

Do *all* the problems. Please typeset your solutions.

1. Ultrametric trees

(a) Read Gusfield's notes on $O(n^2)$ algorithm on ultrametric tree algorithm. Prove the claim that for an edge (p, q) with maximal weight on L , a pair of nodes i and j that spans (p, q) must have same weight as (p, q) .

(b) Prove another claim in that notes that L is a minimum spanning tree of G .

(c) We consider a *partially* fill-in matrix M . That is, some entries, say $M[i, j] = *$ (unspecified, or missing). Our goal is to see whether we can fill in the missing values so that the filled-in matrix M' is ultrametric. We consider a weighted graph $G(M)$, where there is a node corresponding to each row for a given matrix M , and the edge weight between nodes i, j is equal to $M[i, j]$ (where $M[i, j] \neq *$, so $G(M)$ is not a complete graph). Prove that M is *not* ultrametric if for each cycle in $G(M)$, the maximal value appears just once on the cycle.

2. Neighbour joining

Read the proof in the Bryant paper (on the class web site) that if δ is additive then the NJ algorithm does correctly pick neighbors x, y in the (unique) underlying additive tree T , as its first step.

(1) Prove the claim on page 2 (Figure 1) in Bryant's paper that minimizing Q is the same as maximizing the length of branch e .

(2) Work out the details of the difference between $r_d(u)$ and $r_d(v)$ on page 6 in the proof of Theorem 1.

(3) As mentioned in class, the proof (assuming it is correct) that when δ is additive, then the chosen (x, y) are neighbors in T , does not seem sufficient to prove that the NJ algorithm constructs the entire tree T . There are two issues: how δ is changed after the selection of (x, y) ; and the edge weights put on the edges touching x and y . Prove that these two steps are correct so that the ultimate tree created by the NJ algorithm is T .

3. Binary perfect phylogeny and the splits-equivalence theorem

(1) Briefly show that that the *no* pair of columns in the SP matrix (generated from the splits of a tree T) is incompatible.

(2) In class, I showed that when the root sequence is not known, we can simply pick one sequence in the input binary matrix as the root. Now here is another way of choosing the root sequence. A sequence is called majority sequence if at every column, the sequence takes on a value (0 or 1) depending on which (of 0 and 1) appears more frequently in the input matrix. For example, if there are 5 zeros and 3 ones in the first column of the input matrix, then the majority sequence has value 0 at the first column. When there is a tie on the number of zeros and ones, we can pick either 0 or 1. Now prove that if there is a perfect phylogeny for matrix M with *some* root, then there is a perfect phylogeny with the majority sequence as the root.

4. Multi-state perfect phylogeny

Read the notes on multi-states perfect phylogeny, and answer the questions stated there.