

ostensibly (fe)male-dominated trade, even if they have a natural predisposition for it [8].

Sources for such distorted views of a profession can be peers or the mass media - yet today, the Web has arguably become one of the main reference points for personal research, including when acquiring general knowledge about professions. And very frequently, Web surfers end up informing themselves from related **articles on Wikipedia**, the world's largest and most popular online encyclopedia. In 2016 for instance, articles in the category "Occupations" on the English Wikipedia received 368,512,005 unique hits together (of 93 billion in total) and the corresponding category "Beruf" on the German Wikipedia accumulated 30,669,129 unique hits (of 12 billion total)². Apparently, *a sizable proportion of Internet users is hence exposed to, and probably influenced by, the presentation of facts about professions on Wikipedia, and the way gender is portrayed in those professionals fields.*

At the same time, the encyclopedia has already been found to exhibit considerable gender biases in past studies [7, 14, 28] and a similar distortion in its profession articles is entirely possible. The article "Journalist" on the German Wikipedia (Figure 1) is a showcase of how visible gender disparities can become on Wikipedia. Throughout the article, there are many images depicting men and only two images depicting women. Further, considerably fewer female than male journalists are mentioned in the text by name. Lastly, the common female form "Journalistin" doesn't even exist as an article, instead redirecting to the male form "Journalist".

The objectives of our study, are hence to reveal (i) how (im)-balanced the gender presentation is on the profession pages of Wikipedia and (ii) test if any imbalance can be explained by underlying labor market data or other facts which could serve to rebut an inherent gender bias of Wikipedia. We focus on the German Wikipedia as it is less studied than its English counterpart and because the language features a large set of female vs. male profession descriptors for analysis, as gender-specific versions exist for almost all profession names.

As illustrated through the "Journalist" example, we focus on three dimensions of profession articles: we (i) study redirections³ between neutral, male and female profession descriptors, and we analyze (ii) the ratios of male and female images as well as (iii) the ratios of mentioned male and female names of professionals. Our work shows that for most professions for which a male and a female title exists in the German language, only the male title has a corresponding article. In the articles, almost four times more images depicting men than women were encountered. Articles about professions also tend to mention more men. In fact, 75% of all articles have proportions of mentioned men from 0.8 to 1.0. Moreover, 83% men and only 17% women were mentioned on average in the articles. To explain the male bias in the German Wikipedia, we compared it with the Google bias and offline imbalances in the labor market, indicating that Wikipedia not only inherits gender inequalities from other sources but also aggravates skewed gender ratios in many cases.

²According to the Wikimedia pageviews API as of 28.02.17

³A redirection automatically forwards from a requested lemma to the redirection target, another lemma/article. At the target page, a tiny notification "redirected from X" (EN) or "weitergeleitet von X" (DE) appears, cf. top of Fig.1.

2 DATA COLLECTION

In order to conduct our analyses, several datasets had to be collected and prepared.

Seed dataset of profession names. An official list of professions from the German Federal Employment Agency ("Bundesagentur für Arbeit")⁴ was retrieved in order to create a comprehensive seed list of corresponding male-female pairs of profession names.

We parsed the profession descriptors using the grammatical rules of the German language including corresponding suffixes in order to create appropriate male and female forms of job titles. 4274 pairs were generated, such as "Lehrer"- "Lehrerin", "Krankenpfleger"- "Krankenschwester", "Entbindungspfleger"- "Hebamme" etc. Further, we extracted 183 gender-neutral profession labels from the initial professions list, e.g., "PR-Fachkraft", "Fotomodell", "Aufsichtsperson". The complete lists were manually cross-validated by two of the authors.

Wikipedia profession articles dataset. To map the seed list of profession names to Wikipedia articles we extracted all articles belonging to the German Wikipedia categories "Profession" (DE⁵: "Beruf"), "(Public) position" (DE: "Amt"), "Person by occupation" (DE: "Person nach Tätigkeit"), and any of their subcategories down to the 5th depth level.⁶ This was done as (i) profession labels can be homonyms and hence match a seed list entry to a non-profession article and (ii) to find a comprehensive, yet still clearly profession-related set of articles to compute a more relaxed matching method with. We then constructed the intersection of the set of profession names from the seed list with the profession article title set, by applying a relaxed string-matching: the Levenshtein distance and ratio (the match proportion of two words) were calculated between each profession name and each article title. If the Levenshtein distance was at most 2 or the Levenshtein ratio was at least 0.8, the corresponding pairs of words were matched. Hence, we were able to find profession articles with titles written in a slightly different manner. In order to avoid inappropriate matches (e.g., profession "Bäcker" and name "Säcker"), all matched pairs were manually validated. We proceeded to add 22 professions from the seed list that we identified by hand to have a matching article in Wikipedia, but lacking the appropriate category label.

For each matched article title, additional information regarding its redirection was stored, i.e., whether the original lemma redirected to another article - and to which one.

As a result, a high-quality list of unique, non-redirect Wikipedia articles (885 entries) about professions was collected.⁷ An additional 820 lemmas that matched with the seed list were redirects.

Gender-specific person names. To identify if and how many female or male persons are mentioned in a profession article, we applied a two-pronged approach.

First, all internal links in the profession articles were collected which point to articles about persons either in the Wikipedia-categories "Woman" (DE: "Frau") or "Man" (DE: "Mann") using

⁴Retrieved from <http://berufenet.arbeitsagentur.de/berufe/berufe-beschreibungen.html> on 15.06.15

⁵Henceforth "DE" is used as an abbreviation for "German".

⁶Accessed on 07.02.16 from <https://de.wikipedia.org/w/api.php>

⁷One reason for the relatively small overlap is that the seed list is exhaustive and contains many specialized professions that seem not to be "notable" enough for the DE Wikipedia.

the MediaWiki API⁶ - these were each counted as one mention of a person with the Category-derived gender.

Second, the Named Entity Recognition method proposed by Al-Rafou et al. [1] was applied to the complete article text content, yielding all Entity Names of class “Person” that were identified. Then, for each found person mention, the gender was identified according to the first name. To make gender identification more accurate, several vocabularies [16, 20] were used. The results obtained by these two methods were combined for each article, producing an overall dataset of 5085 identified person labels (4272 men and 813 women). For cases where the two datasets clashed in the determination of the gender (2.77%), we were able to identify the correct genders automatically since persons from the first dataset are readily gendered. The end-result allowed us to reliably assess the number of female or male individuals mentioned.

Images of persons. The MediaWiki API⁶ was used to retrieve all images contained in the articles. Only images wider than 100 pixels were stored, assuming that small images are either icons or too small to recognize the gender. A manual corroboration of this assumption with about 80 pictures yielded zero false positives for this filter. Files with the following formats were also excluded: “svg”-files are vector images used solely for schemas and icons, and “ogg”, “ogv” are video and multimedia text formats. Thus, 906 images were collected from 345 profession articles. The remaining articles did not contain (suitable) images.

Labor market statistics. Gender-specific employment statistics were obtained from the “Statistics of the Federal Employment Agency” (DE: “Statistik der Bundesagentur für Arbeit”). The statistics consist of absolute numbers of men and women involved in profession subgroups as of June 30, 2015. Some examples of profession subgroups are: “8445; (Fremd-) Sprachenlehrer/innen” and “8442; Berufe in der Religionspädagogik”. Each profession in our seed dataset is annotated with a specific classification number by the Employment Agency, enabling the assignment to its respective profession group according to an accompanying profession directory (DE: “Klassifikation der Berufe 2010 - alphabetisches Verzeichnis der Berufsbenennungen”) E.g., “8445x” is the encoding of all professions in subgroup “(Fremd-) Sprachenlehrer/innen”. In this way we could match 871 of the 885 article-mapped professions to their respective labor market statistics including gender distributions.

Google hits. By using each of the female and male profession labels as a search query term, we collected the amount of Google search results (hits) generated through the Google Web Search API⁸ to gauge the number of Web resources featuring these descriptions. The search scope was restricted to the German language using the corresponding parameter in the API query, as some search terms could also be valid English words, for example.

3 RESEARCH METHOD

To assess whether and to what extent the German Wikipedia exhibits a gender imbalance with respect to professions, the collected

articles have been analyzed along three dimensions: (i) gender inclusiveness in job titles and a corresponding redirection analysis, (ii) analysis of male-female image proportions, and (iii) the balance of male-female mentions. We present these dimensions in the following subsections.

3.1 Redirection analysis

The following analysis was inspired by the scenario of a user wanting to inform herself about a specific profession and entering the respective (fe)male form in Wikipedia’s search bar or directly navigating to `de.wikipedia.org/wiki/<profession_title>`. We were interested in how many cases such a search will lead to either no results or a redirect to the opposite gender form or a neutral page – with redirects happening immediately after visiting a target article address, being indicated with an unobtrusive disclaimer, and hence possibly going unnoticed.

For each of the female-male pairs in our profession list (and the neutral labels), we applied these rules to sort each profession into distinct classes, indicating a potential skewness, or bias⁹, towards men or women:

- (1) If articles/lemmas do not exist at all on Wikipedia for a profession in our seed list – *no evidence*;
- (2) If (non-redirect) profession articles with both gender titles exist (i.e., male and female titles) – *neutral*;
- (3) If a profession article with a male / female title exists without the other gender form existing as an article – *male bias* or *female bias*;
- (4) If only an article with a neutral title exists – *neutral*.
- (5) If a profession lemma exists, but redirects to the other gender – bias in favor of the redirection target (i.e., *male bias* or *female bias*);
- (6) If a profession lemma with a male or female title redirects to the neutral form or field name – *neutral*;
- (7) If a profession lemma exists, but redirects to some article whose title is not in our seed dataset – *other*.

After assigning professions to these bias groups, we can assess whether professions are more likely to be presented via male, female or neutral profession names. However, the non-existence of one gender form can have a host of reasons, based, e.g., on the fact that virtually no (wo)men are employed in a specific profession. Also, we were interested how idiosyncratic the profession representation of Wikipedia is in comparison with the Web in general. For the latter, we compared the amount of results found through the Google search engine for each profession label to Wikipedia. For the former, we made use of the German labor market statistics.

Google hits. As a starting hypothesis, we presume that profession titles that appear on Wikipedia are more popular on the Web, meaning that there are more sources on the Web about them than about the corresponding profession titles of the opposite gender. Consequently, one should observe more search hits for profession titles represented on Wikipedia than for those that are not. As a proxy, we look into the number of hits returned by the Google search engine. For each profession, the number of hits was stored for female and male job titles. In order to assess if the difference

⁸Accessed on 09.02.16 from <http://ajax.googleapis.com/ajax/services/search/web?v=1.0&hl=de&btnG=Google+Search&q=>; the API results are not subject to custom personalization.

⁹Note that we do not associate “bias” here with any deliberate attempt to systematically discriminate, cf. discussion.

between Google hits for female and male titles is significant, a two-sided Mann-Whitney rank-sum test was utilized.

Next, we examined whether one can describe the relationship between the number of Google hits and the redirection bias of a profession. First, for each profession, the normalized difference between hits for male and corresponding female job titles was estimated using the following formula:

$$Normalized_difference_i = \frac{Hits_{male_i} - Hits_{female_i}}{Hits_{male_i} + Hits_{female_i}} \quad (1)$$

where $Hits_{male_i}$ is the number of hits returned for male title of profession i , $Hits_{female_i}$ is the number of hits returned for the corresponding female title of profession i . A positive difference indicates that more search results have been returned for the male profession title than for the corresponding female title, while a negative difference indicates the contrary. The difference approaches zero if both gendered profession titles returned the same amount of search results.

Then, two logistic regression models were fitted in order to predict the redirection bias using Google hits for profession names. The first model uses the state of being in the male bias group as a dependent variable; the second model deals with female bias instead. The regression functions for both models are given by:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i})}} \quad (2)$$

where $x_{1,i}$ is the normalized difference of Google hits of profession i , and $x_{2,i}$ is the Google hits for male title of profession i .

Labor Market. For the next analysis, each profession was coupled with the corresponding number of employed people per gender as per the labor market statistics collected. For each profession, the percentage of women involved was obtained. Then, the dependence between the percentage of women involved in the profession and the redirection bias group of the profession was tested. The null hypothesis is that two sets of measurements are drawn from the same distribution, i.e., the percentage of women involved in profession of the first bias group is drawn from the same distribution as the percentage from the second bias group. The alternative hypothesis is that values in one sample are more likely to be larger than the values in the other sample.

The Wilcoxon rank-sum test was used for each pair of the bias groups (male-female bias, male-neutral bias, neutral-female bias) in order to find profession groups which show significant differences in the percentage of employed women. Bonferroni correction [9] was used in order to control for the family-wise error rate. To describe the relations between the bias groups using the percentage of women involved in professions, logistic regression models were fitted. Analogously to the analysis of Google results, we predict whether a profession is in the female\neutral\male bias group. The explanatory variable is the percentage of women involved in a profession. Since not all bias groups show significant differences between distributions of values, we fit our models only between those which do.

3.2 Images analysis

To identify the gender of people on an extracted picture, a CrowdFlower¹⁰ task was set up. First, CrowdFlower workers were to identify whether the image shows one or several persons, or none. If the image depicted more than one person, workers were asked to identify whether s/he was depicted in a dominant way, as we are interested only in the gender of the main person. For every image with one person or where one person is dominant, workers were asked to identify the gender of that individual. Otherwise, they were asked to identify the gender of the majority of people on the photo. For non-recognizable gender or an equal number of men and women, corresponding options were offered.

Images depicting one (fe)male, one dominant (fe)male, and images with (fe)male majority were assigned the label “female” and “male”, respectively. Images where gender was not recognizable and with equal number of men and women were assigned two separate labels.

Each image was classified by at least three different workers. In order to control the reliability of all responses, the CrowdFlower accuracy threshold for workers was defined as 70% in the setting of the task. We also manually labeled 15% of the images such that one of 10 images shown to a CrowdFlower contributor would be from the labeled set - this was used as an additional worker accuracy control. If the accuracy of a worker fell below the accuracy threshold, the contributor would be removed from the job and her answers would not be taken into account.

Further, for the whole CrowdFlower task the worker agreement per Fleiss’ kappa [23] was estimated. This allowed us to measure the degree to which the observed amount of agreement among workers exceeds what would be expected if all workers made their choices completely randomly. General reliability of agreement between CrowdFlower contributors was achieved with a 0.75 Fleiss’ kappa score. Thus, we can assume that within our task relatively high agreement was achieved.

After gathering all answers for all images, the majority answer was used to label the image.

Labor Market. In order to test whether images from the profession articles reflect labor market statistics, professions were divided into two groups: professions with female (> 50% women) and male majority according to the labor market statistics. Then, the statistical significance of the difference between distributions of image groups were tested using chi-square independence tests with Monte Carlo p-value simulations [22].¹¹ We also explored how distribution of image categories would look like if one restricts professions to those with more than 70% men or women, respectively.

Next, the strength of relation was examined between the number of images depicting a particular gender in the article and the labor market statistics of profession. The Spearman’s rank correlation was utilized.

We also tested whether distributions of image categories are significantly different for articles with male, female and neutral titles. The images were grouped according to the gender of article titles and chi-square tests were applied.

¹⁰<https://www.crowdfunder.com/>

¹¹The simulations were used since some image categories have small numbers, possibly making p-values unreliable.

Analogously, we analyzed whether distributions of image categories are significantly different for professions which were assigned to different redirection bias groups.

3.3 Mentioned people analysis

Articles that did not mention any persons per our name dataset were excluded from the analysis, leaving 411 articles. Then, for each article, the proportion of mentioned men was calculated.

First, the dependency between the proportion of mentioned men and the gender of the article title was tested, by examining if articles with male title have a higher proportion of mentioned men than articles with female titles. To this end, three rank-sum tests (i.e., one for each pair of article groups) with p-value correction [19] were performed.

Second, we tested the dependency between the proportion of mentioned men and redirection bias group of profession. Thus, analogously to the previous analysis, the professions were grouped and then rank-sum tests were applied between proportions of mentioned men in each of the groups.

Labor Market. Third, we examined whether the proportions of mentioned men in the profession articles reflect labor market statistics. The professions were divided in two groups: professions with more than 50% men and more than 50% women. If proportions of mentioned men in the profession articles reflect the labor market, one would observe a high proportion of mentioned men in the group of professions with male majority and low proportion of mentioned men in the group of professions with female majority. Next we analyzed the strength of the relation between the percentage of mentioned men/women in an article and the percentage/number of employed men/women in a profession. To this end, the Spearman's rank correlation was utilized.

4 RESULTS

Our results on gender inequality in Wikipedia's profession presentations are outlined for the three distinct dimensions.

4.1 Redirection analysis

Most of the pages (885) about professions we found on Wikipedia have male titles (831), compared to much fewer female titles (25). A few articles about professions have neutral profession names (36). Hence, at first glance, the Wikipedia community is more male profession-oriented. For corresponding male-female profession title pairs, we encounter eight "neutral" professions (i.e., 16 articles) that have corresponding articles for both gender versions. Articles with a neutral profession title were also assigned to the neutral group.

Among the 820 redirects, 5 are from the male to the female label of a profession, 214 redirect from female to male labels, and 3 redirect from neutral to male labels; other redirects are either to broader fields of professions or titles that are synonyms. For example, going to the Wikipedia article "Sekretärin" (EN: female secretary), will automatically redirect to the article "Sekretär" (EN: male secretary), and thus, one never reaches a Wikipedia page "Sekretärin".

As a bottom line, the combination of redirects and existing pages of corresponding male-female profession names reveals 812 male

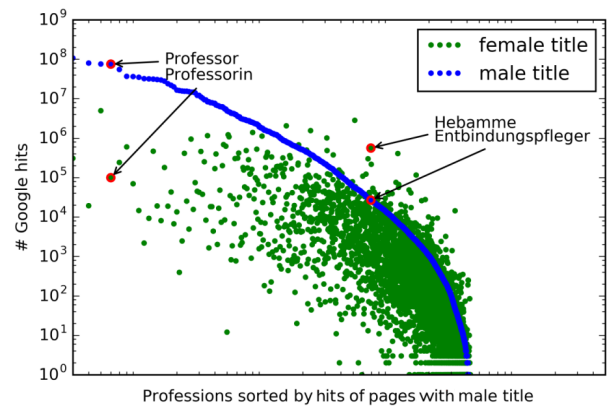


Figure 2: Distribution of Google hits for profession names, sorted according to amount of hits (log) for male profession names. One can see that most female titles have fewer Google results than corresponding male titles. For instance, "Professor" has more Google hits than the female "Professorin". The female profession names are more popular than the male ones for only a few professions, e.g., "Hebamme" (EN: midwife) vs. "Entbindungspfleger" (EN: male midwife).

bias cases, 6 cases of female bias, and 55 neutral cases. I.e., 812 professions have only an article with male title or a female title lemma redirect to an article with the corresponding male title, 6 professions have only an article with female title or a male title lemma redirect to the corresponding female title, 55 professions have either an article with neutral title or articles with both male and female titles. Thus, we observe evidence for gender disparity among article titles chosen by the Wikipedia editors.

Google hits. Figure 2 shows the distribution of Google hits for the profession names in our dataset. One can see that male profession names tend to be more popular than female names ($z = 28.6, p^{***}$) on Google. This indicates that the German speaking Web features more sources for male than female profession names, potentially reflected on Wikipedia.

Next we explored the normalized difference between Google hits of male and female profession titles for three groups separately: all professions with male, female, and neutral redirection bias identified previously. Our results indicate that a logistic regression model (Table 1) can indeed predict if a profession has a male or female redirection bias, based on the normalized difference of Google hits. The coefficients reveal that with a one-unit increase in the normalized difference of Google hits, odds of a profession having a male bias on Wikipedia increases by a factor of 11.48, whereas the probability of having a female bias decreases. In other words, it is more likely that professions have only a Wikipedia article with a female profession title (female bias group) when the female profession name has a greater number of Google hits than the male profession name. This indicates that the male skew of the German speaking Web is indeed mirrored on Wikipedia.

Labor Market. Next we investigated if the gender over- or under-representation in profession titles on Wikipedia can be explained by labor market statistics. We hypothesized that professions which are dominated by women nowadays will be found in the

Table 1: Logistic regression for Google hits. (Results of the best fitted logistic regression models) Model 1 stands for a logit model with binary outcome of a profession being in the female bias group or not. We see an 11.48 factor increase in the odds of being in a female bias group for a one-unit increase in the “Normalized Google difference” (NGD) score, since $\exp(2.44) = 11.48$. Model 2 is the logit model with the binary outcome of a profession being in the male bias group. The coefficient for NGD reveals a 0.0026 factor increase (or $1/0.0026 = 384$ factor decrease) in the odds of being in a male bias group for a one-unit increase in the NGD score since $\exp(-5.93) = 0.002654$. For example, both regression models will predict a female bias when the value of normalized Google difference equals -1 and male bias when the value equals 1.

	coef.	p	95% conf.int.
Model 1		Accuracy: 0.971	Pseudo R-squared:0.21
Normalized Google difference	2.44	0.0000	[1.68, 3.20]
Google hits for male name	0.00	0.9949	[0.00, 0.00]
(Intercept)	2.41	0.0000	[1.92, 2.90]
Model 2		Accuracy: 0.995	Pseudo R-squared:0.62
Normalized Google difference	-5.93	0.006	[-10.13, -1.73]
Google hits for male name	-2.07e-5	0.558	[-9.02e-3, 4.87e-5]
(Intercept)	-5.55	0.001	[-8.87, -2.24]

“female bias” group, whereas professions dominated by men will be found in the “male bias” group we created. To this end, the dependence between the percentage of women involved in a profession and the redirection bias was analyzed. We observed statistically significant differences in terms of percentage of women involved in a profession between: (1) the male and female bias groups ($z = -3.32, p^{**}$), and (2) neutral and female bias groups ($z = -3.35, p^{**}$). In other words, professions which are represented by only articles with female titles (i.e., female bias) show a significantly higher percentage of employed women than other professions, i.e., between 82 and 100 percent of women, whereas professions that are represented by only articles with male title tend to have from 10 to 60 percent of women. There was no significant difference between the underlying distributions of the percentage of women involved in professions which are in the male bias and neutral groups.

Nevertheless, one can glean from Figure 3 that some professions with 80-100 percent of women are in the male bias group. For example, in the profession “Gesundheits- und KrankenpflegerIn” (EN: health and nursing staff) one encounters about 85% women and the Wikipedia page with the female title “Gesundheits- und Krankenpflegerin” automatically redirects to the respective male form “Gesundheits- und Krankenpfleger”, hence the profession is in the male bias group. Another example is profession “Sekretär/in” (EN: secretary) whose workforce is constituted of about 88% women in Germany and only features the male article “Sekretär”. Thus, we cannot claim that the Wikipedia community, e.g., decided to use only female profession names as article titles for professions with a majority of women in the labor market. We can only conclude that existing professions with only a female profession article are more likely to have higher percentage of women in the labor market than professions with only a male profession article.

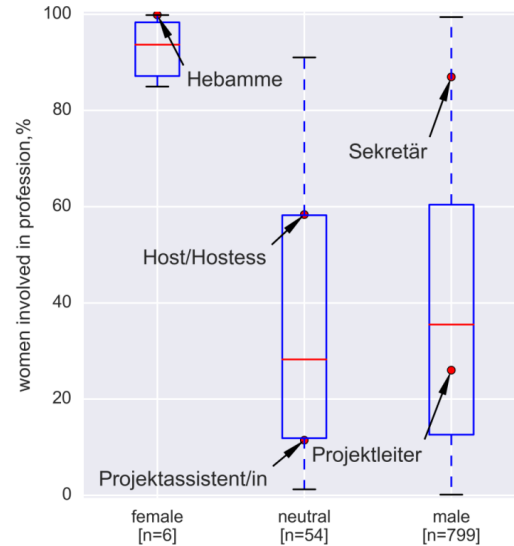


Figure 3: Percentage of women involved in professions (German labor market). Data is grouped by redirection bias of professions. The statistics from the German labor market were associated with 859 professions (799 male bias, 54 neutral, and 6 female bias). For some generic labels (e.g., DE: “Leiter”, EN: “head”, “leader”), no labor market statistics exist. One can see that professions with a female redirection bias on Wikipedia are indeed dominated by women in the German labor market, while there is no clear relation with the labor market statistics for the other two groups.

Second, the logistic regression model was fitted in order to describe relations between the redirection bias groups using the percentage of women involved. Because professions from the neutral and male bias groups do not show significant differences, we fit a logistic regression which will predict whether a profession is in the female bias group or not – i.e., the neutral and male bias groups were processed together. The logistic regression coefficients (Table 2) reveal that a one-unit increase in the percentage of employed women will result in a 44% increase in the odds of being in the female bias group (versus being in the male bias or neutral groups).

Table 2: Logistic regression for the labor market. (Results of the best fitted logistic regression model). The binary dependent variable represents a profession being in the female bias group. The coefficient for “Percentage of involved women” reveals that we will see a 44% increase in the odds of being in a female bias group for a one-unit increase in percentage of involved women, since $\exp(0.364) = 1.44$. For example, the probability of a profession being in the female bias group for a profession with 20% women in Germany equals $5.3e - 13$ and 0.55 if a profession has 98% women.

	coef.	p	95% conf.int.
		Accuracy: 0.99	Pseudo R-squared:0.61
Percentage of involved women	0.36	0.008	[0.10, 0.63]
(Intercept)	-35.53	0.005	[-60.44, -10.62]

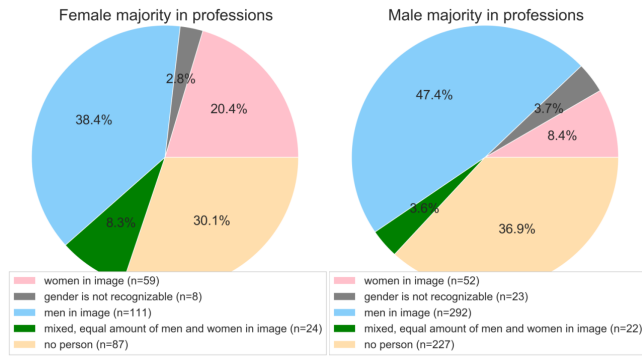


Figure 4: Distribution of image categories, grouped by gender of majority in the labor market. Professions with a female employment majority have almost two times more images depicting men than women, whereas professions with male majority feature even fewer images depicting women, around 5.6 times less.

The output indicates that the percentage of employed women is significantly associated with the probability of being in the female bias group, yet this relation shows itself only at very high values of female employment ratios. For example, we have an estimated probability of 0.001 for being in the female bias group for a profession with 80% employed women. The estimated probability is instead 0.63 for a profession with 99% employed women. That means that only professions with very high (> 97%) percentages of employed women will be likely to have only an article with a female title.

Results of rank-sum tests and fitted logistic regression model support the hypothesis that professions with higher percentages of women have only a female page (i.e., professions from female bias redirection group). At the same time, the threshold of percentage of employed women, where one observes increase in likelihood of profession to be in the female bias, is very high (about 97% employed women). However, we cannot identify a significant difference between the male and neutral bias groups. The relation between these groups cannot be described using the percentage of employed women in the labor market.

4.2 Images analysis

The aggregated results for all 906 labeled images were as follows: (1) almost half (44.8%) of the images show men, whereas only 12.4% of images show women; (2) an equal amount of men and women is observed only in a few images (5.1%); (3) almost a third (34.3%) of the images depicts no person; (4) gender is not recognizable for 3.4% of the images.

The analysis of images, grouped by gender of article title (profession name) where the images were encountered reveals a significant difference in image composition. Articles with a female title have almost 10 times more images depicting women than men, whereas articles with a male title have 4 times more images depicting men than women. Articles with a neutral title have 6 times more images depicting men than women. Thus, articles with female and male titles both show a respective gender inclination. While men are vastly underrepresented in articles with female titles, the underrepresentation of women in articles with male and neutral titles

Table 3: Spearman’s rank correlation coefficients between the number of images in Wikipedia article about a profession and the labor market statistics of the profession.

Images	Labor market	Corr.
percentage of img depicting men	percentage of women	-0.3***
percentage of img depicting women	percentage of women	0.34***
percentage of img depicting women	number of women	0.17**
percentage of img depicting men	number of men	0.013

exists, but is much weaker. We observed the same tendencies when grouping images by redirection bias of profession. Professions with only an article with a female title on Wikipedia have most images depicting women (75% of images). Nevertheless, professions from “male bias” and “neutral bias” redirection groups are not gender neutral in terms of image representations, as the majority of images depicts men (44.4% and 52% of all images, correspondingly).

Labor Market. Professions were divided into two groups: professions with female majority and such with male majority (>50%). Figure 4 shows the distribution of image categories, grouped by gender of majority in the labor market. One can see that there are almost two times more images depicting men (38.4%) than women (20.4%) in professions with more than 50% women in labor market (female majority group). At the same time, there are almost 6 times more images depicting men (47.4%) than women (8.4%) in professions with male majority. In both groups we observe that the majority of images depict men. Therefore, we can conclude that profession images on Wikipedia do not reflect the labor market statistics. We also looked at groups with more than 70% men and women in the labor market.

An analysis of the images, grouped by gender of majority (>50%) according to the labor market statistics, reveals a significant difference in image composition. Professions with more than 50% women in the labor market have 2 times less images depicting women than men, whereas professions with more than 50% men in the labor market have almost 6 times less images depicting women than men. In other words, a majority of one gender in the labor market does not imply a majority of the same gender in images.

Results of Spearman’s rank correlation (Table 3) reveal a moderate positive correlation ($\rho = 0.34$) between the percentage of images depicting women in an article and the percentage of women in the corresponding profession, and a moderate positive correlation ($\rho = 0.3$) between the percentage of images depicting men in an article and the percentage of men in the profession. Thus, professions with higher percentage of images depicting women/men exhibit a moderately higher percentage of women/men in the labor market.

4.3 Analysis of mentioned names

There are 411 articles about professions which mention at least one person. The articles mention overall 5085 persons (4272 men and 813 women). 10.4 men and 1.9 women were mentioned on average in the articles. The mean proportion of mentioned men per article is 0.83 and the median proportion is 0.98. In other words, out of all people mentioned in an article, 83% of them are men and only 17% of them are women on average.

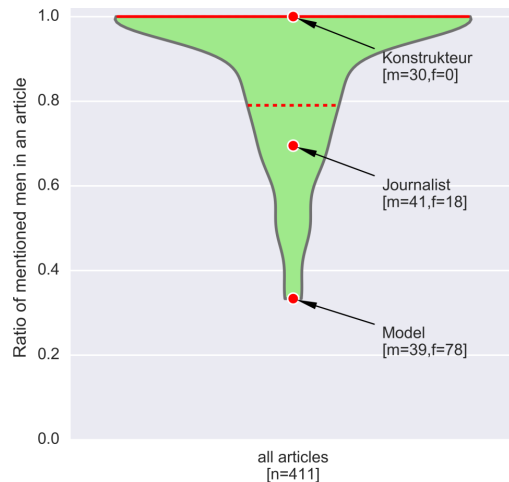


Figure 5: Mentioned men proportions. Violin plot of the distribution of proportional mentions of male names in articles. The violin plot features a kernel density estimation of the underlying distribution. The shape of the violin plot represents the density: the wider the violin plot, the greater the number of articles with a particular proportion of men. The solid red line represents the median; the dotted line shows the 3rd quartile (1st quartile line hidden). The violin plot represents data points which are distributed within two standard deviations of the mean, outliers are removed. One can see that $\frac{1}{2}$ of articles mention only men in the articles, $\frac{3}{4}$ of articles mention 80-100% men and only 0-20% women. Moreover, only a few articles mention either an equal number of men and women or more women than men.

Figure 5 shows the distribution of men proportions. As one can see, 50% of articles have a proportion of mentioned men close to 1.0 and 75% of articles have a proportion of mentioned men from 0.8 to 1.0. The article “Konstrukteur” (EN≈ “design engineer”) for example mentions 30 persons, all of them men, the article “Journalist” mentions 41 men and 18 women, and the article “Model” 39 men and 78 women. Articles which mention an equal number of men and women are considered gender neutral. Consequently, all articles with $> 50\%$ male names are male-biased with the remainder being female-biased in our consideration. One can see from Figure 5 that most articles have a proportion higher than 0.5. In particular, 92.5% of the articles are male-biased, 3.1% of the articles are female-biased, and 4.4% of the articles are gender-equal.

Lastly, we grouped articles according to the gender of the article title and found that the distribution of proportions of mentioned men was significantly different (p^*) between articles with male and female titles. Proportions of mentioned men in articles with male titles (median is 0.98) are more likely to be larger than the proportions in articles with female title (median is 0.65).

Labor Market. The difference in proportions of mentioned men was explored between professions with male and female majorities.

Three quarters of articles among professions with female majority in the labor market mention 68%-100% men and 0-32% women. This implies that the largest part of the articles for professions with female majority expose male bias; this is despite the fact that

Table 4: Spearman’s rank correlation coefficients between the number of mentioned men/women in a Wikipedia article about a profession and the labor market statistics of the profession.

Mentioned people	Labor market	Corr.
percentage of mentioned women	percentage of women	0.27***
percentage of mentioned men	percentage of men	0.27***
percentage of mentioned women	number of women	0.17***

compared to professions with male majority, female majority articles still feature a significantly lower proportion of mentioned men (medians are 0.88 and 0.98, p^{***} for male majority professions).

Results of Spearman’s rank correlation (Table 4) reveal weak positive correlation ($\rho = 0.27$) between the percentage of mentioned women or men and the percentage of women or men in the labor market correspondingly. In other words, the higher is the percentage of women in a profession, the higher is the percentage of mentioned women in the article about the profession, and the other way around. In summary, one can conclude that for some professions the percentage of mentioned men or women reflects the labor market statistics in terms of gender proportions.

5 RELATED WORK

Due to the increasing importance of online media, much research is concerned with an assessment of bias in Wikipedia and on the Web in general. For example, Reagle and Rhue [24] compared biographical articles from the English Wikipedia edition and the online Encyclopedia Britannica with respect to coverage, gender representation, and article length. Authors concluded that Wikipedia provided better coverage and longer articles. While Wikipedia has more articles on women than Britannica in absolute terms, Wikipedia articles on women are missing more often than are articles on men, when compared to Britannica. Wagner et al.[28] studied coverage of famous women in Wikipedia articles and the way women are portrayed in the online encyclopedia. The authors found that, despite good coverage of famous women in many Wikipedia language editions, the ways in which women and men are portrayed differ significantly. For example, romantic relationships and family-related issues are much more frequently discussed about women than men. In [11, 28, 29] differences between descriptions of male and female biographies in terms of network structure, topical focus, structural properties and language were researched. Thus, several studies suggest that gender bias on Wikipedia can be assessed using articles with biographies.

Several studies [7, 18] have tried to explain gender imbalance by studying how conflict-related behaviors (e.g., reverts) affect male and female editors. Collier and Bear [7] studied reasons for why female contributors stop contributing, indicating that the gender contribution gap is due to responses to conflicts. In [13, 15] authors found that significant Internet experiences and beliefs about one’s competence explain a large share of the gender gap. According to [14] only 16.1% of editors on Wikipedia are women. Antin et al.[2] studied differences between men and women’s editing activity in terms of the number and size of the revisions they make.

Bernacchi [27] performed a visual exploration of the gender issue on Wikipedia by studying the articles “Man” and “Woman” in several language editions of Wikipedia. The researchers compared sizes of articles, numbers of edits per article, changes in TOC structures, network structures of related articles and intersection of used concepts, topics coverage, revisions in terms of vandalism and deleted content. However, profession-related gender bias on Wikipedia and corresponding aspects remain uncovered, to the best of our knowledge.

Kay et al. [17] studied the gender bias in image search results and found that manipulated search results have a small significant effect on people’s gender ratio perception (cf. Section 1).

6 DISCUSSION

In general, we observe a strong overrepresentation of male titles, images and mentioned names on the German Wikipedia that cannot be explained in its entirety by labor market statistics.

The analysis of article titles and redirections reveals that most professions are represented only via an article with a male form of the profession. Moreover, most encountered redirections are from female to male titles. This evidence supports the existence of gender disparity along article titles in the professional domain. Choices of article titles mostly reflect the general popularity of male over female profession names on the German speaking Web, with some exceptions. Turning to the labor market statistics, we observe a relation between the percentage of women involved in a profession and the probability of having only a female article on Wikipedia. However, only professions with a distinctly larger female ratio in the workforce are likely to be represented only by a female article. On the other hand, using labor market statistics, we cannot well distinguish between professions with both male and female lemmata versus professions represented only by a male title, as the male redirection/representation bias on Wikipedia does not only appear for “male” professions as measured by true employment.

Regarding depictions, almost four times more images of men than women are used to describe the professions. We find that almost half of the images from the profession articles depict men and only around 12% show women. We observe a significantly higher number of images depicting men in professions, regardless of female majority or male majority in the labor market for that profession. Only for articles under female lemmas and professions which have only a female article on Wikipedia do we find a significantly higher number of images depicting women than men.

Along the mentioned people dimension, an imbalance towards male names (more than 50% men) was observed in 92.5% of the articles. Only very few articles mention exhibit either (a near) equal number of men and women or more women than men. Particularly, out of all mentioned people, we observed $\frac{3}{4}$ of articles with 79-100% men and 0-21% women. Less than 50% men (i.e., female bias) were mentioned in 3.1% of the articles and equal representation was observed in 4.4%. Analyses of the articles with female title and articles of professions with female majority in the labor market revealed significantly lower proportions of mentioned men. Nevertheless, even for groups where the difference of proportions was significant, a male bias was observed in the larger part of articles from these profession groups.

Interpreting these results, one major reason for why male lemmas and redirects dominate in (the profession articles of) the German Wikipedia is certainly the existence of the “Generisches Maskulinum”, i.e., the traditional German language tendency to refer to a group of mixed or unknown gender with the male descriptor. We have also found several discussions on the talk pages of the respective articles studied here that point to a deliberate decision to “simplify” the navigation structure of Wikipedia by agreeing on the guideline to put all profession names under their generic masculine name. This “bias” as we have called it hence doesn’t automatically indicate a deliberate prejudiced ideology against female descriptors. Yet, the effect on a reader is the same: she is being redirected or cannot find results for female profession name forms, which certainly influences her perception of how society views certain professions; also reinforcing the use of the generic masculine form that has been criticized in the German-speaking public discourse. A simple solution would be to host all articles under a gender-neutral lemma (as is practice already for some profession as we have seen), reachable via redirect from the male and female lemmas - the description of the profession throughout the article then has to be simply adapted to a neutral/both-gender form (e.g., with the “/In” suffix).

Regarding images and names, another reason for male overrepresentation is presumably that, especially in long-established professions like “smith”, many older depictions and records exist that contain for the most part men – and as prominence often is achieved post-mortem, names of famous representatives of a trade are more likely to be mentioned by name in reference works, reflecting a historic imbalance towards working men. Nonetheless, the biases we unveiled also appear in some more modern professions. And, if we again consider solely the effect on the reader, in such cases a more recent focus on the occupational field might be useful, for example by concentrating the lede and starting sections on the contemporary state of a profession before covering its history. To control for history effects, future work might filter out entity-matched persons by their date of birth, and specifically identify paragraphs that refer to historical aspects of a profession.

Regarding possible reasons for the observed imbalances, previous research [7, 14] has revealed that a dominant majority of editors in Wikipedia is male. Wikipedia further contains fewer and less extensive articles about women or topics important to women [18], indicating that male editors might not be suited to entirely represent the female world view and interests. Over the last years, the Wikimedia Foundation has made many attempts to attract more female editors [6, 21], although with mixed success. Yet, the gender of editors is likely not the only factor behind the described imbalances. Just as many historical reference works already contain a male-dominant description, Wikipedia might reflect and in some cases inherit biases from other media and the references contained therein. We observed evidence for a male bias on the German speaking Web, i.e., for most of professions one can find much more sources for male than female professions. Moreover, several studies [10, 25] reveal gender bias and stereotypes in mass media. I.e., apart from gender equality of the editor base, having transparent guidelines and rules towards representative equality in article content might prove just as useful to decrease gender disparities of individual articles; defining clear target audience groups

and contemplating suitability of the content for those readers might likewise be helpful.

The method proposed in our research can in principle be applied for the analysis of gender inequalities in different Wikipedia editions. The image and mentioned people analysis methods can be applied to any language; the image analysis could in future work even be fully automated given the availability of reliable algorithmic tools. And while the redirection analysis can only be applied to languages with masculine and feminine grammatical genders, this leaves several candidate languages. A cross-language analysis extension is a promising goal for future work, by matching the respective profession pages, in order to compare biases over editions. Results could be compared to external country-wise data like the Global Gender Gap Index for each country associated with a Wikipedia edition.

7 CONCLUSION

This study presents a computational approach to gender bias assessment along three distinct dimensions using Wikipedia profession articles. We made use of crowdsourcing to complement our analysis with high-quality data and compared against publicly available labor market data.

The results indicate clear gender presentation imbalances on all the three dimensions studied (titles, mentioned persons, images). A notable portion cannot be explained simply by underlying labor market conditions, and many choices of representation seem to be made out of tradition or based on historic reference material.

This work has the potential to provide aides and inform guidelines for the Wikipedia community to identify and address gender disparities. Likewise, the outlined approach could be implemented in a software tool supporting Wikipedia editors in writing articles. The analysis procedure is likely reproducible for other Wikipedia language editions. Lastly, a careful consideration and possible evaluation of the effects of gender imbalance in profession articles especially on younger readers seems worthwhile, given our results.

The data and code are available online on GitHub.¹²

REFERENCES

- [1] Rami Al-Rfou, Vivek Kulkarni, Bryan Perozzi, and Steven Skiena. 2015. PolyglotNER: Massive Multilingual Named Entity Recognition. In *Proceedings of the 2015 SIAM International Conference on Data Mining*. SIAM, 586–594.
- [2] Judd Antin, Raymond Yee, Coye Cheshire, and Oded Nov. 2011. Gender Differences in Wikipedia Editing. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*. ACM, 11–14. <https://doi.org/10.1145/2038558.2038561>
- [3] Florian Arendt and Temple Northup. 2015. Effects of Long-Term Exposure to News Stereotypes on Implicit and Explicit Attitudes. *International Journal of Communication* 9, 0 (2015), 21. <http://ijoc.org/index.php/ijoc/article/view/2691>
- [4] Sherryll Browne Graves. 1999. Television and Prejudice Reduction: When Does Television as a Vicarious Experience Make a Difference? *Journal of Social Issues* 55, 4 (1999), 707–727. <https://doi.org/10.1111/0022-4537.00143>
- [5] David Chambers. 1983. Stereotypic images of the scientist: The draw-a-scientist test. *Science Education* 67, 2 (1983), 255–265. <https://doi.org/10.1002/sce.3730670213>
- [6] Giovanni Luca Ciampaglia and Dario Taraborelli. 2015. MoodBar: Increasing New User Retention in Wikipedia Through Lightweight Socialization. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, 734–742. <https://doi.org/10.1145/2675133.2675181>
- [7] Benjamin Collier and Julia Bear. 2012. Conflict, Criticism, or Confidence: An Empirical Examination of the Gender Gap in Wikipedia Contributions. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, 383–392. <https://doi.org/10.1145/2145204.2145265>
- [8] Shelley J. Correll. 2001. Gender and the Career Choice Process: The Role of Biased Self-Assessments. *Amer. J. Sociology* 106, 6 (May 2001), 1691–1730. <https://doi.org/10.1086/321299>
- [9] Olive Jean Dunn. 1959. Estimation of the Medians for Dependent Variables. *The Annals of Mathematical Statistics* 30, 1 (1959), 192–197.
- [10] Liye Fu, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Tie-breaker: Using language models to quantify gender bias in sports journalism. (July 2016). <http://arxiv.org/abs/1607.03895> arXiv: 1607.03895.
- [11] Eduardo Graells-Garrido, Mounia Lalmas, and Filippo Menczer. 2015. First Women, Second Sex: Gender Bias in Wikipedia. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media (HT '15)*. ACM, 165–174. <https://doi.org/10.1145/2700171.2791036>
- [12] A. G. Greenwald and M. R. Banaji. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review* 102, 1 (1995), 4–27.
- [13] Eszter Hargittai and Aaron Shaw. 2015. Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia. *Information, Communication & Society* 18, 4 (2015), 424–442. <https://doi.org/10.1080/1369118X.2014.957711>
- [14] Benjamin Mako Hill and Aaron Shaw. 2013. The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation. *PLOS ONE* 8, 6 (2013), e65782. <https://doi.org/10.1371/journal.pone.0065782>
- [15] Marit Hinnoosaar. 2015. Gender Inequality in New Media: Evidence from Wikipedia. (May 2015). <https://doi.org/10.2139/ssrn.2617021> Available at SSRN: <https://ssrn.com/abstract=2617021>.
- [16] Fariba Karimi, Claudia Wagner, Florian Lemmerich, Mohsen Jadidi, and Markus Strohmaier. 2016. Inferring Gender from Names on the Web: A Comparative Evaluation of Gender Detection Methods. In *Proceedings of the 25th International Conference Companion on World Wide Web*. ACM, 53–54. <https://doi.org/10.1145/2872518.2889385>
- [17] Matthew Kay, Cynthia Matuszek, and Sean A. Munson. 2015. Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, 3819–3828. <https://doi.org/10.1145/2702123.2702520>
- [18] Shyong (Tony) K. Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R. Musicant, Loren Terveen, and John Riedl. 2011. WP:Clubhouse?: An Exploration of Wikipedia's Gender Imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*. ACM, 1–10. <https://doi.org/10.1145/2038558.2038560>
- [19] Paul L. MacDonald and Robert C. Gardner. 2000. Type I Error Rate Comparisons of Post Hoc Procedures for I x J Chi-Square Tables. *Educational and Psychological Measurement* 60, 5 (2000), 735–54.
- [20] Jörg Michael. 2007. 40000 Namen, Anredebestimmung anhand des Vornamens. *c't* 07, 17 (2007), 182–183. <http://www.heise.de/ct/ftp/07/17/182/>
- [21] Jonathan T. Morgan, Siko Bouterse, Heather Walls, and Sarah Stierch. 2013. Tea and Sympathy: Crafting Positive New User Experiences on Wikipedia. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, 839–848. <https://doi.org/10.1145/2441776.2441871>
- [22] B. V. North, D. Curtis, and P. C. Sham. 2002. A Note on the Calculation of Empirical P Values from Monte Carlo Procedures. *American Journal of Human Genetics* 71, 2 (2002), 439–441. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC379178/>
- [23] D.M.W. Powers. 2011. Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies* 2, 1 (2011), 37–63. <http://en.journals.sid.ir/ViewPaper.aspx?ID=436993>
- [24] Joseph Reagle and Lauren Rhue. 2011. Gender Bias in Wikipedia and Britannica. *International Journal of Communication* 5, 0 (2011), 21. <http://ijoc.org/index.php/ijoc/article/view/777>
- [25] Stacy L. Smith, Marc Choueiti, and Stephanie Gall. 2010. *Gender inequality in popular films: Examining on screen portrayals and behind-the-scenes employment patterns in motion pictures released between 2007–2009*. Los Angeles, CA: Annenberg School for Communication & Journalism.
- [26] U.S. Census Bureau. 1983. 1980, Census of the population. Detailed Occupation and Years of School Completed by Age, for the Civilian Labor Force; by Sex, Race, and Spanish Origin. U.S. Department of Commerce. (1983).
- [27] Viola Bernacchi. 2015. *Gender imbalance and Wikipedia*. Ph.D. Dissertation. Polytechnic University of Milan, Italy.
- [28] Claudia Wagner, David Garcia, Mohsen Jadidi, and Markus Strohmaier. 2015. It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia. In *Ninth International AAAI Conference on Web and Social Media*. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10585>
- [29] Claudia Wagner, Eduardo Graells-Garrido, David Garcia, and Filippo Menczer. 2016. Women through the glass ceiling: gender asymmetries in Wikipedia. *EPJ Data Science* 5, 1 (2016). <https://doi.org/10.1140/epjds/s13688-016-0066-4>

¹²<https://github.com/gesiscss/Wikipedia-Language-Olga-master/>