

# **The Power of Place: Structure, Culture, and Continuity in U.S. Women's Movements**

HTRC ACS Final Report – February 2019

Project Lead: Laura K. Nelson (Northeastern University)  
HTRC Support: Eleanor Dickson and Ryan C. Dubnick

## **Project Objectives**

Why are some ideas and organizations influential, and others insignificant? As social movements such as #BlackLivesMatter and #MeToo dominate the public scene, this question is receiving renewed attention from both social scientists and the general public. In this project I take a comparative historical perspective on this question by comparing influential and insignificant ideas and organizations within the women's movement in Chicago and New York City during both the first and second feminist waves, from 1860 to 1975. Established accounts maintain that the most important ideas of second wave feminism came out of the civil rights and New Left movements. In the first iteration of this project, I challenged this claim through a comparative analysis of place and time. The second wave women's liberation movement was simultaneously founded in Chicago and New York City in 1967. While similar in many ways, the movement in these two cities developed distinct claims-making strategies. I explored whether the differences between the movement in these two cities existed in a similar form during the first wave. To compare both place and time, I used a combination of network analysis to identify the influential organizations in these cities and waves, and computational and qualitative text analysis techniques to identify and explore, in-depth, the movement's claims-making strategies.

Finding more similarities than differences between the waves, I found that second wave ideas were rooted in place-based political logics established during the first wave. In both waves, successful organizations in Chicago sought change by addressing the immediate needs of women, while successful organizations in New York City sought to change individual consciousness. I concluded that collective beliefs become influential when they are aligned with these persistent place-based political logics, or, secondly, when they matched local social structures.

These findings came out of a network analysis of 100 women's movement organizations and additional allied organizations, but a text analysis of only four core organizations. There were both theoretical and practical concerns driving this decision. Movements are diverse, with competing ideas about the best way to organize. Some of the ideas in a movement sector are peripheral, and some are dominant. By doing a focused analysis of the writings produced by the core organizations in Chicago and New York City in both the first and second waves, I was able to inductively identify stability and change within *successful* claims-making strategies, to make claims about why these particular organizations were successful.

Restricting my text analysis to four organizations was also shaped by practical concerns. To collect the writings of the organizations in my study over the entire period, I physically traveled to various archives and I took pictures of the writings (or, in the case where pictures were not allowed, obtained copies of them). Computational text analysis requires machine-readable text. I used the optical character recognition (OCR) software Tesseract to transform the pictures of the text into machine-

readable text. With historical texts, and in particular copies of historical texts, OCR can be quite error-prone. I thus had to hand-correct the raw text in order to be able to analyze it in the detail I required, and this was only feasible for the four organizations. In the second part of this project, I seek to extend this analysis to all organizations in the women's movement in these two cities across the two waves.

### **Corpus-Building with HTRC**

With the help of HTRC, I sought to build a systematic and complete corpus of the writings of the organizations and individuals involved in the women's movement sector in these two cities and wave. This would, I claimed, enable the analysis of the dominant *and* peripheral ideas, and the dynamics of the relationship among all of the ideas circulating in these movement sectors, to better understand the diversity and complexity of recurrent social movements.

With these digitized data I hoped to pursue two analyses. For the first analysis, I planned to use network analysis, computational text analysis, and qualitative deep reading to analyze whether ego-network properties impact the success of an organization. I hypothesize that abstract ideas (e.g., changing hearts and minds) will require decentralized and dense networks in order to be successful, while concrete ideas (e.g., getting a particular policy passed) will require a centralized network to be successful. In the second analysis, I hoped to test three mechanisms that lead to the persistence of political models over time: direct connections between generations of activists – activists purposefully talk to one another and share ideas; direct conferral of culture through stories of successful action indirectly passed down through generations; and indirect conferral of cultural through tradition, patterns of thought, and inherited frames. I planned to use cosine similarity scores between texts over time and phrase and keyword searches to measure the relative weight of each mechanism in producing stability within the women's movement over time.

### **Our Approach**

There were two challenges in compiling the corpus I desired. First, I needed as complete a corpus as possible. Identifying one or two publications from each organization was not going to be complete enough for the comparative analysis I sought, neither would finding documents from only a few organizations. Second, organizations produce many different types of writing, including internal newsletters, meeting and conference notes, legal briefs, policy reports, etc. These different types of writings will use different ideas, language, and frames. To be comparable across organizations, I thus needed a near-complete corpus containing the same type of writing across all organizations.

To collect this corpus, we first started with my compiled list of authors and organizations within the women's movement in these two cities between 1865 and 1975. Out of 82 organizations I had previously identified in Chicago, we were only able to match six to HathiTrust records. We located five organizations in the HathiTrust records out of the 69 I had identified in New York City. Of the organizations we did locate in the HathiTrust records, the collection of documents produced by those organizations was almost always sparse, with just one or two publications from each organization. Thus, I was not able to build the near-complete corpus I desired using HathiTrust's holdings.

The second approach we tried was a vector-space approach. Using a vector space model of all of HathiTrust's collection built by Benjamin Schmidt, we used the text produced by the four core organizations I had already transcribed into machine-readable format as seed texts, and used cosine

similarity to rank the text in HathTrust's collection. According to my qualitative examination of the output, this method was successful at identifying relevant texts, but it was difficult to assess overall accuracy of the output. This approach also did not produce the complete geographic results desired. This method, however, is a good method to use for discovery, as I learned about new authors and organizations to include in my analysis. I recommend this method for those who already have relevant text and would like to identify other similar texts that may be unknown to them.

## **Challenges and Recommendations**

HathiTrust provides a valuable resource to the digital humanities, and as their collection grows, they will provide an increasingly important resource to social scientists. Their historical archive is one of the largest that I know about, and the fact that they open their in-copyright data for text analysis and text mining use is exactly the model that every archive should use. I commend HathiTrust for being forward-looking in this field, and I recommend all social scientists consider using them for comparative-historical analyses.

It remains difficult, however, to draw statistical conclusions from these data in a way that satisfies social science epistemologies and methodologies. There are two ways HathiTrust can improve their holdings for use in the social sciences:

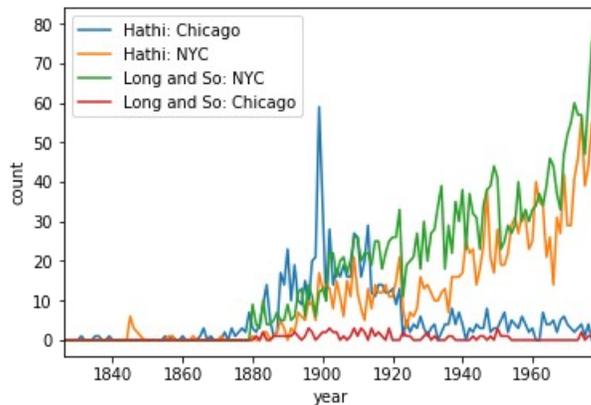
1. HathiTrust has a large collection of novels. Their collection of grey literature and newspapers is smaller, however, and it is not guaranteed to be systematically collected. Social scientists are often more interested in this popular, or mass, culture compared to elite culture. Collecting a representative sample of newspaper data or other grey literature would help make their collection more attractive to social scientists.
2. To make claims about the social world, social scientists need to know to what population a corpus generalizes. HathiTrust could start to build corpora that consist of complete collections of work by specific authors or organizations. Where this is not possible, HathiTrust should strive to provide information on what, out of the complete population of writings, is contained in the HathiTrust collection. This would help enable social scientists to make generalizable claims about the content. This would need to be done on a case-by-case basis.

## **Pivot to Novels**

Leveraging the strength of HathiTrust's collection of novels, I pursued a second corpus-building project after struggling to build the first corpus I desired. From my initial analysis of the four core organizations in New York City and Chicago, I hypothesized that when an organization framed their political argument in a way that resonated with the local culture, they were more likely to be successful. I now seek to test that claim by using novels produced in Chicago and New York City as a way to measure the background culture of a city. I hypothesize that successful women's movement organizations in Chicago will have a similar linguistic style as the novels produced in Chicago, and this style will be different from New York City. I will measure this style using stop words and pronouns – words that most often indicate authorial or linguistic style. To carry out this analysis, I will import my own digitized data that I am currently transcribing into a HathiTrust capsule, to compare against their collection of novels.

To identify novels produced in New York City and Chicago, we used a list of publishing companies housed in these two cities between 1865 and 1975. Using this method, we identified 1,755 novels produced in New York City and 889 novels produced in Chicago. To evaluate whether this was an accurate representation of novels produced in these two cities during this period, I compared the number of novels we identified per year in the HathiTrust records to a dataset compiled by Hoyt Long and Richard So at the University of Chicago. Their list of novels was compiled via a two-part sampling procedure. They obtained a ranked list of library holdings for novels in English based on WorldCat data. Out of that list, they then acquired everything that was available digitally. Figure 1 compares the per-year count of novels in Chicago and New York City from the list obtained through HathiTrust, and the list compiled by Long and So.

**Figure 1: Count of Novels by Year and City from HathiTrust and the Long and So List**



HathiTrust identified more novels produced in Chicago, in particular before 1923, while Long and So identified slightly more novels produced in New York City. Aside from 1900, the number of novels in both lists produced in New York City steadily increased over time, while this number stagnated in Chicago starting around 1922. While these novels do not represent a random sample of all novels produced in this period, they will be broad enough to capture some aspect of the background culture in these two cities. Once my complete data from women’s movement organizations is digitized, I plan to compare my data to these novels, through a HathiTrust capsule.

### Summary of Project Outcomes

1. Through my experience compiling the corpus I desired, I have identified two ways HathiTrust can work to make their collection more usable for social scientists, listed above. I presented these suggestions at HTRC UnCamp 2018 held at the University of California, Berkeley.
2. Using Schmidt’s vector space model I was able to demonstrate how HathiTrust’s collection can be used for data discovery.
3. A collection of novels produced in New York City and Chicago that will be used for an analysis of cultural resonance, and a method to identify novels produced in a city. I plan to present this analysis at the NovelTM conference in October 2019.