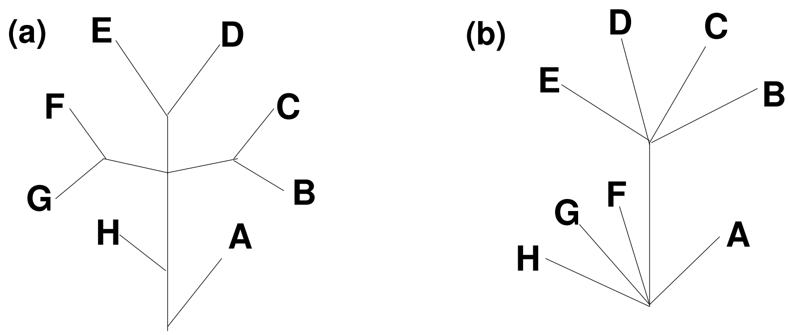
Counting multifurcating trees

In this case, we can't easily see how many placements of a new species are possible.

Example



(a) 8 leaves
6 internal nodes
14 edges

(b) 8 leaves
2 internal nodes
9 edges

New species can be inserted into existing edges and at internal nodes.

Idea

Count the trees with a given number of internal nodes:

$T_{n,m}$: number of trees on n species with m internal nodes ($m \in \{1, \ldots, n-1\}$)

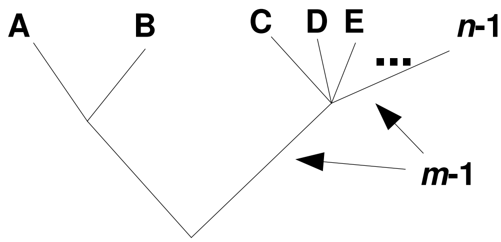
A tree counted in $T_{n,m}$ can be obtained by:

1. adding species n to an internal node of a tree in $T_{n-1,m}$
2. adding species n to an edge of a tree in $T_{n-1,m-1}$

How many possibilities are there in case 1?

Answer: there are m , one for each internal node

How many possibilities are ther in case 2?



Answer: there are $(n-1)+(m-1)$ **:** For each node (including leaves) count edge below it.

We get the following formula:

$$T_{n,m} = \begin{cases} T_{n-1,m} & \text{if } m=1 \\ m \cdot T_{n-1,m} + ((n-1)+(m-1)) \cdot T_{n-1,m-1} \\ = m \cdot T_{n-1,m} + (n+m-2) \cdot T_{n-1,m-1} & \text{if } n>m>1 \\ 0 & \text{if } m \geq n \end{cases}$$

Example

Number of multifurcating trees with 8 labels

	n = # species							
m = # internal nodes	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***
1	0	1	1	1	1	1	1	1
2	0	0	3	10	25	56	119	246
3	0	0		15	105	490	1918	6825
4	0	0						56980
5	0	0						190580
6	0	0						270270
7	0	0						135135
total								660037

Entries obtained using above formula, e.g.:

$T_{3,2} = 2 \cdot T_{2,2} + 3 \cdot T_{2,1} = 2 \cdot 0 + 3 \cdot 1 = 3$

$T_{4,2} = 2 \cdot T_{3,2} + 4 \cdot T_{3,1} = 2 \cdot 3 + 4 \cdot 1 = 7$

Chapter 4: Finding the best tree by heuristic search

Fact: One can't do so by looking at all possible trees, there are too many (see chapter 3).

Fundamental technique

- start with an estimation of the tree
- make small rearrangements of the tree to visit a „neighboring“ tree
- if we find a better tree, explore its neighbors by doing rearrangements
- repeat

Note: the maximum parsimony problem is to find the tree with the minimal number of changes or to maximize $-|\text{changes}|$ for a given set of character sequences.

Tree search

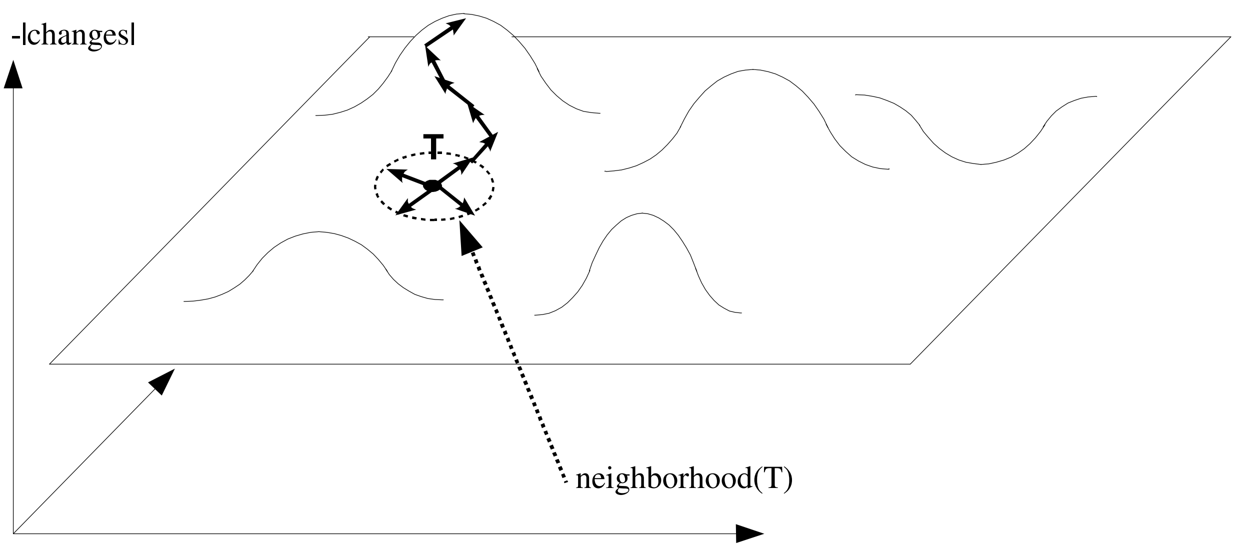


diagram of hillclimbing tree search in tree space

$$\text{neighbor}(T) = \{T' \mid T' \text{ is obtained from } T \text{ using a small rearrangement}\}$$

Hillclimbing

- evaluate the neighborhood of a tree
- move in direction that increases the value

Hillclimbing is an example of a *greedy strategy*.

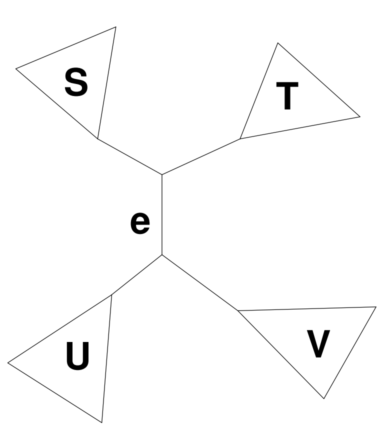
How to define the neighborhood of a tree *T*?

Moves in tree space

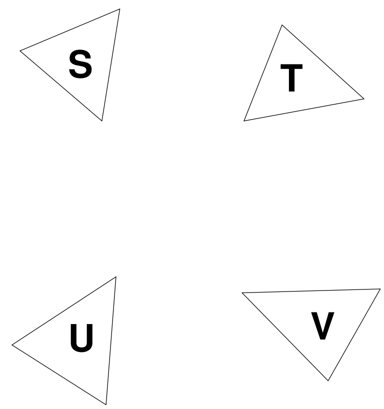
(on bifurcating trees)

1. Nearest neighbor interchange NNI

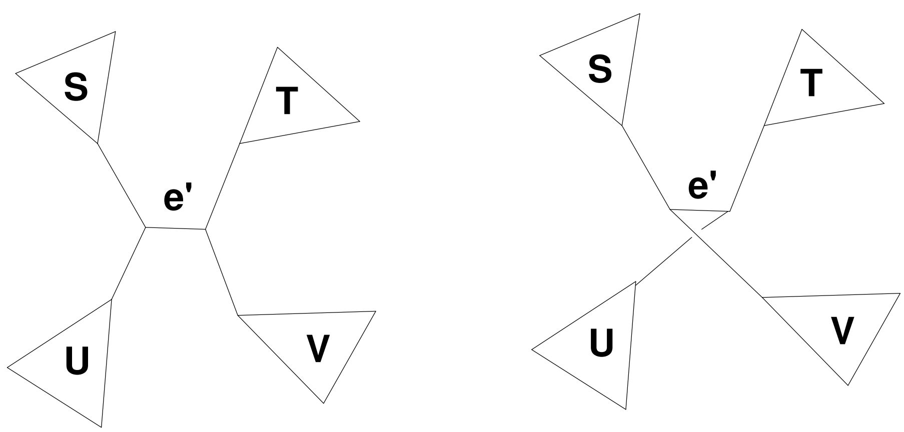
   Example



(a) choose internal edge *e*



(b) delete *e* and adjacent edges



(c) reattach

Look at tree T on n species, how large is
$\mathrm{NNI}(T) = \{T' \mid T' \text{ obtained from } T \text{ using one NNI}\}$?

Answer: NNI can be applied to any internal edge. There are $n-3$ such edges, so the neighborhood of T contains $|\mathrm{NNI}(T)| = 2 \cdot (n-3)$ trees;

e.g. $n = 20$: $|\mathrm{NNI}(T)| = 34$