



UNIVERSITY OF THE BASQUE COUNTRY

FINAL YEAR PROJECT

---

# About Tree Depth

---

*Author:*

Asier MUJICA

*Supervisor:*

Dr. Hubert CHEN

February 5, 2015

# 1 Introduction to Graph Theory

## 1.1 Definition of a graph

A graph is defined as a pair of sets  $G = (V, E)$ , such that  $E \subseteq \{\{a, b\} \mid a, b \in V\}$ . The members of  $V$  are called vertices and the ones of  $E$  edges. Take into account, that the vertices can be anything, they can even be sets themselves. The usual way to draw a graph is by representing the vertices as individual points and for each edge, draw a link between both elements of that edge. The shape in which a graph is drawn is irrelevant, it will contain the same information.

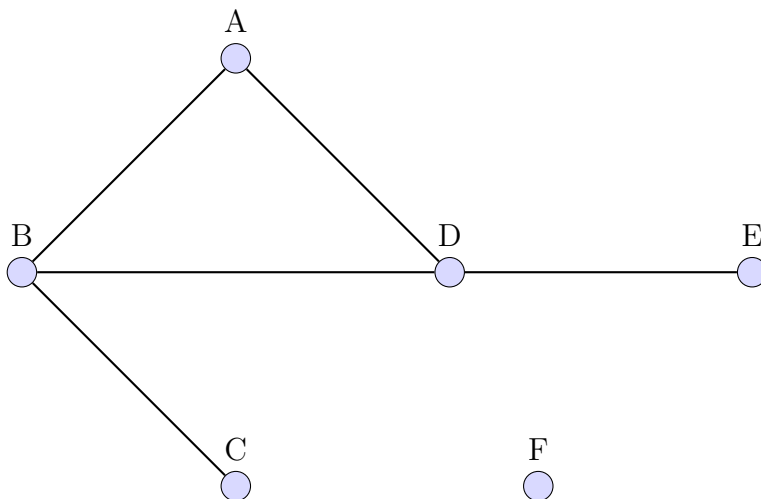


Figure 1: A graph with  $V = \{A, B, C, D, E, F\}$  and  $E = \{\{A, B\}, \{A, D\}, \{B, D\}, \{B, C\}, \{D, E\}\}$

## 1.2 Definitions for undirected graphs

For a graph  $G$ ,  $V(G)$  is its vertex set and  $E(G)$  its edge set.

### Adjacency

$a, b \in V(G)$  are said to be adjacent in  $G$  if  $\{a, b\} \in E(G)$ .

### Path

A path  $a_1, a_2, \dots, a_n$  is a series of vertices in  $V(G)$  such that if  $2 \leq i \leq n$ , then  $a_{i-1}$  is adjacent to  $a_i$  in  $G$ .  $n$  is the length of such a path.

### Cycle

A cycle is a path of the form  $a, \dots, a$  of length greater than 1.

**Subgraph**

$G'$  is a subgraph of  $G$ , expressed as  $G' \subseteq G$ , if  $V(G') \subseteq V(G)$  and  $E(G') \subseteq E(G)$ .  $G' \subset G$  means that  $G' \subseteq G$  but  $V(G') \neq V(G)$  or  $E(G') \neq E(G)$ .

**Connected component**

A connected component  $G'$  of  $G$  is a subgraph of  $G$  such that a path exists between any two vertices of  $G'$  and no  $H$  exists such that  $G' \subset H$  and  $H$  is a connected component.

**Tree**

A tree is a graph with a single connected component and no cycle.

**Forest**

A forest is a graph such that every connected component is a tree.

**Rooted Tree**

A rooted tree is a tree with a special node that is called the root.

**Rooted Forest**

A rooted forest is a graph such that every connected component is a rooted tree.

**Height of a node**

The height of a node in a rooted tree is the length of the path from that node to the root. The height of the root itself is 1. In a rooted forest, the height of a node is its height in the rooted tree it belongs to.

**Height of rooted forest**

The height of a rooted forest is the maximum height of any of its nodes.

## 2 Introduction to Tree Depth

### 2.1 Basic definitions

Vertex  $x$  is said to be the ancestor of  $y$  in a rooted forest  $F$ , if and only if  $x$  belongs to the path between  $y$  and the root of the component to which  $x$  belongs,  $y$  included.

The closure of a rooted forest  $F$ , expressed as  $C = \text{clos}(F)$ , is defined as follows:

- $V(C) = V(F)$

- $E(C) = \{ \{x, y\} : x \text{ is an ancestor of } y \text{ in } F, x \neq y \}$

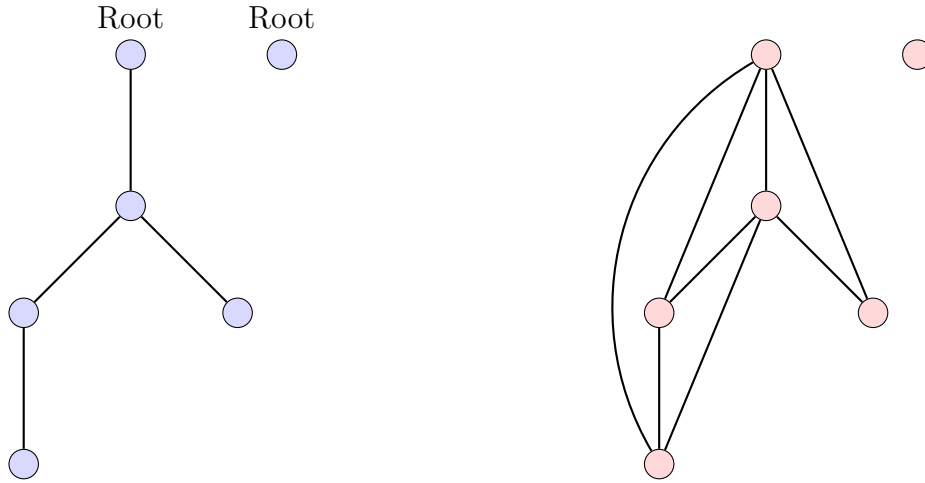


Figure 2: The blue graph at the left is a rooted forest  $F$ , the red graph at the right represents  $\text{clos}(F)$ .

## 2.2 Tree Depth

**Definition 2.1.** *The tree-depth  $td(G)$  of a graph  $G$  is the minimum height of a rooted forest  $F$  such that  $G \subseteq \text{clos}(F)$*

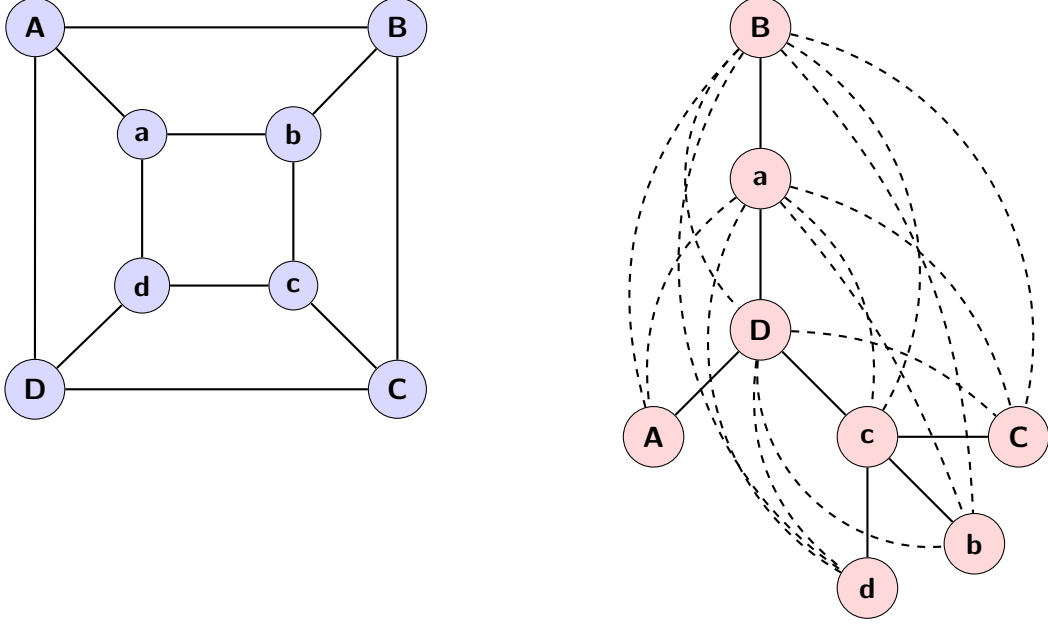


Figure 3: The graph  $G$  and tree  $T$  are in the left and right respectively. The dotted edges in  $T$ , represent the  $\text{clos}(T)$ . Because  $G \subseteq \text{clos}(T)$ ,  $\text{height}(T) = 5$  and the definition of tree depth we just gave, we know that  $\text{td}(G)$  is at most 5.

The tree depth of a graph  $G$  is a numerical invariant of a graph. In other words, the tree depth is a property that depends only on the abstract structure of a graph, not on its representation.

## 2.3 Elimination Forest

An elimination forest  $F$  of a connected graph  $G = (V, E)$  is defined recursively as follows:

- If  $V = \{x\}$  then  $F$  is just  $\{x\}$ .
- If  $G$  is not connected, then  $F$  is the union of the elimination forests of each component of  $G$ .
- Otherwise,  $r \in V$  is chosen as the root of  $F$  and an elimination forest is created for  $G - r$ . The roots of this elimination forest will be the children of  $r$  in  $F$ .

The tree  $T$  in Figure 3 is an elimination forest for the graph  $G$ .

**Lemma 2.2.** *Let  $G$  be a graph and  $F$  a rooted forest such that  $G \subseteq \text{clos}(F)$ . Then,  $Y$  exists, where  $Y$  is an elimination forest of  $G$  and  $\text{height}(Y) \leq \text{height}(F)$ .*

*Proof.*

**Base case:** If  $V(G) = \{v\}$ , then  $V(Y) = \{v\}$  and  $\text{height}(Y) = 1$ . Beware that  $F$  can have nodes that are not in  $G$  but it must contain  $v$ , so  $\text{height}(Y) \leq \text{height}(F)$ .

**Induction:** If  $G$  is connected, set the root of  $F$ ,  $v$ , as the root of  $Y$ . Clearly,  $G-v \subseteq \text{clos}(F-v)$ , so by induction an elimination forest  $Y'$  exists such that  $G-v \subseteq \text{clos}(Y')$  and  $\text{height}(Y') \leq \text{height}(F-v)$ . The roots of  $Y'$  will be the children of  $v$  in  $Y$  and as  $G-v \subseteq \text{clos}(Y')$ , then  $G \subseteq \text{clos}(Y)$  and  $Y$  is an elimination forest. With that we can prove the lemma like this:  $\text{height}(Y) = 1 + \text{height}(Y') \leq 1 + \text{height}(F-v) = \text{height}(F)$ , so  $\text{height}(Y) \leq \text{height}(F)$ .

If  $G$  is not connected, then every component  $G_i$  in  $G$  is contained in the closure of a component  $F_i$  in  $F$ . Otherwise, the edge between two adjacent nodes in  $G$  that are both in  $G_i$  but in two different components of  $F$  wouldn't be in  $\text{clos}(F)$  and that can't happen. By induction we can assume that for every component  $G_i$ , there exists an elimination forest  $Y_i$  such that  $G_i \subseteq \text{clos}(Y_i)$  and  $\text{height}(Y_i) \leq \text{height}(F_i)$ .  $Y$  will be the union of all these  $Y_i$  which is clearly an elimination forest and because for every component of  $Y$  there exists a component in  $F$  with higher or equal height, then  $\text{height}(Y) \leq \text{height}(F)$ .  $\square$

From this lemma we can say that the tree depth of a graph is the minimal height of an elimination forest for that graph. We can now recursively define the tree depth of a graph using the definition of an elimination forest:

**Definition 2.3.** *The tree depth of a graph  $G$  with  $G_1, \dots, G_k$  components is the following:*

$$td(G) = \begin{cases} 1 & \text{if } |G| = 1 \\ \max_{i=1}^p td(G_i) & \text{if } G \text{ is not connected} \\ 1 + \min_{v \in V(G)} td(G - v) & \text{otherwise} \end{cases}$$

## 3 Game Theoretic approach to Tree-Depth

### 3.1 Defining the game

For  $k \geq 0$ , the  $k$ -step selection-deletion game is played by Alice and Bob on a graph. The game is played by turns as follows:

- First, Alice selects a connected component of the graph, and the rest of the components are deleted.
- Then, Bob deletes a node from the remaining graph and the next round is played with this graph.

If Bob deletes the last node at the  $k$ -th round or earlier, he is said to win. Otherwise, Alice wins.

From this definition we can observe that if Bob has a strategy to win in  $k$  rounds that strategy will also guarantee a win in any game that lasts more than  $k$  rounds. Conversely, if Alice has a winning strategy in  $k$ -rounds, that same strategy will also win any game with less than  $k$  rounds.

### 3.2 Bob's winning strategy

**Lemma 3.1.** *Let  $G$  be a graph and let  $F$  be a rooted forest of height  $t$  such that  $G \subseteq \text{clos}(F)$ . Then, Bob has a winning strategy for the  $t$ -step selection-deletion game.*

*Proof.* Because of lemma 2.2 we know an elimination forest  $Y$  exists such that  $\text{height}(Y) \leq \text{height}(F)$ . Consider  $h = \text{height}(Y)$ , we will prove that a winning strategy exists in  $h$  rounds which is also a winning strategy in the  $t$ -step selection-deletion game because  $h \leq t$ .

- **Base case:** If  $h = 1$ , then every component of  $G$  will have a single vertex, so it's clear that Bob will win the 1-step selection-deletion game.
- **Induction:** Let  $G_i \subseteq G$  be the component Alice chooses, then  $Y_i$  exists such that  $Y_i$  is an elimination forest belonging to  $Y$ ,  $G_i \subseteq \text{clos}(Y_i)$  and obviously  $\text{height}(Y_i) \leq h$ . Bob will delete  $v$ , the root of  $Y_i$ . This will leave us with  $G' = G_i - v$  as the new graph. If we consider the children of  $v$  the new roots in  $Y' = Y_i - v$ , then  $G' \subseteq \text{clos}(Y')$  because of how the elimination forests are built. As  $\text{height}(Y') \leq h-1$ , we can assume by induction that Bob has a winning strategy in  $h-1$  rounds for  $G'$ , which together with the strategy for the first round we have just defined makes a winning strategy for Bob in the  $h$ -step selection-deletion game on the graph  $G$ .

□

### 3.3 Alice's winning strategy

**Definition 3.2.** A shelter  $S$  in a graph  $G$  is a set of graphs with the next properties:

- $\forall H \in S, H \subseteq G$  and  $H$  is connected.
- $H$  is said to be minimal if no  $H'$  exists in  $S$  such that  $H' \subset H$ .
- $H$  is said to be maximal if no  $H'$  exists in  $S$  such that  $H \subset H'$ .
- If  $H \in S$  and  $H$  is not minimal, then  $\forall v \in V(H)$ , there exists  $H' \subseteq H - v$  such that  $H$  covers  $H'$ . We will say that  $a \in S$  covers  $b \in S$  if and only if  $b \subset a$ , and no  $c \in S$  exists such that  $b \subset c \subset a$ .

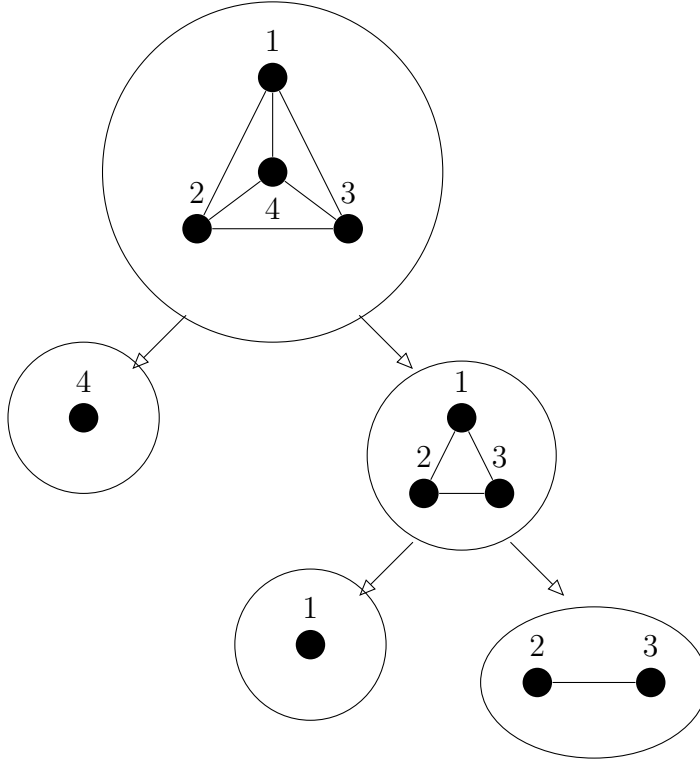


Figure 4: An example of a shelter. The arrows represent the covering relation.

The thickness of a shelter  $S$  is the shortest sequence of elements in  $S$  of the form  $a_1, \dots, a_n$  such that  $a_1$  is maximal and  $a_n$  is minimal and if  $2 \leq i \leq n$ , then  $a_{i-1}$  covers  $a_i$ . The length of a chain is defined as the number of elements



in it. The thickness of the shelter in figure 4 is 2, because of the sequence  $\{1, 2, 3, 4\}, \{4\}$ .

**Lemma 3.3.** *Let  $G$  be a graph,  $S$  a shelter in  $G$ , and  $t$  the thickness of  $S$ . Then, there exists a winning strategy for Alice in the  $(t-1)$ -step selection-deletion game.*

*Proof.* We will proof this by induction over  $t$ .

- **Base case:** If  $t = 1$ , then clearly Alice wins the 0-step selection-deletion game.
- **Induction:** Let  $H$  be a maximal element in  $S$ . Then, Alice picks the connected component  $G_i$  of  $G$ , such that  $H \subseteq G_i$ . Because  $t > 0$ ,  $H$  is not minimal, so for any vertex  $v$  that Bob removes, if  $v \in H$  there exists  $H' \in S$  that is covered by  $H$  and  $v \notin H'$ . Otherwise,  $H$  is still a subgraph of  $G_i - v$ . Let  $S' = \{X \mid X \in S \wedge X \subseteq G_i - v\}$ . It is clear that  $S'$  is a shelter for  $G_i - v$  and that the thickness of  $S'$  is greater than or equal to  $t-1$ . By induction we can assume Alice has a winning strategy in  $t-2$  steps in  $G_i - v$ , which together with the strategy for the first round we have just defined is a winning strategy for the  $(t-1)$ -step selection-deletion game.

□

### 3.4 Relation to Tree-Depth

It is clear that if Alice has a winning strategy in the  $t$ -step selection deletion game, Bob can't have a winning strategy in that same game. Because of this and lemmas 3.1 and 3.3 we can state the following:

**Theorem 3.4.** *Let  $G$  be a graph,  $S$  a shelter in  $G$  of thickness  $x$  and  $F$  a rooted forest of height  $y$  such that  $G \subseteq \text{clos}(F)$ . Then the following is true.*

1. *Alice has a winning strategy in the  $(t-1)$ -step selection-deletion game, for any  $t$  smaller than or equal to  $x$ .*
2. *Bob has a winning strategy in the  $t$ -step selection-deletion game, for any  $t$  greater than or equal to  $y$ .*
3. *Every rooted forest who's closure contains  $G$  has an height higher than or equal to  $x$ . Otherwise, Bob would have a winning strategy in the  $(x-1)$ -step selection-deletion game, which contradicts statement 1.*

4. Every shelter in  $G$  has a thickness smaller than or equal to  $y$ . Otherwise, Alice would have a winning strategy in the  $y$ -step selection-deletion game, which contradicts statement 2.
5. Because we have  $F$ ,  $td(G) \leq y$ . Also, from statement 3 it is clear that  $x \leq td(G)$ . So we can say that  $x \leq td(G) \leq y$ .

With this theorem we can now give an upper-bound and a lower-bound to a graphs tree depth.

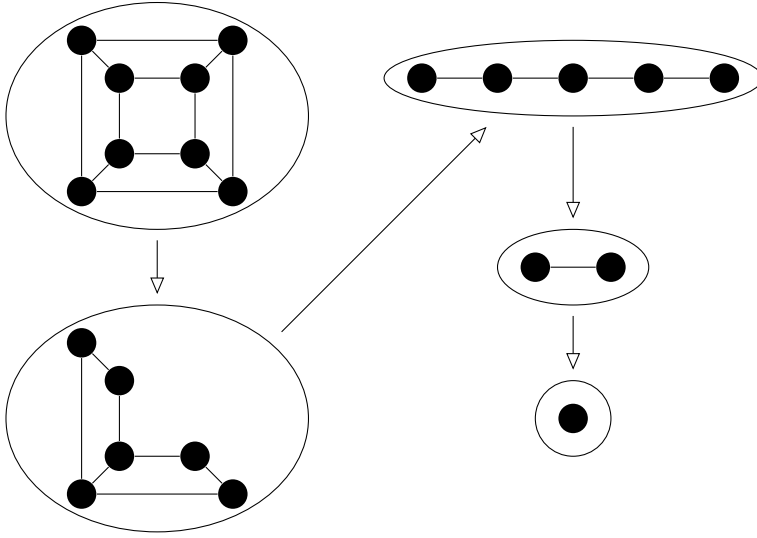


Figure 5: This is a shelter of thickness 5 for the graph in Figure 3. Beware that not all graphs in the shelter are drawn, but every graph in the shelter is isomorphic to these. With this and the rooted forest from Figure 3 we can say that  $td(G) = 5$ .

## 4 Cycle rank

### 4.1 Defining cycle rank

Cycle rank is a numerical invariant in a directed graph which is closely related to the tree depth of an undirected graph.

**Definition 4.1.** *The cycle rank of a digraph  $G$ , denoted by  $r(G)$  is defined as follows:*

- If  $G$  is acyclic, then  $r(G) = 0$ .

- If  $G$  is nontrivial strongly connected, then  $r(G) = 1 + \min_{v \in V(G)} \{r(G - v)\}$ .
- If  $G$  is not strongly connected and contains at least a cycle, then  $r(G)$  is the maximum cycle rank among all nontrivial strongly connected components of  $G$ .

The graph with a single node and no edges is considered trivially strongly connected, while the graph with a single node and a loop is considered nontrivially strongly connected.

## 4.2 Directed elimination forest

Similar to the notion of elimination forests in undirected graphs, we have directed elimination forests on digraphs.

**Definition 4.2.** A directed elimination forest for a digraph  $G$  is a rooted forest  $F$ .  $F$  can be defined recursively as follows:

- For the  $k \geq 0$  non trivially strongly connected components of  $G$ ,  $Y_1, \dots, Y_k$ ,  $(v_i, Y_i)$  are the roots in  $F$ , where  $v_i \in Y_i$  and  $1 \leq i \leq k$ .
- For each  $(v_i, Y_i)$ , a directed elimination forest is created for  $G[Y_i] - v_i$  and the roots of that forest are the children of  $(v_i, Y_i)$  in  $F$ .

**Lemma 4.3.** Let  $F$  be directed elimination forests of minimum height for a digraph  $G = (V, E)$ . Then,  $r(G) = \text{height}(F)$ .

*Proof.* We will proof this by induction on the number of vertices of  $G$ .

- **Base case:** If  $G$  is acyclic, then  $r(G) = 0$ ,  $\text{height}(F)$  is 0 because we assume that the height of the empty tree is 0.
- **Induction:** If  $G$  is nontrivially strongly connected, then  $v \in V$  exists, such that  $r(G) = 1 + r(G - v)$ . Let  $(v, V)$  be the root of  $F$ , then  $\text{height}(F) = 1 + \text{height}(F')$  where  $F'$  is any directed elimination forest of  $G - v$  because of definition 4.2. If we consider  $F'$  to be the directed elimination forest of  $G - v$  of minimum height, by induction we can assume that  $r(G - v) = 1 + \text{height}(F')$ . So,  $r(G) = 1 + r(G - v) = 1 + \text{height}(F') = \text{height}(F)$ .

If  $G$  is not strongly connected but it has at least a cycle, then, for every  $X$  that is a strongly connected component of  $G$  by induction we can assume that  $r(G[X]) = \text{height}(F_X)$  where  $F_X$  is the directed elimination tree of minimum height for  $G[X]$ . Because  $r(G)$  is the maximum among all  $r(G[X])$  and the  $\text{height}(F)$  is the maximum among all  $\text{height}(F_X)$ ,  $r(G) = \text{height}(F)$ .

□

## 5 Game Characterization of Cycle Rank

### 5.1 Definitions

For this section, we will assume all our graphs are directed and contain no self loops. We will also need some basic definitions before we can start talking about the games we will use to define cycle rank.

#### Initial component

$H \subseteq G$  is an initial component of  $G$  if  $H$  is a strongly connected component and there are no edges from  $G \setminus H$  to  $H$ .

#### Successor-closed

$H \subseteq G$  is an successor-closed of  $G$  if  $H$  is there are no edges from  $H$  to  $G \setminus H$ .

#### Power set

The power set of a set  $S$ , denoted by  $\mathcal{P}(S)$ , is the set of all possible subsets of  $S$ .

#### String

A string is a sequence of elements  $a_1, \dots, a_n$  such that all  $a_i$  belong to the same set. That set is called the alphabet.

#### Length

The length of a string  $A = a_1, \dots, a_n$ , denoted by  $|A|$  is  $n$ . The length of the empty string is 0.

#### Concatenation

The concatenation of two string  $A = a_1, \dots, a_n$  and  $B = b_1, \dots, b_k$ , denoted by  $A \cdot B$  is  $a_1, \dots, a_n, b_1, \dots, b_k$

#### $V^*$

$V^*$  is the set of all possible finite words over the set  $V$ , including the empty word.

#### $V^{<k}$

$V^{<k}$  is the set of all possible finite words over the set  $V$  of length smaller than  $k$ .

**Prefix**

$X \in V^*$  is a prefix of  $Y \in V^*$ , denoted by  $X \preceq Y$  if  $Z \in V^*$  exists such that  $Y = X \cdot Z$ .

**String to set**

For a string  $S = a_1, \dots, a_n$ ,  $\{|S|\}$  denotes the set  $\{a_1, \dots, a_n\}$ .

**Symmetric difference**

For two sets  $A$  and  $B$  their symmetric difference, expressed like  $A \Delta B$  is  $(A \cup B) \setminus (A \cap B)$ .

**5.2 Game description**

We will use a cops and robbers game played on a graph  $G$ , where the cops will try to catch a robber. In each step of the game the cops can either place a searcher on a node or remove only the most recently placed searcher. This is why it's called a LIFO search.

For a digraph  $G$ , the state of the game is described by a pair  $(X, R)$ .  $X \in V^*$  is the position of the cops and the order in which they were added. The meaning of  $R$  depends on the game variant we are talking about, but it will always be an induced subgraph of  $G \setminus \{|X|\}$ . We will introduce 4 game variants now:

Invisible