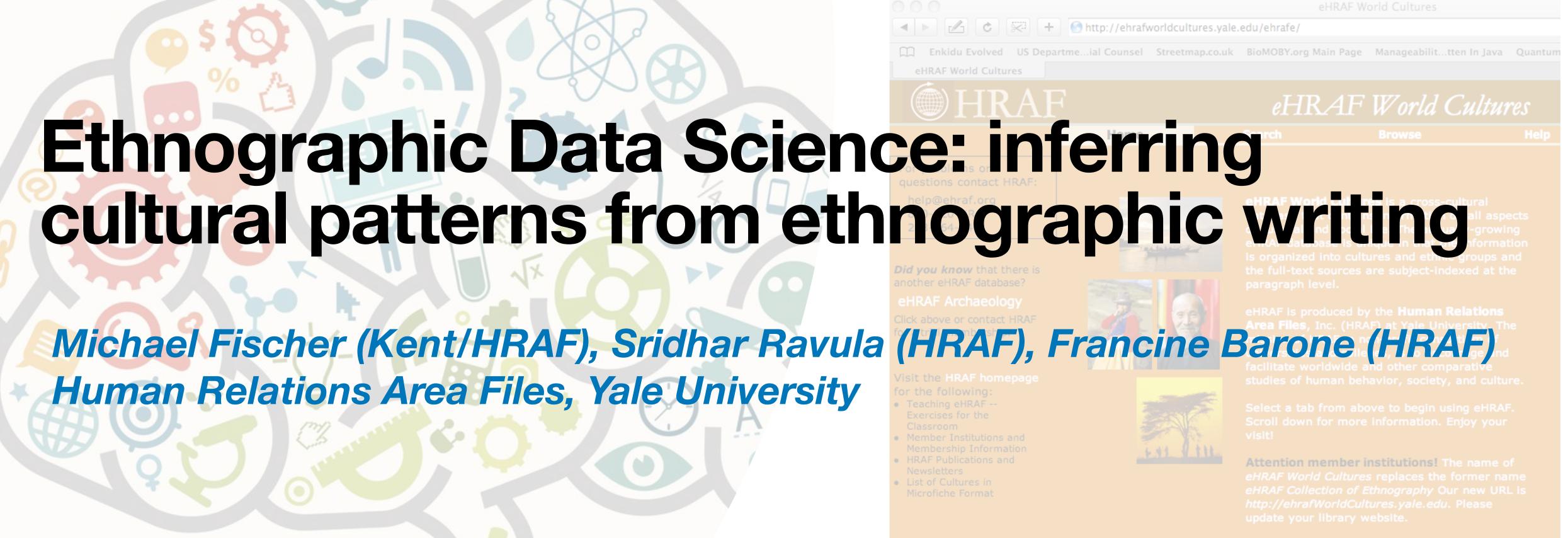
examples/download: https://iklews.hraf.net/

iKLEWS - Infrastructure for Knowledge Linkages from Ethnography of World Societies - National Science Foundation *Human Networks and Data Science Infrastructure* Programme Award#2024286.



HRAF: https://hraf.yale.edu

Michael Fischer: mike@hrafarc.org

Sridhar Ravula: sridhar.ravula@yale.edu

Francine Barone: francine.barone@yale.edu



Unravelling Theories in Ethnography: Data Science in the HRAF Collection of Ethnography

- Culture is a complex concept widely used outside of anthropology, in many related social sciences and humanities subjects. Despite this, anthropologists generally cannot agree on a precise definition of culture, nor do those who do define it agree with each other.
- One reason for this difficulty is that culture is not defined by a single process or system.
 Instead, it is the outcome of the interaction of many aspects of human cognition,
 emotions, and organization. These aspects include communication, learning, adaptation,
 representation, and transformation.
- In short, what anthropologists (and increasingly others) now refer to as culture is an
 emergent phenomenon (or perhaps even an apparent category of phenomena),
 not a causal or explanatory system. It is the result of the interaction of different systems
 that are, at least in part, independent of each other other than through that interaction.
 Many anthropologists have abandoned a concrete concept of culture altogether.

Unravelling Theories in Ethnography: Data Science in the HRAF Collection of Ethnography

- George P. Murdock, in his essay "Anthropology's Mythology," (1971) argued that culture and social structure should not be treated as real things that explain human behaviour. Instead, they are **our** way of describing patterns of interactions between individuals. Murdock was advocating for a program of theory that would focus on the diversity of individual experience and choice, rather than on commonality and conformity.
- Marvin Minsky, in his book The Society of Mind, wrote: "What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle." Minsky was referring to the diversity of principles within a single mind. To represent the diverse principles underlying cultural systems, we might conceptualise culture as "the community of minds."
- All modern anthropologists dispute any conception that culture can be represented in terms of static structures; culture is dynamically enacted, constituted differently by different culture-enacting agents, but with results that are comprehendible, if not acceptable, to other agents. Understanding how cultural systems are distributed within a population in such a way that most people can agree on what aspects people associate with culture are critical.



How does cultural theory help us understand events and allow us to generalise?

- Culture is a complex and contested concept, but it is still useful for understanding human behavior. Culture is not a reified thing, but rather a process of knowledge production and reproduction.
- Fischer (2006) identifies one of the key causes of the tensions between those who see culture as structured and patterned and those who see it as performative and emergent is that culture cannot be understood by treating an instantiation of culture as if it were purely the manifestation of one or more underlying principles.
- Using the deontic logic of permissions and obligations rather than the imperative logic of possibility and necessity, Fischer argues that domain knowledge need not be true, it need only be enabling or effective. Transforming information or experience into knowledge is a role associated with culture, but people embedded in a culture have many ways of carrying out these transformations.

How does cultural theory Help us understand events and allow us to generalise?

- Chit Hlaing (2006) objects to extreme denials of structure or our ability to codify human knowledge because despite the intellectual exercises that unequivocally demonstrate that there is no reified body of culture, every anthropologist who has done fieldwork can confirm that there is indeed some rather large body of [apparently] shared beliefs, practices, and knowledge that somehow maintains consistency and is recognisably the same despite constant change.
- Fischer argues that whatever else anthropologists might argue, we are on very thin ground if we deny the patently obvious fact that collectives of people are able to reproduce something recognisably of the same behavioural domain as themselves and it is difficult to see how that could happen in the absence of a process from which cultural systems emerge.
- Fischer and Hlaing conclude that we should avoid trying to define culture in a precise way, and instead focus on how culture works. They suggest that culture is made up of shared beliefs, practices, and knowledge that are passed down from generation to generation. These shared beliefs, practices, and knowledge are not always consistent or logical, but they are nevertheless real and they have a significant impact on our lives. A focus on the process is necessary.



Identifying patterns acquired by anthropologists ... exploring ethnography

- Ethnography is our main source for retrieving both 'fact' and 'theory' in anthropology. Most ethnography is the product of anthropologists, although there are many contributions from others, including a growing list of academic practitioners, as well as those who produce ethnographic writing and accounts as a part of their work outside academia.
- However much one might contest the fidelity of ethnography in recording the beliefs, understandings and knowledge of the groups of people who are the subject of an ethnography, it is, at least, a record of what a particular ethnographer believed, understood and came to know about that group of people.
- A study of ethnography can then serve as data for understanding ethnographers, if not the indigenous people who were observed by the ethnographer.

Identifying patterns acquired by anthropologists ... exploring ethnography

- At HRAF we have been working for some years to improve the ability of our membership to use ethnographic sources for secondary research.
 Our latest project was to employ developments in representing texts computationally developed over the past 75 or so years, but especially the past two decades which have seen a revolution in computational methods for working with textual sources.
- Our overall approach is to advance secondary ethnographic research to the level that ideas, concepts and hypotheses can be transparently and systematically evaluated by keeping as close as possible to the ethnographic works, while being able to look across the range as well.



IKLEWS

(Infrastructure for Knowledge Linkages from Ethnography of World Societies)

- iKLEWS is a Human Relations Area Files (HRAF) project underwritten by the National Science Foundation *Human Networks and Data Science Infrastructure* programme Award#2024286.
- iKLEWS applies data science to develop semantic infrastructure and ethnographic research services for a growing ethnographic database (eHRAF World Cultures),
 - roughly 800,000 pages from
 - 7,000 ethnographic documents covering
 - 365 world societies (and growing), each at several time points in the ethnographic present.



iKLEWS

(Infrastructure for Knowledge Linkages from Ethnography of World Societies)

- We aim to support researchers who seek to understand the range of possibilities for human understanding, knowledge, belief and behaviour:
 - to address work in anthropological theory as it pertains to ethnography, from humanistic and scientific perspectives
 - to explore the relationship between human evolution and human behaviour,
 - to inform real-world problems we face today, such as: climate change;
 violence; disasters; epidemics; hunger; and war.



iKLEWS

(Infrastructure for Knowledge Linkages from Ethnography of World Societies)

- Improve interoperability between external web resources and databases by creating a HRAF services API (application programming interface).
- Develop services and tools to work with our databases for researchers casual to expert:
 - tools to broaden and narrow search with greater insight into meaning
 - tools to summarise, visualise and navigate the contents of large search results amounting to hundreds or thousands of pages.
 - tools to extract structured data from ethnographic text





Human Relations Area Files

- Founded 1949
- Mission: to encourage and facilitate the crosscultural study of human culture, society, and behavior in the past and present.
- Curates knowledge of day to day life of peoples of different cultures recorded in ethnographic writing.
- Initially using paper now digital.
- Key metadata -
- Ethnonyms Outline of World Cultures OWC
- Descriptors Outline of Cultural Materials OCM



Human Relations Area Files

- Since inception the HRAF collection of ethnography has included topical metadata for each entry in each document.
- These entries roughly correspond to paragraphs, but may include images, figures, lists, tables, etc.
- We refer to entries as Search and Retrieval Elements, or SREs.
- Each SRE in each ethnographic work is assigned classificatory terms by an anthropologist, one or more of 790 drawn from **O**utline of **C**ultural **M**aterials (**OCM**)
- OCMs are organised as major and minor topics.



Human Relations Area Files: Outline of World Cultures

owc \$	EHRAF WORLD CULTURES NAME	REGION \$	SUBREGION \$	SUBSISTENCE + PSF + SRS + TYPE	sccs \$
SI04	Abipón	South America	Southern South America	hunter-gatherers	Yes
RI03	Abkhazians	Asia	Caucasus	pastoralists	Yes
NK04	African Americans	North America	Regional and Ethnic Cultures	commercial economy	
AB06	Ainu	Asia	East Asia	hunter-gatherers	Yes
					H



Human Relations Area Files: Outline of Cultural Materials

→ 150 BEHAVIOR PROCESSES AND PERSONALITY

151 SENSATION AND PERCEPTION

152 DRIVES AND EMOTIONS

153 MODIFICATION OF BEHAVIOR

154 ADJUSTMENT PROCESSES

155 PERSONALITY DEVELOPMENT

156 SOCIAL PERSONALITY

157 PERSONALITY TRAITS

158 PERSONALITY DISORDERS

159 LIFE HISTORY MATERIALS

→ 430 EXCHANGE AND TRANSFERS

431 GIFT GIVING

432 BUYING AND SELLING

433 PRODUCTION AND SUPPLY

434 INCOME AND DEMAND

435 PRICE AND VALUE

436 MEDIUM OF EXCHANGE

437 EXCHANGE TRANSACTIONS

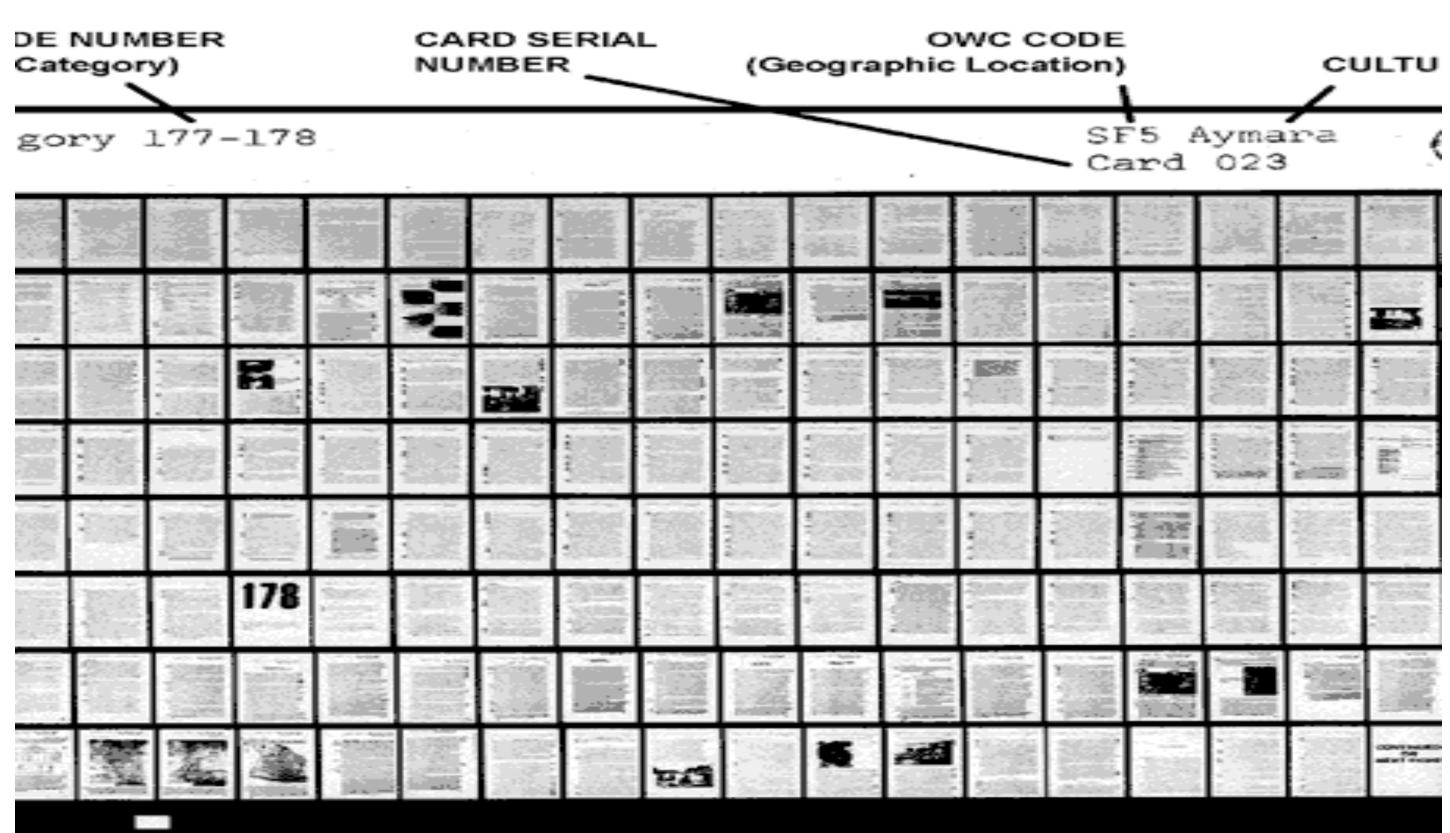
438 INTERNAL TRADE

439 EXTERNAL TRADE

Technological Steps for HRAF



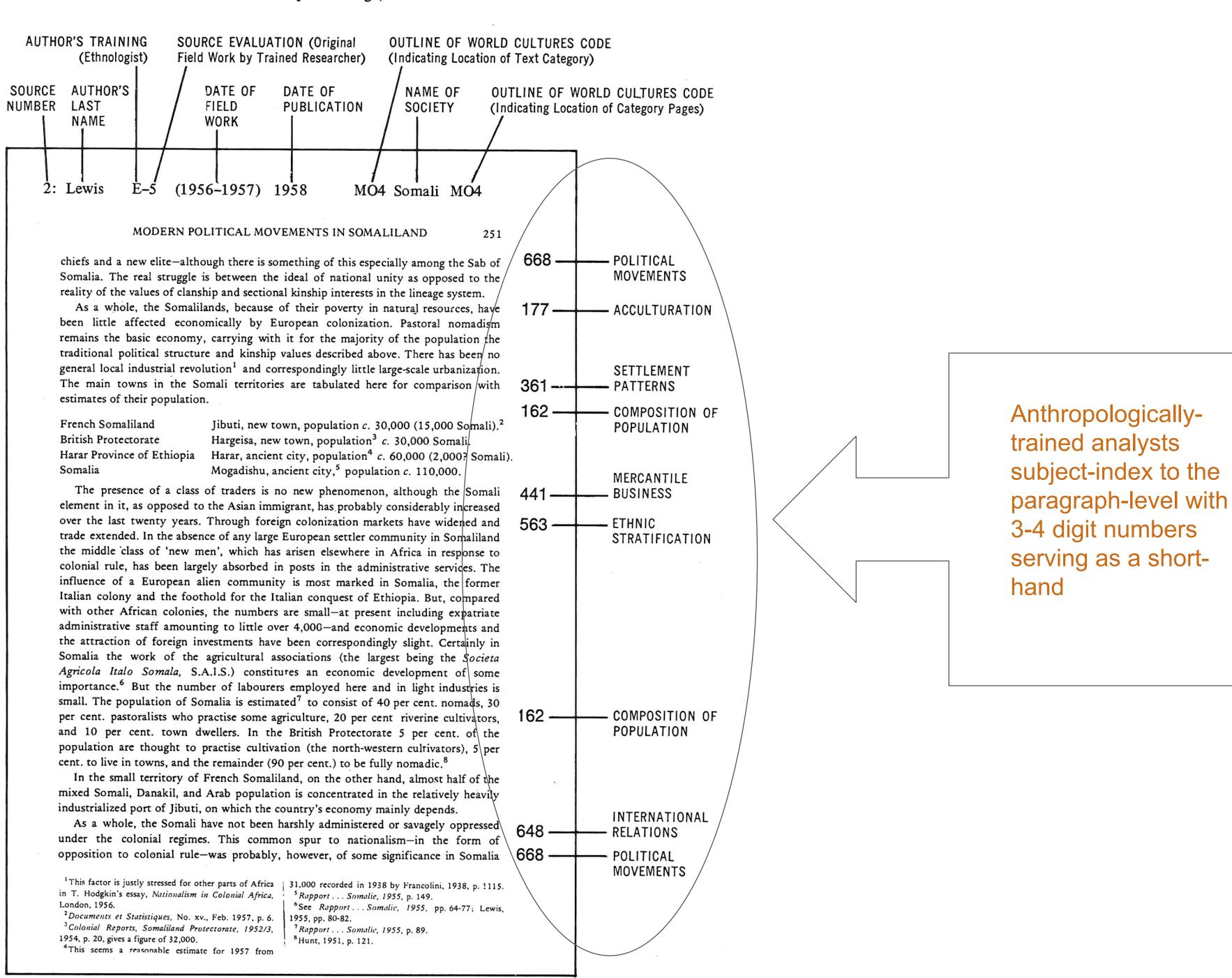
1949-1959. The file drawers organized by OWC contained full-texts and every subject category included all relevant pages from all documents



Microfiche cards-1959-1993



Sample File Page, Annotated



Sample of HRAF Text - Expert Judgements

```
613" dispocms="423 613">
   <p.ocm>
       423 613
   </p.ocm>
   These gifts, in addition to maintaining a balance between
   population and resources, enhance the potential of separate
   groups of children to split apart and form separate
   lineages. Lineages such as these may continue to reside in
   the same district or village and to maintain friendly and
   cooperative relations with each other. Goodenough (1950)
   refers to such a collection of lineages as a ramage. When
   Trukese females migrate to other villages or islands they
   may found separate but related lineages. Members of lineages
   so related may have the option of membership in either
   lineage. In such cases Goodenough (1950) refers to the
   collective entity as a sub-sib.
614 192" dispocms="613 614 192">
   <p.ocm>
       613 614 192
   </p.ocm>
   Finally, all the lineages on different islands which bear
   the same name consider themselves to be somehow related,
   though completely unable to trace the alleged relationship.
   Generally speaking, lineage members so related have tended
   to avoid marriage with each other, but to extend a degree of
   hospitality when visiting one another's home islands.
   Goodenough (1950) refers to each of these large groups of
   lineages as sibs. Traditionally, however, Trukese have not
   distinguished these several levels of lineage organization
   by the use of distinct lables. Although they have borrowed
   the term family (
   <highlight xml:id="or19-025-00726" rend="underline">
       faamenii
   </highlight>
```





• eHRAF:

- Fast at retrieving relevant ethnography
- Uses same methods as HRAF's paper files in 1949
- Few aids to analyzing material once found

• iKLEWS:

- Introduces new methods of working with text
- Deploys analytic tools and improves search
- Tools are open source and free to use (with limitations for nonmember institutions)
- Tools support researchers from beginner to advanced, through web apps and Jupyter notebooks



- Our infrastructure supports investigating a wide range of topics, such as kinship, social organization, and economics.
- We are applying artificial intelligence techniques like natural language processing and deep learning to extract patterns and linguistic analysis.
- We are developing methods to bridge the gap between opaque deep learning outcomes and more transparent logic driven narratives.
- Most public tools are based on pre-computed metadata and infrastructure so that researchers can operate interactively.





- KLEWS Outcomes:
 - Improved relevance of search results through identification of new and finergrained topics in SREs, in addition to OCM-associated topics
 - Semantic representations of SREs in the texts, enabling effective following of topic trails
 - Tools for management, analysis, visualization, and summarization of results, researcher-initiated data mining and pattern identification
 - Assist researchers to identify and test hypotheses about the societies they investigate
- Researcher Access:
 - Access to data and analytic capabilities directly through a Jupyter notebook run on the researcher's computer, or using a web application such as Kaggle or Google's Collaboratory
 - Web-based tools will be released and further developed in the coming years





- One obvious issue with data science and NLP methods is that these are:
 - Based directly on statistical methods, and each thus is associated with a probability of error, often 20% or more
 - Based on non-deterministic 'deep learning' through adaptations of ML, including neural networks
- The latter are more or less opaque with respect to internal operation. Only outcomes are explicable, and these imbue substantial margins of error in application, 10-40%, which is regarded as 'good-enough' by much of the applied community.
- For research a little more care is needed.



- We can understand the process of algorithmic application in general terms, but not in fine detail, due to the complexity of evolving relationships.
- However, we can document and apply the limits of analytic processes and error rates to our work, similar to how expert ethnographers deal with the limits of their own precision.



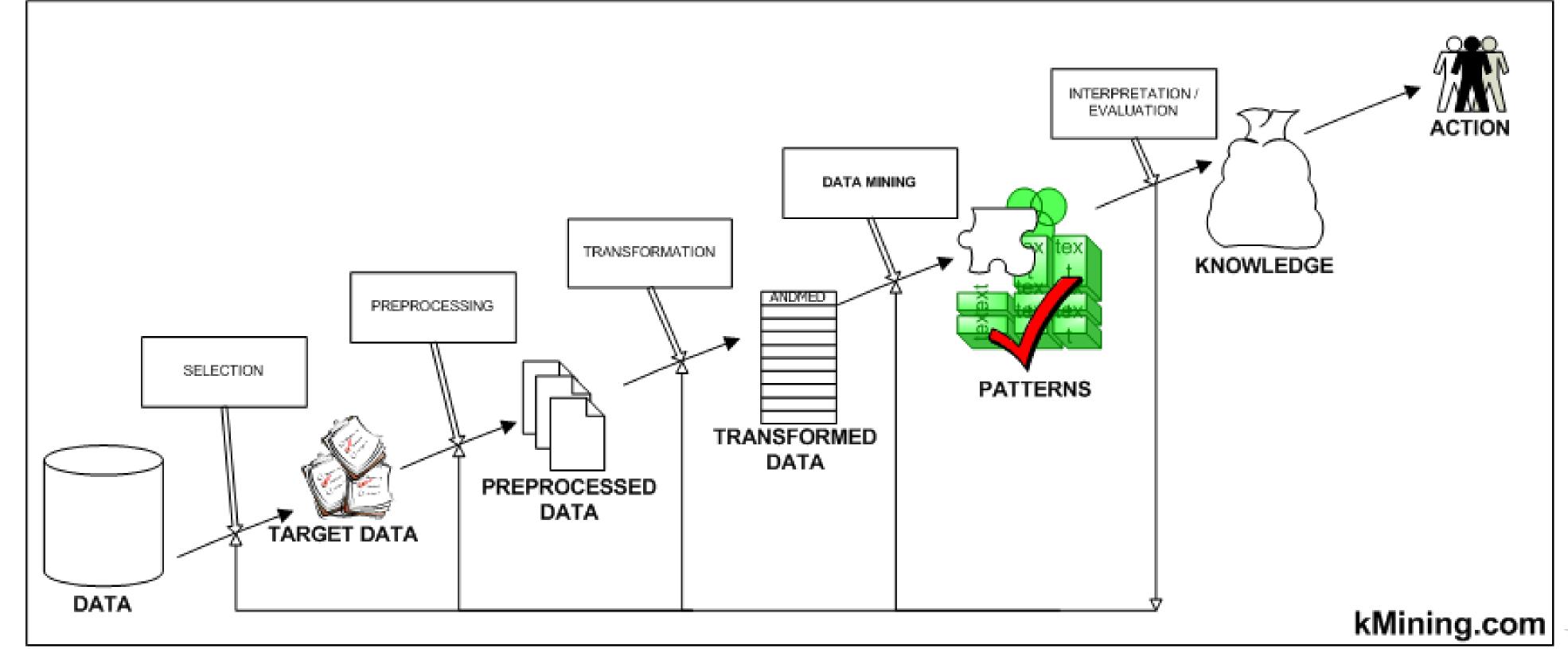


- We must remind ourselves of the limitations of algorithmic procedures and ethnographic content, as these may be overconfident or overly suspicious of the results.
- Ethnographic conceptions and perceptions change over time, and the content of ethnographies can be influenced by the authors' perspectives and biases.
- We can provide researchers with tools and procedures to help them evaluate ethnographic content critically, rather than providing specific guidance on specific documents.



DETECTING ORDER

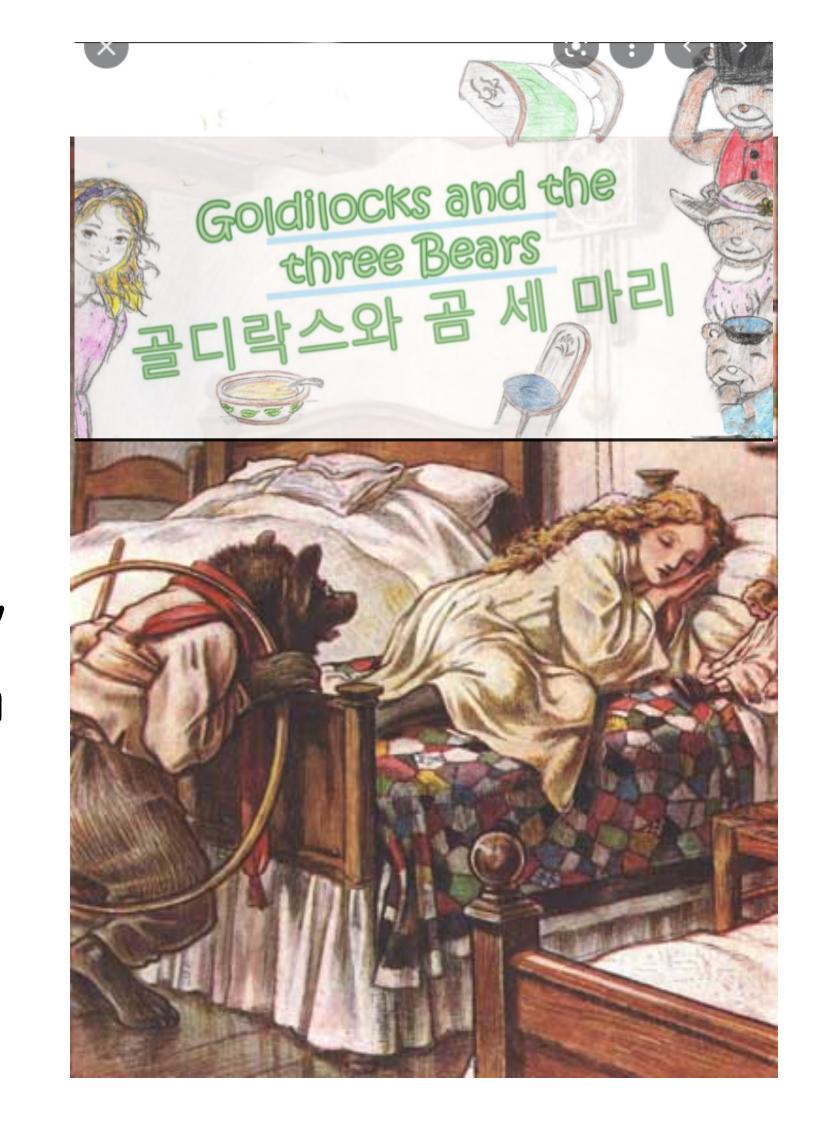
- Data mining knowledge from information
 - Collecting Data
 - Transforming Data
 - Lots of working approaches for identifying patterns in data





Identifying significance

- ML/NLP evaluate the importance of a word is to other words in a collection or corpus.
- Importance increases proportionally to the frequency of word clusters that appear in the ethnographies.
- NLP and ML tools identify patterns that emerge from relations between word clusters.



Goldilocks words - not too common or too rare



Culture

Word2Vec: method based on neural networks Intelligence Myth

Word	Similarity	Word	Similarity
cultural_tradition	0.86	intellect	0.77
cultural_pattern	0.84	capability	0.75
cultural	0.82	shrewdness	0.74
traditional_culture	0.82	aptitude	0.72
other_culture	0.82	talent	0.72
native_culture	0.81	intelligent	0.72
world_view	0.81	cleverness	0.72
subculture	0.80	superior_intelligence	0.71
cultural_system	0.79	intellectual_ability	0.71
western_culture	0.78	wisdom	0.70



Similarity

0.89

0.85

0.84

0.84

0.83

0.82

0.80

0.79

0.77

0.77

Word2Vec: Magic and Power: Good vs Bad

Word	Similarity	Word	Similarity
magical	0.72	black_magic	0.80
charm	0.69	sorcery	0.75
magical_power	0.69	evil	0.74
object	0.65	witchcraft	0.72
magic_power	0.64	sorcerer	0.70
magical_charm	0.64	witch	0.70
supernatural_power	0.64	evil_spirit	0.70
medicine	0.64	spell	0.70
purpose	0.64	malignant	0.69
requisite	0.63	magical	0.69

Word	Similarity
ability	0.75
prestige	0.74
wealth	0.72
capacity	0.72
political_power	0.71
control_over	0.70
not_only	0.69
spiritual_power	0.69
benefit	0.69
advantage	0.69

Word	Similarity
powerless	0.71
evil	0.69
fear	0.68
danger	0.64
evil_power	0.64
cause	0.64
threat	0.62
machination	0.62
malevolence	0.62
supernatural_power	0.61

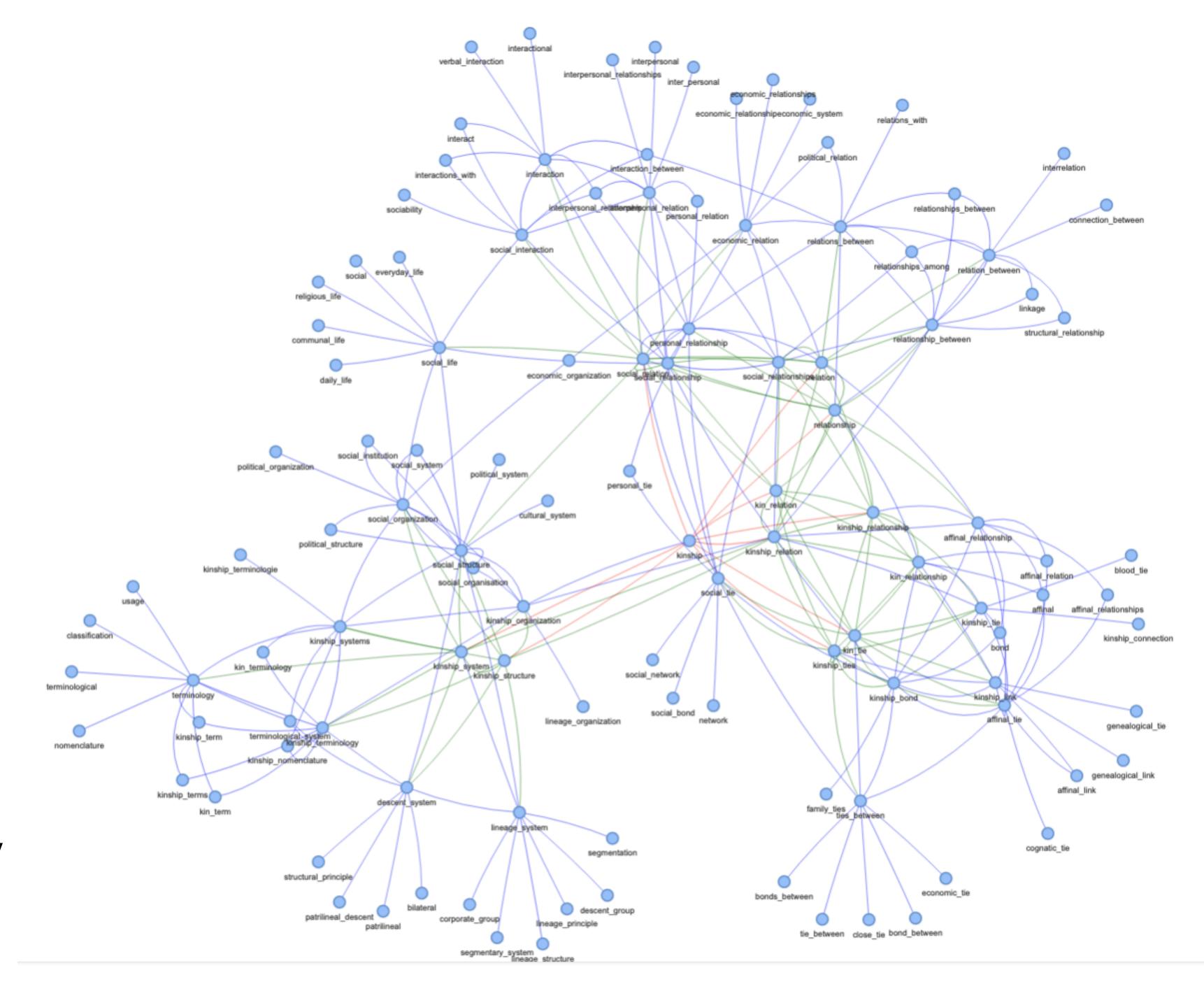
Example: Word2Vec - Love

Love		Positive lov	Positive love			shy love	
Word	Similarity	Word	Similarity	Word	Similarity	Word	Similarity
affection	0.82	affection	0.75	hate	0.78	affection	0.74
he_love	0.78	pleasure	0.72	pity	0.70	like	0.74
she_love	0.77	devotion	0.69	unhappy	0.69	admire	0.72
compassion	0.75	generosity	0.67	anger	0.67	pleasure	0.71
pity	0.75	desire	0.67	grieve	0.66	affectionate	0.71
hate	0.72	enjoyment	0.66	jealous	0.66	happy	0.69
lover	0.71	companionship	0.66	shame	0.66	he_love	0.68
passion	0.71	satisfaction	0.65	angry	0.65	fond	0.68
pleasure	0.70	faithfulness	0.64	wicked	0.65	polite	0.68
happy	0.69	tender_affection	0.64	torment	0.65	good_natured	0.67



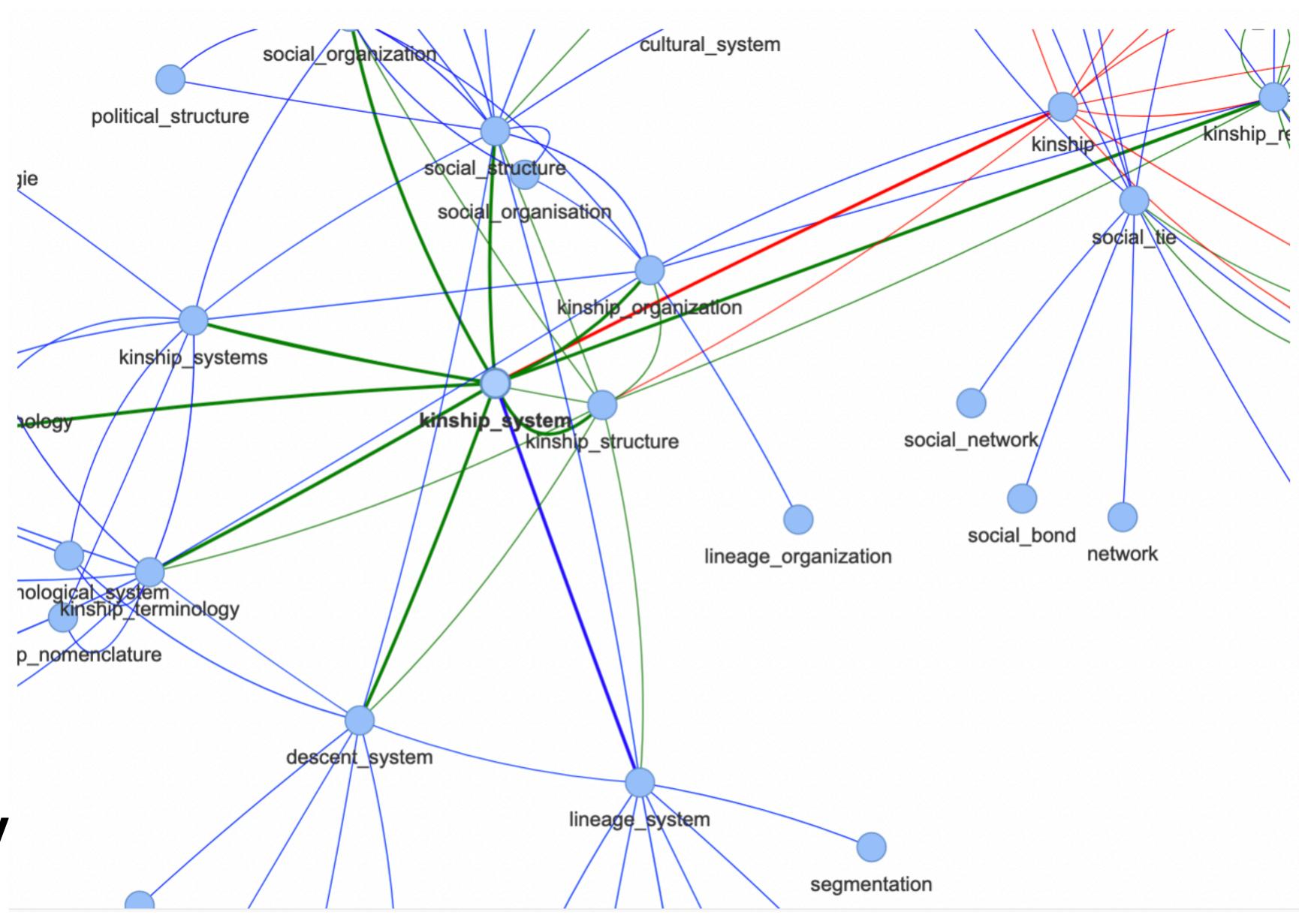
Network of words recursively seeded from Kinship

An ethnographic concept map of how 5000 anthropologists have represented kinship and related terms in ethnography



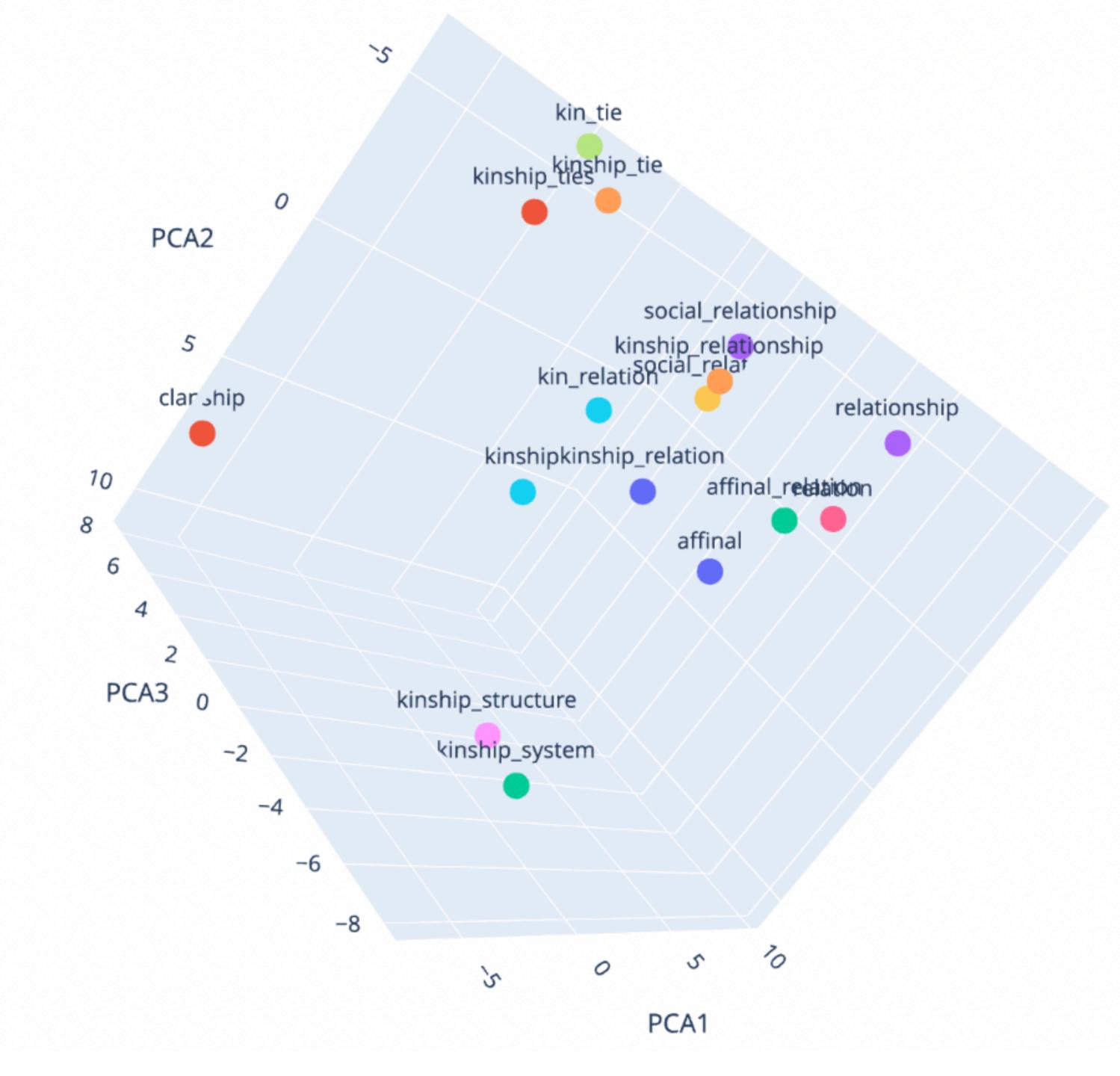
Network of words recursively seeded from "Kinship"

An ethnographic concept map of how 5000 anthropologists have represented kinship and related terms in ethnography



3d Principle Components Analysis Plot of Word2Vec seeded word "Kinship"

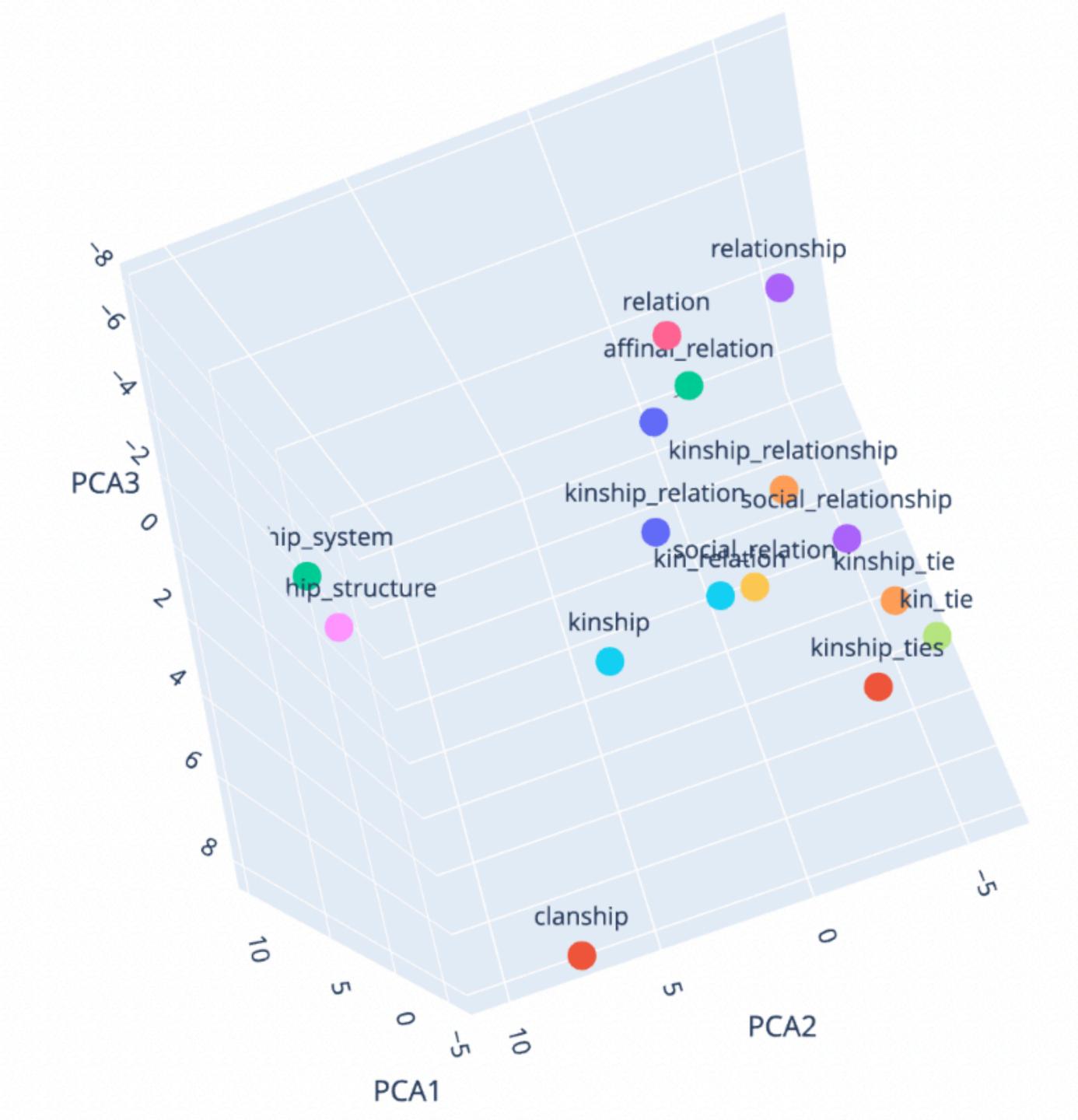
Relations between terms on 3 constructed dimensions.



3d Principle Components Analysis Plot of Word2Vec seeded word "Kinship"

Relations between terms on 3 constructed dimensions.

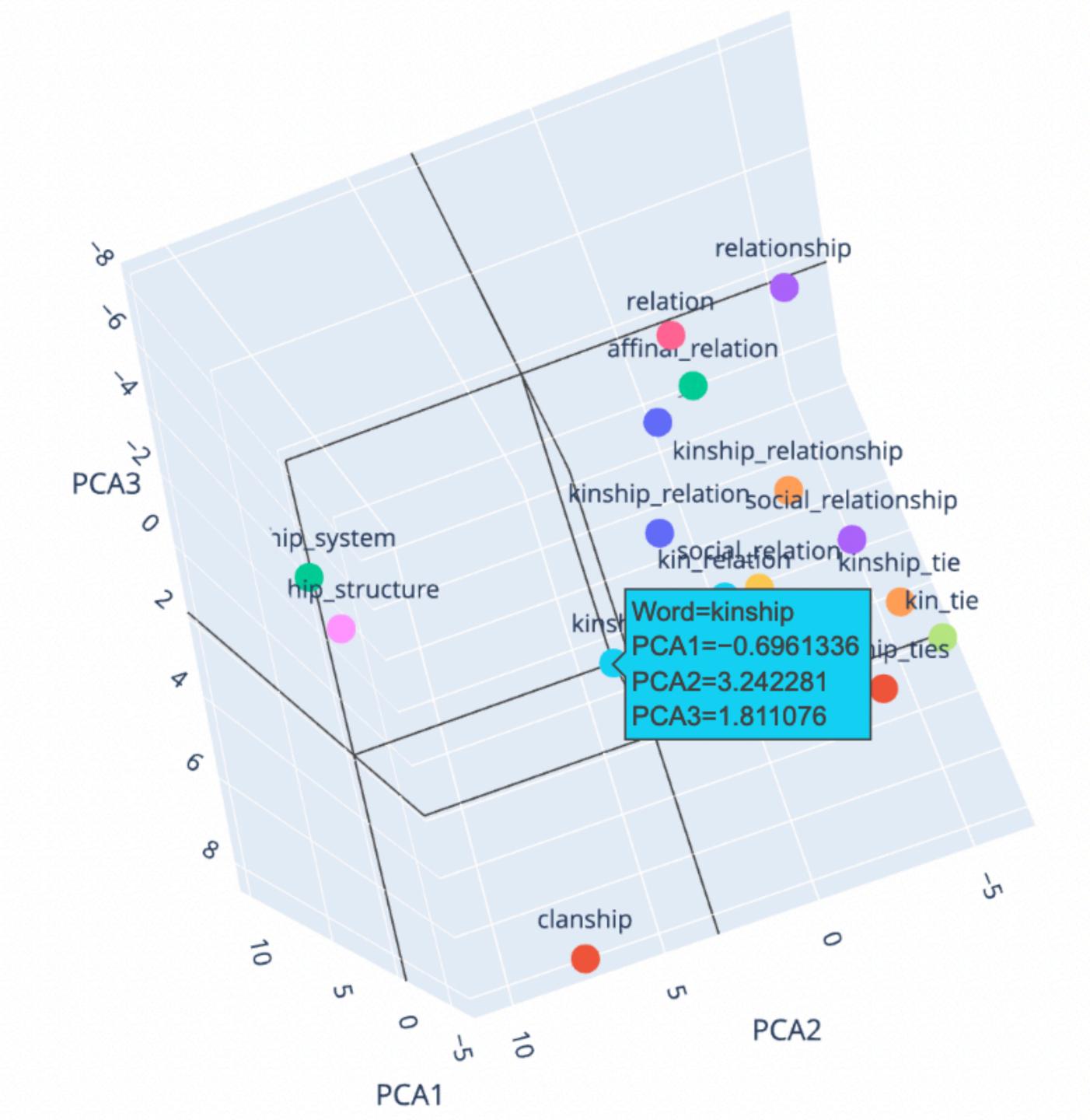
Interactive rotation



3d Principle Components Analysis Plot of Word2Vec seeded word "Kinship"

Relations between terms on 3 constructed dimensions.

Select data on point



Ethnoword: Principal Components Analysis to identify combinative relationships

Filter products of words seeded from "Kinship" by Word2Vec

```
Max
                            Min NegScore
                                                 NegWord PosScore
                                                                          PosWord
o social_structure
                                 0.815142
                                          blood_relationship 0.854422
                                                                   social_organization
                    consanguinity
     relationship
                                 0.755579
                                                           0.83414
                        cognatic
                                                  bilateral
                                                                            relation
        clanship
                       siblingship
                                 0.709837
                                         kinship_relationship
                                                                         kinship_ties
2
                                                          0.696202
                                 0.697407
      siblingship
                                                          0.728855
                                                                         kin_relation
                          affinity
                                                connection
    consanguinity
                   kin_relationship
                                 0.809703
                                             kinship_relation
                                                          0.721294
                                                                            affinity
      siblingship
                  cognatic_kinship
                                 0.751893
                                            bilateral_kinship
                                                           0.71063 kinship_relationship
5
         affinal
                  cognatic_kinship
                                 0.708068
                                          social_organization
6
                                                                             affine
                                                          0.820232
                                                          0.752203 group_membership
       affiliation
7
                    consanguinity
                                  0.65531
                                                   affinity
     relationship kinship_organization
                                 0.753275
                                          social_organization
8
                                                          0.815027
                                                                            relation
                kinship_relationship
                                             kinship_relation
                                                                       consanguinity
9
                                0.806044
                                                          0.669076
               social_structure Δ consanguinity = blood_relationship
                                                                                  0.815 )
kinship
kinship
               consanguinity \Delta social_structure = social_organization
                                                                                 ( 0.854 )
kinship
               relationship \Delta cognatic = bilateral (0.756)
kinship
               cognatic \Delta relationship = relation (0.834)
               clanship \Delta siblingship = kinship_relationship (0.71)
kinship
               siblingship \Delta clanship =
kinship
                                              kinship_ties ( 0.696 )
kinship
               siblingship \Delta affinity =
                                              connection (0.697)
               affinity \Delta siblingship =
                                              kin_relation ( 0.729 )
kinship
kinship
               consanguinity \Delta kin_relationship =
                                                         kinship_relation
kinship
               kin_relationship ∆ consanguinity
                                                     = affinity ( 0.721 )
               siblingship \Delta cognatic_kinship = bilateral_kinship
kinship
                                                                             ( 0.752 )
kinship
               cognatic_kinship ∆ siblingship
                                                      kinship_relationship ( 0.711 )
kinship
               affinal \Delta cognatic_kinship =
                                                  social_organization ( 0.708 )
               cognatic_kinship \Delta affinal =
                                                  affine ( 0.82 )
kinship
                                                   affinity ( 0.655 )
kinship
               affiliation ∆ consanguinity
               consanguinity \Delta affiliation = group_membership (0.752)
kinship
               relationship \Delta kinship_organization = social_organization
kinship
                                                                                     ( 0.753 )
               kinship_organization △ relationship
                                                         = relation (0.815)
kinship
               affinity Δ kinship_relationship =
kinship
                                                        kinship_relation ( 0.806 )
               kinship relationship \Delta affinity =
                                                        consanguinity (0.669)
kinship
```

Word2Vec: Finding ethnographic support from the text.

Clicking in cell lists
SREs from eHRAF by
relevance to the two
categories. Items can
be retained for further
analysis of specific text.

	kinship	marriage	engagement	kin	ancestor	exchange	clan	politics	lineage	descent	alliance	assurance
kinship	35425	7339	386	5495	1685	2058	3703	4050	3345	3154	970	198
marriage	7339	82074	1984	5455	2400	5433	5636	4120	4535	3413	2753	683
engagement	386	1984	22784	449	331	1078	542	1792	356	176	247	255
kin	5495	5455	449	28614	1330	1967	2570	2197	2605	1741	621	268
ancestor	1685	2400	331	1330	32098	634	3638	1674	3514	2612	312	330
exchange	2058	5433	1078	1967	634	32909	1384	2410	1068	635	1012	305
clan	3703	5636	542	2570	3638	1384	48812	3974	5922	2862	787	301
politics	4050	4120	1792	2197	1674	2410	3974	92416	3964	2306	2531	743
lineage	3345	4535	356	2605	3514	1068	5922	3964	29417	3234	744	213
descent	3154	3413	176	1741	2612	635	2862	2306	3234	18372	683	123
alliance	970	2753	247	621	312	1012	787	2531	744	683	10568	156
assurance	198	683	255	268	330	305	301	743	213	123	156	11549

alliance By lineage

Prev - Entries 6 to 10 of 744 - Next

✓ Malays(an05) - #602 #613 #782 - A share of the harvest: kinship, property, and social history among the Malays of Rembau - Michael Gates Peletz (1988) pp. 212 - Rembau district, State of Negeri Sembilan, Malaysia: 1830-1980 [p] - 0.04761905 - Prev - Next Related

Whether these latter concomitants of kenduri were more or less pronounced in decades past is, unfortunately, difficult to gauge. Some villagers claim that invitations to kenduri are more selectively extended now than a few decades ago, and that kenduri have become highly politicized. Although such claims are true, they pertain more to solidifications of cleavages and alliances among lineages (and factions formed around their principal luminaries) than to dissensions within lineages. This fact, too, will become quite apparent when we examine the feud between the Hill and Valley lineages.

✓ Kogi(sc07) - #286 #341 #346 #354 #772 #821 - The Great Mother and the Kogi universe: a concise overview - G. Reichel-Dolmatoff (1987) pp. 108 - Sierra Nevada de Santa Marta: 1950 - 1980 [p] - 0.04761905 - Prev - Next Related

At the apex of some of the principal temples, mainly in the ceremonial centers, there is a small orifice which usually is covered with a potsherd. Occasionally, at a solstice date, the priests will have the orifice uncovered and then a narrow beam of sunlight will penetrate the smoke-filled gloom, and will fall upon one of the hearths; hence the small brilliant disk will wander to the next hearth, in the same half of the temple (northern or southern) and reach it in about six hours. In the course of one solar year the beam of sunlight, after having traveled continuously across the square delimited by the four hearths, will arrive at the other half of the temple. The Kogi see in this the manifestation of Sun the Weaver on the temple floor which, in this imagery, is said to be a loom. It follows, in Kogi logic, that at night Black Sun weaves the "other" side of the fabric, in the invisible, inverted half of the temple. In another image, the Kogi will say that the sun-beams, both visible and invisible, link the different lineages into alliances and that, as the men rotate four times in a year, the solar clock-work traces a network of alliances and oppositions. Less ...

✓ Kogi(sc07) - #772 #776 #788 #821 - The Great Mother and the Kogi universe: a concise overview - G. Reichel-Dolmatoff (1987) pp. 108 - Sierra Nevada de Santa Marta: 1950 - 1980 [p] - 0.1 - Next Related

Lord Mulkuëxe, identified with the sun in the temple, sits in the center and holds in his hand a circular mirror of stone, facing upward. When the Sun Father (Máma Nyúi) stands in the zenith, on the equinoxes, he also holds a mirror, but facing downward. Between these two mirrors, Kogi priests will say "the sun sees everything." In a ritual which is still being performed in a few temples, a priest will sit in the sacred center, holding a mirror and thus establishing a cosmic axis along which his prayers can ascend heavenward while Sun Father can fertilize the temple-womb.

Samai(an06) #121 Society and assmass Chawang of noninsular Malaysia Signa Hawall foreward by Padnay Noodham (1084) nn [n 1] Chawang 10

examples/download: https://iklews.hraf.net/

Conclusions

The presented tools and methods are mainly based on just one of the technologies we have incorporated into iKLEWS. It is, however, a very flexible base technology from which many tools can be constructed. Word2Vec extracts a model of how thousands of ethnographers have related the language they used in ethnography, a model we can use to further interrogate the ethnographies within HRAF. We also have a greatly expanded set of topics, tools for identifying related material based on semantic cues, tools for analysing texts using conventional methods, as well as newer methods based on Large Language Models. For the time being the latter are limited to larger supported projects, but this terrain is rapidly developing so that more and more can be done using standard desktop computers with limited technical assistance. As we make these facilities public over the coming year and beyond, we hope to see greater impact from secondary ethnographic research.