Bioinformatics Advances, 2024, **00**, vbae022
https://doi.org/10.1093/bioadv/vbae022
Advance Access Publication Date: 19 February 2024
Original Article

From Solis-Lemus' talk

OXFORD

Phylogenetics

Novel symmetry-preserving neural network model for phylogenetic inference

Xudong Tang [1,2,†], Leonardo Zepeda-Nuñez[3,†], Shengwen Yang [1,2], Zelin Zhao[3], Claudia Solís-Lemus [1,4,*]

[1]Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI 53706, United States
[2]Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, United States
[3]Department of Mathematics, University of Wisconsin-Madison, Madison, WI 53706, United States
[4]Department of Plant Pathology, University of Wisconsin-Madison, Madison, WI 53706, United States

*Corresponding author. Department of Plant Pathology, University of Wisconsin-Madison, 333 N Orchard St, Madison, WI 53706, United States. E-mail: solislemus@wisc.edu

[†]Equal contribution

Associate Editor: Aida Ouangraoua

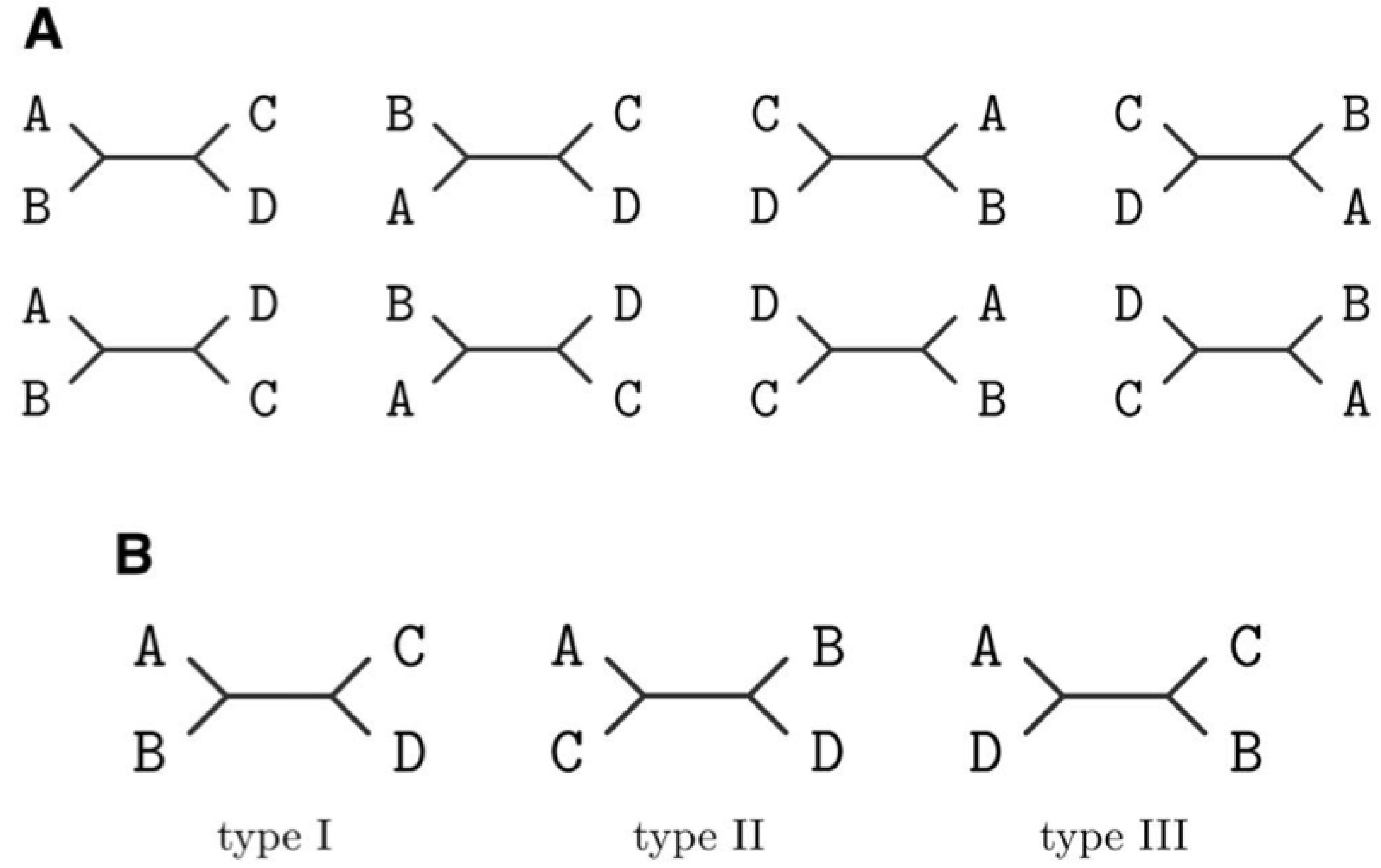

Figure 1. Top: Possible unrooted trees for four species (quartets). Bottom: All the possible permutations on the order of the leaves that are isomorphic to a tree of type I.

G-invariance

Consider a collection of action (mapping) g_i

Given some representation of data x, $f(g_i * x) = f(x)$ for all actions.

Then the representation of data f(x) is G-invariant.

PROJECT 11. Topological Generative Model for Trees

Goal. Build a **generative model** for tree-structured data by representing each tree in a spectral coordinate system induced by the 1-Hodge Laplacian, and use mode-wise permutation of spectral coefficients as a principled null model for resampling and inference. The model will enable core statistical operations on tree populations, including defining an average tree representation, quantifying variability, and testing group differences, while holding the underlying tree topology fixed (or otherwise controlled) and modeling randomness through distributions and permutations on spectral coefficients.

Description. This project introduces a **generative modeling** framework for tree-structured data based on spectral representations of the 1-Hodge Laplacian. Instead of modeling variability through local edge-level modifications, the approach represents each tree in a global spectral coordinate system and introduces randomness through controlled transformations of spectral coefficients. This perspective enables principled simulation, resampling, and statistical inference on trees while preserving their underlying structure and topological constraints.

Learning Outcome. Students will learn how to represent trees in a spectral form, how this representation captures structure and variability, and why trees differ fundamentally from graphs with cycles. By the end of the project, students will be able to build a principled **generative model for tree data**, compute average structure and variability, and perform basic statistical inference using permutation-based resampling.