A. Benoit, B. Depardon
C. Mouilleron, C. Rezvoy

TD n°11
$\mathcal{NP}$-Complétude et Approximation

1 Problème du k-centre

On rappelle quelques définitions :
- Dans un graphe $G = (V, E)$, un *ensemble indépendant* est un sous-ensemble de sommets V' non reliés par des arêtes (si $u \in V'$ et $v \in V'$, alors $(u, v) \notin E$).
- Dans un graphe $G = (V, E)$, un *ensemble dominant* est un sous-ensemble de sommets V' tel que tout sommet de $V \setminus V'$ est adjacent à un sommet de V'. On note $dom(G)$ le cardinal minimal d'un ensemble dominant.

Question 1.1 Montrer que trouver un ensemble indépendant de cardinal maximal est NP-difficile.

Remarque : *Un problème d'optimisation est NP-difficile si son problème de décision associé est NP-complet.*

Solution : *Le problème de décision associé à Ensemble Indépendant de cardinal maximal, est : est-ce qu'il existe un ensemble indépendant de cardinal k.*

Ensemble indépendant est dans NP.

On va faire une réduction à partir de 3-SAT : $F = C_1 \wedge \cdots \wedge C_k$, tel que $C_i = a_1^i \vee a_2^i \vee a_3^i$ ($a_j^i = x_j$ ou $\overline{x}_j$). Pour chaque variable de chaque clause, on créé un nœud, avec pour label le nom de la variable, on peut ainsi avoir plusieurs nœuds avec le même label. Pour chaque clause C_i, on ajoute une arête entre les trois sommets la constituant, on appelle chacun de ces petits graphes des gadgets. Enfin, on ajoute une arête entre chaque pair $(x_i, \overline{x_i})$. On obtient donc un graphe G dans lequel on va essayer de trouver un ensemble indépendant de taille k.

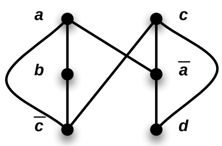


FIG. 1: Réduction pour l'instance de 3-SAT : $(a \vee b \vee \overline{c}) \wedge (c \vee \overline{a} \vee d)$

Si l'instanciation de 3-SAT est bonne, alors au moins une variable de chaque clause est vraie. Soit S un sous-ensemble de nœuds de G. S contient une des variables satisfaites par clause (par gadget), il y a donc k nœuds dans S (puisqu'on a k clauses, et qu'il y a nécessairement au moins une variable à vrai par clause). Comme nous ne prenons qu'un nœud par gadget, l'indépendance est préservée au sein des gadgets. De plus, entre les clauses, l'indépendance est également préservée étant donné que les seules arêtes existantes sont celles entre x_i et $\overline{x}_i$ qui ne peuvent être tous les deux à vrai en même temps. Donc S contient un ensemble indépendant de G de taille k.

Si Ensemble Indépendant est vrai, montrons que 3-SAT l'est aussi. On a un ensemble indépendant S de G. Puisque S est indépendant, au plus un nœud de chaque gadget apparaît dans S. Et vu qu'il y a exactement k gadgets, S doit contenir un nœud par gadget. De plus, vu que S est indépendant, on ne peut avoir x_i et $\overline{x}_i$ dans S, car ils sont reliés par une arête. Une assignation mettant 3-SAT à vrai peut être obtenue en positionnant x_i à vrai s'il appartient à S, et à faux sinon.

Donc Ensemble Indépendant est NP-complet.

Autre solution : *plus rapide, partir de CLIQUE et prendre le graphe complémentaire. L'équivalence est alors immédiate.* □

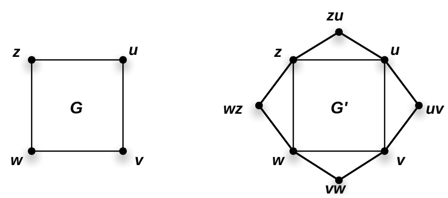
Question 1.2 Montrer que trouver un ensemble dominant de cardinal minimal $dom(G)$ est NP-difficile.

Solution : *Le problème de décision associé est : est-ce qu'il existe un ensemble dominant de taille inférieure ou égale à un entier k. On a un graphe $G = (V, E)$, et un entier positif $k \leq |V|$.*

Pour prouver la NP-complétude de ce problème de décision, on va faire une réduction à partir de Vertex Cover.

Ensemble dominant est dans NP.

Soit $< G, k >$ une instance de Vertex Cover. On construit une instance $< G', k >$ de Ensemble Dominant minimum de la façon suivante : pour chaque arête (u, v) de G on rajoute un sommet uv dans G', ainsi que les arêtes (u, uv) et (uv, v).



Si C est un ensemble solution de Vertex Cover de G de taille k, alors vu que chaque arête de G est incidente à au moins un sommet de C on voit que les sommets de G présents dans G' sont dominés par ces mêmes sommets de C. Les nouveaux sommets de G' représentant les arêtes de G sont aussi dominés, car chaque arête de G est couverte par un sommet de C. Donc, tous ces sommets de G' représentant les arêtes de G sont voisins d'un sommet de C. Donc, n'importe quel Vertex Cover C d'un graphe G est également un ensemble dominant pour G'.

Si D est un ensemble dominant de taille k de G', alors on supprime tout d'abord les sommets correspondants aux arêtes de G. Si $uv \in D$ représentant une arête $(u, v) \in E$, alors on le remplace par u. La taille de D n'est pas modifiée, et on a toujours un ensemble dominant avec des sommets de G. On a ainsi toutes les arêtes de G qui sont couvertes par un sommet G.

Donc Ensemble Dominant est NP-complet. □

Soit $G = (V, E)$ un graphe non-orienté, complet, dont les arêtes sont pondérées par une fonction de poids w qui vérifie l'inégalité triangulaire : $w(u, v) \leq w(u, w) + w(w, v)$ pour tout triplet de sommets (u, v, w). Soit aussi un entier $k \geq 1$.

Pour tout $S \subset V$ et tout $v \in V \setminus S$, on définit $connect(v, S)$ comme le poids minimal d'une arête reliant v à un sommet de S : $connect(v, S) = \min_{s \in S}\{w(v, s)\}$. Le problème est de trouver un k-centre, c'est à dire un sous-ensemble S de cardinal k et tel que $center(S) = \max_{v \in V \setminus S}\{connect(v, S)\}$ soit minimal.

Question 1.3 À quoi peut bien servir de déterminer un k-centre (donner un exemple d'application) ?

Solution : *Prenons un ensemble de villes, avec les distances entre ces villes connues. On veut choisir un sous-ensemble de k villes pour implanter des entrepôts, de façon à minimiser la distance maximum entre une ville et l'entrepôt le plus proche.*

Autres exemples : arrêts de bus, serveur de pages Web, et bien sûr machines à café... □

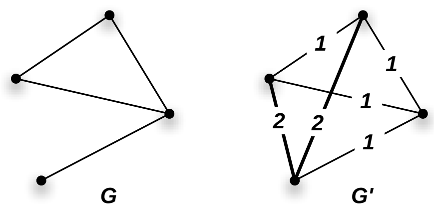
Question 1.4 Montrer que trouver un k-centre est NP-difficile.

Solution : *Nous allons réaliser la réduction à partir d'un ensemble dominant telle que si Ensemble Dominant répond "oui" alors k-centre aura pour solution 1, et sinon (si Ensemble Dominant répond "non") alors on aura 2 pour solution.*

Soit une instance du problème Ensemble Dominant $(G = (V, E), k)$*, on construit une nouvelle instance pour k-centre de la façon suivante. À partir de G on construit un graphe complet* $G' = (V, E')$ *avec pour fonction de poids pour* $e \in E'$ $w(e) = 1$ *si* $e \in E$*, et* $w(e) = 2$ *sinon. Puisque toutes les distances dans* G' *valent 1 ou 2, on a nécessairement que la valeur du k-centre est soit 1 soit 2.*

Si on a un certificat D pour Ensemble Dominant, de taille k, alors il nous suffit de placer un "centre" sur chacun des sommets de D pour avoir que chaque $v \in V \setminus D$ *est à une distance 1 de D dans* G'*, et la solution pour le k-centre est alors 1.*

Inversement, si k-centre retourne 1 comme valeur, alors on a un ensemble $S \subseteq V$ *tel que tout sommet* $v \in V \setminus S$ *soit à une distance 1 de S. Par construction, tous ces sommets doivent avoir une arête* $e \in E$ *vers un sommet de S, et donc S est un ensemble dominant de taille k.*



□

On va chercher une 2-approximation, i.e. un S de cardinal k tel que $center(S) \leq 2 \cdot OPT$, où $OPT = \min_{S \subset V, |S|=k}\{center(S)\}$.

On ordonne les arêtes de E par poids croissant : $w(e_1) \leq w(e_2) \leq \ldots \leq w(e_m)$, où $m = |E|$. On pose $G_i = (V, E_i)$ où $E_i = \{e_1, e_2, \ldots, e_i\}$ est l'ensemble des i premières arêtes.

Question 1.5 Montrer que résoudre le problème du k-centre revient à trouver le plus petit indice i tel que G_i a un ensemble dominant de cardinal au plus k.

Solution : *On veut minimiser* $center(S)$.

On veut montrer qu'avoir S un ensemble dominant de G_i est équivalent à $center(S) \leq w(e_i)$.

$$\begin{aligned} \text{Soit } S \text{ un ensemble dominant de } G_i &\iff \forall v \in V \backslash S, \exists s \in S | (v,s) \in E_i \\ &\iff \forall v \in V \backslash S, \exists s \in S | w(v,s) \leq w(e_i) \\ &\iff \forall v \in V \backslash S, connect(v,S) \leq w(e_i) \\ &\iff center(S) = \max_{v \in V \backslash S}\{connect(v,S)\} \leq w(e_i) \end{aligned}$$

Donc minimiser $center(S)$ avec $|S| \leq k$ revient à trouver le plus petit $w(e_i)$ (c'est à dire l'indice i minimum, car les $w(e_i)$ sont triés par ordre croissant), tel que G_i a un ensemble dominant S de cardinal k. □

Une dernière définition : le carré d'un graphe $G = (V, E)$, noté $G^{(2)} = (V, E^{(2)})$, contient les chemins de longueur au plus deux : $(u, v) \in E^{(2)}$ si $(u, v) \in E$ ou s'il existe $w \in V$ tel que $(u, w) \in E$ et $(w, v) \in E$.

Question 1.6 Étant donné un graphe H, soit I un ensemble indépendant du graphe carré $H^{(2)}$. Montrer que $|I| \leq dom(H)$.

Solution : *On veut montrer que si I est un ensemble indépendant de $H^{(2)}$ (c'est-à-dire $\forall x, y \in I, (x,y) \notin E_H^{(2)}$), alors $|I| \leq dom(H)$.*

Soit D un ensemble dominant de H, alors
- *d'une part, tout sommet de I est "dominé par" (c'est-à-dire relié ou égal à) au moins un sommet de D dans H (car D ensemble dominant de H).*
- *d'autre part, tout sommet de D "domine" (est relié ou égal à) au plus un sommet de I dans H (sinon, soient $x, y \in I$ dominés par un même sommet dans H, alors x et y sont à une distance 1 ou 2 dans H, autrement dit (x, y) est une arête de $H^{(2)}$, d'où une contradiction avec I ensemble indépendant de $H^{(2)}$).*

Donc $|I| \leq dom(H)$. □

Question 1.7 L'algorithme d'approximation du k-centre est le suivant : **début**

Construire $G_1^{(2)}, G_2^{(2)}, \ldots, G_m^{(2)}$ Trouver de manière gloutonne un ensemble indépendant inextensible (auquel on ne peut pas rajouter des sommets) M_i dans chaque graphe $G_i^{(2)}$ Trouver le plus petit indice i tel que $|M_i| \leq k$, soit j cet indice **retourner** M_j

fin

Remarque : *Complexité de l'algorithme (non demandée) : ligne 7 en $O(m \times n^3)$ (calcul de $G_i^{(2)}$ à partir de G_i en $O(n^3)$ avec produit de matrices d'adjacence), ligne 7 en $O(m \times n^2)$ (calcul avec un glouton qui pioche un sommet mis dans M_i, puis retire ce sommet et ses voisins dans $G_i^{(2)}$ et recommence sur le graphe restant, possible en $O(n^2)$, ou $O(m + n)$ avec des listes d'adjacences), ligne 7 en $O(m)$. On a donc un algorithme polynomial.*

1.7.1 Montrer que $w(e_j) \leq OPT$.

Solution : *Montrons que $w(e_j) \leq OPT = \min_{S \subseteq V, |S|=k}\{center(S)\}$.*

On sait d'après la question 5 que $OPT = \min_i\{w(e_i) | dom(G_i) \leq k\}$, or $\forall i < j, |M_i| > k$ (ligne 3 de l'algo)) $\Rightarrow dom(G_i) \geq |M_i| > k$ (d'après la question 6).

On peut donc écrire : $OPT = \min_{i \geq j}\{w(e_i) | dom(G_i) \leq k\} \geq w(e_j)$, car les $w(e_i)$ sont triés par ordre croissant. □

1.7.2 Montrer que l'algorithme est bien une 2-approximation.

Remarque : *À ce stade, on n'a pas encore utilisé le fait que les M_i, et donc M_j, sont maximum pour l'inclusion, ni l'inégalité triangulaire sur les $w(e_i)$.*

Solution : *Montrons que $center(M_j) \leq 2 \cdot OPT$.*

On a $center(M_j) = \max_{v \in V \setminus M_j}\{connect(v, M_j)\}$, soit $v \in V \setminus M_j, \exists s \in M_j$ tel que (v, s) arête de $G_j^{(2)}$ (en effet, sinon M_j ne serait pas un indépendant maximal pour l'inclusion de $G_j^{(2)}$, car on pourrait ajouter v à M_j).

Par conséquent :
– *soit (v, s) arête de G_j, donc $w(v, s) \leq w(e_j)$*
– *soit $\exists$ sommet z tel que (v, z) arête de G_j et (z, s) arête de G_j*
Donc $w(v, s) \leq w(v, z) + w(z, s) \leq 2 \cdot w(e_j)$.

Donc, de toute façon, $connect(v, M_j) \leq 2 \cdot w(e_j)$, et au final $center(M_j) = \max\{connect(v, M_j)\} \leq 2 \cdot w(e_j) \leq 2 \cdot OPT$ □

Question 1.8 Montrer que la borne 2 est stricte : donner un exemple de graphe où l'algorithme réalise effectivement une 2-approximation.

Solution : *Un graphe atteignant la borne 2 est par exemple une roue de $n+1$ sommets : chaque arête incidente au point central de la roue a un poids de 1, et toutes les autres arêtes ont un poids de 2.*

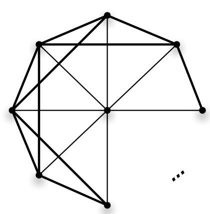


FIG. 2: Roue : les arêtes de poids 1 sont représentés en trait fin, et les arêtes de poids 2 en trait fort.

Pour $k = 1$, la solution optimale est le centre de la roue, et $OPT = 1$. L'algorithme va lui calculer l'indice $j = n$. G_n^2 est une clique, et si la solution renvoyée est un des sommets autres que le bord, alors le coût de la solution est 2. □

Question 1.9 Montrer que si $P \neq NP$, il n'existe pas de $(2 - \varepsilon)$-approximation au problème du k-centre, pour tout $\varepsilon > 0$.

Solution : *On va montrer que si un tel algorithme existait, alors il résoudrait le problème en temps polynomial. On va faire une réduction à partir du problème de décision de l'existence d'un ensemble dominant.*

Soit $G = (V, E)$, k une instance du problème d'ensemble dominant. On construit alors un graphe complet $G' = (V, E')$, avec pour poids des arêtes :

$$cost(u, v) = \begin{cases} 1, \ if \ (u, v) \in E \\ 2 \ sinon \end{cases}$$

G' satisfait bien l'inégalité triangulaire. La réduction satisfait bien les conditions :
- *si $dom(G) \leq k$, alors G' a un k-centre de coût 1*
- *si $dom(G) > k$, alors le coût optimum d'un k-centre de G' est 2.*

Dans ce cas, lorsqu'on utilise l'algorithme de $(2 - \epsilon)$-approximation sur le graphe G', il doit renvoyer une solution de coût 1, puisqu'il ne peut pas utiliser une arête de coût 2. On a donc avec cet algorithme un moyen de déterminer en temps polynomial s'il existe ou non un ensemble dominant. Donc si $P \neq NP$ ce n'est pas possible. $\square$