HLIN608 Algorithmique du texte

TD Problème SSP

Formellement, l'assemblage se modélise par le problème suivant : étant donné un ensemble de mots $\mathcal{F} = \{F_1, \ldots, F_n\}$ (les fragments), trouver une superséquence S de longueur minimale telle que chaque mot de $\mathcal{F}$ est un sous-mot de la superséquence S. C'est le problème SSP (*Shortest Superstring Problem*). SSP est un problème NP-difficile pour plus de trois mots et un alphabet de plus de deux lettres. . .

Ici, on formalise ce problème sous la forme d'un problème d'*optimisation*, c'est-à-dire que l'on cherche à minimiser la longueur de la chaîne. Mais pour étudier la complexité d'un problème, et notamment son caractère NP-complet, on préfère se ramener à un problème de *décision*, c'est-à-dire un problème pour lequel on donne une réponse *oui/non*. La formalisation sous forme d'un problème de décision du problème SSP est la suivante :

Entrée : un ensemble de mots $\mathcal{F} = \{F_1, \ldots, F_n\}$, et un entier strictement positif K.
Question : existe-t-il une superséquence S de $\mathcal{F}$ de longueur $\leq K$ telle que chaque mot de $\mathcal{F}$ est un sous-mot de la superséquence S ?

Pour montrer que le problème de décision est NP-difficile et que son problème de décision associé est NP-complet, on procède par étapes.

— On montre d'abord que le problème d'optimisation est au moins aussi difficile que le problème de décision,
— puis que le problème de décision appartient à la classe NP, c'est-à-dire qu'il existe un algorithme polynomial pour *vérifier* une solution potentielle.
— Ensuite on exhibe une réduction polynomiale depuis un problème connu pour être NP-complet,
— et on prouve que la réduction transporte les solutions d'un problème à l'autre.

Exercice 1 *Montrer que si l'on a un algorithme polynomial pour le problème d'optimisation alors on a un algorithme polynomial pour le problème de décision. En déduire que le problème d'optimisation est "au moins aussi difficile" que le problème de décision.*

Remarque : on peut également montrer que dans ce cas précis, le problème d'optimisation est "au moins aussi facile" que le problème de décision, mais c'est un peu plus compliqué. Désormais, on considère le problème de décision. Pour prouver sa NP-complétude, on va opérer par réduction depuis un problème de base qui est NP-complet. On commence par prouver que le problème est dans la classe NP.

Exercice 2 *Montrer qu'il existe un algorithme polynomial pour vérifier si, pour une instance donnée du problème et une superchaîne, cette chaîne satisfait le problème de décision.*

On va maintenant prouver que le problème SSP est "au moins aussi difficile" qu'un problème NP-complet de base, ce qui prouvera son caractère complet. Le problème de base que l'on considère est le problème appelé "VERTEX COVER", qui nécessite la définition suivante :

Definition Soit $G = (V, E)$ un graphe, et C une partie de V. L'ensemble de sommets C est une *couverture des sommets* (vertex cover) de G si pour toute arête $e = \{u, v\} \in E$, au moins l'une des extrêmités u ou v de l'arête e est dans C.

Le problème de décision VERTEX COVER s'énonce comme suit :

Entrée : un graphe $G = (V, E)$, et un entier strictement positif k.
Question : existe-t-il une couverture des sommets de G, notée C, de taille $\leq k$?

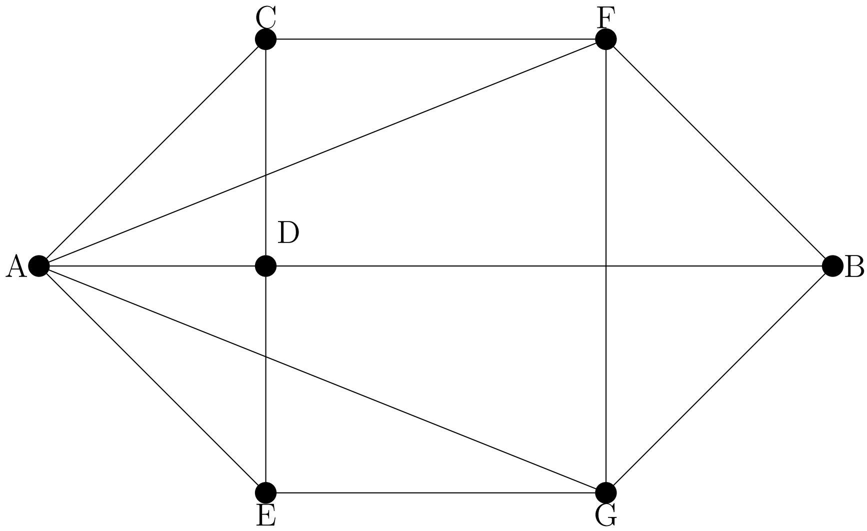

FIGURE 1 – Un graphe non orienté.

Exercice 3 *Exhiber une couverture minimale pour le graphe de la Figure 1.*

Passons maintenant à la réduction : on se donne une instance de VERTEX COVER, et on va la transformer en une instance de SSP à l'aide d'opérations polynomiales.

Exercice 4 *Montrer que si on dispose de cette réduction, alors si SSP admet un algorithme polynomial, on en déduit un algorithme polynomial pour VERTEX COVER.*

La réduction est la suivante : Soit une instance de VERTEX COVER, $[G = (V, E), K]$. On la transforme en instance de SSP par :

1. $\Sigma = V$ (l'alphabet considéré est l'ensemble des sommets)
2. L'ensemble $\mathcal{F}$ est constitué des chaînes *abab* et *baba* pour chaque arête $\{a, b\} \in E$.

Exercice 5 *Transformer le graphe de la Figure 1 en instance de SSP.*

Exercice 6 *(*) On suppose que $|E| = m$. Démontrer que G a une couverture des sommets de taille k si et seulement si son instance tranformée admet une superchaîne de taille $4m + k$. Conclure.*

Voilà, vous avez réalisé votre première preuve de NP-complétude, bravo !