11. Quelques exemples de réductions

Enseignant: Arnaud Casteigts *Exercices: A. Berger, M. De Francesco, L. Heiniger*

Pour rappel, pour montrer qu'un problème est NP-complet, il y a deux étapes à effectuer :

- Montrer qu'il est dans NP,
- Montrer qu'il est NP-difficile.

La stratégie habituelle pour établir le second point est de choisir un autre problème NP-difficile et montrer qu'il peut se réduire en temps polynomial au problème considéré. Voici une liste de réductions classiques. Nous allons en détailler certaines et donner quelques idées pour les autres :

- SAT $\leq_p$ Clique
- Clique $\leq_p$ Independent Set (déjà vu au Cours n°10)
- SAT $\leq_p$ 3-SAT
- 3-SAT $\leq_p$ 3-Coloring
- 3-Coloring $\leq_p$ 4-Coloring $\leq_p$ 5-Coloring $\leq_p \cdots$
- Clique $\leq_p$ Subgraph Isomorphism
- SAT $\leq_p$ Directed Hamiltonian Cycle
- Directed Hamiltonian Cycle $\leq_p$ Hamiltonian Cycle

11.1 Définition des problèmes

Rappelons la définition de ces problèmes, en détaillant un peu plus celle de SAT.

- Clique : Étant donné un graphe G et un entier k, est-ce que G contient une clique de taille k ?
- Independent Set : Étant donné un graphe G et un entier k, est-ce que G contient un ensemble indépendant de taille k ?
- 3-Coloring : Étant donné un graphe G, peut-on colorier ses sommets en utilisant 3 couleurs sans donner la même couleur à des sommets voisins ?
- Graph Isomorphism : Étant donnés deux graphes G_1 et G_2, est-ce que ces graphes sont structurellement identiques ?
- Subgraph Isomorphism : Étant donnés deux graphes G_1 et G_2, est-ce que G_1 contient G_2 comme sous-graphe ?

- Hamiltonian Cycle : Étant donné un graphe G, existe-t-il un chemin qui visite tous les sommets exactement une fois, puis termine au point de départ ?
- Directed Hamiltonian Cycle : Idem, pour un graphe orienté, en respectant l'orientation des arcs.
- SAT : Étant donné une formule booléenne ϕ sous forme normale conjonctive, c.à.d. une conjonction de disjonction de *litéraux* (variables ou leur négation), par exemple :

  $\phi = (x_1 \vee \overline{x_2} \vee x_3) \wedge (x_2 \vee \overline{x_3})$ ($\overline{x_i}$ est synonyme de $\neg x_i$)

  Existe-t-il des valeurs `vrai`/`faux` pour chaque variable x_i telles que la formule est satisfaite ? Chaque disjonction de la formule, par exemple $(x_1 \vee \overline{x_2} \vee x_4)$, est appelée une *clause*. Cette formule a donc 2 clauses, 3 variables, et 5 litéraux.

  Ici, la réponse est oui, par exemple mettre toutes les variables à `vrai` suffit à satisfaire la formule. En revanche, la formule

  $$\phi_2 = (x_1 \vee x_2) \wedge (x_1 \vee \overline{x_2}) \wedge (\overline{x_1} \vee x_2) \wedge (\overline{x_1} \vee \overline{x_2})$$

  n'est pas satisfaisable. On a donc $\phi \in$ SAT et $\phi_2 \notin$ SAT.

- 3-SAT : Même problème, mais où chaque clause est de taille 3, par exemple :

  $$\phi_3 = (x_1 \vee \overline{x_2} \vee x_4) \wedge (\overline{x_1} \vee x_2 \vee \overline{x_4})$$

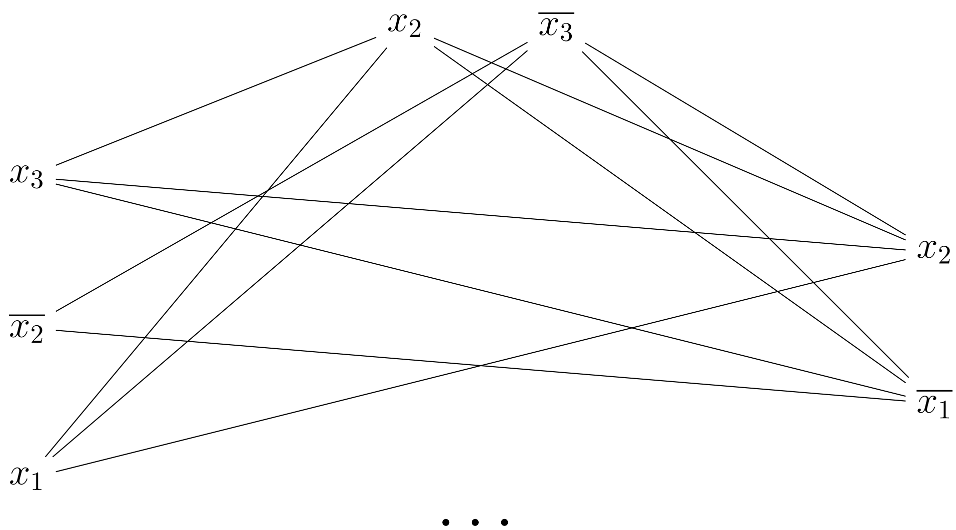
  Notez que si une clause est de taille inférieure à 3, on peut facilement la convertir en une clause de taille 3, en répétant simplement certaines variables. Il s'agit donc du même problème.

11.2 Les réductions

Autant que possible, nous présentons les réductions par le biais d'un exemple, en discutant la façon dont ils peuvent se généraliser.

SAT $\leq_p$ Clique :

Étant donné une formule, par exemple $\phi = (x_1 \vee \overline{x_2} \vee x_3) \wedge (x_2 \vee \overline{x_3}) \wedge (\overline{x_1} \vee x_2) \wedge \cdots$ comportant k clauses, on construit un graphe G comme suit : d'abord, on crée un sommet pour chaque litéral de chaque clause. Puis on relie chaque litéral avec les litéraux des *autres* clauses qui sont *compatibles* (par exemple, on ne relie pas x_1 et $\overline{x_1}$, mais on relie x_1 et x_2, ainsi que x_2 et x_2). Cela donne le graphe suivant :



Pour une formule de taille arbitraire, on procède de la même manière (petits points en bas), avec autant de groupes de sommets que de clauses, et autant de sommets dans le groupe que de litéraux dans la clause.

Théorème : La formule ϕ est satisfaisable si et seulement si G admet une clique de taille k.

Preuve : Pour démontrer ce "si et seulement si", il faut montrer les deux sens : (1) une clique de taille k implique l'existence d'une solution pour ϕ, et (2) une solution pour ϕ implique l'existence d'une clique de taille k. Ici, en réfléchissant un peu, on peut se convaincre facilement que c'est vrai. Mais faisons l'effort de le démontrer (au moins une fois).

(1) : Chaque litéral n'est relié qu'à des litéraux d'autres clauses, donc une clique de taille k doit faire intervenir un litéral dans chaque clause. Par ailleurs, ces litéraux sont tous compatibles entre eux (car ils partagent une arête deux à deux). Une clique de taille k donne donc bien une solution pour ϕ.

Par exemple, la clique x_3 (gauche), x_2 (haut), x_2 (droite) indique que les trois clauses peuvent être satisfaites en mettant x_3 et x_2 à `vrai` (x_1 étant quelconque). De même, la clique x_3 (gauche), x_2 (haut), $\overline{x_1}$ (droite) indique que les trois clauses peuvent être satisfaites en mettant x_3 et x_2 à `vrai` et x_1 à `faux`.

(2) : Si une solution pour ϕ existe, alors il existe une affectation pour les variables qui satisfait chacune des k clauses de ϕ. Chacune de ces clauses est satisfaite via un ou plusieurs litéraux. On peut choisir arbitrairement un litéral satisfait pour chaque clause. Cet ensemble de litéraux est nécessairement compatible, les sommets correspondants ont donc tous une arête entre eux dans G.

Conclusion : SAT est NP-difficile et se réduit à Clique, donc Clique est NP-difficile. Par ailleurs, Clique est aussi dans NP, donc il est NP-complet.

SAT $\leq_p$ 3-SAT :

Principe de la réduction : Étant donné une formule ϕ dont les clauses sont de longueur arbitraire, on peut trouver une formule ϕ' dont les clauses sont de longueur 3 telle que ϕ' est satisfaisable si et seulement si ϕ est satisfaisable. L'idée principale est que chaque clause longue peut être cassée en plusieurs clauses de longueur 3 tout en préservant l'équivalence entre les deux formules. Cela requiert l'ajout de variables supplémentaires.

Prenons l'exemple d'une clause très simple : $(x_1 \vee x_2 \vee x_3 \vee x_4)$. Cette clause est satisfaisable si et seulement si au moins une variable parmi x_1, x_2, x_3, x_4 est à `vrai`. On peut la casser en deux en introduisant une nouvelle variable x_a, en écrivant $(x_1 \vee x_2 \vee x_a) \wedge (\overline{x_a} \vee x_3 \vee x_4)$. Cette nouvelle formule reste bien satisfaisable si et seulement si au moins une variable parmi x_1, x_2, x_3, x_4 est satisfaite (intuitivement, si $x_1 \vee x_2$ n'est pas satisfait, alors $x_3 \vee x_4$ doit l'être, et vice versa).

La même technique fonctionne avec des litéraux artitraires, par exemple $(x_1 \vee \overline{x_2} \vee \overline{x_3} \vee x_4)$ devient $(x_1 \vee \overline{x_2} \vee x_a) \wedge (\overline{x_a} \vee \overline{x_3} \vee x_4)$. Par ailleurs, si la clause est plus grande, on peut chaîner les petites clauses avec d'autres variables intermédiaires, par exemple $(x_1 \vee x_2 \vee x_3 \vee x_4 \vee x_5 \vee x_6)$ devient $(x_1 \vee x_2 \vee x_a) \wedge (\overline{x_a} \vee x_3 \vee x_b) \wedge (\overline{x_b} \vee x_4 \vee x_c) \wedge (\overline{x_c} \vee x_5 \vee x_6)$. Bien sûr, la taille de la formule augmente, mais la nouvelle formule reste de taille polynomiale en l'ancienne, on a donc bien la réduction souhaitée.

Clique $\leq_p$ Subgraph Isomorphism :

Cette réduction est vraiment directe : Étant donné une instance (G, k) pour le problème Clique, on construit deux graphes G_1 et G_2 tels que G_1 est une copie de G et G_2 est une clique de taille k. Clairement, G_2 est un sous-graphe de G_1 si et seulement si G contient une clique de taille k.

3-SAT $\leq_p$ 3-Coloring :

Une réduction assez astucieuse (pour ne pas dire géniale), qui sera faite en exercices.

3-Coloring $\leq_p$ 4-Coloring :

Étant donné un graphe G dont on veut savoir s'il est 3-colorable, on construit un graphe G' comme une copie de G, à laquelle on ajoute un sommet universel (somme relié à tous les autres). Ce sommet étant relié à tous les autres, il doit avoir une couleur unique. On a donc bien que G' est 4-colorable si et seulement si G est 3-colorable.

11.3 Discussions

Le théorème de Ladner nous dit que si $\mathsf{P} \neq \mathsf{NP}$, alors il doit exister des problèmes intermédiaires dans NP, c'est à dire qui ne sont ni NP-complets ni dans P. Cependant, pour des raisons encore énigmatiques, la majorité des problèmes sont soit dans l'un, soit dans l'autre, et leur difficulté opère une transition de phase directement de P à NP-complet lors qu'on fait varier leur définition. Par exemple, 2-SAT est polynomial, mais 3-SAT (et plus généralement, k-SAT, pour tout $k \geq 3$) est NP-complet. De même, 2-Coloring est polynomial, mais k-Coloring est NP-complet pour tout $k \geq 3$. Deux problèmes naturels suspectés intermédiaires sont Graph Isomorphism et Factoring. À ce jour, personne n'a réussi ni à résoudre ces problèmes en temps polynomial, ni à montrer qu'ils sont NP-complet. Il est considéré improbable qu'ils soient NP-complets, car cela aurait des conséquences théoriques importantes (réfutant plusieurs conjectures jugées plausibles). En revanche, ils pourraient s'avérer être dans P sans conséquence théorique majeure (mais bien sûr avec des conséquences pratiques énormes, pour Factoring).