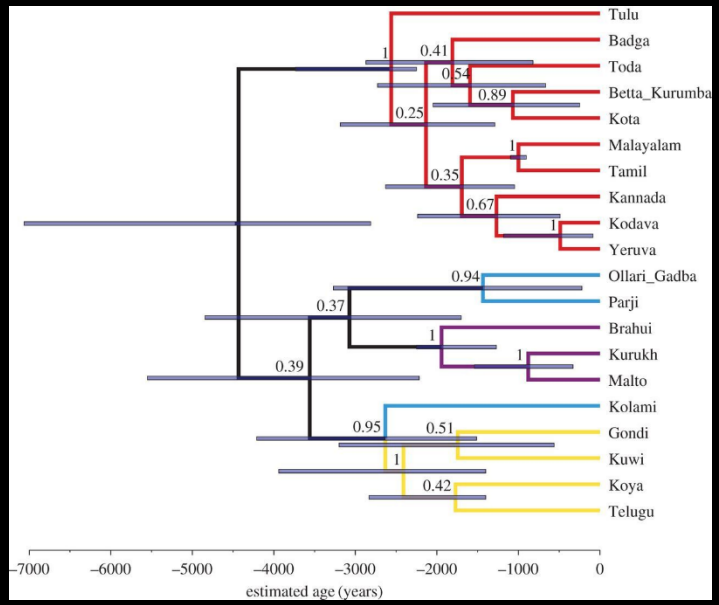
Latent Geographical Factors for Analyzing the Evolution of Dialects in Contact

MURAWAKI Yūgo
Kyoto University

EMNLP2020 @ online 16-20 Nov 2020

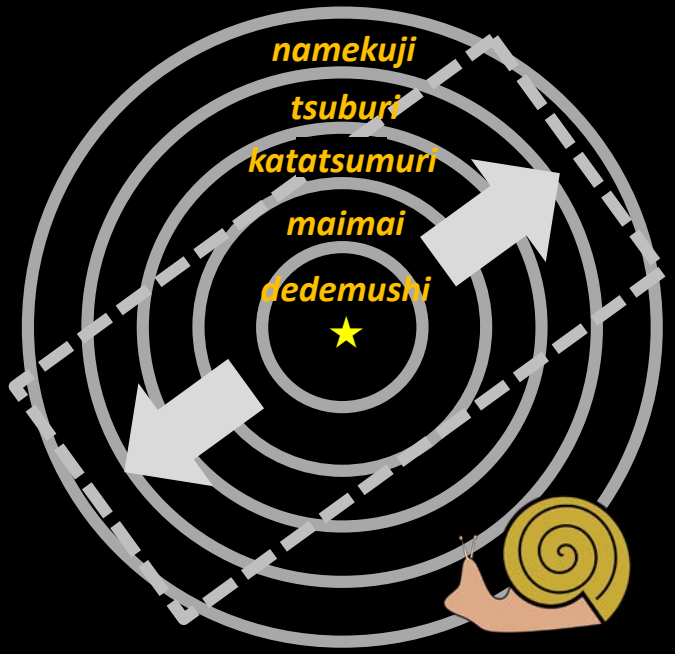
Background: Two modes of evolution

Vertical inheritance (tree-like)



[Kolipakam+, 2018]

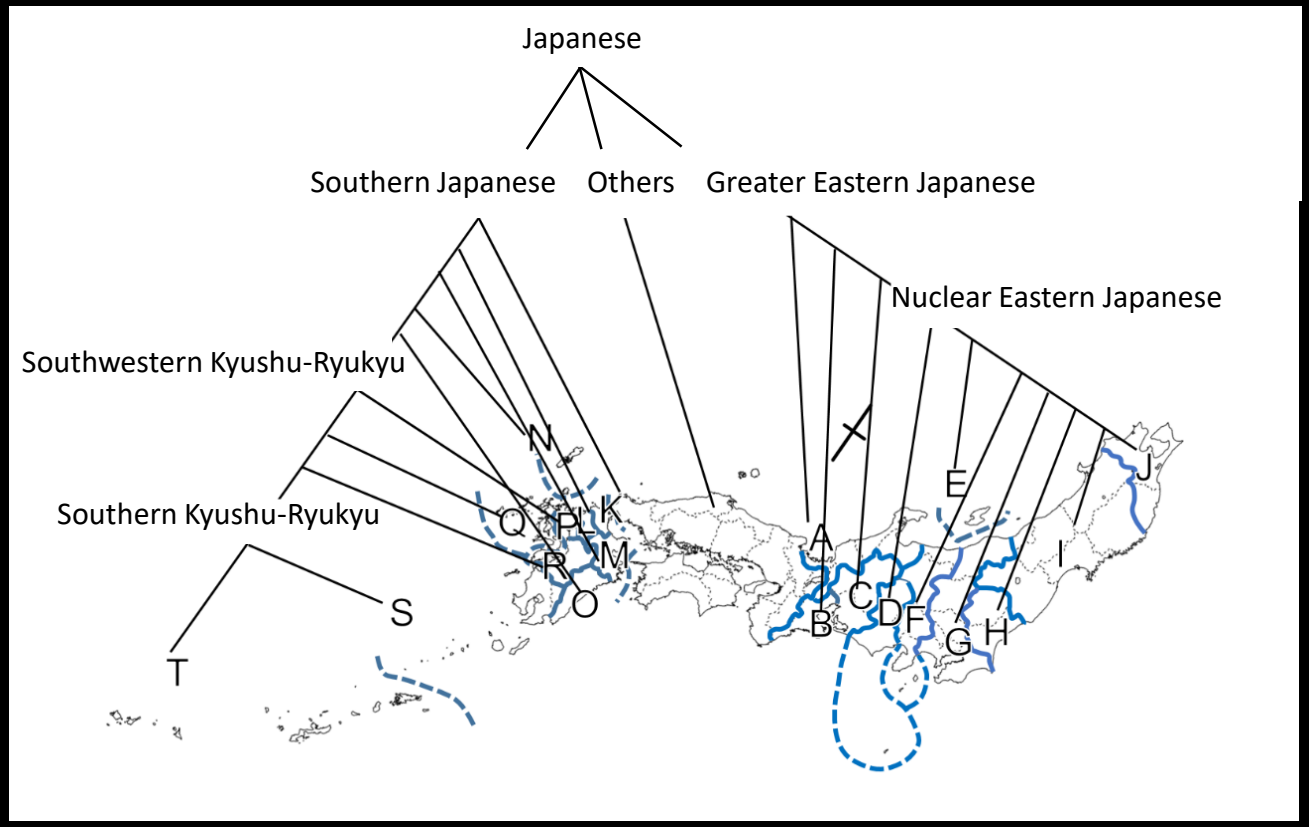
Horizontal transfer (wave-like)



[Yanagita, 1930]

The evolution of dialects in intense contact has mostly been explained in terms of horizontal transfer

Recent studies suggest even dialects show traces of vertical inheritance

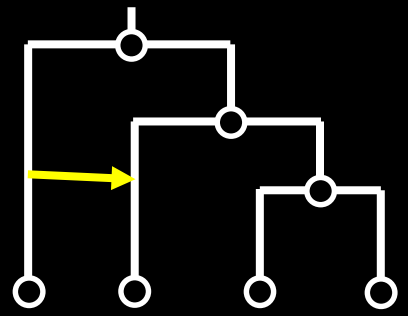


Adopted from [Igarashi, 2018]

Limitations of manual inference and statistical modeling as a solution

- Linguists base their inference on high certainty traits (*shared innovation*)
- But difficulties are inherent in telling signals of vertical inheritance from those of horizontal transfer
- Computer-intensive statistical models are better at handling uncertainty
  - The probabilistic tree model originating from evolutionary biology has been successful

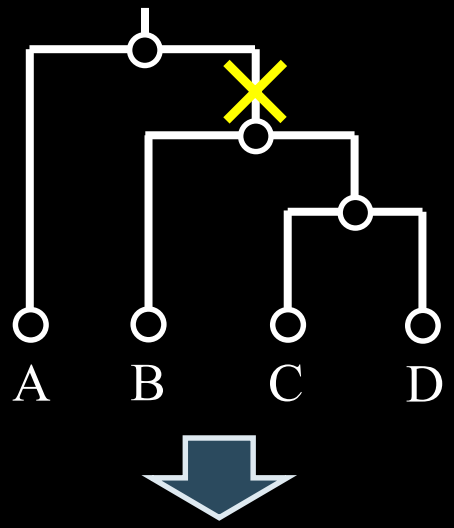
How can we jointly model the two modes of evolution?



- Incorporating horizontal transfer into the tree is impractical
- Some stick with the *vanilla* tree by showing its robustness wrt horizontal transfer to some degree [Greenhill+ 2009]
- Is it possible to develop a computationally feasible, approximate model of vertical inheritance and horizontal transfer?

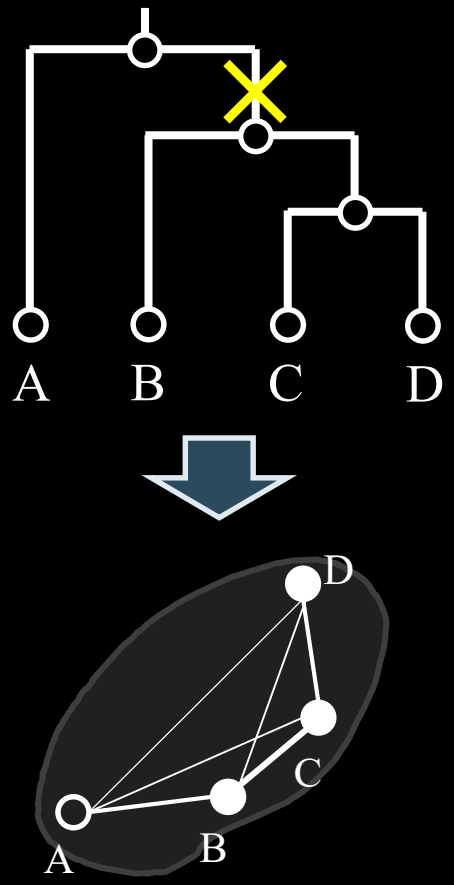
Idea: Both modes result in geographical distributions

Vertical inheritance (tree-like)

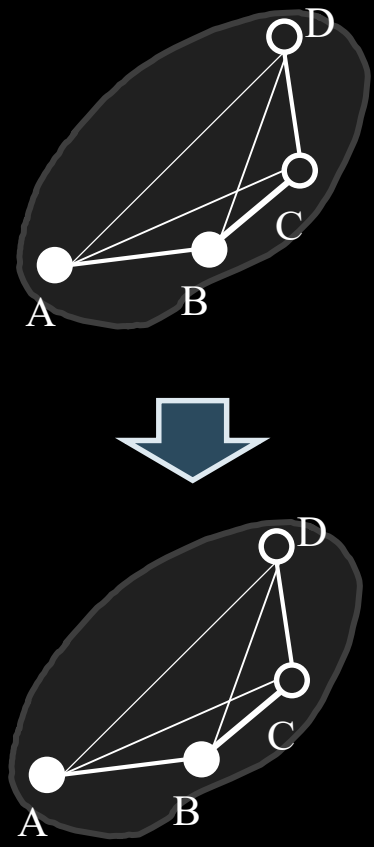


Idea: Both modes result in geographical distributions

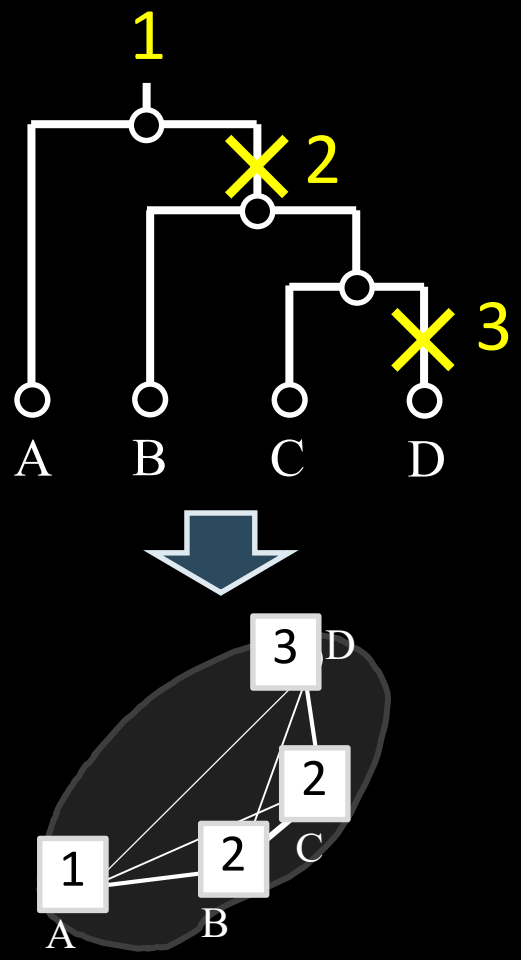
Vertical inheritance (tree-like)



Horizontal transfer (wave-like)

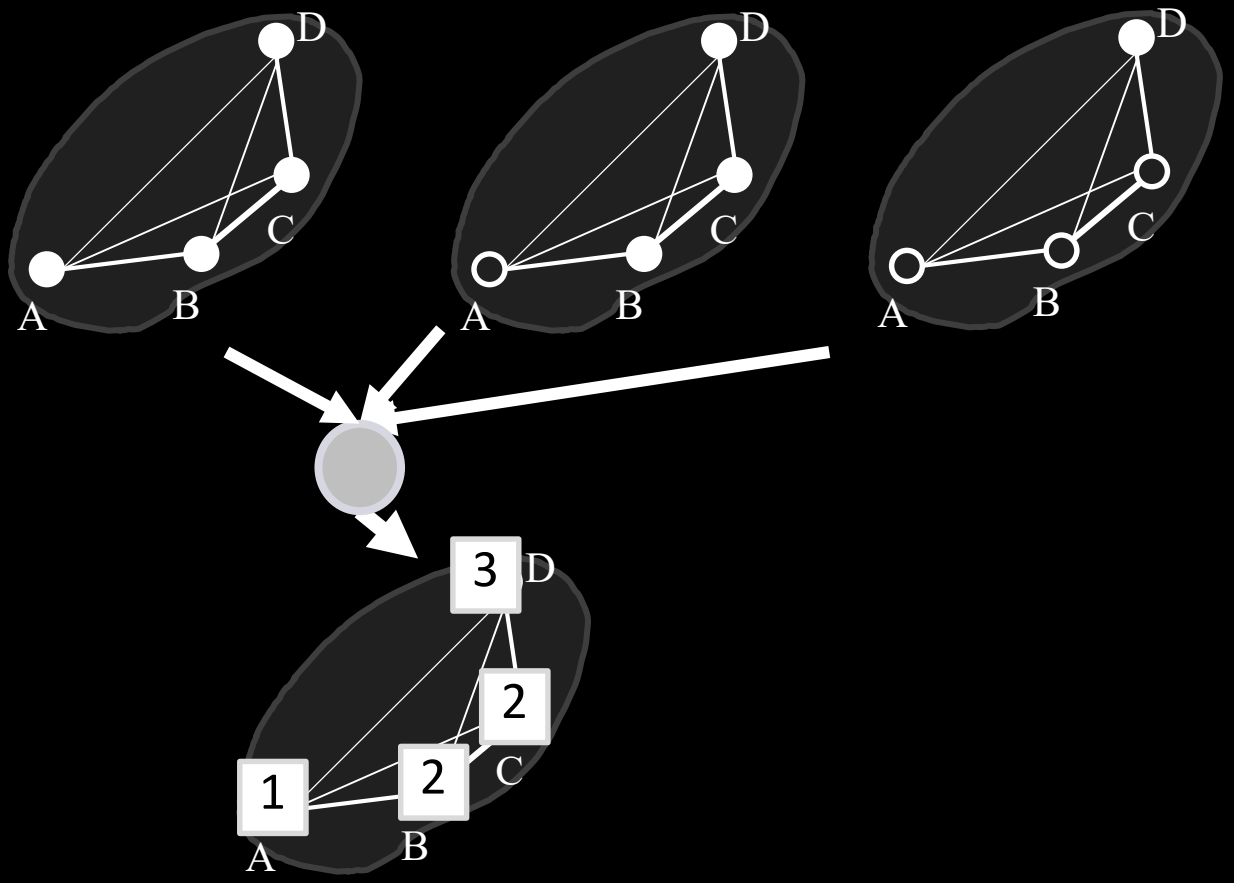


Geographical distributions are not necessarily directly observable; hence *latent*



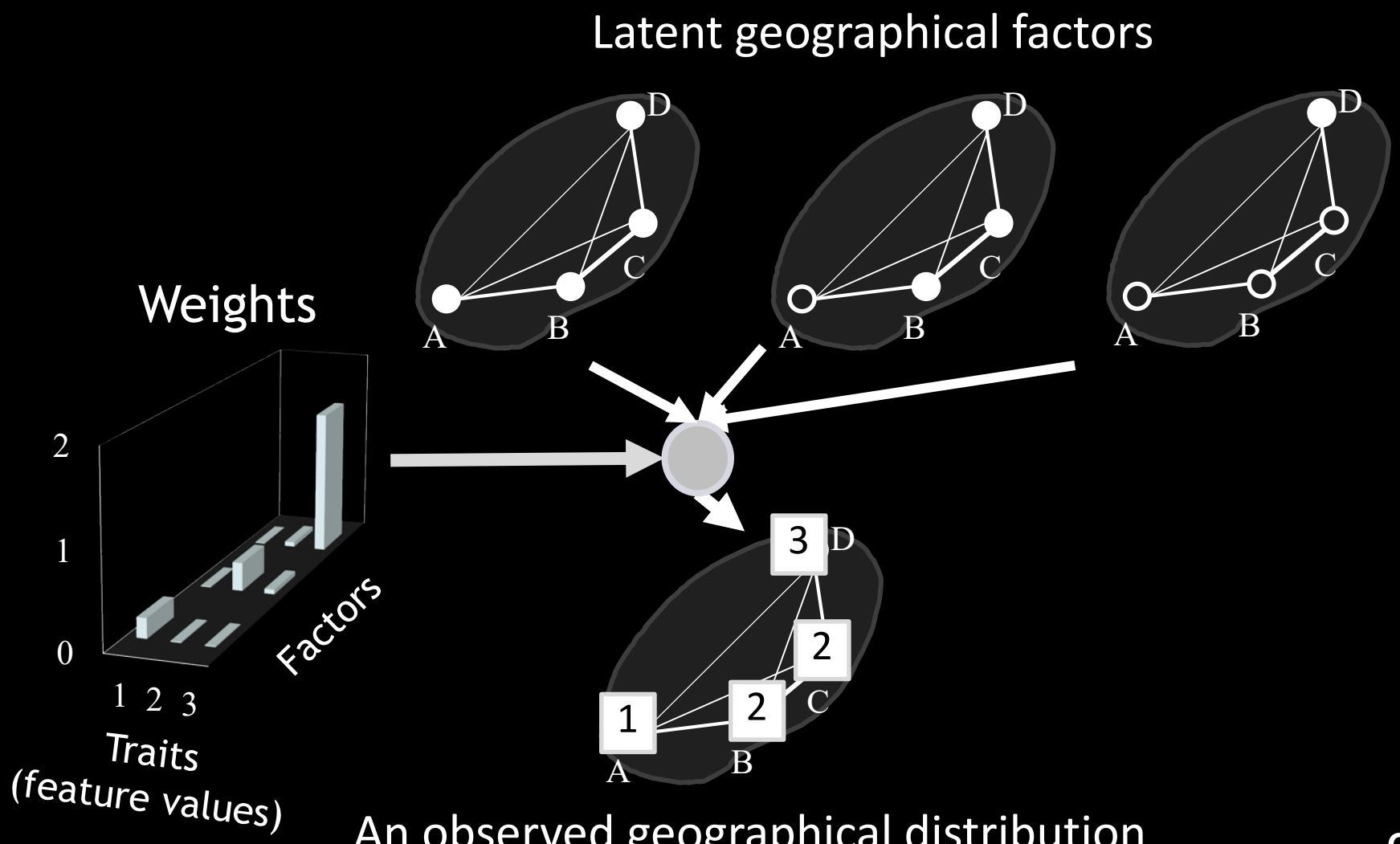
Geographical distributions are not necessarily directly observable; hence *latent*

Latent geographical factors



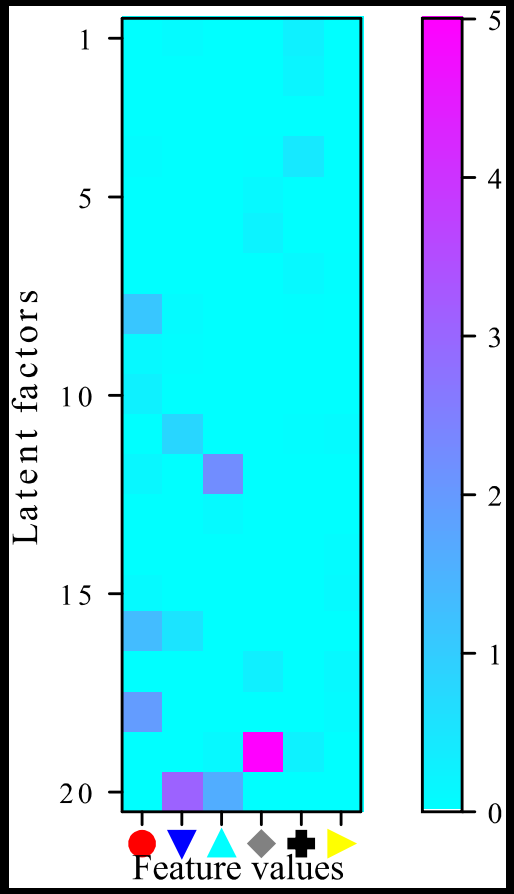
An observed geographical distribution

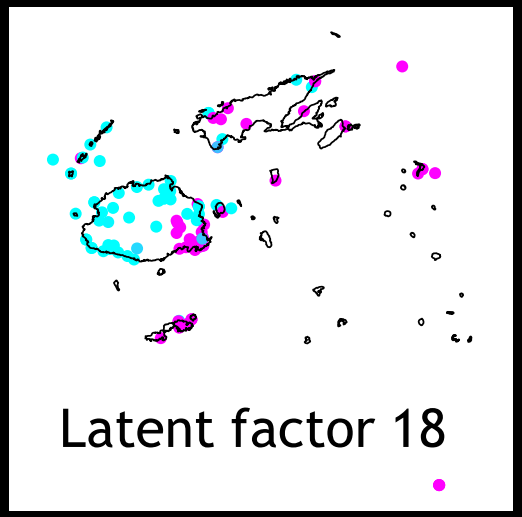
Latent and observed geographical distributions are stochastically connected with positive weights

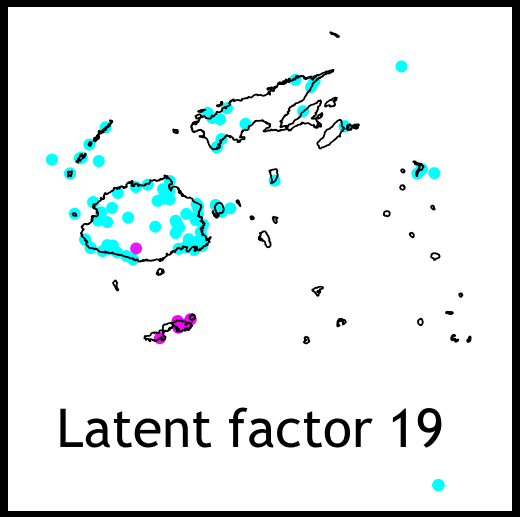


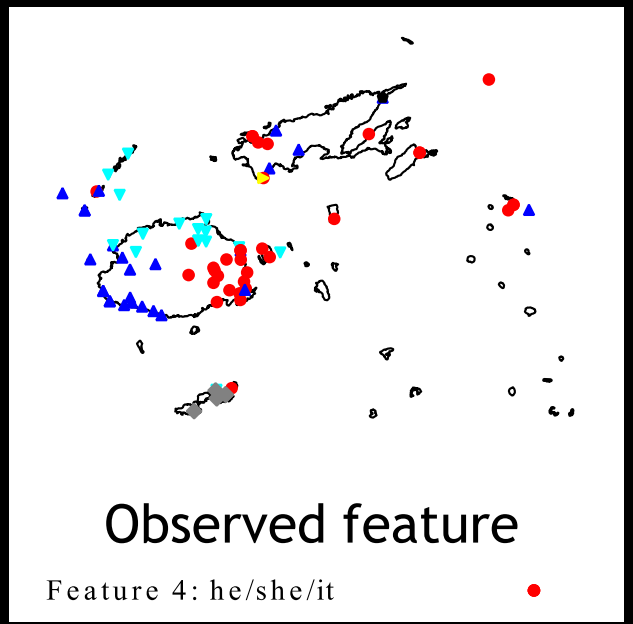
Experiments on Fijian lexical data

Weights



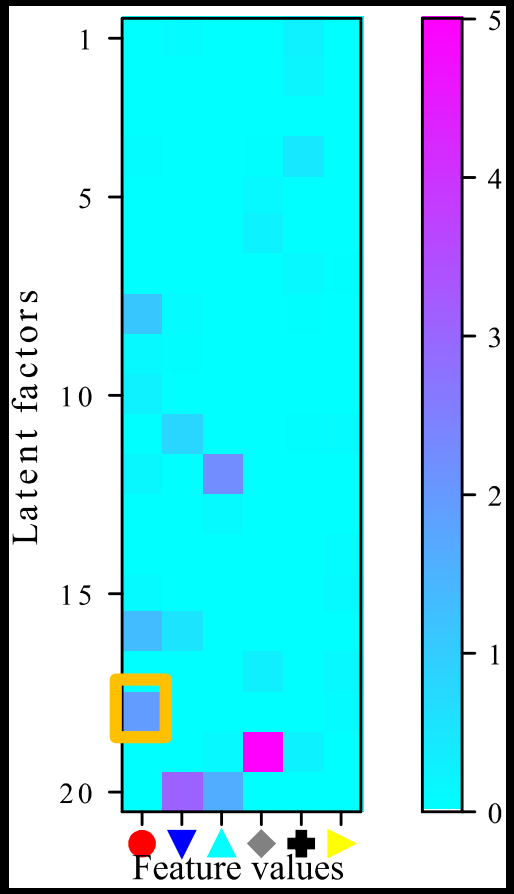


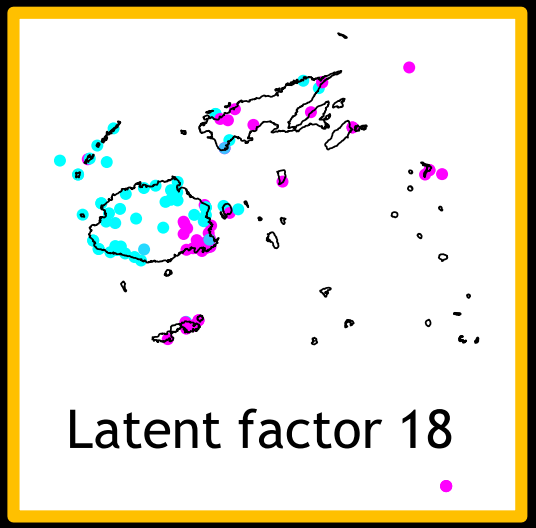


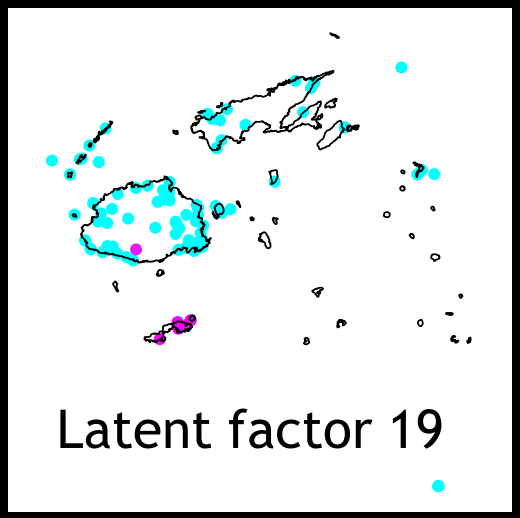


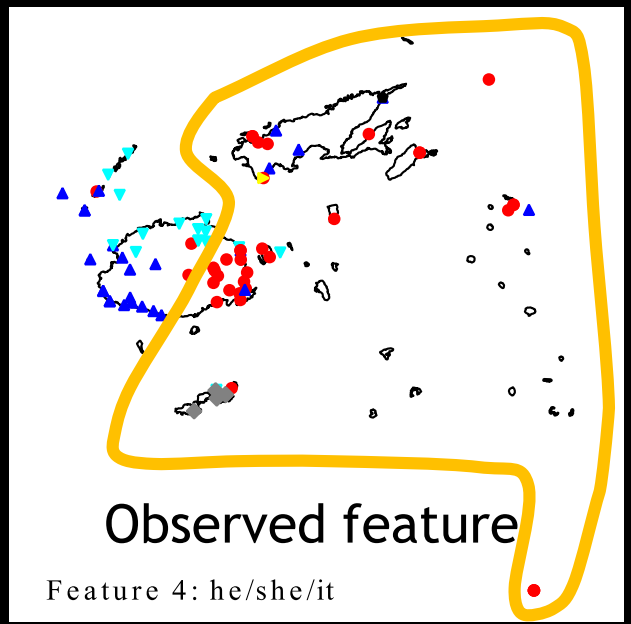
Experiments on Fijian lexical data

Weights



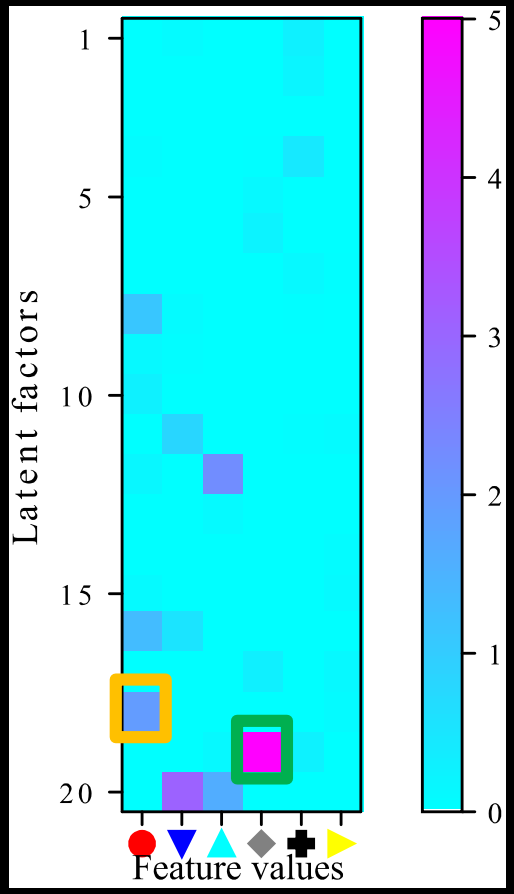


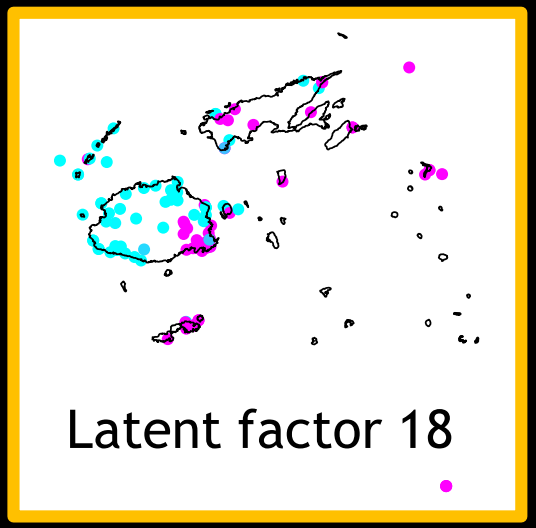


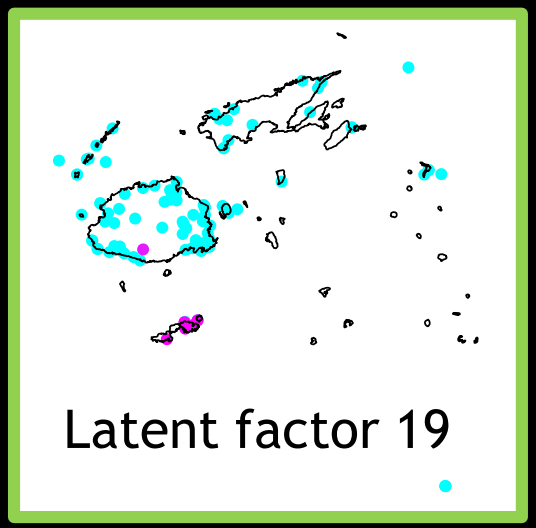


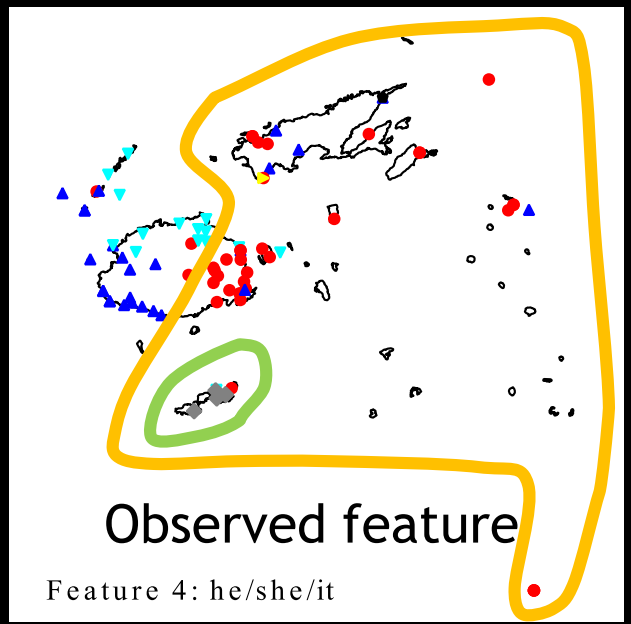
Experiments on Fijian lexical data

Weights









Conclusions and future work

- Proposed latent geographical factors as a means of modeling vertical inheritance and horizontal transfer in a unified way
- Experiments on synthetic and real data show promising results
- In the present model, latent factors are assumed to be independent
- To induce a tree, additional constraints should be imposed on vertical factors