Homework #1

Selected Solutions

2.1 For the first tree, $V = \{a, b, u, v, c, d\}$ and $E = \{\{a, u\}, \{b, u\}, \{u, v\}, \{v, c\}, \{v, d\}\}$.

For the second tree, $V = \{\rho, a, b, c, d, u, v\}$, and $E = \{(\rho, a), (\rho, u), (u, b), (u, v), (v, c), (v, d)\}$.

The edge $\{a, u\}$ was subdivided to create the rooted tree.

2.2 [Lander]

Answer:

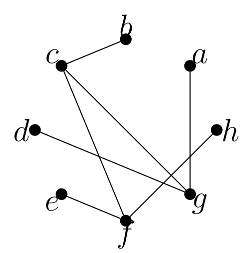

Figure 1: A naïve drawing of the graph specified in Exercise 2.2.

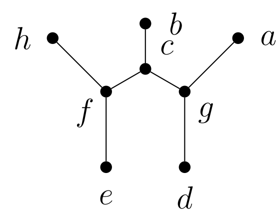

Figure 2: A more informed drawing of the graph in Exercise 2.2.

2.3 *Answer:* By counting the number of edges each vertex appears in, we obtain the following degrees: $d(v_1) = 3$, $d(v_2) = 3$, $d(v_3) = 1$, $d(v_4) = 1$, $d(v_5) = 1$, $d(v_6) = 1$, $d(v_7) = 1$, $d(v_8) = 1$, and $d(v_9) = 4$. The leaves are those vertices of degree 1, so in this case, v_3, v_4, v_5, v_6, v_7, and v_8 are leaves. The tree is not binary, as v_9 has degree 4. The tree is shown in Figure 3.

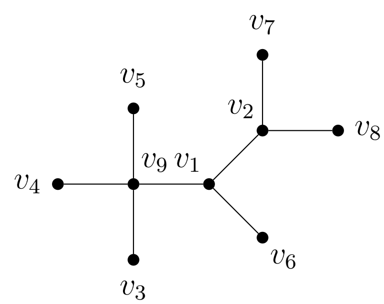

Figure 3: The tree specified in Exercise 2.3.

2.4
- a. T_2, T_3.
- b. T_2, T_3, T_5.
- c. $\{T_2, T_3, T_5\}$ and $\{T_1, T_6\}$.
- d. All six.
- e. T_4, T_6.

2.5 The three rooted trees are shown in Figure 4.

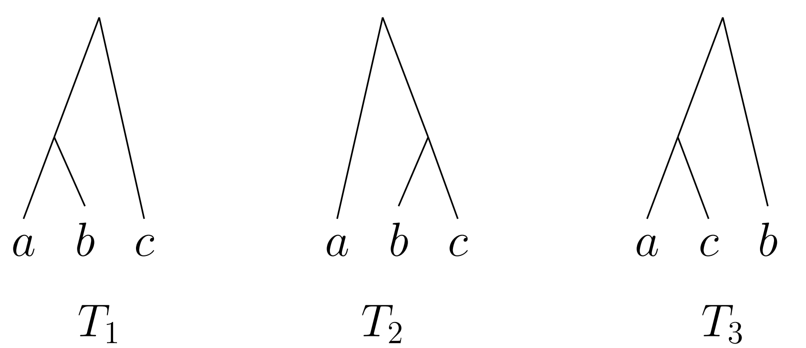

Figure 4: The three rooted trees on 3 taxa.

These three trees relate to the three trees in Figure 2.4 in that we are essentially taking one of those 4 elements (say, element 4) and treating it as an outgroup to root our tree.

2.7

n	$b(n)$
2	$0!! = 1$
3	$1!! = 1$
4	$3!! = 3$
5	$5!! = 15$
6	$7!! = 105$
7	$9!! = 945$
8	$11!! = 10,395$
9	$13!! = 135,135$
10	$15!! = 2,027,025$

2.8 Show that $b(n) = \dfrac{(2n-5)!}{2^{n-3}(n-3)!}$.

Proof. We have already established that $b(n) = (2n-5)!!$, the product of all odd numbers less than or equal to $2n-5$. We can, however, write this as $\frac{(2n-5)!}{(2n-6)!!}$, where we multiply all numbers less than or equal to $2n-5$, then divide out by the even numbers strictly less than $2n-5$. By cancellation, these are equivalent. However, $(2n-6)!! = (2(n-3))!!$. This suggests that $(2n-6)!!$ is in fact the product of all numbers less than or equal to $n-3$, where each number in that product is multiplied by 2. Because there are $n-3$ numbers in that list, we can represent $(2n-6)!!$ as $2^{n-3}(n-3)!$. This gives us the final formula

$$b(n) = \frac{(2n-5)!}{2^{n-3}(n-3)!}.$$

$\square$

2.9 In this case, $n = 147$, and given that this is an ancestry question, it makes sense to talk about rooted trees. Thus, we will compute $b(n+1) = b(148)$, and using the statement in Problem 8, we can say that

$$b(148) = \frac{(2(148)-5)!}{2^{148-3}(148-3)!} = \frac{291!}{2^{145}(145!)}.$$

Using Stirling's formula (and canceling the $\sqrt{2\pi}$ from the top and bottom)

$$b(n) \sim \frac{291^{291.5}e^{-291}}{2^{145}145^{145.5}e^{-145}} = e^{-146}\frac{\sqrt{291}291^{291}}{2^{145}145^{145}\sqrt{145}} = e^{-146}\sqrt{\frac{291}{145}}\frac{291^{291}}{290^{145}}.$$

A numerical approximation of this from Wolfram Alpha gives that it is approximately 4.89×10^{296}.

2.13 a. The correct unrooted phylogenetic tree is shown in Figure 5.

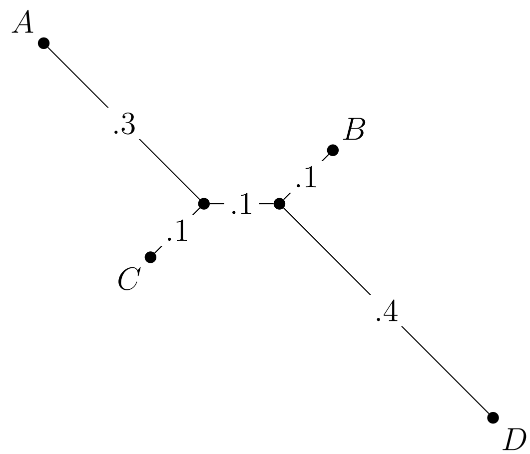

Figure 5: The metric tree (to scale) corresponding to the distance table in Exercise 2.13.

b. This tree is not ultrametric, as the longest path from A is .8, while the longest path from B is only .5. In an ultrametric tree, all those longest paths would have equal length. Thus, an objectively "best" root cannot be determined from the distance table.

c. Most *ad hoc* methods only work well for these quite small trees, and would likely not work out if the numbers were only approximately correct. Thus, better tools are needed.

2.14 The term 'ultrametric' originally was not applied to trees, but rather to a function $d : V \times V \to \mathbb{R}^{\geq 0}$ that satisfied not only the properties of a metric listed in Proposition 5, but also a strong form of the triangle inequality:

$$d(v_1, v_3) \leq \max(d(v_1, v_2), d(v_2, v_3)) \text{ for all } v_1, v_2, v_3.$$

a. Show that this property implies the usual triangle inquality (iii) of Proposition 5.

b. Show that for a tree metric arising from an ultrametric tree, the strong triangle inequality holds when v_1, v_2, v_3 are leaves.

c. Show by a 3-leaf example that the strong triangle inequality on leaves does not hold for all tree metrics.

d. Show that if the strong triangle inequality holds, then for all choices of v_1, v_2, v_3 the two largest of the numbers $d(v_1, v_2), d(v_1, v_3)$ and $d(v_2, v_3)$ are the same. (This is sometimes stated as: An ultrametric implies all triangles are isoceles.)

e. Show that if a tree metric from an unrooted 3-leaf tree satisfies the strong triangle inequality on leaves, then there is a placement of a root for which the underlying tree is ultrametric. (This holds more generally for n-leaf tree metrics satisfying the strong triangle inequality on leaves; with proper placement of a root, they all arise from ultrametric trees.)

Answer:

a. *Proof.* Observe that $\max(d(v_1, v_2), d(v_2, v_3)) \leq d(v_1, v_2)+d(v_2, v_3)$, because distance functions are non-negative. Thus, if $d(v_1, v_3) \leq \max(d(v_1, v_2), d(v_2, v_3))$, it follows from the transitivity of $\leq$ that $d(v_1, v_3) \leq d(v_1, v_2)+d(v_2, v_3)$. $\square$

b. *Proof.* First, it is easy to see that in any rooted ultrametric tree, any subtree induced by some subset of the leaves (so no internal node becomes a leaf) is also ultrametric. If this were not the case, then there would be two leaves v_1 and v_2 that have different distances from the root ρ' of this subtree. But ρ' is an internal vertex of the original tree, so it lies on the unique path from v_1 to ρ and v_2 to ρ, and there is a fixed distance from ρ' to ρ. Thus, v_1 and v_2 cannot have both different distances from ρ' and the same distance from ρ.

   So, choose v_1, v_2 and v_3 from the leaves in some ultrametric tree T, and let ρ' be the root of the minimal subtree T' defined by v_1, v_2, and v_3. Let us define α as the distance from ρ' to any leaf, which is constant because T' is ultrametric. Observe that in all situations, $d(v_1, v_3) \leq 2\alpha$. First suppose the path from v_1 to v_3 does not pass through ρ'. Then at least one of the paths from v_1 to v_2 or from v_3 to v_2 must pass through ρ', or there would be a more recent internal vertex that could have been the root of our subtree. Thus, $\max(d(v_1, v_2), d(v_2, v_3)) = 2\alpha$, and as already established $d(v_1, v_3) \leq 2\alpha$.

   Next, suppose the shortest path from v_1 to v_3 passes through ρ', and thus $d(v_1, v_3) = 2\alpha$. Note that one of three cases occurs: our tree is not binary, and v_1, v_2, and v_3 are all descended directly ρ'; v_1 and v_2 are sisters; or v_2 and v_3 are neighbors. In case 1, $d(v_1, v_3) = d(v_1, v_2) = d(v_2, v_3) = 2\alpha$. If v_1 and v_2 are sisters, then the shortest path from v_2 to v_3 must pass through ρ', and $d(v_2, v_3) = 2\alpha$, and if v_2 and v_3 are sisters, then the shortest path from v_1 to v_2 must pass through ρ' and $d(v_1, v_2) = 2\alpha$. In any of these cases, $\max(d(v_1, v_2), d(v_2, v_3)) = 2\alpha$. $\square$

c. *Proof.* Consider the tree shown in Figure 6. This tree has the property that $d(v_2, v_3) = .5$, while $d(v_1, v_2) = .3$ and $d(v_2, v_3) = .4$. Thus, this tree is not ultrametric. $\square$

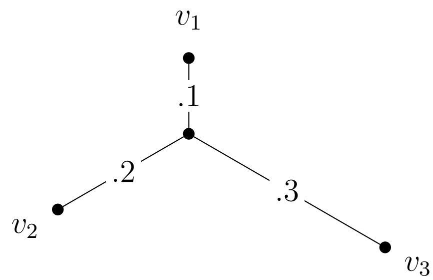

Figure 6: An example showing not all metric trees are ultrametric.

d. *Proof.* Suppose to the contrary that all three distances are distinct. Then there exists a largest; suppose without loss of generality that $d(v_1, v_2)$ is the largest. That is, $d(v_1, v_2) > d(v_1, v_3)$ and $d(v_1, v_2) > d(v_2, v_3)$. But then $d(v_1, v_2) > \max(d(v_1, v_3), d(v_2, v_3))$, which contradicts our assumption that the strong triangle inequality holds (which would imply that $d(v_1, v_2) \leq \max(d(v_1, v_3), d(v_2, v_3))$). $\square$

e. *Proof.* From part (d), we know that the largest two of $d(v_1, v_2)$, $d(v_1, v_3)$, and $d(v_2, v_3)$ are equal. Suppose without loss of generality that $d(v_1, v_2) = d(v_1, v_2)$, a distance which we will call α, and that both are greater than or equal to $d(v_2, v_3)$. Any two of those three sets must have some vertex in common; in this case it is v_1.

   Furthermore, the edge from v_1 to the single internal vertex (let us call it C) in a 3-taxon tree must be greater than or equal to the other two edges. To see this, suppose to the contrary that some other edge is strictly greater than the v_1C edge. Then, because the distances from v_1 to each of v_2 and v_3 are equal, $|v_2C| = |v_3C| > |v_1C|$. But then $d(v_2, v_3) > \max(d(v_1, v_2), d(v_1, v_3))$, which contradicts our assumption that the tree is ultrametric.

   Thus, if we place a vertex ρ a distance $\alpha/2$ from v_1 on the path from v_1 to v_2, it must fall somewhere on the v_1C edge (possibly at C itself). In either case, it also lies on the path from v_1 to v_3. Thus, ρ is distance $\alpha/2$ from v_1 and v_2 (as it is the midpoint of that path) and distance $\alpha/2$ from v_1 and v_3. Therefore, ρ is the root of an ultrametric rooted tree with leaves v_1, v_2, and v_3. $\square$