Comparaison de formulations de K-partitionnement

Zacharie Ales[1,2], Arnaud Knippel[3]

[1] CNAM, Paris, France

[2] UMA, ENSTA Paris, Institut Polytechnique de Paris, Palaiseau, France.

zacharie.ales@ensta-paris.fr

[3] LMI, INSA Rouen Normandie, Rouen, France.

arnaud.knippel@insa-rouen.fr

Mots-clés : *classfication non supervisée, partitionnement de graphes, optimisation combinatoire, approche polyèdrale, branch-and-cut, formulation étendue.*

Étant donné un graphe complet valué $G = (V, E, w)$, le problème de K-partitionnement consiste à partitionner V en exactement K parties non vides tout en minimisant la somme du poids des arêtes à l'intérieur des parties.

La résolution exacte de problèmes de partitionnement dans lesquels le nombre de parties est contraint est souvent effectuée par le biais de formulations dites *sommet-partie* qui associent une variable binaire z_i^k égale à 1 si et seulement si le sommet $i \in V$ est assigné à la partie numéro $k \in \{1, ..., K_{max}\}$. [3, 5, 6, 7, 8, 9]. Un des désavantages de ces formulations est qu'elles peuvent induire un grand nombre de symétries. En effet, il existe un nombre exponentiel de solutions équivalentes à des permutations prêt des numéros des parties. Pour le problème de K-partitionnement, deux formulations sommet-partie sont envisageables suivant que K_{max} est fixé à K (Formulation (F_{nc1})) ou à $|V|$ (Formulation (F_{nc2})). Dans ce second cas, il est possible d'ajouter des inégalités valides retirant la symétrie du problème.

Nous prouvons que ces formulations ont également comme second désavantage de fournir des valeurs optimales de leur relaxation linéaire de mauvaise qualité.

Bound 0.1. *La valeur optimale de la relaxation linéaire de (F_{nc2}) est inférieure ou égale à* $\min\limits_{j \in \{2,...,K\}} \min\limits_{i<j} \frac{w_{ij}}{2^{K-j+1}}$.

Bound 0.2. *La valeur optimale de la relaxation linéaire de (F_{nc1}) est :*

- *dans l'intervalle* $[\min\limits_{i,j \in V} w_{ij} \frac{n-1}{2}, \frac{1}{2} \sum\limits_{i=2}^{n} w_{1i}]$ *si K est égal à 2 ;*
- *inférieure ou égale à* $\min\limits_{j \in \{2,...,K-1\}} \min\limits_{i<j} \frac{w_{ij}}{2^{K-j}}$ *si K est supérieur à 2.*

La Table 1 qui contient le gap à la racine moyen de ces formulations sur 100 graphes à 15 sommets illustre la faible qualité de leur relaxation linéaire.

Pour pallier cela, nous définissons deux formulations *arête-représentant* (F_{er}) [1] et (F_{ext}) [2] qui introduisent la notion de *représentant*. On appel représentant d'une partie de V, le sommet

Formulation	**K**								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
(F_{nc1})	82	97	98	99	99	99	99	100	100
(F_{nc2})	99	99	99	99	100	100	100	100	100
(F_{er})	87	79	71	60	48	34	20	11	5
(F_{ext})	76	70	61	51	39	25	13	7	3

TAB. 1 – Ecart relatif moyen à la racine de graphes aléatoires comportant 15 sommets.

ayant le plus petit indice dans cette partie. Ainsi ces deux formulations considèrent une variable binaire r_i égale à 1 si et seulement le sommet i est un représentant et x_{ij} égale à 1 si et seulement si les sommets i et j sont dans la même partie. L'utilisation de variables de représentants avait préalablement été considérée dans le cas de problème de coloration de graphes [4]. La formulation (F_{ext}) étend (F_{er}) en considérant également des variables $\tilde{x}_{ij}$ égales à 1 et seulement si le sommet j est dans la partie représentée par le sommet i.

Nous prouvons que (F_{ext}) fournit une valeur optimale de relaxation linéaire qui est au moins aussi bonne que celle de (F_{er}). Nous constatons numériquement (voir Table 1) que ces formulations fournissent des relaxations linéaires de bien meilleure qualité que les deux formulations sommet-partie. Nous déterminons les inégalités de ces formulations qui définissent des facettes de l'enveloppe convexe de leurs points entiers. Nous introduisons une famille d'inégalités valides pour (F_{ext}) et déterminons les cas où elles définissent des facettes. Nous présentons des résultats numériques qui montrent l'efficacité de nos formulations arête-représentant comparé aux deux formulations sommet-partie en termes de temps de résolution optimale. Enfin, nous montrons que les performances de (F_{ext}) peuvent être améliorées en considérant un algorithme de *branch-and-cut*.

Références

[1] Z. Ales, A. Knippel, and A. Pauchet. Polyhedral combinatorics of the k-partitioning problem with representative variables. *Discrete Applied Mathematics*, 211 :1–14, 2016.

[2] Zacharie Alès and Arnaud Knippel. The k-partitioning problem : Formulations and branch-and-cut. *Networks*, 76(3) :323–349, 2020.

[3] M. Boulle. Compact mathematical formulation for graph partitioning. *Optimization and Engineering*, 5(3) :315–333, 2004.

[4] M. Campêlo, V.A. Campos, and R.C. Corrêa. On the asymmetric representatives formulation for the vertex coloring problem. *Discrete Applied Mathematics*, 156(7) :1097–1111, 2008.

[5] S. Chopra and M.R. Rao. The partition problem. *Mathematical Programming*, 59(1) :87–115, 1993.

[6] N. Fan, Q.P. Zheng, and P.M. Pardalos. Robust optimization of graph partitioning involving interval uncertainty. *Theoretical Computer Science*, 447 :53–61, 2012.

[7] C.E. Ferreira, A. Martin, C.C. De Souza, R. Weismantel, and L.A. Wolsey. Formulations and valid inequalities for the node capacitated graph partitioning problem. *Mathematical Programming*, 74(3) :247–266, 1996.

[8] C.E. Ferreira, A. Martin, C.C. De Souza, R. Weismantel, and L.A. Wolsey. The node capacitated graph partitioning problem : A computational study. *Mathematical Programming*, 81(2) :229–256, 1998.

[9] V. Kaibel, M. Peinhardt, and M.E. Pfetsch. Orbitopal fixing. *Discrete Optimization*, 8(4) :595–610, 2011.