Symmetry-driven embedding of networks in hyperbolic space

Simon Lizotte[1,2], Jean-Gabriel Young[1,3,4], Antoine Allard[1,2,4]

1. Département de physique, de génie physique et d'optique, Université Laval, Québec (Québec), Canada G1V 0A6
2. Centre interdisciplinaire en modélisation mathématique, Université Laval, Québec (Québec), Canada G1V 0A6
3. Department of Mathematics and Statistics, University of Vermont, Burlington, VT 05405, USA
4. Vermont Complex Systems Center, University of Vermont, Burlington, VT 05405, USA

Network geometry is a versatile yet simple framework that captures several observed properties of empirical networks, such as non-vanishing clustering, sparsity, and power-law degree distribution [1]. This accurate description is achieved by positioning vertices in a metric space (usually hyperbolic) and connecting them according to their proximity. To use this framework inferentially, one must find vertex coordinates that best reproduce the observed topology. Unfortunately, the problem is non-convex, which makes the inference of the coordinates of the vertices a challenging task. Common embedding techniques circumvent this difficulty by using a combination of simplifying heuristics, greedy algorithms, and machine learning techniques [2–5]. These approaches yield a pointwise estimate, which ignores the estimation's uncertainty and the possibility of having many adequate embeddings of the same graph.

We address both issues simultaneously using a Bayesian approach [6] in which the posterior distribution indicates the plausibility of each embedding (see panel a). By sampling the posterior, we can estimate error bars for the coordinates, but also for the graphs and geometric properties that ensue from the embedding. Correctly sampling the posterior is a technically challenging problem because of the large number of local maxima. Indeed, we find that naive Hamiltonian Monte Carlo—the state-of-the-art method for sampling continuous random variables [7]—fails to adequately explore the parameter space. We side-step this issue using cluster-based transformations that leverage natural symmetries of the hyperbolic space (see panel b). The computing cost of sampling is the main drawback of the approach, but it could be mitigated using parallelization and by adapting Hamiltonian Monte Carlo to the embedding geometry. The novelty and effectiveness of this cluster-based approach opens the door to further improvements to the current point-wise methods.

As seen in panels c–h, both approaches are compatible as the coordinates' median is close to the point-wise estimation and the methods perform similarly in a link prediction task. However, we find that the predicted graph transitivity and graph density differ and that the navigability, as measured by the success rate of network greedy routing, is on average lower in the sample than in the maximum likelihood point-wise estimation. Furthermore, we discuss a method to synthetically generate graphs that have two plausible embeddings (two modes of the posterior distribution), a feature that cannot reliably be identified by a point-wise estimator.

[1] M. Boguñá et al., "Network geometry", Nat. Rev. Phys. **3**, 114–135 (2021).

[2] G. García-Pérez et al., "Mercator: uncovering faithful hyperbolic embeddings of complex networks", New J. Phys. **21**, 123033 (2019).

[3] Y.-J. Zhang et al., "Systematic comparison of graph embedding methods in practical tasks", Phys. Rev. E **104**, 044315 (2021).

[4] A. Muscoloni et al., "Machine learning meets complex networks via coalescent embedding in the hyperbolic space", Nat. Commun. **8**, 1615 (2017).

[5] T. Bläsius et al., "Efficient Embedding of Scale-Free Graphs in the Hyperbolic Plane", IEEE/ACM Trans. Netw. **26**, 920–933 (2018).

[6] S. Lizotte et al., "Symmetry-driven embedding of networks in hyperbolic space", arXiv:2406.10711 (2024).

[7] B. Carpenter et al., "*Stan* : A Probabilistic Programming Language", J. Stat. Softw. **76**, 1–32 (2017).

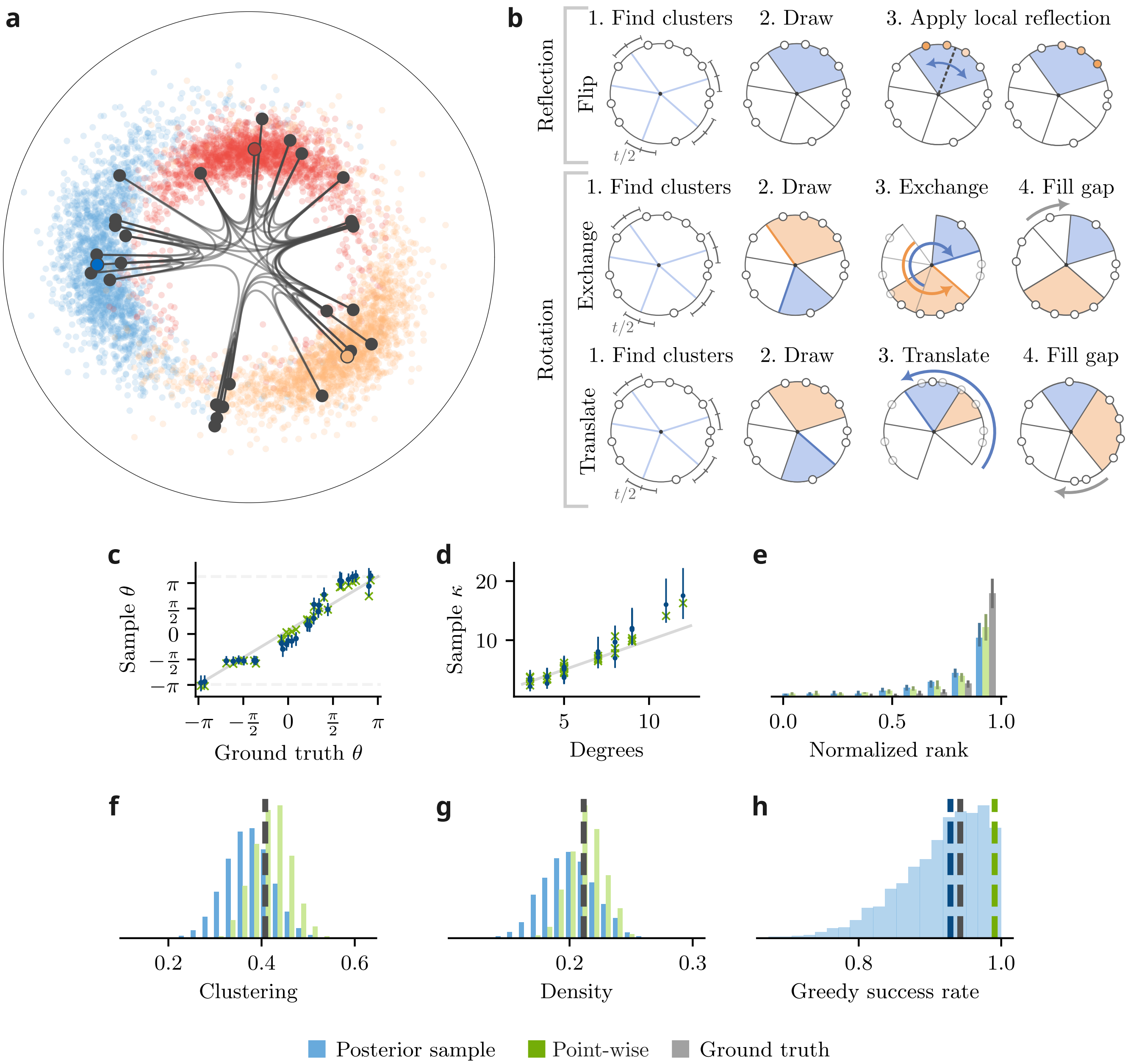

Figure 1: **(a)** Probabilistic hyperbolic embedding of a synthetic graph of 30 vertices in the hyperbolic plane. Black points and dark-colored points are the median coordinates of each vertex. Light-colored points are the positions sampled for the three highlighted vertices. Lines are edges drawn using hyperbolic geodesics. The geodesic polar coordinates—also called native coordinates—are used. The sample contains 2,000 points. **(b)** Cluster-based transformations used in the sampling algorithm. The *flip* transformation targets the reflection symmetry while the *exchange* and *translate* target the rotation symmetries. The radial coordinate of the vertices is not shown in the schematics. **(c-d)** Estimated coordinates, θ and κ, for the network of (a). **(e)** Pairs of embedding and graphs of 30 vertices were sampled from the prior. For each graph, 5% of the edges were then randomly removed and then ranked based on their existence probability among all the absent edges. The normalized ranks approach 1 if the model accurately predicts the missing edges. **(f-h)** Global clustering (transitivity), density and success rate of greedy routing of graphs generated from the inferred embedding of the synthetic network of (a). In the greedy routing algorithm, hyperbolic coordinates act as addresses, and one attempts to reach the target v from a source u by repeatedly following the edge that leads to the neighbor closest to the destination u. Paths in this procedure can sometimes devolve into "greedy loops" that lead nowhere. The success rate is the proportion of paths that reach their destination.