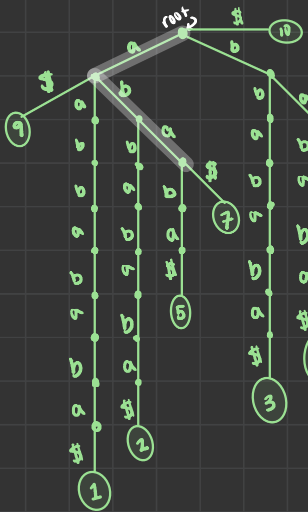
Suffix trees

- input: text $t: t_1, t_2, \dots, t_n$; $t_i \in A$
- output: suffix tree for T

  $O(n)$ space , $O(n^2)$ time

(DFA-like structure)

let's make a tree of all suffixes of T:



ex) $A = \{a, b\}$

end symbol

T = aabbababa\$

suffixes (T):
1. aabbababa\$
2. abbababa\$
3. bbababa\$
4. bababa\$
5. ababa\$
6. baba\$
7. aba\$
8. ba\$
9. a\$
10. \$

$$1 + 2 + \dots + n-1 + n = \frac{n(n+1)}{2} = O(n^2)$$

☆ if you want to know if pattern "aba" is in T:

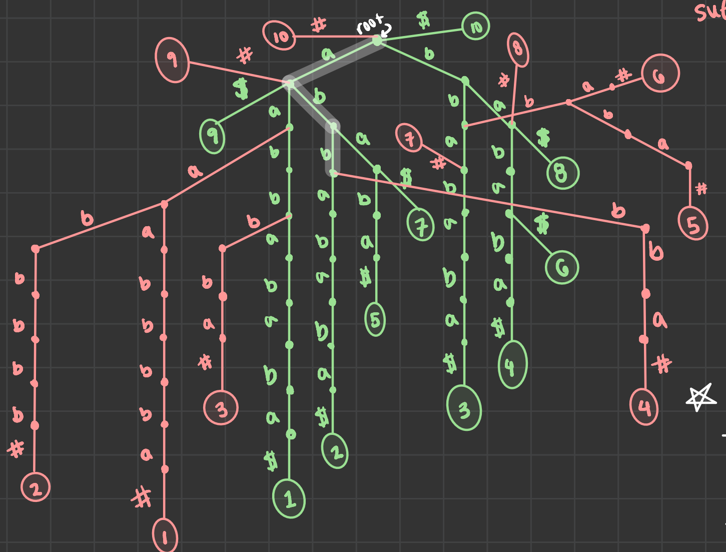
▷ starting at root, see if you can follow edges spelling the pattern (highlighted in white)

→ follow the paths to find the indices at which p occurs in T (5 and 7)

introduce a second string: T' = aaaabbbba

suffixes(T'):
1. aaaabbbba#
2. aaabbbba#
3. aabbbba#
4. abbbba#
5. bbbba#
6. bbba#
7. bba#
8. ba#
9. a#
10. #

add them to the tree



☆ is pattern "aab" in both texts?

→ follow "abb" edges from root (highlighted in white)

→ there is a green path and red path extending from "aab" path, so we know "aab" is in both texts

(@ index 2 in T and 4 in T')

* it is very possible that there are mistakes in my suffix trees. If you find one, lmk!