Lecture 5: Trees and splits

Cuong Van Than

November 14, 2013

Trees and splits. An X-split is a partition of X into two (nonempty) subsets, which are also called blocks of the split. If A and B are the two blocks of the split, we denote the split by $A|B$. We say $A|B$ is a trivial split if either $|A| = 1$ or $|B| = 1$.

Let T be an unrooted X-tree, and let e be an edge of T. The removal of e results in two connected components. Clearly the label sets of the two components form an X-split. The split is said to be induced by e. We denote by $\Sigma(T)$ the set of splits induced by edges of T

Two X-splits $A|B$ and $C|D$ are compatible if at least one of the four intersections $A \cap C$, $A \cap D$, $B \cap C$, $B \cap D$ is empty. It can be seen that if $A|B$ and $C|D$ are distinct, they are compatible if and only if *exactly* one of the four intersections above is empty. Also, a trivial split is compatible with every X-split.

Theorem 1. *Let Σ be a collection of X-splits. Then there exists an X-tree T such that $\Sigma = \Sigma(T)$ if and only if the splits in Σ are pairwise compatible.*

Let T and T' be two X-trees. We write $T \leq T'$, or T' refines T, if $\Sigma(T) \subset \Sigma(T')$. Intuitively, if T' refines T, then T can be obtained from T' by contracting edges. It is easy to see that $\leq$ is a partial order on the set of X-trees.

We say that a collection of X-trees is compatible if there exists an X-tree that refines every element of the collection (such a tree is also called an upper bound under $\leq$).

Corollary 2. *A collection P of X-trees is compatible if and only if $\bigcup_{T \in P} \Sigma(T)$ is a pairwise compatible set of X-splits.*

Tree splits metric. Let T and T' be two X-trees. Define $d(T, T')$ as $|\Sigma(T) \setminus \Sigma(T')| + |\Sigma(T') \setminus \Sigma(T)|$ (i.e. the size of the symmetric difference of $\Sigma(T)$ and $\Sigma(T')$).

Theorem 3. *The function d is a metric on the set of X-trees.*

Hierarchies. A collection H of nonempty subsets of X is called a hierarchy on X if for all A, $B \in H$,

$$A \cap B \in \{\emptyset, A, B\}.$$

Theorem 4. *Let H be a collection of nonempty subsets of H. Then there exists a rooted X-tree such that H is equal to $H(T)$, the set of clusters induced by T, if and only if H is a hierarchy on X.*

Concensus trees. Let P be a collection of X-trees, and let $\Sigma = \bigcup_{T \in P} \Sigma(T)$. For $\sigma \in \Sigma$, let $n(\sigma)$ be the number of elements of T that induce σ. For $0 \leq q < 1$, let

$$\Sigma_q = \{\sigma \in \Sigma \mid n(\sigma) > q\,|P|\}.$$

Theorem 5. *If $q \geq 0.5$, then Σ_q is the set of X-splits induced by an X-tree.*

Adams consensus trees. Let $\{H_1, \ldots, H_k\}$ be a collection of hierarchies on X, and for each $i = 1, \ldots, k$, let H_i^* be the set of maximal elements (under set inclusion) of H_i. Let

$$C^* = \left\{ \bigcap_{i=1}^{k} A_i \mid A_i \in H_i^* \right\} \setminus \{\emptyset\}.$$

The elements of C^* are maximal elements of the hierarchy of the Adams consensus tree for $\{H_1, \ldots, H_k\}$. For each $C \in C^*$, consider the collection of hierarchies $H_1, \ldots, H_k$ restricted on C, and repeat the above procedure to obtain a partition of C. This process continues until no further partition can be formed.