The Gendered Geography of Contributions to OpenStreetMap: Complexities in Self-Focus Bias

Maitraye Das
Northwestern University
Evanston, IL, USA
maitraye@u.northwestern.edu

Brent Hecht
Northwestern University
Evanston, IL, USA
bhecht@northwestern.edu

Darren Gergle
Northwestern University
Evanston, IL, USA
dgergle@northwestern.edu

ABSTRACT

Millions of people worldwide contribute content to peer production repositories that serve human information needs and provide vital world knowledge to prominent artificial intelligence systems. Yet, extreme gender *participation disparities* exist in which men significantly outnumber women. A central concern has been that due to self-focus bias [46], these disparities can lead to corresponding gender *content disparities*, in which content of interest to men is better represented than content of interest to women. This paper investigates the relationship between participation and content disparities in *OpenStreetMap*. We replicate findings that women are dramatically under-represented as OSM contributors, and observe that men and women contribute different types of content and do so about different places. However, the character of these differences confound simple narratives about self-focus bias: we find that on a proportional basis, men produced a higher proportion of contributions in *feminized* spaces compared to women, while women produced a higher proportion of contributions in *masculinized* spaces compared to men. We discuss the implications of these complex results for both theory and practice.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**; Empirical studies in HCI;

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300793

KEYWORDS

Peer production, gender, OpenStreetMap, self-focus bias, urban, rural

ACM Reference Format:
Maitraye Das, Brent Hecht, and Darren Gergle. 2019. The Gendered Geography of Contributions to OpenStreetMap: Complexities in Self-Focus Bias. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3290605.3300793

1 INTRODUCTION

Peer production is a powerful example of the potential of social computing in which communities like Wikipedia and OpenStreetMap (OSM)—the 'Wikipedia of Maps' [32, 73]—create high-quality content at previously unimaginable scales. This content has in turn satisfied billions of human information needs [26, 55, 76] and provided essential world knowledge to countless artificial intelligence systems [38, 45, 71].

Despite the many accomplishments of the peer production model, social computing researchers have also identified structural challenges that may be preventing peer production from reaching an even higher potential. One of the most serious arises from the demographic configurations of peer production communities. High-impact peer production communities tend to have major *participation disparities*, with certain types of people being over-represented and others being under-represented as contributors.

One of the most significant of the peer production participation disparities observed in the literature occurs along the dimension of gender. In particular, both Wikipedia and OSM appear to have a severe under-representation of women [15, 20, 36, 50, 61, 87, 88, 92]. Estimates of women's participation on Wikipedia range from 13-18% [15, 36, 50, 61]. While less is known about OSM, it is believed that this participation imbalance could be even larger, with women's participation in the 3-4% range [20, 87, 88].

A key motivating factor in the literature on gender dynamics in peer production is concern that gender participation disparities will result in corresponding gender *content disparities*. In other words, it has been assumed that the limited representation of women may lead to peer-produced content that is less able to serve women's information needs [51, 65, 92] and that the many AI systems that "learn" from peer-produced content may take on a biased view of the world [49, 56]. This assumption is supported by the notion of "self-focus bias" [46], in which peer production communities produce an out-sized proportion of content in areas of interest to the cultural groups present in the community.

Despite the importance of the assumed relationship between gender participation disparities and content disparities, little work has sought to empirically explore this relationship. For example, research on Wikipedia has observed differences in the characterization and structure of biographical content about women [95, 96, 98] and the quality of content of greater interest to women [61]. Yet, most of this work, which has focused on limited topics in Wikipedia, has not directly linked content differences to the gender of the editors involved. As such, it is unclear whether gender participation disparities led to any observed content disparities. (For a notable exception see [61], discussed in related work).

In this paper, we contribute to a more complete picture of how peer-produced content is affected by gender-based contribution disparities. We do so by examining the content generated by a large sample of male and female power editors in OSM. Our results support the hypothesis that men and women tend to contribute different types of content. We observe this both in terms of the regions where men and women edit (e.g., rural vs. urban) and in terms of the type of content that they contribute (e.g., the specific spatial entities they add to the database). However, our results also reveal critical complexity in these differences that belie a simple gender-based self-focus bias interpretation [65, 68, 92]. In particular, we observed that men disproportionately contribute to entities that critical geographers [65, 92] have identified as being in *feminized spaces* and women disproportionately contribute entities in *masculinized spaces*. Additionally, our analyses point to complicated intersectional dynamics, with contributor gender being associated with a likelihood to exacerbate or mitigate other known content biases in OSM (e.g., those related to the rural-urban spectrum [18, 55, 83]).

Overall, our findings present challenges to overly simple narratives about gender and participation, raise new opportunities for important further research, and have implications for the sociotechnical design of peer production communities. We also contribute to a more nuanced theoretical understanding of the notion of self-focus bias, an important heuristic for understanding the relationship between contributor demographics and the content they produce [46]. An important caveat to note is that the research reported here is limited in that it only considers the genders of men and women. As such, our results suggest that future work that takes a less binary approach will be particularly important.

2 RELATED WORK

Two prevalent areas of social computing research inform our work on gender disparities in OSM. The first is research on gender participation disparities as broadly construed in peer production. The second is research detailing content disparities that exist in peer-produced repositories. We further discuss an implied assumption that resides in much of the literature – namely, that a form of self-focus bias [46] operates in which gender participation disparity leads to associated gender-based content disparity.

Gender Participation Disparities

A number of studies have revealed a substantial participation gender gap in Wikipedia [36, 50], with some studies suggesting that the gap is even more prominent among the most active contributors [15, 61]. Less is known about gender participation in OSM where recent demographic information is more limited [51]. However, earlier surveys suggest that OSM editors are mostly men, well-educated and tech-savvy, with women representing only 3-4% of the community [20, 87, 88]. A survey conducted by Stephens indicated that women were less familiar with OSM than men and that their contribution levels exhibited even greater disparity [92]. While understanding the relationship between participation disparities and content disparities is the main focus of this paper, our findings also add empirical information that bolster existing evidence of a severe participation gap in OSM, specifically among the most active editors.

Another important facet of research focuses on understanding the causes of gender participation gaps. Some that have been identified include the pipeline of skills necessary to edit Wikipedia [43, 90], personal preferences for online collaboration [17, 24, 63, 72] and levels of confidence [17, 24, 82]. Similarly, steep learning curves, insufficient technical feedback, and lack of time are found to be influential factors behind users' inactivity in OSM [67, 87, 88]; however, Schimidt et al. found no gender-related difference with respect to these factors [87]. Steinmann et al. compared women's participation rates across different social media and peer production platforms, and suggested that a lack of social aspects and stringent rules might be responsible for women's lower participation in peer production environments like OSM [91].

Content Disparities

Another important line of research focuses on content disparities in peer-produced repositories. For example, in the case of Wikipedia biographies, men and women appear to be covered equally well [60, 95]. However, biases exist in the ways women are portrayed, with their biographies more likely to explicitly mention family, relationships or gender in comparison to men's biographies [37, 95, 96, 98]. Content differences also manifest in the linguistic choices made by editors in a way that generalize men's successes but not their failures and vice-versa for women [78, 95].

Critically for our analysis, content disparities also have important spatial components. For example, rural areas have lower coverage and/or quality in OSM and Wikipedia compared to urban areas [55, 69, 101]. Similarly, regions with higher levels of education and socioeconomic status (SES) exhibit better coverage [18, 39]. Given that it has been observed that editing outcomes vary along rural/urban and SES spectra, it is important to incorporate these into analyses of contributions in OSM, and we do so here.

Most relevant to the work in this paper is a debate about feminized spaces receiving less attention from men and thus being under-represented in OSM [30, 51, 65, 68, 92]. Although studies have suggested that gender participation disparities may result in a male-oriented worldview in OSM [65, 92], this assumption cannot be directly validated without investigating the mapping behaviors of male and female contributors.

Relationship Between Participation Disparities and Content Disparities

Much less work has been done on the relationship between participation disparities and content disparities than on those individual disparities themselves. The primary theory about this relationship is *self-focus bias*, which describes the phenomena in which contributors focus disproportionately on information that is particularly relevant to dominant cultural groups in the peer production community [1].

Self-focus bias has been empirically observed in a number of peer production contexts. For instance, prior work has seen self-focus bias with respect to geography (i.e., people contribute information about places local to them) [28, 42, 47, 100]. Self-focus bias is also prevalent in peer-produced content in terms of language [46, 48, 89] and politics [52].

The work exploring self-focus bias in a gender context is very limited. Lam et al. [61] found that men and women focused on different content areas in the English Wikipedia, and that a gender participation gap among editors led to a corresponding content disparity whereby articles of interest to women were found to be of lower quality than those of interest to men. On a related note, Antin et al. found no evidence that men and women were interested in different types of Wiki-work such as creating new articles, adding citations, fixing typos; however, they did see evidence of differences in terms of revision size and revision count [15].

Within OSM, Stephens [92] presented evidence of male dominance in shaping OSM's tag ontology where proposals to include spaces associated with feminized skills [65] (e.g., 'childcare' or 'hospice') were debated and rejected by men or abandoned without a vote. In contrast, sexual entertainment venues that reflect "male privilege and female objectification" [65], according to prevalent gender norms, had several variations included. For example, 'swinger club' was accepted without a single downvote, and 'brothel', 'nightclub' and 'stripclub' already existed [92]. However, the existing literature does not explain whether such self-focus bias behavior also prevails during the actual mapping of places as opposed to defining the ontology.

[1]Within the self-focus bias literature, culture is usually defined using the framework from Clark [23], in which cultures are groups that share common knowledge. Clark specifically identifies gender as one of these groups.

3 RESEARCH QUESTIONS

Our overarching goal is to understand to what extent contributor gender plays a role in characterizing the information represented in OSM. Specifically, our research is guided by two research questions:

- RQ1: Are there differences in *where* male and female OSM editors contribute?

This first research question seeks to capture differences in the geographic context of male and female contributions. For instance, from the perspective of what has been edited, are there regions in the country that are characterized more by men than by women? Or, from an individual contributor perspective, do men or women contribute disproportionately to certain types of regions? We examine both simple contiguous regions (e.g., regions of the United States) as well as types of regions (e.g., urban vs. rural), since research has shown that contribution behavior on OSM can vary widely across these human geographic dimensions [18, 39, 55, 83].

- RQ2: Are there differences in *what* male and female OSM editors contribute?

This second research question explores whether men and women disproportionately contribute to different types of spatial entities (e.g., barbershop, childcare, etc.). We are particularly interested in assessing whether men and women contribute differently to *feminized* and *masculinized spaces*, and—to the extent that we see evidence of this—examining whether self-focus bias plays a role in governing the relationship between the two.

4 DATA COLLECTION AND PROCESSING

In this section, we discuss the original sources of our data and detail our steps for data collection and processing. Following prior work [18, 39, 56, 99, 101], we focused on an individual country in this work, specifically the United States.

OSM Datasets

We downloaded OSM history data for the United States from geofabrik.de [3] on February 18, 2018. In OSM, three types of spatial elements can be added to the map: *nodes*, *ways*, and

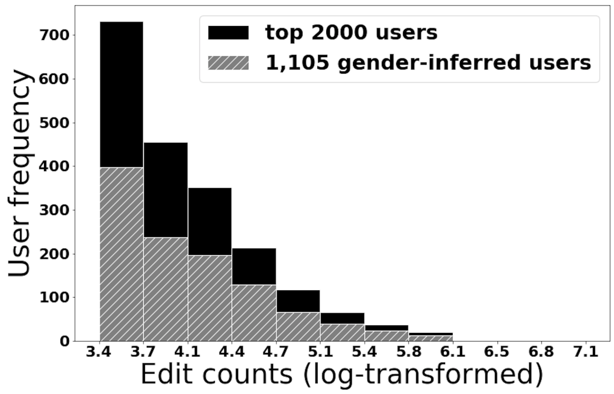

Figure 1: Overlaid histogram of edit counts of top 2,000 users and the 1,105 gender-inferred users. The top 2,000 users contributed 95.36% of all *no-bots* edits.

relations. Nodes can represent point features on their own (e.g., a bench, bus stop, etc.) or they can serve as part of an ordered list of nodes that represent the shape or path of a way. Relations are ordered lists of one or more nodes, ways, and/or other relations that capture important relationships between elements (e.g., bus routes). All types of spatial elements can have a number of tags (i.e., key-value pairs) which describe the features of the particular element to which they are attached. For example, some common tag keys are 'name', 'addr:city', 'amenity', etc.

Node Datasets ("No-bots" and "With-bots"). We consider all node edits in the 48 conterminous U.S. states and District of Columbia. Throughout the paper, by "edit" we refer to different mapping activities, such as adding a node, modifying locations, adding or altering tags, etc. In OSM, editors often use automated software agents (i.e., bots) and automation-assisted batch editors to bulk import pre-existing data. For example, data have been imported in bulk to OSM from the U.S. government's TIGER/Line Street datasets. 'Bot edits' are often not regarded as volunteered human activity [55, 83]. To detect bot edits, we used changeset [2] files downloaded from planet.osm [4]. Following prior work [55, 83, 99, 100], we marked edits as 'bot edits' when they came from a changeset containing edits in very large quantities (more than 4,000 edits) or at a very fast rate (more than one edit per second). Initially, our dataset had 1,019,366,964 node edits. After removing bot imports, there were 91,097,410 node edits done by 62,083 users.

Next, we reviewed the top 2,000 users who generated the most edits to determine their gender identities. In aggregate, this group generated 95.36% of all the edits (without bots) in our dataset. We were able to infer the gender of 1,105 users (57 female, and 1,048 male; over half of the top 2,000 users; see below for our gender inference approach). These 1,105 users contributed to 53.60% and 43.75% of all the edits in the U.S. without and with bot edits, respectively. Figure 1 shows that our gender-identified users appear to be equally distributed across the top 2,000 users in terms of edit counts.

Our primary analysis focuses on the *"no-bots"* dataset; however, we also conducted analyses using the *"with-bots"* dataset to determine whether or not men's and women's mapping patterns change when using bots. The literature suggests that bot operations by editors play an important role for quality control processes as well as creating information about under-represented areas [34, 35, 55]. Thus, analyzing bot activities can provide valuable insights about how men and women variably express their voice through different contribution mechanisms.

Tag Datasets ("Narrow" and "Broad"). We also analyze the specific spatial entities men and women edit in OSM to understand whether they contribute differently to the *feminized* and *masculinized spaces* identified by critical geographers [65, 92]. The 'amenity' [5] tag is of particular interest as it can help identify whether an entity can be associated with *feminized* or *masculinized spaces* (e.g., amenity = 'childcare' or 'brothel'). Initially, our dataset contained 3,134,574 edits on different types of elements with amenity tags. Next, we extracted only the edits done by our 1,105 editors, who contributed to 46.12% of all edits on elements with an amenity tag. These edits contained 1,524 unique amenity values. After normalizing the dataset to account for well-known variations or misspellings, we were left with 867 amenity values.

To guide our formulation of the differences between *feminized* and *masculinized spaces*, we deferred to prior literature in feminist geography. Stephens [92] and Leszczynski and Elwood [65] suggested that places related to nurturing and caregiving are highly *feminized* [64, 81], while public establishments of sexual activities that rely on "female objectification and male privilege" are considered *masculinized spaces* [54][3]. Leszczynski and Elwood write that although sexual venues are not exclusively male spaces, "longstanding gender norms around the expression of sexuality accord men roles as sexual actors and presume women to be passive and submissive recipients of that activity" (p.17, [65]).

We developed two datasets of amenity values for analysis. The first is the *narrow* dataset, based strictly on the amenity types used by critical geographers examining OSM [65, 92]. The amenities included are 'childcare', 'baby-hatch', 'preschool', 'kindergarten' and 'hospice' as *feminized spaces*; 'brothel', 'nightclub' [85], 'strip club' [33], and 'swinger club' [74] as *masculinized spaces*; and the rest as *non-gendered* amenity values.

[2] A changeset consists of group edits (max 10,000) done by a single user over a short period of time (max 24 hours).

[3] Some post-feminist theories view certain sexual entertainment venues as spaces for "female expression, consumption, and autonomy" [19, 29, 41, 93].

Our second dataset—the *broad* dataset—contains amenity types generalized from the initial categories established by Stephens [92] and Leszczynski and Elwood [65]. This dataset also includes caretaking-oriented social facilities such as 'daycare', 'assisted living', 'nursery', 'nursing home', 'retirement home' and 'senior centre' as *feminized spaces*; and sexual venues like 'love hotel', 'sex shop', and 'adult' as *masculinized spaces*. We also drew on established work to include amenities associated with longstanding gender norms such as 'beauty' [86], 'nail salon' [97], 'family_planning' [75, 84] and 'sorority' as *feminized spaces*; and 'sperm bank', 'fraternity', and 'barber shop' [44] as *masculinized spaces*. We further supplemented this dataset by collecting common tag key-value pairs that were automatically suggested by the default OSM editor (iD) when adding gender oriented features to OSM (e.g., 'amenity = clinic', 'healthcare = clinic' and 'healthcare:speciality' = abortion' to describe an abortion clinic). In total, our broad dataset includes amenities for 22 *feminized spaces* and 10 *masculinized spaces*, and the rest are *non-gendered* amenities.

Gender Inference

Since OSM does not collect information about editors' genders, one of the most challenging aspects of our methodological pipeline was performing gender inference. This is a very common obstacle to social computing research that asks important questions related to gender [22, 31, 53, 66, 91, 94]. We based our approach on prior published techniques used to infer the gender of users in Google MapMaker [91], StackOverflow [66], GitHub [94], Resume Search Engines [22], and DBLP Computer Science Bibliography [53].

Our specific procedure was as follows: First, we searched for the profile of a user in the OSM site, OSM Wiki [6] and OSM Help forum [7] and then attempted to infer gender from their profile image, listed real name or text description. We also expanded our search to include information about the user from different social media accounts (Twitter, GitHub, LinkedIn, etc.) [66, 94] or personal websites that we could associate with their OSM profile. When applicable, we used the Gender API [8] to determine gender from user names, following prior work that used similar API services [53, 57].

Applying this procedure, three human coders performed gender inference for the top 2,000 OSM editors in our (*no-bots*) dataset - an extensive process that took over one hundred hours of manual labeling time. The first coder reviewed all 2,000 OSM editors and independently assigned each to a gender category ('male', 'female', 'unable to determine'). To ensure the reliability of our coding procedure, two additional coders independently performed the same task on non-overlapping halves of the set of 2,000 OSM editors. We then assessed inter-rater reliability using Cohen's Kappa and achieved κ = 0.70 and 0.75, which indicate substantial agreement among the coders [62]. The coders then resolved any disagreements through discussion and came to consensus regarding the male or female gender of 1,131 editors. Among them, we determined that 26 editors worked for commercial mapping services such as MapBox and Development Seed. We did not include edits by these editors in our analyses due to concerns that they received remuneration for their contributions and thus their behaviors may not generalize to peer production more broadly. We were left with 1,105 users, among whom 57 were coded as female and 1,048 were coded as male (5.16% female) – values that roughly align with prior studies from 2010 [20] and 2013 [87, 88] indicating that the percentage of female editors in OSM ranges from 3-4%.

Sociodemographic Variables

Prior research has shown that contributions on OSM can vary across different types of regions (e.g., urban vs. rural, poor vs. wealthy, etc.) [18, 55], suggesting that it is important to consider these dimensions when studying OSM editing behavior. To understand the differences in the way men and women contribute to different types of regions, we analyzed the association between editors' gender and the attributes of the regions they mapped. Specifically, we focus on the urban-rural spectrum, socio-economic status (SES) and race and ethnicity.

To capture variation across the urban-rural spectrum, we examined gender contribution ratios in the counties in the top quartile and bottom quartile of the percentage of the population that lives in urban areas (*%pop-urban*) (according to the 2010 U.S. Census [1]). For robustness, we also utilized another metric—2013 Rural-Urban Continuum Codes from U.S. Department of Agriculture [2]—which assigns one of nine codes to each county. We compared the editors' contributions in the counties assigned "1" or "2" (most urban) with those assigned "8" or "9" (most rural). Regarding race and ethnicity, we used the 2011-2015 American Community Survey data on the percentage of each county's population that is both White and not Hispanic or Latino (*%WnHL*) [11]. Specifically, we compared the counties with the highest quartile of *%WnHL* to the lowest quartile. We took a similar approach for SES, but used the 2012-2016 American Community Survey data on median household income (*MHI*) [11]. [4]

5 ANALYSIS METHODS

Following prior work in the peer production domain (e.g. [47, 70], we approach our analysis from two different perspectives. The first perspective is *contribution-centric*: it focuses on the gendered geography of existing contributions to

[4] Although urban-rural divide has correlation with SES, this correlation by no means is perfect; for example, there are 501 counties within the poorest regions that are not rural (e.g., Bronx county in New York) and 504 rural counties that are not among the poorest (e.g., Brown county in Indiana).

OSM. The unit of analysis here is the individual *contribution* (and which gender contributed it). The second perspective is *contributor-centric*: it focuses on editing behavior of the contributors themselves and any differences that may exist along gender lines with regards to where and what they edit. In this case, the unit of analysis is the *contributor* (and their gender). If all contributors made the same number of contributions, these two perspectives would produce identical results. However, in social computing and peer production specifically, this is rarely the case, and certain contributors make orders of magnitude more contributions than others, even among frequent editors [40, 58, 59, 79]. This has made these two analytical perspectives valuable when examining peer production contribution behavior as we do here.

For our *contribution-centric analysis*, we aggregate the edit data into contingency tables, present the total counts and corresponding percentages by gender and sociodemographic dimension of interest (e.g., the urban-rural divide), and apply Chi-square tests to reveal whether, in total, contributions by men and women are distributed differently. Our *contributor-centric analysis* requires more sophisticated methods. The outcome variable here is the count of contributions *aggregated by editor*. We apply negative binomial regression with a dispersion parameter to account for the fact that our outcome variable represents count data that also exhibits over-dispersion (i.e., the variance exceeds the mean). The predictor variables for our models depend on the research question under investigation: For the first research question (detailed in the following section) we include *type of region (two levels)*, *gender (male, female)*, and a *type of region X gender* interaction term as predictors. For our second research question, we replace type of region with a *gendered space (masculinized, feminized, non-gendered)* variable and corresponding interaction term.

6 RESULTS

Recall that our overarching goal is to understand to what extent gender plays a role in characterizing the information in OSM. In the simplest case, if gender did not matter then we would expect similar activity in where and what the women and men in our sample contribute. In what follows, we frame the presentation of our results using the two research questions that guided our exploration:

RQ1: *Where* are the contributions?

We begin our investigation into the geographic regions and types of regions edited by taking a *contribution-centric* perspective that examines the entire collection of edits produced by the men and women in our sample.

Edits from the highly active male editors are more distributed than those from the highly active female editors (see Figure 2). Every county (3,109) in the conterminous U.S.

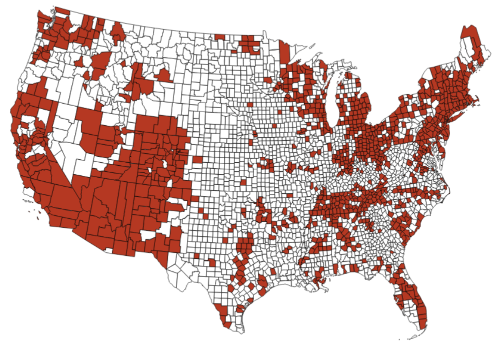

Figure 2: The U.S. counties shown in red are those with at least one edit from the women power editors in our sample in the *no-bots* dataset. A prominent "No Female Edits Belt" (in white) is visible running from the Northern Mountain West down through the Great Plains, Midwest, and Appalachians (note: these counties may be edited by non-power-editors or unidentified female editors).

received at least one edit from the male editors, while about one-third of the counties (1,017) received at least one edit from the female editors (a significant difference in percent coverage based on Fisher's Exact Test of the difference, $p < 0.0001$). In fact, less than three percent (72) of the counties in our dataset have a higher ratio of female edits to male edits. We also looked at the edits by the top 57 male editors as a more direct comparison to the 57 female editors, and found they still made edits in 99.94% (3,107) of the counties.

While Figure 2 illustrates the specific regions edited by the women in our dataset, it does not highlight trends in the *types of regions* that women and men tend to edit. To better understand potential differences along these lines, we analyzed edits across our demographic spectra – urban/rural divide (*%pop-urban* and *rural-urban continuum code*), socio-economic status (*MHI*) and racial/ethnic diversity (*%WnHL*).

In terms of the urban-rural divide, the most rural counties had a higher proportion (nearly 5% greater) of male edits compared to the proportion of female edits, on both measures of *%pop-urban* and *rural-urban continuum code* (see Tables 1a, 1b; all p's < 0.001). In terms of racial and ethnic diversity, the least diverse counties received a higher proportion (8.92% greater) of the male edits compared to that of the female edits, a result that, as noted above, may be associated with the demographics of rural areas (see Table 1c; $p < 0.001$). In terms of socio-economic status, we find a higher proportion (8.41% greater) of female edits in the poorest counties compared to that of male edits in the same regions (see Table 1d; $p < 0.001$). The effect sizes for the overall proportional differences are small when assessed using standardized effect measures such as Cramer's V; however, this aggregate measure masks large and meaningful individual effects especially with respect to female editing patterns. For example, for *%pop-urban*, females

Table 1: Male and Female Edits in Different Types of Regions

(a) Urban-Rural Divide *(%pop-urban)*

County Type	Female Edits	Male Edits
Most Rural	41,382 *(1.87%)*	2,593,746 *(7.37%)*
Most Urban	2,169,078 *(98.13%)*	32,583,189 *(92.63%)*

$\chi^2 = 96076$, $p < 0.0001$

(b) Urban-Rural Divide (*continuum code*)

County Type	Female Edits	Male Edits
Most Rural	21,743 *(1.03%)*	1,878,232 *(5.86%)*
Most Urban	2,097,586 *(98.97%)*	30,160,817 *(94.14%)*

$\chi^2 = 88517$, $p < 0.0001$

(c) Racial/Ethnic Diversity (*%WnHL*)

County Type	Female Edits	Male Edits
Least diverse	71,144 *(3.67%)*	2,875,634 *(12.59%)*
Most diverse	1,868,002 *(96.33%)*	19,956,337 *(87.41%)*

$\chi^2 = 135868$, $p < 0.0001$

(d) Socio-Economic Status (*MHI*)

County Type	Female Edits	Male Edits
Poorest	428,784 *(18.66%)*	2,934,839 *(10.25%)*
Wealthiest	1,869,162 *(81.34%)*	25,710,048 *(89.75%)*

$\chi^2 = 155434$, $p < 0.0001$

exhibit a substantial decrease of 73% in rural edits (a shortfall of 114,415 edits) relative to what would be expected if there were no difference in the proportion of edits produced by gender. Similar magnitudes exist for female edits in rural counties based on *rural-urban continuum* (-82%) and in the least diverse counties (-69%); and a similarly large increase exists in the poorest counties (+72%).

Another way to look at the data is from a *contributor-centric* perspective to understand the editing patterns of a typical male or female editor in our sample. We found that males are less likely to contribute to urban regions vs. rural regions than their female counterparts (for the interaction, Wald $\chi^2 = 7.92, p = 0.0049$ in terms of *%pop-urban* and Wald $\chi^2 = 11.31, p = 0.0008$ in terms of *rural-urban continuum code*). These are large effects as indicated by the incidence rate ratio values [77]: For *%pop-urban*, a male editor produces only 0.81 times what a female produces in urban regions, but produces 4.17 times that in rural regions, 95% CI [1.56, 11.14]. (For *rural-urban continuum* these values are 0.78, 6.01 and 95% CI [2.14, 16.90], respectively).

In regards to racial/ethnic diversity, men concentrate a lower proportion of their edits in the most diverse counties vs. the least diverse counties, compared to women (for the interaction, Wald $\chi^2 = 4.79, p = 0.0286$). This is also a medium to large effect: A male produces only 0.51 times what a female produces in the most diverse counties, but produces nearly 3.78 times that in the least diverse counties, 95% CI [1.18, 12.17]. However—unlike with the rural-urban divide and racial diversity—we did not see evidence that female and male editors differed in how they concentrated their edits in the wealthiest and poorest counties (for the interaction, Wald $\chi^2 = 1.68, p = 0.1955$).

With respect to this paper's central area of inquiry, these findings provide critical evidence supporting a substantial and significant relationship between OSM participation disparities and OSM content disparities. It is also interesting to note that the gendered differences we observed both cut across and align with existing known biases in OSM. Women disproportionately contribute to areas with greater racial and ethnic diversity and poorer areas (contribution-centric only) compared to their male counterparts, with women likely counteracting a bias that has been observed in OSM. However, the reverse is true in the case of the urban/rural spectrum. Of course, interpreting these intersectional results is complex and must be done with caution - a point we return to in the discussion section.

Edits Made with Bots. The above statistics describe the results for our *no-bots* dataset. Examining the results for our *with-bots* dataset, we see similar patterns but at different quantitative scales. For example, the number of counties with female edits in the *with-bots* dataset (1,469) increases by 61% in comparison to the *no-bots* dataset. At the same time, however, the number of counties with a higher ratio of female edits to male edits in the *with-bots* dataset is reduced from 72 (in the *no-bots* dataset) to 9. Furthermore, male editors produced a higher proportion of bot-based contributions than female editors: the number of edits by men in the *with-bots* dataset is 9.51 times as high as in the *no-bots* dataset, while for women it is only 3.55 times as high. Together these results suggest that the male influence on OSM content further increases when we consider bot activity.

We also see differences in the way women and men used bots to edit different *types of regions*. Similar to [55], we see that bots are used extensively in rural areas by both men and women. However, male editors appear to make greater use of bots to map rural, poor and less racially and ethnically diverse regions where increases are ~20 times (compared to growth of 6-7 times for the most urban, wealthy and diverse regions). While female edits exhibit similar growth for rural areas (~24 times), the growth is lower for poor and less diverse regions (~2-6 times). Due to this smaller increase in female bot-based edits, we see a shift in contribution patterns in poor regions using the *with-bots* dataset compared to using the *no-bots* dataset. Without bots, the poorest counties received a higher proportion of female edits (18.66%) compared to that of male edits (10.25%) (see Table 1d). However, when bots are included, this reversed and those same regions received a lower proportion of female edits (13.85%) compared to male edits (30.35%) ($p < 0.0001$).

RQ2: *What* are the contributions?

Our results for RQ1 revealed clear differences in the content generated by male and female editors as well as their individual editing behavior when it comes to *where* contributions are occurring. The goal of our second research question was to determine if there are similar differences in *what* editors are mapping. That is, regardless of whether the geographic context is Wyoming or Washington, D.C., do men and women tend to add different types of entities (e.g. nightclubs, restaurants, or childcare centers)? Furthermore, if men and women contribute different content, do they do so in a way that aligns with a gender-based self-focus bias?

To investigate these questions, we examined whether there were editing differences in gendered space categories that prior work defined as *feminized*, *masculinized* and *non-gendered*. Recall from Section 4 that we have two datasets: (1) a *narrow* focus dataset drawing on Stephens [92]; and (2) a *broad* focus dataset based on generalizations of gendered spaces that have been identified in the critical geography literature [65, 92]. In what follows, we first focus on the *broad* dataset and then report parallel results from the *narrow* dataset.

We begin by taking a *contribution-centric* perspective. We observe clear differences in the proportion of edits produced in the gendered spaces by the men and women in our sample. However, our findings refute simple self-focus based assumptions that suggest that *masculinized* and *feminized spaces* are likely to receive more contributions from the editors whose gender identity align with the spaces [65, 92].

Looking only at edits made in the gendered spaces, we find that 85.90% of male contributions involved *feminized spaces*, while only 68.18% of female contributions involved those same types of spaces. Alternatively, *masculinized spaces* received 31.82% of female edits, but only 14.10% of male edits. Stated another way, males disproportionately edited *feminized spaces* in comparison to females, and vice-versa (see Table 2a; $p < 0.001$). For the *narrow* dataset, we find the same pattern of results; however, the disproportion is even greater (see Table 2b; $p < 0.001$). In terms of simple effect sizes, for the *broad* dataset, females produced 21% less edits in *feminized spaces* and 117% more edits in *masculinized spaces* than would be expected given the null hypothesis (for the *narrow* dataset, these numbers are 32% less and 104% more, respectively).

While the standardized effect sizes (e.g., Cramer's V) for the overall proportional differences are small, it is important to note that the effects are not small from the practical perspective of female editor's edits. Furthermore, they are significant and in the *opposite direction* of the self-focus bias

Table 2: Male and Female Edits in Gendered Spaces

(a) The *Broad* Focus Dataset

Type	Female Edits	Male Edits
Feminized	60 *(68.18%)*	5025 *(85.90%)*
Masculinized	28 *(31.82%)*	825 *(14.10%)*

$\chi^2 = 22.12$, $p < 0.0001$

(b) The *Narrow* Focus Dataset

Type	Female Edits	Male Edits
Feminized	30 *(53.57%)*	2780 *(78.07%)*
Masculinized	26 *(46.43%)*	781 *(21.93%)*

$\chi^2 = 17.70$, $p < 0.0001$

assumption in the literature [46, 65, 92]. This provides a theoretically important new data point in our understanding of the critical relationship between participation and content disparities.

A closer look at the data reveals that the top *gendered* spaces mapped by female editors are (raw counts in parentheses): nightclub (18), childcare (16), kindergarten (14), nursing home (14), and group home (8). The top *gendered* spaces mapped by male editors are: kindergarten (2,388), nursing home (1,564), nightclub (634), childcare (351), group home (281), assisted living (168), and stripclub (141). In the *narrow* dataset, the top *gendered* spaces mapped by female and male editors are kindergarten, nightclub, childcare, etc.

Our results from the *contributor-centric* angle bolstered our conclusions from the *contribution-centric* analyses. We observed a trend suggesting differences in the way male and female editors map *feminized*, *masculinized* and *non-gendered* spaces (for the interaction, Wald $\chi^2(2) = 5.11, p < 0.0774$). Specifically, we saw that women editors were more likely to contribute a higher proportion of information about *masculinized spaces* relative to men, and men were more likely to contribute a higher proportion of information about *feminized spaces* relative to women (Wald $\chi^2 = 3.64, p = 0.0563$). These are large effects as indicated by the incidence rate ratios: A male editor produces 1.56 times what a female produces in the *masculinized spaces*, but produces over 2.84 times that in the *feminized spaces*, 95% CI [0.97, 8.30].

In the *narrow* dataset, we observed a similar trend in the way male and female editors map *feminized*, *masculinized* and *non-gendered* spaces (for the interaction, Wald $\chi^2(2) = 5.22, p < 0.0737$); and we saw that women were more likely to contribute a higher proportion of information about *masculinized spaces* relative to men, and men were more likely to contribute a higher proportion of information about *feminized spaces* relative to women (Wald $\chi^2 = 4.00, p = 0.046$). These effects are similarly large: A male editor produces 1.59 times what a female produces in *masculinized spaces*, but produces nearly 3.08 times that in the *feminized spaces*, 95% CI [1.02, 9.31].

In summary, like our RQ1 results, our RQ2 results show that men and women edit differently, providing more evidence that participation disparities result in content disparities. However, our RQ2 results confound prior assumptions about gender-based self-focus bias and a *direct* relationship between gender participation disparities and content disparities. We see that on a proportional basis, the men in our sample produced a higher proportion of their contributions in the *feminized spaces* compared to women, while the *masculinized spaces* received a higher proportion of their contributions from women compared to men.

7 DISCUSSION

This study set out to examine the assumed relationship between gender-based participation disparities and resulting content disparities. As a step toward answering this question, our results revealed clear differences in male and female editing behavior when it comes to *where* and *what* they are mapping. The results of our analysis have important theoretical implications for understanding the complex relationship between gender participation disparities and associated content disparities, and practical implications for the sociotechnical design of peer production communities.

Theoretical Implications

Complexities of Gender-based Self-focus Bias. The notion of self-focus bias suggests that contributors predominantly add information that caters to the interests of the cultural groups that are prominent in a given peer production community [46]. Strong evidence of self-focus bias has been observed in terms of the localness of geographic contributions [28, 42, 46, 47], politics [52], language-defined cultural groups [46, 47], and others [21, 80]. Critical and feminist GIS literature [65, 68, 92] has suggested that self-focus bias may also exist along gender dimensions. In OSM, this would mean that men would be proportionally more likely to edit *masculinized spaces* and women would be more likely to edit *feminized spaces*.

Our findings depict a different picture in which men tend to contribute more to *feminized spaces* relative to women and women tend to contribute more to *masculinized spaces* relative to men. It is important to think about the potential reasons that female editors might map a lower proportion of the time in *feminized spaces* compared to men and vice versa. One possible explanation is that editors in our dataset have personal interests, hobbies, or skills that do not align with their gender identities according to prevalent gender norms, and thus their interests do not fall within the gendered space categories defined by prior studies and used in our analyses.

Another reason behind the apparent absence of gender-based self-focus bias may lie in the specific knowledge requirements needed to map entities in OSM. In contrast to other peer production communities that show a strong influence of self-focus (e.g., Wikipedia), OSM mapping typically requires less individual knowledge about the entity being contributed. To map a place in OSM, an editor typically needs to add a spatial element (e.g., a node, way or relation) in the appropriate location and specify relevant information (e.g., name, address, etc.). Collecting this information can be relatively easy and low-effort for an editor even if they are not particularly familiar with the entity. For example, a woman can map a barbershop with only cursory knowledge about the place (e.g., location, name, etc.) even if she never visits the shop. However, to write a detailed article about that same place on Wikipedia requires the editor to have extensive knowledge about the place and thus, much greater effort is needed. Consequently, we may see a stronger impact of self-focus bias on platforms where users tend toward contributing rich content catered to their interests in comparison to lower cost or more "opportunistic" peer production activities such as those more often found on OSM.

Consideration of the dimensions of interest and contribution effort as they relate to peer production leaves room for interesting discussions around the role of the self-focus concept. Would gender-based self-focus bias be more apparent if OSM mapping involved more detailed information about an entity? For example, mapping the interior spaces of a women's prayer center or detailing the types of activity stations and services that reside within a childcare center? In such cases, would the increased cost to contribute to a *feminized space* (or vice versa) result in greater gender-based self-focus bias? These are not unrealistic future scenarios; indoor mapping is widely recognized to be an important frontier of spatial data collection [9, 12], and the OSM community has embarked on early indoor mapping efforts [10, 13]. Future work should explore these more complex and multidimensional dynamics of self-focus bias that could shape mapping behaviors in OSM in important ways.

Intersectional Dynamics with Respect to Gender and Demographic Factors. Prior studies have shown that rural regions are under-represented in OSM [55]. We found that the women in our dataset of highly active editors tend to designate a greater proportion of their edits toward urban areas compared to men. This suggests that female editors' mapping tendencies may exacerbate biases in OSM in terms of urban-rural divide to a greater extent than that of male editors. Conversely (and correspondingly), our results suggest that male editors spend less effort in more diverse areas and poorer areas.

It is likely that nuanced intersectional dynamics are at play in the above results, in which multiple facets of identity more accurately define editing interests. A female editor may not be mapping OSM as only a female, rather there might be other important facets of her identity that are missed by looking exclusively at gender. It is possible that she is mapping as a 'local' editor belonging to an urban place, as a hiker interested in conservation spaces, or as politically conservative activist – in other words, other non-gender aspects of identity may play a more important role. An interesting avenue for further research may focus on understanding the multitudes of ways a person identifies and how intersectional dynamics are incorporated into their activities on peer production platforms like OSM. What is self-focus bias in an intersectional world?

Practical Implications

Recruiting Male Contributors as Allies. Existing research has situated male and female OSM editors in a position in which we might expect a misalignment between their interest space. If this were true, a solution to any gender-based content disparity might be easy: attract more female editors. As our results show - this simply is not the case. Instead, our findings reveal that male users are cognizant of at least some of the feminized spaces, and they actively map those facilities. This contradicts the way many prior researchers have been formulating thoughts and discussion of potential improvements to the amount and quality of *feminized spaces* in OSM [51, 65, 68, 92]. In our dataset, female editors tended to map *feminized places* to a lesser extent than their male counterparts. As these spaces characterize important facilities for feminine health and nurturing of others, proper representation is necessary. However, our results point out that a straightforward solution like increasing female participation may not ensure increased representation of *feminized spaces*. We caution that our results should not be interpreted in a way that discourages higher levels of female participation. Rather we need to think critically about ways to increase coverage of under-represented facilities on OSM. One possible approach is the recruitment of male editors as "allies" along with more female participants and informing all editors of the state of the repository. Another solution may be to take the "SuggestBot" approach [25] and design a content recommendation system that will seek contributors based on location, interests, skills, etc. For example, a local person who is probably aware of nearby childcare centers or maternity clinics may be asked to map those places irrespective of their gender.

Bot Activities to Reduce Content Disparities. The men and women in our sample differed in how they made use of bots to support their editing. A key top line result is that male editors created more edits with bots than female editors, suggesting that bots extend male editors' influence beyond that which exists in more traditional manual editing. Exploring this phenomenon in more detail—and better supporting female editors' use of bots—is clearly an important area of future work. Additionally, we identified that male editors' use of bots increased disproportionately in rural, poor and less diverse areas, while female editors' use of bots only saw disproportionate increase in rural areas. In prior research, bot edits have been shown to play an important role in mitigating existing systemic biases in under-represented regions [55]. In this regard, future studies should explore solutions to encourage both genders to gear their bot activities towards all under-represented areas in OSM.

8 LIMITATIONS AND FUTURE WORK

An important limitation of our work is that we only consider male and female genders, and we apply a binary classification of gender, which does not allow us to capture other variations of gender identity. This is a key missing dimension from our research that should be addressed in future studies. Another related aspect is that our work relies on the validity of our technique for gender inference, which rests upon an assumption that OSM editors honestly portray their offline identities in online profiles. Research has shown that women are less likely to disclose sensitive information compared to men [27], and therefore, we may have been unable to identify women editors who have greater interest in *feminized spaces*. Also, albeit unlikely given that female editorship in OSM has been found to be in the 3-4% range by prior surveys [20, 87, 88] (a number close to the 5.16% found in our gender-identified editors), the unidentified editors (43.50% among the top 2000) might contain a higher proportion of females. At the same time, it may so happen that the identified editors in our dataset tend to contribute more about spaces associated with the opposite gender to compensate for their visible gender identities. A potential solution to these problems may involve qualitative interviews and communication with a set of OSM editors while analyzing their mapping activities in a mixed-method approach.

A second potential limitation is that the results presented in this study are based on the contributions of highly active OSM editors. As such, we are not aware of the ways these results may or may not align with the mapping behavior of less active or infrequent editors (i.e., the 'long tail'). Furthermore, we limited our attention in this work to only one country—specifically, the U.S.—and thus, are unable to capture the similarities and differences of our findings across other cultures and geographic contexts [14]. Future work needs to take into account the contributions of editors from different backgrounds and with varying levels of activity to better understand the association between editors' gender and editing behavior across different domains.

Following prior literature [55, 69, 83], we did not distinguish between different types of edits such as addition or modification of entities. Understanding whether and how male and female editors variably focus on different types of edits can be an interesting future direction of research. Also, future work might investigate the coverage and quality of gendered spaces mapped in OSM against a ground truth dataset to further explore the extent of gender content disparities.

Researcher Self-disclosure [16]: Our research team contains a range of gender, race, age and national identities. Data collection, pruning and processing steps—including gender inference—were led by female members of the team.

9 CONCLUSION

Focusing on OSM, this paper investigates the relationship between participation and content disparities along gender dimensions. Our results reveal that there is a substantial gender gap in participation among highly active OSM editors, but we do not see evidence of gender-based self-focus bias in their contributions. Specifically, we observe that men tend to contribute more about *feminized spaces* relative to women and women tend to contribute more about *masculinized spaces* relative to men. In addition, it appears that in comparison to male editors, female editors tend to contribute more to the prevailing content biases in terms of urban-rural divide. We hope that our findings will encourage further investigation of self-focus bias and its implications for peer-produced content, as well as the development of strategies and interventions to address the identified problems.

ACKNOWLEDGMENTS

This work was supported in part by NSF grants 1815507, 1707296, and 1707319. We would like to thank Dana Choi and Oliver Baldwin for their help in the gender inference procedure. We would also like to thank Isaac Johnson for providing resources on OSM data processing and insightful feedback on the paper.

REFERENCES

[1] 2010. U.S. Census: Urban-rural. http://www.census.gov/geo/reference/ua/urban-rural-2010.html
[2] 2013. U.S. Department of Agriculture: Rural-Urban Continuum Code. https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx
[3] Retrieved September 20, 2018. http://download.geofabrik.de/north-america.html.
[4] Retrieved September 20, 2018. https://planet.openstreetmap.org/.
[5] Retrieved September 20, 2018. https://wiki.openstreetmap.org/wiki/Key:amenity.
[6] Retrieved September 20, 2018. https://wiki.openstreetmap.org/wiki/Category:Users_in_the_United_States
[7] Retrieved September 20, 2018. https://help.openstreetmap.org/users/
[8] Retrieved September 20, 2018. https://gender-api.com/en/
[9] Retrieved September 20, 2018. https://www.google.com/maps/about/partners/indoormaps/
[10] Retrieved September 20, 2018. https://wiki.openstreetmap.org/wiki/Indoor_Mapping#active_indoor_maps
[11] Retrieved September 20, 2018. American FactFinder. https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml.
[12] Retrieved September 20, 2018. Esri Announces New Indoor Mapping Product. http://www.esri.com/esri-news/releases/18-3qtr/esri-announces-new-indoor-mapping-product
[13] Retrieved September 20, 2018. OSM goes indoors: Usages, tools and prospects. http://www.esri.com/esri-news/releases/18-3qtr/esri-announces-new-indoor-mapping-product
[14] Judd Antin, Ed H. Chi, James Howison, Sharoda Paul, Aaron Shaw, and Jude Yew. 2011. Apples to Oranges?: Comparing Across Studies of Open Collaboration/Peer Production. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*. ACM, 227–228. https://doi.org/10.1145/2038558.2038610
[15] 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), Mountain View, CA, USA*. 11–14. https://doi.org/10.1145/2038558.2038561
[16] Shaowen Bardzell and Jeffrey Bardzell. 2011. Towards a feminist HCI methodology: social science, feminism, and HCI. In *Proceedings of the ACM International Conference on Human Factors in Computing Systems, (CHI'11), Vancouver, BC*. 675–684. https://doi.org/10.1145/1978942.1979041
[17] Julia B. Bear and Benjamin Collier. 2016. Where are the women in Wikipedia? Understanding the different psychological experiences of men and women in Wikipedia. *Sex Roles* 74, 5–6 (January 2016), 254–265. https://doi.org/10.1007/s11199-015-0573-y
[18] Jonathan Bright, Stefano De Sabbata, and Sumin Lee. 2017. Geodemographic biases in crowdsourced knowledge websites: do neighbours fill in the blanks? *GeoJournal* (2017). https://doi.org/10.1007/s10708-017-9778-7
[19] Felicia M. Brown. 2017. *Women's sexuality in swinging*. Ph.D. Dissertation. Middle Tennessee State University, Tennessee, USA.
[20] Nama R. Budhathoki. 2010. *Participants' motivations to contribute geographic information in an online community*. Ph.D. Dissertation. University of Illinois at Urbana-Champaign, Urbana, Illinois.
[21] Ewa Callahan and Susan C. Herring. 2011. Cultural bias in Wikipedia content on famous persons. *JASIST* 62, 10 (2011), 1899–1915. https://doi.org/10.1002/asi.21577
[22] Le Chen, Ruijun Ma, Anikó Hannák, and Christo Wilson. 2018. Investigating the Impact of Gender on Rank in Resume Search Engines. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, 651:1–651:14. https://doi.org/10.1145/3173574.3174225
[23] Herbert H. Clark. 1996. *Using Language*. Cambridge University Press.
[24] 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 Conference on Computer Supported Cooperative Work (CSCW '12), Seattle, WA, USA*. 383–392. https://doi.org/10.1145/2145204.2145265
[25] Dan Cosley, Dan Frankowski, Loren G. Terveen, and John Riedl. 2007. SuggestBot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces, (IUI'07), Honolulu, Hawaii, USA*. 32–41. https://doi.org/10.1145/1216295.1216309
[26] Martin Dittus, Giovanni Quattrone, and Licia Capra. 2017. Mass Participation During Emergency Response: Event-centric Crowdsourcing in Humanitarian Mapping. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing, (CSCW'17), Portland, OR, USA*. 1290–1303. http://dl.acm.org/citation.cfm?id=2998216

[27] Lauren E. Durant, Michael P. Carey, and Kerstin E. E. Schroder. 2002. Effects of Anonymity, Gender, and Erotophilia on the Quality of Data Obtained From Self-Reports of Socially Sensitive Behaviors. *Journal of Behavioral Medicine* 25, 5 (2002), 439–467.

[28] Melanie Eckle and João Porto de Albuquerque. 2015. Quality assessment of remote mapping in OpenStreetMap for disaster management purposes. In *Proceedings of the 12th Proceedings of the International Conference on Information Systems for Crisis Response and Management (ISCRM'15), Krystiansand, Norway.* http://idl.iscram.org/files/melanieeckle/2015/1218_MelanieEckle+JoaoPortodeAlbuquerque2015.pdf

[29] R. Danielle Egan and Katherine Frank. 2005. Attempts at a feminist and interdisciplinary conversation about strip clubs. *Deviant Behavior* 26 (2005), 297 – 320. https://doi.org/10.1080/016396290931722

[30] Sarah Elwood. 2008. Volunteered geographic information: Future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal* 72 (July 2008), 173–183. https://doi.org/10.1007/s10708-008-9186-0

[31] Eureka Foong, Nicholas Vincent, Elizabeth Gerber, and Brent Hecht. 2018. Women (Still) Ask For Less: Gender Differences in Wage-Setting and Occupation in an Online Labor Marketplace. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing, (CSCW'18).* ACM.

[32] Killian Fox. 2012. OpenStreetMap: 'It's the Wikipedia of maps'. The Guardian. Retrieved September 09, 2018 from https://www.theguardian.com/theobserver/2012/feb/18/openstreetmap-world-map-radicals

[33] Katherine Frank. 2003. "Just trying to relax": masculinity, masculinizing practices, and strip club regulars. *The Journal of Sex Research* 40, 1 (February 2003), 61 – 75.

[34] R. Stuart Geiger. 2014. Bots, bespoke, code and the materiality of software platforms. *Information, Communication & Society* 17, 3 (2014), 342–356. https://doi.org/10.1080/1369118X.2013.873069

[35] R. Stuart Geiger and Aaron Halfaker. 2013. Gender differences in Wikipedia editing. In *Proceedings of the 9th International Symposium on Wikis and Open Collaboration (WikiSym '13), Hong Kong, China.* https://doi.org/10.1145/2038558.2038561

[36] Ruediger Glott, Philipp Schmidt, and Rishab Ghosh. 2010. *Wikipedia survey- Overview of results.* Technical Report. United Nations University - MERIT, Maastricht, Netherlands.

[37] 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), Guzelyurt, TRNC, Cyprus.* 165–174. https://doi.org/10.1145/2700171.2791036

[38] Mark Graham, Matthew Zook, and Andrew Boulton. 2013. Augmented reality in urban places: contested content and the duplicity of code. *Transactions of the Institute of British Geographers* 38, 3 (2013), 464–479. http://doi.org/10.1111/j.1475-5661.2012.00539.x

[39] Muki Haklay. 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design* (2010), 682 – 703. https://doi.org/10.1068/b35097

[40] Muki Haklay. 2012. Nobody wants to do council estates' - digital divide, spatial justice and outliers. In *Annual Meeting of the Association of American Geographers (AAG '12).* https://www.slideshare.net/mukih/nobody-wants-to-do-council-estates-digital-divide-spatial-justice-and-outliers?ref=http://povesham.wordpress.com/.

[41] Sidsel Kirstine Harder and Jakob Demant. 2015. Failing Masculinity at the Club: A Poststructural Alternative to Intoxication Feminism. *Substance Use and Misuse* 50 (2015), 759 – 767. https://doi.org/10.3109/10826084.2015.978617

[42] Darren Hardy, James Frew, and Michael F. Goodchild. 2012. Volunteered geographic information production as a spatial process. *International Journal of Geographical Information Science* 26, 7 (2012), 1191–1212. https://doi.org/10.1080/13658816.2011.629618

[43] 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 and Society* 18, 4 (2015), 424–442. https://doi.org/10.1080/1369118X.2014.957711

[44] Steven A. Harrison. 2005. *"Trim Off the Top and Close on the Sides": America's Disappearing Legacy of Barbershops.* Ph.D. Dissertation. Utah State University, Logan, Utah.

[45] Brent J. Hecht. 2013. *The mining and application of diverse cultural perspectives in user-generated content.* Ph.D. Dissertation. Northwestern University, Evanston, Illinois.

[46] Brent J. Hecht and Darren Gergle. 2009. Measuring self-focus bias in community-maintained knowledge repositories. In *Proceedings of the Fourth International Conference on Communities and Technologies, (C&T'09), University Park, PA, USA.* 11–20. https://doi.org/10.1145/1556460.1556463

[47] Brent J. Hecht and Darren Gergle. 2010. On the "localness" of user-generated content. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, (CSCW'10), Savannah, Georgia, USA.* 229–232. https://doi.org/10.1145/1718918.1718962

[48] Brent J. Hecht and Darren Gergle. 2010. The tower of Babel meets web 2.0: user-generated content and its applications in a multilingual context. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems, (CHI'10), Atlanta, Georgia, USA.* 291–300. https://doi.org/10.1145/1753326.1753370

[49] Brent J. Hecht and Monica Stephens. 2014. A tale of cities: Urban biases in volunteered geographic information. In *Proceedings of the Eighth International Conference on Weblogs and Social Media, (ICWSM'14), Ann Arbor, Michigan, USA.* http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8114

[50] Benjamin M. Hill and Aaron Shaw. 2013. The Wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PLoS ONE* 8, 6, Article e65782 (June 2013). https://doi.org/10.1371/journal.pone.0065782

[51] Sarah Holder. 2018. Why equitable maps are more accurate and humane. https://www.citylab.com/equity/2018/03/who-maps-the-world/555272/

[52] Christoph Hube. 2017. Bias in Wikipedia. In *Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia.* 717–721. https://doi.org/10.1145/3041021.3053375

[53] Mohsen Jadidi, Fariba Karimi, and Claudia Wagner. 2017. Gender disparities in science? Dropout, productivity, collaborations and success of male and female computer scientists. *CoRR* abs/1704.05801 (2017). http://arxiv.org/abs/1704.05801

[54] Sheila Jeffreys. 2008. Keeping Women Down and Out: The Strip Club Boom and the Reinforcement of Male Dominance. *Signs: Journal of Women in Culture and Society* 34, 1 (2008), 151–173. https://doi.org/10.1086/588501

[55] Isaac L. Johnson, Yilun Lin, Toby Jia-Jun Li, Andrew Hall, Aaron Halfaker, Johannes Schöning, and Brent J. Hecht. 2016. Not at home on the range: Peer production and the urban/rural divide. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'15), San Jose, CA, USA.* 13–25. https://doi.org/10.1145/2858036.2858123

[56] Isaac L. Johnson, Connor McMahon, Johannes Schöning, and Brent J. Hecht. 2017. The effect of population and "structural" biases on social media-based algorithms - A case study in geolocation inference across the urban-rural spectrum. In *Proceedings of the ACM Conference on*

Human Factors in Computing Systems (CHI'17), Denver, CO, USA. 1167–1178. https://doi.org/10.1145/3025453.3026015
[57] 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 on World Wide Web (WWW'16), Montreal, Canada*. 53–54. https://doi.org/10.1145/2872518.2889385
[58] Aniket Kittur and Robert E. Kraut. 2008. Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, 37–46. https://doi.org/10.1145/1460563.1460572
[59] Aniket Kittur, Bryan A. Pendleton, Bongwon Suh, and Todd Mytkowicz. 2007. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM.
[60] Maximilian Klein, Harsh Gupta, Vivek Rai, Piotr Konieczny, and Haiyi Zhu. 2016. Monitoring the gender gap with Wikidata Human Gender Indicators. In *Proceedings of the 12th International Symposium on Open Collaboration (OpenSym'16), Berlin, Germany*. Article 16. https://doi.org/10.1145/2957792.2957798
[61] Shyong K. Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R. Musicant, Loren G. 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), Mountain View, CA, USA*. 1–10. https://doi.org/10.1145/2038558.2038560
[62] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174. https://doi.org/10.2307/2529310
[63] David Laniado, Andreas Kaltenbrunner, Carlos Castillo, and Mayo Fuster Morell. 2012. Emotions and dialogue in a peer-production community: The case of Wikipedia. In *Proceedings of the 8th Annual International Symposium on Wikis and Open Collaboration (WikiSym '12), Austria*. Article 9. https://doi.org/10.1145/2462932.2462944
[64] Victoria Lawson. 2007. Geographies of Care and Responsibility. *Annals of the Association of American Geographers* 97, 1 (2007), 1–11. https://doi.org/10.1111/j.1467-8306.2007.00520.x
[65] Agnieszka Leszczynski and Sarah Elwood. 2015. Feminist geographies of new spatial media. *The Canadian Geographer / Le Geographe Canadien* 59 (2015), 12–28. https://doi.org/10.1111/cag.12093
[66] Bin Lin and Alexander Serebrenik. 2016. Recognizing Gender of Stack Overflow Users. In *Proceedings of the 13th International Conference on Mining Software Repositories (MSR '16)*. ACM, 425–429. https://doi.org/10.1145/2901739.2901777
[67] Yuwei Lin and Manuela Schmidt. 2014. A gender-informed curriculum for teaching volunteered geographic information. In *the 8th European Conference on Gender Equality in Higher Education.*
[68] Noemi De Luca, Nicole Ferber, Helena Atteneder, and Thomas Jekel. 2015. Feminist and queer approaches to education for spatial citizenship. *GI_Forum - Journal for Geographic Information Science* (2015), 272–282. https://doi.org/10.1553/giscience2015s272
[69] Afra J. Mashhadi, Giovanni Quattrone, and Licia Capra. 2013. Putting ubiquitous crowd-sourcing into context. In *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work & Social Computing, (CSCW'13), San Antonio, TX, USA*. 611–622. https://doi.org/10.1145/2441776.2441845
[70] Connor McMahon, Isaac Johnson, and Brent Hecht. 2017. The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM '17)*. AAAI.
[71] Olena Medelyan, David N. Milne, Catherine Legg, and Ian H. Witten. 2009. Mining meaning from Wikipedia. *International Journal of Human-Computer Studies* 67, 9 (2009), 716–754. https://doi.org/10.1016/j.ijhcs.2009.05.004
[72] Amanda Menking and Ingrid Erickson. 2015. The heart work of Wikipedia: Gendered, emotional labor in the world's largest online encyclopedia. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, (CHI '15), Seoul, Republic of Korea*. 207–210. https://doi.org/10.1145/2702123.2702514
[73] Christopher Mims. 2011. 'Wikipedia of Maps' Challenges Google. MIT Technology Review. Retrieved September 09, 2018 from https://www.technologyreview.com/s/426481/wikipedia-of-maps-challenges-google/
[74] Elizabeth A. Morton. 2010. *Completely straight, most of the time: An inquiry into the sexuality of male swingers*. Ph.D. Dissertation. California State University, Fullerton, California, USA.
[75] Michelle Murphy. 2012. *Seizing the Means of Reproduction: Entanglements of Feminism, Health, and Technoscience*. Duke University Press.
[76] Pascal Neis and Dennis Zielstra. 2014. Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet* 6, 1 (2014), 76–106. https://doi.org/10.3390/fi6010076
[77] Jake Olivier, Warren L. May, and Melanie L. Bell. 2017. Relative effect sizes for measures of risk. *Communications in Statistics - Theory and Methods* 46, 14 (2017), 6774–6781. https://doi.org/10.1080/03610926.2015.1134575
[78] Jahna Otterbacher. 2015. Linguistic Bias in Collaboratively Produced Biographies: Crowdsourcing Social Stereotypes?. In *Proceedings of the Ninth International Conference on Web and Social Media, (ICWSM'15), Oxford, UK*. 298–307. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10539
[79] Katherine Panciera, Aaron Halfaker, and Loren Terveen. 2009. Wikipedians Are Born, Not Made: A Study of Power Editors on Wikipedia. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work (GROUP '09)*. ACM, 51–60. https://doi.org/10.1145/1531674.1531682
[80] Ulrike Pfeil, Panayiotis Zaphiris, and Chee Siang Ang. 2006. Cultural differences in collaborative authoring of Wikipedia. *Journal of Computer-Mediated Communication* 12, 1 (2006), 88–113. https://doi.org/10.1111/j.1083-6101.2006.00316.x
[81] Geraldine Pratt. 2003. Valuing Childcare: Troubles in Suburbia. *Antipode* 35, 3 (2003), 581–602. https://doi.org/10.1111/1467-8330.00340
[82] Ioannis Protonotarios, Vasiliki Sarimpei, and Jahna Otterbacher. 2016. Similar gaps, different origins? Women readers and editors at Greek Wikipedia. In *Workshops of the 10th International AAAI Conference on Web and Social Media (ICWSM '16), Cologne, Germany*. http://aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13192
[83] Giovanni Quattrone, Licia Capra, and Pasquale De Meo. 2015. There's no such thing as the perfect map: Quantifying bias in spatial crowd-sourcing datasets. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, (CSCW'15), Vancouver, BC, Canada*. 1021–1032. https://doi.org/10.1145/2675133.2675235
[84] Debra Satz. 2017. Feminist Perspectives on Reproduction and the Family. In *The Stanford Encyclopedia of Philosophy* (summer 2017 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.
[85] Dennis M. Savard, Thomas M. Kelley, and David M. Merolla. 2017. Routine Activities and Criminal Victimization: The Significance of Gendered Spaces. *Journal of Interpersonal Violence* (July 2017). https:

//doi.org/10.1177/0886260517721170

[86] Jennifer Scanlon. 2007. 'If My Husband Calls I'm Not Here': The Beauty Parlor as Real and Representational Female Space. *Feminist Studies* 33, 2 (2007), 308–334. https://www.jstor.org/stable/20459141

[87] Manuela Schmidt and Silvia Klettner. 2013. Gender and experience-related motivators for contributing to OpenStreetMap. In *International Workshop on Action and Interaction in Volunteered Geographic Information (ACTIVITY), Leuven*. 13–18.

[88] Manuela Schmidt, Silvia Klettner, and Renate Steinmann. 2013. Barriers for contributing to VGI projects. In *the 26th International Cartographic Conference*.

[89] Shilad W. Sen, Heather Ford, David R. Musicant, Mark Graham, Oliver S.B. Keyes, and Brent Hecht. 2015. Barriers to the Localness of Volunteered Geographic Information. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, 197–206. https://doi.org/10.1145/2702123.2702170

[90] Aaron Shaw and Eszter Hargittai. 2018. The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing. *Journal of Communication* 68, 1 (February 2018), 143–168. https://doi.org/10.1093/joc/jqx003

[91] Renate Steinmann, Elisabeth Hausler, Silvia Klettner, Manuela Schmidt, and Yuwei Lin. 2013. Gender dimensions in UGC and VGI: A desk-based study. In *GI_Forum 2013, Creating the GISociety - Conference Proceedings*. 355–364. https://doi.org/10.1553/giscience2013s355

[92] Monica Stephens. 2013. Gender and the GeoWeb: Divisions in the production of user-generated cartographic information. *GeoJournal* 78 (August 2013), 981–996. https://doi.org/10.1007/s10708-013-9492-z

[93] Qian Hui Tan. 2014. Postfeminist possibilities: unpacking the paradoxical performances of heterosexualized femininity in club spaces. *Social and Cultural Geography* 15, 1 (2014), 23 – 48. http://dx.doi.org/10.1080/14649365.2013.860186

[94] Josh Terrell, Andrew Kofink, Justin Middleton, Clarissa Rainear, Emerson R. Murphy-Hill, Chris Parnin, and Jon Stallings. 2017. Gender differences and bias in open source: pull request acceptance of women versus men. *PeerJ Computer Science* 3, Article e111 (2017). https://doi.org/10.7717/peerj-cs.111

[95] Claudia Wagner, David Garcia, Mohsen Jadidi, and Markus Strohmaier. 2015. It's a man's Wikipedia? Assessing gender inequality in an online encyclopedia. In *Proceedings of the 9th International Conference on Web and Social Media, (ICWSM '15), Oxford, UK*. 454–463. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10585

[96] Claudia Wagner, Eduardo Graells-Garrido, and David Garcia. 2016. Women through the glass-ceiling: Gender asymmetries in Wikipedia. *EPJ Data Science* 5, 5 (2016). 10.1140/epjds/s13688-016-0066-4

[97] Michele White. 2015. *Producing women : the Internet, traditional femininity, queerness, and creativity Women flock to nail art*. Routledge, New York, USA.

[98] Olga Zagovora, Fabian Flöck, and Claudia Wagner. 2017. "(Weitergeleitet von Journalistin)": The gendered presentation of professions on Wikipedia. In *Proceedings of the 2017 ACM on Web Science Conference (WebSci '17), Troy, NY, USA*. 83–92. https://doi.org/10.1145/3091478.3091488

[99] Dennis Zielstra, Hartwig H. Hochmair, and Pascal Neis. 2013. Assessing the effect of data imports on the completeness of OpenStreetMap - A United States case study. *Transactions in GIS* 17, 3 (2013), 315–334. https://doi.org/10.1111/tgis.12037

[100] Dennis Zielstra, Hartwig H. Hochmair, Pascal Neis, and Francesco Tonini. 2014. Areal delineation of home regions from contribution and editing patterns in OpenStreetMap. *ISPRS International Journal of GeoInformation* 3, 4 (2014), 1211–1233. https://doi.org/10.3390/ijgi3041211

[101] Dennis Zielstra and Alexander Zipf. 2010. A comparative study of proprietary geo-data and volunteered geographic information for Germany. In *Proceedings of the 13th AGILE International Conference on Geographic Information Science*.