A model for large scale graph partitioning and efficient upper/lower bound computation via cutting-planes

Viet Hung Nguyen[1], Dang Phuong Nguyen[2], Michel Minoux[3]

[1] INP Clermont Auvergne, Univ Clermont Auvergne, Mines Saint-Etienne, CNRS, UMR 6158 LIMOS, 1 Rue de la Chebarde, Aubiere Cedex, France
viet-hung.nguyen@uca.fr

[2] CEA, LIST, Point Courrier 172, 91191 Gif-sur-Yvette, France

[3] Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, 4 place Jussieu, Paris, France

Mots-clés : *graph partitioning, integer programming.*

1 Introduction

The graph partitioning problem under knapsack constraint (GPKC) investigated in the present paper can be defined as follows. Given an undirected connected graph $G = (V, E)$ where $V = \{1, \ldots, n\}$, $|E| = m$. The nodes in V are weighted by a weight vector $w \in \mathbb{Z}_+^{|V|}$ and the edges in E are valued by a length vector $t \in \mathbb{Z}_+^{|E|}$ ($t_e, e \in E$ will be referred to as the "length" of the edge e). We want to find a partition of V into disjoint sets (or clusters) such that :

— every cluster $C \subset V$ satisfies a node weight constraint of the form $\sum_{v \in C} w_v \leq W$ (a knapsack-type constraint) where W is a given upper limit on the total node weight of the cluster.

— the sum of the lengths of the edges having end-nodes in different clusters is minimized.

GPKC is the classical version of the graph partitioning problem defined in Garey and Johnson [3]. Note that in this definition of GPKC, the number of clusters is not imposed nor bounded and it is part of the output of the optimization process. Graph partitioning problems with unbounded number of clusters was considered by several authors such as Sørensen [4], Labbé et al. [**?**]. Due to the fact that all the lengths $t_e,\ e \in E$ are positive, the objective of minimization of the sum of the lengths of the edges having end-nodes in different clusters already tends to limit the number of clusters in the optimal partition. Variants of GPKC with an upper bound constraint on the number of clusters have also been considered by several authors, see e.g. [**?**], [6], [8], [1].

GPKC is NP-hard [**?**] and lot of efforts have been spent in the development of fast and good heuristics for the problem (e.g. see survey in [**?**]). These heuristics often can handle rather large graphs with thousand vertices and deliver good solutions though not certified. In contrast to the development of heuristics, only limited effort has been devoted to the development of exact algorithms [6], [4], [8]. it is clear that generally problems involving only relatively small graphs can be solved exactly. Nevertheless, exact solutions are of interest for applications if only they can be used for the validation of heuristics.

Exact solutions to the graph partitioning problem are usually obtained using branch-and-cut algorithms based on Integer Programming (IP) formulations with 0/1 variables. In general, there are two types of IP formulations based on the meaning of 0/1 variables : the Node-Cluster model where 0/1 variables x_{ij} determine whether a node i belongs to a cluster j and the Node-Node model where 0/1 variables x_{ij} determine whether two nodes i and j belongs to the same cluster or not. Linear programming relaxations are solved at branch-and-bound nodes The quality of the linear relaxation of the IP formulations is used to obtain lower bounds for the problem. There are

Among the existing models, the Node-Node model is known to provide a good quality bounds [5], [4], [7]. This model involves decision variables x_{ij} which are equal to 1 iff node i and node j are not in the same cluster, this result in a 0/1 linear program featuring $O(n^2)$ variables and $O(n^3)$ triangle inequalities [2], [5]. It has been shown in [7] that one can reduce the number of triangle inequalities in the Node-Node model to $O(nm)$ without weakening the linear relaxation. This reduction is especially significant for sparse graphs in which $m = O(n)$. However, in such a model the number of variables is still in order $O(n^2)$, it quickly becomes quite large as n increases and even the linear relaxation turns out to be difficult to solve. It is therefore interesting to try and reduce the number of variables of the problem while preserving the quality of linear relaxation.
We consider in this paper an alternative model that makes use of only m binary decision variables $(x_e)_{e\in E}$ which are equal to 1 iff the end-nodes of e are in different clusters, together with the so-called *cycle inequalities* defined on them [2], We prove that the node partitions and thus the knapsack constraints can be defined using all pair shortest path of G with respect to the weights $(x_e)_{e\in E}$. The result is a m 0/1 variable programming model for GPKC. This model is referred to as the *cycle model* and it is shown to be equivalent to the Node-Node model. From our computational experiments, it will be shown that this model yields considerable improvement in computational efficiency in case of sparse graphs where $m \ll \frac{n(n-1)}{2}$. Note that since there is a priori no known polynomial upper bound (in terms of n and m) on the number of cycles and of paths of the graph G, the cycle model has a priori an exponential number of inequalities thus a cutting-plane algorithm is needed to make the computations possible. We introduce such an algorithm which can determine both violated cycle inequalities and violated knapsack constraints via the computations of all pair shortest path of G with respect to the weights $(x_e)_{e\in E}$.
Computational experiments show that, when applying cutting-plane algorithm, the cycle model outperforms the Node-Node model with triangle inequalities in terms of computation time. With the new model, the linear relaxation of GPKC can be solved optimally for large size graphs typically featuring 1000 to 2000 nodes.

Références

[1] Alès, Z. & Knippel, A. The K-partitioning problem : Formulations and branch-and-cut. *Networks*. **76**, 323-349 (2020)

[2] Barahona, F. & Mahjoub, A. On the cut polytope. *Mathematical Programming*. **36**, 157-173 (1986) bibitemBonamiNKM12Bonami, P., Nguyen, V., Klein, M. & Minoux, M. On the Solution of a Graph Partitioning Problem under Capacity Constraints.. *ISCO*. **7422** pp. 285-296 (2012)

[3] Ferreira, C., Martin, A., Souza, C., Weismantel, R. & Wolsey, L. The node capacitated graph partitioning problem : A computational study. *Mathematical Programming*. **81**, 229-256 (1998)

[4] Garey, M. and Johnson, D. Computers and Intractability : A Guide to the Theory of NP-Completeness. (W. H. Freeman and Co.,1979)

[5] Kaibel, V., Peinhardt, M. & Pfetsch, M. Orbitopal fixing . *Discrete Optimization* . **8**, 595 - 610 (2011)

[6] Nguyen, D., Minoux, M., Nguyen, V., Nguyen, T. & Sirdey, R. Improved compact formulations for a wide class of graph partitioning problems in sparse graphs. *Discrete Optimization*. **25** pp. 175-188 (2017)

[7] Sorensen, M. Facet-defining inequalities for the simple graph partitioning polytope . Discrete Optimization . **4**, 221 - 231 (2007)