DS Complexité et Calculabilité : proposition de correction

REMI.MORVAN@U-BORDEAUX.FR

Année 2021–2022

Modifications du sujet :

– Dans le problème, au lieu de « un graphe orienté G », lire « un graphe non-orienté G ».
– En question 4, au lieu de « s'il existe $(v,w) \in E$ tel que $d(w,D) \leq l-1$, alors $d(v,D) \leq l$ » lire « il existe $(v,w) \in E$ tel que $d(w,D) \leq l-1$ si et seulement si $d(v,D) \leq l$ ».
– En question 7, au lieu de « $\varphi_1 \wedge \varphi_2 \wedge \varphi_3$ » lire « $\varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4$, où φ_4 est une formule exprimant la propriété "tout sommet de V est à distance au plus d de D" ».

Remarque : Cette correction contient plus de détails que ce qu'il était nécessaire d'écrire pour obtenir la note maximale (21/20).

Question 1

Points attendus : certificat ; vérificateur ; complexité.

Un certificat d'une instance positive (G,k,d) du problème Proxy est un ensemble de sommets D de taille au plus k tel que tout sommet de G est à distance au plus d de D.

On peut prendre comme vérificateur du problème Proxy l'algorithme dont l'entrée est une instance (G,k,d) du problème Proxy ainsi qu'un ensemble de sommets $D \subseteq V$, et vérifiant que D est bien un ensemble de taille au plus k et que tout sommet de G est à distance au plus d de D, par exemple en effectuant un parcours en profondeur depuis chaque sommet de D.

Le certificat est de taille au plus n, et le temps de calcul du vérificateur est un $\mathcal{O}(n+nm) = \mathcal{O}(nm)$—en effet, la vérification de $|D| \leq k$ est linéaire en n, et chaque parcours en largeur se fait en temps $\mathcal{O}(m)$.

Un certificat d'une instance positive n'est pas n'importe quel sous-ensemble D de sommets : c'est un sous-ensemble $D \subseteq V$ de taille au plus k tel que tout sommet de V est à distance au plus d de D.

Question 2

Le certificat est de taille polynomiale et le vérificateur a un temps de calcul polynomial, donc Proxy est dans NP.

Question 3

Pour que les $x_{u,i}$ décrivent un ensemble de taille au plus k, il faut et il suffit que chaque indice $i \in [\![1,k]\!]$ ne soit associé qu'à au plus un sommet. On pose donc :

$$\varphi_1 = \bigwedge_{1 \leq i \leq k} \bigwedge_{u \neq v \in V} \neg x_{u,i} \vee \neg x_{v,i}.$$

Dans la deuxième conjonction, il est important de seulement quantifier sur les paires de sommets $(u, v) \in V^2$ telles que $u \neq v$ pour exprimer le fait que « deux sommets **distincts** ne peuvent pas avoir le même indice ». Si on quantifiait sur toutes les paires $(u, v) \in V^2$, la formule φ_1 demanderait que toutes les variables $x_{u,i}$ (avec $u \in V$ et $i \in [\![1, k]\!]$) soient fausses.

Question 4

On pose :

$$\varphi_2 = \bigwedge_{v \in V} \bigwedge_{1 \leq l \leq d} \Big(\bigvee_{\substack{w \in V \text{ tq} \\ (v,w) \in E}} y_{w,l-1} \Big) \leftrightarrow y_{v,l}.$$

Question 5

La propriété « tout sommet $v \in V$ est à distance au plus 0 de D si et seulement si $v \in D$ » s'exprime par la formule

$$\bigwedge_{v \in V} y_{v,0} \leftrightarrow \Big(\bigvee_{1 \leq i \leq k} x_{v,i} \Big)$$

qui n'est pas en CNF, mais qui est équivalente à

$$\begin{aligned} \bigwedge_{v \in V} \Big(y_{v,0} \to \bigvee_{1 \leq i \leq k} x_{v,i} \Big) \wedge \Big(\Big[\bigvee_{1 \leq i \leq k} x_{v,i} \Big] \to y_{v,0} \Big) &\equiv \bigwedge_{v \in V} \Big(\neg y_{v,0} \vee \bigvee_{1 \leq i \leq k} x_{v,i} \Big) \wedge \Big(\Big[\bigwedge_{1 \leq i \leq k} \neg x_{v,i} \Big] \vee y_{v,0} \Big) \\ &\equiv \bigwedge_{v \in V} \Big(\neg y_{v,0} \vee \bigvee_{1 \leq i \leq k} x_{v,i} \Big) \wedge \bigwedge_{1 \leq i \leq k} \Big(\neg x_{v,i} \vee y_{v,0} \Big). \end{aligned}$$

Cette dernière formule étant en CNF, on pose donc :

$$\varphi_3 = \bigwedge_{v \in V} \Big(\neg y_{v,0} \vee \bigvee_{1 \leq i \leq k} x_{v,i} \Big) \wedge \bigwedge_{1 \leq i \leq k} \Big(\neg x_{v,i} \vee y_{v,0} \Big).$$

Question 6

On a :

$$\begin{aligned} \varphi_2 &\equiv \bigwedge_{v \in V} \bigwedge_{1 \leq l \leq d} \Bigg[\Big(\bigvee_{\substack{w \in V \text{ tq} \\ (v,w) \in E}} y_{w,l-1} \Big) \to y_{v,l} \Bigg] \wedge \Bigg[y_{v,l} \to \Big(\bigvee_{\substack{w \in V \text{ tq} \\ (v,w) \in E}} y_{w,l-1} \Big) \Bigg] \\ &\equiv \bigwedge_{v \in V} \bigwedge_{1 \leq l \leq d} \Bigg[\bigwedge_{\substack{w \in V \text{ tq} \\ (v,w) \in E}} \Big(\neg y_{w,l-1} \vee y_{v,l} \Big) \Bigg] \wedge \Bigg[\neg y_{v,l} \vee \Big(\bigvee_{\substack{w \in V \text{ tq} \\ (v,w) \in E}} y_{w,l-1} \Big) \Bigg]. \end{aligned}$$

Par construction, cette dernière formule est CNF et équivalente à φ_2.

Question 7

> Points attendus : structure de la preuve (instance positive de Proxy ssi formule satisfaisable) ; construction (et justification) d'une valuation à partir d'un certificat ; construction (et justification) d'un certificat à partir d'une valuation ; calcul en temps polynomial.

Posons

$$\varphi_4 = \bigwedge_{u \in V} y_{u,d},$$

et montrons que (G, k, d) est une instance positive de Proxy si et seulement si $\varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4$ est satisfaisable.

Si $\varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4$ est satisfaisable, soit ν une valuation satisfaisant cette formule. Posons $D = \{u \in V \mid \exists i \in [\![1, k]\!], \nu(x_{u,i}) = \texttt{vrai}\}$, et montrons que D est un ensemble d'au plus k sommets tel que tout sommet de V est à distance au plus d de D. Puisque ν satisfait φ_1, pour tout indice $i \in [\![1, k]\!]$, il y a au plus un sommet u tel que $\nu(x_{u,i}) = \texttt{vrai}$, donc D est sous-ensemble de V de taille au plus k. Montrons désormais par récurrence sur $l \in [\![0, d]\!]$ que

$$(H_l) : \text{pour tout sommet } v \in V,\ d(v, D) \leq l \text{ si et seulement si } \nu(y_{v,l}) = \texttt{vrai}.$$

Si $l = 0$, $d(v, D) \leq 0$ si et seulement si $v \in D$ càd, puisque ν satisfait φ_3, $\nu(y_{v,0}) = \texttt{vrai}$, donc (H_0). Si $l \in [\![0, d-1]\!]$ est tel que (H_l), montrons que la propriété est vraie au rang $l+1$. Soit $v \in V$: on a $d(v, D) \leq l+1$ si et seulement s'il existe un sommet $u \in V$, voisin de v, tel que $d(u, D) \leq l$, càd, en appliquant (H_l) à u, $\nu(y_{u,l}) = \texttt{vrai}$ i.e., puisque ν satisfait φ_2, $\nu(y_{v,l+1}) = \texttt{vrai}$. Donc (H_l), ce qui achève la récurrence.

> Résumé de la récurrence : grâce aux contraintes φ_2 et φ_3, les variables $y_{v,l}$ peuvent effectivement être interprétées comme « v est à distance au plus l de D ». La récurrence n'était pas attendue—mais aurait été appréciée.

Puisque ν satisfait φ_4, pour tout $v \in V$, $\nu(y_{v,d}) = \texttt{vrai}$ et donc, puisque (H_d), $d(v, D) \leq d$. Ainsi, tout sommet de V est à distance au plus d de D.

Réciproquement, si D est un ensemble d'au plus k sommets tel que tout sommet de V est à distance au plus d de D, considérons une énumération $(u_1, \dots, u_{|D|})$ des sommets de D (avec $|D| \leq k$) et posons pour tout $v \in V$, $i \in [\![1, k]\!]$:

$$\nu(x_{v,i}) = \begin{cases} \texttt{vrai} & \text{si } i \leq |D| \text{ et } v = u_i, \\ \texttt{faux} & \text{sinon,} \end{cases}$$

et, pour tout $v \in V$ et $l \in [\![1, d]\!]$, posons $\nu(y_{v,l}) = \texttt{vrai}$ si et seulement si v est à distance au plus l de D. Par construction de ν, cette valuation satisfait φ_1. Elle satisfait φ_2 et φ_3 car la fonction $d(-,-)$ est une distance. Finalement, elle satisfait φ_4 car tout sommet de v est à distance au plus d de D. Ainsi, ν satisfait $\varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4$, donc cette dernière formule est satisfaisable.

Ainsi, (G, k, d) est une instance positive de Proxy si et seulement si $\varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4$, qui est une formule CNF, est satisfaisable, càd $\varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4$ est une instance positive de Sat.

De plus, la formule $\varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4$ contient $\mathcal{O}(k|V|^2 + d|E| + k|V| + |V|) = \mathcal{O}(|V|^3)$ littéraux (sous l'hypothèse $d \leq |V|$), ce qui est polynomial en la taille de l'entrée. Puisque chaque étape élementaire pour construire cette formule (énumérer les sommets $v \in V$, tester si $(v, w) \in E$, etc.) se fait en temps polynomial en la taille de l'entrée, donc la fonction $(G, k, d) \mapsto \varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4$ est calculable en temps polynomial.

Ainsi, $(G, k, d) \mapsto \varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4$ est une réduction polynomiale de Proxy vers Sat.

> Souvent, la justification du fait que la fonction $(G, k, d) \mapsto \varphi_1 \wedge \varphi_2 \wedge \varphi_3 \wedge \varphi_4$ est calculable en temps polynomial a été oubliée.

Question 8

Puisque Sat est NP et puisque Proxy se réduit polynomialement à Sat, on en déduit que Proxy est NP.

Annexe : écrire des formules

Comment traduire une phrase en une formule ? C'est presque algorithmique :
- « pour tout $x \in X$, $P(x)$ » devient $\bigwedge_{x \in X} P(x)$,
- « il existe $x \in X$ tel que $P(x)$ » devient $\bigvee_{x \in X} P(x)$,
- « si P alors Q » devient $P \to Q$ ou encore $\neg P \vee Q$,
- « P ssi Q » devient $P \leftrightarrow Q$, ou encore $(P \to Q) \wedge (Q \to P)$, ou, en CNF : $(\neg P \vee Q) \wedge (\neg Q \vee P)$.

Par exemple, « pour tout sommet $v \in V$, pour tout $l \in [\![1, d]\!]$, il existe $w \in V$ tel que $(v, w) \in E$ et w est à distance au plus $l - 1$ de D si et seulement si v est à distance au plus l de D » s'exprime par :

$$\varphi_2 = \bigwedge_{v \in V} \bigwedge_{1 \leq l \leq d} \Bigg(\bigvee_{\substack{w \in V \text{ tq} \\ (v,w) \in E}} y_{w,l-1} \Bigg) \leftrightarrow y_{v,l}$$

sous l'hypothèse que les variables $y_{v,l}$ ($v \in V$, $l \in [\![0, d]\!]$) s'interprètent comme « v est à distance au plus l de D ».