1. Find the compatibility graph and derive the compatibility tree for the species with the following character data set.

<table>
<thead>
<tr>
<th>Species</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platypus</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Elephant</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tiger</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Horse</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Guinea Pig</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cat</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

2. Note that the compatibility tree you provided in the previous question was not the same as the most-parsimonious tree for the same set of species. Show that, however, with 5 or fewer species and 0/1 data with unknown ancestral states, the parsimony and compatibility trees will always be the same.

3. Prove that a tree-derived distance satisfies the following 4 properties. Let $S$ be a set of points. For all points $x$ and $y$ in $S$:
   
   (a) $d(x, y) \geq 0$
   
   (b) $d(x, y) = 0$ if and only if $x = y$
   
   (c) $d(x, y) = d(y, x)$
   
   (d) for all $x, y, and z$ in $S$, $d(x, y) \leq d(x, z) + d(z, y)$

4. Give a method for computing the trimming parameter $\delta$ from the additive phylogeny algorithm presented in class.

5. Prove the following statement. If an $n \times n$ distance matrix is ultrametric, then it is additive.

6. Consider a character $\chi$ on a set of species. We say $\chi$ is trivial if there is at most one state of the character that is assigned to two or more species. Otherwise, $\chi$ is non-trivial.

   (a) How many non-trivial binary characters are there on a set of size $n$?
   
   (b) Let $X = \{A, B, C, D, E\}$. Show that the compatibility graph of non-trivial binary characters on $X$ is isomorphic to the Petersen graph given below.
7. For five species \(a, b, c, d, \) and \(e\) with distances given by

\[
\begin{array}{c|ccccc}
 & a & b & c & d & e \\
\hline
a & 0 & 9 & 8 & 7 & 8 \\
b & 0 & 3 & 6 & 7 & \\
c & 0 & 5 & 6 & & \\
d & 0 & 3 & & & \\
e & 0 & & & & \\
\end{array}
\]

reconstruct the tree using the neighbor joining algorithm and the UPGMA algorithm. Compare your answers.

8. Suppose we have two nucleotide sequences:

\[
\begin{align*}
\text{CCGGCCGCGCG} \\
\text{CGGGCCGCGCG}
\end{align*}
\]

Using the Jukes-Cantor substitution probabilities \(r_t = \frac{1}{4}(1+3e^{-4at})\) is the probability that a character does not change in time \(t\), and \(s_t = \frac{1}{4}(1 - e^{-4at})\) is the probability of a change to any other character in time \(t\), show that the maximum likelihood solution is given by

\[
t_1 + t_2 = -\frac{3}{4} \ln \frac{3n_1 - n_2}{3n_1 + 3n_2},
\]

where \(t_1\) and \(t_2\) are the maximum likelihood edge lengths, \(n_1\) is the number of sites where the residues in the two sequences are identical and \(n_2\) is the number of sites where a substitution occurs. (Recall that for two sequences, there is only one possible tree, namely the one with two branches and a root node which represents the hypothetical common ancestor.)

9. Find the split distance and nearest neighbor interchange distance between the trees
10. Find two trees \( T, S \in \mathcal{T}_n \) such that the splits metric is \( \rho(T, S) = 2n - 6 \).