MATHEMATICAL ANTHROPOLOGY AND CULTURAL THEORY:
AN INTERNATIONAL JOURNAL
VOLUME 15 NO. 1 DECEMBER 2020

ARTIFICIAL INTELLIGENCE / MACHINE LEARNING RESEARCH USING THE AUSTRALIAN ABORIGINAL ALYAWARRA KINSHIP DATASET: PARTIAL BIBLIOGRAPHY 2004-2020

WOODROW W. DENHAM, Ph. D.
RETIRED INDEPENDENT SCHOLAR
WWDENHAM@GMAIL.COM

SUBMITTED: October 1, 2020 ACCEPTED: December 2, 2020

MATHEMATICAL ANTHROPOLOGY AND CULTURAL THEORY: AN INTERNATIONAL JOURNAL
ISSN 1544-5879

ARTIFICIAL INTELLIGENCE / MACHINE LEARNING RESEARCH USING THE AUSTRALIAN ABORIGINAL ALYAWARRA KINSHIP DATASET: PARTIAL BIBLIOGRAPHY 2004-2020

Woodrow W. Denham, Ph. D.

Abstract. *This paper describes methods used at the interface between anthropology and machine learning research. Charles Kemp, a graduate student at MIT in 2004, discovered my numerically coded Alyawarra kinship term applications data (Denham 1973; Denham, McDaniel and Atkins 1979; Denham and White 2005) and received my permission to use the data in his machine learning research. Since then, his co-authored papers (Kemp et al. 2004, 2006, 2010), and other works that cite his papers and mine, have played significant roles in the development of unsupervised pattern detection and machine learning technology as subsets of Artificial Intelligence research. Part 1 of the paper outlines how I produced the Alyawarra (Alyawara) kinship term applications dataset and introduces the structure and content of the dataset and supporting files. Part 2 briefly describes some simple ways to analyze the dataset either manually or with machine learning technology. Minimally these examples demonstrate some ways in which the ethnographic dataset is useful to the machine learning community now. More speculatively, the machine learning technology introduced here may enhance ethnographic research in the future. Part 3 provides links to a sample of 24 papers by Kemp et al. and other AI colleagues, all of which utilize the Alyawarra Kinship dataset. Part 4 contains links to some of my Alyawarra kinship data and documentation files that are available online. Part 5 briefly acknowledges support that I have received for this project over the last half-century.*

1. **Constructing the Alyawarra Kinship dataset.** In 1971-72. I used the following process to collect numerically coded kinship data with Alyawarra speaking people of Central Australia. a) I photographed 225 members of my 366-member research population, most of whom were closely related to each other by descent or marriage or both. b) I elicited and recorded all known genealogical relations among all those people, plus their sexes, ages, sections, descent lines, marital statuses, residential group compositions and other demographics. c) I collected and operationally defined all 26 of the kinship terms (equivalent to mother, father, sister, myself, etc.) that they could use to address or refer to each other. The 26 numerically coded kinship terms are defined in Denham (1973) and Denham et al. (1979). d) I asked 104 egos (i.e.,

carefully selected representatives of all descent lines and demographic categories) to tell me the single "best" kinship term[1] that they could use to refer to each of the other 224 alters (i.e., other photographed members of the population), thus yielding a matrix of 104 Egos X 225 Alters = 23,400 kinship term applications. e) From this large matrix, I selected a smaller square matrix of 104 X 104 = 10,816 data points, containing complete sets of kinship term reciprocals for all pairs. Table 1a (data) and Table 1b (key) together constitute Table 1, the Alyawarra Kinship dataset in use here. Please refer to these files as needed below; search by filenames if links are broken.

Table 1. Alyawarra Kinship dataset (select links to access files).

Table 1a. Alyawarra1971KinData.xls - Alyawarra kinship data in Excel format
https://www.kinsources.net/kidarep/dataset_attachment-/49/184/Alyawarra1971KinData.xls
Table 1b. Alyawarra1971KinshipDataKey.pdf – Brief explanation of dataset structure, content and operation.
https://www.kinsources.net/kidarep/dataset_attachment-/49/180/Alyawarra1971KinDataKey.pdf

2. **Relations between Alyawarra kinship data and machine learning algorithms.** Broadly speaking, machine learning research deals with computer algorithms whose performance improves automatically through experience. Specialists in the field have used the Alyawarra Kinship dataset, in conjunction with other sets having similar or significantly different characteristics, to develop unsupervised pattern detection algorithms, and to test the accuracy and speed of competing pattern detection algorithms that have been developed by many independent research teams. To refresh your knowledge of the theory and practice of supervised and unsupervised pattern detection, see the brief review article by Das, Sumit, Day, Pall and Roy (2015) or similar articles at Wikipedia and elsewhere on the web.

The patterns themselves, and the accuracy and speed with which they can be detected, are best appreciated when the performance of recent machine learning algorithms with 21st century hardware are viewed against the predominantly manual data processing performed by my

[1] Among anthropologists who are interested primarily in the language of kinship (the purity of definitions and logical relations among kinship terms), this step in my work has been highly contentious. Those who focus on kinship linguistics seek error-free data from one or a few key informants that optimally displays the logic of the relations to perfection. Those who, like myself, focus on kinship applications seek data that show the same basic logic of relations as a background for the diversity – modest or almost chaotic - that reflects the flexibility and ambiguity of ways in which living people use kinship terms, genealogies and highly complex life histories of ever-shifting relationships with each other. Specialists in kinship linguistics and kinship applications often talk past each other or do not talk to each other at all. The Alyawarra Kinship dataset described here has developed a strong following in the machine learning community, but no following within mainstream anthropology.

colleagues, John Atkins and Chad McDaniel, and me in 1976-77. The basic procedures - then and now, manual and electronic - are similar whether analytical activities spanned days, weeks or months a half-century ago or nanoseconds now. However, the algorithms were tested only once or a very few times each in the 1970s, whereas they may be tested hundreds or thousands of times each with 21st century technology. The differences are not trivial.

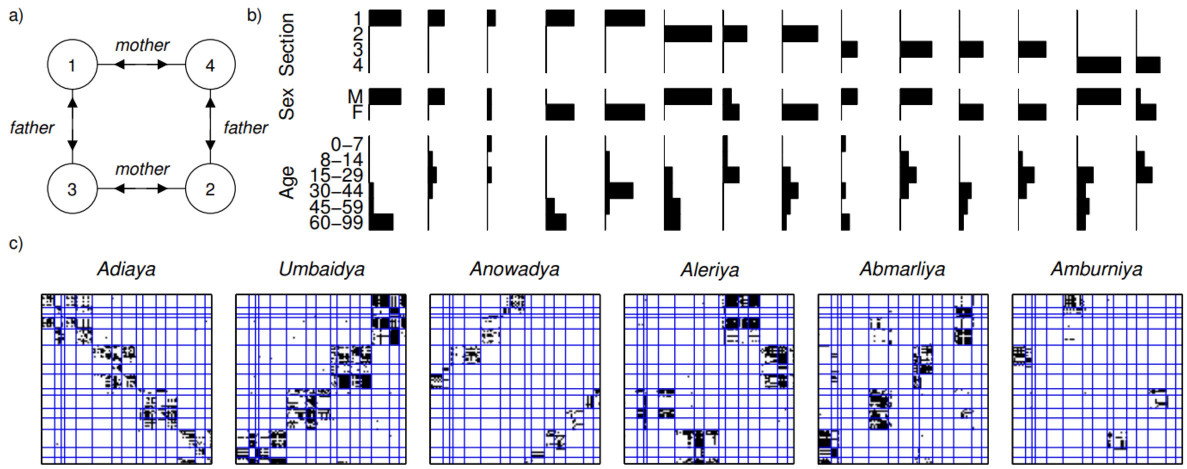
To evaluate the performance of various computational procedures, experiments are conducted using a single algorithm or a set of competing algorithms in conjunction with one or an appropriate assortment of common benchmark datasets such as the Alyawarra Kinship dataset. The publications cited below in Part 3 of this article contain lengthy and detailed mathematical statements of the problems addressed, analytical methods used, and conclusions reached. Since I lack the technical expertise required to succinctly summarize those selected papers, I urge you to explore them for yourself.

To understand similarities and important differences between the Alyawarra Kinship dataset and Geoff Hinton's popular and much older Kinship Data Set, see the following articles: Hinton (1986, 1990), Quinlan (1990), Cunningham (1996). To distinguish succinctly between Hinton's dataset and the Alyawarra dataset, researchers sometimes call Hinton's the "Kinship" dataset and Denham's the "Kinships" dataset.

Here I present four figures and tables included in papers cited below as possibly interesting examples of computerized processing of the data and comment briefly on these graphics. But to understand what they mean, you should read the papers from which I extracted them.

My objectives here are to put the introduced materials in a slightly broader context than that provided by machine learning technicalities. By discussing possible connections between anthropology and machine learning and seeking a two-way bridge that benefits both disciplines, I suggest that the history of the Alyawarra kinship dataset demonstrates the value of interdisciplinary cooperation between unlikely partners.

Figure 1a is a standard representation of an Australian Aboriginal Kariera 4-section kinship system that characterizes all parent-child relationships in the dataset. Figure 1b shows section, sex and age clustering of the 104 people in the sample. Figure 1c shows the distribution of 6 sets of kinship term applications sorted by the clusters in Figure 1b. All these patterns were detected by unsupervised pattern detection using the Infinite Relational Model (Kemp et al. 2006).



(a) The Kariera kinship system. Each person belongs to one of four kinship sections, and the section of any person predicts the sections of his or her parents. (b) Composition of the 15 clusters found by the IRM. The six age categories were chosen by Denham, and are based in part on Alyawarra terms for age groupings (Denham 1973). (c) Data for six Alyawarra kinship terms. The 104 individuals are sorted by the clusters shown in (b).

Figure 1. Learning kinship systems; IRM = Infinite Relational Model (from Kemp et al. 2006:386-87).

Table 2 evaluates the performance of 13 different algorithms (models listed in column 1) against various components of three datasets including the Alyawarra Kinship dataset. Precisely what is being measured is not important for my purposes here, but the fact that two UniKER models achieve the highest scores (**in bold type**) indicates that those models are superior to their competitors.

Model	Kinship			FB15k-237			WN18RR		
	Hit@1	Hit@10	MRR	Hit@1	Hit@10	MRR	Hit@1	Hit@10	MRR
RESCAL (Nickel et al., 2011)	0.489	0.894	0.639	0.108	0.322	0.179	0.123	0.239	0.162
SimplE (Kazemi & Poole, 2018)	0.335	0.888	0.528	0.150	0.443	0.249	0.290	0.351	0.311
KALE (Guo et al., 2016)	0.433	0.869	0.598	0.131	0.424	0.230	0.032	0.353	0.172
RUGE (Guo et al., 2017)	0.495	0.962	0.677	0.098	0.376	0.191	0.251	0.327	0.280
BLP (De Raedt & Kersting, 2008)†	-	-	-	0.062	0.150	0.092	0.187	0.358	0.254
MLN (Richardson & Domingos, 2006)†	0.655	0.732	0.694	0.067	0.160	0.098	0.191	0.361	0.259
ExpressGNN (Zhang et al., 2019)	0.105	0.282	0.164	0.150	0.317	0.207	0.036	0.093	0.054
pLogicNet (Qu & Tang, 2019)†	0.683	0.874	0.768	0.237	0.524	0.332	0.398	0.537	0.441
pGAT (Harsha Vardhan et al., 2020)‡	-	-	-	0.377	0.609	0.457	0.395	**0.578**	0.459
TransE (Bordes et al., 2013)†	0.221	0.874	0.453	0.198	0.441	0.279	0.013	0.531	0.223
UniKER-TransE	**0.866**	**0.968**	**0.910**	0.463	**0.630**	0.522	0.040	0.561	0.307
DistMult (Toutanova et al., 2015)†	0.360	0.885	0.543	0.199	0.446	0.281	0.390	0.490	0.430
UniKER-DistMult	0.770	0.945	0.823	**0.507**	0.587	**0.533**	**0.432**	0.538	**0.485**

† Results on FB15k-237 and WN18RR are taken from (Qu & Tang, 2019).
‡ Results are taken from (Harsha Vardhan et al., 2020).

Table 2. Results of reasoning on Kinship, FB15K-237 and WN18RR datasets (from Kewei et al. 2020:4).

In the following paragraphs, I quote and paraphrase Galbrun and Kimmig (2012:1-16). Their paper, from which I extracted and renumbered Figure 2 on the following page, contains excellent and accessible descriptions of their work.

Their Abstract says: “The paper introduces *relational redescription mining*, that is, the task of finding two structurally different patterns that describe nearly the same set of object tuples in a relational dataset. … it provides a powerful tool to match different relational descriptions of the same concept. … Experiments in the domain of explaining kinship terms [from Denham’s Alyawarra Kinship dataset] show that this approach can produce complex descriptions that match explanations by domain experts, while being much faster than a direct relational query mining approach.”

The Experiments section of their paper begins by saying: “ [The] Alyawarra Ethnographic Database provides genealogical information about individual members of an indigenous community of Australia, the Alyawarra, as well as the kinship terms they use for their relationships to other persons. A glossary of kinship terms is available, to which we can compare our findings.”

After discussing Figure 2(a) and 2(b), they conclude with this description of Figure 2(c). “As a final example, our algorithm found three definitions for the *Umbaidya* term, suggesting that this term is used by mothers to refer to their child (g17.1), and by male and female speakers alike to refer to daughters of their sister (g17.2) or the children of their maternal uncle’s daughter (g17.3). The first clause matches the ethnographic explanation provided for this term. The second clause differs from the second glossary entry, which restricts this structure to male speakers. The third clause has the same level of complexity as the last glossary entry, but a different structure. For most terms, our algorithm returned a pattern containing one or several clauses corresponding to the main definition provided for the term. In some cases, it found matching supplementary usage. In other cases, the additional usage found deviated from the provided explanation. Frequently, the deviation was an intermediate genealogical level or a difference in gender of some individual in the relation, as in the second clause above.”

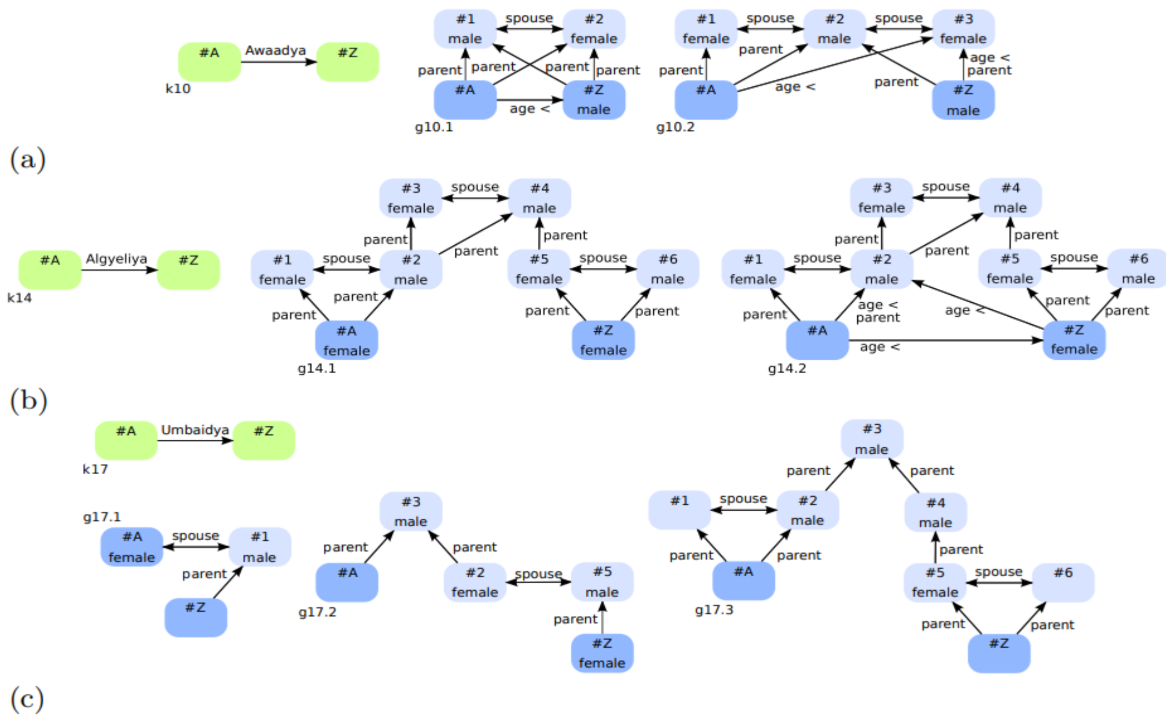


Figure 2. Examples of kinship terms (graphs labeled k10, k14, k17) for pairs (#A, #Z) described in terms of attributes and genealogical relations (remaining graphs). (From Galbrun and Kimmig 2012:14)

Figure 3 is an example of the enormous number and complexity of structural and behavioral relationships that are embedded in the multidimensional Alyawarra dataset. While sitting on top of my Land Rover for 191 hours spanning 51 observation days, I made 41,814 observational behavioral records (BEVRECS). The collection includes 1439 records of 71 different people carrying 24 different infants and children whose ages ranged from birth to 8 years. These records are fine-grained behavioral observations each of which contains 11 kinds of data (File#, ID#, Location1, Actor, Behavior, Orientation, Object, Location2, Continue, Time, Day) that show what specific people did with which other specific people during recording sessions averaging about an hour each. I made these records inside the camp where visibility was excellent, but outside of residences where visibility was limited.

Data used to generate Figure 3 and many other diagrams like it include the following: Alyawarra Kinship dataset, genealogical data, demographic data, camp maps, census data, portraits of all members of the population, and 41,814 behavior records.

Figure 3 is a simple summary of the carrying that one child (♂115) experienced. My data show that this 6-year-old boy was carried 81 times by 38 different people during my observational hours, and this child was not exceptional.

The genealogical diagram in the background of Figure 3 represents the people who lived in Gurlanda camp while I recorded the behavioral observations. The diagram contains four quadrants, each corresponding to one of the four subcommunities in the camp. Squares with arrows and letters such as A are links that connect relationships bridging gaps between subcommunities. People represented by red circles (♀) and red triangles (♂) are the 24 infants and children who were carried during my observation sessions. Blue arrows indicate that the person at the flat end of the arrow was recorded at least once as a carrier of the person at the pointed end of the arrow. (People who carried the child only once are omitted from this diagram.) ♂115 lived in the upper-left quadrant with his siblings, parents, grandparents, and members of his MFFBS's family, all of whom appear on the genealogical background.

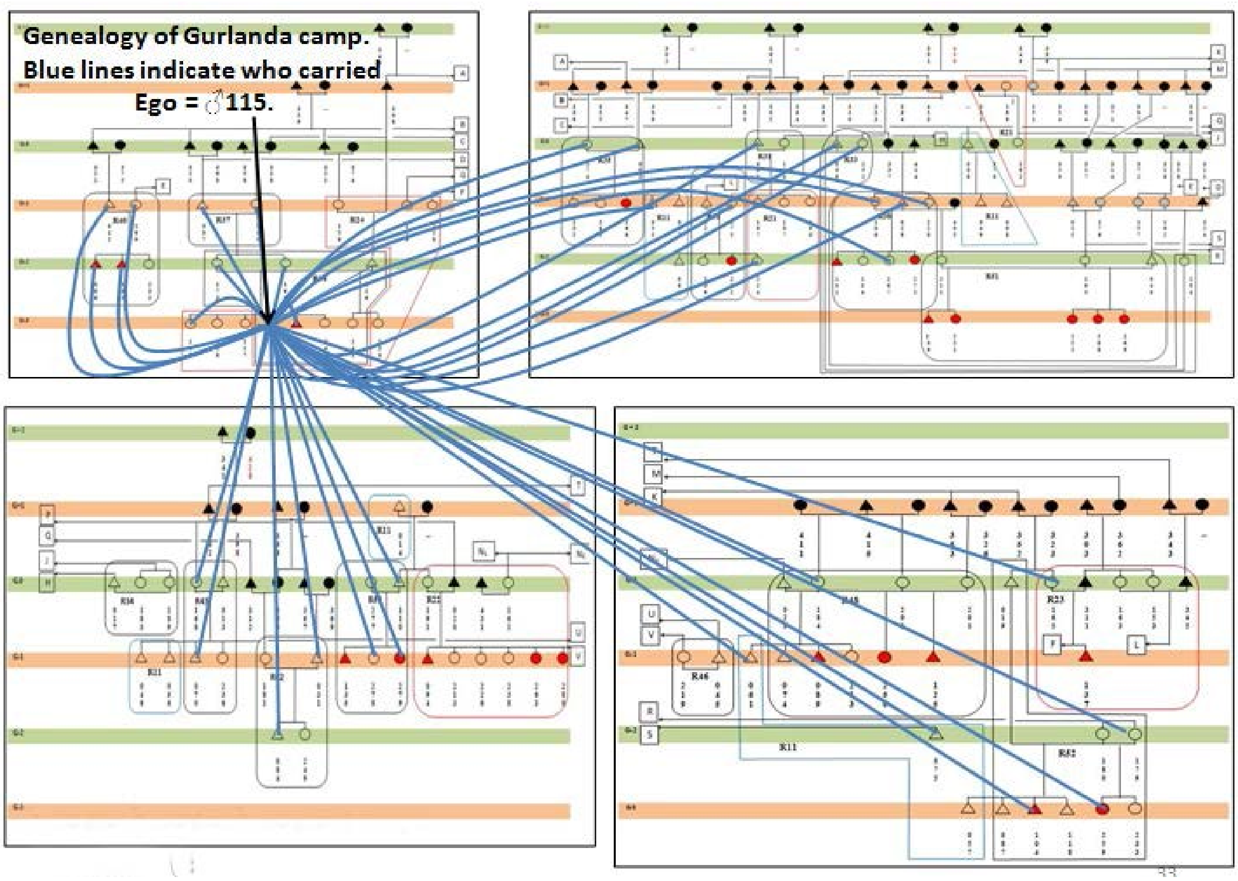


Figure 3. A simple graphical summary of social relations between a 6-year-old boy and 38 people who carried him 81 times within Gurlanda camp during my 191 observation hours (single carries omitted here).

I conclude by citing two additional articles that use machine learning algorithms to seek community structure in social networks (Denham and White 2005:83; Newman 2006), one that detects interesting events in complex streams of behavioral data such as my BEVRECS data (Margineantu, Wong and Dash 2010), and another that deals broadly with the evolution of culture (Kobayashi et al 2019).

To access a huge collection of diverse competing and complementary datasets, you can search the web for "machine learning datasets". "Datasets [such as this one] are an integral part of the field of machine learning [and] high-quality datasets for unsupervised learning can [be] difficult and costly to produce" (Wikipedia 9/13/2020). Because of the enormous amount of time and effort required to collect and edit the Alyawarra Kinship dataset and supporting files, it fits this description perfectly. Building it has entailed a lot of work, but the end-product is a modest contribution to machine learning and anthropological research that has developed a vigorous life of its own.

To the best of my knowledge, anthropologists have made little use of machine learning methods in their kinship research. Certainly I am not conversant with all or most kinship research that has been conducted in recent decades, but I believe that other examples of data such as that stored in the Alyawarra Kinship dataset and supporting files remains rare in the discipline. A notable exception to this generalization is the large collection of genealogical data and PUCK software at the KinSources Genealogical Archive.[2] Just as we can evaluate competing algorithms against standardized datasets, it is equally important to evaluate competing datasets against standardized algorithms.

This article may seem to be off target for MACT's audience, but I suggest that it is not. I offer it as an example of a kind of kinship research that has found favor with a worldwide audience. I believe that both machine learning research and ethnographic research, now and in the future, would benefit from having access to more datasets of this kind. Perhaps this paper will encourage some readers to publish similar field data from other societies or other species.

[2] The KinSources Genealogical Archive at https://www.kinsources.net/ contains 128 genealogical datasets generally lacking kinship applications data and other supporting data. However, the collection includes two linked sets of genealogical data, similar to the Alyawarra Kinship dataset, but with a different kind of kinship terminology. Both sets are from the Wanindiljaugwa people of Groote Eylandt, Arnhem Land, Northern Territory, Australia. The first was collected by Frederick Rose in 1941 and published in Rose (1960); the second was collected by Peter Worsley in 1954 and remains unpublished (Worsley 1954). Rose's fieldwork in 1941 was the precedent for my fieldwork in 1971.

3. **Artificial Intelligence / Machine Learning papers using the Alyawarra Kinship Dataset: 2004 – 2020.** Part 3 provides links to a sample of papers by Kemp et al. (2004, 2006, 2010) and his AI colleagues around the world, all of whom utilize the Alyawarra kinship data, citing Denham 1973 and/or 1979 as their source. The sample demonstrates diversity of styles while minimizing needless redundancy. The entries are listed in approximately chronological order. Most of the earlier entries have a "cited by # papers" tally that suggests the level of activity in this subfield of AI research during these years. When you find broken links here and elsewhere in the paper, use your browser to search for the items by title and author.

Kemp, Charles, Thomas Griffiths and Joshua Tenenbaum.
2004 Discovering latent classes in relational data. *MIT Computer Science and Artificial Intelligence Laboratory Technical Reports* MIT-CSAIL-TR-2004-050. https://cocosci.princeton.edu/tom/papers/blockTR.pdf
Cited by 86.

Xuerui Wang, Natasha Mohanty and Andrew Kachites McCallum.
2005 Group and topic discovery from relations and text. *LinkKDD '05: Proceedings of the 3rd international workshop on Link discovery.* August 2005:28-35; https://doi.org/10.1145/1134271.1134276 . Cited by 161.

McCallum, Andrew, Xuerui Wang, and Natasha Mohanty.
2006 Joint group and topic discovery from relations and text. In: Airoldi E., Blei D.M., Fienberg S.E., Goldenberg A., Xing E.P., Zheng A.X. (eds) Statistical Network Analysis: Models, Issues, and New Directions. ICML 2006. *Lecture Notes in Computer Science,* vol 4503. Springer, Berlin, Heidelberg. DOI https://doi.org/10.1007/978-3-540-73133-7_3

Kemp, Charles, Joshua Tenenbaum, Thomas Griffiths, Takeshi Yamada and Naonori Ueda.
2006 Learning systems of concepts with an infinite relational model. AAAI *Proceedings: 21st National Conference on Artificial Intelligence*. 2006: 381-388 https://www.aaai.org/Papers/AAAI/2006/AAAI06-061.pdf . Cited by 845.

Kok, Stanley and Pedro Domingos.
2007 Statistical predicate invention. *Proceedings of the 24th International Conference on Machine Learning: ICML* 2007:433-440 https://doi.org/10.1145/1273496.1273551
Cited by 173.

Roy, Daniel M., Charles Kemp, Vikash K. Mansinghka, and Joshua Tenenbaum.
2007 Learning annotated hierarchies from relational data. *Advances in Neural Information Processing Systems 19*: NIPS 2007:1-8 http://danroy.org/papers/RoyKemManTen-NIPS-2007.pdf Cited by 66.

Kemp, Charles, Joshua B. Tenenbaum, Sourabh Niyogi, Thomas L. Griffiths.
2009 A probabilistic model of theory formation. *Cognition* 114 (2010) 165–196. https://cocosci.princeton.edu/tom/papers/LabPublications/ProbModelTheoryForm.pdf Cited by 89.

Miller, Kurt T., Thomas Griffiths and Michael Jordan.
2009 Nonparametric latent feature models for link prediction. *Advances in Neural Information Processing Systems 22* (NIPS 2009):1-9. https://cocosci.princeton.edu/tom/papers/linkpred.pdf Cited by 411.

Sutskever, I., Salakhutdinov, R., Tenenbaum, J.B.
2009 Modelling relational data using Bayesian clustered tensor factorization. *Advances in Neural Information Processing Systems 22* (NIPS 2009):1821-1828. http://papers.nips.cc/paper/3863-modelling-relational-data-using-bayesian-clustered-tensor-factorization.pdf. Cited by 242.

Kemp, Charles, Joshua B. Tenenbaum, Sourabh Niyogi, Thomas L. Griffiths.
2010 A probabilistic model of theory formation. *Cognition* 114(2):165–196. http://cocosci.princeton.edu/tom/papers/LabPublications/ProbModelTheoryForm.pdf; PubMed https://pubmed.ncbi.nlm.nih.gov/19892328/; DOI: 10.1016/j.cognition.2009.09.003. Cited by 90.

Menon, A.K and C. Elkan.
2010 Dyadic prediction using a latent feature log-linear model. *arXiv*:1006.2156v1 [cs.LG] 10 Jun 2010. http://arxiv.org/pdf/1006.2156.pdf

Jenatton, Rodolphe; Le Roux, Nicolas; Bordes, Antoine; Obozinski, Guillaume.
2012 A latent factor model for highly multi-relational data. *Advances in Neural Information Processing Systems 25* (NIPS 2012): http://papers.nips.cc/paper/4744-a-latent-factor-model-for-highly-multi-relational-data
Cited by 329.

Kemp, Charles and Terry Regier.
2012 Kinship categories across languages reflect general communicative principles. *Science* 336, 1049 (2012): 1049-1054. https://science.sciencemag.org/content/336/6084/1049.abstract. Cited by 180.

Galbrun, Esther and Kimmig, Angelika.
2012 Towards Finding Relational Redescriptions. *Proceedings of the 15th International Conference on Discovery Science,* DS'12, Oct 2012, Lyon, France; pp.1-16. https://hal.archives-ouvertes.fr/hal-01399271/document
2014 Finding relational redescriptions. *Machine Learning*, September 2014, Volume 96, Issue 3, pp. 225-248. https://link.springer.com/article/10.1007/s10994-013-5402-3

Bordes, Antoine; Glorot, Xavier; Weston, Jason; Bengio, Yoshua.
2014 A semantic matching energy function for learning with multi-relational data. *Machine Learning,* February 2014, 94(2):233-259. https://arxiv.org/abs/1301.3485. https://link.springer.com/article/10.1007%2Fs10994-013-5363-6

Wang, William Yang, Kathryn Mazaitis, William W. Cohen.
2014 Structure learning via parameter learning. Proceedings of the 23rd ACM International Conference on Information and Knowledge Management, November 2014, pp. 1199-1208. https://doi.org/10.1145/2661829.2662022
2014 ProPPR: efficient first-order probabilistic logic programming for structure discovery, parameter learning, and scalable inference. *Statistical Relational AI: Papers from the AAAI-14 Workshop*. https://sites.cs.ucsb.edu/~william/papers/starAI.pdf
2017 Differentiable learning of logical rules for knowledge base reasoning. *31st Conference on Neural Information Processing Systems* (NIPS 2017), Long Beach, CA, USA. https://papers.nips.cc/paper/2017/file/0e55666a4ad822e0e34299df3591d979-Paper.pdf

Garcia Duran, Alberto.
2016 Learning representations in multi-relational graphs: algorithms and applications. Université de Technologie de Compiègne, France; thesis. https://tel.archives-ouvertes.fr/tel-01513058 .

Fozi, Amin.
2017 Algorithmic reconstruction of inverted relationships using the Alyawarra kinship glossary. Undergraduate Honors Thesis, Mathematics Department, University of California San Diego. http://www.math.ucsd.edu/_files/undergraduate/honors-program/honors-program-presentations/2017-2018/Amin_Fozi_Honors_Thesis.pdf.

Yuyu Zhang, Xinshi Chen, Yuan Yang, Arun Ramamurthy, Bo Li, Yuan Qi, Le Song.
2019 Can graph neural networks help logic reasoning? https://arxiv.org/pdf/1906.02111.pdf .

Ye Liu.
2020 Computational methods for complex models with latent structure. North Carolina State University: doctoral dissertation, May 2020. https://repository.lib.ncsu.edu/handle/1840.20/37507

Kewei Cheng, Ziqing Yang, Ming Zhang, Yizhou Sun.
2020 UniKER: A unified framework for combining embedding and horn rules for knowledge graph inference. Proceedings of the 37th International Conference on Machine Learning, Vienna, Austria: *Proceedings of Machine Learning Research,* Vol 119, 2020, pp1-7. https://grlplus.github.io/papers/84.pdf

4. **Alyawarra kinship term applications data and documentation.** Part 4 contains links to some of my Alyawarra kinship data files that are available online.

1973 Denham, Woodrow W. The detection of patterns in Alyawarra nonverbal behavior. Seattle: University of Washington, doctoral dissertation. https://pdfs.semanticscholar.org/87a1/3d8a8d0b59c29fcde32aee7d887f6d13beb7.pdf or https://www.researchgate.net/publication/247830886_The_Detection_of_Patterns_in_Alyawara_Nonverbal_Behaviour

1979 Denham, Woodrow W., Chad McDaniel and John R. Atkins. Aranda and Alyawara kinship: a quantitative argument for a double helix model. *American Ethnologist* 6(1):1-24. https://anthrosource.onlinelibrary.wiley.com/doi/abs/10.1525/ae.1979.6.1.02a00010.

2001 Denham, Woodrow W. Alyawarra ethnographic database: Numerical data documentation file, version 7 https://www.researchgate.net/publication/340256315_AU01NumF00_Alyawarra1971-72_Data_Documentation_File (revised 1977, 2019).

2005 Denham, Woodrow W. and Douglas R. White. Multiple measures of Alyawarra kinship. *Field Methods* 17(1):70-101. https://escholarship.org/uc/item/9xs4j0kg

2006 Denham, Woodrow W. Alyawarra1971KinData.xls. Alyawarra kinship data in Excel format https://www.kinsources.net/kidarep/dataset_attachment-/49/184/Alyawarra1971KinData.xls

2006 Denham, Woodrow W. Alyawarra kinship data in Matlab format (maintained by Charles Kemp) http://charleskemp.com/code/irm.html

2010 Denham, Woodrow W. Alyawarra1971KinshipDataKey.pdf. https://www.kinsources.net/kidarep/dataset_attachment-/49/180/Alyawarra1971KinDataKey.pdf

2010 Denham, Woodrow W. Alyawarra1971GenDiag.pdf. Alyawarra1971 genealogical diagrams https://www.kinsources.net/kidarep/dataset_attachment-/49/182/Alyawarra1971GenDiag.pdf

2010 Denham, Woodrow W. Alyawarra1971ManualREADFIRST.pdf. Alyawarra 1971 User Guide and Reference Manual https://www.kinsources.net/kidarep/dataset_attachment-/49/181/Alyawarra1971ManualREADFIRST.pdf

2011 to 2018 Denham, Woodrow W. This link provides access to eight of my recent articles on Alyawarra kinship in *Mathematical Anthropology and Cultural Theory* at http://www.mathematicalanthropology.org/ and/or https://escholarship.org/uc/hcs_MACT/search?

5. **Acknowledgements**. I designed the Alyawarra project in 1969 as my dissertation research and have paid attention to it ever since. I thank the following for various kinds of support at various times over the past half-century.

In the United States, John Atkins at the University of Washington, Seattle, was my dissertation supervisor. Gene Hammel at the University of California, Berkeley, supervised my NIMH Postdoctoral Traineeship. Doug White at the University of California, Irvine, and David Damas at McMaster University, Canada, generously shared their knowledge and expertise with me.

Paul Ballonoff at *Mathematical Anthropology and Cultural Theory* has been my editor for the last ten years. Charles Kemp, then at MIT, discovered my kinship data, realized its potential value for machine learning research, made it available for the development of unsupervised pattern detection algorithms, established it as a benchmark dataset for product testing, and gave me valuable advice in writing this article.

In Australia, Les Hiatt at Sydney University was my mentor during my year as a visiting postgraduate student. In the Northern Territory, Mac and Rose Chalmers were my generous and supportive sponsors at MacDonald Downs Station where I conducted my field research. Most importantly, for a year the Alyawarra speaking Aboriginal people who belonged to that beautiful but difficult land accepted me in their community and supported me in every conceivable way.

Financial support came from various sources including the US National Science Foundation, US National Institutes of Mental Health, Australian Institute of Aboriginal Studies (AIATSIS) and Canada Council.

Finally, I acknowledge those who developed the Internet, the World Wide Web and ResearchGate, plus the multi-ethnic teams of machine learning specialists cited above. Together they have made the Alyawarra Kinship dataset, supporting files and documentation available worldwide.

My thanks to all.

6. **References** (to items cited in the body of the text)

Cunningham, Sally Jo 1996. Machine learning applications in anthropology: automated discovery over kinship structures. *Computers and the Humanities* 30(6):401-406. https://www.researchgate.net/publication/225263285_Machine_Learning_Applications_in_Anthropology_Automated_Discovery_over_Kinship_Structures

Das, Sumit, Aritra Dey, Akash Pal and Nabamita Roy 2015 Applications of artificial intelligence in machine learning: review and prospect. *International Journal of Computer Applications,* V.115(9):31-41. https://www.researchgate.net/publication/276178017_Applications_of_Artificial_Intelligence_in_Machine_Learning_Review_and_Prospect

Denham, Woodrow W. 1973. The detection of patterns in Alyawarra nonverbal behavior. Seattle: University of Washington, doctoral dissertation. https://pdfs.semanticscholar.org/87a1/3d8a8d0b59c29fcde32aee7d887f6d13beb7.pdf or https://www.researchgate.net/publication/247830886_The_Detection_of_Patterns_in_Alyawara_Nonverbal_Behaviour

Denham, Woodrow W., Chad McDaniel and John R. Atkins 1979. Aranda and Alyawara kinship: a quantitative argument for a double helix model. *American Ethnologist* 6(1):1-24. https://anthrosource.onlinelibrary.wiley.com/doi/abs/10.1525/ae.1979.6.1.02a00010.

Denham, Woodrow W. and Douglas R. White 2005. Multiple measures of Alyawarra kinship. *Field Methods* 17(1):70-101. https://escholarship.org/uc/item/9xs4j0kg

Denham, Woodrow W. 2015. Alyawarra kinship, infant carrying, and alloparenting. *Mathematical Anthropology and Cultural* Theory 8(1):1-101. http://mathematicalanthropology.org/Pdf/MACTDenham1015.pdf

Hinton, Geoff 1986. Learning distributed representations of concepts. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society,*1986:1-12. Amherst, MA. https://www.researchgate.net/publication/3296950_Learning_distributed_representations_of_concepts_using_Linear_Relational_Embedding

- 1990 Kinship Data Set. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/Kinship

Kemp, Charles, Thomas Griffiths and Joshua Tenenbaum 2004. Discovering latent classes in relational data. *MIT Computer Science and Artificial Intelligence Laboratory Technical Reports* MIT-CSAIL-TR-2004-050. https://cocosci.princeton.edu/tom/papers/blockTR.pdf

Kemp, Charles, Joshua Tenenbaum, Thomas Griffiths, Takeshi Yamada and Naonori Ueda 2006. Learning systems of concepts with an infinite relational model. AAAI *Proceedings: 21st National Conference on Artificial Intelligence*. 2006: 381-388 https://www.aaai.org/Papers/AAAI/2006/AAAI06-061.pdf

Kemp, Charles, Joshua B. Tenenbaum, Sourabh Niyogi, Thomas L. Griffiths 2010. A probabilistic model of theory formation. *Cognition* 114(2):165–196. DOI: 10.1016/j.cognition.2009.09.003.

Kewei Cheng, Ziqing Yang, Ming Zhang, Yizhou Sun 2020. UniKER: A unified framework for combining embedding and horn rules for knowledge graph inference. Proceedings of the 37th International Conference on Machine Learning, Vienna, Austria: *Proceedings of Machine Learning Research,* Vol 119, 2020, pp1-7. https://grlplus.github.io/papers/84.pdf

KinSources Genealogical Archive 2020. https://www.kinsources.net/.

Kobayashi, Y., J.Y. Wakano and H. Ohtsuki 2019. Evolution of cumulative culture for niche construction. *Journal of Theoretical Biology* 472 (2019) 67–76 https://doi.org/10.1016/j.jtbi.2019.04.013.

Margineantu, D., Wong, W. K., & Dash, D. (2010). Machine learning algorithms for event detection. *Machine Learning,* 79(3), 257. https://link.springer.com/article/10.1007/s10994-010-5184-9

Newman, M. E. 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23), 8577-8582. https://www.pnas.org/content/103/23/8577

Quinlan, J.R. 1990. Learning logical definitions from relations. *Machine Learning* 5:239-266. https://link.springer.com/content/pdf/10.1007/BF00117105.pdf

Rose, F.G.G. 1960. *Classification of Kin, Age Structure and Marriage amongst the Groote Eylandt Aborigines: A Study in Method and a Theory of Australian Kinship.* Berlin: Akademie-Verlag.

Wikipedia 9/13/2020. List of datasets for machine learning research.

Worsley, Peter 1954. *The changing social structure of the Wanindiljaugwa.* Unpublished doctoral dissertation. Australian National University, June 1954.