TD1 - Arbres

Graphes et programmation dynamique

M1 Informatique - Semestre 1 - Année 2022-2023

Université Côte d'Azur

Christophe Crespelle
christophe.crespelle@univ-cotedazur.fr

Le sujet sera traite sur une seance d'1h30. Il est recommande que vous utilisiez les exercices que vous n'avez pas eu le temps de faire en seance comme entrainement en vue de l'examen final. Les exercices les plus importants sont les 1, 4 et 6 : commencez par ceux la.

Definition [Chemin simple et cycle simple]
Un chemin simple (resp. un cycle simple) est un chemin (resp. un cycle) sans repetition de sommet.

Exercice 1. *Dans chaque poisson y a une arete... et combien dans un arbre ?*

Definition [Arbre]
Un arbre est un graphe connexe et sans cycle simple.

a. Dessinez trois exemples d'arbre differents.

Solution. Un peu d'imagination...

b. L'union disjointe de deux arbres est-elle un arbre ?

Solution. Non, car l'union disjointe de deux graphes (ayant chacun au moins un sommet) n'est jamais connexe. Un graphe dont chaque composante connexe est un arbre est appele une **foret**.

c. Soit G un graphe sans cycle a $n \geq 1$ sommets et m aretes. Montrez que $m \leq n-1$.

Indication. Prouvez le par induction sur le nombre d'aretes.

Solution. C'est vrai pour $m = 0$, car $n \geq 1$.
Supposons que ce soit vrai pour $m \leq k$ et montrons que c'est vrai pour $m = k + 1$.

1. On le montre d'abord pour les graphes G a $k + 1$ aretes qui sont connexes.
   Soit donc G un graphe a $k + 1$ aretes qui est sans cycle et connexe, montrons que $m \leq n - 1$. Pour cela, supprimons une arete uv quelconque de G.

— Comme G ne contient aucun cycle, $G - uv$ est non connexe, car u et v ne sont pas dans la meme composante connexe de $G - uv$. En effet, si u et v sont dans la meme composante connexe de $G - uv$, alors il existe un chemin simple $u, x_1, \ldots, x_k, v$ de u a v dans $G - uv$, avec $k \geq 1$. Ainsi, G contient le cycle simple $u, x_1, \ldots, x_k, v, u$. Par contrapose, on a que u et v ne sont pas dans la meme composante connexe de $G - uv$.

— Montrons que $G - uv$ possede exactement deux composantes connexes. Pour cela considerons un sommet $x \in V \setminus \{u, v\}$ et montrons qu'il appartient soit a la composante connexe de u, soit a celle de v. Comme G est connexe, il existe un chemin simple P de x a u dans G. Si P ne contient pas l'arete uv, alors P est aussi un chemin de $G - uv$ et x est donc dans la composante connexe de u dans $G - uv$. Si au contraire P contient l'arete uv, comme P est simple, il ne la contient qu'une seule fois, a la fin. Par consequent, le debut de P est un chemin de x a v qui ne contient pas l'arete uv. Ainsi, x est dans la composante connexe de v dans $G - uv$. $G - uv$ a donc exactement deux composantes connexes.

On les note C_u et C_v et elles ont m_u aretes et n_u sommets, pour C_u, et m_v aretes et n_v sommets, pour C_v. On a $m = m_u + m_v + 1$ et $n = n_u + n_v$. Par hypothese de recurrence, on a aussi, comme $m_u < m$ et $m_v < m$, $m_u \leq n_u - 1$ et $m_v \leq n_v - 1$. Ce qui donne $m = m_u + m_v + 1 \leq n_u + n_v - 1 = n - 1$. En conclusion, pour un graphe G a $k + 1$ aretes qui est sans cycle et connexe, on a $m \leq n - 1$.

2. Montrons maintenant qu'un graphe G a $k + 1$ aretes qui est sans cycle mais non connexe satisfait aussi $m \leq n - 1$.

   (a) Si toutes les aretes de G non connexe sont dans la meme composante connexe C_1 de G (c.a.d. si G est fait de C_1 plus au moins un sommet isole), d'apres ce qu'on a montre ci-dessus pour $G[C_1]$, on a $m_1 \leq n_1 - 1$. Or, comme $m = m_1$ et $n > n_1$, on a bien $m \leq n - 1$.

   (b) Sinon, les k composantes connexes C_i, $1 \leq i \leq k$ de G, avec $k \geq 2$, verifient toutes $m_i < m$. Ainsi, par hypothese d'induction on a $m_i \leq n_i - 1$ pour tout $i \in [\![1, k]\!]$. Ainsi, en sommant, comme $m = \sum_{i \in [\![1,k]\!]} m_i$ et $n = \sum_{i \in [\![1,k]\!]} n_i$ on obtient $m \leq n - k$. Et comme $k \geq 2$, alors $m \leq k - 1$.

   Ainsi, lorsque G a $k+1$ aretes et qu'il est sans cycle et non connexe, on a $m \leq n-1$.

Comme on a montre que c'est aussi le cas lorsque G est connexe, cela termine l'induction et la preuve.

d. Soit G un graphe connexe a $n \geq 1$ sommets et m aretes. Montrez que $m \geq n - 1$.

Indication. Prouvez le encore par induction sur le nombre d'aretes.

Solution. C'est vrai pour $m = 0$, car si G est connexe et sans arete alors il a exactement 1 sommet.
Supposons que ce soit vrai pour $m \leq k$ et montrons que c'est vrai pour $m = k + 1$. Soit donc G un graphe a $k + 1$ aretes et connexe, montrons que $m \geq n - 1$. Pour cela, supprimons une arete e quelconque de G. Deux cas de figure : $G - e$ reste connexe

ou $G - e$ est non-connexe. Si $G - e$ est connexe, par hypothese de recurrence, on a $m - 1 \geq n - 1$ et donc $m \geq n - 1$. Si $G - e$ est non connexe, alors $G - e$ a exactement 2 composantes connexes (on l'a montre a la question 3 de l'exo 1). Notons m_1 et n_1 le nombre d'aretes et de sommets de la premiere, et notons m_2 et n_2 le nombre d'aretes et de sommets de la seconde. On a $m = m_1 + m_2 + 1$ et donc $m_1 < m$ et $m_2 < m$. En applicant l'hypothese de recurrence a chacune des deux composantes connexes, on obtient $m_1 \geq n_1 - 1$ et $m_2 \geq n_2 - 1$. Ce qui donne $m \geq n_1 + n_2 - 1$. Et comme $n = n_1 + n_2$, on a bien $m \geq n - 1$. Ce qui acheve la recurrence et la preuve.

e. Quel est le nombre d'aretes d'un arbre a n sommets ?

Solution. D'apres les questions precedantes, comme un arbre est sans cycle, il a au plus $n - 1$ aretes. Et comme il est connexe, il a au moins $n - 1$ aretes. Un arbre a n sommets a donc exactement $n - 1$ aretes.

Exercice 2.

Dis moi combien d'aretes tu as et je te dirai...

a. Un graphe a n sommets et $n - 1$ aretes est-il toujours un arbre ?
Si oui, demonstration, si non, contre exemple.

Solution. Non, voir contrexemple de la figure 1

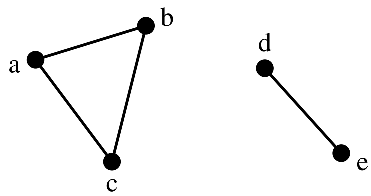

FIGURE 1 – Un graphe a 5 sommets et 4 aretes qui n'est pas un arbre. Remarquez qu'il est a la fois non connexe et possede un cycle. Est-ce toujours le cas d'un tel contre exemple ? (Prouvez le.)

b. Montrez que si un graphe possede $n - 1$ aretes et qu'il est connexe, alors c'est un arbre.

Solution. Soit G un graphe connexe a n sommets. Si G a un cycle simple C, en retirant une arete ab quelconque de C on obtient le graphe $G - ab$ qui est toujours connexe. En effet, on peut remplacer ab dans tous les chemins qui la contiennent par $C - ab$, qui est un chemin de a a b. Ainsi, la connectivite de G n'a pas change en retirant ab.
Comme $G - ab$ est connexe, d'apres la question 4 de l'exo 1, $G - ab$ a au moins $n - 1$ aretes et G en a donc au moins n. Ainsi, si G est connexe et a $n - 1$ aretes, alors G n'a pas de cycle simple : G est un arbre.

c. Montrez que si un graphe possede $n - 1$ aretes et ne contient aucun cycle, alors c'est un arbre.

Solution. Soit G un graphe sans cycle a n sommets. Si G n'est pas connexe, alors G est l'union disjointe de deux graphes G_1 et G_2 ayant chacun m_1 aretes et n_1 sommets, pour G_1, et m_2 aretes et n_2 aretes pour G_2. On a alors $m = m_1 + m_2$ et $n = n_1 + n_2$. Comme G_1 et G_2 sont des sous-graphes de G sans cycle, alors G_1 et G_2 sont sans cycle. D'apres la question 3 de l'exo 1, on a donc $m_1 \leq n_1 - 1$ et $m_2 \leq n_2 - 1$. Ce qui donne $m = m_1 + m_2 \leq n_1 + n_2 - 2 = n - 2$. En conclusion, si G est sans cycle et non connexe, G a au plus $n - 2$ aretes. Ainsi, si G est sans cycle et a $n - 1$ aretes, alors G est connexe et donc G est un arbre.

Exercice 3. *Extremalite des arbres*

a. Montrez que si un graphe G ne contient pas de cycle et qu'on ne peut pas lui ajouter une arete en preservant cette propriete, alors c'est un arbre.

Solution. Si G sans cycle n'est pas connexe, on peut ajouter une arete ab entre deux de ses sommets a et b qui ne sont pas dans la meme composante connexe, cela ne cree pas de cycle.
Montrons le par l'absurde. Supposons que $G + ab$ contienne un cycle C. Alors C contient necessairement l'arete ab, car G n'a pas de cycle. Or on a montre a la question 2 de l'exo 2 que retirer une arete d'un cycle dans un graphe connexe laisse ce graphe connexe. Ainsi, a et b sont dans la meme composante connexe de G : absurde. Donc $G + ab$ ne contient pas de cycle. D'apres, l'hypothese de la question, il est impossible d'ajouter une arete a G sans creer de cyle. Par consequent, G est donc connexe et G est un arbre.

b. Montrez que si un graphe G est connexe et qu'on ne peut pas lui retirer une arete en preservant cette propriete, alors c'est un arbre.

Solution. On a montre a la question 2 de l'exo 2 que retirer une arete d'un cycle dans un graphe connexe laisse ce graphe connexe. Comme cela n'est pas possible dans G, on en deduit que G n'a pas de cycle. Et comme G est connexe, G est un arbre.

Exercice 4. *Plus courts chemins dans les arbres*

Soit $T = (V, E)$ un arbre. Le but de cet exercice est de montrer qu'entre toute paire de sommets de T, il existe un unique plus court chemin (et meme un unique chemin simple).

a. Soit $ab \in E$ une arete de T. Montrez que a et b ne sont pas dans la meme composante connexe de $T - ab$.

Solution. La preuve a ete faite a la question 3 de l'exo 1. Cela repose sur le fait que T n'a pas de cycle. Ainsi si dans T sans l'arete ab, il y avait encore un chemin entre a et b, alors ce chemin plus l'arete ab formerait un cycle dans T.

b. Soit $u, v \in V$ deux sommets de T et soit P un chemin **simple** entre u et v. Montrez que pour toute arete ab de P et pour tout chemin P' entre u et v, ab est une arete de P'.

Solution. Par l'absurde, supposons qu'il existe un chemin P' entre u et v qui ne contienne pas l'arete ab. Cela implique que u et v sont dans la meme composante connexe de $T - ab$. Or, on a montre a la question 3 de l'exo 1 que a et b ne sont pas dans la meme composante connexe de $T - ab$. Ainsi, comme $P = u.P_1.a.b.P_2.v$ est simple, P_1 et P_2 ne contiennent pas d'occurence de l'arete ab. En consequent, dans $T - ab$, u appartient a la composante connexe de a et v a celle de b. u et v ne sont donc pas dans la meme composante connexe de $T - ab$: contradiction. En conclusion, tout chemin P' entre u et v contient l'arete ab.

c. La proposition de la question precedante est-elle toujours vraie si P n'est pas simple ? Prouvez votre reponse.

Solution. Non, cela n'est pas vrai si P n'est pas simple. D'ailleurs, la preuve precedante utilise bien la simplicite de P. Si P n'est pas simple, on peut construire un contre exemple sur l'arbre de la figure 2, en prenant $u = 4$, $v = 10$, $P = 4, 5, 4, 7, 10$, $P' = 4, 7, 10$, $a = 4$ et $b = 5$. P contient l'arete ab mais P' ne la contient pas.

d. Montrez que quelque soit $u, v \in V(T)$, il existe un unique chemin simple de u a v.

Solution. Soient P et P' deux chemins simples entre u et v. D'apres la question 2 de l'exo 4, comme P est simple, toutes les aretes de P se retrouvent dans P'. Et comme P' est simple, toutes les aretes de P' se retrouvent dans P. On a donc que les ensembles d'aretes de P et P' sont les memes. Comme P et P' sont simples et ont les memes sommets de depart et d'arrivee, cela implique que P et P' sont identiques (exercice : montrez le, par recurrence sur la longueur du chemin).

e. Montrez que dans un graphe G, un plus court chemin entre deux noeuds est toujours simple. Deduisez en que dans un arbre, le plus court chemin entre deux noeuds est unique.

Solution. Montrons qu'un chemin P entre u et v qui n'est pas simple n'est pas un plus court chemin entre u et v. Comme P n'est pas simple, il contient une repetition de sommet, disons x : $P = P_1.x.P_2.x.P_3$. Ainsi le chemin $P' = P_1.x.P_3$ est aussi entre les sommets u et v et est strictement plus court que P. Par contraposee, un plus court chemin est simple. Et comme on a montre a la question 4 de l'exo 4 que, dans un arbre, il existe un unique chemin simple entre deux noeuds quelconques, alors il existe un unique plus court chemin entre deux noeuds quelconques.
Dans un arbre, on dira donc **LE** plus court chemin entre u et v... et seulement dans un arbre ;)

Exercice 5. *Reconnaitre un arbre*

a. Donnez un algorithme pour decider si un graphe est connexe. Quelle est sa complexite ?

Solution. On peut faire un parcours en largeur (BFS) a partir d'un sommet quelconque et verifier que tous les sommets du graphe ont ete visites par le parcours. La complexite de BFS est $O(n + m)$.

b. Donnez un algorithme pour decider si un graphe connexe contient un cycle. Quelle est sa complexite ?

Solution. Si on sait que le graphe est connexe, on peut simplement verifier que tous les arcs traverses dans un parcours en largeur a partir d'un sommet quelconque vont vers un sommet non encore visite. La complexite de BFS est $O(n + m)$ mais si on s'arete au premier arc "retour" rencontre, la complexite devient $O(n)$.

c. Donnez un algorithme pour decider si un graphe est un arbre. Quel est sa complexite ?

Solution. On peut obtenir un algorithme dont la complexite est $O(n)$ en procedant comme suit. On effectue un BFS en s'aretant au premier arc retour rencontre, car s'il y a un arc retour alors G contient un cycle. Si le parcours termine sans avoir rencontre d'arc retour, on verifie si le nombre de sommets decouverts par le parcours et le nombre de sommets dans le graphe sont egaux.

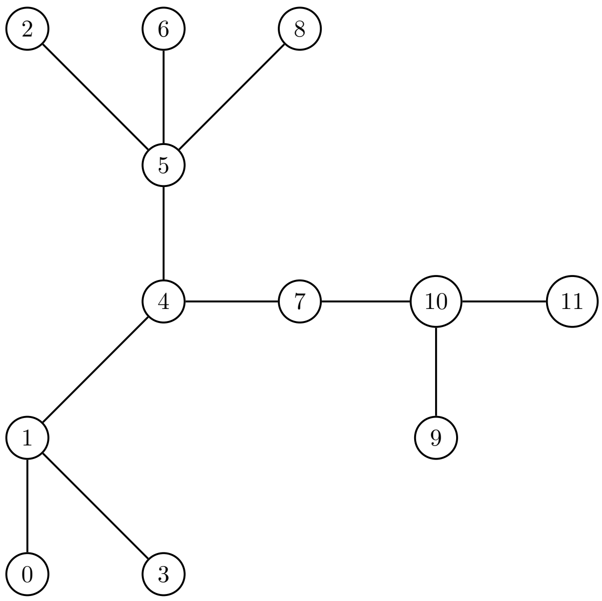

FIGURE 2 – Un arbre a 12 sommets.

Exercice 6.

Par la racine.

Definition [Arbre enracine]
Un **arbre enracine** est un arbre dans lequel on a choisi un noeud **racine** r. Dans un tel arbre, tous les noeuds u distincts de la racine r ont un **pere**, note $p(u)$, qui est le premier noeud intermediaire sur le plus court chemin entre u et la racine. On dit aussi que u est un **fils** de son pere $p(u)$.

a. Comment representer un arbre a n sommets par un tableau de n entiers ?

Solution. On peut donner un identifiant entre 0 et $n-1$ a chacun des n sommets et reserver un tableau *pere* indice par les entiers u de 0 a $n-1$. Dans la case indicee u du tableau pere, on stocke l'identifiant $pere[u]$ du pere $p(u)$ de u, si u est different de la racine r de l'arbre.
Pour la racine r, qui n'a pas de pere, plusieurs conventions sont possibles. On peut par exemple prendre pour $pere[r]$ un entier hors de l'intervalle $[\![0, n-1]\!]$, comme n ou -1 par exemple. Ou on peut prendre la convention que $pere[r] = r$ (r est alors l'unique sommet u tel que $pere[u] = u$).

b. Dessinez ce tableau pour l'arbre de la figure 2 en choisissant le sommet 7 comme racine, puis en choisissant le sommet 2 comme racine.

Solution.

1	4	5	1	7	4	5	–1	5	10	7	10
0	1	2	3	4	5	6	7	8	9	10	11

FIGURE 3 – Representation en tableau de l'arbre de la figure 2 enracine en 7, avec la convention $pere[r] = -1$.

1	4	2	1	5	2	5	4	5	10	7	10
0	1	2	3	4	5	6	7	8	9	10	11

FIGURE 4 – Representation en tableau de l'arbre de la figure 2 enracine en 2, avec la convention $pere[r] = r$.

c. Dans la representation d'un arbre enracine par un tableau, comment decider si deux sommets a et b donnes sont adjacents ? Quelle est la complexite de cette operation ?

Solution. Si a et b sont adjacents, alors a est le pere de b ou b est le pere de a (la reciproque est en evidence). On peut donc verifier $pere[a]$ et $pere[b]$, cela prend un temps constant.

d. Toujours dans cette meme representation, comment lister tous les fils d'un sommet a donne ? Quelle est la complexite de cette operation ?

Solution. Cette operation est plus couteuse. Il faut trouver tous les neouds u tels que $pere[u] = a$. Une facon de faire est de parcourir entierement le tableau et de retenir la liste des indices u tels que $pere[u] = a$. Cela prend un temps $\Theta(n)$.

Exercice 7. *Listes d'adjacence*

a. Donnez la representation par listes d'adjacence de l'arbre de la figure 2.

Solution.

b. Quelle est la taille de la representation par listes d'adjacence d'un arbre a n sommets ?

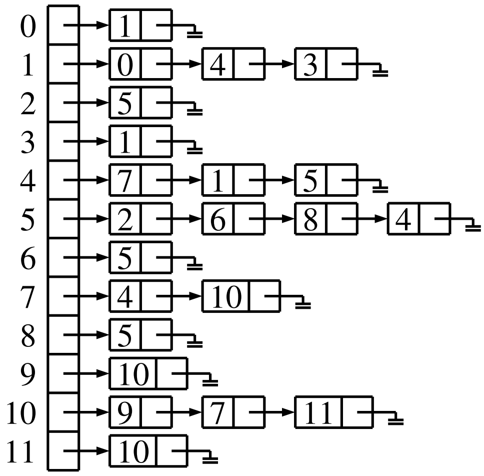

FIGURE 5 – Listes d'adjacence de l'arbre de la figure 2.

Solution. Dans la representation d'un graphe par listes d'adjacence, il y a toujours un tableau de taille n et pour chaque sommet u une liste de taille $d^\circ(u)$, le degre de u. La taille totale de la representation est donc $n + \sum_{u \in V} d^\circ(u) = n + 2m$. Comme dans un arbre, $m = n - 1$, on obtient que la taille des listes d'ajacences d'un arbre est $3n - 2 = \Theta(n)$ entiers. Remarquez qu'a une constante pres, cela est equivalent a la taille de la representation sous forme de tableau.

c. Quel est le degre moyen d'un arbre a n sommets lorsque n tend vers $+\infty$?

Solution. Le degre moyen dans un graphe est $\frac{\sum_{u \in V} d^\circ(u)}{n} = \frac{2m}{n}$. Comme dans un arbre, $m = n - 1$, on obtient que le degre moyen est $2 - \frac{2}{n} \to 2$ quand $n \to +\infty$.

d. Quelle est la densite d'un arbre a n sommets lorsque n tend vers $+\infty$?

Solution.
La densite δ dans un graphe est defini comme $\frac{m}{\frac{n(n-1)}{2}}$ (exercice : montrez que cela est egal a $\frac{d^\circ_m oy}{n-1}$, ou $d^\circ_m oy$ est le degre moyen). Dans le cas d'un arbre, comme $m = n - 1$, on obtient $\delta = \frac{2}{n} \to 0$ quand $n \to +\infty$.

Exercice 8. *Et le vent du nord les emporte.*

Definition [Feuille]
Une **feuille** dans un arbre enracine est un noeud qui n'a aucun fils. Une **feuille** dans un arbre non-enracine est un noeud qui a exactement un voisin. Dans les deux sortes d'arbres, un noeud interne est un noeud qui n'est pas une feuille.

a. Quel est le nombre minimum de feuilles dans un arbre enracine a n sommets ? Et le nombre maximum ?

Solution. Le nombre minimum de feuilles dans un arbre enracine est 1 (un chemin enracine en une de ses extremites) et le nombre maximum est $n - 1$ (une racine avec tous ses fils qui sont des feuilles).

b. Memes questions pour un arbre non-enracine ?

Solution. Si T a au moins 2 noeuds, alors le nombre minimum de feuilles est 2 (un chemin). Le nombre maximum de feuilles est encore $n-1$ (un noeud central relie a tous les autres).

On considere maintenant un arbre enracine T ayant f feuilles et dont tout noeud interne a au moins deux fils.

c. Quel est le nombre maximum possible de noeuds dans T ? Demontrez votre reponse.

Indication. Vous pourrez determiner l'expression exacte de ce nombre sur un cas particulier.

Indication. Pour prouver cette expression dans le cas general, vous pourrez proceder par induction sur la hauteur de l'arbre.

Solution. Puisque le nombre de feuilles est fixe, maximiser le nombre de noeuds est equivalent a maximiser le nombre de noeuds internes. Pour voir comment faire cela, il faut remarquer que chaque noeud interne unifie tous ses sous-arbres fils, jusqu'a ce que tout soit unifie, a la racine. Ainsi, afin de maximiser le nombre de noeuds internes, il faut unifier le minimum possible de sous arbre a chaque noeud interne : c'est a dire 2, puisque chaque noeud interne doit avoir au moins deux fils. Lorsque chaque neoud interne a exactement deux fils, on obtient toujours exactement $f-1$ noeuds internes. Cela donne le maximum du nombre de neouds possible : $f+f-1=2f-1$.

Faisons une preuve plus formelle. D'abord, montrons qu'un arbre a f feuilles dans lequel chaque noeud interne a au plus deux fils, on dit au moins binaire, a au plus $2f-1$ noeuds. On procede par recurrence sur la heuteur h de l'arbre. C'est vrai pour $h=0$: un arbre de hauteur 0 a $f=1$ feuille et $2f-1=1$ noeud.
Supposons que la propriete est vrai pour $h \leq k$ et montrons qu'elle l'est encoe pour $h=k+1$. Soit T un arbre au moins binaire de hauteur $k+1$. Soit $T_1, T_2, \ldots, T_k$ les sous arbres enracines en chacun des k fils, avec $k \geq 2$, de la racine de T. Pour tout $i \in [\![1, k]\!]$, T_i est de hauteur au plus k, donc par hypothese de recurrence, $|V(T_i)| \leq 2f(T_i)-1$, avec $f(T_i)$ le nombre de feuilles de T_i. Ainsi, on obtient $|V(T)| = 1+\sum_{1\leq i\leq k} |V(T_i)| \leq 1+2\sum_{1\leq i\leq k} f(T_i)-k = 1+2f(T)-k \leq 2f(T)-1$, car $k \geq 2$. Cela acheve la recurrence et prouve que pour tout arbre au moins binaire T, on a $|V(T)| \leq 2f(T)-1$, avec $f(T)$ le nombre de feuilles de T.
Pour repondre a la question posee en exercice, il nous reste a exhiber une famille infinie d'arbres T_j realisant l'egalite $|V(T_j)| = 2f(T_j)-1$. Une telle famille est par exemple obtenue en prenant pour T_j l'arbre binaire complet de hauteur j. T_j possede $f(T_j) = 2^j$ feuilles et $\sum_{0\leq i\leq j} 2^i = 2^{j+1}-1 = 2\cdot 2^j - 1 = 2f(T_j)-1$ feuilles.