Cours 2. Partie 1: Introduction à l'algorithmique Graphes. Arbres.

Olivier Bournez

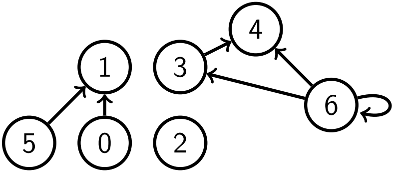
bournez@lix.polytechnique.fr
LIX, Ecole Polytechnique

Aujourd'hui

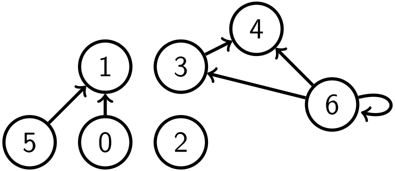
Un graphe orienté

- Un *graphe orienté* (digraph) est donné par un couple $G = (V, E)$, où
  - V est un ensemble.
  - $E \subset V \times V$.

- Exemple:
  - $V = \{0, 1, \ldots, 6\}$.
  - $E = \{(0,1), (3,4), (5,1), (6,3), (6,4), (6,6)\}$.



Vocabulaire

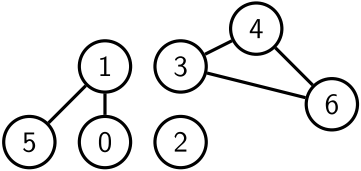


- Les éléments de V sont appelés des *sommets* (parfois aussi des *nœuds*).
- Les éléments e de E sont appelés des *arcs*.
- Si $e = (u, v)$, u est appelé *la source* de e, v est appelé *la destination* de e.
- Remarque:
  - Les boucles (les arcs (u, u)) sont autorisées.

Un graphe

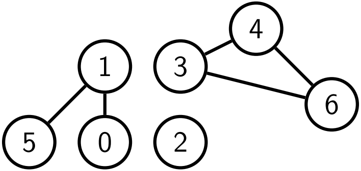
- Un *graphe*[1] est donné par un couple $G = (V, E)$, où
  - V est un ensemble.
  - E est un ensemble de paires $\{u, v\}$ avec $u, v \in V$.
- On convient de représenter une paire $\{u, v\}$ par (u, v) ou (v, u).
- Autrement dit, (u, v) et (v, u) dénotent la même arête.

- Exemple:
  - $V = \{0, 1, \ldots, 6\}$
  - $E = \{(0, 1), (3, 4), (5, 1), (6, 3), (6, 4)\}$.



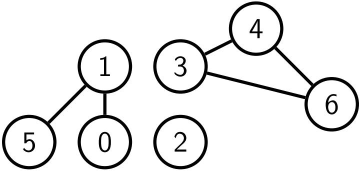
[1] Lorsqu'on ne précise pas, par défaut, un graphe est non-orienté.

Vocabulaire



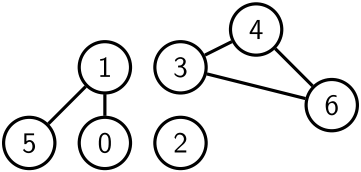
- Les éléments e de E sont appelés des *arêtes*. Si $e = (u, v)$, u et v sont appelés les *extrémités* de e.
- Remarque: (sauf autre convention explicite)
  - Les boucles ne sont pas autorisées.

Vocabulaire (cas non-orienté)



- u et v sont dits *voisins* s'il y a une arête entre u et v.
- Le degré de u est le nombre de voisins de u.

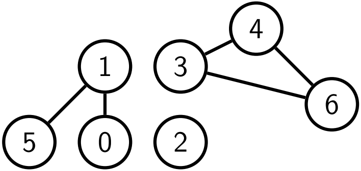
Vocabulaire (cas non-orienté)



- Un *chemin du sommet s vers le sommet t* est une suite $e_0, e_1, \cdots, e_n$ de sommets telle que $e_0 = s$, $e_n = t$, $(e_{i-1}, e_i) \in E$, pour tout $1 \leq i \leq n$.
  - n est appelé la *longueur* du chemin, et on dit que t *est joignable* à partir de s.
  - Le chemin est dit *simple* si les e_i sont distincts deux-à-deux.
  - Un *cycle* est un chemin de longueur non-nulle avec $e_0 = e_n$.
- Le sommet s est dit *à distance* n de t s'il existe un chemin de longueur n entre s et t, mais aucun chemin de longueur inférieure.

Composantes connexes (cas non-orienté)

- Prop. La relation “être joignable” est une relation d’équivalence.
- Les classes d’équivalence sont appelées les *composantes connexes*.

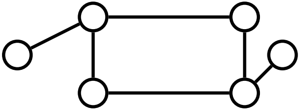
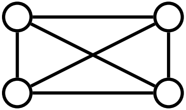
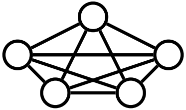
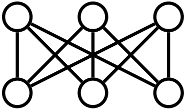


- Un graphe est dit *connexe* s’il n’y a qu’une seule classe d’équivalence.
  - Autrement dit, tout sommet est joignable à partir de tout sommet.

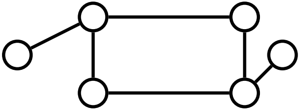
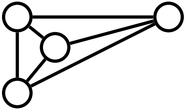
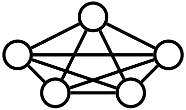
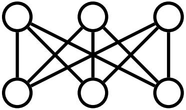
Les graphes sont partout!

- Beaucoup de problèmes se modélisent par des objets et des relations entre objets.
- Exemples:
  - Le graphe routier.
  - Les réseaux informatiques.
  - Le graphe du web.
- Beaucoup de problèmes se ramènent à des problèmes sur les graphes.
- Théorie des graphes:
  - Euler, Hamilton, Kirchhoff, König, Edmonds, Berge, Lovász, Seymour,. . .
- Les graphes sont omniprésents en informatique.

Théorie des graphes. Exemple 1: graphes planaires.

- Un graphe est dit *planaire* s'il peut se représenter sur un plan sans qu'aucune arête n'en croise une autre.
- Planaire:  
- Non-planaire: K_5  $K_{3,3}$ 
- Théorème [Kuratowski-Wagner] Un graphe fini est planaire ssi il ne contient pas de sous-graphe qui soit une expansion de K_5 ou de $K_{3,3}$.
  - Une expansion consiste à ajouter un ou plusieurs sommets sur une ou plusieurs arêtes (exemple:  devient )

Théorie des graphes. Exemple 1: graphes planaires.

- Un graphe est dit *planaire* s'il peut se représenter sur un plan sans qu'aucune arête n'en croise une autre.
- Planaire:  
- Non-planaire: K_5  $K_{3,3}$ 
- Théorème [Kuratowski-Wagner] Un graphe fini est planaire ssi il ne contient pas de sous-graphe qui soit une expansion de K_5 ou de $K_{3,3}$.
  - Une expansion consiste à ajouter un ou plusieurs sommets sur une ou plusieurs arêtes (exemple:  devient )

Théorie des graphes. Exemple 2: Coloriage de graphe.

- Allouer des fréquences GSM correspond à colorier les sommets d'un graphe.
  - sommets: des émetteurs radio.
  - arête entre u et v: le signal de u perturbe v ou réciproquement.
  - couleur: fréquence radio.
- Le problème de *coloriage d'un graphe*: colorier les sommets d'un graphe de telle sorte qu'il n'y ait aucune arête entre deux sommets d'une même couleur.

Un coloriage avec 4 couleurs

Théorie des graphes. Exemple 2: Coloriage de graphe.

- Théorème: Appel et Haken (76): tout graphe planaire est coloriable avec 4 couleurs (preuve avec 1478 cas critiques).
  - Robertson, Sanders, Seymour, Thomas, Gonthier, Werner.

- On ne connaît aucun algorithme pour déterminer si un graphe général est coloriable avec 4 couleurs (même avec 3) qui fonctionne en temps polynomial (le problème est NP-complet).

Applications des graphes. Exemple 3: évaluer une page de la toile

- Brevet "Method for Node Ranking in a Linked Database" (Université Stanford, Janvier 1997).
- En simplifiant.
  - On considère le graphe des liens entre pages.
  - Modèle théorique:
    - Avec probabilité $0 < c < 1$, un surfeur abandonne la page actuelle et recommence sur une des n pages du web, choisie de manière équiprobable.
    - Avec probabilité $1 - c$, le surfeur suit un des liens de la page actuelle j, choisie de manière équiprobable parmi tous les liens l_j émis.
  - Quelle que soit la distribution initiale, on convergera vers une unique distribution stationnaire.
  - On évalue une page par la probabilité de cette page dans la distribution stationnaire.

Aujourd'hui

Les graphes

Les arbres

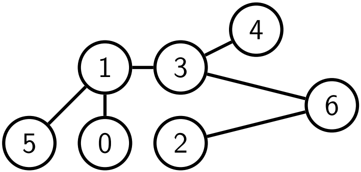
Les arbres binaires

Arbres binaires: hauteurs vs nombre de sommets

Arbres AVL

Les arbres

- Un graphe[2] connexe sans cycle est appelé un *arbre* (libre).
- Un graphe[2] sans-cycle est appelé une *forêt*:
  - chacune de ses composantes connexes est un arbre.

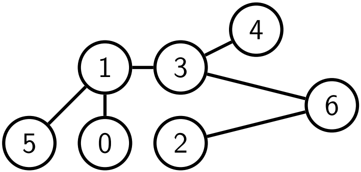


[2]non-orienté.

Les arbres sont partout!

- Un graphe[2] connexe sans cycle est appelé un *arbre* (libre).
- Un graphe[2] sans-cycle est appelé une *forêt*:
  - chacune de ses composantes connexes est un arbre.

- Dès qu'on a des objets, des relations entre objets, et pas de cycle, on a donc un arbre ou une forêt.

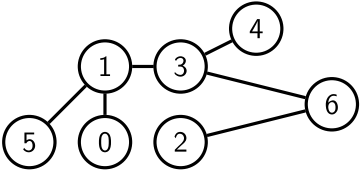


[2] non-orienté.

Les arbres sont partout!

- Un graphe[2] connexe sans cycle est appelé un *arbre* (libre).
- Un graphe[2] sans-cycle est appelé une *forêt*:
  - chacune de ses composantes connexes est un arbre.

- Dès qu'on a des objets, des relations entre objets, et pas de cycle, on a donc un arbre ou une forêt.



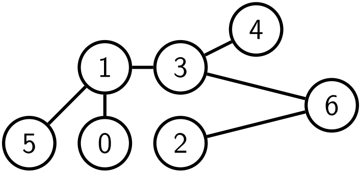
- Les arbres sont omniprésents en informatique.

[2] non-orienté.

Une caractérisation & Quelques propriétés

Soit $G = (V, E)$ un graphe[3]. Les propriétés suivantes sont équivalentes:

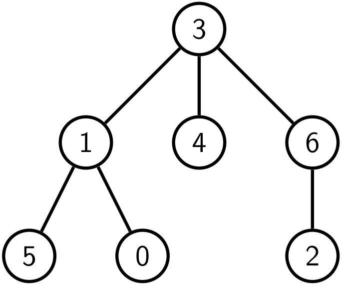
- G est un arbre (libre).
- Deux sommets quelconques de V sont connectés par un unique chemin simple.
- G est connexe, mais ne l'est plus si on enlève n'importe laquelle de ses arêtes.
- G est connexe, et $|E| = |V| - 1$.
- G est sans cycle, et $|E| = |V| - 1$.
- G est sans cycle, mais ne l'est plus si l'on ajoute n'importe quelle arête.



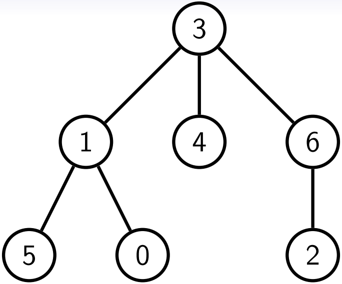
[3] non-orienté.

Les arbres poussent de haut en bas en informatique

- On distingue souvent un sommet que l'on appelle *sa racine*.
- On dessine un arbre
  - en plaçant la racine tout en haut.
  - puis en plaçant les sommets à distance i de la racine à la ligne i.
- Exemple: pour l'arbre libre précédent, en prenant le sommet d'étiquette 3 comme racine.

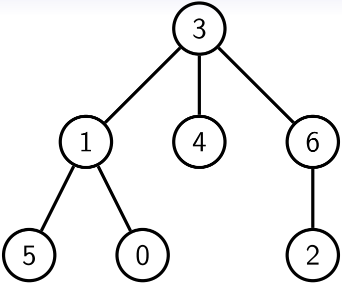


Encore du vocabulaire



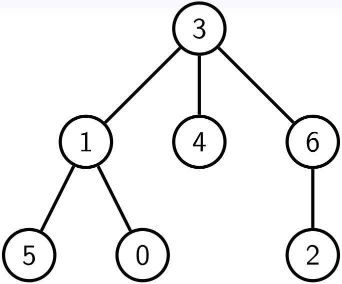
- Pour tout sommet u, il existe un unique chemin (simple) entre la racine r et u.

Encore du vocabulaire



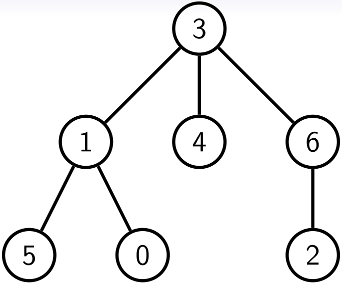
- Pour tout sommet u, il existe un unique chemin (simple) entre la racine r et u.
  - Exemple:
    - pour le sommet 0, le chemin est $3, 1, 0$.

Encore du vocabulaire



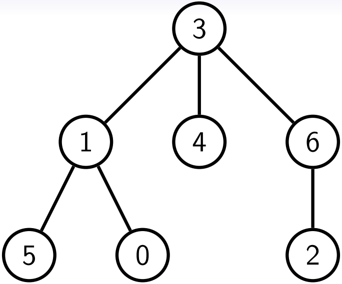
- Tout sommet a sur ce chemin est appelé *un ancêtre* de u. u est dit un *descendant* de a.

Encore du vocabulaire



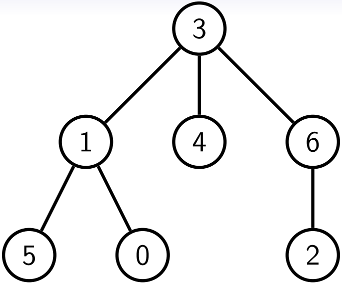
- Tout sommet a sur ce chemin est appelé *un ancêtre* de u. u est dit un *descendant* de a.
  - Exemple:
    - 1,3 sont des ancêtres de 0.
    - 5,0 sont des descendants de 1.

Encore du vocabulaire



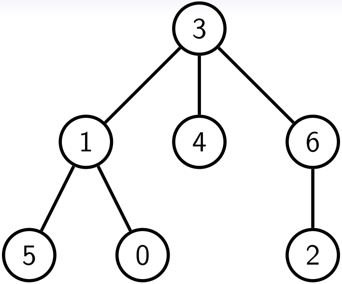
- L'avant dernier sommet p sur ce chemin est appelé *le père* de u. u est appelé *un fils* de p.

Encore du vocabulaire



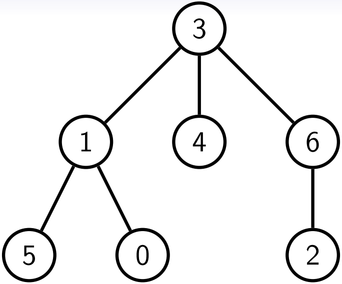
- L'avant dernier sommet p sur ce chemin est appelé *le père* de u. u est appelé *un fils* de p.
  - Exemple:
    - Le père de 0 est 1.
    - Le père de 4 est 3.

Encore du vocabulaire



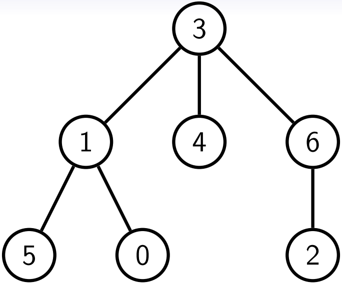
- Le *sous-arbre* de racine u est l'arbre induit par les descendants de u.

Encore du vocabulaire



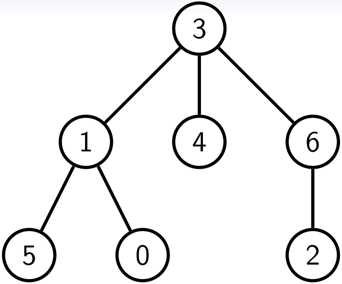
- Le *sous-arbre* de racine u est l'arbre induit par les descendants de u.
  - Exemple:
    - Le sous-arbre de racine 1 est constitué du sous-graphe $1, 5, 0$.
    - Le sous-arbre de racine 6 est le sous-graphe 6,2.

Encore du vocabulaire



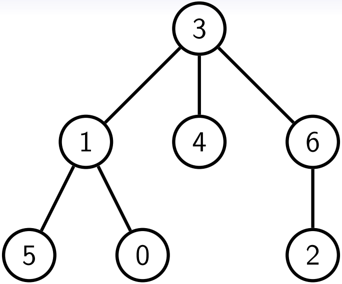
- *L'arité* d'un sommet est le nombre de ses fils.

Encore du vocabulaire



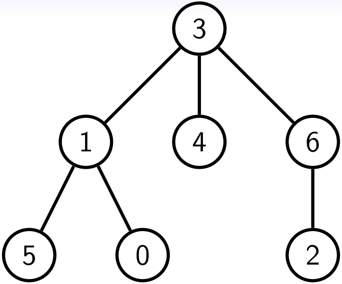
- *L'arité* d'un sommet est le nombre de ses fils.
  - Exemple:
    - 1 est d'arité 2.
    - 4 est d'arité 0.
    - 6 est d'arité 1.

Encore du vocabulaire



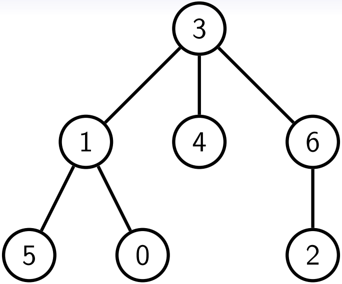
- Un sommet qui n'a pas de fils est appelé *une feuille*.

Encore du vocabulaire



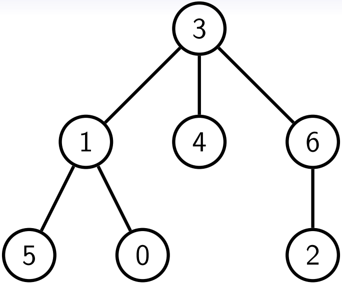
- Un sommet qui n'a pas de fils est appelé *une feuille*.
  - Exemple:
    - Les feuilles de l'arbre sont 5,0,4, et 2.

Encore du vocabulaire



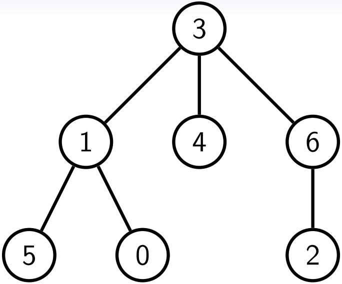
- La *hauteur d'un sommet* est sa distance à la racine r.

Encore du vocabulaire



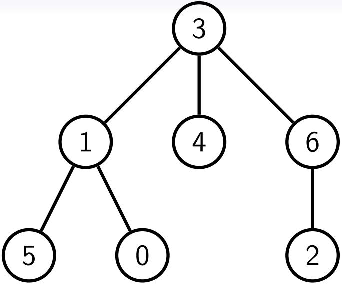
- La *hauteur d'un sommet* est sa distance à la racine r.
  - Exemple:
    - 3 est à hauteur 0.
    - 1,4, et 6 sont à hauteur 1.
    - Les autres sommets sont à hauteur 2.

Encore du vocabulaire



- La *hauteur d'un arbre* est la hauteur de la feuille la plus haute.

Encore du vocabulaire



- La *hauteur d'un arbre* est la hauteur de la feuille la plus haute.
  - Exemple:
    - L'arbre est de hauteur 2.
    - Le sous-arbre de racine 6 est de hauteur 1.

Une vision récursive: cas général

- Un arbre correspond à un couple formé
  1. d'un sommet particulier, appelé sa racine,
  2. et d'une partition des sommets restants en un ensemble d'arbres.

- Cette vision permet de faire des preuves inductives.

  - Exemple: montrer qu'un arbre dont tous les sommets sont d'arité au moins 2 possède plus de feuilles que de sommets internes.

- Mais reste informatiquement encore un peu compliquée.

Aujourd'hui

Les graphes

Les arbres

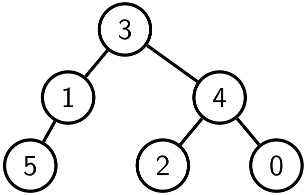
Les arbres binaires

Arbres binaires: hauteurs vs nombre de sommets

Arbres AVL

Une vision récursive: arbres binaires

- Plus simple: Les arbres binaires.



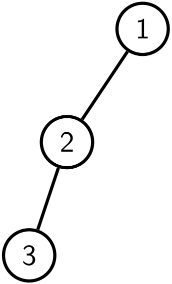
- Un arbre binaire est
  - soit vide
  - soit l'union disjointe d'un sommet, appelé *sa racine*, d'un arbre binaire, appelé *sous-arbre gauche*, et d'un arbre binaire, appelé *sous-arbre droit*.

- Autrement dit, un arbre binaire d'entiers est solution de l'équation:

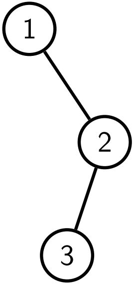
$$A = \textbf{null} \uplus (A \times \textbf{int} \times A)$$

Attention: Arbre binaire $\neq$ Arbre d'arité ≤ 2.

- Un arbre *binaire* a tous ses sommets d'arité 0, 1 ou 2.
- Mais les concept d'arbre de degré ≤ 2 et d'arbre binaire sont différents.
- En fait,



et



ne sont pas les même arbres binaires, car

$$(((\emptyset, 3, \emptyset), 2, \emptyset), 1, \emptyset) \neq (\emptyset, 1, ((\emptyset, 3, \emptyset), 2, \emptyset))$$

Aujourd'hui

Les graphes

Les arbres

Les arbres binaires

Arbres binaires: hauteurs vs nombre de sommets

Arbres AVL

Bornes sur la hauteur et nombre de sommets

- Dans un arbre binaire de hauteur h contenant n sommets, on a

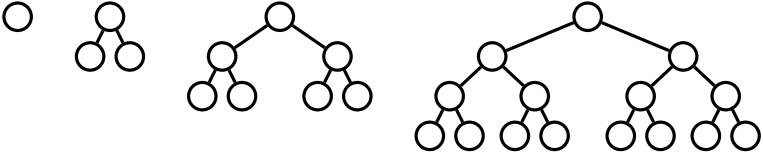
$$\lfloor \log_2 n \rfloor \leq h \leq n - 1.$$

- Si on préfère:

$$h + 1 \leq n \leq 2^{h+1} - 1.$$

- Arbres avec h maximal.

- Arbres avec n maximal.



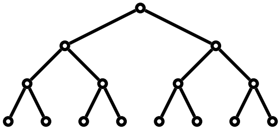
Remarques d'informaticien

- Une façon de comprendre $\log_2 n$:
  - $\lfloor \log_2 n \rfloor$:
    - la hauteur minimale d'un arbre binaire à n sommets.
    - le nombre de questions qu'il faut poser pour déterminer $1 \leq i \leq n$.
    - l'entropie d'une loi uniforme sur $1, 2, \dots, n$.

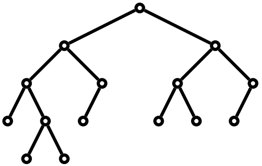
- Une façon de comprendre 2^k:
  - 2^k:
    - à 1 près, le nombre de sommets d'un arbre complet de hauteur $k + 1$.

Arbres avec h de l'ordre de $\log n$

- Un arbre (très bien) équilibré: $h = \log n$.

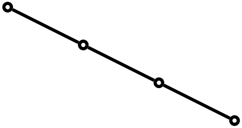


- Un arbre à peu près équilibré: $h \leq 2 \log n$.

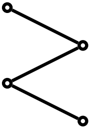


Arbres avec h de l'ordre de n

- Un arbre très mal équilibré: $h = n - 1$.



- Un autre arbre très mal équilibré: $h = n - 1$.



Aujourd'hui

Les graphes

Les arbres

Les arbres binaires

Arbres binaires: hauteurs vs nombre de sommets

Arbres AVL

Principe des arbres AVL

- Un arbre AVL (Adelson-Velskii et Landis) est

  1. un arbre binaire (de recherche[4])
  2. tel qu'en tout sommet, la hauteur du fils droit et du fils gauche diffèrent *au plus* d'un.

- Intérêt majeur:

  - Cette propriété garantie que la hauteur d'un arbre AVL est en $O(\log n)$.

[4] voir cours suivants

Pourquoi?

- Plus précisément, cela découle de

$$\log_2(1+n) \leq 1+h \leq 1.44\log_2(2+n).$$

Pourquoi?

- Plus précisément, cela découle de

$$\log_2(1+n) \leq 1+h \leq 1.44 \log_2(2+n).$$

- La borne $\log_2(1+n) \leq 1+h$ est vraie pour tout arbre binaire (cf ce qui précède): un arbre binaire de hauteur h a au plus $2^{h+1}-1$ sommets.

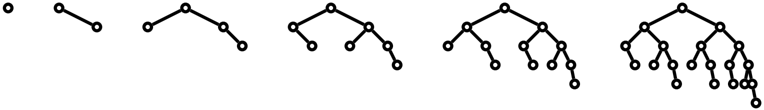
Pourquoi?

- Plus précisément, cela découle de

$$\log_2(1+n) \leq 1+h \leq 1.44\log_2(2+n).$$

- La borne $\log_2(1+n) \leq 1+h$ est vraie pour tout arbre binaire (cf ce qui précède): un arbre binaire de hauteur h a au plus $2^{h+1}-1$ sommets.
- Pour l'autre borne, on considère le nombre minimal $N(h)$ de sommets dans un arbre AVL de hauteur h.
  - Hauteur 0: $N(0) = 1$.
  - Hauteur 1: $N(1) = 2$.
  - Hauteur h, $h \geq 2$: une racine + un sous-arbre de hauteur $h-1$ + un sous-arbre de hauteur $h-2$.

$$N(h) = 1 + N(h-1) + N(h-2).$$



- Posons $F(h) = 1 + N(h)$,
  - On a $F(0) = 2$, $F(1) = 3$, et
  - $F(h) = F(h-1) + F(h-2)$, pour $h \geq 2$.
- C'est la suite de Fibonacci "décalée".
- On peut alors écrire:

$$F(h) = \frac{1}{\sqrt{5}}(\Phi^{h+3} - \Phi^{-(h+3)}),$$

où $\phi = \frac{1+\sqrt{5}}{2}$ est le nombre d'or.

- On a donc $1 + n \geq F(h) > \frac{1}{\sqrt{5}}(\Phi^{h+3} - 1)$ ce qui donne

$$h + 3 < \log_\Phi(\sqrt{5}(2+n)) < \frac{\log_2(2+n)}{\log_2(\Phi)} + 2.$$

- On observe que $\frac{1}{\log_2(\Phi))} \leq 1.44$.