An Interesting Tree and the Number of it's Edges

Statement:

We define two functions :

$$\pi : \mathbb{R}_{>0} \longrightarrow \mathbb{Z}$$

$$\pi(x) := \text{The number of primes in } (0, x]$$

$$p : \mathbb{Z}_{>0} \longrightarrow \mathbb{Z}$$

$$p(1) = 1$$
$$p(n) \text{ is the largest prime divisor of } n, \ n \geq 2$$

Thus, for example, $p(20) = 5$, $\pi(1) = 0$, $\pi(13) = 6$, and so on. Prove that for any $n \in \mathbb{Z}_{>0}$, we have

$$\sum_{k=1}^{n} \max(\pi(\frac{n}{k}) - \pi(p(k)) + 1, 0) = n$$

First Proof:

Consider a tree with n vertices, containing the numbers 1, 2, ..., n amongst them. The tree satisfies these properties :

- The root of the tree is the vertex containing 1.

- The parent of the vertex containing the number $m > 1$ is the vertex containing the number $m/p(m)$.

Since $\frac{m}{p(m)} < m \ \forall m > 1$, we can iteratively build a unique tree with the above rules.

Then for any vertex containing the number $k > 1$, note that it's child (assuming it has one) can have a maximum *possible* value of kq, where q is the largest prime $\leq n/k$, while the minimum *possible* value of it's child is $kp(k)$ (note that if $\exists t$ a child of k such that $t = kt' < kp(k)$, where t' is a prime, then the largest prime divisor of t is $p(k)$, which contradicts the definition that the ratio of a child to it's parent must be the greatest prime divisor of the child).

Thus the number of children k has (assuming it has at-least one child), is the

number of primes between $p(k)$ and q, which is $\pi(\frac{n}{k}) - \pi(p(k)) + 1$. Since some vertices don't have children (such as say $k = n-1$ for $n > 2$), we can summarize the number of children as $\max(\pi(\frac{n}{k}) - \pi(p(k)) + 1, 0)$; $k > 1$.

Finally, for $k = 1$, we note that it has $\pi(n)$ children (all the primes between 2 and n are the children of 1), while the expression $\max(\pi(\frac{n}{k}) - \pi(p(k)) + 1, 0)$ evaluates to $1 + \pi(n)$ at $k = 1$, ie:- over-counts by 1 (note that this exception arises merely due to the definition of $\pi(p(1))$ as 0. We can get the exact count by redefining $\pi(p(1))$ to be 1).

Hence,

$$\sum_{k=1}^{n} \max(\pi(\frac{n}{k}) - \pi(p(k)) + 1, 0)$$

$$= 1 + \text{Total number of children of all the vertices in the tree}$$

$$= 1 + \text{Total number of edges in the tree} = 1 + (n - 1) = n$$

since the number of edges in a tree with n vertices is $n - 1$.

Hence proved.

Second Proof:

We prove the statement by induction. To that end, let $S(n) := \sum_{k=1}^{n} \max(\pi(\frac{n}{k}) - \pi(p(k)) + 1, 0)$. We have to prove that $S(n) = n \; \forall n \in \mathbb{Z}_{>0}$.

We can verify that $S(1)$ is indeed 1. Assume that $S(n) = n$ for some $n \geq 1$, and consider $S(n + 1)$. Then,

$$S(n + 1) = \sum_{k=1}^{n+1} \max(\pi(\frac{n+1}{k}) - \pi(p(k)) + 1, 0)$$

$$= \max(\pi(1) - \pi(p(n+1)) + 1, 0) + \sum_{k=1}^{n} \max(\pi(\frac{n+1}{k}) - \pi(p(k)) + 1, 0)$$

Since $n + 1 \geq 2$, we have $p(n+1) \geq 2 \implies \pi(p(n+1)) \geq 2 \implies \pi(1) - \pi(p(n+1)) + 1 = 1 - \pi(p(n+1)) \leq 0 \implies \max(\pi(1) - \pi(p(n+1)) + 1, 0) = 0$

Thus,

$$S(n + 1) = \sum_{k=1}^{n} \max(\pi(\frac{n+1}{k}) - \pi(p(k)) + 1, 0)$$

$$= \sum_{k=1}^{n} \max(\pi(\frac{n+1}{k}) - \pi(\frac{n}{k}) + \pi(\frac{n}{k}) - \pi(p(k)) + 1, 0)$$

Before proceeding, observe that the term $\pi(\frac{n+1}{k}) - \pi(\frac{n}{k})$ is always non-negative (but atmost 1, as will be shown in the following paragraphs), but for it to actually contribute to $S(n + 1) - S(n)$ (which we desire to show to be 1), it actually has to be positive. But being positive alone is not guarantee enough that $\pi(\frac{n+1}{k}) - \pi(\frac{n}{k})$ will contribute : It also has to be the case that $(\pi(\frac{n}{k}) - \pi(p(k)) + 1)$ **is also non-negative**, as otherwise, since a negative $(\pi(\frac{n}{k}) - \pi(p(k)) + 1)$ would imply that it was atmost -1, and $\pi(\frac{n+1}{k}) - \pi(\frac{n}{k})$ is anyways

2

atmost 1 (to be shown below), that would effectively mean that the $\pi(\frac{n+1}{k}) - \pi(\frac{n}{k})$ term had no contribution at all, due to the maximum clause.

In that light, note that since $\frac{n+1}{k} - \frac{n}{k} \leq 1 \ \forall \ 1 \leq k$, and also since $\nexists t \in (\frac{n}{k}, \frac{n+1}{k})$ (as that would imply $\mathbb{Z} \ni kt \in (n, n+1)$ which is absurd), we have that either both $\frac{n}{k}$ and $\frac{n+1}{k}$ lie completely between two integers, or that one of them is an integer itself (and the other isn't). In any case $\pi(\frac{n+1}{k}) - \pi(\frac{n}{k})$ is atmost 1, and it becomes 1 only when $\frac{n+1}{k}$ is an integer prime.

To that end let $k = \frac{n+1}{q}$, where q is a prime divisor of $n + 1$. Also note that $\max(\pi(\frac{n}{k}) - \pi(p(k)) + 1, 0) = \max(\pi(q) - \pi(p(k)), 0)$ (we wish to ascertain the non-negativity of $(\pi(\frac{n}{k}) - \pi(p(k)) + 1)$).

Now note that the above expression is non-negative if and only if $q = p(n + 1)$, since if q were some other prime divisor of $n + 1$, then $p(k) = p(n + 1)$ and consequently $\pi(q) - \pi(p(k)) \leq -1$.

Thus, for all those k for which $\pi(\frac{n+1}{k}) - \pi(\frac{n}{k})$ actually can contribute to the sum, for exactly one k (ie:- for $k = \frac{n+1}{p(n+1)}$) is it actually allowed to do so.

Thus, we have $S(n + 1) = S(n) + 1 \ \forall \ n \geq 1$, and combined with the fact that $S(1) = 1$, we conclude that $S(n) = n \ \forall \ n \in \mathbb{Z}_{>0}$.

Hence proved.

The Motivation behind this Problem:- Let $S_n = \{1, 2, ..., n\}$. Consider the *strict* divisibility poset $(S_n, |)$, ie:- for any two elements $x \neq y \in S_n$ we have $x \prec y$ iff $x | y$. Consider the Hasse diagram of this poset, and for each vertex $k > 1$ keep only it's smallest (numerically) parent, ie:- delete the edges of that node to it's other parents. Thus, for example, in the Hasse diagram of $(S_6, |)$, both 2 and 3 would be parents of 6, but now we keep only the 2-6 edge and delete the 3-6 one since 2 is it's smallest parent. The diagram then becomes a tree where each non-root vertex has a unique parent. That tree is the one we have described and analysed in the problem.

Some more Interesting Observations:-

We would also like to point out some very interesting properties of the tree defined in the first proof, which aren't necessarily limited only to the context of this theorem. In that vein, we present the paragraphs below.

Let the depth of any vertex in a tree be defined as the length of the path from the root of the tree to that vertex. Thus, the depth of the root is zero, and so on.

By the nature of the tree, it's fairly easy to see that if a vertex contains the number k, where $k = \prod_{i=1}^{\omega(k)} p_i^{l_i}$ is the prime factorization of the number k, then the depth of that vertex is $\sum_{i=1}^{\omega(k)} l_i$.

The largest depth in the tree is that of the vertex containing 2^l, where 2^l is the largest power of $2 \leq n$. Thus the maximum depth of the tree is $\lfloor \log_2 n \rfloor = \Theta(\log n)$.

One also notices that the tree rooted at any node is **similar** to the entire tree itself due to the fact that if the elements in that sub-tree are divide by their root, then one in fact re-obtains a smaller version of the original tree itself (smaller because only primes $\leq n/k$ were available to it). Our tree, thus, is self-similar to some extent.