

# Citational politics: Quantifying the influence of gender on citation in *Digital Scholarship in the Humanities*

Amy E. Earhart,

Department of English, Texas A&M University, USA

Roopika Risam 

Departments of Secondary and Higher Education and English, Salem State University, USA

Matthew Bruno

Department of English, Salem State University, USA

## Abstract

Using citation analysis, we consider the role of gender in citation practices in conference special issues of *Digital Scholarship in the Humanities*. Our examination of citations in Digital Humanities conference special issues from 2006 to 2015 demonstrates gender bias in citational practices. This bias is consistent with broader trends in citational politics across the academy more broadly but is a threat to equity and justice within the scholarly community. We further offer proposals for improving citational practices to resist gender bias. Quantifying the impact of gender on citations, we argue, is one approach to understanding gender inequalities within digital humanities communities and to generating solutions to promote the broadest representation of digital humanities scholarship in scholarly communications.

## Correspondence:

Roopika Risam, 352 Lafayette Street, Salem, MA 01970, USA.

## E-mail:

rrisam@salemstate.edu

Why is nineteenth-century author Willa Cather the second most cited woman in conference issues of *Digital Scholarship in the Humanities (DSH)*? As digital humanities scholarship has expanded in volume and scope, questions of equity and inclusion have come to the fore. While there has been some attention to the composition of digital humanities conference programs and to citation networks, the influence of gender on scholarship is an area that remains particularly under-examined. Quantifying the effects of gender on citations, we argue, is one approach to understanding gender inequities within digital humanities

communities and to generating solutions to promote the broadest representation of digital humanities scholarship within scholarly communications.

Gender bias in academic citations is an endemic problem in research. While this is an area that has received little attention in the context of digital humanities, scholars have found clear evidence of gender bias among citations in the sciences (Cronin and Sugimoto, 2014; Cronin and Sugimoto, 2015, Macaluso *et al.*, 2016). Larivière *et al.*'s (2013) bibliometric study of nearly 5.5 million research papers and review articles in the Thompson Reuters Web

of Science database published between 2008 and 2012 demonstrated that ‘articles with women in dominant author positions receive fewer citations than those with men in the same positions’ (Larivière *et al.*, 2013, n.p.). Weingart and Eichmann-Kalwara’s (2017) careful analysis of abstracts submitted for the Alliance of Digital Humanities Organizations (ADHO) annual conference<sup>2</sup> has prompted the international scholarly community to consider how selections for this conference construct particular boundaries in digital humanities scholarship. Weingart and Eichmann’s work, which examines submitters’ country of origin, gender, home discipline, language, topic of presentation, and other identifiers, is part of a growing trend of citation analysis in digital humanities, which has made an important intervention in scholarly communications. For example, digital humanities has facilitated quantitative analysis of citations within humanities disciplines, illuminated its citational networks, and created workflows and tools for interpreting citations (Sula, 2012; Crymble and Flanders, 2013; Blaney and Siefrig, 2017; Nyhan and Duke-Williams, 2014; Sula and Miller, 2014; Romanello, 2016). Much of this work focuses on collaboration in digital humanities, which is an important consideration. Such analysis has a great deal to offer how we understand the confluence of citation, power, and privilege within academic communities of practice.

Scholarly analysis of citation practices within digital humanities is entwined with recent discussions of community formation, including issues of race, language, nation, method, and access. The digital humanities community fostered by ADHO, which we define as the community engaged with the annual digital humanities conference and with the conference special issues published in the journal *DSH* (formerly *Literary and Linguistic Computing* or *LLC*), has struggled to understand how digital humanities is practiced as growing numbers of scholars undertake digital humanities scholarship. Given this, it is important to examine various markers of community formation and equity within the scholarly community. Terras (2006) was among the first in digital humanities to examine such interactions. Over the last 10 years, scholars have expanded investigations of community and scholarship.<sup>3</sup> For example, Weingart and Eichmann-Kalwara (2017) and

Pino-Díaz and Fiorimonte (2018) focus on conference participation; Nyhan and Duke-Williams (2014), Gao *et al.* (2018), and de la Cruz *et al.* (2015) examine authorship networks; and Palermo (2017) and Leydesdorff and Alkim Almila (2010) consider citation patterns. This foundational work offers better understanding of community formation, yet there remains room for expanded investigation of the dominant trends in citational practice within digital humanities, particularly around gender.

Thus, there is a dire need for scholarship that offers insight into the citational politics that reinforce homogeneous scholarly practices and illuminates the way that gender, race, and nation are understood in digital humanities. As bibliometrics scholarship demonstrates, citations provide insights on the boundaries of scholarly communities (Cronin and Sugimoto, 2014; Cronin and Sugimoto, 2015). Likewise, digital humanities citational practices further reveal the contours of the community. One of the few truly international scholarly communities, digital humanities moves across nations, languages, and institutional structures. However, it is overdetermined by the influence of dominant scholarly practices (Risam, 2018). Therefore, we examine what citational practices might teach us about how digital humanists interact and how inclusive—or exclusive—digital humanities scholarship is. This line of inquiry answers Galina’s (2013) challenge for greater inclusion in digital humanities scholarship by embracing her mandate:

We have a combination of scholars who can provide important insights to do this properly. Cultural theory, postcolonial studies, feminist perspectives and other forms of critical theory can make us aware of the problem. But DHers’ capacity and willingness to build things can allow us to create projects and tools that help us to be more inclusive. (n.p.)

As such, this article presents the method, results, and conclusions of our analysis of citational politics of gender in conference special issues of the journal *DSH/LLC*. We focus on gender to draw attention to one specific measure of equity in the community. As the digital humanities community expands, we propose, we need to pay further attention to how citational practices influence scholarship. Moreover, we use these results to present recommendations for a new

politics of citation that encourages increased diversity, equity, and justice in digital humanities scholarship.

## 1 Citation Practices in Digital Humanities

In recent years, there has been emerging interest in using citation analysis to address questions of power and privilege within digital humanities. [Fiormonte \(2015\)](#) has taken up these questions in relation to multilingualism in digital humanities scholarship. He argues that Anglophone citations are overrepresented, producing a ‘monoculture’ in digital humanities that devalues scholarly contributions from languages other than English ([Fiormonte, 2015](#)). [Stutsman \(2015\)](#) has explored the distribution of digital humanities scholarship in pedagogy, suggesting that the same, narrow list of digital humanities practitioners and theorists from the USA and the UK—Steven Ramsay, Matthew Kirschenbaum, Lev Manovich, Dan Cohen, Franco Moretti, and Susan Hockey—populate syllabi. [Shrout et al.’s \(2018\)](#) DH2018 panel, ‘Global perspectives on Decolonizing Digital Pedagogy’, further confirms the overwhelming influence of Anglophone digital humanities scholarship on pedagogy in multiple cultural contexts. Such studies raise the important question of which factors shape the citational practices of the digital humanities. While [Fiormonte](#) focuses on monolingualism in scholarship and [Stutsman’s](#) and [Shrout et al.’s \(2018\)](#) contributions indirectly hint at the influence of nation on pedagogy, the effect of gender on citations in digital humanities journals remains a mystery. Attention to it is necessary, however, because of the reputational and academic currency that citations produce and the growing focus on citations as metrics for excellence in humanities scholarship. As a result, further analysis of citational practices in digital humanities holds the possibility of revealing which communities are privileged and disadvantaged by citational practices and how this correlates to other ways of conceptualizing the relationship between access, power, and knowledge in the context of digital humanities scholarship, such as geographical divides between Global North and South and representation within ADHO.

The influence of gender in digital humanities scholarship has been largely unexamined, with the exception of the analysis of gender representation on the programs of the annual digital humanities conference. In one such study, [Weingart and Eichmann-Kalwara \(2017\)](#) find that women were ‘about a third of all authors [at the international digital humanities conference] from 2004–2013’, and at DH2015, women were 35% of authors while making up 46% of attendees (p. 12). Our study examines whether similar disjunctions are apparent in citational practices in the journal *DSH/LLC*. Gender-based inequality of authorship and citations within the scholarship has very real effects on the community and its scholars. As [Pearse et al. \(2019\)](#) make clear, ‘The persistent and powerful gendered norms of authoritative knowledge . . . act to marginalise different forms of understanding in their fields’ (p. 110). When particular forms of knowledge inquiry are marginalized and individual scholars are excluded from positions of power, particularly in an academic environment that uses citational counts as markers of scholarly excellence, which affects hiring, tenure, promotion, and funding. It is critical to note, however, that gender is just one axis of identity that influences citational practices, and that further analysis is also necessary to understand the role of intersectionality ([Crenshaw, 1989](#)), such as the compounded harm of multiple axes of identity and oppression, such as race and gender. Thus, in highlighting gendered citational practices, we are also indebted to recent movements such as Cite Black Women (n.d.) and the Digital Feminist Collective (n.d.) who ask us to reconsider power and citation practices and emphasize that making transparent citational choices might shift our practices.

Despite the value of using citation analysis to explore the politics of knowledge production, citation practices have limitations built into the system. As academia moves toward citations as a metric for scholarly achievement, it is important to note that citation, as currently practiced, is a highly controversial and imperfect measure of the reach of humanities scholarship. Resistance to using citation counts as metrics of success has been rightfully growing across academia. For example, the American Association of University Professors (AAUP) in the ‘Statement on “Academic Analytics” and Research Metrics’ noted, ‘There is, however, good reason to doubt the utility

of such metrics in tenure and promotion decisions and/or in judgments affecting hiring, compensation or working conditions' (AAUP, 2016). The AAUP supports this claim with research from a 2015 study by the Higher Education Funding Council for England, indicating that resistance to simple citation counts are not simply a national concern. For humanities work, especially, citation collection is limited by a number of factors, including lack of inclusion of essay collections and books, which are unlikely to be indexed; limited time frames of collection; lack of inclusion of a large number of journals and presses; and lack of citation scraping for non-digital forms. Simply put, citation metrics, as currently determined, are inaccurate for humanities scholarship. When we add the problem of inaccurate metrics to gendered inequality within citations, we have a system that is especially unbalanced.

Though we have privileged citations as a means to tell one story of digital humanities community in this essay, we also resist a simple narrative that citations are a sole marker of scholarly success. We have both called for a broader understanding of digital humanities, with Earhart (2015) arguing for the development of multiple genealogies of digital humanities and Risam (2016) for the illumination of the knowledge diaspora of the digital humanities. A broader understanding of digital humanities necessitates a rethinking of the foundational scholarship that frames scholarly communications, which includes a careful examination of why we cite particular scholarship and how that scholarship might bias our construction of the community. Finally, we recognize that the increasing speed of scholarly communications in digital humanities, accelerated publication rates, and growth in aggregation of citations has shifted the landscape of academia. It is our hope, with this article, that digital humanities scholars recognize that citations provide one critical lens into complicated and multi-variate questions of scholarship and community.

## 2 Method

For the purposes of this study, we focused on gender identification in conference special issues of *DSH/LLC* from DH 2006 to DH 2015 (published between 2008 and 2017). Recognizing the importance of accurate

citations, we were compelled to create a hand-coded dataset of citations for analysis by researching how article authors self-identify. We supplemented this data by researching the affiliations of scholars represented in the citation data for future research on the influence of nation and the intersection of nation and gender. We subsequently analyzed the data to identify rates of citation by gender.

### 2.1 Journal selection

We were interested in carefully curated citational data about the ADHO community, so we limited our inquiry to what we believe to be the most prestigious journal of the community and thus selected conference issues of *DSH/LLC* as the subject of our study. Although publications from other constituent organizations of ADHO, such as *Digital Humanities Quarterly* (Association for Computers and the Humanities) or *Digital Studies/Le champ numérique* (Canadian Society for Digital Humanities/Société canadienne des humanités numériques), are open access and have significant influence on digital humanities scholarship, we chose to examine *DSH/LLC* because it is 'the longest standing journal in the field', represents the broadest international constituency of ADHO, and, arguably, is the most influential, in part because of its long association with Oxford University Press and its production of ADHO conference special issues (Vanhoutte, n.d., n.p.). Its reputation is further bolstered by its print publication, not available for many other digital humanities journals, as well as by its international scope. We recognize that selecting one journal as a source of insight on the ADHO community is limited. It might also seem counterproductive to choose a print publication in a scholarly community that is driven by scholarly conversations that occur in nonprint environments such as online journals, Twitter conversations, and conferences. We maintain, however, that while such conversations are central to the scholarly community, the imprimatur of Oxford and placement of an article in a selective print journal continues to carry scholarly weight within hiring and tenure and promotion decisions. For this reason, it is important that the journal fairly and equivalently represent the community. Systemic exclusion from the status-making journal will create systemic exclusion within the community. Furthermore, we believe that a study of *DSH/LLC* conference special

issues captures the broader international scholarly participation in digital humanities precisely because the conference has increasingly been focused on encouraging participation from digital humanists beyond the USA, Europe, and Canada, which are over-represented within ADHO.

To create a carefully curated dataset, we limited our focus on *DSH/LLC* to the conference issues, believing that the conference issue is the most visible form of the ADHO community. The conference issue casts a net across all conference participants in the ADHO conference, aiming to be representative of scholarship presented at the conference. Furthermore, by focusing on conference issues, we were able to compare our citational data with Weingart and Eichmann's (2017) data on conference acceptance and attendance. The focus on the conference issue will also allow us to eventually compare the citational practices in issues that were constructed by open calls, such as DH 2014 in Lausanne, to those that were invitation-only submission, such as DH 2016 in Krakow, Poland, where organizers invited paper submissions. We limited our scope to begin with the DH 2006 conference because it was the first Digital Humanities conference organized under the ADHO umbrella institution. Therefore, we believe that the special issues for the annual conferences from 2006 forward illuminate the citational practices of the international ADHO community. Accordingly, we scraped data from the journal to create a dataset of titles, authors, affiliations, abstracts, and citations of articles in *DSH/LLC* published between 2008 and 2017, which corresponded to DH conferences from 2006 to 2015.<sup>4</sup>

There are, of course, limitations to using conferences and conference special issues as markers of community because a number of barriers delimit publication in a special issue for a conference. To be published in an issue, one must be accepted to the conference. Acceptance itself is subject to gatekeeping, such as conference reviewers and program committees, as well as to the priority topics identified in the calls for papers. Attendance is dependent on the resources to attend. Then, one must feel emboldened to submit an essay to the conference special issue—or be encouraged to do so. That is to say, simply to be in a position to submit an essay to the conference issue is mediated by a range of factors, from structural inequalities to self-selection. In turn, these barriers

influence diversity of citations. Despite these limitations, we focus on the conference special issue because it does not represent an ideal community of practitioners; rather, it draws on the community that ADHO brings together at conferences. Furthermore, we undertake our analysis with the goal of shedding light on its challenges and potential solutions.

The initial data revealed the challenges of working with *DSH/LLC* citation data and reinforced our decision to concentrate on a small selection of journal issues. As we examined the broader scope of all *DSH/LLC* issues, we discovered inconsistency in citation formats: some articles contained bibliographies, while others used notes for citation. Thus, while the citation guidelines require all papers to be use citation formatted to Harvard Style referencing, there was significant variability in citation styles. Furthermore, the variability in notes and references made extracting the data onerous. The messiness of the citations led to a great number of difficulties with scraping. There was also a substantial lack of consistency in the citations. This does not appear to be unique to *DSH/LLC* or digital humanities, as Brown *et al.* (2017) found similar inconsistencies in their analysis of early modern book history citations in the JSTOR database. There were also a good number of errors in citation format which made scraping very difficult. Once we had scraped the citations we parsed them with AnyStyle Parser. We chose AnyStyle Parser for its specific ability to parse academic references, its customizability, and its ease of use, as Matthew Bruno, an undergraduate student from Salem State University, was compiling the dataset. As citations were parsed, they were added to a spreadsheet.<sup>5</sup> Given the issues that we found with errors and consistency, we spent a significant amount of time hand-cleaning the dataset to remediate errors.

## 2.2 Coding gender

Recent work analyzing digital humanities citations has used digital methods to code for gender. Bibliometrics typically relies on human guessing and/or gender inference programs to code gender in a dataset. Weingart *et al.* (2016) describe their method as 'gender guessing', based on a combination of hand-coding and automated inference. To identify gender, Eichmann-Kalwara and Weingart (2017) deployed an R package that 'uses historical datasets from the U.S. Social Security Administration, the US Census

Bureau (via IPUMS USA), and the North Atlantic Population Project to provide predictions of gender for first names for particular countries and time periods' (Mullen, 2018). Identifying gender from names in this way creates several issues of accuracy. For example, Hackney (2017) pointed to problems of name/gender mapping for non-binary individuals. They note in a tweet, 'As a nb person w a culturally gendered name, use of name/gender mapping makes me v uncomfortable, regardless of method/interpretation' (Hackney, 2017, n.p.) In addition, assumptions about the relationship between gender and names are likely to occur and skew the data. As humanists who privilege nuance and complexity, we wanted our data to encompass current understandings of gender construction. As Posner (2015) notes of efforts to use the Union List of Artist Names, which uses simplistic gender binaries, in her work, 'No self-respecting humanities scholar would ever get away with such a crude representation of gender'. Eichmann-Kalwara *et al.* (2018) concur that there are limitations to automatic generation of gender, noting 'the gross and problematic simplifications involved in this process of gendering authors without their consent or input' (p. 78). However, they still believe their results to be useful as the peer review process of the conference appears to be influenced by reviewers' perception of gender based on names—rather than the actual gender of the named individuals—and their results provide quantitative evidence of a gap, which is likely to be more convincing of inequity to a scholarly community that privileges quantitative analysis. This work is important, but it must also be balanced with citation analysis of carefully curated datasets targeted to particular questions, like ours.

We emphasize curation of data because we posit that citation practices in digital humanities may disproportionately disadvantage communities that have been historically marginalized in the academy. We understand that there is a tradeoff between curated data that we privilege and faster matching approaches, such as those used by Eichmann-Kalwara *et al.* (2018), but doing better involves individualized research and hand-coding, which necessarily limits dataset size. We further follow the practice of Martin and Runyon (2016), who also hand-coded their data for gender. Recognizing the limitations of binary coding, they 'adapted the codes

to include transgender and a multiple category to accommodate projects treating mixed groups' (Martin and Runyon, 2016, p. 23). We wanted to have accurate data, so we undertook research using authors' biographies published with journal articles or book chapters, biographies published on their institutional and personal websites, and pronouns listed in their social media profiles to code for gender. Although much of current bibliometrics scholarship focuses on binary gender, we chose to use the categories 'men' and 'women' (which include men and women who are transgender); 'non-binary' for those who explicitly do not identify themselves using binary gender; and 'unknown' for those who did not specifically identify themselves with gendered markers. By creating a non-binary category, we hoped to better represent the fluidity of gender so often neglected by bibliometrics, though we recognize that coding gender does not properly articulate the fluidity of gender nor does it demonstrate the danger inherent in being out as non-binary. At the same time, hand-coding gender does present its own issues. We relied on a variety of methods for identification but privileged the author's own identification on their personal webpage and other self-authored media. The best method would, of course, be to survey authors and give them the opportunity to contribute their gender data, if they wish. We also position this data as a snapshot in time, recognizing that how gender is represented and authors' genders may shift over time.

In addition to the original data collected from the citations, we created another spreadsheet to which we added pertinent information to aid our analysis, including first names and institutional affiliation and country of institution at the time of publication. To locate gender and institutional affiliation, we triangulated bios in journal articles, personal and university websites, Twitter profiles, and Facebook pages. Following Slyder *et al.* (2011), we also used the WorldCat Dissertations Database, which includes gender and year of doctorate completion. Our method did, of course, present limitations. We relied on sources written at various times, which introduced the possibility of errors in our data because gender can change over time. The method we adopted is time-intensive but necessary to ensure our data are as accurate as possible. However, for future studies,

outreach to authors to ascertain self-identification would be best.

### 2.3 Dominant authorship

While many of the papers published by *DSH/LLC* are multiauthored, we chose to analyze only the first author with the intention of starting the conversation on gender bias in citations by focusing on dominant author positions. As our study is an opening salvo intended to open up the question of gender bias in citations for further examination, we focused on first authors rather than weighting authorship or presuming equivalence among all authors seemed premature. We recognize that we are making an assumption about authorship by privileging the first name listed in citations. For many humanities fields, first authorship represents the person with the greatest intellectual contribution to the publication. This is not consistent in all fields. For some science fields, the last name in the list is the person who has contributed the most to the scholarship. In others, the first name represents the person who contributed the most funding or laboratory space to a project. Some fields, like mathematics, list authors alphabetically. Our presumption in this dataset is that authorship is most likely to follow the humanities model, given the community's privileging of humanities-based intellectual approaches, but there is the potential for variability depending upon the authors' disciplinary moorings. In a scholarly community as collaborative as digital humanities but that does not have a clear set of expectations of how authorship is attributed, an alternative approach in an early study seemed untenable. The complexity of weighting authorship for studies of citation in digital humanities is an important future area of research on gender bias in digital humanities publication.

As digital humanities is collaborative and further studies in this vein should examine all authors' names, our study examines first authorship intentionally to contend with the assumptions about prestige, effort, and power denoted by the authorship order in a humanities model where first authorship is equivalent to intellectual leadership. Given the limits of our study, we feel that authorship order is too complex to be represented by including each author as equally weighted. In this regard, our study takes up the question of how gender bias operates in dominant

authorship positions, an approach taken in the sciences by Larivière et al. (2013). As we expand our study of citations, it is probable that inclusion of additional names will show more women participate in scholarly publication, but that citation remains skewed. As Leydesdorff and Alkim Almila (2010) find in their analysis of citations in science, no matter which author position women held, whether sole authorship, first or last, 'a paper attracted fewer citations than in cases in which a man was in one of these roles' (p. 213). For this reason, as well, the decision to focus on first authors for our analysis will offer an important point of comparison for future studies that focus specifically gender bias on authorship order in digital humanities.

## 3 Results and Discussion

The dataset built using *DSH/LLC* special issues from 2006 to 2015 included ten conference issues, which contained a total of 128 articles. The articles had 109 unique first authors; 65 were men, 44 were women, and 0 were non-binary.<sup>6</sup> Notably, on average, the conference issues feature a slightly higher percentage of women first authors publishing articles (40%) than giving conference presentations, where they comprised 36.1% of the 3,239 authors who presented at DH conferences, counting every unique author only once (Eichmann-Kalwara et al. 2018, p. 80).

The citations themselves show a more marked gender difference. The dataset contained a total of 3,219 citations. The total number of citations of men first authors is 2,314 with the total number of citations of women first authors at only 705. The total citations of non-binary first authors were one. Organizations, predominantly the TEI, were cited as first author eighty-six times. We had 108 citations for which we were unable to identify gender ('unknown'). First author men made up 71.9% of the citations, first author women comprised 22%, organizations were 2.7%, and 3.4% were unknown (Fig. 1). Our findings are consistent with citation analysis across academia, which finds that men enjoy a positive gender citation effect or GCE (Dion et al., 2018; Larivière et al., 2013; Moss-Racusin et al., 2012; West et al., 2013). There was also a slight difference based on gender in the number of citations in articles published in the

conference special issues. Articles with women as first authors of conference had an average of twenty-three citations per article, while articles with men as first authors had an average of 26.77.

We also see that conference special issues reveal slight differences in percentages of women cited, with the numbers of women cited growing in recent years (Fig. 2).

Recent digital humanities conferences have emphasized diversity, which may account for growing numbers of women first authors cited. For example, the September 2013 journal issue theme was ‘Digital Diversity: Cultures, Languages and Methods’ and September 2014 was ‘Freedom to Explore’. However, the September 2012 issues did not have a

diversity- or inclusivity-related theme and still included a greater percentage of women than previous issues. It may be that the community is increasing in numbers of women. Regardless of increased women cited, it is important to note that even special issues that are specifically about diversity and inclusion do not guarantee citation equity.

Our work also examined the most cited men and women within our dataset. Examination of the most cited individuals is telling. Of the top ten most cited first authors, eight were men (total number of citations is in parentheses):

- (1) Patrick Juola (26)
- (2) John F. Burrows (22)
- (3) Willard McCarty (21)
- (4) Maciej Eder (21)
- (5) David Hoover (18)
- (6) Matthew L. Jockers (17)
- (7) Claire Warwick (16)
- (8) Willa Cather (16)
- (9) John Unsworth (14)
- (10) Jan Rybicki (13)

Of the two women, only one is a scholar (Claire Warwick), while the other is nineteenth-century literary writer Willa Cather. The citation of Cather is not just surprising but may be a bit of a smoke screen, as the Cather citations reference the work of the Willa Cather Archive, a project directed by Andrew Jewell.

A comparison by the ten most cited men and ten most cited women is revealing.

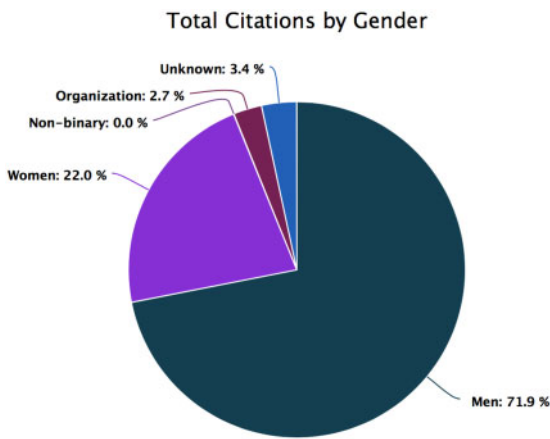


Fig. 1 Total citations by gender

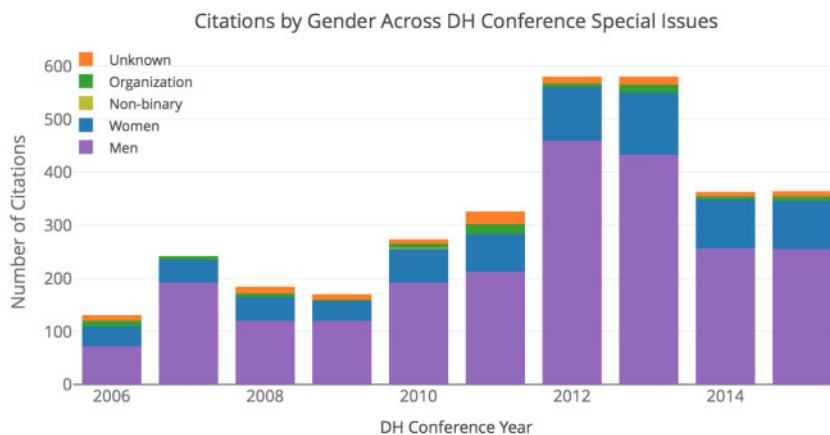


Fig. 2 Citations by gender across DH conference special issues



## Top ten men most cited:

- (1) Patrick Juola (26)
- (2) John F. Burrows (22)
- (3) Willard McCarty (21)
- (4) Maciej Eder (21)
- (5) David Hoover (18)
- (6) Matthew L. Jockers (17)
- (7) John Unsworth (14)
- (8) Jan Rybicki (13)
- (9) Steve Ramsay (12)
- (10) Allen H. Renear (11)

## Top ten women most cited:

- (1) Claire Warwick (16)
- (2) Willa Cather (16)
- (3) Lynne Siemens (12)
- (4) Melissa Terras (11)
- (5) Johanna Drucker (11)
- (6) Arianna Ciula (10)
- (7) Karina Van-Dalen-Oskam (8)
- (8) Gertrude Stein (8)
- (9) Rachel Panckhurst (7)
- (10) Bethany Nowviskie (7)
- (11) Tara L. Andrews (7)

Once again the pattern of citation of a woman literary figure, rather than a scholar, is striking as now we see Gertrude Stein in addition to Willa Cather. Both Stein and Cather are included as subject matter, not authors, with Stein examined in Tanya E. Clement's work on digital reading in Stein's *The Making of Americans* (2008) and Cather the subject of the *Cather Archive* (Jewell, 2012).

We also looked to see if there was a difference in gender of citation based on gender of author and found fairly apparent distinctions. In articles by women, 32.2% of citations were of women (Fig. 3), while only 15% of citations in articles by men authors were citations of women (Fig. 4). Women authors' citations of men were 60% (Fig. 3), while men's citations of other men were 79.4% (Fig. 4). Both, however, include a majority of citations by men.

Our findings are consistent with scholarship that examines gender in citational practices in other disciplines. Ferber's (1986) citational research in economics found that there was a statistically significant difference in how women and men cite, with women more likely to cite women than men. It was also no

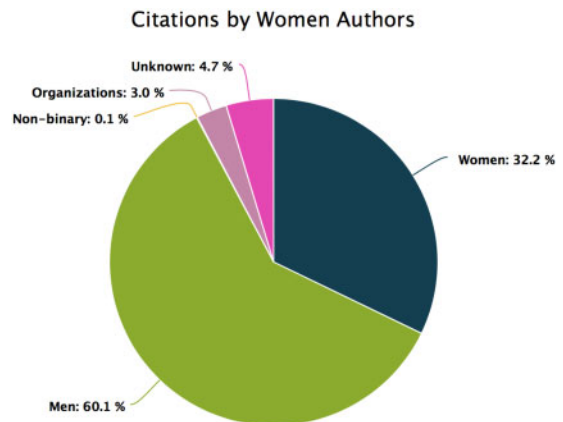


Fig. 3 Citations by women authors

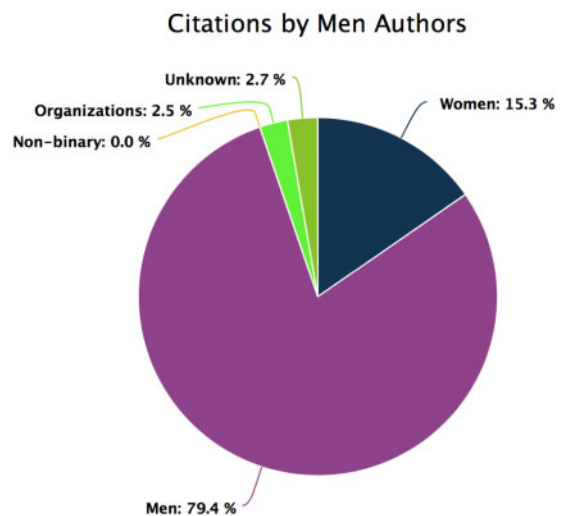


Fig. 4 Citations by men authors

surprise to find that men first authors were cited more than women (26.77 versus 23) as this was a finding consistent to those of Dion *et al.*, (2018), West *et al.* (2013), and others, based on the fact that the numbers of women in academic fields has historically been low. West *et al.* (2013) suggest that some of the reasons for low citation reporting are that 'women historically have been underrepresented in the first author position' (p. 5). Our reliance on first authors may help to explain some of the disparity. Furthermore, Larivière *et al.* (2013) studied 'prominent author positions — sole authorship, first-authorship and last-authorship'

and ‘discovered that when a woman was in any of these roles, a paper attracted fewer citations than in cases in which a man was in one of these roles . . . . The gender disparity holds for national and international collaborations’ (p. 213). Thus, our findings are consistent with the larger body of scholarship analyzing gender and citations in academia.

We posit that recognizing such citational practices might help to diffuse tensions between the many methods and approaches that are subsumed under ‘digital humanities’. One of the explanations for the citation patterns in conference issues may be explained by what is called ‘invisible colleges’, a term coined by Derek de Solla Price (1963, 1986) that viewed informal relationships between scientists as crucial to knowledge production. There is a long history of scholarship that describes information gathering in humanities as driven by community rather than use of databases. Brockman *et al.* (2001) note that much of humanities research is conducted through networks, footnote chaining, which we argue will reinforce any bias introduced by citations. Furthermore, they note that ‘The maintenance of collegial networks for correspondence and collaboration’ plays heavily into the way information is collected (Brockman *et al.*, 2001, p. 11). Given the propensity to work through such networks, which are, in themselves forms of invisible colleges, humanities scholars are privileging personally formed relationships that limit the numbers of people that participate, which are exclusionary and not easily expanded. Long-standing systems of invisible colleges disproportionately affect women, people of color, and less represented nations. Digital humanities is a community not unlike others, where individuals read their friends’ works, identify leaders in research through shared acquaintances and conferences, and share work within communities, a direct result of our collaborative work. And while academia has perhaps expanded since the 1980s, networks remain confined. One positive challenge to such invisible colleges is Academic Twitter, and, more specifically, the digital humanities Twitter network. Twitter has allowed for additional voices to enter the conversation and represents a welcome challenge to limitations of invisible colleges. In their study of digital humanities networks on twitter, Quan-Haase *et al.* (2015) note, ‘The invisible college formed on Twitter is messy, consisting of overlapping social

contexts (professional, personal, and public), scholars with different habits of engagement and both formal and informal ties’ (p. 1). Such shifts allow for possible expansion of influences that have not yet been fully understood. While invisible colleges can encourage a collaborative team to produce better work, we also need to consider how the formation of extra institutional networks might skew our work and our citations.

Citation scholarship also has found that there is a correlation between the number of women in a field and citation of women:

The “citation gap” appears to decrease as the proportion of women in the field, and of articles written by women, increases. The larger the proportion of women in the field the less invisible they are; first, because there are more articles written by women, in which women are more often cited; second, because men cite them more frequently. (Ferber 1988, p. 86)

Therefore, we might expect that as the number of women in digital humanities, and specifically women who present at the digital humanities conference, increases that we will see more equitable citations. To grow equitable citation practices, then, we need to grow women’s participation in DH conferences as authors beyond the 36% participation rate that Eichmann-Kalwara *et al.* (2018) note.

## 4 Conclusions

In the meantime, there are practical ways that the digital humanities community under consideration here might respond to inequity in citational practices. From our study of *DSH/LLC*, it is clear that the digital humanities community needs to pay greater attention to citation, both in terms of accuracy of content within citations and in terms of gender representation among citations. McCarty’s (2014) ‘Getting There from Here: Remembering the Future of Digital Humanities, Robert Busa Award Lecture of 2013’, included in our dataset, reveals one such practice of citational generosity. McCarty includes a works cited of 245, suggesting the breadth of his intellectual community and his indebtedness to that community. If we all encode the scholarly influences in our work

through citational practices, our citations will more accurately reflect the shared values of community and collaboration of digital humanities. We must also set, as a central value to digital humanities, a goal to work outside the invisible colleges that so often hem in citation practices. One method to resist such systems is a review, by authors, of their own citations prior to publication. We have enacted this approach in our own essay. A quick survey of the works cited page of an article, prior to submission, which considers the diversity of citations, would allow the author to see where invisible colleges limit the robustness of intellectual work. Such a survey also forces us to rethink our scholarly networks and to take on ideas of prestige.

We also have come to realize that citations need to be considered a shared responsibility between writers, editors, and peer reviewers. Rather than expect the editor to correct errors in citations, writers must be aware of how errors are exclusionary and work toward accuracy. Peer reviewers must practice generosity and inclusion by pointing to scholarly products that should be included in citations and bibliographies. If we imagine citations as part of a broader collaborative process, then citation practices will be improved.

Recognizing that digital humanities citations are poorly formed and poorly constructed, with very few accurate means of collecting citations across our community, we need work that develops more accurate citation harvesting as well as alternative means of measuring scholarly excellence. *DHQ: Digital Humanities Quarterly* has provided leadership by expanding their indexing into Thomson Reuters' Web of Science, Scopus, and Google Scholar. We know that Twitter and other forms of social media, for instance, have had repercussions for our scholarship. Altmetric platforms have begun to measure the effects of Twitter on scholarly products, and digital humanities should be engaged with such approaches. Digital humanists might find creative ways of using a variety of metrics to locate trends, influential ideas, and emerging concepts through a combination of citation collection, altmetrics, and other bibliometric means. This provides a way to resist the narrow understanding of the significance of scholarly communications based on citation counts, which is increasingly forced on academia. We, as a community, must find

our own ways to measure our ideas, something that digital humanists are uniquely situated to study.

Our study has led us to ask additional questions about citation practices. By using Harvard style, *DSH/LLC*, for example, uses author first initial and last name, which masks gender. We wonder if using first names could produce less equitable citations and should exacerbate gender bias. We also were left to wonder how multiauthored work is affected by gender. Is it possible that gender correlates to likeliness to coauthor? Are women better at sharing credit than men? Are certain subject areas marked by a greater degree of difference? Such questions have been explored in other scholarly communities with varying results, and these questions deserve examination within the digital humanities.

Future work in this area includes the expansion of the dataset to include scholarship from *Digital Humanities Quarterly*, *Digital Studies/Le champ numérique*, and the *Journal of the Japanese Association of Digital Humanities*. One question that would be useful to address is citational practices in relationship to open access. There is a tension between the paywalled *DSH/LLC* journal and the OA journals, such as *Digital Humanities Quarterly*, *Digital Studies/Le champ numérique*, and the *Journal of the Japanese Association of Digital Humanities*, and the conference that *DSH/LLC* helps to support financially. While digital humanities is increasingly concerned with opening our scholarship, our main journal remains paywalled. This tension is further exacerbated by citation research that suggests that open access practices make citations more egalitarian (Atchinson, 2017). In the future, it would be interesting to look at this, though it is outside the scope of our current work. To support inquiry, we are depositing our dataset into our institutional repositories. Adding data to account for ADHO subgroup and for disciplinary home fields will likewise enrich our study.

Finally, we have begun to encode our dataset for nation and language, which offer productive areas of intersectionality for study. To create an accurate dataset for language, we believe that tagging language of paper in relation to language of author (based on dissertation language) might prove helpful. We would also like to consider race as an intersectional factor in citational practices. We believe this is best managed through a survey to identify racial self-identification of authors and expand the dataset to explore the





presented at the meeting of the *Humanities, Arts, Science, and Technology Alliance and Collaboratory (HASTAC)*, East Lansing, MI.

**Sula, C. A.** (2012). Visualizing social connections in the humanities: beyond bibliometrics. *Bulletin of the American Society for Information Science and Technology*, **38**(4): 31–35.

**Sula, C. A. and Miller, M.** (2014). Citations, contexts, and humanistic discourse: toward automatic extraction and classification. *Literary and Linguistic Computing*, **29**(3): 452–64.

**Slyder, J. B., Stein, B. R., Sams, B. S.** et al. (2011). Citation pattern and lifespan: a comparison of discipline, institution, and individual. *Scientometrics*, **89**(3): 955–66.

**Terras, M.** (2006). Disciplined: using educational studies to analyse ‘humanities computing’. *Literary and Linguistic Computing*, **21**(2): 229–46.

**Vanhoutte, E.** (n.d.) *The Journal Is Dead, Long Live The Journal! DSH: Digital Scholarship in the Humanities*. [https://academic.oup.com/dsh/pages/DSH\\_name\\_change](https://academic.oup.com/dsh/pages/DSH_name_change) (accessed 19 June 2019).

**Weingart, S. B. and Eichmann-Kalwara, N.** (2017). What’s under the big tent?: a study of ADHO conference abstracts. *Digital Studies/Le champ numérique*, **7**(1). <https://www.digitalstudies.org/articles/10.16995/dscn.284/> (accessed 19 June 2019).

**Weingart, S. B., Eichmann-Kalwara, N., and Jorgensen, J.** (2016). Gender & centrality: DH 2000–2016. Paper presented at the meeting of *Keystone DH*, Pittsburgh, PA.

**West, J. D., Jacquet, J., King, M. M., Correll, S. J., and Bergstrom, C. T.** (2013). The role of gender in scholarly authorship. *PLoS ONE*, **8**(7): e66212.

## Notes

- 1 In the spirit of author order transparency, we have listed the authors in order of what we perceive to be work contributed to the paper. We also include M.B., an undergraduate student at Salem State University, to recognize his work in the initial data scraping, compilation, and dataset development.
- 2 Weingart wrote a series of blog posts on his analysis of the DH conference from 2013 to 2016 (<http://scottbot.net/dh-quantified/>). Weingart and Eichmann-Kalwara (2017) further analyzed the conference between 2004 and 2015. Eichmann-Kalwara, Jorgensen, and Weingart (2018) also published a study of representation at the conference from 2000 to 2015.
- 3 See Weingart’s ‘dh quantified’ for a helpful review of ongoing work: <http://scottbot.net/dh-quantified/>.
- 4 At the time of writing, the special issue from the 2015 conference is the most recently published special issue (published in 2017).
- 5 All data are deposited in institutional repositories at Texas A&M University and Salem State University. Data are embargoed until the article is published.
- 6 Preliminary studies often begin with small sample sizes, as does ours. For example, Gao et al. (2017) focus on ‘the 200 most cited scholars’ in their analysis of the DH. Additionally, our limiting of the analysis to conference issues constrained our sample size, as did our decision to show that there was a limited pipeline on citations that moved from the conference through the publication cycle. We would like to see our work expanded to cover all *DSH/LLC* issues and additional DH journals.