VISUALIZING TOPIC MODELS ABOUT AFRICAN AMERICAN WOMEN’S EXPERIENCES AND STANDPOINTS

Computational Social Sciences Workshop
January 30, 2016

Ruby Mendenhall
Sociology, African American Studies, Urban and Regional Planning, and Social Work
Faculty Affiliate Woese Institute for Genomic Biology and Institution for Computing in the Humanities, Arts, and Social Sciences
COLLABORATORS

Mark Van Moer - NCSA

Mike Black - Research assistant at ICHASS and graduate student in Department of English

Harriett Green - English and Digital Humanities Librarian

Nicole Brown - Research Assistant in Sociology

Chengxiang Zhai - Professor of Computer Science

Ismini Lourentzou - Research Assistant in Computer Science

Michael Simeone - Former I-CHASS/NCSA, Arizona State University
PRESENTATION OVERVIEW

First Exposure to NCSA – Kevin Franklin

Start of Project - Michael Simone

Motivation for Study – Recovering Black Women’s History

Challenges

Lessons Learned
MY CURIOSITY AND BIG DATA

First Exposure to NCSA – Kevin Franklin

• Presentation to faculty in African American Studies – image came to life
• Blank sheet of paper and an image that came to life

Planned an interdisciplinary and international violence symposium “Youth Violence in Chicago and the Global Community: Historical and Current Causes and What to Do about It.”
Co-author on an interdisciplinary white paper to the Social, Behavioral & Economic Sciences Directorate at the National Science Foundation about foundational and transformative grand challenge questions. Our paper, “Rethinking 21st Century Urban Transformations: Race and the Ecology of Violence” proposed cyber infrastructure, proposed to capture unheard stories.

Worked with Alex Yahja on K01 proposal to use network analysis to visually map Black mothers’ social networks affected by the violence (e.g., murders, shootings, rapes, etc.) and where it occurs.
MY CURIOSITY AND BIG DATA CONT.

DREAM Project
Developing Responses to Poverty through Education And Meaning

Chicago Earned Income Tax Credit Periodic Payment Project Evaluation
Thinking about a Big Data Project - Michael Simeone

- Gave a lecture in methods class twice
- After each lecture, we talked for about 30 minutes about how big data could relate to my sociological research questions
INSTITUTE FOR ADVANCED COMPUTING APPLICATIONS AND TECHNOLOGIES (IACAT) FACULTY FELLOWS PROGRAM PROPOSAL

2012 – Black Women in Social and Neighborhood Context: Social Network Analysis of Protective Mechanisms

Goal: Text Mining and Social Network Analysis of 25 in-depth interviews

2013 - Visualizing Topic Models about African American Women’s Everyday Experiences and Standpoints

Goal: Search hundreds of thousands periodicals, books and newspapers in JSTOR AND the HATHI Trust Digital Library to identify the type of conversations that emerge about African American women’s shared experience over time and the resulting knowledge that developed (standpoint).
MOTIVATION FOR STUDY
RECOVERING BLACK WOMEN’S HISTORY

- Often, literature by and about African American women is inaccessible.
- Project’s goal - Recover what was written about their ideas, challenges, actions/agency, and accomplishments
CHALLENGE – TIMELY DATA SECURITY AGREEMENTS

HathiTrust Digital Library

• Case study between Illinois and HathiTrust Research Center to make the digital content accessible and usable for research
• Partnership of academic & research institutions, offering a collection of millions of titles digitized from libraries around the world.
• Public Domain – prior to 1923
GROUP A: Race
- Black
- Afr* American
- neger*
- colored
- nig*

GROUP B: Gender
- wom?n
- female?
- girl?
- lady
- ladies

Conducted proximity searches (w/5) in the Solr index metadata for the HathiTrust Research Center corpus: Searched for all combinations and variants of Group A and Group B terms.
EXAMPLES OF VOLUMES RESULTING FROM THE SEARCH 1870 - 2014 ~800,000 DOCUMENTS

JSTOR

HathiTrust
*The Crisis* - DuBois
• *Journal of the National Medical Association*
• *The Negro at Work during the War and during Reconstruction* by U.S. Department of Labor 1921.
Topics 20- legal battles played out in court. Are Black female slaves taking cases to court?

Freedom Suits. 575 by 1846, and ~60 of the time slaves won

Unclear if property/estate is slaves or land (historical period will tell)

Sojourner Truth
Son, illegally sold
Went to court & won

Frequency Range
19,858 to 4,935
TIME FRAMES OF INTEREST

1808 – Abolition of slave importation
1831 – Nat Turner Rebellion
1851 - Uncle Tom’s Cabin published
LESSON LEARNED – TRUST MY INSTINCT

• RA with a strong theoretical background and research interest in Black feminist thought.
• Could take this beyond what I was thinking about it.
• Nicole has exceeded what I imagined. Grateful that she is a part of the project.
• Grateful for the other team members as well (Mike, Mark, Ismini, Karen, Assata, Malaika, and Chengxiang)
Data agreement delays affected our ability to complete the study within the fellowship period.

XSEDE: Extreme Science and Engineering Discovery Environment

Extended Collaborative Support Service (ECSS)

Chengxiang Zhai in Computer Science offered to help – grad student, Ismini Lourentzou – Comparative Text Mining – compare the sets of common topics produced across the entire corpus and 2) variations in topics across specific time periods.
LESSON LEARNED – NECESSITY BREEDS INVENTION

Developed a New Method Called Intermediate Readings (Mike Black and team)

In the process, we identified a metadata revision methodology that served to recognize rescue documents for or about Black women that were not previously categorized/catalogued.

This methodology, which algorithmically mimics the African American tradition of ‘call and response’, trains the LDA topic model using texts known to be written by African American women in order to identify discursive patterns present within other texts not previously identified as being for or about Black women.

Using a technique we developed called “intermediate reading”, documents are verified for their applicability.

This strategy of search, recognition, rescue and recovery (SeRRR) may aid curators of information in making Black women’s voices more accessible within the digitized record.
LESSON LEARNED - PUBLISH

Thinking about the right journals and the right audience

Critical Essay in *International Journal of Humanities and Arts Computing. Mechanized Margin to Digitized Center: Black Feminism’s Contributions to Combatting Erasure within the Digital Humanities*

Publishing Methods Paper – Rejected. We will focus on how the intermediate readings helps us to recover and rescue black women’s stories – SeRRR technique.

Publish Substantive Paper
CREATION OF NEW KNOWLEDGE

How inequality is expressed (or hidden) in the everyday lives of African American women?

How they seek to change entrenched interlocking systems of oppression (racism, classism, sexism, etc.)