Bioinformatics – Lecture Notes

Announcements

Next week is Spring Break (March 11-15). NO CLASS.

Class 15 – March 5, 2002 -

1. Evolutionary Problems – Clustering Methods – How do we determine the best fit phylogenetic data?

   Cannot determine the minimum number of mutations for all possible tree topologies having n leaves since the number of binary trees is exponential. Instead we will use some concepts from graph theory

   Definitions

   taxa – entities whose distance from other entities can be measured

   A *directed graph* G(V, E) consists of a set V of *nodes* or *vertices* and a set $E \subseteq V \times V$ of *directed edges*. Then $(i,j) \in E$ means that there is a directed edge from i to j.

   A graph is *undirected* if the edge relation is *symmetric*, that is, $(i,j) \in E$ iff $(j,i) \in E$.

   A directed graph is *connected* if there is a directed path between any two nodes.

   A directed graph is *acyclic* if it does not contain a cycle. (ie. (i,j), (j,k), and (k,i) all belong to E.

   A *tree* is a undirected, connected, acyclic graph.

   A *rooted tree* has a starting node called a *root*.

   The *parent node* is immediately before a node on the path from the root.

   The *child node* is a node that is follows a node.

   An *ancestor* is any node that came before a node on the path from a root.

   A *leaf* or *external node* is a node that had no children.

   Non-leaf nodes are called *internal nodes*.

   The *depth* of a tree is one less than the maximal number of nodes on a path from the root to a leaf.

   An *ordered tree* is a tree where the children of internal nodes are numbered.

   A *binary tree* is a tree where each node has at most two children. Otherwise it is *multifurcating*.

   Question: Draw all binary trees on 1, 2, and 3 taxa.

   A *phylogenetic tree* on n taxa is a tree with leaves labeled by 1,…,n.

Let T_1, and T_2 be phylogenetic trees on n taxa. Then $T_1=(V_1,E_1)$ and $T_2=(V_2,E_2)$, $E_1 \subseteq V_1 \times V_1$ and $E_2 \subseteq V_2 \times V_2$, and the leaves of both T_1 and T_2 are labeled 1,2,…,n. A function $f: V_1 \to V_2$ is an isomorphism from T_1 to T_2 if the following condtions are satisfied.

1) f is one-to-one and onto
2) $x \in V_1$ is a leaf of T_1 labeled by I iff $f(x) \in V_2$ is a leaf of T_2 labeled by i.
3) $(x,y) \in E_1$ iff $(f(x), f(y)) \in E_2$

a) ultrameric trees
b) additive metric
c) estimating branch lengths