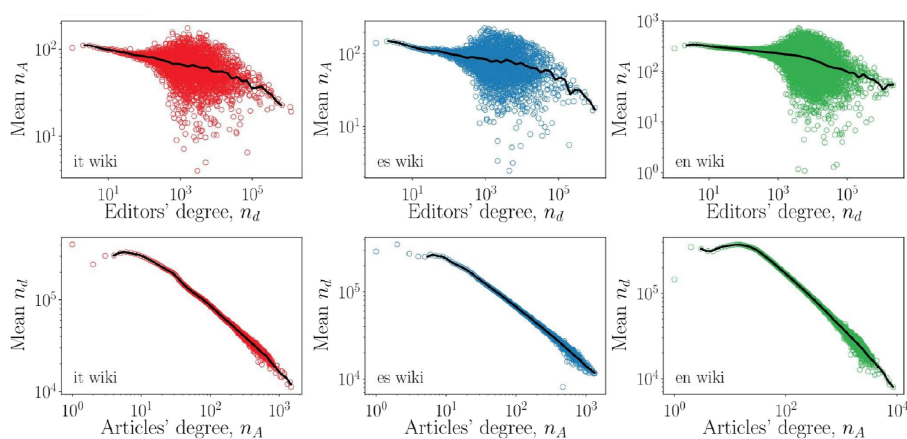


Figure 4. Mean distributions of editors and articles in the bipartite networks. Source: Authors' own work

5. Discussion

Our initial finding verifies the presence of significant disparities in content generation and engagement, which were previously emphasized in literature, with a specific focus on content editing. Nonetheless, we have also demonstrated that despite the curatorial leadership of the central group of contributors, the majority of edits are conducted in a highly cooperative manner. Thanks to such large-scale, flexible collaboration, new edits are often carried out by new individual users who are contributing for the first time to that specific article. In other words, our analysis reveals that differences in participation levels and extensive collaboration are not in opposition, but instead, they complement each other. This is apparent in the curation (i.e. edits) of the articles. On the one hand, we see that very few users dominate the curatorial activity through the editing process, but at the same time the same small, highly engaged “élite” works in an extremely collaborative way, either together with other highly engaged users on the many, less edited articles (that are in fact the backbone of Wikipedia), but also with the occasional users engaged with the editing of the most visible ones.

These results can be interpreted in the light of what we already know about Wikipedia (Reagle, 2010), but also in the light of the broader dynamic of digital participation that has quickly escalated with the advent of social media (Hargittai and Walejko, 2008; Hargittai and Jennrich, 2016). In comparison to other websites and social media platforms, Wikipedia indeed exhibits the widest chasm between content producers and those who read content (Chen et al., 2019). It has been pointed out many times that high levels of participation do not automatically amount to equivalent levels of engagement. The inequalities of agency that characterize the creation and editing of Wikipedia articles are based upon a high barrier to access to curatorship (Konieczny, 2009; Shaw and Hill, 2014). However, such inequality is the driver of the self-selection of engaged editors, whose solid and continued commitment is a pillar for the effectiveness of large-scale collaboration (Lin and Wang, 2020). We could therefore conclude that, rather than invoking a generic wisdom of the crowd to explain the functioning of Wikipedia, we should rather refer to a collaborative culture of co-creation that informs the Wikipedia ecosystem at all levels (Jemielniak, 2014). Creating accurate content on such a scale and level of variety requires not only a foundational model of collective intelligence that can seamlessly adapt to the level and complexity of the epistemic challenge posed by any specific article (Lichtenstein and Parker, 2009; Livingstone, 2016), but also shared goals, sense of community and social capital, so that users are driven not by the desire of individual recognition but by the social incentives to anonymously cooperate with the other encyclopedia members in a subtle, ever changing variable geometry (Shachaf and Hara, 2010; Lageard and Paternotte, 2021).

Our analysis also has many limitations. We provided a very aggregate characterization of the editing dynamics over the history of Wikipedia, but do not consider how specific thematic areas influence the levels and intensity of participation both within the small circle of very active users and the larger pool of more occasional contributors. Also, we do not consider the level of development of the single articles. Whereas some of them may be just short and not particularly detailed texts, others may be extremely lengthy and detailed. At the current level of analysis, we are unable to distinguish between these two very different kinds of articles and to check to what extent they elicit different dynamics of editing and by what kinds of contributors.

Furthermore, we have not addressed the role of bots, which are very present in Wikipedia’s editing process. Bots can range from fully automated entities to those that are partially supervised by human editors. Accurately categorizing and differentiating between human users and bots, and assessing the extent of their automation versus human oversight, is a difficult and often impractical task (Ferrara, 2023). This is a significant limitation that readers should keep in mind when interpreting our findings. Rather than attempting to separately categorize and analyze human users and bots in Wikipedia – a process that would involve still unresolved methodological challenges (e.g. Stein, 2014) – we have chosen to consider the interconnected roles of all elements engaged in the editing process. However, it is important to

recognize that this decision to treat bots and human contributors as part of a unified analytical framework does represent a limitation of our study. Future research could benefit from more targeted investigations into the distinct roles of bots within Wikipedia's complex environment.

Finally, we considered only three versions of Wikipedia in three languages. It would be important to extend the same analysis to a much larger number of languages to check whether there are languages (and corresponding cultural spheres) which are characterized by major structural differences with respect to the common trends found here. All these lines of research, and most likely many more, need to be thoroughly investigated in the future to help us build a more solid knowledge of how Wikipedia actually works and how it is likely to evolve.

6. Conclusion

Wikipedia has often been taken as a virtuous example of a digital platform that unleashes the power of the “wisdom of the crowd” as enabled by Web 2.0 affordances. While the unprecedented large-scale participation in the foundation and evolution of Wikipedia is a matter of fact, research has shown that, despite the massive collective effort, Wikipedia is also characterized by a highly selective curatorial leadership, consisting of editors (i.e. the “Wikipedians”), which show very high levels of commitment in terms of time and effort. Our paper moved from the recognition of an open question in the scientific literature about Wikipedia, that is, the request for a theoretically founded and empirically grounded analysis of the structural logic behind two apparently contradictory trends observed over the years: peer-leveraged crowdsourcing and gatekeeping of core participation. Our first result has been to confirm the existence of strong inequalities in content production and participation that were already highlighted in the literature, focusing in particular on content editing. However, we have also shown how, despite such forms of gatekeeping of the core group of contributors, Wikipedia's users who manage the vast majority of the edits carry out this task in a largely collaborative manner. That is to say, our analysis suggests that inequalities in the level of participation and high levels of collaboration are not antithetical, but rather mutually reinforcing building blocks of Wikipedia.

Our results also offer a complementary perspective on the argument that the unrestricted and massively decentralized access to knowledge production enabled by the digital revolution would lead to the “end of expertise” and the takeover of “amateur culture” (e.g. [Sanger, 2009](#)). We contend that the critique regarding the “end of expertise” often targets a specific form of expertise, typically associated with academic or professional qualifications, rather than expertise in a general sense. The expertise richly on display within Wikipedia is often of a different nature – a deep understanding of the platform's norms, policies, and collaborative dynamics. It is crucial to distinguish between the expertise required to effectively navigate and contribute to the Wikipedia community and the subject-matter expertise necessary to add high-quality content. Indeed, these two forms of expertise can sometimes be in tension. Our findings suggest how, in an open knowledge ecosystem like Wikipedia, the role of so-called “networked expertise” remains vital ([Pfister, 2011](#)). The high levels of large-scale cooperation characteristic of Wikipedia allow for leveraging the untapped cognitive surplus ([Shirky, 2010](#)) of numerous individuals typically excluded from more restricted systems ([Yang and Lai, 2010](#)). Consequently, expertise can become less personalized and more participatory, leading to what has been termed “networked expertise” ([Pfister, 2011](#)). This perspective prompts a reflection on the evolving social mechanisms for certifying and legitimizing expertise in societies where knowledge is produced and disseminated at unprecedented scale and speed ([Hartelius, 2020](#)). In this light, it might be argued that Wikipedia fosters a more efficient utilization of human cognitive resources at a societal level ([Restivo and van de Rijt, 2014](#)) compared to analogous knowledge ecosystems with restricted access, a subject deserving of further exploration in future research.

The insights gleaned from our analysis of Wikipedia's structural logic hold several implications for teaching, public policy, and real-world practice. For teaching, our research

provides a compelling case study of how collective intelligence and collaborative cultures function in a digital environment. It can be used to illustrate concepts of online communities, distributed work, the evolution of expertise, and the challenges and opportunities associated with large-scale digital collaboration, fostering critical thinking about the nature of online knowledge production among students.

From a public policy perspective, our study highlights the complex interplay between open access to information and the crucial role of dedicated communities in ensuring the reliability and accuracy of collaboratively produced knowledge. This is particularly relevant in discussions surrounding digital literacy, media regulation, and the governance of online information ecosystems.

Finally, for practitioners involved in building and managing large-scale collaborative platforms, our findings underscore the importance of acknowledging and potentially fostering a core group of highly engaged contributors while simultaneously enabling and supporting the contributions of more occasional users. Understanding the dynamics between gatekeeping and collaboration can inform the design of platform features and moderation strategies aimed at maximizing both the quality and breadth of user contributions.

References

- Antin, J. and Cheshire, C. (2010), "Readers are not free-riders: reading as a form of participation on Wikipedia", *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, pp. 127-130, doi: [10.1145/1718918.1718942](https://doi.org/10.1145/1718918.1718942).
- Barabási, A.L. (2005), "The origin of bursts and heavy tails in human dynamics", *Nature*, Vol. 435 No. 7039, pp. 207-211, doi: [10.1038/nature03459](https://doi.org/10.1038/nature03459).
- Beniger, J. (2009), *The Control Revolution: Technological and Economic Origins of the Information Society*, Harvard University Press, Cambridge MA.
- Benkler, Y. (2006), *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, Yale University Press, New Haven CT.
- Benkler, Y., Shaw, A. and Hill, B.M. (2015), "Peer production: a form of collective intelligence", in *Handbook of Collective Intelligence*, Vol. 175.
- Bernelius, V. and Vaattovaara, M. (2016), "Choice and segregation in the 'most egalitarian' schools: cumulative decline in urban schools and neighborhoods of Helsinki, Finland", *Urban Studies*, Vol. 53 No. 1, pp. 3155-3171, doi: [10.1177/0042098015621441](https://doi.org/10.1177/0042098015621441).
- Butler, B., Joyce, E. and Pike, J. (2008), "Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in Wikipedia", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1101-1110, doi: [10.1145/1357054.1357227](https://doi.org/10.1145/1357054.1357227).
- Castells, M. and Cardoso, G. (2005), *The Network Society: From Knowledge to Policy*, Center for Transatlantic Relations, Paul H. Nitze School of Advanced International Studies, Johns Hopkins University, Washington, DC.
- Chen, X., Li, X., Yao, D. and Zhou, Z. (2019), "Seeking the support of the silent majority: are lurking users valuable to UGC platforms?", *Journal of the Academy of Marketing Science*, Vol. 47 No. 6, pp. 986-1004, doi: [10.1007/s11747-018-00624-8](https://doi.org/10.1007/s11747-018-00624-8).
- Consonni, C., Laniado, D. and Montresor, A. (2019), "WikiLinkGraphs: a complete, longitudinal and multi-language dataset of the Wikipedia link networks", *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13, pp. 598-607, doi: [10.1609/icwsm.v13i01.3257](https://doi.org/10.1609/icwsm.v13i01.3257).
- Ferrara, E. (2023), "Social bot detection in the age of ChatGPT: challenges and opportunities", *First Monday*. doi: [10.5210/fm.v28i6.13185](https://doi.org/10.5210/fm.v28i6.13185).
- Fidler, M. and Lavbič, D. (2017), "Improving information quality of Wikipedia articles with cooperative principle", *Online Information Review*, Vol. 41 No. 6, pp. 797-811, doi: [10.1108/oir-01-2016-0003](https://doi.org/10.1108/oir-01-2016-0003).
- Forte, A., Larco, V. and Bruckman, A. (2009), "Decentralization in Wikipedia governance", *Journal of Management Information Systems*, Vol. 26 No. 1, pp. 49-72, doi: [10.2753/mis0742-1222260103](https://doi.org/10.2753/mis0742-1222260103).

- Geiger, R.S. (2017), "Beyond opening up the black box: investigating the role of algorithmic systems in Wikipedian organizational culture", *Big Data and Society*, Vol. 4 No. 2, doi: [10.1177/2053951717730735](https://doi.org/10.1177/2053951717730735).
- Geiger, R.S. and Ribes, D. (2011), "Trace ethnography: following coordination through documentary practices", *2011 44th Hawaii International Conference on System Sciences*, IEEE, pp. 1-10.
- Gere, C. (2009), *Digital Culture, Expanded*, 2nd ed., Reaktion Books, London.
- Gershenson, C. and Fernández, N. (2012), "Complexity and information: measuring emergence, self-organization, and homeostasis at multiple scales", *Complexity*, Vol. 18 No. 2, pp. 29-44, doi: [10.1002/cplx.21424](https://doi.org/10.1002/cplx.21424).
- Halfaker, A., Geiger, R.S., Morgan, J.T. and Riedl, J. (2013), "The rise and decline of an open collaboration system: how Wikipedia's reaction to popularity is causing its decline", *American Behavioral Scientist*, Vol. 57 No. 5, pp. 664-688, doi: [10.1177/0002764212469365](https://doi.org/10.1177/0002764212469365).
- Hargittai, E. and Jennrich, K. (2016), "The online participation divide", in *The Communication Crisis in America, and How to Fix it*, Palgrave Macmillan, New York, pp. 199-213.
- Hargittai, E. and Walejko, G. (2008), "The participation divide: content creation and sharing in the digital age", *Information, Communication and Society*, Vol. 11 No. 2, pp. 239-256, doi: [10.1080/13691180801946150](https://doi.org/10.1080/13691180801946150).
- Hartelius, E.J. (2020), *The Rhetoric of Expertise*, Lexington Books, Lanham MD.
- Hobbs, W.R. and Roberts, M.E. (2018), "How sudden censorship can increase access to information", *American Political Science Review*, Vol. 112 No. 3, pp. 621-636, doi: [10.1017/s0003055418000084](https://doi.org/10.1017/s0003055418000084).
- Ingawale, M., Dutta, A., Roy, R. and Seetharaman, P. (2013), "Network analysis of user generated content quality in Wikipedia", *Online Information Review*, Vol. 37 No. 4, pp. 602-619, doi: [10.1108/oir-03-2011-0182](https://doi.org/10.1108/oir-03-2011-0182).
- Jankowski, S. (2023), "The Wikipedia imaginaire: a new media history beyond Wikipedia. org (2001–2022)", *Internet Histories*, Vol. 7 No. 4, pp. 333-353.
- Jemielniak, D. (2014), *Common Knowledge?: An Ethnography of Wikipedia*, Stanford University Press, Stanford CA.
- Karsai, M., Jo, H.H. and Kaski, K. (2018), *Bursty Human Dynamics*, Springer International Publishing, Cham.
- Kittur, A., Chi, E.H. and Suh, B. (2009), "What's in Wikipedia? Mapping topics and conflict using socially annotated category structure", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1509-1512.
- Konieczny, P. (2009), "Governance, organization, and democracy on the Internet: the iron law and the evolution of Wikipedia", *Sociological Forum*, Vol. 24 No. 1, pp. 162-192, doi: [10.1111/j.1573-7861.2008.01090.x](https://doi.org/10.1111/j.1573-7861.2008.01090.x).
- Lageard, V. and Paternotte, C. (2021), "Trolls, bans and reverts: simulating Wikipedia", *Synthese*, Vol. 198 No. 1, pp. 451-470, doi: [10.1007/s11229-018-02029-0](https://doi.org/10.1007/s11229-018-02029-0).
- Lichtenstein, S. and Parker, C.M. (2009), "Wikipedia model for collective intelligence: a review of information quality", *International Journal of Knowledge and Learning*, Vol. 5 No. 3, pp. 254-272, doi: [10.1504/ijkl.2009.031199](https://doi.org/10.1504/ijkl.2009.031199).
- Lin, Y. and Wang, C. (2020), "Wisdom of crowds: the effect of participant composition and contribution behavior on Wikipedia article quality", *Journal of Knowledge Management*, Vol. 24 No. 2, pp. 324-345, doi: [10.1108/jkm-08-2019-0416](https://doi.org/10.1108/jkm-08-2019-0416).
- Livingstone, R.M. (2016), "Models for understanding collective intelligence on Wikipedia", *Social Science Computer Review*, Vol. 34 No. 4, pp. 497-508, doi: [10.1177/0894439315591136](https://doi.org/10.1177/0894439315591136).
- Logan, D.W., Sandal, M., Gardner, P.P., Manske, M. and Bateman, A. (2010), "Ten simple rules for editing Wikipedia", *PLoS Computational Biology*, Vol. 6 No. 9, e1000941, doi: [10.1371/journal.pcbi.1000941](https://doi.org/10.1371/journal.pcbi.1000941).
- Loveland, J. and Reagle, J. (2013), "Wikipedia and encyclopedic production", *New Media and Society*, Vol. 15 No. 8, pp. 1294-1311, doi: [10.1177/1461444812470428](https://doi.org/10.1177/1461444812470428).

- Michels, R. (1911), *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy*, Hearst's International Library, New York.
- Miller, V. (2020), *Understanding Digital Culture*, 2nd ed., Sage, London.
- Niederer, S. and Van Dijck, J. (2010), "Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system", *New Media and Society*, Vol. 12 No. 8, pp. 1368-1387, doi: [10.1177/1461444810365297](https://doi.org/10.1177/1461444810365297).
- O'Sullivan, D. (2016), *Wikipedia: A New Community of Practice?*, Routledge, London.
- Pentzold, C. (2011), "Imagining the Wikipedia community: what do Wikipedia authors mean when they write about their 'community'", *New Media and Society*, Vol. 13 No. 5, pp. 704-721, doi: [10.1177/1461444810378364](https://doi.org/10.1177/1461444810378364).
- Pfister, D.S. (2011), "Networked expertise in the era of many-to-many communication: on Wikipedia and invention", *Social Epistemology*, Vol. 25 No. 3, pp. 217-231, doi: [10.1080/02691728.2011.578306](https://doi.org/10.1080/02691728.2011.578306).
- Radichchi, F. (2009), "Human activity in the web", *Physical Review E*, Vol. 80 No. 2, 026118, doi: [10.1103/physreve.80.026118](https://doi.org/10.1103/physreve.80.026118).
- Reagle, J.M. (2010), *Good Faith Collaboration: The Culture of Wikipedia*, MIT Press, Cambridge MA.
- Reagle, J.M. and Koerner, J. (2020), *Wikipedia @ 20. Stories of an Incomplete Revolution*, MIT Press, Cambridge MA.
- Restivo, M. and van de Rijt, A. (2014), "No praise without effort: experimental evidence on how rewards affect Wikipedia's contributor community", *Information, Communication and Society*, Vol. 17 No. 4, pp. 451-462, doi: [10.1080/1369118x.2014.888459](https://doi.org/10.1080/1369118x.2014.888459).
- Rijshouwer, E., Uitermark, J. and de Koster, W. (2021), "Wikipedia: a self-organizing bureaucracy", *Information, Communication and Society*, Vol. 26 No. 7, pp. 1285-1302, doi: [10.1080/1369118X.2021.1994633](https://doi.org/10.1080/1369118X.2021.1994633).
- Sanger, L.M. (2009), "The fate of expertise after Wikipedia", *Episteme*, Vol. 6 No. 1, pp. 52-73, doi: [10.3366/e1742360008000543](https://doi.org/10.3366/e1742360008000543).
- Schroeder, R. and Taylor, L. (2015), "Big data and Wikipedia research: social science knowledge across disciplinary divides", *Information, Communication and Society*, Vol. 18 No. 9, pp. 1039-1056, doi: [10.1080/1369118x.2015.1008538](https://doi.org/10.1080/1369118x.2015.1008538).
- Shachaf, P. and Hara, N. (2010), "Beyond vandalism: Wikipedia trolls", *Journal of Information Science*, Vol. 36 No. 3, pp. 357-370, doi: [10.1177/0165551510365390](https://doi.org/10.1177/0165551510365390).
- Shaw, A. and Hill, B.M. (2014), "Laboratories of oligarchy? How the iron law extends to peer production", *Journal of Communication*, Vol. 64 No. 2, pp. 215-238, doi: [10.1111/jcom.12082](https://doi.org/10.1111/jcom.12082).
- Shirky, C. (2010), *Cognitive Surplus. Creativity and Generosity in a Connected Age*, Allen Lane, London.
- Steiner, T. (2014), "Bots vs Wikipedians, anons vs logged-ins (redux) a global study of edit activity on Wikipedia and Wikidata", *Proceedings of the International Symposium on Open Collaboration*, pp. 1-7, doi: [10.1145/2641580.2641613](https://doi.org/10.1145/2641580.2641613).
- Tarnoff, B. (2022), *Internet for the People. The Fight for Our Digital Future*, Verso Books, London.
- Van Dijck, J. and Nieborg, D. (2009), "Wikinomics and its discontents: a critical analysis of Web 2.0 business manifestos", *New Media and Society*, Vol. 11 No. 5, pp. 855-874, doi: [10.1177/1461444809105356](https://doi.org/10.1177/1461444809105356).
- Viégas, F.B., Wattenberg, M., Kriss, J. and Van Ham, F. (2007), "Talk before you type: coordination in Wikipedia", *40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, IEEE, p. 78.
- Wilkinson, D.M. and Huberman, B.A. (2007), "Cooperation and quality in Wikipedia", *Proceedings of the 2007 International Symposium on Wikis*, pp. 157-164, doi: [10.1145/1296951.1296968](https://doi.org/10.1145/1296951.1296968).
- Williams, A.D. and Tapscott, D. (2011), *Wikinomics*, Atlantic Books, London.
- Yang, P. and Colavizza, G. (2024), "Polarization and reliability of news sources in Wikipedia", *Online Information Review*, Vol. 48 No. 5, pp. 908-925, doi: [10.1108/oir-02-2023-0084](https://doi.org/10.1108/oir-02-2023-0084).

- Yang, H.L. and Lai, C.Y. (2010), "Motivations of Wikipedia content contributors", *Computers in Human Behavior*, Vol. 26 No. 6, pp. 1377-1383, doi: [10.1016/j.chb.2010.04.011](https://doi.org/10.1016/j.chb.2010.04.011).
- Yasseri, T., Sumi, R., Rung, A., Kornai, A. and Kertész, J. (2012), "Dynamics of conflicts in Wikipedia", *PLoS One*, Vol. 7 No. 6, e38869, doi: [10.1371/journal.pone.0038869](https://doi.org/10.1371/journal.pone.0038869).
- Zha, Y., Zhou, T. and Zhou, C. (2016), "Unfolding large-scale online collaborative human dynamics", *Proceedings of the National Academy of Sciences*, Vol. 113 No. 51, pp. 14627-14632, doi: [10.1073/pnas.1601670113](https://doi.org/10.1073/pnas.1601670113).

Corresponding author

Federico Pilati can be contacted at: federico.pilati@unimib.it