

Cultural macroevolution - types of questions

Relatedness

Given samples of cultural traits, can we infer the pattern of homology and cultural inheritance among populations and whole lineages?

Social Networks

If we have data on cultural variants from many communities across a region, over multiple time periods, can we infer the general structure of social interaction (the “network”)?

Technological Change

What can we say about the processes of social and individual learning by which suites of artifacts become more complex, specialized, and effective? What patterns do such processes leave in large scale data?



Today's talk will focus on the middle question: inference of the temporal and topological structure of regional-scale social networks

Using ML/ABC to infer social networks

1

Choose a method for modeling social networks that change structure and membership

Identify classes of social network structures that we want to distinguish from macroevolutionary data

2

3

Choose an “observable” (e.g., seriations) that can be measured in real data, along with simulated data, and summary statistics on them

Simulate cultural transmission across the evolving social network models, saving samples of trait frequencies, and aggregating them in realistic ways

4

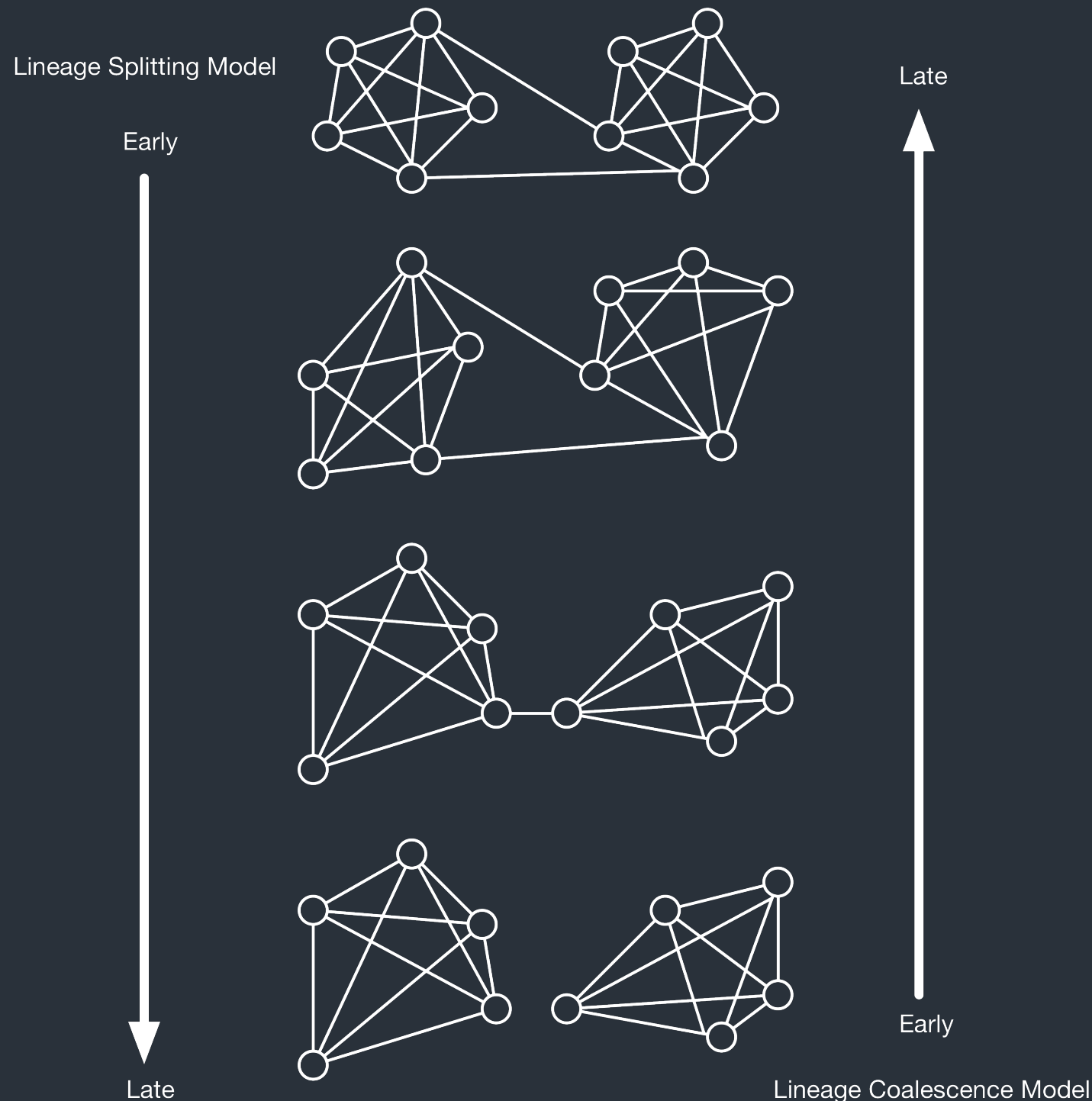
5

Construct seriations and calculate summary statistics for each simulation and social network structure

Build a ML classifier model (e.g., random forests or GB trees) to predict which social network structure produced each simulated data set

6

Modeling Social Network Evolution



Representation

Vertices represent populations

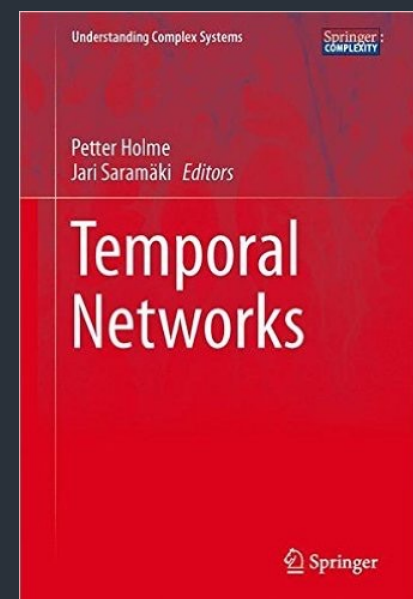
Edges represent migration and sharing of cultural traits

Evolution

Both the populations, and the pattern of linkages, change over time.

Interval Temporal Graph

Represents change as a sequence of graphs, indexed by time, each recording a change to an edge, addition of a community, or loss of a population.



Classes of evolving social network structures

K Complete Networks

Populations in a region all exchange traits, and migration among all populations

L “Lineage Splitting”

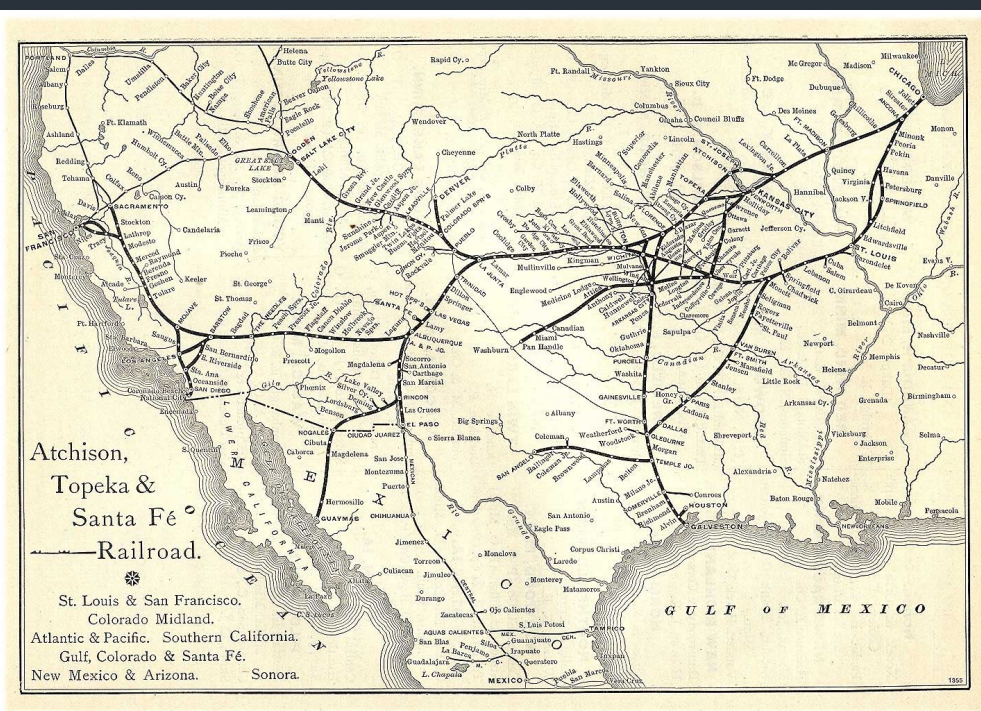
Sets of populations which formerly shared traits and migrants split into two or more groups which no longer communicate

C Lineage Coalescence

Sets of populations which were formerly separate gain ties and begin sharing traits and migration occurs between communities

N “Nearest Neighbor” Bias

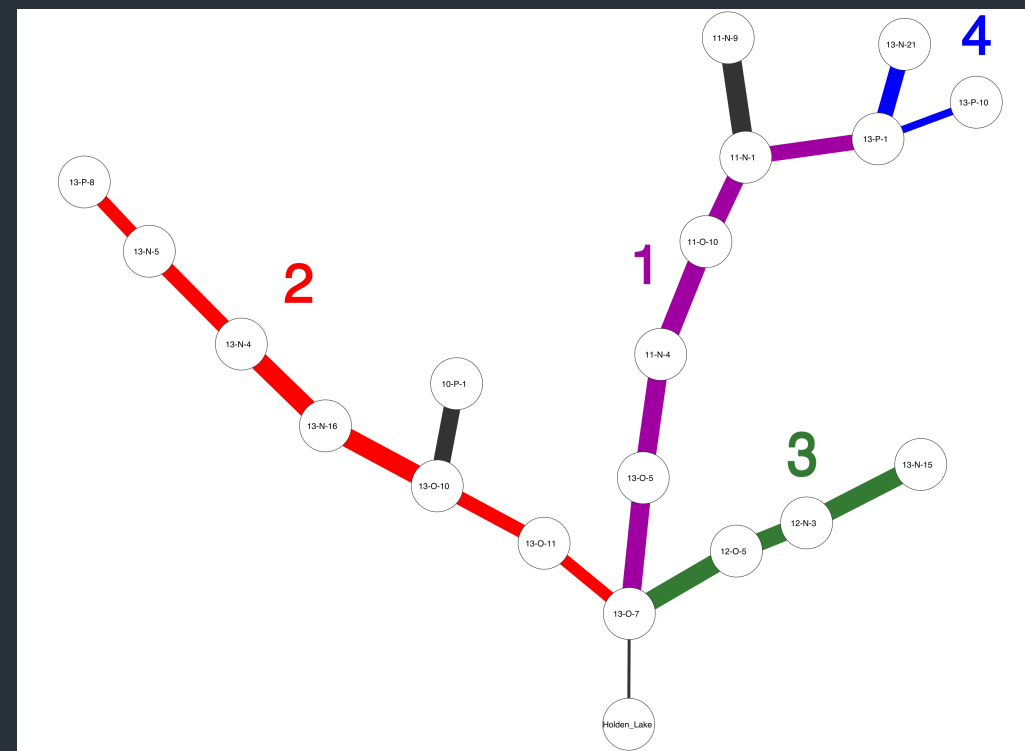
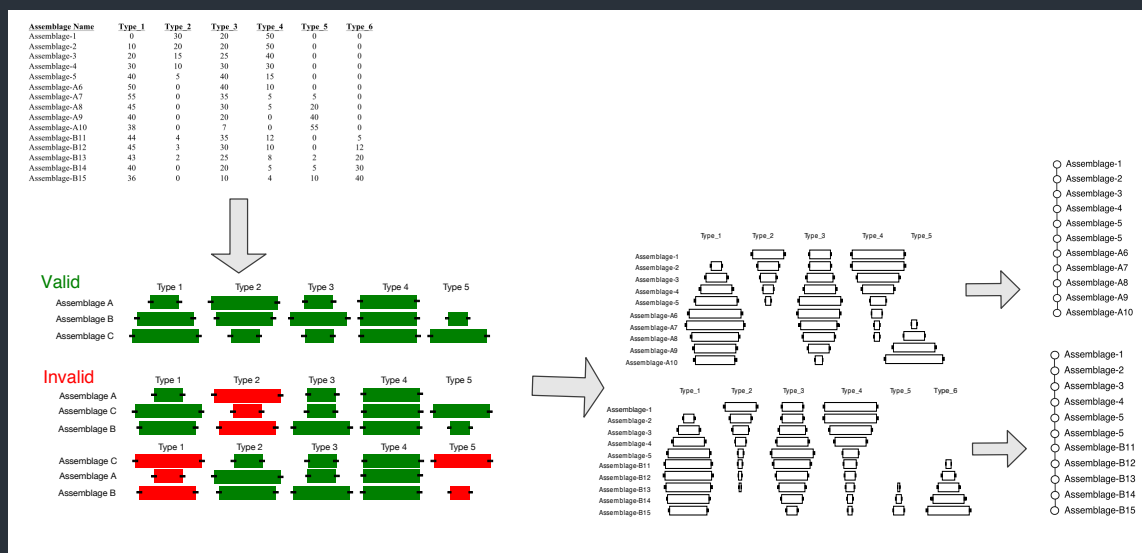
Interaction is biased toward immediate neighbors (either weakly or strongly), with smaller numbers of long distance links



Macroevolutionary “observable”: seriations

Seriation

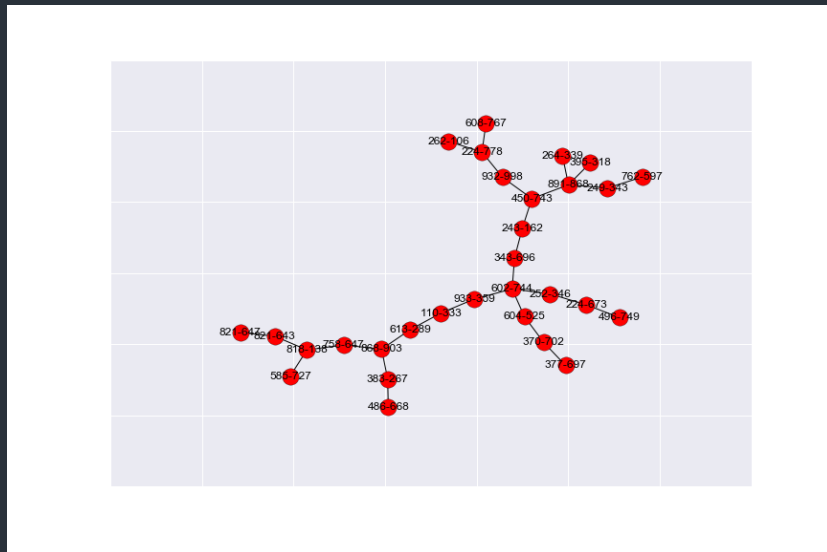
Combinatorial method of ordering samples in space and time given an inter-sample distance measure.



Traditional seriation solutions tended to try to force orders to be linear, instead of recognizing that the simultaneous effects of spatial difference and temporal change cause “branching” when samples are fit together in the optimal order.

Lipo and Madsen (2015) introduced seriation graphs as a tool to capture spatiotemporal ordering in a single structure.

Quantifying Seriation “Structure”



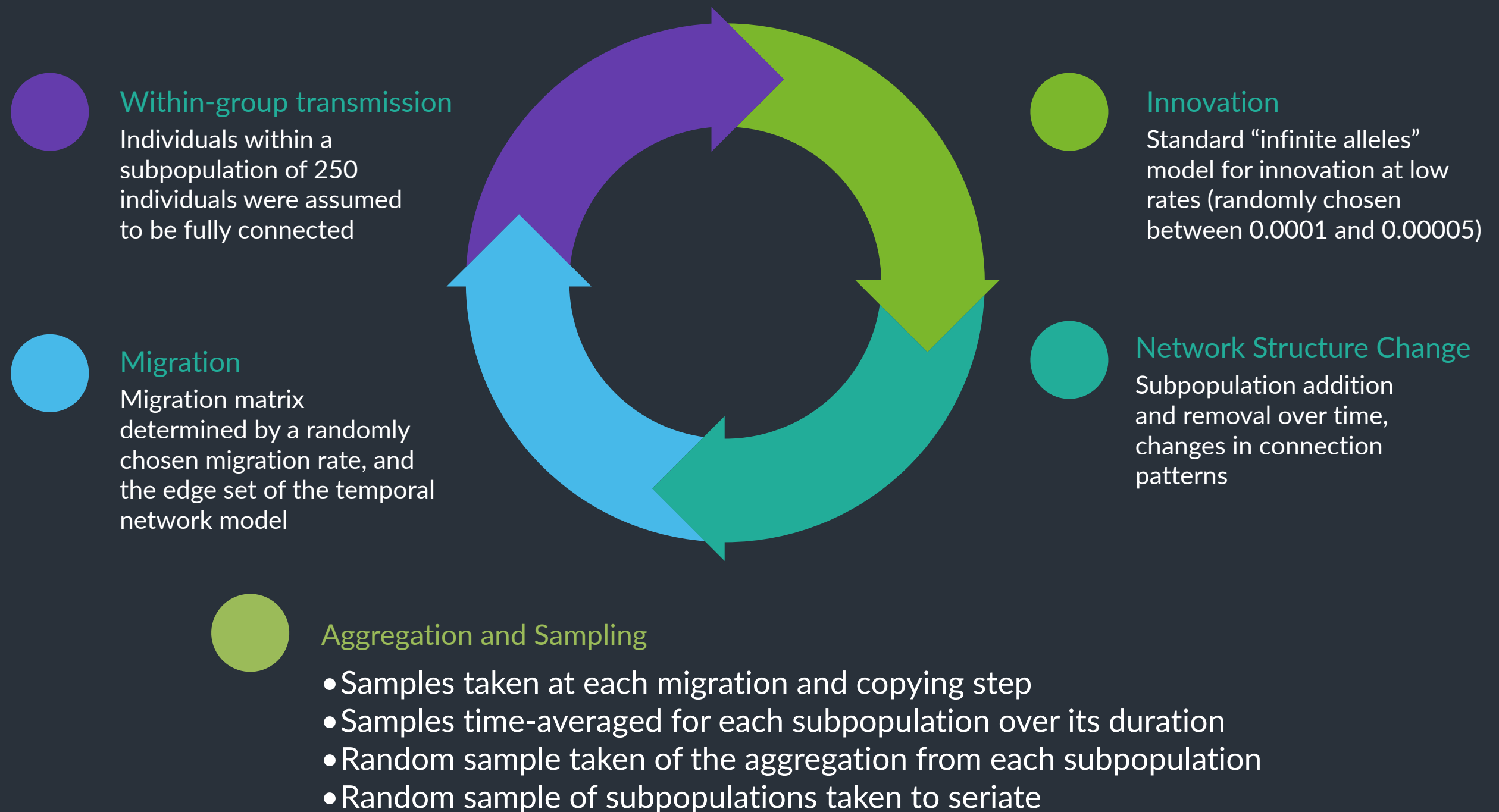
Example seriation solution from simulated cultural evolution on a social network which “split” into two lineages after evolving as one large lineage

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 eigenvalue spectrum becomes the *independent variables* in a machine learning classifier, aimed at correctly predicting the network model from seriation output.

Simulating Cultural Evolution on Temporal Network



Hold-out test set performance good!

Model: Gradient Boosted Decision Trees, hyper parameter optimization over tree depth, number of trees, and learning rate

Actual Model	complete	lineage-split	rect-nn	square-nn	
	31	2	0	0	
	5	56	0	0	
	0	0	33	45	
	0	0	11	11	
		Predicted Model			
	complete	lineage-split	rect-nn	square-nn	

1

Can differentiate complete from nearest neighbor from splitting/coalescence models

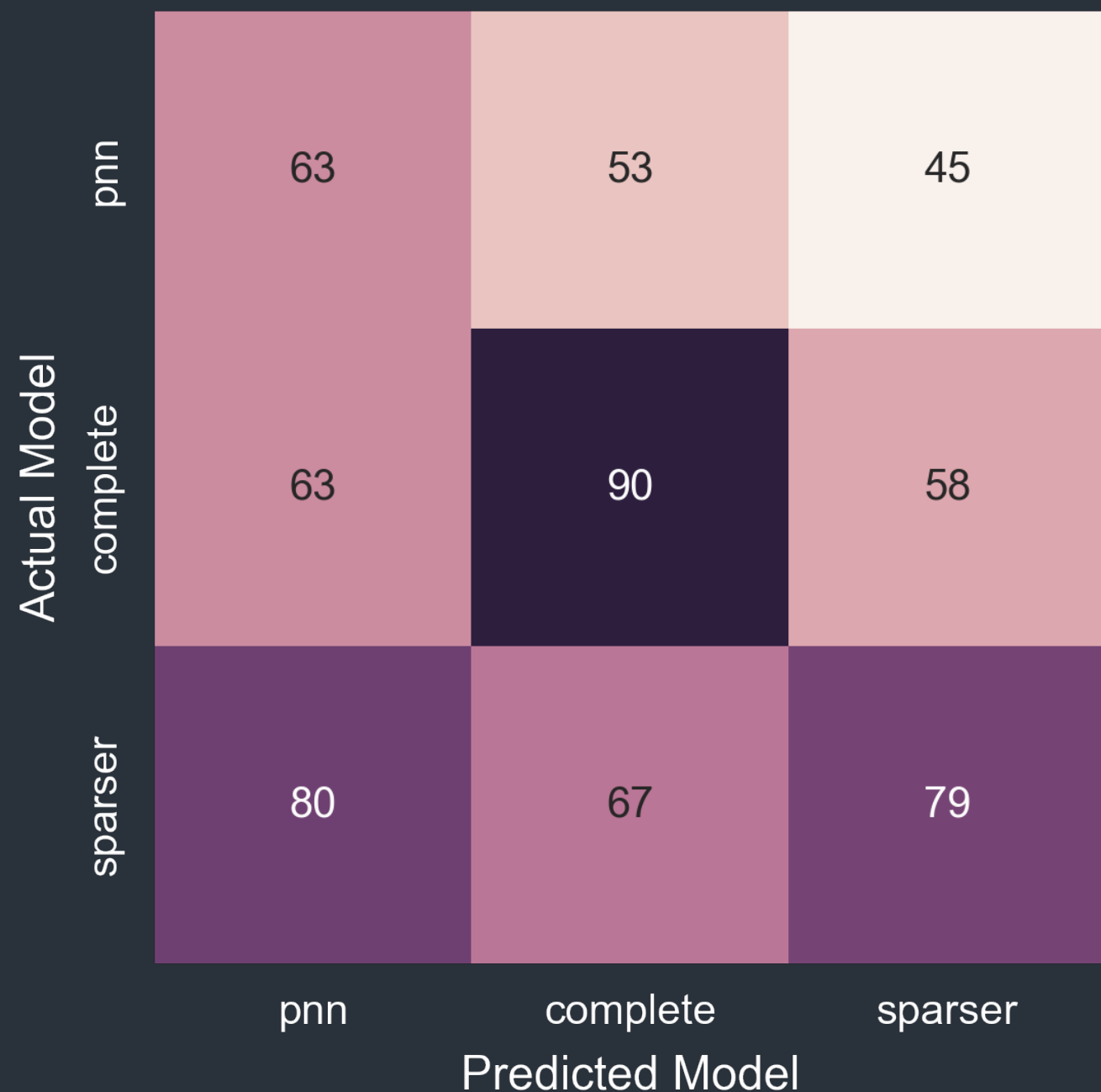
2

Not able to tell different spatial configurations of the same gross network structure apart

3

One limitation here was limited randomization over the network structure itself

Performance poor when network structures similar...



1

“PNN” model here randomizes further over the “nearest neighbor” model and I ran two groups, one with sparser and shorter connections than the other

2

No real ability to distinguish these “probabilistic” nearest neighbor networks from completely connected networks

Identifying empirical cases

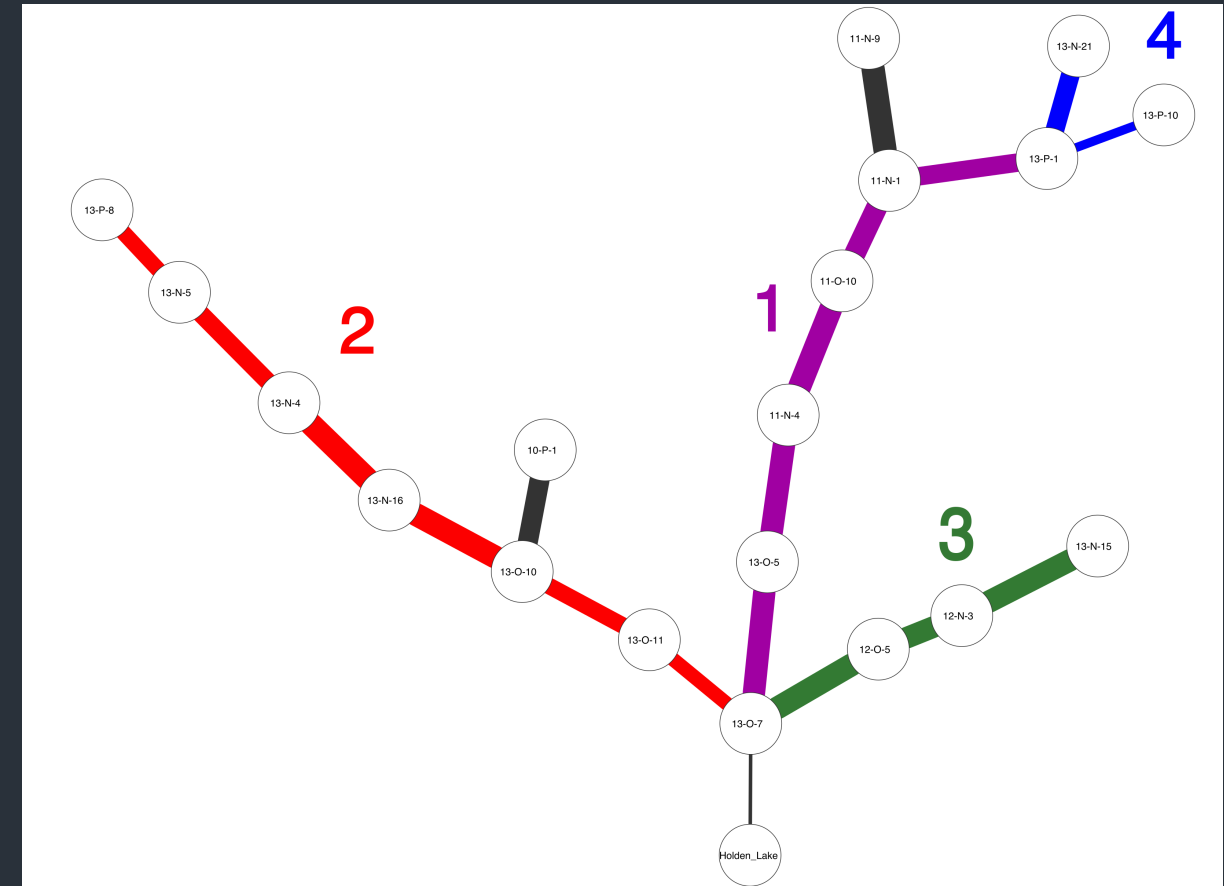
11

1

Given the trained GB model, we can identify which network model an empirical case is most closely associated with.

2

Of the existing models, Late Prehistoric communities in the central Mississippi River valley underwent a lineage splitting event. It is not possible with these models to determine more about internal structure (yet!)



Complete:	0.0667
Lineage Split:	0.9332
Rect-NN:	0.0
Square-NN:	0.0001

Next steps in refining the approach...

- 1 Explore additional summary statistics for seriation graphs — may not be enough variation in the Laplacian eigenvalues given how strong the chronological component is in most seriations.
- 2 Implement more classes of temporal network models and get larger samples of randomized network structures
- 3 Explore whether observables in addition to seriation would be useful, including cladogram topologies, diversity indices on trait frequencies

Thank you!

Mark E. Madsen
University of Washington

<http://notebook.madsenlab.org>
[@mmadsensji](mailto:mmadsensji@notebook.madsenlab.org)
mark@madsenlab.org

