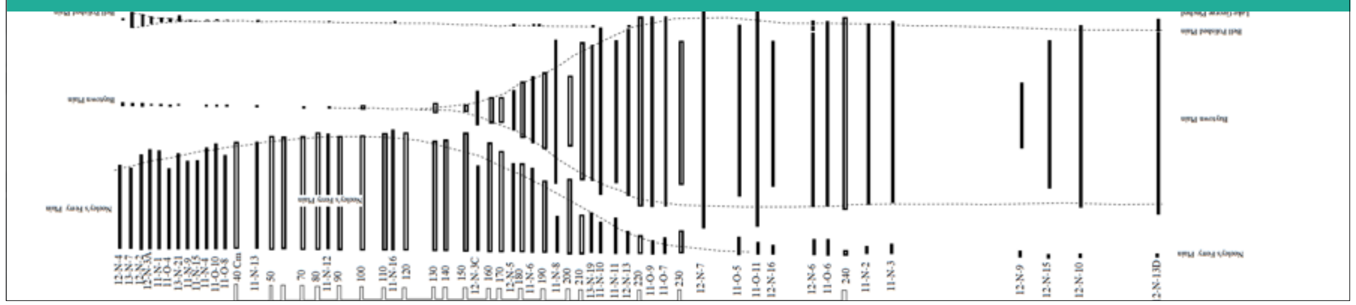
Measuring Cultural Transmission At Archaeological Scales: How Can We Improve Empirical Sufficiency?

Mark E. Madsen

Dissertation Defense
June 11, 2020

UNIVERSITY *of* WASHINGTON



A bit of context...and plan

Context

Syst. Zool., 37(2):142-155, 1988

HOMAGE TO CLIO, OR, TOWARD AN HISTORICAL PHILOSOPHY FOR EVOLUTIONARY BIOLOGY

ROBERT J. O'HARA

Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138

Evolutionary archaeology addresses many questions

But it has a common core task:
documenting evolutionary history

This means **tracing cultural transmission**

O'Hara (1988) -
Chronicle: Evidence for what/when
History: Explanation for chronicle

Model: A proposed **history**

Plan

Dissertation comprised of 4 independent studies

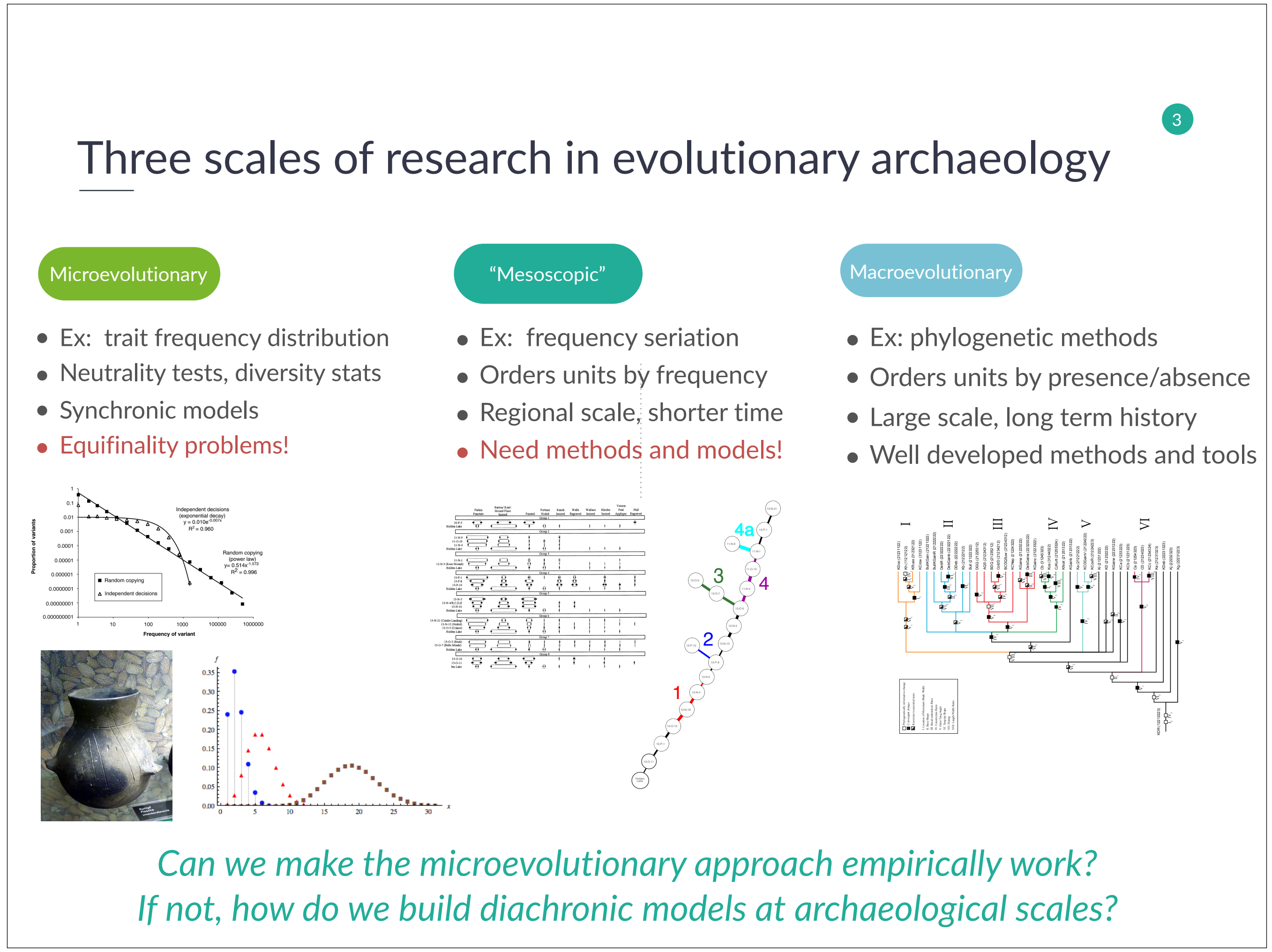
Each is methodological

Goal is to improve empirical sufficiency for documenting CT

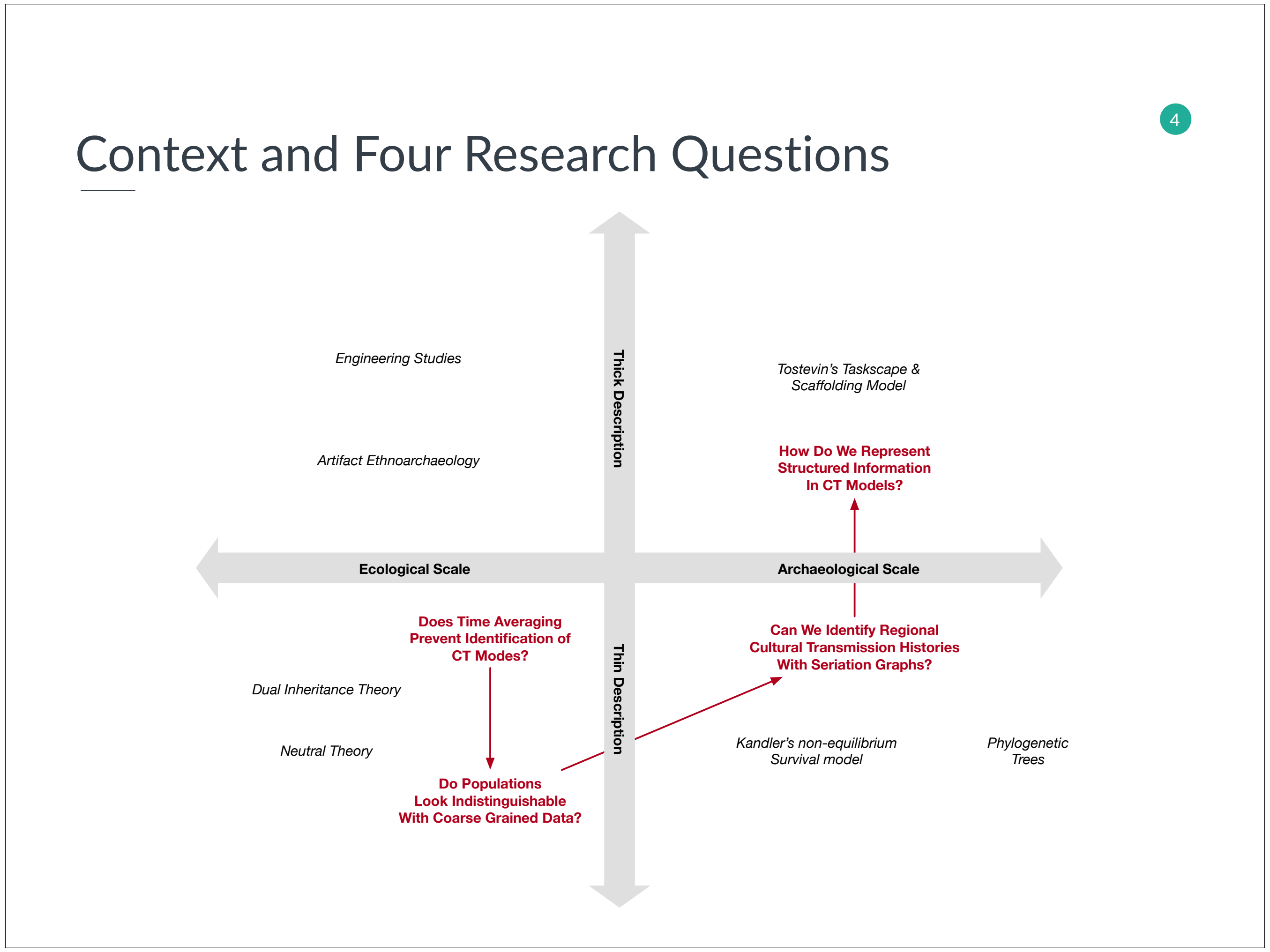
Tasks:

- Develop ways to study **equifinality**
- Develop **better chronicles** and better ways of fitting **models** against **chronicles**

Tools: simulation, machine learning

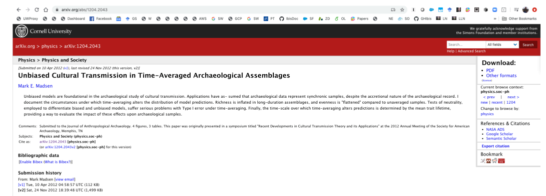


- Reanalyzing LBK, clear that we get different answers - equifinality
- Macro well established but the nature of the method leads to large scales in space and time
- Can we make microevolutionary approaches work by "correcting" equifinalities with analytic methods? I doubt it but it's important to establish that
- Can we build something between the micro and the macro scales? Much of what we study in Holocene arch, arch of complex societies demands a "mesoscopic" approach but we need methods and tools and we need to know what a mesoscopic CT hypothesis "looks like"



- Horizontal axis is time scale over which we study
- Vertical axis is the degree to which models are highly formal (purely about process), versus modeling content
- Thick/thin is not "good" vs "bad" — but we have been operating very thinly, but many questions about the evolution of technologies and even social learning capabilities themselves require thicker descriptions
- Two studies on the left address whether microevolutionary approaches can be made to work
- Many of us involved in early microevolutionary modeling (Bentley, myself, etc) were moving back and forth between ecological and archaeological times, with the same models, same methods; this should have been a red flag because we weren't adapting those methods to the needs of the time scale.
- Two studies on the right address mesoscopic and diachronic models, one thin (spatiotemporal history), one thick (modeling dependencies between cultural traits)

Does Time Averaging Prevent Identification of Cultural Transmission Modes? (Ch. 2)


Physics > Physics and Society
Unbiased Cultural Transmission in Time-Averaged Archaeological Assemblages

Summary

Literature on identifying "biased" CT relies on departure from neutral expectations.

How does **temporal aggregation** affect statistical tests and diversity statistics used to test for departure from neutrality?

Methods

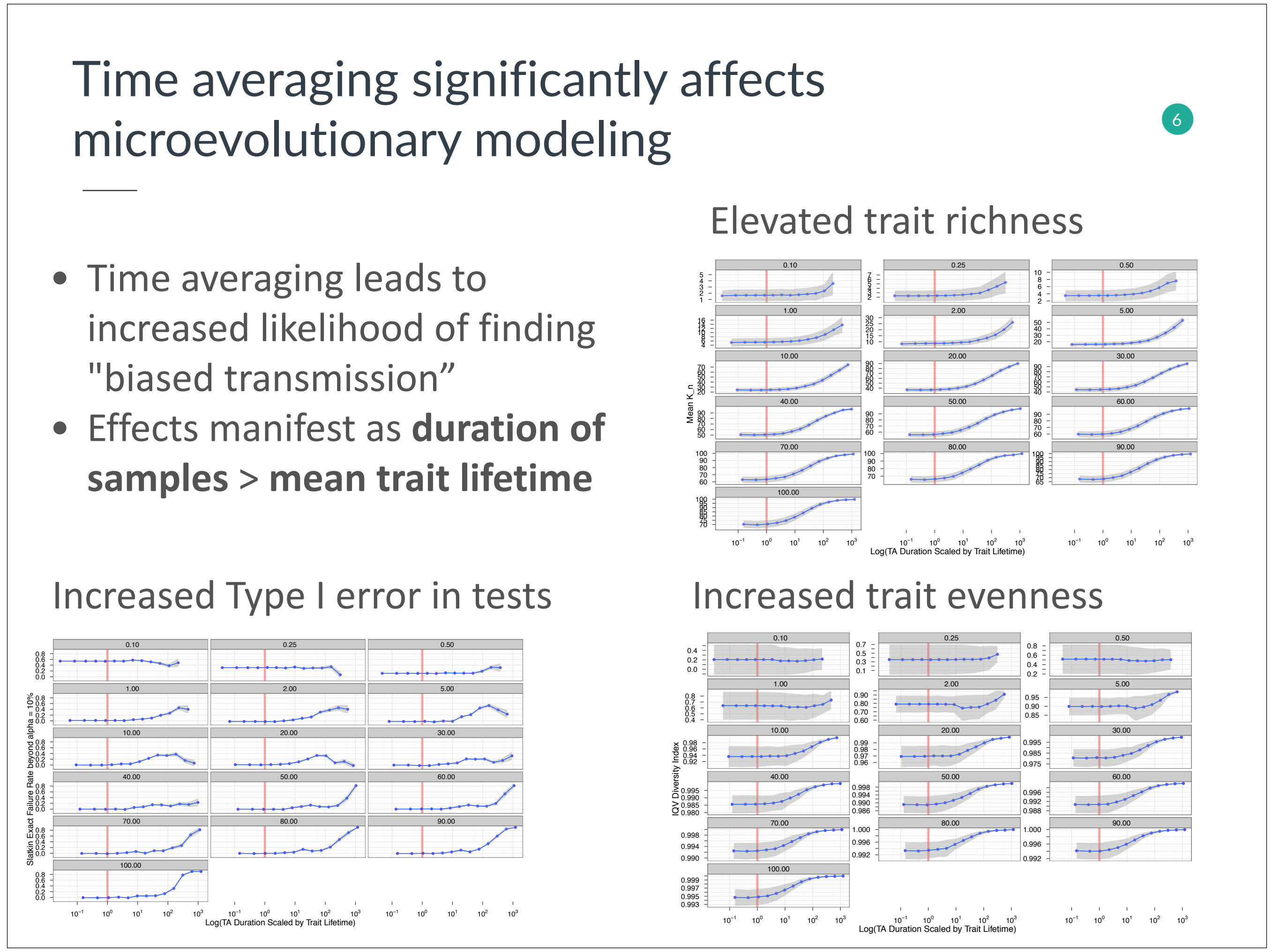
Simulation of unbiased CT (Wright Fisher infinite alleles) with 16 different innovation rates.

Simulator tracked unaggregated and aggregated trait counts (12 window sizes), trait lifetime.

Analysis looked at:

- Excess Type I error in Slatkin neutrality test
- Elevation in trait richness
- Ability to estimate innovation rate (theta)
- Evenness among trait frequencies

arxiv.org, 2012 and SAA conf



- One note about the relationship between assemblage duration and mean trait lifetime: the latter is partially under our control because we build the classifications being tallied. So we can partially adjust the way temporal aggregation affects our quantification, but within limits.
- We can’t fully “correct” for time averaging, which says that fitting synchronic models to diachronic data is something we should stop doing.

Do Populations Look Indistinguishable With Coarse Grained Data ? (Ch. 3)

Summary

Real populations are heterogeneous in their social learning modes and social vs. individual learning.

Do the effects of multiple CT modes “cancel” out in coarse grained, population level data? How do sampling and time averaging affect this?

These questions are about **detecting equifinality** when formulating sets of models. Can we statistically “screen” for equifinality?

Methods

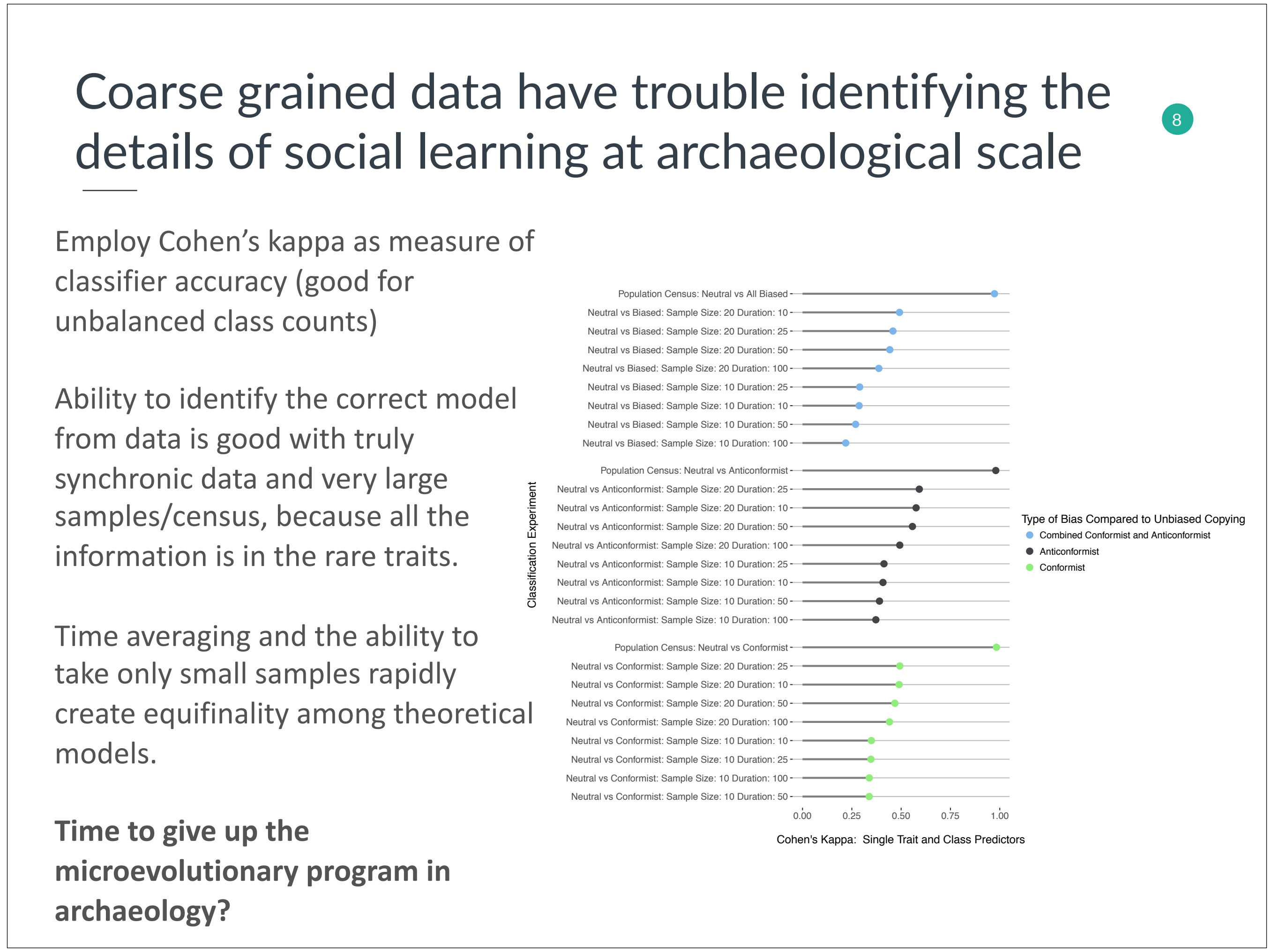
Simulation of mixtures of conformists, unbiased copiers, novelty seekers, and pure unbiased population.

Simulator tracked raw and time averaged trait counts, attribute vs. class, different sample sizes.

Equifinality measured as approximation to Bayes error rate classifying samples compared to true DG process.

Classifier used 23 summary statistics as predictor variables, measured on class and per-attribute counts.

- It’s clear you can identify social learning modes from transmission chain data and experiments — fine grained data. Less clear that you can with population-level, coarse-grain data. Add TA?
- Heterogeneity in the population might make this problem even worse!
- Whether particular CT models are equifinal with coarse grained data
- AND concerned to establish a method for examining equifinality among models ahead of time
- AND reflect data collection — sampling, duration — affect whether data with certain characteristics underdetermine a set of models



- Shown are the classifier accuracy results when we examine statistics at both the class and single attribute level; results for just per-attribute statistics show the same pattern; the effect of classification here is not apparent
- So far, we've looked at pure time averaging, and the effect of only having population-level data to identify individual level models. Kandler and others have suggested diachronic measures (but I included Kandler's diachronic trait survival in this classification model!). This, to me, is clear evidence that we shouldn't be trying to fix the microevolutionary approach — for archy. Perfectly good for contemporary data and living populations, but we need to move on.
- The classifier approach shown here is a good way of exploring issues of equifinality and underdetermination when planning a study; and will be useful even if we "scale up" and stop trying to look at individual populations and move to meso and macro scales.

Can We Identify Regional Cultural Transmission Histories with Seriation Graphs? (Ch. 4-6)

9

Summary

Mesoscopic scale: regional variation, assemblage durations of 10s or 100s years. How do we formalize **models** about evolutionary history at this scale?

What **chronicles** (units) can we fit to historical **models**?

Explored how modern **seriation graphs** provide the **chronicles**; time-varying graphs or “**temporal networks**” formalize a **model** of evolutionary history to test.

What kinds of histories can we distinguish (**equifinality**)?

Methods

Interval temporal networks model a metapopulation; forms hypothesis about population structure and history.

Simulate CT on the ITN; perform TA and sampling; seriate results as graphs

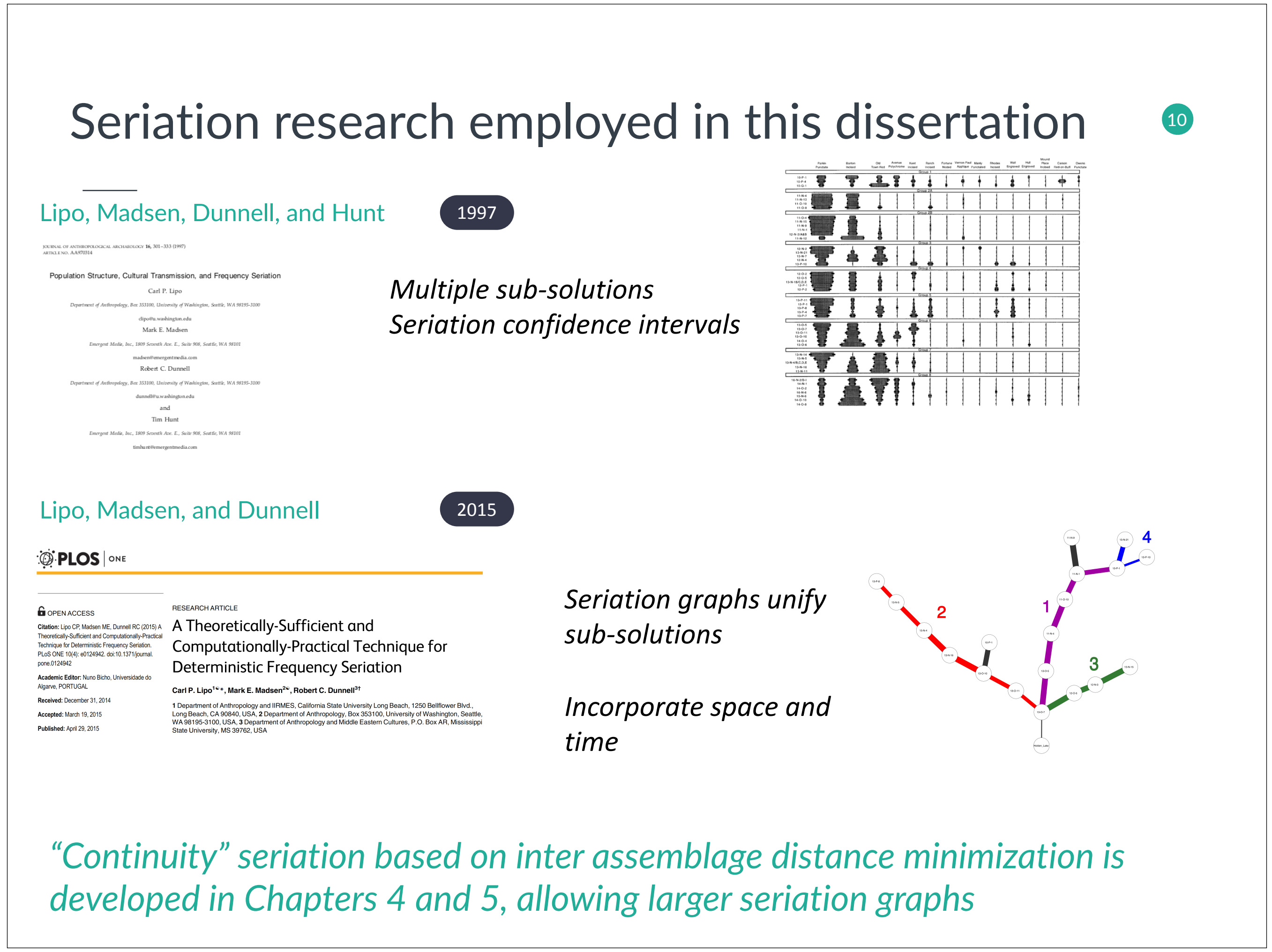
Characterize seriation structure as Laplacian spectrum of the solution graph

Train classifier on seriation spectra from 4 regional models and evaluate equifinality

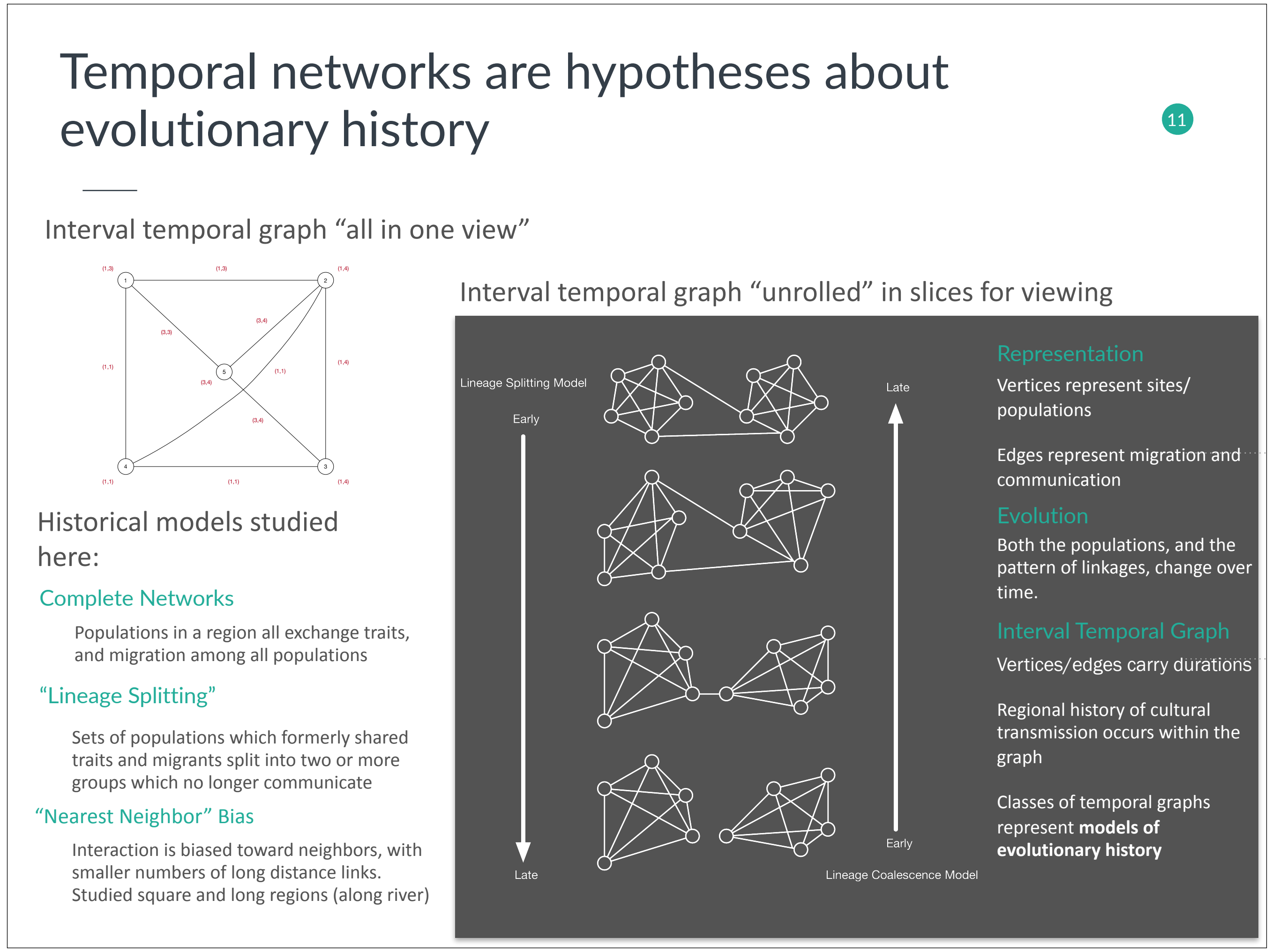
Use trained classifier on Mississippian ceramic seriations

SAA 2015; HBES 2016

- Going to spend more time on this study since it represents the bulk of my new method development over a number of years
- In the microevolutionary literature our fundamental observable were frequencies of types/classes (often a small number) in an assemblage. Here we are going to design new observables with a large state space, with time and space baked in, so we can fit diachronic models to these new observables.
- On the theoretical side, we’re going to use the same classifier approach to assess equifinality among models, using seriation graphs as the observable
- On the empirical side, we then use the trained classifier to generate posterior predictive distributions of seriations for each class and in ABC fashion, see where the empirical data fit.

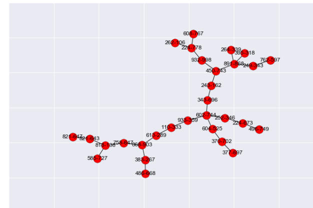


- Because we have a lot to cover, I've omitted details about some of the seriation innovations. The work in Chapters 4 and 5 builds from my previous work in collaboration with Dr. Lipo. In chapter 4, I look at the combinatorics of breaking seriations into multiple sub-solutions, and how it affects the scale of the problem (a lot!). In Chapter 5, I worked with Dr. Lipo on replacing unimodality with inter assemblage distance minimization (previously proposed by Kadane 1971), to achieve better scaling properties. We demonstrated equivalence to uni-modality results.
- Hugely important because we need rich structure in seriation graphs if we're going to use them to fit statistical models, and you get the potential for more structure in a graph with more vertices — the number of distinguishable trees for example is n^(n-2). (Cayley's theorem).



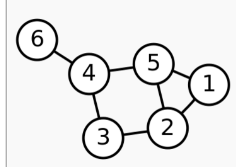
- After generating sets of random temporal graphs corresponding to each of these model classes
- Use the temporal graphs as metapopulation models
- Simulate CT in the metapopulation; time average the results over each vertex duration
- Sample the assemblages and seriate using our IDSS algorithm with continuity
- Collect the resulting seriations with their originating temporal network model as a “label” for machine learning

Quantifying seriation graph structure



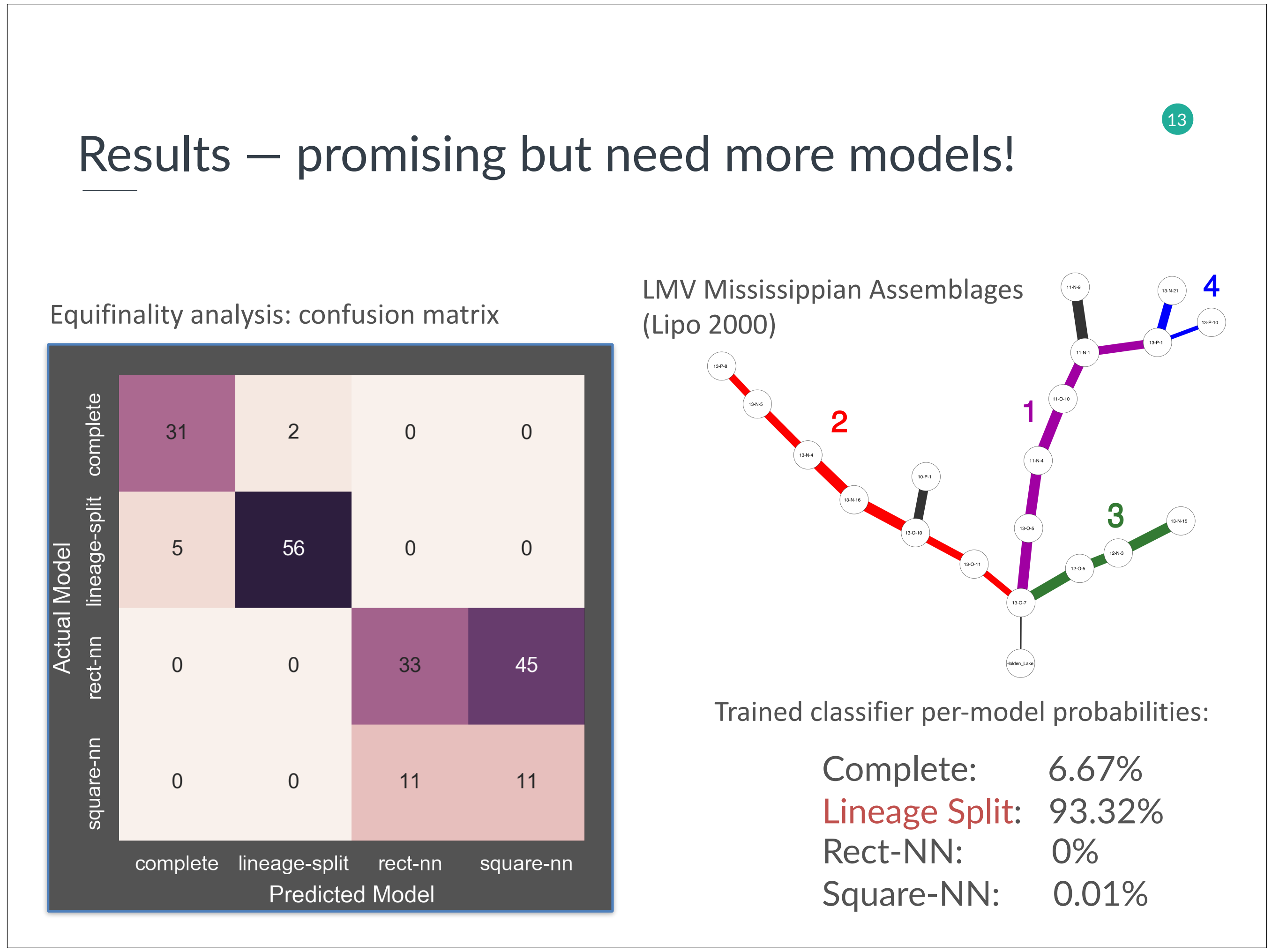
Example seriation solution from simulated cultural evolution on a social network which "split" into two lineages after evolving as one large lineage

The "structure" of edges and vertices in a graph is given by its Laplacian matrix, which can be summarized by its eigenvalue spectrum

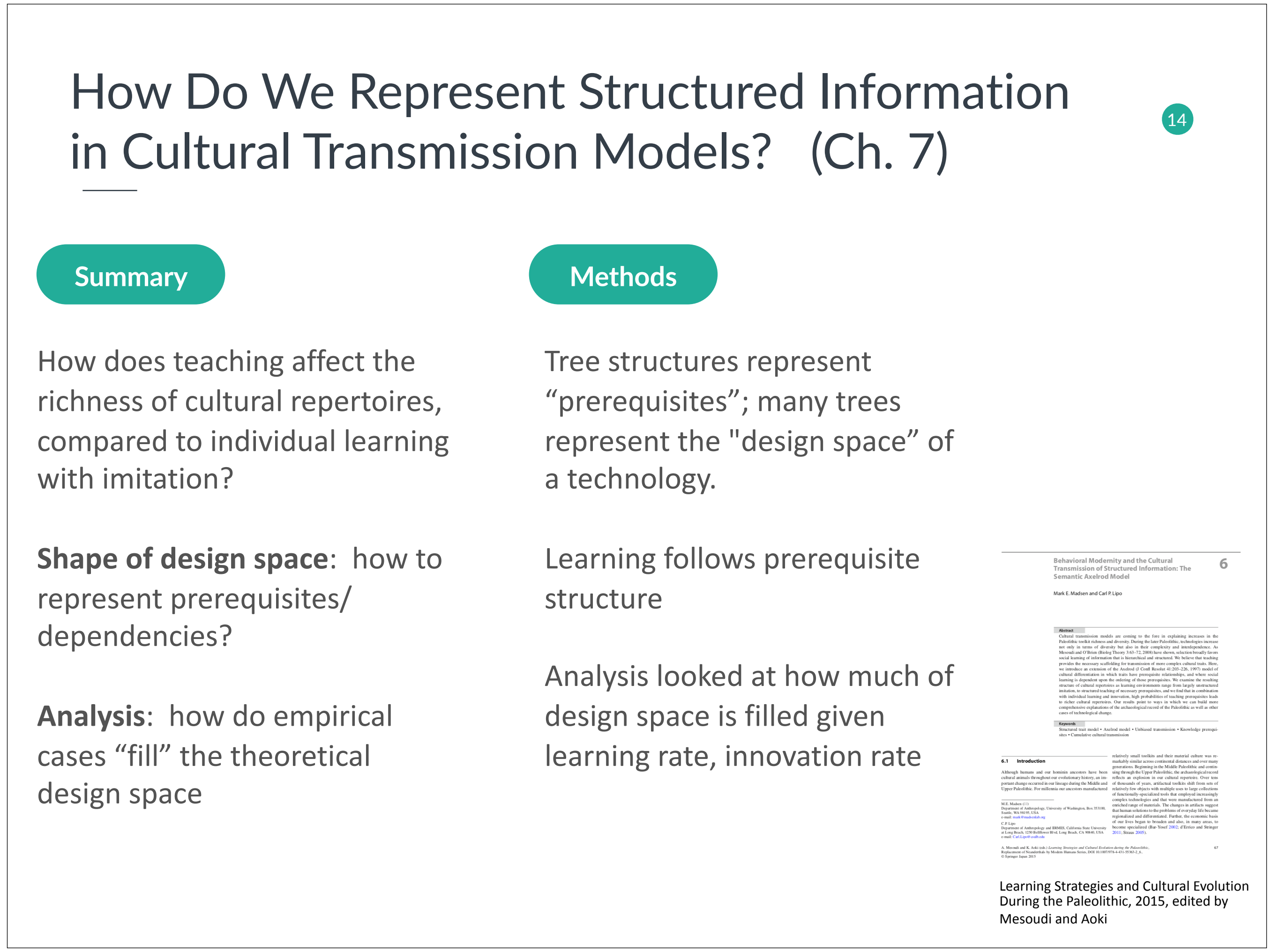
Labeled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2&0&0&0&0&0\\ 0&3&0&0&0&0\\ 0&0&2&0&0&0\\ 0&0&0&3&0&0\\ 0&0&0&0&3&0\\ 0&0&0&0&0&1 \end{pmatrix}$	$\begin{pmatrix} 0&1&0&0&1&0\\ 1&0&1&0&1&0\\ 0&1&0&1&0&0\\ 0&0&1&0&1&1\\ 1&1&0&1&0&0\\ 0&0&0&1&0&0 \end{pmatrix}$	$\begin{pmatrix} 2&-1&0&0&-1&0\\ -1&3&-1&0&-1&0\\ 0&-1&2&-1&0&0\\ 0&0&-1&3&-1&-1\\ -1&-1&0&-1&3&0\\ 0&0&0&-1&0&1 \end{pmatrix}$

The eigenvalues are then used as predictor variables in a classifier model, to predict the correct evolutionary model.

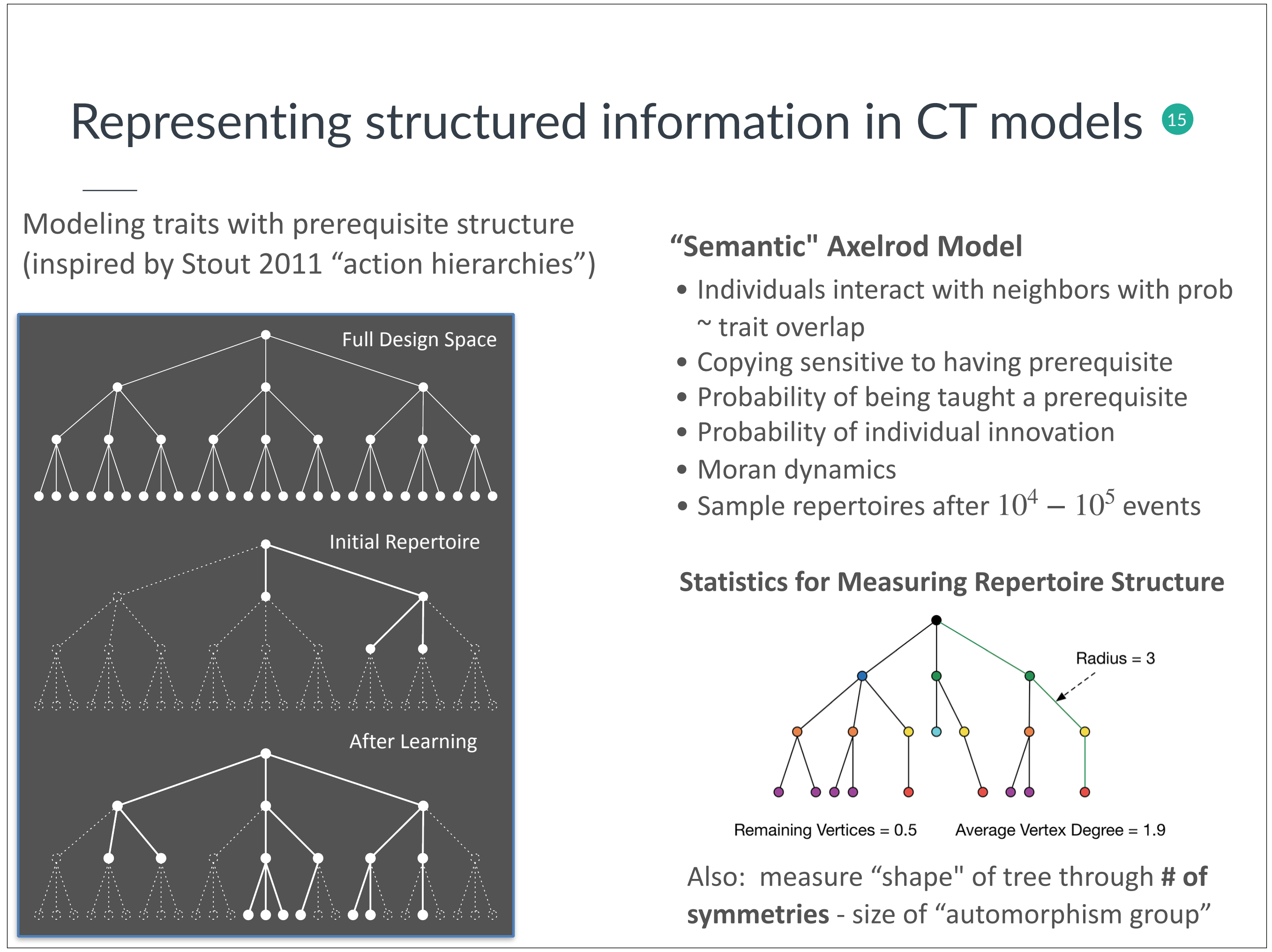
- In order to treat seriation graphs as statistical objects, we measure summary statistics about their "structure" — branching, "shape", etc.
- Graph theory shows us that (especially for trees), all the information about structure — and what equivalence classes of structure — is encoded by Laplacian matrices. The simplest Laplacian shown here is just the degree matrix minus the adjacency matrix.
- A variety of results in algebraic graph theory then tell us that the "spectrum" of eigenvalues (and their multiplicities) encode the info available in the Laplacian
- So the eigenvalues are our summary statistics about seriation graphs, and become predictor variables in our training set for a classifier to distinguish equivalence classes of seriation graphs



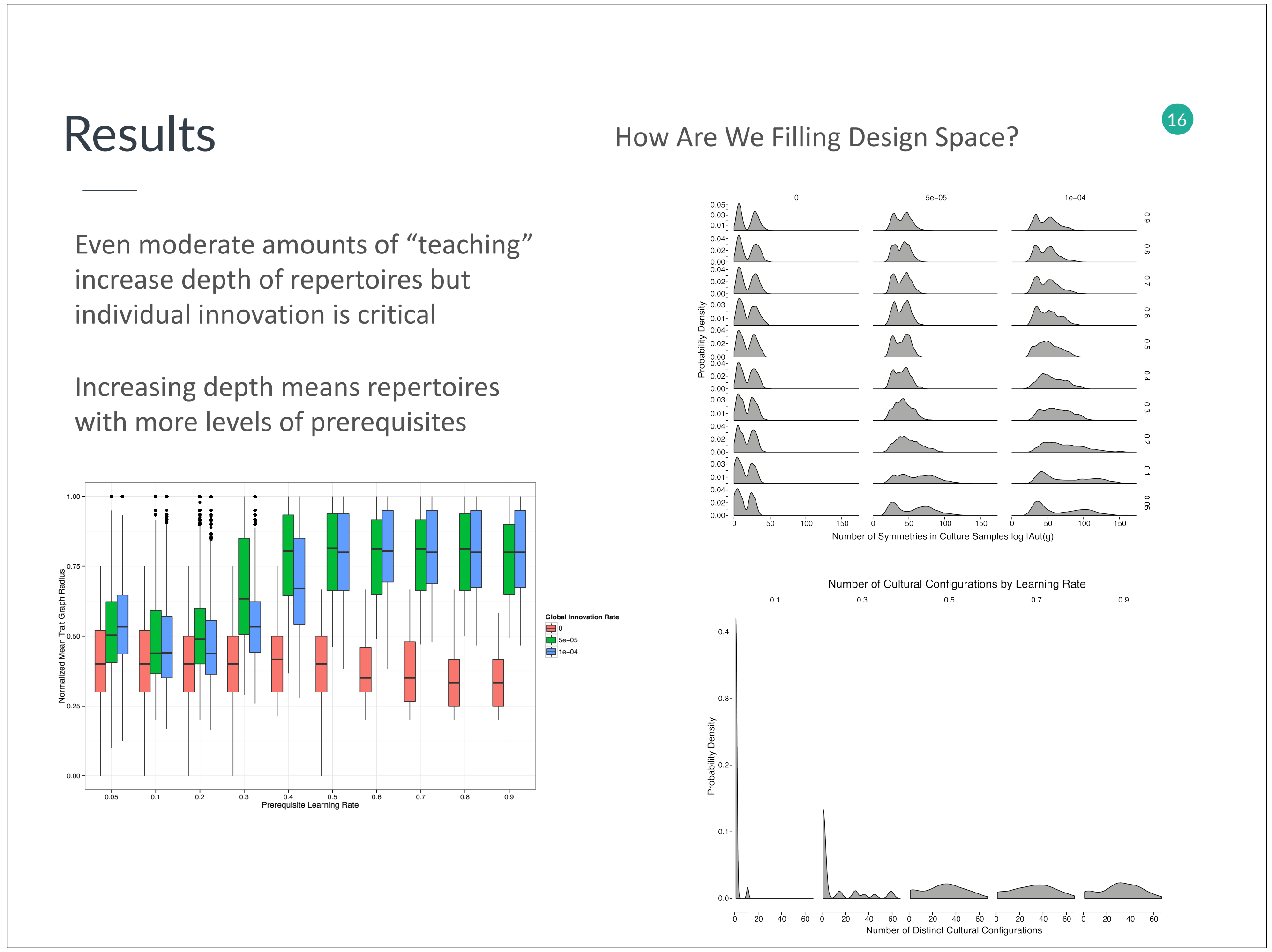
- Other models initially prototyped include a hierarchical model but I did not feel confident about how to represent hierarchical relationships, so I omitted that from final experiments
- I regard this as really a proof of concept - there's so much that can be done. The incorporation of space with different grid shapes needs to be refined into real geograph. The simulations need to be much less computationally expensive so that we can achieve better randomization across model realizations. And there is huge room to study the time-varying models themselves for equifinality rather than using simulated seriations as a probe — first find models which are distinguishable directly, THEN spend the time and money to create predictive distributions of seriations for empirical ABC fitting.



- And now for something completely different
- Up to now, I've been looking at spatiotemporal observables and spatiotemporal models
- But mapping evolutionary history isn't just about flow, it's about the content of cultural traits and how they're modified through evolutionary time.
- So let's get thicker in our descriptions. The goal here is to understand the tools we'll need to get thicker and still have dynamical models of how things evolve

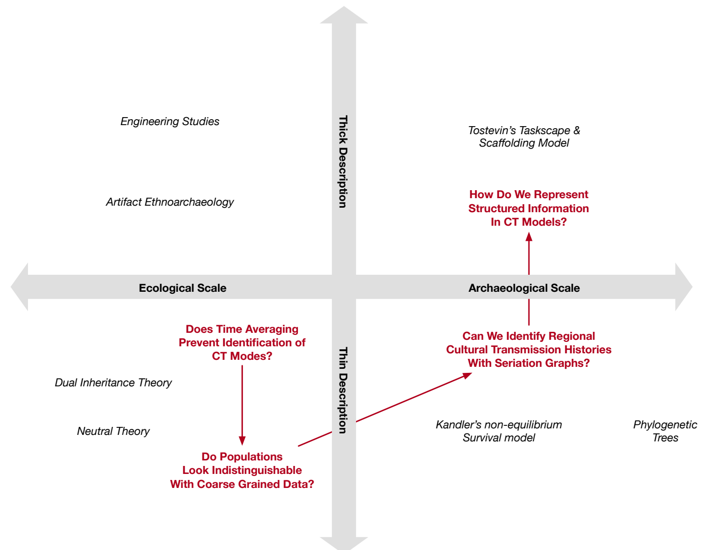


- Modeling dependencies as a tree is a modeling choice; more generally we should expect relations to be multigraphs. I started with trees because they provided a clear way to measure “how much of design space” are we occupying with different social learning parameters
- This was inspired also by Stout’s “action hierarchies”
- Note on symmetries - an abstract way to measure “how many interchangeable” patterns there are, and thus the degree to which we’re filling design space — the full design space here has 1.3 x 10^10 symmetries. A tree with no repeating patterns has 1 symmetry. Automorphism groups are collections of those symmetries.



- This, again, is a good start. Next steps are to look at more general kinds of dependency relationships
- Work with somebody like Gilbert Tostevin, who’s working on thicker models of social learning for lithic reduction. Can we build representations for lithic reduction sequences so we can look at the evolutionary history of meaningful units of knowledge, not just atomistic traits?
- What makes this worth following up is that thicker models cut through many of the divisions in cultural evolution - between the “popgen-like modelers” — B&R, neutral models, and the folks who claim that such models are sterile and non-reflective of how cultural transmission really works (Sperber, Cladiere, etc). Thicker description of transmission integrate what we know about the physics of a technology with models of how it is used and how skills are learned — this is evolutionary developmental (evo-devo) for culture — which is what Wimsatt and Greisemer have been telling us we need for a decade.

Conclusions and Contributions



- Effects of time averaging, particularly relationship to trait turnover
- Method for screening equifinality via classifier models and predictive data distributions
- Temporal networks as diachronic models for CT in space/time
- Seriation graphs as the diachronic observable at the mesoscale
- Dependency trees as the observable for content of transmitted traits

- The first two studies were cautionary — does it work and should we do this? The answers were essentially both no.
- The second two studies were about building the new — one thin and spatiotemporally focused, one thicker and content focused.
- A key step is also to understand the relationship of these methods to true macro methods and questions — seriation graphs to me relate to phylogenetic trees through classification scale — frequencies vs. presence/absence — this gives phylogenetic methods their characteristic "macro" scale
- At the end of the day, our job finding methods to document evolutionary history and heritable continuity is just getting started. My own approach matches the best parts of traditional culture historical method and the tools we get from machine learning, graph theory, simulation toolkits, and approximate Bayesian fitting for empirical data.

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