Notes 5 : Quartet Theorem

MATH 833 - Fall 2012 *Lecturer: Sebastien Roch*

References: [SS03, Chapter 6]

1 Restricted subtrees

Here, an X-tree is also called a *semi-labelled tree*. If two X-trees $\mathcal{T}$, $\mathcal{T}'$ are isomorphic, we write $\mathcal{T} \cong \mathcal{T}'$.

DEF 5.1 (Restricted subtree) *Let $\mathcal{T}$ be an X-tree and $X' \subseteq X$. The* restriction $\mathcal{T}$ to X', *denoted $\mathcal{T}|X'$, is the X'-tree such that*

$$\Sigma(\mathcal{T}|X') = \{A \cap X' | B \cap X' \,:\, A|B \in \Sigma(\mathcal{T})\}.$$

$\mathcal{T}|X'$ is obtained from $\mathcal{T} = (T, \phi)$ by taking the minimal subtree of T including $\phi(X')$ and suppressing degree-two vertices not in $\phi(X')$.

DEF 5.2 (Displaying a semi-labelled tree) *Let $X' \subseteq X$. An X-tree $\mathcal{T}$* displays *an X'-tree $\mathcal{T}'$ if $\mathcal{T}' \leq \mathcal{T}|X'$. Similarly, $\mathcal{T}$ displays a collection $\mathcal{P}$ of semi-labelled tree if it display every tree in $\mathcal{P}$.*

2 Quartet theorem

The following theorem indicates that X-trees are characterized by their restricted subtrees on sets of size at most 4.

THM 5.3 (Quartet theorem) *Let $\mathcal{T}, \mathcal{T}'$ be X-trees. Then, $\mathcal{T} \cong \mathcal{T}'$ if and only if $\mathcal{T}|S \cong \mathcal{T}'|S$ for all $S \subseteq X$ with $|S| \leq 4$.*

Proof: We prove a slightly more general statement:

$$\mathcal{T} \leq \mathcal{T}' \text{ if and only if } \mathcal{T}|S \leq \mathcal{T}'|S \text{ for all } S \subseteq X \text{ with } |S| \leq 4.$$

Note that $\mathcal{T} \cong \mathcal{T}'$ if and only if $\mathcal{T} \leq \mathcal{T}'$ and $\mathcal{T}' \leq \mathcal{T}$.

One direction is trivial. We prove the other direction. Assume $\mathcal{T}|S \leq \mathcal{T}'|S$ for all $S \subseteq X$ with $|S| \leq 4$. Let $A|B \in \Sigma(\mathcal{T})$. We seek to prove that $A|B \in \Sigma(\mathcal{T}')$. First, note that for all $a, a' \in A$ and $b, b' \in B$, by definition of restriction we have

$$\{a, a'\}|\{b, b'\} \in \Sigma(\mathcal{T}|\{a, a', b, b'\}) \subseteq \Sigma(\mathcal{T}'|\{a, a', b, b'\}),$$

where the inclusion follows by assumption. We proceed by contradiction. Suppose $A|B$ is not a split of $\mathcal{T}'$. There are two cases:

- *There is a split $A'|B'$ of $\mathcal{T}'$ incompatible with $A|B$.* Then there is $a \in A \cap A'$, $a' \in A \cap B'$, $b \in B \cap A'$, $b' \in B \cap B'$. But then $\{a, b\}|\{a', b'\} \in \Sigma(\mathcal{T}'|\{a, a', b, b'\})$, contradicting the pairwise compatibility of the latter.

- *$A|B$ is compatible with every split in $\mathcal{T}'$.* Recall the following lemma from the proof of the Splits-Equivalence theorem:

  LEM 5.4 (Label Painting Lemma) *Let $\mathcal{T}' = (T', \phi')$ be an X-tree with $T' = (V', E')$. Let $\sigma = A|B$ be an X-split such that $\sigma \notin \Sigma(\mathcal{T}')$ but σ is compatible with all splits in $\Sigma(\mathcal{T}')$. Colour red (respectively green) the vertices of $\mathcal{T}'$ in $\phi(A)$ (respectively $\phi(B)$). Then, there is a unique vertex $w \in V'$ such that the connected components of $T' \backslash w'$ are monochromatic.*

  Let w be as in the statement of the lemma. There are three cases:

  1. *w has labels from A and B.* Let $a \in \phi^{-1}(w) \cap A$ and $b \in \phi^{-1}(w) \cap B$. Then, $\{a\}|\{b\} \notin \Sigma(\mathcal{T}'|\{a, b\})$, a contradiction. (Take $a = a'$ and $b = b'$ above.)
  2. *w has a label from A but not from B (and the symmetric case).* Let $a \in \phi^{-1}(w) \cap A$. By the uniqueness of w, there must be $b, b' \in \phi(B)$ in two distinct components of $\mathcal{T}' \backslash w$. But then, $\{a\}|\{b, b'\} \notin \Sigma(\mathcal{T}'|\{a, b, b'\})$, a contradiction.
  3. *w is unlabelled.* Again by the uniqueness of w, there must be $a, a' \in \phi(A)$ (resp. $b, b' \in \phi(B)$) in two distinct components of $\mathcal{T}' \backslash w$. But then, $\{a, a'\}|\{b, b'\} \notin \Sigma(\mathcal{T}'|\{a, a', b, b'\})$, a contradiction.

■

3 Strong Quartet Evidence

We now restrict our attention to phylogenetic trees. Let Σ^0 be the set of trivial splits on X (i.e., where one side of the partition is a singleton).

DEF 5.5 (Quartet trees) *A* quartet tree $q = ab|cd$ *is a binary phylogenetic tree on four distinct labels* $\{a, b, c, d\}$ *where* $\{a, b\}|\{c, d\}$ *is the corresponding nontrivial split. For a collection* $\mathcal{Q}$ *of quartet trees, let*

$$\Sigma(\mathcal{Q}) = \{A|B \,:\, q = aa'|bb' \in \mathcal{Q},\ \forall a \neq a' \in A, b \neq b' \in B\}.$$

For an X*-split* $\sigma = A|B$*, let*

$$\mathcal{Q}_\sigma = \{aa'|bb' \,:\, a \neq a' \in A, b \neq b' \in B\}.$$

THM 5.6 *Let* $\mathcal{Q}$ *be a collection of quartet trees on* X *such that for all* S *with* $|S| = 4$ *at most one quartet tree of* $\mathcal{Q}$ *has label set* S*. Then there is a phylogenetic tree* $\mathcal{T}$ *on* X *such that* $\Sigma(\mathcal{Q}) \cup \Sigma^0 = \Sigma(\mathcal{T})$*.*

Proof: It suffices to show that $\Sigma(\mathcal{Q})$ is pairwise compatible. We proceed by contradiction. Assume $A_1|B_1, A_2|B_2 \in \Sigma(\mathcal{Q})$ are incompatible. Then there is $a \in A_1 \cap A_2$, $b \in A_1 \cap B_2$, $c \in B_1|A_2$ and $d \in B_1 \cap B_2$. But then $ab|cd \in \mathcal{Q}$ and $ac|bd \in \mathcal{Q}$, a contradiction. ■

There is an efficient algorithm for computing $\Sigma(\mathcal{Q}) \cup \Sigma^0$ given a collection $\mathcal{Q}$ satisfying the assumptions of the previous theorem. W.l.o.g. assume $X = [n]$. For $i \in [n]$, let

$$\mathcal{Q}_{[i]} = \{ab|cd \in \mathcal{Q} \,:\, a, b, c, d \in [i]\},$$

and $\Sigma_{[i]} = \Sigma(\mathcal{Q}_{[i]}) \cup \Sigma^0_{[i]}$ where $\Sigma^0_{[i]}$ is the set of trivial splits on $[i]$. The algorithm constructs the tree by adding labels in X one by one. It is based on the following observation:

LEM 5.7 *For all nontrivial* $\sigma = A|B \in \Sigma_{[i]}$*, there is* $\sigma' = A'|B' \in \Sigma_{[i-1]}$ *such that either* $\sigma = A' \cup \{i\}|B'$ *or* $\sigma = A'|B' \cup \{i\}$*.*

Proof: W.l.o.g., assume $i \in A$. If $|A - \{i\}| = 1$, σ' is trivial and we are done. Otherwise, by assumption for all distinct $i \neq a, a' \in A$ and $b, b' \in B$, we have $aa'|bb' \in \mathcal{Q}_{[i]}$ and hence $\in \mathcal{Q}_{[i-1]}$. ■

Let $\mathcal{Q}_i = \mathcal{Q}_{[i]} \backslash \mathcal{Q}_{[i-1]}$. The algorithm to build $\Sigma_{[n]}$ proceeds as follows:

- *Initialization.* $\Sigma_{[4]}$ is the union of $\Sigma^0_{[4]}$ and the split corresponding to the unique quartet tree in $\mathcal{Q}_{[4]}$ if any.

- *Main loop.* For $i = 5$ to n:
  - Set $\Sigma_{[i]} = \Sigma^0_{[i]}$.
  - For each $\sigma' = A'|B' \in \Sigma_{[i-1]}$, add $\sigma = A' \cup \{i\}|B'$ to $\Sigma_{[i]}$ if $\mathcal{Q}_\sigma \setminus \mathcal{Q}_{\sigma'} \subseteq \mathcal{Q}_i$ and similarly for $\sigma = A'|B' \cup \{i\}$.

Further reading

The definitions and results discussed here were taken from Chapter 6 of [SS03]. Much more on the subject can be found in that excellent monograph. See also [SS03] for the relevant bibliographic references.

References

[SS03] Charles Semple and Mike Steel. *Phylogenetics*, volume 24 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2003.