Lecture # 4 - Friday, Aug 27th

In this lecture I, went over the homework # 1 in detail and then introduced the following topics

Split compatibility

Two splits are compatible if and only if they can both be "put" onto the same tree. Recall that a split is a bipartition of the set of leaves. So if the set of leaves is $\{1, 2, 3, 4, 5, 6\}$, then a possible split, A, could be: $A = \{1, 6\}|\{2, 3, 4, 5\}$. Note that for splits that are not oriented (in other words splits that are not rooted), it is only the presence of the partition that matters not the left and right side *per se*. Thus, $A = \{2, 3, 4, 5\}|\{1, 6\}$ is also a valid statement.

It is convenient to label one subset of leaves defined by a split as group 1 and the other as group 2 (once again for unrooted splits we can reverse the labels without changing the content. So for the example above, we could say $A_1 = \{1, 6\}$ and $A_2 = \{2, 3, 4, 5\}$

Test for split compatibility

Consider two splits, A and B. Construct the following four intersections:

$$A_1 \cap B_1 \tag{1}$$
$$A_1 \cap B_2 \tag{2}$$
$$A_2 \cap B_1 \tag{3}$$
$$A_2 \cap B_2 \tag{4}$$

If *at least one* of the four intersections is the empty set (which could be denoted $\{\}$, but is usually denoted $\emptyset$) then the two splits are compatible.

Example test for split compatibility

Consider the splits $A = \{1, 6\}|\{2, 3, 4, 5\}$, and $B = \{2, 3, 4, 6\}|\{1, 5\}$, and $C = \{1, 4, 5, 6\}|\{2, 3\}$. For convenience we will number the subsets of each split with a 1 for left of the | symbol and 2 for right of the | symbol. A and B are incompatible because:

$$A_1 \cap B_1 = \{6\} \tag{5}$$
$$A_1 \cap B_2 = \{1\} \tag{6}$$
$$A_2 \cap B_1 = \{2, 3, 4\} \tag{7}$$
$$A_2 \cap B_2 = \{5\} \tag{8}$$

Note that none of the intersections is empty.

C is compatible with A (because $A_1 \cap C_2 = \{1, 4, 5, 6\} \cap \{2, 3\} = \emptyset$), and C is also compatible with B (because $B_1 \cap C_2 = \{1, 5\} \cap \{2, 3\} = \emptyset$).

Split Equivalence Theorem (Buneman, 1971)

A collection of splits can be equivalent to a tree in the sense that there can be one tree that displays that collection of splits and no other splits. From the tree you can obtain the collection of splits, and from the collection of splits you can build the tree. We can denote set of splits displayed by a tree as $S(T)$.

But not all collections of splits can be built into a tree.

The split equivalence theorem states that: A set of splits, $\mathbb{S}$ can be assembled into a tree T such that $\mathbb{S} = S(T)$ if and only if all of the pairs of splits in $\mathbb{S}$ are compatible pairs.

In other words, we only have to test for pairwise compatibility to find out whether a collection of splits can be combined into a tree.

A common use of the theorem is to take a collection of splits and identify the largest collection that can be built into a tree. To do this, we can construct a split compatibility graph (use nodes to represent the splits and introduce an edge between every pair of splits that is compatible), and then identify the largest clique in the graph. A clique is a graph (or subgraph) in which all nodes are adjacent to all other nodes (there is an edge between each pair of nodes).