Structures de données (LU2IN006)

Graphe et parcours

Nawal Benabbou

Licence Informatique - Sorbonne Université

2022-2023

Structure non-linéaire et cyclique : les graphes

Question

Comment manipuler des structures non-linéaires et pouvant posséder des cycles ?

Exemples

- La représentation d'un catalogue de formations, ou des appartenances d'une association (cf TD5).
- Une carte routière, un schéma des transports publics...
- Le plan de transport d'une entreprise entre ses usines, ses entrepôts et ses clients.
- La représentation des hyper-liens d'internet.
- La représentation des liaisons électriques entre les composants électroniques d'un circuit intégré.
- Les liens logiques entre les idées dans un texte.
- Les déplacements possibles d'un robot.

Graphe orienté (objet mathématique - théorie des graphes)

Définition : graphe orienté

Un *graphe orienté* est un couple $G = (S, A)$ où :
- S est un ensemble d'éléments appelés *sommets*,
- A est un ensemble de paires (orientées) de sommets de $S \times S$ appelées *arcs*.

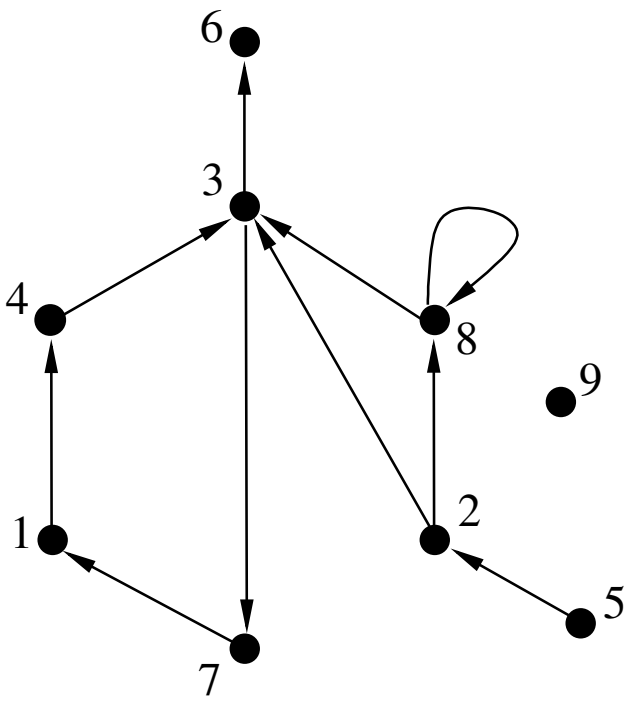
Définition : sommet

Dans un graphe $G = (S, A)$, l'ensemble S est l'ensemble des sommets. Le nombre de sommets est souvent noté n (c-à-d $n = |S|$).
- Un sommet est représenté graphiquement par un rond ou un carré numéroté.
- Le numéro (ou la valeur) d'un nœud est donné à l'intérieur du rond ou à côté.

Définition : arc

Dans un graphe $G = (S, A)$, l'ensemble A est l'ensemble des arcs. Le nombre d'arcs est souvent noté m (c-à-d $m = |A|$).
- Un arc est une paire (orientée) de sommets (appelés les *extrémités* de l'arc).
- Un arc $a \in A$ peut être noté (s_i, s_j) où s_i et s_j sont les extrémités de l'arc.
- Si $(s_i, s_j) \in A$, alors on dit que s_i est un *prédécesseur de* s_j dans le graphe, et que s_j est un *successeur* de s_i.
- L'arc (s_i, s_j) est représenté graphiquement par une flèche allant de s_i vers s_j.

Exemple



Dans cet exemple, on voit que :

- le sommet 5 n'a pas de prédécesseur.
- le sommet 6 n'a pas de successeur.
- le sommet 3 a trois prédécesseurs et deux successeurs.
- le sommet 8 est son propre prédécesseur et successeur.
- le sommet 9 n'a ni prédécesseur, ni successeur (on dit qu'il est *isolé*).

Interprétations

Sommet

Les sommets représentent souvent des lieux (villes, usines) ou des objets (pièces à manipuler en usine, ordinateur). Un sommet est très souvent associé à des données :

- un entier (ou un nom) pour le repérer.
- un ensemble d'informations sur sa description, comme une couleur, des coordonnées, ou encore l'objet qu'il représente.
- un valeur numérique pouvant représenter un coût, un poids, une utilité...

Arc

Les arcs représentent souvent des déplacements, des choix, des liens ou des distances. Un arc est parfois associé à des données :

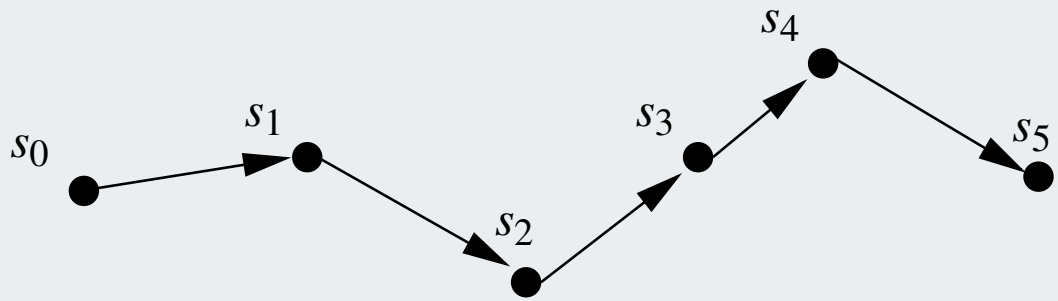
- un entier (ou un nom) pour le repérer.
- un ensemble d'informations sur sa description.
- une valeur numérique représentant une distance, un coût, une capacité...

Chemin dans un graphe orienté

Définition : chemin

Dans un graphe $G = (S, A)$, un chemin est une suite d'arcs reliant des sommets successifs. Plus formellement, un chemin P est une séquence d'arcs $P = (a_1, a_2, \ldots, a_k)$ qui vérifie :

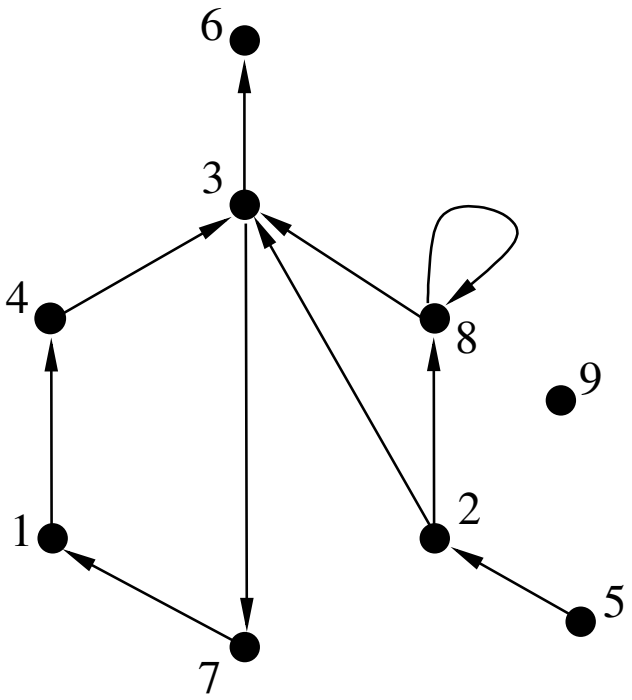
- $a_1 = (s_0, s_1)$,
- $a_2 = (s_1, s_2)$,
- $\ldots$
- $a_k = (s_{k-1}, s_k)$.



Définitions : descendant, ascendant, sommet inaccessible

- Le sommet s_i est un *descendant* du sommet s_j s'il existe un chemin allant de s_j à s_i dans le graphe.
- Le sommet s_i est un *ascendant* (ou ancêtre) du sommet s_j s'il existe un chemin allant de s_i à s_j dans le graphe.
- Le sommet s_i est dit *inaccessible* depuis s_j s'il n'existe aucun chemin allant de s_j à s_i dans le graphe.

Exemple



Dans cet exemple, il y a plusieurs chemins.
Par exemple :

- $P_1 = ((5,2),(2,8),(8,3),(3,6))$
- $P_2 = ((4,3),(3,7))$

On peut voir aussi que :

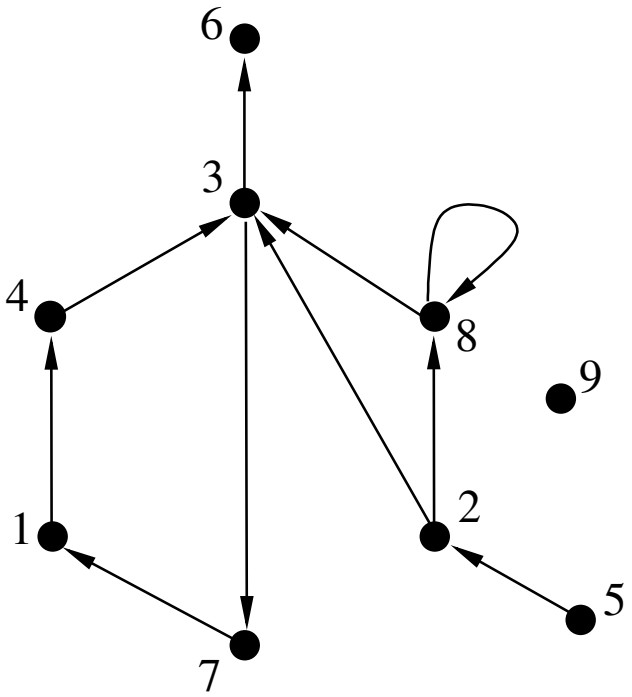
- 6 est descendant de 5 (chemin P_1), mais que l'inverse n'est pas vrai.
- 2 est inaccessible depuis 3.

Circuit

Définition

Un circuit est un chemin reliant un sommet à lui-même.

Remarque : Les sommets d'un circuit sont donc tous ascendants et descendants les uns des autres !



Dans cet exemple, il existe un circuit :
$C = ((7,1),(1,4),(4,3),(3,7))$

Type de données abstrait : Graphe

Définition

Le type de données abstrait “graphe” permet de représenter l'objet mathématique appelé “graphe orienté” (que l'on vient de présenter). Ce type de données contient les sommets et les arcs (ainsi que leurs données associées), et possède différentes opérations :

- ajouter ou de supprimer un arc ou un sommet.
- tester si un sommet est successeur d'un autre,
- lister tous les sommets successeurs d'un sommet,
- etc.

Observation

Le type abstrait “graphe” permet de donner une description d'une structure de données non linéaire et pouvant posséder des cycles. Ce type de données abstrait est donc la généralisation de toutes les structures de données possibles en mémoire d'un ordinateur.

Passage à la structure de données

Remarque : l'implémentation dépend des opérations à réaliser

Le but du type de données abstrait "graphe" n'est pas simplement de stocker des données, mais plutôt de répondre efficacement à des questions précises sur sa structure. On doit donc choisir une implémentation de manière à pouvoir répondre efficacement aux questions posées.

Par exemple, pour trouver rapidement les plus courts chemins, il n'est pas possible de stocker tous les chemins du graphe dans la structure de données pour les comparer, car ils peuvent être en nombre exponentiel.

On peut distinguer deux cas, suivants que les données sont :

- dynamiques (ajout et suppression de sommets/arcs fréquents). Par exemple, si le graphe représente un réseau social, alors les sommets et les arcs changent relativement beaucoup.
- peu dynamiques (c'est le plus fréquent). Par exemple, si les sommets représentent des stations de métro, cet ensemble change peu.

Implémentation

Quels points d'accès ?

Dans une liste chaînée, il suffit d'avoir un accès au premier élément de la liste, pour pouvoir ensuite atteindre tous les autres éléments (pareil pour les arbres). Dans le cas d'un graphe, on ne peut pas se contenter d'une entrée unique car certains sommets sont inaccessibles à partir d'autres.

Plusieurs implémentations possibles

Dans un graphe, il faut pouvoir accéder directement à chaque sommet. On va donc utiliser une structure (matrice, tableau, liste...) pointant sur chaque sommet du graphe. Il existe deux implémentations classiques :

- Matrice d'adjacence.
- Liste d'adjacence.

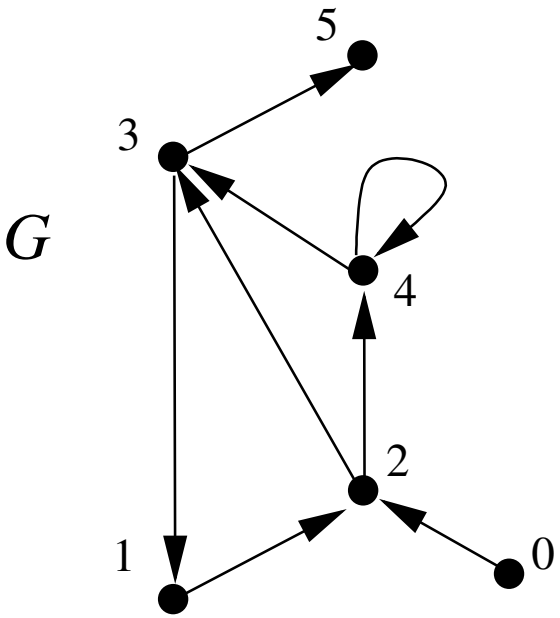
Il existe d'autres implémentations (matrice d'incidence, liste d'arcs) qui peuvent être préférables lors de l'implémentation de certains algorithmes.

Implémentation par matrice d'adjacence

Matrice d'adjacence

Une matrice d'adjacence est une matrice où chaque case (i,j) représente un emplacement possible pour un arc entre les sommets (s_i, s_j). Selon les données à stocker, on peut avoir dans chaque case :

- un booléen qui a pour valeur `vrai` si l'arc existe, et `faux` sinon.
- un `struct arc` permettant de stocker toutes les données de l'arc.



Matrice d'adjacence de G :

	0	1	2	3	4	5
0	0	0	1	0	0	0
1	0	0	1	0	0	0
2	0	0	0	1	1	0
3	0	1	0	0	0	1
4	0	0	0	1	1	0
5	0	0	0	0	0	0

Implémentation par matrice d'adjacence

Avantages

Cette implémentation possède des avantages :

- Tester si s_i est successeur de s_j est en

Implémentation par matrice d'adjacence

Avantages

Cette implémentation possède des avantages :

- Tester si s_i est successeur de s_j est en $\Theta(1)$.
- Tester si s_i prédécesseur de s_j est en

Implémentation par matrice d'adjacence

Avantages

Cette implémentation possède des avantages :

- Tester si s_i est successeur de s_j est en $\Theta(1)$.
- Tester si s_i prédécesseur de s_j est en $\Theta(1)$.
- Lister les successeurs (ou prédécesseurs) de s_i est en

Implémentation par matrice d'adjacence

Avantages

Cette implémentation possède des avantages :

- Tester si s_i est successeur de s_j est en $\Theta(1)$.
- Tester si s_i prédécesseur de s_j est en $\Theta(1)$.
- Lister les successeurs (ou prédécesseurs) de s_i est en $\Theta(n)$.

Défauts

Cette implémentation possède de nombreux défauts :

Implémentation par matrice d'adjacence

Avantages

Cette implémentation possède des avantages :

- Tester si s_i est successeur de s_j est en $\Theta(1)$.
- Tester si s_i prédécesseur de s_j est en $\Theta(1)$.
- Lister les successeurs (ou prédécesseurs) de s_i est en $\Theta(n)$.

Défauts

Cette implémentation possède de nombreux défauts :

- n^2 cases en mémoire même pour les matrices creuses (c-à-d contenant beaucoup de zéro). Pas très efficace auand $m << n^2$.

Implémentation par matrice d'adjacence

Avantages

Cette implémentation possède des avantages :

- Tester si s_i est successeur de s_j est en $\Theta(1)$.
- Tester si s_i prédécesseur de s_j est en $\Theta(1)$.
- Lister les successeurs (ou prédécesseurs) de s_i est en $\Theta(n)$.

Défauts

Cette implémentation possède de nombreux défauts :

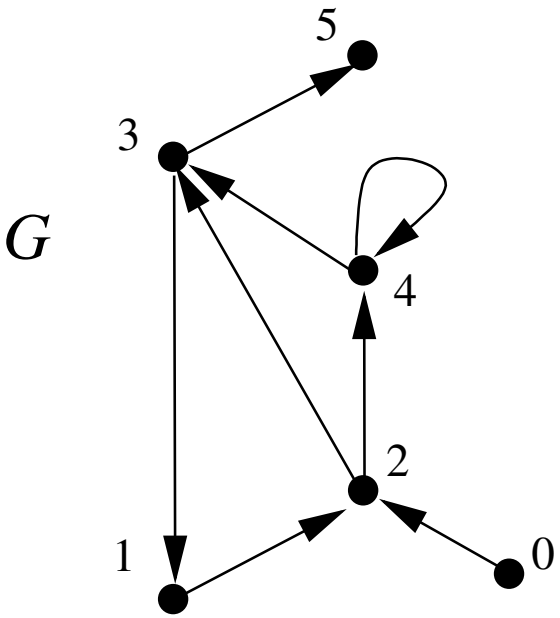
- n^2 cases en mémoire même pour les matrices creuses (c-à-d contenant beaucoup de zéro). Pas très efficace auand $m << n^2$.
- Les successeurs de s_i sont donnés par la ligne i de la matrice et cette ligne contient toujours n cases, même pour les sommets avec peu de successeurs.

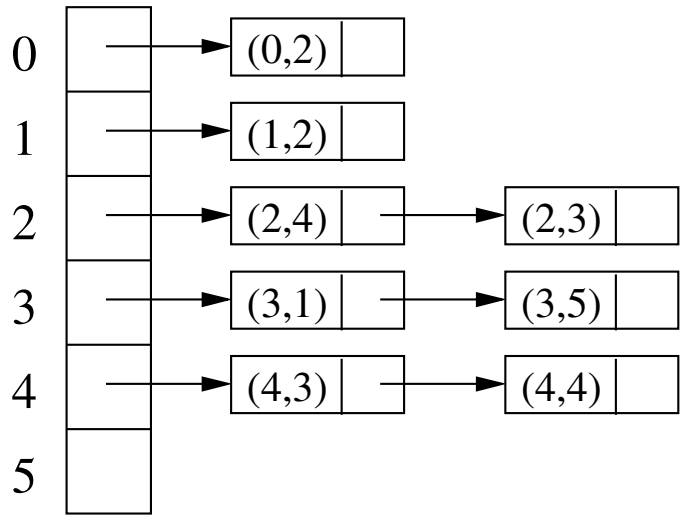
Implémentation par liste d'adjacence

Liste d'adjacence

Une liste d'adjacence est un tableau dont chaque case correspond à un sommet, et contient une liste chaînée de ses successeurs. Selon les données à stocker, on peut avoir :

- un `struct sommet` dans chaque case du tableau, pour stocker toutes informations relatives au sommet.
- des listes chaînées de `struct arc`, pour pouvoir stocker toutes les données relatives aux arcs.





Implémentation par liste d'adjacence

Avantages

- Lister les successeurs de s_i est en

Implémentation par liste d'adjacence

Avantages

- Lister les successeurs de s_i est en $O(n)$.
- Pas d'espace mémoire inutilisé (n cases "sommets" et m cases "arcs").

Défauts

- Tester si s_i est successeur de s_j est en $O(n)$. En pratique, ce n'est pas grave, car on cherche souvent à lister tous les successeurs (même complexité).
- Tester si s_i est prédécesseur de s_j est en $O(n)$, et lister tous les prédécesseurs de s_j est en $O(n+m)$. En effet, il faut parcourir tous les arcs pour connaître tous les prédécesseurs d'un sommet.

Remarque : Si on a réellement besoin de connaître les prédécesseurs, on peut stocker dans chaque sommet la liste chaînée de ses prédécesseurs. Dans ce cas, chaque arc est stockée deux fois, ce qui ne change pas la complexité spatiale.

Cas d'utilisation

- On privilégiera les matrices d'adjacence pour les petits graphes ou pour des graphes denses (pas creux), ou bien lorsque tester "s_i prédécesseur de s_j" est souvent réalisé.
- On privilégiera les listes d'adjacence pour les grands graphes, de manière à utiliser moins de place en mémoire, mais surtout pour parcourir rapidement les successeurs d'un sommet.

Implémentation par liste d'adjacence

```
1  typedef struct arc {
2      char* nom;
3      float poids;
4      int i;
5      int j;
6  } Arc;
7
8  typedef struct elementListeA{
9      Arc* a;
10     struct elementListeA* suiv;
11 } ElementListeA;
12
13 typedef ElementListeA* ListeA;
14
15 typedef struct sommet {
16     char* nom;
17     int num;        // position dans le tableau des sommets
18     float x,y;      // coordonnees du sommet
19     ListeA L_adj; // liste des arcs sortant de ce sommet
20     int nbA ;       // nb arcs sortant
21 } Sommet ;
22
23 typedef struct graphe {
24     int n;           // nb de sommets
25     int m;           // nb d'arcs
26     Sommet** tabS; // tableau de pointeurs sur sommets
27 } Graphe ;
```

Graphe non orienté (objet mathématique - théorie des graphes)

Quand on l'enlève l'orientation des arcs, on obtient ce que l'on appelle un graphe non orienté.

Définitions et terminologie

Un graphe non orienté est un couple $G = (S, A)$ où :

- S est un ensemble de n éléments appelés *sommets*.
- A est un ensemble de m paires non orientées de $S \times S$, appelés *arêtes*. Une arête peut être notée $\{s_i, s_j\}$ ou $\{s_j, s_i\}$ indifféremment.

Dans un graphe non orienté :

- un chemin est appelé une *chaîne*.
- un circuit est appelé un *cycle*.
- deux sommets liés par une arête sont dits *voisins* ou *adjacents* (pas de notions de successeur, prédécesseur, descendant et ascendant ici).
- une arrête $a = \{s_i, s_j\}$ est dite *incidente* aux sommets s_i et s_j.
- deux sommets sont dits *inaccessibles* l'un de l'autre s'il n'existe aucune chaîne entre les deux.

Connexité et arborescence

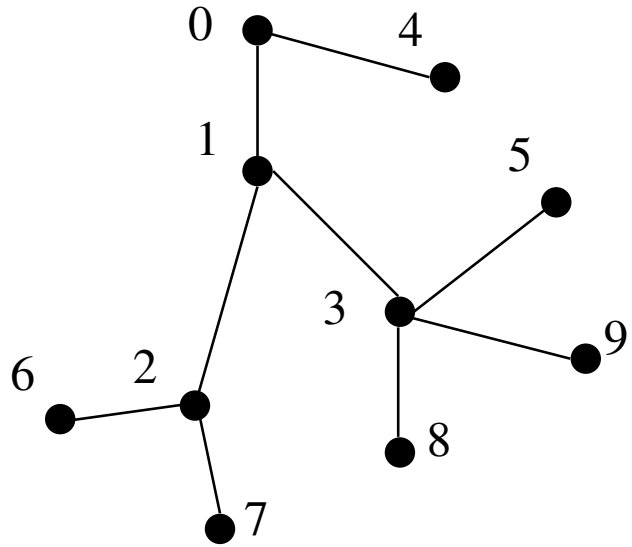
Définition : graphe connexe

Un graphe est dit *connexe* s'il n'existe aucune paire de sommets inaccessibles l'un de le l'autre. Autrement dit, un graphe est connexe s'il existe une chaîne entre toute paire de sommets du graphe.

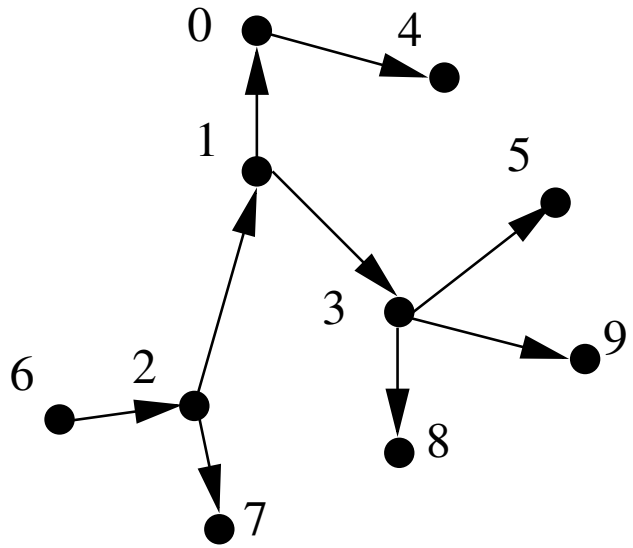
Arbre et arborescence

Un *arbre* est un graphe connexe sans cycle. Lorsque l'on choisit un sommet r, et que l'on oriente les arrêtes de r vers les autres, on obtient un graphe orienté dont r est l'unique racine (tous les sommets sont accessibles depuis r). Cet arbre orienté est appelé une *arborescence*, enracinée en r.

Arbre :



Arborescence enracinée en 6 :



Implémentation des graphes non orientés

Codage par matrice d'adjacence

Le principe du codage par une matrice (pour les graphes orientés) peut s'appliquer au cas non orienté. Dans ce cas, la matrice obtenue est symétrique (on peut alors se contenter d'une matrice triangulaire).

Codage par liste d'adjacence

Il n'existe pas directement d'implémentation d'un graphe non orienté par une structure avec des pointeurs : en effet, les pointeurs sont par nature "orientés". On peut quand même coder un graphe non orienté avec des pointeurs en utilisant une sorte d'équivalence entre graphe orienté et non orienté : une arête entre deux sommets s_i et s_j correspond à mettre à la fois un arc (s_i, s_j) et un arc (s_j, s_i). Dans cette implémentation, il y a deux fois plus de "cases" que dans le cas orienté, mais les complexités des opérations restent inchangées.

Remarque : s'il y a des données à stocker sur les arêtes, il est important que les deux cases d'une même arête pointent sur les mêmes données.

Sur la recherche d'accessibilité

Autres opérations possibles sur les graphes

Dans un graphe (orienté ou non-orienté), on se pose souvent les questions suivantes :

- Existe-t-il un chemin (ou une chaîne) entre deux sommets ?
- Quels sont les sommets accessibles à partir d'un sommet donné ?

Lien entre recherche d'accessibilité et parcours

Chercher si un sommet s est accessible depuis un sommet r revient à explorer tous les sommets accessibles à partir de r. En effet, comme l'on ignore quel chemin peut mener vers s depuis r, on doit explorer toutes les possibilités de chemins issus de r. Les pires-cas sont :

- de ne pas trouver s.
- de trouver s en dernier.

Cette recherche revient donc à parcourir le graphe à partir de r. Le sommet r est alors appelé la *racine* de la recherche.

Sous-parcours d'un graphe

Un *sous-parcours* partant d'un sommet est une exploration des "sommets de proche en proche", c'est-à-dire un déplacement de sommet en sommet en utilisant les arcs (ou les arêtes) du graphe. Plus formellement :

Sous-parcours à partir d'un sommet

On appelle *sous-parcours* d'un graphe à partir d'un sommet r une liste de sommets $L = (s_1, \ldots, s_k)$ telle que :

- $s_1 = r$.
- $s_1, \ldots, s_k$ est l'ensemble de tous les sommets accessibles à partir de r.

Remarque : les sommets sont numérotés de sorte que, pour tout $i \in \{2, \ldots, k\}$, il existe $j \in \{1, \ldots, i-1\}$ tel que (s_j, s_i) est un arc (ou une arête) du graphe (on parlera de bordure).

Sous-parcours et exploration

Le mot *sous-parcours* désigne une liste de sommets, mais désigne aussi le fait de parcourir le graphe. Parcourir un graphe depuis r, c'est explorer un à un les sommets de manière à former un sous-parcours.

Sous-parcours et recherche d'accessibilité

Terminologie

Au cours d'un (sous-)parcours, on appelle :

- **Sommet visité** : un sommet qui a été ajouté au sous-parcours.
- **Sommet non-visité** : un sommet qui n'a pas (encore) été ajouté au sous-parcours.
- **Sommet ouvert** : un sommet visité dont tous les descendants (ou sommets accessibles) n'ont pas encore été visités.
- **Sommet fermé** : un sommet visité dont tous les descendants (ou sommets accessibles) ont été visités.
- **Bordure d'une liste de sommet** : étant donné une liste L de sommets visités, la bordure $B(L)$ est l'ensemble des sommets non-visités qui sont successeurs (ou voisins) d'au moins un sommet dans la liste. Formellement, dans le graphe $G = (S,A)$, on a :

$$B(L) = \{s \in S \backslash L : \exists (s',s) \in A \text{ avec } s' \in L\} \text{ (cas orienté)}$$

$$B(L) = \{s \in S \backslash L : \exists \{s',s\} \in A \text{ avec } s' \in L\} \text{ (cas non-orienté)}$$

Sous-parcours et recherche d'accessibilité

Pseudo-code : sous-parcours à partir d'un sommet r

$L = (r)$
Tant que $B(L) \neq \emptyset$:
    Choisir un sommet $s \in B(L)$
    Ajouter s à la fin de la liste L
Fin Tant que

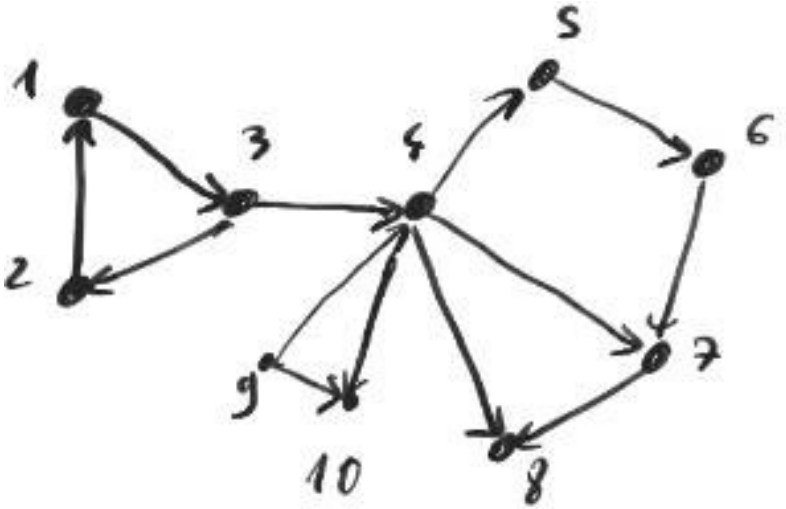
Remarque

À la fin du sous-parcours :

- tous les sommets visités sont fermés.
- un sommet non-visité (c'est-à-dire qui n'est pas dans le sous-parcours) est inaccessible depuis r.

Le deuxième point donne comment répondre à la question de la recherche d'accessibilité à l'aide d'un sous-parcours de graphe.

Exemple : graphe orienté



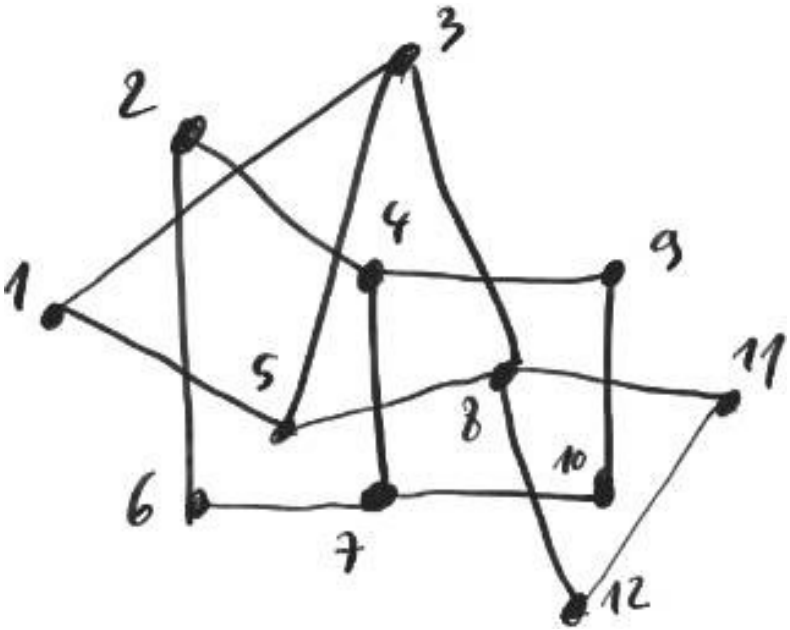
Parcours à partir du sommet 9 :

Itération	Liste des sommets visités
1	$L_1 = (9)$
2	$L_2 = (9,4)$
3	$L_3 = (9,4,8)$
4	$L_4 = (9,4,8,5)$
5	$L_5 = (9,4,8,5,6)$
6	$L_6 = (9,4,8,5,6,7)$
7	$L_7 = (9,4,8,5,6,7,10)$

À la fin de la procédure :

- on obtient un sous-parcours où les sommets 1, 2 et 3 n'ont pas été visités.
- tous les sommets accessibles à partir de 9 ont été trouvés.

Exemple : graphe non-orienté



Parcours à partir du sommet 1 :

Itération	Liste des sommets visités
1	$L_1 = (1)$
2	$L_2 = (1,3)$
3	$L_3 = (1,3,8)$
4	$L_4 = (1,3,8,11)$
5	$L_5 = (1,3,8,11,5)$
6	$L_6 = (1,3,8,11,5,12)$

À la fin de la procédure :

- on obtient un sous-parcours où six sommets n'ont pas été visités.
- tous les sommets accessibles à partir de 1 ont été trouvés.

Parcours et recherche d'accessibilité

A la fin d'un sous-parcours, on peut ne pas avoir visité tous les sommets (on s'est limité aux sommets accessibles à partir de la racine du parcours).

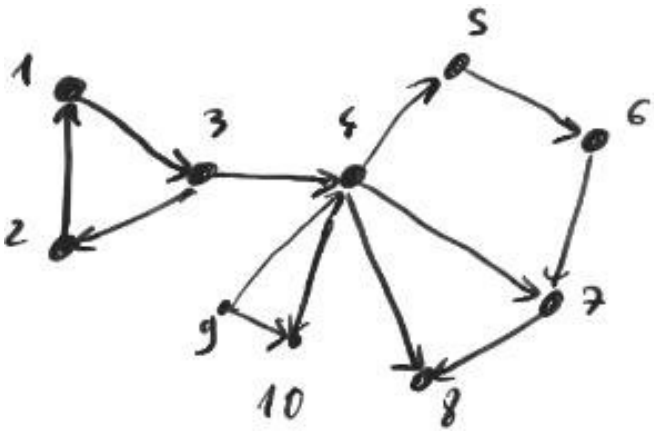
Parcours de graphe et points de régénération

Dans un *parcours*, on doit rencontrer tous les sommets du graphe. Quand ils ne sont pas tous accessibles depuis la racine, on doit "relancer" un sous-parcours à partir d'une nouvelle racine (non-visitée). Les racines de tous les sous-parcours sont appelés des *points de régénération*.

Pseudo-code : parcours de graphe à partir d'un sommet r

$L = (r)$
Tant qu'il existe un sommet non-visité :
    Si $B(L) \neq \emptyset$:
        Choisir un sommet $s \in B(L)$.
    Sinon :
        Choisir un sommet s non-visité.
    Fin Si
    Ajouter s à la fin de la liste L
Fin Tant que

Exemple : graphe orienté



Parcours à partir du sommet 9 :

Itération	Liste des sommets visités
1	$L_1 = (9)$
2	$L_2 = (9,4)$
3	$L_3 = (9,4,8)$
4	$L_4 = (9,4,8,5)$
5	$L_5 = (9,4,8,5,6)$
6	$L_6 = (9,4,8,5,6,7)$
7	$L_7 = (9,4,8,5,6,7,10)$
8	$L_8 = (9,4,8,5,6,7,10,2)$
9	$L_9 = (9,4,8,5,6,7,10,2,1)$
10	$L_{10} = (9,4,8,5,6,7,10,2,1,3)$

À la fin du parcours :

- tous les sommets ont été visités (pas de surprise).
- on a deux points de régénération : 9 et 2.