Let DH Be Sociological!

Andrew Goldstone
Rutgers University

July 11, 2014
Digital Humanities 2014, Lausanne
Outline

contextualize topic modeling
outline

contextualize topic modeling

(see title)
Matthew L. Jockers, Macroanalysis: Digital Methods and Literary History (Urbana: University of Illinois Press, 2013)
John Laudun and Jonathan Goodwin, “Computing Folklore Studies: Mapping over a Century of Scholarly Production through Topics,” Journal of American Folklore 126, no. 502 (Fall 2013): 455–75
A larger literary history requires other skills: sampling; statistics; work with series, titles, concordances, incipits. (Moretti, “The Slaughterhouse of Literature,” 208)

Researchers must conduct searches that not only have high precision…but that also have high recall. (Tangherlini and Leonard, “Trawling in the Sea of the Great Unread,” 726)
motivations

A larger literary history requires other skills: sampling; statistics; work with series, titles, concordances, incipits. (Moretti, “The Slaughterhouse of Literature,” 208)

Researchers must conduct searches that not only have high precision…but that also have high recall. (Tangherlini and Leonard, “Trawling in the Sea of the Great Unread,” 726)

a turn away from the singularity and richness of literary texts and a concomitant refusal of the ethical charisma of the literary translator or messenger. (Love, “Close but not Deep,” 374)

Meaning is clearly a metavalue and also metaproblem for the digital humanities. (Liu, “The Meaning of the Digital Humanities,” 411)
a longer time-scale

corpus of generalist literary studies journals

PMLA (1889–2007)
Modern Philology (1903–2013)
The Review of English Studies (1925–2012)
ELH (1934–2013)

1-gram counts from JSTOR Data for Research http://dfr.jstor.org
21367 “full-length article” documents, 1000 words or longer
9045 word types retained
43,773,430 tokens

Model: MALLET LDA with 120 topics, hyperparameter optimization.
credit and blame where due

AG and Ted Underwood,
“The Quiet Transformations of Literary Studies”
forthcoming in New Literary History
www.rci.rutgers.edu/~ag978/quiet

credit: TU
blame: AG
a broad context: recent topics

naive sense of “recent”

1. take topic proportions by decade
2. choose cutoff date: 1980
3. “recent” if topic proportions of each decade after cutoff are all be greater than proportions of each decade before cutoff (after 1920)
Figure: Time series of topics, orange if “recent”
A transformation of scholarship: topic most-frequent words

- cultural world culture national european
- body desire form power subject
- narrative story narrator events stories
- jewish jews jew hebrew israel
- early modern university use different
- derrida first two way question
- black white african racial slave
- war political national state military
- body freud medical psychoanalysis unconscious
- death pastoral dead life elegy
- science scientific human knowledge natural
- social society public individual class
- film cinema films camera photography

- new media time world modern
- economic money labor value economy
- history historical time past present
- human philosophy philosophical ethical life
- law legal justice rights trial
- father family mother child children
- city place land country journey
- studies university academic research scholars
- new political cultural social theory
- text reading writing reader texts
- women female sexual woman men
- image object space representation world
- way make just another point
coding scheme

S: Social or political topics, including national, ethnic, sexual, or gender identities;
T: Other thematic material, including religion, moral philosophy, love, nature, etc.;
F: Formal topics, including form, language, style, and genre;
NT: Non-thematic topics, including other languages, proper names, organizational labels, topics classifying textual studies, and clearly methodological discourses.
We risk adding the digital humanities to our proliferating disciplinary menus without any meaningful and substantial engagement with fields such as gender studies or critical race theory. (McPherson, “Why Are the Digital Humanities So White?”)

<table>
<thead>
<tr>
<th></th>
<th>not recent</th>
<th>recent</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>NT</td>
<td>56</td>
<td>4</td>
</tr>
<tr>
<td>S</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>T</td>
<td>21</td>
<td>5</td>
</tr>
</tbody>
</table>
We risk adding the digital humanities to our proliferating disciplinary menus without any meaningful and substantial engagement with fields such as gender studies or critical race theory. (McPherson, “Why Are the Digital Humanities So White?”)
(aside: does the cutoff for “recent” matter?)

<table>
<thead>
<tr>
<th>cutoff year</th>
<th>total recent</th>
<th>S not recent</th>
<th>S recent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950</td>
<td>36</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>1960</td>
<td>35</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>1970</td>
<td>32</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>1980</td>
<td>26</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>1990</td>
<td>26</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>2000</td>
<td>21</td>
<td>10</td>
<td>8</td>
</tr>
</tbody>
</table>
polemic: no returns
polemic: no returns

dedicating the digital humanities to the soul of the humanities (Liu, “The Meaning of the Digital Humanities,” 42)
Content analysis is a research technique for making replicable and valid inferences from texts...to the contexts of their use. (Krippendorff, *Content Analysis*, 24)
If the themes are sociological and the methods are sociological...