

Recherche de codes attribuants dans un graphe biparti

Emmanuel CHARBIT

3 juillet 2006

Un exemple de notre problème en des termes concrets

Considérons un ensemble d'individus que l'on parvient tous à différencier grâce à un ensemble d'attributs (couleur des cheveux, taille, QI, marque de voiture possédée). La question est de savoir quel est le sous-ensemble de plus petit cardinal de cet ensemble initial d'attributs, qui nous permette encore de différencier tous nos individus.

1 Cadre général du problème

1.1 Enoncé du problème d'identification par des attributs

On considère ici un graphe biparti

$$G = (X = I \cup A, E)$$

où I et A sont les deux parties du graphe, et E l'ensemble de ses arêtes :

$$E = \{\{i, a\}\} \subseteq I \times A.$$

On notera $V(i)$ l'ensemble des voisins de $i \in I$: $V(i) = \{j \in X \text{ t.q. } \{i, j\} \in E\}$.

On dira que notre graphe est sans jumeaux, si et seulement si

$$\forall i \neq j, V(i) \neq V(j)$$

On s'intéresse à un graphe biparti sans jumeaux tel que défini ci-dessus et on cherche alors $B \subseteq A$ tel que $\forall (i, j) \in I^2, i \neq j, V(i) \cap B \neq V(j) \cap B$.

Théorème 1.1 *Un sous-ensemble B vérifiant ce qui précède existe si et seulement si G est sans jumeaux.*

Si G est sans jumeaux, on cherche finalement B de cardinal minimal.

Ainsi, pour un graphe biparti $G = (I \cup A, E)$ donné, on cherche donc $B \subseteq A$ de cardinal minimal permettant de n'avoir toujours aucun jumeau dans I .

Par soucis de concision, un tel B sera, dans la suite, appelé IDIMIN pour G .

1.2 Exemple

L'exemple suivant a été choisi au hasard. Nous allons vérifier que ce graphe biparti est bien sans jumeaux pour I , et nous allons extraire de A un sous-ensemble, dont nous allons vérifier qu'il est de cardinal minimal vérifiant que I est toujours sans jumeaux.

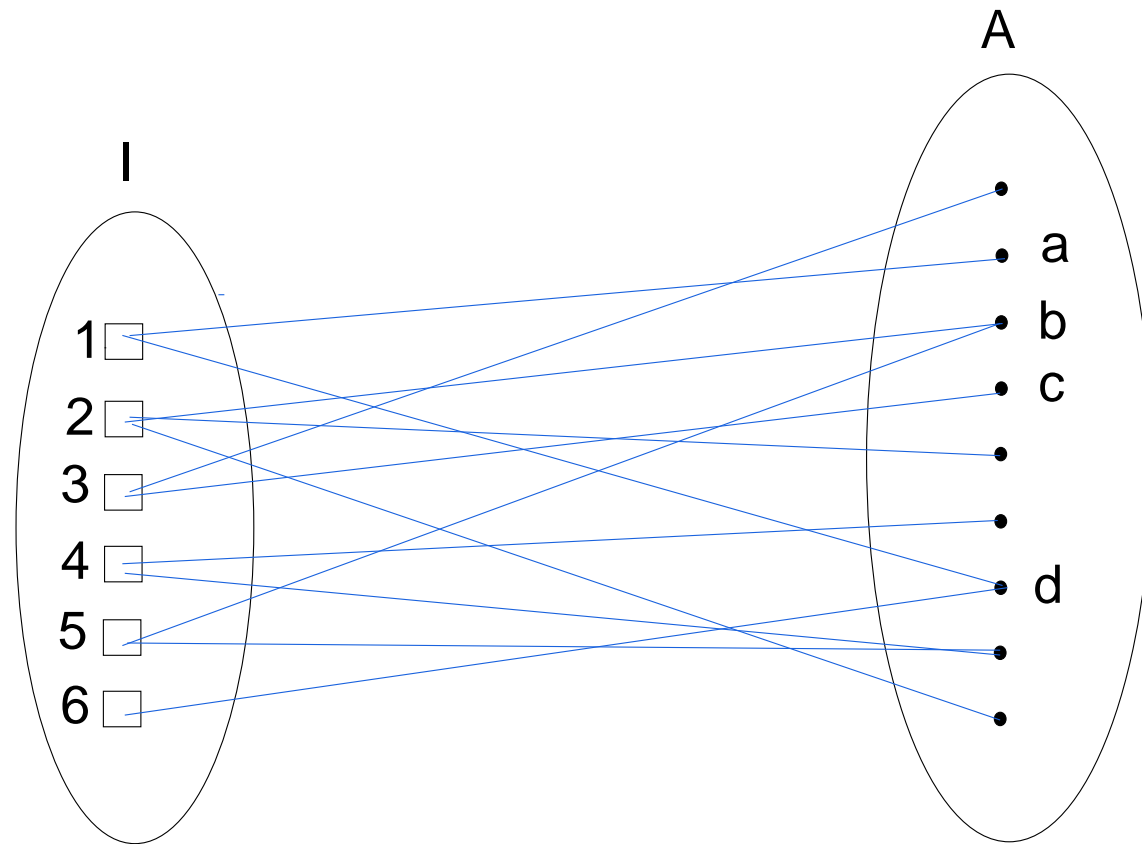


FIG. 1 – *Graphe biparti à identifier*

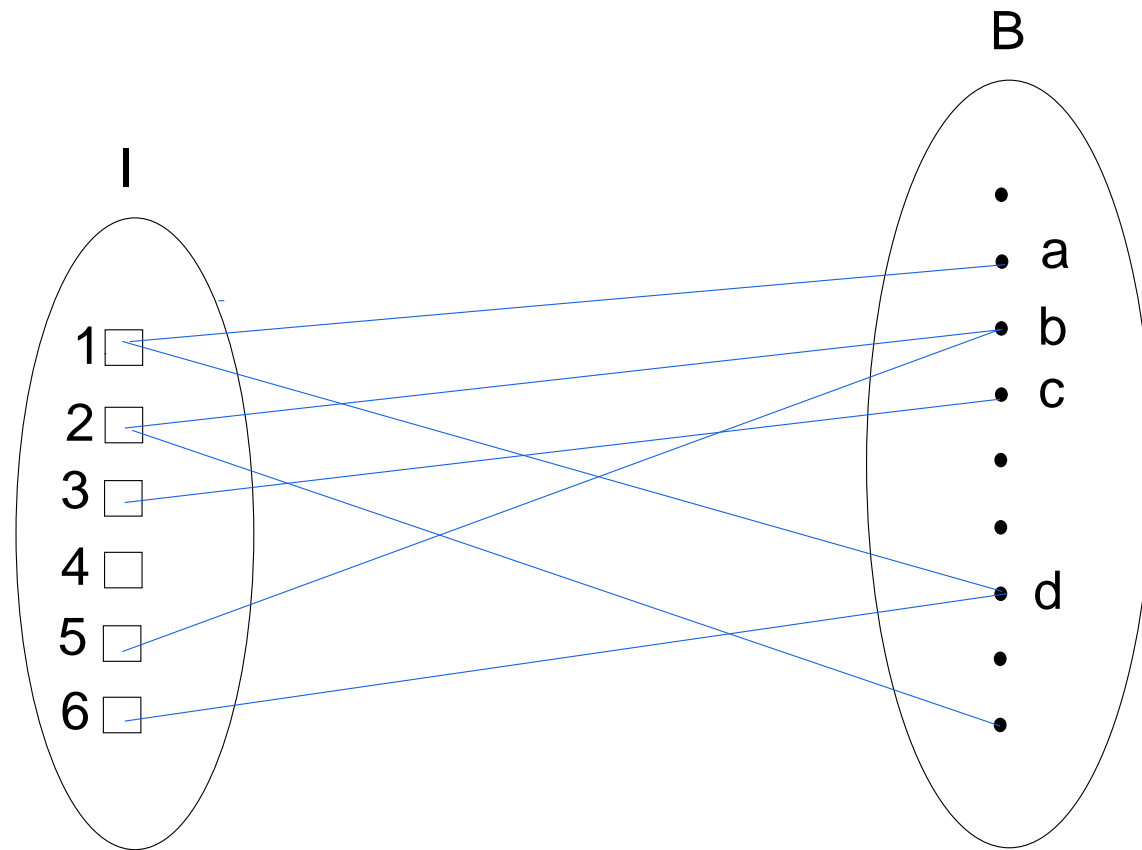


FIG. 2 – *Un choix pour B IDIMIN*

1.3 Premiers résultats d'importance

Lemme de distinction

Lemme 1.1 *Considérons un graphe G biparti sans jumeaux $G = (I \cup A, E)$ et B un code identifiant de cardinal minimal. On a alors :*

$$\forall b \in B, \exists i \text{ et } j \in I \text{ t.q. } V(i) \cap B = (V(j) \cap B) \cup \{b\}$$

Une des conséquences de ce lemme est la construction d'un graphe particulier, que l'on appellera Graphe de distinction associé à un codage B IDIMIN :

Définition 1.1 *Les sommets de ce graphe sont les éléments de I et pour chaque élément b de B , on crée l'arête reliant deux éléments i et j vérifiant $V(i) \cap B = (V(j) \cap B) \cup \{b\}$.*

Ce graphe possède $|B|$ arêtes et aura des propriétés très intéressantes dans la suite.

Un exemple de construction du graphe de distinction pour l'exemple précédent :

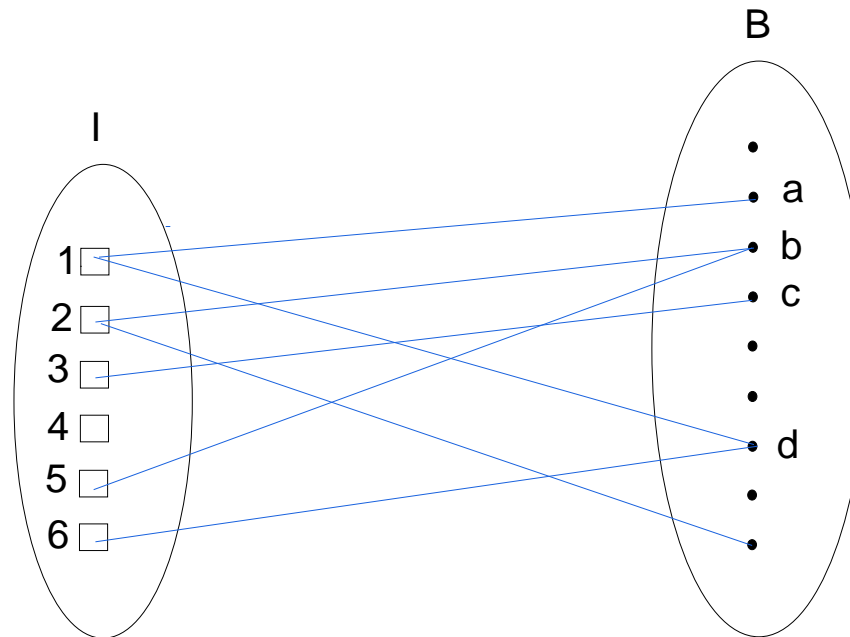


FIG. 3 – *Un choix pour B IDIMIN*

Sur cet exemple, on voit par exemple que l'attribut *a* permet de distinguer exactement les individus 1 et 6, l'attribut *b* les individus 5 et 4, etc...

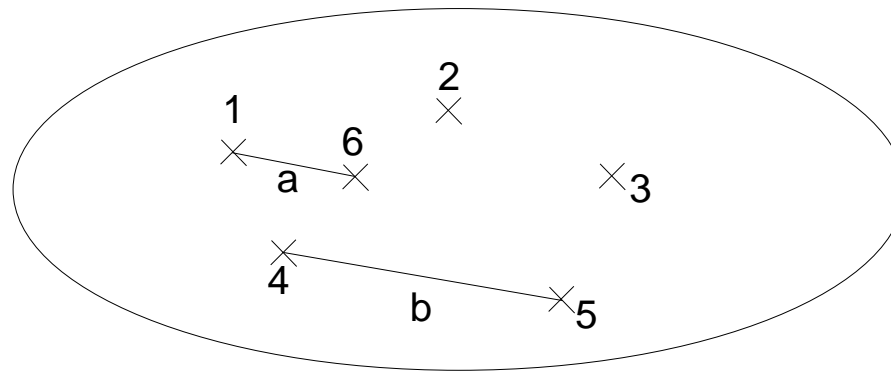


FIG. 4 – *Une partie du graphe de distinction pour l'exemple précédent*

1.3.1 Résultat de complémentarité pour le graphe biparti

Considérons un graphe $G = (I \cup A, E)$ et construisons $G' = (I \cup A, E')$ où E' est le “complémentaire” de E , au sens où E' contient exactement toutes les arêtes de $I \times A$ qui ne sont pas dans E .

Sa matrice d'adjacence est obtenue en remplaçant tous les 0 de la matrice d'adjacence de G par des 1 et tous les 1 par des 0.

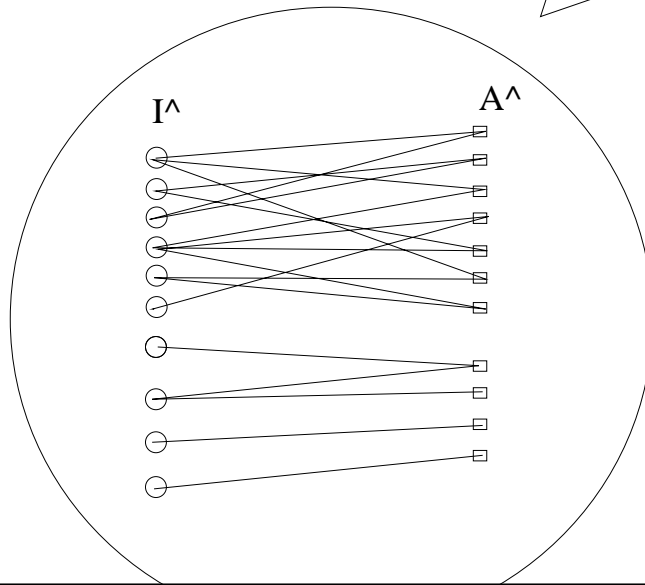
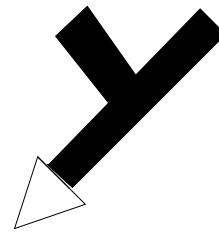
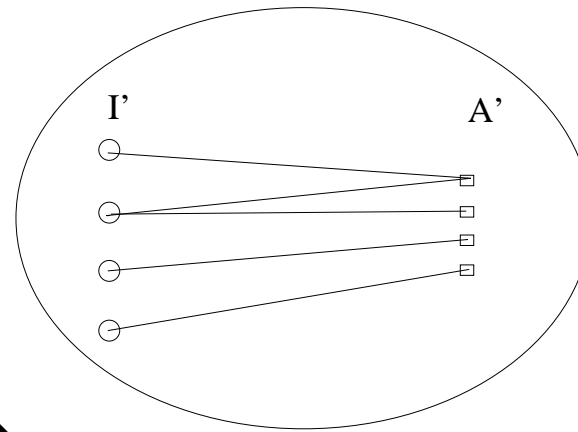
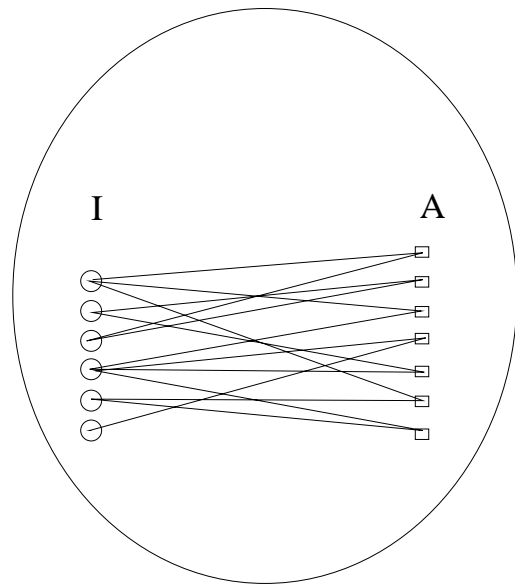
Propriété 1.1 *Nous avons alors le résultat suivant : Si B est IDIMIN pour G , alors il l'est aussi pour G' .*

1.3.2 Concaténation de graphes bipartis

Si l'on considère $G = (I \cup A, E)$ et $G' = (I' \cup A', E')$, nous définissons le nouveau graphe biparti $H = (\hat{I} \cup \hat{A}, \hat{E})$ avec $\hat{I} = I \cup I'$, $\hat{A} = A \cup A'$ et $\hat{E} = E \cup E'$.

Alors si B est IDIMIN1 pour G , B' IDIMIN1 pour G' , alors $B \cup B'$ est IDIMIN1 pour H . Nous imposons uniquement la condition qu'aucun $V(i)$ ne soit réduit à \emptyset pour éviter d'avoir en même temps l'élément sans attributs de I et l'élément sans attributs de I' , qui seraient donc jumeaux.

Un exemple de concaténation de graphes est représenté sur la figure 5



$$B \cup B' = B^{\wedge}$$

2 Valeurs extrémales de B dans le cas général

Considérons dans toute cette partie un graphe biparti G sans jumeaux.

Nous allons donner ici un certain nombre d'inégalités sur le cardinal de B IDIMIN pour G .

2.1 Etude de la borne inférieure de B

La valeur de la borne inférieure de B est donnée par la propriété suivante :

Théorème 2.1 *La valeur minimale que peut prendre B dans le cas général est $\log_2(|I|)$. Cette valeur est atteinte dans le cas du graphe $G = (I \cup B, E)$ où chaque élément i de I est vu comme un élément de $[1, |I|]$ et chaque b de B comme un bit, i étant relié à b si et seulement si le bit b apparaît dans l'écriture binaire de i .*

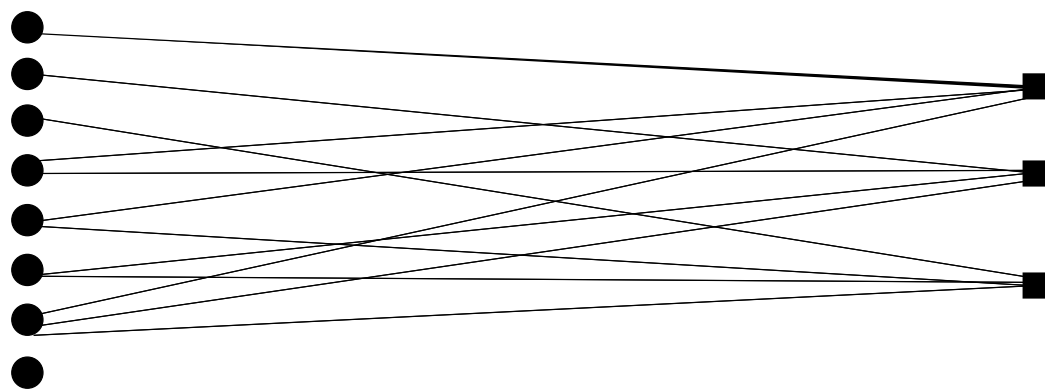


FIG. 6 – *Cas où $\log_2(|I|) = |B| = 3$*

On peut améliorer cette borne de manière tout à fait naturelle, en considérant le caractère entier de $|B|$:

Théorème 2.2 *L'inégalité du théorème 2.1 peut naturellement être passée à la partie entière par excès. On obtient donc le théorème suivant :*

$$|B| \geq \lceil \log_2(|I|) \rceil$$

De plus, cette borne peut être atteinte par un ensemble d'attributs A pour toute valeur de $|I|$.

2.2 Etude de la borne supérieure de B

Nous allons dans cette partie nous intéresser à la valeur maximale que peut prendre le cardinal de B IDIMIN pour G , pour I de cardinal fixe. Le résultat est donné par le théorème suivant :

Théorème 2.3 *Soit G un graphe biparti sans jumeaux et B IDIMIN pour G . Alors $|B| \leq |I| - 1$. De plus cette valeur est la meilleure borne supérieure pour B puisqu'elle est atteinte par l'exemple du couplage parfait.*

1. Démonstration par partitionnement de I

Pour $a \in A$, on note $X_a = \{i \in I \text{ t.q. } \{i, a\} \in E\} (=V(a))$ et $X'_a = I \setminus X_a$.

On définit $N(a_1, a_2, \dots, a_p)$ comme l'ensemble des couples de $I \times I$ que $\{a_1, a_2, \dots, a_p\}$ ne parvient pas à distinguer, i.e., tous les couples de jumeaux pour $\{a_1\} \cup \{a_2\} \cup \dots \cup \{a_p\}$.

Alors $N(a_1, a_2, \dots, a_p)$ est une union de différents $D \times D$, où l'ensemble de tous les D forme une partition de I .

De plus, si B est IDIMIN pour G , alors cela signifie que si l'on considère successivement $N(\emptyset)$, $N(b_1)$, $N(b_1, b_2)$, \dots , $N(b_1, b_2, \dots, b_{|B|})$, à chaque élément de B rajouté, la partition est strictement affinée, grâce à la propriété identifiante minimale de B .

Or, $N(B) = \{i_1\} \times \{i_1\} \cup \{i_2\} \times \{i_2\} \cup \dots \cup \{i_N\} \times \{i_N\}$, qui correspond à la partition maximale, réalisée avec N ensembles de cardinal 1. Donc :

$$|B| \leq N - 1$$

□

Preuve que $N(a_1, a_2, \dots, a_p)$ est une union de différents $D \times D$, où l'ensemble de tous les D forme une partition de I .

On a de manière évidente

$$N(a) = (X_a \times X_a) \cup (X'_a \times X'_a)$$

Et :

$$N(a_1, a_2, \dots, a_p, a_{p+1}) = (A \times A) \cup (B \times B)$$

$$\text{où : } A = N(a_1, a_2, \dots, a_p) \cap X_{a_{p+1}}$$

$$B = N(a_1, a_2, \dots, a_p) \cap \overline{X_{a_{p+1}}}$$

Et le résultat est démontré par récurrence.

2. Démonstration par le graphe de distinction

Nous utilisons maintenant le graphe de distinction défini à la partie 1.1

Supposons que ce graphe contienne un cycle : Il existerait donc un chemin reliant i_1, i_2, \dots, i_p avec $i_p = i_1$. Ainsi, $\exists(b_{i_1}, b_{i_2}, \dots, b_{i_{p-1}})$ tq :

$$(V(i_1) \cap B) \Delta (V(i_2) \cap B) = \{b_{i_1}\}$$

$$(V(i_2) \cap B) \Delta (V(i_3) \cap B) = \{b_{i_2}\}$$

\vdots

$$(V(i_{p-1}) \cap B) \Delta (V(i_1) \cap B) = \{b_{i_{p-1}}\}$$

En combinant une à une toutes ces équations, nous obtiendrions

$$(V(i_1) \cap B) \Delta (V(i_1) \cap B) = \{b_{i_1}\} \cup \{b_{i_2}\} \cup \dots \cup \{b_{i_{p-1}}\},$$

On en déduit que le “graphe de distinction” que nous avons construit est sans cycles. Il contient $|I|$ sommets, il a donc moins de $|I| - 1$ arêtes, et $|B| \leq |I| - 1$. \square

D'autres démonstrations possibles, par récurrence, et en utilisant le rang de la matrice d'adjacence.

De plus cette borne sup est la meilleure, car atteinte dans le cas du graphe de la figure 9, appelé couplage parfait.

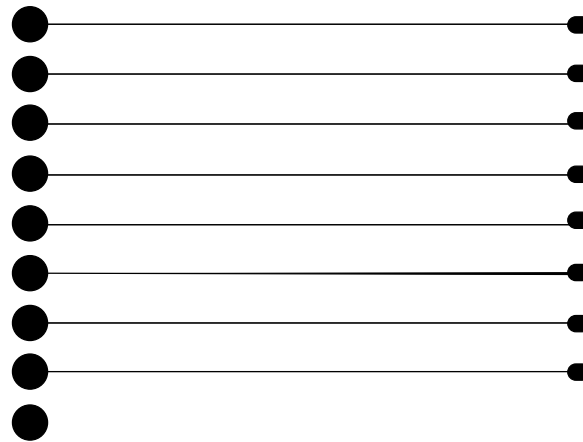


FIG. 7 – Cas où $|I| - 1 = |B| = 8$

2.3 Autres bornes intéressantes pour le cardinal de B

2.3.1 Borne du “covering”

Si B est un sous-ensemble de A vérifiant : $\forall (i \neq j) \in I^2, V(i) \cap B \neq V(j) \cap B$ alors B , sera *à fortiori* “couvrant” pour I , à l’exception éventuellement d’un élément, i.e. :

$$\forall i \in I, V(i) \cap B \neq \emptyset$$

Si $deg(b)$ représente le nombre d’arêtes partant de b , on en déduit

$$\sum_{b \in B} deg(b) \geq |I|.$$

Et si l’on note $\eta = \max_{a \in A} deg(a)$, on obtient

$$|B| \geq \frac{|I|}{\eta}$$

2.3.2 Borne liée au rang de la matrice d'incidence de G

Soit \check{M} la matrice extraite de la matrice d'incidence M de G , en ne conservant que $|B|$ lignes sur les $|A|$ qu'elle possédait au départ, ces lignes correspondant à toutes les lignes représentant le voisinage des éléments $b \in B$.

On montre que la matrice \check{M} est de rang plein et on a donc $rg(\check{M}) = |B|$.

Or clairement puisque \check{M} est extraite de la matrice d'incidence totale, nous avons $rg(\check{M}) \leq rg(M)$.

On en déduit

$$|B| \leq rg(M)$$

3 Cas des graphes réguliers

3.1 Préambule

Soit G un graphe biparti $G = (I \cup A, E)$. On dit que G est régulier à droite si et seulement si le degré de chaque élément de A est constant :

$$\exists \alpha \ \forall a \in A, \text{Card}(V(a)) = \alpha.$$

α sera appelé le degré de régularité à droite de G .

Un exemple de graphe régulier à droite est représenté sur la figure 8

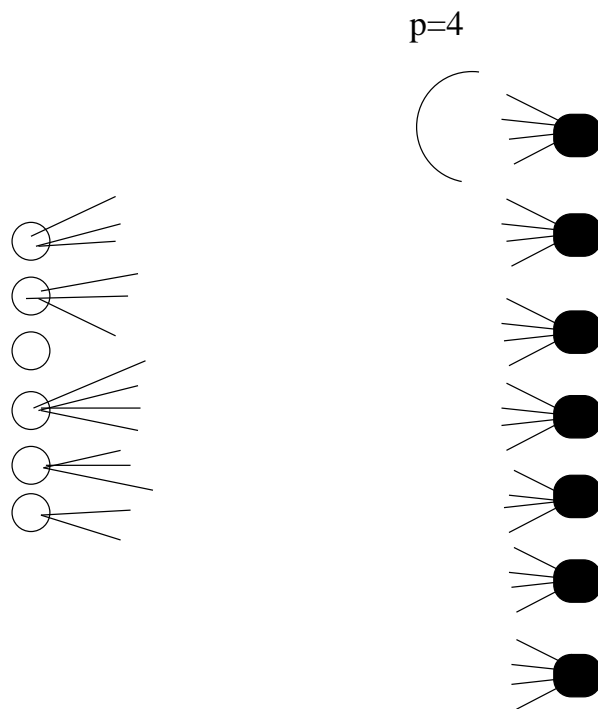


FIG. 8 – *Exemple de graphe biparti régulier à droite de degré 4*

Remarques utiles

Une première remarque sur la propriété de régularité à droite s'impose : si $G = (I \cup A, E)$ est régulier à droite, alors, $\forall B \subseteq A$, $G' = (I \cup B, E)$ est encore régulier à droite. En particulier, ceci est vrai pour B de cardinal minimal vérifiant $\forall (i, j) \in I, i \neq j, V(i) \cap B \neq V(j) \cap B$.

De plus, le complémentaire d'un graphe biparti régulier à droite de degré α , est encore un graphe biparti régulier à droite et de degré $|I| - \alpha$.

3.2 Nouvelles bornes remarquables pour B

3.2.1 Amélioration de la borne du covering

On reprend ici le début du raisonnement établi dans la partie 2.3.1.

Ici le degré de chaque élément $b \in B$ vaut α et on obtient $|B| \times \alpha \geq |I|$, que l'on peut écrire sous la forme suivante :

$$|B| \geq \frac{|I|}{\alpha}$$

Cette borne peut être améliorée en prolongeant encore le raisonnement de couverture des éléments de I .

Si l'on réalise un double comptage des arêtes, à gauche il y a $|B|$ éléments de $|I|$ au maximum de degré 1 et $|I| - |B|$ de degré supérieur à 2.

On obtient donc maintenant :

$$|B| \times \alpha \geq |B| + (|I| - |B|) \times 2$$

Donc, cela nous donne après simplification :

$$|B| \geq \frac{2}{\alpha + 1} \times |I| \quad (1)$$

En passant à la partie entière supérieure, et au complémentaire, on obtient finalement :

$$|B| \geq \max\left(\lceil \left(\frac{2}{\alpha + 1} \times |I|\right) \rceil, \lceil \left(\frac{2}{|I| - \alpha + 1} \times |I|\right) \rceil\right)$$

3.3 Tentatives d'améliorations de la borne sup et conjectures démolies

On remarque tout d'abord que les deux exemples de graphe bipartis permettant d'atteindre les deux bornes de B , i.e, $\log_2 |I|$ et $|I| - 1$ étaient deux exemples de graphes réguliers à droite respectivement de degré $\frac{|I|}{2}$ et 1.

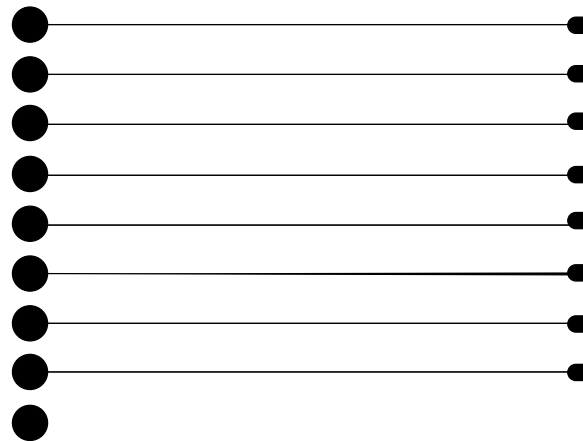
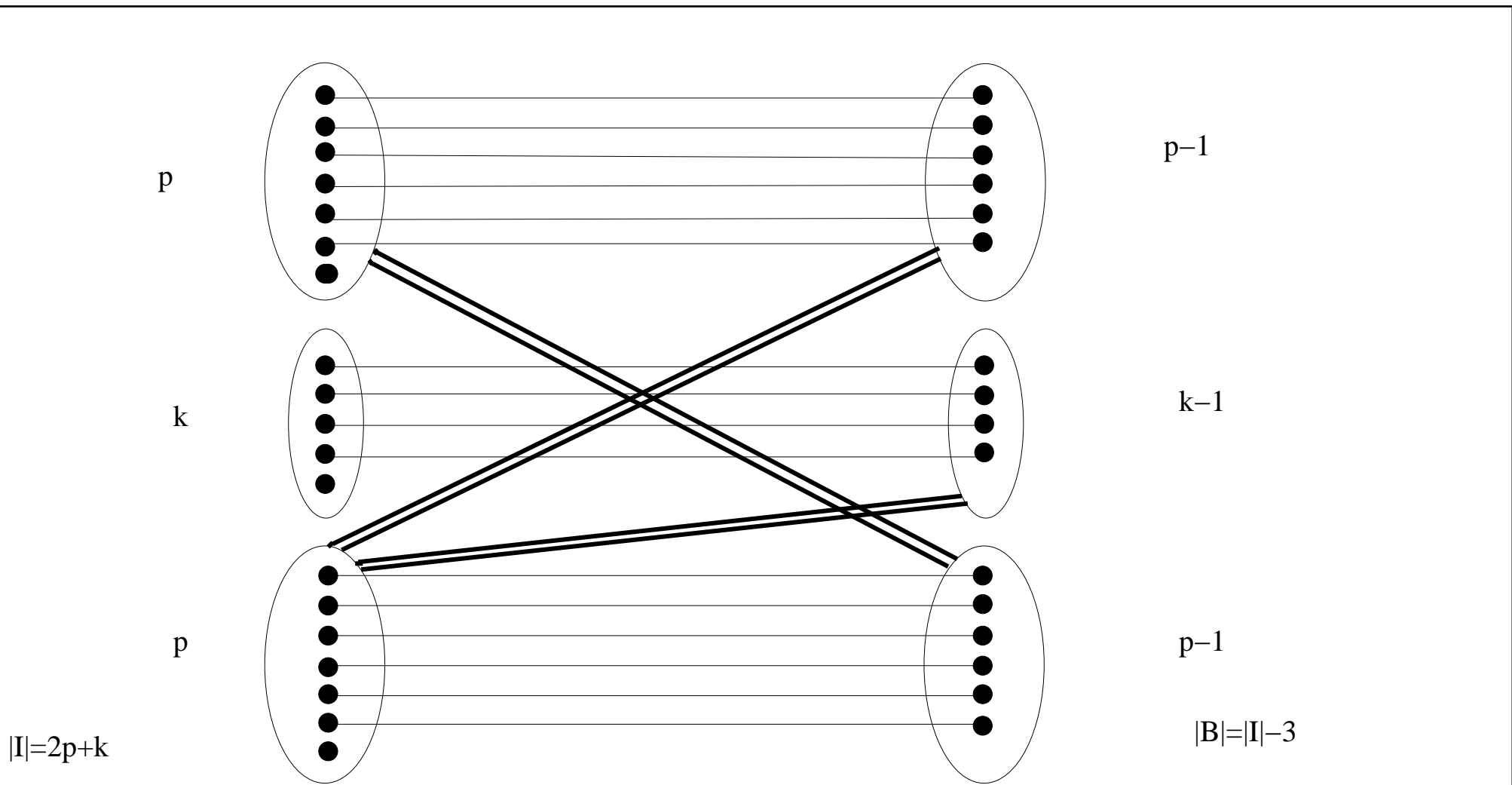


FIG. 9 – Cas où $|I| - 1 = |B| = 8$

Le cas particulier des graphes bipartis réguliers à droite semblait pouvoir permettre d'affiner les résultats quant à la recherche de notre sous-ensemble minimal identifiant. En l'occurrence l'idée était de chercher à considérer que le cas du couplage parfait était particulier, car de degré 1, et que dans le cas de degrés plus grands, la borne supérieure de B pouvait diminuer.

L'exemple de la figure suivante contredit définitivement toute conjecture de la forme $\frac{|I|}{f(\alpha)}$ ou $\max(N-\alpha, \alpha)$:



== signifie que chaque individu au départ est relié à tous les individus à l'arrivée.

FIG. 10 – Graphe biparti régulier à droite de degré quelconque vérifiant $|B| = |I| - 3$

Dans ce graphe, nous avons $|I| = 2 \times p + k$, $|B| = 2 \times p + k - 3$, et $\alpha = p + 1$.
On peut donc avoir un exemple de cas graphes bipartis réguliers de degré $p + 1$
avec $|B| = |I| - 3 = 2 \times p + k - 3$; on peut donc construire pour toutes les
valeurs de α un exemple aussi grand que voulu qui contredit notre conjecture.
Si on pensait pouvoir beaucoup améliorer le cas le pire ($|B| = |I| - 1$), c'est
raté...

3.4 Résolution du cas $|B| > |I| - 3$

Rappelons que d'après le théorème 2.3 de la partie 2.2, nous savons que forcément $|B| \leq |I| - 1$. Nous pouvons améliorer dès à présent cette borne :

Théorème 3.1 *Dans le cas d'un graphe biparti sans jumeaux et régulier à droite de degré $\neq 1$ et $\neq (|I| - 1)$, si B est IDIMIN pour ce graphe, on a*

$$|B| \leq |I| - 2$$

Démonstration Nous considérons B un code IDIMIN pour G régulier à droite de degré α , ainsi que son graphe de distinction.

Si l'on avait $|B| = |I| - 1$, ce graphe serait forcément de connexité égale à 1, et tous les sommets de ce graphe pourraient être reliés entre eux par un chemin. De plus, si l'on choisit une feuille i_1 de ce graphe sans cycles, il existera i_2 et b_1 tel que $V(i_1)\Delta V(i_2) = b_1$.

Observons ceci sur la figure suivante :

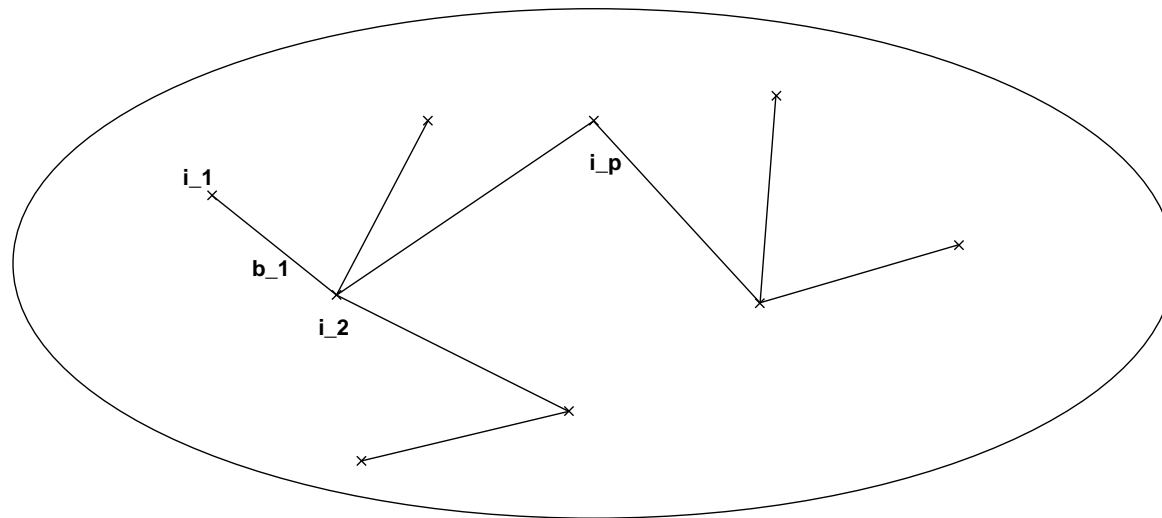


FIG. 11 – *Graphe de distinction lorsque $|B| = |I| - 1$*

Deux cas se présentent :

- Si $b_1 \in V(i_1)$, alors il y a forcément un autre élément $i_p \in I$ relié à b_1 (car le degré de b_1 est ≥ 2) et il existe un chemin reliant i_1 à i_p . On en déduit que $V(i_1) \Delta V(i_p) = b_1 \cup b_2 \cup \dots \cup b_p$ ce qui est contradictoire car $b_1 \in V(i_1) \cap V(i_p)$.

- Si $b_1 \notin V(i_1)$ alors il y a forcément un autre élément $i_p \in I$ tel que $b_1 \notin V(i_p)$ (car le degré de b_1 est $\leq |I| - 2$) et il existe un chemin reliant i_1 à i_p et ce chemin contiendra forcément l'arête b_1 puisque i_1 est une feuille. On en déduit que $V(i_1) \Delta V(i_p) = b_1 \cup b_2 \cup \dots \cup b_p$ ce qui est contradictoire.

Propriété On en déduit que le seul cas de $|B| > |I| - 3$ possible pour un graphe biparti sans jumeaux régulier à droite (de degré de régularité différent de 1) est $|B| = |I| - 2$.

Cherchons maintenant à exhiber tous les cas de graphes bipartis réguliers à droite aboutissant à $|B| = |I| - 2$

Énonçons tout d'abord un lemme utile pour la suite de l'étude :

Lemme 3.1 *Si l'on a une arête a reliant les sommets i et j dans une composante connexe du graphe de distinction, alors elle sépare en deux parties cette composante connexe : de part et d'autre de cette arête, tous les éléments sont dans le voisinage de a "d'un côté", et aucun n'y est de l'autre.*

De plus, pour les autres composantes connexes du graphe de distinction, tous les sommets d'une même composante connexe, seront dans le même état d'appartenance ou non au voisinage de a .

Observons ce lemme sur la figure suivante :

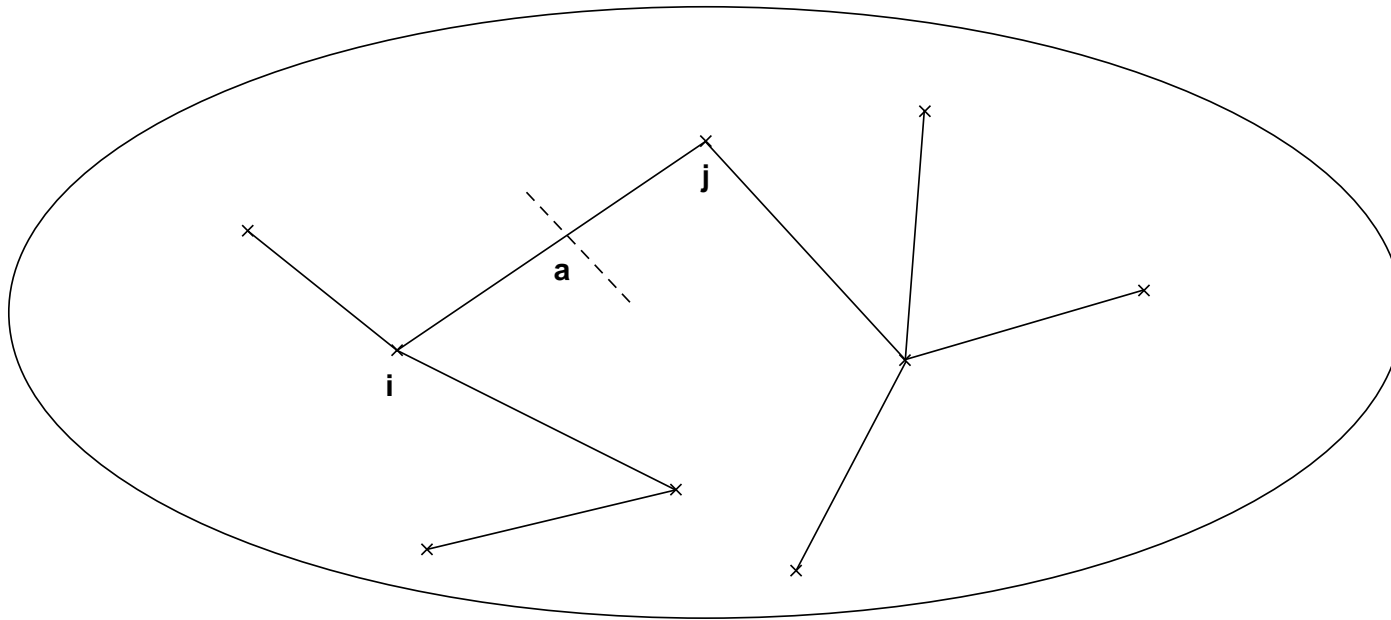


FIG. 12 - i et j vérifient $V(i) - V(j) = a$

Etude du cas $|B| = |I| - 2$:

Concentrons nous d'abord sur le cas $\alpha \geq 3$

Soit $G = (I \cup B, E)$ un graphe biparti sans jumeaux et régulier à droite de degré $\notin \{1, |I| - 1\}$, et B un codage IDIMIN. Considérons alors le graphe de distinction habituel. Il est nécessairement composé de deux composantes connexes. Soient S et T chacune de ces deux composantes. Soit i une feuille de S et a l'arête reliant i à j . On a donc, on le rappelle,

$$V(i)\Delta V(j) = a$$

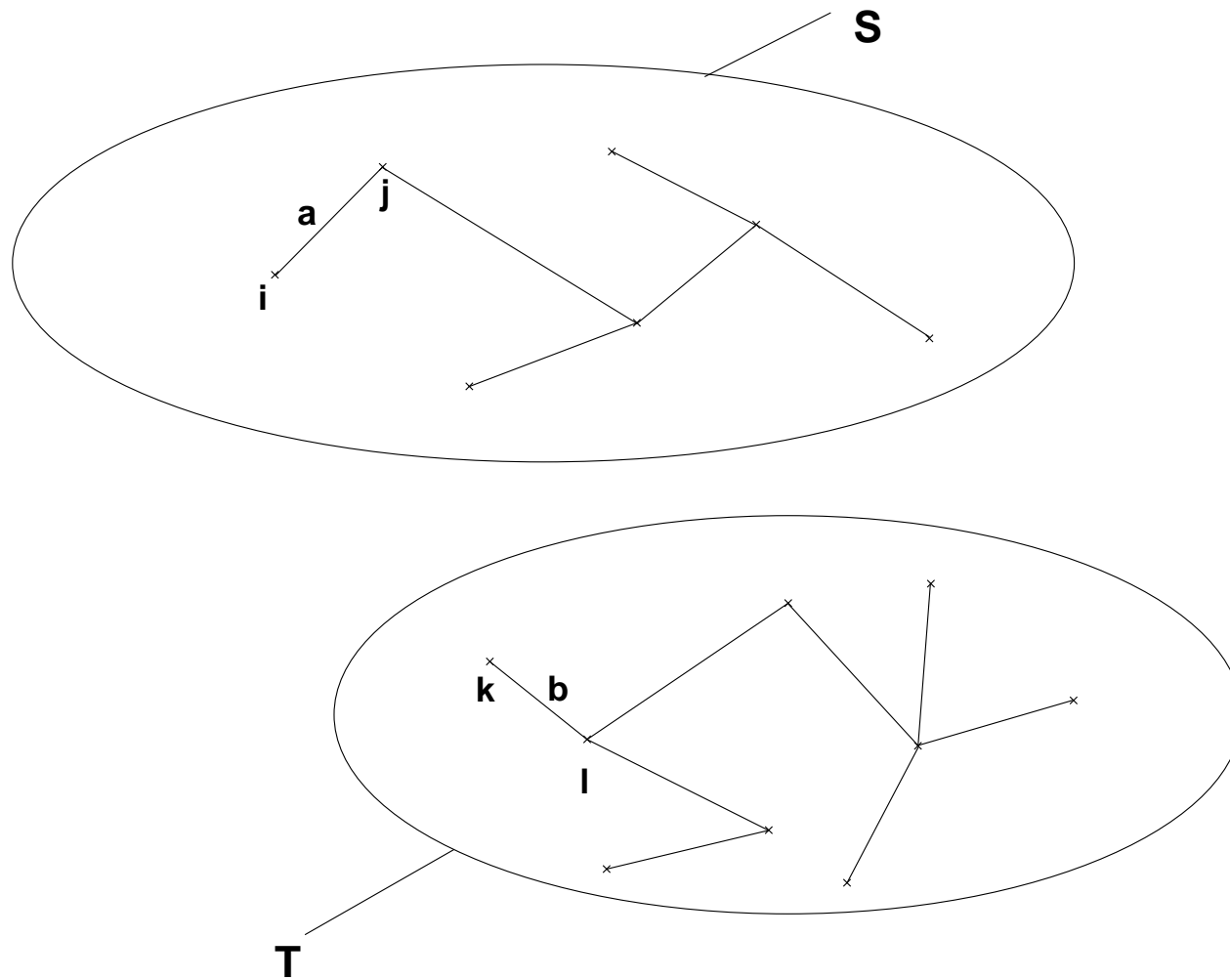


FIG. 13 – *graphe de distinction avec les deux composantes connexes S et T*

Deux cas se présentent alors :

- si $a \in V(i)$, on obtient après différentes étapes un graphe de distinction qui a l'allure suivante :

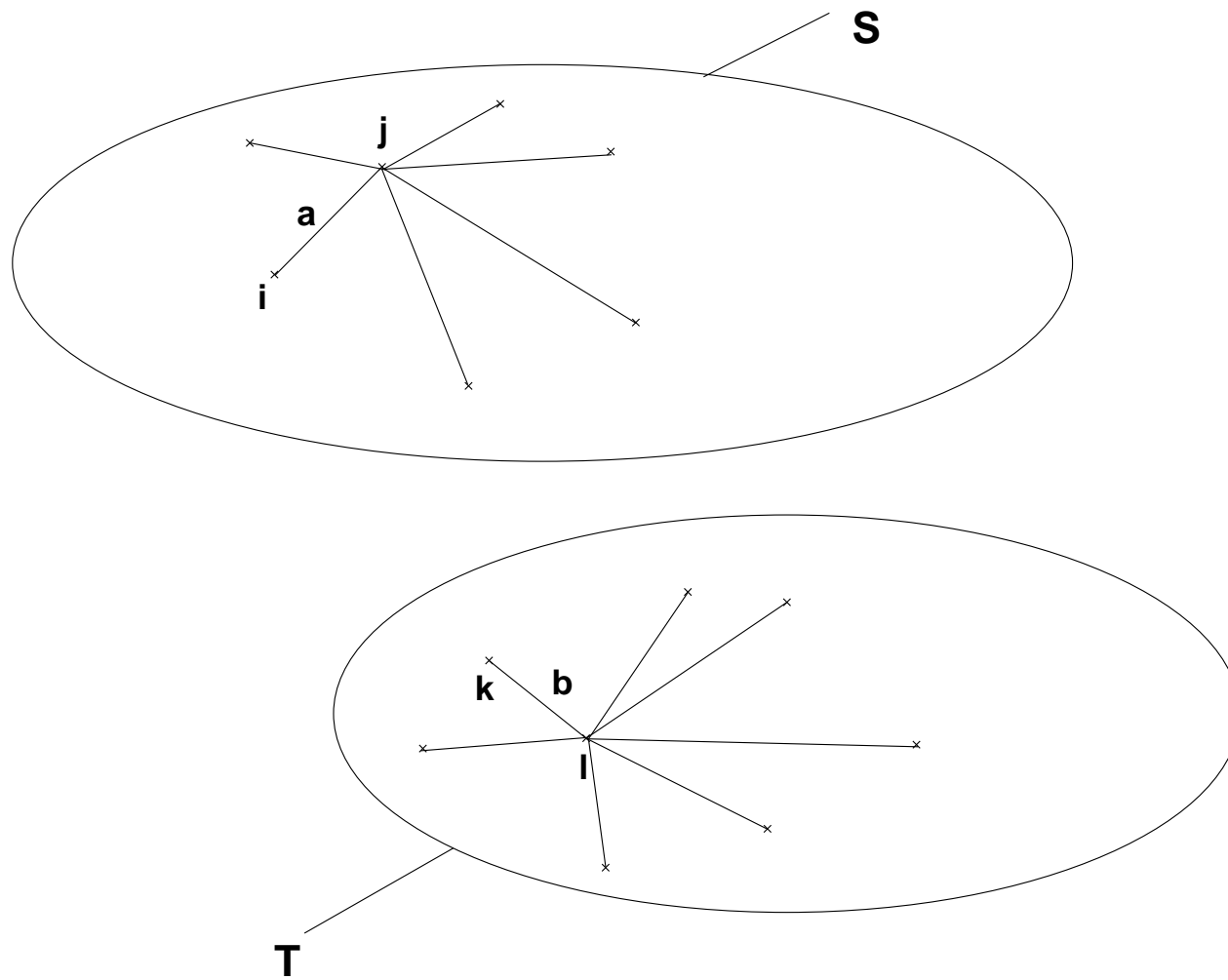
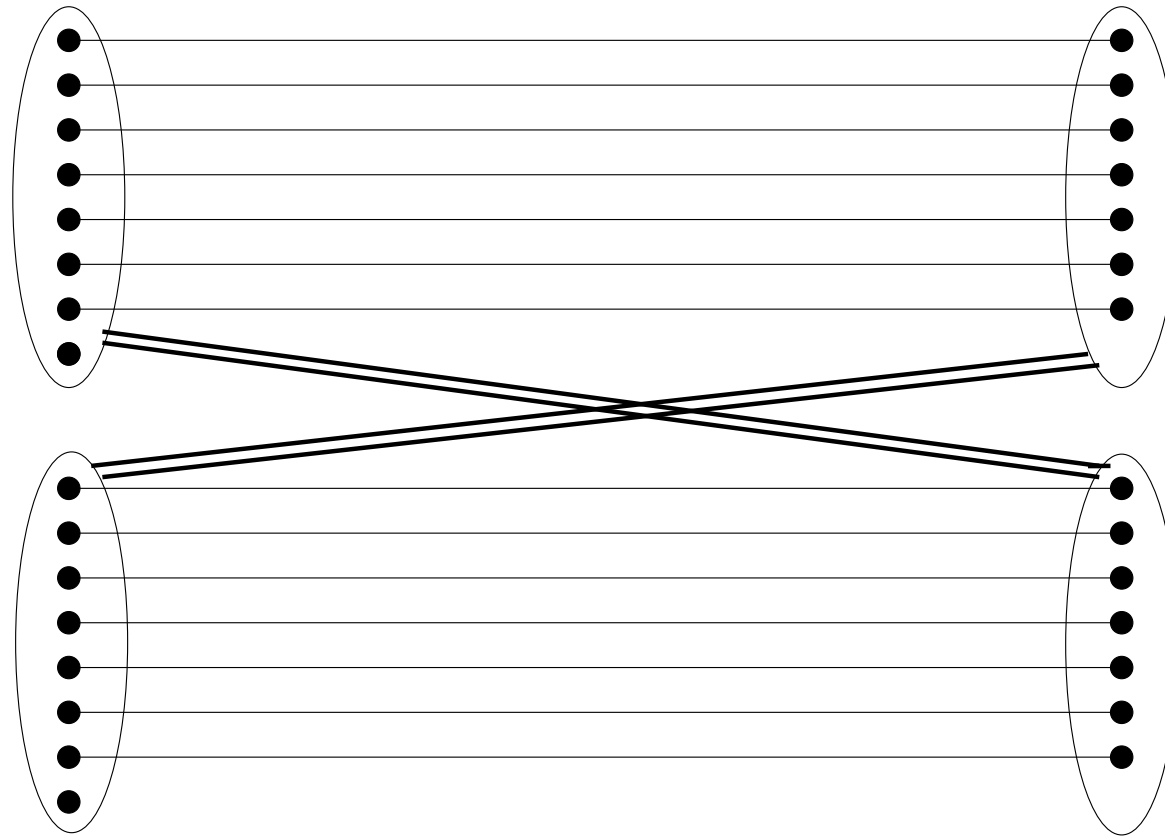


FIG. 14 – *graphe de distinction final avec S et T en étoile*

On remonte maintenant de ce graphe de distinction au graphe de départ et on en déduit que l'on sera alors forcément dans le cas de la figure 15.



=====
Signifiant que chaque individu au départ est relié à tous les attributs à l'arrivée

FIG. 15 – *Graphe biparti régulier à droite de degré supérieur à 2 aboutissant à $|B| = |I| - 2$*

- Si $a \notin V(i)$ alors nous aurions abouti au complémentaire du graphe ci-dessus.

Et si $\alpha = 2$? On considère $|B| > 1$ (le cas $|B| = 1$ ne peut aboutir à $|I| = 3$).

Dans ce cas, en reprenant le raisonnement précédent, une des deux composantes connexes a bien deux éléments ou plus, et l'on considère une feuille de S par exemple.

Soit a l'arête qui en part et j l'autre extrémité : $V(i)\Delta V(j) = a$.

Deux cas se présentent de nouveau :

- Si $a \in V(i)$, on aboutit pour le graphe de distinction au graphe de la figure 16.

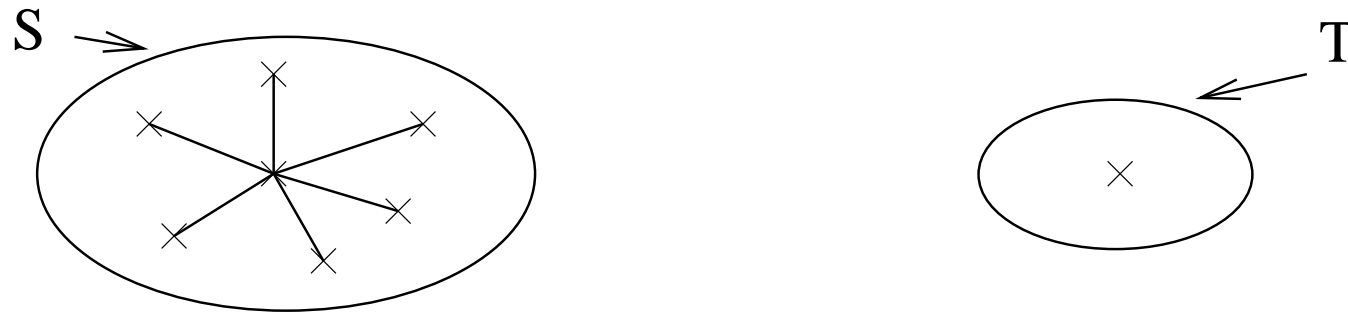


FIG. 16 – Graphe biparti régulier à droite de degré 2 aboutissant à $|B| = |I| - 2$

On remonte au graphe biparti du départ qui est donc inévitablement de l'allure du graphe de la fig suivante :

degré 2

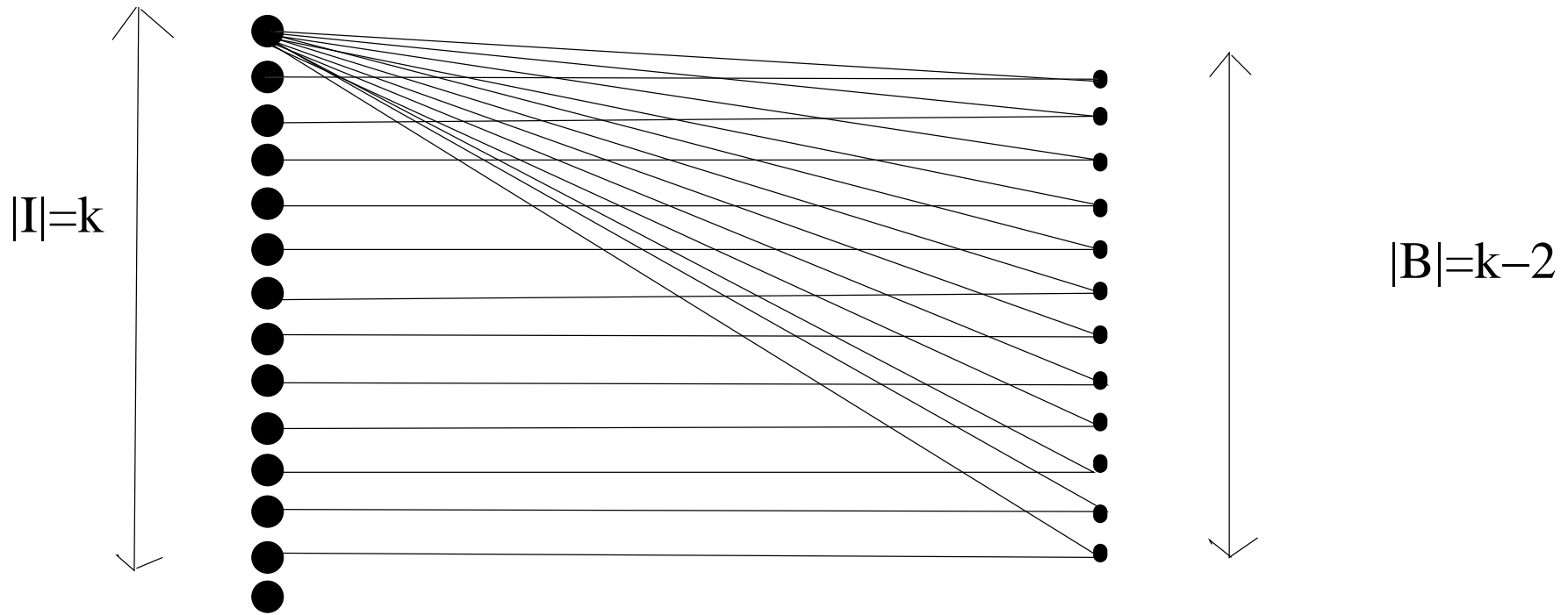


FIG. 17 – *Graphe biparti régulier à droite vérifiant $|B| = |I| - 2$*

- Si $a \notin V(i)$, on aboutit à un graphe de distinction qui a l'allure de la figure 18.

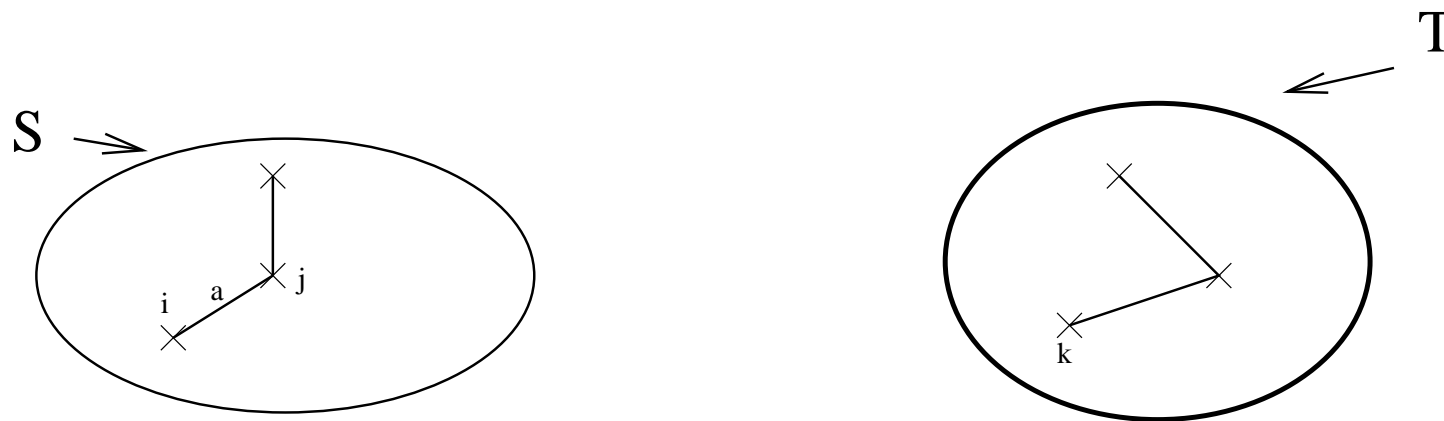


FIG. 18 – *Graphe de distinction pour un graphe biparti régulier à droite de degré 2 et vérifiant $|B| = |I| - 2$*

Le graphe de départ avait donc la forme du graphe de la figure 19.

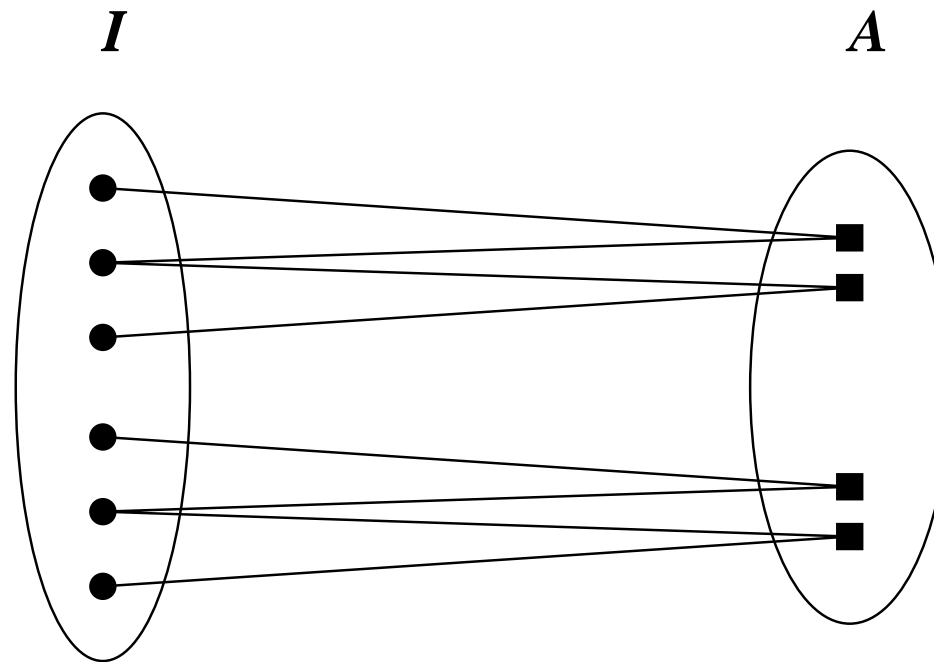


FIG. 19 – *Graphe biparti régulier à droite de degré 2 aboutissant à $|B| = |I| - 2$*

3.5 Dans tous les autres cas : $|I| - 3$

On en déduit donc qu'à l'exception du cas de degré 1 (ou $|I| - 1$), et des graphes décrits ci-dessus, nous avons $|B| \leq |I| - 3$.

Cette amélioration liée au graphe régulier à droite est la meilleure puisque l'on a vu que l'on pouvait obtenir un graphe biparti régulier à droite de degré quelconque, et vérifiant $|B| = |I| - 3$.

4 Cas des arbres

4.1 Quelques rappels terminologiques sur les arbres

Un arbre est un graphe connexe sans cycles.

La hauteur d'un arbre, est la longueur du plus long chemin de la racine à une feuille.

Un arbre p -aire est un graphe où chaque sommet possède au maximum p fils.

Un arbre p -aire parfait est un arbre p -aire dont toutes les feuilles sont à égale distance de la racine.

Un arbre est aussi un graphe biparti, en considérant l'ensemble des sommets de l'arbre à distance paire de la racine d'une part, et ceux à distance impaire d'autre part.

Le graphe binaire parfait de profondeur 4 est représenté ci-dessous sur le schéma 20 .

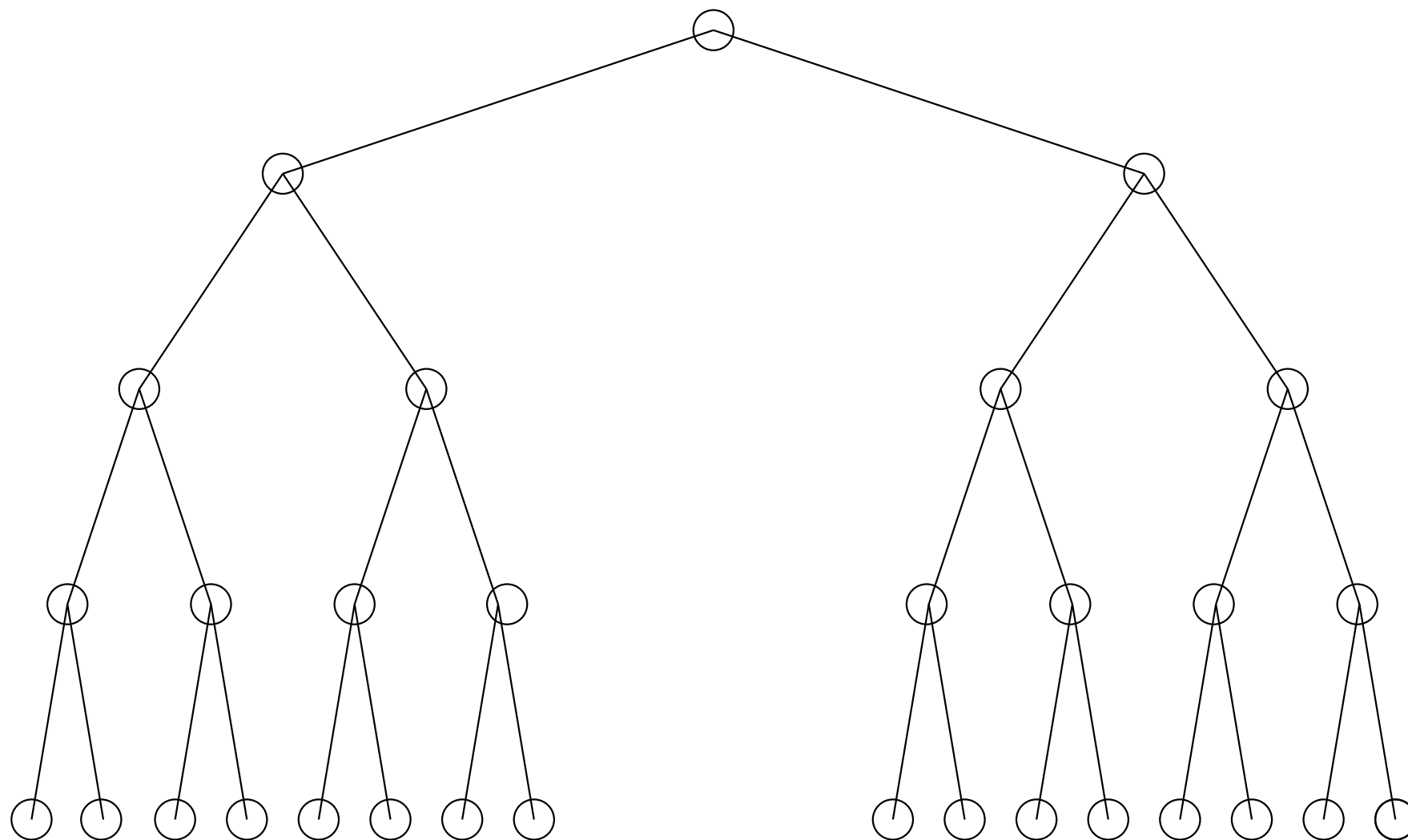


FIG. 20 – *Arbre Parfait de hauteur 4*

4.2 Cas de l'arbre binaire parfait

Cardinal de I et de A

En considérant cet arbre comme un graphe biparti, nous cherchons $B \subseteq A$ de cardinal minimal permettant d'identifier tous les éléments de I .

La première remarque d'importance est que la dernière ligne est forcé d'être une ligne d'attributs puisque sinon deux éléments de I seraient jumeaux et la recherche de B serait close.

Un exemple d'arbre binaire parfait vu comme notre graphe biparti est décrit sur la figure 21

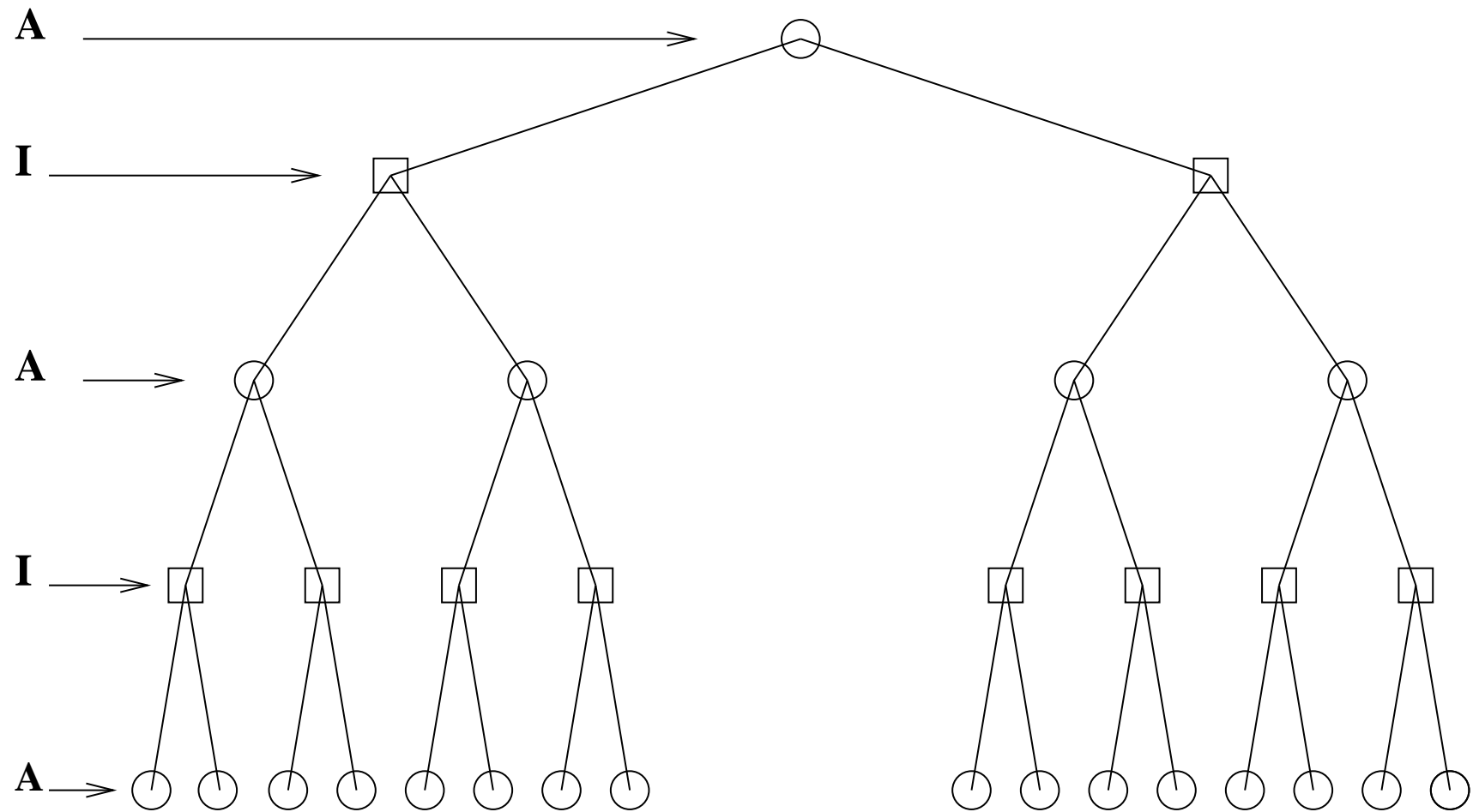


FIG. 21 – *Arbre parfait*

On calcule aisément par récurrence les cardinaux de I et A :

Si l est impair : on trouve $|I| = \frac{2^{(l+1)}-1}{3}$ et $|A| = 2 \times \frac{2^{(l+1)}-1}{3}$

Si l est pair : on trouve $|I| = 2 \times \frac{2^l-1}{3}$ et $|A| = \frac{2^{(l+2)}-1}{3}$

Recherche de B optimal Notons que l'on s'impose la condition suivante $\forall i \in I, V(i) \neq \emptyset$, qui, on le sait, ne modifie pas le résultat de plus d'un en cardinalité pour B .

Un premier affinement se fait en considérant la dernière ligne, pour laquelle clairement une moitié d'attributs est déjà inutile puisque, pris deux par deux dans l'ordre, les attributs ont exactement le même voisin de I . En particulier, les trois dernières lignes sont finalement représentés sur la figure 22 .

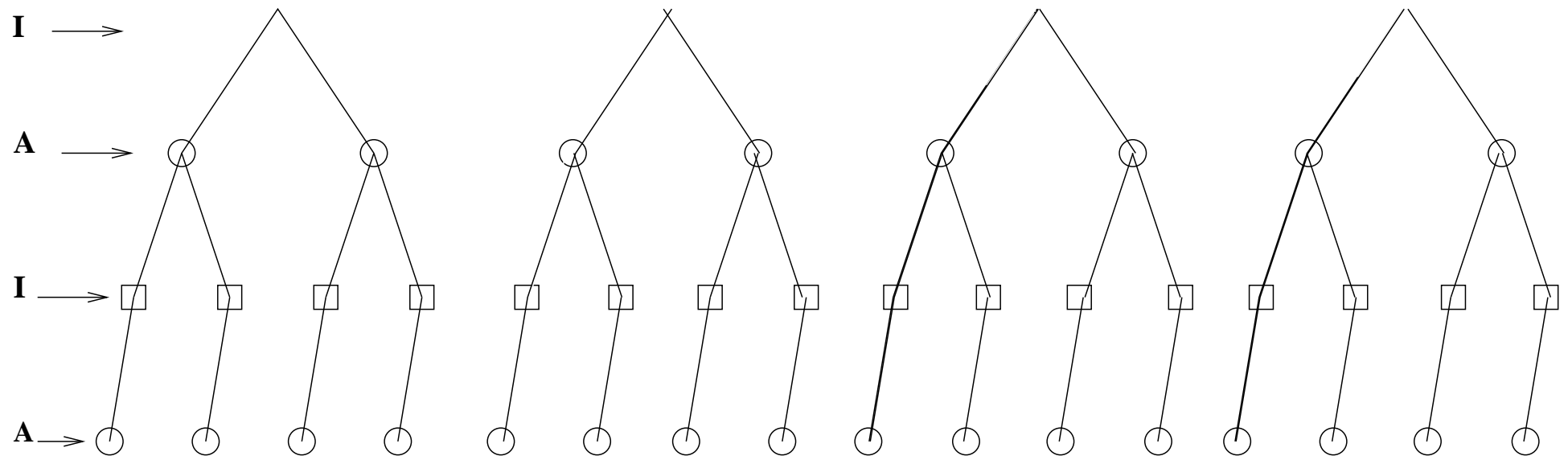


FIG. 22 – *Trois dernières ligne de l'arbre avec A'*

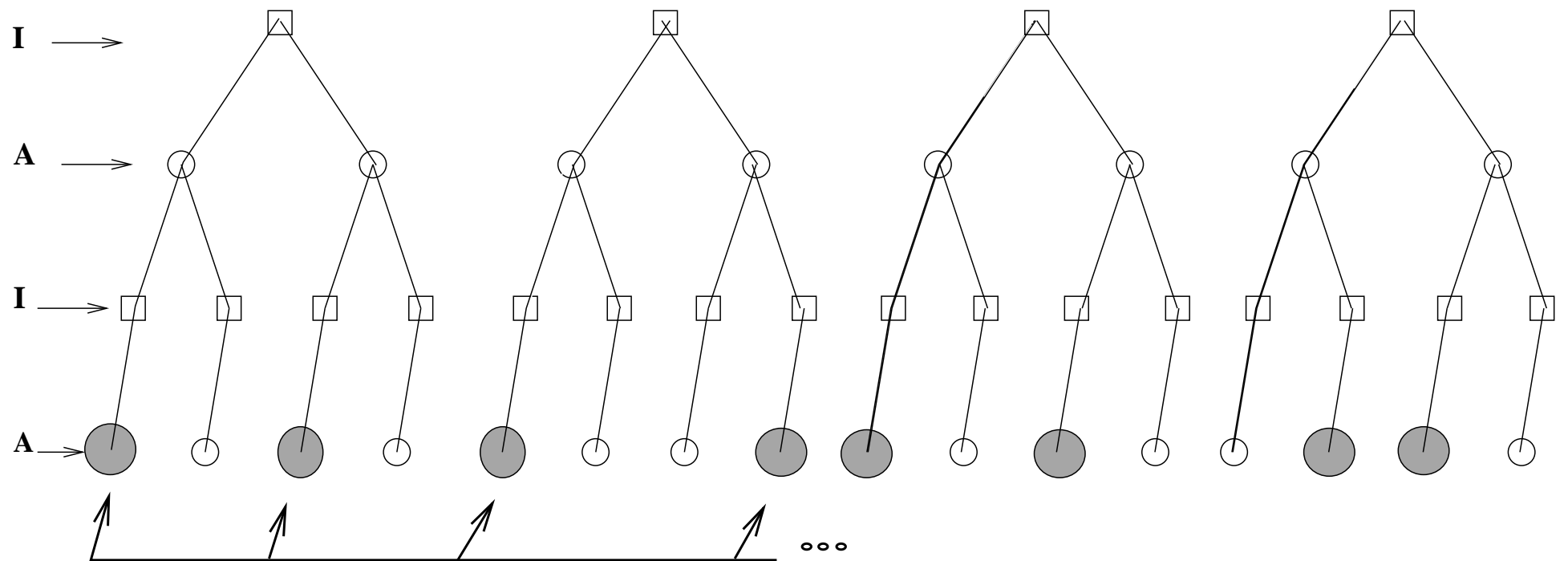
On appellera maintenant A' dans la suite le nouvel ensemble d'attributs dans lequel nous avons supprimé la moitié inutile des attributs de la dernière ligne.

Clairement :

Si l est impair : on trouve $|A'| = 2 \times \frac{2^{(l+1)}-1}{3} - 2^{l-1}$

Si l est pair : on trouve $|A'| = \frac{2^{(l+2)}-1}{3} - 2^{l-1}$

Recherche de B à proprement dite Si l'on veut réussir à identifier tous les individus de l'avant dernière ligne, en particulier si l'on veut réussir à différencier chacun des individus deux par deux dans l'ordre, nous sommes forcés de prendre déjà dans B un individu sur deux de A' , comme représenté sur la figure 23 :



Nous sommes forcés de les choisir pour différencier leur père et leur oncle.

FIG. 23 – Dernière ligne d'individus

La moitié des individus a donc un attribut de la dernière ligne dans son voisinage, mais une moitié des individus de l'avant dernière ligne a pour l'instant un voisinage vide. Pour chacun de ces individus là, il faut leur associer un voisin dans B , et il n'y a que deux possibilités, le père ou le fils.

Le choix du père pour chacun de ces individus est de toute évidence beaucoup plus judicieux, car il permet immédiatement l'identification de tous les individus de ligne d'individus immédiatement au-dessus, ainsi que l'on peut l'observer sur la figure 24.

Tous les éléments de I de cette figure sont identifiés.

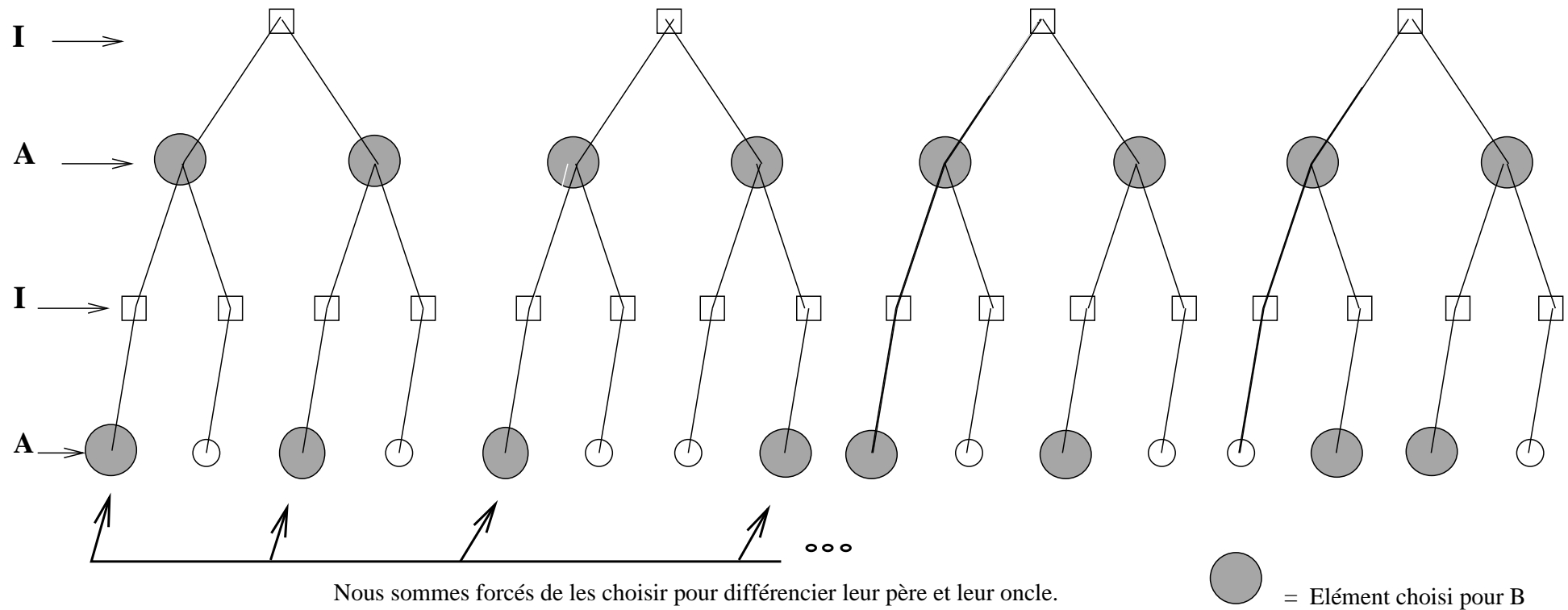


FIG. 24 – Deux dernières lignes d'individus

Il fallait forcément que les individus de la dernière ligne aient un voisinage non vide, et étant forcés de prendre un attribut sur deux de A' pour distinguer les couples, nous devons choisir soit le père soit le fils pour la moitié des individus de la dernière ligne, ce qui imposait au minimum un total de 2^{l-2} attributs, uniquement pour distinguer les individus de la dernière ligne.

Le choix de la figure ci-dessus pour ces 2^{l-2} attributs est donc forcément le meilleur puisqu'il permet de distinguer tous les individus de la ligne immédiatement au-dessus.

La situation à laquelle on aboutit est visible sur la figure 25.

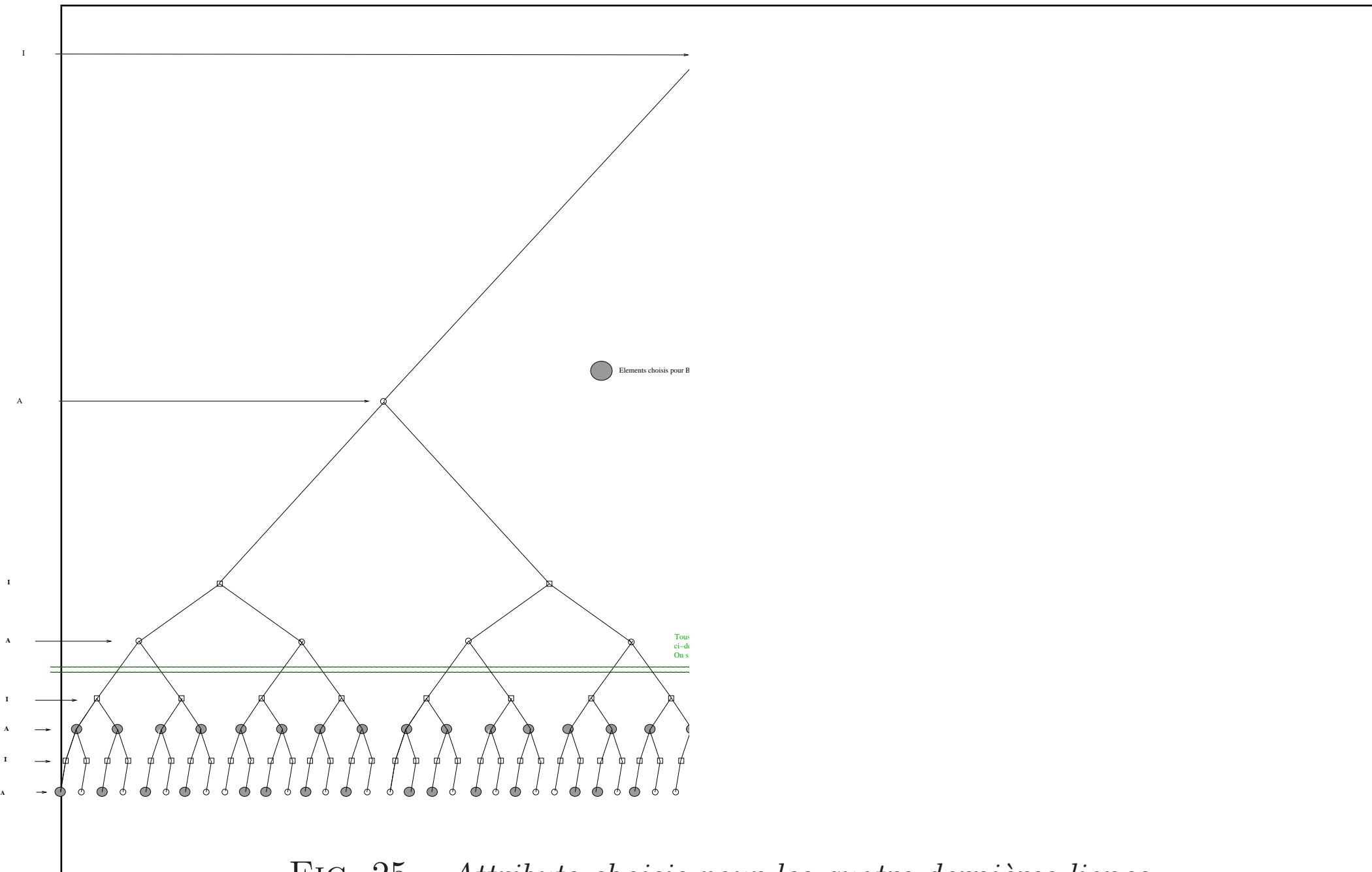
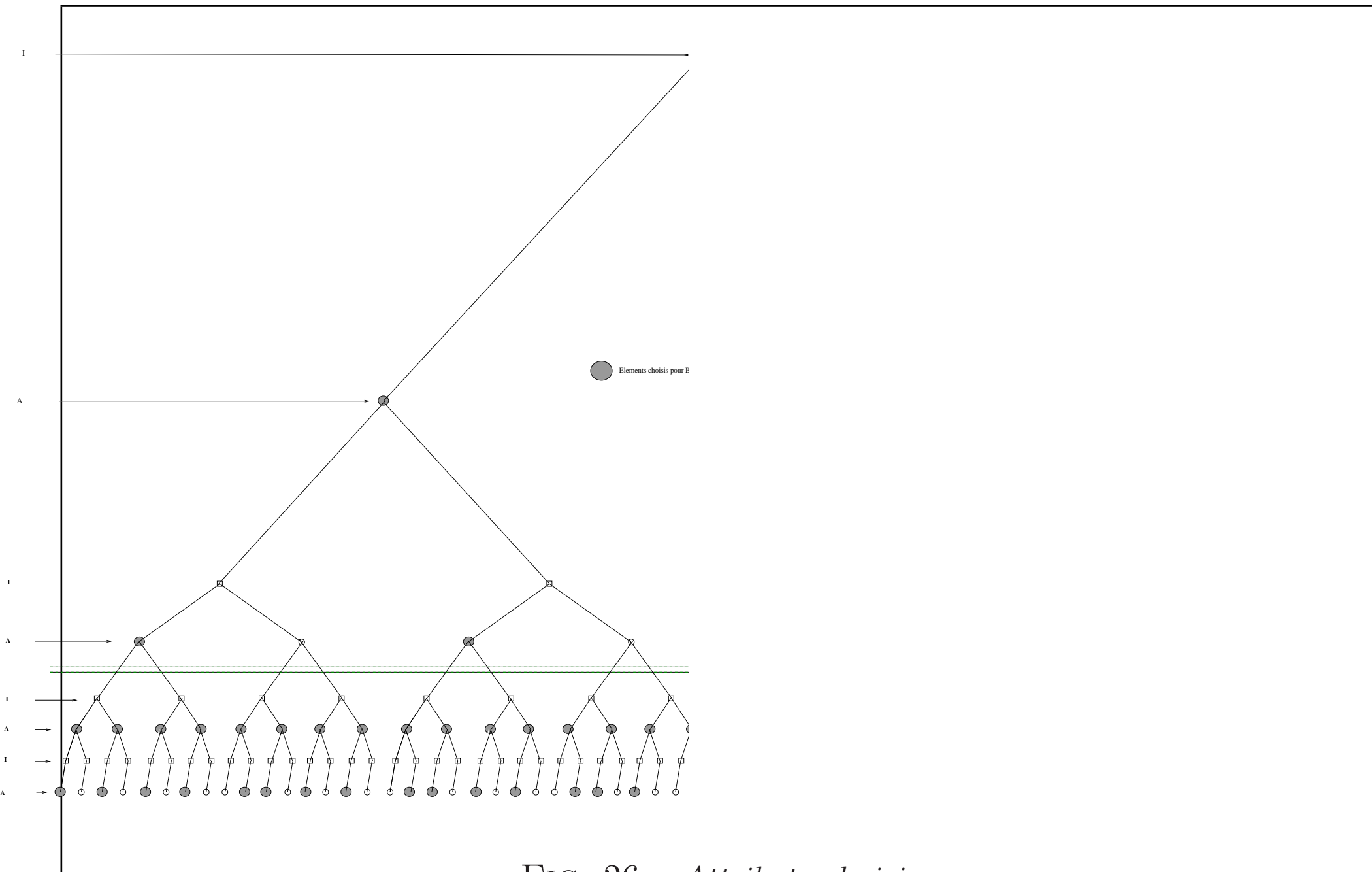


FIG. 25 — *Attributs choisis pour les quatre dernières lignes*
ENST – Recherche de codes attribuants dans un graphe biparti

On se ramène donc maintenant à la recherche de sous-ensemble d'attributs de cardinal minimal pour le nouvel arbre binaire parfait de profondeur $l - 4!!$

Par le même raisonnement, on aboutit obligatoirement au schéma 26, qui fait office d'algorithme :



ENST – Recherche de codes attribuants dans un graphe biparti

FIG. 26 – *Attributs choisis*

En répétant cette opération récursivement, on obtient bien sûr la valeur exacte du cardinal de B , et même un algorithme permettant de le construire, ainsi que l'ensemble B dans son détail exactement.

La valeur de $\text{card}(B)$ est clairement : $2^{l-1} + 2^{l-5} + 2^{l-9} + \dots 2^{l-1-4 \times k} + \dots$

Il faut distinguer quatre cas, selon la valeur de la hauteur de l'arbre modulo 4.

Ainsi :

– si $l \equiv 0 [4]$, nous obtenons par exemple

$|B| = 2^{l-1} + 2^{l-5} + 2^{l-9} + \dots 2^{l-1-4 \times k} + \dots 2^7 + 2^3$, ce qui donne après calcul

$$|B| = \frac{2^{l+3} - 8}{15}$$

– si $l \equiv 1 [4]$, nous obtenons après calcul

$$|B| = \frac{2^{l+3} - 1}{15}$$

– si $l \equiv 2 [4]$, nous obtenons après calcul

$$|B| = \frac{2^{l+3} - 2}{15}$$

– si $l \equiv 3 [4]$, nous obtenons après calcul

$$|B| = \frac{2^{l+3} - 4}{15}$$

Si l'on reprend maintenant les valeurs de I obtenues pour les quatre cas, on en déduit dans le cas de l'arbre binaire parfait :

– si $l \equiv 0 [4]$,

$$\frac{|B|}{|I|} = \frac{2^{l+3} - 8}{5 \times (2^{l+1} - 2)}$$

,

– si $l \equiv 1 [4]$,

$$\frac{|B|}{|I|} = \frac{2^{l+3} - 1}{5 \times (2^{l+1} - 2)}$$

– si $l \equiv 2 [4]$,

$$\frac{|B|}{|I|} = \frac{2^{l+3} - 2}{5 \times (2^{l+1} - 2)}$$

– si $l \equiv 3 [4]$,

$$\frac{|B|}{|I|} = \frac{2^{l+3} - 4}{5 \times (2^{l+1} - 2)}$$

On en déduit que lorsque la hauteur grandit, la valeur de $\frac{|B|}{|I|}$ tend vers $\frac{2^2}{5}$.

Théorème 4.1 *Si l'on considère un arbre binaire parfait et l sa profondeur, alors*

$$\lim_{l \rightarrow \infty} \frac{|B|}{|I|} = \frac{4}{5}$$

4.3 Arbre p-aire parfait

Nous avons résolu dans son intégralité la recherche de B IDIMIN dans le cas de l'arbre binaire parfait. Résolvons maintenant celui de l'arbre p-aire parfait. Les raisonnements du cas binaire sont valables pour le calcul de $|A|$ et $|I|$ et l'on obtient directement :

Si l est impair : on trouve $|I| = \frac{p^{l+1}-1}{(p-1)(p+1)}$ et $|A| = 2 \times \frac{p^{l+1}-1}{(p-1)(p+1)}$

Si l est pair : on trouve $|I| = 2 \times \frac{p^{l+1}-p}{(p-1)(p+1)}$ et $|A| = \frac{p^{l+2}-1}{(p-1)(p+1)}$

De même que dans le cas binaire, on va introduire A' , extrait de A , obtenu en ne gardant qu'un des p fils des individus de la dernière ligne. A' est représenté sur la figure 27 .

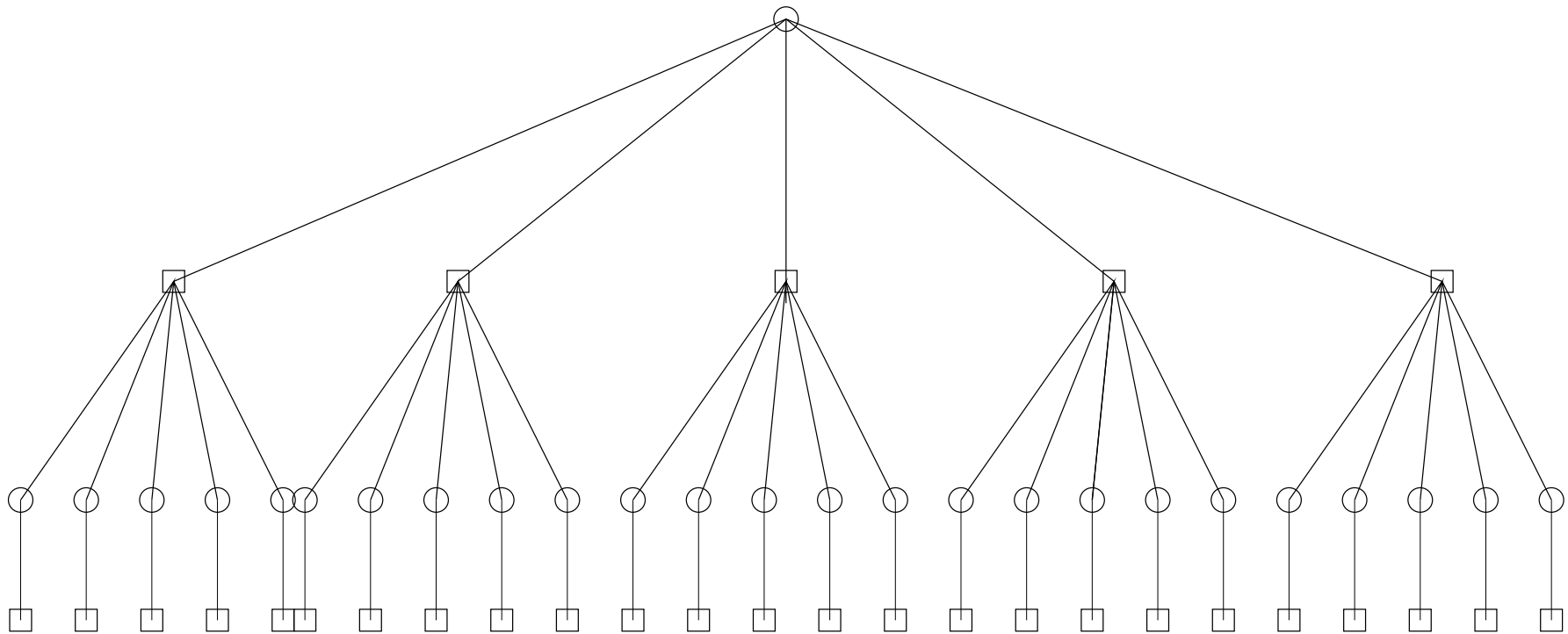


FIG. 27 – *Arbre p -aire avec A' à la dernière ligne*

Et l'on obtient :

$$\text{Si } l \text{ est impair} : |A'| = 2 \times \frac{p^{l+1} - 1}{(p-1) \times (p+1)} - p^{l-1}$$

$$\text{Si } l \text{ est pair} : |A'| = \frac{p^{l+2} - 1}{(p-1) \times (p+1)} - p^{l-1}$$

Le raisonnement permettant dans le cas binaire d'aboutir à l'algorithme de recherche de B , est adaptable au cas p -aire : si l'on considère la dernière ligne d'individus, et les familles de p frères de cette ligne, il va être obligatoire de choisir le fils de $p - 1$ d'entre eux. En ce qui concerne le dernier, qui ne doit pas avoir un voisinage vide, il reste la possibilité de prendre son fils, ou son père, et l'on choisit tous les pères car ils permettent également d'identifier directement tous les individus de la ligne immédiatement au dessus. Le raisonnement nous permet d'avoir identifié les deux dernières lignes d'individus avec le choix de p^{l-2} attributs de A' qui était le nombre minimal d'attributs nécessaires pour différencier uniquement la dernière ligne. En réappliquant ce raisonnement au nouvel arbre obtenu en écartant les quatre dernières lignes, on obtient par récurrence le résultat suivant :

$$p^{l-1} + p^{l-5} + p^{l-9} + \dots p^{l-1-4 \times k} + \dots$$

Et, en distinguant les quatre cas, selon la hauteur de l'arbre modulo 4.

– si $l \equiv 0 [4]$, nous obtenons par exemple

$|B| = p^{l-1} + p^{l-5} + p^{l-9} + \dots p^{l-1-4 \times k} + \dots p^7 + p^3$, ce qui donne après calcul

$$|B| = \frac{p^{l+3} - p^3}{p^4 - 1}$$

– si $l \equiv 1 [4]$,

$$|B| = \frac{p^{l+3} - 1}{p^4 - 1}$$

– si $l \equiv 2 [4]$,

$$|B| = \frac{p^{l+3} - p}{p^4 - 1}$$

– si $l \equiv 3 [4]$,

$$|B| = \frac{p^{l+3} - p^2}{p^4 - 1}$$

Si l'on reprend maintenant les valeurs de I obtenues pour les quatre cas, on en déduit dans le cas de l'arbre binaire parfait :

– si $l \equiv 0 [4]$,

$$\frac{|B|}{|I|} = \frac{p^{l+3} - p^3}{(p^2 + 1) \times (p^{l+1} - 1)}$$

,

– si $l \equiv 1 [4]$,

$$\frac{|B|}{|I|} = \frac{p^{l+3} - 1}{(p^2 + 1) \times (p^{l+1} - p)}$$

– si $l \equiv 2 [4]$,

$$\frac{|B|}{|I|} = \frac{p^{l+3} - p}{(p^2 + 1) \times (p^{l+1} - 1)}$$

– si $l \equiv 3 [4]$,

$$\frac{|B|}{|I|} = \frac{p^{l+3} - p^2}{(p^2 + 1) \times (p^{l+1} - p)}$$

Théorème 4.2 *Si l'on considère un arbre binaire parfait et l sa profondeur, alors*

$$\lim_{l \rightarrow \infty} \frac{|B|}{|I|} = \frac{p^2}{p^2 + 1}$$

4.4 Résultats sur le cas de l'arbre dans le cas général

Nous allons maintenant nous intéresser au cas d'un arbre quelconque, que l'on considère toujours comme un graphe biparti. On supposera donc la donnée d'un arbre enraciné quelconque et le choix de l'appartenance à A ou à I de la racine de l'arbre. Pour la recherche de B IDIMIN, il faut naturellement s'assurer que ces données nous donnent un graphe biparti sans jumeaux.

4.4.1 Bornes min et max prises par B

Nous allons maintenant déterminer quelles sont les bornes prises par $|B|$ en fonction du cardinal de I , de même que nous l'avons fait dans le cas des graphes bipartis quelconques ou des graphes bipartis réguliers à droite.

Théorème 4.3 *Dans le cas d'un arbre enraciné quelconque, vu comme un graphe biparti, le cardinal de B IDIMIN vérifie :*

$$\frac{I}{2} \leq |B| \leq |I|$$

et ces bornes sont les meilleurs possibles, puisqu'atteintes, et même atteintes pour un arbre de profondeur ou de largeur aussi grandes que souhaitées.

Démonstration La borne de droite était déjà connue. On voit facilement qu'elle est atteinte par l'arbre du schéma 28, que l'on peut rendre aussi large qu'on veut ou aussi profond que voulu.

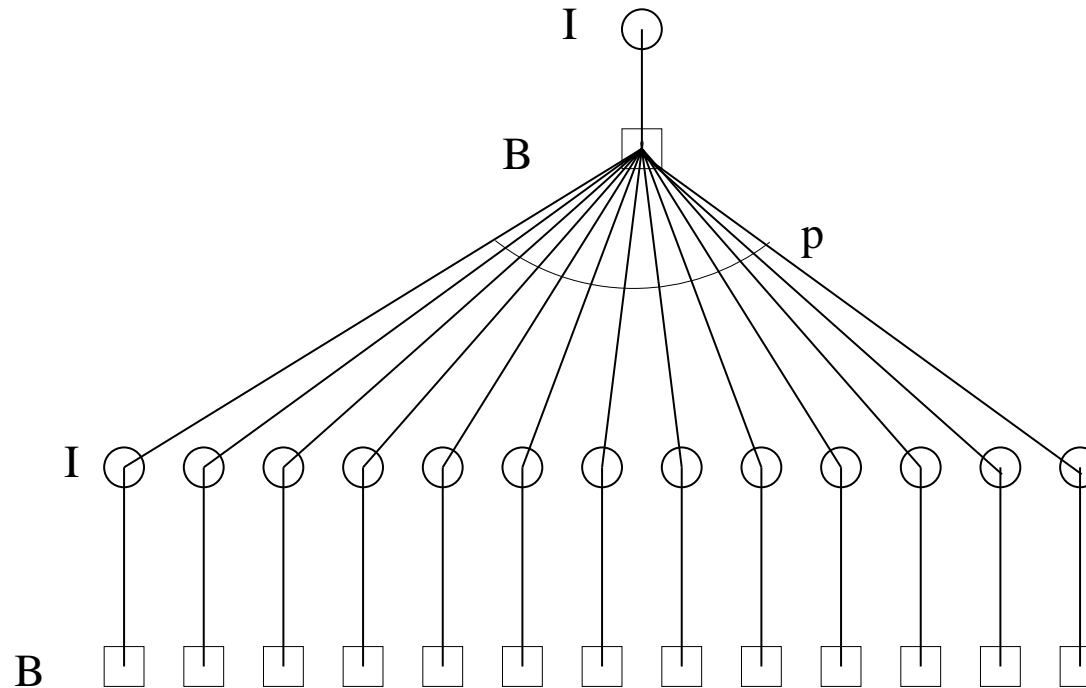


FIG. 28 - Cas d'un arbre avec $|B| = |I| = p + 1$

Intéressons nous donc maintenant à la borne de gauche (qui devient linéaire, alors qu'elle était logarithmique dans le cas général).

Considérons une ligne quelconque d'individus de l'arbre. Cette ligne est composée de plusieurs groupes de frères. Prenons l'un quelconque de ces groupes de frères, disons de cardinal p . Alors il faudra nécessairement choisir p éléments distincts de A qui leur sont voisins, pour les identifier tous.

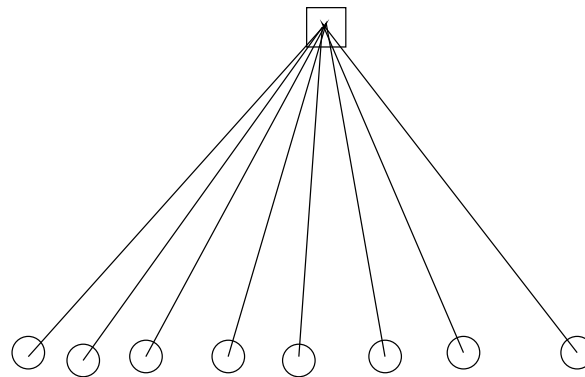


FIG. 29 – *Famille d'individus ou grappe*

A partir de là, si notre ligne contient N individus, alors B contiendra au moins N éléments répartis sur la ligne d'individus immédiatement supérieure et la ligne d'attributs immédiatement inférieure.

Ce qui nous donne, si l'on considère la ligne L_k :

$$|(\cup_{i \in L_k} V(i)) \cap B| \geq |L_k|$$

et donc : $\sum_k |(\cup_{i \in L_k} V(i)) \cap B| \geq \sum_k |L_k| = |I|$

Le comptage de gauche est tel qu'on ne compte les éléments de B qu'au maximum deux fois (un fois par la ligne d'individus supérieure, une fois par la ligne d'individus inférieure), et l'on en déduit

$$2 \times |B| \geq \sum_k \cup_{i \in L_k} |V(i) \cap B| , \text{ donc :}$$

$$|B| \geq \lceil \frac{|I|}{2} \rceil$$

De plus, cette borne est la meilleure comme nous le montre l'exemple de la figure 30, où l'on a $|I| = 2k + 1$ et $|B| = k + 1$

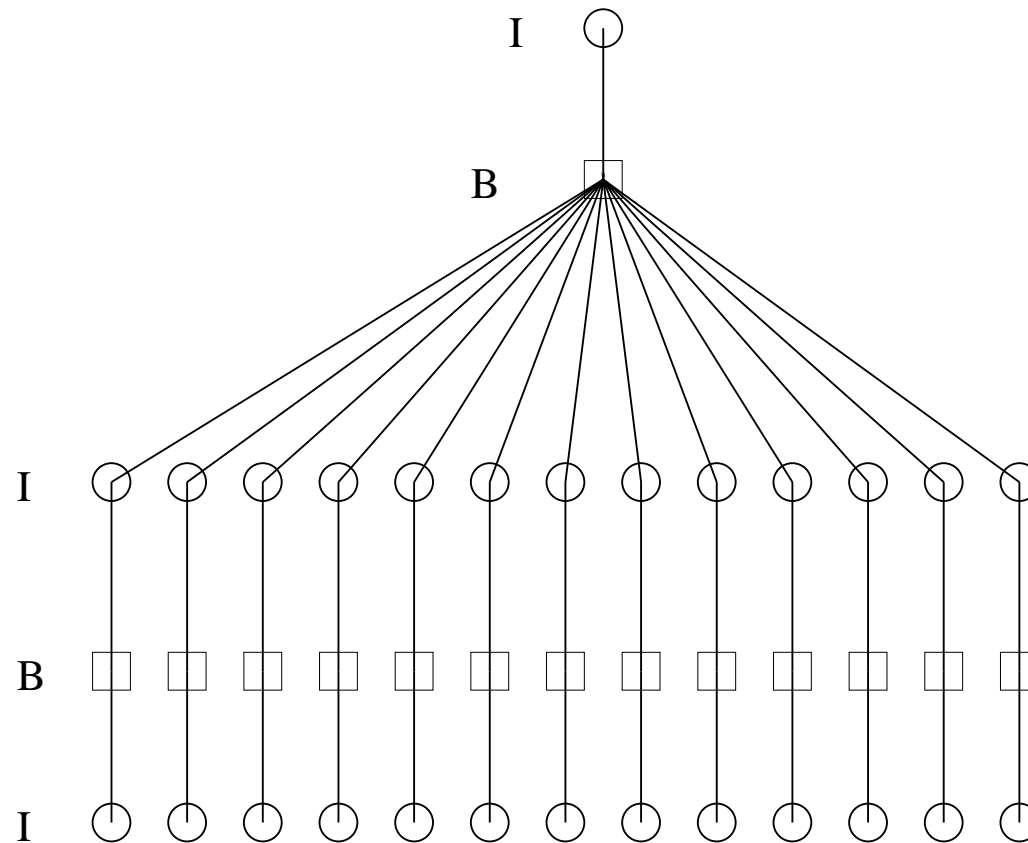


FIG. 30 – Cas où $|B|$ est minimal pour l'arbre quelconque

5 Complexité logique du problème

Nous allons montrer maintenant que la recherche d'un sous-ensemble B IDIMIN est un problème NP-difficile.

Théorème 5.1 :

Le problème suivant est NP-complet :

NOM : BID

ENTREE : Un graphe biparti $G = (I \cup A, E)$ sans jumeaux et un entier k .

QUESTION : G admet-il $B \subseteq A$, avec $|B| \leq k$ tel que

$$\forall i \neq j \in I, V(i) \cap B \neq V(j) \cap B ?$$

De manière évidente, $BID \in NP$: si l'on se donne $B \subseteq A, |B| \leq k$, alors on observe que la vérification de la propriété $\forall i \neq j \in I, V(i) \cap B \neq V(j) \cap B$ se fait aisément en un nombre d'opérations inférieur à $k \times \frac{|I| \times (|I| - 1)}{2}$ puisqu'il suffit de considérer les voisinages de chaque couple de I . Ce nombre d'opérations est bien polynômial par rapport à la taille de l'instance.

Nous allons maintenant montrer que BID est NP-complet en lui réduisant le problème 1-ICD, qui est NP-complet.

NOM : 1-ICD

ENTREE : Un graphe $H = (X, Z)$ sans jumeaux et un entier $m \leq |X|$.

QUESTION : Existe-t'il un code 1-identifiant de H qui soit de taille inférieure ou égale à m ?

Soit donc une entrée de 1-ICD, i.e, un graphe sans jumeaux $H = (X, Z)$ et un entier $m \leq |X|$. A partir de cela, nous voulons construire un graphe biparti sans jumeaux G et un entier k tel qu'il existe un code 1-identifiant de H de taille inférieure ou égale à m si et seulement s'il existe B IDIMIN pour G de taille inférieure ou égale à k .

Nous construisons donc le graphe G suivant : l'ensemble des sommets de G est égal à $I \cup A$ où $I = X$ et $A = \cup_{i \in I} \{A_i\}$ avec $A_i = \{i\} \cup V_H(i)$. On considère maintenant l'ensemble des arêtes suivant : chaque A_i (qui est un sous-ensemble de X) est relié à l'ensemble des éléments de X qu'il contient.

On pose finalement $k = m$.

La construction est représenté sur la figure suivante :

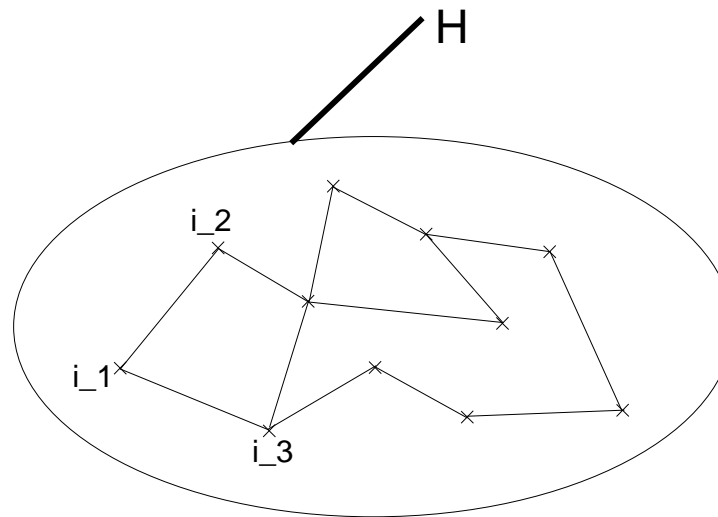


FIG. 31 – *Grappe H*

Ici, on a $X = I = \{i_1, i_2, \dots, \}$ et par exemple $A_{i_1} = \{i_1\} \cup \{i_2\} \cup \{i_3\}$

On représente le graphe G obtenu à partir de H , ainsi que l'arête correspondant à A_{i_1} sur la figure suivante.

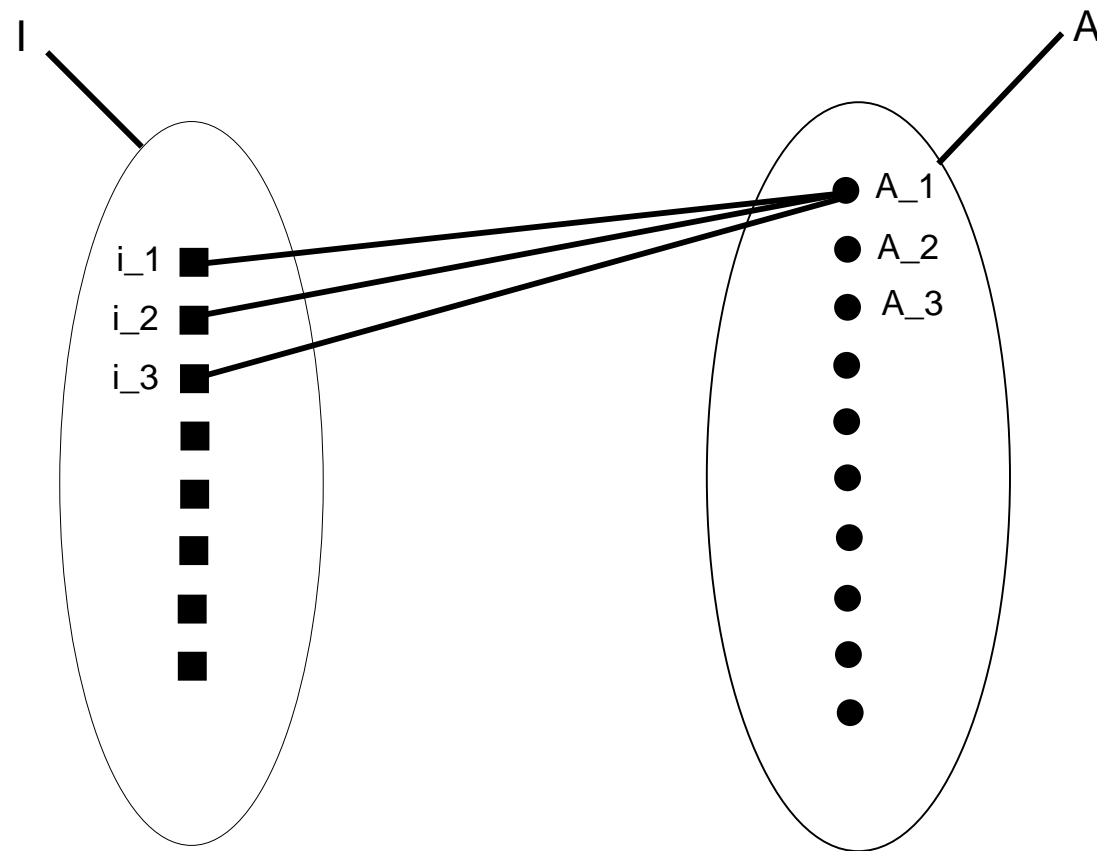


FIG. 32 – Graphe G avec le voisinage de A_1

Supposons maintenant que H possède un code 1-identifiant de taille inférieure ou égale à m . Soit $C = \{i_1, i_2, \dots, i_m\}$ ce code. Alors si l'on considère $B = \{A_{i_1}, A_{i_2}, \dots, A_{i_m}\}$, alors B est de taille égale à m , et vérifions que l'on a $\forall i \neq j \in I, V_G(i) \cap B \neq V_G(j) \cap B$.

Pour $i \in I$, $V_G(i) = \{A_i\} \cup \{A_j / j \in V_H(i)\} = \{A_j / j \in B_{1H}(i)\}$.

On en déduit

$$V_G(i) \cap B = \{A_j / j \in B_{1H}(i) \cap C\}$$

et ils sont donc tous bien distincts et non vides car C est un code 1-identifiant et que tous les $B_{1H}(i) \cap C$ sont distincts-non vides. Deux éléments de I ne pourront donc posséder la même famille de $\{A_i\}$.

On a donc bien trouvé un code B avec $|B| \leq m$ permettant d'identifier tous les éléments de I .

Réciproquement supposons maintenant que G admette $B \subseteq A$, avec $|B| \leq k$ et tel que

$$\forall i \neq j \in I, V_G(i) \cap B \neq V_G(j) \cap B$$

Montrons que H admet un code 1-identifiant de taille inférieure à k . Soit B de nouveau égal à $\{A_{i_1}, A_{i_2}, \dots, A_{i_m}\}$.

On considère donc $C = \{i_1, i_2, \dots, i_m\} \subseteq I$ et nous allons montrer qu'il est 1-identifiant. Soit $i \neq j \in I$, et soit $A_{i_p} \in (V_G(i) \cap B) \Delta (V_G(j) \cap B)$. Supposons par exemple $A_{i_p} \in (V_G(i) \cap B)$ et $A_{i_p} \notin (V_G(j) \cap B)$. Forcément i_p est dans C .

Si l'on considère maintenant

$$V_H(i) \cap C = (i \cup V(i)) \cap C = A_i \cap C \text{ et } V_H(j) \cap C = A_j \cap C.$$

Nous en déduisons clairement que $i_p \in V_H(i) \cap C$ mais que $i_p \notin V_H(j) \cap C$.

La transformation permettant de passer de G à H étant polynômiale, ceci conclut la démonstration de la NP-complétude de notre problème. \square

C'est fini