

OLA encoding of binary trees

Harry Richman, joint with Erick Matsen, Cheng Zhang

Fred Hutch Cancer Center

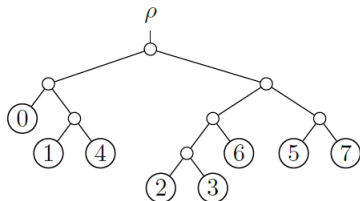


ICERM: Current Methods and Open Problems in Phylogenetics
17 September 2024



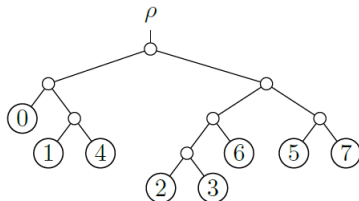
How to tell your computer about phylogenetic trees?

- Newick format: ID's and parentheses



↔ ((0,(1,4)),(((2,3),6),(5,7)));

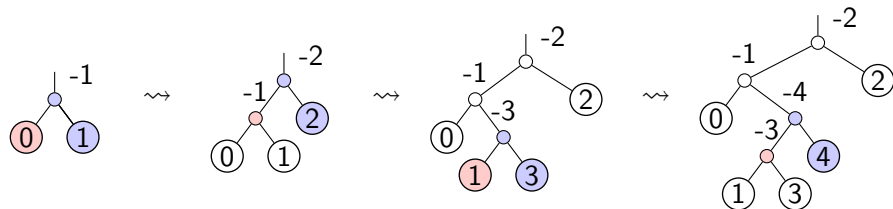
-  OLA encoding: list of integers



↔ (0,-1,2,1,-3,-3,5)

OLA = "Ordered Leaf Attachment"

Example: tree with encoding $(0, -1, 1, -3)$



Pros:

- linear-time encoding and decoding
- convex region in encoding space
- somewhat-low distortion vs. SPR distance