

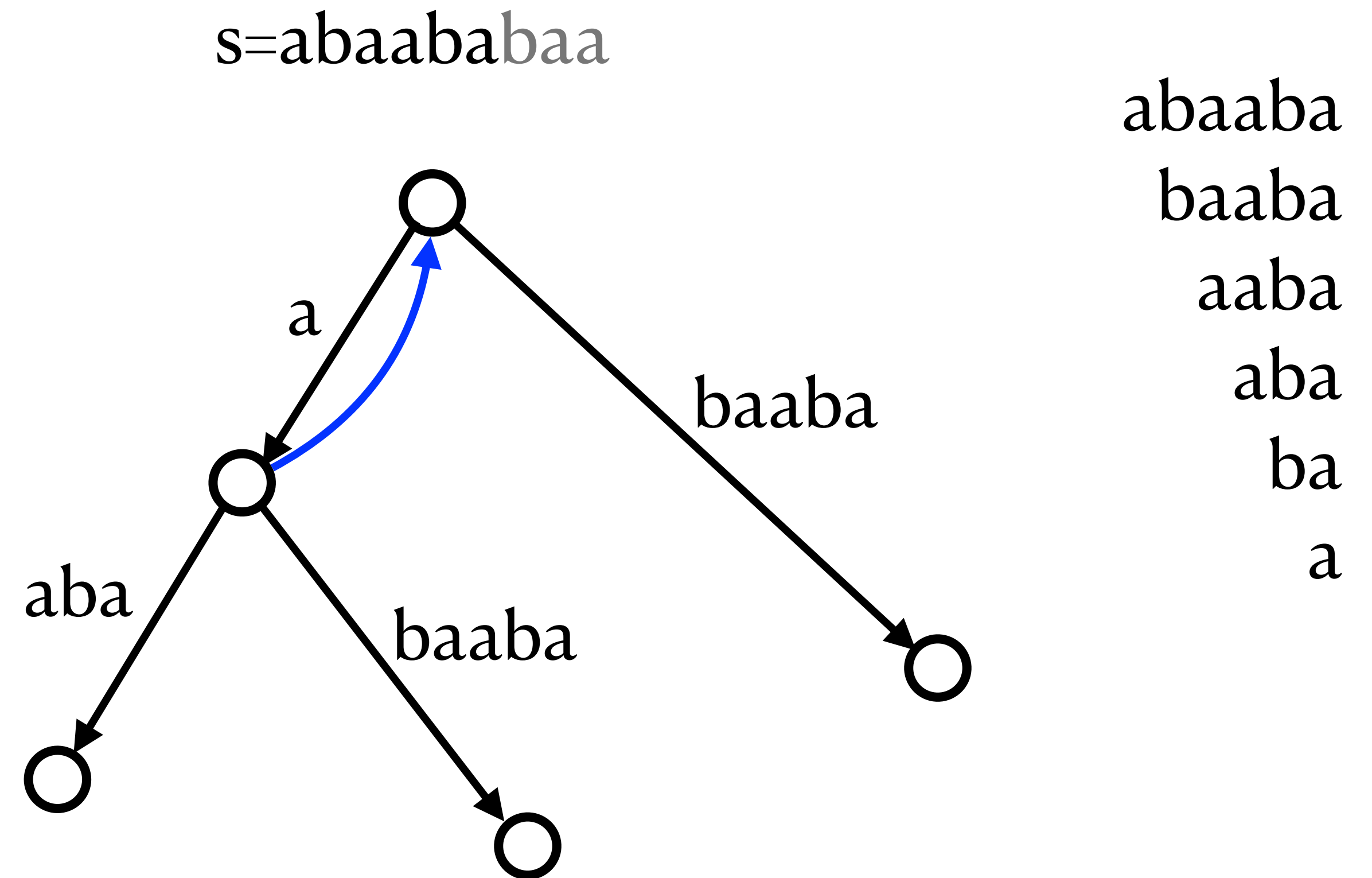
CSE 566 Spring 2023

Ukkonen's Algorithm

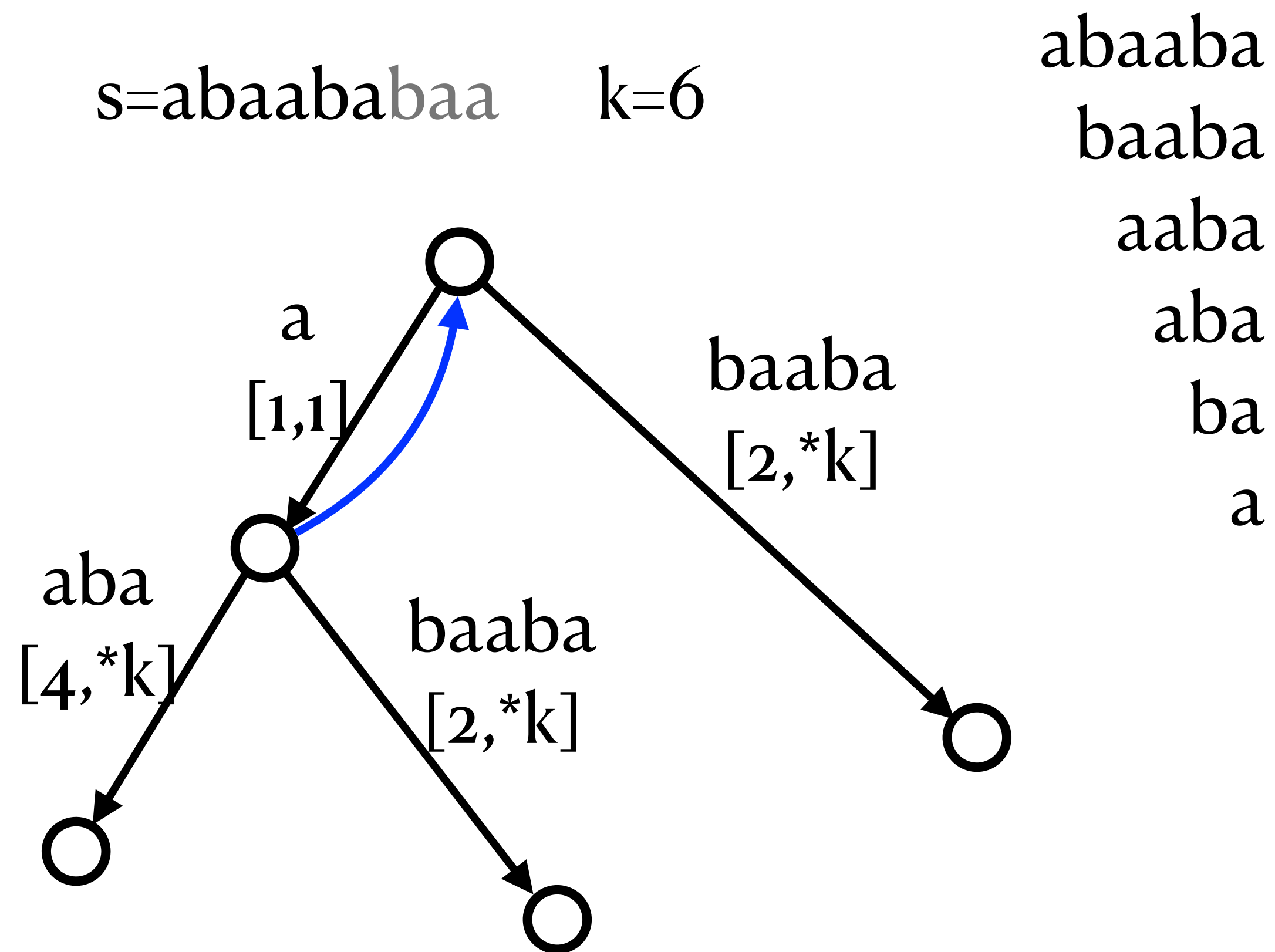
Instructor: Mingfu Shao

Ukkonen's Algorithm

- The same algorithmic **framework** with constructing suffix trie: constructing the tree for $s[1..k]$ given the tree for $s[1..k-1]$.
- The **implicit suffix tree**: each suffix can be spelled out from the root, but may not end up in a leaf
- Throughout the Ukkonen's algorithm, only internal nodes have suffix links



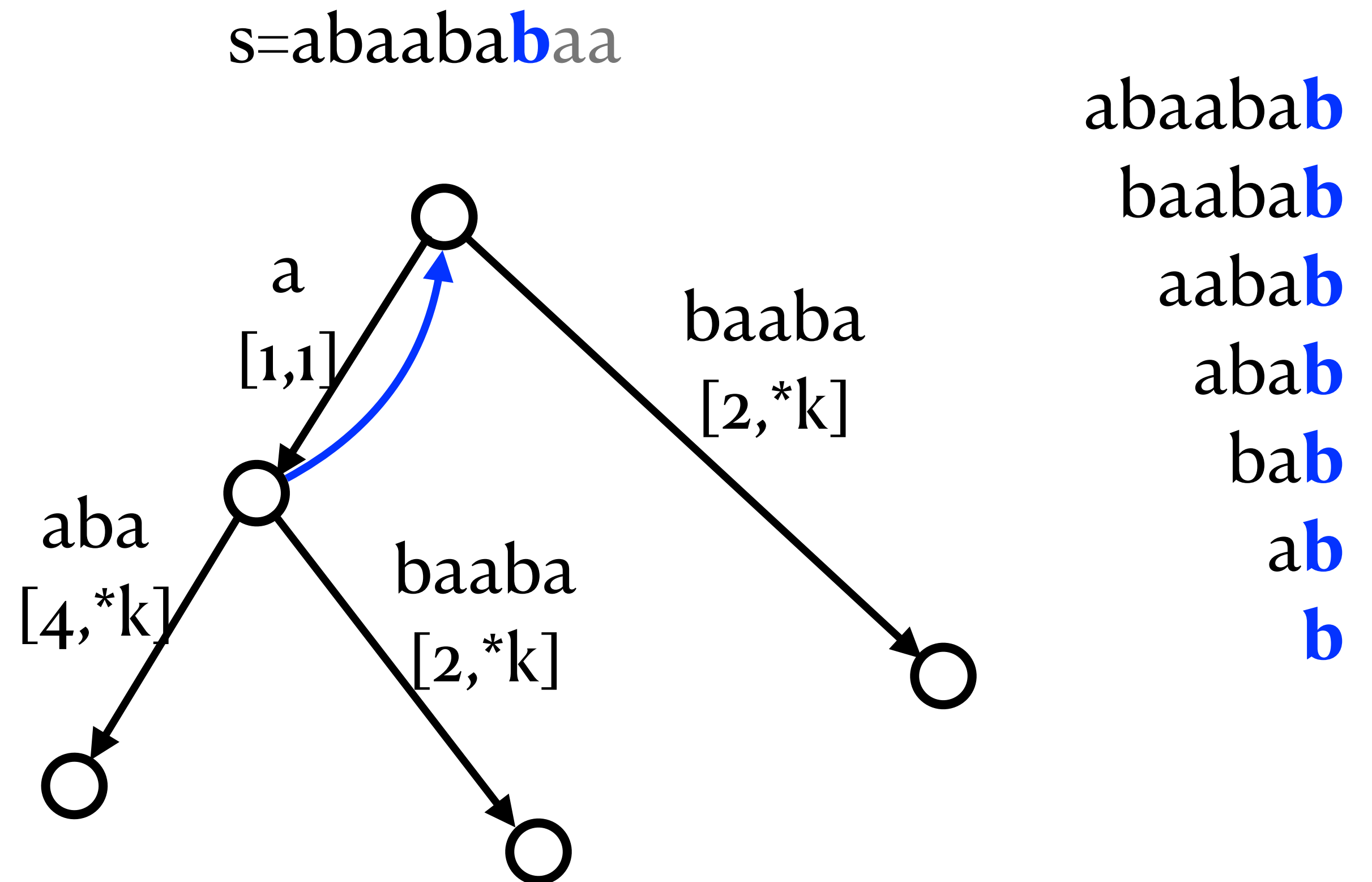
Using Indices to Label Edges



- Leaf nodes will be labeled like $[4,*k]$, where $*k$ is a pointer points to a global variable k .
- This will allow leaf nodes for getting updated automatically!
- Also, to save space.

A Single Phase

- **Q1:** how to find the next suffix
 - Still suffix links, but more complicated
- **Q2:** how to update the tree
 - Add the current letter $s[k]$
 - Add suffix links (for internal nodes)
- **Q3:** how to jump to next phase
 - Wherever we stops at the current phase!



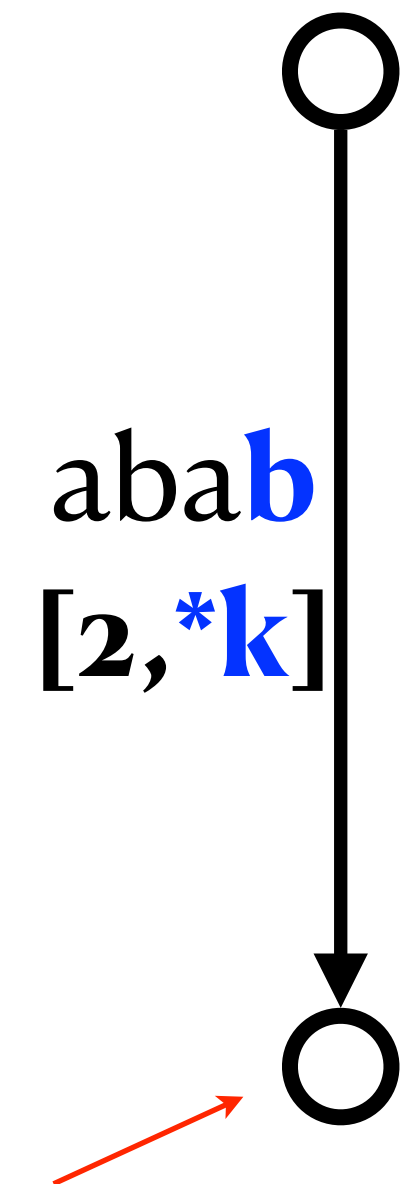
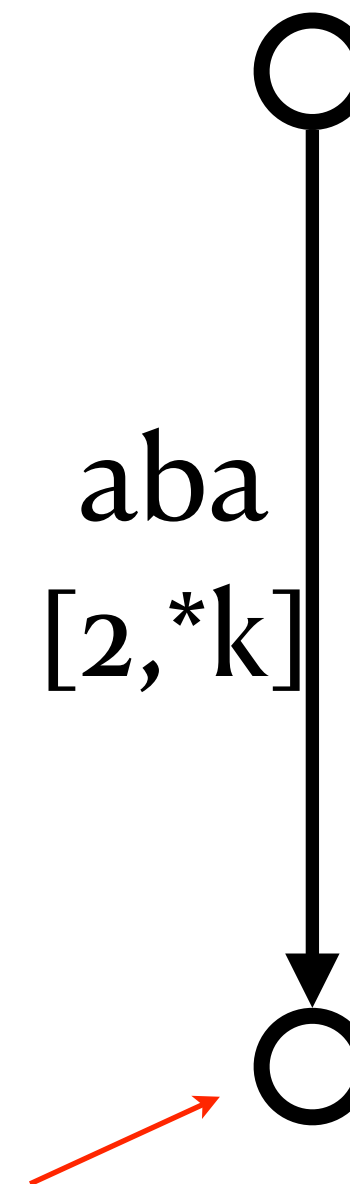
Update A suffix

- Pointer to the current suffix: node, or edge plus offset
 1. leaf node
 2. internal node, and one child labels $s[k]$
 3. internal node, but no one child labels $s[k]$
 4. edge, and next letter matches $s[k]$
 5. edge, but next letter does not match $s[k]$

Update A Suffix (Case 1)

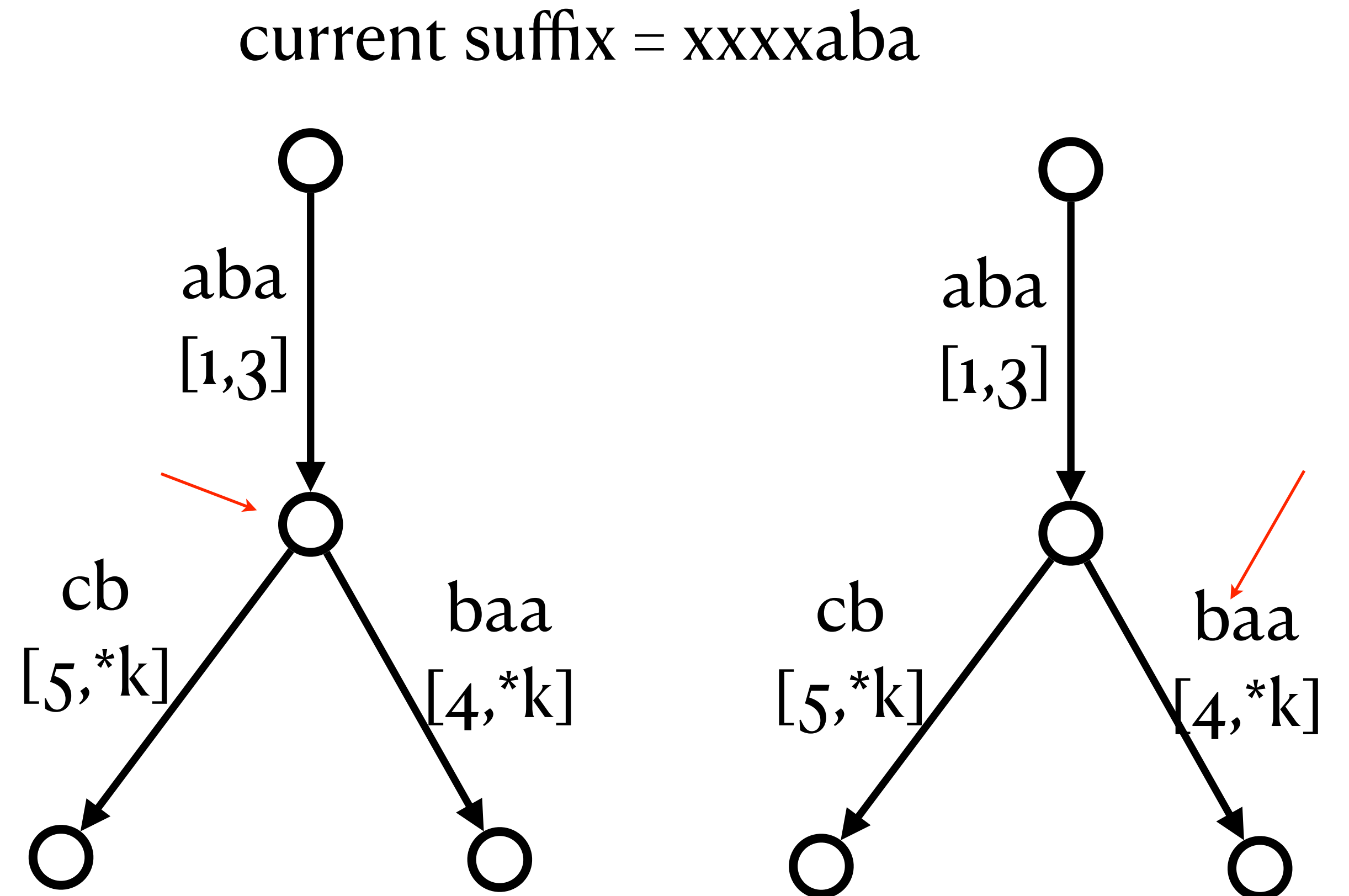
- Current suffix: leaf node, $s[k] = b$
- No action needed (in fact, this case will be taken care of automatically).

current suffix = xxxxaba



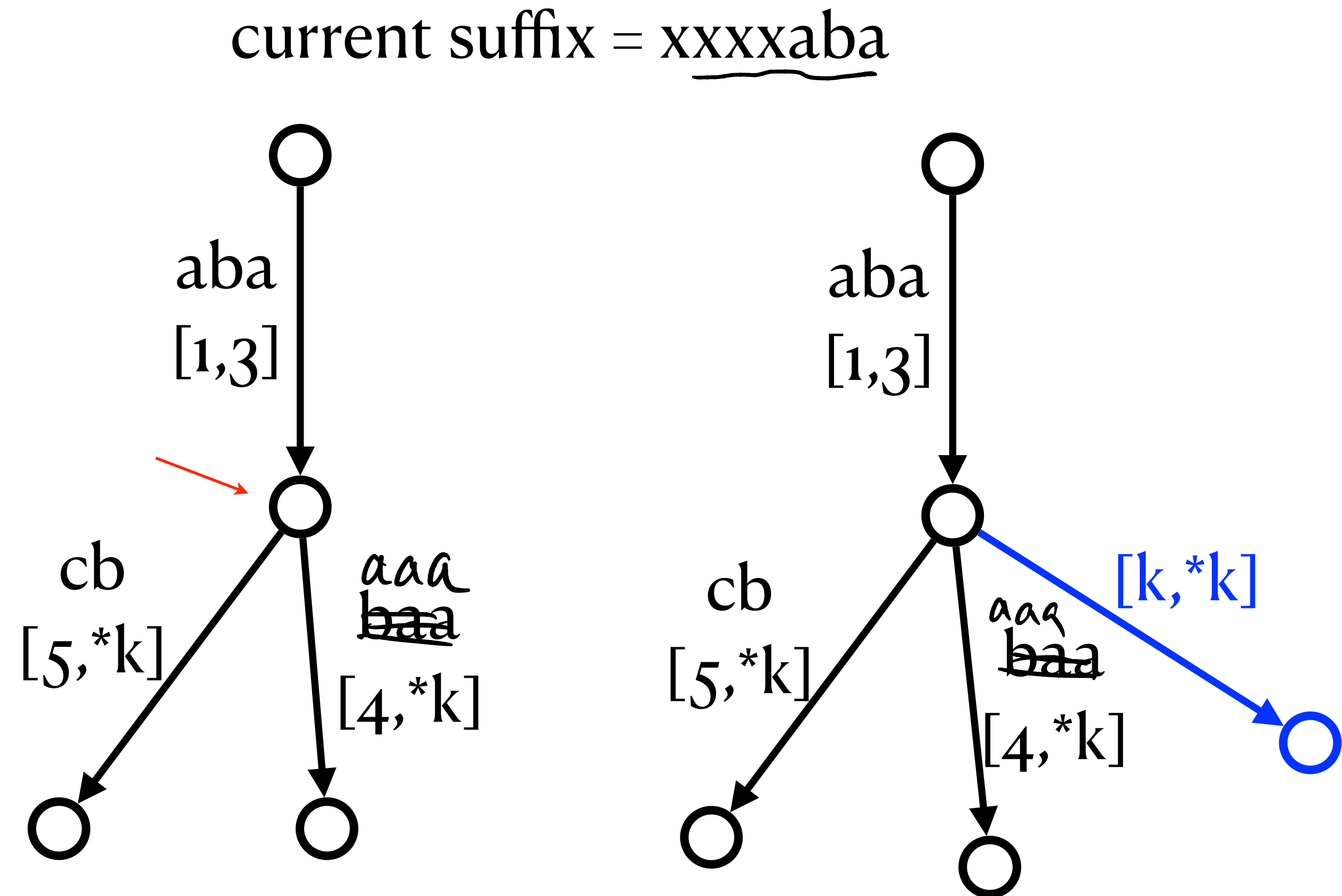
Update A Suffix (Case 2)

- Current suffix: internal node, with one child labels $s[k] = b$
- Actions:
 - This **phase terminates**.
 - Record where we stop (the starting point of next phase)



Update A Suffix (Case 3)

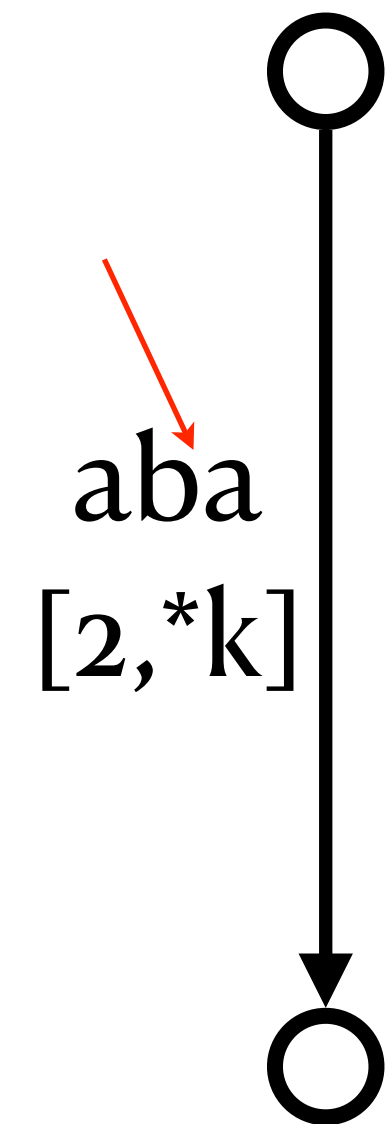
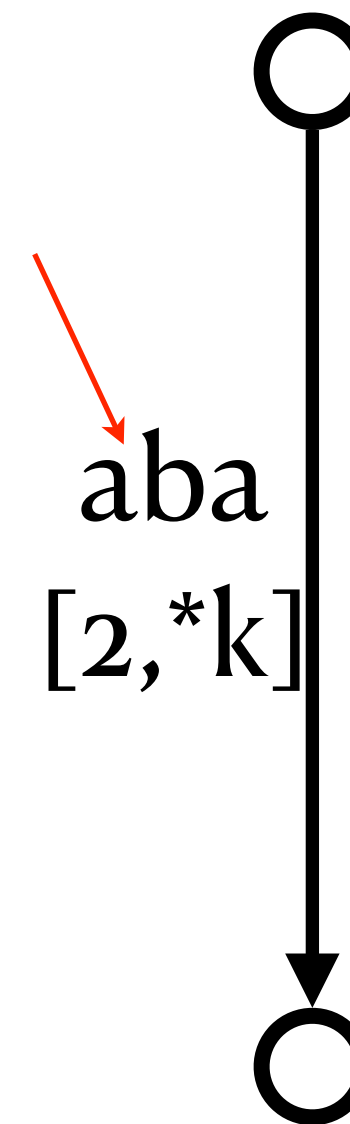
- Current suffix: internal node, but one child labels $s[k] = b$
- Actions:
 - Add a new leaf node labeled as $[k, *k]$
 - If the internal node is not the root: follow the suffix link of the internal node to the next suffix; otherwise, phase terminates.



Update A Suffix (Case 4)

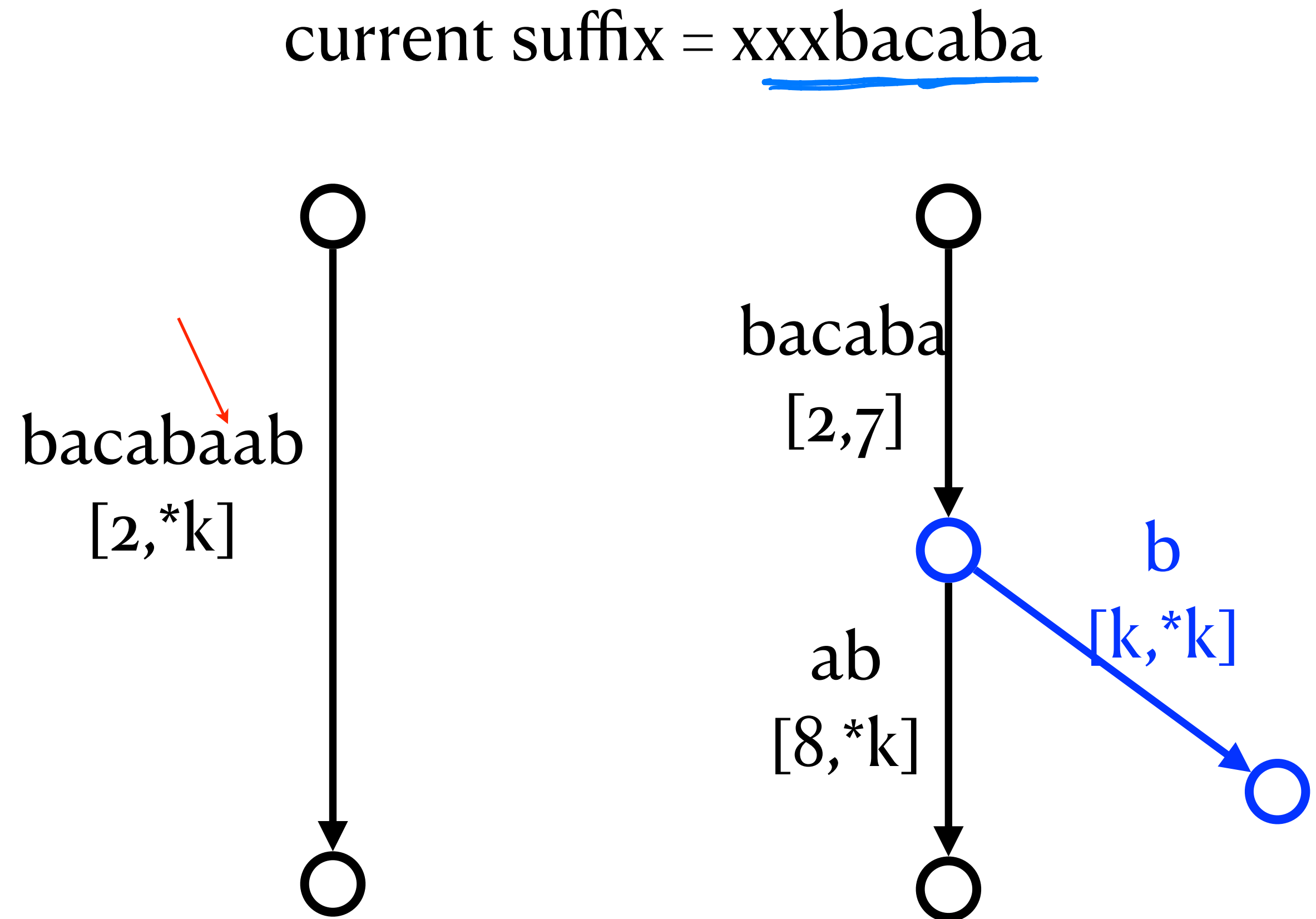
- Current suffix: edge, and the next letter matches $s[k] = b$
- Actions:
 - This **phase terminates!**
 - Record where we stop (the starting point of next phase)

current suffix = xxxxa



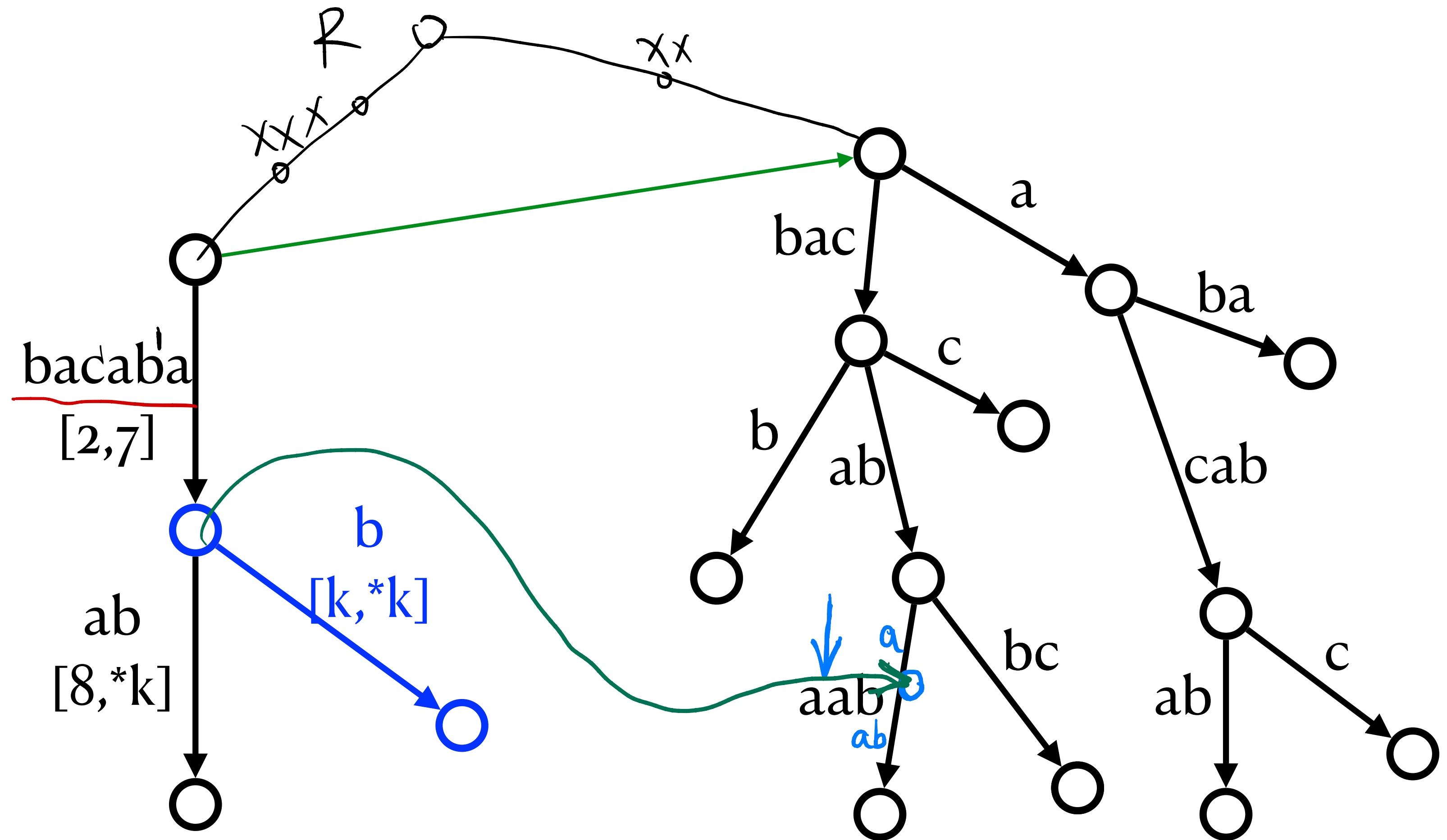
Update A Suffix (Case 5)

- Current suffix: edge, but the next letter does not match $s[k] = b$
- Actions:
 - Create a new internal node by splitting the edge
 - Create a new leaf node.
 - *Move to the next suffix and create the suffix link for the new internal node*



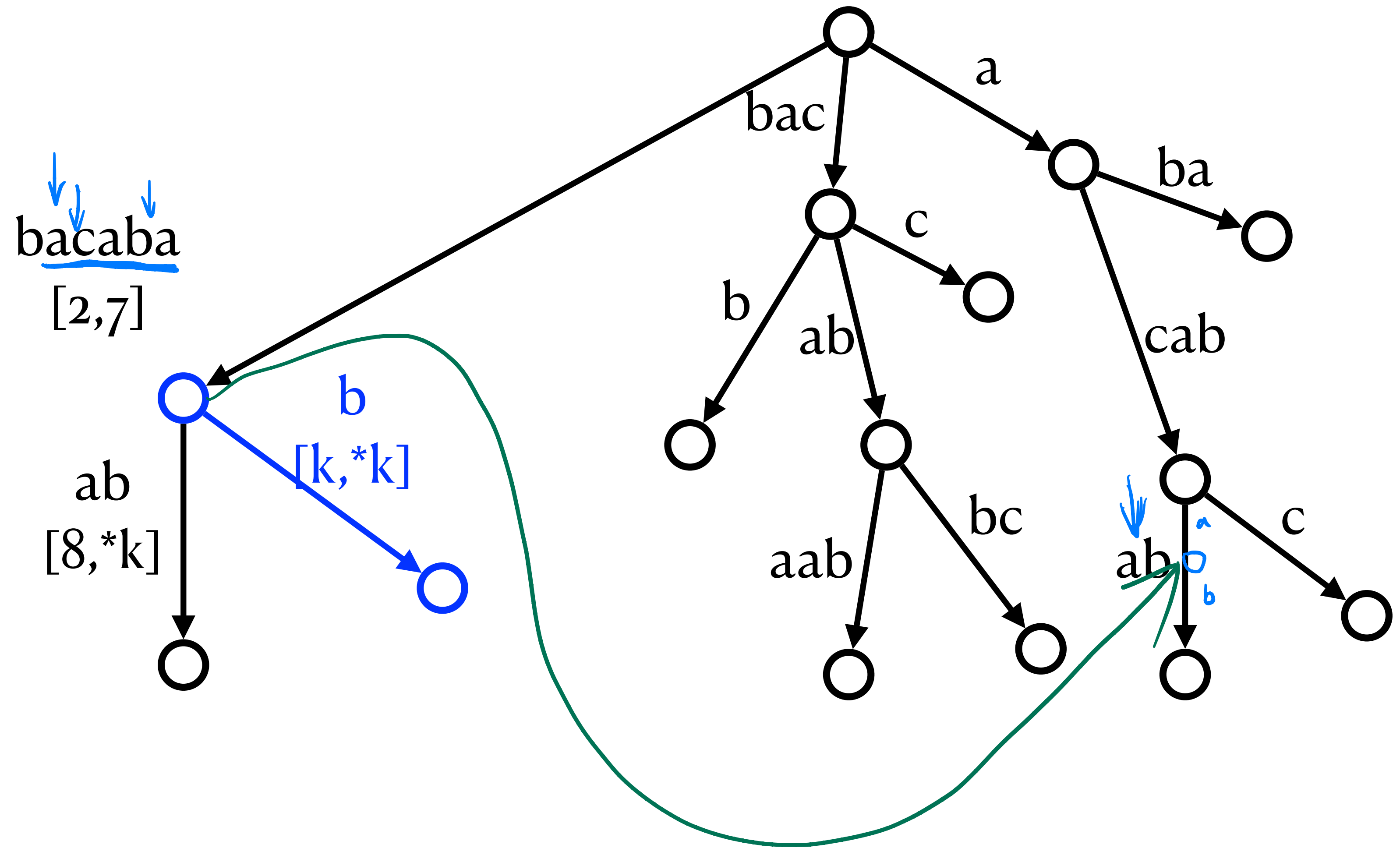
Case 5: Next Suffix and New Suffix Link

- In case the parent node is not the root:
- Follow the suffix link of the parent node
- Hop down the tree **by nodes**, by calculating the lengths as needed
- Add the suffix link
- Running time:
 $O(\text{\#nodes hopped})$

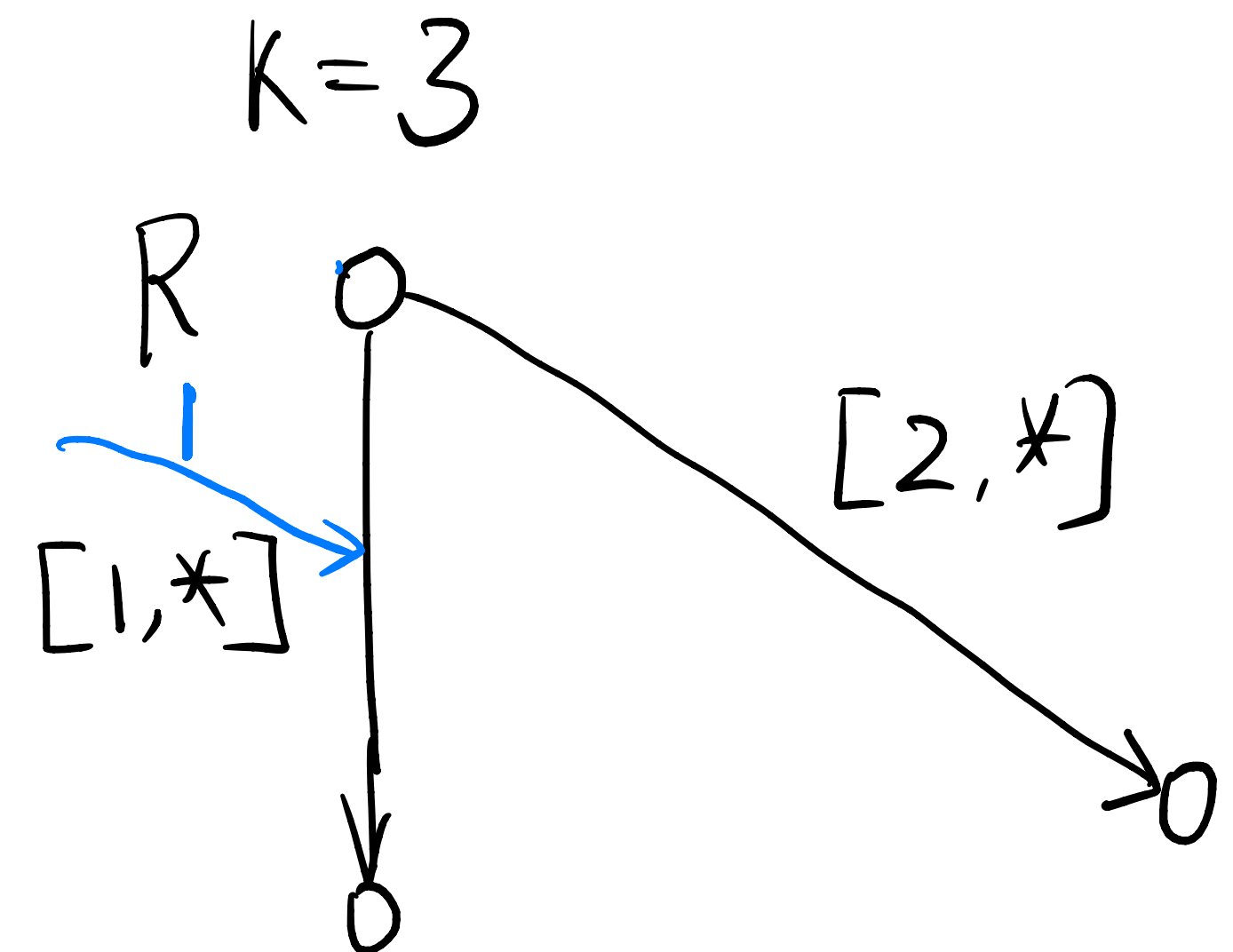
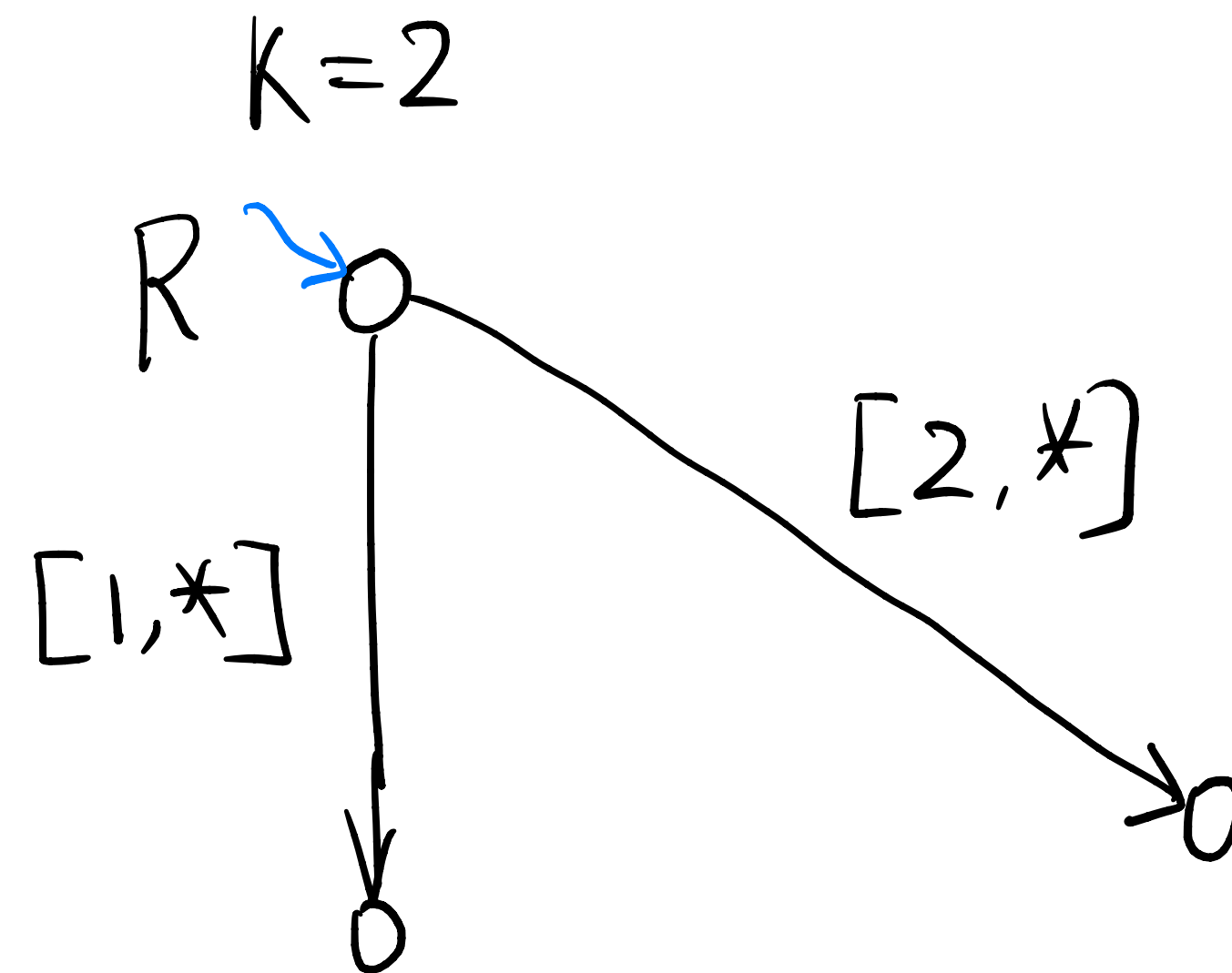
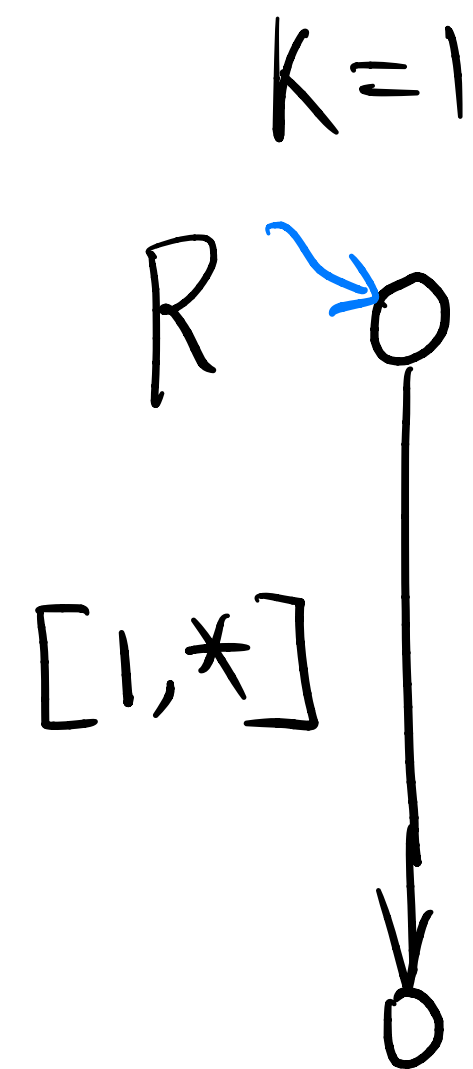


Case 5: Next Suffix and New Suffix Link

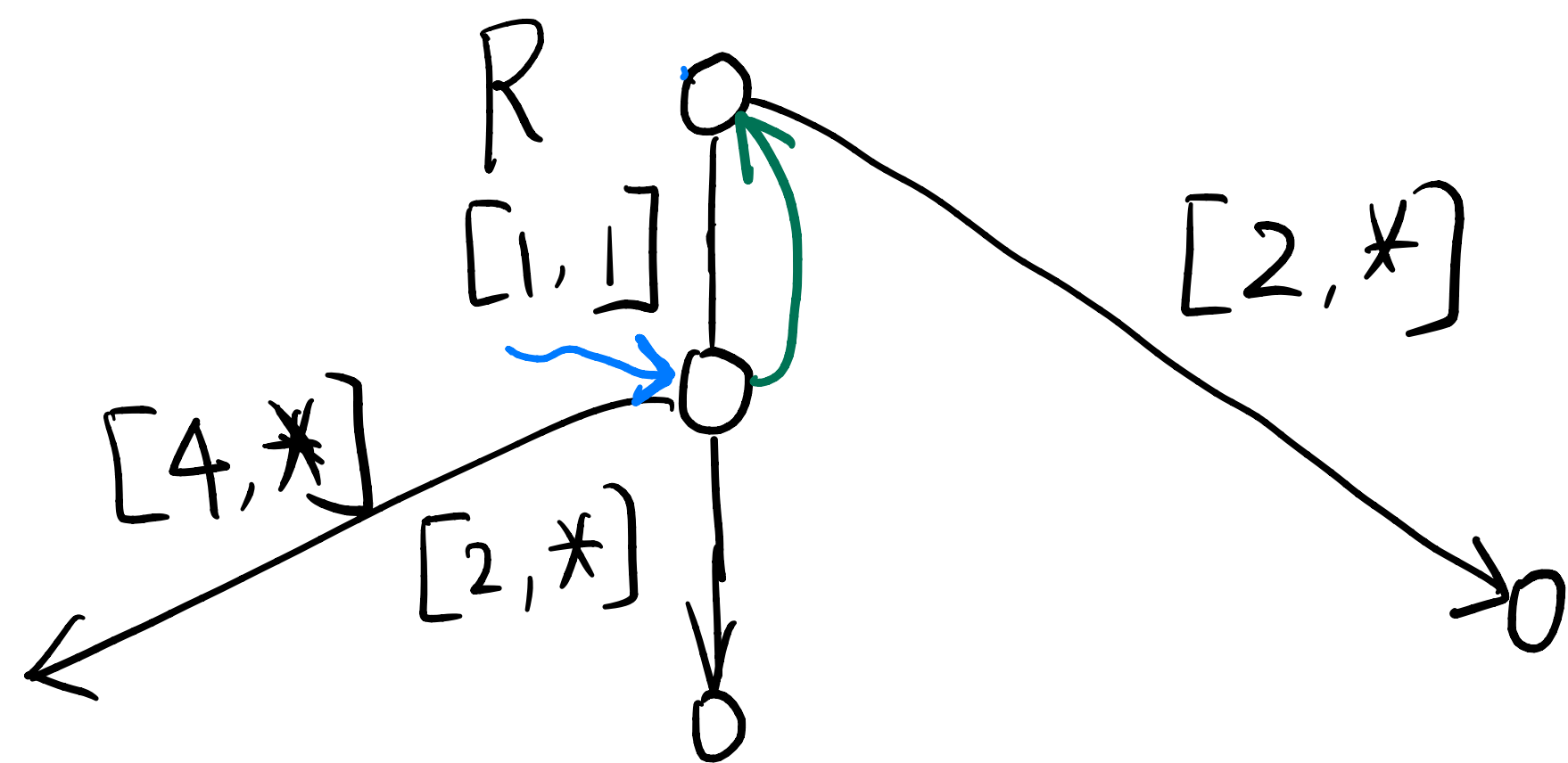
- In case the parent node is the root:
- Starting from the next letter, hop down the tree **by nodes**, by calculating the lengths as needed
- Add the suffix link
- Running time:
 $O(\# \text{nodes hopped})$



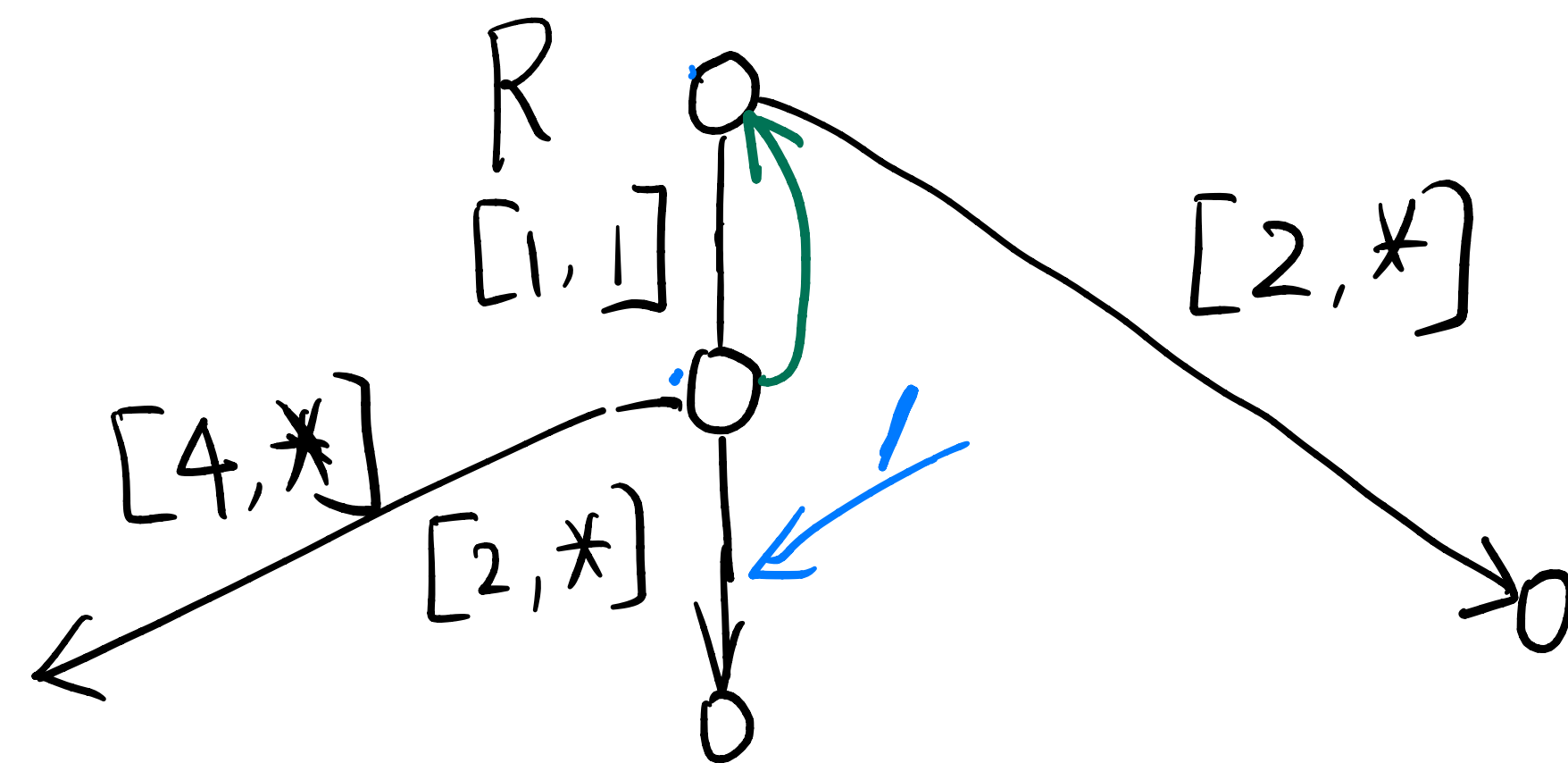
Example : $S = a b a a b b a \$$
 1 2 3 4 5 6 7 8



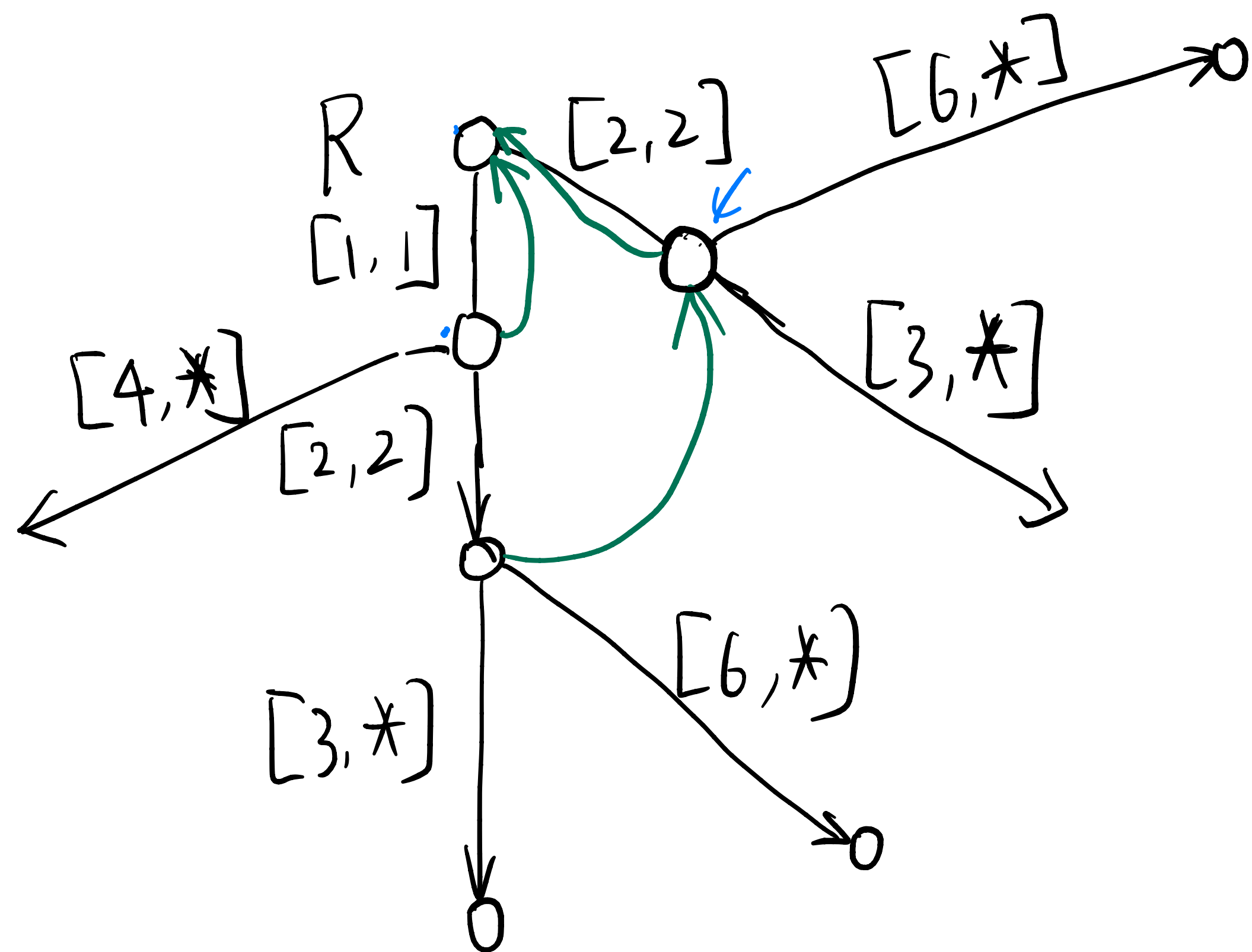
$k=4$



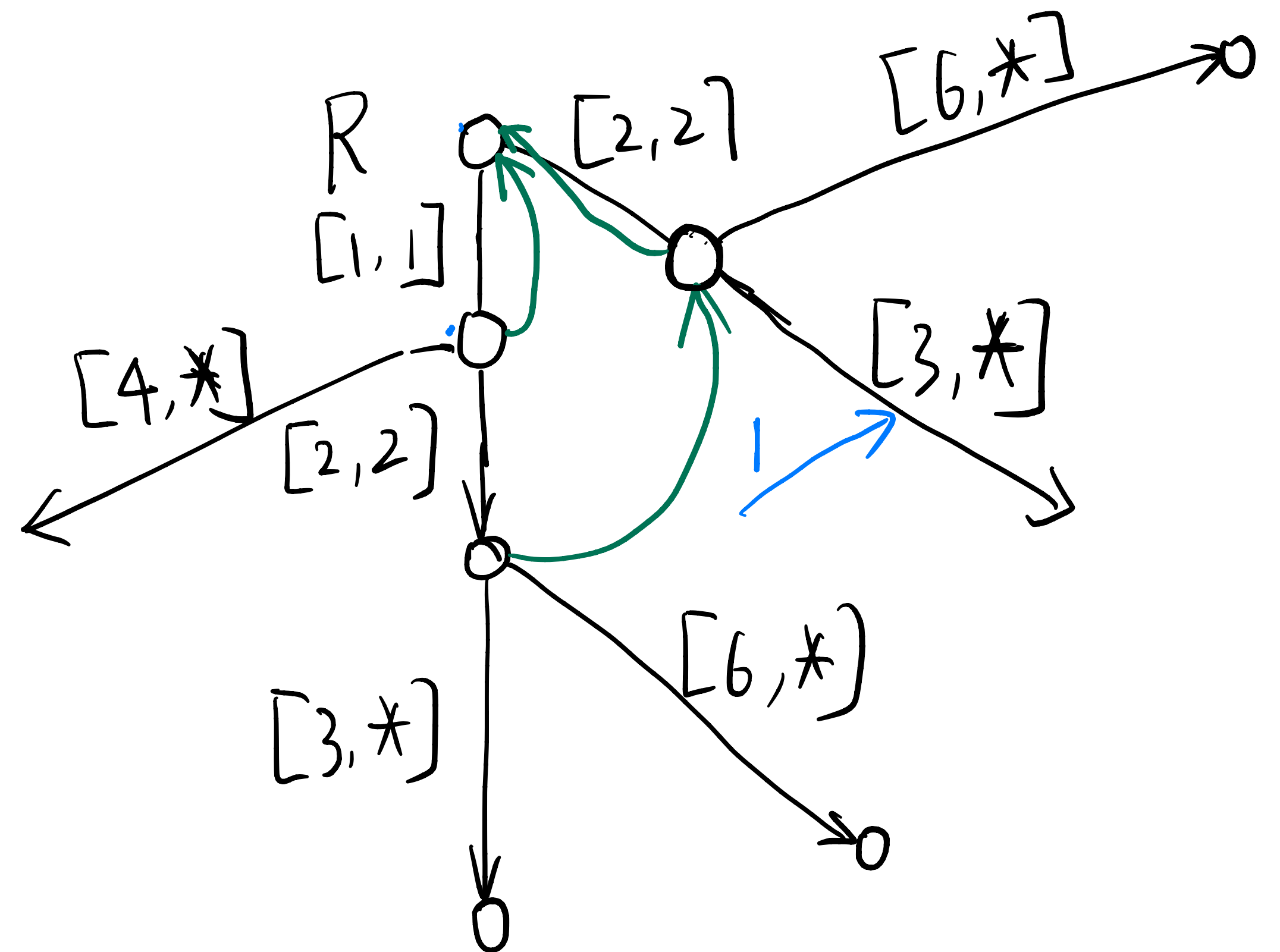
$k=5$

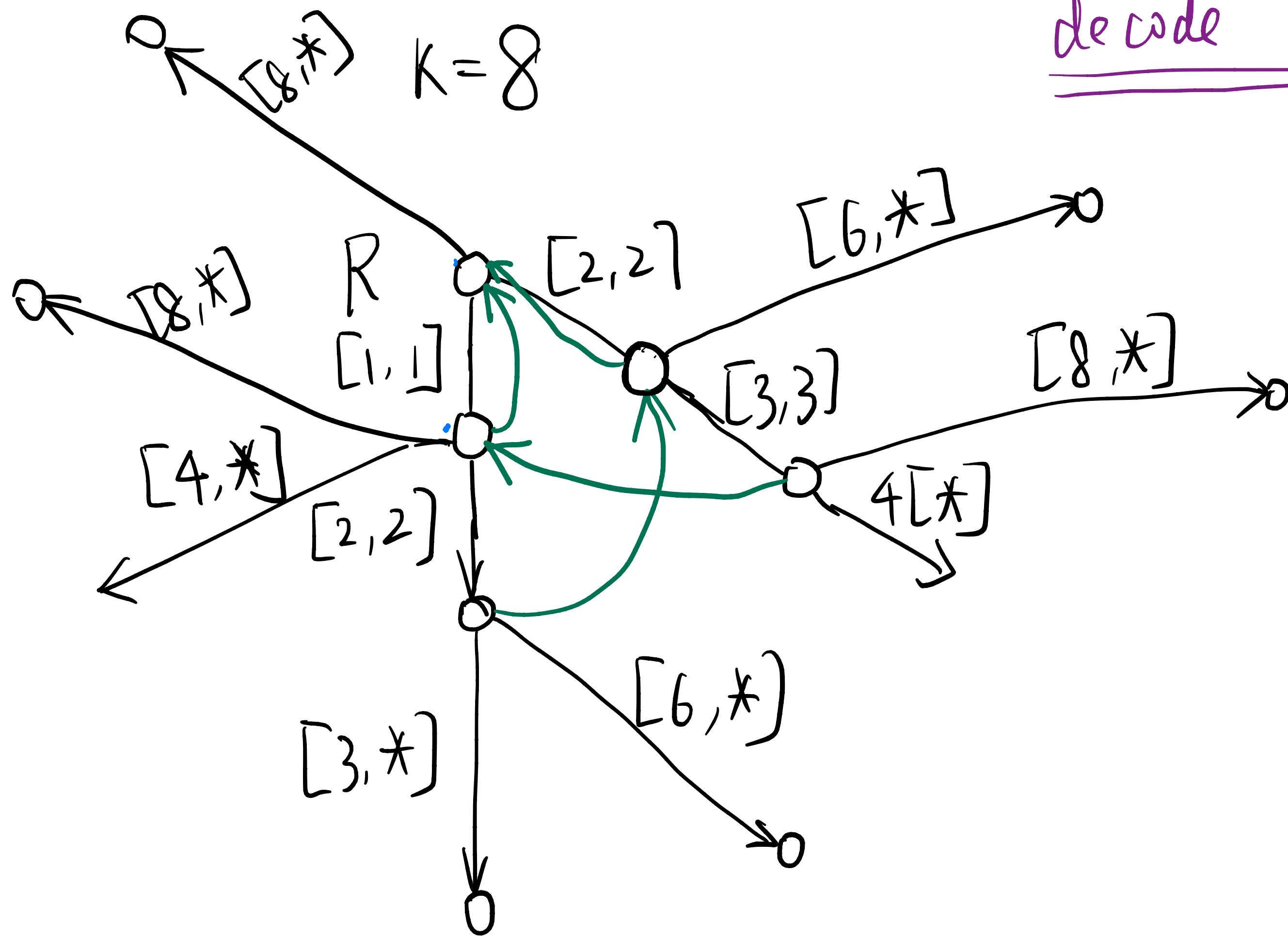


$k=6$



$k=7$





decode * →

$S = \text{abaabbba}\$$
 1 2 3 4 5 6 7 8

