

# WP:Clubhouse? An Exploration of Wikipedia's Gender Imbalance

Shyong (Tony) K. Lam<sup>1</sup>, Anuradha Uduwage<sup>1</sup>, Zhenhua Dong<sup>2</sup>, Shilad Sen<sup>3</sup>, David R. Musicant<sup>4</sup>, Loren Terveen<sup>1</sup>, John Riedl<sup>1</sup>

<sup>1</sup>GroupLens Research  
Dept. of Computer Science and Engineering  
University of Minnesota  
Minneapolis, Minnesota  
{lam,uduwage,terveen,riedl}@cs.umn.edu

<sup>2</sup>Dept. of Information Technical Science  
Nankai University  
Tianjin, China  
dongzh@mail.nankai.edu.cn

<sup>3</sup>Math, Statistics, and Computer Science Dept.  
Macalester College  
St. Paul, Minnesota  
ssen@macalester.edu

<sup>4</sup>Dept. of Computer Science  
Carleton College  
Northfield, Minnesota  
dmusican@carleton.edu

## ABSTRACT

Wikipedia has rapidly become an invaluable destination for millions of information-seeking users. However, media reports suggest an important challenge: only a small fraction of Wikipedia's legion of volunteer editors are female. In the current work, we present a scientific exploration of the gender imbalance in the English Wikipedia's population of editors. We look at the nature of the imbalance itself, its effects on the quality of the encyclopedia, and several conflict-related factors that may be contributing to the gender gap. Our findings confirm the presence of a large gender gap among editors and a corresponding gender-oriented disparity in the content of Wikipedia's articles. Further, we find evidence hinting at a culture that may be resistant to female participation.

## Categories and Subject Descriptors

H.3.4 [Information Systems]: Systems and Software—*Information networks*; H.5.3 [Information Systems]: Group and Organization Interfaces—*computer-supported collaborative work*

## General Terms

Human Factors, Measurement

## Keywords

Wikipedia, collaboration, gender gap, content coverage

## 1. INTRODUCTION

Over the past decade, Wikipedia has become a premier information resource on the web. Remarkably, it was not built by experts,

but by harnessing the collective effort of millions of volunteer editors. However, not all is well with Wikipedia. Researchers have identified and studied several factors that represent challenges for Wikipedia, including increased vandalism [21], increased overhead in resolving editor conflict and performing other coordination activities [13], and an overall stagnation in growth rate [24].

More recently, in a January 2011 New York Times article, Noam Cohen described another challenge: a wide gender gap amongst Wikipedia's editors [7]. Cohen observes that just 13% of Wikipedia's contributors are female, according to a 2009 Wikimedia Foundation survey. Furthermore, he suggests that this disparity has led to deficiencies in Wikipedia's coverage of "female" topics, as evidenced by a series of anecdotal examples (e.g., Wikipedia's coverage of topics like friendship bracelets or "Sex and the City" pales in comparison to that of toy soldiers or "The Sopranos").

The Wikimedia Foundation has established a goal of increasing the female share in editors to 25% by 2015. While ambitious, such an accomplishment is certainly not out of reach. In *Unlocking the Clubhouse: Women in Computing* [17], Margolis and Fisher describe a series of studies and educational reforms that helped Carnegie Mellon University address a wide gender gap in their undergraduate Computer Science program. Over the course of five years, female enrollment rose from 7% in 1995 to 42% in 2000.

In the current work, we present a quantitative exploration of the gender imbalance in English Wikipedia's volunteer editor population. Cohen's article presents a compelling argument, but we believe there is need for more rigorous analysis that expands on the reported survey results and anecdotal evidence. We believe our work represents a crucial next step in understanding the nature of the gender gap and deciding what should be done to address it.

### 1.1 Related Work

Research from the volunteering literature and technology adoption literature offer reasons not to expect a large gender gap among Wikipedia's editors. Taniguchi finds that females are more likely to volunteer than males, and that females do more volunteer work than males [26]. In [31], Wilson cites four underlying reasons for females' increased volunteerism: they exhibit greater empathy and altruism, they place more value in helping others, they perceive a gender-specific norm that they should take care of others, and they

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WikiSym'11, October 3–5, 2011, Mountain View, California  
Copyright 2011 ACM 978-1-4503-0909-7/11/10 ...\$10.00.

view volunteering as part of their “social life.” Overall, these findings suggest that females may be more likely to volunteer their time to edit Wikipedia, though they may edit less if they lack a social connection to the Wikipedia community.

The technology adoption literature suggests that females may lag behind males in adopting new technology. Broadly, Venkatesh et al. find that females are less likely to adopt new technologies than males, and that females are more heavily influenced by social norms related to a technology and the perceived difficulty of the new technology [29]. However, studies of gender differences in adoption of the Internet and social media offer more encouraging findings. Periodic Pew Research Center surveys of the general populace show that Internet usage between 2000 to 2004 was skewed toward males, but that the gap has since dissolved [19]. Females are now *more* likely than males to participate in some social media sites such as Facebook or MySpace [28]. In addition, females are more likely to tweet (10% of females, 7% of males), and teenage girls are more likely to blog (25% of girls, 15% of boys) [23, 15]. Even online gaming, which is traditionally seen as a male-dominated activity [34], shows signs of a sea change; market research surveys indicate that females and males are on par with each other in online social gaming [12, 25].

Together, these observations suggest that Wikipedia, a community of volunteers collaborating to build an online encyclopedia, ought to have a reasonable gender balance. However, multiple studies have indicated an apparent disparity. Lim’s surveys of college students find that while all respondents had used Wikipedia, females visited it less frequently and perceived it to be of lower quality than males did [16]. A 2011 Pew Research survey finds a small gender gap in readership (50% of female Internet users, and 56% of male ones) [35]. The Wikimedia Foundation commissioned a survey of Wikipedia users in 2009, and its results show a large gap among readers (75% male, 25% female), and an even larger gap among editors (87% male, 13% female) [10]. However, because users self-selected to participate in the survey, the report authors acknowledge that it is “hard to evaluate whether the shares we found in our survey are representative.” Furthermore, 75% of the users who took the survey were using a non-English Wikipedia. Thus, it is uncertain what the gender gap is in English Wikipedia.

## 1.2 Contributions

Our work seeks to explore more carefully the state of gender imbalance among the English Wikipedia’s volunteer editors. We extend existing research on Wikipedia and gender in three key ways. First, we conduct a high-level study of gender and editing behavior in order to measure and characterize the editor gender gap. Second, we explore how the imbalance affects Wikipedia as a way of showing why it is an important issue to address. Finally, we analyze how conflict-related behaviors such as reverts and blocks affect editors of each gender to help understand why an imbalance might exist.

## 2. RESEARCH QUESTIONS

We begin our exploration of Wikipedia’s gender gap by posing three overarching research questions and nine hypotheses.

### 2.1 RQ1: Gap-Overall

**What is the extent of Wikipedia’s gender gap, and how has it changed over time?**

We are interested in measuring Wikipedia’s gender gap and determining whether the imbalance is growing or shrinking. Based on the survey results presented in [10], we hypothesize that Wikipedia does have a wide gender gap. Note that the remainder of our hy-

potheses and research questions provisionally assume that this hypothesis will be supported.

**H1a *Gap-Exists*:** Wikipedia has a substantial editor gender gap.

Periodic surveys of the general populace indicate that there was a modest gender gap in Internet use in the early 2000s, but that it has been shrinking steadily [19]. We hypothesize that a similar trend has been taking place in Wikipedia.

**H1b *Gap-Shrinking*:** Wikipedia’s gender gap is shrinking.

## 2.2 RQ2: Gap-Matters

**How is Wikipedia affected by the gender gap?**

In this research question, we wish to explore whether the gender gap is causing some parts of Wikipedia to receive less attention than other parts. Our next hypotheses are inspired by Noam Cohen’s New York Times article [7], which anecdotally demonstrates a large disparity in the depth of coverage between “female” and “male” topics in Wikipedia. We systematically study this phenomenon in large-scale data-driven ways that do not depend on anecdotal evidence and invocation of gender stereotypes to determine which topics are “female” or “male.” Formally:

**H2a *Focus-Differences*:** Male and female editors focus on different content areas.

**H2b *F-Coverage-Worse*:** Coverage of topics with particular interest to females is inferior to topics with particular interest to males.

We further hypothesize that due to gender differences in extraversion, empathy, and altruism [9, 31], females will tend to be more active than males in social- or community-oriented areas of Wikipedia that offer increased interaction with other editors and opportunity to build interpersonal relationships. If this hypothesis is supported, addressing the gender gap might lead to a healthier community in which there are more resources available for community-oriented tasks like helping new editors and organizing editor efforts.

**H2c *F-Social*:** Females are more likely to be involved in social- and community-oriented areas of Wikipedia.

## 2.3 RQ3: Gender-Conflict

**What gender differences exist in conflicts in Wikipedia, and how do those differences relate to the gender gap?**

In our final research question, we look at gender differences in conflict among Wikipedia’s editors in order to learn about how conflict might be contributing to the gender gap. Prior research finds that conflict has been a growing problem for Wikipedia, consuming increasing amounts of editor effort [13]. Studies on gender and personality have shown that females tend to have more agreeable and less aggressive personalities [9, 2], which suggests that they may tend to avoid conflict if possible. Therefore, a possible explanation for the gender gap is that females may find conflict among Wikipedia editors to be distasteful and unappealing, and may simply choose to not edit Wikipedia as a result. As a partial test of this explanation, we hypothesize that Wikipedia’s existing female editors tend to do their work in less controversial areas.

**H3a *F-Uncontentious*:** Females tend to avoid controversial or contentious articles.

Females who do decide to edit Wikipedia may find it difficult to make contributions that are accepted by the community. Drawing upon years of research in gender and computer-mediated communication, Herring finds that “gender differences in on-line com-

munication tend to disfavor women” and that females who participate in mixed-gender online environments tend to be marginalized [11]. Furthermore, Herring notes that even if participant gender is not made salient, features of a participant’s discourse style can often reveal gender information. Recent research has shown that males and females have measurably different editing behaviors on Wikipedia [1]. Thus, we believe there may be a systemic bias against females that cause their edits to be more likely to be reverted (undone) by another editor, particularly early on in their Wikipedia tenure. Furthermore, since females may prefer to avoid conflict, we believe they are more likely than males to lose interest and leave Wikipedia if their early contributions are reverted.

**H3b F-Reverted-More:** Female editors are more likely to have their early edits reverted.

**H3c F-Reverted-Leave:** Female editors are more likely to stop editing and leave Wikipedia when being reverted as newcomers.

Another action used in Wikipedia conflict resolution is a block, which prevents the affected user from editing for a period of time. Blocks are much less common than reverts, and their use is limited to cases of disruptive or anti-social behavior where other forms of conflict resolution have been ineffective. Research in multiple domains indicates that males are more likely to violate rules [30, 11, 27] and that they are more aggressive, especially when provoked [2]. Thus, we hypothesize that females are less likely to be blocked than males, and that conflicts that escalate to the point where a block is required are *not* a contributor to the gender gap.

**H3d F-Blocked-Less:** Female editors are less likely to be blocked.

### 3. DATA

To test our hypotheses and answer our research questions, we perform a variety of quantitative statistical analyses on publicly-available English Wikipedia data. The majority of our data is either from the January 2011 data dump<sup>1</sup> or from the Wikipedia website itself (collected during February and March 2011 via the public API or screen-scraping). We also drew upon several other sources of data during our analyses, including data derived from older full-text Wikipedia dumps released in January 2010 or January 2008. Because each of these additional sources of data is specific to one of our hypotheses, we will defer discussion about these other data until the relevant section.

#### 3.1 Editor Gender Data

A key piece of information we need for our analyses is editor gender. In Wikipedia, there are several ways that an editor can publicly disclose his or her gender:

1. They can specify whether they are male or female in their account’s preference settings. The gender setting is described as being “used for gender-correct addressing by the software,” and is public information available from Wikipedia’s API.
2. They can place a gender userbox on their User page to openly announce and display their gender in a de facto standard way.
3. They can mention their gender while describing themselves on their User page or during a discussion with fellow editors in one of Wikipedia’s discussion areas.

The editor gender data used in our analysis includes users who disclosed their gender via preference setting (#1) and a subset of those who used a userbox (#2). Figures 1 and 2 shows screen shots of these two disclosure methods. For #1, we queried the Wikipedia

<sup>1</sup><http://dumps.wikimedia.org/enwiki/20110115/>

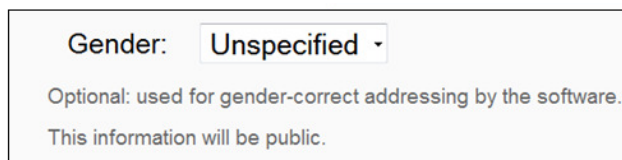


Figure 1: Gender preference setting in its default state.



Figure 2: Userboxes used by Wikipedia editors to display and announce their gender.

API to obtain the gender setting for all users with at least one edit, and for #2, we identified users who have either the *User:UBX/male* or *User:UBX/female* gender userbox displayed on their User page. There exist several other userboxes that can be used to denote one’s gender (including ones for transgender and other alternative genders), but since they have been used by too few users to give statistically reliable data, we chose to not include them in our analysis. We also chose to not collect gender disclosures that occurred only via method #3 because doing so would require substantially more advanced natural language processing techniques.

We collected data for 106,698 users that have disclosed their gender using the user preference setting and 8,630 users who have done so using a userbox. Of these, 1,478 users specified their gender using both methods. 1,476 of these users specified the same gender in both places. The remaining two users have a different gender specified using each method, and are excluded from our analyses. In total, our editor gender data set contains 113,848 users who have collectively made over 67 million edits.

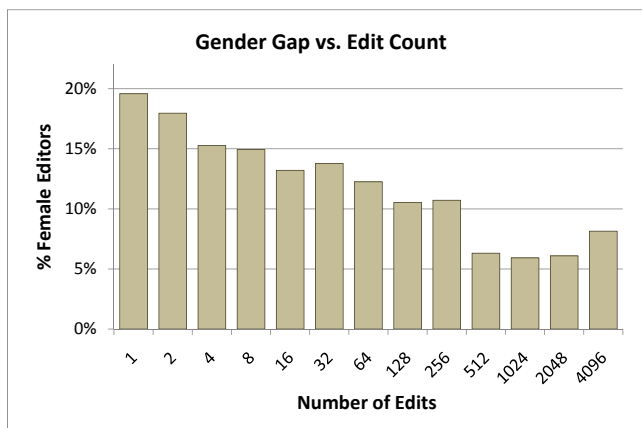
#### 3.2 Assumptions and Limitations

This gender data is comprised of self-reports from self-selected users, and thus has limitations. The validity and generalizability of our results are subject to several assumptions about this data. First, we assume that users are mostly honest in reporting their gender. Second, we assume that users who do not report their gender behave similarly to those who do report their gender (which is just 2.8% of editors). Finally, we assume that self-report rates are similar between males and females at similar stages of their Wikipedia life-cycle. In our data, we find that self-report rates increase dramatically for dedicated editors: we have gender information for 6.5% of editors with at least ten edits, 14.1% of those with at least 100 edits, and 34.7% of those with at least 1,000 edits.

While necessary to enable our analysis, these assumptions are difficult to confirm. We do note that at a high level, our results are comparable to those obtained from the Wikimedia Foundation’s recent survey of Wikipedia users [10], which provides limited support for our assumption about truthful self-reports. However, we are unable to provide evidence for the other assumptions. Unfortunately, this problem is fundamental for any Wikipedia research that depends on existing publicly-available data. Nearly all the results in this paper are subject to these assumptions and limitations.

### 4. RESULTS AND ANALYSIS

We now step through each of our research questions and hypotheses, describing our methods for testing each hypothesis, the data we used, and our results.



**Figure 3: Wikipedia’s gender gap as a function of editor activity for editors first editing Wikipedia during 2009. The gender gap is more pronounced when looking at high-activity editors.**

## 4.1 RQ1: Gap-Overall

**H1a Gap-Exists:** Wikipedia has a substantial editor gender gap.

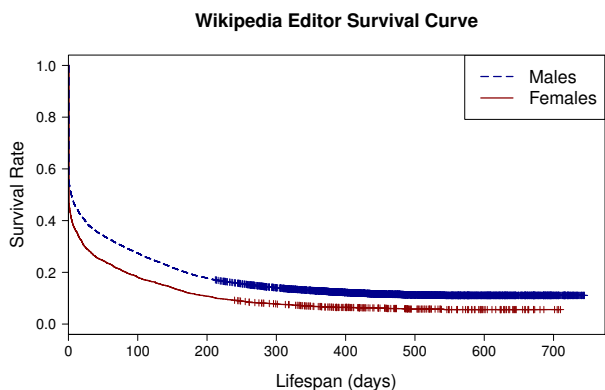
To characterize the gender gap at a high level, we compared male and female editors using three broad metrics: editor count, edit count, and activity lifespan. We found that females comprised 16.1% of the 38,497 editors who started editing Wikipedia during 2009 and who specified their gender. This is indicative of a substantial gender gap in Wikipedia editors – males outnumber females by over 5 to 1. However, this is not the end of the story. Our other two metrics suggest that the gender gap is even deeper than indicated by editor count.

We found that despite females being 16.1% of the new editors in 2009, they only accounted for 9.0% of edits made by this cohort of editors. On average, a male editor made almost twice as many edits as a female editor. Figure 3 depicts the gender gap at various levels of edit count. We see that the gender gap widens when looking at editors with many edits, and does not appear to stabilize until the percentage of female editors drops to around 6% for editors making more than about 500 edits. This observation points to the possibility that females leave Wikipedia earlier than males in their editing tenures. Our third metric, activity lifespan, directly examines this possibility.

Activity lifespan measures how long a user is active in a system before leaving. We used a definition of activity lifespan similar to Yang et al.’s in their work studying user survival in social question-answering systems [33]. We considered an editor’s “birth” to be his or her first edit date and a “death” to be a period of edit inactivity exceeding six months (the death is recorded as the beginning of the period). Using this notion of user birth and death, we applied standard survival analysis techniques [8]. Figure 4 shows the estimated male and female survival curves for users joining during 2009. We see a distinct difference between the two curves. Females stop editing Wikipedia sooner than males, and the ratio of males remaining to females remaining for this cohort increases steadily as time passes. Therefore, one of the factors contributing to Wikipedia’s gender gap is a lower retention rate for female editors compared to male editors. We will return to exploring this survival data in greater detail later in section 4.3.

**H1b Gap-Shrinking:** Wikipedia’s gender gap is shrinking.

To examine how the overall gender gap has evolved over time, we looked for a trend in the gender breakdown of new editors join-



**Figure 4: Activity lifespan of new female and male Wikipedia editors in 2009 expressed as Kaplan-Meier survival curves. The difference between the curves is statistically significant (log-rank test,  $p < 0.001$ ).**

ing Wikipedia each month. However, there are confounds in the data that made this a tricky task. The two methods for disclosing one’s gender that we consider in our data (userboxes and preference setting) were introduced at different times, so simply looking at the trend over all of Wikipedia’s existence is not a fair analysis. Users who started editing Wikipedia before a gender disclosure method was introduced could not specify their gender using that method until after its introduction date. The survival analysis presented above shows that males tend to have longer activity lifespans. Therefore, males who joined Wikipedia before a gender disclosure method was introduced are more likely than females to still be active once the method is made available (and thus, be able to specify a gender using it).

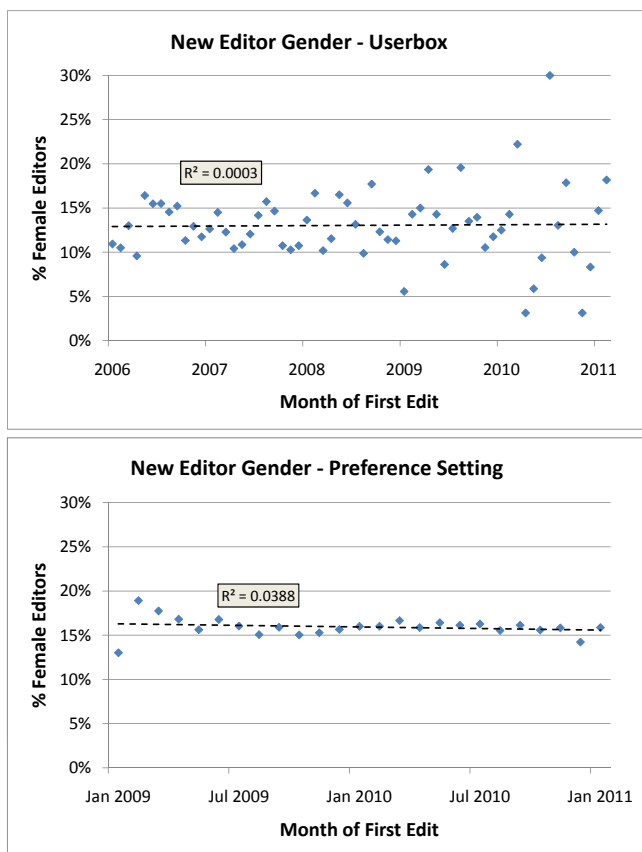
Due this confound, we could only make valid comparisons for users joining Wikipedia after the introduction of a gender disclosure method, and only for the subset of users who used that disclosure method. The gender userboxes were introduced in December 2005, and the gender preference setting was introduced in January 2009. Figure 5 shows two charts, each depicting the gender gap over time for one of the gender disclosure methods. The trends in both charts are flat with nearly zero slope. Therefore, Wikipedia’s gender gap appears to have remained approximately constant since December 2005, which is surprising given that other online gender gaps have been shrinking over time. Note that the two charts indicate different female editor percentages. This may be because the userbox gender disclosure method requires more Wikipedia-specific knowledge to use. Thus, there may be fewer female userbox users due to the survival differences shown earlier (that is, females might be more likely to stop editing Wikipedia before learning about the userboxes).

Our findings yield evidence to support **H1a Gap-Exists**, but not **H1b Gap-Shrinking**. Wikipedia suffers from a substantial gender gap, and the gender gap has *not* been closing over time. Next, we turn to an exploration of why the gender gap matters to Wikipedia.

## 4.2 RQ2: Gap-Matters

**H2a Focus-Differences:** Male and female editors focus on different content areas.

To test this hypothesis, we used data and methods described by Chen et al. in [6] to determine editors’ interest areas. Based on the January 2008 data dump, this method uses Wikipedia’s category



**Figure 5: Wikipedia’s gender gap in new editors as a function of time for users specifying their gender via userbox (top) and users specifying their gender via preference setting (bottom). Best-fit lines are plotted as black dotted lines, and have slopes of 0.0043% (top) and -0.029% (bottom) per month.**

structure to assign articles to eight primary interest areas: Arts, Geography, Health, History, Science, People, Philosophy and Religion. Each editor is then placed into zero or more of these eight areas based on his or her editing activity and how focused it is in specific areas’ articles.

Using Chen et al.’s data, we determined how many male and female editors are in each of the eight interest areas and compared each area’s gender distribution with the Wikipedia-wide one. Table 1 shows our results, which indicate that males and females are focused on disparate content areas within Wikipedia. There is a greater concentration of females in the People and Arts areas, while males focus more on Geography and Science. These findings are consistent with gender differences found in the National Science Foundation’s biennial survey of college graduates: men generally skew toward science and engineering fields, while women skew toward other fields including arts and humanities [18]. These results, when combined with the existence of an editor gender gap, are suggestive of a deficiency in some areas.

**H2b F-Coverage-Worse:** Coverage of topics with particular interest to females is inferior to topics with particular interest to males.

To more directly examine the effect of the gender gap on content coverage quality, we performed two analyses: a general one based on editor activity, and a domain-specific one using an external data source that allowed us to more carefully control for possible confounds. To measure coverage quality, we used article length as

Area	Females	Males	% Female
People	441	3,673	10.7% ***
Arts	382	3,282	10.4% ***
Philosophy	31	344	8.3%
Religion	37	484	7.1%
Health	24	313	7.1%
History	113	1,578	6.7%
Science	235	4,299	5.2% ***
Geography	71	1,856	3.7% ***
<b>All Editors</b>	<b>1,915</b>	<b>23,430</b>	<b>7.6%<sup>a</sup></b>

**Table 1: Gender distribution of editors in eight interest areas as of January 2008. Not all editors are in an interest area, and some editors are in multiple areas. Statistical comparisons were performed using a Chi-Square goodness-of-fit test using the overall gender breakdown of editors in this data as the expected distribution (\*\*\*)  $p < 0.001$ .**

<sup>a</sup>7.6% of editors being female is very low compared to figures reported earlier. This is a manifestation of the survival-related confounds described in section 4.1. The data issues do not affect this analysis because the gender distributions we are comparing are all drawn from the same data.

a proxy. While length is a very simple metric, existing research shows that it is an excellent predictor of a Wikipedia article’s assessment level [3, 32], which is a community-assigned rating of an article’s overall quality.

**Wikipedia-wide comparison.** In our first analysis, we used the gender of an article’s editors to determine whether the article topic is of male or female interest. Specifically, we defined an article’s  $f_e$  as the proportion of its known-gender editors that are female. As we found earlier in section 4.1, edits by female editors turn out to be more rare than one would expect, so this metric is subject to high relative variance and noise. To help reduce the effect of noise, we limited this analysis to high-activity articles where we knew the gender of at least 30 editors. In addition, we excluded any articles that are less than 100 bytes long because such “articles” are likely to be redirects that point to other articles. Applying these constraints left us with 59,579 articles in the *Main* namespace, which is where all of Wikipedia’s encyclopedic content is located.

We considered an article topic to be “male” if it is in the bottom quintile of  $f_e$ , “neutral” if it is in the third (center) quintile, and “female” if it is in the top quintile. We found that the average male article is 33,301 bytes long, the average female article is 28,434 bytes long, and the average neutral article is 36,511 bytes long. All differences are statistically significant ( $t$ -Test,  $p < 0.001$ ). So, on average, male articles are significantly longer than female articles, which indicates that coverage quality of topics with particular interest to females is indeed lacking. We also see that neutral articles are longer than both male and female articles. This is perhaps because gender-neutral topics appeal to both genders, and thus, are likely to garner the most overall attention from Wikipedia’s editors.

**Movie article comparison.** The second analysis for this hypothesis looked at one specific domain of topics: movies. We chose to focus on this particular domain because it allowed us to use a method that let us more carefully control for other factors that may affect an article’s length such as the popularity or age of the topic. Furthermore, this method is *not* subject to the Wikipedia data assumptions described in section 3.2.

For this analysis, we used data from our movie recommender web site *MovieLens*<sup>2</sup>. MovieLens users can assign ratings to movies

<sup>2</sup><http://movielens.umn.edu>

in order to receive personalized movie recommendations. To date, MovieLens has collected over 15 million movie ratings from its 150,000 users. A key feature of this data is that it contains self-reported gender information from over 80% of users who started using MovieLens before May 2003 (MovieLens stopped requesting demographic information upon registration in May 2003). While MovieLens also appears to be affected by a gender gap (32% of its users are female), it is less imbalanced than Wikipedia, which allowed us to compute a gender metric even for relatively obscure movies. We defined a movie's  $f_r$  as the proportion of its known-gender raters that are female. To help avoid confounds due to gender differences in long-term MovieLens usage, we limited our analysis to movies that existed in the system as of May 2003.

We mapped each MovieLens movie to its corresponding article by scanning Wikipedia for articles that have a link to the movie's *IMDb* page and then applying basic heuristics to compare the movie name and article titles. We hand-checked 100 randomly-selected mappings and found no errors. We excluded any movies with fewer than ten known-gender raters, as well as movies that we could not locate a Wikipedia article for (3.7% of movies). The resulting movie data set contained 5,850 movies.

For this analysis, we built a linear regression model that predicts article length using  $f_r$  and several movie properties that may affect article length. The regression variables are summarized below. All variables except *Movie Age* are standardized with a z-score transformation. VIF values for these variables are below 1.3, so multicollinearity is not an issue.

- **Article Length** is our dependent variable and is defined as the length (in characters) of the Wikipedia article about the movie (log-transformed for normality).
- **Movie Gender** is our independent variable and is defined as  $f_r$ . We additionally include its quadratic term (labelled as “*Movie Gender Sq.*”) because our previous analysis suggests that article length may have a non-linear relationship with topic gender.
- **Movie Popularity** is a control variable defined as the number of MovieLens ratings that the movie has (log-transformed for normality). Articles about often-rated movies may draw more attention from editors, and thus, may be longer.
- **Movie Quality** is a control variable defined as the average MovieLens rating assigned to the movie. Articles about well-liked movies may be longer than ones about poorly-rated movies, again due to increased editor attention.
- **Movie Age** is a control variable defined as the age of the movie in years. Article length may vary with movie age due to better availability of information about newer movies.

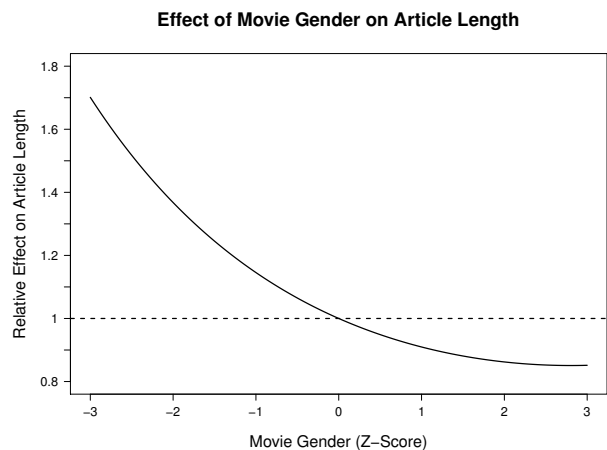
The results of the regression model are shown in table 2. We see that even when controlling for variables that we would expect to affect article length, both  $f_r$  and its quadratic term are significantly associated with article length. A plot of the effect size of movie gender is shown in figure 6. All else being equal, articles about “female” movies are shorter than ones about “male” movies<sup>3</sup>. We also built a similar regression model using WikiProject Film’s article assessment ratings<sup>4</sup> as the dependent variable (coded as an equally spaced ordinal variable) and obtained qualitatively similar results; that is, “female” movie articles have lower assessment ratings than “male” movie articles.

<sup>3</sup>The Movie Popularity control variable has the expected effect. The Movie Quality variable has no significant effect. The Movie Age variable has the opposite of the expected effect: older movies tend to have longer articles. Perhaps only very noteworthy older movies were in our data set.

<sup>4</sup><http://en.wikipedia.org/wiki/WP:FILMA>

Variable	Coef.	SE
Intercept	-0.335 ***	0.0190
Movie Popularity	0.676 ***	0.0106
Movie Quality	0.00521	0.0107
Movie Age	0.0123 ***	0.000604
Movie Gender	-0.144 ***	0.0107
Movie Gender Sq.	0.0257 ***	0.00692
Adj. $R^2 = 0.4704$ , $F(5, 5843) = 1040$ , $p < 0.001$		

**Table 2: Results of multiple linear regression model with movie article length as the dependent variable. Positive coefficients indicate variables associated with increased article length, and negative coefficients indicate variables associated with decreased article length (\*\*\*)  $p < 0.001$ .**



**Figure 6: Effect size of movie gender ( $f_r$ ) on Wikipedia article length. Movie gender is shown as standardized z-score, where negative values are more “male” and positive values are more “female”. Effect is shown as a multiplicative factor. For example, a movie that is at -2 on the x axis is two standard deviations more “male” than the average movie, and its Wikipedia article is expected to be 1.4 times as long as the average movie’s article.**

The results of these two analyses show the same thing: there are measurable gender-associated imbalances in Wikipedia’s content coverage quality. There is a silver lining though: in follow-up analyses of two other domains – Nobel Prize winners, and recipients of the Academy Award for Best Actor/Actress – we found that the average length of articles about female subjects is comparable to that of articles about male subjects. The effects of Wikipedia’s gender gap do not seem to apply when it comes to coverage of very important and notable topics (these award winners are arguably at the pinnacle of scientific, social, and/or cultural achievement).

**H2c F-Social:** Females are more likely to be involved in social- and community-oriented areas of Wikipedia.

To test this hypothesis, we performed comparisons of male and female involvement in three different areas of Wikipedia that represent increased social or community engagement. We describe each of the three in turn below.

**Editing Behavior.** First, we looked at social engagement at a broad level by examining activity within the *User* and *User Talk* namespaces. Wikipedia’s guideline<sup>5</sup> on these namespaces states: “There is no fixed use for user pages, except that usually one’s

<sup>5</sup><http://en.wikipedia.org/wiki/WP:UPYES>



Namespace	Females	Males
<i>User &amp; User Talk</i> ***	<b>25.2%</b>	19.1%
<i>Main &amp; Talk</i> ***	69.1%	<b>74.5%</b>
Other namespaces ***	1.88%	<b>2.82%</b>

**Table 3: Comparisons of Wikipedia editing behavior across different namespaces. The figures are the percentage of editing behavior in each namespace, averaged across female or male editors. Differences were compared using a *t*-Test (\*\*\*)  $p < 0.001$ .**

	Females	Males
<b>Menteeships (Jan. 2010 data)</b>		
Adopted users ***	<b>0.54%</b> (51/9511)	0.30% (200/67252)
<b>Administrators (Jan. 2011 data)</b>		
Administrators ***	0.33% (50/15362)	<b>0.59%</b> (579/98486)
" $\geq$ 2,000 edits *	<b>18.6%</b> (49/263)	13.4% (575/4283)

**Table 4: Comparisons of Wikipedia participation in mentoring and administration. The figures are the percentage of each gender to have participated. Proportions were compared using a Chi-square test (\*\*\*)  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .**

user page has something about oneself, and one’s talk page is used for messaging.” Since pages in these namespaces are usually used for self-expression and interpersonal communication, we interpret editing activity within them to be indicative of social engagement. We looked at users with at least ten edits (4,990 females and 43,850 males), computed the percentage of each user’s edits that are in each namespace, and compared the male and female means. Table 3 shows our findings for three groups of namespaces: 1) *User* and *User Talk*, 2) *Main* and *Talk*, and 3) all other namespaces. We see that on average, a female makes a significantly higher concentration of her edits in the *User* and *User Talk* namespaces, mostly at the cost of fewer edits in *Main* and *Talk*, which contain encyclopedic content and discussions about the content, respectively.

**Menteeships.** Second, we looked at activity in one of Wikipedia’s mentoring programs, Adopt-a-User<sup>6</sup>. Specifically, we looked at all editors who obtained a mentor, referred to as an “adopter,” through this program. We view participation in this program as a form of social engagement because the mentoring process is inherently interpersonal and social [22]: an adoptee chooses to seek personalized help and advice from a more experienced peer instead of going it alone. In the Adopt-a-User program, once an adoption relationship has formed, the editor who has been adopted places an *Adoptee* template on his or her User page. To identify program participants, we scanned all User pages in the January 2010 data dump for editors with an *Adoptee* template on their User page at any point in their history. Table 4 shows our results, which indicate that females are significantly more likely than males to participate in this program as mentees.

**Administrators.** Lastly, we looked at how often editors have become Wikipedia administrators. Being an administrator provides a user with additional capabilities such as protecting pages, hiding revisions from public view, and blocking others from editing. Besides their usual editing tasks, administrators have the additional responsibility of being Wikipedia’s janitors and custodians<sup>7</sup>. Administrators help clean up after vandals, resolve conflicts between editors, and maintain order in the community. Since becoming an administrator is a major form of civic duty in Wikipedia, we consider it an indicator of increased community engagement.

<sup>6</sup><http://en.wikipedia.org/wiki/WP:ADOPT>

<sup>7</sup>Wikipedia’s symbol for an administrator is a janitorial mop!

Returning to table 4, we see that overall, a greater proportion of males than females become administrators. However, one major factor in successfully becoming an administrator is having a substantial edit count [5], and we know from our earlier results that females are more likely to leave Wikipedia before accumulating many edits. If we restrict our analysis to users who have at least 2,000 edits (all but five administrators have over 2,000 edits), the tables are turned. Within this group of dedicated Wikipedians, females are actually significantly *more* likely to become administrators than their male counterparts. We also performed the same comparison using thresholds of 1,000 and 4,000 edits and obtained the same qualitative result.

The data support all three hypotheses under **RQ2: Gap-Matters**. In **H2a Focus-Differences** and **H2b F-Coverage-Worse**, we find that the gender gap appears to have a detrimental effect on content coverage of topics with particular interest to females. Our results for **H2c F-Social** suggest that addressing the gender gap could help Wikipedia better address its needs in social- and community-oriented areas.

### 4.3 RQ3: Gender-Conflict

**H3a F-Uncontentious:** Females tend to avoid controversial or contentious articles.

We address this hypothesis by looking at the edit protection status of *Main* namespace articles that have a high concentration of either female editors or male editors. An edit-protected article cannot be edited by certain classes of users, depending on the level of protection (typically new or anonymous editors). Wikipedia’s protection policy<sup>8</sup> states that articles that are subject to content disputes, vandalism, or other forms of disruption are candidates for protection. Therefore, protected articles tend to be ones that are about controversial or contentious topics.

We found that 5.20% of the “female” articles described in section 4.2 are protected, while just 2.39% of the “male” articles are protected,  $\chi^2(1, N = 23989) = 129.1, p < 0.001$ . Thus, articles that have a higher concentration of female editorship are actually *more* likely to be contentious than those with more males.

**H3b F-Reverted-More:** Female editors are more likely to have their early edits reverted.

Our analysis of reverts uses the January 2010 data dump, which is the most recent dump that we had processed revert information for<sup>9</sup> We used the method described in Priedhorsky et al. in [21] to detect reverts and to classify whether they are for damage repair (specifically, we use the “D-Loose” classification, which is an imperfect heuristic, but identifies many common vandalism repair patterns). Because we were only interested in reverts of good-faith attempts to improve the encyclopedia, we only considered reverts in the *Main* namespace that were *not* for the purposes of repairing damage or vandalism. To limit the effect of right truncation, we also only considered reverts that occurred within one week of an edit (this was the case for over 95% of the reverts in our data set).

Using this data, we took each user’s chronological sequence of *Main* namespace edits, partitioned the edits into bins of increasing size to represent different stages of editor tenure, and determined what percentage of edits in each bin were reverted for non-vandalism reasons. Table 5 shows the results of this analysis aggregate

<sup>8</sup><http://en.wikipedia.org/wiki/WP:PROTECT>

<sup>9</sup>This dump was generated with about five months of data missing (<http://blog.wikimedia.org/2010/05/29/>). Because this issue only affects a small subset of edits made by 2.7% of users in our analysis, we do not believe it materially affects the reported results.

Revert Rates During Editor Tenure				
Edit #	Females		Males	
	N	Mean	N	Mean
1	6,305	6.99% ***	53,738	4.96%
2-3	4,989	6.04% ***	45,514	4.44%
4-7	3,798	4.69% **	37,272	3.98%
8-15	2,871	3.47%	30,066	3.12%
16-31	2,080	2.56%	23,798	2.66%
32-63	1,490	2.33%	18,270	2.49%
64-127	1,039	1.97%	13,850	2.27%
128-255	749	2.29%	10,355	2.07%

**Table 5: Average rate at which editors are reverted for non-vandalism-related reasons, by gender and by stage of editor tenure (the first row shows the reverted rates for users’ first edits, the second row shows the rates for users’ second and third edits, and so on). Averages were compared using a *t*-Test (\*\*\*)  $p < 0.001$ , \*\*  $p < 0.01$ ).**

gated by gender. We see that in the first three bins, which consist of users’ first seven edits, the average reverted edit percentage for females is significantly higher than that for males. Therefore, females are indeed significantly more likely than males to have their edits reverted during the early parts of their tenure.

Interestingly, beyond this initial handful of edits, we see little statistical difference between females and males in how often they are reverted. This suggests that females and males who manage to reach a modest level of Wikipedia experience are on par with each other with respect to community-perceived contribution quality.

**H3c *F-Reverted-Leave*:** Female editors are more likely to stop editing and leave Wikipedia when being reverted as newcomers.

Now, we look at how editors react to being reverted. The survival analysis in section 4.1 indicated that females appear to stop editing Wikipedia sooner than males. We investigate this phenomenon more deeply here, looking at whether female newcomers are more likely than their male counterparts to stop editing if reverted. To do so, we developed a Cox regression model [8] to determine which factors are associated with longer (or shorter) activity lifespan. The variables in our model are limited to those that describe a user during his or her first 24 hours of editing Wikipedia, and are summarized below. VIF values for these variables are below 1.4.

- ***Gender*** is the editor’s gender, dummy-coded with females as 1 and males as 0.
- ***Edits24H*** is the number of edits made in the first 24 hours of editing Wikipedia (log-transformed for normality).
- ***%RvVandal*** is the proportion of edits made in the first 24 hours that were reverted for vandalism-related reasons.
- ***%RvNonVandal*** is the proportion of edits made in the first 24 hours that were reverted, but not for vandalism-related reasons.
- ***%RvNV × Gen*** is an interaction term between *%RvNonVandal* and *Gender*, and is used to study the interaction effect between gender and being reverted for non-vandalism reasons.

The results of the regression model are shown in table 6. The model has limited predictive power, but nonetheless, we see that all the variables *except* the interaction term have a significant association with activity lifespan. Making more edits during one’s first 24 hours as a Wikipedia editor is associated with a longer activity lifespan, while having one’s early edits reverted for any reason, vandalism-related or otherwise, is associated with a shorter lifespan. Note that even after taking these factors into account, being female *still* has a strong association with shorter activity lifespan.

Variable	Coef.	Hazard Ratio	95% CI
Gender (female)	0.248	1.281 ***	1.229-1.335
Edits24H	-0.164	0.849 ***	0.838-0.861
%RvVandal	0.486	1.626 ***	1.498-1.766
%RvNonVandal	0.332	1.394 ***	1.314-1.478
%RvNV × Gen	0.0393	1.040	0.904-1.197
Adj. $R^2 = 0.043$ , $p < 0.001$			

**Table 6: Results of Cox proportional hazards regression model predicting activity lifespan for editors who started editing during 2009. Variables with hazard ratios above 1 are associated with shorter activity lifespans, while those with ratios below 1 are associated with longer lifespans (\*\*\*)  $p < 0.001$ .**

However, contrary to our expectation, there is no interaction effect between gender and being reverted for non-vandalism reasons. It appears that males and females are affected similarly when their edits are not accepted by the Wikipedia community. This point deserves elucidation. Although early reverts appear to drive editors away from Wikipedia, and although females are more likely to be reverted early in their tenure, if a revert happens to a female, the likelihood of her departure is not affected more than that of a male in a similar situation. Therefore, the gender gap appears to be due more to females being reverted disproportionately, rather than to females reacting more strongly when they are reverted.

**H3d *F-Blocked-Less*:** Female editors are less likely to be blocked.

Our final hypothesis looks at blocks imposed by Wikipedia’s administrators. We counted how many males and females have ever been blocked and determined which of those blocks were designated as being for an indefinite length of time. Indefinite-length blocks are typically reserved for cases of “significant disruption or threats of disruption, or major breaches of policy”<sup>10</sup>, so we expected to see fewer females affected by such blocks.

We found that 4.39% of female users (673) and 4.52% of male users (4,449) have been blocked at some point in their Wikipedia tenures, which is *not* a significant difference,  $\chi^2(1, N = 113848) = 0.545$ ,  $p = 0.460$ . Looking only at users who were subject to an indefinite-length block, we found a 3.85% rate for females (592), and 3.32% for males (3,274),  $\chi^2(1, N = 113848) = 11.2$ ,  $p < 0.001$ . So, while males and females appear to be blocked at similar rates, females are significantly *more* likely to be blocked indefinitely.

Intrigued by this result, we performed a follow-up analysis looking at another form of anti-social behavior on Wikipedia: vandalizing articles. Using an approach similar to the one depicted in table 5, we looked at how often males and females have their edits reverted specifically for vandalism-related reasons early in their tenure. The results here are qualitatively consistent with the block results presented above: females are significantly *more* likely to be reverted for vandalizing Wikipedia’s articles (female and male reverted-for-vandalism rate over each user’s first seven edits: 3.26% and 2.11% respectively,  $t(60041) = 7.28$ ,  $p < 0.001$ ).

A limitation of the data available for testing this hypothesis is that of users who are blocked or who are reverted for vandalism, the proportion who have self-reported their gender is even smaller than the baseline. It is possible that gender differences in self-reporting rate or other phenomena among these troubled users are distorting the results presented above. For example, perhaps some people who participate in Wikipedia for anti-social purposes will intentionally mis-report their gender to elicit a different reaction from the community [4]. (We have no data available to test speculations of this sort.) Taken at face value, the results are quite surprising

<sup>10</sup><http://en.wikipedia.org/wiki/WP:BLOCK>



Hypothesis	Supported?	Description
<b>H1a Gap-Exists</b>	Yes	Wikipedia has relatively few female editors, and they leave Wikipedia sooner than males
<b>H1b Gap-Shrinking</b>	No	The gender gap has <i>not</i> been shrinking over time
<b>H2a Focus-Differences</b>	Yes	Females and males focus on different broad content areas
<b>H2b F-Coverage-Worse</b>	Yes	Coverage of “female” topics is inferior to coverage of “male” topics
<b>H2c F-Social</b>	Yes	Females are more likely to participate in social- or community-oriented areas of Wikipedia
<b>H3a F-Uncontentious</b>	Reversed	Articles with high female editor concentrations are <i>more</i> contentious
<b>H3b F-Reverted-More</b>	Yes	Female newcomers are reverted more than males
<b>H3c F-Reverted-Leave</b>	No	Being reverted as newcomers has the same apparent effect on males and females
<b>H3d F-Blocked-Less</b>	Reversed	Females are <i>more</i> likely to be indefinitely blocked

**Table 7: A summary of our hypotheses and findings.**

given the substantial body of literature indicating that males are more likely to act out or violate rules (e.g., [30, 11, 27]).

Our findings for **RQ3: Gender-Conflict** are mixed. We find that female newcomers have a harder time getting good-faith contributions to be accepted by the community (**H3b F-Reverted-More**). However, our findings for **H3c F-Reverted-Leave** indicate that the effect of having an edit reverted is no worse for females than it is for males. Unexpectedly, we find that female editors are *more* concentrated in areas with high controversy (**H3a F-Uncontentious**), and are *more* likely than males to draw corrective actions from fellow editors (**H3d F-Blocked-Less**). In summary, the available data indicate that female editors experience more adversity than male editors in all the areas that we studied.

## 5. DISCUSSION

Table 7 shows a summary of our hypotheses, whether we found support for each hypothesis, and a brief statement of our results. One way of interpreting our results is in terms of Preece and Shneiderman’s Reader-to-Leader Framework for social media participation [20]. Preece and Shneiderman describe a process in which users of a social media system move through four levels of participation: reader, contributor, collaborator, and leader. We systematically look at female participation in Wikipedia through the lens of this framework.

**Readers.** Becoming a consumer of social media is a typical first step toward active participation. Existing survey research yields mixed findings about the share of females in Wikipedia’s readership [16, 35, 10]. The most accurate and unbiased of these appears to be [35], which used random telephone dialing along with statistical correction techniques to account for non-response bias. It indicates a female readership share of approximately 47%, which suggests that Wikipedia is effective at drawing a relatively gender-balanced population of readers. The present research does not consider readership directly, but this figure will serve as a useful reference point.

**Contributors.** Perhaps the most challenging task for a social media system is to convince a reader to start giving back to the community – that is, to turn readers into contributors, or in Wikipedia’s case, editors. Our results for **H1a Gap-Exists** show that Wikipedia is much less successful in “converting” female readers than male readers, dropping from a 47% female share in readership to a 16% share in editors, a figure that has shown little to no change for years according to our results for **H1b Gap-Shrinking**.

**Collaborators.** When multiple contributors come together to work toward a common goal, they become collaborators. In researching **H2c F-Social** we found that females edit more in the *User* and *User Talk* namespaces, indicating a potential interest in collaboration. Our data do not contain metrics that directly describe collaboration activity, but a reasonable proxy is simply the presence of sustained editing activity. Our survival analyses in

**H1a Gap-Exists** and **H3b F-Reverted-More** indicate that females who become contributors stop editing Wikipedia sooner than males. Furthermore, both **H3b F-Reverted-More** and **H3d F-Blocked-Less** (reversed to F-Blocked-More!) suggest that females encounter more adversity in Wikipedia. Together, these data suggest that while females appear interested in becoming collaborators, they have more difficulty in making the transition for a variety of reasons.

**Leaders.** Effective collaborators who are passionate about their work and who are interested in the system at a meta-level emerge as community leaders. Preece and Shneiderman specifically examine Wikipedia administration as an exemplar of a leadership role in online communities. In our analysis for **H2c F-Social**, we see that females who reach a high level of participation are more likely than their male counterparts to take on a leadership and administration role. However, as we saw in **H1a Gap-Exists** only 6% of the editors who have contributed more than 2,000 edits are female<sup>11</sup>. Some Wikipedians have observed that an administrator shortage may be looming as current administrators “retire,” but few new administrators are emerging to fill their shoes<sup>12</sup>. Addressing the gender gap in high-participation editors might be an opportunity to meet this demand for more administrators.

**Implications.** Overall, our findings indicate that there is a substantial male-skewed gender imbalance in English Wikipedia editors that does not appear to be closing at any appreciable rate. This is at odds with observed participation rates in other forms of online social media that are gender-balanced or that are even female-skewed [15, 19]. Furthermore, we find that the gender gap matters to Wikipedia: there are gender-associated imbalances in coverage quality, which impinges on Wikipedia’s goal of producing a high-quality encyclopedia. Not only would addressing the gender gap help resolve the quality disparity, it would also help increase diversity within Wikipedia’s collaborations, which prior research has shown to improve group productivity and retention rates [6] as well as decision-making quality [14].

The problem is subtle, and simple attempts at solution without detailed understanding are likely to fail. Of our nine hypotheses, all of which seemed plausible before we began the study, nearly half were not supported by the data. How can it be that the gender gap in Wikipedia is not closing, though overall Internet usage has become gender-balanced? Taken together, our results for **RQ3: Gender-Conflict** hint at a culture that may be resistant to female participation. More research, including interviews, surveys, and focus groups, is needed to determine the underlying causes of the problems evidenced by our findings, and to determine what can be done to improve the situation. We hope this research is a first step toward addressing the gender imbalance – and the problems it causes – in Wikipedia.

<sup>11</sup>Nearly all administrators have at least 2,000 edits.

<sup>12</sup><http://en.wikipedia.org/w/index.php?oldid=393297323>

## 6. ACKNOWLEDGEMENTS

We are grateful for the support of the members of GroupLens Research, particularly Angela Brandt and Katherine Panciera for fruitful discussions about the gender differences literature. We thank the Wikimedia Foundation for facilitating access to public Wikipedia data. This work is funded by the NSF (grants IIS 08-08692, 09-68483, 10-38277) and by the China Scholarship Council.

## 7. REFERENCES

- [1] J. Antin, R. Yee, C. Cheshire, and O. Nov. Gender differences in Wikipedia editing. In *Proc. WikiSym 2011*, Mountain View, CA. ACM.
- [2] B. A. Bettencourt and N. Miller. Gender differences in aggression as a function of provocation: A meta-analysis. *Psychol Bull*, 119(3):422–447, May 1996.
- [3] J. E. Blumenstock. Size matters: Word count as a measure of quality on Wikipedia. In *Proc. WWW 2008*. ACM.
- [4] A. S. Bruckman. Gender swapping on the Internet. In *Proc. INET 2003*, San Francisco, CA. The Internet Society.
- [5] M. Burke and R. Kraut. Mopping up: Modeling Wikipedia promotion decisions. In *Proc. CSCW 2008*, San Diego, CA. ACM.
- [6] J. Chen, Y. Ren, and J. Riedl. The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proc. CHI 2010*, Atlanta, GA. ACM.
- [7] N. Cohen. Define gender gap? Look up Wikipedia’s contributor list. *The New York Times*, Jan. 2011.
- [8] D. Cox and D. Oakes. *Analysis of Survival Data*. Chapman and Hall/CRC, June 1984.
- [9] A. Feingold. Gender differences in personality: A meta-analysis. *Psychol Bull*, 116(3):429–456, Nov. 1994.
- [10] R. Glott, P. Schmidt, and R. Ghosh. Wikipedia survey – overview of results. Technical report, United Nations University MERIT, Mar. 2010.
- [11] S. C. Herring. Gender and power in on-line communication. In J. Holmes and M. Meyerhoff, editors, *The Handbook of Language and Gender*, pages 202–228. Blackwell, 2003.
- [12] Information Solutions Group. 2010 social gaming research. Technical report, 2010.
- [13] A. Kittur, B. Suh, B. A. Pendleton, and E. H. Chi. He says, she says: Conflict and coordination in Wikipedia. In *Proc. CHI 2007*, San Jose, CA. ACM.
- [14] S. K. Lam, J. Karim, and J. Riedl. The effects of group composition on decision quality in a social production community. In *Proc. GROUP 2010*, Sanibel Is., FL. ACM.
- [15] A. Lenhart, K. Purcell, A. Smith, and K. Zickuhr. Social media and young adults. <http://pewinternet.org/Reports/2010/Social-Media-and-Young-Adults.aspx>, February 2010. Accessed March 19, 2011.
- [16] S. Lim and N. Kwon. Gender differences in information behavior concerning Wikipedia, an unorthodox information source? *Library & Information Science Research*, 32(3):212–220, 2010.
- [17] J. Margolis and A. Fisher. *Unlocking the clubhouse: Women in computing*. MIT Press, 2002.
- [18] National Science Foundation. Science and engineering degrees: 1966–2006. Detailed statistical tables NSF 08–321. Technical report, 2008.
- [19] Pew Research Center. Internet Usage Over Time Trends Dataset. <http://www.pewinternet.org/Static-Pages/Trend-Data/Usage-Over-Time.aspx>, November 2010. Accessed March 19, 2011.
- [20] J. Preece and B. Shneiderman. The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, 1(1):13–32, Mar. 2009.
- [21] R. Priedhorsky, J. Chen, S. K. Lam, K. Panciera, L. Terveen, and J. Riedl. Creating, destroying, and restoring value in Wikipedia. In *Proc. GROUP 2007*, Sanibel Is., FL. ACM.
- [22] J. E. A. Russell and D. M. Adams. The changing nature of mentoring in organizations: An introduction to the special issue on mentoring in organizations. *Journal of Vocational Behavior*, 51(1):1–14, Aug. 1997.
- [23] A. Smith and L. Rainie. 8% of online Americans use Twitter. <http://www.pewinternet.org/Reports/2010/Twitter-Update-2010/Findings.aspx>, December 2010. Accessed March 19, 2011.
- [24] B. Suh, G. Convertino, E. H. Chi, and P. Pirolli. The singularity is not near: Slowing growth of Wikipedia. In *Proc. WikiSym 2009*, Orlando, FL. ACM.
- [25] K. Sweet. 57M U.S. consumers playing social network games: Survey. *Fox Business*, August 23 2010.
- [26] H. Taniguchi. Men’s and women’s volunteering: Gender differences in the effects of employment and family characteristics. *Nonprofit and Voluntary Sector Quarterly*, 35(1):83–101, 2006.
- [27] The Social Issues Research Centre. Sex differences in driving and insurance risk. Technical report, 2004.
- [28] Z. Tufekci. Grooming, Gossip, Facebook and MySpace – what can we learn about these sites from those who won’t assimilate? *Information, Communication & Society*, 11(4):544–564, 2008.
- [29] V. Venkatesh, M. Morris, and P. Ackerman. A longitudinal field investigation of gender differences in individual technology adoption decision-making processes. *Organizational Behavior and Human Decision Processes*, 83(1):33–60, 2000.
- [30] B. E. Whitley, A. B. Nelson, and C. J. Jones. Gender differences in cheating attitudes and classroom cheating behavior: A Meta-Analysis. *Sex Roles: A Journal of Research*, 41:657–80, 1999.
- [31] J. Wilson. Volunteering. *Annual Review of Sociology*, 26:215–240, 2000.
- [32] T. Wöhner and R. Peters. Assessing the quality of Wikipedia articles with lifecycle based metrics. In *Proc. WikiSym 2009*, New York, NY. ACM.
- [33] J. Yang, X. Wei, M. S. Ackerman, and L. A. Adamic. Activity lifespan: An analysis of user survival patterns in online knowledge sharing communities. In *Proc. ICWSM 2010*, Washington DC, USA. AAAI.
- [34] N. Yee. Maps of digital desires: Exploring the topography of gender and play in online games. In Y. B. Kafai, C. Heeter, J. Denner, and J. Y. Sun, editors, *Beyond Barbie® and Mortal Kombat: New Perspectives on Gender and Gaming*, pages 83–96. The MIT Press, Sept. 2008.
- [35] K. Zickuhr and L. Rainie. Wikipedia, past and present. <http://www.pewinternet.org/Reports/2011/Wikipedia.aspx>, January 2011. Accessed March 19, 2011.