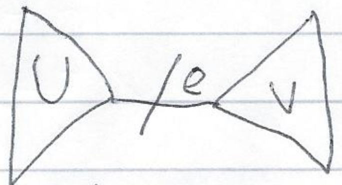
2 Distances between trees

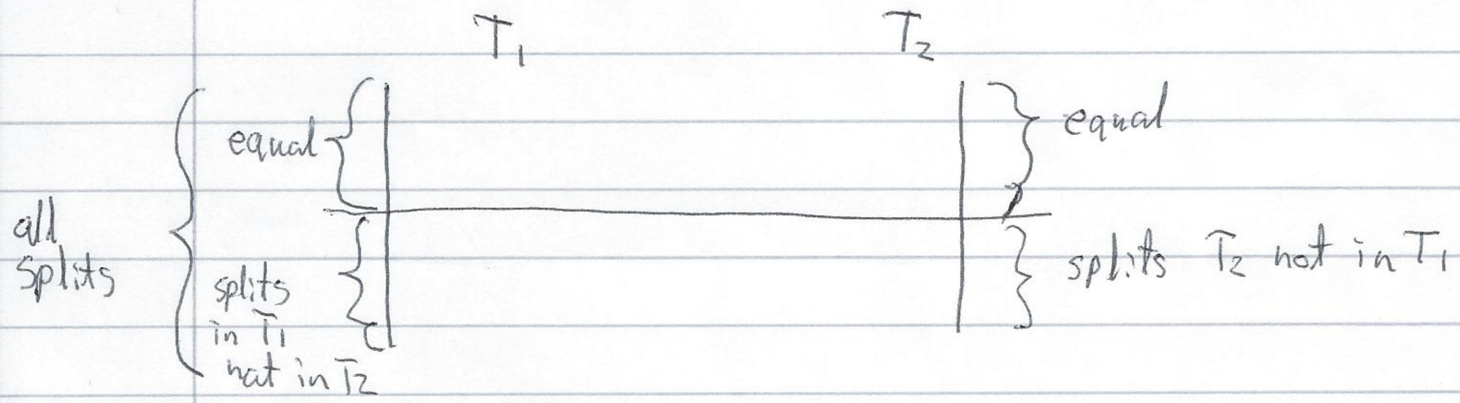
Split distance



$U|V$

Leaf edges are trivial (always exist in other tree)
Internal edges are non-trivial

$SD(T_1, T_2)$ = splits in T_1 not in T_2
Generally $SD(T_1, T_2) \neq SD(T_2, T_1)$ however for binary trees it is equal:
Binary unrooted trees have $2n-3$ edges = splits



Robinson-Foulds distance

$$RF(T_1, T_2) = SD(T_1, T_2) + SD(T_2, T_1)$$
$$= \text{"Splits not in both trees"}$$

Sensitive for changes in a tree.

Naive

If we can find shared splits we can calculate RF

shared $= 0$

For each pair of split $U|V$ in T_1 and $U'|V'$ in T_2 $\quad O(n^2)$

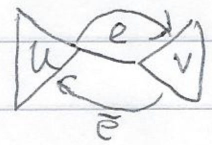
    if $U|V == U'|V'$

        shared++;

return $|T_1| + |T_2| - 2 \cdot$ shared;

Using preproccessing reducing time for $U|V == U'|V'$

$O(n^2)$ construction time. $O(1)$ query time.



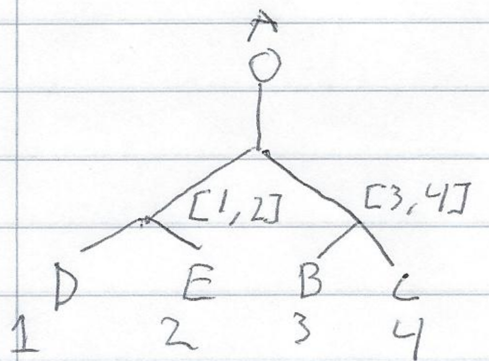
$$T(\underrightarrow{e}, \underrightarrow{e'}) = |V \cap V'|$$

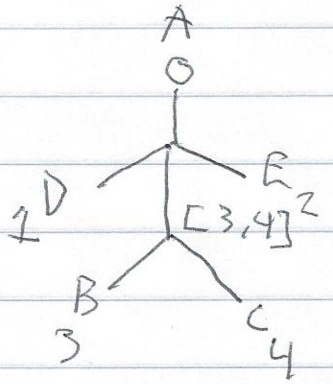
$$U|V == U'|V' \iff |U \cap U'| + |V \cap V'| = n$$

Days

1) Root by same leaf
2) Number nodes in T_1 using depth-first
3) Number nodes in T_2 according to numbering in T_1
4) Create intervals and find shared intervals using Radix sort

Takes $O(n)$





[1,2]
[3,4] $\Rightarrow$ shared $= 1$
[3,4]

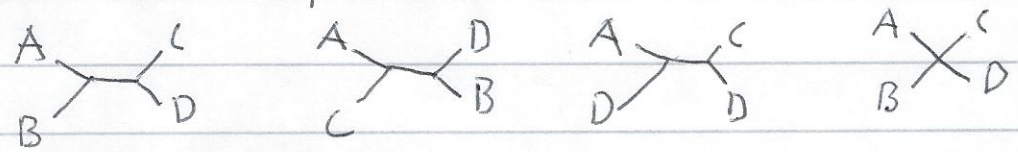
Internal splits

RF-dist $= |T_1| + |T_2| - 2 \cdot$ shared

$= 2 + 1 - 2 \cdot 1 = 1$

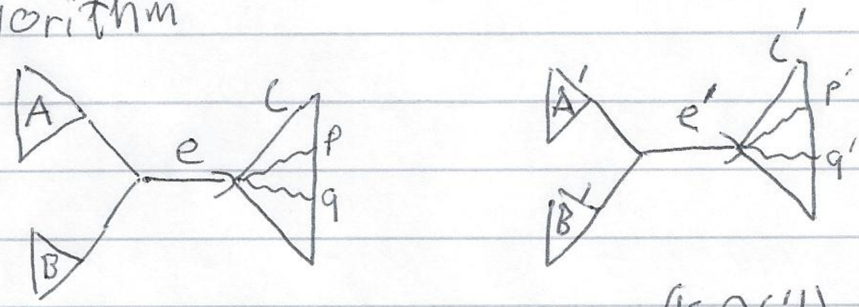
Quartet Distance

Four different quartets:



$$QD(T_1, T_2) = \binom{n}{4} - \text{shared}$$

Algorithm



$$\text{count}(\vec{e}, \vec{e}') = |A \cap A'| \cdot |B \cap B'| \cdot \binom{|C \cap C'|}{2} + |A \cap B'| \cdot |B \cap A'| \cdot \binom{|C \cap C'|}{2}$$

$$QD(T_1, T_2) = \binom{n}{4} - \frac{1}{2} \sum_{\substack{\vec{e} \in T_1 \\ \vec{e}' \in T_2}} \text{count}(\vec{e}, \vec{e}') \qquad \text{Takes } O(n^2)$$

Divide by 2 since each is counted twice: