Problème de Correlation Clustering avec Médiateurs

Zacharie Ales[1], Céline Engelbeen[2], Rosa Figueiredo[3]

[1] ENSTA-ParisTech / UMA, 91762 Palaiseau and Laboratoire CEDRIC, Paris, France,
zacharie.ales@ensta-paristech.fr

[2] QUARESMI Laboratory, ICHEC Brussels Management School, Brussels, Belgium,
Celine.Engelbeen@ichec.be

[3] Laboratoire Informatique d'Avignon LIA EA 4128, Avignon France,
rosa.figueiredo@univ-avignon.fr

Mots-clés : *Correlation clustering, algorithme d'énumération, MIP formulation, graphe signé.*

Le problème de correlation clustering (CCP) [1] consiste à partitionner les sommets d'un graphe afin que la somme des poids positifs entre les parties et la somme des poids négatifs à l'intérieur des parties soient minimales. Diverses variantes de ce problème ont été étudiées dans la littérature [2, 3]. Dans ce travail, nous introduisons un nouveau problème de ce type dans lequel un ensemble de sommets appelés médiateurs doit également être identifié. Les médiateurs doivent avoir de bonnes relations les uns avec les autres, mais aussi de suffisamment bonnes relations avec les autres sommets du graphe. Par exemple, dans un contexte politique, un ensemble de médiateurs pourrait représenter un ensemble de députés chargés d'évaluer une loi : ceux-ci doivent donc s'entendre suffisamment entre eux pour pouvoir converger vers une décision, mais aussi avec le reste des députés pour les représenter correctement.

Considérons un digraphe $D = (V, A, s)$, dans lequel V est un ensemble de sommets, A est un ensemble d'arcs et $s : A \rightarrow \{+, -\}$ est une fonction indiquant si la relation entre deux sommets est positive ou négative. L'intensité (en valeur absolue) de la relation entre deux sommets i et j est représentée par w_{ij}. Soit $\delta[S] = \{(i,j) \in A : i \in S, j \in V \setminus S\} \cup \{(i,j) \in A : i \in V \setminus S, j \in S\}$ et $A[S] = \{(i,j) \in A : i,j \in S\}$. Nous notons également A^+ (resp. A^-) l'ensemble des arcs ayant un signe positif (resp. négatif). Soient α et β deux entiers positifs. Un ensemble de médiateurs est un ensemble S qui satisfait les deux propriétés suivantes :

$$\sum_{(i,j)\in A^- \cap A[S]} w_{ij} \leq \beta \sum_{(i,j)\in A^+ \cap A[S]} w_{ij} \qquad (1)$$

$$\sum_{(i,j)\in A^- \cap \delta[S]} w_{ij} \leq \alpha \sum_{(i,j)\in A^+ \cap \delta[S]} w_{ij} \qquad (2)$$

La contrainte (1) indique que les médiateurs doivent s'entendre relativement bien entre eux. Le paramètre α permet de modifier l'intensité de cette entente : si $\alpha = 0$, il ne peut y avoir que des relations positives dans l'ensemble des médiateurs, et au fur et à mesure qu'on augmente la valeur de α on se permet d'inclure des relations négatives dans l'ensemble des médiateurs. La contrainte (2) signifie que l'ensemble des médiateurs doit avoir de relativement bonnes relations avec les sommets non-médiateurs. A nouveau, le paramètre β nous permet d'apporter une certaine flexibilité à cette définition.

Le problème de correlation clustering avec médiation positive (CCPM) consiste à trouver une partition des sommets $\{P_1, P_2, \cdots\}$ telle que P_1 est un ensemble de médiateurs et telle que la fonction suivante est minimisée :

$$\sum_{(i,j)\in A^- : i,j \in P_k,\ k>1} w_{ij} + \sum_{(i,j)\in A^+ : i \in P_k,\ j \in P_l,\ k,l>1} w_{ij}$$

Les propriétés définies par les contraintes (1) et (2) sont non héréditaires. Le CCP est NP-difficile [1] ainsi que le CCPM. Nous présentons une formulation de CCPM sous forme d'un programme linéaire en nombres entiers. Nous développons également un algorithme d'énumération pour résoudre CCPM qui repose sur la particularité que les ensembles de sommets respectant la contrainte (1) forment un système accessible. Enfin, nous présenterons les résultats de cet algorithme d'énumération sur des instances aléatoires ainsi que sur des instances réelles de différents groupes politiques.

Références

[1] N. Bansal, A. Blum, and S. Chawla. 2002. Correlation Clustering. In 43rd Annual IEEE Symposium on Foundations of Computer Science. 238–247.

[2] P. Doreian and A. Mrvar. 2009. Partitioning signed social networks. Social Networks 31, 1 (2009), 1–11.

[3] R. Figueiredo and G. Moura. 2013. Mixed integer programming formulations for clustering problems related to structural balance. Social Networks 35, 4 (2013), 639–651.