I. La NP-Complétude

A. Préliminaires

Définition : problème de décision
Un *problème de décision* A est la donnée d'un ensemble E d'*instances* et d'un sous-ensemble $E_+ \subset E$ d'*instances positives.* On associe en général à un ensemble d'instances E un codage $\langle . \rangle : E \to \Sigma^*$. Par abus de notation, on note aussi A le langage $\{\langle p \rangle : p \in E_+\}$.
Exemples
SAT est le problème de décision dont les instances sont les formules du calcul propositionnel et dont les instances positives sont les formules satisfiables. 3-SAT est le problème de décision dont les instances sont les formules de la forme $c_1 \wedge c_2 \wedge \dots \wedge c_l$ où c_i est une clause contenant 3 littéraux. Les instances positives sont les formules satisfiables.

Définition : les classes P et NP
Un problème de décision A est dans la classe P (resp. $NP)$ si une MT déterministe (resp. non déterministe) décide A en temps au plus polynomial par rapport à la taille de l'entrée.
Exemples
SAT et 3-SAT sont dans la classe NP

Théorème de Cook
Si SAT est dans NP, alors $P = NP$.

Cela motive la définition de la classe des problèmes qui vérifient cette propriété :

B. Problèmes NP-complets

Définition : réduction polynomiale
Soient A et B deux problèmes de décision. On dit que A se *réduit polynomialement à* B et on note $A \leq_P B$ si il existe une fonction $f: \Sigma^* \to \Sigma^*$ calculable en temps polynomial telle que $x \in A$ ssi $f(x) \in B$.
Exemple
Il est clair que 3-SAT se réduit polynomialement à SAT. 3-SAT se réduit aussi polynomialement à SAT.

Définition : problème NP-complet
Un problème de décision A est NP-complet si : • A est NP ; • Pour tout problème B dans NP, $B \leq_P A$.

Proposition
Si A est un problème NP-complet et B est un problème NP tel que $A \leq_P B$ alors B est NP-complet.

Par ce qu'on a dit avant, SAT et 3-SAT sont donc NP-complets.

II. Problèmes de graphes NP-complets

A. Réductions depuis 3-SAT

On peut réduire polynomialement 3-SAT à de nombreux problèmes de graphes. Ces réductions se font à l'aide de gadgets, structures se répétant dans les graphes qu'on construit et représentant les variables et clauses des formules de 3-SAT.

Définition-proposition
Le problème CLIQUE est le problème défini par : • Les instances sont la donnée d'un graphe fini orienté G et d'un entier k ; • Les instances positives sont les graphes dont il existe un sous-graphe de G complet à k sommets. CLIQUE est NP-complet.

Définition-proposition (DEVELOPPEMENT)
Le problème HAMPATH est le problème défini par : • Les instances sont la donnée d'un graphe fini orienté G et de sommets s et t de G ; • Les instances positives sont les graphes dont il existe un chemin de s à t passant une et une seule fois par chacun des sommets de G. HAMPATH est NP-complet

Définition-proposition (DEVELOPPEMENT)
Le problème VERTEX-COVER est le problème défini par : • Les instances sont la donnée d'un graphe fini non orienté G et d'un entier k ; • Les instances positives sont les graphes G dont il existe un sous-graphe à k sommets contenant toutes les arêtes de G. VERTEX-COVER est NP-complet

B. Réductions plus avancées

Définition-proposition
Le problème HAMCYCLE est le problème défini par : • Les instances sont la donnée d'un graphe orienté G • Les instances positives sont les graphes G dont il existe un cycle passant une et une seule fois pas chaque arête. HAMCYCLE est NP-complet : la réduction se fait depuis HAMPATH Idée : on ajoute un sommet u à l'instance de HAMPATH et deux arêtes $u \rightarrow s$ et $t \rightarrow u$. On a alors existence d'un cycle hamiltonien dans le nouveau graphe si et seulement si un chemin hamiltonien reliait s à t.

Définition-proposition
Le problème du voyageur de commerce TS est le problème défini par : • Les instances sont la donnée d'un entier n, de $d: \{0, \dots, n\}^2 \rightarrow \mathbb{N}$, et d'un entier k ; • Les instances positives sont les graphes G dont il existe une suite de sommets $x_0, \dots, x_n$ tel que chaque entier i apparaît une et une seule fois et tel que $\sum d(x_i, x_{i+1}) + d(x_n, x_0) \leq k$. TS est NP-complet : la réduction se fait depuis HAMCYCLE : Etant donnée un graphe G, on pose $d(i,j) = 1$ si (i,j) est une arête de G et 0 sinon. On pose de plus $k = n$ et le problème TS pour cette instance est équivalent au problème HAMCYCLE pour G.

III. Problèmes ensemblistes/sur les entiers

A. SET-COVER

Définition-proposition
Le problème SET-COVER est le problème défini par : • Les instances sont la donnée d'un univers fini U, d'un ensemble de parties $S = \{S_1, \dots, S_p\}$ de U et d'un entier k. • Une instance est positive s'il existe un sous-ensemble de S à k éléments qui recouvre U. SET-COVER est NP-complet : la réduction se fait depuis VERTEX-COVER : Idée de la preuve : Etant donnée une instance de VERTEX-COVER, on prend pour univers l'ensemble A des arêtes de G et on pose S_i l'ensemble des arêtes incidentes au sommet i. On voit facilement que c'est une réduction de VERTEX-COVER à SET-COVER.

B. SUBSET-SUM

Définition-proposition
Le problème SUBSET-SUM est le problème défini par : • Les instances sont la donnée d'un ensemble d'entiers et d'un entier cible t; • Les instances positives sont les ensembles E dont il existe un sous-ensemble de somme égale à t. TS est NP-complet : la réduction se fait depuis 3-SAT

IV. La programmation linéaire en entiers

A. Définitions

La programmation linéaire est à l'origine un problème d'optimisation de la forme

$$\begin{cases} \text{minimiser}\, c \cdot x \\ Ax \leq b \end{cases}$$

où c est un vecteur et A une matrice et b un vecteur. Ce problème est dans la classe P si x est une variable continue, mais si on restreint x à des valeurs entières, il devient difficile.

Définition-proposition

Le problème de décision PLE est le problème dont les instances sont la donnée de $A \in \mathcal{M}_{m,n}(\mathbb{Z}), b \in \mathbb{Z}^m, c \in \mathbb{Z}^n, k \in \mathbb{Z}$. Une instance est positive si il existe $x \in \mathbb{Z}^n$ tel que :

$$\begin{cases} c \cdot x \leq k \\ Ax \leq b \end{cases}$$

Comme on va le voir, PLE est NP-complet car il permet d'exprimer de nombreux problèmes NP-complets.

B. Réductions de problèmes à PLE

Le problème SET-COVER

On se place sur $\mathbb{Z}^S$ et on cherche $x \in \mathbb{Z}^S$ tel que :

$$\sum x_s \leq k$$

$\sum_{s:s\ est\ un\ sommet\ de\ a} x_s \geq 1$ pour tout arête a

$x_s \leq 1$ et $x_s \geq 0$ pour tout $s \in S$.

Le problème SET-COVER est équivalent à l'existence d'un tel x.

Le problème KNAPSACK

Le problème du sac à dos KNAPSACK est le problème dont les instances sont la donnée d'un volume V et de couples $(v_1, s_1), \dots, (v_n, s_n)$ et d'un entier k. Le problème est de savoir s'il existe un sous-ensemble S de $\{1, \dots, n\}$ tel que $\sum_{i \in S} s_i \geq k$ et $\sum_{i \in S} v_i \leq V$.

On admet que ce problème est NP-complet et il s'exprime naturellement à l'aide de PLE.

Remarque : le problème du sac à dos a un algorithme polynomial en V et n : il est dont dans la classe P si l'entrée V est donnée en écriture polynomiale. On dit aussi que ce problème est *pseudo-polynomial.*

Le problème TS

Etant donné une instance du problème du voyageur de commerce, on choisit pour variables $x_{i,j}, i, j \in \{0, \dots, n\}$ et on considère le problème

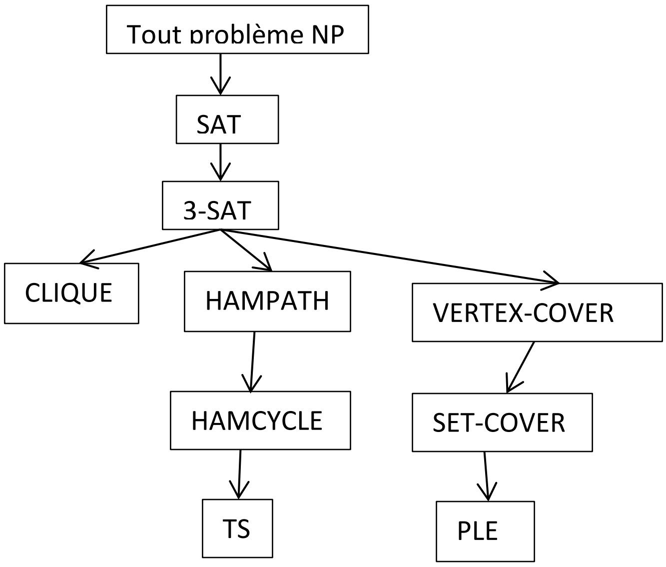
$$\sum x_{i,j} d(i,j) \leq k$$

Pour tout i, j $0 \leq x_{i,j} \leq 1$: $x_{i,j}$ sont booléennes

Pour tout i, $\sum_{j \neq i} x_{i,j} = 1$: une seule arête sortante

Pour tout j, $\sum_{i\neq j} x_{i,j} = 1$: une seule arête entrante
Pour tout sous-ensemble $S \subset \{0, .. n\}$,$S \neq \emptyset$, $\sum_{i\in S, j\notin S} x_{i,j} \geq 2$: on impose ici qu'on a un seul cycle et non des cycles disjoints
Attention : cette réduction n'est pas polynomiale par rapport à la taille de l'entrée : le nombre de contraintes est exponentiel! Il existe bien une réduction polynomiale, mais c'est plus compliqué ! (cf. Wikipedia)

Références : surtout Sipser, sinon Wolper, Carton, Papadimitriou algorithmes, programmation mathématique.



(1)

Rappel: Soit G un graphe. Un chemin hamiltonien est un chemin qui passe une et une seule fois par chaque sommet de G:

Th: Hampath.
Le problème de savoir si un graphe G donné ~~contient~~ admet un chemin hamiltonien est NP-complet.

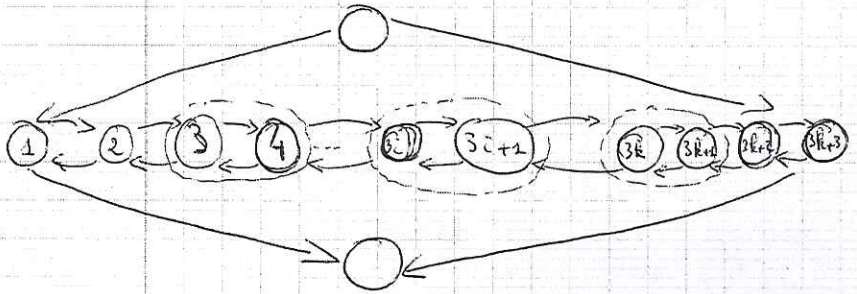
Preuve: Nous allons procéder par réduction de SAT vers Hampath.

Soit φ une formule du calcul propositionnel sous forme normale conjonctive. $\varphi = C_1 \wedge \dots \wedge C_k$ avec C_i une clause.

Notons $\{x_1, \dots, x_l\}$ les variables de φ.

Nous construisons le graphe G suivant :

- Pour chaque clause, on a un sommet C_i
- Pour chaque variable x_j, on associe le gadget suivant:



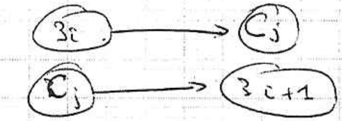
Les sommets $3i$, $3i+1$ représentent la clause C_i

Nous allons nous en servir pour relier les sommets C_i aux gadgets des "x_j"

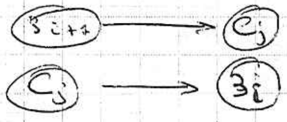
(2)

. On rajoute les arête suivante:

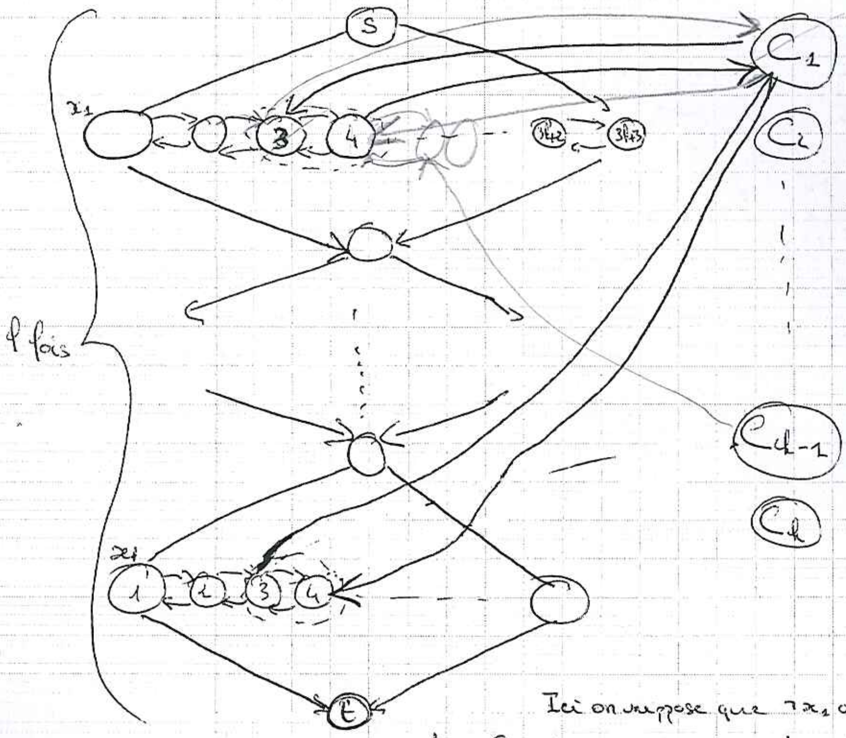
- si x_j apparait dans C_i, on rajoute.



- Si $\neg x_j$ apparait dans C_i; on rajoute



On obtient donc un graphe comme ceci:



Ici on suppose que $\neg x_1$ apparait dans C_1, et que x_l apparait dans C_1

Juste pour le dessin.

(3)

Nous remarquons deux choses;

- Un chemin hamiltonien doit commencer par (s), car il n'y aucune arête allant vers ce sommet.
- Un chemin hamiltonien doit finir par (t), car il n'y a aucune arête sortante.

(*) Supposons φ satisfaisable, alors il existe une distribution de vérité I qui satisfait φ.

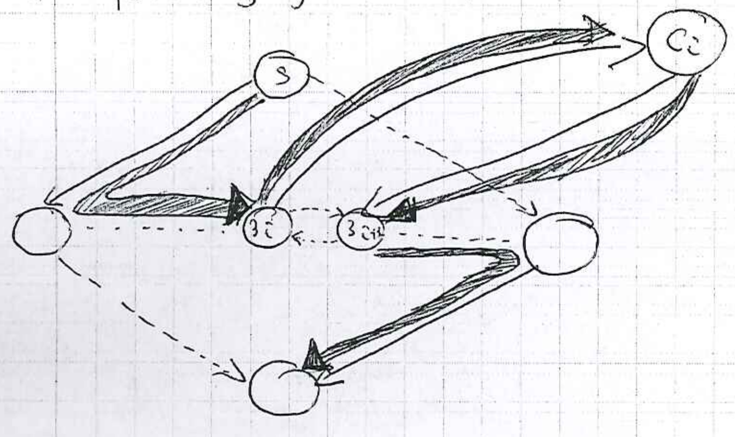
1) On part de (S)

2) si $I(x_1) = 1$, alors on prend la flèche de gauche sinon, la flèche de droite

3) Supposons que $I(x_1) = 1$, alors on parcourt les sommets du milieu de gauche à droite.

Si x_1 apparait dans C_i, alors on visite ce sommet en passant, si il n'a pas déjà été visité.

4) On passe au gadget suivant.



(4)

Si $I(x_1) = 0$ et

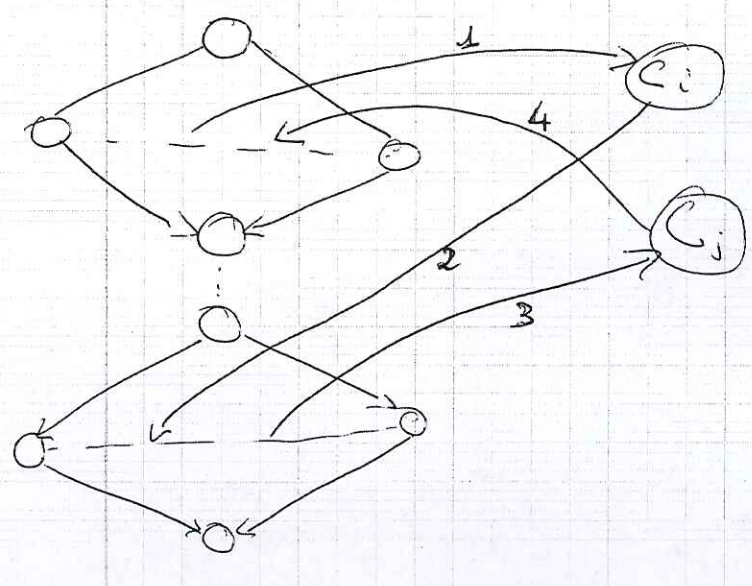
Si $(\neg x_1)$ apparait dans C_i, alors on transpose la méthode.

Ce procédé nous fournit bien un chemin Hamiltonien :

- On visite tous les sommets de tous les gadget associés aux "x_j".
- Toutes les clause C_i sont vraies, donc il existe une variable x_j tq x_j ou $\neg x_j$ rend cette clause vraie: ainsi on aura visité (C_i) pendant le parcours du gadget associé à x_j.

(*) Supposons que G admette un chemin Hamiltonien. Alors il ne peut être que de la forme précédente.

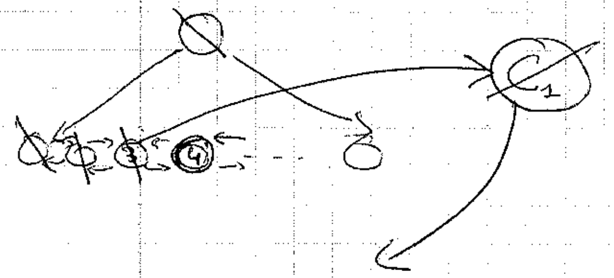
Le seul accident qui peut se produire est le suivant.



5

Cela revient à passer d'un gadget à un autre en passant par les sommets C_i.

Supposons qu'en "x_1", on a fait la transition

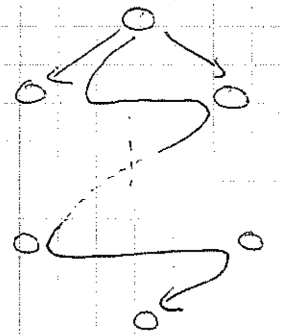


Il faudra forcément visiter 4.

on ne peut pas utiliser la transition $C_2 \longrightarrow 4$ car C_1 a déjà été visité.
On ne peut qu'utiliser la transition $3 \longrightarrow 4$

Or une fois arrivée en 4, on ne peut plus rien faire car tous ses voisins ont été visités. Notre chemin doit donc finir par 4 : Ceci contredit le fait qu'un chemin hamiltonien de G doit finir par F.

Un chemin hamiltonien de G est donc de la forme



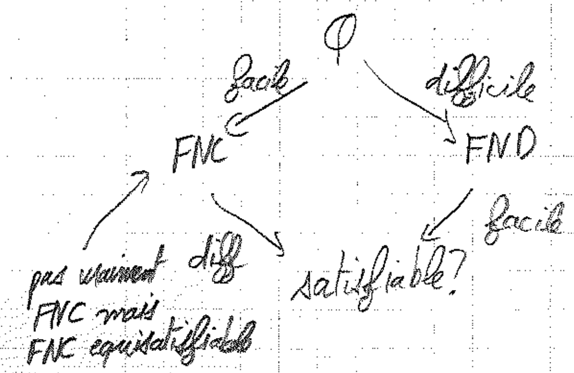
6

Si on emprunte la ligne associée à x_j de la droite vers la gauche, on pose $I(x_j) = 0$, sinon $I(x_j) = 1$.

I satisfait donc φ.

On a doit bien réduit SAT à Hampath, or SAT est NP-Complet. De plus, pour construire G, on a dû générer $\ell \times (3k+4) + 1$ sommet, et moins de $(\ell \times (3k+4) + 1)^2$ arêtes puisque notre graphe n'est pas complet.
Notre réduction est donc calculable en temps polynomial :

Hampath est donc NP-Complet.



Th: (Vertex-Cover) Le problème de savoir si pour un graphe $G=(S,E)$ donné et une constante j donnée, il existe un sous-ensemble $S' \subset S$ tq $|S| \leq j$ et pour toute arête $(u,v) \subset E$, $u \in S'$ ou $v \in S'$, est NP-complet.

Preuve: Nous allons faire une réduction de 3-SAT, dont nous savons qu'il est NP-complet, vers Vertex-Cover.

Soit φ une instance de 3-SAT: $\varphi = E_1 \wedge \dots \wedge E_k$
avec $E_i := x_{i1} \vee x_{i2} \vee x_{i3}$ où x_{ij} est un littéral.

Notons $P=(p_1,\dots,p_\ell)$ les variables de φ.
On construit le graphe $G=(S,E)$ suivant.

- $S = \left\{ \begin{array}{l} p_i \text{ et } \neg p_i, \ \forall i \in [\![1,\ell]\!] \\ x_{i1}, x_{i2} \text{ et } x_{i3}, \ \forall i \in [\![1,k]\!] \end{array} \right\}$ $2\ell + 3k$ sommets

- $E = \{(p_i, \neg p_i) \text{ pour tout } i \in [\![1,\ell]\!]\}$
$\cup \{(x_{i1},x_{i2}),(x_{i2},x_{i3}),(x_{i3},x_{i1}), \forall i \in [\![1,k]\!]\}$
$\cup \{(x_{ij}, p_k) \text{ si } x_{ij} = p_k \text{ ou } (x_{ij}, \neg p_k) \text{ si } x_{ij} = \neg p_k\}$

Nous avons donc des arêtes du type
- p_i — $\neg p_i$ $\quad 2\ell$
- x_{i1} — x_{i2} — x_{i3} — x_{i1} $\quad + 3k$
- x_{ij} — p_k ou x_{ij} — $\neg p_k$ $\quad + 3k$

$= \ell + 6k$ arêtes

Nous pouvons donc construire notre graphe en temps polynomial.

Mq φ est satisfaisable ssi G admet une couverture de sommet de taille au plus $\ell + 2k$.

(Le choix de $j = \ell + 2k$ donne tout son sens au graphe choisi).

- Supposons φ satisfaisable: Il existe I une distribution de vérité satisfaisant φ.

 * On prend donc dans S', tous les p_i ou $\neg p_i$ qui ont pour valeur de vérité 1 via I. Il y en a exactement ℓ car on ne peut pas avoir p_i et $\neg p_i$ en même temps.

 * Pour tout $i \in [\![1,k]\!]$, on choisit un x_{ij} qui a pour valeur 1: ce choix est toujours possible puisque $x_{i1} \vee x_{i2} \vee x_{i3}$ doit être vrai pour tout i. On rajoute les 2 autres sommets dans S'.

On a donc un sous-ensemble de sommets de taille $\ell + 2k$.

De plus, pour toute arête:
- p_i — $\neg p_i$, soit $p_i \in S'$, soit $\neg p_i \in S'$.
- x_{i1} — x_{i2} — x_{i3}, on a toujours 2 sommets sur 3, donc toutes les arêtes ont une extrémité dans S'.
- x_{ij} — p_k: • si x_{ij} n'est pas dans S', alors $x_{ij} = 1 = p_k$ donc p_k est dans S'.
 • si x_{ij} est dans S', c'est bon.

S' est donc une couverture de sommet de taille $l+2k$.

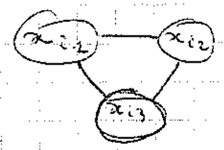
- Supposons que G admet une couverture de taille inférieure ou égale à $l+2k$.

1) $\forall i \in [\![1, l]\!]$, on a une arête p_i — $\neg p_i$

donc on doit prendre l'un des deux.

$\longrightarrow$ l sommets.

2) $\forall i \in [\![1, k]\!]$,



on doit forcément prendre 2 sommets sur 3.

$\rightarrow 2k$ sommet.

Si à l'étape 1, on a pris p_i, et qu'à l'étape 2 on a exclu x_{ij} tq on ait x_{ij} — $\neg p_i$

alors on peut prendre $\neg p_i$ à la place de p_i.

Ainsi on n'augmente pas le nombre de sommets.

En prenant I la distribution qui attribut 1 à p_i ssi il est dans S' est une distribution qui satisfait φ.

On a donc bien réduit 3-SAT à Vertex-Cover,

donc Vertex-Cover est NP-complet $\blacksquare$