Recherche de chemins courts dans des graphes réalistes

Evripidis Bampis [1], Matthieu Latapy [2] et Fanny Pascual [1]

On rencontre dans de nombreux domaines des graphes de très grande taille, qui ont plusieurs centaines de milliers de sommets, voire plusieurs millions de sommets. Citons en particulier : le graphe d'Internet, le graphe du web, les réseaux d'interaction entre proteines, et de nombreux graphes sociaux. Outre leur grande taille, ces graphes ont plusieurs propriétés communes :

- la distance moyenne entre les sommets est courte ;
- la probabilité que deux voisins d'un sommet soient eux-même voisins est forte : ce sont des *graphes clusterisés* ;
- la distribution des degrés suit une loi de puissance, ce qui implique qu'il existe un nombre faible mais non négligeable de sommets de très haut degré : ce sont des *graphes sans-échelle*.

Nous appellerons les graphes possédant ces trois propriétés des *graphes réalistes*.

Trouver la distance, ou un plus court chemin, entre deux sommets est un problème important en algorithmique. En effet, non seulement cette information peut être utile en tant que telle (trouver un plus court chemin dans un réseau par exemple) mais elle est aussi utilisée dans de nombreux autres problèmes algorithmiques. Il existe des algorithmes, comme le parcours en largeur, qui permettent de calculer un plus court chemin entre tout couple de sommets d'un graphe non valué en un temps $O(n \cdot m)$, où n est le nombre de sommets, et m le nombre d'arêtes du graphe. La place nécessaire pour stocker ces distances est en $O(n^2)$. Pour de très gros graphes, il serait beaucoup trop coûteux en temps et en espace d'utiliser de tels algorithmes. Nous allons donc utiliser les propriétés des graphes réalistes afin de pouvoir trouver rapidement une structure de données ayant une taille raisonnable en mémoire et permettant de donner à la demande des chemins les plus courts possible entre les sommets du graphe.

L'idée de l'algorithme est la suivante : si on veut trouver un plus court chemin entre a et b, on se déplace de a vers son voisin de plus haut degré, c, puis de c vers son voisin de plus haut degré, etc. De même, on se déplace de b vers son voisin de plus haut degré, et on continue de se déplacer vers son voisin de plus haut degré jusqu'à rencontrer le chemin partit de a. On obtient alors un chemin entre a et b. On peut ensuite souvent raccourcir ce chemin, car les graphes réalistes sont des graphes clusterisés.

[1] LaMI – CNRS – Université d'Évry – 523 place des Terrasses ,91000 Evry, France. Mail :{bampis, fpascual}@lami.univ-evry.fr

[2] LIAFA – CNRS – Université Paris 7 – 2 place Jussieu, 75005 Paris, France. Mail : latapy@liafa.jussieu.fr

Il n'est en réalité pas toujours possible que les chemins partis de a et de b se rencontrent. Cela est du au fait qu'il n'est pas possible de construire un arbre couvrant dont chaque arête serait une arête joignant un sommet à son voisin de plus haut degré. Nous allons en fait obtenir des forêts couvrantes ayant cette propriété, puis nous relierons les arbres de la forêt en effectuant des plus courts chemins entre les racines des arbres.

Cet algorithme, ainsi que quelques variantes, donnent dans le cas de graphes réalistes des chemins qui ne sont pas beaucoup plus longs que les plus court chemins. Ils ont l'avantage d'utiliser une place raisonnable en mémoire (en $O(n)$) et d'avoir une complexité faible, ce qui permet de traiter rapidement des graphes de plusieurs centaines de milliers de sommets.

Plus généralement, il existe actuellement très peu d'algorithmes qui utilisent les propriétés statistiques des graphes réels, comme le fait qu'ils soient clusterisés et sans-échelle. Une perspective intéressante serait de tenir compte de ces caractéristiques pour trouver, pour divers problèmes, des algorithmes performants que l'on puisse utiliser sur des graphes réalistes de très grande taille.