CIS 575. Introduction to Algorithm Analysis
Material for March 27, 2024

Union by Rank

The topic of this note is covered in *Cormen*'s Section 19.3.

1 Union by Rank

Recall that we have seen that UNION and FIND can be implemented to run in time linear in the height of the tree. It should not be hard to guess how to ensure that the trees we construct have reasonable height. One such way is to do *Union by Rank* where we always:

> let (the root of) the *shortest* tree be the *child* of (the root of) the *tallest* tree
> (if the trees have the same height, either root could be the new root)

You may wonder why this is not called "Union by Height" but for now, just think of "rank" as another word for "height".

It turns out that if UNION is always done **by Rank** then for any tree T constructed by a sequence of unions we will have

$$\mathbf{H}(\mathbf{T}) \leq \lg(\mathbf{N}(\mathbf{T}))$$

where $H(T)$ is the height of T, and $N(T)$ is the number of nodes in T.

Proof: We shall prove the claim by induction in the size of T. If T has just one node, then is has height zero and the claim follows since $0 = \lg(1)$.

Otherwise, assume that T is formed by taking the Union (by Rank) of the smaller trees T_1 and T_2. Inductively, we can assume

$$H(T_1) \leq \lg(N(T_1)) \text{ and } H(T_2) \leq \lg(N(T_2)).$$

We split into 3 cases:

1. If $H(T_1) > H(T_2)$, we know (from Union being by Rank) that the root of T_1 has received one more child, the root of T_2, but that will *not* increase its height. Hence
$$H(T) = H(T_1) \leq \lg(N(T_1)) < \lg(N(T)).$$
2. If $H(T_2) > H(T_1)$, we proceed as in the previous case.
3. If $H(T_1) = H(T_2)$, in which case any of the two roots may be the new root, the resulting tree will have a height that is one more. Assume that $N(T_1) \leq N(T_2)$ (the case $N(T_2) < N(T_1)$ is symmetric). Then the desired result follows from the calculation
$$H(T) = H(T_1) + 1 \leq \lg(N(T_1)) + \lg(2) = \lg(2 \cdot N(T_1)) \leq \lg(N(T))$$

We conclude that when using **Union by Rank**, FIND and UNION will both run in time $\mathbf{O}(\lg(\mathbf{n}))$.