Compressed Sensing : trouver le k-sous-graphe le plus dense

Dorian Lagadec, Martin Mugnier

27 mars 2019

Polynomial integrality gaps for strong SDP relaxations of Densest k-subgraph, Bhaskara, Charikar, Guruswami [2011]

Sommaire

Théorie des graphes : rappels et présentation du problème du k-sous-graphe le plus dense

- Graphe : ensemble (E,V) de noeuds et d'arêtes pouvant représenter toutes sortes de réseaux ;
- De nombreux problèmes de graphes, tous liés à une même complexité : la croissance exponentielle de la difficulté ;
- Notre problème : trouver le sous-graphe à k noeuds le plus dense dans un graphe non pondéré non orienté. Un problème NP-difficile.

Relaxation convexe d'un problème NP-hard

Problème IP

$$
\begin{aligned}
\max \quad & \sum_{(i,j)\in E(G)} x_{ij} \\
\text{subject to} \quad & \sum_{i\in V} x_i \leq k, \\
& \forall i,j \in V,\ x_{ij} \in \{0,1\}
\end{aligned}
$$

- Problème NP-difficile $\rightarrow$ relaxation convexe et solution approximante
- LP : $\forall i,j \in V,\ 0 \leq x_{ij} \leq 1$
- Relaxation convexe SDP et hiérarchies : introduire des contraintes linéaires et SDP locales de plus en plus importantes
  $\rightarrow$ hiérarchies de Sherali-Adams, Lasserre.

Performance des algorithmes d'approximation ?

- Integrality Gap :

$$IG = \frac{\text{Solution approchée}}{\text{Solution du problème en nombre entier}}$$

- Trade-off dans les niveaux de hierarchies : précision / coût de calcul

But ultime : **trouver une borne inférieure la plus fine possible sur l'integrality gap pour un certain type de relaxations SDP du problème originel**.

$\rightarrow$ Avantage : non conditionnel à $\mathscr{P} = \mathscr{N}\mathscr{P}$ pour éliminer potentiellment une large classe d'algorithmes.

Résultats et questions

- Peu de résultats d'innaproximabilité parlants : seulement en $\mathcal{O}(1)$.
- Peut-on donner une borne inférieure qui dépende du niveau dans les hierarchies SA et LAS ?
- Qui dépende de n ?

Rappel : Pour un graphe à n sommets, n niveaux de SA ou LA conduisent à la solution exacte du pb.

Borne inférieure pour Sherali-Adams

Theorem (Bhaskara et. al, 2011)

Soit $L \leq \frac{\log n}{10 \log\log n}$. *L'écart d'approximation d'un* SA_L *est au moins* $\Omega\left(\frac{n^{1/4}}{L \log^2 n}\right)$.

Idées de la preuve :

- Propriété de graphes d'Erdos-Renyi bien choisis
- Arbre de Steiner minimal

Borne inférieure pour Lasserre

Theorem (Bhaskara et. al, 2011)

Pour un $\kappa > 0$ suffisamment petit, on peut trouver un écart d'approximation en $N^{2/53-O(\kappa)}$ pour le N^{κ}-niveau du Lasserre SDP appliqué au Min-degree.

Idée de la preuve :

- Réduction d'un problème de Max-CSP.

Une application : la contrainte sphérique de Malick et Roupin

Problème initial

$$\begin{aligned} \max \quad & y^T W y \\ \text{subject to} \quad & \langle y, (1,1,\ldots,1) \rangle = k \\ & y \in \{0,1\}^n \end{aligned}$$

Problème final : un lagrangien $\theta(\alpha)$

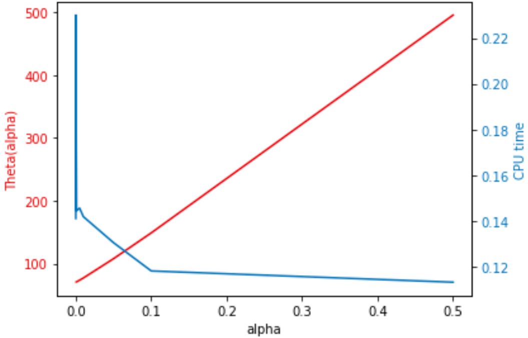
$$\begin{aligned} \max \langle Q, X \rangle - \alpha(||X||^2 - (n+1)^2) \\ \text{subject to} \langle Q_j, X \rangle = 4k - 2n, j \in \{0,\ldots,n\}, \\ \langle E_j, X \rangle = 1, i \in \{0,\ldots,n\}, \\ X \succcurlyeq 0 \end{aligned}$$

Transformations nécessaires :

- Ajout de contraintes redondantes ;
- Changement de variable affine de y vers x $(x = 2y - 1)$;
- Homogénéisation de la contrainte quadratique ;
- Réécriture du problème sous forme matricielle $(X = xx^T)$;
- Relaxation de la contrainte de rang $(rg(X) = 1$ par définition$)$ et définition du lagrangien.

Simulations et résultats

Simulation sur un graphe Erdös-Renyi $\mathscr{G}(30, 0.4)$ et une recherche du sous-graphe le plus dense à 10 noeuds :

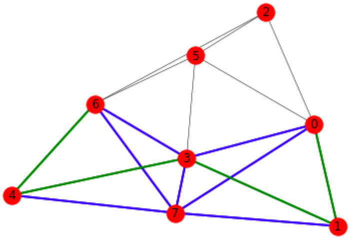


Le temps CPU décroit exponentiellement avec α alors que la borne supérieure augmente linéairement. Autrement dit, avec un α faible mais bien choisi, on a une approximation correcte et extrêmement économe en terme de CPU.

Extension : revenir au problème du sous-graphe le plus dense à k noeuds

Algorithme proposé

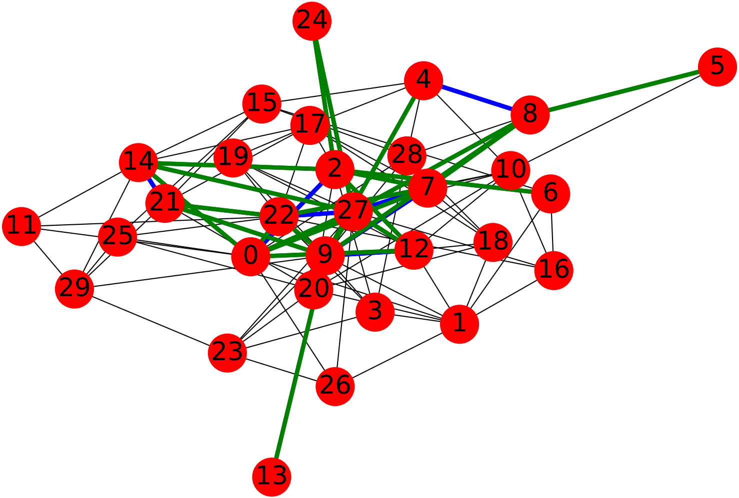
- A partir de la matrice X, inférer les noeuds les plus connectés jusqu'à en obtenir k (seuil décroissant) ;
- Obtenir une nouvelle matrice d'adjacence "approximée" ;
- Comparer cette dernière à la première, retirer les arêtes artificielles et ajouter les arêtes manquantes.



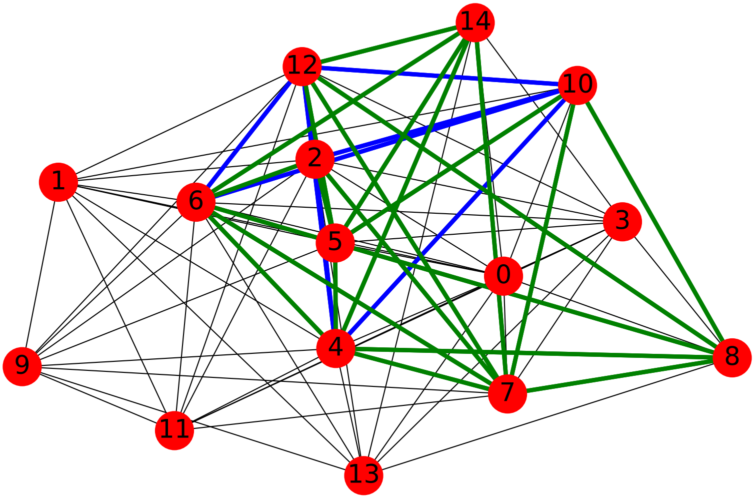
Résultats :

- Une intuition fondée ;
- Des résultats théoriques manquants ;
- Une validation correcte sur des graphes de petite taille.

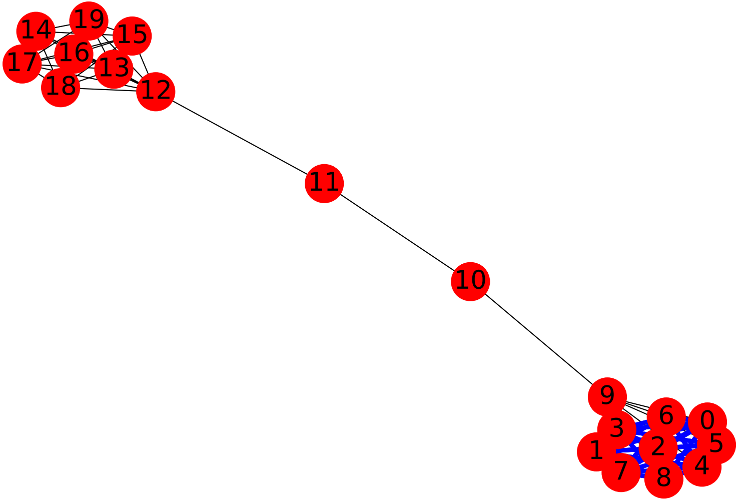
15 densest subgraph de $\mathscr{G}(30, 0.2)$



10 densest subgraph de $\mathscr{G}(15, 0.7)$



10-densest subgraph sur un graphe déterministe avec plusieurs "cliques"



Conclusions, limites et extensions

- Quelle borne pour des régimes plus rapides de relaxation (ex : polynomiaux de n) ?
- Peu de proposition de Branch and bound algorithms,
- Instabilité de nos algorithmes quand $n \to \infty$