Introduction to Computational Biology
Homework 4

Problem 1: Perfect phylogeny
Consider binary characters and let 1_i be the set of objects that have state 1 for character i. Define 0_i similarly.

(a) Consider the following character state matrix:

	c_1	c_2
A	0	1
B	1	1
C	1	0

According to the condition stated in class, namely that 1_i and 1_j are either disjoint or one is a subset of the other, a perfect phylogeny does not exist. However, the colored graph produced by the state matrix is acyclic, implying that a perfect phylogeny does exist. Which interpretation is correct?

(b) Show that there exists a perfect phylogeny for a collection of objects on two binary undirected characters i and j if and only if 1_i, 1_j or 0_i, 0_j satisfy the above condition. This is why the problem can be solved by directing the character in such a way that the ancestral state is larger.

(c) Consider ordered undirected (the state tree is unrooted) non-binary characters. We have argued in class that the problem of finding a perfect phylogeny is equivalent to finding one for a set of directed binary factors. Show that for each character i, the state tree can be rooted in such a way that for each binary factor j, $|0_j| \geq |1_j|$ (this means that the problem can be solved as an instance of directed binary characters). *Hint*: root the tree arbitrarily, then reverse edges that violate the condition.

Problem 2: Parsimony
The following dynamic programming algorithm solves the maximum parsimony problem for a general distance criterion and one character.

$$f_v(a) = 0, v \text{ is a leaf labeled } a$$

$$f_v(a) = \infty, v \text{ is a leaf not labeled } a$$

$$f_v(a) = \sum_{w \in \delta(v)} \min_{b \in A} [f_w(b) + d(a,b)]$$

$$M_w(a) = \{b : f_w(b) + d(a,b) \text{ is minimal}\}$$

We seek $\min_a \in A f_{root}(a)$ and M can be used for backtracking.

(a) What is the running time and space requirement for this algorithm? Assume $|A| = r$ (r states), we have n leaves (objects), and m characters.

(b) Adapt this algorithm to the special case when the distance is defined as:

$$d(a,b) = \begin{cases} 0 & a = b \\ 1 & a \neq b \end{cases}$$