Codes Identifiants dans les Graphes

Julien Moncel

Université Joseph Fourier – Grenoble I

Codes Identifiants dans les Graphes

Julien Moncel

Thèse présentée pour l'obtention du titre de Docteur de l'Université Joseph Fourier
École Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique
Formation Recherche Opérationnelle, Combinatoire et Optimisation

Préparée au Laboratoire Leibniz-IMAG, UMR 5522
Soutenue publiquement le 27 juin 2005

Composition du jury :

Brigitte Plateau	Présidente
Simon Litsyn	Rapporteur
Antoine Lobstein	Rapporteur
Gerd Finke	Examinateur
André Raspaud	Examinateur
Sylvain Gravier	Directeur de thèse

Remerciements

Je voudrais exprimer toute ma gratitude à mon directeur de thèse Sylvain Gravier, qui m'a accompagné et soutenu pendant mon travail de recherche ces trois dernières années.

Mes remerciements les plus sincères aux membres du jury : Brigitte Plateau, qui m'a fait l'honneur de présider le jury ; Simon Litsyn et Antoine Lobstein, qui ont été deux rapporteurs patients et attentifs — merci pour toutes les corrections et améliorations qu'ils ont suggérées, merci pour leurs encouragements aussi — enfin merci à Gerd Finke et à André Raspaud, qui ont bien voulu examiner ma thèse.

Je voudrais saluer et remercier tous ceux avec qui j'ai eu l'occasion de travailler pendant ces trois années : Gábor Bacsó, Nadia Brauner, Irène Charon, Marc Daniel, Alan Frieze, András Gyárfás, Sándor Győri, Olivier Hudry, Vincent Jost, Antoine Lobstein, Ryan Martin, Michel Mollard, Charles Payan, Miklós Ruszinkó, András Sebő, Ahmed Semri, et Cliff Smyth.

Un clin d'œil aux thésards et stagiaires, ex-stagiaires et ex-thésards du laboratoire Leibniz, qui font que la vie du laboratoire ne se réduit pas à ce qu'elle serait sans eux : Pablo Arrighi, Attila Bernáth, Prakash Countcham, Marc Daniel, David Defossez, Paul Dorbec, Éric Duchêne, Mnacho Echenim, Karine Godot, Vincent Jost, Yann Kieffer, Marie Lalire, Pierre Lemaire, Benjamin Lévêque, Mehdi Mhalla, Cécile Ouvrier-Buffet, Simon Perdrix, Gábor Salamon, Éric Tannier, Nicolas Trotignon.

Mes pensées vont aussi vers ceux qui, en amont de ce travail de thèse, ont, d'une façon ou d'une autre, contribué à ce que je prenne la voie que j'ai prise : Jean-Marie Gadebois, Denise Grenier, Jean-Paul Leroux, Gérard Massine, Claude Maumet, et Charles Payan.

Un grand merci à la communauté du logiciel libre, qui fournit des outils de qualité – tels le langage LaTeX, les logiciels OpenOffice.org et GSview – permettant la réalisation d'un document comme celui-ci.

Je tiens à remercier le réseau ADONET, le programme EURODOC de la région Rhône-Alpes, l'ERTé "Maths à Modeler", et le CIES de Grenoble pour les opportunités qu'ils m'ont données de vivre des expériences plus enrichissantes les unes que les autres.

Enfin, je voudrais remercier tous ceux qui me rendent la vie plus belle chaque jour : toute ma famille, Léa, mes amis d'ici ou d'ailleurs, et tous ceux que j'oublie...

Ce travail est dédié à la mémoire de mon grand-père Harvey Cazier-Charpentier.

Table des matières

Introduction

Entre janvier et juin 1998, j'ai suivi à l'Université Joseph Fourier, à Grenoble, un module optionnel intitulé "Jeux Combinatoires et Raisonnement Mathématique". Les enseignants de ce module, Denise Grenier et Charles Payan, nous proposaient des sortes de casse-têtes qui, dans un sens, avaient à voir avec les mathématiques — après tout, l'intitulé du module ne contenait-il pas le mot "Mathématique" ? — mais n'avaient en tout cas rien à voir avec le programme de mathématiques du DEUG. Les énoncés de ces problèmes étaient très simples à comprendre, mais ils étaient difficiles dans le sens où il était malaisé de deviner quels théorèmes du cours il fallait appliquer pour résoudre ces problèmes. Tout compte fait, il s'avérait d'ailleurs qu'aucun théorème d'aucun cours n'était utilisable...

En essayant de paver des rectangles avec des dominos, de dessiner des figures sans passer deux fois par le même point, ou encore en essayant d'identifier la fausse pièce en un nombre minimum de pesées, je venais, avec mes camarades ébahis, de découvrir le monde merveilleux des Mathématiques Discrètes. J'étais loin de soupçonner ce que ces deux heures hebdomadaires de "Jeux Combinatoires" allaient impliquer par la suite.

Les années suivantes, j'étudiai avec le plus grand sérieux l'Analyse, la Topologie, l'Algèbre, et la Géométrie Différentielle à l'Université Joseph Fourier, mais je n'arrivais pas à me sortir ces Mathématiques Discrètes de la tête. En licence et en maîtrise, je fis deux stages au Laboratoire Leibniz, avec Sylvain Gravier et Charles Payan, qui me proposèrent des problèmes ouverts dans la droite lignée du module "Jeux Combinatoires et Raisonnement Mathématique". En été 2000, ils m'offrirent la possibilité de partir étudier la théorie des graphes en Hongrie, avec Gábor Bacso et András Gyárfás, qui m'apprirent énormément de choses.

Je fus surpris et déçu de ne jamais me voir proposer les Mathématiques Discrètes comme choix possible de module au cours de mes études de mathématiques. Les Mathématiques Discrètes n'étaient-elles donc pas des mathé-

matiques ? On finit par me le dire : les Mathématiques Discrètes, c'était en fait de l'informatique. Qu'à cela ne tienne ! Je m'infiltrai donc dans une école d'informatique environnante, l'ENSIMAG[1], où je pus en effet suivre quelques cours qui avaient bien l'air d'être des Mathématiques Discrètes, tels "Algorithmique", "Optimisation Combinatoire" ou "Management de la Production et des Services". À l'ENSIMAG, j'appris entre autres que les Mathématiques Discrètes n'étaient pas de l'informatique, mais des mathématiques...

En 2002, je m'inscrivis au DEA de Recherche Opérationnelle, Combinatoire et Optimisation de l'Institut National Polytechnique de Grenoble, à l'issue duquel je commençais une thèse avec Sylvain Gravier.

Sept ans après le module "Jeux Combinatoires et Raisonnement Mathématique", je suis toujours fasciné et amoureux de la combinatoire et de la théorie des graphes. En général, aucun théorème d'aucun cours ne me permet de résoudre les problèmes que je rencontre, et l'activité de recherche consiste essentiellement à se retrousser les manches et à comprendre les problèmes dans leur originalité. Les outils utilisés sont simples : des points, des traits, des patates, et, de temps à autre, quelques "cochonneries" comme $\forall$, $\exists$, x, δ, ou G — que celui qui n'a jamais écrit de cochonneries me jette la première patate. La simplicité des énoncés et des outils utilisés n'empêche pas — voire contribue — à la richesse des problèmes et à l'esthétique de la matière.

En conclusion de ces années de thèse, je pourrais dire que les Mathématiques Discrètes sont, à mes yeux, une discipline qui mérite son existence au moins autant que d'autres disciplines mieux reconnues. Les Mathématiques Discrètes sont des mathématiques belles et difficiles. Chose à ne pas négliger, elles possèdent de nombreuses applications pratiques, souvent liées au développement des nouvelles technologies.

Cette thèse constitue ma contribution à la compréhension d'un sujet récent tout à fait passionnant : les *codes identifiants.*

Les codes identifiants sont présentés de façon détaillée dans le premier chapitre de cette thèse. Ils modélisent un problème de détection de défaillance dans les réseaux qui fait partie de la grande famille des problèmes de couverture par tests. Ces problèmes possèdent de nombreuses applications pratiques, dans des domaines variés comme ceux de la reconnaissance de formes, de séquençage d'ADN, d'aide au diagnostic médical ou encore de communication dans des réseaux multi-utilisateurs.

[1]École Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble.

On peut aussi rattacher les codes identifiants aux problèmes de tests groupés, qui contiennent le problème des fausses pièces du module "Jeux Combinatoires et Raisonnement Mathématique".

Les questions que nous étudions dans cette thèse sont en lien avec le problème fondamental suivant :

Étant donné un graphe G, quels sont les codes identifiants C de G de cardinalité minimum ?

La tentation première est de chercher une procédure automatique — un *algorithme* — qui, étant donné un graphe quelconque G, nous fournit un code identifiant optimum C de G après un nombre fini d'étapes de calcul. Une telle procédure existe — considérer tous les sous-ensembles de sommets possibles de G — mais nécessite un nombre d'étapes de calcul exponentiel en la taille du graphe de départ G. Cette procédure ne permet donc en pratique que de résoudre le problème dans les cas où G n'a pas trop de sommets.

Il a été montré que le problème précédent était NP-difficile, ce qui, dans un sens, revient à dire que nous ne pouvons espérer guère mieux qu'un algorithme d'énumération exhaustive pour résoudre ce problème dans le cas général. Une solution consiste à contraindre le graphe de départ G, *i.e.* à supposer que G a une structure particulière, et à chercher un algorithme efficace adapté à la structure de G. Par efficace nous entendons en temps polynomial par rapport à la taille de G.

Dans le Chapitre 2, nous abordons donc ce problème du point de vue algorithmique, que nous étudions dans deux classes particulières de graphes : les arbres orientés et les fasciagraphes (qui généralisent la grille). Pour ces deux classes de graphes nous proposons des algorithmes polynomiaux et même linéaires.

Pour certaines classes de graphes particulières, nous n'avons pas besoin d'algorithme puisque nous pouvons directement, "à la main", déterminer des codes identifiants optimums de ces graphes.

Dans le Chapitre 3, nous donnons, par exemple, des codes identifiants optimums pour des cycles et des grilles. Dans certains cas, nous ne sommes pas capables de déterminer l'optimum, et nous donnons des bornes inférieures et supérieures pour la cardinalité minimum d'un code identifiant.

Dans le Chapitre 4, nous nous intéressons à des questions structurelles, qui consistent essentiellement à déterminer des graphes extrémaux vis-à-vis de

certaines propriétés des codes identifiants. Par exemple, nous nous intéressons à déterminer le degré minimum, le nombre minimum de sommets, ou encore le nombre maximum d'arêtes d'un graphe admettant un code identifiant. La construction de graphes admettant des codes identifiants de faible cardinalité sera abordée, ainsi que celle de graphes n'admettant que des codes identifiants de grande cardinalité.

Enfin, dans le Chapitre 5, nous étudions les codes identifiants dans les graphes aléatoires. Cette étude nous permet d'obtenir aussi des résultats déterministes d'existence. Les outils (élémentaires) de probabilités nécessaires seront rappelés au début de ce chapitre, qui se veut auto-suffisant.

Des rappels élémentaires de la théorie des graphes, destinés principalement à fixer les notations, ont été placés en annexe à la fin de ce document.

Chapitre 1

Codes identifiants

Dans ce chapitre nous présentons les codes identifiants comme un cas particulier du problème général de couverture par tests. L'application pratique ayant motivé l'introduction de cette notion est également présentée. Nous donnons plusieurs variantes et généralisations possibles de la notion de code identifiant. Des liens avec d'autres types de codes sont également présentés. Nous donnons ensuite des résultats préliminaires consistant essentiellement en bornes inférieures sur la cardinalité minimum d'un code identifiant dans les graphes. Enfin, nous fixons quelques notations utilisées dans l'ensemble de ce document.

1.1 Codes identifiants dans les graphes

1.1.1 Problèmes de couverture par tests

Les codes identifiants font partie de la famille des problèmes de couverture par tests, définis de façon générique comme suit :

> *Étant donnée une matrice M, quels sont les sous-ensembles de lignes de M telles que les colonnes résultantes soient toutes différentes ?*

La Figure 1.1 présente un exemple de ce problème dans le cas d'une matrice M à coefficients entiers.

Ce problème a été, à l'origine, formulé pour modéliser un problème d'aide au diagnostic médical, dans lequel les lignes de la matrice M correspondent

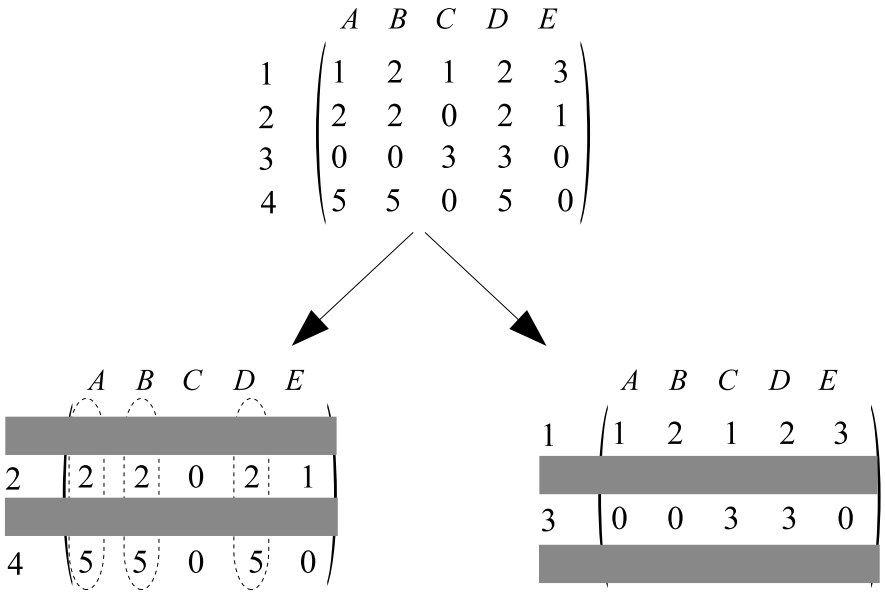


FIG. 1.1 – *Problème de couverture par tests : il faut déterminer un ensemble de lignes de la matrice permettant de différencier les colonnes entre elles. Ici, les restrictions aux lignes 2 et 4 des colonnes A, B et D sont les mêmes : les lignes 2 et 4 ne permettent donc pas de différencier les colonnes de la matrice. Les lignes 1 et 3, quant à elles, permettent d'identifier les colonnes A, B, C, D, E : les traces des colonnes A, B, C, D, E sur les lignes 1 et 3 sont toutes différentes.*

à des symptômes et les colonnes à des maladies. Les coefficients de M correspondent à l'intensité des symptômes dans les maladies. Le problème consistait à déterminer un sous-ensemble de symptômes qui identifiait chaque maladie de façon unique [PP80].

Les problèmes d'identification par tests sont très fréquents dans des domaines variés comme ceux de la reconnaissance de formes, de tests groupés en biologie, de détection de pannes, de localisation, ou encore de recherche de clés dans les bases de données (voir [CM02, HM76, Kog95, PP80]).

Nous étudions dans cette thèse un problème de couverture par tests dans le cas où la matrice M est la matrice d'adjacence d'un graphe, qui correspond à la détection de pannes dans les réseaux multiprocesseurs [KCL98].

1.1.2 Définition des codes identifiants

Soit $G = (V, E)$ un graphe non-orienté, et soit C un sous-ensemble de sommets de G. Si C est tel que tout sommet v de $V \setminus C$ est voisin d'au moins un sommet de C, alors on dit que C est un *code couvrant* de G (voir Figure 1.2). Un code couvrant est aussi appelé *dominant* du graphe. Les codes couvrants ont été largement étudiés dans la littérature [CHLL97].

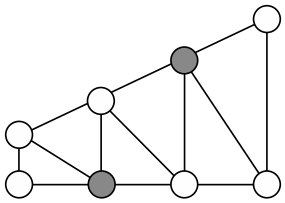


FIG. 1.2 – *Le sous-ensemble des sommets grisés est un code couvrant du graphe : tout sommet non grisé est voisin d'au moins un sommet grisé.*

Pour un sommet v de G, on définit le *voisinage étendu* de v comme l'ensemble

$$N[v] \ := \ N(v) \cup \{v\}.$$

Un sous-ensemble de sommets C est un code couvrant de G si et seulement si pour tout $v \in V$ on a

$$N[v] \cap C \ \neq \ \emptyset.$$

On dira qu'un sommet $c \in C$ *couvre* le sommet v s'il appartient au voisinage étendu de v.

Un sous-ensemble C' de sommets de G est un *code séparateur* de G si et seulement si pour toute paire de sommets distincts u, v de G on a

$$N[u] \cap C' \ \neq \ N[v] \cap C',$$

ou, de façon équivalente,

$$(N[u] \cap C') \, \Delta \, (N[v] \cap C') \ \neq \ \emptyset,$$

où $A \Delta B$ désigne la différence symétrique de A et B :

$$A \Delta B \ = \ (A \setminus B) \cup (B \setminus A).$$

On dira qu'un sommet $c \in C'$ *sépare* les sommets u et v s'il appartient à la différence symétrique de $N[u] \cap C'$ et $N[v] \cap C'$. Le sous-ensemble C' est

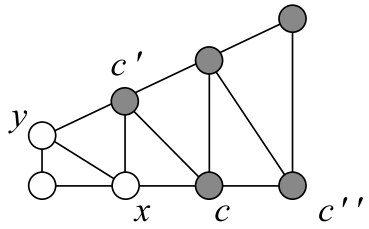


FIG. 1.3 – *L'ensemble des sommets grisés est un code séparateur du graphe : toute les paires de sommets du graphe sont séparées par un sommet grisé. Par exemple, le sommet c sépare x et y, puisque x et c sont voisins alors que y et c ne le sont pas. Le sommet c′, lui, ne sépare pas x et y puisqu'il est voisin des deux à la fois. Il sépare cependant c de c″.*

un code séparateur de G s'il sépare toutes les paires de sommets distincts de G (voir Figure 1.3).

La recherche d'un code séparateur d'un graphe G revient à résoudre un problème de couverture par tests dont l'instance est la matrice d'adjacence de G.

Un sous-ensemble de sommets de G qui est à la fois un code couvrant et un code séparateur de G est appelé un *code identifiant* de G (voir Figure 1.4). Ainsi, un sous-ensemble de sommets C d'un graphe G est un code identifiant de G si et seulement si tous les sommet de G sont couverts et séparés par C : $N[v] \cap C \neq \emptyset$ pour tout $v \in V$, et $N[u] \cap C \neq N[v] \cap C$ pour toute paire de sommets distincts u, v.

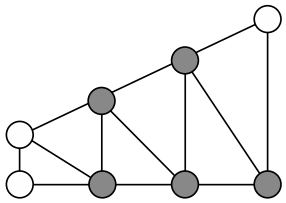


FIG. 1.4 – *L'ensemble des sommets grisés forme un code identifiant du graphe.*

L'ensemble $N[v] \cap C$ est appelé *ensemble identifiant* de v, on le note $I(v, C)$ ou simplement $I(v)$ s'il n'y a pas d'ambiguïté. C est un code identifiant de G si et seulement si l'application

$$v \mapsto I(v, C)$$

est une injection dont l'image ne contient pas l'ensemble vide : les ensembles identifiants des sommets de G sont non vides et distincts deux à deux.

1.1.3 Application pratique

Les codes identifiants ont été introduits pour modéliser un problème pratique d'identification de processeurs défectueux dans des réseaux multiprocesseurs. Nous détaillons ce problème dans cette section.

Supposons que chaque processeur `p` d'un réseau soit capable d'exécuter une procédure `test(p)`, qui s'applique à `p` ainsi qu'aux processeurs voisins de `p`. Cette procédure teste le bon fonctionnement de `p` et de ses voisins, et ne retourne qu'une information de type binaire : par exemple, 0 si une défaillance a été détectée sur `p` ou sur l'un de ses voisins, et 1 sinon. En supposant qu'à tout moment, au plus un processeur du réseau soit défectueux, le problème est de déterminer un sous-ensemble de processeurs `C` tel que :

- si au moins un des processeurs de `C` renvoie 0 après l'exécution de `test`, alors il y a un unique processeur défectueux dans le réseau, que nous sommes en mesure de localiser d'après les résultats des exécutions de `test` sur `C`,
- si tous les processeurs de `C` renvoient 1 après l'exécution de `test`, alors tous les processeurs du réseau sont en bon état de marche.

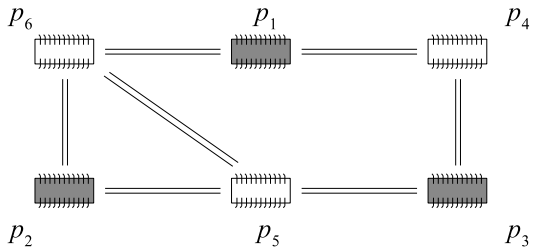


FIG. 1.5 – *Les processeurs p_1, p_2, p_3 permettent d'identifier de façon unique chaque processeur du réseau. Les processeurs p_4, p_5, p_6 couvrent bien le réseau mais ne permettent pas d'identifier de façon unique ses processeurs : nous ne pouvons pas distinguer le cas p_5 défectueux du cas p_6 défectueux si l'on prend $C = \{p_4, p_5, p_6\}$.*

Il est facile de voir qu'un sous-ensemble de processeurs `C` vérifie ces deux conditions si et seulement si l'ensemble de sommets correspondant C est un code identifiant du graphe associé au réseau (voir Figure 1.5).

En effet, la deuxième condition garantit le fait que si aucun processeur de C ne détecte de défaillance alors aucun processeur du réseau n'est défectueux : cela équivaut à dire que C est un code couvrant du graphe associé au réseau. La première condition équivaut, elle, à dire que C est un code séparateur du graphe. En effet, le sous-ensemble I de processeurs de `C` ayant renvoyé 0 après l'exécution de `test` est l'ensemble des processeurs de `C` voisins du processeur défectueux `p` : c'est l'ensemble identifiant de `p`. Dire que $I = I(\mathtt{p}, \mathtt{C})$ détermine de façon unique `p` équivaut à dire que C est un code séparateur du graphe du réseau.

Les codes identifiants ont été définis en 1998 par M. Karpovsky, K. Chakrabarty et L. Levitin dans [KCL98] pour modéliser ce problème. Dans [RSTU04] les codes identifiants sont également utilisés pour modéliser un problème de localisation par des réseaux de capteurs.

Dans la section suivante nous donnons plusieurs généralisations possibles et variantes des codes identifiants, ainsi que des liens avec d'autres types de codes. Ces généralisations sont elles aussi motivées par l'application pratique que nous venons de décrire.

1.2 Quelques généralisations et variantes possibles, liens avec d'autres types de codes

1.2.1 Identification à distance $t \geq 1$

Nous pouvons imaginer que la procédure `test` puisse tester tous les processeurs à distance au plus t du processeur exécutant la procédure, où t est une constante supérieure ou égale à 1, et où la distance entre deux processeurs est définie comme le nombre minimum d'arêtes d'un chemin entre ces deux processeurs. On parle dans ce cas de *codes t-identifiants* (voir Figure 1.6) ; et nous dirons simplement *code identifiant* pour code 1-identifiant.

Au niveau graphique, cette généralisation consiste simplement à considérer le problème dans la fermeture t-transitive de G : C est un code t-identifiant de $G = (V, E)$ si et seulement si C est un code 1-identifiant de $G^t = (V, E')$, deux sommets de G^t étant adjacents si et seulement si ils sont à distance inférieure ou égale à t dans G.

Alternativement, C est un code t-identifiant de G si et seulement si on a

$$B_t(u) \cap C \neq \emptyset$$

pour tout sommet u de G, et

$$B_t(u) \cap C \neq B_t(v) \cap C$$

pour toute paire u, v de sommets distincts de G, où $B_t(v)$ dénote la boule de rayon t centrée en v : $B_t(v)$ est l'ensemble des sommets à distance au plus t de v.

L'ensemble identifiant de v, $B_t(v) \cap C$, est en général désigné par $I_t(v, C)$ (voir Figure 1.6).

Il est à noter qu'un graphe admettant un code t-identifiant admet aussi un code t'-identifiant pour tout $t' \leq t$.

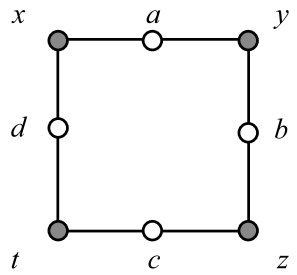


FIG. 1.6 – *Les processeurs x, y, z, t forment un code 1- et 2-identifiant du graphe, mais pas un code 3-identifiant car a, b, c et d sont tous à distance au plus 3 de x, y, z et t. Le graphe admet cependant un code 3-identifiant — par exemple constitué des sommets x, y, z, t, a, b, c, d — mais pas de code t-identifiant avec $t \geq 4$, car pour tout $t \geq 4$ et pour tout sommet v on a $B_t(v) = \{x, y, z, t, a, b, c, d\}$.*

Cette généralisation est donnée dans [KCL98], elle est largement étudiée dans la littérature (voir [BCHL04, BHL01, CHLa, CHLb, CHL02a, CHHL01, CHLZ99, GMS, HL02b, HL02c, KCL98, KCLA99] par exemple). Nous considérons cette généralisation dans les Chapitres 2, 3 et 4.

1.2.2 Identification d'ensembles de sommets

Nous pouvons, de plus, supposer qu'à tout moment il y ait au plus ℓ processeurs défectueux dans le réseau, où ℓ est une constante fixée. Pour faire face à cette éventualité, nous définissons les codes identifiant les ensembles d'au plus ℓ sommets de G.

Formellement, on dit que C est un code identifiant les ensembles d'au plus ℓ sommets de $G = (V, E)$ si et seulement si, les ensembles identifiants

$I(X,C)$ sont distincts pour tous les sous-ensembles X de cardinalité au plus ℓ de V, où $I(X,C)$ est défini comme l'union des ensembles identifiants des sommets de X :

$$I(X,C) \ := \ \bigcup_{x\in X} I(x,C) \ = \ \bigcup_{x\in X} N[x]\cap C.$$

On parle alors de *code* $(1,\leq \ell)$-*identifiant* (voir Figure 1.7).

On peut bien entendu combiner ceci avec l'identification à distance $t \geq 1$; on parle alors de *code* $(t,\leq \ell)$-*identifiant* : un code $(t,\leq \ell)$-identifiant d'un graphe G est un sous-ensemble de sommets C tel que les ensembles $I_t(X,C)$ sont distincts pour tous les sous-ensembles d'au plus ℓ sommets X de G, où $I_t(X,C)$ est défini comme $\bigcup_{x\in X} B_t(x)\cap C$.

Notons qu'un graphe admettant un code $(t,\leq \ell)$-identifiant admet aussi un code $(t',\leq \ell')$-identifiant pour tout $t' \leq t$ et pour tout $\ell' \leq \ell$.

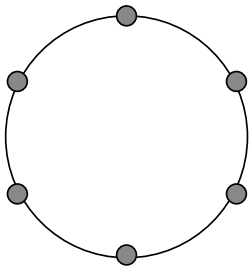


FIG. 1.7 – *L'ensemble des sommets du cycle forme un code* $(1,\leq 2)$-*identifiant du cycle.*

Cette généralisation a été considérée dans [KCL98], et a été étudiée dans [FMMRS, GM05, HL03a, HLR01, KCL98, KCLA99, Lai02a, LR01]. Nous étudions les codes $(t,\leq \ell)$-identifiants dans les Chapitres 4 et 5.

Au niveau matriciel, chercher un code $(1,\leq \ell)$-identifiant d'un graphe G revient à chercher un sous-ensemble de lignes $\mathcal{L}$ de la matrice d'adjacence de G tel que, si $M_{\mathcal{L}}$ désigne la sous-matrice de la matrice d'adjacence définie par $\mathcal{L}$, alors on a :

> Le `OU` bit-à-bit d'au plus ℓ colonnes de $M_{\mathcal{L}}$ est différent du `OU` bit-à-bit d'au plus ℓ autres colonnes de $M_{\mathcal{L}}$. (1.1)

On rappelle que le `OU` bit-à-bit de deux vecteurs $u,\ v \in \{0,1\}^n$ est un vecteur $w \in \{0,1\}^n$ tel que $w_i = u_i$ `OU` v_i pour tout $i = 1,\ldots,n$ (on rappelle que a `OU` $b = 0$ si et seulement si a et b sont nuls).

Un ensemble de vecteurs satisfaisant (1.1) est connu sous le nom de *code ℓ-superimposé* ou *code UD_ℓ*, notion introduite par W. Kautz et R. Singleton en 1964 dans [KS64] pour modéliser un problème d'identification dans des canaux de communication multi-utilisateurs et des problèmes de recherche dans des bases de données.

Ces codes sont rattachés à la famille des problèmes de *tests groupés* — *group testing* en anglais [DH00] — qui, dans leur version générique, s'énoncent comme suit :

> *Dans un ensemble E de n objets parmi lesquels au plus k sont défectueux, minimiser le nombre de tests à effectuer pour déterminer les objets défectueux de E.*

Un exemple fameux de test groupé est le problème dit des fausses pièces, dans lequel il faut, en un nombre minimum de pesées, déterminer quelles sont les fausses pièces parmi un ensemble de n pièces. Le nombre de fausses pièces, p, est connu à l'avance, mais nous ignorons si les fausses pièces sont plus lourdes ou plus légères que les pièces normales. Les problèmes de tests groupés ont de nombreuses autres applications, parmi lesquelles le séquençage d'ADN en génomique [CM02, CM04].

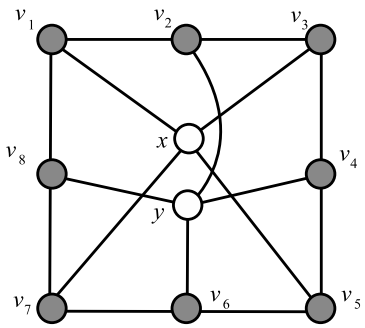
Grâce à la propriété (1.1), un code $(1, \leq \ell)$-identifiant nous fournit donc trivialement un code ℓ-superimposé (voir Figure 1.8). Dans le Chapitre 4, nous verrons comment obtenir un code $(1, \leq \ell)$-identifiant à partir d'un code ℓ-superimposé maximal. Dans le paragraphe suivant, nous montrons que dans les cas des graphes orientés il y a un lien fort entre les codes identifiants et les codes superimposés.

1.2.3 Cas orienté

Nous pouvons imaginer que les liens entre les processeurs du réseau soient directionnels, c'est-à-dire que le graphe du réseau considéré soit orienté. En ce cas, un processeur `p` peut tester un processeur `p'` si et seulement si il y a un arc de `p` vers `p'`. Le fait que `p` puisse tester `p'` n'implique pas que `p'` puisse tester `p`.

Au niveau graphique, il suffit dans ce cas de remplacer $N[v]$ par $\Gamma^-[v]$ dans la définition d'un code identifiant, où $\Gamma^-[v]$ désigne l'ensemble fermé des voisins entrants de v :

$$\Gamma^-[v] \ := \ \Gamma^-(v) \cup \{v\}.$$



	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	x	y
v_1	1	1	0	0	0	0	0	1	1	0
v_2	1	1	1	0	0	0	0	0	0	1
v_3	0	1	1	1	0	0	0	0	1	0
v_4	0	0	1	1	1	0	0	0	0	1
v_5	0	0	0	1	1	1	0	0	1	0
v_6	0	0	0	0	1	1	1	0	0	1
v_7	0	0	0	0	0	1	1	1	1	0
v_8	1	0	0	0	0	0	1	1	0	1

FIG. 1.8 – *Un graphe G admettant un code $(1, \leq 2)$-identifiant C (sommets en grisé). La matrice représentée est la sous-matrice de la matrice d'adjacence de G engendrée par les sommets de C. Les colonnes de la matrice forment un code 2-superimposé.*

Ainsi, un sous-ensemble de sommets C d'un graphe orienté G est un code identifiant de G si et seulement si

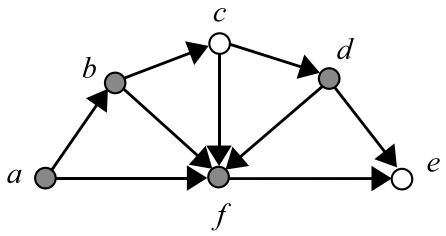
$$\Gamma^-[v] \cap C \neq \emptyset$$

pour tout sommet v, et

$$\Gamma^-[u] \cap C \neq \Gamma^-[v] \cap C$$

pour toute paire de sommets distincts u, v de G (voir Figure 1.9).

Cette variante a été considérée en [CGHLMM, CHL02b], et nous l'étudions dans le Chapitre 2. L'ensemble identifiant de v, $\Gamma^-(v) \cap C$, est noté $I^-(v, C)$.



$C = \{a, b, d, f\}$

$I^-(a, C) = \{a\}$ $\quad I^-(b, C) = \{a, b\}$

$I^-(c, C) = \{b\}$ $\quad I^-(d, C) = \{d\}$

$I^-(e, C) = \{d, f\}$ $\quad I^-(f, C) = \{a, b, f\}$

FIG. 1.9 – *Exemple de graphe orienté muni d'un code identifiant. On remarque que, si l'on fait abstraction des orientations, alors l'ensemble C spécifié n'est pas un code identifiant du graphe non-orienté résultant.*

Nous pouvons également identifier des sommets à distance $t \geq 1$ ou identifier des ensembles d'au plus ℓ sommets dans les graphes orientés. Dans ce dernier cas, nous pouvons établir un lien fort entre les codes $(1, \leq \ell)$-identifiants dans les graphes orientés et les codes ℓ-superimposés.

Théorème 1.1 ([FMMRS])
Il est possible de réaliser un code ℓ-superimposé maximum de t vecteurs de $\{0,1\}^N$ comme un graphe orienté à t sommets muni d'un code $(1, \leq \ell)$-identifiant de cardinalité N.

Preuve : Soit $\{v_1, \ldots, v_t\}$ un code ℓ-superimposé maximal de $\{0,1\}^N$, et soit M la matrice $N \times t$ dont les colonnes sont $v_1, \ldots, v_t$. À partir d'une sous-matrice $N \times N$ de M' n'ayant que des 1 sur sa diagonale, on peut facilement construire un graphe à t sommets $G = (V, E)$, muni d'un code ℓ-identifiant C de cardinalité N : chaque colonne de M' correspond à un sommet de C, et les autres colonnes de M correspondent aux sommets de $V \setminus C$. Les arcs de G sont déterminés par les coefficients de M.

Soit $\{A, B\}$ le graphe biparti "colonnes-coordonnées" associé à M : $A = \{1, \ldots, N\}$ et $B = \{v_1, \ldots, v_t\}$, et il y a une arête entre i et v_j si et seulement si la i-ème coordonnée de v_j est égale à 1. Nous montrons qu'il existe un couplage de $\{A, B\}$ couvrant A. En effet, en utilisant le Théorème de Hall, s'il n'existe pas de couplage de $\{A, B\}$ qui couvre A, alors il existe $X \subseteq A$ tel que $|N(X)| < |X|$. En remplaçant ces $|N(X)|$ vecteurs par les $|X|$ vecteurs unité sur l'ensemble de coordonnées X, alors on obtient un code ℓ-superimposé de cardinalité strictement supérieure à celle du code original, ce qui est une contradiction. Donc il existe un couplage de $\{A, B\}$ qui couvre A. Ce couplage correspond à une sous-matrice $N \times N$ de M' n'ayant que des 1 sur sa diagonale. □

1.2.4 Cas où les sommets du code n'ont pas à être identifiés

Nous pouvons également supposer que les sommets de C n'ont pas à être identifiés, ce qui revient à faire l'hypothèse que les processeurs qui exécutent la procédure `test` ne seront jamais eux-mêmes défectueux. En ce cas, il nous suffit de vérifier la condition

$$N[u] \cap C \neq N[v] \cap C$$

pour toute paire de sommets distincts de $V \setminus C$.

On dit que C est un *code localisateur-dominateur* de G s'il couvre et sépare tous les sommets de $V \setminus C$ (voir Figure 1.10). Cette notion est antérieure à celle de code identifiant et a été largement étudiée (voir par exemple [BCHL04, CHLa, CHL02b, CSS87, RS84]).

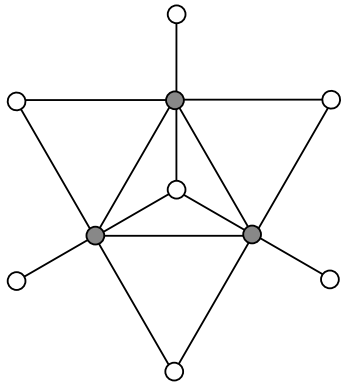


FIG. 1.10 – *Un graphe muni d'un code localisateur-dominateur : les sommets du code n'ont pas à être séparés les uns des autres. Un code identifiant est toujours un code localisateur-dominateur, mais la réciproque est fausse comme dans cet exemple.*

1.2.5 Cas où les sommets défectueux peuvent renvoyer une information erronée

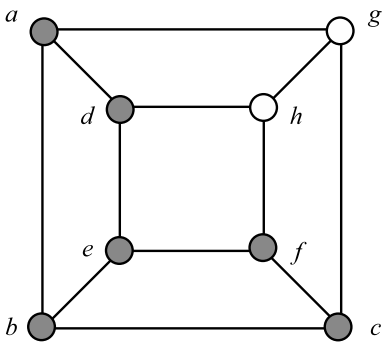
Il est légitime de se demander si un processeur défectueux ne risque pas de renvoyer une information erronée après exécution de la procédure `test`. Afin de procéder à l'identification de processeurs défectueux sous cette nouvelle hypothèse, on définit, pour tout sous-ensemble d'au plus ℓ sommets X, l'ensemble $\mathcal{I}(X, C)$ comme suit :

$$\mathcal{I}(X, C) \ := \ \{U \mid I(X, C) \setminus (X \cap C) \subseteq U \subseteq I(X, C)\},$$

où $I(X, C)$ est égal à $\bigcup_{x \in X} N[x] \cap C$.

On dit alors qu'un sous-ensemble de sommets est un code identifiant *au sens fort* [HLR02] si et seulement si $\mathcal{I}(X, C) \neq \mathcal{I}(Y, C)$ pour toute paire X, Y de sous-ensembles distincts d'au plus ℓ sommets (voir Figure 1.11). On peut aussi combiner ceci avec l'identification à distance $t \geq 1$ et généraliser aux graphes orientés en remplaçant $N[v]$ par $\Gamma^-[v] = \Gamma^-(v) \cup \{v\}$.

Cette généralisation a été considérée dans [HLR02, Lai02b, LR02].



$C=\{a,b,c,d,e,f\}$

$\mathbf{I}(a,C)=\{\{a,b,d\},\{b,d\}\}$
$\mathbf{I}(b,C)=\{\{a,b,c,e\},\{a,c,e\}\}$
$\mathbf{I}(c,C)=\{\{b,c,f\},\{b,f\}\}$
$\mathbf{I}(d,C)=\{\{a,d,e\},\{a,e\}\}$
$\mathbf{I}(e,C)=\{\{b,d,e,f\},\{b,d,f\}\}$
$\mathbf{I}(f,C)=\{\{c,e,f\},\{c,e\}\}$
$\mathbf{I}(g,C)=\{\{d,f\}\}$
$\mathbf{I}(h,C)=\{\{a,c\}\}$

FIG. 1.11 – *Exemple de graphe muni d'un code* $(1,\leq 1)$*-identifiant au sens fort : on peut localiser le sommet défectueux même si celui-ci est un mot du code qui renvoie une information éventuellement erronée.*

1.2.6 Codes identifiants dans les hypergraphes

On peut généraliser de façon naturelle les codes identifiants au cas des hypergraphes de la façon suivante : soit $\mathcal{H} = (V, E)$ un hypergraphe et soit C un sous-ensemble de sommets de $\mathcal{H}$. On dit que C est un code identifiant de $\mathcal{H}$ si et seulement si on a

$$e \cap C \neq \emptyset$$

pour toute hyperarête e, et

$$e \cap C \neq f \cap C$$

pour toute paire d'arêtes distinctes $e, f \in E$. Le code identifiant C permet alors d'identifier de façon unique les hyperarêtes de $\mathcal{H}$ (voir Figure 1.12).

Soit $t \geq 1$. Étant donné un graphe $G = (V, E)$, soit $\mathcal{H}_t(G) = (V, E')$ l'hypergraphe dont les hyperarêtes correspondent aux boules de rayon t de G : on a $B_t(v) \in E'$ pour tout $v \in V$. L'hypergraphe $\mathcal{H}_t(G)$ a n sommets et n hyperarêtes. La définition précédente généralise celle de code identifiant dans les graphes, dans le sens où un sous-ensemble C de V est un code t-identifiant de G si et seulement si c'est un code identifiant de $\mathcal{H}_t(G)$. De même, dans le cas d'un graphe orienté G, on peut construire un hypergraphe dont les hyperarêtes correspondent aux voisinages entrants des sommets de G : $\Gamma^-[v]$ est une hyperarête pour tout sommet v de G.

On peut également généraliser la notion de code identifiant des ensembles de sommets à l'aide des hypergraphes : C est un code identifiant les ensembles

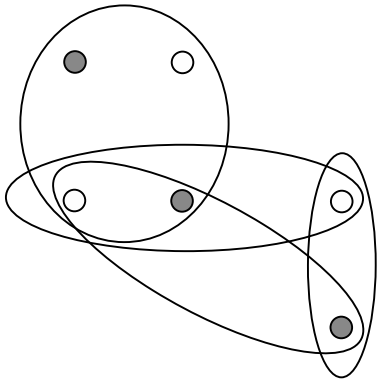


FIG. 1.12 – *Exemple d'hypergraphe muni d'un code identifiant : les hyperarêtes sont identifiées de façon unique par la trace du code.*

d'au plus ℓ sommets de $\mathcal{H}$, si et seulement si

$$\bigcup_{e \in X} e \cap C \neq \bigcup_{f \in Y} f \cap C$$

pour tous les sous-ensembles X et Y d'au plus ℓ hyperarêtes de $\mathcal{H}$.

Cette généralisation a un lien fort avec la notion de code correcteur d'erreurs au sens classique du terme. Rappelons qu'un sous-ensemble de sommets C d'un graphe G est un *code correcteur* de G si et seulement si tous les voisinages étendus $N[v]$, $v \in C$, sont d'intersection vide (voir Figure 1.14).

Soit $\mathcal{H}$ un hypergraphe 2-uniforme[1] sur un ensemble de sommets V, et soit C un code identifiant de $\mathcal{H}$. Comme $\mathcal{H}$ est 2-uniforme, alors $\mathcal{H}$ peut être vu comme un graphe $G(\mathcal{H})$ sur l'ensemble de sommets V, les arêtes de $G(\mathcal{H})$ correspondant aux hyperarêtes de $\mathcal{H}$. Un code identifiant C de $\mathcal{H}$ est alors un code identifiant les arêtes de $G(\mathcal{H})$, dans le sens où toute arête uv est identifiée de façon unique par la trace de C sur uv. La notion d'identification des arêtes d'un graphe a déjà été considérée dans [HKL01, HKL03].

Théorème 1.2
Soit C un sous-ensemble de sommets d'un graphe G. Alors C est un code identifiant les arêtes de G si et seulement si $\overline{C}$, le complémentaire de C, est un code correcteur de G.

[1] On rappelle qu'un hypergraphe dont toutes les hyperarêtes sont de cardinalité $r \geq 1$ est dit *r-uniforme*.

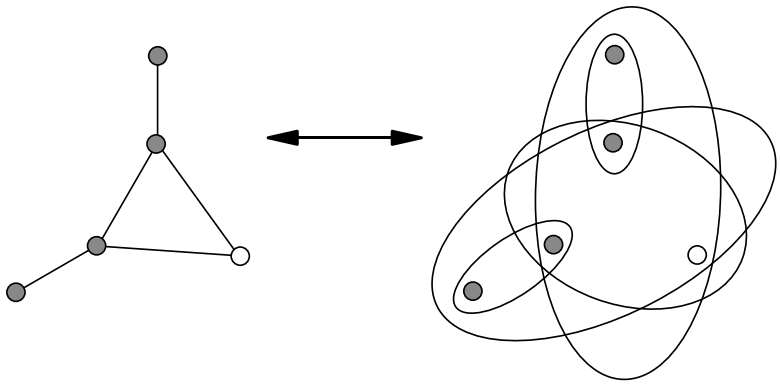


FIG. 1.13 – *Exemple de graphe G muni d'un code 1-identifiant C et son hypergraphe associé $\mathcal{H}_1(G)$. Les hyperarêtes de $\mathcal{H}_1(G)$ sont les voisinages étendus des sommets de G. L'ensemble C est un code identifiant de $\mathcal{H}_1(G)$.*

Preuve : Soit C un code identifiant les arêtes d'un graphe G. Il est facile de voir que $\overline{C}$ est un code correcteur de G. En effet, soient u et v deux sommets distincts de $\overline{C}$. Ils ne sont pas voisins, sinon l'arête uv ne serait pas couverte par C. Ils ne sont pas non plus à distance 2, car sinon il existe $w \in C$ tel que w est voisin de u et v, et les arêtes uw et vw ne sont pas séparées : elles sont toutes les deux couvertes par w. Les deux sommets u et v sont donc à distance au moins 3, ce qui est équivalent à dire que $\overline{C}$ est un code correcteur de G.

Réciproquement, soit C un code correcteur de G. On montre que $\overline{C}$ est un code identifiant les arêtes de G. En effet, il suffit de remarquer que deux arêtes incidentes à un même sommet sont identifiées. Soient uv et uw deux telles arêtes : puisque C est un code correcteur de G alors au plus un sommet parmi u, v et w appartient à C, *i.e.* au moins deux sommets parmi u, v et w appartiennent à $\overline{C}$. Il est facile de voir que ceci implique que les deux arêtes uv et uw sont identifiées. Deux arêtes non incidentes sont nécessairement séparées, et enfin toute arête est nécessairement couverte par $\overline{C}$. Ceci montre que $\overline{C}$ est un code identifiant les arêtes de G. □

La notion de code identifiant dans les hypergraphes nous permet donc d'unifier beaucoup de généralisations et variantes des codes identifiants : elle englobe les notions de codes identifiants dans les graphes orientés, de codes identifiants à distance $t \geq 1$, de codes identifiant des ensembles de sommets, ou de codes identifiant les arêtes d'un graphe. I. Honkala et A. Lobstein avaient également proposé un problème d'identification de sommets de $\mathbb{Z}^2$ par des polyminos dans [HL03b] : ce problème est lui aussi facilement exprimable

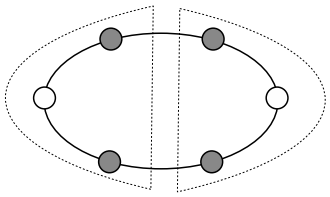


FIG. 1.14 – *Exemple de graphe muni d'un code identifiant les arêtes C. Le complémentaire de C est un code correcteur du graphe. On peut noter que le complémentaire de C est de cardinalité maximum, ce qui implique l'optimalité de C.*

en termes d'hypergraphe.

Dans cette thèse nous n'étudions pas les codes identifiants dans les hypergraphes, mais nous envisageons l'étude des codes identifiants dans ce cadre-là comme poursuite possible de cette thèse.

1.2.7 Jeux et stratégies

Oublions un instant la condition *"$N[v] \cap C \neq \emptyset$ pour tout sommet v du graphe"*. Dans ce cas, nous pouvons voir le problème d'identification de sommets dans un graphe comme un jeu à deux joueurs consistant, pour l'un des joueurs, à deviner le sommet inconnu x choisi en secret par l'autre joueur. À l'image de nombreux jeux de devinettes populaires ou de jeux commerciaux comme le "Qui est-ce ?", l'objectif du premier joueur est de minimiser le nombre de coups nécessaires pour déterminer le sommet inconnu x. Un jeu de ce type s'appelle *jeu de Rényi*[2] [Ren62]. Ajoutons qu'une règle inviolable de ce jeu est que le second joueur n'a pas le droit de mentir[3].

L'ajout de la condition *"$N[v] \cap C \neq \emptyset$ pour tout sommet v du graphe"* revient à considérer que le second joueur a la possibilité de choisir "aucun sommet". Ceci ne change pas fondamentalement le jeu si l'on rajoute arti-

[2] A. Rényi (1921–1970), mathématicien hongrois, fondateur de l'Institut des Mathématiques de Budapest, auteur de nombreux résultats en théorie des nombres, probabilités et théorie de l'information. Il serait l'auteur du bon mot :

"un mathématicien est une machine à convertir le café en théorèmes",

plus souvent attribué à P. Erdős.

[3] Il existe des généralisations de ce jeu dans lesquelles le second joueur est autorisé à mentir un certain nombre de fois ; cette variante est connue sous le nom de *jeu de Rényi-Ulam* [Ula76].

ficiellement un sommet isolé ε au graphe, tel que ε n'appartient pas à C : choisir un sommet inconnu dans $G \cup \{\varepsilon\}$ revient alors au problème de code identifiant dans G.

La présentation que nous avons faite de ce problème consiste, d'un point de vue stratégique, à déterminer à l'avance un ensemble de questions du type *"est-ce que x appartient au voisinage étendu de c ?"* pour c appartenant à un code identifiant C du graphe considéré au début du jeu — le graphe en question étant connu des deux joueurs. Le fait que C soit un code identifiant du graphe du jeu garantit que le premier joueur pourra toujours trouver le sommet inconnu x quelles que soient les réponses du second joueur : l'ensemble des sommets c tels que la réponse à la question *"est-ce que $x \in N[c]$?"* forme un sous-ensemble de C déterminant de façon unique le sommet inconnu x (rappelons que $x \mapsto I(x, C)$ est injective). L'ordre des questions n'a donc aucune importance dans la présentation que nous avons faite de ce problème. Ce point de vue sera parfois adopté pour résoudre des questions combinatoires liées aux codes identifiants, comme par exemple dans la Proposition 4.3.

Dans la plupart des jeux de devinettes de ce type, le fonctionnement est un peu différent. En particulier, l'*ordre* des questions a souvent une importance stratégique. En effet, dans ces jeux de questions-réponses, le premier joueur pose ses questions les unes après les autres, chaque question tenant compte des réponses aux questions posées précédemment. La liberté laissée au joueur d'adapter ses questions en fonction des réponses à ses questions précédentes lui permet en général de trouver la réponse plus rapidement.

On parle de *stratégie adaptative* lorsque nous autorisons le premier joueur à choisir ses questions au fur et à mesure. Lorsque l'ensemble des questions posées est déterminé à l'avance on parle de *stratégie non-adaptative* ou encore de *stratégie linéaire*.

Une stratégie adaptative peut être représentée par un arbre enraciné, où la racine de l'arbre est la première question posée par le joueur, et les fils d'un nœud α représentent les questions possibles selon la réponse à la question du nœud α. Dans le cas courant où les réponses sont de type binaire ("oui-non" pour le "Qui est-ce ?", 0-1 pour les codes identifiants), l'arbre de la stratégie est un arbre binaire (voir Figure 1.15).

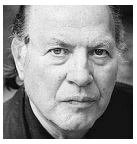
Dans le cas des codes identifiants nous pourrions également considérer des stratégies adaptatives pour la détermination du sommet inconnu x. Au lieu de déterminer un ensemble de sommets C à l'avance, nous pourrions déterminer un arbre binaire de décision enraciné, dont chaque nœud serait

Pál Erdős († 1996)
Mathématicien hongrois

John Steinbeck († 1968)
Écrivain américain

Vera Sós
Mathématicienne hongroise

Arundhati Roy
Écrivaine indienne



Imre Kertész
Écrivain hongrois

Srinivasa Ramanujan († 1920)
Mathématicien indien

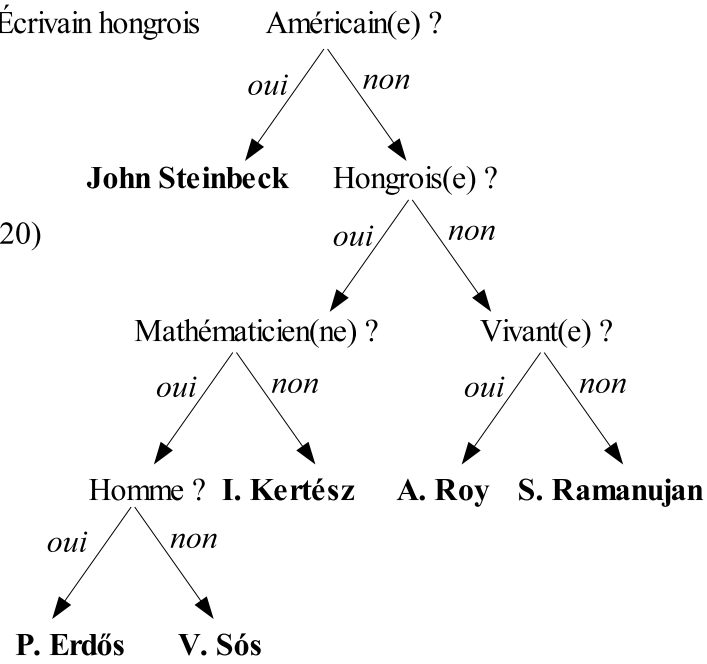


FIG. 1.15 – *Exemple de stratégie adaptative pour un jeu de devinettes du type "Qui est-ce ?". L'arbre binaire obtenu est de profondeur 4. Peut-on trouver une stratégie ayant une profondeur plus petite ? Quel est le minimum ?*

une question du type *"est-ce que x appartient au voisinage étendu de y ?"*, tel qu'à chaque feuille nous soyons en mesure de déterminer quel est le sommet inconnu x (éventuellement ce sommet inconnu est ε, *i.e.* aucun processeur du réseau n'est défectueux).

Dans ce contexte, il serait pertinent d'optimiser plusieurs critères, comme

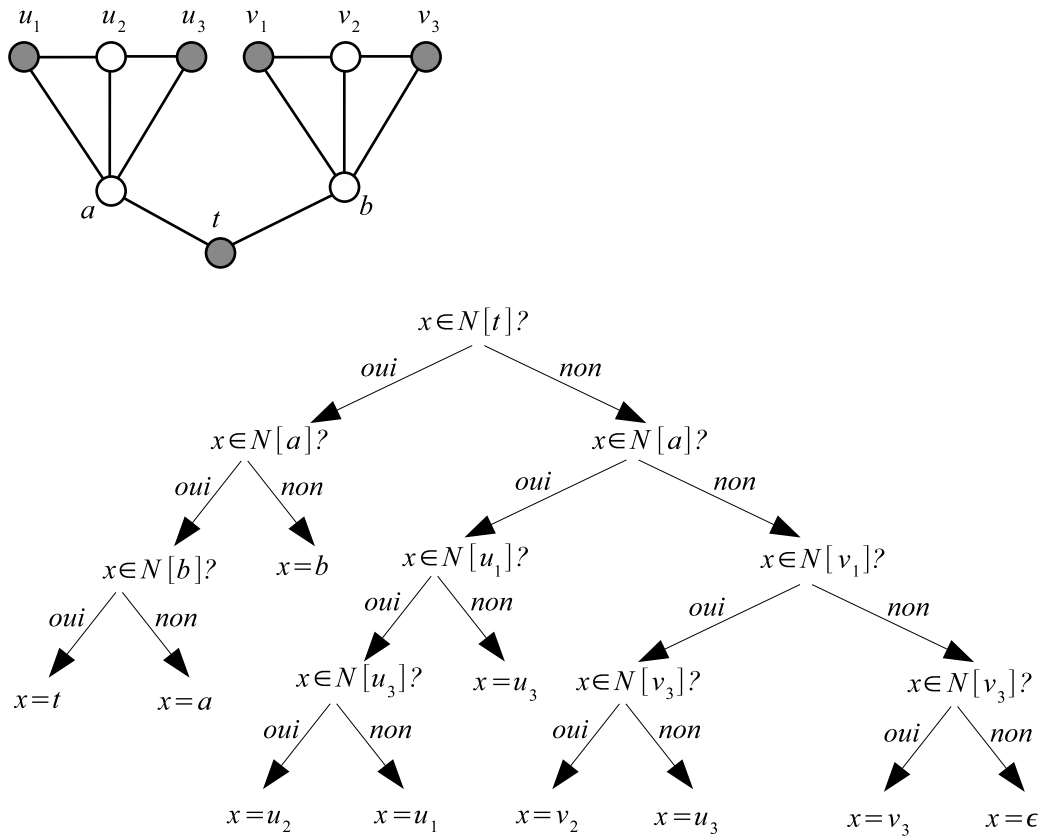


FIG. 1.16 – *Exemple de stratégie adaptative dans le cas déterministe : nous pouvons concevoir une stratégie adaptative nécessitant au plus quatre questions, alors que la cardinalité minimum d'un code identifiant du graphe considéré est cinq.*

la profondeur de l'arbre (voir Figure 1.16) ou encore la distance moyenne d'une feuille à la racine. Nous pourrions aussi introduire des probabilités de défection pour chaque processeur, et minimiser l'espérance du nombre de coups nécessaires pour déterminer le processeur défectueux (voir Figure 1.17). D'un point de vue pratique, il nous paraît en effet sensé de considérer que les processeurs n'aient pas tous la même fiabilité.

Il est à remarquer que, même sans introduire de probabilité de défection, une stratégie adaptative peut permettre de déterminer le sommet inconnu en un nombre maximum de questions qui soit inférieur à la cardinalité minimum d'un code identifiant du graphe considéré (voir figure 1.16).

Ce point de vue n'a, à notre connaissance, jamais été considéré dans la

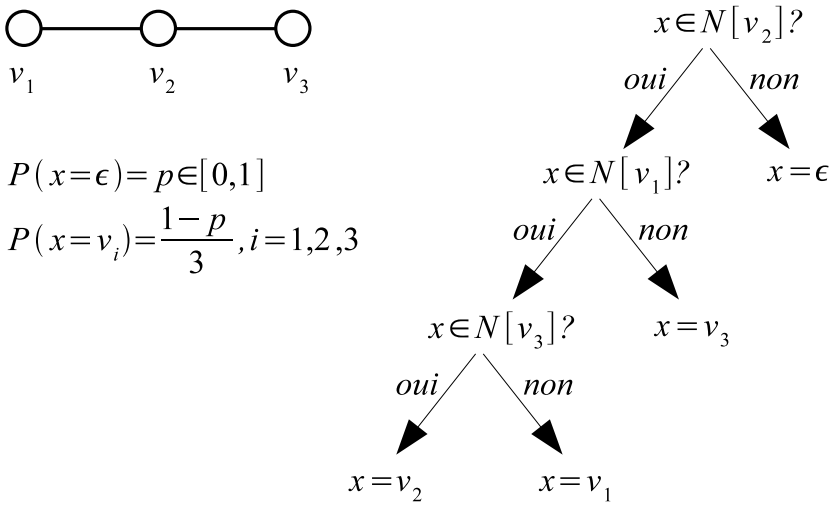


FIG. 1.17 – *Exemple de stratégie adaptative pour l'identification du sommet inconnu dans le chemin à trois sommets P_3. On sait que la cardinalité minimum d'un code identifiant de P_3 est 2, ainsi la stratégie non-adaptative optimum utilise deux questions du type "est-ce que $x \in N[v]$?" pour trouver le sommet inconnu x. En supposant que la probabilité que le sommet inconnu soit ε est $p \in [0,1]$, et que la probabilité que le sommet inconnu soit v_i est $\frac{1-p}{3}$ pour $i = 1,2,3$, cette stratégie adaptative utilise en moyenne $\frac{8}{3} - \frac{5}{3}p$ questions du type "est-ce que $x \in N[v]$?", ce qui est inférieur à 2 dès que $p > \frac{2}{5}$.*

littérature concernant les codes identifiants, mais est fréquemment étudié pour d'autres types de codes (voir par exemple [Ber68, DR02, RV97] ou l'article de *survey* [Pel02] concernant les jeux de type Rényi-Ulam). Dans cette thèse nous ne considérons pas de stratégies adaptatives pour les codes identifiants, mais il nous semblerait pertinent, dans l'avenir, de considérer ce point de vue, qui nous semble riche en possibilités.

1.3 Premiers résultats

Dans cette section nous ne considérons que des graphes non-orientés.

Il est à noter que tous les graphes n'admettent pas de code identifiant. Par exemple, si un graphe G contient deux sommets u et v tels que

$$N[u] = N[v],$$

alors pour tout sous-ensemble de sommets C de G on aura

$$N[u] \cap C \ = \ N[v] \cap C,$$

et G n'admet pas de code 1-identifiant. De même, dans le cas des codes identifiant des ensembles de sommets, si G contient deux sous-ensembles distincts d'au plus ℓ sommets X, Y, tels que

$$N[X] \ = \ N[Y],$$

alors G n'admet pas de code $(1, \leq \ell)$-identifiant. Il est facile de voir que ces conditions sont nécessaires et suffisantes :

Théorème 1.3 ([KCL98])
Un graphe G admet un code $(t, \leq \ell)$-identifiant si et seulement si on a :

$$\bigcup_{x \in X} B_t(x) \ \neq \ \bigcup_{y \in Y} B_t(y)$$

pour toute paire X, Y de sous-ensembles d'au plus ℓ sommets de G, où $B_t(x)$ désigne l'ensemble des sommets à distance au plus t de x. Par conséquent, G admet un code $(t, \leq \ell)$-identifiant si et seulement si $V(G)$ est un code $(t, \leq \ell)$-identifiant de G.

Preuve : S'il existe deux sous-ensembles d'au plus ℓ sommets X, Y tels que $\bigcup_{x \in X} B_t(x) = \bigcup_{y \in Y} B_t(y)$, alors pour tout sous-ensemble de sommets C de G on aura $\bigcup_{x \in X} B_t(x) \cap C = \bigcup_{y \in Y} B_t(y) \cap C$, et G n'admet pas de code $(t, \leq \ell)$-identifiant. Réciproquement, si

$$\bigcup_{x \in X} B_t(x) \ \neq \ \bigcup_{y \in Y} B_t(y) \tag{1.2}$$

pour toute paire X, Y de sous-ensembles d'au plus ℓ sommets de G, alors $C = V(G)$ est trivialement un code $(t, \leq \ell)$-identifiant de G. □

Remarquons que, si ℓ est fixé, alors la vérification de la condition (1.2) se fait en temps polynomial (si t n'est pas fixé, considérer le problème dans G^t, la fermeture t-transitive du graphe). Nous ignorons la complexité du problème suivant :

IDENTIFIABLE
Instance : Un graphe G, deux entiers $t \geq 1$ et $\ell \geq 1$.
Question : Est-ce que G admet un code $(t, \leq \ell)$-identifiant ?

Deux sommets u, v tels que $B_t(u) = B_t(v)$ seront parfois appelés *sommets jumeaux*. Ainsi, un graphe admet un code t-identifiant si et seulement si il ne contient pas de sommets jumeaux.

Lorsqu'un graphe G admet un code identifiant, alors $V(G)$ est toujours un code identifiant de G, ainsi le problème d'optimisation associé est de *trouver un code identifiant de cardinalité minimum*. Ce problème est NP-difficile, aussi bien dans les graphes orientés [CHL02b] que dans les graphes non-orientés [CHL03]. Dans le deuxième chapitre de cette thèse nous étudierons la complexité de ce problème dans des classes restreintes de graphes.

Dans le cas général on peut établir des bornes inférieures pour la cardinalité minimum d'un code $(t, \leq \ell)$-identifiant dans un graphe.

Théorème 1.4 ([KCL98])
Soit $G = (V, E)$ un graphe muni d'un code $(t, \leq \ell)$-identifiant C. Alors la cardinalité de C vérifie :

$$|C| \ \geq \ \log\left(\sum_{i=0}^{\ell} \binom{|V|}{i}\right) \ = \ \Omega(\ell \log |V|).$$

Preuve : Les ensembles identifiants $I(X, C)$ étant des sous-ensembles distincts de C, l'application $X \mapsto I(X, C)$ est une injection des sous-ensembles d'au plus ℓ sommets de G dans 2^C. Ceci nous donne

$$\sum_{i=0}^{\ell} \binom{|V|}{i} \ \leq \ 2^{|C|},$$

ce qui conduit au résultat annoncé en considérant que $\sum_{i=0}^{\ell} \binom{|V|}{i} = \Theta(|V|^{\ell})$. □

On peut remarquer que la borne du théorème précédent ne dépend pas de t. Dans le cas $\ell = 1$, cette borne nous dit que $|C| \geq \lceil \log(|V| + 1) \rceil$. Dans le Chapitre 4 nous montrons que cette borne est serrée : pour tout $t \geq 1$, et pour tout n assez grand, nous construisons un graphe $G_{n,t}$ à n sommets admettant un code t-identifiant de cardinalité $\lceil \log(n + 1) \rceil$ (Théorèmes 4.1 et 4.2).

Le lien que nous avons établi avec les codes superimposés nous permet en fait d'améliorer le Théorème 1.4. Rappelons que $K \subseteq \{0, 1\}^N$ est un code ℓ-superimposé si et seulement si le `OU` binaire d'au plus ℓ vecteurs de K est différent du `OU` binaire d'au plus ℓ autres vecteurs de K. Les codes superimposés ont été largement étudiés dans la littérature, et nous disposons des bornes suivantes :

Théorème 1.5 ([DR83, KS64])
Soit K un code ℓ-superimposé de $\{0,1\}^N$. Alors il existe deux constantes c_1 et c_2, indépendantes de n et ℓ, telles que :

$$2^{c_1 N/\ell^2} \leq |K| \leq 2^{c_2 N \log \ell/\ell^2}.$$

De plus, la borne inférieure est constructive : il existe un algorithme qui, étant donné un entier N, construit un code ℓ-superimposé de $\{0,1\}^N$ de cardinalité $2^{c_1 N/\ell^2}$.

La borne inférieure était déjà donnée dans l'article introduisant les codes superimposés [KS64], et des preuves combinatoires de la borne supérieure, établie à l'origine en [DR83], peuvent être trouvées dans [Rusz94, Fur96]. Un algorithme glouton construisant un code ℓ-superimposé de cardinalité $2^{c_1 N/\ell^2}$ est décrit en [HS87].

Dans le cas où l'on identifie des ensembles de sommets (paragraphe 1.2.2), nous avions mentionné le fait qu'un graphe muni d'un code $(t, \leq \ell)$-identifiant nous fournissait un code ℓ-superimposé de $\{0,1\}^{|C|}$. Ce lien nous permet de déduire une nouvelle borne sur la cardinalité minimum de C :

Théorème 1.6 ([FMMRS])
Soit $G = (V, E)$ un graphe muni d'un code $(t, \leq \ell)$-identifiant C. Alors la cardinalité de C vérifie :

$$|C| \geq \Omega\left(\frac{\ell^2}{\log \ell} \log n\right).$$

Preuve : Soit G un graphe muni d'un code $(t, \leq \ell)$-identifiant C. Les vecteurs caractéristiques des ensembles identifiants des sommets de G formant un code ℓ-superimposé de $\{0,1\}^{|C|}$, le résultat découle du Théorème 1.5. □

Nous ignorons si cette borne est serrée. Dans le Chapitre 4, nous donnons une construction explicite d'une famille de graphes à n sommets admettant un code $(1, \leq \ell)$-identifiant de cardinalité $O(\ell^4 \log n)$ (Théorème 4.3), et dans le Chapitre 5 nous donnons une construction probabiliste montrant qu'il existe une famille de graphes à n sommets admettant un code $(1, \leq \ell)$-identifiant de cardinalité $O(\ell^2 \log n)$ (Proposition 5.5).

Dans le cas des graphes réguliers, nous pouvons donner une borne en fonction de la cardinalité des boules :

Théorème 1.7 ([KCL98])
Soit G un graphe régulier à n sommets, et C un code t-identifiant de G. Soit

$V(t)$ la cardinalité d'une boule de rayon t dans G. Alors la cardinalité de C vérifie :

$$|C| \ \geq \ \frac{2n}{V(t)+1}.$$

Preuve : Soit G un graphe muni d'un code t-identifiant C, et soit H le graphe biparti dont l'un des stables est C et l'autre V, et tel qu'il y a une arête entre $c \in C$ et $v \in V$ si et seulement si $c \in B_t(v)$ dans G. On conclut par un argument classique de double-comptage : le nombre d'arêtes de H est, d'une part, trivialement égal à $|C|V(t)$, et d'autre part il y a au plus $|C|$ sommets de V de degré 1 et au moins $n - |C|$ sommets de V de degré au moins 2, d'où $|C|V(t) \geq 2(n - |C|) + |C|$, *i.e.* $|C| \geq \frac{2n}{V(t)+1}$. □

1.4 Notations utilisées

Sauf mention contraire explicite, les graphes considérés seront non-orientés. Les problèmes d'identification d'ensembles d'au plus ℓ sommets à distance $t \geq 1$ seront considérés. Nous les désignerons par *codes $(t, \leq \ell)$-identifiants.* Pour plus de simplicité, le terme code t-identifiant sera utilisé à la place de code $(t, \leq 1)$-identifiant.

Dans cette thèse nous présentons essentiellement des résultats personnels. La référence de l'article ou de l'ouvrage correspondant est donnée pour chaque résultat. Lorsqu'aucune référence n'est donnée c'est qu'il s'agit d'un résultat personnel non publié.

La fin des preuves est marquée par le signe □. L'énoncé d'un résultat donné sans preuve sera immédiatement suivi de □.

Le vocabulaire et les notions de base de la théorie des graphes sont donnés en annexe à la fin de ce document. Les notions utilisées sont élémentaires, et l'annexe a pour principale fonction de fixer les notations.

Lorsque aucune base n'est spécifiée, $\log x$ désigne le logarithme de x en base 2, qui est la base "naturelle" en théorie de l'information et en théorie des codes.

Les notations o, O, Ω et Θ seront utilisées de la façon conventionnelle : on dira que $f = o(g)$ si $f(n)/g(n) \to \infty$ lorsque n tend vers l'infini ; $f = O(g)$ s'il existe une constante c telle que $f(n) \leq cg(n)$ pour tout n assez grand ; $f = \Omega(g)$ s'il existe une constante c' telle que $f(n) \geq c'g(n)$ pour tout n assez grand ; et enfin $f = \Theta(g)$ si à la fois $f = O(g)$ et $f = \Omega(g)$.

Chapitre 2

Aspects algorithmiques

Dans ce chapitre, nous abordons l'aspect algorithmique des codes identifiants dans les graphes. Le problème d'optimisation associé est le suivant :

ID-Code(t,ℓ) :
Instance : Un graphe G admettant un code $(t, \leq \ell)$-identifiant.
Question : Quelle est la cardinalité minimum d'un code $(t, \leq \ell)$-identifiant de G ?

Pour $\ell = 1$, il a été montré dans [CHL03] que ce problème était NP-difficile pour tout $t \geq 1$, et ce même lorsque l'on se restreignait aux graphes bipartis. Ceci implique que ce problème est NP-difficile pour tout $\ell \geq 1$.

Dans le cas des graphes orientés, le problème est aussi NP-difficile pour tout $\ell \geq 1$ et tout $t \geq 1$ [CHL02b]. La complexité de problèmes connexes dans l'hypercube a été étudiée en [HL02a, HL02c].

Nous étudions ce problème sur deux classes restreintes de graphes : les fasciagraphes (qui sont une généralisation des grilles) et les arbres orientés. Pour ces deux classes de graphes nous fournissons, sous certaines hypothèses, des algorithmes polynomiaux très performants, puisque fonctionnant, respectivement, en temps constant et linéaire.

Dans ces deux cas, nous résolvons non seulement le problème de la détermination de la cardinalité minimum d'un code identifiant, mais également celui de la *construction* d'un code identifiant de cardinalité minimum.

La notion de largeur d'arborescence (*tree-width* en anglais) a été introduite dans les années quatre-vingts par N. Robertson et P. D. Seymour pour

résoudre la conjecture de Wagner[1]. La largeur d'arborescence d'un graphe G est un entier, supérieur ou égal à 0, qui exprime la complexité structurelle de G par rapport à celle d'un arbre — les arbres ayant une largeur d'arborescence inférieure ou égale à 1. Nous ne souhaitons pas donner ici plus de détails sur la largeur d'arborescence, au sujet de laquelle il existe de nombreux articles de *survey*, tels [Bod93] ou encore [Die00, Chapitre 12]. Cette notion est néanmoins très importante, elle a notamment de nombreuses conséquences algorithmiques, parmi lesquelles :

Théorème 2.1 ([ALS91])
Soit Π *un problème de décision exprimable dans le langage MS, qui est le langage obtenu en utilisant :*

- *des quantifications* $\forall$ *et* $\exists$ *sur des sommets, des arêtes, ou sur des ensembles de sommets ou d'arêtes du graphe,*
- *les opérateurs logique* `ET`*,* `OU`*,* `NON`*,*
- *la relation d'adjacence des sommets.*

Soit $\mathcal{K}$ *une classe de graphes de largeur d'arborescence bornée par une constante. Alors il existe un algorithme linéaire résolvant le problème* Π *sur la classe de graphes* $\mathcal{K}$*. De plus, tout problème d'optimisation consistant à maximiser ou minimiser la cardinalité d'un ensemble de sommets ou d'arêtes satisfaisant des conditions exprimables dans le langage MS, est linéaire sur une telle classe de graphes* $\mathcal{K}$*.* □

Or, étant donné un graphe G et un entier $t \geq 1$, il est très facile d'exprimer la condition "C est un code t-identifiant de G" dans le langage décrit ci-dessus. Par exemple, pour $t = 1$, on peut écrire :

$\exists C \subseteq V(T)$,
  $\forall x \in V(T), x \in C$ `OU` $\exists y \in C, xy \in E(G)$
  `ET`
  $\forall x \in V(G), \forall y \in V(G)$,
    $x = y$
    `OU`
    $\exists z \in C, (xz \in E(G)$ `ET NON` $yz \in E(G))$ `OU` $(yz \in E(G)$ `ET NON` $xz \in E(G))$

Puisque les arbres (orientés ou non-orientés) ont une largeur d'arborescence bornée (par 1), alors le Théorème 2.1 implique donc directement que la

[1]K. Wagner conjectura en 1970 que toute séquence infinie de graphes en contient nécessairement deux dont l'un est un mineur de l'autre [Wag70, page 61]. Nous rappelons que H est un *mineur* de G si H peut être obtenu à partir de G par une séquence de délétions de sommets et de contractions d'arêtes de G. Ce résultat a été démontré par N. Robertson et P. D. Seymour [RS04]. D'aucuns, tels L. Lovász [Lov98], estiment que ce résultat est l'un des plus profonds en théorie des graphes.

recherche de la cardinalité minimum d'un code t-identifiant est linéaire dans les arbres. Par rapport à ce résultat, l'algorithme *ad hoc* présenté dans la suite (pour $t = 1$) a l'avantage de donner *explicitement* un code 1-identifiant de cardinalité minimum en temps linéaire, ce que ne fait pas l'algorithme issu du théorème précédent.

En ce qui concerne les grilles (et les fasciagraphes), nous donnons un algorithme en temps constant, ce qui est bien évidemment meilleur que ce que nous donne le théorème précédent (temps linéaire).

2.1 Grilles et fasciagraphes

La grille bidimensionnelle de taille $k \times n$ est le graphe ayant

$$\{0, \ldots, k-1\} \times \{0, \ldots, n-1\}$$

pour ensemble de sommets et

$$\{uv \mid |u_1 - v_1| + |u_2 - v_2| = 1\}$$

pour ensemble d'arêtes.

Nous donnons des bornes générales sur la cardinalité d'un code identifiant dans les grilles dans le Chapitre 3, paragraphe 3.2.

Lorsque k est fixé, nous avons donné dans [DGM04] un algorithme logarithmique en n, calculant la cardinalité minimum d'un code t-identifiant de la grille bidimensionnelle de taille $k \times n$. Dans cette section, nous donnons une nouvelle présentation de ce résultat, que nous améliorons significativement : nous donnons un algorithme qui retourne en temps constant une *formule* pour la cardinalité minimum d'un code t-identifiant de la grille bidimensionnelle de taille $k \times n$, pour tous k et t fixés. L'algorithme retourne également une formule décrivant un code t-identifiant de cardinalité minimum, toujours en temps constant si k et t sont fixés.

Les fasciagraphes sont une classe de graphes généralisant la grille, que nous définissons dans le paragraphe suivant. Tous les résultats que nous venons de mentionner sont également vrais pour les fasciagraphes, et c'est dans ce cadre que nous allons présenter notre formulation du problème et les résultats qui en découlent.

Nous commençons donc par définir les fasciagraphes, puis nous donnons une formulation du problème dans ce cadre plus général. De notre formulation découlent des résultats de pseudo-périodicité structurelle des codes

t-identifiants dans les fasciagraphes qui entraînent les résultats annoncés. Dans le dernier paragraphe nous discutons de l'extension de ces résultats à d'autres types de problèmes et à d'autres classes de graphes.

2.1.1 Fasciagraphes

Soit G un graphe et soit X un ensemble d'arêtes entre deux copies de G. Soit n un entier positif. Le *rotagraphe* d'ordre n et de *fibre* G est le graphe $\rho_n(G, X)$ constitué de n copies $G_1, \ldots, G_n$ de G, telles que deux copies consécutives G_i et G_{i+1} sont jointes par l'ensemble d'arêtes X, pour $i = 1, \ldots, n$, les indices étant considérés modulo n (voir Figure 2.1). Il n'y a pas d'arêtes entre deux copies G_i et G_j si $|i - j| \neq 1$. Un rotagraphe $\rho_n(G, X)$ privé des arêtes entre G_1 et G_n est appelé *fasciagraphe*, on le note $\Psi_n(G, X)$ (voir Figure 2.1).

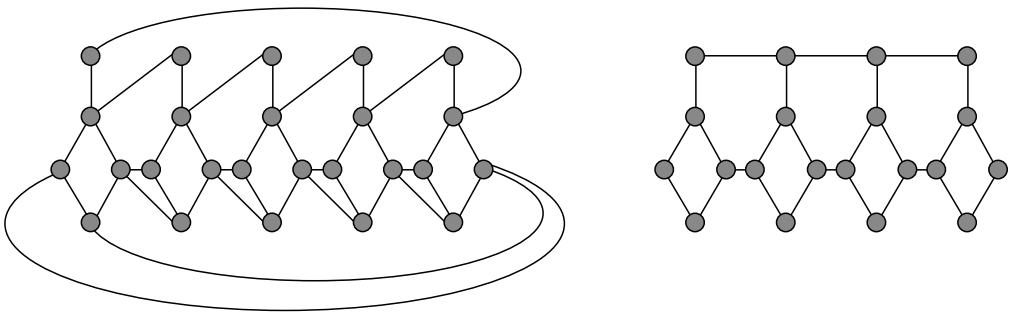


FIG. 2.1 – *Exemples de rotagraphe (à gauche) et de fasciagraphe (à droite), d'ordres 5 et 4 respectivement.*

Les fasciagraphes et les rotagraphes ont été à l'origine définis pour modéliser des molécules en chimie des polymères [BGMP86]. On peut voir la grille bidimensionnelle de taille $k \times n$ comme un fasciagraphe $\Psi_n(P_k, X)$, où P_k est le chemin sur k sommets. Les fasciagraphes généralisent les grilles dans le sens où ils peuvent être construits inductivement par ajouts successifs de fibres (équivalents aux colonnes de la grille) jointes par des ensembles d'arêtes (équivalentes à la structure "en ligne" des grilles).

2.1.2 Reformulation du problème

Nous montrons que la recherche de la cardinalité minimum d'un code t-identifiant dans un fasciagraphe $\Psi_n(G, X)$ est équivalente à la recherche d'un chemin de poids minimum dans un graphe orienté $\Gamma(G, X)$. Pour la simplicité de l'exposé, nous présentons cette équivalence dans le cas des codes 1-identifiants, et nous espérons que notre construction sera assez claire pour que le lecteur en déduise la construction dans le cas général $t \geq 2$. À la fin de ce paragraphe nous donnerons les éléments essentiels permettant d'adapter cette construction au cas général.

Étant donné un fasciagraphe $\Psi_n(G, X)$, soit $\Gamma(G, X)$ le graphe orienté pondéré défini de la façon suivante :

- les sommets de $\Gamma(G, X)$ sont tous les couples $(\Psi_4(G, X), C)$, où C est un sous-ensemble de sommets de $\Psi_4(G, X)$ qui 1-sépare et 1-couvre tous les sommets de G_2 et G_3, les deux fibres centrales de $\Psi_4(G, X)$. Un sommet de $\Gamma(G, X)$ peut être vu comme un $\Psi_4(G, X)$ muni d'un code qui 1-identifie les sommets des deux colonnes centrales de $\Psi_4(G, X)$.
- $\Gamma(G, X)$ comporte de plus deux sommets additionnels D et F, qui sont appelés, respectivement, sommets *début* et *fin*,
- il y a un arc entre deux sommets $(\Psi_4(G, X), C)$ et $(\Psi_4(G, X), C')$ s'ils sont *compatibles*, dans le sens où (voir Figure 2.2) :
  - il existe C'' un code qui 1-identifie les sommets des trois colonnes centrales de $\Psi_5(G, X)$,
  - C peut être vu comme la trace de C'' sur les quatre premières colonnes de $\Psi_5(G, X)$,
  - C' peut être vu comme la trace de C'' sur les quatre dernières colonnes de $\Psi_5(G, X)$.
- il y a un arc entre D et tout sommet $(\Psi_4(G, X), C)$ tel que C est un code qui 1-identifie les sommets des trois premières fibres de $\Psi_4(G, X)$,
- il y a un arc entre $(\Psi_4(G, X), C)$ et F si et seulement si C est un code qui 1-identifie les sommets des trois dernières fibres de $\Psi_4(G, X)$,
- le poids d'un arc entre $(\Psi_4(G, X), C)$ et $(\Psi_4(G, X), C')$ est le nombre de sommets de C' appartenant à la dernière fibre de $\Psi_4(G, X)$,
- le poids d'un arc entre D et $(\Psi_4(G, X), C)$ est le nombre de sommets de C,
- les arcs entrants en F ont un poids nul.

Le poids d'un arc (U, V) sera noté $\omega(U, V)$. On étend cette pondération

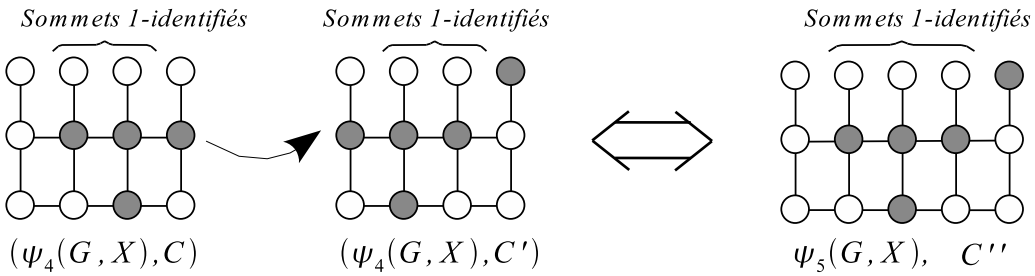


FIG. 2.2 – *Exemple d'arc dans* $\Gamma(G,X)$.

aux chemins de $\Gamma(G,X)$ de la façon suivante :

$$\omega(P) = \sum_{i=1}^{k-1} \omega(U_i, U_{i+1})$$

pour tout chemin $P = (U_1, \ldots, U_k)$ de $\Gamma(G,X)$. Il existe alors une correspondance naturelle entre les codes 1-identifiants de $\Psi_n(G,X)$ et les chemins dans $\Gamma(G,X)$:

Théorème 2.2 ([DGM04])
Pour tout $n \geq 4$*, il existe une correspondance biunivoque* φ *entre l'ensemble des codes* 1*-identifiants de* $\Psi_n(G,X)$ *et l'ensemble des chemins de* $n-2$ *arcs allant de* D *à* F *dans* $\Gamma(G,X)$*. De plus, cette correspondance est telle que* $|C| = \omega(P)$ *pour tout code* 1*-identifiant* C *et tout chemin* P *tels que* $\varphi(C) = P$*.*

Preuve : Par récurrence sur n, ce résultat découle facilement des définitions de $\Gamma(X,G)$ et ω (voir Figure 2.3). □

La recherche de la cardinalité minimum d'un code 1-identifiant de $\Psi_n(G,X)$ revient donc à chercher un chemin de poids minimum de $n-2$ arcs entre D et F. Dans le paragraphe suivant nous allons voir comment, à l'aide de la programmation dynamique, nous pouvons résoudre ce problème en temps constant lorsque la taille de G est fixée. Ceci provient du fait que la taille de $\Gamma(G,X)$ est bornée par une constante lorsque celle de G est fixée. La méthode utilisée nous permet de plus de *construire* un code 1-identifiant de cardinalité minimum.

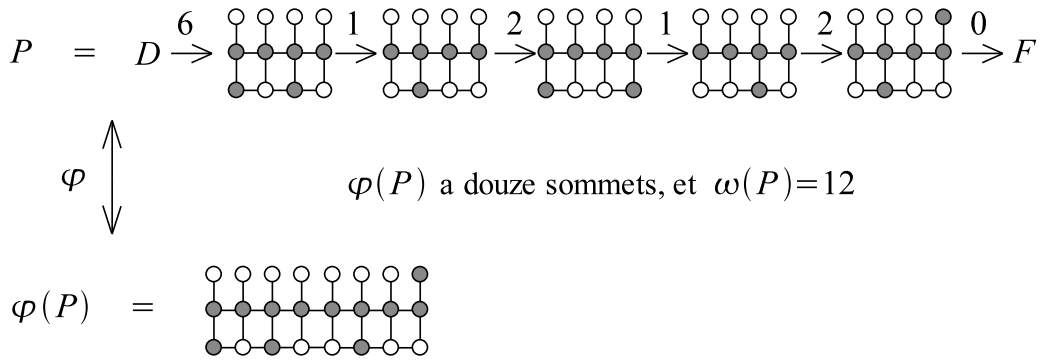


FIG. 2.3 – *Un chemin sur 6 arcs de D à F dans $\Gamma(G,X)$ correspond à un code 1-identifiant de $\Psi_8(G,X)$.*

2.1.3 Programmation dynamique

Dans ce paragraphe soit G un graphe de taille fixée et soit $\Psi_n(G,X)$ un fasciagraphe de fibre G tel que, pour n assez grand, $\Psi_n(G,X)$ admette un code 1-identifiant. Pour déterminer si $\Psi_n(G,X)$ admet un code 1-identifiant il faut et il suffit de vérifier que $\Psi_5(G,X)$ admet un code 1-identifiant, ce qui se fait en temps constant. Soient $\sigma_1,\ldots,\sigma_K$ les sommets de $\Gamma(G,X)$; on a $K \leq 2^{4|G|}$: le nombre de sommets de $\Gamma(G,X)$ est borné par une constante indépendante de n.

2.1.3.1 Formulation matricielle

Pour tout $n \geq 1$, nous définissons deux matrices $M^{(n)}$ et $N^{(n)}$ telles que :
- $M^{(n)}_{i,j}$ contient le poids minimum d'un chemin sur n arcs de σ_i à σ_j dans $\Gamma(G,X)$ ($M^{(n)}_{i,j} = +\infty$ si un tel chemin n'existe pas),
- $N^{(n)}_{i,j}$ contient le k minimum tel que σ_k soit l'antécédent de σ_j dans un chemin sur n arcs de poids minimum de σ_i à σ_j dans $\Gamma(G,X)$ ($N^{(n)}_{i,j} = 0$ si un tel chemin n'existe pas).

Les matrices $M^{(n+1)}$ et $N^{(n+1)}$ peuvent être obtenues récursivement à partir de $M^{(n)}$ par des fonctions f et g telles que :

$$M^{(n+1)} = f(M^{(n)}) \quad \text{et} \quad N^{(n+1)} = g(M^{(n)}) \tag{2.1}$$

Par exemple, pour n assez grand, on peut calculer $M^{(n+1)}$ de la façon suivante :

$$M_{i,j}^{(n+1)} \;=\; \min_{\sigma_k \in Pred(\sigma_j)} \{M_{i,k}^{(n)} + \omega(k,j)\}$$

pour tout $i, j = 1, \ldots, K$, où $Pred(\sigma_j)$ désigne l'ensemble des prédécesseurs de σ_j dans $\Gamma(G,X)$: $Pred(\sigma_j) = \{\sigma_k \mid \sigma_k\sigma_j \text{ arc de } \Gamma(G,X)\}$. De même, pour n assez grand, $N^{(n+1)}$ ne dépend que de $M^{(n)}$.

La programmation dynamique consiste à partir d'une matrice initiale $M^{(n_0)}$ et à itérer les fonctions f et g pour calculer $M^{(n)}$ et $N^{(n)}$. Pour obtenir un chemin de n arcs de poids minimum dans $\Gamma(G,X)$, il suffit alors de considérer $M_{i_0,j_0}^{(n)}$, où i_0 et j_0 sont les indices correspondant aux sommets D et F respectivement. Les sommets du chemin sont obtenus récursivement à partir des matrices $N^{(n)}$.

Comme les matrices $M^{(n)}$ et $N^{(n)}$ sont de taille bornée par une constante, ceci est réalisable en temps linéaire en n. Dans [DGM04] nous avions montré que nous pouvions en fait réaliser ceci en temps logarithmique, et dans la suite nous montrons que nous pouvons en fait réduire au temps constant.

Remarque : Le choix des notations $M^{(n)}$ et $N^{(n)}$ n'est pas innocent : on peut montrer que $M^{(n)}$ est la puissance n-ième de $M^{(1)}$ dans un certain semi-anneau idempotent $\mathcal{M}_K(\mathbb{N} \cup \{\infty\}, \min, +, \infty, 0)$, appelé parfois semi-anneau *tropical* (voir [KZ96, Zer99, Zer] pour la définition d'un semi-anneau idempotent).

2.1.3.2 Propriétés des matrices $M^{(n)}$ et $N^{(n)}$

Soit J la matrice ne comportant que des 1. Alors les fonctions f et g satisfont la propriété suivante :

$$f(M + cJ) \;=\; f(M) + cJ \quad \text{et} \quad g(M + cJ) \;=\; g(M) \text{ pour tout } c \in \mathbb{Z} \quad (2.2)$$

Autrement dit, si l'on définit

$$\widetilde{M}^{(n)} \;:=\; M^{(n)} - \alpha_n J,$$

où

$$\alpha_n \;=\; \min_{i,j=1,\ldots,K} M_{i,j}^{(n)},$$

alors pour n assez grand on a :

$$M^{(n+1)} \;=\; f(\widetilde{M}^{(n)}) + \alpha_n J \quad \text{et} \quad N^{(n+1)} \;=\; g(\widetilde{M}^{(n)}) \qquad (2.3)$$

Ceci implique :

Lemme 2.1
Soient $p > q$ tels que $\widetilde{M}^{(p)} = \widetilde{M}^{(q)}$, et soit $T = p - q$. Alors il existe $c \in \mathbb{Z}$ tel que pour tout $a, b \geq 0$ on ait

$$M^{(a+bT)} \;=\; M^{(a)} + bcJ \quad \textit{et} \quad N^{(a+bT)} \;=\; N^{(a)} \tag{2.4}$$

Preuve : Soient $p > q$ tels que $\widetilde{M}^{(p)} = \widetilde{M}^{(q)}$, et soit $T = p - q$. On a alors $M^{(q+T)} = M^{(q)} + (\alpha_p - \alpha_q)J$. En utilisant (2.1) et (2.2) on montre par récurrence que $M^{(a+bT)} = M^{(a)} + bcJ$ pour tout $a, b \geq 0$, ce qui implique $N^{(a+bT)} = N^{(a)}$ par (2.2). □

Dans la suite nous montrons que les hypothèses du Lemme précédent sont satisfaites pour un certain p et un certain q qui sont bornés par une constante. Ceci implique la résolution du problème en temps constant : on calcule $M^{(n)}$ jusqu'à trouver p, q et c tels que

$$M^{(p)} = M^{(q)} + cJ \tag{2.5}$$

Les formules pour la cardinalité minimum d'un code 1-identifiant ou décrivant un code 1-identifiant de cardinalité minimum sont obtenues à partir de (2.4) et (2.5) (voir paragraphe 2.1.4).

2.1.3.3 Existence d'une pseudo-période

Dans ce paragraphe nous montrons que les hypothèses du Lemme 2.1 sont satisfaites pour un p et un q bornés par une constante.

Lemme 2.2
Soit K le nombre de sommets de $\Gamma(G, X)$. Alors pour tout n assez grand, deux éléments de $M^{(n)}$ diffèrent d'au plus $4K$.

Preuve : Soit $\kappa(n)$ la cardinalité minimum d'un sous-ensemble de $\Psi_n(G, X)$ qui 1-sépare et 1-couvre tous les sommets des $n - 1$ premières fibres de $\Psi_n(G, X)$, et soit $\nu(n)$ la cardinalité minimum d'un code 1-identifiant de $\Psi_n(G, X)$. Alors on a

$$\kappa(n) \;\leq\; \nu(n) \;\leq\; \kappa(n) + 4K.$$

En effet, un code 1-identifiant de $\Psi_n(G, X)$ est en particulier un sous-ensemble de $\Psi_n(G, X)$ qui 1-sépare et 1-couvre tous les sommets des $n - 1$ premières fibres de $\Psi_n(G, X)$, d'où la première inégalité. Pour la deuxième inégalité, il

suffit de voir que l'union des sommets des quatre dernières fibres de $\Psi_n(G,X)$ et d'un code qui 1-sépare et 1-couvre tous les sommets des $n-1$ premières fibres de $\Psi_n(G,X)$ est un code 1-identifiant de $\Psi_n(G,X)$. Ces inégalités entraînent le résultat annoncé. □

Ceci a pour conséquence que les coefficients de la matrice $\widetilde{M}^{(n)}$ sont à valeur dans $\{0,\ldots,4K\}$. Comme $N^{(n)}$ a ses coefficients à valeur dans $\{1,\ldots,K\}$, ceci entraîne :

Proposition 2.1
Il existe $T \leq [K(4K+1)]^{K^2}$ et une constante c tels que

$$M^{(a+bT)} \;=\; M^{(a)} + bcJ$$

et

$$N^{(a+bT)} \;=\; N^{(a)}$$

pour tout $a,b \in \mathbb{N}$.

Preuve : Par le principe dit des cages à pigeons, il y a $(4K+1)^{K^2}$ valeurs possibles de $\widetilde{M}^{(n)}$ et K^{K^2} valeurs possibles de $N^{(n)}$, donc parmi $(4K+1)^{K^2} \times K^{K^2} + 1 = [K(4K+1)]^{K^2} + 1$ couples de matrices $(\widetilde{M}^{(n)}, N^{(n)})$ il y en a au moins deux de même valeur. On conclut en utilisant le Lemme 2.1. □

2.1.4 Résultats

Dans ce paragraphe soient G et X fixés et soit $\nu(n)$ la cardinalité minimum d'un code 1-identifiant de $\Psi_n(G,X)$. Soit $\mathbb{A}$ l'alphabet dont les lettres sont les fibres de $\Psi_n(G,X)$ ayant leurs sommets étiquetés par 0 ou 1. On désigne par $\mathbb{A}^*$ l'ensemble des mots finis sur l'alphabet $\mathbb{A}$, et pour tout $\alpha \in \mathbb{A}$ et pour tout $n \in \mathbb{N}$, α^n désigne la concaténation de n copies de α. Tout code 1-identifiant de $\Psi_n(G,X)$ correspond à un unique mot de n lettres de $\mathbb{A}$. Alors on a :

Théorème 2.3
Soient G et X fixés. Alors il existe $T, \nu, \nu_0, \ldots, \nu_{T-1} \in \mathbb{N}$ tels que, pour tout n assez grand, on ait

$$\nu(n) \;=\; \nu_i + \nu\left\lfloor \frac{n}{T} \right\rfloor$$

si $n \equiv i \bmod T$.
De plus, il existe $\zeta_0,\ldots,\zeta_{T-1},\theta_0,\ldots,\theta_{T-1} \in \mathbb{A}^*$ tels que, pour tout $i = 0,\ldots,T-1$, ζ_i ait i mots et θ_i ait T mots, et, pour tout n assez grand

$$\zeta_i(\theta_i)^{\lfloor \frac{n}{T} \rfloor}$$

corresponde à un code 1*-identifiant de* $\Psi_n(G,X)$ *de cardinalité minimum, avec* $n \equiv i \ mod\ T$.

Preuve : D'après le Théorème 2.2, et en utilisant les notations des paragraphes précédents, on sait qu'il existe i_0 et j_0 tels que pour tout n assez grand on ait

$$\nu(n) \ = \ M_{i_0,j_0}^{(n-2)}.$$

Le résultat annoncé découle de la Proposition 2.1 appliquée avec $a = n \bmod T$ et $b = \lfloor \frac{n}{T} \rfloor$. La formule $\zeta_i(\theta_i)^{\lfloor \frac{n}{T} \rfloor}$ est obtenue récursivement à partir des matrices $N^{(n)}$ (on commence avec $N_{i_0,j_0}^{(n-2)}$). □

Théorème 2.4
Soient G *et* X *fixés et soit* $\nu(n)$ *la cardinalité minimum d'un code* 1*-identifiant de* $\Psi_n(G,X)$*. Il existe un algorithme qui, en temps constant, détermine une formule close pour* $\nu(n)$*, de la forme*

$$\nu(n) \ = \ \nu_i + \nu \left\lfloor \frac{n}{T} \right\rfloor$$

si $n \equiv i \ mod\ T$*, avec* T *constante ne dépendant que de* G *et* X*. De plus l'algorithme fournit une formule close de la forme*

$$\zeta_i(\theta_i)^{\lfloor \frac{n}{T} \rfloor}$$

si $n \equiv i \ mod\ T$*, décrivant un code* 1*-identifiant de* $\Psi_n(G,X)$ *de cardinalité minimum.*

Preuve : L'algorithme est le suivant : on part de deux matrices $M^{(n_0)}$ et $N^{(n_0)}$, auxquelles on applique f et g de sorte à calculer $M^{(n_0+1)}, M^{(n_0+2)}, \ldots$ et $N^{(n_0+1)}, N^{(n_0+2)}, \ldots$, jusqu'à un rang p tel qu'il existe un rang $q < p$ tel que la matrice $M^{(p)} - M^{(q)}$ soit proportionnelle à J (matrice dont tous les coefficients sont égaux à 1) et $N^{(p)} = N^{(q)}$. D'après la Proposition 2.1 on sait que l'on va rencontrer un tel p en temps constant, et par le Lemme 2.1 on sait que ceci entraîne l'existence d'un code 1-identifiant pseudo-périodique de $\Psi_n(G,X)$. La description d'un tel code et de sa cardinalité se fait en temps constant à partir des matrices $M^{(n_0)}, \ldots, M^{(p)}$ et $N^{(n_0)}, \ldots, N^{(p)}$ comme dans le Théorème 2.3. □

2.1.5 Extensions des résultats

La technique décrite ci-dessus est beaucoup plus générale et peut être appliquée à d'autres types de problèmes. On peut par exemple montrer le résultat suivant :

Théorème 2.5
Soient G et X fixés et soit $t \geq 1$. Soit $\nu^t(n)$ la cardinalité minimum d'un code t-identifiant de $\Psi_n(G,X)$. Alors il existe un algorithme qui, en temps constant, détermine une formule close pour $\nu^t(n)$, de la forme

$$\nu^t(n) \;=\; \nu_i + \nu \left\lfloor \frac{n}{T} \right\rfloor$$

si $n \equiv i \bmod T$, avec T constante ne dépendant que de G, X et t. De plus l'algorithme fournit une formule close de la forme

$$\zeta_i(\theta_i)^{\left\lfloor \frac{n}{T} \right\rfloor}$$

si $n \equiv i \bmod T$, décrivant un code t-identifiant de $\Psi_n(G,X)$ de cardinalité minimum.

Idée de Preuve : On construit un graphe auxiliaire orienté pondéré $\Gamma_t(G,X)$, dont les sommets sont des $\Psi_{2t+2}(G,X)$ munis d'un code qui t-identifie les sommets des $2t$ colonnes centrales de $\Psi_{2t+2}(G,X)$. On ajoute deux sommets *début* et *fin* D et F, et on pondère les arcs de façon similaire. Comme dans le Théorème 2.2 il y a une correspondance biunivoque entre les chemins de $n-2$ arcs de D à F et les codes t-identifiants de $\Psi_n(G,X)$, telle que le poids d'un tel chemin corresponde au nombre de sommets du code t-identifiant correspondant. On résout le problème de chemins par la programmation dynamique : de la même façon que dans le Lemme 2.2, les coefficients de la matrice des plus courts chemins diffèrent d'au plus $(2t+2)K$, ce qui implique l'existence d'un code t-identifiant pseudo-périodique comme dans la Proposition 2.1 et le Théorème 2.3.
Ce résultat n'est pas surprenant si l'on remarque que, pour tout $t \geq 1$, C est un code t-identifiant de $\Psi_n(G,X)$ si et seulement si c'est un code t-identifiant de $\Psi_n(G,X)^t$, la fermeture t-transitive de $\Psi_n(G,X)$. Or, pour un entier n multiple de t, la fermeture t-transitive d'un fasciagraphe est encore un fasciagraphe, de fibre $\Psi_t(G,X)$: $\Psi_n(G,X)^t = \Psi_{n/t}(\Psi_t(G,X),X')$. Dans ce cas on peut directement appliquer le Théorème 2.4. □

Cette méthode est également valide pour des structures plus générales que les fasciagraphes. On peut par exemple faire la même construction pour calculer la cardinalité minimum d'un code t-identifiant d'un rotagraphe :

Théorème 2.6
Soient G et X fixés et soit $t \geq 1$. Soit $\mu^t(n)$ la cardinalité minimum d'un code t-identifiant du rotagraphe $\rho_n(G,X)$. Alors il existe un algorithme qui, en temps constant, détermine une formule close pour $\mu^t(n)$, de la forme

$$\mu^t(n) \;=\; \mu_i + \mu \left\lfloor \frac{n}{T} \right\rfloor$$

si $n \equiv i \bmod T$, avec T constante ne dépendant que de G, X et t. De plus l'algorithme fournit une formule close de la forme

$$\zeta_i(\theta_i)^{\lfloor \frac{n}{T} \rfloor}$$

si $n \equiv i \bmod T$, décrivant un code t-identifiant de $\rho_n(G, X)$ de cardinalité minimum.

Preuve : On définit $\Gamma_t(G, X)$, $M^{(n)}$ et $N^{(n)}$ comme précédemment. Tout se passe comme dans le théorème précédent, si ce n'est que $\mu^t(n)$ est obtenu de la façon suivante :

$$\mu^t(n) \ = \ \min_{i=1,\ldots,K} M_{i,i}^{(n-1)}$$

En effet, il est facile de voir qu'un circuit sur $n-1$ arcs dans $\Gamma_t(G, X)$ correspond à un code t-identifiant dans $\rho_n(G, X)$. □

Il est également facile de généraliser tous ces résultats au cas des graphes orientés.

Dans [KZ96, LS94, Zer99] cette méthode est appliquée au problème de domination, et dans [KV03] une équivalence similaire à celle du Théorème 2.2 permet de donner des résultats pour les (2, 1)-coloriages ainsi que pour le nombre de stabilité de produits de cycles. Dans [JMGSZ95] cette méthode est appliquée pour calculer l'index de Wiener des fasciagraphes et des rotagraphes.

Dans [GMP] nous donnions une généralisation du problème d'exclusion des pentominos de S. W. Golomb [Gol94], à laquelle nous avons appliqué cette méthode pour obtenir un algorithme en temps logarithmique. J. Žerovnik améliore notre résultat [Zer] en utilisant un lemme de pseudo-périodicité similaire à celui de la Proposition 2.1.

Comme les auteurs de tous ces articles, nous prétendons que notre méthode fonctionne pour une très large classe de problèmes combinatoires, et nous espérons pouvoir un jour donner une caractérisation simple des problèmes pouvant être résolus par cette méthode.

Nous donnons ici quelques questions similaires à celles de [Sto] qui nous semblent pertinentes :

Question 2.1 *À quels types de problèmes et à quels types de familles de graphes $\mathcal{F}_n$ peut-on adapter la méthode décrite au-dessus ?*

Question 2.2 *Dans le cas de la grille multidimensionnelle de dimension d, existe-t-il aussi une formule close pour la cardinalité minimum d'un code*

t-identifiant, de la forme

$$\nu_t(n_1, \ldots, n_d) \;=\; f_{i_2,\ldots,i_d}(n_2, \ldots, n_d)$$

où, pour tout $(i_2, \ldots, i_d) \in \{0, \ldots, T_2 - 1\} \times \ldots \times \{0, \ldots, T_d - 1\}$, $f_{i_2,\ldots,i_d}$ *est une fonction calculable, avec* $T_2, \ldots, T_d$ *constantes ne dépendant que de* n_1 *?*

2.2 Arbres orientés

Nous considérons des arbres orientés *enracinés*, c'est-à-dire que nous spécifions un sommet f qui est considéré comme la *racine* de l'arbre orienté $T = (V, A)$. La racine f est un sommet de T arbitrairement choisi. Un arbre T sera représenté comme en Figure 2.4.

Oublions temporairement les orientations afin de définir la terminologie employée. La racine f est le *père* de $s_1, \ldots, s_6$ et le *grand-père* de $g_1, \ldots, g_7$. De façon symétrique $s_1, \ldots, s_6$ sont les *fils* de f, et $g_1, \ldots, g_7$ sont les *petits-fils* de f. Les sommets de degré 1, $h_1, h_2, g_2, \ldots, g_7, s_2, \ldots, s_5$, sont les *feuilles* de T. La *profondeur* de T est la plus grande distance à f d'une feuille de T. Les feuilles à distance maximum de f, h_1 et h_2, sont les sommets de *plus grande profondeur* de T. Enfin, pour tenir compte des orientations, nous utilisons les termes *voisin entrant* et *voisin sortant* de la façon standard : f est un voisin sortant de $s_1, \ldots, s_4$ et un voisin entrant de s_5 et s_6. Nous utiliserons *fils entrant* pour désigner un fils qui est aussi un voisin entrant, de même pour *fils sortant*, *père entrant* et *père sortant*. Pour un sommet v de T, l'ensemble des voisins entrants de v est noté $\Gamma^-(v)$. Étant donné un sous-ensemble C de sommets de T, on définit

$$I^-(v, C) \;:=\; (\Gamma^-(v) \cup \{v\}) \cap C.$$

L'ensemble $I^-(v, C)$ sera appelé l'*ensemble identifiant* de v. C est un code 1-identifiant de T si et seulement si tous les $I^-(v, C)$ sont distincts et non vides.

Dans cette section nous donnons un algorithme linéaire qui fournit un code 1-identifiant de cardinalité minimum de T. Nous commençons par décrire l'algorithme, dont nous prouvons ensuite la validité et la complexité.

Les résultats présentés dans cette section proviennent de [CGHLMM], qui est le fruit d'une collaboration avec I. Charon, O. Hudry et A. Lobstein, que nous avions invités pour un *workshop* d'une semaine à Grenoble avec S. Gravier et M. Mollard.

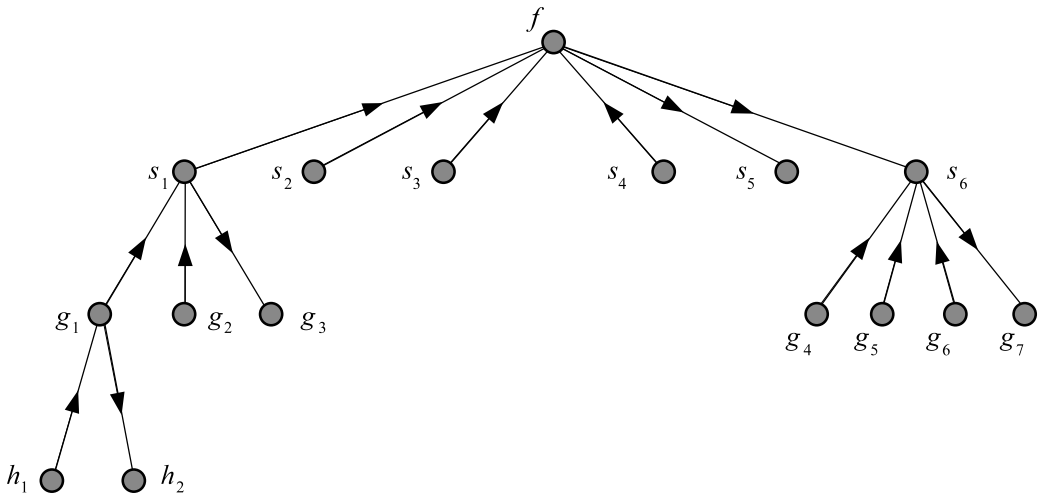


FIG. 2.4 – *Représentation d'un arbre orienté $T = (V, A)$ enraciné en f.*

2.2.1 Description d'un algorithme linéaire

L'algorithme ID-TREE est récursif ; il a pour paramètres un arbre T_0 enraciné en un sommet f, et un sous-ensemble de sommets C_0, à partir duquel on construit un code 1-identifiant de T_0. L'algorithme fournit un code 1-identifiant de cardinalité minimum C contenant le code partiel C_0. Le premier appel de ID-TREE se fait avec $C_0 = \emptyset$, mais en prenant C_0 quelconque l'algorithme résout le problème plus général suivant :

MIN-ID-CODE/ARBRE-OR/SOMMETS IMPOSÉS

Instance : Un arbre orienté $T_0 = (V_0, A_0)$, et un sous-ensemble de sommets $C_0 \subseteq V$.

Sortie : Un code 1-identifiant de T_0, contenant C_0, de cardinalité minimum parmi les codes 1-identifiants de T_0 contenant C_0.

Dans la suite T et C désignent l'arbre et le code 1-identifiant partiel d'un appel récursif, l'arbre et le code initiaux étant désignés par T_0 et C_0. C est un sous-ensemble de sommets de T_0, contenant C_0. T est un arbre obtenu par réduction de T_0, soit en enlevant des sommets et des arcs à T_0, soit en réduisant le nombre de sommets de profondeur maximum de T_0 en supprimant des arcs existants et en les remplaçant par d'autres. Nous ne faisons qu'ajouter des sommets à C, et nous ne faisons qu'ôter des sommets à T.

Algorithme ID-Tree
Paramètres : T, C.
Cas 1 : T est de profondeur au plus 1.

- Si f a des voisins sortants qui sont dans C, alors on enlève ces sommets de V (voir Figure 2.5.0).
- Sinon, nous sommes dans l'un de ces cinq cas :
  1. Si f n'a pas de voisins entrants, alors on ajoute f et tous ses voisins sortants à C (voir Figure 2.5.1).
  2. Si f a un seul voisin entrant s et au moins un voisin sortant, alors à C on ajoute f, s et tous les voisins sortants de f sauf un (voir Figure 2.5.2).
  3. Si f a un seul voisin entrant s et aucun voisin sortant, alors on ajoute f et s à C (voir Figure 2.5.3).
  4. Si f a au moins deux voisins entrants et au moins un voisin sortant, alors à C on ajoute f, tous les voisins entrants de f, et tous les voisins sortants de f sauf un (voir Figure 2.5.4).
  5. Si f a au moins deux voisins entrants et aucun voisin sortant, alors on ajoute tous les voisins entrants de f à C (voir Figure 2.5.5).
- **Arrêt** : fin de l'algorithme.

Cas 2 : T est de profondeur au moins 2.

- **Étape 1 :** On choisit un sommet $x \in V$ dont tous les fils sont des feuilles ; x sera appelé sommet *porte-feuilles*.
- **Étape 2 :**
  - **Opération** α **:** On enlève de V tous les fils sortants de x qui sont des sommets de C (voir Figure 2.5.0).
  - **Opération** β **:** S'il reste au moins un fils à x, alors nous sommes dans l'un des neuf cas suivants :
    1. (β.1) Si x n'a aucun fils entrant, et a un père entrant y, alors :
       - on ajoute x, y, et tous les fils de x sauf un à C,
       - on enlève x et tous ses fils de V (voir Figure 2.6.1).
    2. (β.2) Si x n'a aucun fils entrant, et a un père sortant y, alors :
       - on ajoute x et tous ses fils à C,
       - on enlève tous les fils de x de V (voir Figure 2.6.2).
    3. (β.3) Si x a au moins fils entrant et au moins un fils sortant, alors :
       - on ajoute x et tous ses fils sauf un fils sortant à C,

– on enlève tous les fils de x de V (voir Figure 2.6.3).

4. (β.4) Si x n'a aucun fils sortant, a au moins deux fils entrants, et a un père entrant, alors :
   – on ajoute tous les fils de x à C,
   – on enlève x et tous ses fils de V (voir Figure 2.6.4).

5. (β.5) Si $x \in C$, x n'a aucun fils sortant, a au moins deux fils entrants, et a un père sortant y, alors :
   – on ajoute tous les fils de x à C,
   – on enlève tous les fils de x sauf un, t, de V,
   – on enlève l'arc (t, x) de A et on ajoute l'arc (t, y) à A (voir Figure 2.6.5).

6. (β.6) Si $x \notin C$, x n'a aucun fils sortant, a au moins deux fils entrants, et a un père sortant, alors :
   – on ajoute tous les fils de x à C,
   – on enlève x et tous ses fils de V (voir Figure 2.6.6).

7. (β.7) Si $x \notin C$, x n'a aucun fils sortant, a un seul fils entrant, et a un père entrant y, alors :
   – on ajoute x et y à C,
   – on enlève x et le fils de x à V (voir Figure 2.6.7).

8. (β.8) Si $x \in C$, x n'a aucun fils sortant, a un seul fils entrant, et a un père entrant, alors :
   – on ajoute le fils de x à C,
   – on enlève x et le fils de x de V (voir Figure 2.6.8).

9. (β.9) Si x n'a aucun fils sortant, a un seul fils entrant t, et a un père sortant y, alors :
   – on ajoute x et t à C,
   – on enlève l'arc (t, x) de A et on ajoute l'arc (t, y) à A (voir Figure 2.6.9).

– **Étape 3 :** Appel récursif de ID-Tree.

Fin de l'algorithme ID-Tree

Remarquer que les sous-cas 1.-5. du Cas 1 sont disjoints et complets, de même que les sous-cas β.1-β.9 du Cas 2. La sélection d'un porte-feuilles x est détaillée dans le paragraphe 2.2.3.

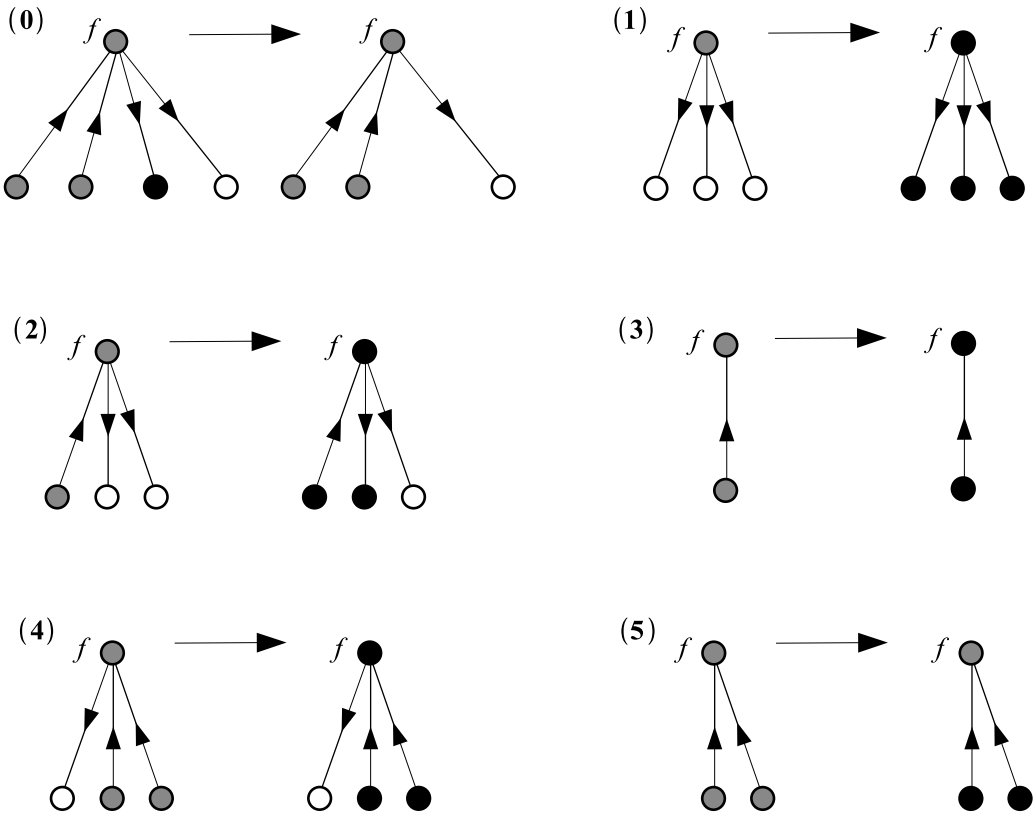


FIG. 2.5 – *Cas où T est de profondeur 1 dans l'algorithme* ID-TREE. *Les sommets noirs sont dans C, les blancs dans $V \setminus C$, et le statut des sommets gris est indéterminé. Un sommet gris restant gris après la transformation n'a pas changé de statut.*

2.2.2 Preuve de la validité de l'algorithme

Nous prouvons la validité de l'algorithme ID-TREE décrit ci-dessus. Nous montrons que si T_0 et C_0 sont les paramètres d'entrée de l'algorithme, alors ID-TREE retourne un code 1-identifiant de cardinalité minimum C de T_0 contenant C_0.

Tout d'abord nous remarquons qu'un sommet x peut être choisi au plus une fois comme porte-feuilles (début du Cas 2). En effet, une fois les opérations (α) et (β) effectuées, x n'est plus un porte-feuilles de T. Ceci montre que l'algorithme se termine. Alternativement, chaque opération réduit soit le nombre de sommets de T, soit le nombre de sommets de profondeur maximum de T ; il ne fait jamais augmenter ces paramètres. C'est dans ce sens que l'algorithme

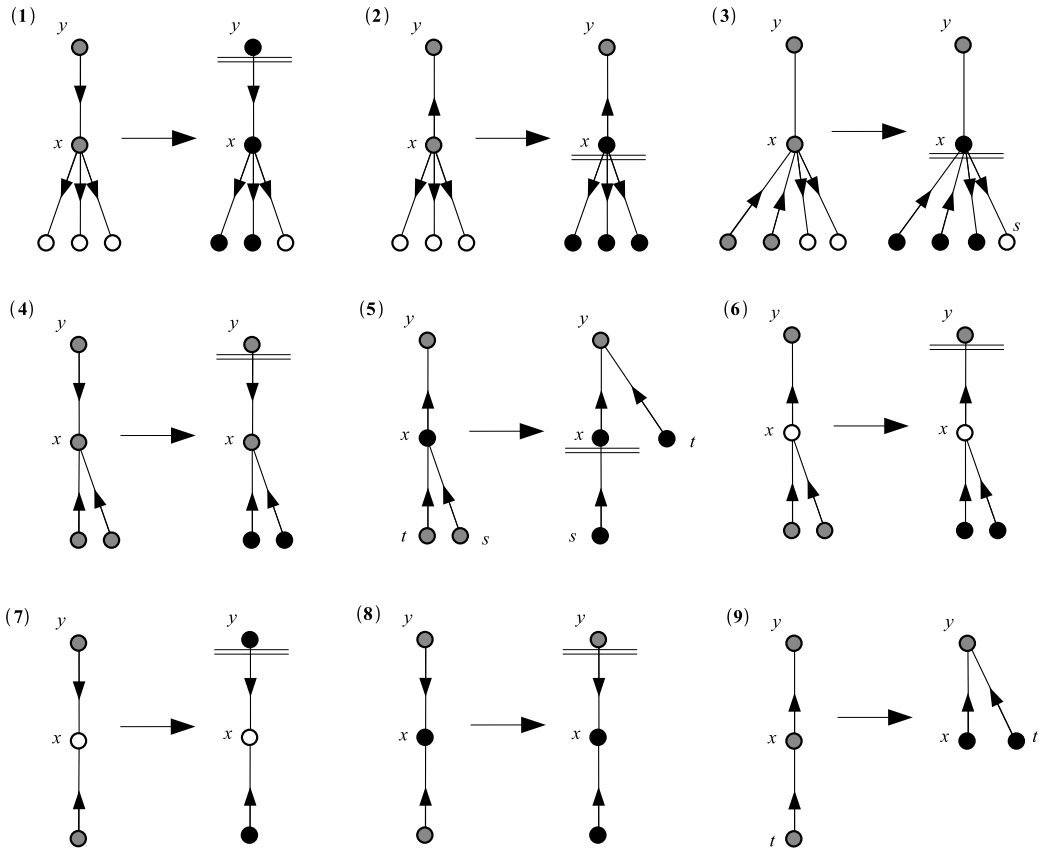


FIG. 2.6 – *Les cas β.1 à β.9 de l'algorithme* ID-TREE. *Les sommets noirs sont dans C, les blancs dans $V \setminus C$, et le statut des sommets gris est indéterminé. Un sommet gris restant gris après la transformation n'a pas changé de statut. Les sommets situés en-deçà de la double barre sont enlevés de V.*

réduit la taille du problème.

Il n'y a un seul appel récursif de l'algorithme, à la fin du Cas 2. Numérotons les appels récursifs par un entier p. On attribue le numéro $p = 1$ à la dernière invocation de l'algorithme, et nous attribuons le numéro $p + 1$ à l'invocation appelant l'invocation portant le numéro p. Ainsi, la première invocation de l'algorithme porte le plus grand numéro $p_{\max}$; ce nombre n'est pas connu à l'avance. La preuve de la validité de l'algorithme se fait par induction sur $p_{\max}$.

Si $p_{\max} = 1$, alors nous sommes dans le Cas 1, où la profondeur de T est

au plus 1. Il est facile de voir que l'algorithme fournit un code 1-identifiant de cardinalité minimum contenant $C \cap V$ (voir Figure 2.5).

Supposons que $p_{\max} \geq 2$, et considérons la première invocation d'ID-TREE (elle porte le numéro $p_{\max}$). Posons $C' = C \cap V$.

Si, avant l'opération (α), il existe dans V une feuille $\ell \in C$ ayant un père entrant, alors pour obtenir un code 1-identifiant de T contenant C' il faut et il suffit de trouver un code 1-identifiant de $T \setminus \{\ell\}$ contenant $C' \setminus \{\ell\}$, d'où la validité de l'opération (α).

De la même façon, nous prouvons la validité de l'opération (β) en étudiant les neuf cas β.1-β.9.

Cas β.1 **:** x n'a pas de fils entrant et a un père entrant y (voir Figure 2.6.1). Nous montrons que, parmi les codes 1-identifiants de T contenant C', il en existe un contenant $C' \cup \{x, y\}$.
Soit K un code 1-identifiant de T contenant C'.
Tout d'abord, supposons que $x \notin K$. Ceci entraîne $y \in K$, et tous les fils de x sont dans K, parce que x et ses fils doivent être 1-couverts par K. Puisque l'opération (α) n'impose pas que les fils de x soient dans C', alors on peut remplacer un des fils de x par x lui-même. On peut donc supposer que $x \in K$.
De même, supposons que $y \notin K$. Pour 1-séparer x de ses fils il faut que tous les fils de x soient dans K. Nous pouvons encore remplacer un de fils de x par y, et supposer que y aussi est dans K.
Soit K un code 1-identifiant de cardinalité minimum de T contenant $C' \cup \{x, y\}$. K contient nécessairement tous les fils de x sauf un. Ainsi, on ne change pas la cardinalité minimum d'un code 1-identifiant de T contenant C' en ajoutant x et tous ses fils sauf un à C.
Ainsi, dans la partie droite de la Figure 2.5.1, x et ses fils sont 1-couverts et 1-séparés par C, et ni x ni aucun de ses fils ne couvre ni y ni aucun de ses ancêtres. Ces sommets peuvent donc être enlevés de T, ce qui prouve la validité de l'opération β.1.
Nous faisons ensuite un appel récursif pour résoudre le problème, ce qui valide l'opération β.1 par hypothèse de récurrence.

Cas β.2 **:** x n'a pas de fils entrant et a un père sortant y (voir Figure 2.6.2). Comme x et ses fils doivent être 1-couverts et 1-séparés par le code, alors nécessairement x et tous ses fils sont dans le code. Puisque les fils de x ne 1-couvrent qu'eux-mêmes, alors on peut les enlever de V, d'où la validité de l'opération β.2.

Cas β.3 **:** x au moins un fils entrant et un fils sortant (voir Figure 2.6.3). Le père de x, y, peut être entrant ou sortant. Un code 1-identifiant de T

contient nécessairement tous les fils entrants de x, et tous les fils sortants de x sauf un. Nous pouvons donc mettre x et tous ses fils sauf un fils sortant s dans C – nous rappelons que l'opération (α) ne requiert pas que les fils sortants de x soient dans C. Les fils de x sont donc tous 1-couverts et deux à deux 1-séparés. Ils sont également 1-séparés de x, qui lui est 1-couvert et 1-séparé de son père y.
Si y est un voisin entrant de x, alors y, dès lors qu'il sera 1-couvert, sera 1-séparé de x et de ses fils.
Si y est un voisin sortant de x, alors y et s sont tous deux 1-couverts par x. Pour les 1-séparer, il faut et il suffit de 1-couvrir y avec un sommet autre que x. Si nous enlevons les fils de x de V, alors pour 1-séparer x de y il est également nécessaire et suffisant de 1-couvrir y avec un sommet autre que x. Nous avons reporté sur x l'information qu'il y avait un conflit entre y et s.
Dans les deux cas la situation de y est inchangée en enlevant les fils de x de V, ce qui montre la validité de l'opération $\beta.3$.

Cas $\beta.4$ **:** x n'a pas de fils sortant, a au moins deux fils entrants, et a un père entrant y (voir Figure 2.6.4). Tout code 1-identifiant de T contient nécessairement tous les fils de x, qui 1-couvrent et 1-séparent x de tout autre sommet de T. Comme y est le père entrant de x, alors y n'est pas 1-couvert par x, et la situation de y est inchangée si l'on enlève x et tous ses fils de V, ce qui montre la validité de l'opération $\beta.4$.

Cas $\beta.5$ **:** $x \in C$ n'a aucun fils sortant, a au moins deux fils entrants, et a un père sortant y (voir Figure 2.6.5). Tout code 1-identifiant de T contient nécessairement tous les fils de x, et y est 1-couvert et 1-séparé de x et s. Ceci restant vrai après l'opération $\beta.5$, cette opération est valide.

Cas $\beta.6$ **:** $x \notin C$ n'a aucun fils sortant, a au moins deux fils entrants, et a un père sortant y (voir Figure 2.6.6). Tout code 1-identifiant de T contient nécessairement tous les fils de x.
Nous montrons que parmi tous les codes 1-identifiants de cardinalité minimum de T contenant C', il en existe un ne contenant pas x. Soit K un code 1-identifiant de T contenant $C' \cup \{x\}$. Nous enlevons x de K et nous considérons deux cas :
– Si $y \notin K$, alors tout sommet de $V \setminus \{y\}$, hormis x et ses fils, est 1-couvert par un sommet de K qui n'est pas y. En ajoutant y à K nous obtenons un code 1-identifiant de T ne contenant pas x, de même cardinalité que K.
– Si $y \in K$, le seul problème à considérer est le cas d'un sommet z étant 1-couvert seulement par y : y et z ne sont plus 1-séparés par x. Mais dans ce cas il suffit de rajouter z à K pour obtenir un code 1-identifiant de T de même taille que K ne contenant pas x.

Ainsi la situation de y est inchangée en enlevant x et ses fils de V, ce qui prouve la validité de l'opération $\beta.6$.

Cas $\beta.7$ **:** $x \notin C$ n'a aucun fils sortant, a un seul fils entrant, et a un père entrant y (voir Figure 2.6.7). Tout code 1-identifiant de T contient nécessairement le fils de x. Soit K un code 1-identifiant de cardinalité minimum de T contenant C'. Comme x et son fils sont 1-séparés par K, alors soit x soit y est dans K. Ils ne peuvent être tous deux dans K, parce que sinon nous pourrions enlever x de K. De même si K contenait x mais pas y, alors nous pourrions remplacer x par y dans K. Ainsi il existe un code 1-identifiant de cardinalité minimum de T contenant C', y, et ne contenant pas x. Nous pouvons donc enlever x et son fils de V, ce qui valide l'opération $\beta.7$.

Cas $\beta.8$ **:** $x \in C$ n'a aucun fils sortant, a un seul fils entrant, et a un père entrant y (voir Figure 2.6.8). Le fils de x est nécessairement dans C. Comme x et son fils ne 1-couvrent aucun autre sommet de T, alors nous pouvons les enlever de V, ce qui valide l'opération $\beta.8$.

Cas $\beta.9$ **:** x n'a aucun fils sortant, a un seul fils entrant t, et a un père sortant y (voir Figure 2.6.9). Tout code 1-identifiant de T contient t et x, qui 1-couvre et 1-sépare y de x et t. Ceci reste vrai après l'opération $\beta.9$, qui est donc valide.

Ceci montre que l'algorithme nous fournit un code 1-identifiant de cardinalité minimum de T, C, contenant le code partiel initial C_0. Dans le paragraphe suivant nous étudions la complexité de l'algorithme.

2.2.3 Preuve de la linéarité de l'algorithme

L'arbre T_0 est enraciné en un sommet f qui restera la racine de T jusqu'à la fin de l'algorithme. Un parcours de T_0 en largeur d'abord nous permet de décrire T_0 en donnant, pour chaque sommet, son père et ses fils, ainsi que la direction des arcs. Ce parcours est fait en un temps linéaire par rapport au nombre de sommets de T_0.

Dans le Cas 1 (où la profondeur de T est au plus un), ainsi que dans le cas des opérations (α) et $\beta.1$-$\beta.4$, $\beta.6$-$\beta.8$, chaque arc entrant ou sortant de f ou du porte-feuilles x n'est traité qu'une seule fois. Le traitement d'un arc nécessite un nombre d'opérations élémentaires borné par une constante (voir Figures 2.5, 2.6.1-2.6.4 et 2.6.6-2.6.8).

Les opérations $\beta.5$ et $\beta.9$ sont plus délicates, puisque nous créons de

nouveaux arcs. À moins que nous n'appliquions l'opération $\beta.5$ deux fois sur le même arc (avec y jouant le rôle de x), après ces opérations les arcs créés seront effacés lors de leur prochain traitement. Ainsi, afin de garantir que nous n'appliquons pas l'opération $\beta.5$ deux fois sur le même arc, nous décidons qu'à chaque application de l'opération $\beta.5$:

- l'arc (t, y) est créé en utilisant un sommet t tel que l'arc initial (t, x) existait déjà dans T_0 (on peut toujours trouver un tel t) ; et, par conséquent :
- tous les arcs éventuellement créés par des opérations $\beta.5$ et $\beta.9$ appliquées précédemment sont effacés.

Ceci nous garantit qu'à chaque application d'une opération $\beta.5$ ou $\beta.9$:

- tout arc entrant en x n'est traité qu'une seule fois,
- le seul traitement que puisse recevoir un arc créé par une opération $\beta.5$ ou $\beta.9$ est la suppression.

Nous pouvons donc conclure que le traitement de tout arc ne requiert qu'un nombre d'opérations élémentaires borné par une constante, ce qui montre que la complexité de l'algorithme est linéaire par rapport au nombre d'arcs de T_0.

En fait, le point crucial est le choix d'un porte-feuilles x. Nous allons toujours choisir un sommet qui n'est pas une feuille et a une profondeur maximum sous cette contrainte. Ceci peut être fait en temps linéaire de la façon suivante.
Avant de démarrer l'algorithme, numérotons les sommets et construisons un tableau $\mathcal{A}$ donnant les sommets par profondeur croissante. Le tableau $\mathcal{A}$ peut être construit en temps linéaire par un parcours en largeur d'abord. L'appartenance à T est aussi stockée dans $\mathcal{A}$. Plaçons de plus un curseur χ qui, au début de l'algorithme, pointe sur le dernier élément de $\mathcal{A}$: c'est une feuille de plus grande profondeur, et son père est un porte-feuilles de profondeur maximum. Le curseur χ pointe sur le sommet courant traité par l'algorithme. À chaque fois que nous voulons trouver un porte-feuilles, nous remontons depuis la position courante χ dans $\mathcal{A}$, et nous cherchons le premier sommet appartenant à l'arbre courant : nous prenons son père comme porte-feuilles. Après chaque opération effectuée dans l'algorithme ID-Tree, nous mettons à jour l'appartenance à T et déplaçons le curseur χ d'une unité dans $\mathcal{A}$. Dans le cas des opérations $\beta.5$ et $\beta.9$ (voir Figures 2.6.5 et 2.6.9), nous n'avons pas besoin de considérer le changement de profondeur de la feuille t : en effet, nous traiterons son nouveau père y lorsque le curseur χ pointera sur x dans $\mathcal{A}$. Dans ces cas, aucune information utile n'est perdue en laissant derrière nous dans $\mathcal{A}$ le sommet t. Les opérations $\beta.5$ et $\beta.9$ sont les deux seules pour lesquelles nous modifions la profondeur d'un sommet dans T.

Nous parcourons donc les porte-feuilles en un nombre d'opérations élémentaires qui est linéaire par rapport au nombre de sommets de T_0. Ainsi la complexité globale de l'algorithme est linéaire.

Dans ce paragraphe et le paragraphe précédent, nous avons donc montré :

Théorème 2.7 ([CGHLMM])
Étant donné un arbre orienté $T = (V, A)$ et un sous-ensemble de sommets $C_0 \subseteq V$, l'algorithme ID-Tree *décrit au paragraphe 2.2.1 résout le problème* Min-ID-Code/Arbre-Or/Sommets Imposés *en temps linéaire par rapport au nombre de sommets de T : il retourne C un code* 1*-identifiant de T, contenant C_0, de cardinalité minimum parmi les codes* 1*-identifiants de T contenant C_0.*
Dans le cas particulier où $C_0 = \emptyset$, l'algorithme ID-Tree *retourne un code* 1*-identifiant de cardinalité minimum de T.*

En ce qui concerne les codes t-identifiants, il est facile de voir que la propriété "C est un code t-identifiant de G" est une propriété exprimable dans le langage MS pour tout $t \geq 1$ fixé, où le langage MS est celui décrit dans le Théorème 2.1. Comme conséquence, nous obtenons directement que la recherche de la cardinalité minimum d'un code $(t, \leq \ell)$-identifiant est linéaire si l'on se restreint aux arbres — et, de façon générale, ce problème est linéaire pour toute classe de graphes de largeur d'arborescence bornée par une constante. Aucun algorithme *ad hoc* de construction d'un code t-identifiant optimum n'est connu pour ce problème.

Chapitre 3

Classes de graphes particulières

Dans ce chapitre nous étudions les codes identifiants pour trois classes de graphes particulières : les hypercubes, les grilles, et les cycles. Pour chacune de ces familles nous donnons des bornes inférieures et supérieures sur la cardinalité minimum d'un code identifiant. En début de chaque section nous rappelons la définition des familles de graphes étudiées.

3.1 Hypercubes

L'*hypercube* de dimension n, noté Q_n, est le graphe ayant pour ensemble de sommets $\{0,1\}^n$, tel que uv est une arête de Q_n si et seulement si u et v diffèrent sur exactement une coordonnée. Alternativement, deux mots u et v sont voisins dans Q_n si et seulement si leur distance de Hamming est égale à 1, où la distance de Hamming entre deux vecteurs u, v est définie comme le nombre de coordonnées sur lesquelles u et v diffèrent. Dans cette section $d(u,v)$ désignera la distance de Hamming entre u et v.

Nous pouvons construire récursivement Q_{n+1} à partir de deux copies de Q_n, Q_n^1 et Q_n^2, en les joignant par un couplage connectant chaque $u^1 \in Q_n^1$ avec son unique voisin $u^2 \in Q_n^2$. Dans la suite nous verrons Q_{n+1} comme deux copies jointes de Q_n, ce que nous noterons

$$Q_{n+1} \;=\; Q_n^1 \square Q_n^2.$$

L'unique voisin $v^1 \in Q_n^1$ d'un sommet $v^2 \in Q_n^2$ sera appelé le *jumeau* de v^2. La cardinalité minimum d'un code t-identifiant de Q_n sera notée $M_t(Q_n)$.

Les hypercubes sont des graphes très étudiés. En particulier, ils ont des

propriétés qui intéressent les constructeurs de machines dédiées au calcul parallèle. Par exemple, la compagnie nCUBE construit des serveurs basés sur cette architecture [NCube]. La Connection Machine CM-2 de Thinking Machines Corporation était également basée sur une topologie en hypercube [Hil85]. Dans le cas des codes identifiants, il existe une littérature relativement vaste sur le sujet. Historiquement, les premiers articles parus au sujet des codes identifiants donnaient une large place à l'étude des hypercubes (voir par exemple [BHL00, BHL01, KCL98, KCLA99]).

Cependant, de nombreuses questions sur les codes identifiants dans les hypercubes restent sans réponses. Par exemple, $M_t(Q_n)$, la cardinalité minimum d'un code t-identifiant de Q_n, n'est pas connue pour tout $n \geq 1$. Une boule de rayon t ayant pour cardinalité

$$\sum_{i=0}^{t} \binom{n}{i} \;=\; \Theta\left(n^t\right)$$

dans Q_n, le Théorème 1.7 nous dit que, pour tout $t \geq 1$ fixé, la cardinalité minimum d'un code t-identifiant de Q_n est exponentielle en n :

$$M_t(Q_n) \;\geq\; \frac{2^{n+1}}{\sum_{i=0}^{t}\binom{i}{n}+1} \;=\; \Theta\left(\frac{2^n}{n^t}\right).$$

M. G. Karpovsky, K. Chakrabarty et L. Levitin [KCL98] ont affiné cette borne dans le cas $t = 1$:

Théorème 3.1 ([KCL98])

$$M_1(Q_n) \;\geq\; \frac{n2^{n+1}}{n(n+1)+2}. \qquad \square$$

Ils proposent de plus la construction suivante :

Théorème 3.2 ([KCLA99])
Soient $n \geq 3$ et $t < n/2$. Soit $C^\star$ un code couvrant optimum de Q_n de rayon de recouvrement $2t$. Alors le code $C = \{w |\, \exists v \in C^\star,\, d(v,w) = t\}$ est un code t-identifiant de Q_n. $\square$

Ceci implique :

Corollaire 3.1 ([KCLA99])
Pour $n \geq 3$ et $t < n/2$, soit $K(n,t)$ la cardinalité minimum d'un code couvrant de rayon t de Q_n. Alors $M_t(Q_n)$ satisfait l'inégalité suivante :

$$M_t(Q_n) \;\leq\; \binom{n}{t} K(n,2t). \qquad \square$$

Des tables de valeurs de $K(n,t)$ pour $n \leq 33, t \leq 10$ peuvent être trouvées dans [CHLL97].

Une autre question naturelle concernant les codes identifiants dans les hypercubes est la conjecture que font U. Blass, I. Honkala et S. Litsyn dans [BHL00] :

Conjecture 3.1 ([BHL00])
Pour tout $t \geq 1$, la fonction $n \mapsto M_t(Q_n)$ est monotone, i.e. pour tout $t, n \geq 2t+1$ on a

$$M_t(Q_{n+1}) \geq M_t(Q_n).$$

Dans le cas des codes couvrants, la question précédente est triviale : si C est un code couvrant de $Q_{n+1} = Q_n^1 \equiv Q_n^2$, il suffit de projeter tout sommet $u^2 \in Q_n^2 \cap C$ sur son jumeau $u^1 \in Q_n^1$ pour obtenir un code couvrant de Q_n^1. Dans le cas des codes identifiants, cet argument ne fonctionne plus : projeter tout $u^2 \in Q_n^2 \cap C$ sur son jumeau u^1 ne nous fournit pas nécessairement un code identifiant de Q_n^1 (voir par exemple le cas de la Figure 3.1).

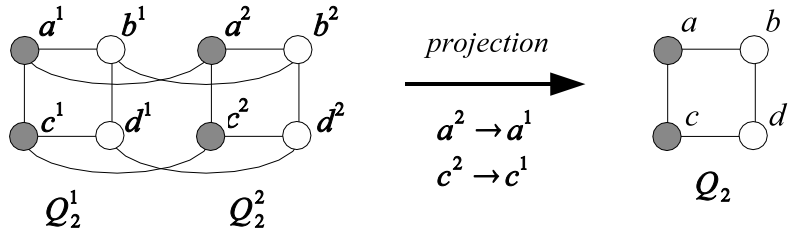


FIG. 3.1 – *On projette a^2 et c^2 sur leurs jumeaux respectifs a^1 et c^1, mais cela ne nous fournit pas un code 1-identifiant de Q_2^1 : a et c ne sont pas 1-séparés.*

Dans le cas $t = 1$, nous pouvons cependant adapter cet argument et faire tout de même une projection. La différence est que l'on ne projete plus nécessairement v^2 sur son jumeau v^1, mais éventuellement sur un sommet de $N(v^1) \cap Q_n^1$. Le théorème suivant répond à la conjecture 3.1 dans le cas $t = 1$.

Théorème 3.3 ([Monb])
Pour tout $n \geq 2$ on a

$$M_1(Q_n) \leq M_1(Q_{n+1}).$$

Preuve : Soit C un code 1-identifiant de Q_{n+1}. À partir de C nous allons construire un code $\widetilde{C}$ qui identifie les sommets de Q_n^1 entre eux. $\widetilde{C}$ sera tel que $|\widetilde{C}| \leq |C|$ et $|\widetilde{C} \cap Q_n^2| = 0$.

Nous commençons par poser $\widetilde{C} = C$, et procédons par induction sur $|\widetilde{C} \cap Q_n^2|$. Si $|\widetilde{C} \cap Q_n^2| = 0$, alors il n'y a rien à faire. Sinon, soit $u^2 \in \widetilde{C}$. Nous projetons alors u^2 sur un $\pi(u^2) \in Q_n^1$ comme suit :

(a) Si $I_t(u^1, \widetilde{C}) \setminus \{u^2\}$ est encore distinct de tous les autres ensembles identifiants $I_t(v^1, \widetilde{C})$, $v_1 \neq u_1$, alors $\pi(u^2) = u^1$

(b) Sinon il existe un et un seul $v^1 \neq u^1$ tel que $I_t(u^1, \widetilde{C}) \setminus \{u^2\} = I_t(v^1, \widetilde{C})$, et nous posons dans ce cas $\pi(u^2) = w^1$, où w^1 est un sommet arbitraire choisi parmi les sommets de $\Big(I_t(u^1, \widetilde{C}) \Delta I_t(v^1, \widetilde{C})\Big) \cap Q_n^1$.

Nous affirmons que $\widetilde{C} \leftarrow \widetilde{C} \cup \{\pi(u^2)\} \setminus \{u^2\}$ est un code qui identifie les sommets de Q_n^1. Par définition de π, les sommets de Q_n^1 ont des ensembles identifiants $I_t(v, \widetilde{C})$ non vides et distincts. Il reste seulement à vérifier que dans le cas (b) il n'y a qu'*un et un seul* $v^1 \neq u^1$ tel que $I_t(u^1, \widetilde{C}) \setminus \{u^2\} = I_t(v^1, \widetilde{C})$. S'il y avait un autre w^1 tel que $I_t(u^1, \widetilde{C}) \setminus \{u^2\} = I_t(w^1, \widetilde{C})$, nous aurions $I_t(w^1, \widetilde{C}) = I_t(v^1, \widetilde{C})$, et $\widetilde{C}$ ne serait pas un code 1-identifiant de Q_{n+1}. Clairement, nous avons $|\widetilde{C}| \leq |C|$, et $|\widetilde{C} \cap Q_{n+1}| = |C \cap Q_{n+1}| - 1$, et nous pouvons appliquer l'hypothèse de récurrence sur $\widetilde{C}$. □

Il nous semble étonnant que la Conjecture 3.1 ne soit pas entièrement résolue à ce jour, tellement le résultat annoncé paraît naturel.

3.2 Grilles et Bandes

La *grille n-dimensionnelle* (infinie) est définie comme le graphe ayant pour ensemble de sommets $\mathbb{Z}^n$ et pour ensemble d'arêtes $\{uv \mid d_1(u,v) = 1\}$, où $d_1(u,v) = \sum_{i=1}^{n} |v_i - u_i|$. Alternativement, $d_1(u,v)$ est la longueur d'un plus court chemin entre u et v. La distance d_1 est parfois appelée *distance de Manhattan*, ou encore *distance de Lee*.

Ici, nous désignons simplement par *grille* la grille bidimensionnelle. La *bande* de hauteur k désigne le sous-graphe $\mathcal{S}_k$ de la grille induit par le sous-ensemble de sommets $\{1, \ldots, k\} \times \mathbb{Z}$, et *demi-bande* de hauteur k le sous-graphe $\mathcal{S}_k^+$ de la grille induit par le sous-ensemble de sommets $\{1, \ldots, k\} \times \mathbb{N}$. Nous appelons *grille* $k \times n$ (finie) le sous-graphe $\mathcal{G}_{k \times n}$ de la grille induit par le sous-ensemble de sommets $\{1, \ldots, k\} \times \{1, \ldots, n\}$.

Les grilles finies sont des structures de graphes très étudiées, en particu-

lier elles sont souvent utilisées comme structure de machines parallèles. Par exemple, le PARAGON d'Intel [IC91] ou le DAP d'AMT [BLP90] sont basés sur des topologies de grille bidimensionnelle.

Nous rappelons que, pour un graphe fini ou infini G, la *densité* d'un code t-identifiant de G est définie comme suit.

Soit v_0 un sommet quelconque de G, et pour tout $n \in \mathbb{N}$ soit $B_n(v_0)$ la boule de rayon n centrée en v_0 : $B_n(v_0) = \{x \mid d_1(x, v_0) \leq n\}$. La densité $d(C, G)$ d'un code t-identifiant C de G est définie comme étant la limite :

$$d(C, G) \;=\; \limsup_{n\to\infty} \frac{|C \cap B_n(v_0)|}{|B_n(v_0)|}.$$

On note $d_t^*(G)$ l'infimum de la densité d'un code t-identifiant de G. Dans le cas où G est fini, on a $d_t^*(G) = |C^*|/|G|$, où $|G|$ désigne le nombre de sommets de G et où C^* est un code t-identifiant de cardinalité minimum de G.

Il est facile de voir que

$$d_t^*(\mathcal{S}_k^+) \;=\; d_t^*(\mathcal{S}_k)$$

pour tout $k \geq 1$. En effet, si C est un code t-identifiant de la bande $\mathcal{S}_k$, alors, après un nombre fini de modifications sur la trace de C sur $\{1, \ldots, k\} \times \mathbb{N}$, on peut obtenir un code t-identifiant de la demi-bande $\mathcal{S}_k^+$ (voir Figure 3.2). Ainsi $d_t^*(\mathcal{S}_k^+) \leq d_t^*(\mathcal{S}_k)$.

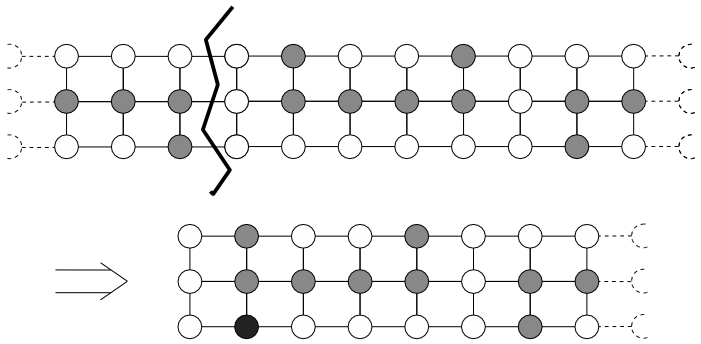


FIG. 3.2 – *Créer un code 1-identifiant de $\mathcal{S}_3^+$ à partir de la trace sur $\mathcal{S}_3^+$ d'un code 1-identifiant de $\mathcal{S}_3$. Il suffit d'ajouter un sommet à la trace de C sur $\{1, \ldots, 3\} \times \mathbb{N}$ pour obtenir un code t-identifiant de la demi-bande $\mathcal{S}_3^+$.*

D'autre part, soit C un code t-identifiant de la demi-bande $\mathcal{S}_k^+$. En recollant deux copies de $\mathcal{S}_k^+$ par leurs premières colonnes on obtient une bande

$\mathcal{S}_k$. Il suffit alors de faire un nombre fini de modifications sur les deux copies de C pour obtenir un code t-identifiant de $\mathcal{S}_k$ (voir Figure 3.3). Ainsi $d_t^*(\mathcal{S}_k^+) \geq d_t^*(\mathcal{S}_k)$.

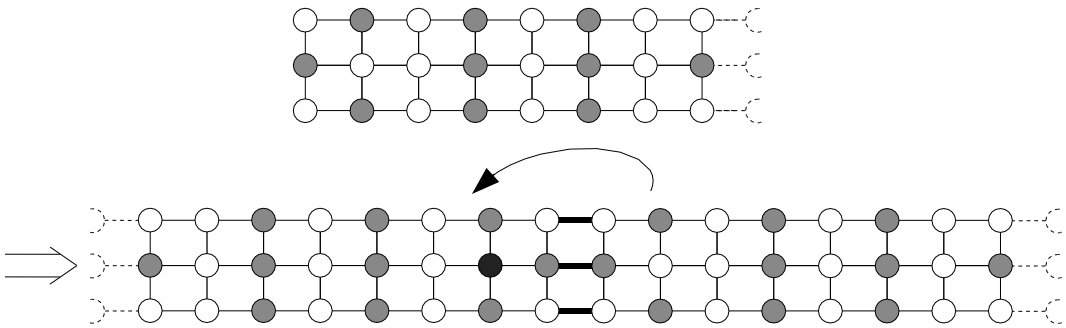


FIG. 3.3 – *Code 1-identifiant de $\mathcal{S}_3$ obtenu à partir de deux copies d'un code 1-identifiant de $\mathcal{S}_3^+$.*

Nous avons donc montré :

Proposition 3.1 ([BCHL04])
Pour tout $k \geq 1$ on a

$$d_t^*(\mathcal{S}_k^+) \;=\; d_t^*(\mathcal{S}_k). \qquad \square$$

Dans la suite nous n'étudierons donc que la bande, les résultats sur la demi-bande se déduisant trivialement de la proposition précédente.

Dans la sous-section suivante nous donnons les valeurs exactes de $d_1^*(\mathcal{S}_1)$ et $d_1^*(\mathcal{S}_2)$. Dans la deuxième sous-section nous donnons, dans le cas général, des bornes inférieures et supérieures sur $d_1^*(\mathcal{S}_k)$ et $d_1^*(\mathcal{G}_{k\times n})$. Puisque nous ne nous intéressons ici qu'aux codes 1-identifiants, pour plus de simplicité le terme "code identifiant" sera utilisé à la place de "code 1-identifiant" dans cette section. De même nous dirons, respectivement, "couvrir" et "séparer" au lieu de "1-couvrir" et "1-séparer".

Les résultats présentés dans cette section proviennent de [DGM04]. Ils sont le fruit d'un travail réalisé avec M. Daniel lors de son stage de DEA en 2003 [Dan03].

3.2.1 Bandes de petite taille

Proposition 3.2 ([DGM04])
Nous avons $d_1^*(\mathcal{S}_1) = \frac{1}{2}$ *et* $d_1^*(\mathcal{S}_2) = \frac{3}{7}$.

Preuve : Soit C un code identifiant de $\mathcal{S}_1$. Pour tout ensemble $X = \{u, v, w, t\}$ de quatre sommets consécutifs de $\mathcal{S}_1$, on a $|C \cap X| \geq 2$. En effet, si $|C \cap X| = 1$, alors soit les sommets v, w ont le même ensemble identifiant (cas $C \cap X = \{v\}$ ou $C \cap X = \{w\}$), ou un des sommets v, w n'est pas couvert (cas $C \cap X = \{u\}$ ou $C \cap X = \{t\}$). Si $C \cap X = \emptyset$, alors v et w ne sont pas couverts. Ainsi $d_1^*(\mathcal{S}_1) \geq \frac{1}{2}$. Pour conclure nous exhibons en Figure 3.4 un code identifiant de $\mathcal{S}_1$ de densité $\frac{1}{2}$. Soit C un code identifiant optimum

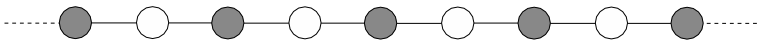


FIG. 3.4 – *Un code identifiant périodique de* $\mathcal{S}_1$ *de densité* $\frac{1}{2}$.

de $\mathcal{S}_2$. Nous pouvons supposer que sa densité est strictement inférieure à $\frac{1}{2}$. Il est en effet facile d'exhiber un code identifiant de $\mathcal{S}_2$ de densité $< \frac{1}{2}$ (voir par exemple la Figure 3.8).

Appelons *colonne* de $\mathcal{S}_2$ toute paire de sommets (u, v) telle que $u - v = (0, \pm 1)$. Une colonne (u, v) est dite *de type* k si $|C \cap \{u, v\}| = k$, pour $k = 0, 1, 2$. Nous disons qu'une colonne de $\mathcal{S}_2$ est *isolée* si c'est une colonne de type 0 qui n'est adjacente à aucune colonne de type 2.

Nous pouvons supposer que :

Il n'y a pas trois colonnes consécutives de type 2 dans $\mathcal{S}_2$. (3.1)

Supposons le contraire. Soit (u, v) une colonne de type 2 adjacente à deux colonnes de type 2. En ce cas, nous pouvons enlever le sommet u de C et obtenir un autre code identifiant de $\mathcal{S}_2$ ayant la même densité que C. En faisant cela pour toute telle colonne de type 2, nous obtenons un code identifiant de $\mathcal{S}_2$, qui satisfait (3.1), et qui a une densité inférieure ou égale à celle de C.

Nous pouvons de plus supposer que :

Aucune colonne de type 2 n'est adjacente à une colonne de type 2 et à une colonne de type 1. (3.2)

Supposons le contraire. Soit (u, v) une colonne de type 2 adjacente à une colonne (x, y) de type 2 et à une colonne (z, a) de type 1. Sans perte de généralité supposons que $a \notin C$ et que $a - v = (1, 0)$. Selon le type de l'autre colonne adjacente à (z, a), nous pouvons soit déplacer, soit supprimer v de C, et obtenir un autre code identifiant de $\mathcal{S}_2$ ayant la même densité que C. En effet, soit (b, c) l'autre colonne adjacente à (z, a), et supposons que $b - z = (1, 0)$. Si $b \in C$, alors $C \setminus \{v\}$ reste un code identifiant de $\mathcal{S}_2$. Sinon, alors $C \setminus \{v\} \cup \{a\}$ est un code identifiant de $\mathcal{S}_2$. Remarquons que ce procédé n'ajoute aucune colonne de type 2 adjacente à une colonne de type 2 et à une colonne de type 1. En répétant ceci pour toute telle colonne de type 2, nous obtenons un autre code identifiant de $\mathcal{S}_2$ satisfaisant (3.2) ayant une densité inférieure ou égale à celle de C.

Enfin, nous pouvons supposer que :

Toute colonne de type 2 est adjacente à une colonne de type 0. (3.3)

Grâce à (3.1) et (3.2), nous savons que le seul cas à traiter est le cas où une colonne de type 2 est adjacente à deux colonnes de type 1. Soit (u, v) une telle colonne, et soit (x, y) la colonne adjacente à (u, v) telle que $y \in C$ et $y - v = (1, 0)$. Nous allons procéder à ce que nous appelons le *principe de décalage à droite*, qui consiste à enlever y de C, et à éventuellement ajouter à C un nouveau sommet appartenant à la partie droite de la bande $\mathcal{S}_2$ (voir l'illustration de ce principe en Figure 3.5). Cette procédure transforme (x, y) en une colonne de type 0 qui est adjacente à (u, v).

Soient a, z, b, c, r, d, s, t les sommets des quatre colonnes consécutives suivant (x, y) comme en Figure 3.5. Si $C \setminus \{y\}$ est encore un code identifiant de $\mathcal{S}_2$, alors nous pouvons enlever y à C et nous avons terminé.

Sinon, nous affirmons que, soit $C \setminus \{y\} \cup \{z\}$, soit $C \setminus \{y\} \cup \{b\}$, est un code identifiant de $\mathcal{S}_2$. En effet, si ni $C \setminus \{y\}$ ni $C \setminus \{y\} \cup \{z\}$ n'est un code identifiant de $\mathcal{S}_2$, cela signifie que y est nécessaire pour séparer z de l'un des sommets a ou c. Dans ce cas, remplacer y par b sépare z de a et c, et nous obtenons un code identifiant de $\mathcal{S}_2$ ayant une densité égale à celle de C.

Nous voulons répéter cette procédure pour toute colonne de type 2 voisine de deux colonnes de type 1, pour obtenir un code identifiant de $\mathcal{S}_2$ satisfaisant 3.3. Cependant, ce procédé peut éventuellement créer de nouvelles colonnes de type 2 voisines de deux colonnes de type 1 (voir par exemple le cas décrit en Figure 3.5). En ce cas, nous appliquons le procédé de nouveau, et ainsi de suite tant que, en appliquant le procédé, nous créons de nouvelles colonnes de type 2 voisines de deux colonnes de type 1. Nous affirmons que nous

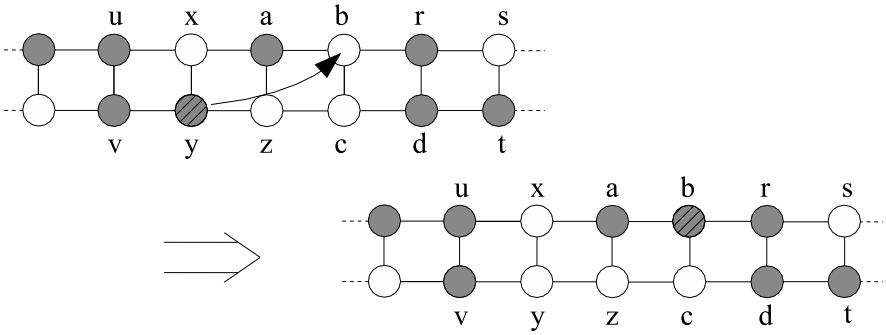


FIG. 3.5 – *Illustration du principe de décalage à droite.*

n'appliquons le procédé qu'un nombre fini de fois.

En effet, il suffit de remarquer que ce procédé crée une colonne de type 2 voisine de deux colonnes de type 1 si et seulement si nous sommes dans le cas décrit en Figure 3.5, où l'on a remplacé y par b pour séparer z de a (en effet, remplacer y par z ne peut pas créer une telle colonne, et si nous voulons séparer z de c alors nous savons que $d \notin C$). Dans ce cas, (r, d) devient une nouvelle colonne de type 2 voisine de deux colonnes de type 1.

Si l'application répétée de ce procédé ne termine pas, cela signifie qu'il y a une succession infinie de tels "blocs" u, v, x, y, a, z, b, c. Comme la densité d'un tel bloc est supérieure ou égale à $\frac{1}{2}$, alors nous obtenons une demi-bande $\mathcal{S}_2^+ \subseteq \mathcal{S}_2$ qui est telle que $C \cap \mathcal{S}_2^+$ a une densité supérieure ou égale à $\frac{1}{2}$. Ceci est une contradiction puisque nous pouvions supposer que C avait une densité strictement inférieure à $\frac{1}{2}$ (en effet, cela signifie que le complémentaire de $\mathcal{S}_2^+$ a une densité $d < d(C, \mathcal{S}_2) = d_1^*(\mathcal{S}_2)$, ce qui est absurde puisque $d_1^*(\mathcal{S}_2) = d_1^*(\mathcal{S}_2^+)$).

Ainsi, le voisinage $\mathcal{N}$ d'une colonne isolée est nécessairement identique (aux symétries près) à celui décrit en Figure 3.6. Ceci peut être obtenu par une étude exhaustive (les premières et dernières colonnes sont obtenues en utilisant (3.3)).

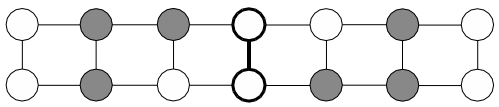


FIG. 3.6 – *Voisinage $\mathcal{N}$ d'une colonne isolée.*

Pour finir, nous avons besoin de prouver que :

> Aucune colonne de type 2 n'est adjacente à deux colonnes de type 0. (3.4)

En effet, dans le cas contraire, les deux sommets u, v d'une telle colonne auraient le même ensemble identifiant $\{u, v\}$.

Grâce à (3.4), nous savons qu'il est possible de partitionner $\mathcal{S}_2$ en des blocs $\mathcal{N}$, et en d'autres blocs, décrits en Figure 3.7.

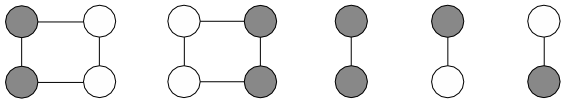


FIG. 3.7 – *Blocs partitionnant $\mathcal{S}_2 \setminus \{$toutes les copies du bloc $\mathcal{N}\}$.*

Comme $d(C, \mathcal{N}) = \frac{3}{7}$ et $d(C, B) \geq \frac{1}{2}$ pour tout bloc B décrit en Figure 3.7, alors on a $d(C, \mathcal{S}_2) \geq \frac{3}{7}$.

Pour conclure nous donnons en Figure 3.8 un code identifiant de $\mathcal{S}_2$ de densité $\frac{3}{7}$. □

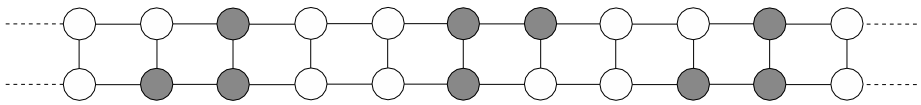


FIG. 3.8 – *Un code identifiant périodique de $\mathcal{S}_2$, de densité $\frac{3}{7}$.*

Nous pouvons utiliser la même technique pour trouver la densité optimum d'un code identifiant du chemin :

Proposition 3.3 ([BCHL04, DGM04])
Soit $n \geq 2$ et soit P_n le chemin de longueur n. Alors on a $d_1^*(P_n) = \left\lceil \frac{n+1}{2} \right\rceil / n$.□

3.2.2 Bornes générales

Nous allons donner des bornes inférieures et supérieures de $d_1^*(\mathcal{G}_{k\times n})$ et $d_1^*(\mathcal{S}_k)$ valables pour tout k et n assez grands. Nos calculs sont basés sur le résultat suivant :

Théorème 3.4 ([CGHLMPZ99, LM])
Nous avons

$$d_1^*(\mathbb{Z}^2) \;=\; \frac{7}{20}. \qquad \square$$

Ce résultat est dû, d'une part, à G. Cohen, S. Gravier, I. Honkala, A. Lobstein, M. Mollard, C. Payan et G. Zémor [CGHLMPZ99] – qui ont montré que $d_1^*(\mathbb{Z}^2) \leq \frac{7}{20}$ en exhibant un code ayant cette densité – et, d'autre part, à S. Litsyn et Y. Merksamer [LM] – qui ont montré que $d_1^*(\mathbb{Z}^2) \geq \frac{7}{20}$ en affinant la méthode d'investigation initiée en [CHLZ99].

Pour obtenir des bornes sur, par exemple, $d_1^*(\mathcal{G}_{k\times n})$, l'idée est la suivante. Considérons un code identifiant $C_{k\times n}$ de la grille $\mathcal{G}_{k\times n}$. Si l'on pave la grille infinie $\mathbb{Z}^2$ avec des copies de $\mathcal{G}_{k\times n}$ et de $C_{k\times n}$, alors nous obtenons C un code couvrant de $\mathbb{Z}^2$. C n'est pas nécessairement un code identifiant de $\mathbb{Z}^2$: les sommets appartenant aux "bords" de deux copies de $\mathcal{G}_{k\times n}$ ne sont éventuellement pas séparés par C (voir par exemple la Figure 3.9). Mais il suffit alors de faire un nombre raisonnable de corrections (c'est-à-dire un nombre linéaire en $k+n$) sur chaque copie de $C_{k\times n}$ pour obtenir C' qui soit un code identifiant de $\mathbb{Z}^2$ (voir Figure 3.9). Ceci nous fournit une borne inférieure sur $d_1^*(\mathcal{G}_{k\times n})$.

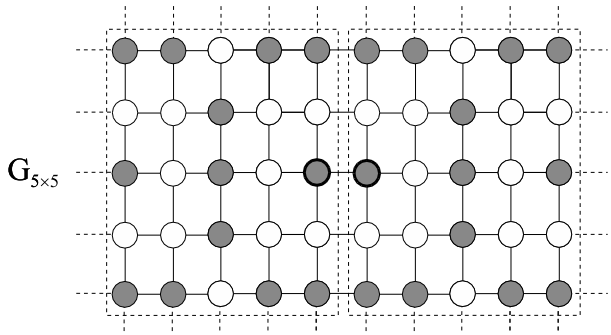


FIG. 3.9 – *Soit C le code couvrant de la grille infinie $\mathbb{Z}^2$ obtenu par translations d'un code identifiant de la grille finie $\mathcal{G}_{5\times 5}$. Les deux sommets en gras, appartenant aux bords de deux copies de $\mathcal{G}_{5\times 5}$, ne sont pas séparés par C. Cependant il suffit d'ajouter un voisin quelconque de l'un de ces deux sommets à C pour obtenir un code qui, par translation, est un code identifiant de la grille infinie $\mathbb{Z}^2$.*

Pour obtenir une borne supérieure sur $d_1^*(\mathcal{G}_{k\times n})$, nous commençons par considérer C un code identifiant de la grille infinie $\mathbb{Z}^2$. La trace de C sur une

sous-grille $\mathcal{G}_{k\times n}$ de $\mathbb{Z}^2$ nous donne un code $C_{k\times n}$ qui identifie les sommets de l'"intérieur" de $\mathcal{G}_{k\times n}$, mais éventuellement ne couvre ou ne sépare pas certains sommets du "bord" de $\mathcal{G}_{k\times n}$. En faisant un nombre linéaire (en $k+n$) de modifications sur $C_{k\times n}$, nous pouvons donc obtenir un code identifiant $C'_{k\times n}$ de $\mathcal{G}_{k\times n}$. Ceci nous fournit une borne supérieure sur $d_1^*(\mathcal{G}_{k\times n})$.

Cette technique intuitive de "couper-coller" nous avait déjà permis de trouver des bornes générales pour un autre problème, le problème de Δ-dislocation dans les graphes [GMP].

Théorème 3.5 ([DGM04])
Pour tout $k \geq 3$ on a

$$\frac{7}{20} - \frac{1}{2k} \;\leq\; d_1^*(\mathcal{S}_k) \;\leq\; \min\left(\frac{2}{5}, \frac{7}{20} + \frac{2}{k}\right)$$

et pour tout $k, n \geq 2$ on a

$$\frac{7}{20} - \frac{1}{2}\frac{k+n}{kn} \;\leq\; d_1^*(\mathcal{G}_{k\times n}) \;\leq\; \frac{7}{20} + 2\frac{k+n-2}{kn}.$$

Preuve : Comme nous utilisons la même technique pour $\mathcal{S}_k$ et $\mathcal{G}_{k\times n}$, nous allons ici seulement donner la preuve pour $\mathcal{S}_k$. La preuve pour $\mathcal{G}_{k\times n}$ peut être facilement déduite de celle pour $\mathcal{S}_k$.

La borne supérieure $\frac{2}{5}$ provient du fait que l'ensemble de sommets $\{(u,v) \in \{1,\ldots,k\} \times \mathbb{Z} \mid v \bmod 5 \in \{1,3\}\}$ est un code identifiant de $\mathcal{S}_k$ pour tout $k \geq 3$ (voir Figure 3.10).

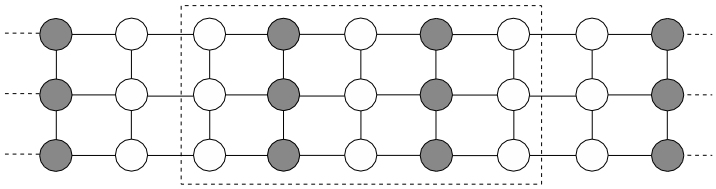


FIG. 3.10 – *Un code identifiant périodique de $\mathcal{S}_3$, de densité $\frac{2}{5}$.*

Pour l'autre borne supérieure, soit C un code identifiant optimum de $\mathbb{Z}^2$. Par le Théorème 3.4 nous savons que la densité de C est égale à $\frac{7}{20}$. Donc pour tout $k \geq 3$ il existe une bande $\mathcal{S}_k$ incluse dans $\mathbb{Z}^2$ telle que la trace de C sur $\mathcal{S}_k$ a une densité inférieure ou égale à $\frac{7}{20}$. Noter que $C_k := C \cap \mathcal{S}_k$, la trace de C sur $\mathcal{S}_k$, n'est peut-être pas un code identifiant de $\mathcal{S}_k$. Soit alors

$C'_k := C_k \bigcup \{1,k\} \times \mathbb{Z}$: nous affirmons que C'_k est un code identifiant de $\mathcal{S}_k$. En effet, il suffit de remarquer que tous les sommets de $\mathcal{S}_k \cap \{2,\ldots,k-1\} \times \mathbb{Z}$ sont déjà couverts et séparés entre eux par C_k, et séparés des sommets du bord $\mathcal{S}_k \cap \{1,k\} \times \mathbb{Z}$. Restent donc à couvrir et séparer les sommets de $\mathcal{S}_k \cap \{1,k\} \times \mathbb{Z}$, ce que l'on peut faire, par exemple, en ajoutant $\{1,k\} \times \mathbb{Z}$ à C_k. On obtient alors C'_k qui est un code identifiant de $\mathcal{S}_k$. La densité de C'_k étant inférieure ou égale à $\frac{1}{k}\frac{7}{20} \times (k+2)$, on obtient donc la borne escomptée :

$$d_1^*(\mathcal{S}_k) \;\leq\; d(C'_k, \mathcal{S}_k) \;\leq\; \frac{7}{20} + \frac{2}{k}.$$

Pour la borne inférieure, considérons C_k un code identifiant optimum de $\mathcal{S}_k$. Lorsque l'on pave $\mathbb{Z}^2$ par translations de $\mathcal{S}_k$ et C_k, on obtient C un code couvrant de $\mathbb{Z}^2$. Nous allons modifier C_k en un code C'_k tel que la translation de C'_k est un code identifiant de $\mathbb{Z}^2$. Il est facile de voir que, lorsque l'on translate C_k, les seuls cas dans lesquels deux sommets de deux copies de $\mathcal{S}_k$ ne sont pas séparés, sont les cas décrits en Figure 3.11.

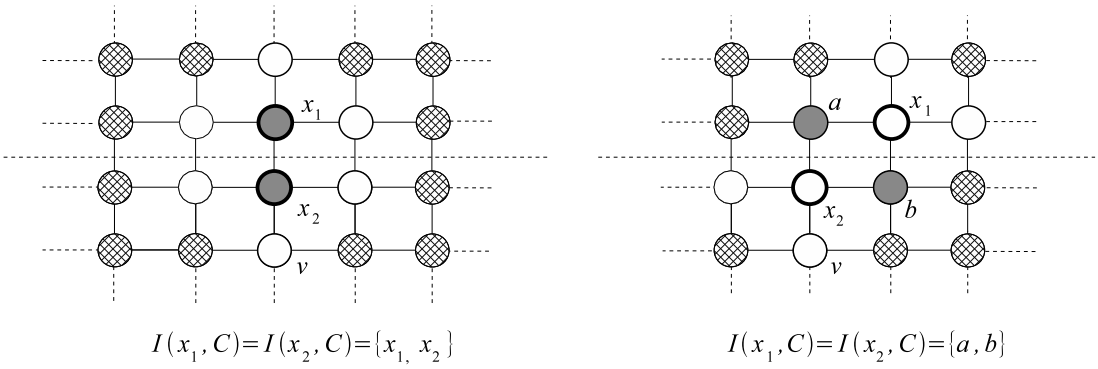


FIG. 3.11 – *Cas où deux sommets de deux copies de $\mathcal{S}_k$ ne sont pas séparés par les translations de C_k. Les sommets hachurés sont des sommets dont l'appartenance à C_k n'est pas spécifiée.*

Pour toute paire de sommets (x_1, x_2) qui ne sont pas séparés après translation de C_k, il suffit alors d'ajouter v à C_k pour obtenir un code couvrant de $\mathcal{S}_k$ qui sépare x_1 et x_2, où v est le sommet décrit en Figure 3.11. Nous obtenons alors C'_k, un code identifiant de $\mathcal{S}_k$ tel que sa translation de vecteur $(0,k)$ est un code identifiant C' de $\mathbb{Z}^2$. On a alors

$$d(C_k, \mathcal{S}_k) + \frac{1}{2k} \;\geq\; d(C'_k, \mathcal{S}_k) \;=\; d(C', \mathbb{Z}^2) \;\geq\; \frac{7}{20},$$

ce qui conduit à l'inégalité annoncée. □

Ces méthodes peuvent être facilement transposées à d'autres types de réseaux, en particulier aux grilles triangulaire, hexagonale et royale, déjà considérés dans [CHHL01, CHHL04, CHL02a, CHLZ99]. Nous pourrions également étendre nos résultats au cas des codes t-identifiants, avec $t \geq 2$. La seule différence est qu'il faudrait considérer t lignes sur le bord des grilles au lieu de une dans le cas des codes 1-identifiants. Nous trouverions alors des bornes en $d_t^*(\mathbb{Z}^2) \pm \Theta(t(k+n))$. Le point ennuyeux dans ceci est que la densité optimum d'un code t-identifiant de $\mathbb{Z}^2$ n'est pas connue dans le cas général $t \geq 2$. Nous disposons des bornes suivantes :

$$d_t^*(\mathbb{Z}^2) \ \geq \ \frac{3}{8t+4} \text{ pour tout } t \geq 2,$$

$$d_t^*(\mathbb{Z}^2) \ \leq \ \frac{2}{5t} \text{ pour tout } t \geq 2, \ t \text{ pair},$$

$$d_t^*(\mathbb{Z}^2) \ \leq \ \frac{2t}{5t^2-2t+1} \text{ pour tout } t \geq 2, \ t \text{ impair}.$$

La borne inférieure provient de [CHHL01], et les bornes supérieures de [HL02b]. Pour des petites valeurs de t il existe de meilleures bornes supérieures provenant de constructions *ad hoc* [CHL02a].

Les codes $(1, \leq \ell)$-identifiants ont aussi été étudiés dans $\mathbb{Z}^2$, pour $\ell = 2, 3$ [HL03a]. Les méthodes que nous venons de présenter sont bien entendu susceptibles d'être appliquées dans ces cas-là aussi.

Nous savons qu'il existe des codes t-identifiants optimaux *périodiques* de la bande $\mathcal{S}_k$, et nous savons que la période est bornée par une certaine constante $C = C(t)$ (voir le Théorème 2.3). Ainsi, pour rechercher un code t-identifiant optimum périodique dans la bande $\mathcal{S}_k$, nous pouvons nous ramener à rechercher des codes t-identifiants (non nécessairement périodiques) dans les tores $\mathcal{T}_l$ pour tout $l \leq C$. Un tore de longueur inférieure ou égale à $C(t)$ ayant un code t-identifiant de densité minimum nous fournira alors un code t-identifiant périodique de la bande $\mathcal{S}_k$. L'étude des codes identifiants dans les tores est donc une activité qui est très liée à celle de l'étude de tels codes dans les bandes.

Les tores étant des produits de cycles, l'étude des cycles est donc une voie possible pour comprendre les codes identifiants dans les bandes et les grilles, ce qui nous amène à la section suivante...

3.3 Cycles

Dans cette section nous étudions les codes t-identifiants dans les cycles non-orientés pour tout $t \geq 1$. Le cycle à n sommets sera désigné par $\mathcal{C}_n$ et ses sommets par $v_0, v_1, \ldots, v_{n-1}$. Pour n assez grand, nous allons résoudre complètement le cas n pair pour tout $t \geq 1$, et nous allons donner des bornes pour le cas n impair, qui s'avère plus difficile. La cardinalité minimum d'un code t-identifiant de $\mathcal{C}_n$ sera noté $M_t(\mathcal{C}_n)$.

Les résultats présentés ici proviennent de [GMS] et [BCHL04]. Les résultats de [GMS] ont été en partie obtenus en collaboration avec A. Semri lors de son séjour à Grenoble en hiver 2002-2003.

3.3.1 Borne inférieure générale

Pour commencer, nous allons donner une borne inférieure générale sur $M_t(\mathcal{C}_n)$, basée sur le fait qu'un code t-identifiant est nécessairement un code t-séparant deux sommets consécutifs du cycle.

Théorème 3.6 ([GMS])
Pour tout $t \geq 1$ et tout $n \geq 2t+2$, on a

$$M_t(\mathcal{C}_n) \;\geq\; \operatorname{pgcd}(2t+1, n) \left\lceil \frac{n}{2\operatorname{pgcd}(2t+1, n)} \right\rceil .$$

Preuve : Soient $t \geq 1$ et $n \geq 2t+2$, et soit C un code t-identifiant de $\mathcal{C}_n$. Nous savons que C est aussi un code t-séparateur de $\mathcal{C}_n$. Puisque pour tout $i \in \mathbb{Z}_n$ nous avons $B_t(v_i) \Delta B_t(v_{i+1}) = \{v_{i-t}, v_{i+1+t}\}$, alors pour tout $i \in \mathbb{Z}_n$ l'un des deux sommets v_{i-t} ou v_{i+1+t} doit appartenir à C.

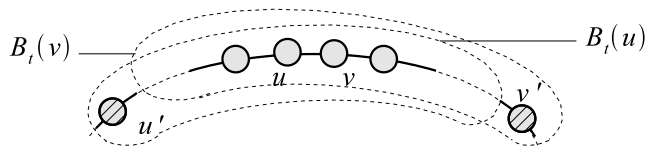


FIG. 3.12 – *Un des deux sommets v_{i-t} ou v_{i+1+t} appartient à C.*

Soit $\mathcal{C}'_{(n,t)}$ le graphe ayant pour ensemble de sommets $\{v_i \mid i \in \mathbb{Z}_n\}$ tel que $v_{i-t}v_{i+t+1}$ soit une arête de $\mathcal{C}'_{(n,t)}$ pour tout $i \in \mathbb{Z}_n$. D'après ce qui précède,

l'ensemble C couvre toutes les arêtes de $\mathcal{C}'_{(n,t)}$: C est donc un transversal de $\mathcal{C}'_{(n,t)}$. Ainsi, la cardinalité minimum d'un transversal de $\mathcal{C}'_{(n,t)}$ est une borne inférieure pour la cardinalité de C.

Soit $a = \text{pgcd}(2t+1, n)$, et soit $n' = \frac{n}{a}$. $\mathcal{C}'_{(n,t)}$ est l'union disjointe de a cycles de n' sommets. La cardinalité minimum d'un transversal de chacun de ces cycles étant $\lceil \frac{n'}{2} \rceil$, on obtient

$$|C| \;\geq\; a \left\lceil \frac{n'}{2} \right\rceil,$$

qui est l'inégalité annoncée. □

Ce résultat généralise le Théorème 9 de [BCHL04]. En général, un code t-séparateur de $\mathcal{C}_n$ n'est pas un code t-identifiant de $\mathcal{C}_n$. Il peut par exemple arriver qu'un des sommets de $\mathcal{C}_n$ ne soit pas t-couvert par un transversal de $\mathcal{C}'_{(n,t)}$ (voir Figure 3.13). Et même dans le cas où tous les sommets de $\mathcal{C}'_{(n,t)}$ sont t-couverts, il peut arriver que deux sommets (non-consécutifs) de $\mathcal{C}'_{(n,t)}$ ne soient pas t-séparés (voir Figure 3.13).

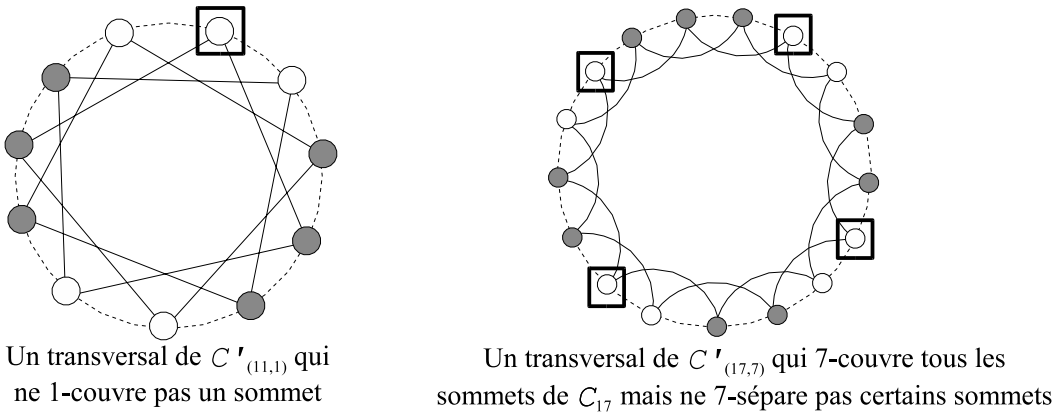


FIG. 3.13 – *Un transversal de $\mathcal{C}'_{(n,t)}$ n'est en général pas un code t-identifiant de $\mathcal{C}_n$.*

3.3.2 Cas n pair

Dans le cas où $n \geq 2t+4$ est pair, nous avons de la chance car les choses décrites en Figure 3.13 ne se produisent pas. En effet, la borne inférieure du Théorème 3.6 donne simplement $\frac{n}{2}$, et l'ensemble $\{v_i \mid 1 \leq i \leq n, i \text{ impair}\}$ est un code t-identifiant trivial de $\mathcal{C}_n$. On retrouve ainsi le résultat suivant :

Théorème 3.7 ([BCHL04])
Pour tout $t \geq 1$ et $n \geq 2t+4$, n pair, on a

$$M_t(\mathcal{C}_n) \;=\; \frac{n}{2}. \qquad \square$$

Dans la situation extrême où $n = 2t+2$, la situation est un peu différente. Pour tout $v_i \in V$, $B_t(v_i)$ consiste en tous les sommets du cycle sauf le sommet v_{i+t+1}. Comme au plus un sommet du cycle peut être couvert par C tout entier, alors ceci implique que $M_t(\mathcal{C}_{2t+2}) \geq n-1$, et comme pour tout $v \in V$, $V \setminus \{v\}$ est un code t-identifiant de $\mathcal{C}_n$, alors on a :

Proposition 3.4 ([BCHL04])
Pour tout $t \geq 1$, on a

$$M_t(\mathcal{C}_{2t+2}) \;=\; 2t+1. \qquad \square$$

3.3.3 Cas n impair

Dans le cas où n est impair, la borne inférieure du Théorème 3.6 n'est en général pas atteinte. Dans le paragraphe suivant nous donnons une borne supérieure de $M_t(\mathcal{C}_n)$ qui diffère d'au plus t de la borne inférieure. Ensuite, nous étudions les cas particuliers $t = 1$, $n = 2t+3$ et $2t+1 \mid n$. Enfin, nous donnons des conditions sur t et n pour que la borne inférieure soit atteinte.

3.3.3.1 Une borne supérieure

Lemme 3.1 ([GMS])
Pour tout $t \geq 1$ et $n \geq 2t+3$ impair on a

$$M_t(\mathcal{C}_n) \;\leq\; \frac{n+1}{2} + t.$$

Preuve : Nous montrons que l'ensemble $C := \{v_i \mid i = 0, \ldots, 2t+1\} \cup \{v_i \mid i \text{ impair }, 2t+3 \leq i \leq n-1\}$ est un code t-identifiant de $\mathcal{C}_n$. Appelons *barrière* l'ensemble $B := \{v_i \mid i = 0, \ldots, 2t+1\}$. On dira qu'un sommet $v_i \in V$ *touche* la barrière si et seulement si $B_t(v_i) \cap B \neq \emptyset$. Il est facile de voir que tous les sommets de $\mathcal{C}_n$ sont t-couverts par C. Il nous reste donc à montrer qu'ils sont tous t-séparés par C. Tout d'abord, montrons qu'un sommet v_i touchant la barrière est t-séparé de tout sommet $v_j \neq v_i$. En effet, si v_j ne touche pas la barrière alors v_i et v_j sont trivialement t-séparés, car la barrière est un

sous-ensemble de C. Si v_j touche lui aussi la barrière, alors par symétrie il suffit de considérer le cas $t+1 \leq i < j \leq n-t-1$, où v_i et v_j sont t-séparés par v_{i-t}, et le cas $t+1 \leq i \leq n-t-1, -t \leq j \leq t$, où v_i et v_j sont t-séparés par v_0. Enfin, il nous reste à montrer que deux sommets v_i et v_j ne touchant ni l'un ni l'autre la barrière sont t-séparés, ce qui est facile : supposant $i < j$, si $j = i+1$ alors ils sont, soit séparés par v_{i-t}, soit par v_{i+1+t} (soit $i-t$ soit $i+t+1$ est impair) ; et dans les autres cas ils sont t-séparés, soit par v_{i-t}, soit par v_{i-t+1}. □

Ce résultat nous dit que $M_t(\mathcal{C}_n)$ se trouve dans un intervalle d'amplitude au plus t. Pour mettre ceci en évidence, on peut reformuler le Théorème 3.6 avec ce dernier lemme :

Théorème 3.8 ([GMS])
Pout tout $t \geq 1$ et $n \geq 2t+3$ impair, on a

$$\frac{n+1}{2} + \frac{\text{pgcd}(2t+1, n) - 1}{2} \;\leq\; M_t(\mathcal{C}_n) \;\leq\; \frac{n+1}{2} + t. \qquad \square$$

Parfois c'est la borne supérieure qui sera atteinte et non la borne inférieure (par exemple lorsque $t = 1$), éventuellement les deux à la fois (lorsque $2t+1$ divise n), et parfois ni l'une ni l'autre (par exemple dans le cas où $n = 2t+3$). Ce sont ces cas que nous allons étudier dans la suite.

3.3.3.2 Cas particuliers

3.3.3.2.1 Cas $t = 1$

Si $t = 1$, alors pour tout n multiple de 3 les deux bornes du Théorème 3.8 sont égales ; dans le cas général où $n \not\equiv 0$ [3] alors c'est la borne supérieure qui est atteinte :

Théorème 3.9 ([GMS])
Pour tout $n \geq 7$, n impair, on a

$$M_1(\mathcal{C}_n) \;=\; \frac{n+1}{2} + 1.$$

Preuve : Soit $n \geq 7$ impair. Par le Théorème 3.8 il suffit de montrer qu'il n'existe pas de code 1-identifiant de $\mathcal{C}_n$ de cardinalité $\frac{n+1}{2}$. Par l'absurde, supposons qu'il existe C un code 1-identifiant de $\mathcal{C}_n$ de cardinalité $\frac{n+1}{2}$. Dans ce cas, il y a au moins deux sommets de C qui sont consécutifs sur le cycle,

disons v_1 et v_2 sans perte de généralité. Pour séparer v_1 de v_2, soit v_0 soit v_3 appartient aussi à C ; disons $v_3 \in C$, toujours sans perte de généralité. Nous avons alors $|I_1(v_2, C)| = 3$, et pour tout $c \in C \setminus \{v_2\}$ il y a au plus un sommet ayant $\{c\}$ comme ensemble identifiant. Ceci nous donne

$$3|C| \ \geq \ \sum_{i=0}^{n-1} |I_1(v_i, C)| \ \geq \ 1 \times 3 + (|C| - 1) \times 2 + (n - |C|) \times 1.$$

Puisque $|C| = \frac{n+1}{2}$, alors il y a égalité, et on a :

> Pour tout $c \in C \setminus \{v_2\}$ il y a un sommet ayant $\{c\}$ comme ensemble identifiant. (3.5)

En particulier, $v_0 \notin C$ (considérer $c = v_1$), $v_4 \notin C$, et $v_5 \notin C$ (prendre $c = v_3$). Pour 1-couvrir v_5 il nous faut alors $v_6 \in C$. Si $n = 7$, alors il y a contradiction car $I_1(v_5, C) = I_1(v_6, C) = \{v_6\}$. Sinon, $n \geq 9$ et $v_7 \in C$ pour 1-séparer v_5 de v_6. Ensuite, $v_8 \in C$ pour 1-séparer v_6 de v_7, ce qui reporte la contradiction sur v_7, qui viole la condition (3.5). □

Pour compléter l'étude des cycles impairs dans le cas où $t = 1$, il est utile de préciser que $M_1(\mathcal{C}_5) = 3$ ($\mathcal{C}_3$ n'admet pas de code 1-identifiant).

3.3.3.2.2 Cas $n = 2t + 3$

Dans ce cas aucune des bornes du Théorème 3.8 n'est atteinte, et la valeur de $M_t(\mathcal{C}_n)$ est $\left\lfloor \frac{2n}{3} \right\rfloor$:

Théorème 3.10 ([GMS])
Pour tout $t \geq 1$, on a

$$M_t(\mathcal{C}_{2t+3}) \ = \ \left\lfloor \frac{4t}{3} \right\rfloor + 2.$$

Preuve : Soit C un code t-identifiant de $\mathcal{C}_{2t+3}$. Pour tout i on a alors

$$I_t(v_{i-t-1}, C) = C \setminus \{v_i, v_{i+1}\}. \tag{3.6}$$

Par conséquent :
(a) il y a au plus une paire $\{v_i, v_{i+1}\}$ telle que $v_i \notin C$ et $v_{i+1} \notin C$;
(b) il n'y a aucune paire $\{v_i, v_{i+2}\}$ telle que $v_i \notin C$ et $v_{i+2} \notin C$.
En effet, (a) s'ensuit de (3.6), et (b) provient du fait que $I_t(v_{i-t-1}, C) \neq I_t(v_{i-t}, C)$. Partitionnons alors les sommets de $\mathcal{C}_{2t+3}$ en ensembles de trois

sommets consécutifs, plus éventuellement un ensemble de un ou deux sommets. S'il y a une paire $\{v_i, v_{i+1}\}$ telle que $v_i \notin C$ et $v_{i+1} \notin C$ alors on peut partitionner les sommets de sorte que v_i et v_{i+1} ne soient pas dans la même partie (par (a) il existe au plus une telle paire). Par (b), toutes les parties à trois sommets de la partition contiennent au moins deux éléments de C. L'inégalité $M_t(\mathcal{C}_{2t+3}) \geq \left\lfloor \frac{2n}{3} \right\rfloor$ en découle, et pour conclure il suffit d'exhiber des codes t-identifiants ayant la cardinalité désirée (voir Figure 3.14). □

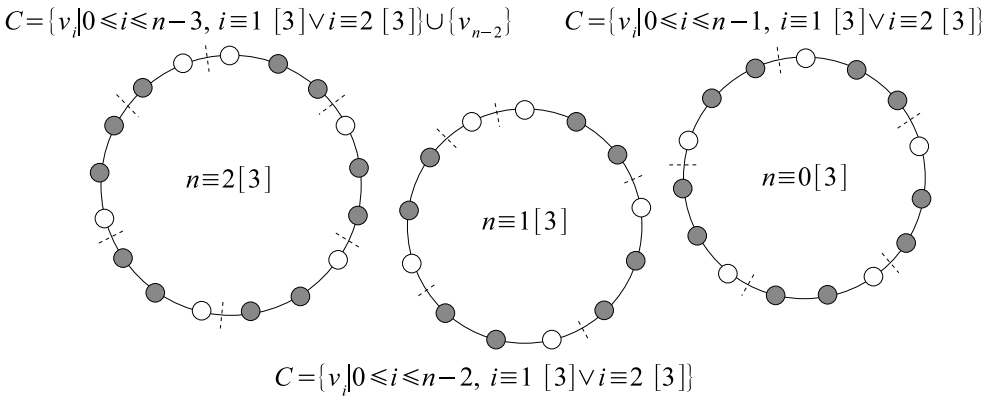


FIG. 3.14 – *Codes t-identifiants de cycles de $2t+3$ sommets.*

3.3.3.2.3 Cas où $2t+1$ divise n

Les deux bornes du Théorème 3.8 sont égales si et seulement si $\text{pgcd}(2t+1, n) = 2t+1$, c'est-à-dire $2t+1$ divise n, d'où :

Théorème 3.11 ([GMS])
Soit $t \geq 1$, et soit $n \geq 2t+3$ un entier impair tel que $2t+1$ divise n. Alors

$$M_t(\mathcal{C}_n) \;=\; \frac{n+1}{2} + t. \qquad \square$$

3.3.3.3 Quelques conditions sur t et n pour atteindre la borne inférieure

Dans le cas $t \geq 2$ nous pouvons donner quelques conditions sur t et n pour atteindre la borne inférieure du Théorème 3.8. Comme dans la preuve du Théorème 3.6, nous notons $\mathcal{C}'_{(n,t)}$ le graphe ayant pour sommets $v_0, v_1, \ldots, v_{n-1}$

et tel que $v_{i-t}v_{i+t+1}$ soit une arête de $\mathcal{C}'_{(n,t)}$ pour tout i. Pour commencer, nous montrons que si n est assez grand devant t alors tout transversal de $\mathcal{C}'_{(n,t)}$ t-sépare les sommets de $\mathcal{C}_n$ à distance au plus $2t$:

Lemme 3.2 ([GMS])
Soit $t \geq 1$, soit n un entier impair tel que $n \geq 3t+2$, et soit C un transversal de $\mathcal{C}'_{(n,t)}$. Alors tous les sommets u, v de $\mathcal{C}_n$ tels que $d(u,v) \leq 2t$ sont t-séparés par C.

Preuve : Comme C est un transversal de $\mathcal{C}'_{(n,t)}$, alors deux sommets consécutifs de $\mathcal{C}_n$ sont t-séparés (par définition de $\mathcal{C}'_{(n,t)}$). Soient u et v deux sommets non-consécutifs de $\mathcal{C}_n$ tels que $d(u,v) \leq 2t$. Sans perte de généralité, on peut supposer que $u = v_0$ et $v = v_i$, avec $2 \leq i \leq 2t$. Soit $j = \left\lfloor \frac{i}{2} \right\rfloor$, et soient $x = v_j$ et $y = v_{j+1}$. Soient $x' = v_{j-t}$ et $y' = v_{j+1+t}$: x' et y' sont tels que $B_t(x) \Delta B_t(y) = \{x', y'\}$, et $x'y'$ est une arête de $\mathcal{C}'_{(n,t)}$. Comme C est un transversal de $\mathcal{C}'_{(n,t)}$, alors x' ou y' appartient à C. Pour conclure il suffit donc de montrer que $x' \in B_t(u) \setminus B_t(v)$ et $y' \in B_t(v) \setminus B_t(u)$.

Comme $d(u,x') = |j-t|$ et $d(v,y') = |i-j-1-t|$ alors $x' \in B_t(u)$ et $y' \in B_t(v)$. Comme $n \geq 3t+2$ alors le plus court chemin entre x' et v passe par x : $d(x',v) = d(x',x) + d(x,v) = t + d(x,v) \geq t+1$, donc $x' \notin B_t(v)$. De même $d(y',u) = d(y',y) + d(y,u) = t + d(y,u) \geq t+1$, donc $y' \notin B_t(u)$. □

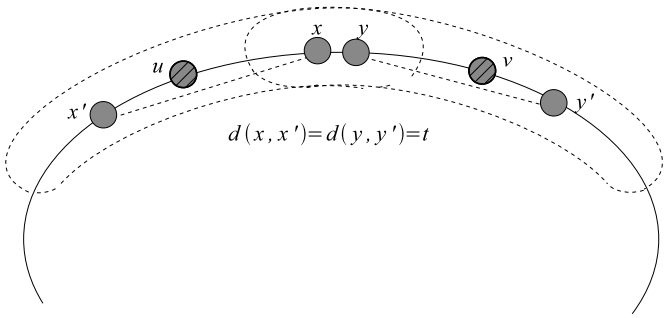


FIG. 3.15 – *Les sommets u, v, x, y, x', y', dans le cas où $t+1 \leq d(u,v) \leq 2t$.*

Pour obtenir le résultat suivant, il suffit de remarquer qu'il suffit de t-couvrir deux sommets éloignés pour les t-séparer. Ainsi, pour n assez grand, tout transversal de $\mathcal{C}'_{(n,t)}$ qui t-couvre les sommets de $\mathcal{C}_n$ est un code t-identifiant de $\mathcal{C}_n$.

Lemme 3.3 ([GMS])
Soient $t \geq 1$ et n un entier impair tel que $n \geq 3t+2$. Soit C un transversal

de ${\mathcal{C}'}_{(n,t)}$ *qui* t*-couvre les sommets de* $\mathcal{C}_n$*. Alors* C *est un code* t*-identifiant de* $\mathcal{C}_n$*.*

Preuve : Soit C un transversal de ${\mathcal{C}'}_{(n,t)}$ qui t-couvre les sommets de $\mathcal{C}_n$. Pour prouver que C est un code t-identifiant de $\mathcal{C}_n$, il suffit de montrer que deux sommets distincts de $\mathcal{C}_n$ sont t-séparés par C. Soient u et v deux sommets distincts de $\mathcal{C}_n$. Si $d(u,v) \leq 2t$, alors par le Lemme 3.2 u et v sont t-séparés. Si $d(u,v) \geq 2t+1$, alors les boules $B_t(u)$ et $B_t(v)$ sont disjointes. Comme u et v sont t-couverts, alors ils sont t-séparés. □

Maintenant, nous montrons que si n n'est pas trop grand par rapport à t, alors tout transversal de ${\mathcal{C}'}_{(n,t)}$ t-couvre les sommets de $\mathcal{C}_n$. Par le Lemme précédent, nous en déduisons que, dans ces conditions, tout transversal de ${\mathcal{C}'}_{(n,t)}$ est un code t-identifiant de $\mathcal{C}_n$.

Théorème 3.12 ([GMS])
Soient $t \geq 1$ *et* n *un entier impair tel que* $3t+2 \leq n \leq 4t+1$*. Soit* C *un transversal de* ${\mathcal{C}'}_{(n,t)}$*. Alors* C *est un code* t*-identifiant de* $\mathcal{C}_n$*.*

Preuve : Nous montrons que l'existence d'un transversal de ${\mathcal{C}'}_{(n,t)}$ ne t-couvrant pas un sommet de $\mathcal{C}_n$ implique $n \geq 4t+2$. Le Lemme 3.3 nous permet alors de conclure. Soit C un tel transversal, et supposons que v_0 n'est pas t-couvert par C (voir Figure 3.16). En ce cas $v_i \notin C$ pour tout $i = -t, \ldots, t$. Mais pour tout $i = 1, \ldots, t$ les sommets v_i et v_{i-1} sont t-séparés, puisqu'un transversal de ${\mathcal{C}'}_{(n,t)}$ t-sépare précisément toutes les paires de sommets consécutifs de $\mathcal{C}_n$. Puisque $v_{i-1-t} \notin C$, alors nécessairement $v_{i+t} \in C$ pour tout $i = 1, \ldots, t$. De même, pour t-séparer v_j et v_{j-1}, il faut que $v_{j+t} \in C$ pour tout $j = t+1, \ldots, 2t$. Enfin, pour t-séparer v_{2t} de v_{2t+1}, le sommet v_{3t+1} doit être dans C. Ainsi le cycle $\mathcal{C}_n$ a au moins $4t+2$ sommets. □

Nous pouvons aussi utiliser cet argument dans l'autre sens de parcours du cycle. Dans le cas où $\text{pgcd}(2t+1, n) = 1$, cela nous donne :

Théorème 3.13 ([GMS])
Soient $t \geq 1$ *et* n *impair tels que* $\text{pgcd}(2t+1, n) = 1$ *et* $4t+5 \leq n \leq 8t+1$*. Alors tout transversal de* ${\mathcal{C}'}_{(n,t)}$ *est un code* t*-identifiant de* $\mathcal{C}_n$*.*

Preuve : Soit C un transversal optimum de $\mathcal{C}'_{(n,t)}$. Comme $\text{pgcd}(2t+1, n) = 1$, alors $|C| = \frac{n+1}{2}$. Comme dans la preuve du théorème précédent, supposons que C ne t-couvre pas v_0. En utilisant le même argument que précédemment, les sommets $v_{t+1}, \ldots, v_{3t+1}$ appartiennent tous à C. En parcourant le cycle

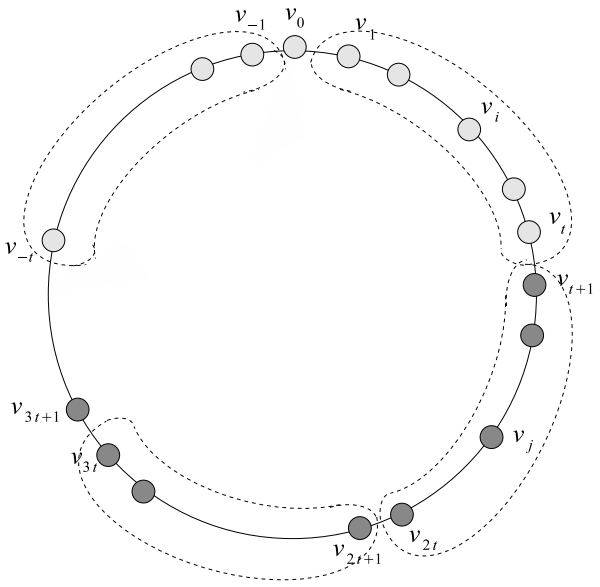


FIG. 3.16 – *Si $v_{-t}, \ldots, v_t$ n'appartiennent pas à C, alors C doit contenir $\{v_{t+1}, \ldots, v_{3t+1}\}$.*

dans l'autre sens, nous avons que $v_{-(t+1)}, \ldots, v_{-(3t+1)}$ sont aussi dans C. Si $4t+5 \leq n \leq 6t+3$, alors tous les sommets de $\mathcal{C}_n$, hormis les sommets $v_{-t}, \ldots, v_0, \ldots, v_t$, sont dans C. Ceci implique $|C| = n - (2t+1) > \frac{n+1}{2}$: contradiction. Si $6t+5 \leq n \leq 8t+1$, alors C contient au moins $4t+2$ sommets, ce qui contredit également $|C| = \frac{n+1}{2}$. □

Dans le cas $n = 4t+3$, alors $\text{pgcd}(2t+1, n) = 1$, et il suffit d'ajouter un sommet à un transversal optimum de $\mathcal{C}'_{(n,t)}$ pour obtenir un code t-identifiant de $\mathcal{C}_n$.

Proposition 3.5 ([GMS])
Soit $t \geq 1$. Alors on a :

$$M_t(\mathcal{C}_{4t+3}) \;=\; 2t+3.$$

Preuve : Il y a essentiellement un seul transversal optimum C de $\mathcal{C}'_{(n,t)}$, de cardinalité $2t+2$. En commençant par un sommet qui n'est pas dans C, parcourons $\mathcal{C}'_{(n,t)}$: nous alternons entre des sommets de C et des sommets qui ne sont pas dans C, jusqu'à atteindre les deux derniers sommets, qui sont tous deux dans C. Sans perte de généralité, nous pouvons supposer que $v_t \notin C$: ainsi $v_{3t+1} \in C$, puis $v_{t-1} \notin C$, et ainsi de suite, jusqu'à atteindre

v_{t+1} et v_{3t+2} qui sont tous deux dans C. Nous obtenons $C = \{v_{t+1}, \ldots, v_{3t+2}\}$, qui ne t-couvre pas v_0, donc C n'est pas un code t-identifiant de $\mathcal{C}_n$. Puisque $\mathcal{C}_n$ admet un code t-identifiant de cardinalité $2t+3$ (par exemple en ajoutant v_0 au C précédent), nous obtenons le résultat escompté. □

Si n est assez grand, nous montrons que si ${\mathcal{C}'}_{(n,t)}$ est l'union disjointe de plus de deux cycles (*i.e.* si $\text{pgcd}(2t+1, n) \neq 1$), alors nous pouvons choisir un transversal optimum de ${\mathcal{C}'}_{(n,t)}$ qui t-couvre les sommets de $\mathcal{C}_n$. Par le Lemme 3.3, ce transversal est alors un code t-identifiant de $\mathcal{C}_n$.

Théorème 3.14 ([GMS])
Soient $t \geq 1$ et n un entier impair tels que $n \geq 3t+2$ et $\text{pgcd}(2t+1, n) \neq 1$. *Alors il existe un transversal optimum de ${\mathcal{C}'}_{(n,t)}$ qui est un code t-identifiant de $\mathcal{C}_n$.*

Preuve : Soient $a = \text{pgcd}(2t+1, n)$ et $n' = \frac{n}{a}$. Nous savons que $\mathcal{C}'_{(n,t)}$ consiste en l'union disjointe de a cycles de n' sommets. Notons $v_k^{(j)}$ le sommet numéro k du cycle numéro j de $\mathcal{C}'_{(n,t)}$, de sorte que $v_k^{(j)} = v_{j+(2t+1)k}$ pour tout $j \in \mathbb{Z}_a$ et $k \in \mathbb{Z}_{n'}$. Pour tout $j \in \mathbb{Z}_a$, les ensembles $T^{(j)}_{\texttt{pair}} := \{v_k^{(j)} \mid k \text{ pair}\}$ et $T^{(j)}_{\texttt{impair}} := \{v_k^{(j)} \mid k \text{ impair}\} \cup \{v_0^{(j)}\}$ sont des transversaux optimums du cycle numéro j de $\mathcal{C}'_{(n,t)}$. Ainsi,

$$T := T^{(0)}_{\texttt{pair}} \bigcup T^{(1)}_{\texttt{impair}} \bigcup \left(\bigcup_{j=2}^{a-1} T^{(j)}_{\texttt{pair}} \right)$$

est un transversal optimum de $\mathcal{C}'_{(n,t)}$. Nous affirmons que T t-couvre tous les sommets de $\mathcal{C}_n$. En effet, considérons $2t+1$ sommets consécutifs du cycle $\mathcal{C}_n$. Comme $a \leq 2t+1$, alors pour un certain $k \in \mathbb{Z}_{n'}$, le sommet $v_k^{(1)}$ et l'un des deux sommets $v_k^{(0)}$ ou $v_k^{(2)}$, apparaissent dans les $2t+1$ sommets considérés. Si k est pair alors $v_k^{(0)}$ ou $v_k^{(2)}$ est dans T, et si k est impair alors $v_k^{(1)} \in T$. Comme $n \geq 3t+2$, nous concluons alors avec le Lemme 3.3. □

3.3.4 Pour conclure sur les cycles

Dans cette section nous avons donné quelques résultats sur les cycles. Le cas où n, la longueur du cycle, est pair est complètement résolu (il l'avait déjà été dans [BCHL04]), et le cas où n est impair a été étudié en détail.

Dans le cas n pair, nous rappelons les résultats de [BCHL04] :

– $M_t(\mathcal{C}_n) = \frac{n}{2}$ pour tout $t \geq 1$ et tout $n \geq 2t+4$, n pair (Théorème 3.7),
– $M_t(\mathcal{C}_{2t+2}) = 2t+1$ pour tout $t \geq 1$ (Proposition 3.4).

Dans le cas n impair, nous avons obtenu une borne supérieure et une borne inférieure qui diffèrent d'au plus t (Théorème 3.8) :

$$\frac{n+1}{2} + \frac{\text{pgcd}(2t+1,n)-1}{2} \;\leq\; M_t(\mathcal{C}_n) \;\leq\; \frac{n+1}{2} + t$$

pour tout $t \geq 1$, $n \geq 2t+3$, n impair.

Dans certains cas il nous a été possible de calculer $M_t(\mathcal{C}_n)$. Nous avons obtenu les résultats suivants :

– $M_1(\mathcal{C}_5) = 3$ (Théorème 3.10),
– $M_1(\mathcal{C}_n) = \frac{n+1}{2} + 1$ pour tout $n \geq 7$, n impair (Théorème 3.9),
– $M_t(\mathcal{C}_{2t+3}) = \left\lfloor \frac{4t}{3} \right\rfloor + 2$ pour tout $t \geq 1$ (Théorème 3.10),
– $M_t(\mathcal{C}_n) = \frac{n+1}{2} + t$ pour tout $t \geq 1$, $n \geq 2t+3$ tels que $2t+1$ divise n, n impair (Théorème 3.11),
– $M_t(\mathcal{C}_n) = \text{pgcd}(2t+1,n) \left\lceil \frac{n}{2\,\text{pgcd}(2t+1,n)} \right\rceil$ pour tout $t \geq 1$ et n impair tels que $3t+2 \leq n \leq 4t+1$ (Théorème 3.12),
– $M_t(\mathcal{C}_n) = \frac{n+1}{2}$ pour tout $t \geq 1$ et n impair tels que $4t+5 \leq n \leq 8t+1$ et $\text{pgcd}(2t+1,n) = 1$ (Théorème 3.13),
– $M_t(\mathcal{C}_{4t+3}) = 2t+3$ (Proposition 3.5),
– $M_t(\mathcal{C}_n) = \text{pgcd}(2t+1,n) \left\lceil \frac{n}{2\,\text{pgcd}(2t+1,n)} \right\rceil$ pour tout $t \geq 1$, $n \geq 3t+2$, n impair, tels que $\text{pgcd}(2t+1,n) \neq 1$ (Théorème 3.14).

Que reste-t-il à faire sur les cycles ? Le cas $\text{pgcd}(2t+1,n) \neq 1$ est presque résolu, puisque seul le cas particulier où $\text{pgcd}(2t+1,n) \neq 1$ et $2t+5 \leq n \leq 3t+1$ reste à étudier. C'est le cas où $\text{pgcd}(2t+1,n) = 1$ qui reste le moins connu, lorsque la borne inférieure et la borne supérieure diffèrent d'exactement t. En ce cas, nous savons que le graphe $\mathcal{C}'_{(n,t)}$ consiste en un seul cycle de cardinalité n, et il y a essentiellement un seul transversal optimum de $\mathcal{C}'_{(n,t)}$. Combien de sommets faut-il ajouter à ce transversal pour obtenir un code t-identifiant de $\mathcal{C}_n$?

Nous suspectons qu'un hypergraphe pourrait être utilisé pour répondre à cette question. En effet, en remarquant qu'un code t-identifiant t-séparait en particulier tous les sommet consécutifs de $\mathcal{C}_n$, nous avons défini $\mathcal{C}'_{(n,t)}$ comme le graphe des différences symétriques des voisinages étendus des paires de sommets consécutifs de $\mathcal{C}_n$. Un code t-identifiant de $\mathcal{C}_n$ étant un transversal de $\mathcal{C}'_{(n,t)}$, nous nous sommes ramenés à la recherche d'un transversal de $\mathcal{C}'_{(n,t)}$.

De la même façon, nous pourrions définir $\mathcal{H}_{(n,t)}$ comme l'hypergraphe

dont les sommets sont les sommets de $\mathcal{C}_n$ et les hyperarêtes sont les ensembles $B_t(u)\Delta B_t(v) \cup B_t(v)\Delta B_t(w)$, où u, v, w sont trois sommets consécutifs de $\mathcal{C}_n$ (voir Figure 3.17). Comme tout code t-identifiant de $\mathcal{C}_n$ t-sépare en particulier tous les triplets de sommets consécutifs de $\mathcal{C}_n$, alors un code t-identifiant C de $\mathcal{C}_n$ est un transversal de $\mathcal{H}_{(n,t)}$ tel que toute hyperarête de $\mathcal{H}_{(n,t)}$ contient au moins deux sommets de C.

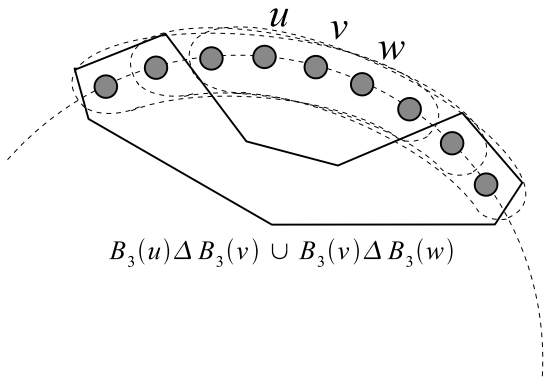


FIG. 3.17 – *Une hyperarête de l'hypergraphe $\mathcal{H}_{(n,3)}$. Pour 3-séparer u, v et w, il faut au moins que deux sommets parmi $B_t(u)\Delta B_t(v) \cup B_t(v)\Delta B_t(w)$ soient dans le code 3-identifiant de C.*

L'hypergraphe $\mathcal{H}_{(n,t)}$ serait une extension de notre stratégie utilisant $\mathcal{C}'_{(n,t)}$. Des hypergraphes de différences symétriques de voisinages étendus d'ensembles de sommets ont déjà été utilisés, de façon implicite, dans [CHHL01, CHHL04] par exemple.

Chapitre 4

Quelques questions sur des problèmes extrémaux

Dans ce chapitre, nous proposons quelques questions extrémales qui nous ont semblé naturelles, et nous exposons les réponses que nous avons pu y apporter. Ces questions ont été aussi abordées par d'autres auteurs, elles correspondent à des préoccupations de la communauté des chercheurs s'intéressant aux codes identifiants.

4.1 Existence d'un code identifiant

Dans cette section nous étudions des questions extrémales ayant trait à l'*existence* d'un code identifiant.

Nous savons que, à la différence des codes couvrants par exemple, les codes identifiants ont la propriété de ne pas exister dans tous les graphes. Nous rappelons que, pour tous $t, \ell \geq 1$, une condition nécessaire et suffisante pour qu'un graphe $G = (V, E)$ admette un code $(t, \leq \ell)$-identifiant est la suivante :

$$\forall X, Y \subseteq V,\ X \neq Y,\ |X| \leq \ell,\ |Y| \leq \ell,\ N_t[X] \neq N_t[Y]$$

où $N_t[X]$ désigne $\bigcup_{x \in X} B_t(x)$, avec $B_t(x) = \{y \mid d(x, y) \leq t\}$.

Un graphe admet un code identifiant si et seulement si toutes ses composantes connexes admettent un tel code. Ainsi, dans cette section, nous considérons uniquement le cas des graphes connexes.

4.1.1 Minimiser le nombre de sommets

Intuitivement, un graphe ayant trop peu de sommets comportera inévitablement des ensembles de sommets ayant le même voisinage étendu. Par exemple, un graphe ayant moins de t sommets sera tel que tous ses sommets auront un voisinage étendu à distance t égal à l'ensemble des sommets du graphe : un tel graphe n'admet pas de code t-identifiant.

4.1.1.1 Cas $\ell = 1$

Étant donné un entier $t \geq 1$, soit $n(t)$ le plus petit entier n tel qu'il existe un graphe connexe à n sommets admettant un code t-identifiant.

Proposition 4.1
Pour tout $t \geq 1$ on a $n(t) = 2t + 1$. De plus, le chemin P_{2t+1} est l'unique graphe connexe à $n(t)$ sommets admettant un code t-identifiant.

Preuve : Soit G un graphe connexe admettant un code t-identifiant, montrons que G a au moins $2t + 1$ sommets. Pour un chemin P dans G, soit $l(P)$ le nombre d'arêtes de P. Soit $P^* = x_0x_1 \dots x_l$ un chemin de cardinalité maximum dans G, et supposons que $l = l(P^*) \leq 2t - 1$. Soient $u := x_{\lfloor l/2 \rfloor}$ et $v = x_{\lfloor l/2 \rfloor + 1}$. On montre que $B_t(u) = B_t(v) = V(G)$. En effet, pour tout $i = 0, \dots, l$ on a $x_i \in B_t(u) \cap B_t(v)$, donc $P^* \subseteq B_t(u) \cap B_t(v)$. Maintenant, soit $x \in V(G) \setminus P^*$, et soit P' un plus court chemin de x à P^*. Sans perte de généralité on peut supposer que P' relie x à un sommet $x_{i'}$ avec $i' \leq \lfloor l/2 \rfloor$. Pour montrer que $x \in B_t(u)$ il suffit de montrer que $x \in B_t(v)$. Par l'absurde, si $x \notin B_t(v)$, alors la concaténation de P' et de $x_{i'}Pu$ est un chemin de longueur supérieure ou égale à $t + 1$, et on aurait $P'x_{i'}Px_l$ de longueur strictement supérieure à P^* (voir Figure 4.1). Donc $x \in B_t(v) \cap B_t(u)$ pour tout $x \in V(G) \setminus P^*$. Ceci montre que $l = l(P^*) \geq 2t$, *i.e.* $n(t) \geq 2t + 1$.

Le fait que P_{2t+1} admette un code t-identifiant montre que $n(t) = 2t + 1$. Maintenant, soit G un graphe à $n(t) = 2t + 1$ sommets admettant un code t-identifiant. Montrons que $G = P_{2t+1}$. Soit P^* un chemin de longueur maximum dans G. Comme précédemment on montre que $l(P^*) \geq 2t$, donc $l(P^*) = 2t$ et tous les sommets de G sont contenus dans $P^* = x_0x_1 \dots x_{2t}$. Soient $u = x_t$ et $v = x_{t+1}$. Si P^* comporte une corde x_ix_j, $|i-j| \geq 2$, alors on a $B_t(u) = B_t(v) = \{x_0, x_1, \dots, x_{2t}\}$. Donc P^* est sans corde, ce qui montre que $G = P^* = P_{2t+1}$. □

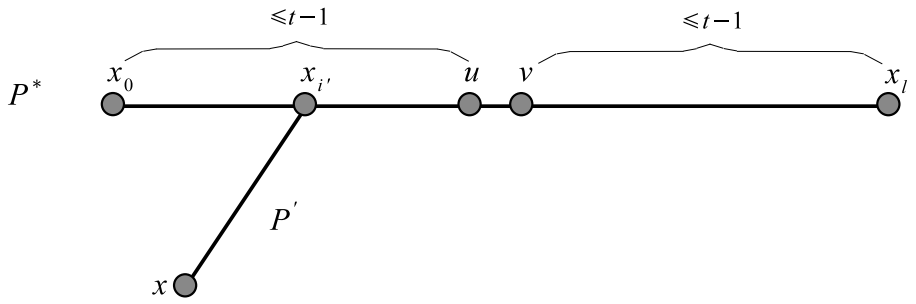


FIG. 4.1 – *Tout sommet $x \notin P^*$ est à distance inférieure ou égale à t de u et v.*

4.1.1.2 Cas $\ell \geq 2$

Le cas général $\ell \geq 2$ est peu connu. Pour l'instant nous connaissons très peu de familles de graphes connexes admettant un code $(t, \leq \ell)$-identifiant, et nous ignorons le n minimum tel qu'il existe un graphe connexe à n sommets admettant un code $(t, \leq \ell)$-identifiant. Dans cette section nous donnons quelques familles de graphes admettant un code $(1, \leq \ell)$-identifiant.

Nous rappelons qu'étant donnés deux graphes H et G, le *produit cartésien* de H et G, noté $H\Box G$, est le graphe ayant pour ensemble de sommets $V(H) \times V(G)$ tel que $(u_1, v_1)(u_2, v_2) \in E(H\Box G)$ si $u_1 = u_2$ et $v_1v_2 \in E(G)$ ou si $u_1u_2 \in E(H)$ et $v_1 = v_2$. On note $\Box^n G$ le produit cartésien

$$\underbrace{G\Box G\Box \ldots \Box G}_{n \text{ termes}}.$$

Proposition 4.2
Soit $\mathcal{C}_p$ le cycle sur p sommets, $p \geq 4$. Alors le graphe $\Box^\ell \mathcal{C}_p$ admet un code $(1, \leq \ell)$-identifiant.

Preuve : Notons $0, \ldots, p-1$ les sommets du cycle $\mathcal{C}_p$, $p \geq 4$. Montrons que $V = V(\Box^\ell \mathcal{C}_p)$ est un code $(1, \leq \ell)$-identifiant de $\Box^\ell \mathcal{C}_p$.
Soient X et Y deux sous-ensembles distincts de V satisfaisant $I(X, V) = I(Y, V)$ (*i.e.* $N[X] = N[Y]$). Sans perte de généralité, on peut supposer que $u = (0, \ldots, 0) \in X \setminus Y$. Pour tout $i \in \{1, \ldots, \ell\}$ et tout $\varepsilon = +1$ ou -1, soit $u^{i,\varepsilon}$ le sommet de coordonnées $u_i^{i,\varepsilon} = \varepsilon$ et $u_j^{i,\varepsilon} = 0$ pour tout $j \neq i$. Comme $p \geq 3$ alors tous les $u^{i,\varepsilon}$ sont distincts.

Comme $u \in X \setminus Y$ et $N[X] = N[Y]$, alors on peut supposer que $u^{1,+1} = (+1, 0, \ldots, 0) \in Y$. Comme pour tout i et tout ε on a $u^{i,\varepsilon} \in N[X]$ alors $Y \cap N[u^{i,\varepsilon}] \neq \emptyset$. Comme $p \geq 4$, alors pour tout i et pour tout ε, on a

$u^{1,+1} \in N[u^{i,\varepsilon}]$ si et seulement si $i = 1$ et $\varepsilon = +1$.

Maintenant, on remarque que pour tout $v \neq u$ il y au moins deux sommets $u^{i,\varepsilon}, u^{i',\varepsilon'}$ tels que $v \in N[u^{i,\varepsilon}] \cap N[u^{i',\varepsilon'}]$. Finalement, on obtient $|Y| \geq 1 + \frac{2\ell-1}{2} > \ell$, ce qui est une contradiction. □

De même, le produit de cycles $\mathcal{C}_{p_1} \Box \mathcal{C}_{p_2} \Box \ldots \Box \mathcal{C}_{p_\ell}$ admet un code $(1, \leq \ell)$-identifiant pourvu que $p_i \geq 4$ pour tout $i = 1, \ldots, \ell$. On peut aussi montrer que l'hypercube $Q_{2\ell}$ admet un code $(1, \leq \ell)$-identifiant. Les graphes de ces familles ont un nombre de sommets exponentiel en ℓ.

Étant donné $d \geq 1$, nous ignorons la valeur du plus grand entier ℓ tel que Q_{2d}, ou $\Box^d \mathcal{C}_4$, admette un code $(1, \leq \ell)$-identifiant. Nous soupçonnons que Q_{2d} et $\Box^d \mathcal{C}_4$ admettent un code $(1, \leq \ell)$-identifiant avec $\ell > d$.

Dans le cas où ℓ est la puissance d'un nombre premier, nous pouvons obtenir un résultat bien meilleur. Nous rappelons qu'un *plan projectif* (fini) d'ordre n est un hypergraphe sur $n^2 + n + 1$ sommets tel que :

- toute paire de sommets est contenue dans une unique hyperarête,
- deux hyperarêtes s'intersectent en un unique sommet,
- chaque sommet est contenu dans exactement $n + 1$ hyperarêtes, et
- chaque hyperarête contient exactement $n + 1$ sommets,

certaines de ces propriétés étant redondantes. On note $\mathbb{P}_n$ le plan projectif d'ordre n. Il est connu que $\mathbb{P}_n$ existe si n est la puissance d'un nombre premier. $\mathbb{P}_n$ est également connu sous le nom de 2-$(n^2 + n + 1, n + 1, 1)$ design, ou système de Steiner $S(2, n+1, n^2+n+1)$. Nous renvoyons le lecteur à [vLW92, Chapitre 19] pour une introduction aux designs.

Proposition 4.3 ([GM05])
Si q est la puissance d'un nombre premier, alors il existe un graphe connexe G_q à $2(q^2 + q + 1)$ sommets admettant un code $(1, \leq q)$-identifiant. De plus, G_q est $(q + 1)$-régulier.

Preuve : Supposons que q soit la puissance d'un nombre premier. Soit $\mathbb{P}_q$ un plan projectif d'ordre q. Nous rappelons que $\mathbb{P}_q$ a un ensemble de $q^2 + q + 1$ sommets, possède $q^2 + q + 1$ hyperarêtes de cardinalité $q + 1$. Tout sommet est contenu dans exactement $q + 1$ hyperarêtes, toute paire de sommets est contenue dans une unique hyperarête, et deux hyperarêtes distinctes s'intersectent en un unique sommet. Soit A la matrice d'incidence de $\mathbb{P}_q$: les lignes de A sont indexées par les sommets, les colonnes de A sont indexées par les hyperarêtes, et A_{ij} vaut 1 si le i-ième sommet de $\mathbb{P}_q$ est contenu dans la j-ième hyperarête de $\mathbb{P}_q$, et 0 sinon. Chaque ligne (resp. colonne) de A a exactement $q + 1$ uns, et chaque paire de lignes (resp. de colonnes) de A a exactement un 1 en commun.

Nous construisons un graphe G_q à partir de A comme suit. Soit B la matrice définie par

$$B = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix},$$

et soit G_q le graphe (simple, non orienté) dont la matrice d'adjacence est B (*i.e.* i et j sont adjacents dans G_q si et seulement si $B_{ij} = 1$). G_q est bien défini puisque B est une matrice symétrique n'ayant que des 0 sur sa diagonale.

Par définition, G_q a $2(q^2+q+1)$ sommets et est $(q+1)$-régulier. De plus, G_q est biparti, puisque toutes les arêtes sont entre les q^2+q+1 premiers et les q^2+q+1 derniers sommets de G_q. G_q est de plus connexe : pour toute paire de sommets (u, v) parmi les q^2+q+1 premiers, il y a un unique sommet parmi les q^2+q+1 derniers adjacents à u et v, et vice-versa.

Nous montrons que l'ensemble des sommets de G_q est un code $(1, \leq q)$-identifiant de G_q. Soit X un sous-ensemble d'au plus q sommets de G_q. Supposons que nous ignorons X, mais que nous connaissons $I(X)$. Soit v un sommet de G_q. Nous avons $|I(v)| = q+2$, et

> pour tout sommet $u \neq v$, l'ensemble $I(u)$ contient au plus un élément de $I(v) \setminus \{v\}$. (4.1)

En effet, pour les sommets u qui, dans la bipartition, sont du même côté que v, (4.1) s'ensuit des propriétés des plans projectifs ; et pour les autres sommets (4.1) est trivial par construction. Ainsi, pour un sommet quelconque $v \in X$, les $q+2$ sommets de $I(v)$ sont contenus dans $I(X)$; mais si $v \notin X$, alors au plus $q+1$ sommets de $I(v)$ sont dans $I(X)$. Ainsi il suffit de regarder si $I(v) \subseteq I(X)$ pour déterminer si $v \in X$, et ce pour tout $v \in X$, ce qui achève la preuve. □

4.1.2 Maximiser le nombre d'arêtes

De la même façon, le nombre d'arêtes d'un graphe est un paramètre critique pour admettre un code identifiant. Par exemple, à n fixé, si un graphe connexe G à n sommets contient trop d'arêtes (plus de $\binom{n}{2} - \lfloor n/2 \rfloor$), alors il contiendra nécessairement deux sommets voisins u et v dont le voisinage est $V(G)$. Donc un graphe connexe à n sommets, admettant un code 1-identifiant, a au plus $\binom{n}{2} - \lfloor n/2 \rfloor$ arêtes. Dans la suite nous montrons que $\binom{n}{2} - \lfloor n/2 \rfloor$ est en effet le nombre maximum d'arêtes que peut avoir un graphe admettant un code 1-identifiant, et nous montrons que ce maximum est at-

teint pour une clique privée d'un couplage maximum. Nous ne savons pas résoudre ce problème dans le cas général d'un code $(t, \leq \ell)$-identifiant avec ℓ et t quelconques.

Proposition 4.4 ([Mona])
Soit $m_{\max}(n)$ le plus grand entier m tel qu'il existe un graphe à n sommets et m arêtes admettant un code 1-identifiant. Alors on a

$$m_{\max}(n) \;=\; \binom{n}{2} - \left\lfloor \frac{n}{2} \right\rfloor .$$

De plus les seuls graphes à n sommets et $m_{\max}(n)$ arêtes admettant un code 1-identifiant sont les cliques privées d'un couplage maximum.

Preuve : On commence par montrer que $m_{\max}(n) \leq \binom{n}{2} - \left\lfloor \frac{n}{2} \right\rfloor$. Soit G un graphe à n sommets et $m > \binom{n}{2} - \left\lfloor \frac{n}{2} \right\rfloor$ arêtes. Il existe au plus un sommet $v \in V(G)$ tel que $B_1(v) = V(G)$, donc il y a au plus un sommet de degré $n-1$ et au moins $n-1$ sommets de degré inférieur ou égal à $n-2$ dans G. Ceci nous conduit à l'inégalité :

$$m \;=\; \frac{1}{2} \sum_{v \in V(G)} d(v) \;\leq\; \frac{n(n-1)}{2} - \frac{n-1}{2}.$$

Comme m est entier, on a en fait

$$m \;\leq\; \frac{n(n-1)}{2} - \left\lceil \frac{n-1}{2} \right\rceil .$$

Comme $m > \binom{n}{2} - \left\lfloor \frac{n-1}{2} \right\rfloor$, ceci conduit à $\left\lceil \frac{n-1}{2} \right\rceil < \left\lfloor \frac{n-1}{2} \right\rfloor$, ce qui est absurde. Donc $m_{\max}(n) \leq \binom{n}{2} - \left\lfloor \frac{n}{2} \right\rfloor$.

Maintenant, soit G un graphe admettant un code 1-identifiant à n sommets et $\binom{n}{2} - \left\lfloor \frac{n}{2} \right\rfloor$ arêtes, et soit $\overline{G}$ le graphe des non-arêtes de G ($\overline{G}$ est le graphe complémentaire de G, il a $\left\lfloor \frac{n}{2} \right\rfloor$ arêtes). On montre que $\overline{G}$ est un couplage. On sait que G comporte au plus un sommet de degré $n-1$, donc $\overline{G}$ comporte au plus un sommet de degré nul. Si $\overline{G}$ n'est pas un couplage, alors il contient au moins un sommet de degré 2, et on a ainsi

$$n \;=\; 2 + (n-2) \;\leq\; \sum_{x \in V(\overline{G})} d(x) \;=\; 2 \left\lfloor \frac{n}{2} \right\rfloor ,$$

Ceci est impossible si n est impair. Si n est pair, l'inégalité au-dessus est une égalité, et il y a exactement un sommet x_0 de degré 2 dans $\overline{G}$. Soient y et

z les deux voisins de x_0 dans $\overline{G}$: y et z sont de degré 1, et ont x_0 comme seul voisin commun. Ceci contredit le fait que $\overline{G}$ (et donc, dans ce cas, G) admette un code 1-identifiant. Donc $\overline{G}$ est un couplage.
Il nous reste donc à montrer que $G^* = K_n - M_n^*$ admet un code 1-identifiant, avec M_n^* un couplage maximum de K_n. Soient u et v deux sommets distincts de G^*. Si $uv \in M_n^*$, alors u et v ne sont pas voisins dans G^*, donc $B_1(u) \neq B_1(v)$. Comme M_n^* est un couplage maximum de K_n, alors si $uv \notin M_n^*$ il existe w tel que $w \in B_1(u)\Delta B_1(v)$. G^* est donc un graphe tel que pour tout $u, v \in V(G^*)$, $u \neq v$, on a $B_1(u) \neq B_1(v)$: G^* admet un code 1-identifiant. □

4.1.3 Lien avec le degré minimum

Dans [LR01] il est montré qu'un graphe connexe admettant un code $(1, \leq \ell)$-identifiant a son degré minimum supérieur ou égal à ℓ. En effet, soit G un graphe admettant un sommet v de degré inférieur ou égal à $\ell - 1$. Alors $X := N(v)$ et $Y := \{v\} \cup N(v)$ sont tels que $N[X] = N[Y]$, donc G n'admet pas de code $(1, \leq \ell)$-identifiant. Dans [LR01] ils mentionnent les chemins comme exemples de graphes de degré minimum 1 admettant un code 1-identifiant. Ici nous montrons que pour tout $\ell \geq 1$ il existe des graphes de degré minimum ℓ admettant un code $(1, \leq \ell)$-identifiant. Nous donnons une construction explicite, qui est basée sur la Proposition 4.2.

Lemme 4.1 ([GM05])
Soit C un code $(1, \leq \ell)$-identifiant d'un graphe G, et soient X et Y deux sous-ensembles d'au plus ℓ sommets de G.
Alors nous avons soit

$$|X| + |I(X)\Delta I(Y)| \ > \ \ell$$

soit

$$|Y| + |I(X)\Delta I(Y)| \ > \ \ell.$$

Preuve : Soit $X' := X \cup I(X, C)\Delta I(Y, C)$ et soit $Y' := Y \cup I(X, C)\Delta I(Y, C)$. Il est facile de vérifier que $I(X', C)\Delta I(Y', C) = \emptyset$. Comme C est un code $(1, \leq \ell)$-identifiant de G, ceci implique que $|X'| > \ell$ ou $|Y'| > \ell$. □

Proposition 4.5 ([GM05])
Pour tout $\ell \geq 1$ il existe un graphe G_ℓ de degré minimum ℓ admettant un code $(1, \leq \ell)$-identifiant.

Preuve : Soit H un graphe connexe admettant un code $(1, \leq \ell)$-identifiant (d'après la Proposition 4.2 un tel H existe). Prenons ℓ copies $H_1, \ldots, H_\ell$ de H, dans lesquelles nous spécifions ℓ sommets $h_i \in H_i$, $i = 1, \ldots, \ell$. On construit alors un graphe G en joignant les H_i avec un nouveau sommet u tel que $N(u) = \{h_1, \ldots, h_\ell\}$. Il est facile de voir que G admet un code $(1, \leq \ell)$-identifiant. En effet, soient X et Y deux sous-ensembles d'au plus ℓ sommets de G. Si $u \notin X \cup Y$, alors clairement $N[X] \neq N[Y]$ puisque H admet un code $(1, \leq \ell)$-identifiant. Si $u \in X \cap Y$, alors soit i tel que $X \cap H_i =: X_i \neq Y_i := Y \cap H_i$. Comme $|X_i| \leq \ell - 1$ et $|Y_i| \leq \ell - 1$, alors par le Lemme 4.1 on sait que $|N[X_i] \Delta N[Y_i]| \geq 2$. Comme u a un unique voisin h_i dans H_i, alors $N[X] \neq N[Y]$. Enfin, si, par exemple, $u \in X \setminus Y$, alors Y doit avoir une intersection non vide avec chacun des $H_1, \ldots, H_\ell$. Donc $|Y| = \ell$ et pour tout $i = 1, \ldots, \ell$ on a $|Y \cap H_i| = 1$. Comme H admet un code $(1, \leq \ell)$-identifiant alors $\delta(H) \geq \ell \geq 1$ et ainsi $|N[Y] \cap H_i| \geq 2$ pour tout $i = 1, \ldots, \ell$. Ceci implique que pour tout $i = 1, \ldots, \ell$ il existe $x_i \in X \cap H_i$. Comme X contient déjà u, ceci contredit $|X| \leq \ell$. □

Une question naturelle est de savoir s'il existe des graphes *ℓ-réguliers* admettant un code $(1, \leq \ell)$-identifiant. Nous rappelons que la Proposition 4.3 a pour conséquence l'existence de graphes $(\ell+1)$-réguliers admettant un code $(1, \leq \ell)$-identifiant.

4.2 Graphes ayant des codes identifiants de faible cardinalité

Le problème d'optimisation associé aux codes identifiants est un problème de minimisation : étant donné un graphe G, nous souhaitons trouver un code identifiant de G de cardinalité minimum. C'est une question extrémale naturelle que de se demander quels sont les graphes à n sommets pour lesquels la cardinalité minimum d'un code est la plus petite possible.

Du point de vue des applications, cette question a également un intérêt : si nous avions à construire un réseau de n processeurs sans aucune autre contrainte que de faire en sorte qu'il soit "pratique" du point de vue de la détection de processeurs défectueux, alors nous choisirions une structure de réseau pour laquelle le nombre de processeurs nécessaire à la détection soit le plus petit possible.

4.2.1 Cas où l'on identifie un seul sommet

Nous rappelons du Chapitre 1 (Théorème 1.4) la borne inférieure triviale suivante : pour tout $t \geq 1$, étant donné un graphe G à n sommets admettant un code t-identifiant C, on a

$$|C| \geq \lceil \log(n+1) \rceil. \tag{4.2}$$

On montre que cette borne est serrée, c'est-à-dire que pour tous $t \geq 1$ et $n \geq 1$ il existe un graphe à n sommets admettant un code t-identifiant de cardinalité égale à $\lceil \log(n+1) \rceil$.

Dans le cas $t = 1$, nous sommes même en mesure de donner *tous* les graphes à n sommets admettant un code 1-identifiant de cardinalité égale à $\lceil \log(n+1) \rceil$. Ce cas est donc traité à part du cas général $t \geq 2$ dans la suite.

Dans la suite, un graphe à n sommets sera dit *t-optimal* (ou, s'il n'y a pas d'ambiguïté, *optimal*) s'il admet un code t-identifiant de cardinalité $\lceil \log(n+1) \rceil$.

4.2.1.1 Identification des sommets à distance 1

Nous connaissons tous les graphes à n sommets admettant un code 1-identifiant de cardinalité égale à $\lceil \log(n+1) \rceil$. Cette construction est donnée dans le premier paragraphe de cette section. Dans le paragraphe suivant nous essayons de déterminer les graphes optimaux ayant un nombre minimum d'arêtes. Du point de vue pratique, cette préoccupation est sensée : nous pouvons imaginer que les liaisons entre les processeurs d'un réseau ont un coût non-nul, et dans ce contexte il est tout à fait justifié de vouloir concevoir un réseau qui à la fois minimise le nombre de processeurs nécessaire à la détection et le nombre de liaisons dans le réseau. Enfin dans le dernier paragraphe nous étudions deux paramètres des graphes optimaux : leur nombre de domination ainsi que la cardinalité maximum d'un code 1-identifiant minimal.

4.2.1.1.1 Construction de graphes optimaux

Soit $n \geq 1$ un entier et soit $p = \lceil \log(n+1) \rceil$. Soit H un graphe à p sommets $x_1, \ldots, x_p$ admettant un code 1-identifiant. À partir de H nous construisons un graphe $\mathcal{G}(H)$ à n sommets, admettant un code 1-identifiant de cardinalité p, que l'on obtient par la construction suivante :

1. On commence par prendre H, qui admet $\{x_1, \ldots, x_p\}$ comme code 1-identifiant. Le graphe $\mathcal{G}(H)$ est obtenu en rajoutant des sommets connectés à H, il aura $\{x_1, \ldots, x_p\}$ comme code 1-identifiant.
2. À chaque sommet x_j de H on associe le vecteur caractéristique de l'ensemble 1-identifiant de x_j : c'est un vecteur en $0-1$ $v(x_j)$ tel que $v(x_j)_i = 1$ si et seulement si $x_i \in B_1(x_j)$. Soit $\mathcal{V} := \{v(x_j) \mid j = 1, \ldots, p\}$. Comme H admet un code 1-identifiant alors $\mathcal{V}$ a exactement p éléments et ne contient pas le vecteur $(0, 0, \ldots, 0)$.
3. Soit $\mathcal{W}$ un sous-ensemble de $n-p$ éléments de $\{0,1\}^p \setminus (\mathcal{V} \cup \{(0, 0, \ldots, 0)\})$. Pour tout $w \in \mathcal{W}$, on ajoute un sommet y_w à $\mathcal{G}(H)$ tel que

$$N(y_w) \;=\; \{x_i \mid w_i = 1\}.$$

   Le sommet y_w est tel que w est le vecteur caractéristique de l'ensemble 1-identifiant de y_w (voir Figure 4.2).
4. On pose $C = \{x_1, \ldots, x_p\}$: C est un code 1-identifiant de $\mathcal{G}(H)$. On peut rajouter à volonté des arêtes entre les vecteurs y_w, $w \in \mathcal{W}$: cela ne change pas le fait que $I(y_w, C) = \{x_i \mid w_i = 1\}$, $w \in \mathcal{W}$.

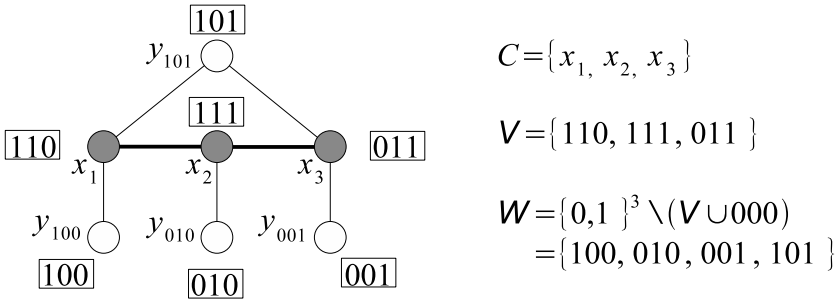


FIG. 4.2 – *Construction d'un graphe optimal $\mathcal{G}(H)$ à 7 sommets à partir de $H = P_3$. Le vecteur caractéristique de chaque sommet est indiqué.*

Nous prouvons maintenant la validité de cette construction ainsi que son caractère universel, dans le sens où tout graphe optimal G peut être réalisé comme un $\mathcal{G}(H)$ par la construction ci-dessus.

Théorème 4.1 ([Mona])
Soit $n \geq 1$ et soit H un graphe à $p := \lceil \log(n+1) \rceil$ sommets admettant un code 1-identifiant et soit $\mathcal{G}(H)$ un graphe construit par le procédé décrit ci-dessus. Alors $\mathcal{G}(H)$ est un graphe 1-optimal, i.e. il admet un code 1-identifiant de cardinalité $\lceil \log(n+1) \rceil$.

De plus, soit G un graphe à n sommets admettant un code 1-identifiant C de cardinalité $\lceil \log(n+1) \rceil$. Alors G peut être réalisé comme un $\mathcal{G}(H)$ par la construction décrite ci-dessus en prenant $H = G[C]$ le sous-graphe de G induit par C.

Preuve : Soit $n \geq 1$ et soit H un graphe à $p := \lceil \log(n+1) \rceil$ sommets $v_1, \ldots, v_p$ admettant un code 1-identifiant. Soit $\mathcal{G}(H)$ un graphe construit par le procédé décrit ci-dessus. Montrons que $C := \{v_1, \ldots, v_p\}$ est un code 1-identifiant de $\mathcal{G}(H)$. Comme $\mathcal{G}(H)$ a n sommets et que $p = |C| = \lceil \log(n+1) \rceil$, ceci montre que $\mathcal{G}(H)$ est un graphe 1-optimal. Il suffit de vérifier que les ensembles identifiants $I(x, C)$ sont tous distincts et non vides, ce qui est immédiat si l'on remarque que par construction les vecteurs caractéristiques de ces ensembles sont tous distincts et différents de $(0, 0, \ldots, 0)$.

Maintenant, soit G un graphe à n sommets admettant un code 1-identifiant C de cardinalité $p := \lceil \log(n+1) \rceil$. Montrons que G peut être réalisé comme un $\mathcal{G}(H)$ par la construction décrite ci-dessus en prenant $H = G[C]$. Tout d'abord, comme C est un code 1-identifiant de G alors le graphe $H = G[C]$ admet un code 1-identifiant : il suffit de remarquer que pour tout $x \in H$, $N[x]$ est égal à l'ensemble identifiant de x dans G. Soit $\mathcal{W}'$ l'ensemble des vecteurs caractéristiques des ensembles identifiants des sommets de $V(G) \setminus C$: en prenant $\mathcal{W} = \mathcal{W}'$ à l'étape 3 de la construction on construit un graphe $\mathcal{G}(H)$ égal à G privé d'arêtes entre des sommets de $V(G) \setminus C$. En rajoutant ces arêtes à l'étape 4 de la construction on obtient $\mathcal{G}(H) = G$. □

4.2.1.1.2 Minimiser le nombre d'arêtes d'un graphe optimal

Ici nous nous intéressons aux graphes optimaux ayant un nombre minimum d'arêtes, ce qui est une préoccupation justifiée du point de vue des applications. Dans le cas où le nombre de sommets du graphe est $2^p - 1$, $p \geq 1$, nous allons voir que ce problème revient à maximiser le nombre d'arêtes d'un graphe à p sommets admettant un code 1-identifiant, qui est une question que nous avons abordée dans le paragraphe 4.1.2.

Dans cette section soient $p \geq 1$ et $n = 2^p - 1$. Nous savons que tous les graphes optimaux à n sommets peuvent être obtenus comme des $\mathcal{G}(H)$ par la construction décrite au paragraphe précédent. Comme $n = 2^p - 1$, alors, nécessairement, l'ensemble des vecteurs caractéristiques des ensembles 1-identifiants des sommets de $\mathcal{G}(H)$ est égal à $\{0,1\}^p \setminus \{(0, \ldots, 0)\}$. Autrement dit, l'ensemble $\mathcal{W}$ choisi à l'étape 3 est égal à $\{0,1\}^p \setminus (\mathcal{V} \cup \{(0, \ldots, 0)\})$. Ensuite, si l'on veut minimiser le nombre d'arêtes de $\mathcal{G}(H)$, alors on n'ajoutera aucune arête superflue à l'étape 4 de la construction. Ceci étant dit, on

remarque que le nombre d'arêtes de $\mathcal{G}(H)$ dépend du graphe H choisi au départ (voir Figure 4.3).

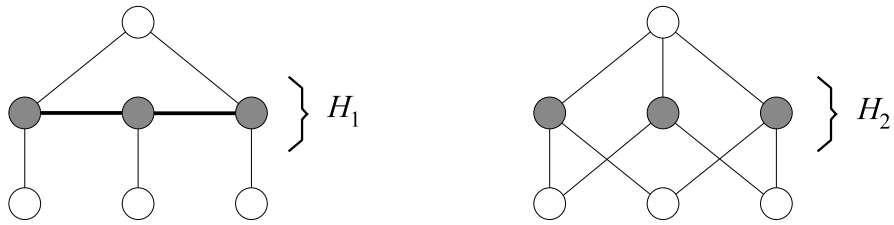


FIG. 4.3 – *Deux graphes optimaux $\mathcal{G}(H_1)$ et $\mathcal{G}(H_2)$ avec $H_1 = P_3$ et H_2 un stable. $\mathcal{G}(H_1)$ a 7 arêtes tandis que $\mathcal{G}(H_2)$ en a 9.*

En fait, le résultat suivant nous dit que le nombre d'arêtes de $\mathcal{G}(H)$ diminue lorsque celui de H augmente : plus H a d'arêtes, moins $\mathcal{G}(H)$ en a.

Lemme 4.2 ([Mona])
Soit H un graphe à p sommets et m arêtes admettant un code 1-identifiant et soit $\mathcal{G}(H)$ le graphe à $2^p - 1$ sommets construit par le procédé décrit au paragraphe 4.2.1.1.1 sans ajouter aucune arête à l'étape 4 de la construction. Alors $\mathcal{G}(H)$ a $2^p - 1 - p - m$ arêtes.

Preuve : Pour tout sommet x_i de H soit α_i la cardinalité de l'ensemble identifiant de x_i : pour tout $i = 1, \ldots, p$ α_i est égal au degré de x_i plus 1. Comme le nombre de sommets de $\mathcal{G}(H)$ est $2^p - 1$ et que l'on n'ajoute pas d'arêtes superflues à l'étape 4 de la construction, alors le nombre d'arêtes de $\mathcal{G}(H)$ est égal à

$$\sum_{j=1}^{p} \binom{p}{j} - \sum_{i=1}^{p} \alpha_i + m.$$

Or $\sum_{i=1}^{p} \alpha_i = \sum_{i=1}^{p} (d(x_i) + 1) = 2m + p$, d'où le nombre d'arêtes de $\mathcal{G}(H)$ est

$$\sum_{j=1}^{p} \binom{p}{j} - p - m \;=\; 2^p - 1 - p - m.$$

□

Ceci entraîne :

Proposition 4.6 ([Mona])
Soit $p \geq 1$. Alors le nombre minimum d'arêtes d'un graphe optimal G à $2^p - 1$

sommets est égal à

$$2^p - p - 1 - \binom{p}{2} + \left\lfloor \frac{p}{2} \right\rfloor .$$

De plus, G peut être obtenu comme un $\mathcal{G}(H)$ par la construction décrite au paragraphe 4.2.1.1.1 en prenant $H = K_p - M_p^*$, où M_p^* est un couplage maximum de la clique K_p.

Preuve : Par le Théorème 4.1, on sait que tout graphe optimal peut être obtenu comme un $\mathcal{G}(H)$ par la construction décrite au paragraphe 4.2.1.1.1. Par le Lemme 4.2, on sait que minimiser le nombre d'arêtes de $\mathcal{G}(H)$ revient à maximiser le nombre d'arêtes de H. La Proposition 4.4 nous dit que le nombre maximum d'arêtes d'un graphe à p sommets admettant un code 1-identifiant est $\binom{p}{2} - \left\lfloor \frac{p}{2} \right\rfloor$, obtenu en prenant $H = K_p - M_p^*$, où M_p^* est un couplage maximum de la clique K_p. □

Dans le cas où le nombre de sommets du graphe n'est plus de la forme $2^p - 1$, $p \geq 1$, nous ne savons pas comment obtenir les graphes optimaux ayant un nombre minimum d'arêtes. L'exemple de la Figure 4.4 montre qu'en général il ne suffit pas de construire un $\mathcal{G}(H)$ avec H ayant un nombre maximum d'arêtes.

FIG. 4.4 – *Cas où un graphe optimal à $n \neq 2^p - 1$ sommets est obtenu comme un $\mathcal{G}(H)$ à partir d'un H n'ayant pas un nombre maximum d'arêtes : $\mathcal{G}(H_1)$ a 3 arêtes contre 2 pour $\mathcal{G}(H_2)$, bien qu'H_1 ait 2 arêtes et H_2 aucune.*

4.2.1.1.3 Deux paramètres des graphes optimaux

Dans ce paragraphe nous calculons le nombre de domination de certains graphes optimaux ainsi que la cardinalité maximum d'un code 1-identifiant minimal de ces graphes. Le calcul de ce dernier paramètre nous permet de prouver que l'algorithme glouton de construction de code 1-identifiant minimal n'offre pas de garantie de performance.

Nous rappelons qu'un *dominant* d'un graphe G est un sous-ensemble de sommets $D \subseteq V(G)$ qui couvre tous les sommets de G : pour tout $u \in$

$V(G) \setminus D$, il existe $x \in D$ tel que $v \in N(x)$. Un code 1-identifiant C (resp. un dominant D) de G est dit *minimal* s'il n'existe aucun C', $C' \subset C$, $C' \neq C$, tel que C' est un code 1-identifiant de G (resp. il n'existe aucun D', $D' \subset D$, $D' \neq D$, tel que D' est un dominant de G). Dans cette section $\gamma(G)$ désigne la cardinalité minimum d'un dominant de G et $M(G)$ désigne la cardinalité minimum d'un code 1-identifiant de G. $\widetilde{\gamma}(G)$ et $\widetilde{M}(G)$, quant à eux, désignent la cardinalité maximum d'un dominant minimal de G (resp. la cardinalité maximum d'un code 1-identifiant minimal de G).

Proposition 4.7 ([Mona])
Soit H un graphe à p sommets admettant un code 1-identifiant et soit $\mathcal{G}(H)$ le graphe à $2^p - 1$ sommets construit par le procédé décrit au paragraphe 4.2.1.1.1 sans ajouter aucune arête à l'étape 4 de la construction. Si H a au moins deux sommets isolés alors on a

$$\gamma(\mathcal{G}(H)) \ = \ p - 1$$

et

$$\gamma(\mathcal{G}(H)) \ = \ p$$

sinon.

Preuve : Soient $x_1, \ldots, x_p$ les sommets de H et soit $\mathcal{G}(H)$ le graphe à $2^p - 1$ sommets construit par le procédé décrit au paragraphe 4.2.1.1.1 sans ajouter aucune arête à l'étape 4 de la construction. On rappelle que $C = \{x_1, \ldots, x_p\}$ est un code 1-identifiant de $\mathcal{G}(H)$, c'est donc en particulier un dominant de $\mathcal{G}(H)$ donc

$$\gamma(\mathcal{G}(H)) \leq p. \tag{4.3}$$

Si H a au moins deux sommets isolés x_i et x_j, soit $y_{i,j}$ le sommet de $V(\mathcal{G}(H)) \setminus C$ tel que $N(y_{i,j}) = \{x_i, x_j\}$ — $y_{i,j}$ existe puisque $\mathcal{G}(H)$ a $2^p - 1$ sommets. Il est facile de voir que

$$D : = \ \{x_1, \ldots, x_p\} \cup \{y_{i,j}\} \setminus \{x_i, x_j\}$$

est un dominant de $\mathcal{G}(H)$, d'où :

$$\text{si } H \text{ a au moins deux sommets isolés, alors } \gamma(\mathcal{G}(H)) \leq p - 1. \tag{4.4}$$

Soit D un dominant de $\mathcal{G}(H)$. Soit k le nombre de sommets de

$$A \ := \ \{x_1, \ldots, x_p\} \setminus D,$$

et soit

$$B \ := \ \{x_1, \ldots, x_p\} \setminus A.$$

On a $B \subseteq D$ et $A \cap D = \emptyset$. Si $k = 0$, alors $|D| \geq p$, ce qui montre que $\gamma(\mathcal{G}(H)) = p$ par (4.3). Dans ce cas, H a au plus un sommet isolé d'après (4.4).

Supposons maintenant que $k \geq 1$. S'il existe un sommet $a_0 \in A$ voisin d'un sommet $b_0 \in B$, alors la cardinalité de

$$\{a \in A \mid I(a, C) \subseteq A\}$$

est inférieure ou égale à $k-1$ (au moins a_0 est tel que $I(a_0, C) \not\subseteq A$)). Comme $\mathcal{G}(H)$ a $2^p - 1$ sommets, alors il y a au moins $2^k - 1 - (k-1)$ sommets de $V(\mathcal{G}(H)) \setminus C$ ayant leur ensemble identifiant inclus dans A. Ces sommets appartiennent nécessairement à D, donc

$$|D| \ \geq \ 2^k - 1 - (k-1) + (p-k) \ = \ 2^k + p - 2k,$$

qui est supérieur ou égal à p pour tout $k \geq 1$. Dans ce cas, on a donc $\gamma(\mathcal{G}(H)) = p$ et H a au plus un sommet isolé, toujours d'après (4.3) et (4.4).

S'il n'existe aucune arête entre un sommet de A et un sommet de B, alors la cardinalité de

$$\{a \in A \mid I(a, C) \subseteq A\}$$

est inférieure ou égale à k, et il y a au moins $2^k - 1 - k$ sommets de $V(\mathcal{G}(H)) \setminus C$ ayant leur ensemble identifiant inclus dans A. Ces sommets appartenant nécessairement à D (car $A \cap D = \emptyset$, et aucune arête n'a été ajoutée à l'étape 4), on a

$$|D| \ \geq \ 2^k - 1 - k + (p-k) \ = \ 2^k + p - 1 - 2k.$$

Si $k \neq 1, 2$, alors cette quantité est supérieure ou égale à p, et l'on a donc encore $\gamma(\mathcal{G}(H)) = p$ et H a au plus un sommet isolé par (4.3) et (4.4).

Si $k = 1$, alors $A = \{a\}$, et il faut au moins un sommet de $V(\mathcal{G}(H)) \setminus C$ pour couvrir a, donc $|D| \geq p - 1 + 1 = p$, d'où $\gamma(\mathcal{G}(H)) = p$ et H a au plus un sommet isolé par (4.3) et (4.4).

Si $k = 2$, alors $A = \{a_1, a_2\}$, et $a_1 a_2 \notin E(H)$ (sinon H n'admet pas de code 1-identifiant) : H a au moins deux sommets isolés. Il faut alors au moins un sommet de $V(\mathcal{G}(H)) \setminus C$ pour couvrir a_1 et a_2, donc $|D| \geq p-2+1 = p-1$, d'où $\gamma(\mathcal{G}(H)) = p - 1$ par (4.4). □

La motivation de ce résultat réside dans la question suivante : quels sont les graphes G tels que

$$M(G) = \gamma(G) \ ? \tag{4.5}$$

D'une certaine façon, de tels graphes ont une "bonne" structure pour le problème d'identification de sommets puisqu'on peut trouver un dominant minimum de G pour obtenir un code 1-identifiant de G. Nous ne connaissons

pas l'ensemble des graphes satisfaisant (4.5), mais la proposition précédente montre qu'il existe de tels graphes.

Dans la proposition suivante nous calculons de façon asymptotique l'écart entre la cardinalité minimum d'un code 1-identifiant et la cardinalité maximum d'un code 1-identifiant minimal de certains graphes optimaux.

Proposition 4.8 ([Mona])
Soit H un graphe à p sommets admettant un code 1-identifiant et soit $\mathcal{G}(H)$ le graphe à $2^p - 1$ sommets construit par le procédé décrit au paragraphe 4.2.1.1.1 sans ajouter aucune arête à l'étape 4 de la construction. Alors on a

$$\lim_{p\to\infty} \frac{\widetilde{M}(\mathcal{G}(H))}{M(\mathcal{G}(H))} \;=\; +\infty.$$

Preuve : Soit $C = \mathcal{G}[H]$: C est un code 1-identifiant de $\mathcal{G}$. Il est facile de vérifier que $V(\mathcal{G}(H)) \setminus C$ est un code 1-identifiant minimal de $\mathcal{G}(H)$. Comme $|V(\mathcal{G}(H)) \setminus C| = 2^p - p - 1$, alors on a $\widetilde{M}(\mathcal{G}(H)) \geq 2^p - p - 1$, donc $\frac{\widetilde{M}(\mathcal{G}(H))}{|C|} \to +\infty$. □

La motivation de ce résultat est algorithmique : cela montre que l'écart entre la cardinalité minimum d'un code 1-identifiant et la cardinalité maximum d'un code 1-identifiant minimal de certains graphes optimaux n'est bornée par aucune constante, ce qui signifie que l'algorithme glouton de construction d'un code 1-identifiant minimal n'admet aucune garantie.

4.2.1.2 Identification des sommets à distance $t \geq 2$

Soit $t \geq 1$. On donne dans cette section une construction pour obtenir des graphes à n sommets ayant un code t-identifiant de cardinalité $\lceil \log(n+1) \rceil$. Comme dans la section précédente on commence par prendre un graphe H admettant un code t-identifiant (ici on prendra H un cycle à $\lceil \log(n+1) \rceil$ sommets). Ensuite on fait pendre de chacun des sommets du cycle un chemin de longueur $t-2$, et on ajoute finalement des sommets que l'on branche aux extrémités de ces chemins. La construction détaillée est la suivante :

1. Soient $t \geq 1$ et $n \geq 2^{2t+1}$. Soit $p := \lceil \log(n+1) \rceil$, $p \geq 2t+2$. Soit $\mathcal{C}_p$ un cycle ayant pour ensemble de sommets $\{x_1^0, \ldots, x_p^0\}$. On construit un graphe à n sommets $\mathcal{G}(\mathcal{C}_p)$ contenant $\mathcal{C}_p$ ayant pour code t-identifiant $C = \{x_1^0, \ldots, x_p^0\}$.
2. Pour tout $i = 1, \ldots, p$ on attache à x_i une chaîne $P_i = x_i^1, \ldots, x_i^{t-1}$ de longueur $t-2$ (voir Figure 4.5), de sorte que x_i^0 est voisin de x_i^1 pour

tout $i = 1, \ldots, p$. L'ensemble C est toujours un code t-identifiant du graphe construit jusqu'ici.

3. Soit $\mathcal{V}$ l'ensemble des vecteurs caractéristiques des ensembles identifiants (par C) des x_i^j, $i = 1, \ldots, p$, $j = 1, \ldots, t-1$. Soit $\mathcal{W}$ un sous-ensemble de cardinalité $n-p$ de $\{0,1\}^p \setminus (\mathcal{V} \cup \{(0, \ldots, 0)\})$. Pour tout $w \in \mathcal{W}$, on ajoute un sommet y_w tel que $N(y_w) = \{x_i^{t-1} \mid w_i = 1\}$. Autrement dit w est le vecteur caractéristique de l'ensemble identifiant de y_w (voir Figure 4.5).
4. On peut rajouter des arêtes à volonté entre les sommets y_w ajoutés à l'étape 3 sans changer leurs ensembles identifiants. $\{x_1^0, \ldots, x_p^0\}$ est alors un code t-identifiant de $\mathcal{G}(\mathcal{C}_p)$.

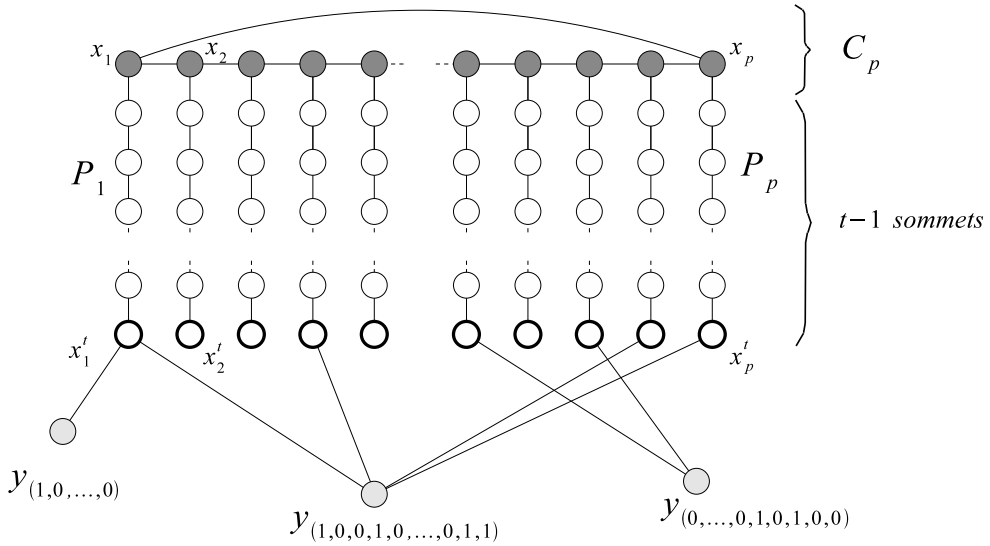


FIG. 4.5 – *Construction de $\mathcal{G}(\mathcal{C}_p)$.*

Théorème 4.2 ([CHLa])
Soient $t \geq 1$ et $n \geq 2^{2t+1}$. Soit $\mathcal{C}_p$ un cycle à $p := \lceil \log(n+1) \rceil$ sommets et soit $\mathcal{G}(\mathcal{C}_p)$ un graphe construit par le procédé décrit ci-dessus. Alors $\mathcal{G}(\mathcal{C}_p)$ est un graphe t-optimal, i.e. il admet un code t-identifiant de cardinalité $\lceil \log(n+1) \rceil$.

Preuve : Soient $t \geq 1$ et $n \geq 2^{2t+1}$. Soit $\mathcal{C}_p$ un cycle à $p := \lceil \log(n+1) \rceil$ sommets $x_1^0, \ldots, x_p^0$ et soit $\mathcal{G}(\mathcal{C}_p)$ un graphe construit par le procédé décrit ci-dessus. On montre que $\{x_1^0, \ldots, x_p^0\}$ est un code t-identifiant de $\mathcal{G}(\mathcal{C}_p)$.

Soit $0 \leq j \leq t-1$. Il est facile de voir que pour tout $i \in \{1, \ldots, p\}$, le

vecteur caractéristique de l'ensemble t-identifiant de x_i^j est une permutation circulaire de

$$(0,\ldots,0,\underbrace{1,\ldots,1}_{t-j},\underline{1},\underbrace{1,\ldots,1}_{t-j},0,\ldots,0)$$

où le 1 souligné est à l'indice i. Comme $p > 2t+1$, alors tous ces vecteurs sont différents. Ensuite, il est facile de voir que le vecteur caractéristique de l'ensemble t-identifiant d'un sommet y_w, $w \in \mathcal{W}$, est égal à w comme annoncé au-dessus. Par construction tous les sommets de $\mathcal{G}(\mathcal{C}_p)$ ont donc des ensembles t-identifiants distincts et non-vides : $\{x_1^0,\ldots,x_p^0\}$ est un code t-identifiant de $\mathcal{G}(\mathcal{C}_p)$. □

Dans le cas général $t \geq 2$ nous pouvons donc construire pour tout n des graphes admettant un code t-identifiant de cardinalité $\lceil \log(n+1) \rceil$, ce qui montre que la borne (4.2) est serrée. Au contraire du cas $t = 1$, nous ne savons cependant pas construire tous les graphes optimaux pour $t \geq 2$.

4.2.2 Cas où l'on identifie des ensembles de sommets

Dans le cas des codes $(1,\leq \ell)$-identifiants, on rappelle du Chapitre 1 (Théorème 1.6) que la cardinalité d'un code $(1,\leq \ell)$-identifiant est supérieure ou égale à

$$\Omega\left(\frac{\ell^2}{\log \ell}\log n\right).$$

Nous ne savons pas s'il existe des graphes admettant un code $(1,\leq \ell)$-identifiant d'une telle cardinalité. Au Chapitre 5 (Proposition 5.5) nous donnerons une construction probabiliste permettant de construire des graphes admettant des codes $(1,\leq \ell)$-identifiants de cardinalité

$$O\left(\ell^2 \log n\right).$$

Cependant cette construction probabiliste a l'inconvénient de ne pas avoir de pendant algorithmique : elle ne nous fournit pas de procédé explicite pour construire de tels graphes. Ici nous donnons une construction explicite d'une famille de graphes ayant un code $(1,\leq \ell)$-identifiant de cardinalité $O\left(\ell^4 \log n\right)$. Cette construction utilise des codes superimposés, dont nous avons déjà parlé au Chapitre 1 (paragraphe 1.2.2).

Nous rappelons qu'un code ℓ-superimposé de $\{0,1\}^N$ est un ensemble K de vecteurs de $\{0,1\}^N$ satisfaisant la propriété suivante :

la somme d'au plus ℓ vecteurs de K est différente de la somme d'au plus ℓ autres vecteurs de K. (4.6)

Nous avions mentionné le fait qu'à partir d'un code $(1, \leq \ell)$-identifiant C d'un graphe G il était facile d'obtenir un code ℓ-superimposé en prenant les vecteurs caractéristiques des ensembles identifiants $I(X, C)$ pour tout $X \subseteq V$ tel que $|X| \leq \ell$ (paragraphe 1.2.2). Réciproquement, nous avons montré qu'il était possible de construire un graphe orienté muni d'un code $(1, \leq \ell)$-identifiant à partir d'un code ℓ-superimposé optimum (Théorème 1.1). Dans cette section, nous construisons un graphe non-orienté muni d'un code $(1, \leq \ell)$-identifiant à partir d'un code ℓ-superimposé maximal, et nous prouvons le théorème suivant :

Théorème 4.3 ([GM05])
Il existe une fonction $c(n) = O(\ell^4 \log n)$ et une famille de graphes $(\mathcal{G}_i)_{i \in \mathbb{N}}$, telle que, pour tout $i \in \mathbb{N}$, $\mathcal{G}_i$ a n_i sommets, avec $n_i \to \infty$ quand $i \to \infty$, et telle que G_i admet un code $(1, \leq \ell)$-identifiant de cardinalité $c(n_i)$. De plus, il existe un algorithme construisant de tels graphes.

Dans la suite nous décrivons notre construction, dont nous prouvons la validité dans le paragraphe 4.2.2.2.

4.2.2.1 Construction explicite

Soit $\ell \geq 2$. Dans ce paragraphe nous décrivons les étapes nécessaires à la construction d'un graphe $\mathcal{G}$ muni d'un code $(1, \leq \ell)$-identifiant C. La validité de cette construction est discutée dans le paragraphe suivant.

1. Soit $N = \lceil \ell^2 \log n \rceil$ et soit K un code ℓ-superimposé maximal de $\{0,1\}^N$, *i.e.* il n'existe pas de $K' \neq K$, $K' \supset K$, tel que K' soit un code ℓ-superimposé. Soit k la cardinalité de K : $K = \{V_1, \ldots, V_k\}$.
2. Soit M la matrice $N \times k$ dont les colonnes sont les vecteurs de K. Soit M' une sous-matrice $N \times N$ de M telle qu'il y ait au moins un 1 sur chaque ligne de M'.
3. Soit H un graphe connexe admettant un code $(1, \leq \ell)$-identifiant. À partir de M et M', construisons un graphe $\mathcal{G} = \mathcal{G}(M, M')$ et un code $(1, \leq \ell)$-identifiant $C = C(M, M')$ de $\mathcal{G}$ comme suit. Le sous-graphe de $\mathcal{G}$ induit par le code, $\mathcal{G}[C]$, est l'union disjointe de N copies de H. Dans chaque copie H_i de H nous spécifions un sommet h_i, $i = 1, \ldots, N$. Ces

sommets $h_1, \ldots, h_N$ vont être tels que

$$N(V(\mathcal{G}) \setminus C) \ = \ \{h_1, \ldots, h_N\}.$$

À chaque colonne V_j de $M \setminus M'$ on associe un sommet $v_j = \phi(V_j)$ de $\mathcal{G}$, dont les voisins sont tous les h_i pour lesquels i est tel que la i-ème coordonnée de V_j est égale à 1 (voir Figure 4.6). Il n'y a pas d'arêtes entre les v_j, donc le vecteur V_j est le vecteur caractéristique de l'ensemble identifiant de v_j, qui est égal au voisinage de v_j.

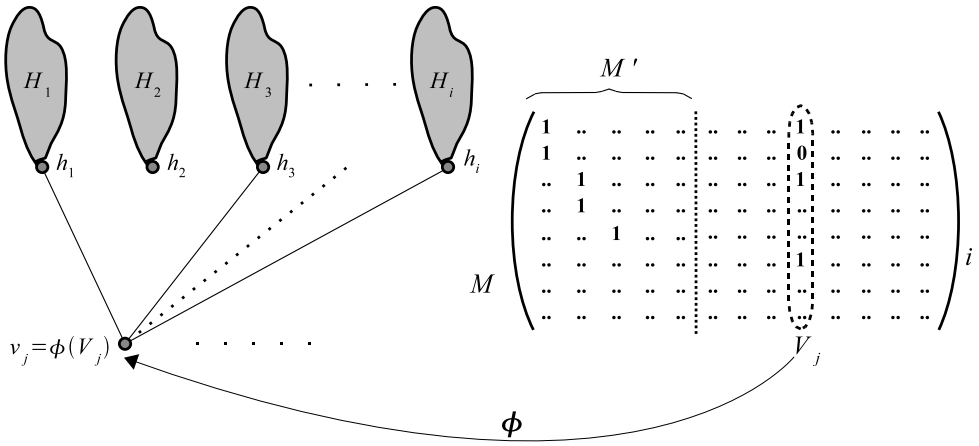


FIG. 4.6 – *Construction d'un graphe $\mathcal{G} = \mathcal{G}(M, M')$ muni d'un code $(1, \leq \ell)$-identifiant $C = C(M, M')$ à partir de M et M'.*

4.2.2.2 Preuve de la validité de la construction

Nous montrons la validité de la construction décrite au-dessus et nous prouvons le Théorème 4.3.

À l'étape 2 de la construction, nous avions besoin du résultat suivant :

Lemme 4.3 ([GM05])
Soit M une matrice $n \times m$ $(n \leq m)$ en $0-1$ ne comportant pas de ligne composée uniquement de 0. Alors il existe une sous-matrice $n \times n$ M' de M telle qu'il y ait au moins un 1 sur chaque ligne de M'.

Preuve : Soit M une matrice satisfaisant les conditions de l'énoncé. Soient $M_1, \ldots, M_m$ les colonnes de M.

On raisonne par induction sur n. Sans perte de généralité, on peut supposer qu'il existe $p \leq n$ tel que $M_{i,1} = 1$ pour tout $i \leq p$ et $M_{j,1} = 0$ pour tout $j > p$. Si $p = n$ alors on prend $M' = M_1$. Sinon, soit P la matrice obtenue comme la restriction à $M_2, \ldots, M_m$ des lignes de M indexées par $p+1, \ldots, n$. Par induction, il existe une sous-matrice P' de P telle qu'il y ait au moins un 1 sur chaque ligne de P'. On prend alors M' la sous-matrice de M définie par les colonnes de P' plus M_1. □

Comme la matrice d'un code ℓ-superimposé maximal de $\{0,1\}^N$ est une matrice $0-1$ dont aucune ligne ne contient que des 0, nous avons alors :

Lemme 4.4 ([GM05])
Soit M une matrice $N \times k$ dont les colonnes sont les vecteurs d'un code ℓ-superimposé maximal de $\{0,1\}^N$. Alors il existe M' une sous-matrice $N \times N$ de M telle qu'il y ait au moins un 1 sur chaque ligne de M'.

Dans la suite nous aurons aussi besoin du lemme suivant :

Lemme 4.5 ([GM05])
Soit M une sous-matrice $N \times k$ dont les colonnes sont les vecteurs de K un code ℓ-superimposé maximal de $\{0,1\}^N$, et soit M' une sous-matrice $N \times N$ de M telle qu'il y ait au moins un 1 sur chaque ligne de M' (d'après le lemme précédent une telle sous-matrice existe). Alors chaque colonne de $M \setminus M'$ comporte au moins ℓ coordonnées non nulles.

Preuve : Soit V une colonne de $M \setminus M'$ ayant strictement moins de ℓ coordonnées non nulles. Comme il y a au moins un 1 sur chaque ligne de M' alors nous pouvons trouver un ensemble d'au plus $\ell - 1$ colonnes de M', $\{V_1, \ldots, V_l\}$, $l \leq \ell - 1$, tel que

$$V \leq \sum_{i=1}^{l} V_i,$$

où $\sum$ désigne le `OU` bit-à-bit. Ceci implique $\sum_{i=1}^{l} V_i + V = \sum_{i=1}^{l} V_i$, ce qui contredit le fait que K soit un code ℓ-superimposé. □

Maintenant nous pouvons prouver la validité de la construction décrite au-dessus et prouver le Théorème 4.3.

Preuve du Théorème 4.3 : Le cas $\ell = 1$ est déjà bien connu (voir par exemple le Théorème 4.1 du paragraphe 4.2.1). Soit donc $\ell \geq 2$. Soit $N = \lceil \ell^2 \log n \rceil$ et soit K un code ℓ-superimposé maximal de $\{0,1\}^N$. Par le Théorème 1.5 nous savons qu'il existe un tel K satisfaisant $|K| \geq \Omega(n)$.

Soit M la matrice dont les colonnes sont les vecteurs de K. À l'étape 2 de la construction nous avions besoin de trouver une sous-matrice $N \times N$ de M ayant au moins un 1 sur chacun de ses colonnes : comme K est maximal, alors par le Lemme 4.4 une telle sous-matrice existe. À l'étape 3 de la construction nous avions également besoin d'un graphe connexe H à $\Theta(\ell^2)$ sommets admettant un code $(1, \leq \ell)$-identifiant : par la Proposition 4.3 nous savons que si ℓ est la puissance d'un nombre premier alors il existe un tel H (prendre $H = G_\ell$ comme construit dans la Proposition 4.3). Si ℓ n'est pas la puissance d'un nombre premier, alors par la Conjecture de Bertrand – prouvée, entre autres[1], par P. Erdős en 1932[2] [Erd32] – nous savons qu'il existe un nombre premier p dans l'intervalle $[\ell, 2\ell]$, et nous prenons $H = G_p$ comme construit dans la Proposition 4.3. Comme $p \geq \ell$, alors G_p admet un code $(1, \leq p)$-identifiant implique que G_p admet un code $(1, \leq \ell)$-identifiant. G_p a également $\Theta(\ell^2)$ sommets.

Maintenant, soient $\mathcal{G}$ et C comme décrits à l'étape 3 de la construction. On montre que C est un code $(1, \leq \ell)$-identifiant de $\mathcal{G}$. Soient X et Y deux sous-ensembles de sommets distincts de $\mathcal{G}$ de cardinalité inférieure ou égale à ℓ. On montre que $I(X) = I(Y)$ si et seulement si $X = Y$. La preuve se décompose en deux parties : nous montrons d'abord que $I(X) = I(Y) \Rightarrow X \cap C = Y \cap C$, et ensuite nous montrons que $I(X) = I(Y) \Rightarrow X \setminus C = Y \setminus C$. Dans la suite on suppose que $I(X) = I(Y)$.

(a) Par l'absurde, supposons que $X \cap C \neq Y \cap C$, et soit H_i une composante connexe de $\mathcal{G}[C]$ sur laquelle X et Y diffèrent. En notant $X_i = X \cap H_i$ et $Y_i = Y \cap H_i$, on a $X_i \neq Y_i$. Comme $V(H_i) \subset C$ et $V(H_i)$ est un code $(1, \leq \ell)$-identifiant de H_i, alors on a $I(X_i) \neq I(Y_i)$. S'il existe un $h \in H_i$, $h \neq h_i$, tel que $h \in I(X_i) \Delta I(Y_i)$, alors on obtient une contradiction puisque $h \notin N(X \setminus X_i) \cup N(Y \setminus Y_i)$: le voisinage de $h \neq h_i$ est contenu dans H_i, et par conséquent $h \in I(X_i) \Delta I(Y_i) \Rightarrow h \in I(X) \Delta I(Y)$. Ainsi $I(X_i) \Delta I(Y_i) = \{h_i\}$. Par le Lemme 4.1 nous pouvons supposer que $|X_i| = \ell$, *i.e.* $X = X_i \subseteq H_i$ et $h_i \in I(X) \setminus I(Y_i)$. Comme nous avons supposé que $I(X) = I(Y)$, cela signifie qu'il existe y un voisin de h_i dans $Y \setminus C$. Par le Lemme 4.5, y est voisin d'au moins ℓ sommets de C (rappelons qu'à chaque vecteur colonne w de $M - M'$ nous avons associé un sommet voisin de h_i pour tout i tel que la i-ième coordonnée de w soit 1). Comme $\ell \geq 2$, alors il existe $h_j \in C$, $h_j \neq h_i$, tel que $h_j \in I(Y) \setminus I(X)$: ceci contredit $I(X) = I(Y)$.

(b) Soient $X' = X \setminus C$ et $Y' = Y \setminus C$. À chaque $h_i \in I(X') \Delta I(Y')$ on peut associer un unique $h'_i \in X \cap C = Y \cap C$. En effet, comme $I(X) = I(Y)$, alors

[1] P. L. Chebyshev est le premier à montrer cette conjecture en 1850. S. Ramanujan en donnera également une preuve, plus simple.

[2] Il s'agit en fait de sa première publication, il avait alors 19 ans.

pour tout h_i dans, disons, $I(X') \setminus I(Y')$, il existe un $h'_i \in Y \cap H_i = X \cap H_i$ tel que $h_i \in N(h'_i)$. Nous pouvons donc construire une injection $I(X')\Delta I(Y') \hookrightarrow X \cap C = Y \cap C$. Ceci montre que :

$$|X| \ \geq \ |X'| + |I(X')\Delta I(Y')| \qquad \text{et} \qquad |Y| \ \geq \ |Y'| + |I(X')\Delta I(Y')|. \quad (4.7)$$

On rappelle qu'aux ensembles $X' = \{v_p\}_{p\in P}$ et $Y' = \{v_q\}_{q\in P}$ correspondent $\phi^{-1}(X) = \{V_p\}_{p\in P}$ et $\phi^{-1}(Y) = \{V_q\}_{q\in Q}$ deux ensembles de vecteurs colonnes de la matrice $M - M'$. On note que $|I(X')\Delta I(Y')|$ est égal au nombre de coordonnées sur lesquelles $\{V_p\}_{p\in P}$ et $\{V_q\}_{q\in Q}$ diffèrent. Soit $\mathcal{I}$ l'ensemble de coordonnées sur lequel $\{V_p\}_{p\in P}$ et $\{V_q\}_{q\in Q}$ diffèrent : $|\mathcal{I}| = |I(X')\Delta I(Y')|$. Pour toute coordonnée $i \in \mathcal{I}$, soit $W_{\tau(i)}$ le vecteur colonne de M' ayant sa i-ième coordonnée égale à 1. Par définition de $W_{\tau(i)}$, on a :

$$\sum_{p\in P} V_p + \sum_{i\in\mathcal{I}} W_{\tau(i)} \ = \ \sum_{q\in Q} V_q + \sum_{i\in\mathcal{I}} W_{\tau(i)}.$$

Comme M est la matrice d'un code ℓ-superimposé, ceci implique :

$$|P| + |\mathcal{I}| \ > \ \ell \qquad \text{ou} \qquad |Q| + |\mathcal{I}| \ > \ \ell.$$

Par (4.7), comme $|P| = |X'|$, $|Q| = |Y'|$, et $|\mathcal{I}| = |I(X')\Delta I(Y')|$, alors on a :

$$|X| \ > \ \ell \qquad \text{ou} \qquad |Y| \ > \ \ell,$$

ce qui est une contradiction.
Ainsi C est un code $(1, \leq \ell)$-identifiant de $\mathcal{G}$. C a pour cardinalité $N \times |H|$, et $\mathcal{G}$ a $N \times |H| + (|K| - N)$ sommets. Comme $N = \lceil \ell^2 \log n \rceil$, $|K| \geq \Omega(n)$ et $|H| = \Theta(\ell^2)$, alors on a

$$|C| \ = \ \Theta(\ell^2)\lceil \ell^2 \log n \rceil$$

et

$$|\mathcal{G}| \ = \ \Omega\left(n\right)$$

d'où

$$|C| \ = \ O\left(\ell^4 \log |\mathcal{G}|\right).$$

□

Pour $t \geq 1$, et $\ell = 1, 2$ la construction de codes $(t, \leq \ell)$-identifiants de faible cardinalité a été déjà abordée dans [BHL00, BHL01, HLR01, KCL98, KCLA99, LR01]. Dans ces papiers, les auteurs utilisaient des codes couvrants pour construire des codes identifiant des ensembles de sommets dans l'hypercube. Il nous semble que les codes superimposés sont un bon outil pour construire des codes identifiants.

Dans le Chapitre 1 (paragraphe 1.2.2) nous avions mentionné le fait que nous pouvions facilement obtenir un code ℓ-superimposé à partir d'un code $(1, \leq \ell)$-identifiant. Ici nous avons établi une correspondance dans l'autre sens, qui nous permet de construire un graphe à n sommets admettant un code $(1, \leq \ell)$-identifiant à partir d'un code ℓ-superimposé maximal de dimension $\lceil \ell^2 \log n \rceil$.

Cette construction utilise l'existence de graphes connexes à $\Theta(\ell^2)$ sommets admettant un code $(1, \leq \ell)$-identifiant (Proposition 4.3).

4.3 Graphes ayant des codes identifiants de grande cardinalité

Dans la section précédente nous avons discuté de la construction de graphes ayant un code identifiant de faible cardinalité. À l'inverse, on peut se demander quels sont les "mauvais graphes", tels que la cardinalité minimum d'un code identifiant de ces graphes soit grande. Dans cette section on résout cette question dans les cas des codes t-identifiants : on montre que tout graphe à n sommets possédant un code t-identifiant admet un code t-identifiant de cardinalité inférieure ou égale à $n-1$, et on exhibe une famille de graphes $(G_n)_{n\in\mathbb{N}}$ telle que la cardinalité minimum d'un code t-identifiant de G_n est égale à $n-1$. Nous montrons aussi que pour tout $t \geq 1$ il existe un unique graphe infini $G = (V, E)$ admettant V pour unique code t-identifiant.

On commence par régler le cas des codes 1-identifiants dans le cas des graphes infinis.

Théorème 4.4 ([GM])
Soit $G = (V, E)$ un graphe de degré maximum borné admettant un code 1-identifiant. Alors il existe $x \in V$ tel que $V \setminus \{x\}$ est un code 1-identifiant de G, sauf si G n'a pas d'arêtes.

Preuve : Si G n'a pas d'arêtes alors le seul code 1-identifiant de G est clairement V. Sinon, il suffit de considérer une composante connexe de G de cardinalité au moins deux. Supposons donc que G est un graphe connexe d'au moins deux sommets.

Pour un sommet x de G, comme V est un code 1-identifiant de G, vérifier que $C := V \setminus \{x\}$ est un code 1-identifiant de G est équivalent à vérifier que tous les sommets de $B_1(x)$ sont 1-séparés de tous les sommets de $V \setminus B_1(x)$. En effet, tous les sommets de G sont 1-couverts par un sommet de C

puisque G est connexe. Comme V est un code 1-identifiant de G, alors tous les sommets de $B_1(x)$ sont 1-séparés les uns des autres :

$$\forall u, v \in B_1(x), B_1(u) \neq B_1(v) \;\Rightarrow\; B_1(u) \setminus \{x\} \neq B_1(v) \setminus \{x\}.$$

De même, tous les sommets de $V \setminus B_1(x)$ sont 1-séparés les uns des autres. Donc la seule chose qui puisse se passer est que l'on ait $u \in B_1(x)$ et $v \notin B_1(x)$ non 1-séparés par C. Dans ce cas on a $B_1(u) = B_1(v) \cup \{x\}$. Ainsi, pour tout sommet w différent de x, u, v on a : $uw \in E \iff vw \in E$. En d'autres termes, u 1-sépare w_1 de w_2 si et seulement si v 1-sépare w_1 de w_2, et ce pour tout $w_1, w_2 \neq u, v, w$.

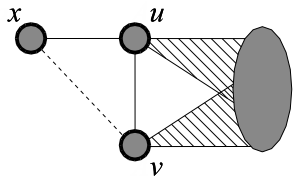


FIG. 4.7 – *Deux sommets u et v tels que $B_1(u) = B_1(v) \cup \{x\}$.*

Soit a un sommet de degré maximum de G. Si $V \setminus \{a\}$ est un code 1-identifiant de G, alors on a fini. Sinon, il existe deux sommets b, b' tels que $B_1(b) = B_1(b') \cup \{a\}$. Nous affirmons que $V \setminus \{b'\}$ est un code 1-identifiant de G. Vérifions que chaque sommet de $B_1(b')$ est 1-séparé de chaque sommet de $V \setminus B_1(b')$:

- b 1-sépare tout sommet de $V \setminus B_1(b') \setminus \{a\}$ de tout sommet de $B_1(b')$ puisque $B_1(b) = B_1(b') \cup \{a\}$.
- a se 1-sépare lui-même de b'
- a est 1-séparé de tout sommet $a' \in B_1(b') \setminus \{b'\}$: s'il existait $a' \in B_1(b') \setminus \{b'\}$ tel que $B_1(a') = B_1(a) \cup \{b'\}$, alors a' serait de degré strictement supérieur à celui de a, une contradiction.

□

Il est facile d'étendre ce résultat au cas des codes t-identifiants, $t \geq 1$.

Corollaire 4.1 ([GM])
Soit $t \geq 1$ et soit $G = (V, E)$ un graphe de degré maximum borné admettant un code t-identifiant. Alors il existe $x \in V$ tel que $V \setminus \{x\}$ est un code t-identifiant de G, sauf si G n'a pas d'arêtes.

Preuve : Étant donné un entier $t \geq 1$, il suffit de remarquer qu'un graphe $G = (V, E)$ admet un code t-identifiant si et seulement si G^t admet un code

1-identifiant, où G^t dénote la fermeture t-transitive de G : $G^t = (V, E')$, avec $uv \in E'$ si et seulement si il existe un chemin d'au plus t arcs entre u et v dans G. De plus, si G est de degré maximum borné, alors G^t l'est aussi. □

Nous pouvons donner un exemple de graphe infini ayant pour unique code 1-identifiant l'ensemble de tous ses sommets :

Soient G_1 et G_2 deux copies du graphe complet sur $\mathbb{Z}$. Joignons $x_i \in V(G_1)$ et $y_j \in V(G_2)$ pour tout i, j tels que $i \geq j$. On obtient un graphe infini $G_\infty = (V, E)$ ayant V pour unique code 1-identifiant (voir Figure 4.8). En effet, pour tout $i \in \mathbb{Z}$, on a $B_1(y_i) = B_1(y_{i-1}) \cup \{x_i\}$, de sorte que $x_i \in C$ pour tout code 1-identifiant C de G_∞. Par symétrie tous les y_i sont aussi dans C. Par ailleurs, il est facile de voir que V est un code 1-identifiant de G_∞.

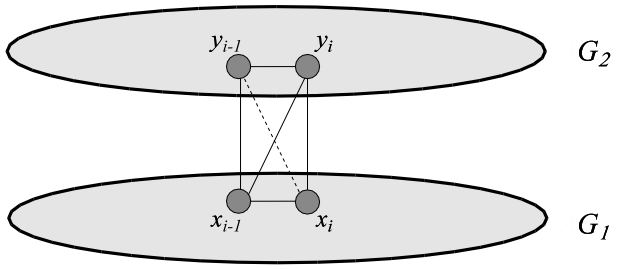


FIG. 4.8 – *Le graphe G_∞, qui a pour unique code 1-identifiant l'ensemble de ses sommets.*

Ce graphe est donné en exemple dans [CHLa]. Nous pouvons montrer que tout graphe $G = (V, E)$ admettant uniquement V comme code 1-identifiant est tel que tout sommet $v \in V$ est contenu dans un sous-graphe isomorphe à G_∞. Ceci montre que le degré minimum de G n'est pas borné.

Si on se restreint aux graphes finis, on retrouve un résultat de [CHLa] :

Corollaire 4.2 ([CHLa])
Soient $t \geq 1$ et $n > 1$ et soit G un graphe connexe à n sommets admettant un code t-identifiant. Alors G admet un code t-identifiant de cardinalité $n - 1$.

Cette borne est serrée, puisque l'on peut pour tout $t \geq 1$ et n assez grand construire un graphe connexe $G_{t,n}$ à n sommets admettant un code t-identifiant minimum de cardinalité $n-1$. Le cas $t = 1$ est trivial : on prend $G_{1,n} = K_{1,n}$. Il est facile de voir que pour tout $n \geq 3$, le graphe $K_n - M_n^*$,

où M_n^* est un couplage maximum de K_n, est aussi un graphe admettant un code 1-identifiant minimum de cardinalité $n-1$. Pour tout $t \geq 2$ et $n \geq 3t^2$, I. Charon, O. Hudry et A. Lobstein [CHLa] ont montré qu'on pouvait réaliser $K_n - M_n^*$ comme la fermeture t-transitive d'un graphe à n sommets $G_{t,n}$, qui est donc un graphe admettant un code 1-identifiant minimum de cardinalité $n-1$. C'est un résultat non trivial que nous citons ici sans preuve :

Théorème 4.5 ([CHLa])
Pour tout $t \geq 1$ et tout $n \geq 3t^2$, il existe un graphe $G_{t,n}$ admettant un code t-identifiant de cardinalité minimum $n-1$. □

Dans le cas des codes $(t, \leq \ell)$-identifiants, il existe des graphes G n'admettant que $V(G)$ comme code $(t, \leq \ell)$-identifiant. Par exemple, dans [HL03a] il est montré que, pour k et n assez grand, le tore $\mathbb{T}_{k,n}$ admet $V(\mathbb{T}_{k,n})$ pour unique code $(1, \leq 3)$-identifiant. Nous ignorons si pour tout $t \geq 1$, $\ell \geq 2$ il existe un graphe $G_{t,\ell}$ admettant uniquement $V(G_{t,\ell})$ comme code $(t, \leq \ell)$-identifiant.

4.4 Récapitulatif

Nous rappelons brièvement les résultats importants présentés dans ce chapitre, et nous donnons quelques questions suscitées par ces résultats. Cette section est volontairement rédigée dans un style télégraphique, qui vise à l'efficacité.

• Pour tout $t \geq 1$, le plus petit entier n tel qu'il existe un graphe à n sommets admettant un code t-identifiant est égal à $2t+1$ (Proposition 4.1).

• Dans le cas $\ell \geq 2$, nous pouvons construire des graphes connexes à $\Theta(\ell^2)$ sommets admettant un code $(1, \leq \ell)$-identifiant (Proposition 4.3). Nous ignorons la valeur du plus petit entier n tel qu'il existe un graphe connexe à n sommets admettant un code $(t, \leq \ell)$-identifiant, et nous posons ceci comme un problème ouvert :

Question 4.1 *Soit $n(t,\ell)$ la valeur du plus petit entier n tel qu'il existe un graphe connexe à n sommets admettant un code $(t, \leq \ell)$-identifiant. Quelle est la valeur de $n(t,\ell)$? On sait déjà que $n(t,1) = 2t+1$ (Proposition 4.1) et que $n(1,\ell) \leq O(\ell^2)$ (Proposition 4.3).*

• Nous avons montré que le nombre maximum d'arêtes que pouvait avoir un graphe admettant un code 1-identifiant était $\binom{n}{2} - \left\lfloor \frac{n}{2} \right\rfloor$ (Proposition 4.4).

Nous ignorons ce qu'il se passe dans le cas général, et nous posons ceci comme un problème ouvert :

Question 4.2 *Soit $m^{t,\ell}_{\max}(n)$ la valeur du plus grand entier m tel qu'il existe un graphe à n sommets admettant un code $(t, \leq \ell)$-identifiant. Quelle est la valeur de $m^{t,\ell}_{\max}(n)$? On sait déjà que $m^{1,1}_{\max}(n) = \binom{n}{2} - \left\lfloor \frac{n}{2} \right\rfloor$ (Proposition 4.4).*

• Pour tout $\ell \geq 1$ il existe un graphe G_ℓ de degré minimum ℓ admettant un code $(1, \leq \ell)$-identifiant (Proposition 4.5). Il est naturel de se demander si l'on peut avoir G_ℓ qui soit ℓ-régulier, et, plus généralement :

Question 4.3 *Soient $t \geq 1$ et $\ell \geq 1$. Soit $\delta(t, \ell)$ la valeur du plus petit entier δ tel qu'il existe un graphe de degré minimum δ admettant un code $(t, \leq \ell)$-identifiant. Quelle est la valeur de $\delta(t, \ell)$? Existe-t-il un graphe $\delta(t, \ell)$-régulier admettant un code $(t, \leq \ell)$-identifiant ? On sait que $\delta(t, \ell) \geq \ell$ ([LR01]), et que $\delta(1, \ell) = \ell$ (Proposition 4.5). Le cas du cycle montre que pour tout $t \geq 1$ il existe un graphe 2-régulier admettant un code $(t, \leq 2)$-identifiant.*

• Pour tout $t \geq 1$ et n assez grand il existe des graphes à n sommets admettant un code t-identifiant de cardinalité $\lceil \log(n+1) \rceil$ (Théorèmes 4.1 et 4.2). De plus dans le cas $t = 1$ et $n = 2^p - 1$, $p \geq 1$, on connaît le nombre minimum d'arêtes que peut avoir un graphe à n sommets ayant un code 1-identifiant de cardinalité p (Proposition 4.6).

• Dans le cas où $\ell \geq 2$, nous savons construire une famille de graphes admettant un code $(1, \leq \ell)$-identifiant de cardinalité $O(\ell^4 \log n)$ (Théorème 4.3). Nous savons par ailleurs que ceci n'est pas le minimum : nous verrons au Chapitre 5 qu'il existe une "construction" probabiliste d'une famille de graphes admettant un code $(1, \leq \ell)$-identifiant de cardinalité $O(\ell^2 \log n)$. Le cas général n'est pas donc connu ; nous posons ceci comme un problème ouvert :

Question 4.4 *Soient $t \geq 1$, $\ell \geq 1$ et $n \geq 1$. Soit $c_{t,\ell}(n)$ la valeur du plus petit entier c tel qu'il existe un graphe à n sommets admettant un code $(t, \leq \ell)$-identifiant de cardinalité c. Quelle est la valeur de $c_{t,\ell}(n)$? On sait déjà que $c_{t,1}(n) = \lceil \log(n+1) \rceil$ pour n assez grand (Théorèmes 4.1 et 4.2), et que $\Omega\left(\frac{\ell^2}{\log \ell} \log n\right) \leq c_{1,\ell}(n) \leq O(\ell^2 \log n)$ pour tout $\ell \geq 1$ et n assez grand (voir Théorème 1.6 pour la borne inférieure et la Proposition 5.5 pour la borne supérieure).*

• Pour tout $t \geq 1$, un graphe $G = (V, E)$ de degré maximum borné (G peut éventuellement être infini) admettant un code 1-identifiant contient

nécessairement un sommet $x \in V$ tel que $V \setminus \{x\}$ soit un code 1-identifiant de G, à moins que E soit vide (Corollaire 4.1). De plus, pour tout $n \geq 3t^2$ il existe un graphe à n sommets admettant un code t-identifiant de cardinalité minimum $n-1$ (Théorème 4.5). Cette question n'est pas résolue dans le cas général :

Question 4.5 *Soient $t \geq 1$, $\ell \geq 1$ et $n \geq 1$. Soit $c'_{t,\ell}(n)$ la valeur du plus grand entier c tel qu'il existe un graphe à n sommets admettant un code $(t, \leq \ell)$-identifiant minimum de cardinalité c. Quelle est la valeur de $c'_{t,\ell}(n)$? On sait déjà que $c'_{t,1}(n) = n-1$ pour n assez grand (Théorème 4.5).*

En complément du Théorème 1.4 et du Corollaire 4.2, il nous semble pertinent de mentionner le résultat suivant :

Théorème 4.6 ([CHLb])
Pour tout $t \geq$ et n assez grand, pour tout $c \in [\lceil \log(n+1) \rceil, n-1]$, il existe un graphe $G_{t,n}(c)$ admettant un code t-identifiant minimum de cardinalité c. □

Chapitre 5

Cas des graphes aléatoires

Dans ce chapitre nous présentons quelques résultats concernant les codes identifiants dans les graphes aléatoires. Nous allons tout d'abord faire quelques rappels de la théorie des probabilités et définir un modèle de graphe aléatoire en section 5.1. Dans la section suivante nous étudierons les codes 1-identifiants dans les graphes aléatoires, et la dernière section sera consacrée au cas des codes $(1, \leq \ell)$-identifiants.

L'étude des codes identifiants dans les graphes aléatoires peut être rapprochée du problème d'identification des sommets d'un graphe à l'aide de la séquence des distances, étudié par B. Bollobás dans le cas des graphes aléatoires [Bol82].

Les résultats présentés ici proviennent de [FMMRS], fruit d'une collaboration avec M. Ruszinkó, qui m'a accueilli à Budapest lors de mon séjour doctoral durant l'année 2004[1], R. Martin, rencontré au *Workshop on Algebraic and Geometric Methods in Combinatorics* en avril 2004 à Budapest, et A. Frieze et C. Smyth, qui ont amélioré certains de nos résultats après le retour de R. Martin aux États-Unis.

5.1 Graphes aléatoires

Nous commençons par faire des rappels élémentaires en probabilités.

[1]De janvier à septembre 2004, grâce à une bourse EURODOC de la Région Rhône-Alpes, j'ai eu la chance de pouvoir travailler avec M. Ruszinkó à SZTAKI, le laboratoire d'automatique et d'informatique de l'académie des sciences de Hongrie.

5.1.1 Rappels de probabilités

Dans ce qui suit nous n'avons besoin que de *probabilités discrètes*, que nous présentons ici.

Un *espace probabilisé* (discret) est un couple (Ω, P), où Ω est un ensemble fini et P une fonction de 2^Ω dans $[0,1]$ telle que

$$\sum_{\omega\in\Omega} \mathrm{P}(\omega) = 1.$$

La *probabilité* d'un *évènement* $A \in 2^\Omega$ est définie comme

$$\mathrm{P}(A) := \sum_{a\in A} \mathrm{P}(a).$$

Dans la suite nous utiliserons des propriétés élémentaires de P, parmi lesquelles :

- $\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$ pour tout évènement A, et
- $\mathrm{P}(\cup_i A_i) \leq \sum_i \mathrm{P}(A_i)$ pour toute collection d'évènements $(A_i)_i$.

Cette dernière propriété est connue sous le nom d'inégalité de Boole. En général il n'est pas vrai que $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$; deux évènements A, B satisfaisant cette égalité sont dits *indépendants*.

Une *variable aléatoire* (réelle) est une fonction $X : \Omega \to \mathbb{R}$. L'*espérance* d'une variable aléatoire est définie comme

$$E(X) \;:=\; \sum_{\omega\in\Omega} X(\omega)\,\mathrm{P}(\omega).$$

L'espérance satisfait l'inégalité

$$\mathrm{P}(|X| \geq a) \;\leq\; E(|X|)/a$$

pour tout a non nul. Dans le cas particulier où X est à valeurs dans $\mathbb{N}$, ceci implique $\mathrm{P}(X > 0) \leq E(X)$. Cette inégalité est connue sous le nom d'inégalité de Markov.

5.1.2 Graphes aléatoires

Il nous arrive parfois de vouloir quantifier la rareté d'apparition d'une propriété sur les graphes. Par exemple, au Chapitre 4 (Théorème 4.1) nous

avons mentionné que pour tout $n \geq 3$ il existait un graphe G_n à n sommets admettant un code 1-identifiant de cardinalité minimum $\lceil \log(n+1) \rceil$, et que pour tout $n \geq 3$ il existait un graphe G_n à n sommets admettant un code 1-identifiant de cardinalité minimum $n-1$. Nous savons par ailleurs que pour n assez grand, et pour tout $c \in [\lceil \log(n+1) \rceil, n-1]$, il existe un graphe $G_n(c)$ admettant un code t-identifiant minimum de cardinalité c (Théorème 4.6).

Dans ce contexte, il nous semble naturel de vouloir déterminer le caractère exceptionnel de la propriété d'admettre un code 1-identifiant de cardinalité minimum, disons, $n-1$: les graphes admettant un code 1-identifiant de cardinalité minimum $n-1$ sont-ils des graphes très spéciaux, ou en existe-t-il "un grand nombre" ? Si l'on "prend un graphe au hasard", quelle "chance" a-t-on de tomber sur un graphe ayant un code 1-identifiant de cardinalité minimum $n-1$? La notion de *graphe aléatoire* donne un sens formel à "prendre un graphe au hasard", et nous permet de montrer que les graphes admettant un code 1-identifiant de cardinalité minimum $n-1$ sont effectivement des graphes très spéciaux (Théorème 5.2).

Il existe plusieurs modèles de graphes aléatoires (voir [Bol85] pour un panorama complet), mais ici nous nous restreignons au modèle $\mathcal{G}(n,p)$, pour lequel l'espace probabilisé $(\Omega(n), \mathrm{P}(p))$ est le suivant :

- $\Omega(n)$ est l'ensemble des graphes étiquetés à n sommets,
- p est une fonction de n : $p : \mathbb{N} \to [0,1]$, et
- pour tout graphe étiqueté à n sommets et m arêtes G, on a $\mathrm{P}(G) = p^m(1-p)^{\binom{n}{2}-m}$.

Un graphe de l'espace probabilisé $\mathcal{G}(n,p) = (\Omega(n), \mathrm{P}(p))$ sera noté $G_{n,p}$. Alternativement, on peut aussi définir $\mathcal{G}(n,p)$ en disant que pour tout $i \neq j$, l'espérance de la variable aléatoire X_{ij} est $E(X_{ij}) = 1$, où $X_{ij}(G_{n,p})$ vaut 1 si i et j sont adjacents dans $G_{n,p}$ et 0 sinon. On peut donc voir tout graphe $G_{n,p} \in \mathcal{G}(n,p)$ comme un graphe construit en $\binom{n}{2}$ étapes, où à chaque étape on décide de l'existence d'une arête ij en tirant à pile ou face avec une pièce donnant "existe" avec probabilité p.

Ceci donne un sens à "prendre un graphe au hasard" : la "chance" de tomber sur un graphe ayant une propriété Π est donc la probabilité que $G_{n,p} \in \mathcal{G}(n,p)$ satisfasse Π.

Le cas spécial de l'espace $\mathcal{G}(n,1/2) = (\Omega(n), \mathrm{P}(p \equiv 1/2))$ a un intérêt particulier. En effet, pour $p = 1/2$, pour un graphe G donné à m arêtes, la probabilité que $G_{n,1/2}$ soit égal à G est

$$\mathrm{P}(G_{n,p} = G) = (1/2)^m(1-1/2)^{\binom{n}{2}-m} = \frac{1}{2^{\binom{n}{2}}} = \frac{1}{|\Omega(n)|} \qquad (5.1)$$

Autrement dit, il y a équiprobabilité : la loi $\mathrm{P}(p)$ est une loi uniforme lorsque $p \equiv 1/2$. Ceci nous est particulièrement utile lorsque nous voulons "prendre un graphe au hasard" de façon uniforme parmi les graphes étiquetés à n sommets. Le cas $p \equiv 1/2$ est donc en général celui que nous considérons en premier lorsque nous étudions une propriété sur les graphes aléatoires. Bien souvent, nous énonçons nos résultats pour tout p constant, mais le cas $p \equiv 1/2$ a une importance particulière à nos yeux, qui tient essentiellement à (5.1)).

Nous dirons qu'une propriété Π est satisfaite pour *presque tout graphe* de $\mathcal{G}(n,p)$ si et seulement si

$$\lim_{n\to\infty} \mathrm{P}(G_{n,p} \text{ a la propriété } \Pi) = 1. \tag{5.2}$$

De même, Π est satisfaite pour presque aucun graphe de $\mathcal{G}(n,p)$ si et seulement si $\mathrm{P}(G_{n,p}$ a la propriété $\Pi)$ tend vers 0 lorsque n tend vers l'infini.

Comme la terminologie le suggère, le fait que presque tout graphe de $\mathcal{G}(n,p)$ satisfasse une propriété Π n'implique pas que tout graphe satisfasse Π, mais la réciproque est vraie. Nous pouvons interpréter (5.2) comme le fait que le nombre de graphes ne satisfaisant pas Π est de plus en plus *négligeable* devant le nombre de graphes à n sommets, ce qui, bien sûr, n'implique pas que tous les graphes à n sommets satisfont Π. Ceci nous permet de quantifier le caractère exceptionnel d'une propriété Π : si Π est telle que presque sûrement aucun graphe de $\mathcal{G}(n,p)$ ne satisfait Π, alors les graphes satisfaisant Π sont des graphes "rares".

Lorsque p est une constante, $p \neq 0, 1$, P. Erdős et A. Rényi [ER60, ER61] se sont rendus compte que, pour un très grand nombre de propriétés fondamentales, la valeur de p n'avait aucune influence sur les calculs : soit Π est vraie pour presque tout graphe de $\mathcal{G}(n,p)$, soit Π est fausse pour presque tout graphe de $\mathcal{G}(n,p)$[2]. Pour qu'il se passe des choses intéressantes, il faut laisser p varier en fonction de n : $p = p(n)$. Dans ce cas, P. Erdős et A. Rényi remarquent que, bien souvent, une propriété fondamentale Π de $G_{n,p}$ apparaît de façon très brusque : il existe une fonction $s(n)$ telle que

– Pour tout $\epsilon > 0$, si $p(n)$ est telle que $\frac{p(n)-\epsilon}{s(n)} \to 0$, alors presque aucun graphe de $\mathcal{G}(n,p)$ ne satisfait la propriété Π,

[2]Plus tard il sera démontré [Fag76] que toute propriété de graphes pouvant être décrite par une expression logique du premier ordre – *i.e.* en utilisant le prédicat d'adjacence, la conjonction (ET), la disjonction (OU), la négation (NON), l'implication, et la quantification sur les sommets – soit est vraie pour presque tout graphe de $\mathcal{G}(n,p)$, soit est fausse pour presque tout graphe de $\mathcal{G}(n,p)$, et ce pour tout p constant, $p \neq 0, 1$. Une preuve relativement lisible de ce phénomène peut être trouvée dans [Win93].

– Pour tout $\epsilon > 0$, si $p(n)$ est telle que $\frac{p(n)+\epsilon}{s(n)} \to \infty$, alors presque tout graphe de $\mathcal{G}(n,p)$ satisfait la propriété Π.

La fonction $s(n)$ est appelée *fonction de seuil* de la propriété Π. Il existe des définitions plus spécifiques de fonctions de seuil, et dans ce chapitre les fonctions de seuil désigneront de façon générale l'étude de l'apparition et de la disparition brusque des propriétés étudiées. Par exemple, pour la propriété d'admettre un code 1-identifiant, nous verrons qu'il existe deux intervalles pour lesquels presque tout graphe de $\mathcal{G}(n,p)$ admet un code 1-identifiant (voir le Théorème 5.6 et la Figure 5.1).

5.1.3 La méthode probabiliste en Théorie des Graphes

P. Erdős fut un des premiers à "prendre un graphe au hasard". Son résultat de 1947 concernant la Théorie de Ramsey [Erd47] est souvent retenu comme point de départ de la méthode probabiliste en Théorie des Graphes :

Théorème 5.1 (Erdős)
Soit $R : \mathbb{N} \to \mathbb{N}$ la fonction telle que, pour tout $k \in \mathbb{N}$, tout graphe G à $R(k)$ sommets contienne soit une clique de cardinalité k, soit un stable de cardinalité k. Alors R vérifie :

$$R(k) \ > \ 2^{k/2}. \qquad \square$$

Depuis, l'utilisation des probabilités en Théorie des Graphes s'est développée et est devenu un outil standard d'appréhension de problèmes combinatoires. De nombreuses monographies existent sur le sujet, parmi lesquelles *Random Graphs* de B. Bollobás [Bol85], *The Probabilistic Method* de N. Alon et J. H. Spencer [AS00], ou encore *Random Graphs* de S. Janson, T. Łuczak et A. Ruciński [JLR00].

Comme le montre le Théorème 5.1, la méthode probabiliste donne parfois des résultats de nature déterministe. Ces résultats ont bien souvent le gros inconvénient de ne pas être constructifs, et c'est la critique principale qui leur est faite. En effet, ils sont souvent obtenus en montrant que

$$\mathrm{P}(G_{n,p} \text{ a la propriété } \Pi) \ > \ 0$$

pour un p bien choisi, ce qui montre qu'il existe des graphes satisfaisant Π. Voir par exemple la Proposition 5.5, pour laquelle il s'avèrera pertinent de poser $p = 1/\ell$ afin de déduire l'existence de graphes à n sommets admettant

un code $(1, \leq \ell)$-identifiant de cardinalité $\Theta(\ell^2 \log n)$.

En compensation, les résultats non-constructifs que l'on obtient sont souvent meilleurs que les résultats constructifs. Par exemple, il n'existe toujours pas de preuve constructive[3] de l'existence d'une constante c telle que $R(k) > (1+c)^k$. C'est aussi le cas pour les codes identifiants : dans le chapitre précédent, pour tout $\ell \geq 1$ et pour tout n assez grand, nous avons explicitement construit un graphe $\mathcal{G}_n$ admettant un code $(1, \leq \ell)$-identifiant de cardinalité $O(\ell^4 \log n)$ (Théorème 4.3). Dans ce chapitre nous allons montrer que pour tout $\ell \geq 1$ et pour tout n assez grand il existe un graphe $\mathcal{G}_n$ admettant un code $(1, \leq \ell)$-identifiant de cardinalité $O(\ell^2 \log n)$ (Proposition 5.5), ce qui améliore d'un coefficient ℓ^2 la construction du Théorème 4.3.

5.2 Cas des codes 1-identifiants

5.2.1 Cardinalité minimum d'un code 1-identifiant

Dans cette section nous déterminons avec précision la cardinalité minimum d'un code 1-identifiant d'un graphe aléatoire. Nous montrons le résultat suivant :

Théorème 5.2 ([FMMRS])
Soit p tel que p et $1-p$ sont tous deux supérieurs ou égaux à $4 \log\log n / \log n$. Pour un graphe G, soit $M(G)$ la cardinalité minimum d'un code 1-identifiant de G. Alors pour presque tout graphe $G_{n,p}$ on a

$$M(G_{n,p}) \ \sim \ \frac{2 \log n}{\log(1/q)},$$

i.e.

$$\lim_{n \to \infty} \mathrm{P}\left(M(G_{n,p}) \cdot \left(\frac{2 \log n}{\log(1/q)} \right)^{-1} \to 1 \right) = 1,$$

où q dénote la quantité $p^2 + (1-p)^2$.

La preuve de ce théorème est divisée en deux parties. Nous montrons d'abord que la cardinalité minimum d'un code 1-identifiant d'un graphe

[3] P. Erdős, qui avait l'habitude "d'intéresser la partie" lorsqu'il était question de mathématiques, a mis 100 dollars en jeu pour ce problème. Pour la récompense, il faut s'adresser à F. Chung et R. Graham, qui perpétuent la tradition, et qui sont par ailleurs auteurs d'une remarquable compilation de conjectures et problèmes laissés par P. Erdős [CG98].

aléatoire est presque sûrement inférieure ou égale à $\frac{(2+\epsilon)\log n}{\log(1/q)}$, avec ϵ petit. C'est la partie facile de la preuve, obtenue simplement en montrant que n'importe quel ensemble de sommets de cardinalité $\frac{(2+\epsilon)\log n}{\log(1/q)}$ est presque sûrement un code 1-identifiant de $G_{n,p}$. Ceci est présenté dans le paragraphe suivant. Ensuite, nous montrons que, presque sûrement, aucun ensemble de sommets de cardinalité inférieure ou égale à $\frac{(2-\epsilon)\log n}{\log(1/q)}$ n'est un code 1-identifiant de $G_{n,p}$. Ce résultat fait appel à l'inégalité de Suen [Sue90], que nous rappellerons en temps voulu.

5.2.1.1 Borne supérieure

Nous avons besoin du résultat élémentaire suivant :

Lemme 5.1 ([FMMRS])
Soit C un sous-ensemble de c sommets de $G_{n,p}$. La probabilité que C ne soit pas un code 1-identifiant de $G_{n,p}$ est bornée par :

$$\begin{aligned} &\mathrm{P}(C \textit{ n'est pas un code 1-identifiant de } G_{n,p}) \\ &\quad \leq \binom{c}{2} pq^{c-2} + c(n-c)pq^{c-1} + \binom{n-c}{2} q^c, \end{aligned}$$

où q dénote la quantité $p^2 + (1-p)^2$.

Preuve : Soit C un sous-ensemble de c sommets de $G_{n,p}$. Pour toute paire de sommets distincts $x \neq y$, notons $A_{x,y}(C)$ l'évènement $\{N[x] \cap C = N[y] \cap C\}$. La probabilité que C ne soit pas un code 1-séparant de $G_{n,p}$ est bornée par :

$$\mathrm{P}(\cup_{x \neq y} A_{x,y}(C)) \leq \sum_{x \neq y} \mathrm{P}(A_{x,y}(C)).$$

Si $x \in C$ et $y \in C$, alors $\mathrm{P}(A_{x,y}(C)) = pq^{c-2}$; si $x \in C$ et $y \notin C$, ou si $x \notin C$ et $y \in C$, alors $\mathrm{P}(A_{x,y}(C)) = pq^{c-1}$; et si $x \notin C$ et $y \notin C$, alors $\mathrm{P}(A_{x,y}(C)) = q^c$. Ainsi on a :

$$\mathrm{P}(\cup_{x \neq y} A_{x,y}(C)) \leq \binom{c}{2} pq^{c-2} + c(n-c)pq^{c-1} + \binom{n-c}{2} q^c.$$

□

La borne supérieure découle de façon immédiate du Lemme précédent.

Lemme 5.2 ([FMMRS])
Soit ϵ tel que $n^\epsilon \to +\infty$, et soit p tel que $p = \omega((\log n)^{-1})$ et $1 - p =$

$\omega\left((\log n)^{-1}\right)$. *Alors presque tout graphe de* $\mathcal{G}(n,p)$ *admet un code* 1*-identifiant de cardinalité inférieure ou égale à*

$$\frac{(2+\epsilon)\log n}{\log(1/q)},$$

où q *dénote la quantité* $p^2+(1-p)^2$.

Preuve : Par le Lemme 5.1 on sait que

$$\begin{aligned}&\mathrm{P}(C \text{ n'est pas un code 1-identifiant de } G_{n,p})\\ &\quad\leq \binom{c}{2}pq^{c-2}+c(n-c)pq^{c-1}+\binom{n-c}{2}q^c\end{aligned}$$

pour tout sous-ensemble de sommets C de cardinalité $c=c(n)$. Soit $c(n)=\frac{(2+\epsilon)\log n}{\log(1/q)}$. Comme p et $1-p$ sont tous deux des $\omega\left((\log n)^{-1}\right)$, alors on a $c=o(n)$. Nous pouvons réécrire cette probabilité :

$$\begin{aligned}&\mathrm{P}\left(C \text{ n'est pas un code 1-identifiant de } G_{n,p}\right)\\ &\quad\leq (n-c)^2q^c\left[1+\frac{2c}{n-c-1}\frac{p}{q}+\frac{c(c-1)}{(n-c)(n-c-1)}\frac{p}{q^2}\right],\end{aligned}$$

$c(n)$ et p étant tels que le terme entre crochets tend vers 1, il nous reste à montrer que $(n-c)^2q^c$ tend vers 0 :

$$\begin{aligned}(n-c)^2q^c &= \exp\left\{2\log(n-c)+c\log q\right\}\\ &\leq \exp\left\{2\log(n-c)-2\log n-\epsilon\log n\right\}\\ &\leq \exp\left\{2\log\frac{n-c}{n}-\log n^{\epsilon}\right\}.\end{aligned}$$

Comme $\frac{n-c}{n}\leq 1$ et $n^{\epsilon}\to+\infty$, alors l'exposant tend vers $-\infty$, donc $(n-c)^2q^{c-2}p$ tend vers 0, ce qui conduit au résultat désiré. □

Puisque de façon déterministe la cardinalité minimum d'un code 1-identifiant est bornée inférieurement par $\lceil\log(n+1)\rceil$ (voir Théorème 1.4), ceci montre que presque sûrement la cardinalité minimum d'un code 1-identifiant de $G_{n,p}$ est en $\Theta(\log n)$. Comme nous l'avions annoncé, nous sommes en fait capables de montrer que la valeur de la constante est 2, ce qui est fait dans le paragraphe suivant.

5.2.1.2 Borne inférieure

Dans ce paragraphe nous avons besoin de l'inégalité de Suen [Sue90], améliorée par S. Janson dans [Jan98]. Cette inégalité ressemble au Lemme

Local de Lovász [EL75], dans le sens où l'on utilise un graphe de dépendance d'évènements. Nous présentons ce résultat avec les conventions de notation de [Jan98].

Soit $(A_i)_{i\in\mathcal{I}}$ un ensemble d'évènements, et pour tout $i \in \mathcal{I}$ soit $\mathbb{I}_i$ la fonction indicatrice de l'évènement A_i : pour un évènement élémentaire ω, $\mathbb{I}_i(\omega) = 1$ si et seulement si $\omega \in A_i$. Soit p_i l'espérance de $\mathbb{I}_i$ pour tout $i \in \mathcal{I}$, et soit $X = \sum_{i\in\mathcal{I}} \mathbb{I}_i$.

Le *graphe de dépendance* associé aux évènements $(A_i)_{i\in\mathcal{I}}$ a pour ensemble de sommets $\mathcal{I}$. Il n'y a pas d'arêtes entre deux sous-ensembles disjoints de sommets I_1 et I_2 si toute combinaison booléenne des $A_i, i \in I_1$ est indépendante de toute combinaison booléenne des $A_j, j \in I_2$. Par combinaison booléenne d'évènements on entend une combinaison de ces évènements liés par les opérateurs standard de la théorie des ensembles (union, intersection ou complémentation) : par exemple $A_1 \cup (A_2 \setminus A_3)$ est une combinaison booléenne de A_1, A_2, A_3, puisque on peut réécrire $A_2 \setminus A_3 = A_2 \cap \overline{A_3}$.
L'adjacence sera notée $\sim$, et on notera $k \sim \{i, j\}$ pour signifier que le sommet k est adjacent à i ou j (ou aux deux).

Cette notion de graphe de dépendance est plus forte que celle du Lemme Local de Lovász, pour lequel on exige seulement que le fait qu'il n'y ait pas d'arêtes entre i et $I \subseteq \mathcal{I}$, $i \notin I$, implique que A_i soit indépendant de toute combinaison booléenne des A_j, $j \in I$.

Pour la version de l'inégalité de Suen dont nous avons besoin, il nous faut définir trois paramètres :
- $\mu := \sum_{i\in\mathcal{I}} p_i$,
- $\Delta := \sum_{\{\{i,j\}|i\sim j\}} E(\mathbb{I}_i\mathbb{I}_j)$,
- $\delta := \max_{i\in\mathcal{I}} \sum_{j\sim i} p_j$.

L'inégalité de Suen s'énonce alors comme suit :

Théorème 5.3 (Suen)
Avec les notations précédentes, soit $X = \sum_{i\in\mathcal{I}} \mathbb{I}_i$. Alors on a :

$$\mathrm{P}(X = 0) \;\leq\; \exp\left\{-\mu + \Delta e^{2\delta}\right\}. \qquad \square$$

Cette inégalité nous permet de donner une estimation de la probabilité qu'aucun évènement ne se réalise parmi une collection d'évènements dépendants les uns des autres. Dans le cas où les évènements sont tous indépendants cela est trivial :

$$\mathrm{P}(\cap_{i\in\mathcal{I}} A_i \mid A_i \text{ indépendant de } A_j \text{ pour tout } i \neq j) \;=\; \prod_{i\in\mathcal{I}} \mathrm{P}(A_i).$$

Dans le cas qui nous intéresse, nous voulons obtenir une borne supérieure de la probabilité P(C est un code 1-identifiant), et nous allons utiliser le fait que C est un code 1-identifiant si et seulement si aucune paire de sommets distincts u, v ne vérifie $N[u] \cap C = N[v] \cap C$: l'ensemble $\mathcal{I}$ sera donc égal à l'ensemble des paires de sommets distincts de $V \setminus C$.

Lemme 5.3 ([FMMRS])
Soit p tel que

$$2p(1-p) \ \geq \ \frac{4\log\log n}{\log n},$$

et soit $\epsilon = \frac{3\log\log n}{\log n}$. Alors, presque sûrement, il n'existe pas de code 1-identifiant de $G_{n,p}$ de cardinalité inférieure ou égale à

$$\frac{(2-\epsilon)\log n}{\log(1/q)},$$

où $q = p^2 + (1-p)^2$.

Preuve : Soit C un sous-ensemble de sommets de $G_{n,p}$ de cardinalité

$$c \ := \ \left\lfloor \frac{(2-\epsilon)\log n}{\log(1/q)} \right\rfloor.$$

Ceci implique que

$$n^{\epsilon-2} \ \leq \ q^c \leq n^{\epsilon-2}/q \leq 2n^{\epsilon-2}.$$

On utilise l'inégalité de Suen pour borner la probabilité que C soit un code 1-identifiant de $G_{n,p}$. Soit $\mathcal{I}$ l'ensemble des paires de sommets de $V \setminus C$. Soit A_i l'évènement

$$A_i\{N[u] \cap C \ = \ N[v] \cap C\},$$

où u et v sont les deux sommets correspondant à i. Soit X la variable aléatoire définie par

$$X \ = \ \sum_{i \in \mathcal{I}} \mathbb{I}_i.$$

Ainsi on a

$$C \text{ est un code 1-identifiant de } G_{n,p} \implies X = 0.$$

Le graphe de dépendance est défini par $i \sim j$ si et seulement si les paires de sommets correspondants ont une intersection non vide. Ainsi, on a $p_i = q^c$ pour tout $i \in \mathcal{I}$. De même, si $i \sim j$ alors on a

$$E(\mathbb{I}_i\mathbb{I}_j) \ = \ \left(p^3 + (1-p)^3\right)^c.$$

Comme $|\mathcal{I}| = \binom{n-c}{2}$, $|\{\{i,j\} : i \sim j\}| = 3\binom{n-c}{3}$ et $|\{j : j \sim i\}| = 2(n-c-2)$ pour tout $i \in \mathcal{I}$, alors ceci nous donne

– $\mu = \binom{n-c}{2} q^c$,
– $\Delta = 3\binom{n-c}{3} (p^3 + (1-p)^3)^c$, et
– $\delta = 2(n-c-2)q^c$.

On peut alors appliquer l'inégalité de Suen (Théorème 5.3) :

$$\begin{aligned}
&\mathrm{P}\left(C \text{ est un code 1-identifiant de } G_{n,p}\right) \\
&\quad \leq \mathrm{P}(X=0) \\
&\quad \leq \exp\left\{-\binom{n-c}{2} q^c + 3\binom{n-c}{3}(p^3+(1-p)^3)^c e^{4nq^c}\right\} \\
&\quad \leq \exp\left\{-\binom{n-c}{2} q^c \left(1 - n\left(1-\frac{1-q}{2q}\right)^c e^{4nq^c}\right)\right\} \\
&\quad \leq \exp\left\{-\frac{n^\epsilon}{5}\left(1 - n\exp\left\{-\frac{1-q}{2q}\frac{(2-\epsilon)\log n}{\log 1/q} + O(n^{\epsilon-1})\right)\right\}\right\}.
\end{aligned}$$

Comme la fonction $x \mapsto \frac{x-1}{x \log x}$ est décroissante sur l'intervalle $[0,1]$, on peut borner la probabilité qu'il existe un code 1-identifiant de cardinalité c dans $G_{n,p}$:

$$\begin{aligned}
&\mathrm{P}\left(\text{il existe un code 1-identifiant de } G_{n,p} \text{ de cardinalité } c\right) \\
&\quad \leq \binom{n}{c} \exp\left\{-\frac{n^\epsilon}{5}\left(1 - n\exp\left\{\frac{\theta}{2(1-\theta)}\frac{(2-\epsilon)\log n}{\log(1-\theta)} + O(n^{\epsilon-1})\right)\right\}\right\} \\
&\quad \leq \binom{n}{c} \exp\left\{-\frac{n^\epsilon}{5}\left(1 - n^{-\epsilon/2+\theta/2+O(\theta^2)}\right)\right\}.
\end{aligned}$$

Comme

$$\binom{n}{c} \;=\; e^{O((\log n)^3/\log\log n)}$$

et

$$n^\epsilon \;=\; \Omega((\log n)^3)$$

alors on a

$$\mathrm{P}\left(\text{il existe un code 1-identifiant de } G_{n,p} \text{ de cardinalité } c\right) \;\to\; 0$$

quand $n \to \infty$, ce qui achève la preuve de ce lemme et du Théorème 5.2. □

Dans le cas particulier $p = 1/2$, ceci nous donne :

Corollaire 5.1
Presque tout graphe de $\mathcal{G}(n, 1/2)$ admet un code 1-identifiant de cardinalité minimum égale à $c(G_{n,1/2}) \sim 2\log n$.

5.2.2 Fonctions de seuil pour l'existence d'un code 1-identifiant

Afin de calculer les fonctions de seuil pour l'existence d'un code 1-identifiant, nous avons besoin de deux résultats fondamentaux de P. Erdős et A. Rényi [ER60, ER61], que nous présentons ici comme dans [Bol85]. Ces théorèmes donnent les fonctions de seuil pour le nombre de composantes connexes de $G_{n,p}$ qui sont des arbres.

Théorème 5.4 (Erdős-Rényi)
Soit X la variable aléatoire égale au nombre de sommets isolés de $G_{n,p}$. Alors on a :

1. *Si $pn - \log n \to -\infty$ alors pour tout $L \in \mathbb{R}$ on a $\mathrm{P}(X \geq L) \to 1$.*
2. *Si $pn - \log n \to x$ pour un certain $x \in \mathbb{R}$ alors X converge en probabilité vers la distribution de Poisson de moyenne $\lambda := e^{-x}$, i.e. $\mathrm{P}(X = r)$ tend vers $e^{-\lambda}\frac{\lambda^r}{r!}$ pour tout $r \geq 0$.*
3. *Si $pn - \log n \to +\infty$ alors $X = 0$ pour presque tout graphe de $\mathcal{G}(n,p)$.*

□

Théorème 5.5 (Erdős-Rényi)
Pour tout $k \geq 2$, soit T_k la variable aléatoire égale au nombre de composantes connexes de $G_{n,p}$ qui sont des arbres à k sommets. Alors on a :

1. *Si $p = o(n^{-\frac{k}{k-1}})$ alors $T_k = 0$ pour presque tout graphe de $\mathcal{G}(n,p)$.*
2. *Si $p \sim cn^{-\frac{k}{k-1}}$ pour une certaine constante $c > 0$ alors T_k converge en probabilité vers la distribution de Poisson de moyenne $\lambda := c^{k-1}\frac{k^{k-2}}{k!}$, i.e. $\mathrm{P}(T_k = r)$ tend vers $e^{-\lambda}\frac{\lambda^r}{r!}$ pour tout $r \geq 0$.*
3. *Si $pn^{\frac{k}{k-1}} \to +\infty$ et $pkn - \log n - (k-1)\log\log n \to -\infty$ alors pour tout $L \in \mathbb{R}$ on a $\mathrm{P}(T_k \geq L) \to 1$.*
4. *Si $pkn - \log n - (k-1)\log\log n \to x$ pour un certain $x \in \mathbb{R}$ alors T_k converge en probabilité vers la distribution de Poisson de moyenne $\frac{e^{-x}}{k \times k!}$.*
5. *Si $pkn - \log n - (k-1)\log\log n \to +\infty$ alors $T_k = 0$ pour presque tout graphe de $\mathcal{G}(n,p)$.*

□

Si $p \neq 1$ est constante, le Lemme 5.1 nous dit que presque tout graphe de $\mathcal{G}(n,p)$ admet un code 1-identifiant. Mais si p est maintenant une fonction de n, alors cela n'est plus nécessairement vrai. Par exemple, si $p = p(n)$ est trop grand, alors $G_{n,p}$ contient presque sûrement deux sommets universels

(*i.e.* deux sommets adjacents à tous les sommets du graphe), ce qui empêche $G_{n,p}$ d'admettre un code 1-identifiant. D'autre part, si p est si petit que $G_{n,p}$ n'a presque sûrement aucune arête, alors $G_{n,p}$ a un code 1-identifiant. Mais si p est petit de telle sorte qu'il existe presque sûrement des arêtes isolées dans $G_{n,p}$, alors $G_{n,p}$ n'a pas de code 1-identifiant. En fait, nous montrons que les sommets universels et les arêtes isolées sont les seuls obstacles pour admettre un code 1-identifiant dans $\mathcal{G}(n,p)$.

Théorème 5.6 ([FMMRS])
Pour tout $\epsilon > 0$ on a :

- *Si $p = o(n^{-2})$, alors presque tout graphe de $\mathcal{G}(n,p)$ admet un code 1-identifiant (et, presque sûrement, ce code 1-identifiant est unique : l'ensemble de tous les sommets de $G_{n,p}$),*
- *si $pn^2 \to +\infty$ et $p \leq \frac{1}{2n}(\log n + (1-\epsilon)\log\log n)$, alors presque sûrement aucun graphe de $\mathcal{G}(n,p)$ n'admet de code 1-identifiant,*
- *si $\frac{1}{2n}(\log n + (1+\epsilon)\log\log n) \leq p \leq 1 - \frac{1}{n}(\log n + \epsilon\log\log n)$, alors presque tout graphe de $\mathcal{G}(n,p)$ admet un code 1-identifiant,*
- *si $p \geq 1 - \frac{1}{n}(\log n - \epsilon\log\log n)$, alors presque aucun graphe de $\mathcal{G}(n,p)$ n'admet de code 1-identifiant.*

On remarque que l'intervalle non-trivial d'existence d'un code 1-identifiant est asymétrique, puisque sa borne inférieure est asymptotiquement égale à $\frac{\log n}{2n}$, alors que sa borne supérieure est de l'ordre de $1 - \frac{\log n}{n}$. Ceci provient du fait que nous devons 1-séparer toutes les paires de sommets *adjacents*. En effet, dès lors qu'ils sont 1-couverts, deux sommets non-adjacents sont automatiquement 1-séparés. Intuitivement, dans un graphe dense (*i.e.* lorsque p tend vers 1), deux sommets quelconques sont presque sûrement adjacents, et nous avons donc à 1-séparer un grand nombre de sommets. Lorsque p tend vers 0, la plupart des sommets de $G_{n,p}$ sont non-adjacents, et nous n'avons à considérer qu'un petit nombre de paires de sommets.

Nous décomposons la preuve du Théorème 5.6 en les quatre propositions qui suivent.

Proposition 5.1 ([FMMRS])
Si $p = o(n^{-2})$, alors presque tout graphe de $\mathcal{G}(n,p)$ admet un code 1-identifiant.

Preuve : Pour un tel p, $G_{n,p}$ n'a presque sûrement aucune arête, et admet donc $V(G_{n,p})$ comme unique code 1-identifiant. □

Proposition 5.2 ([FMMRS])
Pour tout $\epsilon > 0$, si $pn^2 \to +\infty$ et $p \leq \frac{1}{2n}(\log n + (1-\epsilon)\log\log n)$, alors presque sûrement aucun graphe de $\mathcal{G}(n,p)$ n'admet de code 1-identifiant.

Preuve : On applique le point 3 du Théorème 5.5 avec $k = 2$: pour un tel p presque tout graphe de $\mathcal{G}(n,p)$ admet un arbre à deux sommets comme composante connexe, *i.e.* une arête isolée. Un graphe contenant une arête isolée n'admet pas de code 1-identifiant. □

Proposition 5.3 ([FMMRS])
Pour tout $\epsilon > 0$, si

$$p \;\geq\; \frac{1}{2n}\left(\log n + (1+\epsilon)\log\log n\right)$$

et

$$p \;\leq\; 1 - \frac{1}{n}\left(\log n + \epsilon \log\log n\right),$$

alors presque tout graphe de $\mathcal{G}(n,p)$ admet un code 1-identifiant.

Preuve : L'ensemble de tous les sommets de $G_{n,p}$ est un code 1-identifiant si et seulement si c'est un code 1-séparateur. Par le Lemme 5.1, la probabilité que l'ensemble de tous les sommets de $G_{n,p}$ ne soit pas un code 1-séparateur est inférieure ou égale à

$$\binom{n}{2} p\left(p^2 + (1-p)^2\right)^{n-2} = f^n(p).$$

La fonction

$$f^n : x \mapsto \binom{n}{2} x\left(x^2 + (1-x)^2\right)^{n-2}$$

croît de 0 jusqu'à un $\alpha_n = \Theta\left(n^{-\frac{1}{2}}\right)$, puis décroît jusqu'à un $\beta_n = \frac{1}{2} - \Theta\left(n^{-\frac{1}{2}}\right)$, et croît de nouveau jusqu'à 1. Comme pour n assez grand $n^{-\frac{1}{2}}$ tend vers 0 moins vite que $\frac{1}{2n}\left(\log n + (1+\epsilon)\log\log n\right)$, alors le maximum de $f^n(p)$ sur l'intervalle

$$\left[\frac{1}{2n}\left(\log n + (1+\epsilon)\log\log n\right), 1 - \frac{1}{n}\left(\log n + \epsilon\log\log n\right)\right]$$

est atteint pour

$$p \;=\; \frac{1}{2n}\left(\log n + (1+\epsilon)\log\log n\right)$$

ou

$$p \;=\; 1 - \frac{1}{n}\left(\log n + \epsilon\log\log n\right).$$

Il suffit alors de montrer que

$$f^n\left(\frac{1}{2n}\left(\log n + (1+\epsilon)\log\log n\right)\right)$$

et

$$f^n\left(1 - \frac{1}{n}\left(\log n + \epsilon \log\log n\right)\right)$$

tendent tous deux vers 0 lorsque n tend vers l'infini, ce qui est facile à vérifier. □

Proposition 5.4 ([FMMRS])
Pour tout $\epsilon > 0$, si $p \geq 1 - \frac{1}{n}\left(\log n - \epsilon \log\log n\right)$, alors presque aucun graphe de $\mathcal{G}(n,p)$ n'admet de code 1-identifiant.

Preuve : Nous utilisons le fait que le nombre de sommets universels (*i.e* des sommets voisins de tous les autres) de $\mathcal{G}(n,p)$ est égal au nombre de sommets isolés de $\mathcal{G}_{n,1-p}$. Par le point 1 du Théorème 5.4, il existe presque sûrement deux sommets universels dans $\mathcal{G}(n,p)$ pour un tel p. Un graphe admettant deux sommets universels n'admet pas de code 1-identifiant. □

On peut représenter les résultats du Théorème 5.6 par la Figure 5.1, où nous représentons l'évolution de P($G_{n,p}$ admet un code 1-identifiant) comme une fonction de $p(n)$. Il est à noter qu'il existe deux intervalles d'existence presque sûre d'un code 1-identifiant.

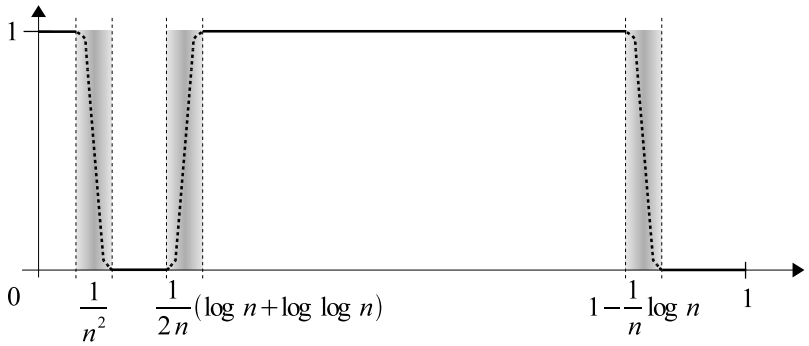


FIG. 5.1 – *Représentation graphique des fonctions de seuil pour la propriété d'admettre un code 1-identifiant. La valeur asymptotique de* P($G_{n,p}$ *admet un code 1-identifiant*) *est représentée sur l'axe vertical, et $p(n)$ sur l'axe horizontal.*

Les Théorèmes 5.4 et 5.5 d'Erdős et Rényi sont d'une précision telle que nous pouvons dire ce qu'il se passe *aux* seuils, *i.e.* lorsque p est dans l'une des trois zones grisées de la Figure 5.1.

Théorème 5.7 ([FMMRS])
Pour toute constante $c > 0$, si $p \sim cn^{-2}$, alors la probabilité qu'un graphe de $\mathcal{G}(n,p)$ admette un code 1-identifiant tend vers $e^{-\frac{c}{2}}$ lorsque n tend vers l'infini.

Preuve : On sait que $G_{n,p}$ n'admet pas de code 1-identifiant si et seulement si il existe une paire de sommets distincts $u \neq v$ tels que $N[u] = N[v]$, mais nous pouvons en fait nous restreindre aux sommets $u \neq v$ tels que $N[u] = N[v] = \{u, v\}$, c'est-à-dire aux arêtes isolées. En effet, la présence de $u \neq v$ tels que $N[u] = N[v]$ et $|N[u]| \geq 3$ implique la présence d'un triangle dans $G_{n,p}$, et pour un tel p la probabilité que $G_{n,p}$ contienne un triangle est bornée par $\binom{n}{3}p^3$, qui tend vers 0 lorsque n tend vers l'infini. Ainsi, pour un n grand, on a

$$\mathrm{P}\left(G_{n,p} \text{ n'admet pas de code 1-identifiant}\right) \sim \mathrm{P}\left(G_{n,p} \text{ contient une arête isolée}\right).$$

Grâce au point 2 du Théorème 5.5, nous savons que le nombre d'arêtes isolées converge en probabilité vers la distribution de Poisson de moyenne $\frac{c}{2}$. □

Théorème 5.8 ([FMMRS])
Pour toute constante $x \in \mathbb{R}$, si $2np - (\log n + \log\log n)$ tend vers x lorsque n tend vers l'infini, alors la probabilité qu'un graphe de $\mathcal{G}(n,p)$ admette un code 1-identifiant tend vers $e^{-e^{-\frac{x}{4}}}$ lorsque n tend vers l'infini.

Preuve : Comme dans le théorème précédent, il suffit de regarder les arêtes isolées. En effet, nous allons voir que la présence de deux sommets $u \neq v$ tels que $N[u] = N[v]$ et $|N[u]| \geq 4$ implique la présence d'un sous-graphe isomorphe à H_4 dans $G_{n,p}$, où H_4 est l'unique graphe à 4 sommets et 5 arêtes (voir Figure 5.2).

L'espérance du nombre de H_4 contenus dans $G_{n,p}$ est égal à $6\binom{n}{4}p^5$, qui tend vers 0 lorsque n tend vers l'infini pour un tel p. Ainsi, la probabilité que $G_{n,p}$ contienne deux sommets $u \neq v$ tels que $N[u] = N[v]$ et $|N[u]| \geq 4$ tend vers 0 lorsque n tend vers l'infini. Maintenant, regardons la probabilité que $G_{n,p}$ contienne deux sommets $u \neq v$ tels que $N[u] = N[v]$ et $|N[u]| = 3$. L'espérance du nombre de telles paires de sommets est $3\binom{n}{3}p^3(1-p)^{2(n-3)}$, qui tend vers 0 lorsque n tend vers l'infini. Presque sûrement, $G_{n,p}$ ne contient donc pas de paires de sommets $u \neq v$ telle que $N[u] = N[v]$ et $|N[u]| = 3$.

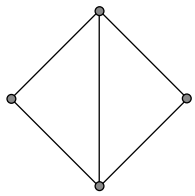


FIG. 5.2 – *L'unique graphe (à isomorphisme près) à 4 sommets et 5 arêtes. Ce graphe, que nous désignons par H_4 dans notre preuve, est parfois appelé diamant.*

Ainsi, pour n grand, on a

$$\mathrm{P}\left(G_{n,p} \text{ n'a pas de code 1-identifiant}\right) \sim \mathrm{P}\left(G_{n,p} \text{ a une arête isolée}\right).$$

Nous concluons en utilisant le Théorème 5.5. □

Théorème 5.9 ([FMMRS])
Pour toute constante $x \in \mathbb{R}$, si $n(1-p) - \log n$ tend vers x lorsque n tend vers l'infini, alors la probabilité qu'un graphe de $\mathcal{G}(n,p)$ admette un code 1-identifiant tend vers $e^{-e^{-x}}(1+e^{-x})$ lorsque n tend vers l'infini.

Preuve : On sait que $G_{n,p}$ n'admet pas de code 1-identifiant si et seulement si il existe deux sommets $u \neq v$ tels que $N[u] = N[v]$, mais nous pouvons en fait nous restreindre aux sommets universels, *i.e.* au cas $N[u] = N[v] = V(G_{n,p})$. En effet, le nombre de paires de sommets $u \neq v$ tels que $N[u] = N[v]$ et $|N[u]| \leq n-1$ est $\binom{n}{2} p\left((p^2+(1-p)^2)^{n-2} - p^{2(n-2)}\right)$, qui tend vers 0 lorsque n tend vers l'infini pour un tel p. Ainsi, pour n grand, on a

$$\begin{aligned}\mathrm{P}\left(G_{n,p} \text{ n'a pas de code 1-identifiant}\right) \\ \sim \mathrm{P}\left(G_{n,p} \text{ a deux sommets universels}\right).\end{aligned}$$

On conclut avec le point 2 du Théorème 5.4, en utilisant le fait que le nombre de sommets universels de $G_{n,p}$ est égal au nombre de sommets isolés de $G_{n,1-p}$. □

5.3 Cas des codes $(1, \leq \ell)$-identifiants

Dans un graphe G, soit (X,Y) une paire de sous-ensembles d'au plus ℓ sommets de G. La paire (X,Y) est dite *ℓ-maximale* si et seulement si on a :

– soit $|X| = \ell - 1$, $|Y| = \ell$, et $X \subseteq Y$;
– soit $|Y| = \ell - 1$, $|X| = \ell$, et $Y \subseteq X$;
– soit $|X| = \ell$ et $|Y| = \ell$.

Nous montrons que pour qu'un sous-ensemble C de sommets de G soit un code $(1, \leq \ell)$-identifiant de G, on peut se restreindre à séparer les paires ℓ-maximales de sous-ensembles de sommets de G :

Lemme 5.4 ([FMMRS])
Soit C un sous-ensemble de sommets d'un graphe $G = (V, E)$. Alors C est un code $(1, \leq \ell)$-identifiant de G si et seulement si $I(X, C) \neq \emptyset$ pour tout $X \subseteq V$ tel que $|X| \leq \ell$, et la condition

$$I(X, C) \neq I(Y, C)$$

est vraie pour toute paire ℓ-maximale (X, Y).

Preuve : Soit G un graphe. Par définition, un code $(1, \leq \ell)$-identifiant C de G ℓ-sépare les paires ℓ-maximales de G. Réciproquement, montrons qu'un sous-ensemble de sommets C satisfaisant les conditions de l'énoncé est un code $(1, \leq \ell)$-identifiant de G. Soit (X, Y) une paire de sous-ensembles d'au plus ℓ sommets de G. Si (X, Y) est ℓ-maximale alors il n'y a rien à faire. Si (X, Y) n'est pas ℓ-maximale, alors, sans perte de généralité, supposons que $|X| \leq |Y|$.

(a) Si $X \subseteq Y$, soit $Z \subseteq V \setminus Y$ de cardinalité $\ell - |Y|$ et soit $y_0 \in Y \setminus X$. Soient $X' := Y \cup Z \setminus \{y_0\}$ et $Y' := Y \cup Z$. Si C ne sépare pas X et Y, alors C ne sépare pas non plus X' et Y'. Comme la paire (X', Y') est ℓ-maximale, ceci implique que la paire (X, Y) est ℓ-séparée.

(b) si $X \not\subseteq Y$, soit $Z \subseteq V \setminus Y$ de cardinalité $\ell - |Y|$ et soit $T \subseteq Y \setminus X$ tel que $|X| + |T| + |Z| = \ell$. Soient $X' := X \cup T \cup Z$ et $Y' := Y \cup Z$. Si C ne sépare pas X et Y, alors C ne sépare pas non plus X' et Y'. Comme la paire (X', Y') est ℓ-maximale, ceci implique que la paire (X, Y) est ℓ-séparée.

□

5.3.1 Cardinalité d'un code $(1, \leq \ell)$-identifiant

Nous commençons par donner une estimation de la probabilité qu'un sous-ensemble arbitraire de sommets de $G_{n,p}$ soit un code $(1, \leq \ell)$-identifiant de $G_{n,p}$. Le résultat suivant est analogue au Lemme 5.1.

Lemme 5.5 ([FMMRS])
Soit $C \neq V$ un sous-ensemble de sommets de $G_{n,p}$. La probabilité que C ne soit pas un code $(1, \leq \ell)$-identifiant de $G_{n,p}$ est bornée par :

$$\begin{aligned} &\mathrm{P}(C \textit{ n'est pas un code } (1, \leq \ell)\textit{-identifiant}) \\ &\qquad \leq \; n^{2\ell} \left(1 - \min\{p, 2p(1-p)\}(1-p)^{\ell-1}\right)^{|C|-2\ell}. \end{aligned}$$

Dans le cas où $C = V$, on a :

$$\begin{aligned} &\mathrm{P}(V \textit{ n'est pas un code } (1, \leq \ell)\textit{-identifiant}) \\ &\qquad \leq \; n^{2\ell} \left(1 - (1-p)^{\ell}\right) \left(1 - \min\{p, 2p(1-p)\}(1-p)^{\ell-1}\right)^{n-2\ell}. \end{aligned}$$

Preuve : Soit $C \neq V$ un sous-ensemble de sommets de $G_{n,p}$. Soit $\mathcal{S}$ l'ensemble $\{(X,Y) \mid X \subseteq V, Y \subseteq V, X \neq Y, |X| \leq \ell, |Y| \leq \ell\}$, et soit $\mathcal{S}'$ l'ensemble des paires ℓ-maximales de $\mathcal{S}$. Pour toute paire ℓ-maximale $(X,Y) \in \mathcal{S}'$, soit $A_{X,Y}$ l'évènement $\{I(X,C) = I(Y,C)\}$. On a alors

$$\mathrm{P}(C \text{ n'est pas un code } (1, \leq \ell)\text{-identifiant de } G_{n,p}) \;\leq \sum_{(X,Y)\in\mathcal{S}'} \mathrm{P}(A_{X,Y}).$$

Comme $|\mathcal{S}'| = \Theta(n^{2\ell})$, il nous faut maintenant calculer une borne supérieure de $\mathrm{P}(A_{X,Y}) = \mathrm{P}(\cap_{z\in C} A_{X,Y}(z))$, où $A_{X,Y}(z)$ désigne l'évènement

$$\{z \in I(X,C) \cap I(Y,C)\} \cup \{z \notin (I(X,C) \cup I(Y,C))\}.$$

Nous pouvons décomposer cette quantité comme suit :

$$\mathrm{P}(A_{X,Y}) \leq \left\{ \prod_{z\in C\smallsetminus(X\cup Y)} \mathrm{P}(A_{X,Y}(z)) \right\} \times \mathrm{P}\left(\bigcap_{z\in C\cap(X\cup Y)} A_{X,Y}(z) \right) \qquad (5.3)$$

En effet, pour tout $z \in C \smallsetminus (X \cup Y)$, les évènements $A_{X,Y}(z)$ sont indépendants les uns des autres, et sont indépendants de l'intersection $\bigcap_{z\in C\cap(X\cup Y)} A_{X,Y}(z)$: pour deux z_1, z_2 distincts n'appartenant pas à $X \cup Y$, l'existence des arêtes entre z_1 et $X \cup Y$ est indépendante de l'existence des arêtes entre z_2 et $X \cup Y$. Ceci n'est plus vrai si l'on prend z_1 et z_2 appartenant à $X \cup Y$, ainsi nous ne pouvons pas développer le deuxième terme de (5.3). Nous allons donner une borne supérieure de chacun des termes de ce produit :

(a) Borne pour $\prod_{z\in C\smallsetminus(X\cup Y)} \mathrm{P}(A_{X,Y}(z))$:
On décompose $A_{X,Y}(z)$ comme suit :

$$\begin{aligned} A_{X,Y}(z) \;=\; & \{z \in I(X \cap Y)\} \\ & \cup \{z \in I(X \smallsetminus Y) \texttt{ ET } z \in I(Y \smallsetminus X) \texttt{ ET } z \notin I(X \cap Y)\} \\ & \cup \{z \notin I(X \cup Y)\}. \end{aligned}$$

Ceci conduit à :

$$\begin{aligned}\mathrm{P}(A_{X,Y}(z)) \leq f(X,Y) \quad := \quad & 1-(1-p)^{|X\cap Y|} \\ & +\left(1-(1-p)^{|X\smallsetminus Y|}\right)\left(1-(1-p)^{|Y\smallsetminus X|}\right) \\ & \quad (1-p)^{|X\cap Y|} \\ & +(1-p)^{|X\cup Y|}.\end{aligned}$$

Supposons sans perte de généralité que $|X| \leq |Y|$. Il y a alors deux cas à traiter :

(a.1) $X \subseteq Y$ **:**
Comme (X,Y) est ℓ-maximale, alors on a $|X| = \ell - 1$ et $|Y| = \ell$, donc :

$$f(X,Y) \;=\; 1-(1-p)^{\ell-1}+(1-p)^{\ell} \;=\; 1-p(1-p)^{\ell-1}.$$

(a.2) $X \not\subseteq Y$ **:**
Comme (X,Y) est ℓ-maximale, alors on a $|X| = |Y| = \ell$. Si $|X\cap Y| < \ell - 1$, soient alors $x_0 \in X \smallsetminus Y$ et $y_0 \in Y \smallsetminus X$. Soient $X' := X \smallsetminus \{x_0\} \cup \{y_0\}$ et $Y' = Y$. Comme (X,Y) est ℓ-maximale, alors (X',Y') est aussi ℓ-maximale, et il est facile de vérifier que $f(X',Y') > f(X,Y)$. En itérant ceci, nous voyons que le maximum de f est atteint dans le cas $|X\cap Y| = \ell - 1$, pour lequel on a :

$$\begin{aligned}\mathrm{P}(A_{X,Y}(z)) \;&\leq\; 1-(1-p)^{\ell-1}+(1-(1-p))^2(1-p)^{\ell-1} \\ &\quad +(1-p)^{\ell+1} \\ &=\; 1-2p(1-p)^{\ell}.\end{aligned}$$

Comme $|C\smallsetminus(X\cup Y)| \geq |C|-2\ell$, ceci nous donne la borne suivante :

$$\prod_{z\in C\smallsetminus(X\cup Y)} \mathrm{P}(A_{X,Y}(z)) \;\leq\; \left(1-\min\{p,2p(1-p)\}(1-p)^{\ell-1}\right)^{|C|-2\ell}.$$

(b) Borne pour $\mathrm{P}\left(\bigcap_{z\in C\cap(X\cup Y)} A_{X,Y}(z)\right)$ **:**
Dans le cas où $|C| < n$, nous bornons simplement cette quantité par 1 et nous obtenons le résultat annoncé. Si $C = V$, alors pour toute paire $(X,Y) \in \mathcal{S}'$ il existe un sommet $z_0 \in X\Delta Y$. Sans perte de généralité, supposons que $z_0 \in Y \smallsetminus X$. Pour un tel sommet z_0, nous avons simplement $A_{X,Y}(z_0) = \{z_0 \in N(X)\}$, qui a pour probabilité $1-(1-p)^{|X|}$. Comme $|X| \leq \ell$, on a alors :

$$\begin{aligned}\mathrm{P}\left(\bigcap_{z\in C\cap(X\cup Y)} A_{X,Y}(z)\right) \;&\leq\; \mathrm{P}(A_{X,Y}(z_0)) \\ &\leq\; 1-(1-p)^{\ell},\end{aligned}$$

ce qui nous donne la borne annoncée.

□

On cherche alors $|C|$ tel que la borne supérieure du Lemme précédent tende vers 0.

Théorème 5.10 ([FMMRS])
Soit ϵ tel que $n^\epsilon \to +\infty$, et p constant, $p \neq 0, 1$. Alors presque tout graphe de $\mathcal{G}(n,p)$ admet un code $(1, \leq \ell)$-identifiant C de cardinalité

$$|C| \ \leq \ \frac{2(\ell+\epsilon)\log n}{\log(1/q_\ell)},$$

où $q_\ell = 1 - \min\{p, 2p(1-p)\}(1-p)^{\ell-1}$.

Preuve : Avec les hypothèses précédentes, on sait par le Lemme 5.5 que

$$\mathrm{P}(C \text{ n'est pas un code } (1,\leq \ell)\text{-identifiant de } G_{n,p}) \ \leq \ n^{2\ell} q_\ell^{|C|-2\ell}.$$

Il suffit alors de vérifier que $n^{2\ell} q_\ell^{|C|-2\ell} \to 0$ lorsque $|C| = \frac{2(\ell+\epsilon)\log n}{\log(1/q_\ell)}$, ce qui est facile. □

Remarquer que ceci ne signifie pas que la cardinalité minimum d'un code $(1, \leq \ell)$-identifiant de $G_{n,p}$ est presque sûrement $O(\ell \log n)$, parce que $\frac{1}{\log(1/q_\ell)}$ est en fait une quantité en $O(2^\ell)$. Le théorème précédent montre en fait que la cardinalité minimum d'un code $(1, \leq \ell)$-identifiant de $G_{n,p}$ est presque sûrement $O(\ell 2^\ell \log n)$.

Le théorème précédent est analogue à la borne supérieure du Théorème 5.2, mais nous ignorons si cette borne est serrée. Peut-on aussi utiliser l'inégalité de Suen pour obtenir une borne inférieure similaire à celle du Lemme 5.3 ? Nous posons ceci comme un problème ouvert.

Nous rappelons que nous disposons d'une borne inférieure déterministe, provenant de la théorie des codes superimposés (voir Théorème 1.5) :

$$|C| \ \geq \ \frac{\ell^2}{\log \ell} \log n$$

pour tout code $(1, \leq \ell)$-identifiant C d'un graphe à n sommets.

Nous pouvons également déduire un résultat d'existence du Lemme 5.5, comme cela arrive souvent avec la méthode probabiliste.

Proposition 5.5 ([FMMRS])
Soit ϵ tel que $n^\epsilon \to +\infty$. Alors pour tout $n \in \mathbb{N}$ il existe un graphe G^n admettant un code $(1, \leq \ell)$-identifiant C^n de cardinalité

$$|C^n| \ \leq \ \sqrt{2}(\ell^2 + \epsilon) \log n.$$

Preuve : Soit $p = \frac{1}{\ell}$. Nous savons par le Lemme 5.5 que

$$\begin{aligned} &\mathrm{P}(C \text{ n'est pas un code } (1, \leq \ell)\text{-identifiant de } G_{n,p}) \\ &\leq \ n^{2\ell} \left(1 - 2\frac{1}{\ell}\left(1 - \frac{1}{\ell}\right)^\ell\right)^{|C|-2\ell}. \end{aligned}$$

Il existe une constante K_ℓ (qui dépend de ℓ mais pas de n) telle que

$$\begin{aligned} &\mathrm{P}(C \text{ n'est pas un code } (1, \leq \ell)\text{-identifiant de } G_{n,p}) \\ &\leq \ K_\ell n^{2\ell} \left(1 - 2\frac{1}{\ell}\left(1 - \frac{1}{\ell}\right)^\ell\right)^{|C|} \\ &\leq \ K_\ell \exp\left(2\ell \log n - 2\frac{|C|}{\ell}\left(1 - \frac{1}{\ell}\right)^\ell\right) \\ &\leq \ K_\ell \exp\left(2\ell \log n - \sqrt{2}\frac{|C|}{\ell}\right), \end{aligned}$$

puisque $\left(1 - \frac{1}{\ell}\right)^\ell \geq \frac{1}{\sqrt{2}}$ pour $\ell > 1$. Nous obtenons alors

$$\begin{aligned} &\mathrm{P}(C \text{ n'est pas un code } (1, \leq \ell)\text{-identifiant de } G_{n,p}) \\ &\leq \ K_\ell \exp\left(\left(2\ell - 2\ell - 2\frac{\epsilon}{\ell}\right)\log n\right) \\ &\leq \ K_\ell \exp\left(-2\frac{\epsilon}{\ell}\log n)\right) \\ &\leq \ K_\ell n^{-2\epsilon/\ell}. \end{aligned}$$

Puisque $n^\epsilon \to +\infty$, nous avons

$$\mathrm{P}(C \text{ n'est pas un code } (1, \leq \ell)\text{-identifiant de } G_{n,p}) \to 0.$$

Ainsi, pour un tel p, presque tout graphe de $\mathcal{G}(n,p)$ admet un code $(1, \leq \ell)$-identifiant de cardinalité $\sqrt{2}(\ell^2 + \epsilon)\log n$, en particulier il existe un graphe à n sommets G^n admettant un code $(1, \leq \ell)$-identifiant C^n de cardinalité $|C^n| \leq \sqrt{2}(\ell^2 + \epsilon)\log n$. □

Ce résultat d'existence est non-constructif, il ne nous dit rien sur la structure des graphes admettant un code $(1, \leq \ell)$-identifiant de cardinalité $\sqrt{2}\ell^2 \log n$ et ne nous fournit pas de procédé explicite pour en construire. Nous rappelons que nous avions donné en Chapitre 4 un résultat d'existence constructif de graphes à n sommets admettant un code $(1, \leq \ell)$-identifiant de cardinalité $\Theta(\ell^4 \log n)$ (Théorème 4.3). Nous posons comme un problème ouvert la construction *explicite* de graphes à n sommets admettant un code $(1, \leq \ell)$-identifiant de cardinalité $O(\ell^2 \log n)$.

5.3.2 Fonctions de seuil pour l'existence d'un code $(1, \leq \ell)$-identifiant

À la différence du cas $\ell = 1$, nous ne disposons que de résultats partiels sur les fonctions de seuil pour la propriété d'admettre un code $(1, \leq \ell)$-identifiant dans le cas $\ell > 1$. Nous présentons ici les résultats obtenus.

Certains arguments que nous avions invoqués dans le cas des codes 1-identifiants sont encore valables dans le cas général :

Proposition 5.6 ([FMMRS])
Si $p = o(n^{-2})$, alors presque tout graphe de $\mathcal{G}(n,p)$ admet un code $(1, \leq \ell)$-identifiant. □

En effet, pour un tel p, $G_{n,p}$ n'a presque sûrement aucune arête, et admet donc un unique code $(1, \leq \ell)$-identifiant $C = V$.

Proposition 5.7 ([FMMRS])
Soit $\epsilon > 0$. Si $p \geq 1 - \frac{1}{n}(\log n - \epsilon \log\log n)$ presque sûrement aucun graphe de $\mathcal{G}(n,p)$ n'admet de code $(1, \leq \ell)$-identifiant. □

Nous rappelons que pour un tel p le graphe $G_{n,p}$ contient presque sûrement deux sommets universels.

Maintenant, en utilisant le Lemme 5.5, nous pouvons obtenir un intervalle d'existence d'un code $(1, \leq \ell)$-identifiant dans $G_{n,p}$:

Proposition 5.8 ([FMMRS])
Soit $\epsilon > 0$. Si $p = p(n)$ vérifie

$$\frac{\ell 2^{\ell-1}}{n}(\log n + \epsilon \log\log n) \quad \leq \quad p \leq 1 - \left(\frac{1}{n}(\log n + \epsilon \log\log n)\right)^{1/\ell},$$

alors presque tout graphe de $\mathcal{G}(n,p)$ admet un code $(1, \leq \ell)$-identifiant.

Preuve : D'après le Lemme 5.5, nous avons :

$$\begin{aligned}
&\mathrm{P}(V \text{ n'est pas un code } (1,\leq \ell)\text{-identifiant de } G_{n,p}) \\
&\quad\leq \quad n^{2\ell}\left(1-(1-p)^{\ell}\right)\left(1-\min\{p,2p(1-p)\}(1-p)^{\ell-1}\right)^{n-2\ell} \\
&\quad\leq \quad \begin{cases} K_\ell n^{2\ell}\left(1-(1-p)^{\ell}\right)\exp\left(-pn(1-p)^{\ell-1}\right), & \text{si } p\leq 1/2\,; \\ K'_\ell n^{2\ell}\left(1-(1-p)^{\ell}\right)\exp\left(-2pn(1-p)^{\ell}\right), & \text{si } p\geq 1/2. \end{cases}
\end{aligned}$$

où K_ℓ et K'_ℓ sont deux constantes dépendant seulement de ℓ. Il est facile de vérifier que ces deux quantités tendent vers 0 lorsque p satisfait les inégalités de l'énoncé. □

Pour la proposition suivante nous avons besoin d'un résultat de Bollobás sur la séquence des degrés d'un graphe aléatoire, que nous citons ici comme dans [Bol85].

Théorème 5.11 (Bollobás)
Soit $\epsilon > 0$ fixé et soit p tel que $\epsilon n^{-\frac{3}{2}} \leq p \leq 1-\epsilon n^{-\frac{3}{2}}$. Soit k un nombre naturel et soit X_k la variable aléatoire égale au nombre de sommets de degré k dans $G_{n,p}$. Soit

$$\lambda_k \;:=\; \lambda_k(n) = n\binom{n-1}{k}p^k(1-p)^{n-k}.$$

Alors, pour tout $t \geq 0$ fixé, on a :

$$\lim \lambda_k(n) = +\infty \implies \lim \mathrm{P}(X_k \geq t) = 1. \qquad \square$$

Ceci nous est très utile puisque qu'un graphe contenant un sommet v de degré $1 \leq d(v) \leq \ell - 1$ n'admet pas de code $(1,\leq \ell)$-identifiant. En effet, il est impossible de séparer $X := N(v)$ de $Y := N(v) \cup \{v\} = N[v]$, puisque ces deux ensembles ont le même voisinage étendu : $N[X] = N[Y]$.

Proposition 5.9 ([FMMRS])
Pour tout $\epsilon > 0$, si p est tel que $pn^2 \to \infty$ et $p \leq \frac{1}{n}(\log n + (\ell - 1 - \epsilon)\log\log n)$, alors presque aucun graphe de $\mathcal{G}(n,p)$ n'admet de code $(1,\leq \ell)$-identifiant.

Preuve : On utilise le Théorème 5.11 avec $k = \ell - 1$ et $t = 1$. Il est facile de voir que si $pn^2 \to \infty$ et $p \leq \frac{1}{n}(\log n + (\ell - 1 - \epsilon)\log\log n)$, alors on a $\lambda_k(n) = n\binom{n-1}{k}p^k(1-p)^{n-k} \to +\infty$. Par conséquent, $G_{n,p}$ contient presque sûrement un sommet v de degré $\ell - 1$. Maintenant, soient $X := N(v)$ et $Y := N(v) \cup \{v\}$: X et Y sont tous deux de cardinalité inférieure ou égale à ℓ, et vérifient $N[X] = N[Y]$, donc $G_{n,p}$ n'admet pas de code $(1,\leq \ell)$-identifiant. □

Annexe : Graphes

Dans cette annexe nous donnons de brefs rappels de théorie des graphes et d'algorithmique. Les définitions et les termes employés proviennent d'ouvrages de références tels [Die00, GJ79, vLW92].

Graphes, sous-graphes

Un *graphe non-orienté* G est un couple (V, E), où E est un ensemble de paires (non-ordonnées) de points de V : $E \subseteq \{uv \mid u \in V, v \in V\}$. Noter que, possiblement, $uu \in E$: on dit en ce cas que uu est une *boucle* de G. Un graphe sans boucle est un graphe *simple*. Un *graphe orienté* G est un couple (V, A), où A est un ensemble de couples de points de V : $A \subseteq \{(u, v) \mid u \in V, v \in V\}$. Un élément de A du type (u, u) est également appelé *boucle* de G. Lorsque V est fini on parle de *graphe fini*. L'ensemble V est appelé ensemble des *sommets* du graphe et l'ensemble E est appelé ensemble des *arêtes* du graphe. Dans le cas d'un graphe orienté, l'ensemble A est appelé ensemble d'*arcs* du graphe.

Deux sommets u, v tels que $uv \in E$ (resp. $(u, v) \in A$) seront dit *adjacents* ou encore *voisins*. Pour un arc $(u, v) \in A$, u sera dit *voisin entrant* de v, et v sera dit *voisin sortant* de u. Deux arêtes ayant un sommet commun $uv \in E$ et $uw \in E$ (resp. $(u, v) \in A$ et $(u, w) \in A$; ou $(u, v) \in A$ et $(w, u) \in A$) seront dites *incidentes* en u. Le *degré* d'un sommet v d'un graphe non-orienté est le nombre de voisins de v, on le note $d(v)$. Dans le cas de graphes orientés, le *degré entrant* de v désigne le nombre de voisins entrants de v, noté $d^-(v)$, et le *degré sortant* de v désigne le nombre de voisins sortants de v, noté $d^+(v)$. L'ensemble des voisins de v est noté $N(v)$, l'ensemble des voisins entrants (resp. sortants) de v est noté $\Gamma^-(v)$ (resp. $\Gamma^+(v)$). Le *degré maximum* de G, noté $\Delta(G)$, est le degré maximum d'un sommet de G. Le *degré minimum* de G, noté $\delta(G)$, est le degré minimum d'un sommet de G. Lorsque $\delta(G) = \Delta(G) = r$ on dit que G est *r-régulier* :

tous les sommets de G ont le même degré.

Le théorème suivant lie le nombre d'arêtes d'un graphe avec la somme des degrés de ses sommets :

Théorème A.1 (Folklore)
Soit $G = (V, E)$ un graphe non-orienté sans boucle à m arêtes. Alors on a :

$$\sum_{v \in V} d(v) = 2m.$$

Preuve : Soit $G = (V, E)$ un graphe non-orienté sans boucle à m arêtes. Considérons le graphe biparti $H(G) = \{A, B\}$, où $A = V$ et $B = E$, tel que $v \in A$ et $e \in B$ sont voisins dans $H(G)$ si et seulement si e est une arête incidente à v dans G. Soit k le nombre d'arêtes de $H(G)$. Comme il y a $d(v)$ arêtes à chaque sommet v, alors on a $k = \sum_{v \in V} d(v)$. Par ailleurs, comme chaque arête est incidente à deux sommets, alors on a $k = 2m$. En rassemblant, $k = 2m = \sum_{v \in V} d(v)$. □

La terminologie provient des représentation graphiques habituelles des graphes, où les éléments de V sont des points du plan et les éléments de E ou de A sont des traits reliant ces points (voir Figure A.1). À ce titre, le terme *point* sera parfois utilisé à la place de sommet.

Dans le cas des graphes finis, le nombre de sommets sera souvent dénoté n et le nombre d'arêtes m. La notation $|G|$ sera parfois utilisée pour désigner le nombre de sommets de G. Les notations $V(G)$, $E(G)$ et $A(G)$ désignent, respectivement, l'ensemble des sommets, des arêtes et des arcs d'un graphe G.

On peut également voir un graphe $G = (V, E)$ comme une relation binaire sur l'ensemble V. Dans le cas où G est fini, on a l'habitude de représenter G comme une matrice $M \in \mathcal{M}_{n,n}(\{0, 1\})$, où $n = |V|$, et pour tout $i, j = 1, \ldots, n$ on a :

$$M_{ij} = 1 \text{ si et seulement si } v_i v_j \in E \text{ (resp. } (v_i, v_j) \in A)$$

La matrice M est appelée *matrice d'adjacence* de G. Dans le cas où G est non-orienté M est symétrique : $M_{ij} = M_{ji}$ pour tout i, j.

Un graphe H est un *sous-graphe* d'un graphe G s'il existe $V' \subseteq V$ et $E' \subseteq E_{|_{V'}}$ tel que $H = (V', E')$, où $E_{|_{V'}}$ désigne la restriction de E à V'^2 : $E_{|_{V'}} = \{uv \in E \mid u \in V', v \in V'\}$ (*cf.* Figure A.2). Le graphe G est parfois

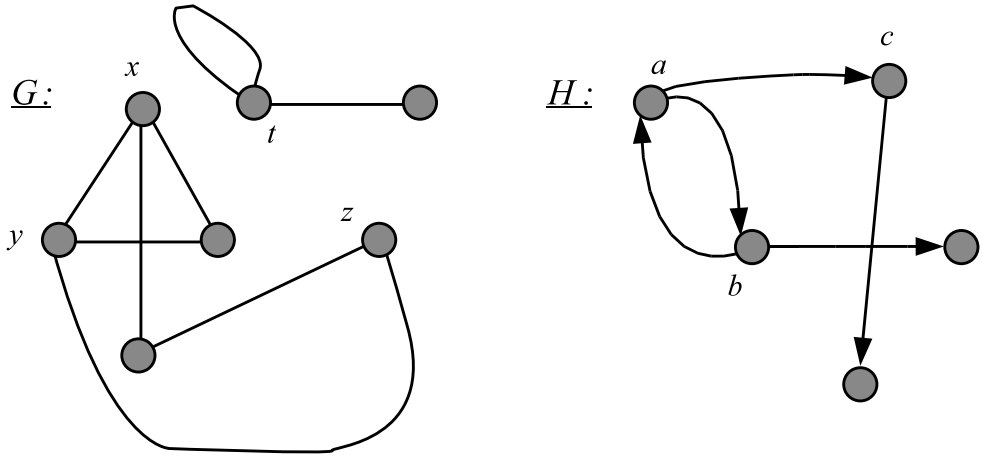


FIG. A.1 – *À gauche, un graphe non-orienté ; et un graphe orienté à droite. les sommets x et y sont adjacents, mais pas les sommets x et z. Le sommet x est de degré 3. L'arête tt est une boucle de G. Les arêtes xyet yz sont incidentes en y. Les sommets b et c sont les deux voisins sortants de a, b est aussi un voisin entrant de a : le voisinage sortant de a est donc $\{b, c\}$, et le voisinage entrant de a est $\{b\}$.*

appelé *supergraphe* de H. Lorsque $E = E_{|_{V'}}$ alors on dit que H est un *sous-graphe induit* de G. En ce cas on dit que H est *engendré* par V' et on écrit $H = G[V']$.

De façon algébrique, un sous-graphe H de G peut être défini par une application $f : V(H) \to V(G)$ telle que $uv \in E(H) \to f(u)f(v) \in E(G)$. Une telle application est appelée homomorphisme de graphes, c'est une application des sommets préservant la relation d'adjacence. Si $f : V(H) \to V(G)$ est bijective, on parle d'isomorphisme de graphes et on dira que H et G sont *isomorphes*. Autrement dit, deux graphes sont isomorphes s'ils ne diffèrent que par les noms qu'ont leurs sommets. Ainsi nous étudions souvent les graphes à isomorphisme près puisque nous nous intéressons aux propriétés combinatoires des graphes. Lorsque les sommets sont numérotés et que cet ordre a une importance on parle de *graphes étiquetés*.

Soient X et Y deux ensembles et soit G un graphe. Lorsque G est muni d'une fonction $w : V(G) \to X$ on dit que G est *valué*, et lorsque G est muni d'une fonction $w' : E(G) \to Y$ on dit que G est *pondéré*. Dans ce contexte, la *valuation* d'un sommet v désigne $w(v)$, et le *poids* d'une arête e désigne $w'(e)$. Parfois, les arêtes d'un graphe G sont pondérées afin de redéfinir la

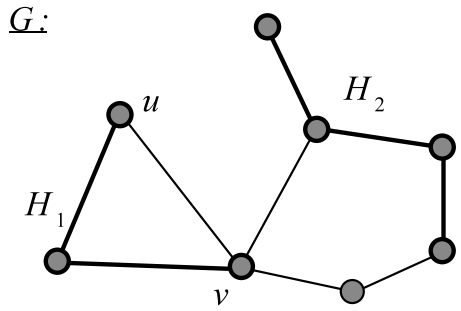


FIG. A.2 – *H_1 et H_2 sont tous les deux des sous-graphes de G. Seul H_2 est un sous-graphe induit, car u et v sont adjacents dans G mais pas dans H_1.*

distance entre deux sommets u, v : $d(u, v)$ est égale au poids minimum d'un chemin d'extrémités u et v, où le poids d'un chemin $u_1u_2 \dots u_k$ est égal à $u_1 \star u_2 \star \dots \star u_k$, avec $\star : Y^2 \to Y$ loi de composition associative de l'ensemble Y. Typiquement, $Y = \mathbb{N}$ et $x \star y$ est la somme de x et y.

Opérations sur les graphes

L'*union disjointe* de deux graphes G et H est le graphe $G \cup H := (V(G) \cup V(H), E(G) \cup E(H))$.

Étant donnés deux graphes H et G, le *produit cartésien* de H et G, noté $H \square G$, est le graphe ayant pour ensemble de sommets $V(H) \times V(G)$ tel que $(u_1, v_1)(u_2, v_2) \in E(H \square G)$ si $u_1 = u_2$ et $v_1v_2 \in E(G)$ ou si $u_1u_2 \in E(H)$ et $v_1 = v_2$. On note $\square^n G$ le produit cartésien $G \square G \square \dots \square G$, avec $n - 1$ carrés dans la formule.

Soit G un graphe et soit $t \geq 1$ un entier. La *fermeture t-transitive* de G, notée G^t, est le graphe ayant pour ensemble de sommets $V(G)$ tel que deux sommets u et v sont adjacents dans G^t si et seulement si ils sont à distance inférieure ou égale à t dans G (voir Figure A.4).

Graphes et sous-graphes remarquables

Un graphe G dont l'ensemble d'arêtes (resp. d'arcs) est vide est appelé *graphe vide* ou *stable*. Un sous-graphe vide d'un graphe G est appelé *stable*

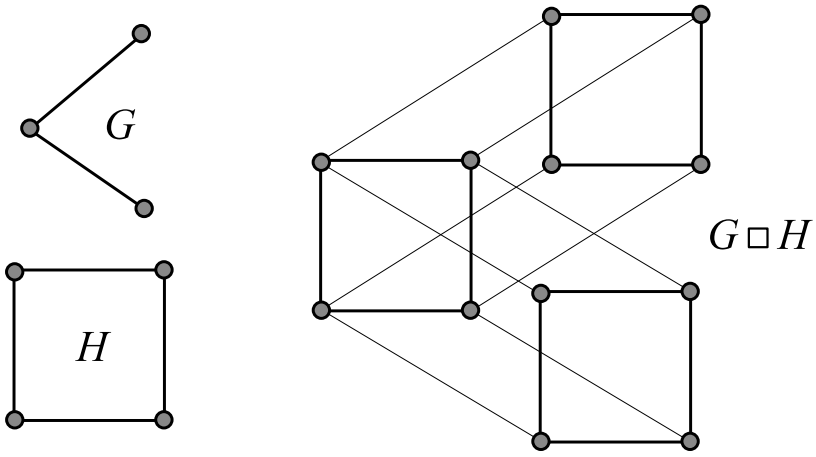


FIG. A.3 – *À gauche, union disjointe des graphes G et H ; à droite, le produit cartésien de G et H.*

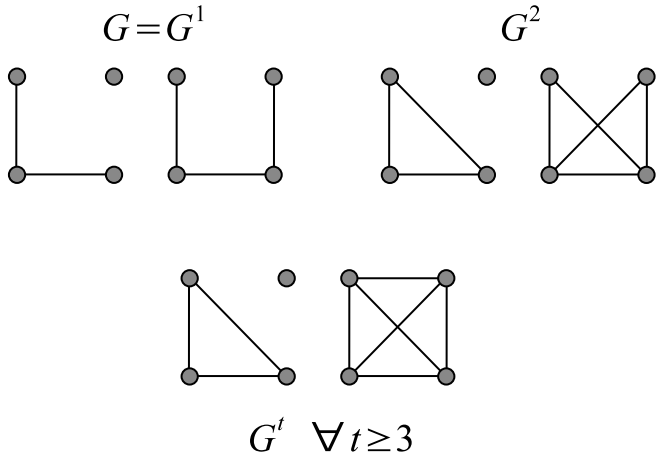


FIG. A.4 – *Exemple de fermetures t-transitives d'un graphe G pour tout $t \geq 1$.*

de G.

Un graphe G dont l'ensemble des sommets peut être partitionné en deux sous-ensembles, $V(G) = A \cup B$, tels qu'il n'y ait pas d'arêtes (resp. pas d'arcs) entre deux sommets de A et pas d'arêtes (resp. pas d'arcs) entre deux sommets de B, est appelé *graphe biparti*. Un graphe biparti sera souvent (et abusivement) noté $G = \{A, B\}$.

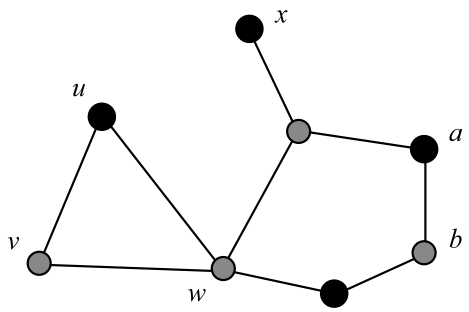


FIG. A.5 – *L'ensemble des sommets noirs forme un stable de G. Les sous-graphes $G[x]$, $G[a,b]$ et $G[u,v,w]$ sont des cliques de G.*

Un graphe biparti ayant tous ses sommets de degré un est appelé *couplage*. Un sous-graphe de d'un graphe G ayant pour ensemble de sommets $V(G)$ est appelé un *couplage parfait* de G.

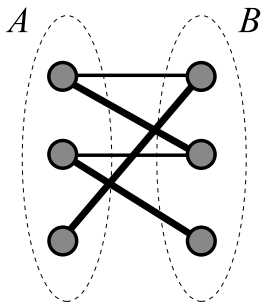


FIG. A.6 – *Exemple d'un graphe biparti $\{A, B\}$ muni d'un couplage parfait.*

Un graphe G tel que $E(G) = \{uv \mid u \in V(G), v \in V(G), u \neq v\}$ est appelé *graphe complet*. Un graphe complet à n sommets sera souvent dénoté K_n. Un sous-graphe de G qui est un graphe complet est appelé une *clique* de G.

Un *chemin* est un graphe G dont l'ensemble des sommets est $\{v_1, \ldots, v_n\}$ et l'ensemble des arêtes est $\{v_i v_{i+1} \mid i = 1, \ldots, n-1\}$. Les sommets v_1 et v_n sont appelées *extrémités* de G. Un chemin de n sommets sera souvent noté P_n, la *longueur* d'un chemin désigne son nombre d'arêtes. Le sous-graphe de P_n induit par les sommets $v_i, i_1 \leq i \leq i_2$, où $1 \leq i_1 \leq i_2 \leq n$, est encore un chemin, dénoté parfois $v_{i_1} P v_{i_2}$.

Un *cycle* est un graphe dont l'ensemble des sommets est $\{v_1, \ldots, v_n\}$ et

l'ensemble des arêtes est $\{u_i u_{i+1} \mid i = 1, \ldots, n-1\} \cup \{u_n u_1\}$. Un cycle de n sommets sera souvent noté $\mathcal{C}_n$, la *longueur* d'un cycle désigne son nombre d'arêtes.

Un *cheminement* d'un graphe G est un sous-graphe H de G ayant pour ensemble d'arêtes $E(H) = \{u_{i_j} u_{i_{j+1}} \mid u_{i_j} \in V(H)\ \forall j = 1, \ldots, k\}$. Noter qu'il peut exister $j \neq j'$ tel que $u_{i_j} = u_{i'_j}$: le cheminement emprunte éventuellement plusieurs fois le même sommet. Les sommets u_{i_1} et u_{i_k} sont les extrémités du cheminement. La longueur du cheminement est son nombre d'arêtes, $k-1$.

Un *circuit* d'un graphe G est un sous-graphe H de G ayant pour ensemble d'arêtes $E(H) = \{u_{i_j} u_{i_{j+1}} \mid u_{i_j} \in V(H)\ \forall j = 1, \ldots, k\} \cup \{u_{j_k} u_{j_1}\}$. La longueur du circuit est son nombre d'arêtes, k.

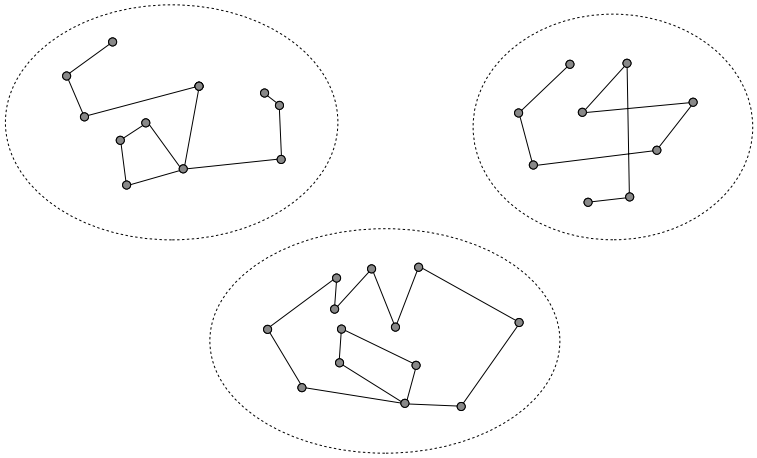


FIG. A.7 – *Exemples de cheminements et de circuit. Noter que le deuxième cheminement est en fait un chemin.*

Un graphe G est dit *connexe* si pour toute paire de sommets distincts u, v de G il existe un chemin d'extrémités u et v dans G. La *distance* entre deux sommets u et v d'un graphe connexe G est la longueur minimum d'un chemin d'extrémités u et v dans G. On la note parfois $d(u, v)$. L'application $d : V(G)^2 \to \mathbb{N}$ est une distance au sens topologique du terme, elle vérifie les propriétés suivantes :

- $d(u, u) = 0$ pour tout $u \in V(G)$,
- $d(u, v) = d(v, u)$ pour tout $u, v \in V(G)$,
- $d(u, v) \leq d(u, w) + d(w, v)$ pour tout $u, v, w \in V(G)$.

La *boule* de rayon $r \geq 1$ centrée en v, notée $B_r(v)$, est le sous-ensemble des

sommets à distance au plus r de v : $B_r(v) = \{u \mid d(u,v) \leq r\}$. La boule de rayon un, $B_1(v)$, sera souvent dénotée $N[v]$. On l'appelle *voisinage étendu* de v, car elle est l'union du voisinage de v et de v lui-même : $N[v] = N(v) \cup \{v\}$.

Un sous-graphe connexe maximal (par inclusion) de G est appelé *composante connexe* de G : tout graphe est l'union disjointe de ses composantes connexes.

Un graphe non-orienté connexe ayant n sommets et $n-1$ arêtes est un *arbre*. Un graphe dont les composantes connexes sont des arbres est une *forêt*.

L'*hypercube* de dimension n est le graphe ayant pour ensemble de sommets $V = \{0,1\}^n$ tel que deux vecteurs de V sont adjacents si et seulement si ils diffèrent en exactement une coordonnée. On note Q_n l'hypercube de dimension n. On peut voir l'hypercube comme le produit cartésien de K_2 par lui-même : $Q_n = \square^n K_2$ pour tout $n \geq 1$. On peut remarquer que la distance entre deux sommets de l'hypercube est égale au nombre de coordonnées sur lesquelles ces deux sommets diffèrent.

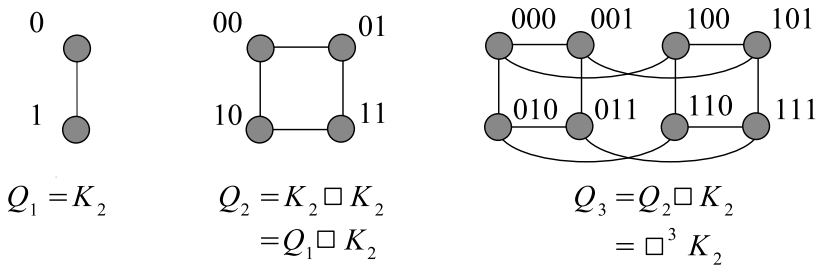


FIG. A.8 – *Les hypercubes de dimension 1, 2 et 3.*

Hypergraphes

Un *hypergraphe* est un couple (V, E), où V est un ensemble et E est un sous-ensemble de l'ensemble des parties de V (voir Figure A.9). Les éléments de V sont appelés *sommets* de l'hypergraphe, et les éléments de E sont appelés *hyperarêtes* de l'hypergraphe. Un graphe non-orienté peut être vu comme un hypergraphe dont toutes les hyperarêtes sont de cardinalité deux.

Un *plan projectif* (fini) d'ordre n est un hypergraphe sur $n^2 + n + 1$ sommets tel que :

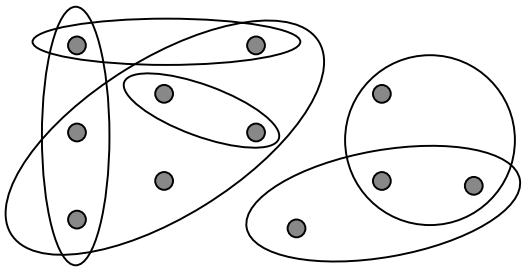


FIG. A.9 – *Exemple d'hypergraphe.*

- toute paire de sommets est contenue dans une unique hyperarête,
- deux hyperarêtes s'intersectent en un unique sommet,
- chaque sommet est contenu dans exactement $n+1$ hyperarêtes, et
- chaque hyperarête contient exactement $n+1$ sommets.

On note $\mathbb{P}_n$ le plan projectif d'ordre n. On sait que $\mathbb{P}_n$ existe si n est la puissance d'un nombre premier. $\mathbb{P}_n$ est également connu sous le nom de 2-$(n^2+n+1, n+1, 1)$ design, ou système de Steiner $S(2, n+1, n^2+n+1)$. Nous renvoyons le lecteur à [vLW92, Chapitre 19] pour une introduction plus complète aux designs.

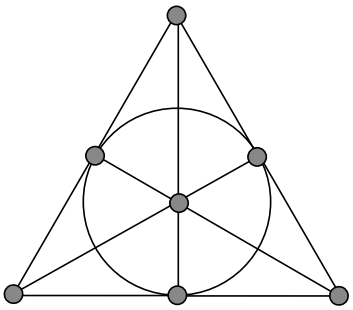


FIG. A.10 – *Plan projectif d'ordre 2. Les hyperarêtes sont les six segments plus le cercle central qui détermine la septième hyperarête. Chaque hyperarête contient trois sommets, chaque paire de sommets détermine une unique hyperarête, et deux hyperarêtes s'intersectent en exactement un point.*

Algorithmes

Beaucoup de problèmes de combinatoire dans les graphes consistent à chercher un sous-ensemble de sommets et/ou d'arêtes X satisfaisant certaines propriétés. Les problèmes de ce type sont appelés *problèmes de décision*, ils sont de la forme :

Problème de décision :
Instance : Un graphe G.
Question : Existe-t-il un sous-ensemble X satisfaisant l'ensemble de propriétés $\mathcal{P}$ dans G ?

Souvent, nous nous intéressons de plus à déterminer un tel ensemble X qui soit *optimum* vis-à-vis d'une fonction objectif. Le plus souvent, l'optimisation consiste à déterminer un tel X de cardinalité minimum/maximum, mais les fonctions objectif peuvent être éventuellement plus compliquées. En particulier, les graphes sont parfois pondérés et/ou valués afin de définir des fonctions objectif plus complexes. Les problèmes de ce type sont appelés *problèmes d'optimisation*, ils sont de la forme :

Problème d'optimisation :
Instance : Un graphe G.
Question : Quel est l'optimum de la fonction objectif $f(X)$, où X est un sous-ensemble satisfaisant l'ensemble de propriétés $\mathcal{P}$?

Dans ce contexte, il est souvent utile de distinguer deux notions d'optimalité. Un ensemble X sera dit *optimum* (ou, selon le cas, *maximum* ou *minimum*) si et seulement si c'est un ensemble satisfaisant les propriétés requises et optimisant la fonction objectif considérée. Typiquement, il s'agira d'un ensemble X de cardinalité minimum/maximum. D'autre part, un ensemble X sera dit *minimal* (resp. *maximal*) si et seulement si c'est un ensemble satisfaisant les propriétés requises minimal par inclusion (resp. maximal par inclusion), *i.e.* tel qu'il n'existe pas d'ensemble X' satisfaisant les propriétés requises inclus dans X (resp. contenant X) mais distinct de X.

Par exemple, un sous-ensemble de sommets T d'un graphe G est un *transversal* de G, si et seulement si toute arête de G a au moins une des ses extrémités dans T. Un transversal minimum de G est un transversal de cardinalité minimum de G, alors qu'un transversal minimal de G est un transversal ne contenant aucun autre transversal de G. Un transversal minimum de G est aussi un transversal minimal, mais la réciproque est en général fausse.

Un *algorithme* est un procédé automatique de calcul manipulant uniquement des structures discrètes (comme des valeurs booléennes `VRAI/FAUX`, des entiers, des graphes finis...). En général, un algorithme a des *paramètres*

d'entrée (par exemple, l'instance d'un problème de décision) et des *paramètres de sortie* retournés à la fin de l'exécution de l'algorithme (par exemple, la réponse à la question du problème de décision considéré). Le *temps d'exécution* d'un algorithme est le nombre maximum d'opérations élémentaires (comparaisons, opérations de lecture/écriture en mémoire...) nécessaires à la terminaison de l'algorithme.

Un algorithme est dit *polynomial* si son temps d'exécution est polynomial en la *taille* des paramètres d'entrées, où la taille d'un paramètre est le nombre de bits (information binaire du type 0-1 ou `VRAI/FAUX`) nécessaires pour coder ce paramètre. Par extension, un problème de décision est dit polynomial s'il existe un algorithme polynomial résolvant ce problème. Un problème polynomial est aussi parfois dit *de classe P*.

Un *oracle* est une opération faisant un choix non-déterministe parmi un ensemble fini de possibilités. Par exemple, l'opération 'choisir un sous-ensemble de sommets de G' est un oracle. Un algorithme non-déterministe est un procédé automatique de calcul utilisant un oracle. Un oracle est considéré comme une opération élémentaire s'exécutant en un temps constant. Un problème de décision est dit *de classe NP* s'il existe un algorithme non-déterministe utilisant un oracle $\mathcal{O}$, tel que, pour toute instance x de Π telle que $\Pi(x) =$`VRAI`, il existe un choix de $\mathcal{O}$ tel que l'algorithme réponde `VRAI` en un temps polynomial.

Par exemple, le problème de la recherche d'un transversal de taille inférieure ou égale à un entier k fixé est un problème de classe NP, admettant un algorithme non-déterministe polynomial en deux passes :

- ORACLE : choisir un sous-ensemble T de k sommets de G,
- renvoyer `VRAI` si et seulement si T est un transversal de G.

La première passe est l'oracle, qui, en un temps constant, choisit un sous-ensemble T parmi $\binom{n}{k}$ sous-ensembles possibles. La deuxième passe consiste à vérifier que le sous-ensemble T choisi est bien un transversal du graphe, ce qui est faisable en temps polynomial : pour chaque arête uv du graphe, il faut vérifier que u ou v est dans T. Ceci est réalisable en un nombre d'opérations égal à $O(m)$, où m est le nombre d'arêtes du graphe considéré.

Trivialement, un problème de classe P est un problème de classe NP. La réciproque de ceci est certainement l'une des questions ouvertes les plus importantes qui soient[4]. La communauté scientifique s'accorde à penser que la réponse à cette question est non, à savoir P $\neq$ NP : il existe des problèmes

[4]Depuis le 24 mai 2000, le *Clay Mathematics Institute* offre un million de dollars pour une preuve de P $\neq$ NP [CMI].

de classe NP qui ne sont pas de classe P.

On dit qu'un problème de décision Π *se réduit* à un problème de décision Π' s'il existe une fonction f qui transforme une instance de Π en une instance de Π', telle que :

- la réponse de $\Pi(x)$ est `VRAI` si et seulement si la réponse de $\Pi(f(u))$ est `VRAI`,
- f est calculable en temps polynomial.

Un argument qui va dans le sens de la conjecture P $\neq$ NP est l'existence de problèmes de décision Π^* tels que tout problème de NP se réduise à Π^*. De tels problèmes sont dits *NP-complets*. L'existence de problèmes NP-complets a été montrée par S. A. Cook dans les années 70 [Coo71]. Depuis, nombre de problèmes combinatoires importants se sont révélés NP-complets.

Pour un problème d'optimisation Π dont la fonction objectif f est à valeurs dans $\mathbb{Z}$, on peut considérer, pour tout entier $k \in \mathbb{Z}$, le problème de décision associé Π_k qui retourne `VRAI` si et seulement si $f \leq k$. On dit que Π est *NP-difficile* si le problème d'optimisation associé Π_k est NP-complet.

Conclusion

Dans cette thèse nous avons abordé divers problèmes concernant les codes identifiants dans les graphes. La notion de code identifiant a tout d'abord été replacée dans le cadre plus général des problèmes de couverture par tests et de tests groupés. Nous avons aussi établi de nombreux liens avec d'autres types de codes mieux connus, tels les codes correcteurs et les codes superimposés (Chapitre 1). Après avoir donné deux classes de graphes pour lesquelles la recherche de la cardinalité minimum d'un code identifiant était polynomiale (Chapitre 2), nous avons étudié d'autres classes de graphes pour lesquelles nous avons pu directement donner des codes identifiants optimaux (Chapitre 3). De nombreuses questions extrémales ont été abordées (Chapitre 4). Nous avons également étudié ce problème dans le cas des graphes aléatoires (Chapitre 5).

Ce travail répond à des préoccupations de la communauté scientifique internationale et a pu à ce titre donner lieu à des publications (à ce jour cinq articles acceptés dans des revues internationales avec comité de lecture et quatre articles soumis). Nous avons également pu diffuser nos résultats lors de conférences et séminaires (participation à quatre conférences nationales et une conférence internationale).

Certains de nos résultats sont des résultats partiels, et laissent en suspens des questions auxquelles il nous semble pertinent d'essayer de répondre. Comme perspective de ce travail de thèse nous suggérons quelques pistes sur lesquelles poursuivre, parmi lesquelles :

- résoudre la **conjecture de Blass, Honkala et Litsyn** (Conjecture 3.1) en adaptant l'argument de projection donné dans le Théorème 3.3, qui résout cette conjecture dans le cas $t = 1$,
- renforcer le lien entre les codes identifiants et les **codes superimposés** (donner d'autres constructions similaires à celle du Théorème 4.3),
- au sujet des **codes identifiant des ensembles de sommets**, essayer de déterminer une borne inférieure serrée sur la cardinalité minimum

d'un code $(1, \leq \ell)$-identifiant d'un graphe à n sommets (améliorer le Théorème 1.6),
- **généraliser l'algorithme sur les fasciagraphes** (Théorème 2.4), et proposer une caractérisation de classes de problèmes combinatoires solvables par cet algorithme,
- chercher des **algorithmes d'approximation** pour les codes identifiants : si des heuristiques ont été déjà utilisées dans la littérature afin d'établir des bornes supérieures sur la cardinalité minimum d'un code identifiant [CHL02a], la performance de ces heuristiques n'a à notre connaissance jamais été étudiée,
- étudier les **codes identifiants dans les hypergraphes** (paragraphe 1.2.6),
- enfin, déterminer des **stratégies adaptatives** pour les codes identifiants dans les graphes, comme décrit dans le paragraphe 1.2.7.

J'accorde une une importance particulière à la collaboration scientifique, comme peuvent le témoigner onze coauteurs de quatre nationalités différentes. J'espère pouvoir, dans un avenir proche, étudier certaines de ces questions en collaboration avec les chercheurs de l'équipe Coding Theory de l'Université de Turku, dirigée par Iiro Honkala. Je souhaiterais également poursuivre la collaboration avec l'équipe *Discrete Structures* de Budapest, en particulier pour renforcer les liens entre les codes identifiants et les codes superimposés.

À plus long terme, je souhaiterais élargir mes connaissance et étudier d'autres questions combinatoires issues de problèmes de communication et de gestion de l'information.

Bibliographie

[ALS91] S. Arnborg, J. Lagergren, D. Seese, *Easy Problems for Tree-Decomposable Graphs*, Journal of Algorithms **12(2)** (1991), 308–340.

[AS00] N. Alon, J. H. Spencer, *The Probabilistic Method,* Wiley-Interscience [John Wiley & Sons] (2000).

[BCHL04] N. Bertrand, I. Charon, O. Hudry, A. Lobstein, *Identifying and Locating-Dominating Codes on Chains and Cycles*, European Journal of Combinatorics **25(7)** (2004), 969–987.

[Ber68] E. R. Berlekamp, *Block Coding for the Binary Symmetric Channel with Noiseless, Delayless Feedback*, in Error-Correcting Codes, Wiley, New York (1968), 61–85.

[BGMP86] D. Babić, A. Graovac, B. Mohar, T. Pisanski, *The Matching Polynomial of a Polygraph*, Discrete Applied Mathematics **15** (1986), 11–24.

[BHL00] U. Blass, I. Honkala, S. Litsyn, *On Binary Codes for Identification*, Journal of Combinatorial Designs **8** (2000), 151–156.

[BHL01] U. Blass, I. Honkala, S. Litsyn, *Bounds on Identifying Codes*, Discrete Mathematics **241** (2001), 119–128.

[BLP90] A. G. Bale, J. Litt, J. Pavelin, *The* AMT DAP *500 system*, Elsevier Science Publishers, North-Holland (1990).

[Bod93] H. L. Bodlaender, *A Tourist Guide through Treewidth*, Acta Cybernetica **11(1-2)** (1993), 1–21.

[Bol82] B. Bollobás, *Distinguishing Vertices of Random Graphs*, Annals of Discrete Mathematics **13** (1982), 33–50.

[Bol85] B. Bollobás, *Random Graphs*, Academic Press (1985).

[CHLL97] G. Cohen, I. Honkala, S. Litsyn, A. Lobstein, *Covering Codes*, Elsevier, North-Holland Mathematical Library (1997).

[CHHL01] I. Charon, I. Honkala, O. Hudry, A. Lobstein, *General Bounds for Identifying Codes in some Infinite Regular Graphs*, Electronic Journal of Combinatorics **8(1)** R39 (2001).

[CHHL04] I. Charon, I. Honkala, O. Hudry, A. Lobstein, *The Minimum Density of an Identifying Code in the King Lattice*, Discrete Mathematics **276(1-3)** (2004), 95–109.

[CHL02a] I. Charon, O. Hudry, A. Lobstein, *Identifying Codes with Small Radius in some Infinite Regular Graphs*, Electronic Journal of Combinatorics **9(1)** R11 (2002).

[CHL02b] I .Charon, O. Hudry, A. Lobstein, *Identifying and Locating-Dominating Codes : NP-completeness Results for Directed Graphs*, IEEE Transactions on Information Theory **48(8)** (2002), 2192–2200.

[CHL03] I. Charon, O. Hudry, A. Lobstein, *Minimizing the Size of an Identifying or Locating-Dominating Code in a Graph is NP-hard*, Theoretical Computer Science **290(3)** (2003), 2109–2120.

[CHLa] I. Charon, O. Hudry, A. Lobstein, *Extremal Cardinalities for Identifying and Locating-Dominating Codes*, à paraître dans Discrete Mathematics.

[CHLb] I. Charon, O. Hudry, A. Lobstein, *Possible Cardinalities for Identifying Codes in Graphs*, à paraître dans Australasian Journal of Combinatorics.

[CG98] F. Chung, R. Graham, *Erdős on Graphs — His Legacy of Unsolved Problems*, A. K. Peters, Wellesley, Massachusetts (1998).

[CGHLMM] I. Charon, S. Gravier, O. Hudry, A. Lobstein, M. Mollard, J. Moncel, *A Linear Algorithm for Minimum 1-Identifying Codes in Oriented Trees*, soumis.

[CGHLMPZ99] G. Cohen, S. Gravier, I. Honkala, A. Lobstein, M. Mollard, C. Payan, G. Zémor, *Improved Identifying Codes for the Grid*, Electronic Journal of Combinatorics, comments to **6(1)** R19 (1999).

[CHLZ99] G. Cohen, I. Honkala, A. Lobstein, G. Zémor, *New Bounds for Codes Identifying Vertices in Graphs*, Electronic Journal of Combinatorics, **6(1)** R19 (1999).

[CM02] M. Csűrös, A. Milosavljevic, *Pooled Genomic Indexing (PGI) : Mathematical Analysis and Experiment Design*, Lecture Notes in Computer Science **2452** (2002), 10–28.

[CM04] M. Csűrös, A. Milosavljevic, *Pooled Genomic Indexing (PGI) : Analysis and Design of Experiments*, Journal of Computational Biology **11(5)** (2004), 1001–1021.

[CMI] `http://www.claymath.org/millennium/P_vs_NP/`

[Coo71] S. A. Cook, *The Complexity of Theorem-Proving Procedures*, Proceedings of the Third Annual ACM Symposium on the Theory of Computing (1971), 151–158.

[CSS87] C. J. Colbourn, P. J. Slater, L. K. Stewart, *Locating Dominating Sets in Series Parallel Networks*, Congressus Numerantium **56** (1987), 135–162.

[Dan03] M. Daniel, *Codes Identifiants*, Mémoire pour le DEA ROCO, Université Joseph Fourier, Grenoble (2003).

[DGM04] M. Daniel, S. Gravier, J. Moncel, *Identifying Codes in some Subgraphs of the Square Lattice*, Theoretical Computer Science **319** (2004), 411–421.

[DH00] D.-Z. Du, F. K. Hwang, *Combinatorial Group Testing and its Applications*, Series on Applied Mathematics, Singapour (2000).

[Die00] R. Diestel, *Graph Theory*, Springer, Berlin (2000).

[DR83] A. G. D'yachkov, V. V. Rykov, *Bounds on the Length of Disjunctive Codes*, Problems of Information Transmission **18** (1983), 166–171.

[DR02] A. G. D'yachkov, V. V. Rykov, *Optimal Superimposed Codes and Designs for Rényi's Search Model*, Journal of Statistical Planning and Inference **100(2)** (2002), 281–302.

[EL75] P. Erdős, L. Lovász, *Problems and Results on 3-chromatic Hypergraphs and some Related Questions,* in Infinite and Finite Sets (to Paul Erdős on his 60th birthday), North-Holland, Amsterdam (1975), 609–627.

[ER60] P. Erdős, A. Rényi, *On the Evolution of Random Graphs,* Publications of the Mathematical Institute of the Hungarian Academy of Sciences **5** (1960), 17–61.

[ER61] P. Erdős, A. Rényi, *On the Evolution of Random Graphs,* Bulletin de l'Institut International de Statistique **38(4)** (1961), 343–347.

[Erd32] P. Erdős, *Beweis eines Satzes von Tschebyschef*, Acta Litterarum ac Scientiarum, Szeged **5** (1932), 194–198.

[Erd47] P. Erdős, *Some Remarks on the Theory of Graphs*, Bulletin of the American Mathematical Society **53** (1947), 292–294.

[Erd59] P. Erdős, *Graph Theory and Probability*, Canadian Journal of Mathematics **11** (1959), 34–38.

[FMMRS] A. Frieze, R. Martin, J. Moncel, M. Ruszinkó, C. Smyth, *Codes Identifying Sets of Vertices in Random Networks*, soumis.

[Fag76] R. Fagin, *Probabilities on Finite Models*, Journal of Symbolic Logic **41**(1976), 50–58.

[Fur96] Z. Füredi, *On r-cover free Families*, Journal of Combinatorial Theory Series A **73(1)** (1996), 172–173.

[GJ79] M. R. Garey, D. S. Johnson, *Computers and Intractability : A Guide to the Theory of NP-completeness*, Freeman, San Francisco (1979).

[GM05] S. Gravier, J. Moncel, *Construction of Codes Identifying Sets of Vertices*, Electronic Journal of Combinatorics **12(1)** R13 (2005).

[GM] S. Gravier, J. Moncel, *On Graphs Having a $V \setminus \{x\}$ Set as an Identifying Code*, à paraître dans Discrete Mathematics.

[GMP] S. Gravier, J. Moncel, C. Payan, *A Generalization of the Pentomino Exclusion Problem : the Δ-Dislocation in Graphs*, à paraître dans Discrete Mathematics.

[GMS] S. Gravier, J. Moncel, A. Semri, *Identifying Codes of Cycles*, à paraître dans European Journal of Combinatorics.

[Gol94] S. W. Golomb, *Polyominoes : Puzzles, Patterns, Problems, and Packings*, Princeton University Press (1994).

[Hil85] W. D. Hillis, *The Connection Machine*, MIT press (1985).

[HKL01] I. Honkala, M. G. Karpovsky, S. Litsyn, *On the Identification of Vertices and Edges Using Cycles*, Lecture Notes in Computer Science **2227** (2001), 308–314.

[HKL03] I. Honkala, M. G. Karpovsky, S. Litsyn, *Cycles Identifying Vertices and Edges in Binary Hypercubes and 2-dimensional Tori*, Discrete Applied Mathematics **129(2-3)** (2003), 409–419.

[HL02a] I. Honkala, A. Lobstein, *On the Complexity of the Identification Problem in Hamming Spaces*, Acta Informatica **38(11-12)** (2002), 839–845.

[HL02b] I. Honkala, A. Lobstein, *On the Density of Identifying Codes in the Square Lattice*, Journal of Combinatorial Theory Series B **85** (2002), 297–306.

[HL02c] I. Honkala, A. Lobstein, *On Identifying Codes in Binary Hamming Spaces*, Journal of Combinatorial Theory Series A **99** (2002), 232–243.

[HL03a] I. Honkala, T. Laihonen, *On the Identification of Sets of Points in the Square Lattice*, Discrete and Computational Geometry **29** (2003), 139–152.

[HL03b] I. Honkala, A. Lobstein, *On Identification in $\mathbb{Z}^2$ Using Translates of Given Patterns*, Journal of Universal Computer Science, **9(10)** (2003), 1204–1219.

[HLR01] I. Honkala, T. Laihonen, S. Ranto, *On Codes Identifying Sets of Vertices in Hamming Spaces*, Designs, Codes and Cryptography **24(2)** (2001), 193–204.

[HLR02] I. Honkala, T. Laihonen, S. Ranto, *On Strongly Identifying Codes*, Discrete Mathematics **254(1-3)** (2002), 191–205.

[HM76] F. Harary, R. A. Melter, *On the Metric Dimension of a Graph*, Ars Combinatoria **2** (1976), 191–195.

[HS87] F. K. Hwang, V. Sós, *Non-adaptive Hypergeometric Group Testing*, Studia Scientiarum Mathematicarum Hungaricae **22(1-4)** (1987), 257–263.

[IC91] Intel Corporation, PARAGON XP/S *Product Overview*, Supercomputer Systems Division, Intel Corporation, Beaverton, (1991).

[Jan98] S. Janson, *New Versions of Suen's Correlation Inequality,* Random Structures Algorithms **13(3-4)** (1998), 467–483.

[JLR00] S. Janson, T. Łuczak, A. Ruciński, *Random Graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, New York (2000).

[JMGSZ95] M. Juvan, B. Mohar, A. Graovac, S. Klavžar, J. Žerovnik, *Fast Computation of the Wiener Index of Fasciagraphs and Rotagraphs*, Journal of Chemical Information and Computer Sciences **35** (1995), 834–840.

[JMZ97] M. Juvan, B. Mohar, J. Žerovnik, *Distance-Related Invariants on Polygraphs*, Discrete Applied Mathematics **80(1)** (1997), 57–71.

[KCL98] M. G. Karpovsky, K. Chakrabarty, L. B. Levitin, *On a New Class of Codes for Identifying Vertices in Graphs*, IEEE Transactions on Information Theory **44(2)** (1998), 599–611.

[KCLA99] M. G. Karpovsky, K. Chakrabarty, L. B. Levitin, D. R. Avreky, *On the Covering of Vertices for Fault Diagnosis in Hypercubes*, Information Processing Letters, **69** (1999), 99–103.

[Kog95] A. Kogan, *On the Essential Test Sets of Discrete Matrices*, Discrete Applied Mathematics **60(1-3)** (1995), 249–255.

[KS64] W. H. Kautz, R. C. Singleton, *Nonrandom Binary Superimposed Codes*, IEEE Transactions on Information Theory **10** (1964), 363–377.

[KV03] S, Klavžar, A. Vesel, *Computing Graph Invariants on Rotagraphs using Dynamic Algorithm Approach : The case of (2, 1)-colorings and independence numbers*, Discrete Applied Mathematics **129(2-3)**(2003), 449–460.

[KZ96] S. Klavžar, A. Žerovnik, *Algebraic Approach to Fasciagraphs and Rotagraphs*, Discrete Applied Mathematics **68(1-2)** (1996), 93–100.

[Lai02a] T. Laihonen, *Sequences of Optimal Identifying Codes*, IEEE Transactions on Information Theory **48(3)** (2002), 774–776.

[Lai02b] T. Laihonen, *Optimal codes for strong identification*, European Journal of Combinatorics **23(3)** (2002), 307–313.

[LM] S. Litsyn, Y. Merksamer, *Exact Minimum Density of Codes Identifying Vertices in the Square Grid*, soumis.

[Lov98] L. Lovász, *One Mathematics*, The Berliner Intelligencer, Berlin (1998), 10–15.

[LR01] T. Laihonen, S. Ranto, *Codes Identifying Sets of Vertices*, Lecture Notes in Computer Science **2227** (2001), 82–91.

[LR02] T. Laihonen, S. Ranto, *Families of Optimal Codes for Strong Identification*, Discrete Applied Mathematics **121(1-3)** (2002), 203–213.

[LS94] M. Livingston, Q. F. Stout, *Constant Time Computation of Minimum Dominating Sets*, Congressus Numerantium **105** (1994), 116–128.

[Mona] J. Moncel, *Optimal Graphs for Identification of Vertices in Networks*, soumis.

[Monb] J. Moncel, *On the Monotonicity of the Minimum Cardinality of an Identifying Code in the Hypercube*, soumis.

[NCube] `http://www.ncube.com/`

[Pel02] A. Pelc, *Searching Games with Errors — Fifty Years of Coping With Liars*, Theoretical Computer Science **270(1-2)** (2002), 71–109.

[PP80] R. W. Payne, D. A. Preece, *Identification Keys and Diagnostic Tables : A Review*, Journal of the Royal Statistical Society, Series A **143** (1980), 253–292.

[Ram19] S. Ramanujan, *A Proof of Bertrand's Postulate*, Journal of the Indian Mathematical Society **11** (1919), 181–182.

[Ren62] A. Rényi, *On a Problem of Information Theory*, Publications of the Mathematical Institute of the Hungarian Academy of Sciences, **6B** (1962), 505–516.

[RS84] D. F. Rall, P. J. Slater, *On Location-Domination Numbers for Certain Classes of Graphs*, Congressus Numerantium **45** (1984), 97–106.

[RS04] N. Robertson, P. D. Seymour, *Graph Minors XX — Wagner's Conjecture*, Journal of Combinatorial Theory Series B **92(2)** (2004), 325–357.

[RSTU04] S. Ray, D. Starobinski, A. Trachtenberg, R. Ungrangsi, *Robust Location Detection with Sensor Networks*, IEEE Journal on Selected Areas in Communications **22(6)** (2004), 1016–1025.

[Rusz94] M. Ruszinkó, *On the Upper Bound of the Size of the r-cover-free Families*, Journal of Combinatorial Theory Series A **66(2)** (1994), 302–310.

[RV97] M. Ruszinkó, P. Vanroose, *How an Erdős-Rényi-type Search Approach Gives an Explicit Code Construction of Rate 1 for Random Access with Multiplicity Feedback*, IEEE Transactions on Information Theory **43(1)** (1997), 368–373.

[Sto] `http://www.eecs.umich.edu/~qstout/constantques.html`

[Sue90] W.-C. S. Suen, *A Correlation Inequality and a Poisson Limit Theorem for Nonoverlapping Balanced Subgraphs of a Random Graph,* Random Structures Algorithms **1(2)** (1990), 231–242.

[Ula76] S. M. Ulam, *Adventures of a Mathematician*, Charles Scribner's Sons, New York (1976).

[vLW92] J. H. van Lint, R. M. Wilson, *A Course in Combinatorics*, Cambridge University Press (1992)

[Wag70] K. Wagner, *Graphentheorie*, Mannheim Bibliographisches Institut (1970).

[Win93] P. Winkler, *Random Structures and Zero-One Laws*, Proceedings of the NATO Advanced Study Institute, NATO ASI Series C **411** (1993), 399–420.

[Zer99] J. Žerovnik, *Deriving Formulas for Domination Numbers of Fasciagraphs and Rotagraphs*, Lecture Notes in Computer Science **1684** (1999), 559–568.

[Zer] J. Žerovnik, *Deriving Formulas for the Pentomino Exclusion Problem*, soumis.

Index