Sur la difficulté de séparer un graphe par des plus courts chemins

Emilie Diot[†], Cyril Gavoille[*,‡] et Pascal Ochem[§]

Les schémas de routage et de calcul de distances les plus efficaces sont conçus à partir de décompositions hiérarchiques de la topologie en plus courts chemins. Ces constructions sont calculables efficacement pour de nombreuses topologies, comme les graphes planaires par exemple. Dans cet article nous montrons cependant que la décomposition d'une topologie arbitraire en k plus courts chemins est NP-difficile.

Mots-clefs : routage, séparateurs, plus courts chemins, graphes, NP-complétude

1 Introduction

Motivations. Le nombre d'entrées des tables de routage d'Internet (on parle ici des tables de routage FIB – *Forwarding Information Base* – de BGP) augmente d'un facteur estimé entre $1,2$ et $1,3$ par an. De 175k en 2007, on pense que le nombre d'entrées franchira le cap des 500k courant 2011 pour certains routeurs. La taille des tables a un impact sur la latence des routeurs et plus généralement sur le délai de transmission des messages[¶]. Malheureusement le taux d'accroissement des performances matériels des routeurs ne contre-balance pas complètement la croissance des tables, si bien qu'une augmentation de la latence est envisagée si à terme la structure des tables n'est pas réorganisée.

Sur le plan théorique, on peut concevoir des schémas de routage où, contrairement au protocole existant, la taille des tables est sous-linéaire voir poly-logarithmique au nombre d'AS (qui sont vus comme les sommets d'un graphe). Une technique bien connue consiste à décomposer hiéarchiquement le graphe en composantes grâce à de « petits » séparateurs. Le routage dans une même composante est résolu récursivement, alors que le routage d'un sommet vers ceux des autres composantes est réalisé en stockant dans la table une entrée par sommet du séparateur seulement. Les tables résultantes sont proportionnelles au produit de la taille des séparateurs et de la hauteur de la hiéarchie.

Lorsque les séparateurs sont de grande taille, comme les topologies planaires ou les réseaux routiers, la technique n'est plus efficace. Il est cependant possible d'outre-passer la taille des séparateurs si leur « structure » est particulière, quitte à relaxer l'optimalité de la longueur des routes. Par exemple, si le diamètre des séparateurs[‖] est faible, alors une approximation de la meilleure route est possible en utilisant un seul représentant du séparateur. En terme d'informations, on a réduit ainsi sa taille à un seul élément par séparateur.

Une autre situation favorable est lorsque les séparateurs sont des plus courts chemins. Dans ce cas, un routage s'appuyant sur quelques sommets bien choisis de ces chemins (on parle de *landmarks*) permet de s'approcher arbitrairement près des longueurs optimales de route tout en contrôlant la taille des informations de routage. Ainsi, Thorup [Tho04] a montré grâce à cette technique que les graphes planaires

[†]Université de Bordeaux, LaBRI, France, {diot,gavoille}@labri.fr

[‡]Le second auteur est également membre de l'Institut Universitaire de France. Les deux premiers auteurs sont supportés par les projets ANR Blanc "ALADDIN", et l'équipe-projet commune LaBRI-INRIA Bordeaux Sud-Ouest "CÉPAGE".

[§]Université Paris-Sud & CNRS, LRI, France, ochem@lri.fr. Supporté par le projet ANR Blanc "GRATEL".

[¶] Par exemple, la mise à jour de plus en plus fréquente des tables FBI entraine des périodes instabilités du routage.

[‖] Se reporter aux travaux autour des concepts de *tree-length* et d'*hyperbolicité*.

possèdent des schémas de routage dont les tables sont de taille poly-logarithmique et dont l'étirement¶ est majoré par $1+\varepsilon$, pour tout $\varepsilon > 0$ fixé.

En fait, cette technique permet de résoudre d'autres problèmes et sur des familles de graphes plus larges encore. Abraham et Gavoille [AG06] ont par exemple montré que les résultats des graphes planaires s'étendent à tous les graphes excluant un mineur fixé (comme les graphes de genre borné par exemple), et que ces plus courts chemins pouvaient servir à transformer ces mêmes graphes en « réseaux petit-monde » (cf. [Kle06] pour une introduction au domaine), et donc servir de topologies sous-jacentes à des tables de hachage distribuées (DHT) efficaces.

Une question naturelle est de savoir comment de tels séparateurs peuvent être déterminés en pratique, et quels sont les graphes qui possèdent des séparateurs composés d'un nombre limité de plus courts chemins, le nombre de plus courts chemins (plutôt que leur taille) étant un paramètre contrôlant de manière importante les performances des applications.

Avant de présenter nos résultats, nous définissons plus formellement les notions utilisées.

Définitions. Une *pondération* est une fonction à valeurs réelles positives (ou nulles) définie sur les sommets et les arêtes d'un graphe. Un graphe *pondéré* est un graphe G muni d'une pondération ω. On le note (G,ω). Le *poids* d'un sous-graphe H de G, noté $\omega(H)$, est la somme des poids associés aux sommets de H. Un *plus court chemin* entre deux sommets u,v de G est un chemin connectant u à v dont la somme des poids de ses arêtes est minimum.

Un *demi-séparateur* d'un graphe pondéré (G,ω) est un sous-ensemble de sommets S tel que chaque composantes connexes de $G \setminus S$ a un poids $\leqslant \frac{1}{2}\omega(G)$.

Un *k-chemins séparateur* d'un graphe pondéré G est un demi-séparateur de G qui induit un sous-graphe $P_0 \cup P_1 \cup \ldots$ où chaque P_i est composé d'une union de k_i plus courts chemins de $G \setminus \bigcup_{j<i} P_j$ avec $\sum_i k_i \leqslant k$. Un k-chemins séparateur S est dit *fort* si $S = P_0$, *i.e.* s'il est seulement composé de k plus courts chemins de G. Un graphe pondéré est *k-chemins séparable* si tout sous-graphe induit possède un k-chemins séparateur. Et, il est *fortement k-chemins séparable* si tout sous-graphe induit possède un k-chemins séparateur fort. Bien sûr, lorsque $k = 1$, les notions de k-chemins séparateur et de k-chemins séparateur fort sont identiques.

État de l'art. Tous les graphes que nous allons considérer dans la suite seront pondérés (sur les sommets et les arêtes). Les graphes de largeur arborescente k (qui sont *grosso modo* les graphes ayant un séparateur d'au plus k sommets) sont $\lceil (k+1)/2 \rceil$-chemins séparables ‖. À ce titre, la chemin-séparabilité est une généralisation de la largeur arborescence. De manière plus intéressante, tout graphe planaire est 3-chemins séparable, alors qu'il en existe de largeur arborescente arbitrairement grande. En fait, tout graphe sans mineur H est k_H-chemins séparables pour une certaine constante k_H, (cf. [AG06]). De plus, les graphes qui sont k-chemins séparables (quelle que soit la pondération) forment une famille close par mineurs (et donc caractérisable par un nombre fini de mineurs interdits). Pour $k = 1$, une liste (non exhaustive) de mineurs interdits a été donnée dans [DG10].

Contributions. Dans cet article nous montrons que les trois problèmes suivants sont NP-complets ou NP-difficiles. Parmi ces résultats négatifs, on montre qu'il n'y a peu d'espoir de calculer un plus court chemin séparateur ($k = 1$) dans un graphe par des techniques de programmation dynamique basées sur les décompositions arborescentes, puisque que le problème est déjà NP-complet même pour les graphes de largeur aborescence 3.

Théorème 1 *Soient (G,ω) un graphe à pondération entière et k un entier.*

- k-CHEMINS SÉPARATEUR*: Déterminer si (G,ω) possède un k-chemins séparateur est faiblement NP-complet pour $k = 1$, même si ω est unitaire pour les arêtes, G est planaire, de largeur arborescente* 3 *et de degré maximum* 3.
- k-CHEMINS SÉPARABLE*: Déterminer si (G,ω) est k-chemins séparable est NP-difficile, même si ω*

¶ Le ratio maximum entre la longueur de la route et celle d'un plus court chemin.

‖ Tout sous-graphe contient toujours $k+1$ sommets formant un demi-séparateur de la composante connexe la plus lourde, donc du sous-graphe. Un plus court chemin peut alors couvrir au moins deux sommets de cet ensemble, et donc $\lceil (k+1)/2 \rceil$ chemins suffisent à séparer le sous-graphe [DG10].

est unitaire pour les sommets et bornée par 2k pour les arêtes.

- FORTEMENT k-CHEMINS SÉPARABLE*: Déterminer si* (G,ω) *est fortement k-chemins séparable est NP-difficile, même si* ω *est unitaire pour les sommets et les arêtes, et G est de diamètre deux.*

2 Réductions

k-CHEMINS SÉPARATEUR. Il se réduit au problème PARTITION, dont le but est de déterminer si une suite de n entiers > 0 peut être partitionnée en deux sous-suites de même somme. Le problème est *faiblement* NP-complet car les entiers écrits en binaires peuvent avoir une valeur exponentielle en n [GJ79].

Soit $A = (a_i)_{i=1}^{n}$ une instance de PARTITION. On note $\sigma(X) = \sum_{x \in X} x$ la somme des éléments d'une suite X. On transforme l'instance A en une autre $B = (b_i)_{i=1}^{n+4} = (a_1, \ldots, a_n, \sigma(A), \sigma(A), \sigma(A), \sigma(A))$. La somme $\sigma(B) = 5\sigma(A)$. On peut vérifier que A possède une solution si et seulement si B en possède une¶. Dans la suite, on utilisera le fait que la suite B ne possède pas deux éléments dont la somme est $\geqslant \frac{1}{2}\sigma(B)$, ce qui n'est pas forcément vrai pour A.

À partir de B, on contruit un graphe planaire pondéré, noté (G_B, ω), avec $O(n)$ sommets (cf. figure 1). Le degré maximum est 3, et on peut vérifier que la largeur arborescente est également de 3. Les $n+4$ sommets « centraux » de G_B sont distingués (en noirs sur la figure) et correspondent aux éléments de B. La pondération des arêtes est unitaire, et celle des sommets est nulle sauf pour les sommets centraux où le i-ème sommet a pour poids b_i. Le poids de G_B est donc $\omega(G_B) = \sigma(B)$.

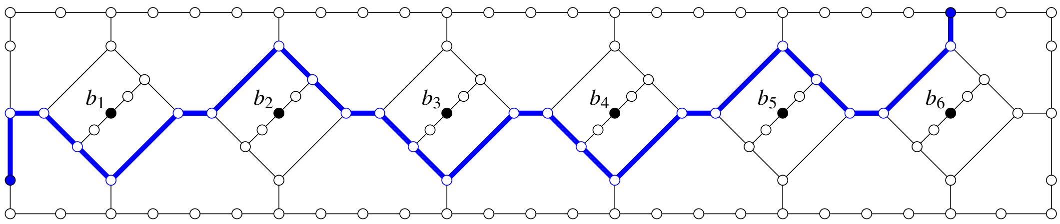

Fig. 1: Construction du graphe G_B à partir de l'instance $B = (b_1, \ldots, b_6)$ de PARTITION. Le plus court chemin séparateur (en bleu) correspond à la partition $(b_2, b_5, b_6), (b_1, b_3, b_4)$. D'autres plus courts chemins définissant la même partition sont possibles.

Dans la version longue de l'article, on montre que l'instance B (et donc l'instance A) admet une solution si et seulement si le graphe (G_B, ω) est possède un 1-chemin séparateur.

k-CHEMINS SÉPARABLE. Contrairement au problème précédant, les problèmes k-CHEMINS SÉPARABLE et FORTEMENT k-CHEMINS SÉPARABLE évoqués dans le théorème 1 ne sont pas trivialement dans NP. Il n'est pas évident non plus qu'ils soit dans PSPACE. On va donc seulement montrer qu'ils sont NP-difficiles.

Ils se réduisent au problème H-PARTITION dont le but est de déterminer si les sommets d'un graphe G peuvent être partitionnés de sorte que chaque partie induise un sous-graphe isomorphe à H. D'après [GJ79], ce problème est NP-complet pour tout graphe H fixé ayant au moins trois sommets. Dans toute la suite, nous allons fixer $H = P_3$, un chemin à trois sommets. On va utiliser le fait que si P_3 est un sous-graphe induit de G, alors c'est aussi un plus court chemin de G.

On ne présente que la réduction de FORTEMENT k-CHEMINS SÉPARABLE. La construction, et surtout la preuve, pour le problème k-CHEMINS SÉPARABLE est complexifiée par le fait que les chemins d'un k-chemins séparateur S ne sont pas tous forcément des plus courts chemins du graphe G, et dépendent les uns des autres.

Soit I_k un graphe à $3(k-1)$ sommets, instance‖ du problème P_3-PARTITION. À partir de I_k, on construit

¶ Si A possède une solution, B aussi. Ensuite, les parts de toute solution pour B doivent contenir chacune exactement deux termes $\sigma(A)$, sinon une des parts a une somme $\geqslant 3\sigma(A) > \sigma(B)/2$.

‖ Sans perte de généralité, le problème reste NP-complet même si l'on suppose que le nombre de sommets de l'instance est un multiple de trois, puisque sinon I_k est trivialement fausse.

comme suit le graphe $G(I_k)$ qui définira l'instance du problème FORTEMENT k-CHEMINS SÉPARABLE.

On crée trois copies disjointes de I_k, disons I_k^1, I_k^2, I_k^3, et pour chacune d'elles on ajoute trois sommets: x_i, y_i, z_i. Pour chaque $i = 1,2,3$, on connecte le sommet z_i à x_i, y_i, et à tous les autres sommets de I_k^i, et on note G_i ce graphe. Finalement, on forme un graphe biparti complet entre les sommets de G_1 et de G_2, puis entre ceux de G_2 et de G_3. Le graphe résultant est le graphe $G(I_k)$ (voir la figure 2 pour un exemple). Il possède $9k$ sommets.

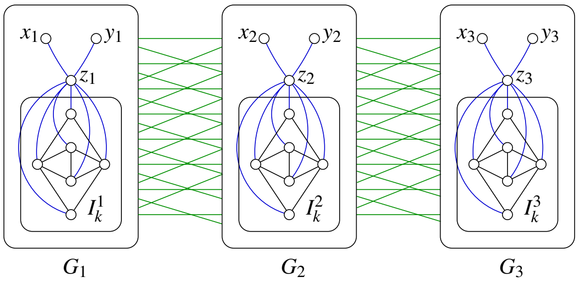

Fig. 2: Construction du graphe $G(I_k)$ à partir de l'instance I_k avec $k = 3$. Il y a un biparti complet entre $V(G_1)$ et $V(G_2)$, et $V(G_2)$ et $V(G_3)$ (toutes les arêtes ne sont pas représentées). Ce graphe est fortement 3-chemins séparable.

La pondération de $G(I_k)$ est unitaire, *i.e.* $\omega(e) = \omega(v) = 1$ pour toute arête e et sommet v de $G(I_k)$. Il est facile de vérifier que $G(I_k)$ est de diamètre deux (z_2 est connecté à tous les sommets du graphe). Une autre propriété de $G(I_k)$ est liée au fait que chaque sommet de G_2 est connecté à tous ceux de $G_1 \cup G_3$: pour tout sous-graphe induit H de $G(I_k)$ ne contenant pas tous les sommets de G_2, $G(I_k) \setminus H$ est composée d'une seule componsante connexe de $9k - |V(H)|$ sommets.

3 Conclusion

Nous avons montré que les problèmes k-CHEMINS SÉPARATEUR, k-CHEMINS SÉPARABLE et FORTEMENT k-CHEMINS SÉPARABLE sont NP-complets ou NP-difficiles. La complexité n'est pas connue pour k fixé, par exemple lorsque $k = 2$. C'est un problème intéressant, puisqu'en pratique on souhaite utiliser ces techniques de séparation par des plus courts chemins lorsque le nombre de chemins est faible.

D'autres problèmes intéressants sont de déterminer si une topologie G est (fortement) k-chemins séparable pour n'importe quelle pondération de ses sommets ou de ses arêtes. Ces problèmes sont de complexité paramétrique $O(f(k) \cdot n^3)$, pour une certaine fonction de k, d'après la « Théorie des Mineurs de Graphes » de Robertson et Seymour. Cependant, ces algorithmes ne sont pas explicites car ils dépendent de listes de mineurs minimaux interdits qui ne sont pas connues.

La conception d'algorithmes d'approximation du nombre de chemins k est une direction intéressante de recherche. À l'heure actuelle, aucun algorithme d'approximation ni aucune heuristique satisfaisante n'ont été proposés.

References

[AG06] Ittai Abraham and Cyril Gavoille. Object location using path separators. In 25^{th} *Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 188–197. ACM Press, July 2006.

[DG10] Émilie Diot and Cyril Gavoille. Path separability of graphs. In 4^{th} *International Frontiers of Algorithmics Workshop (FAW)*, volume 6213 of Lecture Notes in Computer Science, pages 262–273. Springer, 2010.

[GJ79] Michael R. Garey and David S. Johnson. *Computers and Intractability - A Guide to the Theory of NP-Completeness*. W.H. Freeman, 1979.

[Kle06] Jon Kleinberg. Complex networks and decentralized search algorithms. In *International Congress of Mathematicians (ICM)*. European Mathematical Society Publishing House (EMS Ph), August 2006.

[Tho04] Mikkel Thorup. Compact oracles for reachability and approximate distances in planar digraphs. *Journal of the ACM*, 51(6):993–1024, November 2004.