Purpose:

Tumors are the result of an evolutionary process (Nowell 1976). Clonal trees are one way to describe the evolution of a tumor, and it is often helpful to visually and quantitatively compare these trees. We have implemented a visualization tool that can take two input trees and output a visual comparison between them based on a user-chosen distance metric. These distance metrics are fundamental to the comparisons and are described below.

Parent-Child

The traditional parent-child metric counts the number of parent-child relationships that are unique to exactly one of the trees (with the parent and child connected by exactly one edge). In our implementation, every parent-child pair contributes 1 to the child node if it does not appear in the other tree, and 0 otherwise.

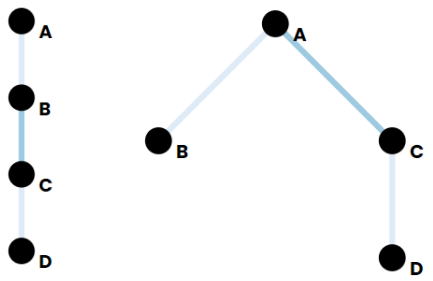


Fig 1. The visualization of the parent-child distance of these two trees show us that the edge (B, C) in tree 1 (shown on the left) and (A, C) in tree 2 (shown on the right) are the edges contributing to the difference between them

Ancestor-Descendant

The ancestor-descendant measure generalizes the parent-child measure by allowing a contributing pair of nodes to be connected by a directed path rather than an edge. In our implementation, each ancestor-descendant pair contributes 1 if it does not appear in the other tree, and 0 otherwise. The contribution of each node is determined by the number of contributing ancestor-descendant pairs it appears in.

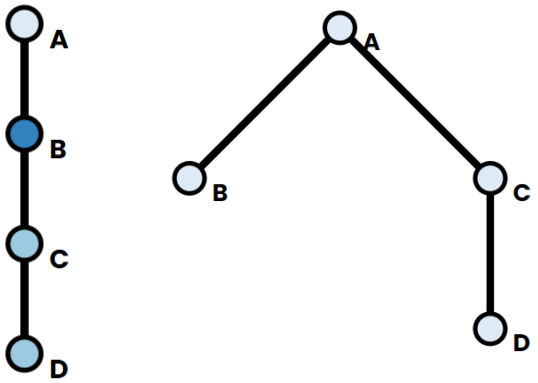


Fig 2. The above visualization shows that the ancestor-descendant relationships (B, C) and (B, D) in the left tree are responsible for the distance between the trees. In this case, B contributes 2 while C and D each contribute 1.

CASet

Overview: CASet emphasizes mutation differences closer to the root of the tree.

Details: If mutations i and j appear in a tree, the *common ancestor set* of i and j (denoted $C(i,j)$) consists of the set of mutations that are ancestors of both i and j.

The *Jaccard distance* between two sets A and B is given by

$$Jacc\,(A,\,B) \;=\; \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

In other words, $Jacc(A,B)$ yields a normalized symmetric difference between A and B.
The *CASet distance* between trees T_k and T_l is computed as follows

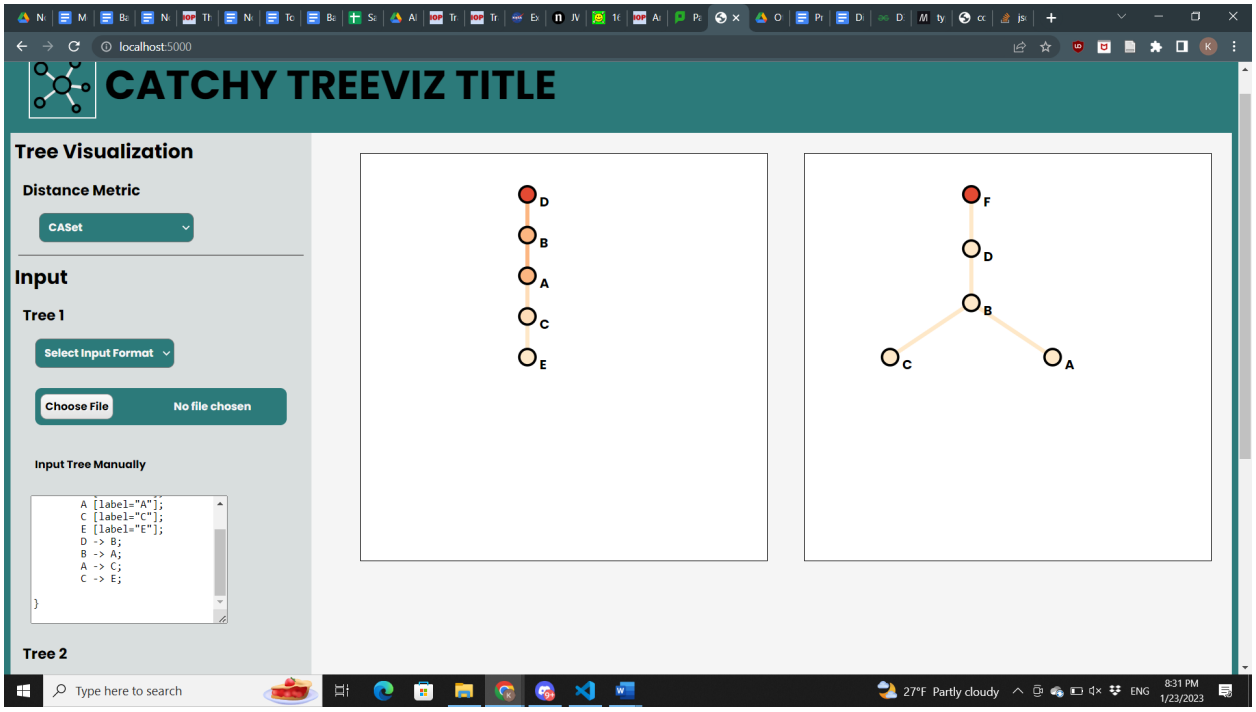
$$\mathrm{CASet}\,(T_k, T_\ell) = \frac{1}{\binom{m}{2}} \sum_{\{i,j\}\subseteq[m]} \mathrm{Jacc}\,(C_k(i,j), C_\ell(i,j)).$$

That is, the CASet distance measure looks at each pair of mutations present in the trees and computes the common ancestor sets of these mutations in each tree. It takes the Jaccard distance between the common ancestor sets in each tree, and then averages this across all mutation pairs.

From a visualization perspective, the difficulty lies in visually encoding information about the distance between two trees. The key to achieving this task lies in the observation that, much like the ancestor-descendant distance measure, CASet functionally assigns a contribution to each ancestor-descendant pair of mutations, then takes the sum of those contributions. We simply need

to extract the contributions of each ancestor-descendant pair, then depict them visually in the same way we depicted the ancestor-descendant distance.

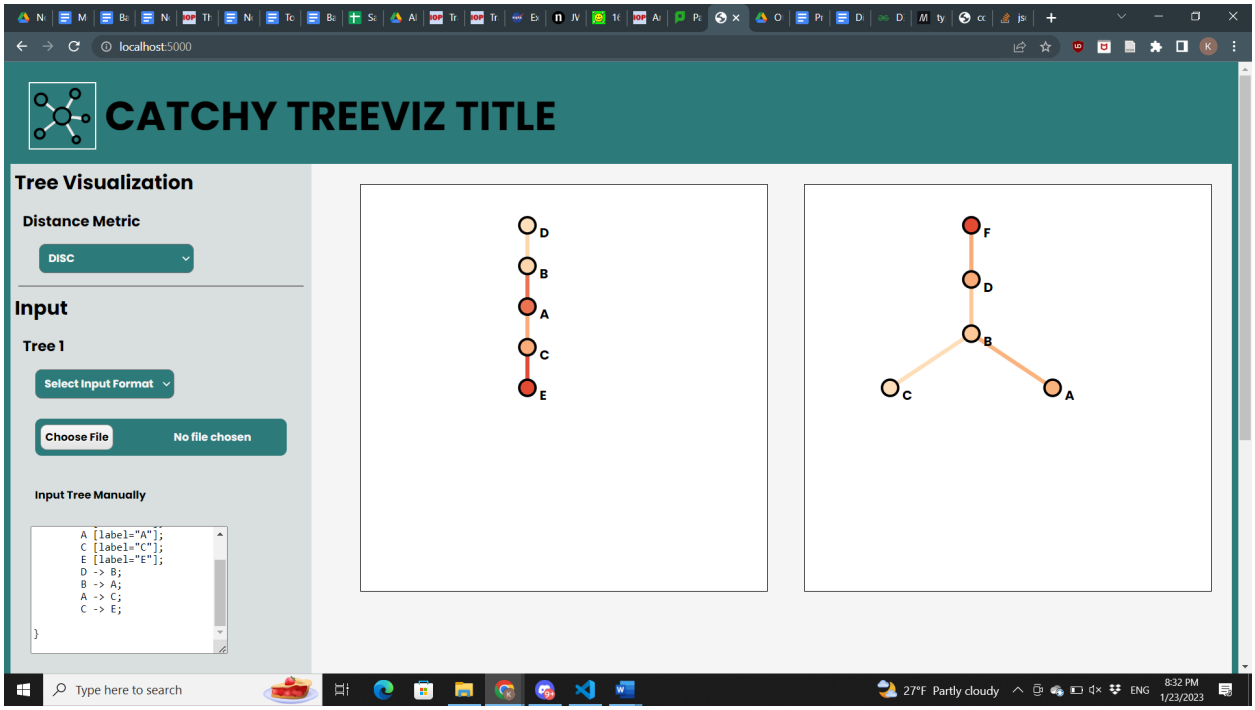
How does CASet assign contributions to ancestor-descendant pairs? If a mutation appears near the root in the left tree, and near a leaf in the right, it will appear in many common-ancestor sets of mutations in the left tree, and few on the right. Each time such a mutation appears in a common ancestor set of two mutations lower down on a tree, its ancestor-descendant relationship with each of the lower mutations contributes towards the distance measure.



DISC

If mutations A and B appear in the same tree, the distinctly inherited ancestor set of A and B consists of the set of mutations that are distinct ancestors of only either A or B. The DISC distance measure is the average Jaccard distance between all corresponding inherited ancestor sets between trees T_k and T_ℓ:

$$\mathrm{DISC}\,(T_k, T_\ell) = \frac{1}{m(m-1)} \sum_{\substack{(i,j) \in [m]^2 \\ i \neq j}} \mathrm{Jacc}\,(D_k(i,j), D_\ell(i,j)).$$


localhost:5000
CATCHY TREEVIZ TITLE
Tree Visualization
Distance Metric
DISC
Input
Tree 1
Select Input Format
Choose File
No file chosen
Input Tree Manually
A [label="A"];
C [label="C"];
E [label="E"];
D -> B;
B -> A;
A -> C;
C -> E;
}
D
B
A
C
E
F
D
B
C
A
Type here to search
27°F Partly cloudy
ENG
8:32 PM
1/23/2023