



Majeure INFO - INF 435 - TP2

Graphes, arbres et diffusion

Cécile Bothorel
Printemps 2014

1 Objectif

Cette séance de TP s'intéresse à un problème de théorie des graphes qui a un intérêt applicatif important : la diffusion d'un message dans un réseau. Ici, le graphe est valué, la structure de données du TP1 ne peut donc pas être réutilisée telle quelle. Ne vous jetez pas sur la programmation, pensez à bien réfléchir avant.

Ce que vous allez développer est un élément de base du mini-projet, tout comme l'était le premier TP.

Au terme de ce TP, vous serez capables :

- de décrire le principe, le résultat, les avantages et inconvénients d'un algorithme de calcul d'arbre couvrant minimal,
- d'appliquer cet algorithme à une problématique de diffusion de messages dans un réseau,
- d'écrire un algorithme de calcul d'arbre couvrant minimal (euh maximal!)
- de choisir la structure de données permettant de manipuler un graphe valué,
- ... et vous aurez bâti la première étape de l'algorithme final de votre mini-projet !

2 Introduction au mini-projet

Le mini-projet consiste à implémenter une solution de ciblage marketing sur les réseaux sociaux. Cette solution a été proposée par des chercheurs en 2012. Leur publication s'intitule **Spread of Information in a Social Network Using Influential Nodes** [1].

Pour commencer la séance de TP, nous demanderons à des volontaires de venir au tableau pour décrire rapidement le mini-projet. Pour préparer, voici 3 questions :

Question 1 : Décrire l'objectif de l'algorithme de l'article [1].

Question 2 : Avez-vous vu pour quelle tâche DFS est utilisé ?

Question 3 : Décrire à quoi sert l'arbre couvrant maximum (maximum spanning tree).

3 Problème de diffusion dans un réseau

Vous êtes l'organisateur du WEI, et vous êtes très fier : ce week-end s'est très bien déroulé, c'est dimanche, vous savourez. Votre sens de l'organisation n'a jamais été pris en défaut. Or, on vient de vous annoncer que le car doit partir une heure avant l'horaire prévu, c'est-à-dire dans quelques minutes. Évidemment, pas de connexion correcte à internet dans votre centre de vacances. Il faut prévenir toute la promotion dans un temps record, vous n'y arriverez pas tout seul, il faut qu'on vous aide. Heureusement, vous connaissez bien toute la promotion et, comme vous êtes un peu maniaque, vous avez pris soin de noter tous les étudiants qui se connaissent. Avec des idées derrière la tête, vous avez également estimé leur relation de 100 (grosse amitié) à 1 (franche antipathie). Ces informations sont contenues dans le fichier `promotion.graph`.

Vous êtes un fainéant malin et vous vous dites que le bouche-à-oreilles est sans doute un moyen efficace de diffuser l'information : il suffirait de donner l'information à un seul étudiant, qui pourrait ensuite la donner à d'autres étudiants qui, eux-mêmes, pourraient également la diffuser, etc. Vous êtes également très organisé et vous connaissez les lascars avinés qui vous entourent. Il faut faire des choses simples : vous vous engagez à donner une feuille qui permet à chaque étudiant de connaître la liste des étudiants qu'il doit alerter.

3.1 Diffusion de l'information

Vous souhaiteriez que la communication ne concerne que des étudiants qui s'entendent bien. Vous souhaitez donc que la somme des "niveaux d'amitiés" de tous les liens utilisés pour diffuser l'information soit maximale.

Question 4 : Parmi tous les liens du graphe `promotion.graph`, quel est l'ensemble des liens qui vous garantit que (1) tous les étudiants recevront l'information et (2) le "coût amical total" est maximal?

3.2 Et si un étudiant était oublié ?

Si des élèves n'ont pas été prévenus, il faut pouvoir retrouver le coupable ou les coupables qui n'ont pas transmis le message comme prévu.

Question 6 : A partir du résultat de votre code précédent, créez une structure de données qui permette de retrouver n'importe quelle chaîne de diffusion, de n'importe quelle source à n'importe quelle cible.

3.3 Optimisation de l'arbre de diffusion

Vous avez mis en place votre système et vous vous avez remarqué un problème inattendu : le téléphone arabe ! Le message se déforme au fur et à mesure qu'il est retransmis par les étudiants.

Question 5 (optionnelle) : A qui devez-vous donner l'information pour réduire l'impact de ce téléphone arabe, c'est-à-dire que le nombre de retransmission du message soit minimal?

4 Instructions

Vous trouverez sur Moodle des fichiers à télécharger et sauvegarder dans votre espace personnel :

- un fichier `tp2.py` sur lequel vous allez travailler
- un fichier `graphviz.py` que vous n'avez pas besoin de modifier
- un fichier `promotion.graph` qui contient les données du graphe

Lorsque vous exécutez `python tp2.py`, un nouveau fichier est généré. Il s'agit d'une image `promotion.png` que vous pouvez visualiser (en utilisant par exemple `eog` sous Linux). Votre mission consiste à compléter le fichier `tp2.py`.

Références

- [1] Arpan Chaudhury, Partha Basuchowdhuri and Subhashis Majumder. Spread of Information in a Social Network Using Influential Nodes. PAKDD 2012, Part II, LNAI 7302, pp. 121-132, 2012