

Massive data analysis helps uncover black women's experiences

February 25 2016



Harriet Tubman, famous as an abolitionist, Underground Railroad leader and women's suffrage pioneer. Credit: H. B. Lindsley – National Portrait Gallery, Smithsonian Institution, Public Domain (PD-1875)

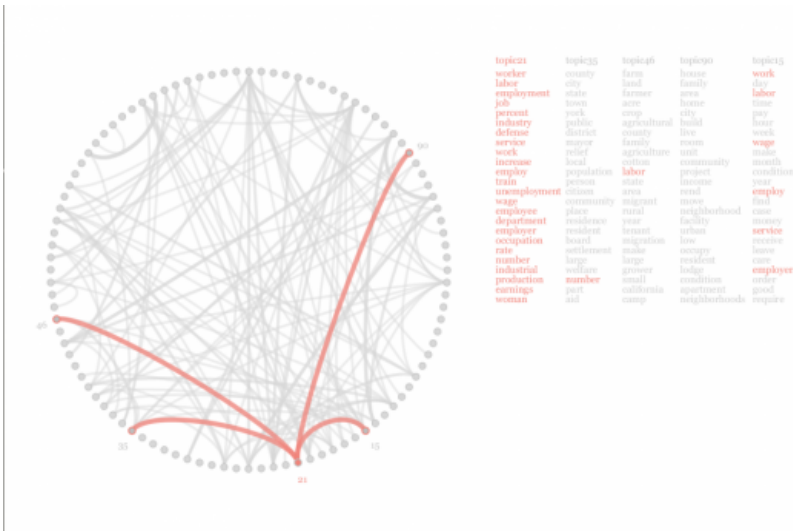
It is often said that history is written by the victors. But it's probably more true to say it is written by the people who have the opportunity to write.

One example of this is the study of black women, their lives and their experiences. Documents recording the lives of black women are often historically obscure, hidden away in vast library collections and unintentionally misleadingly titled or cataloged. Other historical documents don't mention black women directly but may still offer clues. Until recently, researchers had no good way of recovering this "lost history" from either of these categories of documents.

Ruby Mendenhall, an associate professor of sociology, African American studies and urban and regional planning at the University of Illinois at Urbana-Champaign, is leading a collaboration of [social scientists](#), humanities scholars and digital researchers that hopes to harness the power of high-performance computing to find and understand the historical experiences of black women by searching two massive databases of written works from the 18th through 20th centuries. The team also is developing a common toolbox that can help other digital humanities projects.

"With a Big Data approach we get a chance to make use of hundreds of thousands of texts—journals, books, periodicals," Mendenhall says. "The number is greater than what you would normally be able to look at

during an entire career."



Network graphs such as this one are another visualization method that shows how strongly topics correlate across documents with a topic of interest — in this case, the team's "Topic 21," a cluster of concepts spanning labor, employment and industry. Credit: Ruby Mendenhall, University of Illinois Urbana-Champaign

Powering up

Mendenhall's team realized that to search tens or even hundreds of thousands of books, articles and letters, they'd need considerably more computing power than available on a typical university computer cluster. They consulted with colleagues on campus who were members of the National Science Foundation (NSF)-supported [Extreme Science and Engineering Discovery Environment](#) (XSEDE), the most advanced collection of integrated advanced digital resources and services in the world. Those colleagues helped them identify the Blacklight

supercomputer at the Pittsburgh Supercomputer Center (PSC) as a good fit for their project.

Blacklight (now retired) allowed the researchers to analyze 20,000 documents from the [HathiTrust](#) and [JSTOR](#) databases that were known to contain information about black women and to create a computational model based on this corpus of document. They are now using this model to study the entire 800,000 documents in both databases.

Words translated into numbers, graphics



Treemaps such as this one help researchers view the statistics relating to word frequency in a way that spurs insight. Credit: Ruby Mendenhall, University of Illinois at Urbana-Champaign

To make sense of the huge datasets, the investigators turned to two sets of computational techniques: topic modeling and data visualization.

Topic modeling looks at how often certain keywords appear in connection with other terms. For example, a book that contains the word "negro"—at the time considered the most respectful term to describe black men and women—the word "vote" and the word "women" might offer clues about black women's participation in the women's suffrage movement. Mike Black, formerly at the University of Illinois and currently at the University of Massachusetts, headed the team's topic modeling project.

"We're hoping, in the next stage, to ramp up and check these topics against the larger corpus of works," Mendenhall adds.



Sculptor Edmondia Lewis (1844-1907) was the first woman of African- and Native-American descent to achieve notoriety in the fine arts world. She spent most of her career in Rome. Credit: Henry Rocher – National Portrait Gallery, Smithsonian Institution, Public Domain

Mark Van Moer, an XSEDE staff member at the University of Illinois's National Center for Supercomputing Applications, worked as the team's visualization specialist.

As part of the project, he built ways of displaying results that help make more intuitive sense of the data. For instance, a "tree map" displays key words in boxes that correspond to each word's frequency, whereas a "network graph" charts how often key words appear close to each other, also offering insight into how those words are being used and what they mean in context. Yet another visualization technique plots key terms in histograms that allow users to track the emergence and prominence of a given topic over time.

Making sense of the numbers

One aspect of the research involved explorations of the post-World War I Black Women's Club and the New Negro movement. A keyword search revealed that many of the documents that referenced one topic also referenced the other, confirming Mendenhall's prediction that these historical activities were linked. The finding raises interesting questions about how the two movements, which historians knew were contemporaneous, may have interacted. The Illinois researchers hope to begin answering these questions in their on-going work at PSC, as well as their proposed work on Bridges, an NSF-funded supercomputer

coming online later this year.

"The beauty of computation and Big Data lies in how it complements the traditional close reading," says Nicole Brown, a postdoctoral fellow in Mendenhall's group who is interpreting the computational results in light of black feminist theory. "The two methods complement each other to give you a full picture of what's going on."

Van Moer adds that working with social science and humanities researchers "has been a real eye opener in a lot of ways. In the previous seven years, I pretty much worked with physical scientists. Humanities and social science researchers have to be worried about not just what the numbers mean at a surface level. They have a whole theory behind how you go about interpreting things as they relate to the larger society—that's really an interesting aspect of the project for me."

Another group goal is to create a set of computational tools that researchers in many fields will be able to help search various texts for topics of interest—and to understand how those topics interrelate. Topic modeling and visualization methods can be modules in a larger toolbox for digital humanities research.

"We're generally interested in [black women](#) and their life experience," Mendenhall says. "But we also see this as a tool that social scientists and people in the humanities can use to study many topics."

More information: Pittsburgh Supercomputing Center: www.psc.edu

PSC's new Bridges system: psc.edu/index.php/resources-foreign-resources/bridges

Provided by National Science Foundation

Citation: Massive data analysis helps uncover black women's experiences (2016, February 25)
retrieved 2 May 2024 from

<https://phys.org/news/2016-02-massive-analysis-uncover-black-women.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--