Exploring Native American Texts using Text Mining methods

Kun Lu, Raina Heaton, Raymond Orr

Presentation at the TDM@ Univ of Toronto May 3rd, 2023

SCWAReD projects

- Scholar-Curated Worksets for Analysis, Reuse & Dissemination (SCWAReD)

- Funded by Andrew W. Mellon Foundation

- Seek to build HathiTrust Research Center (HTRC) worksets drawn from materials related to historically under-resourced and marginalized textual communities and identify gaps in the HT collection.

OU research team

Dr. Raina Heaton
Native American Studies

Dr. Raymond Orr
Native American Studies

Dr. Kun Lu
Library and Information studies

Research questions

- How is the coverage of Native American authored works in HathiTrust?

- What are the characteristics of the Native works corpus? What does this tell us about representation among Native authors, genres?

- What are the characteristics of the writing styles of Native authors? How do they compare with non-Native authors when writing on similar subjects?
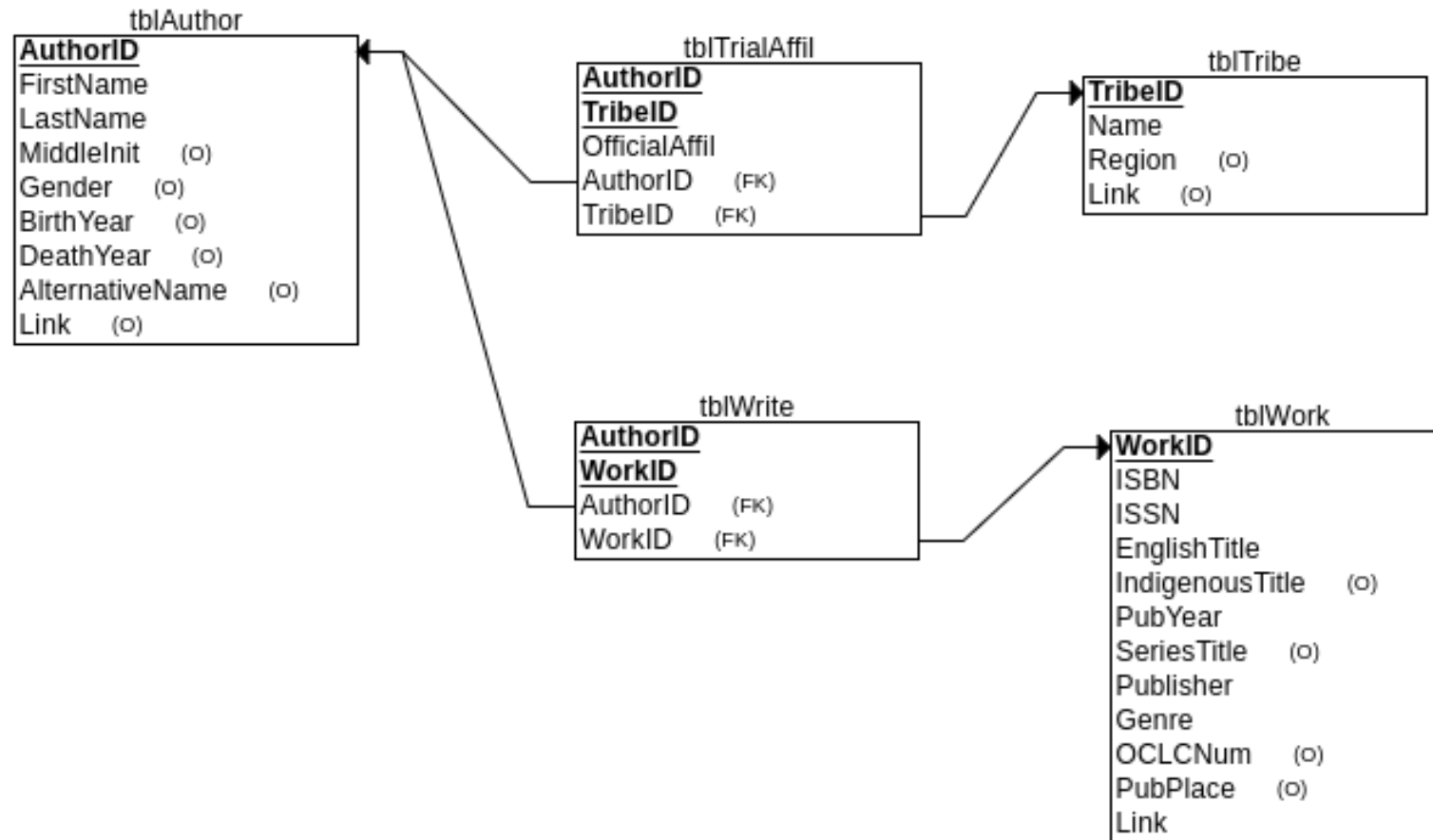
Method

- Create a relational database to store Native author/work information
- Identify Native American authors and their works
- Search in HathiTrust to create a HTRC workset of Native American authored works
- Identify gaps where Native American works are not covered in HathiTrust
- Apply text mining on the workset
- Compare the works generated by this textual community with other communities of interest

What do we mean by 'Native authored works'?

- Geographically limited to the continental US and Canada
- Authors needed to self-identify as a member of a federally recognized tribe
- Included authors of books, chapters in edited volumes, dissertations, reports (but not journal articles)
- Included works with non-Native co-authors
- *No claims to be exhaustive! Hope to continue expand the database.

Create a database

tblAuthor
AuthorID
FirstName
LastName
MiddleInit (O)
Gender (O)
BirthYear (O)
DeathYear (O)
AlternativeName (O)
Link (O)

tblTrialAffil
AuthorID
TribeID
OfficialAffil
AuthorID (FK)
TribeID (FK)

tblTribe
TribeID
Name
Region (O)
Link (O)

tblWrite
AuthorID
WorkID
AuthorID (FK)
WorkID (FK)

tblWork
WorkID
ISBN
ISSN
EnglishTitle
IndigenousTitle (O)
PubYear
SeriesTitle (O)
Publisher
Genre
OCLCNum (O)
PubPlace (O)
Link

Identify Native-authored works

- Internet Public Library (IPL) Native American Authors collection
  - 631 native authors and 2017 work titles (4 titles without author information)
- Native American Languages (NAL) collection
  - 122 titles
- Native American Studies journals over last 30 years
  - 51 volumes
- The LSA's Natives4Linguistics list of Indigenous linguists
  - 34 titles

Identify Native-authored works – cont.

- Indigenous studies departments
- OU listservs
- [Goodreads Native American Authors Shelve](#)
- [Electric Lit Webpage](#)
- Library websites
- [First Nations Development Institute](#)
- Personal blogs
- Publisher websites

Search HathiTrust for matches

- HTRC staff helped search the authors and titles in the HathiTrust index

- The search process relied on a tool built for identifying volumes available in HT digital library based on a list of titles/authors.

- Search results from automated searches were further manually screened for additional matches that were missed due to variations in author names or titles.

Coverage of Native American Authored Works

Table 1: Coverage of native authored works in HathiTrust

Dataset	# of titles searched	# of titles found in HT	Coverage
IPL	2013	900	44.71%
NAL	122	37	30.33%
LSA native linguists	34	0	0%

We also searched a prominent journal in the domain "American Indian Culture and Research Journal", which is covered by HT digital library

Basic statistics of the Native-authored workset

Table 2: A summary of statistics in the Native-authored worksets.

	# of titles	# of Hathi records	# of Hathi items	Publication date range	# of Native authors
IPL	900	1061	1304	1826 to 2009	358
NAL	36	45	56	1900 to 2006	15
Total (after removing overlap)	919	1083	1322	1826 to 2009	369

There is overlap between the NAL and IPL worksets: Out of the 36 matched titles, 45 matched records, and 56 matched items in NAL, 17 titles, 23 records and 27 items are also matched in the IPL workset, respectively. Four of the 15 NAL authors are also included in the IPL workset. The total number of matches after removing overlap is listed in the last row of Table 2.
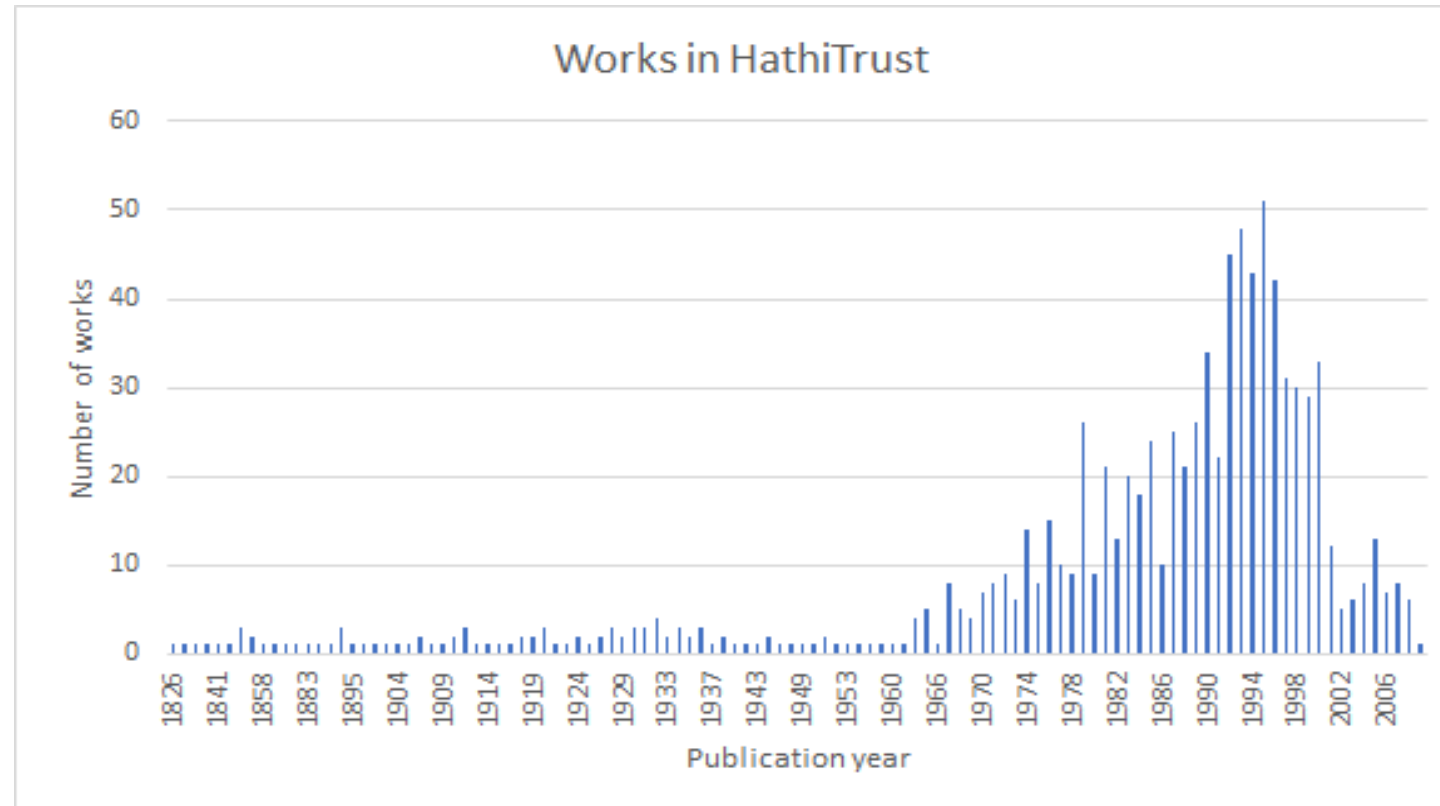
Search HathiTrust for matches

Comparative workset

- Searching HT digital library for similar subjects
  - "Indians of North America", "Indians", "Indian", "Indiens", "Indiens d'Amérique"

- Control the timer period to 1826 to 2009

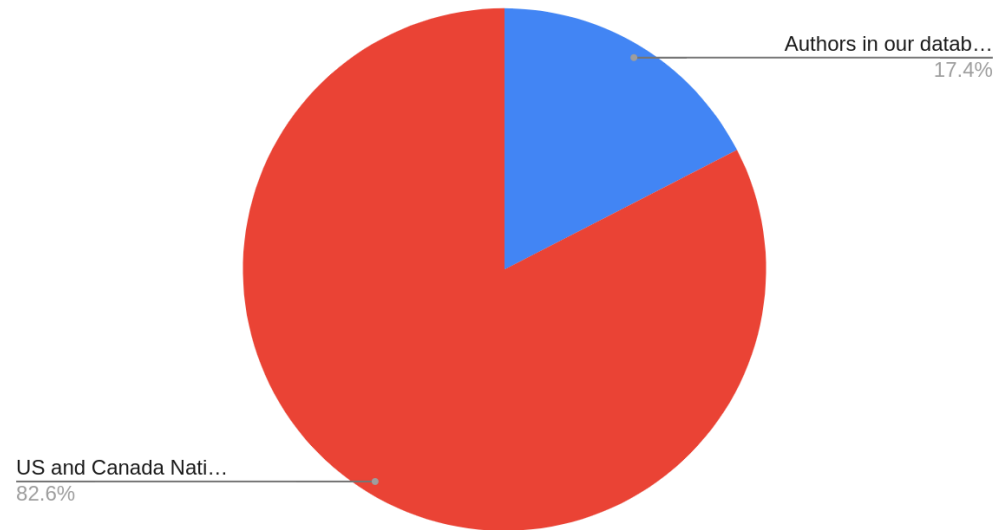- Random sample 1000 works, and adjusted language presentation

- 927 volumes

Native authorship was very limited prior to 1960

Works in HathiTrust
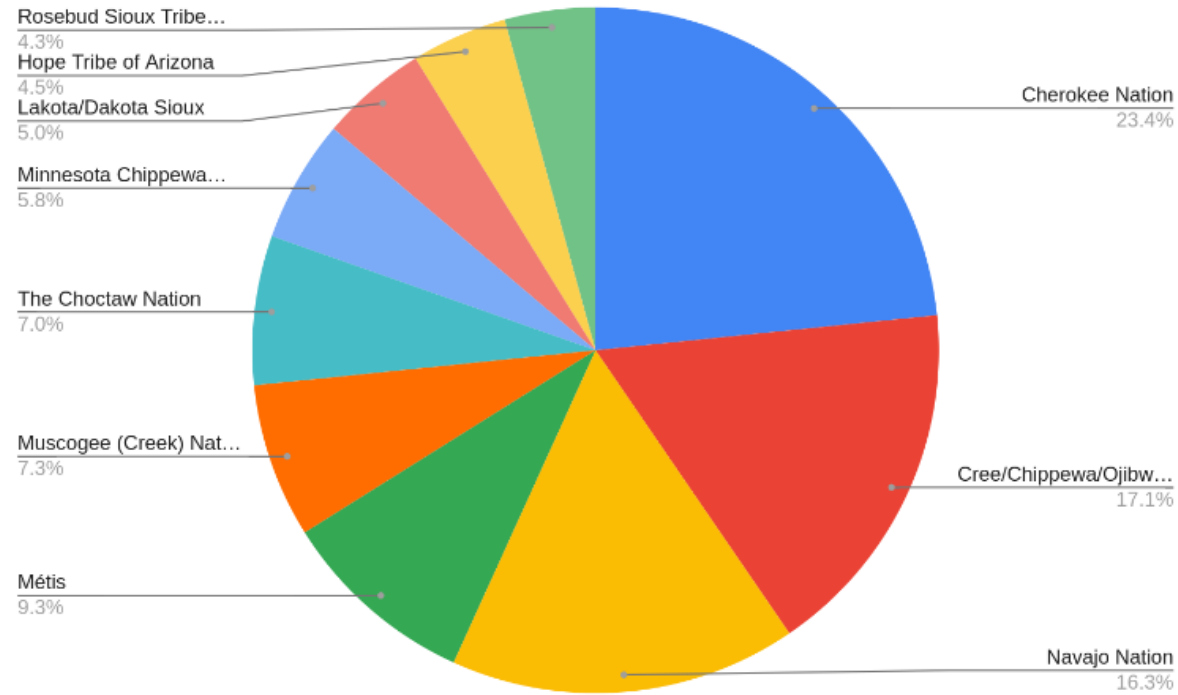
Tribal representation

- Authors from ~255 federally recognized tribes are represented in the database, vs. ~1208 recognized Native Nations and First Nations in the US and Canada (17.4%)
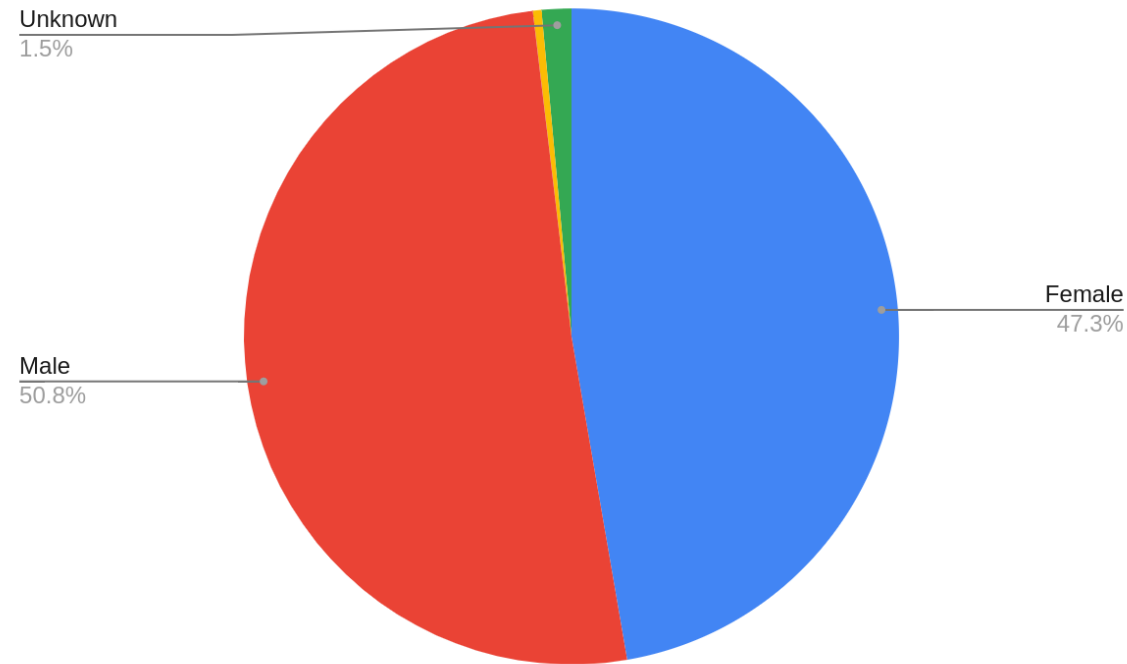
Tribal representation

Authors in our datab…
17.4%

US and Canada Nati…
82.6%

Top tribes

Rosebud Sioux Tribe…
4.3%

Hope Tribe of Arizona
4.5%

Lakota/Dakota Sioux
5.0%

Minnesota Chippewa…
5.8%

The Choctaw Nation
7.0%

Muscogee (Creek) Nat…
7.3%

Métis
9.3%

Cherokee Nation
23.4%

Cree/Chippewa/Ojibw…
17.1%

Navajo Nation
16.3%

Representation in terms number of authors from top 10 tribes

# Representational parity between female and male Native authors

452 female authors, 486 male authors, 4 Two Spirit, 14 unknown

HathiTrust gap analysis

- Lack of representation of more recent works (post-2009)
- Restriction by genre (mostly fiction, history, scholarly)
- Limited coverage of non-English works (issues with language IDs)
- Less diversity in tribal representation
- *Restrictions for discoverability (identify authors by sub-unit, i.e. article, chapter)

Textual analysis
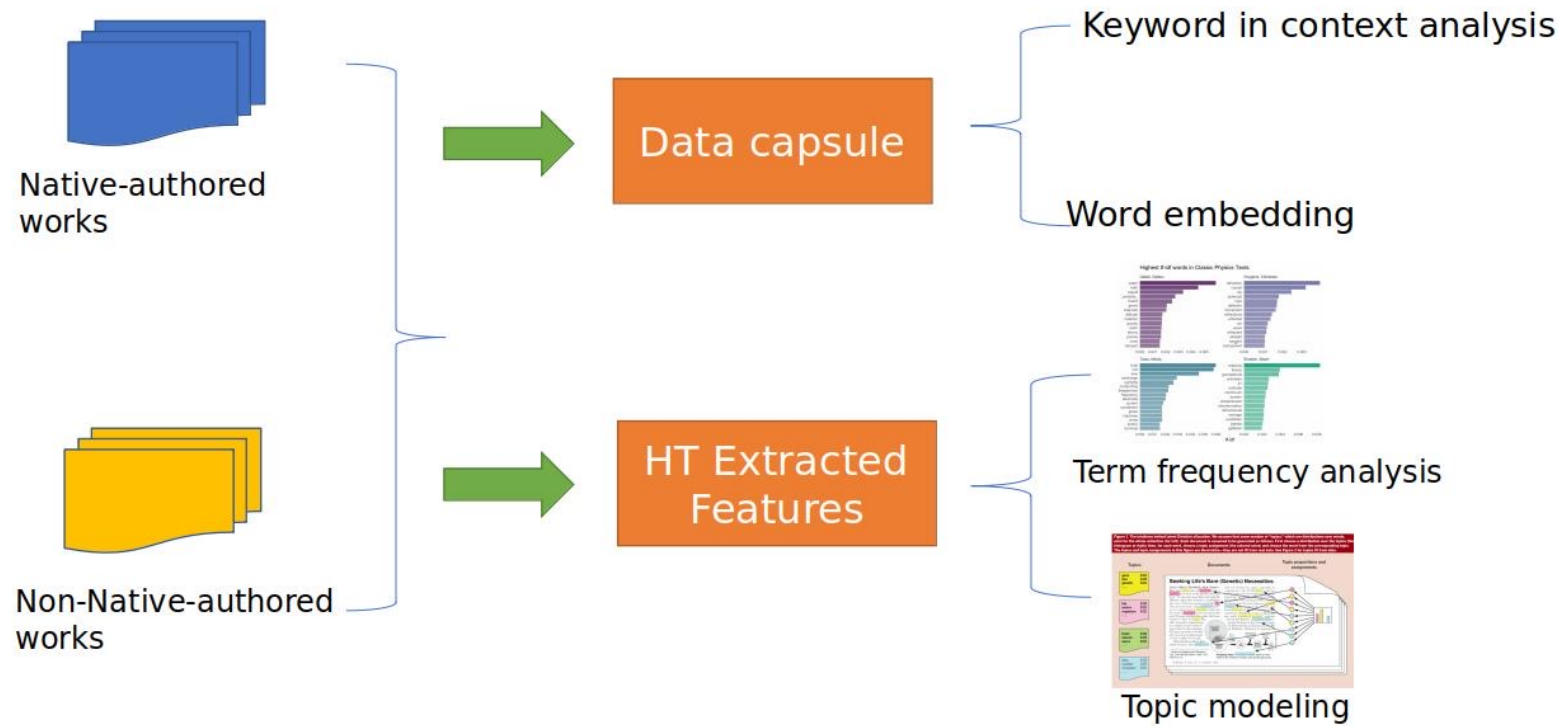
Topic modeling for exploratory analysis
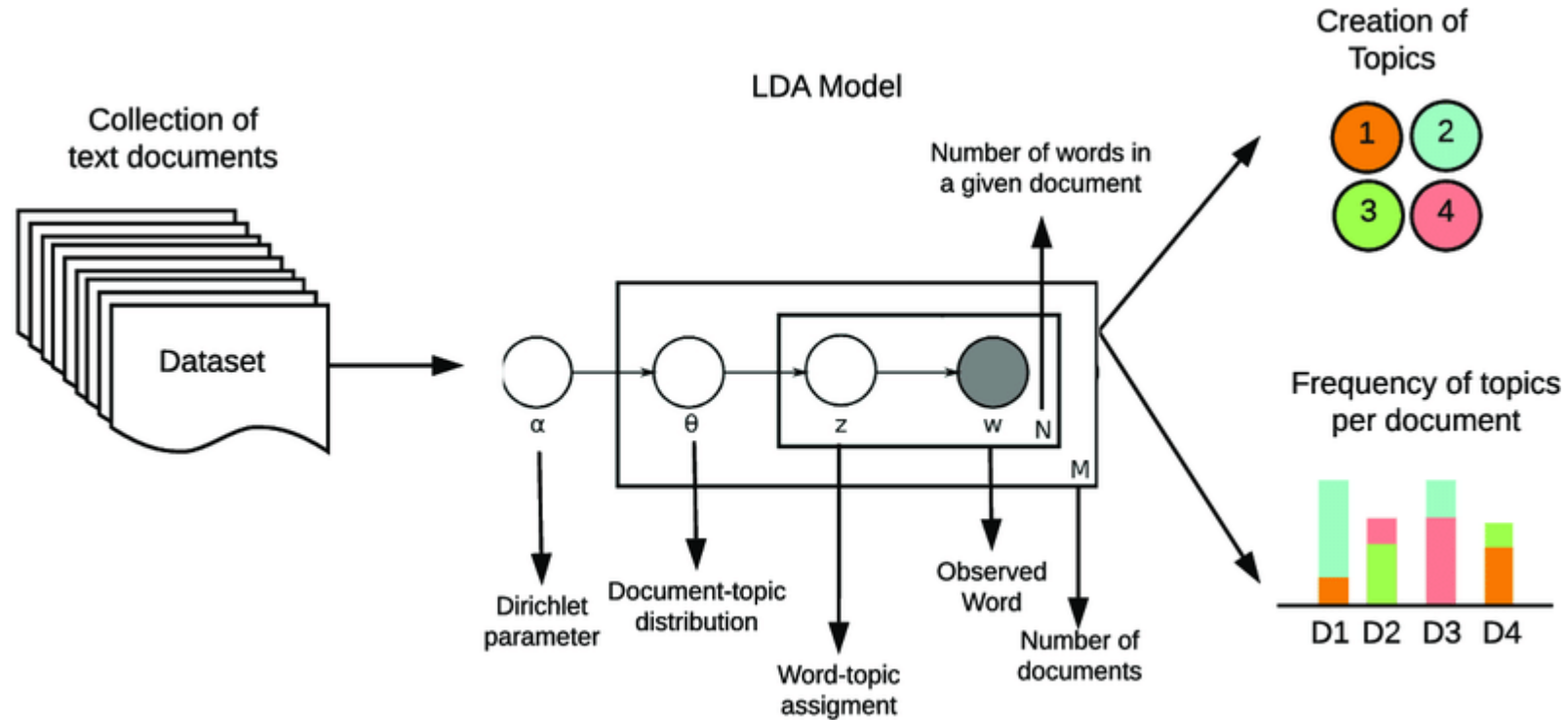
Topic Modeling: raw output

```
|(0, '0.004*"hors" + 0.003*"sat" + 0.003*"car" + 0.003*"smile" + 0.002*"window" + 0.002*"billi" + 0.002*"mayb" + 0.002*"yes" + 0.002*"charg" +
0.002*"town" + 0.002*"tabl" + 0.002*"shoulder" + 0.002*"street" + 0.001*"money" + 0.001*"stare"')
(1, '0.014*"yaqui" + 0.010*"deer" + 0.008*"adv" + 0.006*"adj" + 0.005*"adam" + 0.004*"cheryl" + 0.004*"flower" + 0.003*"tucson" + 0.003*"tobacco" +
0.003*"arizona" + 0.003*"sdi" + 0.003*"kansa" + 0.003*"enchant" + 0.003*"soul" + 0.003*"brush"')
(2, '0.023*"site" + 0.013*"museum" + 0.009*"villag" + 0.008*"pit" + 0.007*"bead" + 0.006*"grave" + 0.006*"iroquoi" + 0.006*"creek" + 0.005*"seneca" +
0.005*"potteri" + 0.005*"york" + 0.005*"burial" + 0.004*"mile" + 0.004*"pipe" + 0.004*"counti"')
(3, '0.016*"hors" + 0.010*"camp" + 0.009*"buffalo" + 0.009*"lakota" + 0.008*"elk" + 0.006*"sioux" + 0.005*"dakota" + 0.005*"tipi" + 0.005*"arrow" +
0.004*"soldier" + 0.004*"warrior" + 0.004*"eya" + 0.003*"meat" + 0.003*"grandfath" + 0.003*"shall"')
(4, '0.091*"yaw" + 0.051*"pay" + 0.045*"pam" + 0.021*"puma" + 0.020*"piw" + 0.019*"oovi" + 0.015*"noq" + 0.014*"pas" + 0.011*"paasat" + 0.009*"pumuy" +
0.008*"kita" + 0.008*"hopi" + 0.007*"kur" + 0.006*"hiita" + 0.006*"hak"')
(5, '0.007*"medicin" + 0.007*"lodg" + 0.005*"ceremoni" + 0.005*"hunter" + 0.004*"coyot" + 0.004*"chief" + 0.004*"villag" + 0.004*"sacr" + 0.004*"pipe" +
0.004*"hors" + 0.003*"camp" + 0.003*"sat" + 0.003*"feather" + 0.003*"eagl" + 0.003*"crow"')
(6, '0.011*"chief" + 0.010*"lake" + 0.005*"ojibway" + 0.005*"warrior" + 0.004*"peac" + 0.004*"council" + 0.004*"shall" + 0.004*"seneca" + 0.003*"villag"
+ 0.003*"parti" + 0.003*"general" + 0.003*"iroquoi" + 0.003*"enemi" + 0.003*"cano" + 0.002*"island"')
(7, '0.019*"tribal" + 0.013*"servic" + 0.010*"communiti" + 0.009*"provid" + 0.009*"villag" + 0.009*"area" + 0.008*"acr" + 0.008*"health" + 0.007*"locat"
+ 0.007*"govern" + 0.007*"employ" + 0.007*"member" + 0.007*"develop" + 0.006*"program" + 0.006*"alaska"')
(8, '0.080*"bom" + 0.020*"mari" + 0.018*"jame" + 0.014*"john" + 0.013*"william" + 0.012*"jess" + 0.011*"joseph" + 0.011*"hinckley" + 0.010*"buri" +
0.010*"crocker" + 0.009*"sept" + 0.009*"oct" + 0.009*"charl" + 0.008*"june" + 0.008*"list"')
(9, '0.014*"fort" + 0.011*"soldier" + 0.011*"june" + 0.011*"juli" + 0.011*"august" + 0.010*"march" + 0.010*"septemb" + 0.010*"octob" + 0.009*"april" +
0.009*"januari" + 0.009*"creek" + 0.008*"novemb" + 0.008*"armi" + 0.008*"treati" + 0.008*"decemb"')
(10, '0.010*"cheyenn" + 0.009*"creek" + 0.008*"oklahoma" + 0.006*"warrior" + 0.006*"chief" + 0.005*"territori" + 0.005*"battl" + 0.005*"thoma" +
0.005*"fight" + 0.004*"steinbeck" + 0.004*"veteran" + 0.004*"victor" + 0.004*"john" + 0.004*"hors" + 0.004*"soldier"')
(11, '0.020*"les" + 0.014*"des" + 0.014*"negro" + 0.014*"que" + 0.013*"est" + 0.010*"une" + 0.010*"dan" + 0.009*"term" + 0.009*"nous" + 0.008*"pas" +
0.008*"slave" + 0.008*"qui" + 0.006*"african" + 0.006*"pour" + 0.005*"mulatto"')
(12, '0.012*"chinook" + 0.009*"wellston" + 0.009*"crash" + 0.008*"plane" + 0.007*"pilot" + 0.007*"report" + 0.006*"mexicano" + 0.006*"investig" +
0.005*"powhatan" + 0.005*"flight" + 0.005*"senat" + 0.005*"fbi" + 0.005*"evid" + 0.004*"assassin" + 0.004*"paul"')
(13, '0.013*"california" + 0.009*"mexico" + 0.007*"mexican" + 0.006*"san" + 0.006*"spanish" + 0.005*"apach" + 0.005*"group" + 0.005*"area" +
0.004*"mission" + 0.004*"cultur" + 0.004*"valley" + 0.003*"jicarilla" + 0.003*"spaniard" + 0.003*"villag" + 0.003*"santa"')
(14, '0.019*"shoshon" + 0.016*"social" + 0.014*"health" + 0.014*"titl" + 0.014*"program" + 0.011*"sour" + 0.011*"auth" + 0.011*"iden" + 0.010*"group" +
0.009*"cultur" + 0.009*"mental" + 0.009*"urban" + 0.009*"studi" + 0.008*"servic" + 0.008*"alcohol"')
(15, '0.011*"bit" + 0.010*"pos" + 0.009*"bee" + 0.009*"dah" + 0.008*"object" + 0.008*"lit" + 0.006*"doo" + 0.006*"bik" + 0.005*"baa" + 0.005*"naa" +
0.005*"aah" + 0.004*"car" + 0.004*"fit" + 0.004*"stem" + 0.004*"dii"')
(16, '0.008*"realli" + 0.008*"lot" + 0.006*"john" + 0.006*"mayb" + 0.005*"guy" + 0.005*"kid" + 0.004*"mom" + 0.004*"money" + 0.004*"car" +
```

Common topics

- Colonial govt. - tribe relations
- Linguistics
- Industry
- Native arts and subsistence
- Ceremony, kinship

Different topics

- More culture and tradition **in the Native set**; close association between culture, medicine, community, and government

- More archaeology and history/militarism **in the non-Native set**

- Casinos/gaming, health, boarding schools, and student/education-related topics **in the non-Native set**

Keyword in Context/Term Frequency

- Terms of interest within Native American Studies:
- People/groups
  - E.g. Vine Deloria, Sitting Bull, Quanah Parker, Custer, Black Kettle, AIM …
- Events/places
  - E.g. Sand Creek, Trail of Tears, Wounded Knee, Black Hills, Alcatraz, NAGPRA …
- Academic terms
  - E.g. Relationality, Reciprocity, Survivance, Decolonize, Sovereignty, Revitalization …
- Archaeology terms
  - E.g. prehistory, Mississippian, artifact, excavate, ancient …
- People terms
  - E.g. Indian, Whites, savage, Indigenous, colonist, blood quantum, …
- Misc. relevant things
  - E.g. land, massacre vs. battle, frontier, New World, per cap, assimilate …

Keyword in context data

volume_id	title	year	page	sentence	keyword	context	before_sentence	after_sentence
						Filter	Filter	Filter
mdp.39015050467557	Out there somewhere /	2002	171	12	mountain	"Big Mountain Woman," I say.	The baby's father is holding her. "What's h...	"Big Mountain Woman," I say. Ev
mdp.39015050467557	Out there somewhere /	2002	171	15	Land	For you are the people of this land.	He tells me and adds, "She was born at ...	For you are the people of this lat
mdp.39015050467557	Out there somewhere /	2002	171	18	Spirit	For you are the smiles of a human spirit th...	Everybody is smiling. For you are the ...	For you are the smiles of a huma
mdp.39015050467557	Out there somewhere /	2002	175	1	Community	He is a native of Acoma Pueblo in New ...	ABOUTTHE AUTHOR^ Simon J. Ortiz is a ...	He is a native of Acoma Pueblo i
mdp.39015050467557	Out there somewhere /	2002	175	1	Native	He is a native of Acoma Pueblo in New ...	ABOUTTHE AUTHOR^ Simon J. Ortiz is a ...	He is a native of Acoma Pueblo i
mdp.39015050467557	Out there somewhere /	2002	175	3	Land	As a major Native writer, he insists on telli...	ABOUTTHE AUTHOR^ Simon J. Ortiz is a ...	As a major Native writer, he insi
mdp.39015050467557	Out there somewhere /	2002	175	3	Community	As a major Native writer, he insists on telli...	ABOUTTHE AUTHOR^ Simon J. Ortiz is a ...	As a major Native writer, he insi
mdp.39015050467557	Out there somewhere /	2002	175	3	Native	As a major Native writer, he insists on telli...	ABOUTTHE AUTHOR^ Simon J. Ortiz is a ...	As a major Native writer, he insi
mdp.39015050467557	Out there somewhere /	2002	175	5	Sand Creek	His previous works include Men on the ...	ABOUTTHE AUTHOR^ Simon J. Ortiz is a ...	His previous works include Men
mdp.39015050467557	Out there somewhere /	2002	175	5	power	His previous works include Men on the ...	ABOUTTHE AUTHOR^ Simon J. Ortiz is a ...	His previous works include Men
mdp.39015011414938	Northern breezes /	1917	9	8	Spirit	Waugoshance and Skillagalee Spirit of the ...	Greetings .... My Theme .... Walk With Me...	Waugoshance and Skillagalee Sp
mdp.39015011414938	Northern breezes /	1917	9	25	Indian	On the Death of an Old Indian.	The Fire Whistle... Ship of State......... I Can...	On the Death of an Old Indian. A
mdp.39015011414938	Northern breezes /	1917	23	2	Land	Oh, the charms on every hand In this joy-...	HOME Where the wild goose wings its way...	Oh, the charms on every hand Ir
mdp.39015011414938	Northern breezes /	1917	25	1	Land	In far-away countries over the seas are ...	NORTHERN PICTURES	In far-away countries over the se
mdp.39015011414938	Northern breezes /	1917	25	1	nature	In far-away countries over the seas are ...	NORTHERN PICTURES	In far-away countries over the se
mdp.39015011414938	Northern breezes /	1917	25	2	Land	The rocks of Lake Superior, Niagara's migh...	NORTHERN PICTURES In far-away countrie...	The rocks of Lake Superior, Niag
mdp.39015011414938	Northern breezes /	1917	25	2	nature	The rocks of Lake Superior, Niagara's migh...	NORTHERN PICTURES In far-away countrie...	The rocks of Lake Superior, Niag
mdp.39015011414938	Northern breezes /	1917	26	2	Spirit	And my spirit longs to take Its flight and ...	Piercing the utmost sky And my soul is at ...	And my spirit longs to take Its fli
mdp.39015011414938	Northern breezes /	1917	27	2	Savage	Waugoshance and Skillagalee, The lights ...	WAUGOSHANCE AND SKILLAGALEE ...	Waugoshance and Skillagalee, T

Term frequency by year/decade

keyword	year	num_of_sentences	num_of_works
Filter	Filter	Filter	Filter
Indian	2008	1617	3
Indian	2009	89	2
Indian Removal Act	1951	2	1
Indian Removal Act	1959	5	1
Indian Removal Act	1973	3	1
Indian Removal Act	1980	1	1
Indian Removal Act	1983	4	1
Indian Removal Act	1984	1	1
Indian Removal Act	1986	3	1
Indian Removal Act	1987	1	1
Indian Removal Act	1990	1	1
Indian Removal Act	1991	2	1
Indian Removal Act	1992	9	4
Indian Removal Act	1993	4	2
Indian Removal Act	1994	6	1

Term frequency by year

keyword	decade	sentences	works
Filter	Filter	Filter	Filter
AIM	1940s	1	1
AIM	1970s	5	4
AIM	1980s	95	16
AIM	1990s	372	36
AIM	2000s	156	11
Aboriginal	1850s	2	1
Aboriginal	1880s	9	1
Aboriginal	1900s	8	1
Aboriginal	1910s	8	5
Aboriginal	1920s	76	7
Aboriginal	1930s	58	5
Aboriginal	1940s	1	1
Aboriginal	1950s	10	2
Aboriginal	1960s	55	15

Term frequency by decade

Term frequency normalization

- Term usage = # of works use the term/ total # of works in the decade

term	Date	Works	Totals	Percent
Aboriginal	1830s	0	2	0.00%
Aboriginal	1840s	0	2	0.00%
Aboriginal	1850s	1	4	25.00%
Aboriginal	1880s	1	2	50.00%
Aboriginal	1890s	0	2	0.00%
Aboriginal	1900s	1	7	14.29%
Aboriginal	1910s	5	16	31.25%
Aboriginal	1920s	7	18	38.89%
Aboriginal	1930s	5	22	22.73%
Aboriginal	1940s	1	8	12.50%
Aboriginal	1950s	2	7	28.57%
Aboriginal	1960s	15	31	48.39%
Aboriginal	1970s	24	110	21.82%
Aboriginal	1980s	39	197	19.80%
Aboriginal	1990s	93	363	25.62%
Aboriginal	2000s	25	102	24.51%

The percentage of works used this term in the decade

Nature terms

- The most frequent terms like 'land', 'water', 'rights', 'power', 'spirit(ual)(ity)' were important (frequent) in both corpora.
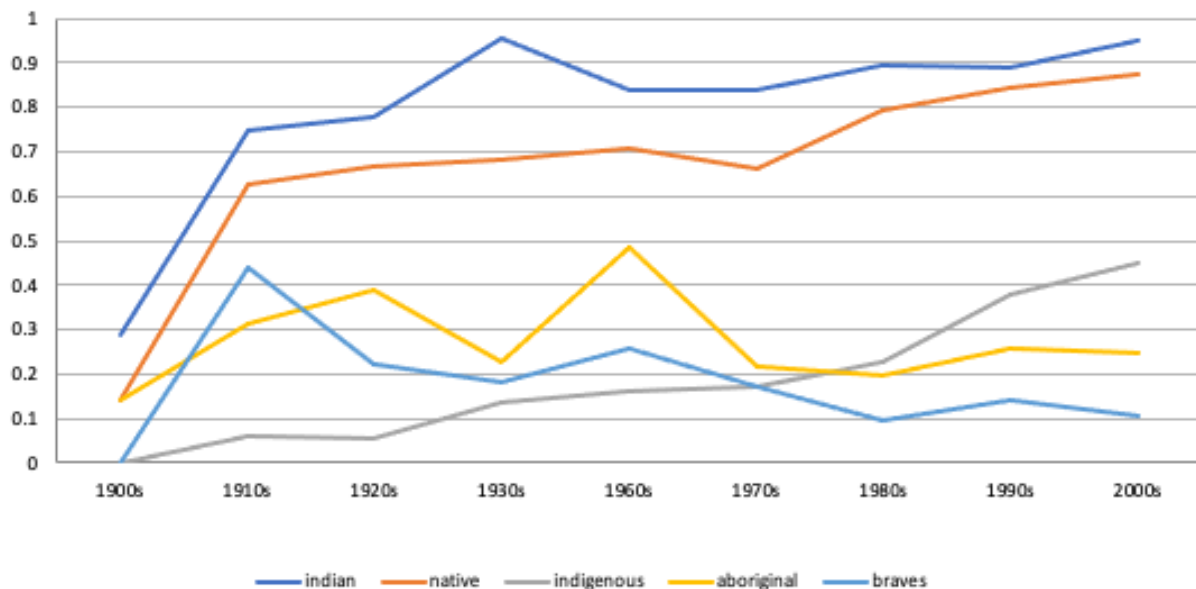
Native workset
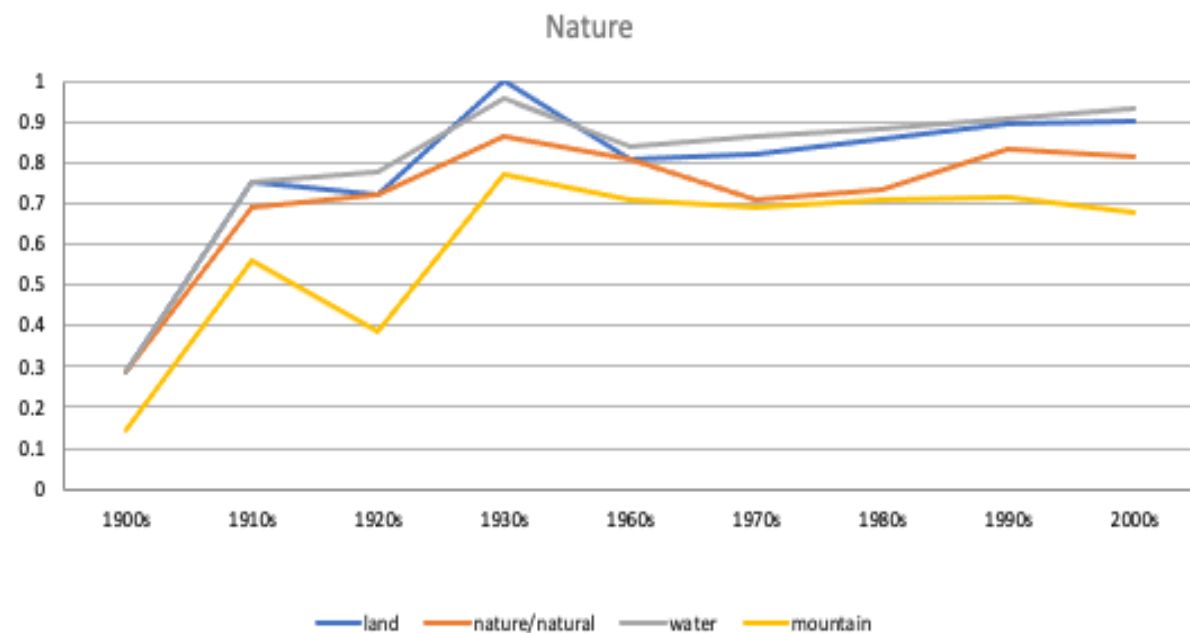
Non-Native workset
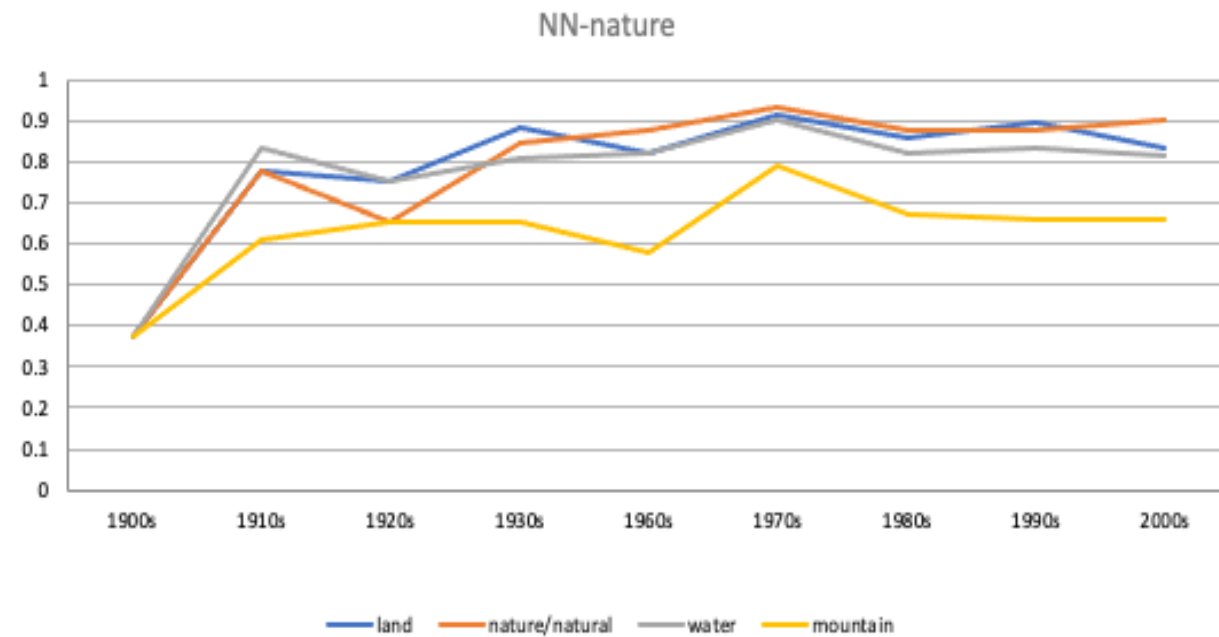
Group terms

- "Indian" is the most widely used term by both groups for all years, followed by 'Native'. 'Indigenous' is surprisingly uncommon, and 'Aboriginal' is used more often in the comparative corpus (statistically significant)
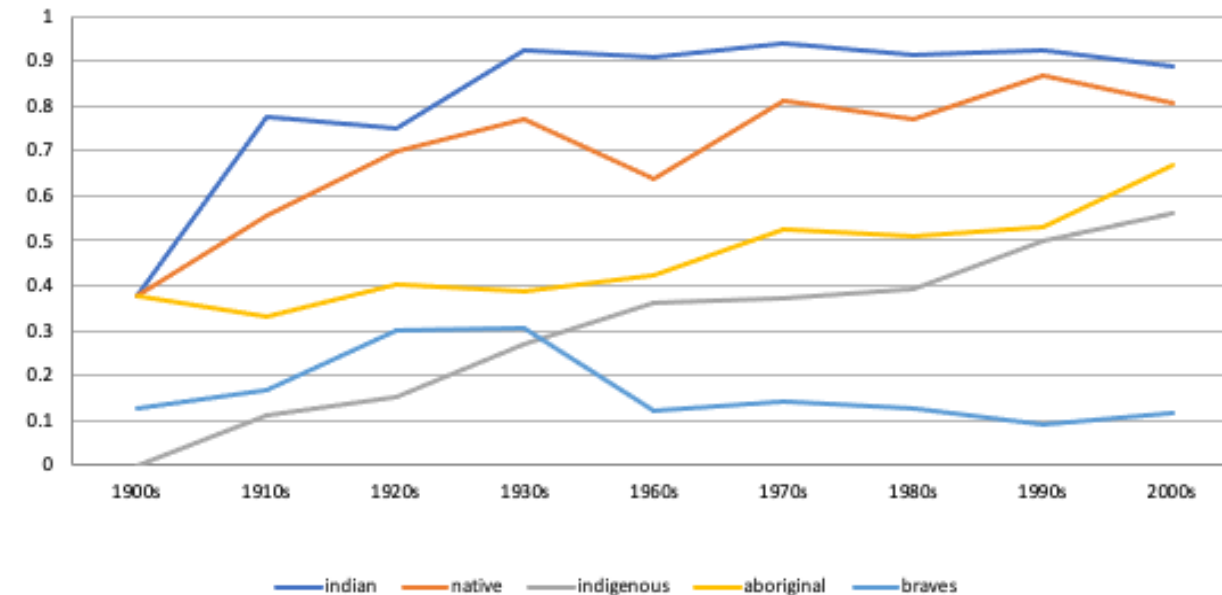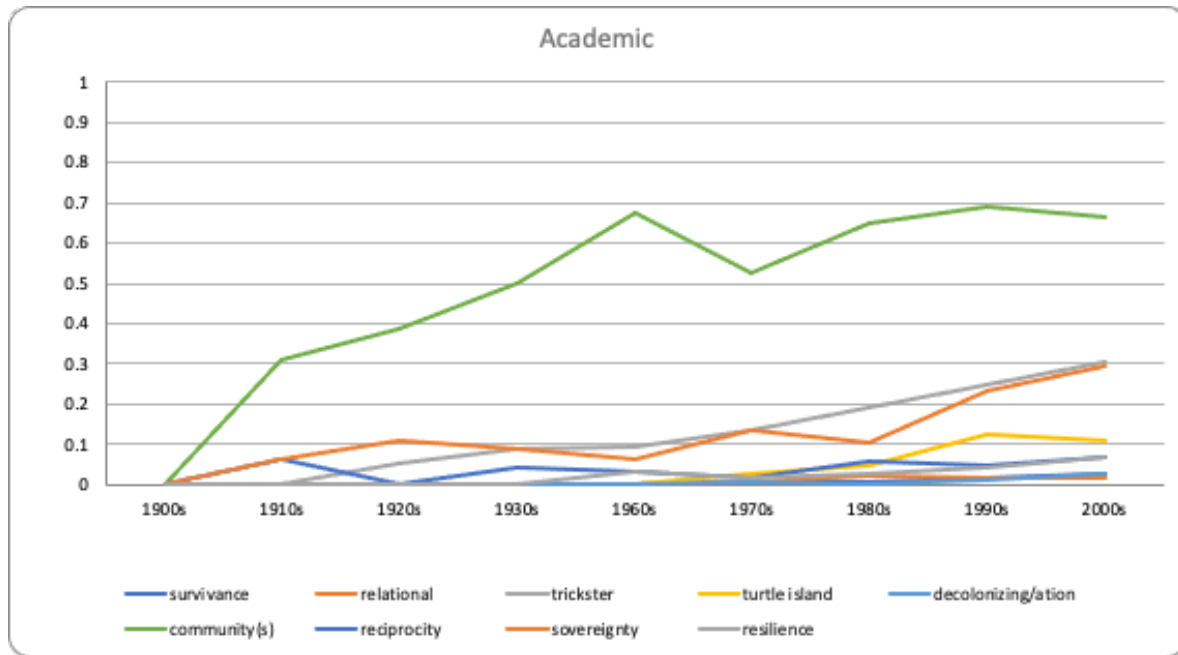
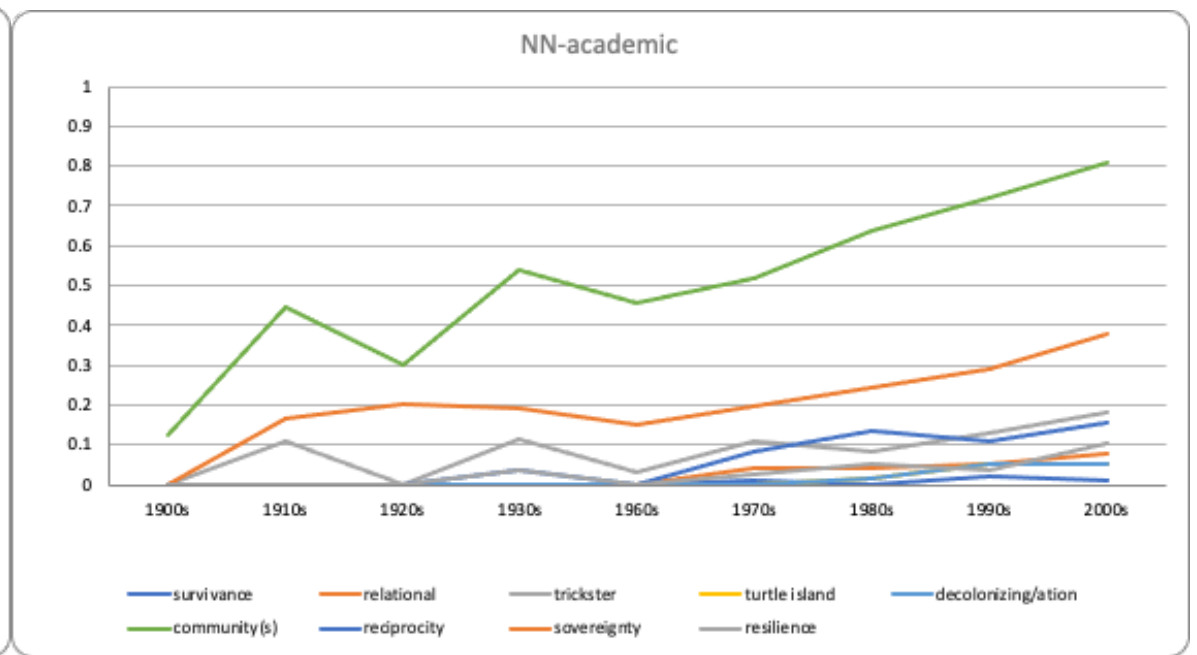Native workset

Non-Native workset

In-group

NN-ingroup

Academic terms

- These terms do not seem to have gained much traction for either set of authors. 'Community' is a popular term in both sets.

Native workset

Non-Native workset

(Surprisingly) infrequent in both:

- Historical figures (Geronimo, Custer, Sequoyah, etc.)
- Contemporary figures and movements (Deloria, AIM, BIA, etc.)
- Legislation (Indian Removal Act, Allotment, Trail of Tears, etc.)
- Conflicts (Wounded Knee, Little Big Horn, etc.)

Native vs. Non-Native sets

- Notable differences related to the US national narrative, shown by uses of some terms:
  - 'New World'
  - 'extinct'
  - 'settler'
  - 'frontier'
  - 'pioneer'

- Authors from the comparative set are more likely to use these **uncritically**; Native authors are generally **critical** of and sometimes play off these concepts.

- Sentiment analyses are underway.

Word Embedding
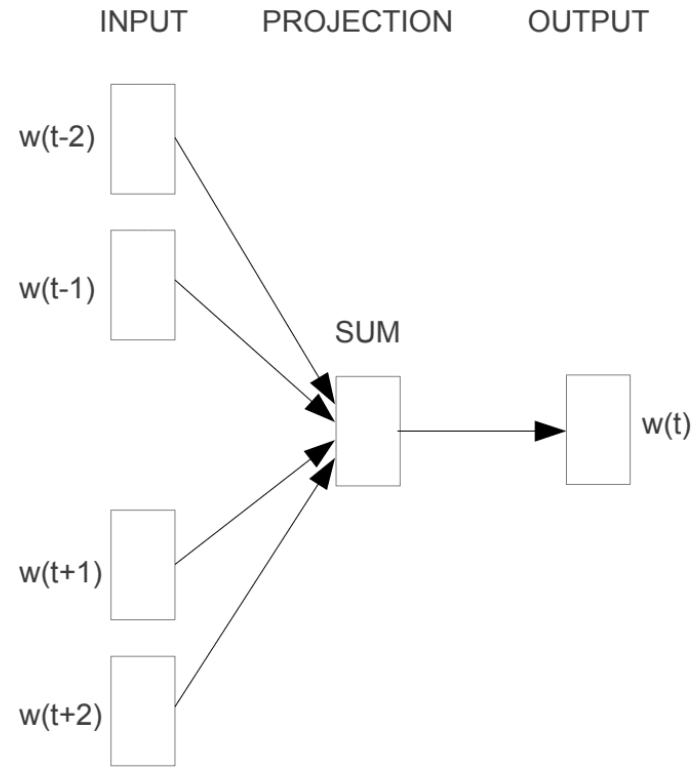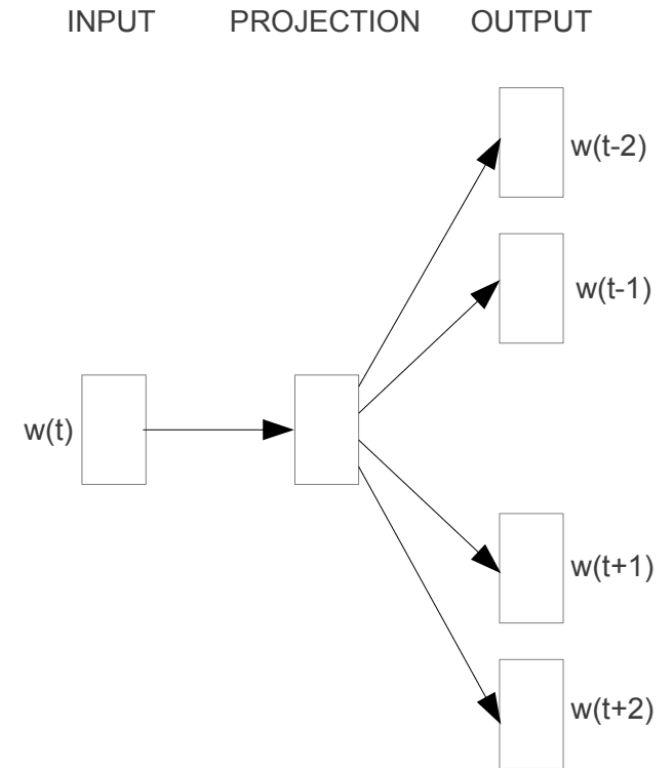
```
most_similar(model_sg,"trail of tears", topn=30)  #
```

```
[('sand creek', 0.6573565006256104),
 ('national historic trail', 0.6313459873199463),
 ('forced removal', 0.6285008788108826),
 ('have weeped', 0.6216064691543579),
 ('de soto expedition', 0.6159104108810425),
 ('death march', 0.6077911853790283),
 ('bering land bridge', 0.6054410338401794),
 ('the great sioux nation', 0.6045249104499817),
 ('dispossession', 0.6041374802589417),
 ('canadian pacific', 0.597732663154602),
 ("custer's last stand", 0.5953289270401001),
 ('removal west', 0.5897485613822937),
 ('hegira', 0.5886384844779968),
 ('page 115', 0.5865357518196106),
 ('european invasion', 0.5844636559486389),
 ('forced migration', 0.5840181708335876),
 ('booting', 0.5830328464508057),
 ('six decades', 0.5822076201438904),
 ('yellow thunder', 0.5815514326095581),
 ('molestaton', 0.5811564326286316),
 ('33595', 0.5797830820083618),
 ('infamous', 0.5794732570648193),
 ('noahvose', 0.5784812569618225),
 ('ancestress', 0.5778813362121582),
 ('camp yellow thunder', 0.5775296688079834),
 ('secret medicine societies', 0.5765159130096436),
 ('chisholm trail', 0.5751395225524902),
 ('brunot agreement', 0.574424147605896),
 ('morrisseau', 0.5742289423942566),
 ("turtle's war party", 0.5734469890594482)]
```

Native-authored workset

```
most_similar(model_sg,"trail of tears", topn=30) # sitting bull,
```

```
[('204 n3', 0.7087621688842773),
 ('70,71', 0.5988435745239258),
 ('wounded knee massacre', 0.5714915990829468),
 ('black hawk war', 0.5575634241104126),
 ('cherokee', 0.5571921467781067),
 ('de soto', 0.5566932559013367),
 ('terrible ordeal', 0.556323766708374),
 ('choctaw indians', 0.5543562173843384),
 ('death march', 0.5524013638496399),
 ('yount made', 0.552379846572876),
 ('cherokee removal', 0.5469951629638672),
 ('donner party', 0.5468500852584839),
 ('westward migrations', 0.5445303320884705),
 ('shawnee', 0.541441023349762),
 ('vernon county', 0.536487340927124),
 ('cherokee nation', 0.5335493683815002),
 ('ordeal', 0.532654881477356),
 ('yatasi were adai', 0.5321779847145081),
 ('chattahoochee river', 0.5320509672164917),
 ('rallies against', 0.5312162041664124),
 ('indian territory', 0.5297058820724487),
 ('forced removal', 0.5295249223709106),
 ("battle of stillman's run", 0.5294533371925354),
 ('few included within', 0.5290290117263794),
 ('having instituted', 0.5285264849662781),
 ('forced relocation', 0.52825528383255),
 ('have noved', 0.5280379056930542),
 ('reverend watt spade and willard walker', 0.5278054475784302),
 ('prominent researchers whose', 0.5272254943847656),
 ('distillery have', 0.5258493423461914)]
```

Comparative workset

Take-aways

- Although Native Studies rarely foregrounds quantitative techniques, we see here that DH quantitative techniques confirm at least some of the things we know about Native writing.

- Computer-generated methods are incredibly helpful for handling larger datasets. However, manual screening is still always necessary to get accurate results.
  - Screening HathiTrust author lists
  - Cleaning Keyword in Context results
  - Standardizing data (tribal affiliation, language ID)
  - Adding missing information (e.g. chapter vs. edited volume)

How does ChatGPT respond to our research questions?

ChatGPT

How is the coverage of Native American authored works in HathiTrust?

As an AI language model, I do not have access to the most current or comprehensive data on the coverage of Native American authored works in HathiTrust, but I can provide some general information.

HathiTrust is a digital repository that provides access to a large number of books, articles, and other materials. While HathiTrust includes many resources related to Native American studies, including books written by Native American authors, it is unclear how many of these works were authored by Native Americans. One challenge in determining the coverage of Native American authored works in HathiTrust is the lack of consistent metadata and the fact that not all works include information on the author's ethnicity or tribal affiliation.

However, some organizations and individuals have conducted research on the representation of Native American authors in HathiTrust and other digital repositories. For example, the Tribal Libraries Program of the Institute of Museum and Library Services has partnered with the American Indian Library Association and other organizations to increase the visibility and accessibility of Native American-authored works in digital collections. Some scholars have also conducted research on the representation of Native American authors in digital repositories, but these studies tend to focus on specific repositories or time periods.

Overall, while it is difficult to determine the precise coverage of Native American-authored works in HathiTrust, efforts are underway to increase the visibility and accessibility of these materials in digital repositories.

What are the characteristics of the Native works corpus? What does this tell us about representation among Native authors, genres?

The Native works corpus in HathiTrust is characterized by a relatively small number of works, with a total of 8,190 items. This is compared to the total corpus in HathiTrust, which contains millions of items.

In terms of genres, the Native works corpus is diverse, with works representing a range of genres including fiction, non-fiction, poetry, and memoir. There is also a strong representation of works related to traditional Native American practices, including tribal histories, language, and culture.

In terms of representation among Native authors, the corpus reveals that the majority of works were authored by non-Native writers, particularly in the early 20th century. However, there is a notable increase in the number of Native authors represented in the corpus from the 1970s onwards, reflecting a growing interest in promoting Native voices and perspectives in literature.

Overall, the characteristics of the Native works corpus in HathiTrust suggest that while there is a significant amount of representation of Native American authors and genres, there is also a need for continued efforts to ensure that Native voices are heard and represented in literature.

What are the characteristics of the writing styles of Native authors? How do they compare with non-Native authors when writing on similar subjects?

There is no single writing style that characterizes all Native authors, as there is significant diversity among Native communities and their literary traditions. However, some common themes and stylistic elements can be identified.

One characteristic of Native literature is the use of storytelling to convey cultural and historical knowledge. Many Native authors draw on traditional oral storytelling techniques, such as repetition, parallelism, and metaphor, to create narratives that connect contemporary experiences to ancestral traditions.

Another characteristic of Native literature is a focus on the relationship between humans and the natural world. Many Native authors use descriptions of the land, animals, and plants to create a sense of place and to convey cultural and spiritual values.

Native authors also often engage with issues of identity, representation, and historical trauma, exploring the complexities of living in a society that has historically marginalized Native peoples.

When compared to non-Native authors writing on similar subjects, Native authors may bring different cultural perspectives and experiences to their writing, which can shape their storytelling styles and thematic concerns. Additionally, Native authors may challenge dominant narratives and bring attention to issues that have been overlooked or ignored in mainstream literature.

Thanks for your attention! Questions & Comments

Contact me at kunlu@ou.edu