We run Kruskal's MST algorithm, and build τ inductively as we go. Each time we merge components C_i and C_j by an edge of length ℓ, we create a new node v to be the parent of the subtrees that (by induction) consist of C_i and C_j, and give v a height of ℓ.

Note that for any p_i and p_j, the quantity $\tau(p_i, p_j)$ is equal to the edge length considered when the components containing p_i and p_j were first joined. We must have $\tau(p_i, p_j) \leq d(p_i, p_j)$, since p_i and p_j will belong to the same component by the time the direct edge (p_i, p_j) is considered.

Now, suppose there were a hierarchical metric τ' such that $\tau'(p_i, p_j) > \tau(p_i, p_j)$. Let T' be the tree associated with τ', v' the least common ancestor of p_i and p_j in T', and T'_i and T'_j the subtrees below v' containing p_i and p_j. If h'_v is the height of v' in T', then $\tau'(p_i, p_j) = h'_v > \tau(p_i, p_j)$.

Consider the p_i-p_j path P in the minimum spanning tree. Since $p_j \notin T'_i$, there is a first node $p' \in P$ that does not belong to T'_i. Let p be the node immediately preceding p on P. Then $d(p, p') \geq h'_v > \tau(p_i, p_j)$, since the least common ancestor of p and p' in T' must lie above the root of T'_i. But by the time Kruskal's algorithm merged the components containing p_i and p_j, all edges of P were present, and hence each has length at most $\tau(p_i, p_j)$, a contradiction.

[1] ex947.542.655