

Reconstructing probabilistic trees of differentiation from single-cell RNA-seq

Miriam Shiffman^{1,2}, Will Stephenson¹, Geoffrey Schiebinger^{1,2}, Jonathan Huggins³, Trevor Campbell⁴, Aviv Regev^{1,2}, Tamara Broderick¹

¹ MIT

² Broad

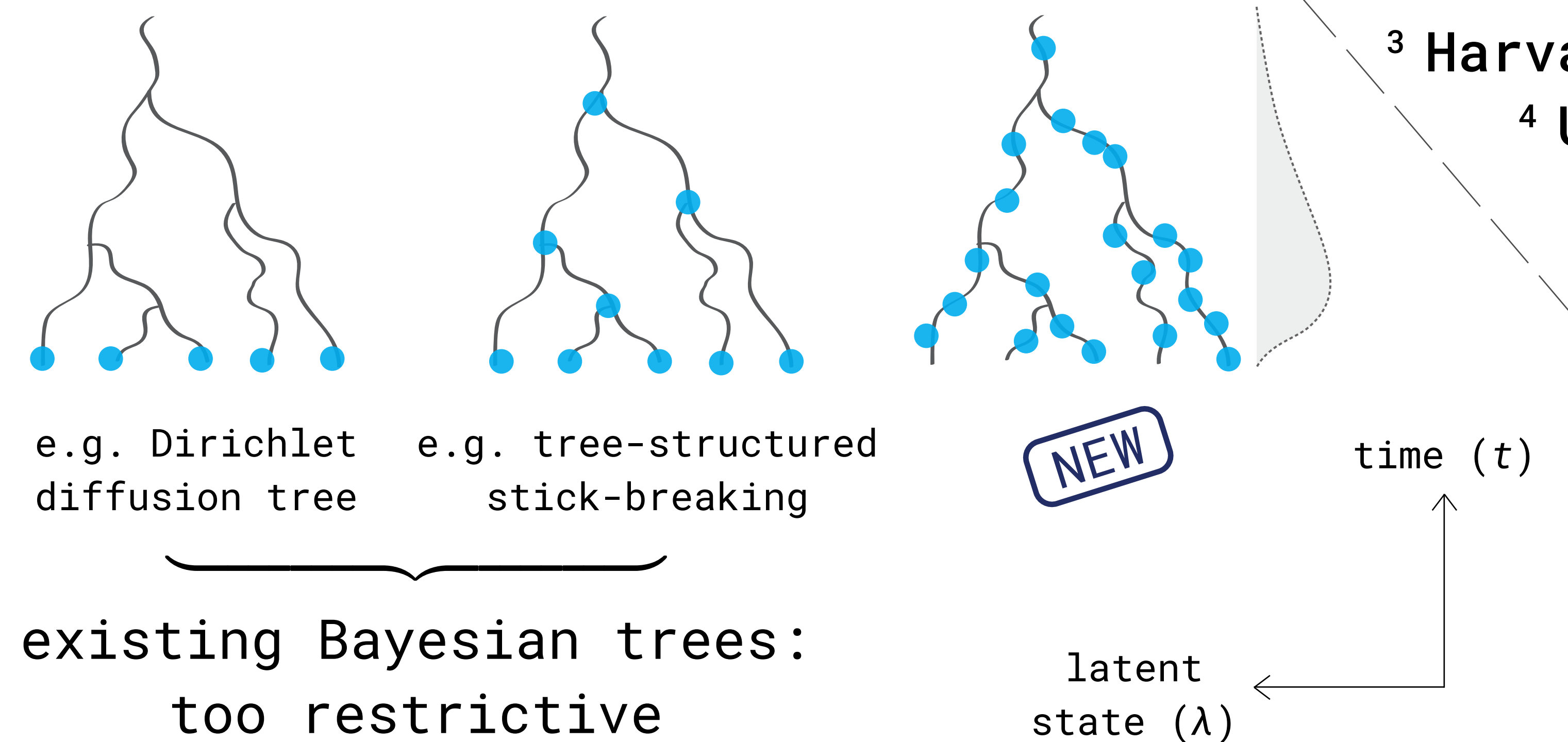
³ Harvard

⁴ UBC

our goal: use noisy snapshots
(single-cell RNA-seq of differentiating cells)
to reconstruct a dynamic process
(latent tree of cellular differentiation)

BUT, more expansive than a canonical “tree”:

- * **continuous**: no discrete (cell) types
- * **probabilistic**: not just one tree... distribution over trees!



the innovation (model + inference): **augmented Dirichlet diffusion tree** (DDT)

- * continuous latent trajectories
- * learn # leaves (cell fates)
- * learn dense latent representations λ (cell state) that parameterize the observation model to generate profiles of sparse counts x (in gene expression space)

$$x \sim \text{Binom} \left(N_{\text{UMI}}, 1 - e^{-q h(\lambda)} \right)$$

max countable transcripts/gene

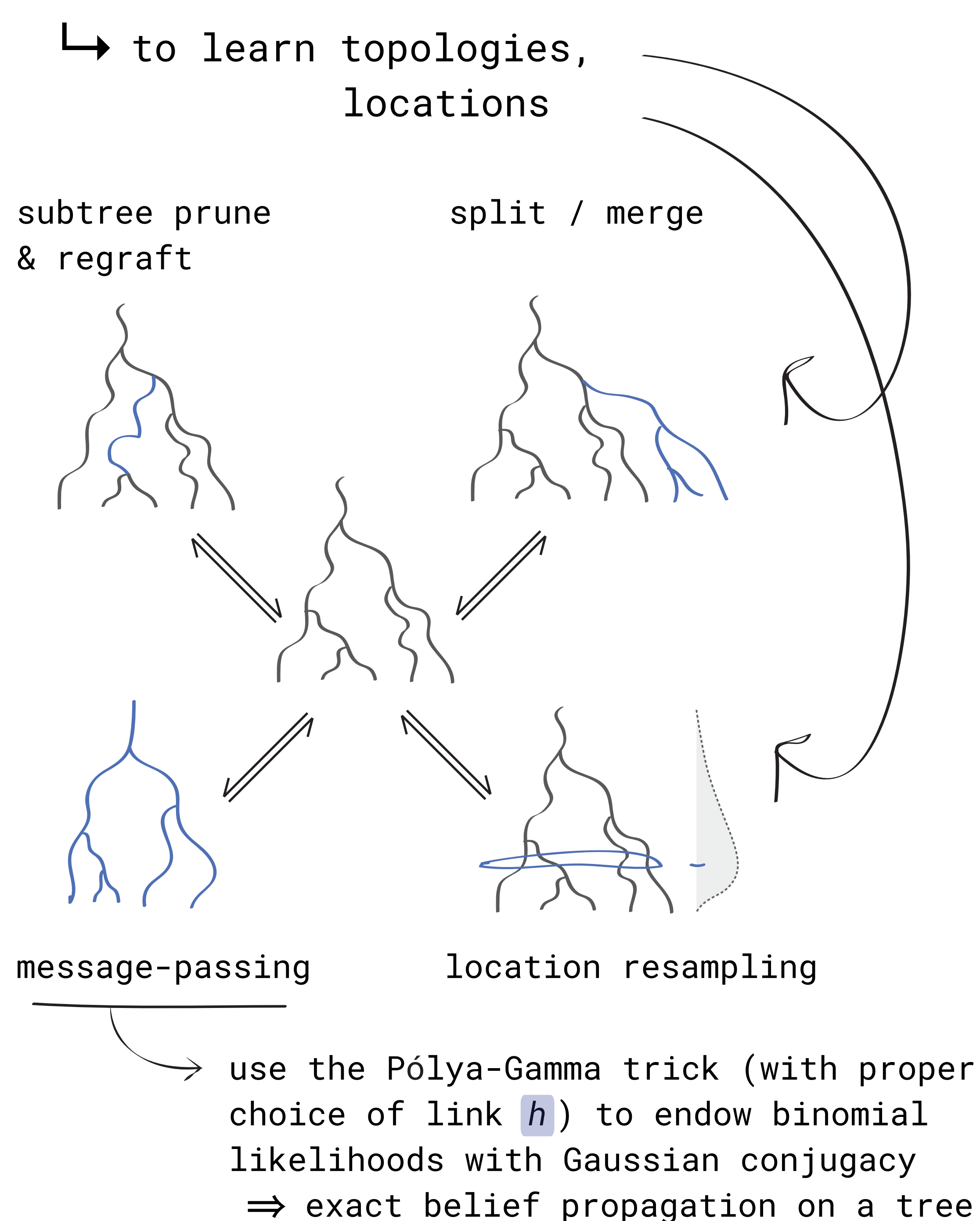
parameter for gene dropout

positive link, $h: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$

from summing over nonzero Poissons

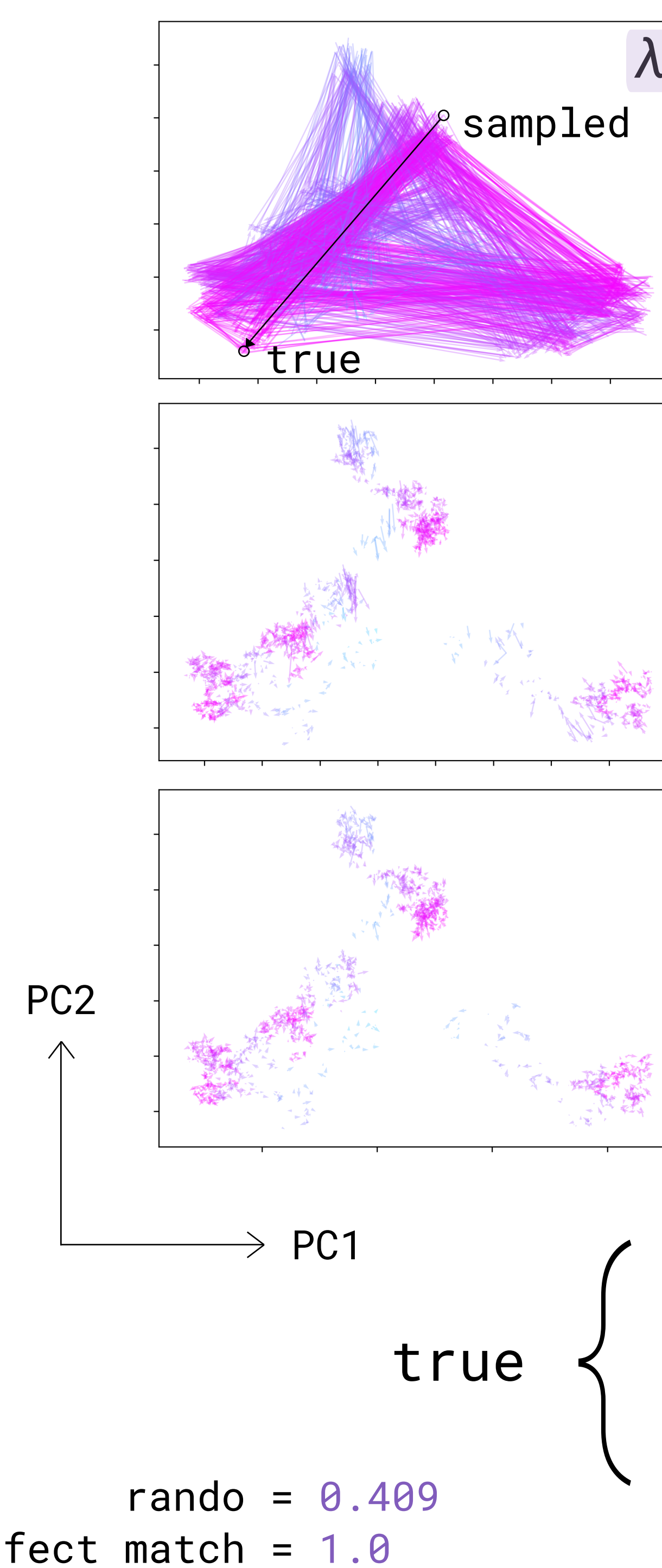
experiment: simulate 2000 cells, 10 genes, 4 leaves fixed topology/times (“time-course experiment”)

inference: novel sampler, interleaving **MCMC** + message-passing



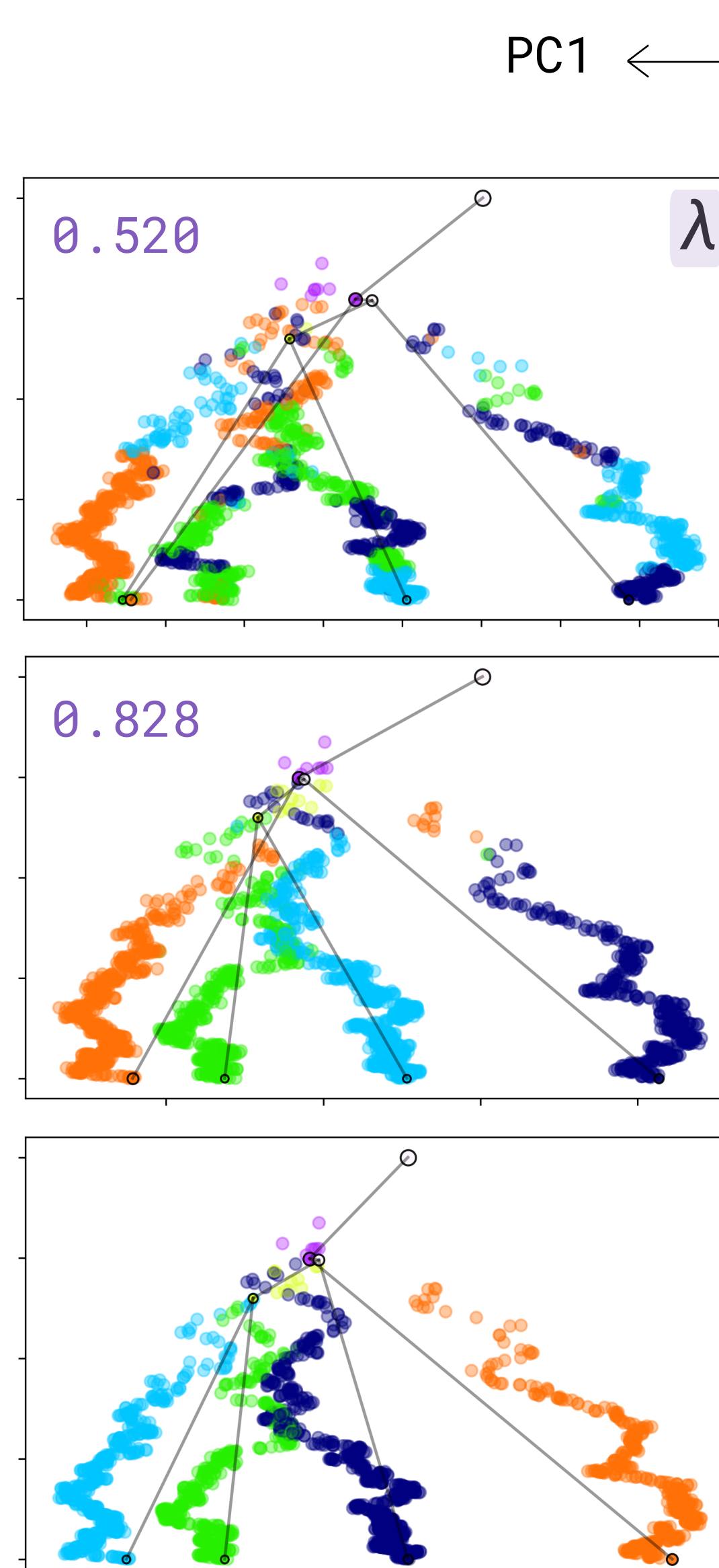
recovery of latent states

(visualized as shifts from ground truth)

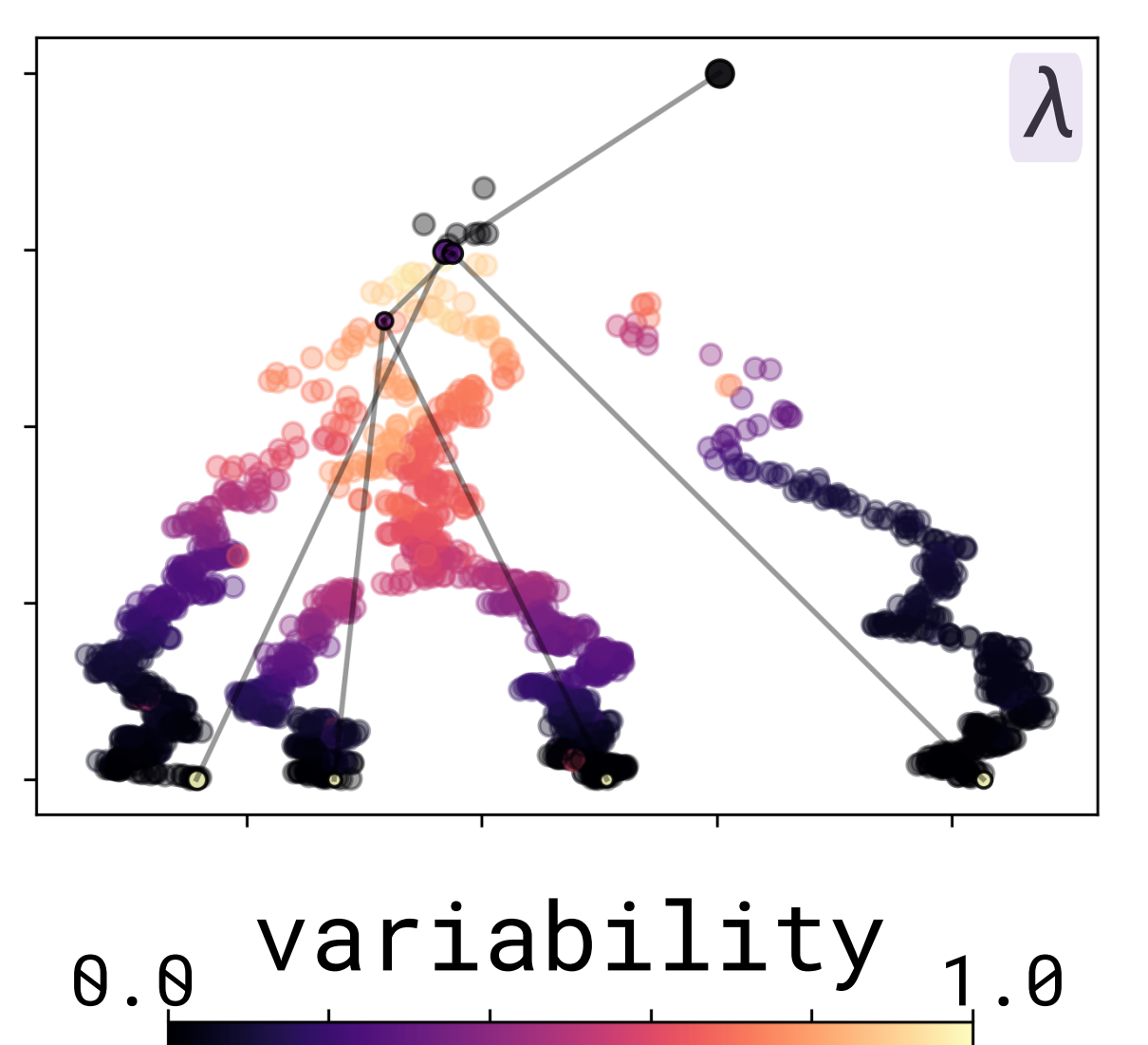


recovery of branch assignments

(quantified by “triplet metric” -- invariant to label-switching)



recovery of expected branch uncertainties



rando = 0.409
perfect match = 1.0