Supplement on UPCEMA
UPGMA in its execution.

topology and edge weights, and UPGMA is identical with

FACT. If tree is ultrametric, then UPGMA correctly computes the

The goal of this appendix is to prove the following basic fact.
Ultrametric tree example

Diagram:

```
      a
     /  
    3    3
   / 
  b   q
 /   /
1  3  3
 /  /
 d  c
 / /
2 1 2
 /  
 p e
```
Given as follows.

The distance matrix for the ultrametric tree example is:

\[
\begin{array}{cccccc}
0 & 6 & 6 & 10 & 10 \\
6 & 0 & 2 & 10 & 10 \\
6 & 2 & 0 & 10 & 10 \\
10 & 10 & 10 & 0 & 6 \\
10 & 10 & 10 & 6 & 0 \\
\end{array}
\]
In combining clusters $c$ into a supercluster $e$, the label of the least index (1 \geq i \geq \delta) in a cluster $c$ as the label of the cluster combined into a supercluster. Our implementation chooses the label of the smaller of the labels of the two clusters.

Each step through the algorithm UPGMA, two clusters are

\[
\begin{array}{cccccc}
0 & 0 & 6 & 10 & 10 \\
0 & 0 & 0 & 0 & 0 \\
9 & 0 & 1 & 10 & 10 \\
10 & 0 & 10 & 0 & 6 \\
10 & 0 & 10 & 6 & 0 \\
\end{array}
\]

c and $p$ will be deleted.
See output from computer program Phys.c.

and columns of the distance matrix $D$. Leaves are labeled 0, 1, 2, 3, 4 corresponding to the original rows of that node (distance from that node to the leaves), while the height of the numerical value of the internal nodes is the height finally.
where additionally \( R_+ \rightarrow \mathcal{A} : M \). 

A tree with positive edge weights is of the form (acyclic).

\[
\begin{align*}
\forall i \neq j \iff i \neq j \in \{ 1, \cdots, m \} \vee i, j \in \{ 0 \} \neq 0, \forall i, j \in \{ 1, \cdots, m \}, \text{ such that } i_0 = m_i = 0 \text{ and for all } \\
\text{there is no cycle } \{ 1 \} \cup \{ 0 \} \\
\text{(connectivity)} \mathcal{A} \in \mathcal{E}^{(+)} \\
\forall m > j \geq 0 \text{ and for all } \forall i, j \text{ such that } x_i = x_j \in \Lambda \in \mathcal{E} \text{ there is a path} \\
\text{(edges undirected)} [\mathcal{A} \in \{ i, j \} \leftrightarrow \mathcal{A} \in \{ j, i \}] (\Lambda \in \mathcal{E} \Lambda) \\
\Lambda \times \Lambda \subseteq \mathcal{A} \\
\text{where } (\mathcal{A}', \Lambda) = \mathcal{L} ; \text{ i.e.} \\
\text{Recall that a tree } \mathcal{L} \text{ is a connected, acyclic, undirected graph.} \\
\text{Formal development}
Recall that an ultrametric tree $T$ is a rooted tree with positive edge weights with the property that for all leaves $i, j \in T$,
\[
\forall i, j \in T \quad \text{with positive edge weights, then for all } x, y \in \Lambda, \text{ define } D(x, y) = \text{ the sum of edge weights along the unique path from } x \text{ to } y.
\]

**Definition 1** Let $T$ be a weighted tree with $\Lambda = \langle \mathcal{F}, \Lambda \rangle$.
\[ (\mathcal{J}, x)D + (x, \mathcal{J})D = (\mathcal{J}, \mathcal{J})D \]

and

\[ (\mathcal{J}, x)D + (x, \mathcal{J})D = (\mathcal{J}, \mathcal{J})D \]

Proof. If not, then by additivity of edge weights and definition of \( D \), \( \mathcal{J} \) leaves \( \mathcal{J} \in \mathcal{J} \). If \( \mathcal{J} \) is an ultrametric tree and \( x \in \mathcal{J} \), then for all \( \mathcal{J} \)
The proof is easy by induction on height of $x$. If the height is 0,
\[
(x)y - (\{x\})M
\]
then the weight $M$ must equal $y$. Otherwise, the height of $x$ is

then $x$ is a leaf and has no children. Otherwise, the height of $x$ is

so the lemma follows.

Lemma 2 Suppose that $y$ is a child of $x$ in the ultrametric tree $T$.

property (see last lemma).

child $y$ of $x$ in above definition), because of the ultrametric

height of a node is well defined (i.e. independent of choice of

[\begin{align*}
(x)y & + (\{x, y\})M \\
0 & \\
\end{align*}\]

height

Definition 2 If $T$ is an ultrametric tree and $x \in T$, then define the
Definition 3. If $\mathcal{T}$ is a tree and $c$ is contained in the set of leaves of $\mathcal{T}$, then $\mathcal{T}$ is the smallest subtree whose set of leaves contains $c$. If $\mathcal{T}$ is a set of leaves, we mean the subtree of $\mathcal{T}$ of the vertex $x$, where $x$ is a vertex of $\mathcal{T}$, where $x \in \mathcal{T}$. There should be no difficulty, as context will distinguish whether this contains $c$. Note the difference with our earlier definition of $\mathcal{T}$, where $x \in \mathcal{T}$.  

subtree of $\mathcal{T}$, defined as the smallest subtree whose set of leaves.
for an edge between clusters \( c \) and \( e \). In class, I had

\[ \forall (c, e) \in \mathcal{E}, \exists \text{ height } \mathcal{H}(e), \text{ for a cluster } e \in \mathcal{C}, \text{ and hence the notion of weight during the execution of UPGMA, we can define the notion of } \]

\[ \text{far. Moreover by the following definition, applied inductively to the set of nodes (clusters) formed together as a subtree so} \]

\[ (\{u, \ldots, i\}) \mathcal{A} = (\Lambda) \mathcal{A} \subset \mathcal{A} \]

\[ (\mathcal{M}, \mathcal{E}, \mathcal{A}) = \mathcal{I} \]

UPGMA, we have a set of three, where each tree

\[ \text{roots of subtrees already defined. Thus, during the execution of } \]

\[ \text{cluster } e \text{ having two children } c, d \text{, which themselves are the } \]

\[ \text{cluster from clusters } c, d \text{, we form a new tree, whose root is the } \]

\[ \text{formed from clusters } c, d \text{, we form a new tree, whose root is the } \]

\[ \text{subcluster } e \text{ as its sole root (and leaf). When subcluster } e \]

\[ \text{initially, there are no trivial trees, the ith tree consisting of the } \]

\[ \text{trees, whose nodes are clusters which have been formed so far. } \]

\[ \text{At any point in the execution of UPGMA, there is a set of } \]
extra notation is more confusing than anything.)
used the notation $c$ for the label of cluster $c$, but I think this
\[(p)\sim - (e)\sim = (\{p',e\})_M\]
\[(e)\sim - (e)\sim = (\{e',e\})_M\]

Inductively define

\[
\frac{2}{\text{dist}(p',e)} = (e)\sim
\]

be the height of e as defined in the algorithm in the text, i.e., in the execution of the $\text{UPCMA}$ algorithm, then define $h(e)$ to

Definition 4

If supercluster $e$ is set equal to $c$ at some step
point in the execution of UPGMA and then Claim A.

The proofs of the claims is by induction on the number of steps,

$\phi$, where induction is on the height of vertices.

Within the definition of within Claim B, since one of the

extend the definition of within Clai$\phi$m $B$, we will turn out to be a leaf of $L$ is mapped to the leaf $i$ of $L$. We

which

additionally respects edge weights. To that end, initially define

the tree $L$ of $UPGMA$ and the original ultrametric tree $\mathcal{L}$, which

We intend to build up an isomorphism between the output
\[(p, c)_{\text{dist}} = (\langle \ell, i \rangle), \text{hca}(\ell) A = (\langle \ell, i \rangle) A = \langle \ell, i \rangle A \cdot 2 = 2 \cdot \langle \ell, i \rangle A = 2.
\]

I. \text{hca}(\ell) = \text{hca}(\ell).

Claim A. Let \( c \) be distinct clusters at any point in the execution of \( \text{UPGMA} \). Then for all \( i, j \) and \( c, \notj \subset D \),
\[(\{c_1\})y = (\{c_1\})y\]

and for all vertices of \(Y,\)

\[((c_2)(\{c_1\}))(M) = (\{c_2, c_1\})(M)\]

weights, in the sense that for all edges \(e, f\) which respects edge \(Y,\)

and is an isomorphism of onto \(U \cup \text{CMA}\) which contains \(e,\) then \(e\)

execution of \(U \cup \text{CMA}\) which contains \(e\).

2. Let \(Y = (\mathcal{M}, \mathcal{L})\) be the unique tree at this point in the

Define \(x = (\{e\})\phi\)

\[e = p \cap c = (\mathcal{L})\text{eavness} T = (x_L)\text{eavness} T\]

1. There is a vertex \(x \in L \in \mathcal{L}\) such that

Claim B. Suppose that the supercluster \(e\) is formed at

some point in the execution of \(U \cup \text{CMA.}\)
\((f', \beta) \Delta = (f', \beta) \Delta\)

cluster \(e\) formed by \(c \cup \{p\}\) and \(f \in f\), then for any \(i \in c, p \in e\) and \(f \in f\),

Then in the execution of UPGMA, if \(f \in f \in c\) is distinct from

\[
\begin{align*}
(f', \beta) & \Delta \sum_{i \in e, j \in e} \frac{|f||\beta|}{1} = \\
& \frac{|p| + |\sigma|}{(f', \rho) \text{dist} |p| + (f', \sigma) \text{dist} |\sigma|} = (f', \sigma) \text{dist}
\end{align*}
\]

WPAGMA for clusters \(e\) and \(f\). Then

\textbf{Proposition 1} Let \(\text{dist}(f', e)\) be the distance defined by

following proposition we have proved.

WPAGMA are identical on ultrametric trees. Namely, recall the

Note that Claim A and Claim B imply that WPAGMA and
$$\frac{2}{(f',p)_{dist} + (f',\mathcal{C})_{dist}} = \frac{(f',p)_{dist}}{(f',p)_{dist} + (f',\mathcal{C})_{dist}} = \frac{|p| + |\mathcal{C}|}{(f',\mathcal{C})_{dist}|p| + (f',\mathcal{C})_{dist}|\mathcal{C}|} = (f',\mathcal{C})_{dist}$$

Hence $$(f',p)_{dist} = (f',\mathcal{C})_{dist}$$ and so