PROBABILISTIC HYPERBOLIC EMBEDDINGS

Simon Lizotte[1], J.-G. Young[2], A. Allard[1,2]

WHY BOTHER?

The $\mathbb{S}^1$ random graph model simultaneously generates [1]

- community structure,
- power law degree distributions,
- sparse connections,
- non-vanishing clustering coefficients.

Algorithms to fit this model exist, but none give error bars on the inferred positions.

BAYESIAN MODEL

The likelihood of the $\mathbb{S}^1$ model is

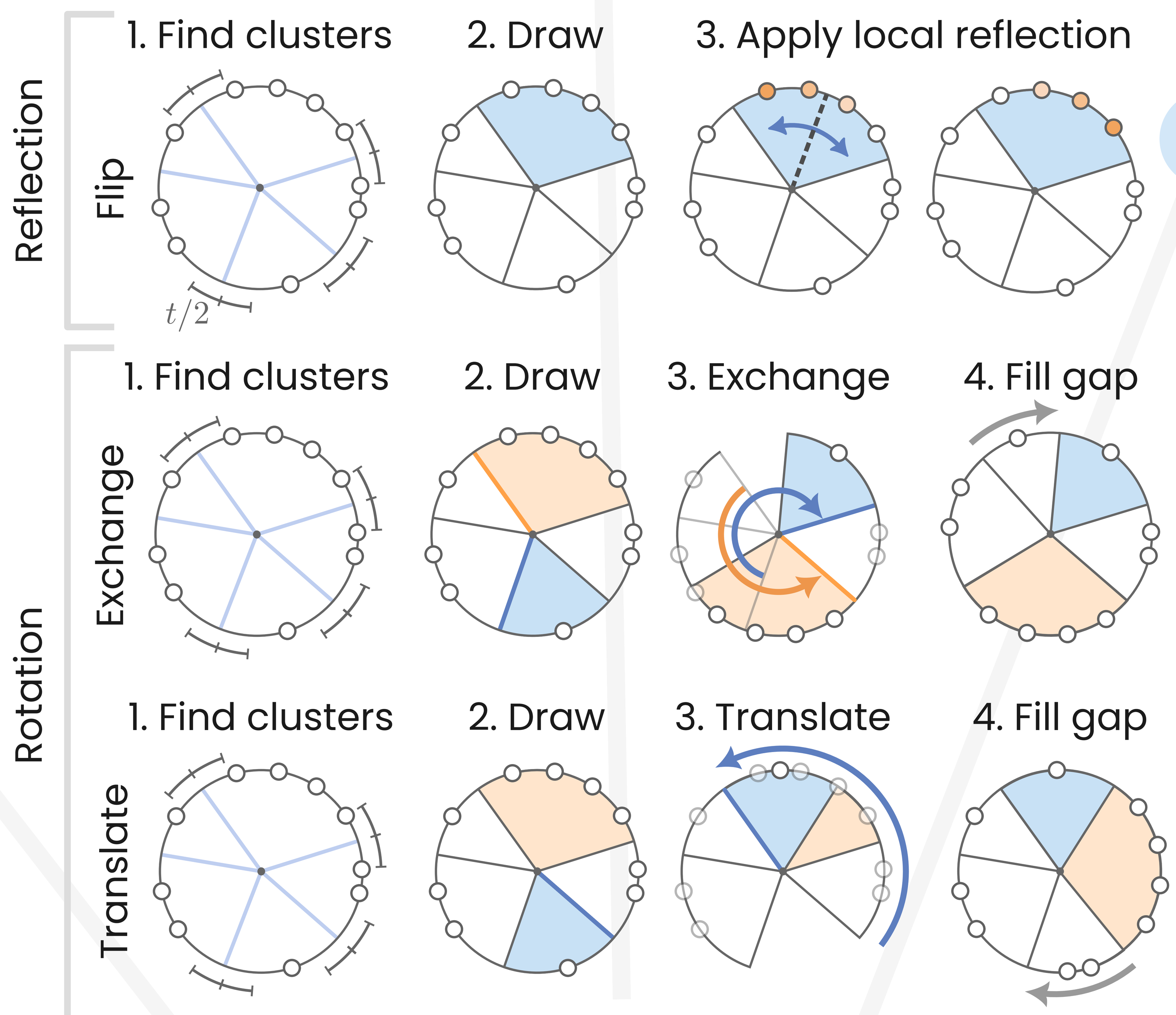
$$\mathbb{P}[\mathcal{G}=G|\theta,\kappa,\beta] = \prod_{u<v}\left(1+\left(\frac{R\Delta(\theta_u,\theta_v)}{\mu\kappa_u\kappa_v}\right)^{\beta^{\pm}_{uv}}\right)^{-1}.$$

Parameters:

- θ_u: angular coordinate of vertex u
- κ_u: expected degree of vertex u
- $\beta^{\pm}_{uv}$: β if u and v are connected and is $-\beta$ otherwise
- $\Delta(\theta_u,\theta_v)=\pi-|\pi-|\theta_u-\theta_v||$: angular separation
- μ: constant set such that the expected degree u is κ_u
- $R=|V|/2\pi$: circle radius.

The likelihood is invariant to graph automorphisms, rotations and reflexions ($\theta \leftrightarrow -\theta$).

We use a uniform prior on each θ_u, a half-Cauchy prior on each κ_u and a normal prior on β.

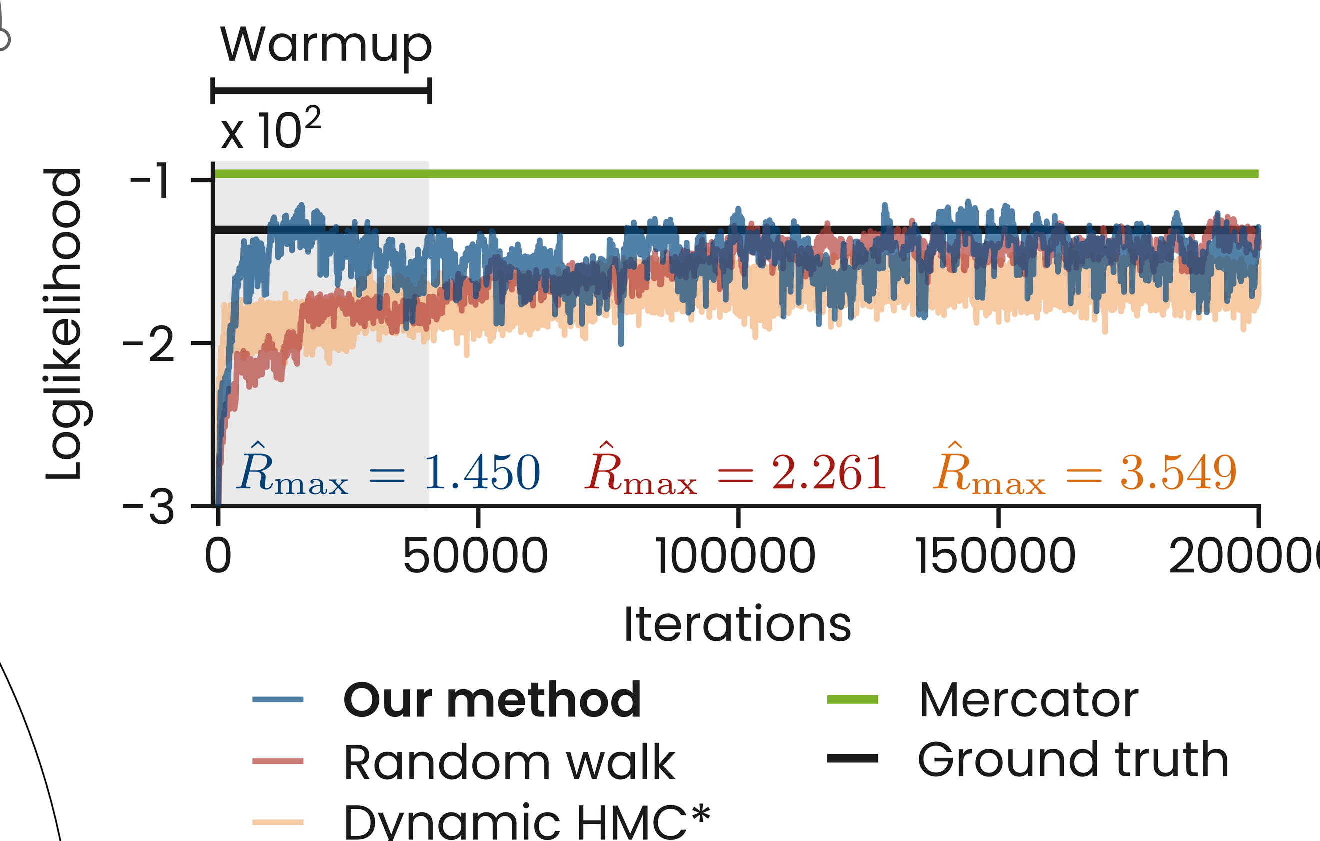


MARKOV CHAIN MONTE CARLO

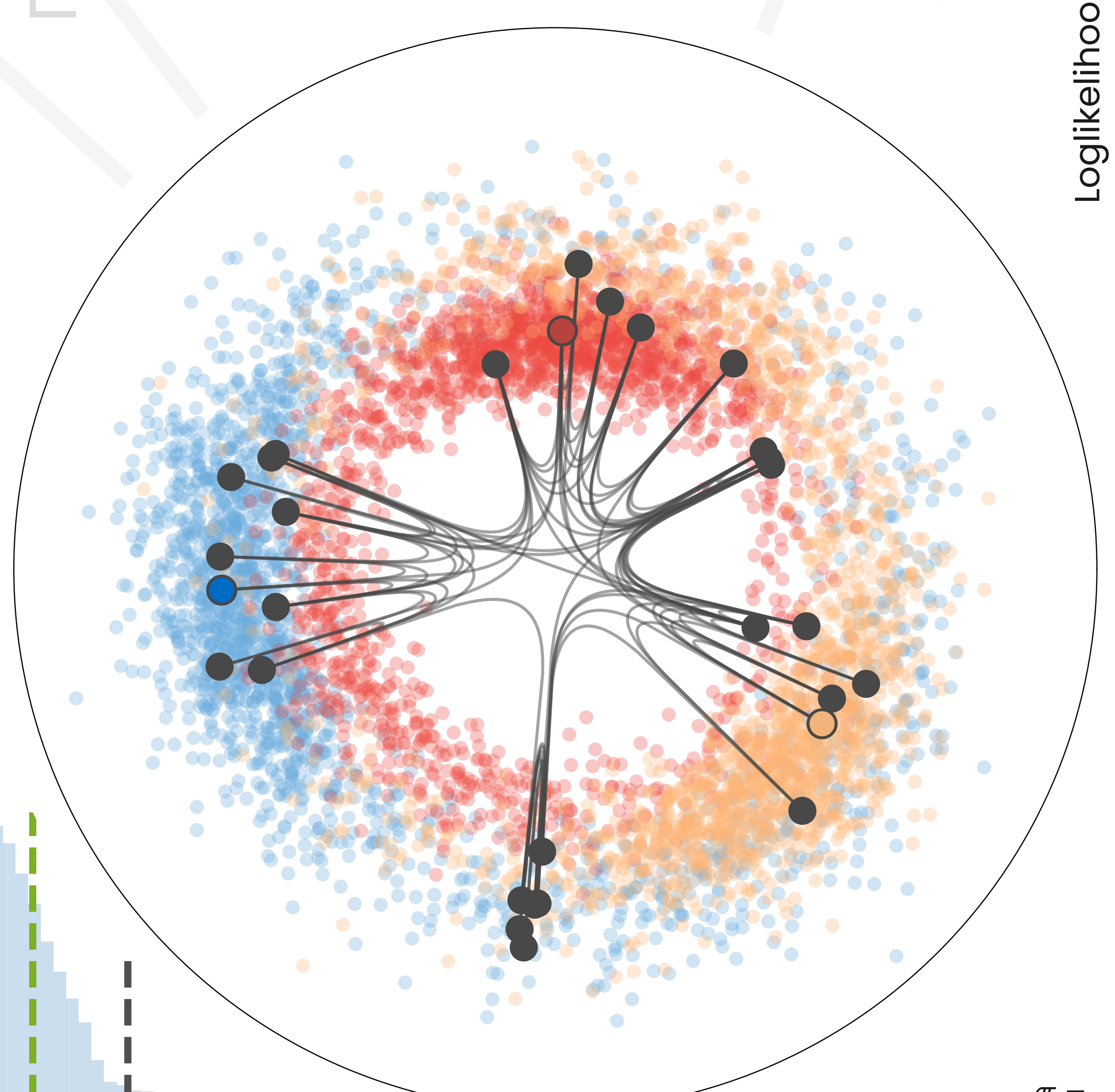
Distant groups of vertices have a weak influence on each other because the connection probability decreases rapidly with the distance. We apply transformations that preserve intra-cluster distances.

Our method is a Metropolis-Hastings algorithm that uses both random walk and cluster-based moves (figure on the left).

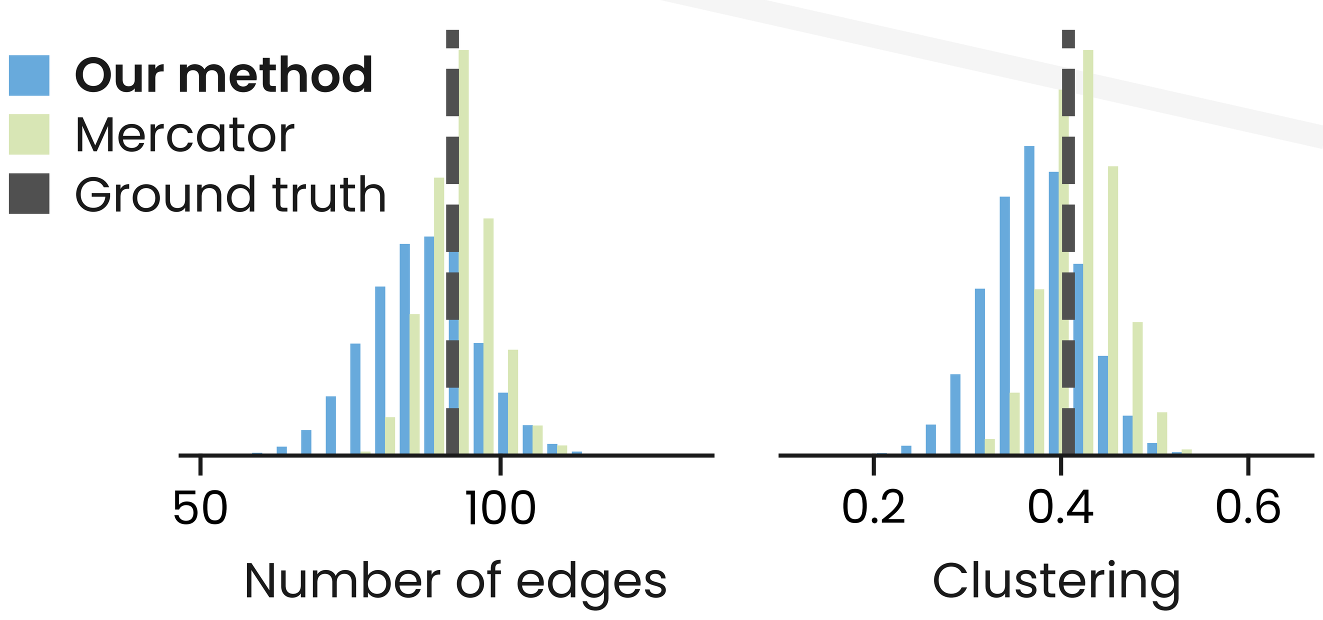
Our method's mixing is significantly better than random walk and HMC based algorithms.

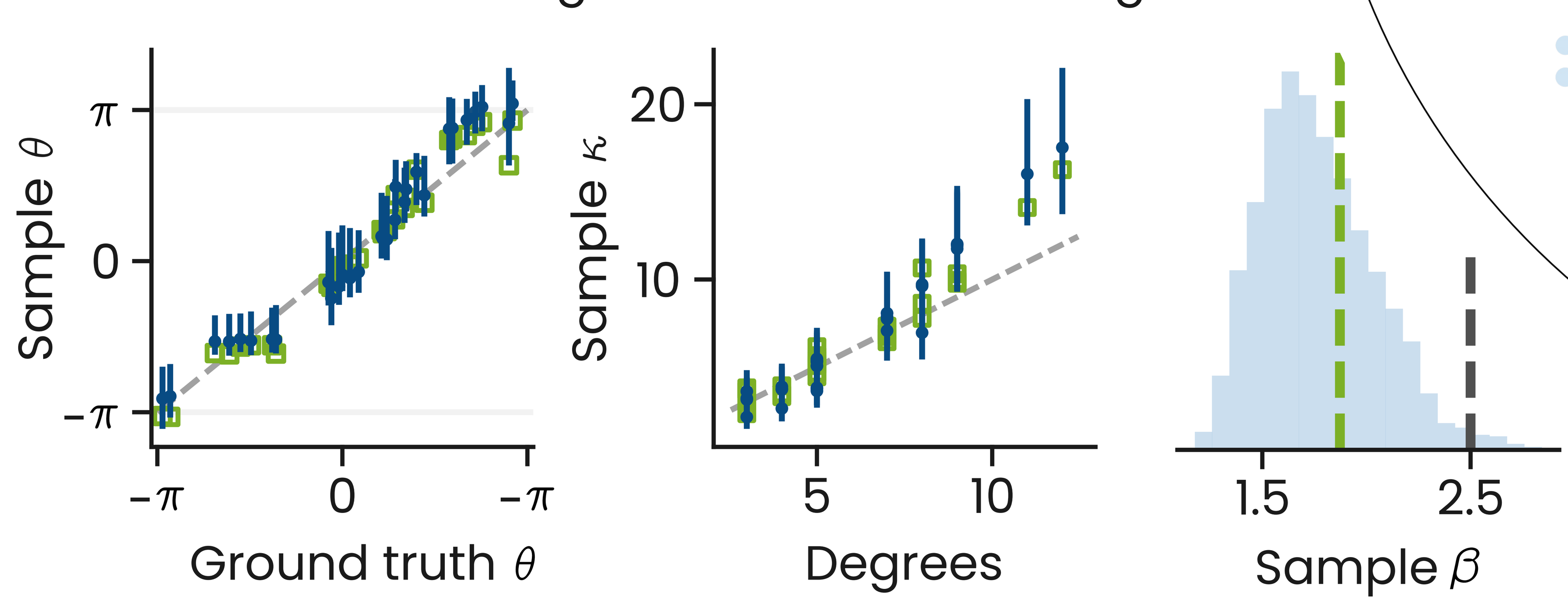


* With a smooth approximation of the likelihood.



SAMPLING IS NOT OPTIMIZATION

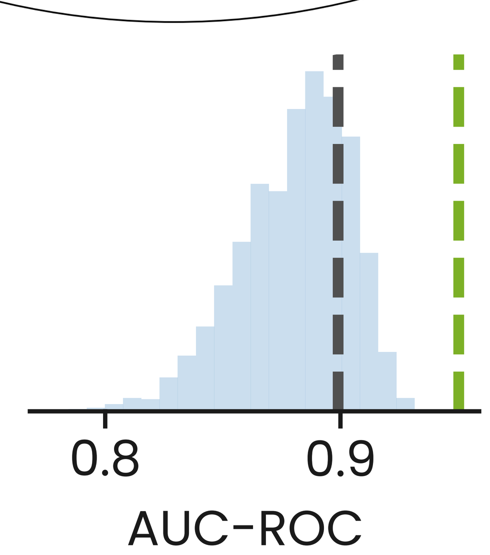




Samples from the posterior give many plausible embeddings of the same graph (here, a synthetic graph — the ground truth). Those provide error bars for every parameter.

Properties of graphs generated from the posterior sometimes differ from those of graphs generated from Mercator's maximum likelihood estimation [2].

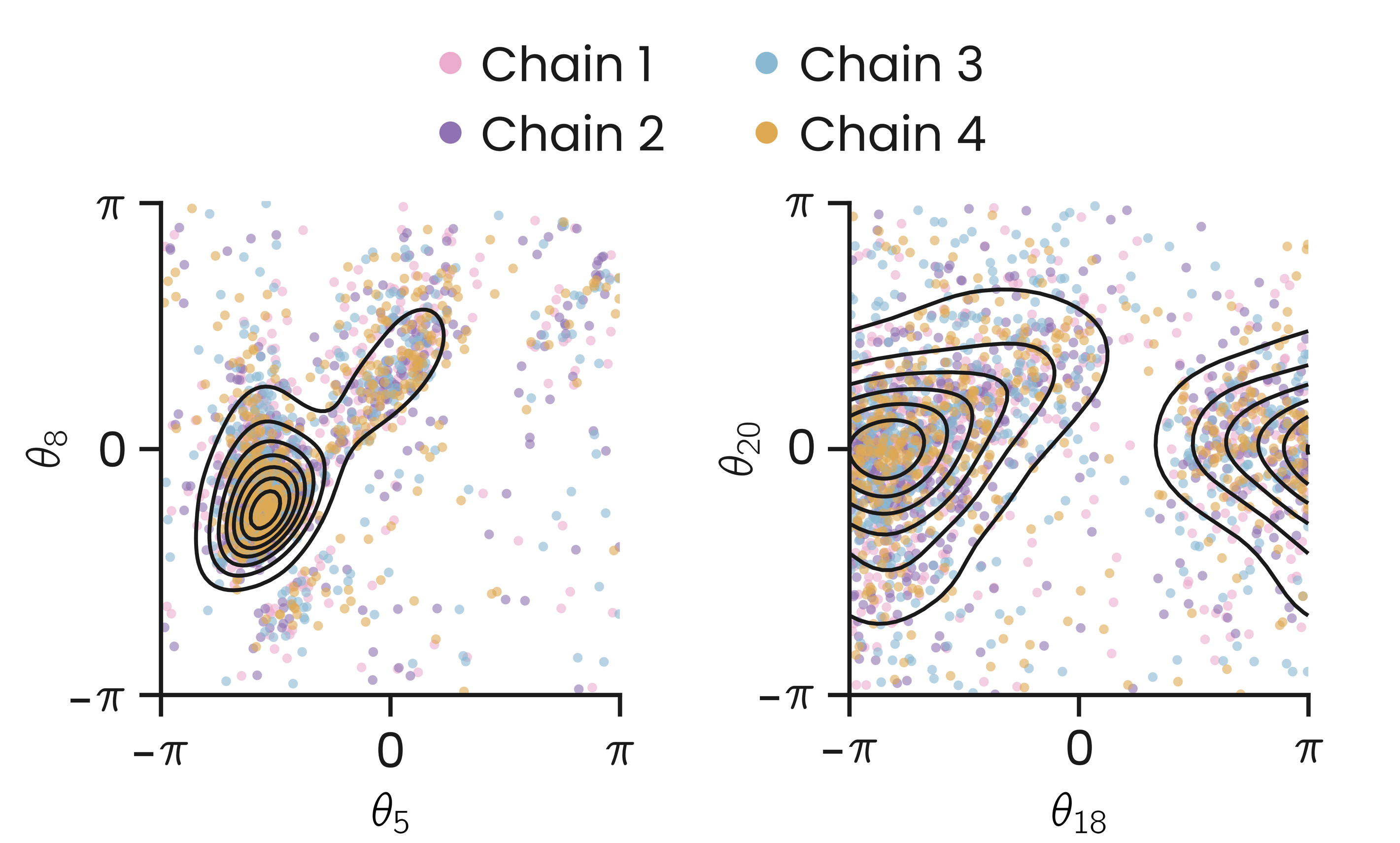
The figure in the center shows the sampled positions for three vertices (colored) of the synthetic graph and the median for the others (black). Black lines are the geodesics of the edges.



Agreement between edge posterior probabilities and their existence.

The posterior is not a multivariate normal both for synthetic graphs (below, left panel) and for Zachary's karate club (below, right panel).

A normal approximation of the posterior is not appropriate.



arXiv coming very soon!

[1] Nat. Rev. Phys. 3, 114–135 (2021)
[2] New J. Phys. 21, 123033 (2019)

1. Université Laval
2. University of Vermont

Dynamica · Sentinel North · Fonds de recherche Nature et technologies Québec · Université Laval · Digital Research Alliance of Canada · Alliance de recherche numérique du Canada