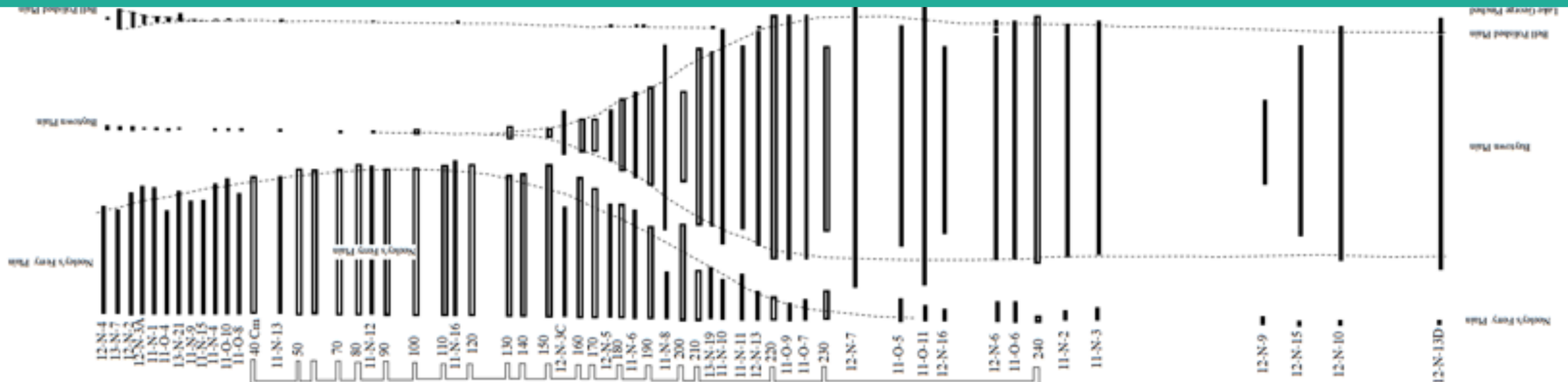
Measuring Cultural Relatedness Using Multiple Seriation Ordering Algorithms

Mark E. Madsen and Carl P. Lipo

SAA 2016 Electronic Symposium:
"Evolutionary Archaeologies: New Approaches, Methods, and Empirical Sufficiency"

UNIVERSITY of WASHINGTON

BINGHAMTON UNIVERSITY
The State University of New York



Frequency Seriation and Unimodality

30s - 40s

Ford and Krieger

Historical types are constructed and tested for chronological significance

Unimodality served as a “signature” for chronologically useful types

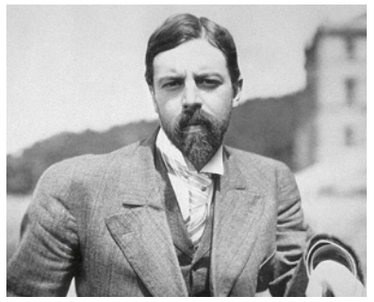
1916

TABLE 5.

PERCENTAGES.

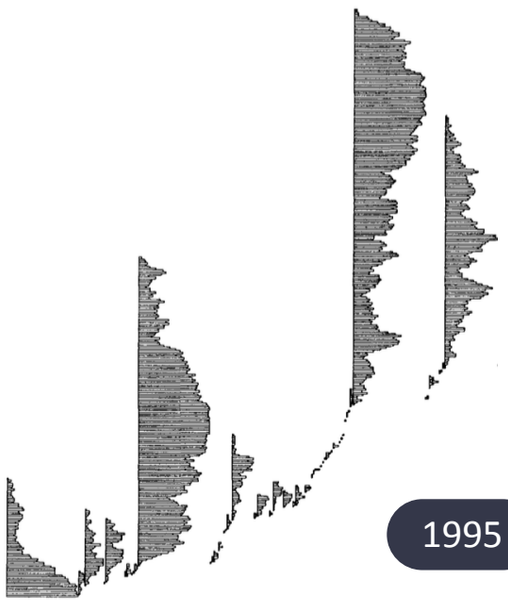
Period	Site	Corrugated	Three Colors	Black on Red	Any Red	"Black" ware
PRESENT	Zuñi	0[a]	0	1		
LATE A	Towwayallanna	1	8	3	22[b]	53[b]
	Kolliwa	—	7	2		
	Shunntekkya	2	7	2		
	Wimmayawa	2	4	1		
	Mattsakya	3	4	3		
	Kyakkima	4	3	2		
EARLY A	Pinnawa	10	1	8		
	Site W	24	—	1		
LATE B	Hattsinawa	27	—	5	10	19
	Kyakkima West	12[c]	—	4	8	—[c]
MIDDLE B	Shoptlawwayala	40	—	2	3	7
	"Hawwikku B"	49	—	6	12	9
EARLY B	Te'allatashshhanna	66	—	—	—	5
	Site X	71	—	—	3	1
	Tetlnatluwwayala	72	—	—	2	—
? — B	He'i'tli'annanna	—	—	—	—	3
	Site Y	—	—	—	—	—

[a] Present, but less than half of one percent.
[b] The variation between sites here lumped seems due more to accident or selection in collecting than to differences typical of period.
[c] Unfortunately only 25 pieces are available from this site.



Kroeber

Stylistic types change according to “normal curves” and the popularity principle



1995

Neiman and others

Unimodality is an outcome of cultural transmission, but one kind of pattern — albeit a useful one!

Lipo, Madsen, and Dunnell

2015

PLOS ONE

OPEN ACCESS

Citation: Lipo CP, Madsen ME, Dunnell RC (2015) A Theoretically-Sufficient and Computationally-Practical Technique for Deterministic Frequency Seriation. PLoS ONE 10(4): e0124942. doi:10.1371/journal.pone.0124942

Academic Editor: Nuno Bicho, Universidade do Algarve, PORTUGAL

Received: December 31, 2014

Accepted: March 19, 2015

Published: April 29, 2015

RESEARCH ARTICLE

A Theoretically-Sufficient and Computationally-Practical Technique for Deterministic Frequency Seriation

Carl P. Lipo[1☯]*, Mark E. Madsen[2☯], Robert C. Dunnell[3†]

1 Department of Anthropology and IIRMES, California State University Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840, USA, **2** Department of Anthropology, Box 353100, University of Washington, Seattle, WA 98195-3100, USA, **3** Department of Anthropology and Middle Eastern Cultures, P.O. Box AR, Mississippi State University, MS 39762, USA

Distance Minimization in Seriation

Method: order assemblages by minimizing pairwise inter-assemblage distance according to some distance metric, forming a graph by agglomeration of assemblages by edges, without enforcing linearity

1 Global criterion

Unimodality was sufficient and convenient, but not necessary, and restricted seriations to a subset of the total variation. It was essential prior to good computing support, but is **optional** today.

2 Theoretically sound

CT models are generally AR(1) processes and Markov chains. Large jumps are possible but rare, spatiotemporal smoothness in frequencies is the rule, not the exception. Assemblages closer together in space and and time will tend to have **smaller** pairwise distance metrics

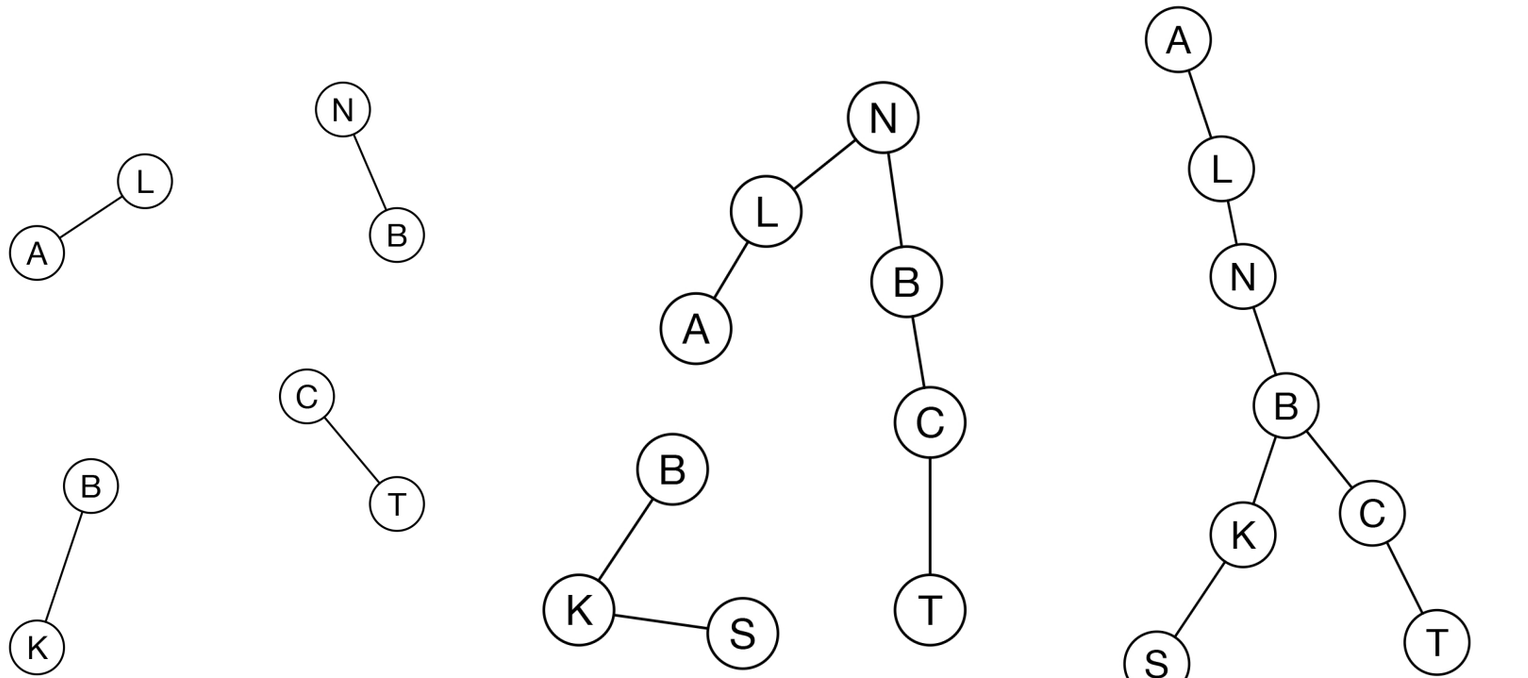
3 Builds on existing work

Kadane (1971) and Shepherdson (2006) examined total path length approaches, but with the requirement of a single linear solution, treating departures as error rather than **information**

4 Full spatiotemporal order

By finding the network with minimum total path distance, whether linear or not, we allow **spatial variation** in trait evolution into the solution along with **temporal ordering**. (we do this in IDSS frequency seriation too, but see #1)

“Continuity” seriation: exact distance minimization

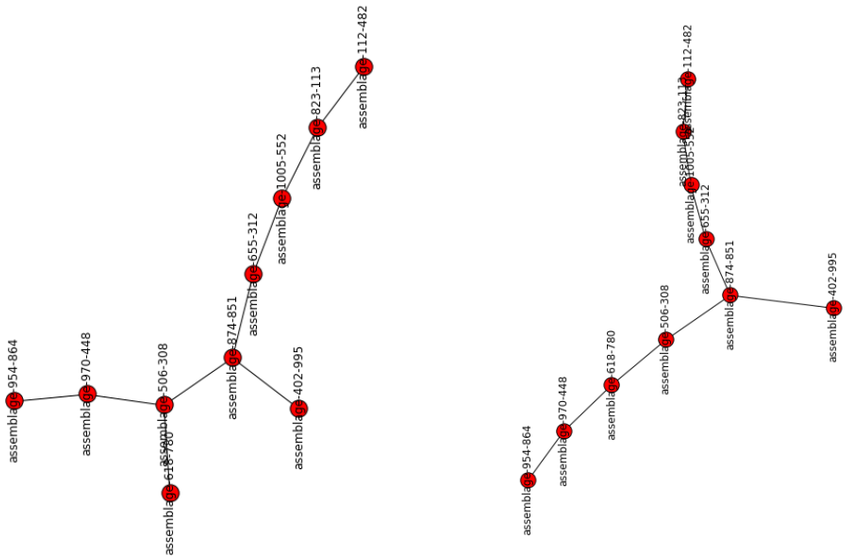


Find shortest pairwise distances	Iteratively add vertices by distance	Coalesce partial solutions
Distance can be any metric, but we began with Euclidean distance (L2 norm). The optimal distance metric may vary depending upon the number of types (Manhattan distance with high dimensional data)	Keep all partial solutions because each may measure change in a different region or direction over time	Overlay partial solutions using overlapping vertices to form seriation networks. Linear orders still occur, but spatial variation or other mesoscale structure induces tree structure

Comparing Frequency and Continuity Seriation

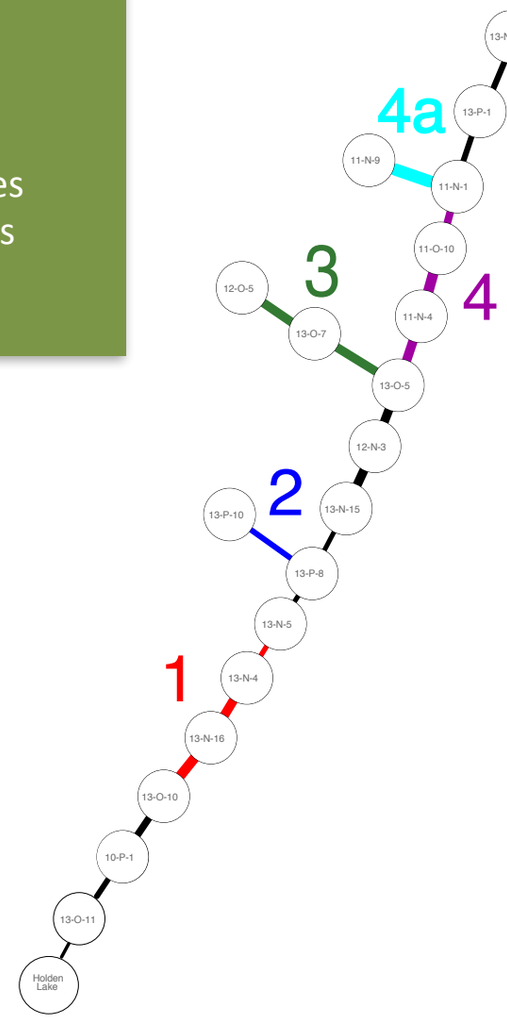
Simulated cultural transmission in a fully spatiotemporal network of communities, with time averaging and typical sampling and filtering techniques.

Frequency (unimodal) and continuity (exact distance minimization) give identical results in most cases, and differ in specific assignments when there are disagreements, given that small differences in directionality vs. distance determine the order.



Empirical

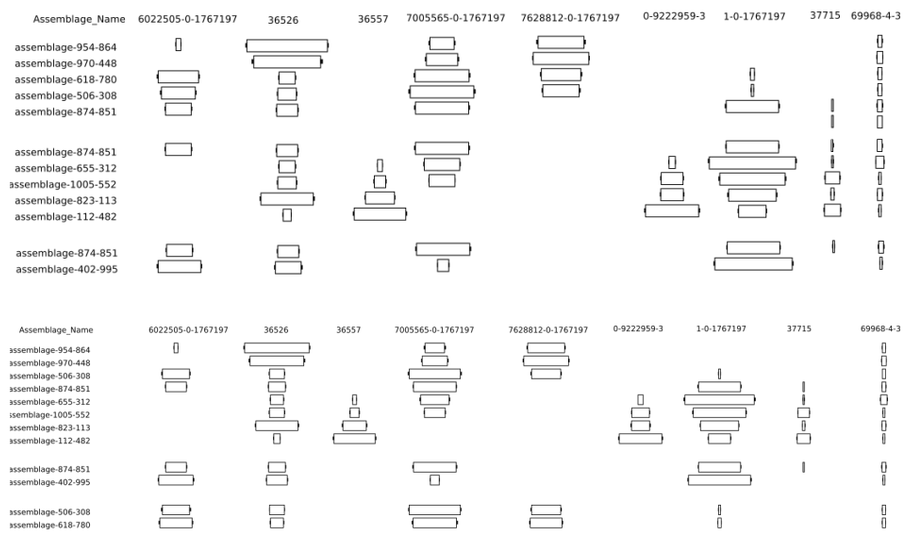
Phillips et al. 1951 pottery assemblages show no differences between methods



Simulation

Continuity (left) and Frequency (right) are identical other than the placement of one community, which forms a separate solution in continuity.

Bar charts: freq (top) & continuity (bottom)



Why do we need better seriation methods?

1 Global criterion

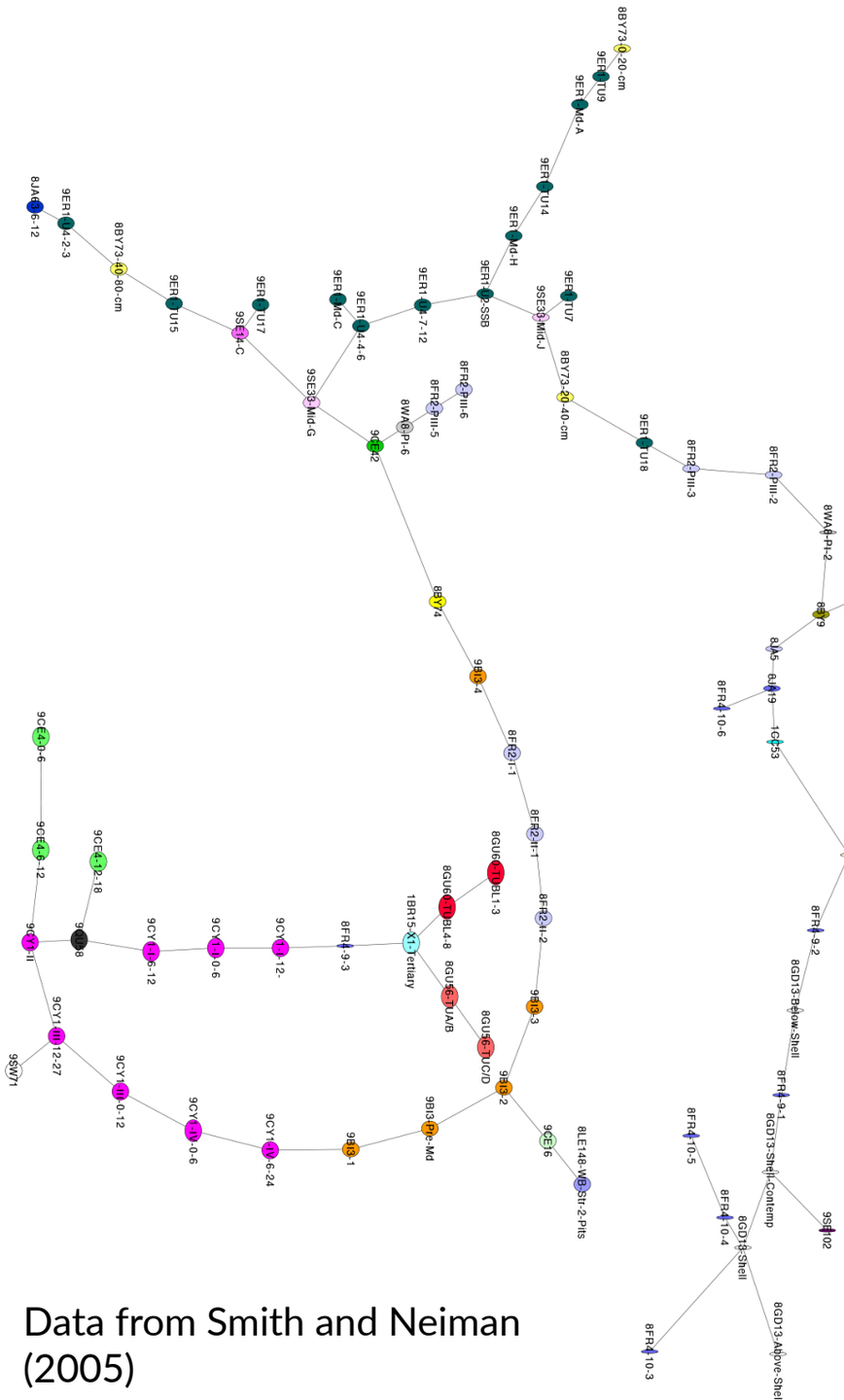
Unimodality was sufficient and convenient, but not necessary, and restricted seriations to a subset of the total variation. We should be able to apply seriation to any cultural transmission problem, whether unimodal types can be found or not.

2 Targets mesoscale problems

Cladistics tends to operate at particular scales given presence/absence of synapomorphic traits. Little work on polymorphic populations has been done, while polymorphism is key to understanding mesoscale, rather than macroscale, evolution given an ability to sort finer-scale events

3 Performance and scalability

Even with advances in technique in our 2015 paper, frequency seriation of large datasets is hard. Exact distance minimization as shown here performed 26x faster than unimodality analysis, while using 1/4th of the CPU count, for an effective speedup of 104x. This allows us to seriate large datasets and attack contemporary, not just archaeological, data.



Data from Smith and Neiman (2005)

Thank You!

For more information, contact:

Mark E. Madsen
Dept. of Anthropology
University of Washington
Seattle, WA 98195

http://notebook.madsenlab.org
ResearchGate/MarkMadsen3
Academia.edu/MarkMadsen
mark@madsenlab.org