CS 224 Fall 2009, HW 4 Due November 5, although that might not be enough time - let me know.

Splits Equivalence

Problems 1, 2: In the notes on the Perfect Phylogeny Problem, on pages 18 and 19 there are two problems concerning Splits Equivalence and its relation to the Four Gametes Theorem. Solve those problems.

Steiner Tree problem on Hypercubes

Let $G = (N, E)$ be an undirected graph on node set N. Let $X \subseteq N$ be a given subset of nodes. A *Steiner tree* ST of G for X is any connected subtree of G that contains all the nodes of X, although it may contain nodes in $N - X$ as well. The *weight* of a Steiner tree ST is the number of edges in ST, and is denoted $W(ST)$. The *Steiner tree problem*, given G and X, is to find the Steiner tree for X in G of minimum weight.

A *hypercube* of dimension k is an undirected graph with 2^k nodes, where the nodes are labeled with the integers between 0 and $2^k - 1$. Two nodes in the hypercube are adjacent if and only if the binary representation of their labels differs in exactly one bit. The Steiner tree problem on hypercubes is the Steiner tree problem where the graphs are hypercubes.

The (two-state) perfect phylogeny problem can be viewed as a question about the weight of minimum weight Steiner Tree.

Problem 3: What question about the minimum weight Steiner Tree can be cast as the perfect phylogeny problem? As a hint, develop a simple lower bound on $W(ST)$ that applies to any G and X. Then relate the perfect phylogeny problem to that lower bound.

Rooted Tree Compatibility

Definition A rooted, leaf-labeled tree T' is a *refinement* of another rooted, leaf-labeled tree T if T can be obtained by a series of contractions of edges of T. When an edge (u, v) touching a leaf v is contracted, the label of v is

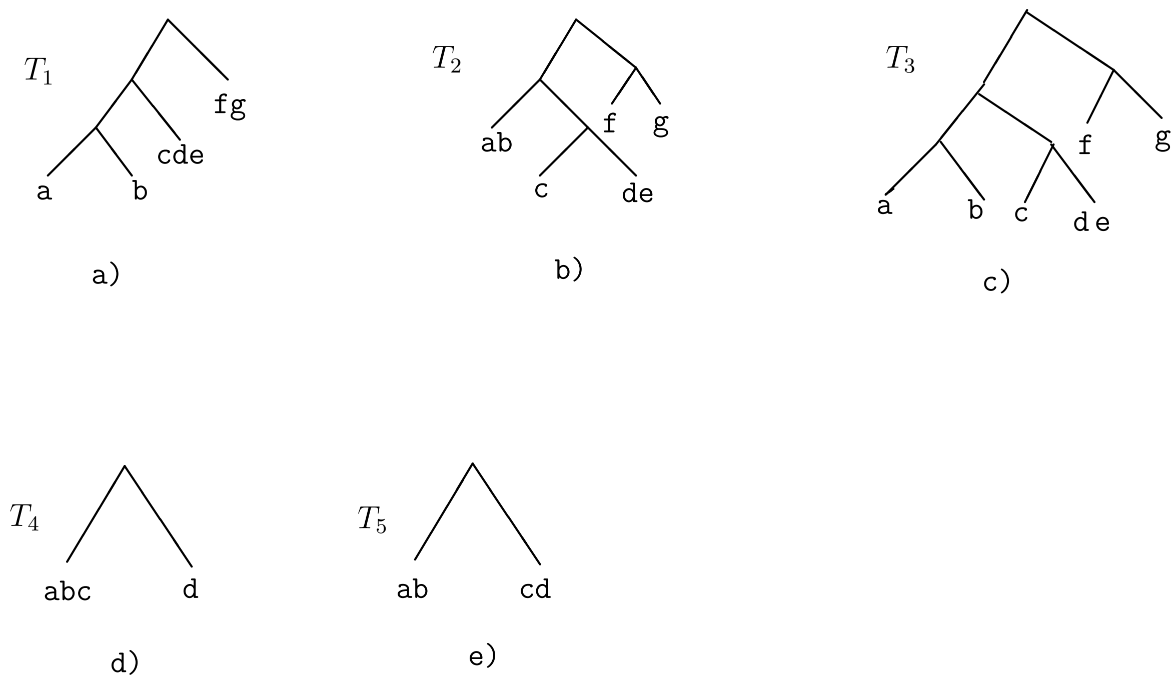

Figure 1: Trees T_1 and T_2 are compatible; they are refined by T_3. Trees T_4 and T_5 are not compatible.

added to any label on u. In general when an edge (u, v) directed from u to v is contracted, any label on v is added to whatever label is on u.

If T' refines T, then T' agrees with all the evolutionary history displayed in T, while displaying additional history not contained in T.

Let T_1 and T_2 be two rooted trees whose leaves are labeled with the same set of labels. But note that T_1 and T_2 can have different numbers of leaves, and a leaf can have more than a single label. We will assume that T_1 and T_2 are both in "reduced form", that is, both are binary trees, and no node except the root can have exactly one child.

Definition Trees T_1 and T_2 are *compatible* if there exists a rooted tree T_3 refining both T_1 and T_2 (see Figure 1).

Tree compatibility problem Given trees rooted, leaf-labeled trees T_1 and T_2 on the same label set, determine whether the two trees are compatible, and if so, produce a refinement tree T_3.

Let M_1 be a 0-1 matrix with one row for each taxon and one column for each node j in T_1, including leaves. Entry (i, j) of M_1 has value one if and only if taxon i is found at or below node j. That is, column j of M_1 records the taxa found in the subtree of T_1 rooted at node j. Matrix M_2 is similarly defined for T_2, and matrix M_3 is the matrix formed by the union of the columns of M_1 and M_2. Then

Theorem T_1 and T_2 are compatible if and only if there is a perfect-phylogeny for M_3 with the all-zero ancestral sequence. Further, a perfect-phylogeny T_3 for M_3 is a refinement of both T_1 and T_2. Note that two columns in M_3 might be identical, and hence T_3 might have a leaf with more than a single label.

Problem 4: Prove (with a good explanation) this theorem.