Résolution exacte du problème k-cluster par optimisation semidéfinie

Jérôme Malick[1] Frédéric Roupin[2]

[1] CNRS, Laboratoire J. Kunztmann, 51 rue des Mathématiques, 38400 Grenoble, France.
jerome.malick@inrialpes.fr

[2] CEDRIC-CNAM, 292 rue Saint-Martin, 75141 Paris Cedex France.
frederic.roupin@cnam.fr

Mots-Clés : *optimisation quadratique 0-1, optimisation semidéfinie, optimisation dans les graphes, dualité lagrangienne, branch-and-bound*

Nous présentons une méthode de résolution exacte du problème k-cluster (recherche d'un sous-graphe le plus dense de taille k fixée), basée sur un branch-and-bound utilisant une famille de bornes originales obtenues par optimisation semidéfinie. Pour ces bornes, un paramètre réel joue le rôle de curseur pour trouver un équilibre entre qualité et temps de calcul. Les premiers résultats numériques montrent que la méthode est compétitive avec les meilleurs algorithmes.

Problème combinatoire. Soient un graphe non orienté de n sommets $G = (V, E)$ et un entier $1 \leq k \leq n$; on cherche un sous-graphe de G de k sommets avec un nombre maximal d'arêtes. Ce problème se formule naturellement comme un programme quadratique en variables 0-1 :

$$(k\text{-cluster}) \begin{cases} \max \quad y^\top W y \\ \quad y_1 + \cdots + y_n = k \\ \quad y \in \{0,1\}^n \end{cases}$$

où W est la matrice d'adjacence de G. Le problème k-cluster est un problème d'optimisation combinatoire NP-difficile, et assez mal résolu en pratique. Plusieurs approches et relaxations par programmation mathématique ont été proposées pour résoudre (de manière exacte ou approchée) ce problème - en particulier par optimisation linéaire [2], semidéfinie [3, 8], et quadratique convexe [1].

Reformulation SDP équivalente. Nous proposons une nouvelle approche qui utilise une relaxation convexe originale de ce problème. Cette relaxation est de type SDP (optimisation semidéfinie) ; pour l'obtenir, nous reformulons au préalable le problème k-cluster avec :

- renforcement des contraintes - par n contraintes-produits supplémentaires,
- changement de variable - pour passer en variables $\{-1, 1\}$,
- homogénisation - pour obtenir un problème purement quadratique dans $\mathbb{R}^{n+1}$.

On applique ensuite le classique "lifting" [4] dans l'espace des matrices symétriques de taille $n+1$ que l'on munit du produit scalaire $\langle X, Y\rangle = \mathrm{trace}(XY)$. On peut ainsi écrire de manière équivalente le problème de k-cluster sous la forme

$$(\text{SDP}) \begin{cases} \max \quad \langle Q, X\rangle \\ \quad \langle Q_j, X\rangle = 4k - 2n, \quad j \in \{0, \ldots, n\} \\ \quad \langle E_i, X\rangle = 1, \quad i \in \{0, \ldots, n\} \\ \quad \mathrm{rang}(X) = 1, \quad X \succeq 0, \end{cases}$$

avec Q, Q_j, E_i des matrices symétriques particulières de taille $n+1$. L'étape finale consiste alors à remarquer que dans cette situation la contrainte de rang 1 est précisement équivalente à la contrainte $\|X\|^2 = (n+1)^2$, appelée "contrainte sphérique" [5].

Dualité lagrangienne et bornes. Nous montrons que la dualisation par un paramêtre réel α de la contrainte sphérique introduit une famille de bornes SDP aux propriétés intéressantes (voir [7]). La borne notée $\Theta(\alpha)$ est toujours moins précise que la borne SDP standard [8], mais elle en est très proche pour α petit. De plus, elle est calculé en pratique en résolvant un problème de projection

$$\left\{\begin{array}{l} \min \quad \|X - Q/\alpha\|^2 \\ \quad \langle Q_j, X\rangle = 4k - 2n, \ j \in \{0, \dots, n\} \\ \quad \langle E_i, X\rangle = 1, \ i \in \{0, \dots, n\} \\ \quad X \succeq 0. \end{array}\right.$$

par des méthodes efficaces (voir [6]). Le paramêtre α joue de plus le rôle de curseur pour réduire le temps de résolution en échange d'une dégradation de la qualité de la borne. Cette possibilité s'avère être particulièrement appréciable dans le contexte d'une résolution exacte.

Résolution exacte par branch-and-bound. Nous proposons alors une méthode de type branch-and-bound pour résoudre le problème k-cluster utilisant :

– pour l'évaluation : les bornes originales $\Theta(\alpha)$, avec un arrêt prématuré éventuel du solver interne en cas d'élagage dans l'abre de recherche ;
– pour le branchement : une stratégie standard, couplée avec des heuristiques pour le calcul de solutions réalisables à des noeuds autres que les feuilles de l'arbre de recherche.

Dans [7], nous comparons notre approche aux meilleures méthodes existantes, dont [1]. Les premiers résultats numériques effectués sur des instances classiques montrent que nous obtenons des résultats comparables et parfois meilleurs. En effet, la borne de [1] est de qualité SDP à la racine de l'arbre de recherche, mais elle se dégrade ensuite lorsqu'on s'enfonce dans l'arbre de recherche, tandis que notre borne est de qualité SDP dans tout l'arbre de recherche sans en payer le prix habituel en temps de calcul. Ainsi, nous énumérons en moyenne 10 fois moins de noeuds, et nous pouvons ainsi résoudre des instances de k-cluster de taille 100 et même plus en des temps de calculs raisonnables.

Références

[1] A. Billionnet, S. Elloumi and M.-C. Plateau. Improving the performance of standard solvers for quadratic 0-1 programs by a tight convex reformulation : the QCR method. *Discrete Applied Mathematics* 157(6):1185-1197, 2009.

[2] A. Billionnet. Different formulations for solving the heaviest k-subgraph problem. *Information Systems and Operational Res.*, 43(3):171-186, 2005.

[3] G. Jäger and A. Srivastav. Improved approximation algorithms for maximum graph partitioning problems. *Journal of Combinatorial Optimization* 10(2):133-167, 2005.

[4] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM* 6:1115-1145, 1995.

[5] J. Malick. The spherical constraint in Boolean quadratic programs. *Journal of Global Optimization* 39(4):609-622, 2007.

[6] J. Malick. A dual approach to semidefinite least-squares problems. *SIAM Journal on Matrix Analysis and Applications* 26(1):272-284, 2004.

[7] J. Malick and F. Roupin. Solving k-cluster to optimality using adjustable semidefinite programming bounds. *Submitted*, 2009.

[8] F. Roupin. From linear to semidefinite programming : an algorithm to obtain semidefinite relaxations for bivalent quadratic problems. *Journal of Combinatorial Optimization*, 8(4), 2004.