

6.047/6.878/HSPH IMI.231/HST.507 Fall 2015

Problem Set 5: Clustering Phylogenetic Trees

Due Thursday, December 3 at 8pm (submit on the course website)

Submit a zip file of a directory named Lastname.Firstname containing:

- A PDF file named Lastname.Firstname.pdf with your written answers, which should include all plots you are referencing.
- A directory named code with all the code you are submitting

In your answers to the questions please refer to the appropriate file name where your code for that problem is located. Unless skeleton code has been provided, feel free to use any programming language you are comfortable with, as long as you structure and comment your code to make it concise and legible.

We've seen that phylogenetic tree algorithms can construct many different "good" putative evolutionary histories. An important and challenging problem is that of reducing a large number of trees to a smaller number of representative solutions. The objective of this assignment is to explore techniques for dealing with large sets of different phylogenetic trees for the same data. This assignment is inspired by the paper "Statistically based postprocessing of phylogenetic analysis by clustering" by Cara Stockham, Li-San Wang, and Tandy Warnow (*Bioinformatics*, Vol 18, Suppl. 1, 2002, pp. S285-S293).

In Part 1, you will implement the Robinson–Foulds distance metric. In Part 2, you will use a clustering algorithm to partition distinct phylogenetic trees into clusters of trees where the trees in a given cluster are similar with respect to Robinson–Foulds distance. In Part 3, you'll implement a general consensus algorithm that will allow you to find a consensus tree for each cluster and measure the quality of the cluster by its "specificity" – how close it is to being a binary tree.

This problem is due to Ran Libeskind-Hadas. The full details are available at the URL below:

<http://www.cs.hmc.edu/~hadas/mitcompbio/treedistance.html>

MIT OpenCourseWare
<http://ocw.mit.edu>

6.047 / 6.878 / HST.507 Computational Biology
Fall 2015

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.