Arbres couvrants

OPTION INFORMATIQUE - TP n° 3.5 - Olivier Reynet

À la fin de ce chapitre, je sais :

- ☞ expliquer le fonctionnement de l'algorithme de Prim
- ☞ expliquer le fonctionnement de l'algorithme de Kruskal
- ☞ programmer dans un style purement fonctionnel en OCaml

Dans tout ce TP, on considère des graphes pondérés non orientés représentés par des listes d'adjacence. L'objectif est d'exploiter au maximum le module `List` d'OCaml et de programmer dans un style purement fonctionnel, c'est à dire sans références.

A Algorithme de Prim

L'algorithme de Prim est un algorithme glouton optimal qui s'applique aux graphes pondérés connexes. L'algorithme part d'un sommet et fait croitre un arbre en choisissant un sommet dont la distance est la plus faible et qui n'appartient pas à l'arbre, garantissant ainsi l'absence de cycle.

Algorithme 1 Algorithme de Prim, arbre recouvrant

```
1 : Fonction PRIM(G = (V, E, w))
2 :     T ← ∅                                  ▷ la sortie : l'ensemble des arêtes de l'arbre recouvrant
3 :     S ← s un sommet quelconque de V
4 :     tant que S ≠ V répéter
5 :         (u, v) ← min (w(u, v), u ∈ S, v ∈ E)          ▷ Choix glouton !
6 :         S ← S ∪ {v}
7 :         T ← T ∪ {(u, v)}
8 :     renvoyer T
```

A1. Montrer que l'algorithme de Prim termine.

A2. Montrer que l'algorithme de Prim est correct, c'est à dire qu'il calcule un arbre couvrant de poids minimal.

Pour faciliter la programmation, on utilise une représentation des graphes par liste d'adjacence. Ce choix n'est pas le fruit du hasard : on observe que les algorithmes de Prim ou de Kruskal construisent des arbres à partir des arêtes. Il est donc important de disposer des arêtes. On définit donc un type graphe comme suit :

```
1  type list_graph = ((int * int) list) list;;
2
3  let lcg = [[(2,1); (1,7)];
```

```
4               [(3,4); (4,2); (2,5); (0,7)];
5               [(5,7); (4,2); (1,5); (0,1)];
6               [(4,5); (1,4)];
7               [(5,3); (3,5); (2,2); (1,2)];
8               [(4,3); (2,7)]];;
9
10  let uclg = [[(1,7)];
11              [(4,1); (3,4); (0,7)];
12              [(5,7)];
13              [(4,5); (1,4)];
14              [(3,5); (1,2)];
15              [(2,7)]];;
```

A3. Dessiner les deux graphes `lcg` et `uclg`. Quelle différence y-a-t-il entre les deux?

A4. Écrire une fonction de signature `graph_order : int list -> int` dont le paramètre est un graphe sous la forme d'une liste d'adjacence qui renvoie l'ordre de ce graphe.

A5. Écrire une fonction dont la signature est :
`make_triplets_list : (int * int)list list -> (int * int * int)list`
qui transforme un graphe sous la forme d'une liste d'adjacence en une liste de type `(i,j,w)`, où `i` et `j` sont des sommets et `w` le poids associé à l'arête `(i,j)`. Par exemple, le graphe `lcg` devient `[(0, 2, 1); (0, 1, 7); (1, 3, 4); (1, 4, 2); (1, 2, 5); (2, 5, 7); (2, 4, 2);(3, 4, 5); (4, 5, 3)]`.
Bien observer qu'on a enlevé les redondances : on considère que les arêtes sont dans les deux sens.

A6. L'algorithme de Prim termine sa boucle lorsque les ensemble S et V sont identiques. On représente ces ensembles par des listes `S` et `V` en OCaml. Écrire une fonction de signature
`same_list : int list -> int list -> bool`
qui renvoie `true` si deux listes contiennent exactement les mêmes éléments et `false` sinon.

A7. Écrire une fonction de signature `compare_edges : int * int * int -> int * int * int -> int` qui compare les poids de deux triplets `(i,j,w1)` et `(k,l,w2)`. La fonction renvoie un entier positif si `w2` est plus grand que `w1`, 0 s'ils sont égaux et négatif sinon.

A8. Écrire une fonction de signature
`sort_triplets : (int * int)list list -> (int * int * int)list`
qui prend en entrée un graphe sous la forme d'une liste d'adjacence et qui renvoie la liste des triplets `(i,j,w)` triés dans l'ordre croissant de poids.

A9. Écrire une fonction de signature
`filter_triplet : int list -> int * int * int -> (int * int * int)`
Le premier paramètre d'entrée est une liste de sommet, le second un triplet `(i,j,w)`. La fonction renvoie des types optionnels :

1. `Some (i,j,w)` si $i \in S$ et $j \notin S$,
2. `Some (j,i,w)` si $j \in S$ et $i \notin S$,
3. `None` sinon.

L'objectif est d'utiliser ensuite cette fonction dans une expression de type `List.filter_map`. Cette fonction renvoie une liste qui n'est pas de la même taille que la liste d'entrée : le filtre écarte certaines valeurs. Ce filtre renvoie donc parfois un triplet, parfois rien.

A10. Écrire une fonction de signature
`select_min_triplet : int list -> (int * int * int)list -> int * int * int`

Le premier paramètre est la liste des sommets S, le second la liste des triplets triés dans l'ordre croissant des poids. Cette fonction renvoie le triplet de poids le plus faible dont un des sommets est dans S.

A11. Écrire une fonction purement fonctionnelle de signature
`pure_prim : (int * int)list list -> (int * int * int)list`
qui implémente l'algorithme de Prim. Le paramètre d'entrée est un graphe sous la forme d'une liste d'adjacence. Cette fonction renvoie la liste des triplets qui constitue l'arbre de poids minimal. Par exemple pour le graphe `lcg`, la sortie vaut :
`[(1, 3, 4); (4, 5, 3); (4, 1, 2); (2, 4, 2); (0, 2, 1)]`

A12. Tester la fonction `pure_prim` sur le graphe `uclg`. Que se passe-t-il ? Est-ce normal ?

A13. Comment pourrait-on utiliser une file de priorités pour améliorer les performances de cet algorithme ?

B Algorithme de Kruskal

L'algorithme de Kruskal est un algorithme glouton optimal qui s'applique aux graphes pondérés. Le graphe peut ne pas être connexe et dans ce cas on obtient un forêt d'arbres recouvrants. Pour construire la forêt et effectuer un choix d'arête, il ordonne les arêtes du graphe par ordre de pondération croissante.

Algorithme 2 Algorithme de Kruskal, forêt d'arbres recouvrants

```
1 : Fonction KRUSKAL(G = (V, E, w))
2 :     T ← ∅                          ▷ la sortie : l'ensemble des arêtes de l'arbre recouvrant
3 :     pour k de 1 à |E| répéter
4 :         e ← l'arête de pondération la plus faible de E          ▷ Choix glouton !
5 :         si (S, T ∪ {e}, w) est un graphe acyclique alors
6 :             T ← T ∪ {e}
7 :     renvoyer T
```

La principale difficulté de l'algorithme de Kruskal réside au niveau du test d'acyclicité : comment savoir si le nouvel arbre qu'on projette de construire est bien acyclique ? On pourrait utiliser un parcours en profondeur car dès qu'on trouve un sommet déjà découvert un cycle est détecté. Cependant ce n'est pas optimal au niveau de la complexité. C'est pourquoi il est préférable d'utiliser une structure de type Union-Find : c'est un structure simple et efficace qui permet de garder la trace des sommets qui sont connexes dans le graphe.

On se place dans le cadre du graphe `lcg` défini plus haut. C'est un graphe à six sommets. On va donc créer une structure Union-Find à partir de la liste `uf = [0,1,2,3,4,5]`. Chaque élément de la liste représente le numéro de l'ensemble connexe auquel appartient le sommet du graphe. Au début de l'algorithme, comme on n'a pas encore sélectionné d'arête, tous les sommets apparaissent déconnectés, chacun dans un ensemble différent. Donc le sommet 0 est dans l'ensemble 0, le sommet 1 dans l'ensemble 1…

Lorsqu'on connecte deux sommets par une arête, les sommets appartiennent au même ensemble connexe. On doit donc modifier `uf` et faire en sorte que les deux sommets portent le même numéro d'ensemble connexe. Par exemple, si on connecte le sommet 1 et le sommet 5, alors on aura `uf = [0,1,2,3,4,1]` ou bien `uf = [0,5,2,3,4,5]`.

Au fur et à mesure de l'algorithme, on va faire donc évoluer à la fois la liste des arêtes de l'arbre et la structure `uf`. À chaque tour de boucle, on choisit une arête et on vérifie que ses sommets n'appartiennent pas au même ensemble connexe, c'est à dire que les valeurs associées à ces sommets dans `uf` sont différentes. Par exemple, si on a `uf = [0,5,2,3,4,5]`, on ne pourra pas ajouter l'arête (1,5).

B1. On dispose d'une liste `uf` de type Union-Find. Écrire une fonction de signature
`union : int -> int -> int list -> int list`
qui procède à l'union des ensembles de numéro `e1` et `e2` donnés en paramètres. Le troisième paramètre est la liste `uf`. La fonction renvoie une nouvelle liste Union-Find dans laquelle `e1` a remplacé `e2`. Par exemple, si `uf = [0,0,2,3,4,4]` et si on souhaite unir les ensembles de numéro 0 et 4, la fonction renverra `uf = [0,0,2,3,0,0]`.

Le résultat de l'algorithme de Kruskal est une forêt d'arbre. On représente une forêt d'arbre comme les arbres précédemment, c'est à dire par une liste de triplets `(i,j,w)` où `i` et `j` sont des sommets et `w` le poids associé à l'arête `(i,j)`. La liste représente donc indifféremment un arbre ou une forêt. La seule différence est que, dans le cas d'une forêt, tous les sommets ne sont pas connexes.

Pour l'algorithme, on choisit un tuple de type `(uf, forest)` où `uf` est de type `int list` et `forest` de type `(int * int * int)list`. Au fur et à mesure de l'algorithme, la forêt s'épaissit, la structure `uf` s'homogénéise.

B2. Écrire une fonction de signature
`set_union_edge : int list * (int * int * int)list -> int * int * int -> int list * (int * int * int)list`
qui :

1. ajoute un triplet à `(uf, forest)` si l'arête ne crée pas un cycle,
2. ne modifie pas `(uf, forest)` sinon.

Les paramètres sont donc constitués d'un tupe `(uf, forest)` et d'un triplet. La fonction renvoie `(uf,forest)` dans tous les cas.

B3. Écrire une fonction purement fonctionnelle de signature
`pure_kruskal : (int * int)list list -> (int * int * int)list`
qui implémente l'algorithme de Kruskal. La fonction renvoie la forêt sous la forme d'une liste de triplets `forest`.

B4. Comment pourrait-on utiliser une file de priorités pour améliorer les performances de cet algorithme ?

B5. Démontrer la correction et la terminaison de l'algorithme de Kruskal.