The Gender Perspective in Wikipedia: A Content and Participation Challenge

NÚRIA FERRAN-FERRER, Universitat de Barcelona, Catalonia
MARC MIQUEL-RIBÉ, Tecnocampus-Universitat Pompeu Fabra, Catalonia
JULIO MENESES, Universitat Oberta de Catalunya, Spain
JULIÀ MINGUILLÓN, Universitat Oberta de Catalunya, Spain

Wikipedia is one of the most widely used information sources in the world. Although one of the guiding pillars of this digital platform is ensuring access to the diversity of human knowledge from a neutral point of view, there is a clear and persistent gender bias in terms of content about or contributed by women. The challenge is to include women as equal partners in the public sphere, in which Wikipedia is developing a central role as the most used educational resource among students, professionals, and many other profiles.

In this paper, we introduce the gender perspective in the analysis of the gender gap in the content and participation of women in Wikipedia. While most studies focus on one of the two dimensions in which the gender gap has been observed, we review both approaches to provide an overview of the available evidence. Firstly, we introduce how the gender gap is framed by the Wikimedia Movement strategy, then we evaluate the gender gap on content and participation, especially regarding editor practices. Finally, we provide some insights to broaden the discussion about the consequences of not addressing the gender gap in Wikipedia, and we provide some research topics that can support the generation of recommendations and guidelines for a community that needs both equity and diversity.

CCS Concepts: • Human-centered computing → Collaborative content creation; • Information systems → Social networks.

Additional Key Words and Phrases: Gender gap, Gender bias, Equality, Editors, Peer production, Participation, Retention, Digital inequality, Open content, Women biographies, Public sphere, Notability

ACM Reference Format:

1 INTRODUCTION

Information and communication technologies seem to have opened up a scenario in which pluralism and new forms of expression are more possible than ever before, giving rise to an interconnected public sphere [5, 10, 11, 48, 53]. Wikipedia is part of the current public sphere and has the unique potential to facilitate a more equitable production of knowledge [23] through commons-based peer production and the provision of virtual spaces for discussion. Wikipedia defends itself as a kind of intellectual democracy because it is ostensibly based on the principles of rational deliberation, consensus, and negotiation [8].

Wikipedia has transformed the way in which information is produced and distributed through open collaboration [51], but doubts have been raised about how decentralized, flexible, and open these new opportunities are, given that they are generated within a pre-existing economic, social, and political model [28]. Wikipedia has content policies
such as neutral point of view, no original research, and verifiability which might pose unintended difficulties to several collectives in order to reach access to the public sphere and include their contribution to knowledge construction in the encyclopedia.

Through agreed-upon rules for debates, it is said that the community of Wikipedians can reach consensus, constituting an ideal speech community in Habermas’s terms [21]. However, this notion of consensus on Wikipedia often results in exclusions, thereby failing to meet Wikipedia’s ideal goals of being the most comprehensive encyclopedia and including a plurality of points of view [17].

Moreover, it is well known that Wikipedia suffers from a strong and persistent gender bias in different ways [27, 37, 54]. This issue has been raised both in the community involved in the editing process (most editors are men) and in its available content (biographies of men outnumber those of women and tend to be more extensive). In this short paper, we will introduce the gaps in terms of the gender perspective, in its contents and its participation, in order to ensure that Wikipedia contents reflect the true composition of society and eliminate stereotypes as recommended by UNESCO in its Gender-Sensitive Indications for Media [20]. In this regard, it must be noted that the only element that appears to be fairly gender equally distributed is the readership of Wikipedia [25].

The organization of this paper is as follows. In section 2, we will explain how the gender gap is framed by the Wikimedia Movement strategy conversations over the years. In section 3, we will deepen into what we call the Gender perspective, which is the gender gap both in terms of content and participation, paying special attention to editor editing practices and the first stages of the editor’s life cycle measured by the metric retention. In section 4, we will conclude with some reflections on the negative consequences of such a gap and the urgency to address it. The aim of this paper is to introduce scholars to an unsolved long-term problem that affects the most consumed knowledge repository and encourage research on both its causes and the ways in which they could be addressed.

2 THE GENDER GAP IN THE WIKIMEDIA MOVEMENT AND THE STRATEGY CONVERSATIONS

Gender inequality, both in Wikipedia’s communities and their available contents, has been recognized by the Wikimedia Foundation, which has been working to address it. In 2011, Sue Gardner, former executive director of the Wikimedia Foundation, suggested some reasons that may prevent women from editing Wikipedia, such as difficulties with the interface, lack of time, lack of self-confidence, encountering a misogynistic atmosphere, and the absence of a culture of social interaction [16].

Currently, the Wikimedia Movement, which includes the Wikimedia Foundation and all the language and territory-based affiliates, has agreed on a new strategic direction towards the 2030 horizon. This agreement sets the goal of delivering knowledge as a service (“become a platform that serves open knowledge to the world across interfaces and communities”) and knowledge equity (“focus our efforts on the knowledge and communities that have been left out by structures of power and privilege [...]. We will break down the social, political, and technical barriers preventing people from accessing and contributing to free knowledge”). Accordingly, the strategic direction has collaboratively set the principles\(^1\) of inclusivity and people-centeredness (i.e., addressing people’s needs) as the core of its recommendations, and has established the goal to eliminate the gender gap and focus on the inclusion of underrepresented groups by 2030\(^2\).


3 THE GENDER PERSPECTIVE IN WIKIPEDIA

Several authors have analyzed the gender gap in Wikipedia, both in terms of content and participation. The gaps in both levels suggest that there is still a lack of gender perspective in Wikipedia. In the next two sections, we will review and describe its aspects.

3.1 The gender gap in content

Regarding the research conducted on the gender gap in content, less than 20% of biographies are from women in the Spanish Wikipedia [15]. One of the handicaps faced to develop new content in the Wikipedia is that an entry must be notable, that is to say, it must comply with the notability requirements: articles should be based on third party, non-affiliated sources, with some degree of editorial overview, as a guarantee of neutrality and quality. In this sense, Wikipedia’s vision of knowledge conflates cultural significance with visibility in secondary media sources [17]. This is especially relevant to women, who do not easily attract mainstream media interest and therefore are less frequently covered. Only 24% of news sources are related to women; that is to say, most of the people seen, heard, or read about in the media are men. Additionally, the news topics in which women are the most visible gather the least coverage [36]. The case of Spain is even worse, only 7.3% of women are represented as main characters in media [38].

Men and women also focus on different content areas in Wikipedia, thus reinforcing content imbalance between ‘male’ and ‘female’ topics [32]. The Wikipedia Gender Inequality Index measures the gender gap in the content of Wikipedia, following the analysis in time and languages [31] of the available biographies in different languages, and provides a quantitative evaluation of the gap. Other initiatives such as the Wikiproject ‘Countering Systematic Bias’ [56] aim at broader goals, including gender bias. There are specific studies about the content gender gap, such as the analysis of the number of female porn actresses’ biographies vs. the number of female poets’ biographies [30], attributing the differences to editors’ interests and showing how the lack of female perspective is present in the “sum of all knowledge”, the Wikipedia’s prime objective. Other analyses compare the born-digital encyclopedia with other traditional encyclopedias such as the Britannica [45], and previous studies aim at the Britannica bias as well [18].

Recently, a theoretical study, from the cyberfeminism approach, has analyzed the empirical studies on content gap in order to provide a framework for understanding this complex phenomenon which produces an asymmetric collective perception of knowledge. This research mentions at least ten types of biases in Wikipedia content: related to length of writing, centrality in the hyperlink network, lexicon used, number of sources, quantity and quality of images, multilingual notability, topics mentioned, classification of articles, among others [7].

3.2 The gender gap in participation

Regarding the gender gap in participation or community composition, Glott et al. [19] analyzed a survey to characterize Wikipedia’s contributors and found that only 12.64% were women. Lam et al. [32] tried to determine the nature of this gender imbalance. Analyzing users’ profiles, they found that female editors only comprised 16.1% of the profiles and only accounted for 9.0% of edits. In a previous study based on a massive content codification, we showed that female editors represent a small minority of the editors involved, accounting for just 11.3% of all the editors of the Spanish Wikipedia [39, 40]. This evidence is similar to that of the English Wikipedia. Although till now the data has been obtained by survey methods, it has been reported that the percentage of female editors is between 10% and 15% or even lower in the English Wikipedia [1, 32]. Nevertheless, in our previous quantitative research, we found that for those editors with a large number of edits over time, such differences were not statistically significant, or even reversed,
with females outperforming male editors \[40\]. According to this evidence, engaging, participating, and persisting in Wikipedia appears to be a much more complex process for most female editors than for men.

### 3.2.1 The gender gap in editing practices

When researching the gender gap in Wikipedia, different editing practices have been observed among men and women. Women are more exposed to conflicts such as reversion and blocking than men, pushing them towards dropping out. In 2017, Ford and Wajcman \[14\] described the gender gap as the result of a knowledge institution governed by power issues. Since power is male-dominated, those who do not fit in with what Wikipedia recognizes as knowledge are excluded.

Access to the public sphere is not equally available to everyone. According to Juliano \[29\], whoever is in the power position in hierarchical societies such as ours determines what constitutes legitimate discourse and what does not, as well as who has the right to express their point of view or must remain silent. Therefore, to understand the production of content, we need to better understand the culture behind it and, particularly, how exclusion practices are performed \[14\]. Juliano reminds us that women have represented the group that has been most systematically denied the right to talk in the public sphere throughout history. Missing individuals and collectives in the public sphere subtract plurality from the human condition. As Arendt \[2\] stated, if someone is prevented from accessing the public space, they are not only deprived of seeing and hearing others, but also of seeing and hearing themselves. For Arendt, the public sphere is the place where plurality concurs and has the double meaning of equality and distinction. The public arena is where individuals are differentiated by each other, and reality is produced from the intertwining of perspectives of all those who fit in the arena occupying different positions. Thus, a common is created where identities are distinguished and recognized.

Habermas defined the public sphere as the realm of our social life, in which public opinion can be formed. Therefore, the public sphere must permit the circulation of information to enable communicative interaction among individuals, to form a public consensus, and to facilitate decision-making. For Habermas, the success of the public sphere was founded on rational-critical discourse. In the public sphere, everyone is an equal participant, and the power of argument is the supreme communication skill \[22\]. Obviously, liberal values were placed upon patriarchal substrates, the subordination of women, who were kept in the private sphere to tend merely to domestic matters. Therefore, achieving the equality needed in the public sphere requires the epistemological and philosophical foundations of liberalism to be broken \[43\].

The Habermasian conception of the public sphere also understands deliberative processes as opportunities for consensus, although agonistic pluralism criticizes this idea, as consensus can only be built at the expense of the dissenters’ voices. Mouffe \[42\] believes that society is irreducibly plural, therefore the idea that all identities may deliberate based on shared communicative rationality is implausible. Also, the idea that an unadulterated and unbiased ideal speech situation should serve as a model is unrealistic. Mouffe’s main critique is that Habermas understands a rational consensus specifically in liberal terms, and this excludes individuals and collective identities who do not fully identify with these values.

In Wikipedia, the power relations, its ideology, and its culture are produced by infrastructures \[14\]. The rationales of the origin of Wikipedia are found in the modern encyclopedic tradition and the free software movement. Wikipedia is built on an installed base \[49\] and represents knowledge and how this knowledge is produced \[57\]. The encyclopedic tradition produces a specific kind of knowledge and has its roots in colonialism according to predefined norms and logics for the scientific knowledge \[55\]. Contrary to an ideal design, this knowledge is not open to everybody.

Furthermore, low female participation rates in Science, Technology, Engineering, and Maths (STEM) are part of the academic tradition \[9\]. Some critics blame this problem on women’s attitudes, rather than pointing to the lack of
welcoming spaces and methods in STEM disciplines and institutions. Works questioning whether technical knowledge and expertise are masculine skills \[3, 6, 35, 46, 55\] reveal the exclusion of women in the academic and encyclopedic tradition. These issues in science and technology extend to computer science and the free software tradition in which Wikipedia is rooted.

On the one hand, according to Ford and Wajcman \[14\], Wikipedia is extending the epistemologies of previously male-dominated technoscientific projects such as D’Alembert and Diderot’s encyclopedia, which contained articles and drawings explaining the world, culture, and society, among others, and were written by philosophers and academics. The gender gap is not just related to Wikipedia, there are also examples of how women’s writings were invisible, hidden or underrated throughout history \[47\].

On the other hand, Wikipedia is based on hacker culture \[34\], follows the hacker ethics \[26\], and it is based on an open culture of collaboration. At the same time, it has been built within a culture of white men, online geeks and programmers \[45\] where the masculine identity has its roots in computational culture \[52\]. The research on this topic suggests that the lack of participation of women is due to their lack of skills, confidence, and fear of criticism and conflict \[24\], but Ford and Wajcman’s point of view identified a culture that denies the space for women’s participation.

As Aaron Swartz, a well-known hacker, wrote: ‘a website is not based only on the technology but on its community’ \[50\].

Bryce Peake suggests that the change needed is not the incorporation of women, but “the creation of a space of multiple points of view” \[44\]. Achieving this change would first require a major cultural shift among Wikipedians. Although the ideal solution could be to expand the public sphere by including women, this strategy is not receiving support from feminist authors such as Landes and Young, who agree that the public sphere already has a gender bias. For them, the exclusion of women from the public sphere is no accident, they think that universality is homogeneous to appeal to impartiality and is therefore a way of silencing differences and plurality \[33, 58\]. In this sense, Benhabib \[4\] asserts that it is only possible to include women in the public sphere if they enter with full rights in the universal dialogue, and their specificity as women is recognized. In contrast to Habermasian’s universal communication community, Benhabib argues for the need for a community that shows solidarity, taking into consideration the needs for both equality and differentiation. This means that the public sphere needs to be feminized. However, the point of view of feminist criticism, women do not identify with the cultural patterns that respond to the male experience of the world, they do not feel comfortable with the requirements of public time and space, which demand exclusive dedication, meaning that others will have to take care of domestic demands \[12\].

3.2.2 The gender gap in retention. Several studies have shown that women seem to stop editing Wikipedia sooner than men \[32, 40\]. In the case of newcomers, the mechanisms used to attract potential editors are not enough to retain them, especially in the case of women. Some authors have pointed out that creating a safe and warm environment is a key factor to attract and retain new editors \[13, 41\]. Participation in WikiProjects\[3\] or edit-a-thons\[4\] are two examples of these spaces. Also being a Wikipedian in a smaller Wikipedia helps, as are said to provide a quiet and safer environment for participation \[13\]. Research in such spaces can be useful to better detect and understand possible signs of attrition among newcomers. This research could provide some basic recommendations and guidance to both editors and communities, fostering such spaces and smaller communities.

\[4\]https://en.wikipedia.org/wiki/Edit-a-thon
4 CONCLUSIONS

Adopting the gender perspective, both in content and participation, is crucial to ensure that Wikipedia reflects the true composition of society and eliminate stereotypes in one of the most used information/learning resources in the world.

In this sense, there is the need that women could effectively access, edit, and transform Wikipedia. To do so, on the one hand, we see the need for measuring their effective participation as well as to characterize their editing practices, observing the gender differences in content creation, modification, and participation in the discussion processes, among others. The assessment of editing practices can provide a set of different types of good and bad practices for the rest of the community that could contribute to the generation of a space of plurality and diversity once implemented.

Additionally, studying the grounds of the notability criteria, that serve as a basis to determine the value of articles through the analysis of their information sources, is critical for women’s biographies. The absence of sources or citations in an article does not always indicate that a subject is not notable. This is the case for women’s biographies, as they suffer an exclusion from mainstream media that makes it harder to provide the traditionally accepted sources. Alternative sources or new evidence for achieving the notability criteria should be suggested.

On the other hand, we foresee the need to deepen our knowledge on the retention of women editors, to be able to identify the critical points for dropping out and provide recommendations to foster their engagement. Using a gender-based perspective, this research could expand on previously examined concepts and theories, translating them into specific social actions. Collaboration between Wikipedian groups and projects, who share aspirations for reducing gender bias, with other entities such as libraries, schools, and universities, could provide fruitful opportunities for positive change.

ACKNOWLEDGMENTS

This study is funded by the Plan Nacional I+D+I of the Ministry of Science and Innovation of Spain (Ref. PID2020-116936RA-I00).

REFERENCES


