Graphes et complexité d'algorithmes

CM nº1 — Algorithmique (AL5)

Matěj Stehlík

22 septembre 2023

Équipe pédagogique

TD INF1+BIO	Roberto Mantaci	<roberto.mantaci@irif.fr>
TD INF2+JAP	Juliusz Chroboczek	<jch@irif.fr>
TD INF3	Emily Clement	<emily.clement@irif.fr>
TD INF4	Yan Jurski	<jurski@irif.fr>
TD INF5	Mónika Csikós	<csikos@irif.fr>
TD MI1	Mikaël Rabie	<mikael.rabie@irif.fr>
CM	Matěj Stehlík	<matej@irif.fr>

Volumes horaires

Cours magistraux (CM)	24h	*2h par semaine*
Travaux dirigés (TD)	24h	*2h par semaine*

Transparents du cours + feuilles de TD seront affichés sur Moodle

Évaluation

Contrôle continu intégral (pas de session 2)

I	note moyenne de 2 interrogations TD
ET	note épreuve terminale (janvier)
SC	note seconde chance (juin)

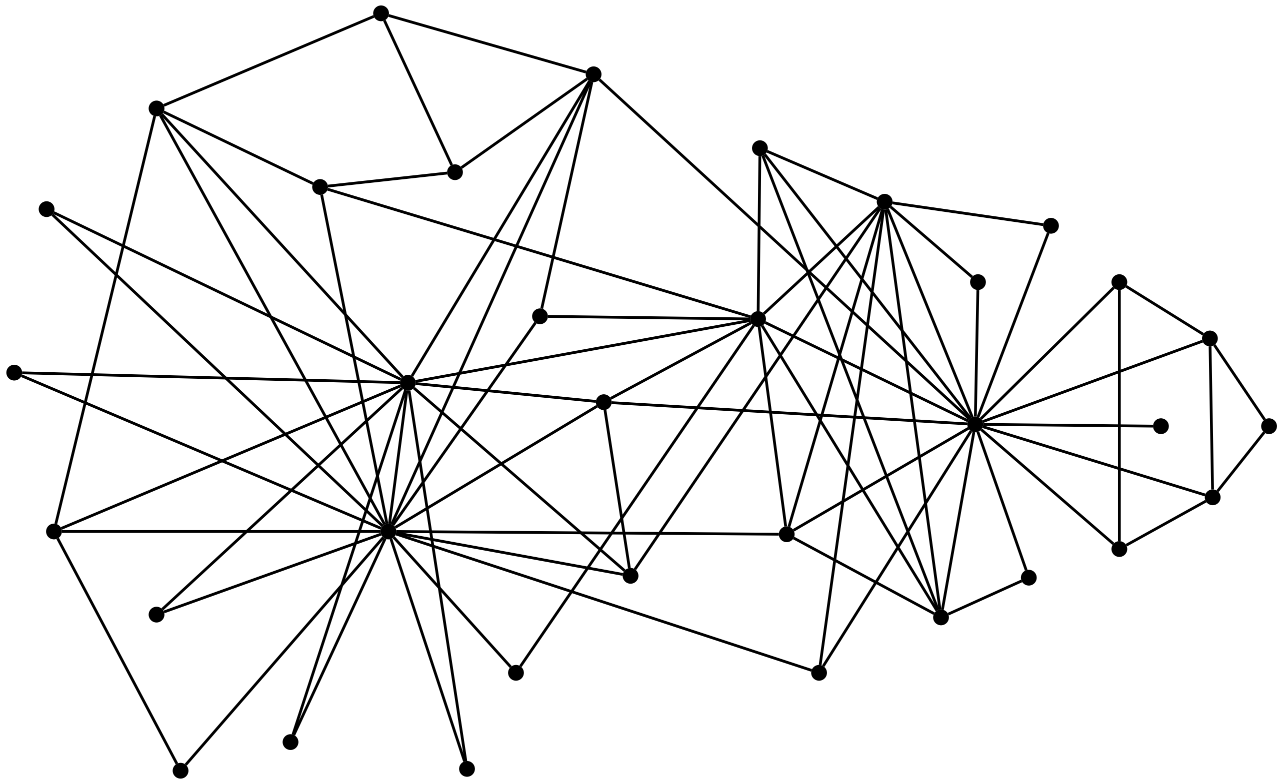
$$\text{note finale} = \max\{\mathrm{SC}, 0.5 \times \mathrm{I} + 0.5 \times \mathrm{ET}\}$$

Objectifs

Le cours présente les algorithmes des graphes, plus particulièrement

- algorithmes d'exploration
  - parcours en largeur
  - parcours en profondeur
- algorithmes d'optimisation
  - arbre couvrant minimum
  - plus court chemin
  - couplage maximum
  - flot maximum. . .

Exemple d’un graphe



Graphes

- Soit X un ensemble.
- On note $\binom{X}{2}$ l'ensemble des parties à deux éléments de X.
- En général, on notera uv la partie $\{u, v\}$.
- L'ordre et les répétitions ne sont pas pris en compte : $12 = 21$.

Exemple

Si $X = \{1, 2, 3\}$, alors $\binom{X}{2} = \{12, 13, 23\}$.

Définition

- Un *graphe* est un couple $G = (V, E)$ formé par un ensemble fini V et un sous-ensemble E de $\binom{V}{2}$.
- V est l'ensemble des *sommets* de G (on le note aussi $V(G)$).
- E est l'ensemble des *arêtes* de G (on le note aussi $E(G)$).

À quoi ça sert ?

- Les sommets modélisent des “objets”
  - personnes
  - pages web
  - neurones
  - aéroports. . .
- Les arêtes modélisent des “relations” (binaires) entre ces objets
  - amitiés
  - hyperliens
  - connexions synaptiques
  - vols. . .
- Les arêtes peuvent être
  - non-orientées
  - orientées (dans ce cas on parle de graphes orientés)

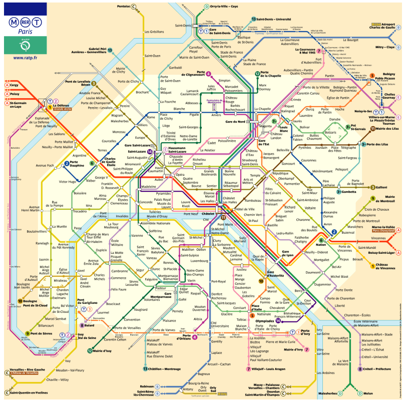
Quelques applications des graphes

Les graphes sont très utilisés dans :

- les problèmes de routage en réseau,
- les problèmes de trafic en transport,
- l'étude des jeux,
- la recherche d'information (graphe du web)
- codage
- ordonnancement et emploi du temps
- . . .

Quelques exemples de graphes

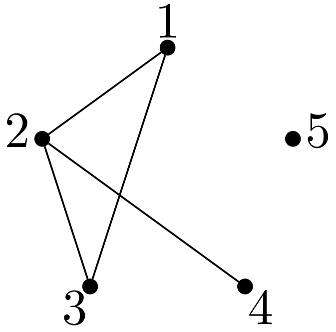
- $(\{1,2,3,4,5\},\{12,13,23,24\})$.
- Le métro : $(\{\text{stations}\},\{\text{stations directement reliées}\})$.



- L'internet : $(\{\text{pages web}\},\text{(hyper-)liens})$.
- Facebook : $(\{\text{utilisateurs}\},\{\text{amitiés}\})$.
- Molécules. $V=\{\text{atomes}\}$, $E=\{\text{atomes partageant des électrons}\}$.

Représentation graphique

- On représente chaque sommet par un disque : ●
- Pour représenter une arête uv, on trace un trait entre les disques correspondants à u et à v.



Remarques

- La forme des « disques » et des « traits » n'a aucune importance (sauf pour la lisibilité de la figure).
- Ce qui compte, c'est de traduire graphiquement s'il y a une arête entre deux sommets ou non.

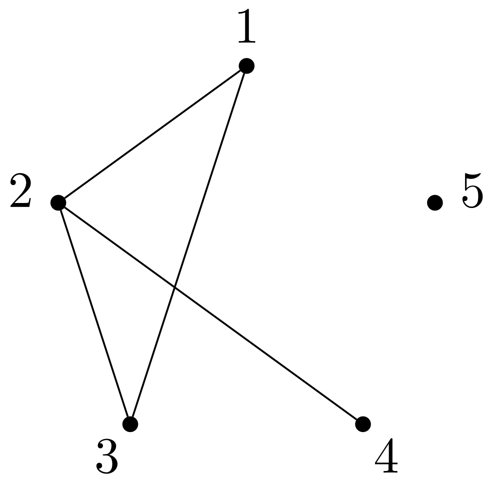
Adjacence et incidence

Définition

Soient G un graphe, u et v deux sommets de G et e une arête de G.

- u et v sont *adjacents* (ou *voisins*) si $uv \in E(G)$;
- e est *incidente* à u si $u \in e$;
- les deux éléments de e sont ses *extrémités* ;
- le *voisinage* de u dans G est l'ensemble $N_G(u)$ des sommets de G adjacents à u ;
- les *voisins* de u sont les éléments de $N_G(u)$;
- l'ensemble des arêtes incidentes à u est noté $\delta_G(u)$.

Exemple



- 1 et 2 sont adjacents
- 4 est voisin de 2
- $N(2) = \{1, 3, 4\}$
- $N(5) = \emptyset$ (5 est *isolé*)
- l'arête 12 est incidente à 2

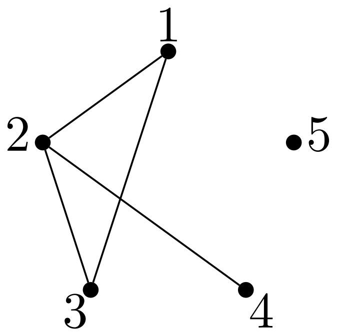
Sous-graphes

Définition

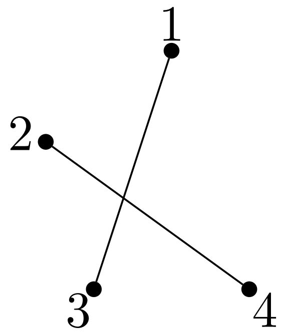
Soient $G = (V, E)$ et $H = (W, F)$ deux graphes.

- H est un *sous-graphe* de G si $W \subseteq V$ et $F \subseteq E$.
- H est un *sous-graphe couvrant* de G si $W = V$ et $F \subseteq E$.
- H est un *sous-graphe induit* de G si $W \subseteq V$ et F contient toutes les arêtes $uv \in E$ où $u, v \in W$. On le note $G[W]$.

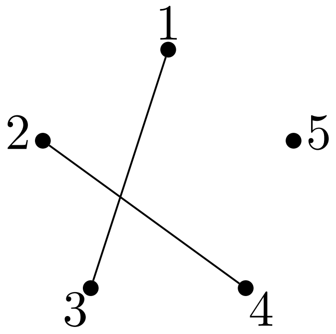
Illustration des différents types de sous-graphe



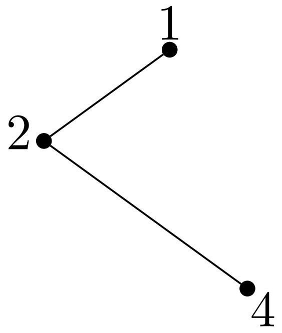
G



sous-graphe de G



sous-graphe couvrant de G



sous-graphe induit de G

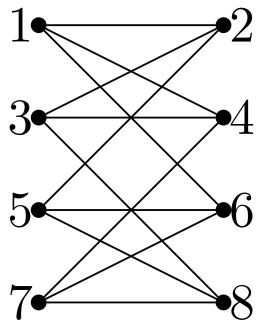
Isomorphismes

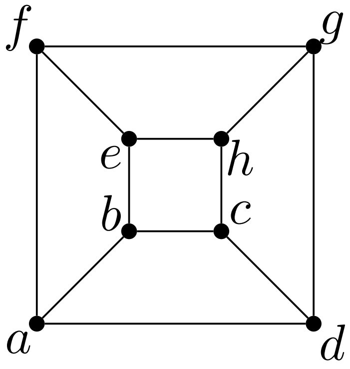
- Souvent, on ne fera pas de distinction entre deux graphes ayant « la même forme », c'est-à-dire : qu'on ne peut les distinguer si l'on oublie les noms de leurs sommets.

Définition

- Soient G, H deux graphes.
- On dit que G est *isomorphe* à H s'il existe une bijection f de $V(G)$ sur $V(H)$ telle que pour toute paire xy de sommets de G, on a $xy \in E(G)$ si et seulement si $f(x)f(y) \in E(H)$.

Illustration





- La relation d'isomorphisme est une relation d'équivalence.

Remarque

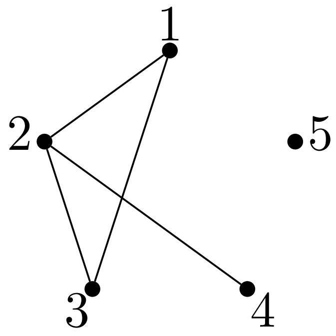
Il n'est pas toujours facile, à partir de représentations graphiques, de décider si deux graphes sont isomorphes.

Graphe complémentaire

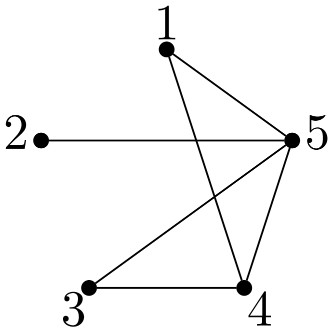
Définition

Soit $G = (V, E)$ un graphe. Le *graphe complémentaire* $\overline{G}$ est défini comme $\overline{G} = (V, \binom{V}{2} \setminus E)$.

- C'est-à-dire, les arêtes de G sont les non-arêtes de $\overline{G}$, et vice versa.



G



$\overline{G}$

Représentation matricielle et par listes

- Il y a plusieurs façons de représenter un graphe en mémoire de l'ordinateur.
- On va en voir trois :
  1. matrice d'adjacence
  2. matrice d'incidence
  3. liste d'adjacence
- En général, chacune est plus ou moins adaptée au problème considéré et possède des avantages/inconvénients notamment par rapport à la densité (en arêtes) du graphe.

Matrices d'adjacence

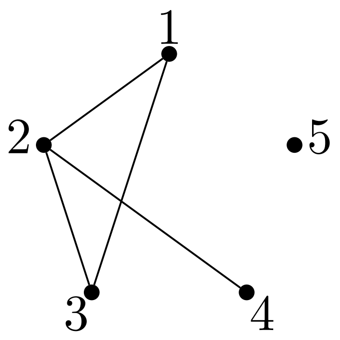
Définition

- Soit G un graphe à n sommets.
- On numérote les sommets $V(G) = \{v_1, \ldots, v_n\}$.
- La *matrice d'adjacence* de G (pour la numérotation choisie) est la matrice M carrée $n \times n$ sur $\{0, 1\}$ définie par :

$$M_{ij} = 1 \text{ si et seulement si } v_i v_j \in E(G).$$

Remarque

- M est *symétrique* et nulle sur la diagonale.



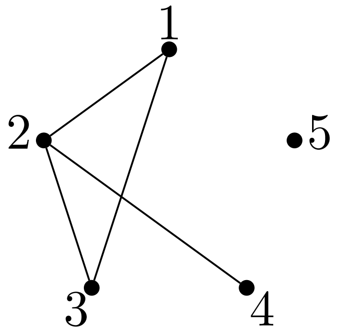
$$M = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Matrices d'incidence

Définition

- Soit G un graphe à n sommets et m arêtes.
- On numérote les sommets $V(G) = \{v_1, \ldots, v_n\}$ et les arêtes $E(G) = \{e_1, \ldots, e_m\}$.
- La *matrice d'incidence* de G est la matrice N sur $\{0,1\}$ de taille $n \times m$ définie par :

$$N_{ij} = 1 \text{ si et seulement si } v_i \in e_j.$$

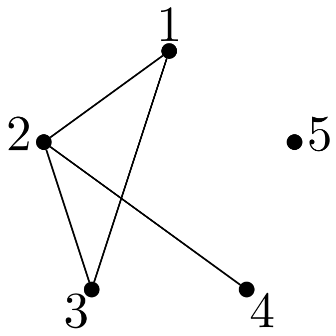


$$N = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Remarque

- La somme de chaque colonne vaut 2.

Listes d'adjacence



Définition

- Soit G un graphe.
- Une représentation en *liste d'adjacence* de G est la donnée, pour chaque sommet v de G, de la liste des voisins de v.

$1 :\ [2, 3]$

$2 :\ [1, 3, 4]$

$3 :\ [1, 2]$

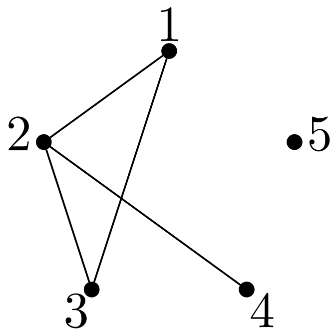
$4 :\ [2]$

$5 :\ []$

Degrés

Définition

- Soient G un graphe et v un sommet de G.
- Le *degré* de v dans G, noté $d_G(v)$, est le nombre d'arêtes de G incidentes à v.
- C'est aussi (par simplicité des graphes définis dans ce cours) le nombre de voisins de v : $d_G(v) = |N_G(v)|$.
- Si $d_G(v) = 0$ on dit que v est *isolé*.
- Si $d_G(v) = 1$ on dit que v est une *feuille*.



- $d(1) = 2$
- $d(2) = 3$
- $d(3) = 2$
- $d(4) = 1$
- $d(5) = 0$

La somme des degrés

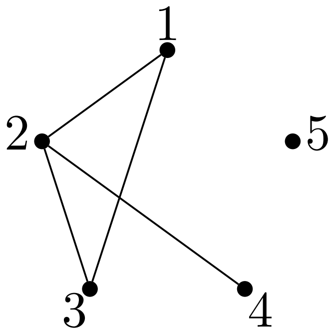
Théorème

Soit G un graphe, alors $\sum_{v \in V(G)} d_G(v) = 2|E(G)|$.

Démonstration

- Soit S la somme de tous les éléments de la matrice d'incidence de G.
- La somme de chaque ligne est égale au degré du sommet correspondant, donc $S = \sum_{v \in V(G)} d_G(v)$.
- La somme de chaque colonne est égale à deux, et on a $|E(G)|$ colonnes, donc $S = 2|E(G)|$.

Illustration



$$N = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$S = \sum_{v \in V(G)} d(v) = 2 + 3 + 2 + 1 + 0 = 8$$

$$S = 2|E(G)| = 2 \cdot 4 = 8$$

Une conséquence

Corollaire

- Soit G un graphe. La somme $\sum_{v \in V(G)} d_G(v)$ est paire.
- Autrement dit, le nombre de sommets de degré impair est pair.

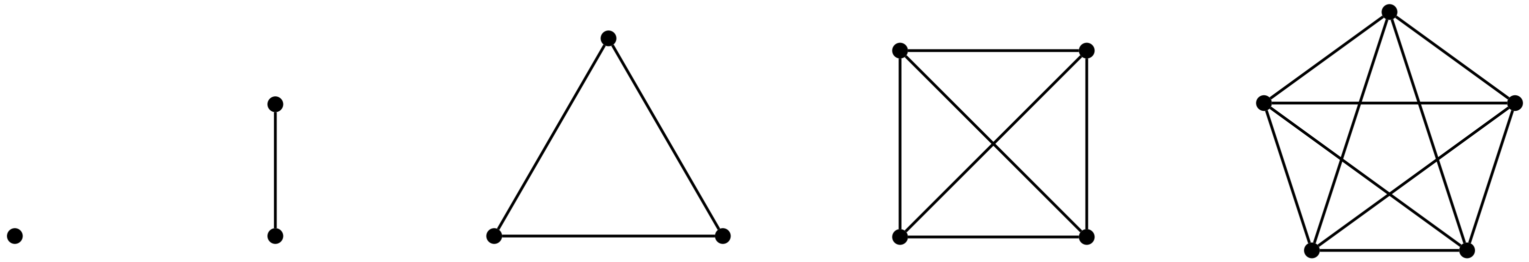
Exemple

Sept personnes participent à une fête. Est-il possible que chacun d'entre eux serre la main de trois autres personnes exactement ?

Graphes complets

Définition

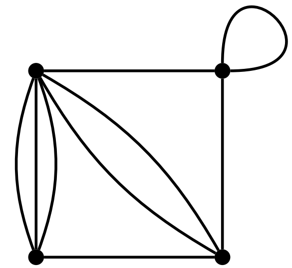
- Soit $n \geq 1$ un entier.
- Le *graphe complet* à n sommets est le graphe $\left(\{1, \ldots, n\}, \binom{\{1,\ldots,n\}}{2}\right)$.
- Il est noté K_n.



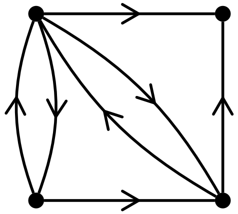
K_1 K_2 K_3 K_4 K_5

Variants des graphes

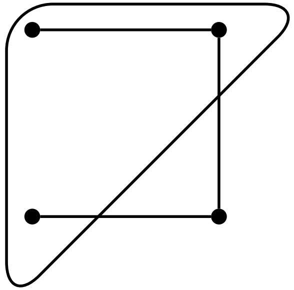
- Un *multigraphe* est un graphe auquel on permet d'avoir plus d'une arête entre deux sommets ou une arête dont les deux extrémités sont identiques (boucles).



- Un graphe *orienté* est obtenu à partir d'un graphe en ordonnant, pour chaque arête, ses extrémités. Autrement dit, chaque arête est dirigée vers une de ses extrémités.



- Dans un *hypergraphe*, les (hyper-)arêtes peuvent être incidentes à un nombre arbitraire de sommets (et pas seulement à deux comme dans le cas des graphes).



Opérations élémentaires

Définition

Une *opération élémentaire* est une opération qui s'effectue en temps constant sur tous les calculateurs usuels.

On considérera les opérations suivantes comme élémentaires :

- Affectation ;
- Comparaisons ;
- Opérations arithmétiques et logiques ;
- Accès à une case d'un tableau ;
- Appel d'une sous-routine ;
- ...

Complexité temporelle

Définition

La *complexité temporelle* (dans le pire cas) d'un algorithme A, noté $T(n)$, est le nombre d'opérations élémentaires maximum que puisse effectuer A avant d'arriver à un résultat, étant donné une entrée de taille n.

- $T(n)$ s'exprime en fonction de la taille n de l'entrée.
- pour un graphe, on compte la complexité en fonction du *nombre de sommets* n, et éventuellement du *nombre d'arêtes* m.
- donc n n'est pas ici exactement la taille de l'entrée, mais les deux sont reliés polynomialement.

Remarques sur la complexité temporelle

- Les études de complexité portent dans la majorité des cas sur le comportement *asymptotique*, lorsque la taille des entrées tend vers l'infini, et l'on utilise couramment les notations grand O.
- La complexité temporelle est la mesure la plus courante en algorithmique ; on parle parfois simplement de la complexité d'un algorithme
- Il existe d'autres mesures comme la complexité spatiale.

Exemple

Entrées : graphe G à n sommets sous forme de matrice d'adjacence A
Sorties : degré moyen de G
début

```
somme_degre ← 0 ;
pour i de 0 à n − 1 faire
    pour j de 0 à n − 1 faire
        somme_degre ← somme_degre+A[i][j] ;
Retourner (somme_degre/n) ;
```

La notation grand O

Définition

- Soient $f(n)$ et $g(n)$ des fonctions de $\mathbb{R}$ vers $\mathbb{R}$.
- On écrit $f \in O(g)$ (ou plus souvent $f = O(g)$) s'il existe une constante $c > 0$ telle que $|f(x)| \leq c \cdot |g(x)|$ pour tout $x \in \mathbb{R}$ suffisamment grand.

Remarques

- $f \in O(g)$ veut dire que f n'augmente pas plus vite que g.
- $f \in O(g)$ est moins fort que $f \leq g$.
- La différence vient de la constante c ; par exemple, $100n \in O(n)$.
- Cette constante nous permet d'ignorer ce qui se passe pour des petites valeurs de n.

La notation grand O : exemple

- Supposons que nous devrons choisir entre deux algorithmes A_1 et A_2 pour une certaine tâche, de complexité $T_1(n) = n^2$ et $T_2(n) = 300n + 700$, respectivement.
- T_2 se comporte mieux quand n augmente ; A_2 est meilleur.
- $T_2 \in O(T_1)$, parce que

$$\frac{T_2(n)}{T_1(n)} = \frac{300n + 700}{n^2} \leq 1000$$

  pour tout $n \geq 1$.
- Par contre, $T_1 \notin O(T_2)$, car

$$\frac{T_1(n)}{T_2(n)} = \frac{n^2}{300n + 700}$$

  tend vers l'infini quand n tend vers l'infini.

La notation grand O : exemple

- Supposons qu'il y a un autre algorithme A_3 de complexité $T_3(n) = n$.
- La différence entre T_2 et T_3 est minuscule comparé à la différence énorme entre T_1 et T_2.
- Donc, on considère deux fonctions comme équivalentes si elles ne diffèrent que par une constante multiplicative.
- On remarque que $T_2 = O(T_3)$:

$$\frac{T_2(n)}{T_3(n)} = \frac{300n + 700}{n} \leq 1000.$$

- On a aussi $T_3 = O(T_2)$, avec $c = 1$.

La notation Ω et Θ

Définition

De la même manière que $O(\cdot)$ est un analogue de $\leq$, nous pouvons aussi définir des analogues de $\geq$ et de $=$ comme suit :

$$f \in \Omega(g) \text{ veut dire } g \in O(f)$$
$$f \in \Theta(g) \text{ veut dire } f \in O(g) \text{ et } f \in \Omega(g).$$

Règles pour simplifier les fonctions dans $O(\cdot)$

Omettre les termes dominés par d'autres termes. En particulier :

- Omettre les constantes multiplicatives : $25n^3$ domine n^3.
- n^a domine n^b si $a > b$: par exemple, n^2 domine n.
- Les fonctions exponentielles dominent les polynômes : 2^n domine n^{100}.
- Les polynômes dominent les logarithmes : n domine $(\log n)^3$.